

Volumen 06 - Número 1 - 2024 ISSN: 2810-8124 (en línea)



Online



Regresion

Generado con Meta AI

DOI: <https://doi.org/10.57166.riqchary.v6.n2.2024>



UNIVERSIDAD NACIONAL
MICAELA BASTIDAS
DE APURÍMAC



ISSN/2810-8124



Comité Editorial

- **Editor General:** Dr. Ecler Mamani Vilca
Universidad Micaela Bastidas Nacional de Apurímac, Perú

Equipo Técnico:

- Dr. José Luis Merma Aroni
Universidad Micaela Bastidas Nacional de Apurímac, Perú
- M.Sc. Mario Aquino Cruz
Universidad Tecnológica de los Andes Perú
- Dr. Manuel J. Ibarra Cabrera
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Dr. Yalmar Ponce Atencio
Micaela Nacional de Juliaca, Perú
- Dr. Leonel Coyla Idme
Universidad Nacional del Altiplano, Perú
- Dr. José Luis Herrera Salazar
Universidad Norbert Wiener, Perú
- Mag. Olga Jaramillo
Santa Fe Community College, Estados Unidos
- Dr. Edwar Ilasaca Cahuata
Universidad Nacional del Altiplano, Perú
- Mag. Yonatan Mamani Coaquira
Universidad Micaela Bastidas Nacional de Apurímac, Perú

Equipo Editorial:

- Mag. Marleny Peralta Ascue
Universidad Privada de Tacna, Perú
- Mag. Rafael Quispe Merma
Universidad Nacional San Antonio de Cusco, Perú
- Ing. Ebert Gómez Aiquipa
Universidad Nacional San Antonio de Cusco, Perú
- MBA. Evelyn Naida Luque Ochoa
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Kevin Arnold Arias Figueroa
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Alejandrina Huaylla Quispe
Universidad Tecnológica de los Andes Perú
- Mag. Betsabe M. Ccolque Ruiz
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Miguel Alfaro
Micaela Bastidas Universidad Nacional de Apurímac, Perú

Periodicidad: Semestral

Vol. 6 Núm. 2 año (2024)

Áreas temáticas: Ciencias exactas, ingenierías y de materiales

Editorial: UNAMBA – Perú

Vice Rectorado de Investigación

<https://doi.org/10.57166/riqchary/V6.n2.2024>

ISSN: 2706-543X- ISSN: 2810-8124 (e)

Editorial

Con gran satisfacción presentamos esta edición, que destaca la importancia del análisis de datos en el desarrollo científico y tecnológico. La imagen de portada, generada mediante inteligencia artificial de Meta IA, representa la integración de modelos de regresión y técnicas de machine learning para analizar y predecir patrones en los datos. En este número, se evidencia cómo el aprendizaje automático, el aprendizaje profundo y la optimización de hiperparámetros se consolidan como herramientas clave para la generación de conocimiento e innovación, contribuyendo al avance de la investigación en la actualidad.

Director: Dr. Ecler Mamani Vilca





C & T Riqchary

Revista de investigación - UNAMBA

Autores

- Carlos Mendoza Contreras Torres
 - German R. Espinoza Rivas
 - Alonso Sanchez Torres C
 - Jorge Campos Malpartida
 - Erech Ordoñez Ramos
 - Alejandrina Huaylla Quispe
 - Mary Luz Nina Palacios
 - Ernesto Nayer Tumi Figueroa
 - Hugo Ticona Salluca
 - Aliana Gamboa Ortega
 - José Castro Soto
 - Reyna Vargas Peña
 - Niki Franklin Flores Pacheco
 - Reyna Vargas Peña
 - Niki Franklin Flores Pacheco
 - Malu Beatriz Silva Zarate
 - Mario Aquino Cruz
 - Darío Dante Sánchez Castillo
 - Pablo Ruben Zuloaga Candia
 - Karina Gamarra Peralta
-
- © Dr. Ecler Mamani Vilca / Director del Comité Editorial

Av. Garcilaso C.U Tamburco, Abancay Telef, +51983747138

Apurímac – Perú eclervirtual@unamba.edu.pe

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2020-10220 ISSN: 2709-8990

ISSN: 2955-8646 (e)- ISSN: 2709-8990

Tiraje: 500 ejemplares

Volumen 6 Número 2 – 2024

Se terminó de imprimir en El Misti EIRL

RUC 20450599108 - Cel: 950-509953, imprentamisti@hotmail.com

01

Pág.

Caracterización geológica y potencial mineralógico de Tumipampa.

07

Pág.

Análisis del sentimiento con NLP en tutoría académica: la vocación como fator de mejora en el rendimiento académico.

14

Pág.

Aplicación de Análisis de Componentes Principales (PCA) y K-Means para la Clasificación de Unidades Estratigráficas en Muestras Geoquímica del Volcán Misti

21

Pág.

Evaluación de parámetros de calidad en diversas macas comerciales de cloro domestico

27

Pág.

Efecto de los tipos de sustrato en almacigo más hidroponía NFT en las características morfológicas de tres variedades de lechuga (*Lactuca sativa* L.) en el INIA Andahuaylas Peru-2024

35

Pág.

Rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) en dos tipos de sustrato en almacigo más hidroponía NFT en INIA Andahuaylas Peru-2024

41

Pág.

Modelo predictivo de machine learning para el servicio de orientación vocacional hacia los estudiantes de secundaria en la Gerencia Regional de Trabajo y Promoción del Empleo Cusco 2024.

49

Pág.

Determinación de la contaminación por metal pesados en los sedimentos superficiales de la subcuenca del rio Andabamba en un periodo de mediano plazo

INDICE

Caracterización geológica y potencial mineralógico de Tumipampa

Geological characterization and mineralogical potential of Tumipampa

Carlos Mendoza-Contreras ^A, German Rafael Espinoza-Rivas ^B, Alonso Sanchez-Torres ^C

Resumen— Este estudio se centra en la evaluación integral del yacimiento Tumipampa, localizado en la región de Apurímac, Perú, reconocido por su notable potencial mineralógico. La investigación se enfoca en la caracterización detallada de las unidades litológicas, la identificación precisa de las estructuras geológicas y la valoración del potencial económico basado en la mineralización presente. Mediante un riguroso mapeo geológico, un exhaustivo muestreo geoquímico y técnicas avanzadas de perforación diamantina, se identificaron intrusiones de granodiorita, formaciones de caliza, mármol, cuarcita y limolitas. Las estructuras geológicas, incluyendo fallas y pliegues, desempeñan un papel crucial en la distribución de la mineralización. Los análisis geoquímicos detallados revelaron concentraciones significativas de oro (Au), plata (Ag) y cobre (Cu) en muestras de afloramiento, con valores que alcanzan hasta 25.10 ppm de Au y 24.8 ppm de Ag. Adicionalmente, los resultados de la perforación diamantina indicaron concentraciones de hasta 12.25 ppm de Au en intervalos de 2.00 metros de ancho y 23.8 ppm en intervalos de 1.20 metros de ancho. Estos hallazgos sugieren un considerable potencial económico para la explotación minera del yacimiento Tumipampa, respaldado por datos cuantitativos de recursos y los resultados obtenidos mediante perforación diamantina.

Palabras claves: caracterización geológica, depósito Tumipampa, mineralización, potencial económico

Abstract— This study focuses on the comprehensive evaluation of the Tumipampa deposit, located in the Apurímac region of Peru, known for its notable mineralogical potential. The research emphasizes detailed lithological characterization, precise identification of geological structures, and the assessment of economic potential based on existing mineralization. Through rigorous geological mapping, exhaustive geochemical sampling, and advanced diamond drilling techniques, granodiorite intrusions, limestone formations, marble, quartzite, and siltstones were identified. Geological structures, including faults and folds, play a crucial role in mineral distribution. Detailed geochemical analyses revealed significant concentrations of gold (Au), silver (Ag), and copper (Cu) in outcrop samples, with values up to 25.10 ppm of Au and 24.8 ppm of Ag. Additionally, diamond drilling results indicated concentrations up to 12.25 ppm of Au over 2.00-meter intervals and 23.8 ppm over 1.20-meter intervals. These findings suggest considerable economic potential for mining exploitation at the Tumipampa deposit, supported by quantitative resource data and diamond drilling results.

Keywords: characterization mineralization, Tumipampa deposit, geological, economic potential

1 INTRODUCCIÓN

El yacimiento Tumipampa, situado en la región de Apurímac, Perú, se encuentra en una zona rica en diversidad geológica y potencial mineralógico. Las exploraciones en esta área han revelado la presencia de minerales metálicos significativos incluyendo oro (Au), plata (Ag), cobre (Cu) y hierro (Fe). Este estudio tiene como objetivo principal identificar y describir las unidades litológicas, analizar las estructuras geológicas y evaluar el potencial mineralógico del área mediante el análisis de 141 muestras. El contexto geológico regional ha sido objeto de numerosas investigaciones debido a su compleja historia tectónica y variada ocurrencia de recursos minerales polimetálicos. Este trabajo integra datos de campo y análisis de

laboratorio para proporcionar una evaluación comprensiva del potencial mineralógico del yacimiento, abordando tanto la caracterización litológica como la interpretación estructural y los estudios geoquímicos necesarios para entender mejor las oportunidades de explotación minera del yacimiento [1], [2], [3]. El centroide de la ubicación geográfica en coordenadas UTM (WGS-84) es 720934 E, 8439266 N.

2 OBJETIVO

El estudio tiene como objetivo caracterizar geológicamente y evaluar el potencial mineralógico del área, analizando 141 muestras para determinar la viabilidad de la explotación minera. Destaca concentraciones significativas de oro, plata y cobre, y se basa en un análisis exhaustivo de las



Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x
Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 22/09/24 - Indexaciones
Número: doi.org/10.57166/riqchary/V6.n2.2024
Páginas: 1- 6 | Recibido 04/07/2024 ; Aceptado 28/07/2024

<https://doi.org/10.57166/riqchary.v6.n2.2024.122>

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0002-6313-2003> Carlos Mendoza Contreras Torres, Universidad Nacional Micaela Bastidas de Apurímac, mendozac@unamba.edu.pe
- B. **ORCID iD** <https://orcid.org/0000-0002-6547-7761> German R. Espinoza-Rivas, Universidad Nacional Micaela Bastidas de Apurímac, gespinoza@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0009-0008-6124-3299> Alonso Sanchez Torres C, Universidad Nacional de Ingeniería, asanchezt@uni.edu.pe

unidades litológicas, estructuras geológicas y datos geoquímicos, sugiriendo un considerable potencial económico para la explotación del yacimiento.

3 METODOLOGÍA

3.1 Mapeo Geológico

El mapeo geológico detallado se llevó a cabo utilizando GPS y brújula geológica sobre una base topográfica precisa, empleando una estación total para localizar y mapear las diferentes unidades litológicas. Esta técnica permitió obtener un conocimiento detallado de la distribución espacial de las litologías en el área de estudio. Se identificaron diversas formaciones rocosas, incluyendo intrusiones ígneas, secuencias sedimentarias y metamórficas, así como estructuras geológicas, a una escala detallada (Fig.1) [4], [5].

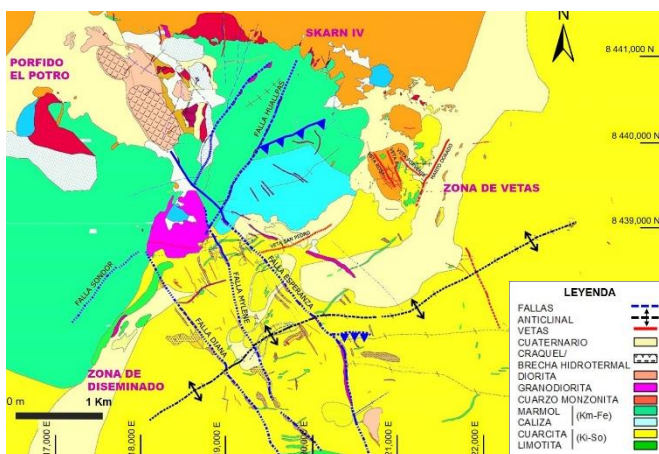


Fig.1. Mapa Geológico y Estructural del área de estudio, mostrando las unidades litológicas y las estructuras geológicas mapeadas. Las fallas representan el sistema andino Noroeste-Sureste y trasandino Noroeste-Suroeste, así como un anticlinal dentro de la unidad de cuarcitas. Estas estructuras juegan un papel fundamental en el control estructural de la mineralización del yacimiento.

3.2 Muestreo Geoquímico

Se recolectaron 141 muestras de minerales mediante técnicas de muestreo y se analizaron en laboratorio acreditado (SGS y CERTIMIN) para determinar su composición química mediante técnicas de espectrometría de masas (ICP-MS) y fluorescencia de rayos X (XRF). Se evaluaron las concentraciones de elementos metálicos en las estructuras geológicas mineralizadas y se documentaron los resultados y conclusiones relevantes para su importancia geológica y científica. En el siguiente diagrama se presenta el flujo de trabajo para el análisis geoquímico de minerales.

Mendoza-Contreras Carlos, Espinoza-Rivas German
Rafael y Sanchez-Torres Alonso

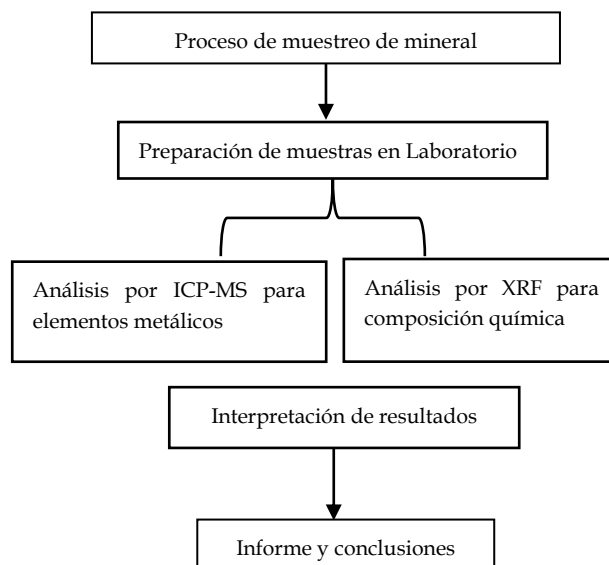


Fig.2. Flujo de trabajo optimizado para el análisis geoquímico de minerales: desde la recolección hasta la presentación de resultados.

3.3 Perforación Diamantina

Se emplearon equipos de perforación diamantina DIAMEC 262 para extraer núcleos de roca desde una profundidad de 300 metros. Estos núcleos fueron analizados mediante logueo para obtener información detallada sobre los sistemas de mineralización y la composición de las rocas en el subsuelo. La Figura 3 muestra los núcleos de roca interceptados con perforación diamantina, así como las unidades litológicas, la mineralización, la alteración y las fallas geológicas, revelando información crucial sobre la geología subterránea del área.

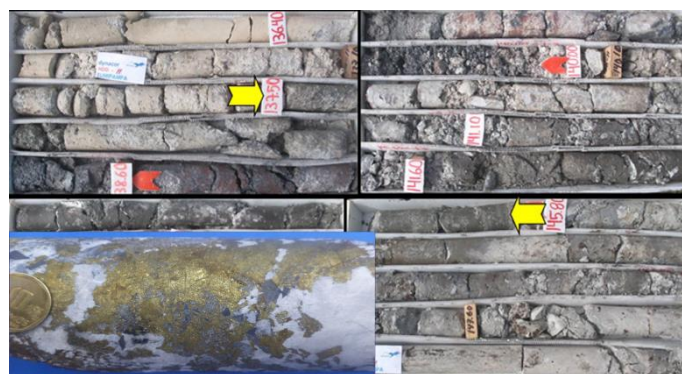


Fig.3. Núcleo de Perforación Diamantina.

4 RESULTADOS Y DISCUSIÓN

4.1 Unidades Litológicas

Las unidades litológicas identificadas en el área de estudio incluyen intrusiones de granodiorita, formaciones de caliza, mármol, cuarcita y limolitas. Estas unidades han sido descritas detalladamente en estudios previos y se correlacionan con los datos obtenidos en este estudio [6], [7]. Las rocas ígneas comprenden intrusiones granodioríticas del batolito de Andahuaylas-Yauri (Eoceno-Oligoceno). Según

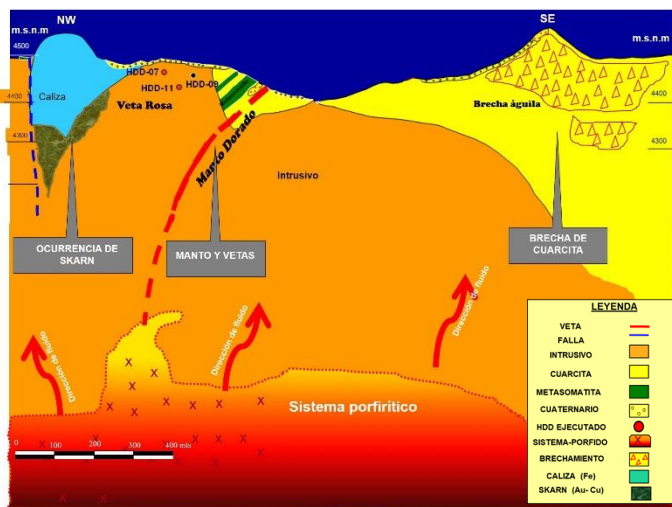


Fig. 6. Propuesta de modelo genético del yacimiento

4.4 Mineralización y Geoquímica

Se han reconocido los siguientes minerales principales: oro (Au), calcopirita (FeCuS₂), pirita (FeS₂) y cuarzo (SiO₂). También se identificaron minerales secundarios como hematita (Fe₂O₃), calcita (CaCO₃), marcasita (FeS₂), siderita (FeCO₃), malaquita (CuCO₃ Cu(OH)₂), pirrotita (Fe_{1-x}S), magnetita (Fe₃O₄) y entre otros minerales. El análisis geoquímico reveló que las estructuras mineralizadas contienen concentraciones significativas de Au, Ag y Cu, cuyos resultados relevantes, analizados en los laboratorios acreditados SGS y CERTIMIN, se presentan en la Tabla 1. Estas concentraciones sugieren un alto potencial minero del yacimiento. La mineralización en las vetas está compuesta por cuarzo y pirita, asociado con el caolín, identificado mediante el Espectrómetro Portátil Infrarrojo de Onda Corta (PIMA). En el skarn, la mineralización incluye calcopirita, pirita, hematita, vetillas de cuarzo, calcita, magnetita, granates y minerales supérgenos como malaquita y crisocola, similar al estudio realizado por [12]. Las brechas de cuarzita presentan mineralización diseminada de pirita con venillas de cuarzo; cuya secuencia paragenética de la mineralización del yacimiento se muestra en la fig.7

TABLA 1

Composición geoquímica relevante en las muestras mineralizadas.

Muestra	Ancho (m)	Ley gr/tn Au	Ley gr/tn Ag	Cu (ppm)	Estructura
1026	0.70	15.65	1.70	17	Skarn
1251	0.60	9.88	140.00	13300	Skarn
55681	C Rock	12.70	5.00		Veta A
55682	0.40	9.63	3.50		Veta B
55737	0.80	0.80	0.60	24	Diseminada
55757	C.Rock	1.00	7.60	134	Diseminada
55763	1.10	4.11	13.90	316	M. Dorado
55765	1.00	6.63	17.30	176	M. Dorado
55766	0.70	25.10	24.80	292	M. Dorado

MINERAL	Estadio de Mineralización - Vetos-Mantos			
	1er Estadio	Estadio Intermedio	Estadio Principal	POSTMINERAL
Cuarzo	██████████	██████████	██████████	
Molibdenita		██████████		
Pirita	██████████	██████████	██████████	
Calcopirita		██████████		
Esfalerita		██████████		
Cobres Grises		██████████		
Galena		██████████		
Sulfosal Plomo			██████████	
Electrum			██████████	
Oro			██████████	
Calcita				██████████
Oxidos				██████████

Fig. 7. Secuencia paragenética de las principales estructuras mineralizadas del yacimiento

4.5 Resultados de Perforación

Basándose en la información litológica, la interpretación del modelo estructural y el análisis geoquímico de muestras de superficie, se llevaron a cabo exploraciones mediante perforaciones diamantinas dirigidas a las estructuras que reportaron valores significativos. Los resultados de estas exploraciones, presentados en la Tabla 2, destacan al Manto Dorado como la principal estructura mineralizada, con valores de hasta 23.80 gr/tn de Au, y a la Veta Rosa con 13.80 gr/tn de Au, además de la mineralización del skarn y otras vetas. Asimismo, en la Tabla 3 se presentan los recursos minerales estimados mediante métodos geológicos y geométricos, basados en los resultados obtenidos durante la exploración del yacimiento. Estos datos son insumos cruciales para una evaluación económica [13].

TABLA 2

Resultados de la exploración con perforación Diamantina.

Muestra	Ancho (m)	Ley gr/tn Au	Ley gr/tn Ag	Cu (ppm)	Estructura
TCD1-02	6.00	7.33	55.23	7058.33	Veta 1
TCD3-02	4.00	8.20	1.69	636.5	Veta Z
TCD6-02	2.00	12.25	0.73	9.8	Veta A
HDD3-08	1.40	12.80	14.2	77	M. Dorado
HDD5-08	0.73	4.24	0.6	210	M. Dorado
HDD6-08	1.20	23.80	2.8	773	M. Dorado
HDD7-08	0.75	13.80	2.8	544	Veta Rosa
HDD9-08	0.35	17.70	1.8	13	Veta Rosa
HD11-08	2.76	6.33	2.28	20	Veta Rosa
HD19-08	0.55	15.00	19.7	27	Skarn
HDD20-08	1.05	1.21	0.50	30	Skarn

TABLA 3
Recursos estimados del yacimiento

Mineralización	Ancho (m)	Ley Oz/tn Au	Recursos (Tn)
Manto Dorado	2.01	0.48	86,671.00
Veta Rosa	1.98	0.21	116,228.00
Veta A	1.57	0.17	49,235.00
Veta 1	1.83	0.36	8,943.00
Skarn	5.45	0.49	305,200.00

5 POTENCIAL ECONÓMICO

Las áreas con alta concentración de elementos metálicos, específicamente el Manto Dorado y la Veta Rosa, presentan un considerable potencial económico para la actividad minera. Los modelos geológicos desarrollados indican que estas zonas son viables para la exploración y extracción de minerales. Los análisis geoquímicos realizados revelaron concentraciones significativas de oro (Au), plata (Ag) y cobre (Cu) en las muestras recogidas, con valores de hasta 25.10 ppm de Au y 24.8 ppm de Ag en afloramientos, lo cual sugiere un alto potencial de mineralización. Las exploraciones mediante perforaciones diamantinas han reportado resultados prometedores, como el sondeo TCD6-02 que registró hasta 12.25 ppm de Au para 2.00 metros de ancho y el sondeo HDD6-08 que registró 23.8 ppm para 1.20 metros de ancho. La estimación de recursos, presentada en la Tabla 4, indica la presencia de una cantidad significativa de recursos minerales en el yacimiento. Estos datos se basan en un análisis exhaustivo de las muestras y en modelos geológicos detallados, lo que proporciona una base sólida para la evaluación económica. La mineralización controlada estructuralmente y las elevadas concentraciones de oro (Au) hacen que el área de estudio sea altamente atractiva para la minería, según indica [14].

5 DISCUSIÓN

El estudio del yacimiento Tumipampa muestra una mineralización controlada estructuralmente, con alta concentración de elementos metálicos principalmente oro (Au). La evaluación geoquímica y las perforaciones diamantinas indican un significativo potencial económico para la explotación minera. Los modelos geológicos y los datos obtenidos de las muestras sugieren que las intrusiones de granodiorita y las calizas del Cretácico Superior son unidades litológicas clave para la formación de skarns de cobre y hierro [15]. La alta concentración de Au y Ag en las muestras de afloramiento y perforación refuerza la viabilidad de futuras exploraciones y explotaciones mineras en el área.

La interpretación del control estructural concluye que el fracturamiento y fallamiento preliminar, emplazados dentro de la cuarcita y el intrusivo, han jugado un papel crucial al servir como conductos para las soluciones hidrotermales mineralizantes y como receptores en las brechas de cuarcita, el Manto Dorado, la Veta Rosa y otras vetas menores. Asimismo, el contacto entre la caliza y el intrusivo ha sido una condición fundamental para la formación del depósito tipo skarn indica [16].

6 CONCLUSIONES

El área de estudio presenta un alto potencial para la actividad minera, particularmente en las áreas del Manto Dorado y la Veta Rosa, debido a las elevadas concentraciones de elementos metálicos como el oro (Au). Las investigaciones detalladas, incluyendo el mapeo geológico, el análisis litológico, las interpretaciones estructurales y los estudios geoquímicos, han permitido identificar estructuras geológicas significativas que controlan la mineralización. Los resultados sugieren que las zonas identificadas son viables para la exploración y extracción de minerales, destacando la importancia de continuar con las actividades de exploración para incrementar los recursos minerales estimados.

7 AGRADECIMIENTOS

Agradezco a la gerencia y equipo de colaboradores de la compañía Tumipamas SAC.

REFERENCIAS BIBLIOGRÁFICAS

- [1] J. Smith y R. Brown, "Geological Mapping Techniques," *Journal of Geology*, vol. 34, no. 2, pp. 123-135, 2020.
- [2] INGEMMET, "Geología del Perú, serie A: carta geológica nacional," Lima, octubre 1995.
- [3] M. Johnson, "Geochemical Analysis Methods," *Geochemical Journal*, vol. 56, no. 4, pp. 223-234, 2021.
- [4] V. Carlotto, L. Cerpa, H. Acosta, y R. Rodríguez, "Evolución tectónica y metalogenética de la litosfera del sur del Perú: Resultado de la acreción de bloques alóctonos," INGEMMET.
- [5] A. White y B. Green, "Diamond Drilling in Mineral Exploration," *Mining Science*, vol. 45, no. 3, pp. 456-467, 2019.
- [6] P. N. Balladares, "Geología económica y estudio geoambiental de la mina magistral," Huaraz, Cusco, 2005.
- [7] B. Bellido y D. E. De Montreuil, "Aspectos generales de la metalogenia del Perú," Ministerio de Energía y Minas, 1972.
- [8] S. E. Cardoso, "Modelo genético preliminar del yacimiento tipo skarn de cobre de Chalcobamba," Lima, 1999.
- [9] G. Ventura, "Perforación diamantina proyecto Tumipampa," Informe interno, Mines Dynacor Sucursal del Perú, septiembre 2003.
- [10] G. Jordan y H. S. Zarate, "Reconocimiento geológico y posibilidades económicas del prospecto minero Nuevo Amanecer," enero 1997.
- [11] K. Steinmuller, "Depósitos metálicos en el Perú: su metalogenia, sus modelos, su exploración y el medio ambiente," INGEMMET, 1999.
- [12] C. Almonacid y A. Tomás, "Geología y exploración del prospecto tipo skarn de Fe (Au) de Pitamarca - Andahuaylas - Apurímac," Lima, 1998.
- [13] N. Lewis, "Economic Evaluation of Mineral Deposits," *Mining Economics*, vol. 50, no. 2, pp. 201-215, 2020.
- [14] L. Black y P. Gray, "Structural Controls on Mineralization," *Geological Society Review*, vol. 59, no. 1, pp. 34-49, 2018.
- [15] Loayza C. Yimi, Evaluación geológica del prospecto aurífero Coriminas Chalhuanca - Aimaraes, agosto, 2002.

- [16] T. A. Sanchez, C. C. Mendoza, G. Oblitas, "Informe ejecutivo exploraciones Greenfields Tumipampa," Informe interno, Minera Dynacor del Perú S.A., junio, 2008.

BIOGRAFÍAS

Carlos Mendoza Contreras es Ingeniero Geólogo (UNSAAC, 2009) con una maestría en Gestión Minera (UNI, 2024). Está especializado en riesgos naturales (CENEPRED) y gestión ambiental (UNALM), además de contar con una diplomatura en gestión pública (PUCP). Con 20 años de experiencia, ha sido jefe de geología en operaciones mineras y jefe de proyectos de exploración minera en yacimientos polimetálicos, de oro, pórfidos y skarn, liderando proyectos de exploración y operaciones mineras. Es fundador de Gepromin SAC y docente invitado en la UNAMBA. Analista del mercado bursátil con énfasis en el sector minero, es miembro de la Sociedad Geológica del Perú, del Instituto de Ingenieros de Minas del Perú y del Colegio de Ingenieros del Perú.

Dr. Germán Rafael Espinoza Rivas es especialista en geología, geotecnia, evaluación de riesgos y medio ambiente. Tiene formación en Ingeniería Geológica, Civil y Topografía en Perú, con una maestría en Ingeniería Civil y Medio Ambiente (Universidad de Utah) y un doctorado en Ciencia, Tecnología y Medio Ambiente. Desde 1984, ha liderado estudios geológicos y de riesgos, y es docente en la Universidad Nacional Micaela Bastidas de Apurímac. Publica sobre geotecnia y evaluación de riesgos.

Ingeniero Alonso Sánchez es CEO de Consultora Trading SAC y Persona Competente según NI 43-101 / JORC, con 27 años de experiencia en la industria minera. Graduado en Geología (UNI) y con maestrías en Administración de Empresas y Finanzas (Universidad de Chicago). Ha trabajado en todas las fases de la minería en Perú, Bolivia, Canadá, Europa y EE. UU. Es miembro del Colegio de Ingenieros del Perú y del American Institute of Professional Geologists, y profesor a tiempo parcial en la UNI.

Análisis del Sentimiento con NLP en tutoría académica: la vocación como factor de mejora en el rendimiento académico

Sentiment analysis with NLP in academic tutoring: vocation as a factor for improvement in academic performance

Jorge Campos-Malpartida^A, Erech Ordoñez-Ramos^B y Alejandrina Huaylla-Quispe^C

Resumen— Las tutorías en la Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA), obedecen a una estructura desplegada en toda la universidad, el cual evalúa y busca fortalecer cuatro factores o variables, entre ellas la vocación y el rendimiento académico los cuales son más afines a los docentes, ya que las otras están más ligadas a las ciencias de la salud mental como la psicología, y el área de personal social, es así que en esta investigación se priorizan dos variables desde un diagnóstico general, así como de una evaluación de datos adquiridos de los tutorados desde la aplicación de librerías de Inteligencia Artificial (IA) que miden el sentimiento utilizando el procesamiento del lenguaje natural (NLP).

Palabras clave: Vocación, Rendimiento Académico, Inteligencia Artificial (IA) y Procesamiento de Lenguaje Natural (NLP)

Abstract— The tutoring program at the Professional School of Computer Science and Systems Engineering of the National University Micaela Bastidas de Apurímac (UNAMBA) follows a structure implemented throughout the university, which evaluates and aims to strengthen four factors or variables. Among these, vocation and academic performance are more closely related to the teachers, while the others are more linked to mental health sciences such as psychology the social personnel area. Therefore, this research prioritizes two variables based on a general diagnosis, as well as an evaluation of data acquired from the tutored students using Artificial Intelligence (AI) libraries that measure sentiment through natural language processing (NLP).

Keywords: Vocation, Academic Performance, Artificial Intelligence (AI) and Natural Language Processing (NLP)

1 INTRODUCCIÓN

Los estudios [1]-[6] citados indican que los docentes universitarios que no pertenecen a profesiones como psicología, psiquiatría, medicina o ciencias del deporte tienen dificultades para abordar adecuadamente temas de salud mental y física en las tutorías universitarias. Esto sugiere la necesidad de replantear las tutorías y sus estructuras, especialmente en las áreas de "Personal Social" y "Salud Corporal y Mental"[7].

Siendo el objetivo de la investigación la evaluación de la variable independiente vocación como factor de mejora del rendimiento académico, los resultados al respecto de dicha variable independiente, muestran una mayoría de respuestas neutrales o "de acuerdo", cuando lo ideal o esperado, sería una mayoría en "muy de acuerdo" concluyendo en consecuencia un bajo rendimiento académico. Sin embargo, el 72.1% de los estudiantes elige ser emprendedor al finalizar la carrera, lo que indicaría la necesidad de incluir el

emprendimiento en el plan de estudios.

Finalmente, la medición del sentimiento con NLP para la vocación revela un 22.5% de respuestas neutrales y un 21.5% negativas sobre esta variable, lo cual en consecuencia estaría afectando negativamente el rendimiento académico en el semestre 2024-I, en conclusión, debemos enfocarnos en esta variable focalizada en esta investigación como es la vocación para las tutorías académicas y de esta manera se estaría mejorando el rendimiento académico.

2 PROCEDIMIENTO

2.1 Etapa de Revisión

Al realizar el estado del arte respecto a las cuatro variables consideradas en la tutoría universitaria aplicada en la Escuela Profesional de Ingeniería Informática y Sistemas que son: vocación, rendimiento académico, salud física y salud mental, se pudo evaluar que los docentes universitarios en general carecen de habilidades formales por formación profesional para abordar idónea y adecuadamente las dos variables: salud



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 22/03/24 - Indexaciones

Número: doi.org/10.57166/riqchary/V6.n2.2024

Páginas: 1- 13 | Recibido 04/07/2024 ; Aceptado 31/07/2024

<https://doi.org/10.57166/riqchary.v6.n2.2024.123>

Autores:

- A. **ORCID iD** <https://orcid.org/0000-0002-6063-0771>
Jorge Campos-Malpartida, Universidad Nacional Micaela Bastidas de Apurímac, jcampos@unamba.edu.pe
- B. **ORCID iD** <https://orcid.org/0000-0002-3058-9683>
Erech Ordoñez-Ramos, Universidad Nacional Micaela Bastidas de Apurímac, eordonhez@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0000-0002-0258-8632>
Alejandrina Huaylla-Quispe, Universidad Nacional Micaela Bastidas de Apurímac, ahuaylla@unamba.edu.pe

mental [1] [2] [3] [4], y salud física [5] [6] [7], asociadas a profesionales de la salud como son los psicólogos, psiquiatras, medicina o ciencias del deporte, y si más bien podrían abordar las variables: vocación y rendimiento académico con mejor posibilidad y solvencia [8][9][10][11][12], así entonces para este estudio se descartan evaluar dichas variables para enfocarnos en dos de ellas: vocación y rendimiento académicos, a la pregunta cual influye como variable principal, estudios han demostrado que es la vocación quien influye en el rendimiento académico [13][14][15][16]. Para evaluar la vocación en los estudiantes tutorados se diseñó una encuesta de diagnóstico basado en estándares de medición en instituciones como la ONU (Encuesta Mundial sobre la Juventud) [17], UNESCO (Global Education Monitoring Report) [18] y Foro Económico Mundial (Future of Jobs Report) [19], dicho instrumento elaborado, consta de 9 preguntas, 7 en escala de liker, 1 en opción cerrada y una pregunta abierta final, la cual será sometida a evaluación de medición del sentimiento con librería de procesamiento de lenguaje natural NPL, dicha librería será elegida en un ejercicio de benchmarking para obtener la idónea en dicha tarea específica que tenemos en este estudio.

2.2 Elaboración del instrumento de recolección

Tomando en cuenta las referencias anunciadas se elabora la siguiente encuesta de 9 preguntas, tomando en cuenta la escala de liker para las 7 primeras preguntas:

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Neutral
- (4) De acuerdo
- (5) Totalmente de acuerdo

Las preguntas son:

1. ¿Estoy entusiasmado/a por aprender más sobre mi campo de estudio?
2. ¿Mi carrera elegida refleja mis intereses personales y profesionales?
3. Me siento preparado/a para enfrentar los desafíos en mi campo de estudio?
4. Estoy seguro/a de que encontraré un empleo en mi campo después de graduarme
5. Tengo acceso a los recursos y apoyo necesarios para tener éxito en mi carrera
6. Considero que la educación que estoy recibiendo es de alta calidad.
7. Creo que mi carrera elegida tendrá un impacto positivo en la sociedad.

La pregunta octava, es cerrada con tres posibles respuestas excluyente, como a continuación se presenta:

8. Cuando termine, quisiera ser:
 - EMPLEADO (busca empleo y piensa trabajar por el sueño de otros hasta jubilarse, sueldo seguro nunca podrá ser millonario).
 - EMPRENDEDOR/EMPRESARIO DIGITAL

(riesgo basado en la Innovación y el emprendimiento, negocio propio, auto empleado o inversionista, se dedica a luchar por sus propios sueños tiene posibilidad de volverse millonario).

- INVESTIGADOR (menos riesgo que ser emprendedor e innovador, con posibilidad de volverse millonario ya que sus descubrimientos y patentes podrían venderse en millones, también vive y trabaja por sus propios sueños).

Finalmente, la pregunta novena es abierta, y se presenta de la siguiente manera:

9. Finalmente, déjanos una autodefinición de tu vocación por tu carrera profesional de ingeniería informática y sistemas, se lo más explícito posible, detallando los verdaderos motivos por los cuales la eligiste.

Calculando el tamaño de la muestra de una población de 440 estudiantes se obtiene:

$$n = 205.4 \approx 206$$

Por lo tanto, se necesitaría una muestra de aproximadamente 206 individuos para tener una muestra representativa con un nivel de confianza del 95% y un margen de error del 5% [27] [28].

Calculando el Alfa de Cronbach

El instrumento diseñado (estadístico) es puesto a prueba, y para ello se calcula el Alfa de Cronbach, se utiliza Python sobre Colab y a continuación se presente los resultados obtenidos, tanto numérico como gráficos.

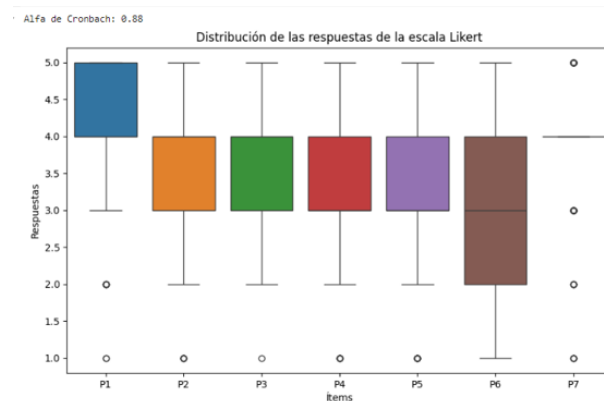


Fig. 1. Distribución de las respuestas de la escala de likert

En la Fig. 1, el gráfico de caja y bigotes ilustra la distribución de las respuestas en una escala Likert para siete ítems (P1-P7). Se observa que el ítem P1 presenta la mediana más alta, alrededor de 4.5, indicando respuestas generalmente positivas, mientras que los ítems P2 a P7 tienen medianas que oscilan entre 3.0 y 4.0. Los ítems P2, P4 y P6 muestran mayor dispersión en las respuestas, con rangos intercuartílicos más amplios. Adicionalmente, se identifican valores atípicos en todos los ítems, destacando P6 y P7 con más outliers, lo que sugiere

variabilidad en las percepciones de los encuestados.

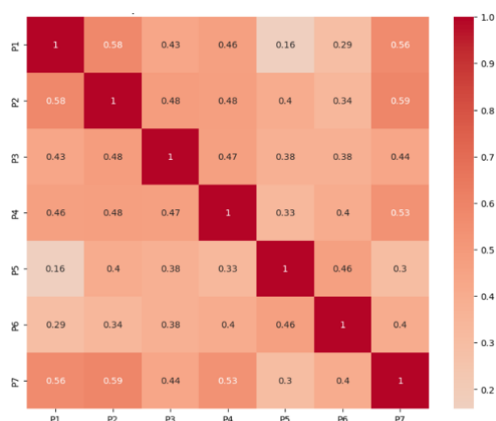


Fig. 2. Mapa de calor de la matriz de correlación

En la Fig. 2, el mapa de calor presentado muestra la matriz de correlación entre siete ítems (P1-P7) en una encuesta con escala Likert. Se observa que las correlaciones más altas se encuentran entre los ítems P2 y P7 (0.59), P2 y P1 (0.58), y P7 y P1 (0.56), lo que indica una relación positiva moderada entre estas respuestas. Los ítems P5 y P1 (0.16), y P5 y P6 (0.33) exhiben las correlaciones más bajas, sugiriendo una menor relación entre ellos. Este análisis permite identificar las relaciones y patrones en las respuestas, facilitando la comprensión de cómo se asocian las percepciones de los encuestados sobre los diferentes ítems.

2.3 Eligiendo librería NLP (Natural Processing Language)

Las librerías para medir el sentimiento desde el procesamiento de lenguaje natural, son muy variadas, y según sean el propósito están pueden ser elegidas, en este caso en particular, estamos frente a la variable vocación, la cual está siendo diagnóstica y medida, en ese sentido se realiza un benchmarking para elegir una idónea para este tipo de precisiones, las que destacan en términos de precisión, facilidad de uso y comunidad de soporte. A continuación, se presenta un benchmarking de las principales librerías de análisis de sentimiento:

1. NLTK (Natural Language Toolkit)
2. TextBlob
3. VADER (Valence Aware Dictionary and Sentiment Reasoner)
4. SpaCy
5. BERT (Bidirectional Encoder Representations from Transformers)

1. NLTK (Natural Language Toolkit) [20]

- **Descripción:** Una de las librerías más antiguas y completas para-NLP en Python. Ofrece herramientas para el análisis léxico, etiquetado, análisis sintáctico, etc.

- **Ventajas:** Gran cantidad de recursos y documentación, altamente configurable.
- **Desventajas:** Puede ser compleja de utilizar para principiantes, menos precisa para análisis de sentimiento comparado con métodos más modernos.

2. TextBlob [21]

- **Descripción:** Basada en NLTK y Pattern, proporciona una API simple para tareas comunes de NLP como el etiquetado, traducción, análisis de sentimiento, etc.
- **Ventajas:** Fácil de usar, buena para prototipos rápidos, integración con NLTK y Pattern.
- **Desventajas:** Menos precisa en comparación con modelos más avanzados como BERT.

3. VADER (Valence Aware Dictionary and Sentiment Reasoner) [22]

- **Descripción:** Diseñada específicamente para el análisis de sentimiento en textos sociales (tweets, comentarios, etc.). Basada en una combinación de reglas y un diccionario específico.
- **Ventajas:** Muy precisa para textos cortos e informales, fácil de implementar.
- **Desventajas:** Menos eficaz para textos más largos y complejos.

4. SpaCy [23]

- **Descripción:** Una librería moderna y eficiente para NLP en Python. Ofrece herramientas de procesamiento de texto, etiquetado, análisis sintáctico y más.
- **Ventajas:** Muy rápida y eficiente, fácil de integrar con otras librerías y herramientas, comunidad activa.
- **Desventajas:** Requiere más recursos computacionales, curva de aprendizaje más pronunciada.

5. BERT (Bidirectional Encoder Representations from Transformers) [24]

- **Descripción:** Un modelo basado en transformadores desarrollado por Google, que ha revolucionado el campo de NLP con su capacidad para comprender el contexto bidireccional de las palabras.
- **Ventajas:** Alta precisión, capaz de manejar contextos complejos, adaptable a diversas tareas de NLP.
- **Desventajas:** Requiere muchos recursos computacionales, implementación y ajuste más complejos.

Tabla 1.
Cuadro Comparativo de Librerías NPL

Librería	Precisión	Facilidad de Uso	Rendimiento	Documentación y Soporte	Contexto Ideal
NLTK	Media	Baja	Media	Alta	Investigación académica
Text-Blob	Media	Alta	Media	Media	Prototipos rápidos
VADER	Alta	Alta	Alta	Media	Textos cortos/informales
SpaCy	Alta	Media	Alta	Alta	Aplicaciones rápidas/robustas
BERT	Muy Alta	Baja	Baja	Alta	Contextos complejos

La elección de la librería adecuada dependerá de los requisitos específicos del proyecto. Si se busca facilidad de uso y rapidez en la implementación, **TextBlob** o **VADER** pueden ser buenas opciones. Como es la precisión la prioridad aun manejando una implementación más compleja, **BERT** es la mejor elección. **SpaCy** ofrece un buen equilibrio entre rendimiento y precisión, mientras que **NLTK** es ideal para un enfoque más académico. Se elige **BERT** [25] [26].

3 RESULTADOS

Resultados por evaluación de las siete primeras preguntas en la escala de liker:

1. ¿Estoy entusiasmado/a por aprender más sobre mi campo de estudio?

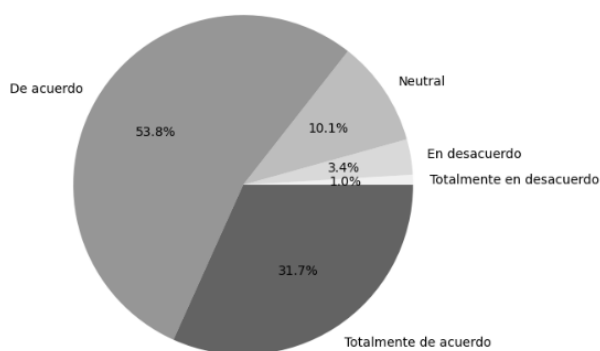


Fig. 3. Resultados de la pregunta N° 1

La Fig. 3 presenta la distribución de respuestas para la Pregunta 1 de la encuesta. El 53.8% de los encuestados se mostró "De acuerdo", seguido por un 31.7% que indicó estar "Totalmente de acuerdo". Un 10.1% se mantuvo "Neutral", mientras que el 3.4% respondió "En desacuerdo" y solo el 1.0% se manifestó "To-

talmente en desacuerdo". Estos resultados evidencian una fuerte tendencia hacia el acuerdo con la afirmación planteada.

2. ¿Mi carrera elegida refleja mis intereses personales y profesionales?

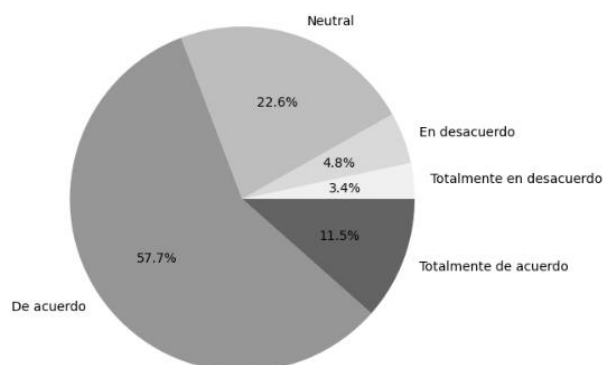


Fig. 4 Distribución de respuestas pregunta N° 2

La Fig. 4 muestra la distribución de respuestas para la Pregunta 2 de la encuesta. El 57.7% de los encuestados se mostró "De acuerdo", seguido de un 22.6% que se mantuvo "Neutral". El 11.5% indicó estar "Totalmente de acuerdo", mientras que el 4.8% respondió "En desacuerdo" y el 3.4% se manifestó "Totalmente en desacuerdo". Estos resultados indican una inclinación significativa hacia el acuerdo con la afirmación evaluada.

3. Me siento preparado/a para enfrentar los desafíos en mi campo de estudio?

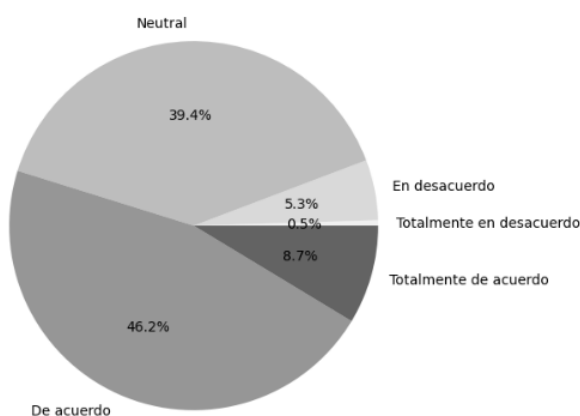


Fig. 5 Distribución de respuestas pregunta N° 3

La Fig. 5, muestra la distribución de respuestas para la Pregunta 3 de la encuesta. El 46.2% de los encuestados se mostró "De acuerdo" con la afirmación, seguido de un 39.4% que se mantuvo "Neutral". Un 8.7% indicó estar "Totalmente de acuerdo", mientras que el 5.3% se manifestó "En desacuerdo" y solo un 0.5% respondió "Totalmente en desacuerdo". Estos resultados sugieren una tendencia general hacia el acuerdo entre los encuestados.

4. Estoy seguro/a de que encontraré un empleo en mi campo después de graduarme

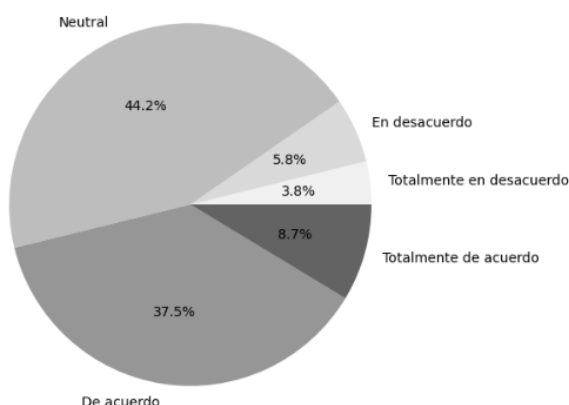


Fig. 6 Distribución de respuestas pregunta N° 4

La Fig. 6, muestra la distribución de respuestas para la Pregunta 4 en un gráfico de pastel. Los datos indican que el 44.2% de los encuestados se sienten neutrales respecto a la pregunta. Un 37.5% de los participantes está de acuerdo, mientras que un 8.7% está totalmente de acuerdo. Por otro lado, el 5.8% está en desacuerdo y el 3.8% está totalmente en desacuerdo. Esta distribución sugiere una tendencia hacia la neutralidad y el acuerdo entre los encuestados, con una menor proporción de respuestas negativas.

5. Tengo acceso a los recursos y apoyo necesarios para tener éxito en mi carrera

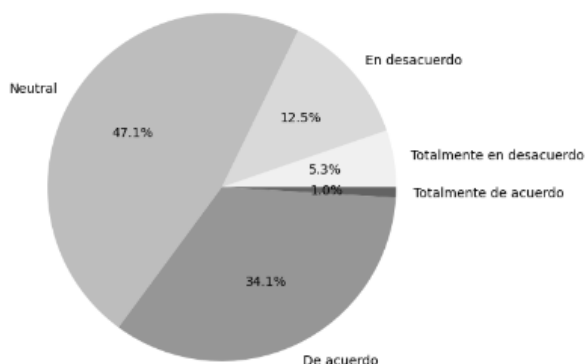


Fig. 7 Distribución de respuestas pregunta N° 5

La Fig. 7, muestra la distribución de respuestas para la Pregunta 5 en un gráfico de pastel. Los resultados indican que el 47.1% de los encuestados se sienten neutrales respecto a la pregunta. Un 34.1% de los participantes está de acuerdo, mientras que solo el 1.0% está totalmente de acuerdo. En contraste, el 12.5% está en desacuerdo y el 5.3% está totalmente en desacuerdo. Esta distribución revela una inclinación significativa hacia la neutralidad y el acuerdo, con una menor proporción de respuestas fuertemente positivas o negativas.

6. Considero que la educación que estoy recibiendo es de alta calidad.

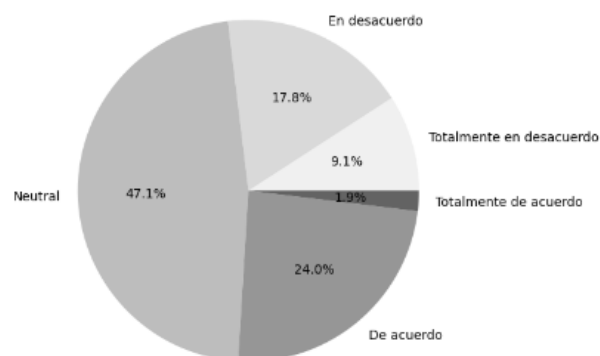


Fig. 8 Distribución de respuestas pregunta N° 6

La Fig. 8 presenta la distribución de respuestas para la Pregunta 6 en un gráfico de pastel. Los datos revelan que el 47.1% de los encuestados se sienten neutrales respecto a la pregunta. Un 24.0% de los participantes está de acuerdo, y un 1.9% está totalmente de acuerdo. Por otro lado, el 17.8% está en desacuerdo y el 9.1% está totalmente en desacuerdo. Esta distribución muestra una tendencia marcada hacia la neutralidad, seguida por el acuerdo, con una presencia significativa de respuestas negativas.

7. Creo que mi carrera elegida tendrá un impacto positivo en la sociedad.

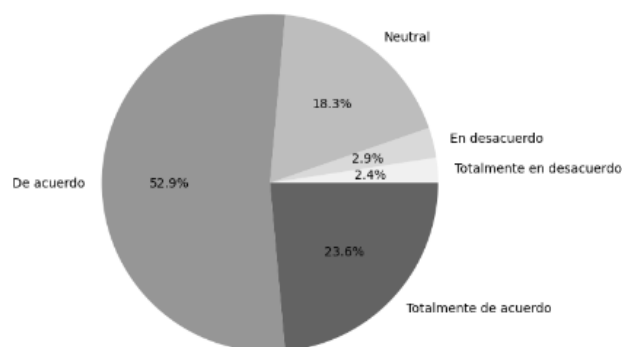


Fig. 9 Distribución de respuestas pregunta N° 7

La Fig. 9, muestra la distribución de respuestas para la Pregunta 7, revelando que el 76.5% de los encuestados se encuentra en las categorías "De acuerdo" (52.9%) y "Totalmente de acuerdo" (23.6%), indicando una fuerte tendencia hacia la concordancia. Un 18.3% se mantuvo neutral, mientras que las respuestas negativas fueron minoritarias, con solo el 5.3% en desacuerdo, distribuidos entre "En desacuerdo" (2.9%) y "Totalmente en desacuerdo" (2.4%). En conjunto, estos resultados sugieren una predominante aprobación entre los participantes.

Para la Pregunta cerrada número 8, se ha evaluado la vocación desde tres importantes sectores posibles de desempeño profesional, como ya se detalló anteriormente, y se muestra el resultado.

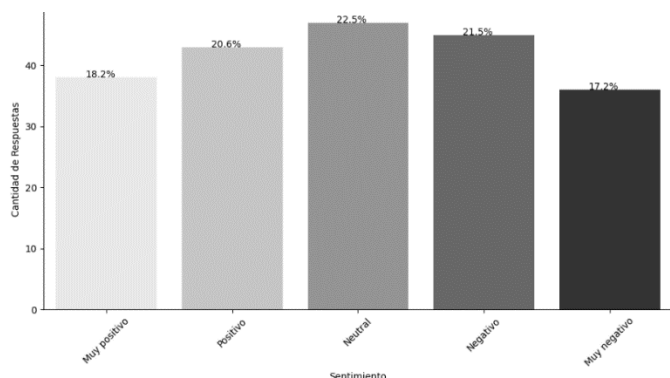


Fig. 10 Distribución de respuestas en porcentajes de la pregunta N° 8

La Fig. 10, muestra la distribución de respuestas para una pregunta cerrada, indicando que la mayoría de los encuestados se identifican como "Emprendedores" con 72.1% (150) respuestas, seguidos por "Investigadores" con 20.3% (43) respuestas, y finalmente "Empleados" con 7.2% (15) respuestas. Estos resultados sugieren una predominancia de emprendedores entre los encuestados, con una notable menor cantidad de empleados e investigadores.

3.2 MEDICIÓN DEL SENTIMIENTO CON NLP (EBERT LIBRARY)

La utilización de la inteligencia artificial a través de librerías como BERT desde Python permite obtener una medición expectante de la veracidad en los sentimientos expuestos por los individuos encuestados en la pregunta abierta correspondiente presentada, "Finalmente, déjanos una autodefinición de tu vocación por tu carrera profesional de ingeniería informática y sistemas, se lo más explícito posible, detallando los verdaderos motivos por los cuales la elegiste", a dicha medición se ha obtenido los resultados siguientes:

Fig. 11 Distribución de sentimientos en respuestas vocacionales

La Fig. 11, muestra la distribución de sentimientos en respuestas vocacionales, donde el 22.5% de los encuestados se posiciona como "Neutral". Las categorías "Negativo" y "Positivo" siguen con 21.5% y 20.6%, respectivamente. Los extremos "Muy positivo" y "Muy negativo" son los menos frecuentes, con 18.2% y 17.2%, respectivamente. Estos resultados sugieren una distribución equilibrada de sentimientos, con una ligera inclinación hacia la neutralidad y sentimientos negativos.

3 DISCUSIONES

A las referencias [1]-[7] citadas y evaluadas cuyos estudios

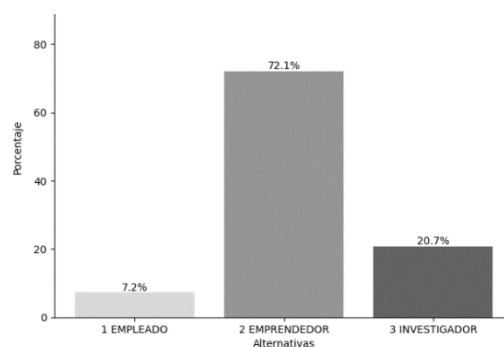
demuestran que los docentes universitarios en general son distantes a profesiones como: la psicología, la psiquiatría, la medicina o ciencias del deporte; por lo cual, indican que es inadecuado abordar áreas concernientes a la salud mental y física consideradas dentro de la tutoría universitaria de acuerdo a [29] como; "PERSONAL SOCIAL (Favorece la construcción de una personalidad saludable y equilibrada) y SALUD CORPORAL Y MENTAL (Propicia la adquisición de estilos de vida saludables. El completo estado de bienestar físico, mental y social)" referidos de [29], lo cual sugiere replantear dichas tutorías y sus estructuras.

Respecto a los resultados en la evaluación de la vocación, estos reflejan que en su mayoría los jóvenes tienen un convencimiento neutral y a favor de la opción "de acuerdo" en función a las preguntas planteadas, cuando lo ideal sería una respuesta mayoritaria o contundente para la opción "Muy de acuerdo"; sin embargo, a la elección de la preferencia de la actividad profesional después del término de su carrera hay un contundente 72.1% que elige ser EMPRENDEDOR, lo cual invita a replantear el plan de estudios, que debe incluir esta necesidad expuesta en esta investigación.

Del resultado de evaluar la vocación desde la medición del sentimiento con NPL, se observan como los estudiantes tienen un convencimiento neutral de su vocación con un 22.5% y Negativo con un 21.5%, que representan los porcentajes más altos, mostrando una vocación débil, y en consecuencia dadas las referencias [13-16] se demuestran debilidades en el rendimiento académico.

4 CONCLUSIONES

Los estudios citados indican que los docentes universitarios que no pertenecen a profesiones relacionadas con la psicología, psiquiatría, medicina o ciencias del deporte no abordan adecuadamente temas de salud mental y física en tutorías universitarias. Esto sugiere la necesidad de replantear estas tutorías y sus estructuras. Además, los resultados de la evaluación de la vocación muestran respuestas mayoritariamente neutrales o "de acuerdo" en lugar de "muy de acuerdo", con un 72.1% de estudiantes eligiendo ser emprendedores al terminar la carrera, lo que sugiere incluir el emprendimiento



en el plan de estudios. Finalmente, la medición del sentimiento con NPL revela un 22.5% de respuestas neutrales y un 21.5% negativas sobre la vocación, lo cual debilita la vocación y afecta negativamente el rendimiento académico, según las

referencias [13-16].

REFERENCIAS

- [1] J. García, M. Fernández, "La preparación de los docentes para abordar la salud mental en contextos universitarios: Una revisión crítica", *Revista Peruana de Psicología*, pp. 14-15, 2023.
- [2] A. López, C. Mendoza, "Capacitación docente para la tutoría universitaria: Enfoques para abordar la salud mental de los estudiantes", *Revista de formación docente*, pp. 28-30, 2022.
- [3] V. Díaz, L. Hernández, "Desafíos en la preparación de los docentes para manejar la salud mental en la tutoría universitaria", *Revista de Psicología y Educación*, pp. 33-34, 2022.
- [4] M. Martínez, R. López, "El papel de los docentes en la promoción de la salud mental: Una evaluación de competencias y preparación", *Revista de Psicología Educativa*, pp. 12-13, 2023.
- [5] C. Fernández, J. Martínez, "Desafíos en la preparación de los docentes para abordar la salud física en contextos de tutoría universitaria", *Educación Física y Deporte*, pp. 20-21, 2022.
- [6] E. Sánchez, P. García, "El rol limitado de los docentes en la promoción de la salud física: Un análisis de la preparación académica y profesional", *Revista de Educación en Ciencias de la Salud*, pp. 40-41, 2022.
- [7] L. Fernández, M. Torres, "La preparación de los docentes para promover la salud física en la universidad: Evaluación de competencias y estrategias", *Educación y Desarrollo en América Latina*, Educación y Desarrollo en América Latina, pp. 22-23, 2023.
- [8] J. Ruiz, M. López, "La influencia de la vocación en el rendimiento académico de los estudiantes universitarios: El papel de los docentes en el proceso de tutoría", *Revista Latinoamericana de Educación Inclusiva*, pp. 45-46, 2023.
- [9] L. Gutiérrez, A. Morales, "El rol de los docentes en el fomento de la vocación académica y su impacto en el rendimiento estudiantil", *Revista de Educación y Desarrollo*, pp. 57-59, 2022.
- [10] M. Fernández, J. García, "La competencia pedagógica de los docentes para influir en la vocación y el rendimiento académico de los estudiantes universitarios", *Revista de Ciencias de la Educación*, pp. 29-30, 2022.
- [11] C. Vargas, A. Pérez, "Capacitación docente para la tutoría universitaria en vocación y rendimiento académico: Un estudio en universidades peruanas", *Revista de Investigación en Educación Superior*, pp. 38-40, 2023.
- [12] E. González, R. Martínez, "Estrategias pedagógicas para mejorar la vocación y el rendimiento académico en la tutoría universitaria", *Educación y Sociedad en América Latina*, pp. 50-52, 2023.
- [13] J. Ruiz, M. López, "La influencia de la vocación en el rendimiento académico de los estudiantes universitarios", *Revista Latinoamericana de Educación Inclusiva*, pp. 45-47, 2023.
- [14] L. Gutiérrez, A. Morales, "El impacto de la vocación en el rendimiento académico de estudiantes universitarios: Un enfoque longitudinal", *Revista de Educación y Desarrollo*, pp. 102-104, 2022.
- [15] M. Fernández, J. García, "Motivación vocacional como predictor del éxito académico en educación superior", *Revista de Ciencias de la Educación*, pp. 56-57, 2022.
- [16] E. González, R. Martínez, "Vocación y desempeño académico: Un estudio comparativo en estudiantes de distintas disciplinas", *Educación y Sociedad en América Latina*, pp. 70-72, 2023.
- [17] Naciones Unidas (ONU), "The Sustainable Development Goals Report 2019.", Nueva York, pp. 34-39, 2019.
- [18] UNESCO, "Global Education Monitoring Report 2020: Inclusion and education: All means all.", Paris, pp. 78-85, 2020.
- [19] World Economic Forum, "The Future of Jobs Report 2020.", Geneva, pp. 55-60, 2020.
- [20] Bird, S., Klein, E., & Loper, E. *Natural Language Processing with Python*. O'Reilly Media. 2009.
- [21] Loria, S. TextBlob: Simplified Text Processing. Available at: textblob.readthedocs.io 2018
- [22] Hutto, C.J., & Gilbert, E.E. Vader: A Parsimonious Rule-based Model for Sentiment Analysis of social media Text. In Eighth International Conference on Weblogs and social media (ICWSM-14). 2014. <https://doi.org/10.1609/icwsml.v8i1.14550>
- [23] Honnibal, M., & Montani, I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Available at, 2017.
- [24] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. x BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint ar-Xiv:1810.04805. 2017
- [25] Manning, C.D., Raghavan, P., & Schütze, H. "Introduction to Information Retrieval. Cambridge University Press" 2008. <https://doi.org/10.1017/CBO9780511809071>
- [26] Jurafsky, D., & Martin, J.H. "Speech and Language Processing (3rd ed. draft). Stanford University" 2021.
- [27] Cochran, W. G. *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons. Páginas 75-80, 1977.
- [28] Daniel, W. W. *Biostatistics: A Foundation for Analysis in the Health Sciences* (7th ed.). New York: John Wiley & Sons. Páginas 189-192, 1999.
- [29] Oficina Psicopedagógica y Psicológica Universitaria, Formato de control individual del tutorado "Universidad Nacional Micaela Bastidas de Apurímac". Tamburco, Abancay.

Aplicación de Análisis de Componentes Principales (PCA) y K-Means para la Clasificación de Unidades Estratigráficas en Muestras Geoquímica del Volcán Misti

Application of Principal Component Analysis (PCA) and K-Means for Stratigraphic Unit Classification in Geochemical Samples from the Misti Volcano

Mary Luz Nina-Palacios ^A, Ernesto Nayer Tumi-Figueroa ^B y Hugo Ticona-Salluca ^C

Resumen— Este estudio presenta un análisis exhaustivo de los depósitos de tefra de la etapa Pacheco del volcán Misti, Arequipa, Perú, utilizando técnicas de reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA) y agrupamiento mediante K-Means. Se analizaron muestras geoquímicas obtenidas entre 2015 y 2023, enfocándonos en elementos mayores y traza como Ti, Fe, MgO, y SiO₂. El PCA permitió reducir la dimensionalidad de los datos, revelando que los dos primeros componentes explican el 34.27% de la varianza total, destacando la influencia de variables como Ti y Fe. A continuación, el algoritmo K-Means identificó cuatro clusters geoquímicamente diferenciados, los cuales se compararon con las unidades estratigráficas conocidas, mostrando una correlación notable con unidades como Ponche Gris y La Rosada. Los resultados sugieren que estos clusters reflejan variaciones en los procesos magmáticos y fases eruptivas, ofreciendo nuevas perspectivas sobre la evolución geoquímica del volcán Misti. Futuros trabajos podrían incluir análisis isotópicos y la integración de técnicas de aprendizaje profundo para mejorar la comprensión de las fuentes magmáticas y su influencia en la estratigrafía volcánica.

Palabras clave: Volcán Misti, PCA, K-Means, Geoquímica, Unidades Estratigráfica

Abstract— This study presents a comprehensive analysis of the Pacheco stage tephra deposits from Misti Volcano, Arequipa, Peru, using dimensionality reduction techniques through Principal Component Analysis (PCA) and clustering via K-Means. Geochemical samples collected between 2015 and 2023 were analyzed, focusing on major and trace elements such as Ti, Fe, MgO, and SiO₂. PCA allowed for dimensionality reduction, revealing that the first two principal components explained 34.27% of the total variance, with variables such as Ti and Fe having the most significant influence. Subsequently, the K-Means algorithm identified four geochemically differentiated clusters, which were compared with known stratigraphic units, showing a notable correlation with units such as Ponche Gris and La Rosada. The results suggest that these clusters reflect variations in magmatic processes and eruptive phases, offering new insights into the geochemical evolution of Misti Volcano. Future work could include isotopic analysis and the integration of deep learning techniques to enhance the understanding of magmatic sources and their influence on volcanic stratigraphy.

Keywords: Misti Volcano, PCA, K-Means, Geochemistry, Stratigraphic Units

1 INTRODUCCIÓN

El estudio de la composición geoquímica de minerales es esencial para comprender los procesos geológicos que han moldeado la Tierra. Los ortopiroxenos, minerales comunes en rocas ígneas y metamórficas, juegan un papel clave en la interpretación de la evolución magmática y tectónica de

diferentes regiones geológicas. Las concentraciones de óxidos metálicos, como FeO, TiO₂, y MnO, son indicadores importantes de las condiciones en las que se formaron estos minerales, como las temperaturas de cristalización y las presiones experimentadas en profundidad [1].

Tradicionalmente, la clasificación y estudio de las rocas y minerales se ha basado en unidades estratigráficas, que agrupan



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 09/10/24 - [Indexaciones](#)

Número: doi.org/10.57166/riqchary/V6.n2.2024

Páginas: 1-13 | Recibido 09/01/2024; Aceptado 09/10/2024



<https://doi.org/10.57166/riqchary.v6.n2.2024.124>

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0008-1773-5894>
Mary Luz Nina-Palacios, Universidad Nacional del Altiplano, Pe. maninap@est.unap.edu.pe.
- B. **ORCID iD** <https://orcid.org/0000-0003-0401-9027>
Ernesto Nayer Tumi-Figueroa, Universidad Nacional del Altiplano, Pe. nayer.tumi@unap.edu.pe.
- C. **ORCID iD** <https://orcid.org/0000-0002-3800-8433>
Hugo Ticona-Salluca, Universidad Nacional del Altiplano, Pe. hts.ez.v@gmail.com.

a las formaciones geológicas según características litológicas y cronológicas [2]. Si bien este enfoque ha sido ampliamente adoptado, las técnicas de Machine Learning están emergiendo como herramientas poderosas para detectar patrones ocultos en datos geológicos multivariados, proporcionando una perspectiva más detallada y objetiva sobre las relaciones entre las propiedades geoquímicas y los procesos geológicos [3].

El uso de Análisis de Componentes Principales (PCA) y algoritmos de clustering, como K-Means, ha demostrado ser efectivo para reducir la dimensionalidad de los datos y revelar patrones importantes que pueden no ser evidentes utilizando métodos tradicionales. Estos algoritmos permiten agrupar muestras en función de sus características geoquímicas, proporcionando información sobre la heterogeneidad geoquímica de las formaciones rocosas [4,5].

En este estudio, aplicamos estas técnicas de Machine Learning a un conjunto de datos de ortopiroxenos recolectados en diversas unidades estratigráficas de la región de [nombre de la región]. Mediante el uso de PCA, identificamos los principales factores que explican la variabilidad en las concentraciones de óxidos metálicos, y posteriormente, utilizamos K-Means para agrupar las muestras en clusters geoquímicos. Finalmente, comparamos estos clusters con las unidades estratigráficas tradicionales para evaluar si existe una correlación entre las agrupaciones detectadas y las clasificaciones geológicas previas.

2 TRABAJOS RELACIONADOS

La aplicación de técnicas de Machine Learning en el análisis de datos geoquímicos ha ganado popularidad debido a su capacidad para manejar grandes volúmenes de datos multivariados y revelar patrones complejos que son difíciles de detectar con métodos convencionales. Estudios recientes han demostrado que algoritmos como el Análisis de Componentes Principales (PCA) y el clustering, incluidos K-Means y otros métodos jerárquicos, son efectivos para identificar agrupamientos geoquímicos que pueden estar relacionados con procesos de formación mineral o evolución tectónica [6,7]. Estos enfoques permiten no solo una reducción dimensional significativa, sino también la detección de patrones geoquímicos que pueden tener una correlación directa con eventos geológicos, como la cristalización magmática y la alteración mineral [8].

Investigaciones como la de Jolliffe y Cadima [9] han destacado el uso de PCA para analizar la variabilidad geoquímica

en muestras de minerales, señalando que las concentraciones de elementos trazan y óxidos metálicos son determinantes clave en la diferenciación de entornos de formación mineral. Otros trabajos, como los de Olden y Jackson [10], han empleado redes neuronales para clasificar unidades geológicas, logrando mejorar las predicciones sobre la base de datos geoquímicos complejos. Sin embargo, estudios recientes de Feng et al. [11] han mostrado que algoritmos de clustering como K-Means son igualmente poderosos cuando se trata de agrupar muestras en clusters geoquímicos y correlacionarlos con unidades geológicas conocidas.

Además, investigaciones como las de Rollinson [12] y Boggs [13] han enfatizado la importancia de utilizar enfoques multivariados en geoquímica aplicada, particularmente en el análisis de formaciones magmáticas y sedimentarias. Estos estudios han demostrado que la combinación de métodos de Machine Learning con herramientas tradicionales de análisis geológico permite mejorar la comprensión de los procesos de formación de minerales y unidades estratigráficas. Estos avances metodológicos abren nuevas posibilidades para reinterpretar datos geoquímicos históricos, proporcionando así un contexto más detallado y profundo para la exploración geológica.

3 METODOLOGÍA

3.1 Ámbito de estudio

Este estudio se centra en los depósitos de tefra de la etapa Pacheco del volcán Misti, ubicado en Arequipa, Perú. Los datos geoquímicos fueron recolectados entre 2015 y 2023 y se refieren a los depósitos de caída de tefra formados por las erupciones del volcán Misti, entre <21 ka y 2 ka, en el sur de Perú. Este período incluye múltiples erupciones volcánicas de índice de explosividad volcánica (VEI) de 2 a 5.

Los datos provienen del análisis de muestras de piedra pómez (roca volcánica ligera y porosa) y escoria (roca volcánica densa y oscura) utilizando una microsonda electrónica que analizó los elementos mayores en minerales como vidrio volcánico, anfíbol, plagioclasa, ortopiroxeno, clinopiroxeno, y óxidos de Fe-Ti. Asimismo, en los casos en que estuvo presente, también se incluyó el análisis de olivino. Para cada unidad estratigráfica, se determinaron las concentraciones de elementos mayores y trazas mediante fluorescencia de rayos X, lo que permitió realizar un análisis petrológico detallado de cada unidad. Estos datos proporcionan un marco para el análisis multivariable de las características geoquímicas de las distintas unidades, ayudando a interpretar los procesos de formación de estos depósitos volcánicos.

3.2 Conjunto de Datos

En este estudio contiene 27 variables geoquímicas obtenidas mediante análisis de microsonda electrónica de ortopiroxenos. Estas variables incluyen concentraciones de óxidos y elementos traza que son fundamentales para la clasificación de las unidades estratigráficas. Las principales variables analizadas fueron:

Óxidos principales (en porcentaje en peso)

- SiO₂ (dióxido de silicio)
- FeO (óxido de hierro)
- MgO (óxido de magnesio)
- Al₂O₃ (óxido de aluminio)
- CaO (óxido de calcio)
- Na₂O (óxido de sodio)
- K₂O (óxido de potasio)
- TiO₂ (óxido de titanio)
- MnO (óxido de manganeso)

Elementos traza (en partes por millón - ppm):

- Cr (cromo)
- Ni (níquel)
- V (vanadio)
- Zr (circonio)
- Rb (rubidio)
- Sr (estroncio)

Estos elementos fueron seleccionados por su relevancia en estudios geoquímicos y su capacidad para proporcionar información sobre las condiciones de formación de los minerales y la evolución de los sistemas geológicos. Cada muestra fue etiquetada con su correspondiente unidad estratigráfica, permitiendo su uso en el modelo supervisado de clasificación.

3.3 Preprocesamiento de Datos

El preprocesamiento es una etapa crítica en cualquier proyecto de aprendizaje automático, ya que garantiza que los datos sean adecuados para el entrenamiento de modelos. Los siguientes pasos se realizaron antes del modelado:

a) Imputación de datos Faltantes:

Algunos valores de las variables geoquímicas estaban ausentes debido a la falta de mediciones completas en ciertas muestras. Se utilizó la imputación basada en la media para rellenar estos valores faltantes. Este método fue seleccionado tras un análisis de sensibilidad, que demostró que la imputación media no afectaba significativamente la calidad del modelo. Además, otras técnicas como la imputación por K-Nearest Neighbors (KNN) fueron evaluadas, pero no aportaron mejoras significativas en este caso.

b) Normalización de Datos:

Dado que las variables presentan escalas muy diferentes (por ejemplo, porcentajes en peso para los óxidos y ppm para

los elementos traza), se utilizó la normalización min-max. Esto permitió escalar todas las variables al rango [0, 1], lo que es esencial para garantizar que algoritmos como XGBoost no se vean afectados negativamente por magnitudes desbalanceadas. También se evaluó la estandarización Z-score, pero se observó que la normalización min-max mejoraba la convergencia del modelo.

3.4 Análisis de Componentes Principales (PCA)

Para reducir la dimensionalidad del conjunto de datos y resaltar las variables más relevantes, se implementó el Análisis de Componentes Principales (PCA). Este método permite transformar las características originales en componentes principales, los cuales explican la mayor parte de la varianza en los datos. El PCA se fundamenta en la descomposición en valores singulares (SVD), cuyo proceso matemático puede representarse de la siguiente manera:

$$Z = XW$$

Donde, X es la matriz de datos original (previamente normalizada), W son los vectores propios o componentes principales que maximizan la varianza en los datos y Z es la matriz de datos transformada en el nuevo espacio de componentes principales.

La selección de los componentes principales se realiza seleccionando aquellos con los valores propios más grandes, ya que estos capturan la mayor cantidad de varianza en los datos.

3.5 Clustering mediante K-Means

Se utilizó el algoritmo de K-Means clustering para agrupar las muestras en función de sus similitudes geoquímicas. Este algoritmo busca minimizar la distancia entre cada punto de datos y el centroide de su cluster. El número óptimo de clusters se determinó utilizando el método del codo (elbow method), resultando en la identificación de cuatro clusters. La función objetivo del algoritmo K-Means se puede describir matemáticamente como:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

El objetivo del algoritmo es minimizar J , la suma de las distancias al cuadrado entre los puntos de datos y sus centroides. Este procedimiento agrupó las muestras en clusters con composiciones químicas similares, permitiendo una interpretación más clara de las relaciones geoquímicas en el área de estudio.

3.6 Comparación con unidades estratigráficas

Las agrupaciones obtenidas mediante K-Means se compararon con las unidades estratigráficas. Se evaluó la concordancia entre los clusters geoquímicos y las divisiones geológicas mediante matrices de contingencia y análisis estadístico de coincidencias. Esta comparación permitió identificar patrones y discrepancias, proporcionando información útil para el refinamiento de la clasificación geológica del área de estudio.

3.7 Herramientas y Software Utilizados

El análisis de los datos se realizó utilizando herramientas de software ampliamente aceptadas en el ámbito de Machine Learning y análisis de datos científicos:

- **Python:** Lenguaje principal para el procesamiento de datos y análisis.
- **Bibliotecas específicas:** Se utilizaron bibliotecas como scikit-learn para la implementación de PCA y K-Means, pandas para la manipulación de los datos, y matplotlib para la visualización de los resultados.
- **Jupyter Notebooks:** Entorno de desarrollo interactivo utilizado para ejecutar y documentar todo el análisis.

Estas herramientas permitieron una integración eficiente de los datos, el preprocesamiento, el análisis multivariable y la generación de visualizaciones para apoyar las conclusiones del estudio.

4 RESULTADOS Y DISCUSIÓN

4.1 Reducción de Dimensionalidad mediante PCA

El Análisis de Componentes Principales (PCA) fue implementado para reducir la dimensionalidad del conjunto de datos, identificando combinaciones lineales que capturan la mayor parte de la varianza. Se seleccionaron los primeros dos componentes principales, explicando el 34.27% de la varianza total. Las principales contribuciones provinieron de Ti, Fe y TiO₂. La Fig. 1 muestra la distribución de los datos en el espacio de los dos primeros componentes principales obtenidos mediante el Análisis de Componentes Principales (PCA). El primer componente principal explica el 22.28% de la varianza total, mientras que el segundo componente explica el 11.99%. Juntos, estos componentes capturan una porción significativa de la variabilidad presente en los datos, lo que nos permite representar la estructura de los datos en dos dimensiones.

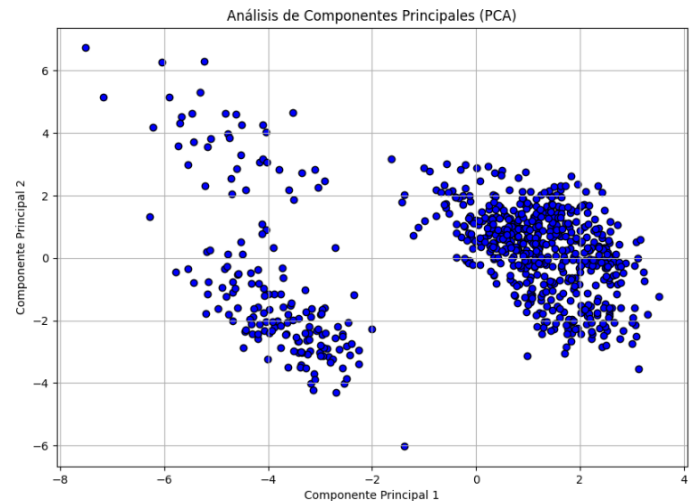


Fig. 1 Análisis de Componentes Principales

En la Tabla 1 se presentan las contribuciones más importantes de cada variable a los dos primeros componentes principales. Los elementos Ti, Fe, y TiO₂ tienen las mayores contribuciones al Componente Principal 1, mientras que MnO, FeO, y Cr contribuyen principalmente al Componente Principal 2.

TABLA 1

Contribución de Variables al Componente Principal 1 y 2

Variable	Componente Principal 1	Componente Principal 2
Ti	0.365	0.167
Fe	0.343	0.159
TiO ₂	0.245	0.038
MgO	0.220	-0.275
Total	0.164	-0.044
Longitude (WGS84)	0.160	0.136
Al ₂ O ₃	0.160	-0.115
Spot number	0.155	0.153
SiO ₂	0.134	-0.230
MnO	0.096	0.363
Cr ₂ O ₃	0.062	-0.108
CaO	0.053	-0.033
NiO	0.015	-0.067
Ni	0.012	-0.202
Mg	0.010	-0.108
Na ₂ O	0.010	-0.055
Crystal	-0.016	0.037
K ₂ O	-0.019	-0.075
Cr	-0.130	0.270
Latitude (WGS84)	-0.131	-0.130
Al	-0.187	-0.300
Ca	-0.214	0.249
FeO	-0.228	0.344
MnO	-0.230	0.363
K	-0.290	0.136
Si	-0.311	-0.312

4.2 Agrupamiento mediante K-Means

El algoritmo K-Means fue aplicado sobre los componentes principales obtenidos en la reducción de dimensionalidad con PCA. El objetivo era agrupar las muestras en clusters basados en su similitud geoquímica. En la Fig. 2, se observa el resultado del agrupamiento, con cuatro clusters identificados y distribuidos de acuerdo a los primeros dos componentes principales. Este análisis permitió identificar patrones de similitud entre las diferentes muestras en función de su composición geoquímica.

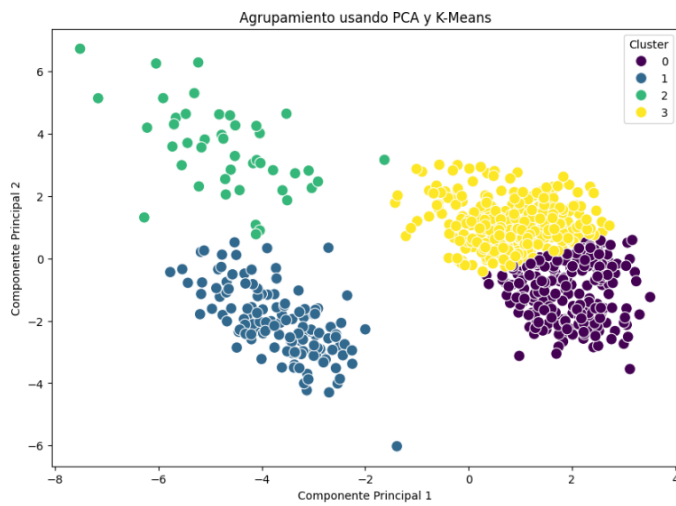


Fig. 2 Agrupamiento de Muestras Geoquímicas Usando K-Means basado en Componentes Principales

4.3 Composición Geoquímica de los Clusters

La Fig. 3 se muestra la distribución promedio de las características geoquímicas en cada uno de los clusters identificados. El cluster 0 se caracteriza por altos niveles de titanio (Ti) y hierro (Fe), lo cual sugiere procesos de cristalización fraccionada relacionados con magmas basálticos. Por otro lado, el cluster 2 exhibe concentraciones elevadas de manganeso (Mn) y cromo (Cr), lo que podría estar vinculado a una mayor presencia de minerales máficos. La variabilidad en las concentraciones de silicio (Si), magnesio (Mg), y aluminio (Al) entre los clusters refleja diferencias en los procesos de diferenciación magmática, mientras que la alta presencia de níquel (Ni) en el cluster 1 podría señalar una mayor influencia de fases metálicas o sulfídicas. Estos patrones sugieren una relación directa entre la composición geoquímica y los procesos volcánicos en la región estudiada.

Mary Luz Nina-Palacios, Ernesto Nayer Tumi-Figueroa y Hugo Ticona-Salluca

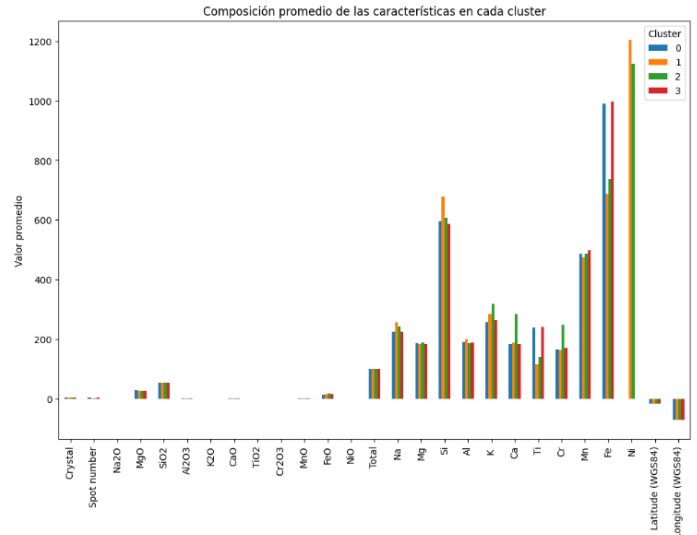


Fig. 3 Composición promedio de las características en cada cluster

4.4 Comparación con Unidades Estratigráficas

La comparación de los clusters con las unidades estratigráficas conocidas aporta una perspectiva más clara sobre cómo los diferentes tipos de muestras geológicas se relacionan con las capas estratigráficas. En la Fig. 4 resalta la relación cuantitativa entre los clusters y las unidades estratigráficas. Los colores más intensos indican una mayor correspondencia. Se observa que el **Cluster 3** tiene una fuerte presencia en las unidades *La Rosada* y *Sandwich Inferior*, mientras que el **Cluster 0** está bien representado en la unidad *Ponche Gris* y la unidad *Apo*. Esto sugiere que las características geoquímicas agrupadas en estos clusters tienen una fuerte correspondencia con las capas geológicas, lo cual puede ser útil para la interpretación geológica futura.

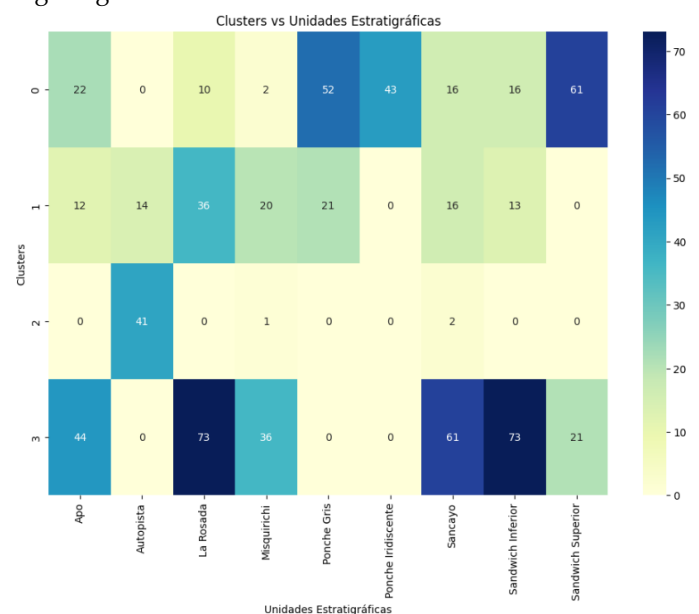


Fig. 4 Mapa de calor entre clusters y unidades estratigráficas

4.5 Discusión de los Resultados

El análisis PCA mostró que los dos primeros componentes principales explican un 34% de la variabilidad total de los datos, lo que, aunque representa una proporción significativa, también sugiere que existe variabilidad adicional que no se capta completamente en los primeros dos componentes. Este resultado refleja que, aunque algunas variables como Ti, Fe y TiO₂ contribuyen en gran medida a la diferenciación de las muestras, el sistema geológico es complejo y puede haber otras variables implicadas en el comportamiento geoquímico. Esta observación es consistente con estudios previos sobre la variabilidad geoquímica en volcanes similares, donde elementos traza y mayores suelen marcar diferencias entre episodios eruptivos y fuentes magmáticas [7].

En cuanto al agrupamiento K-Means, los resultados indicaron la formación de cuatro clusters claramente diferenciados, que parecen representar diferentes fases magmáticas o condiciones eruptivas. El agrupamiento revela una clara segregación entre las muestras, lo que respalda la idea de que las diferencias en la composición química son lo suficientemente significativas como para definir grupos distintos de muestras volcánicas. Este tipo de clasificación es crucial en estudios geológicos, ya que permite identificar diferentes fases eruptivas, magmas de diversas fuentes o variaciones en las condiciones de cristalización [8]. En comparación con investigaciones previas en otros volcanes peruanos, como el Misti o Sabancaya, los resultados obtenidos son similares, mostrando una estructura geológica heterogénea de episodios eruptivos a lo largo del tiempo [9].

La comparación de los clusters obtenidos mediante K-Means con las unidades estratigráficas muestra una correlación notable. Unidades como "Ponche Gris" y "La Rosada" presentaron una correspondencia significativa con clusters específicos, lo que sugiere que la evolución geoquímica refleja fielmente las unidades estratigráficas del volcán estudiado. Este resultado es crucial, ya que apoya la idea de que las diferencias observadas en la composición geoquímica pueden atribuirse a cambios temporales en la actividad volcánica y a la variabilidad en las fuentes magmáticas [10]. Esta correspondencia es un aporte importante a la tefroestratigrafía, proporcionando evidencia adicional de que la geoquímica puede ser utilizada no solo para diferenciar entre tipos de magma, sino también para mejorar la interpretación estratigráfica en volcanes con registros complejos.

5 CONCLUSIONES

El análisis geoquímico y estratigráfico realizado en este

estudio ha permitido establecer una metodología robusta para la caracterización de depósitos volcánicos mediante el uso de técnicas avanzadas como el Análisis de Componentes Principales (PCA) y el agrupamiento K-Means. La aplicación de PCA ha sido crucial para reducir la complejidad inherente a los datos de alta dimensionalidad, identificando las variables que tienen mayor influencia en la composición química de las muestras, particularmente Ti, Fe y MnO. Estas variables no solo han facilitado la diferenciación entre fases magmáticas, sino que también han ofrecido una comprensión más detallada de los procesos subyacentes a las erupciones volcánicas del Misti.

El agrupamiento K-Means, por otro lado, ha sido fundamental para identificar patrones de similitud entre las muestras, permitiendo su clasificación en clusters geológicamente coherentes. La diferenciación clara de los clusters sugiere que existen variaciones significativas en los procesos magmáticos y eruptivos, lo que aporta evidencia cuantitativa a hipótesis geológicas previas sobre la evolución volcánica del Misti. Este enfoque ha superado las limitaciones de los métodos tradicionales, donde la clasificación de las muestras dependía de observaciones cualitativas y subjetivas, ofreciendo en su lugar una herramienta más objetiva y reproducible para la caracterización de los depósitos volcánicos.

El análisis comparativo entre los clusters identificados y las unidades estratigráficas ha mostrado una correlación significativa, destacando particularmente las unidades Ponche Gris y La Rosada. Este hallazgo es un indicio claro de la relación entre la composición geoquímica de las muestras y las características estratigráficas de los depósitos. La combinación de estas técnicas ha permitido no solo una mejor clasificación de los datos geoquímicos, sino también una validación cruzada con las observaciones estratigráficas, lo que refuerza la fiabilidad de los resultados.

El uso de PCA y K-Means en este contexto no es solo una elección metodológica innovadora, sino también una contribución valiosa para el campo de la estratigrafía volcánica. Estas técnicas permiten una evaluación más precisa de la variabilidad geoquímica dentro de los depósitos volcánicos y proporcionan una plataforma para el desarrollo de estudios más detallados sobre la evolución magmática y los procesos eruptivos.

A diferencia de los métodos geológicos más tradicionales, la aplicación de algoritmos de aprendizaje no supervisado facilita la identificación de patrones ocultos en los datos, lo que tiene implicaciones significativas para futuros estudios. Esta investigación sienta las bases para la integración de técnicas de análisis computacional avanzado, abriendo la puerta a estudios que podrían incluir análisis isotópicos o el uso de técnicas de aprendizaje profundo para mejorar aún más la precisión en la

caracterización de muestras geológicas y predecir eventos eruptivos con mayor exactitud.

6 TRABAJOS FUTUROS

A partir de los resultados obtenidos en este estudio, se presentan varias oportunidades para profundizar en la investigación y ampliar el conocimiento sobre la evolución magmática y los procesos eruptivos del volcán Misti. Una de las principales direcciones futuras sería la integración de análisis isotópicos. Estudios previos han demostrado que el uso de isótopos de oxígeno y azufre puede proporcionar información crucial sobre las fuentes magmáticas y los mecanismos de diferenciación geoquímica en sistemas volcánicos complejos [14]. La combinación de estos análisis con los clusters geoquímicos identificados permitiría validar y complementar los resultados obtenidos mediante PCA y K-Means, proporcionando una visión más detallada de las características magmáticas y la historia eruptiva del volcán.

El uso de redes neuronales profundas y otros algoritmos de aprendizaje profundo ha sido cada vez más frecuente en el análisis de datos geológicos debido a su capacidad para identificar patrones complejos y realizar clasificaciones precisas basadas en grandes volúmenes de datos [15]. Estos modelos podrían aplicarse para mejorar la clasificación de las unidades geológicas basadas en composiciones geoquímicas y para predecir con mayor precisión la evolución de los sistemas magmáticos. Investigaciones como las de Caldeira y otros [16] han demostrado el potencial del aprendizaje profundo en la predicción de erupciones volcánicas, lo que sugiere que su aplicación en el volcán Misti podría generar resultados valiosos. Y por supuesto otra dirección interesante sería la inclusión de un mayor número de muestras en el análisis, provenientes de diferentes erupciones y fases eruptivas, para entender de manera más integral la evolución volcánica del Misti. Asimismo, la combinación de datos geoquímicos con información geofísica o geomorfológica podría ofrecer un enfoque multidisciplinario que mejore aún más la caracterización de los depósitos volcánicos y sus implicaciones para la estratigrafía volcánica regional.

REFERENCIAS

- [1] Rollinson, H. R. (2014). Using geochemical data: evaluation, presentation, interpretation. Routledge. <https://doi.org/10.4324/9781315845548>
- [2] Winter, J. D. (2010). Principles of Igneous and Metamorphic Petrology. Pearson Education.
- [3] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- [4] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [5] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1), 100-108. <https://doi.org/10.2307/2346830>
- [6] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [7] Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135-150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- [8] Feng, X., Min, X., Wang, J., & Guo, J. (2020). A comparison study of clustering methods for geo-chemical data analysis. *Geochemistry, Geophysics, Geosystems*, 21(7), e2020GC009274.
- [9] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [10] Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135-150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- [11] Feng, X., Min, X., Wang, J., & Guo, J. (2020). A comparison study of clustering methods for geo-chemical data analysis. *Geochemistry, Geophysics, Geosystems*, 21(7), e2020GC009274.
- [12] Rollinson, H. R. (2014). Using geochemical data: evaluation, presentation, interpretation. Routledge. <https://doi.org/10.4324/9781315845548>
- [13] Boggs, S. (2011). Principles of Sedimentology and Stratigraphy. Pearson Education.
- [14] Bindeman, I. N. (2008). Oxygen isotopes in mantle and crustal magmas as revealed by single crystal analysis. *Reviews in Mineralogy and Geochemistry*, 69(1), 445-478. <https://doi.org/10.2138/rmg.2008.69.12>
- [15] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- [16] Caldeira, K., Hoffmann, N., & Rudiger, M. (2020). Using deep learning for volcanic eruption prediction. *Geophysical Research Letters*, 47(3), e2019GL086650

BIOGRAFÍAS

Mary Luz Nina Palacios, estudiante de la Facultad de Ingeniería Estadística e Informática, Investigador en el Laboratorio Odin (Optimización, Desarrollo e Investigación) de la Universidad Nacional del Altiplano, Becaria Pronabec.

Ernesto Nayer Tumi Figueroa, Docente de la Universidad Nacional del Altiplano, director del Instituto de Investigación en Ciencias de la Computación, Investigador RENACYT.

Hugo Ticona Salluca, Bachiller en Ingeniería Estadística e Informática, desarrollador de Software en el Laboratorio Odin (Optimización, Desarrollo e Investigación) de la Universidad Nacional del Altiplano, Investigador RENACYT.

Evaluación de parámetros de calidad en diversas marcas comerciales de cloro doméstico

Evaluation of quality parameters in various commercial brands of domestic chlorine

Aliana Gamboa-Ortega ^A y José Castro-Soto ^B

Resumen— El cloro doméstico es una solución amarillenta de olor característico, ampliamente utilizada en los hogares, usualmente con una contracción entre 5% y 6% en hipoclorito de sodio (NaClO). Sin embargo, el NaClO es inestable y se degrada rápidamente. Se evaluó seis (6) marcas comerciales de cloro doméstico, con el objetivo de cuantificar los niveles de cloro disponible y su alcalinidad representada por la alcalinidad libre disponible y la alcalinidad total disponible, según la norma venezolana COVENIN 3664 (2001): Productos químicos para uso industrial. Hipocloritos utilizados en el tratamiento de potabilización de aguas. Requisitos y métodos de ensayo. Los resultados muestran que el contenido de cloro disponible osciló entre 0,48% y 3,19%, todas las muestras indican estar por debajo del valor recomendado, la alcalinidad libre disponible entre 0,19% y 1,63%, M₁, M₂, M₃, M₄ y M₅ demuestran estar en norma. Finalmente, la alcalinidad total disponible entre 0,34% y 37,65%, M₄ y M₅ presentan un mayor un efecto blanqueador en los tejidos, en comparación con las otras muestras estudiadas. En conclusión, se encontró que M₆ es el producto de mayor contenido de cloro disponible, asimismo la mayor alcalinidad libre disponible y alcalinidad total disponible.

Palabras clave: alcalinidad, cloro, evaluación, hipoclorito.

Abstract— Household bleach is a yellowish solution with a characteristic odor, widely used in homes, usually with a contraction between 5% and 6% in sodium hypochlorite (NaClO). However, NaClO is unstable and degrades rapidly. Six (6) commercial brands of domestic chlorine were evaluated, with the objective of quantifying the levels of available chlorine and its alkalinity represented by the free available alkalinity and the total available alkalinity, according to the Venezuelan standard COVENIN 3664 (2001): Chemical products for industrial use. Hypochlorites used in water purification treatment. Requirements and test methods. The results show that the available chlorine content ranged between 0.48% and 3.19%, all samples indicate to be below the recommended value, the free available alkalinity between 0.19% and 1.63%, M₁, M₂, M₃, M₄ and M₅ prove to be normal. Finally, the total alkalinity available between 0.34% and 37.65%, M₄ and M₅ present a greater whitening effect on fabrics, compared to the other samples studied. In conclusion, it was found that M₆ is the product with the highest available chlorine content, as well as the highest free available alkalinity and total available alkalinity.

Keywords: alkalinity, chlorine, evaluation, hypochlorite.

1 INTRODUCCIÓN

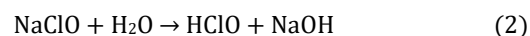
El cloro doméstico también llamado blanqueador, agua de cloro y agua Javelle [1], son soluciones comerciales transparente de color amarillo-verdoso y olor característico, es ampliamente conocido y usado en hogares como agente blanqueador, biocida y en la purificación del agua [1], [2], [3]. El ingrediente activo en el cloro doméstico es el hipoclorito de sodio (NaClO) [4], [5] y su poder desinfectante reside en el alto poder oxidante del ion hipoclorito (ClO⁻) [6],

La producción de NaClO consiste en la reacción del cloro gaseoso (Cl₂) y una solución acuosa de hidróxido de sodio (NaOH). Al electrolizar una solución concentrada de cloruro de sodio (NaCl), en el ánodo se produce el gas cloro (Cl₂) y en el cátodo se produce el gas hidrógeno (H₂) y el ión hidróxido (OH⁻). Con agitación, el Cl₂ y el OH⁻ reaccionan para producir el ión hipoclorito (ClO⁻) en solución, que generalmente está

en forma de la sal hipoclorito de sodio (NaClO) [7] (ver Ecuación 1). Obteniéndose una disolución al 11%, la cual se diluye a una concentración alrededor de 5% [2], [8], [9].



El NaClO reacciona con el agua para formar ácido hipocloroso (HClO) más hidróxido de sodio (NaOH) (ver Ecuación 2). Posteriormente, el HClO se descompone generando cloro gaseoso (Cl₂) y agua (H₂O) [6] (ver Ecuación 3).



Una de las características del cloro doméstico es la inestabilidad que posee en soluciones acuosas y el contenido de éste decrece rápidamente al ser sometida a cambios ambientales. El NaClO es muy reactivo [10], por lo que es poco estable y comienzan a descomponerse en cuanto se preparan y



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 17/10/24 - Indexaciones

Número: doi.org/10.57166/riqchary/V6.n2.2024

Pág.: 21- 26 | Recibido 04/09/2024 ; Aceptado 01/10/2024



doi.org/10.57166/riqchary.v6.n1.2024.125

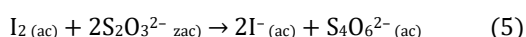
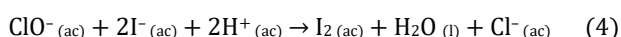
Autores:

- A. **ORCID iD** <https://orcid.org/0009-0003-5999-4441>
Aliana Gamboa-Ortega, está en el Centro de Estudios en Química, Universidad Politécnica Territorial del Estado Bolívar, Ciudad Bolívar, Venezuela, CO 8001. alia-nagamboaortega@gmail.com
- B. **ORCID iD** <https://orcid.org/0009-0009-5786-4743>
José Castro-Soto, está en el Centro de Estudios en Química, Universidad Politécnica Territorial del Estado Bolívar, Ciudad Bolívar, Venezuela, CO 8001. jlcas-tros78@gmail.com

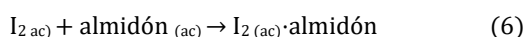
continúan hasta su descomposición total, su inestabilidad depende de cinco (5) factores principales, a) la concentración del hipoclorito, b) su alcalinidad o pH de la solución, c) la temperatura de almacenamiento, d) la presencia de impurezas que catalizan la descomposición, y e) la exposición a la luz solar fuerte; sin embargo, la principal causa tiene que ver con su incorrecto almacenamiento [11], [12], [13] y [14].

Con la finalidad de garantizar y preservar sus propiedades. [14]. Entre los requisitos químicos se encuentran: el cloro disponible (cloro activo), un término usado para expresar el poder oxidante del cloro en el producto. [1], como estos productos con fines domésticos contienen baja concentración de NaClO este se descomponen más lentamente que aquellas que contienen una concentración más elevada, asimismo un pH alcalino (pH≈11) que le da mayor estabilidad en el tiempo [12], y su alcalinidad representada por dos variables: la alcalinidad libre disponible y la alcalinidad total disponible, que brindan estabilidad y reducir la descomposición del NaClO y de mejorar el blanqueamiento, respectivamente. Es necesario destacar que el total de base disponible (como NaOH) no debe exceder de 1,5% [1], [7], [10] y [11].

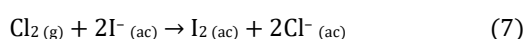
El método comúnmente empleado para evaluar el contenido de cloro disponible es la titulación yodométrica, esto es haciendo reaccionar un volumen conocido de cloro con un exceso de yoduro de potasio (KI) en medio ácido, en solución de ácido acético, el ClO^- forma yodo (I_2) de los yoduros (I^-). Evidenciándose un cambio de color de incolora a un tono marrón-rojizo, debido al I_2 [1], [7] (ver Ecuación 4). Seguidamente el I_2 formado se retrotitula con tiosulfato de sodio ($\text{Na}_2\text{S}_2\text{O}_3$) disminuyendo la concentración de I_2 , por lo que el color cambia a un amarillo claro al acercarse al punto estequiométrico [1], [7] (ver Ecuación 5).



La reacción se completa cuando el I_2 se convierte a I^- , tornándose incolora, pero este cambio de color de amarillo claro a incoloro no es claramente distinguible, se hace difícil establecer el punto final. En consecuencia, en este punto se añade como indicador una solución de almidón al 10%, el I_2 que queda sin reaccionar, formará el complejo I_2 -almidón de color azul oscuro [1], [7] (ver Ecuación 6).



Se continúa la titulación, el $\text{S}_2\text{O}_3^{2-}$ reacciona con el I_2 del complejo hasta que se consume, se rompe el complejo y desaparece el color azul oscuro, señalando claramente el punto final. [1], [7] (ver Ecuación 7).



La alcalinidad libre disponible, que constituye el contenido de hidróxido de sodio (NaOH), para ello se precipita todo el carbonato de sodio (Na_2CO_3) presente (ver Ecuación 8). El NaOH se neutraliza empleando como disolución titulante ácido clorhídrico (HCl) [1], [7] (ver Ecuación 9).



En cuanto a la alcalinidad total disponible, esta representa el contenido de hidróxido de sodio (NaOH) y carbonato de sodio (Na_2CO_3) presente, se neutraliza empleando como disolución titulante ácido clorhídrico (HCl) [1], [7] (ver Ecuación 9).



2 MATERIALES Y MÉTODOS

2.1 Recolección de la muestra

Para la selección de la muestra se realizó bajo un muestreo aleatorio simple. Se recolectaron al azar seis (6) marcas de cloro doméstico, no se consideró la presentación del producto en cuanto a volumen y envase. Las muestras se compraron en locales comerciales de Ciudad Bolívar: a) Super Único Panda C.A., b) Supermercado Latino, c) Supermarket El Económico y d) Mi Bodega Express, Market Place. Las muestras fueron conservadas en un lugar seco a temperatura ambiente. Los nombres comerciales (marcas) fueron sustituidos por un código conformado por una letra en mayúscula que lo identifica y un número M_1 , M_2 , M_3 , M_4 , M_5 y M_6 (ver Figura 1).

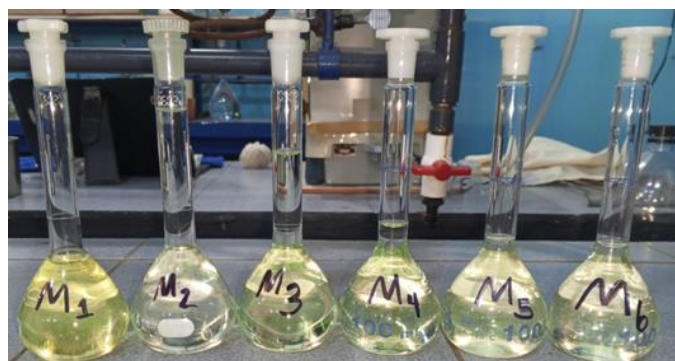


Fig. 1. Muestras de cloro doméstico empleadas en la investigación.

2.2 Determinación del cloro disponible

Se llenó una bureta de 50 mL con solución estándar de tiosulfato de sodio ($\text{Na}_2\text{S}_2\text{O}_3$) a una concentración de 0,1 M. Por otro lado, se tomó 5,0 mL de muestra (cloro doméstico), y se transfirió a un balón aforado de 250 mL, se aforó con agua destilada y se homogenizó. Seguidamente, se tomó una alícuota de 25 mL de la disolución anterior (muestra diluida) y se colocó en un matraz Erlenmeyer de 250 mL. Seguidamente

se adicionó al matraz 1,0 g de yoduro de potasio (KI), 10 mL de ácido acético glacial (CH_3COOH) diluido en una relación 1:2 y se homogenizó. Se inició la titulación añadiendo el titulante ($\text{Na}_2\text{S}_2\text{O}_3$) gota a gota, con agitación constante el matraz hasta que el color del yodo tome una coloración amarillo tenue. Se adicionó 2,0 mL de almidón al 10% y se continuó la titulación hasta que la solución se vuelva incolora. El procedimiento se realiza por triplicado (ver Figura 2).

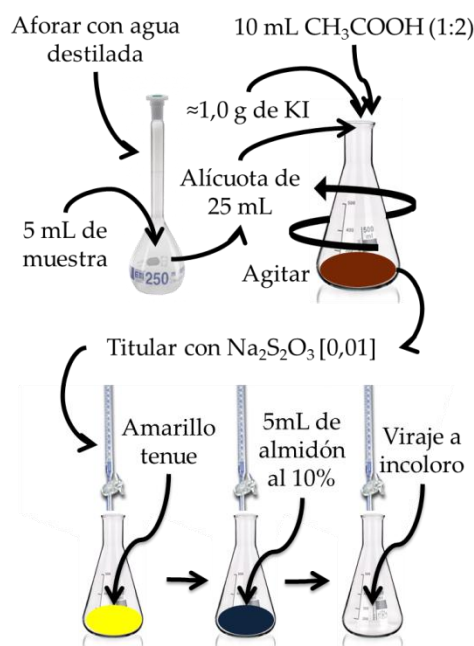


Fig. 2. Diagrama de flujo para la determinación de cloro disponible.

El cloro disponible se calcula como sigue:

$$\%Cl \text{ disponible} = \frac{V \times M \times 0,03545 \times 50}{\text{muestra (mL)}} \times 100 \quad (11)$$

Dónde: V = Volumen de tiosulfato de sodio ($\text{Na}_2\text{S}_2\text{O}_3$) consumido en la titulación de la muestra en mililitros.

M = Molaridad del tiosulfato de sodio ($\text{Na}_2\text{S}_2\text{O}_3$), expresado en equivalente/L.

50 = Factor de dilución $\frac{250}{5}$

0,03545 = Relación a la masa atómica del cloro.

2.3 Determinación de la alcalinidad libre disponible

En un matraz Erlenmeyer de 250 mL, se añadió 50 mL de cloruro bario (BaCl_2) 100g/L, 30 mL de peróxido de hidrógeno (H_2O_2) al 3% y 10 gotas del indicador fenolftaleína al 0,5%, seguidamente la solución se neutralizó con hidróxido de sodio (NaOH) 0,1 M. Por otro lado, en un segundo matraz Erlenmeyer de 250 mL, se pesó 10 g de muestra (cloro doméstico), sin diluir con precisión de 0,1 mg. El contenido del

primer matraz se vertió en el segundo y se homogenizó. Seguidamente, se tituló con ácido clorhídrico (HCl) estandarizado 0,1 M hasta que el color rosa del indicador desaparezca. El procedimiento se realiza por triplicado (ver Figura 3).

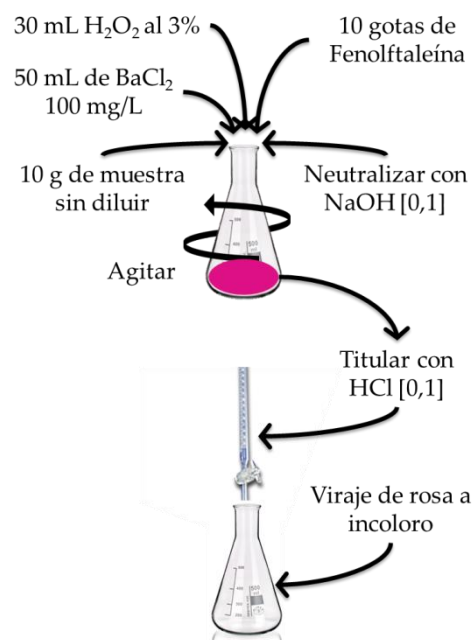


Fig. 3. Diagrama de flujo para la determinación de la alcalinidad libre disponible.

La alcalinidad libre disponible se calcula como sigue:

$$\text{Alcalinidad L.D.} = \frac{V \times M \times 0,04}{\text{muestra (g)}} \times 100 \quad (12)$$

Dónde: V = Volumen de ácido clorhídrico (HCl) consumido en la titulación de la muestra en mililitros.

M = Molaridad del ácido clorhídrico (HCl).

0,04 = Relación a la masa molecular del hidróxido de sodio (NaOH).

2.4 Determinación de la alcalinidad total disponible

Se pesó aproximadamente 10 g de muestra (cloro doméstico) sin diluir, con precisión de 0,1 mg en un matraz Erlenmeyer de 250 mL. Se adicionó solución de peróxido de hidrógeno (H_2O_2) al 3%, hasta el cese la reacción, esto es indicado por la finalización del burbujeo, posteriormente, se agitó vigorosamente durante un minuto. Se añadió cinco (5) gotas del indicador naranja de metilo. Finalmente, se tituló con solución de HCl , hasta el viraje del indicador desde el color amarillo a naranja. El procedimiento se realiza por triplicado (ver Figura 4).

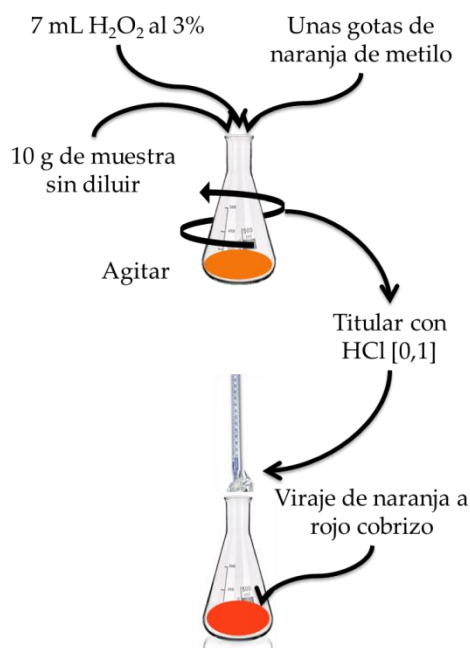


Fig. 4. Diagrama de flujo para la determinación de la alcalinidad total disponible

La alcalinidad total disponible se calcula como sigue:

$$\text{Alcalinidad L.D.} = \frac{V \times M \times 0,04}{\text{muestra (g)}} \times 100 \quad (10)$$

Dónde: V = Volumen de ácido clorhídrico (HCl) consumido en la titulación de la muestra en mililitros.

M = Molaridad del ácido clorhídrico (HCl).

0,04 = Relación a la masa molecular del hidróxido de sodio (NaOH).

3 RESULTADOS Y DISCUSIÓN

3.1 Determinación del cloro disponible

El valor más alto se ve representado por M₆ (3,19%), seguido de M₁ (1,84%). El resto de las muestras disminuyen su concentración, la menor concentración fue presentada por M₃ (0,48%) (ver Figura 5). Según [9] el valor recomendado de NaClO utilizado para la limpieza doméstica es de 5,25 %. En este sentido, todas las muestras indican estar por debajo del valor recomendado.

TABLA 1

Valores del contenido de cloro disponible en diversas marcas comerciales de cloro doméstico.

Muestra	($\bar{x} \pm s$) % _{p/v}
M ₁	(1,84 ± 0,28)
M ₂	(1,47 ± 0,04)
M ₃	(0,48 ± 0,04)
M ₄	(1,40 ± 0,01)
M ₅	(0,52 ± 0,04)
M ₆	(3,19 ± 0,15)

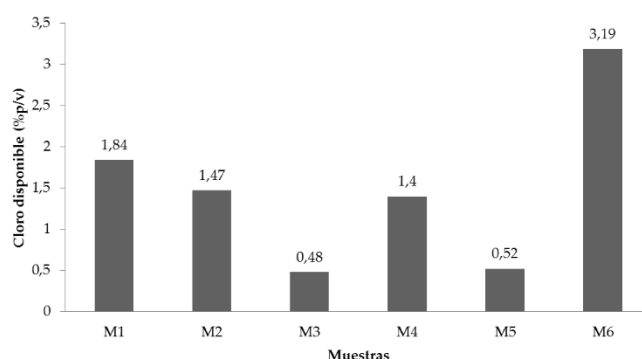


Fig. 5. Valores del contenido de cloro disponible en diversas marcas comerciales de cloro doméstico

3.2 Determinación de la alcalinidad libre disponible.

El valor más alto se ve representado por M₆ (1,63%), según [1] el total de base disponible (como NaOH) no debe exceder de 1,5%. Le siguen M₂ (0,83%) y el resto de las muestras que disminuyen su concentración hasta la menor concentración presentada por M₃ (0,19) (ver Figura 6). En este sentido, las muestras M₁, M₂, M₃, M₄ y M₅ demuestran estar en norma. Por otro lado, la M₆ presentó un valor ligeramente por encima del máximo permisible.

TABLA 2

Valores de la alcalinidad libre disponible en diversas marcas comerciales de cloro doméstico.

Muestra	($\bar{x} \pm s$) % _{p/p}
M ₁	(0,37 ± 0,06)
M ₂	(0,83 ± 0,05)
M ₃	(0,19 ± 0,06)
M ₄	(0,42 ± 0,06)
M ₅	(0,40 ± 0,09)
M ₆	(1,63 ± 0,04)

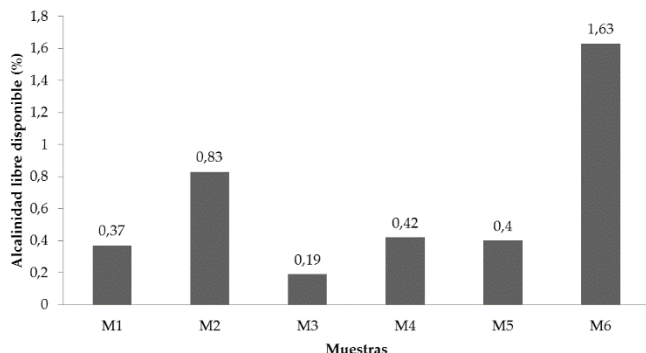


Fig. 6. Valores de la alcalinidad libre disponible en diversas marcas comerciales de cloro doméstico.

3.3 Determinación de la alcalinidad total disponible

El valor más alto se ve representado por M₆ (37,65%) seguido de M₂ (20,12%). Por otro lado, el resto de las muestras (M₁, M₃, M₄ y M₅) disminuyen su concentración considerablemente, la menor concentración es presentada por M₁ (0,34%) (ver Figura 7). Según [15] aquellos productos con mayor alcalinidad total libre disponible presentan un mayor efecto blanqueador en los tejidos. En este sentido, las muestras M₄ y M₅ presentan un mayor poder blanqueador.

TABLA 3

Valores de la alcalinidad total disponible en diversas marcas comerciales de cloro doméstico.

Muestra	$(\bar{x} \pm s) \%_{op/p}$
M ₁	(0,34 ± 0,01)
M ₂	(20,12 ± 0,19)
M ₃	(4,92 ± 0,13)
M ₄	(3,78 ± 0,18)
M ₅	(7,56 ± 0,21)
M ₆	(37,65 ± 0,10)

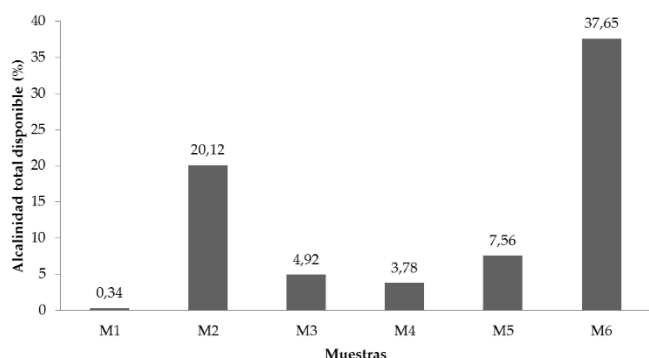


Fig. 7. Valores de la alcalinidad total disponible en diversas marcas comerciales de cloro doméstico.

4 CONCLUSIONES

Los resultados reportan que todas las muestras indican estar por debajo del valor recomendado en las concentraciones en cloro disponible. En cuanto a la alcalinidad libre disponible, las muestras M₁, M₂, M₃, M₄ y M₅ demuestran estar en norma. Finalmente, la alcalinidad total disponible las muestras M₄ y M₅ presentan un mayor efecto blanqueador en los tejidos, en comparación con las otras muestras.

5 REFERENCIAS

- [1] COVENIN 3664, «Productos químicos para uso industrial. Hipocloritos utilizados en el tratamiento de potabilización de aguas. Requisitos y métodos de ensayo.» 2001.
- [2] M. Gusukuma y I. Quispe, «Análisis de Ciclo de Vida del producto Lejía,» Red Peruana Ciclo de Vida y Ecología Industrial, 30 Mayo 2019. [En línea]. Available: https://www.oneplanetnetwork.org/sites/default/files/from-crm/192-6dae808bc2cb757c9006eea3aebd55e4_Actividad_9_ACV_Lejia.pdf. [Último acceso: 26 Julio 2024].
- [3] ATSDR, «ToxFAQs™ - Hipoclorito de calcio e hipoclorito de sodio (Calcium Hypochlorite/Sodium Hypochlorite): ¿Qué son el hipoclorito de sodio y de calcio?,» Agencia para Sustancias Tóxicas y el Registro de Enfermedades (ATSDR), 06 Mayo 2016. [En línea]. Available: https://www.atsdr.cdc.gov/es/toxfaqs/es_tfacts184.html. [Último acceso: 04 Agosto 2024].
- [4] M. Gagliardi, «Qué es el blanqueador y cuáles son sus ingredientes activos,» Clorox, 2024. [En línea]. Available: <https://www.clorox.com/es/aprender/que-es-el-blanqueador-cuales-son-sus-ingredientes-activos/>. [Último acceso: 18 Septiembre 2024].
- [5] L. Muñoz, A. Borrego-Loya, C. Villalba, R. González-Escobedo, N. Orduño, G. Villezas-Villegas, M. Rodríguez-Roque, G. Ávila-Quezada y I. Vargas, «El cloro y su importancia en la inactivación de bacterias, ¿Puede inactivar virus?,» 2021. [En línea]. Available: <https://www.scielo.org.mx/pdf/rmfi/v39nspe/2007-8080-rmfi-39-esp-198.pdf>. [Último acceso: 18 Septiembre 2024].
- [6] G. de la Rosa y S. Figueroa, «El hipoclorito y el cloro: dos tipos de cuidado,» 11 Julio 2020. [En línea]. Available: <https://www.ugto.mx/investigacionposgrado/eugreka/index.php/contribuciones/262-el-hipoclorito-y-el-cloro-dos-tipos-de-cuidado>. [Último acceso: 04 Agosto 2024].
- [7] S. Delgado, L. Solís y Y. Muñoz, Laboratorio de Química General, México, D.F.: McGraw-Hill Interamericana Editores, S.A. de C.V., 2012.
- [8] Unipar, «unipar,» 2023. [En línea]. Available: <https://unipar.com/es/hipoclorito-de-sodio-es/#:~:text=Aplicaciones-,El%20Hipoclorito%20de%20Sodio%20tiene%20propiedades%20oxidantes%2C%20blanqueadoras%20y%20desinfectantes,de%20agua%20potable%20y%20hospitalaria..> [Último acceso: 26 04 2024].
- [9] Amoquimicos, «Amoquimicos,» 17 Marzo 2023. [En línea]. Available: <https://www.amoquimicos.com/noticias/niveles->

de-concentracion-del-hipoclorito-de-sodio. [Último acceso: 30 Mayo 2024].

- [10] Lenntech, «Desinfectantes Hipoclorito de sodio,» [https://www.products.pcc.eu/es/blog/carbonato-de-sodio-un-componente-de-los-agentes-de-lavado-ecologicos/](https://www.lenntech.es/procesos/desinfeccion/quimica/desinfectantes-hipoclorito-de-sodio.htm#:~:text=Alrededor%20del%20a%C3%B1o%201785%20el,especificas%20se%20extendi%C3%B3%20con%20facilidad., South Miami, 2023.[11] G. Peña, J. Caram, A. Rodríguez, A. Anselmi, . M. Gimena, M. S. Barrera y . J. González, «Determinación de cloro activo en diferentes soluciones de hipoclorito de sodio,» <i>UNCuyo</i>, vol. XI, n° 1, p. 21.25, 2017.[12] GBCB, «Experimental 1. Determinación de “cloro activo” en una muestra de agua lavandina mediante titulación yodométrica,» Santa Fe, 2017.[13] A. Londoño, G. Giraldo y Á. Gutiérrez, Métodos analíticos para la evaluación de la calidad físicoquímica del agua, Primera ed., Manizales, Manizales: Blanecolor Ltda, 2010.[14] M. J. Alarcón Lema, «Variación de la concentración del hipoclorito de sodio por cambio de almacenamiento, 2018,» Riobamba – Ecuador, 2019.[15] Grupo PCC, «Group PCP. Fabricante de productos quimicos,» 06 10 2021. [En línea]. Available: <a href=). [Último acceso: 29 04 2024].

6 BIOGRAFÍAS

Aliana Gamboa Ortega. Técnico Superior Universitario en Química en la Universidad Politécnica Territorial del Estado Bolívar (UPTBolívar), actualmente participante del Programa Nacional de Formación en Química (PNFQuímica) en Licenciatura en Química.

José Castro Soto. Licenciado en Química egresado de la Universidad De Oriente (Núcleo de Sucre). Ha sido jefe de la División de Laboratorio (Gerencia de Control de Calidad – HIDROBOLÍVAR C.A). Con más de 15 años de experiencia en educación universitaria en el Instituto Universitario de Tecnología del Mar de Fundación “La Salle” de Ciencias Naturales (IUTEMAR-FLACN) y en la Universidad Politécnica Territorial del Estado Bolívar (UPTBolívar). Actualmente, se desempeña como jefe del Departamento del Programa Nacional de Formación en Química (PNFQuímica) en la UPTBolívar.

Efecto de dos tipos de sustrato en almacigo más hidroponía NFT en las características morfológicas de tres variedades de lechuga (*Lactuca sativa* L.) en el INIA Andahuaylas Perú-2024

Effect of two types of substrate in seedbed plus NFT hydroponics on the morphological characteristics of three varieties of lettuce (*Lactuca sativa* L.) at the INIA Andahuaylas Perú-2024

Reyna Vargas-Peña ^A, Niki Franklin Flores-Pacheco ^B

Resumen—La investigación tuvo como objetivo Evaluar el efecto de dos tipos de sustrato en almácigo más hidroponía NFT en las características morfológicas de tres variedades de lechuga (*Lactuca sativa* L.) en INIA Andahuaylas Perú, el estudio se realizó en la provincia de Andahuaylas, departamento Apurímac. Las variables de estudio fueron: longitud de raíz, altura de planta, diámetro de la cabeza, N° de hojas, con un diseño de bloques completamente al azar. Donde los resultados han evidenciado diferencias significativas (sig. <0.05) reportando una mayor longitud radicular el T3=(var. americana + arena+ HNFT) con 29.86 cm, en cuanto a la altura de planta T4= (var. americana +sustrato tradicional + H-NFT) con 32.37 cm, Diámetro de la cabeza (cm) T1= (var. seda +arena + H-NFT) con 34.30 cm y Número de hojas (ud.) T1= (var. seda +arena + H-NFT) con 24.93 cm. Se concluye que la variedad americana se adapta mejor a esas condiciones.

Palabras clave: Almácigos, Hidroponía, lechuga

The objective of the research was to evaluate the effect of two types of substrate in a seedbed plus NFT hydroponics on the morphological characteristics of three varieties of lettuce (*Lactuca sativa* L.) in INIA Andahuaylas Peru, the study was carried out in the province of Andahuaylas, Apurímac department. . The study variables were: root length, plant height, head diameter, number of leaves, with a completely randomized block design. Where the results have shown significant differences (sig. <0.05) reporting a greater root length T3=(American var. + sand+ HNFT) with 29.86 cm, in terms of plant height T4= (American var. +traditional substrate + H-NFT) with 32.37 cm, . Head diameter (cm) T1= (var. silk +sand + H-NFT) with 34.30 cm and Number of leaves (unit) T1= (var. silk +sand + H-NFT) with 24.93 cm. It is concluded that the American variety is better adapted to these conditions.

Keywords: Seedbeds, Hydroponics, Lettuce

1 INTRODUCCIÓN

Los cultivos hidropónicos se han considerado una progresión tecnológica en el ámbito de la agricultura. Sin embargo, esta metodología se ha empleado desde la antigüedad, ya que naciones como India, China y algunos países estadounidenses han optado por la hidroponía como sustituto de las limitaciones impuestas por el suelo y el agua [1] La lechuga es un producto agrícola que se puede cultivar mediante varios métodos. Entre estos métodos, la producción al aire libre es la más frecuente en el país. El proceso de cultivo comienza con el inicio de un sistema de almacenamiento de agua, mientras que el suministro de tierra fértil y abundante agua son imprescindibles para un crecimiento exitoso. Cabe destacar que la temperatura desempeña un papel clave a la hora de facilitar el desarrollo de este tipo específico de verdura. En particular, la lechuga prospera en regiones caracterizadas por temperaturas que oscilan entre 18 y 25 °C. Cabe mencionar que los avances contemporáneos han llevado a la disponibilidad de variedades de lechuga aptas para el cultivo

de invierno [2]. (Vargas) [3], Afirma que la producción hidropónica no solo tiene la capacidad de aliviar los problemas relacionados con la contaminación del suelo y la escasez de agua en ciertas regiones, sino también de implementarse en entornos residenciales.

1.1. Cultivo de hortalizas mediante la técnica de NFT (Sistema Nutrient Film Techniquel)

El sistema NFT de película nutriente ofrece múltiples ventajas para el cultivo de hortalizas. En este caso particular, una película delgada de agua nutritiva fluye a través de los canales, lo que facilita la absorción de sustancias nutritivas por parte de las raíces de las plantas. Es imprescindible supervisar continuamente este sistema, ya que la regulación de la conductividad eléctrica y el caudal es esencial para evitar la saturación de agua de las bombonas de PVC [1]

1.2. Ventajas

De acuerdo con (Vargas) [3], el sistema Nutrient Film Techniquel, Se logra una mayor productividad en comparación con los métodos de siembra tradicionales, ya que permite



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 04/11/24 - [Indexaciones](#)

Número: doi.org/10.57166/riqchary/V6.n2.2024

Pág. 27- 34 | Recibido 10/08/2024 ; Aceptado 10/21/2024



doi.org/10.57166/riqchary.v6.n1.2024.125

Autores:

A. **ORCID iD** <https://orcid.org/0009-0002-5638-3920>

Reyna Vargas-Peña, Universidad Nacional Micaela Bastidas de Apurímac, 181677@unamba.edu.pe

B. **ORCID iD** <https://orcid.org/0000-0002-2772-0951>

Niki Franklin Flores-Pacheco, Universidad Nacional Micaela Bastidas de Apurímac, nflores@unamba.edu.pe

una mayor densidad de plantas por metro cuadrado y un menor tiempo de cosecha.

Este enfoque también facilita una importante conservación del agua al utilizar una cantidad menor de agua en comparación con los cultivos en tierra.

El cultivo hidropónico evita la infiltración de agua y una parte importante de la evapotranspiración debido a la naturaleza cerrada del cultivo, que mantienen una humedad relativa alta. Este método demuestra ser rentable y eficiente en términos de productividad. Al reducir las enfermedades de los cultivos, también minimiza la necesidad de aplicar agroquímico. En última instancia, el cultivo hidropónico permite a los productores obtener cultivos sanos y de alta calidad aptos para el consumo.

1.3. Desventajas

(Cedeño & Gonzales) [7][8], expresan que el principal inconveniente de implementar un sistema de cultivo hidropónico es la inversión inicial. Al igual que cualquier esfuerzo tecnológico, el establecimiento de un sistema hidropónico requiere la adquisición de diversos materiales y equipos, incluidas bombas de oxigenación y recirculación, contenedores, sustratos, tuberías, filtros de agua, grifos y otros. Sin embargo, es crucial tener en cuenta que, una vez que se establezca el sistema, permanecerá operativo durante varios años. Es importante reconocer que la hidroponía es una tecnología agrícola que exige una atención meticulosa a numerosas complejidades, que a menudo se minimizan en la agricultura tradicional.

Es indispensable el pH en la solución nutritiva la medida es de 6.63, se utilizó el pH metro digital lo cual estaba graduado bajo la escala de 1-14 (Acido, Neutro y base) con un valor neutro de pH 7 lo que significa una sustancia no acida ni alcalina

La conductividad eléctrica (E.C) es una medida del total de la cantidad de sales diluidas en el agua que tuvo un rango de medida de 3.56 MS todo ello nos ayuda a valorar la capacidad del agua para poder transportar la corriente eléctrica.

El caudal máximo 35 por minuto.[4][5] y Motor hp de 0.5.

1.4. Soluciones nutritivas

Son compuestos esenciales para el desarrollo de las lechugas, comprende de macro y micro nutrientes. agua que contiene oxígeno disuelto y minerales no asociados. Estas soluciones proporcionan a los cultivos los nutrientes necesarios y los niveles de agua apropiados para su crecimiento y desarrollo específicos. La solución nutritiva en sí misma consiste en una combinación de sales minerales mezcladas con agua, y su composición se puede ajustar de acuerdo con los requisitos específicos de los diferentes cultivos. Es esencial que el pH de la solución esté entre 5,5 y 6 para garantizar la utilización óptima de los nutrientes [6]. Las lechugas tienen una buena respuesta a las soluciones nutritivas debido a la morfología del sistema radicular por lo que la novedad del estudio es la adaptación. En la investigación se evaluó diferentes variedades de lechuga (*Lactuca Sativa* L) en el cambio de sustrato tradicional a producción hidropónica se evaluó las características formológicas como: longitud de raíz, altura de planta, diámetro de la cabeza, N° de hojas, cada una de estas variables se evaluó cada 15 días hasta cumplir 60 días según su fase fenológica.

Se utilizó el análisis de varianza (ANOVA) y la prueba de Tukey al 95%

2 MÉTODO

La metodología empleada fue de naturaleza experimental, ya que el fenómeno hipotetizado se observó sistemáticamente; se manipularon las variables dependientes asociadas a tres variedades distintas de lechuga (*Lactuca Sativa* L utilizando dos tipos de sustrato diferentes (sustrato tradicional = tierra negra, compost y arena), (sustrato = Arena)

2.1 Tipo y nivel de investigación

El tipo de investigación debido a su alcance es temporal, se caracteriza por ser transversal (dentro de un marco temporal agrícola). A la luz de las características de los datos, se clasifican como cuantitativos (las variables evaluadas son de naturaleza cuantitativa). El nivel de investigación: es de naturaleza descriptiva (los hallazgos se articulan utilizando métodos estadísticos descriptivos). También es experimental (las variables, que comprenden tres variedades distintas de lechuga y dos tipos de sustratos, se manipularon sistemáticamente para evaluar las características morfológicas de la lechuga) conlleva longitud de raíz, altura de planta, diámetro de la cabeza y número de hojas.

2.2 Diseño de investigación

El marco metodológico empleado se refiere a un estudio experimental, en el que se proporcionaron respuestas a las preguntas planteadas durante la evaluación y la corroboración de la hipótesis propuesta. Se tuvieron en cuenta los principios de la aleatorización, la repetición y el control local en el contexto de un diseño de bloques completamente aleatorizado (DBCA).

2.3 Población y muestra

La población lo conforman las plantas de las tres variedades (seda, americana y crespa) de lechuga (*Lactuca sativa* Se tuvo un total de 200 plantas, El tamaño de la muestra se determina mediante la aplicación del enfoque probabilístico con un nivel de confianza del 95%, lo que resulta en la identificación de 132 unidades de planta designadas para la evaluación.

2.4 Análisis estadístico

Los hallazgos del estudio se recopilaron dentro de los instrumentos de evaluación. Para el análisis de los datos, se siguieron las siguientes metodologías: evaluación de la normalidad, evaluación de la homogeneidad de las varianzas, descripción estadística, aplicación del ANOVA, ejecución de la prueba de Tukey, empleo del software Excel y, a continuación, análisis de datos para abordar los objetivos e hipótesis de la investigación utilizando formas tabulares de Excel y SPSS.

2.4.1 Prueba de Normalidad

Para determinar si empleamos métodos estadísticos paramétricos o no paramétricos, se utilizó el estadístico de Shapiro-Wilk para la evaluación de la normalidad

Tabla 1
Prueba de normalidad

Variable	Significancia de los tratamientos al 95%
----------	--

	T1=Var seda +arena +H- NFT	T2=Var seda + Sust Trad+ H-NFT	T3=Var ameri- cana + arena + H- NFT	T4=Var ameri- cana +Sust Trad+ H-NFT	T5=Var crespa + arena+ H- NFT	T6=Var crespa +Sust Trad+ H- NFT
Longitud raíz 1 Evaluación	0.294	0.532	0.659	0.488	0.446	0.823
Longitud raíz 2 Evaluación	0.453	0.976	0.919	0.946	0.787	0.949
Longitud raíz 3 Evaluación	0.230	0.304	0.464	0.339	0.262	0.900
Longitud raíz 4 Evaluación	0.455	0.398	0.987	0.684	0.837	0.970
Altura de planta 1 Evaluación	0.251	0.894	0.992	0.297	0.756	0.147
Altura de planta 2 Evaluación	0.700	0.234	0.886	0.962	0.590	0.126
Altura de planta 3 Evaluación	0.394	0.187	0.931	0.917	0.899	0.992
Altura de planta 4 Evaluación	0.410	0.548	0.436	0.786	0.981	0.355
Diámetro de la cabeza	0.653	0.415	0.871	0.930	0.639	0.405
Nº de hojas 1 Evaluación	0.272	0.683	0.406	0.406	0.850	0.972
Nº de hojas 2 Evaluación	0.650	0.195	0.972	0.224	0.972	0.972
Nº de hojas 3 Evaluación	0.683	0.972	0.972	0.972	0.972	0.103
Nº de hojas 4 Evaluación	0.850	0.195	0.406	0.972	0.850	0.406

En la Tabla 1, se presentan los resultados de la evaluación de la normalidad y los niveles de significación de las variables y los tratamientos se evaluaron con un umbral de confianza del 95%. Los valores obtenidos superaron el 0,05 (Sig.>0,05), por lo que no se rechazó la hipótesis nula (Ho) y, en última instancia, se dedujo que el conjunto de datos proviene de una población con distribución normal.

2.4.2 Prueba de homogeneidad de varianza

Para evaluar la uniformidad de la varianza entre las variables, se empleó el estadístico de Levene.

Tabla 2
Prueba de homogeneidad de varianza

Variable	Estadístico de Levene	gl1	gl2	Sig.
Longitud raíz 1 Evaluación	0.228	5	18	0.946
Longitud raíz 2 Evaluación	0.138	5	18	0.981
Longitud raíz 3 Evaluación	0.894	5	18	0.506
Longitud raíz 4 Evaluación	0.423	5	18	0.826
Altura de planta 1 Evaluación	2.138	5	18	0.107
Altura de planta 2 Evaluación	0.431	5	18	0.821
Altura de planta 3 Evaluación	0.713	5	18	0.622
Altura de planta 4 Evaluación	1.628	5	18	0.203
Diámetro de la cabeza	1.697	5	18	0.186
Nº de hojas 1 Evaluación	0.537	5	18	0.746
Nº de hojas 2 Evaluación	1.937	5	18	0.138
Nº de hojas 3 Evaluación	0.385	5	18	0.853
Nº de hojas 4 Evaluación	0.210	5	18	0.954

Hipótesis de homogeneidad:

Observando la tabla 2, tenemos que el valor de la significancia es mayor a 0.05 (Sig.>0.05), por lo que no se rechazó la hipótesis nula y se aceptó que las variables son homogéneas

3 RESULTADOS

Las características morfológicas de tres variedades de lechuga (*Lactuca sativa* L)

Se realizó las evaluaciones del número de hojas, altura de la

planta, diámetro de la cabeza y longitud radicular en la estación experimental INIA Chumbibamba, Andahuaylas Perú.

3.1 Longitud de raíz

La evaluación de campo de la longitud de raíz de tres variedades distintas de lechuga (*Lactuca sativa* L) se llevó a cabo 60 días después de la siembra, mediante la hidroponía, Técnica de Película Nutriente (NFT).

Tabla 3
Estadístico descriptivo de la longitud de la raíz (cm)

Tratamiento	Longitud raíz (cm) cuarta evaluación			
	Media	Máximo	Mínimo	Desviación estándar
T1-Var seda +arena + H-NFT	26.17	26.63	25.68	0.45
T2-Var seda + Sust Trad+ H- NFT	25.80	26.87	25.10	0.83
T3-Var americana + arena + H- NFT	29.86	30.75	28.90	0.77
T4-Var americana +Sust Trad+ H- NFT	26.61	27.35	25.90	0.64
T5-Var crespa + arena+ H-NFT	23.61	24.36	22.90	0.65
T6-Var crespa +Sust Trad+ H- NFT	21.54	22.25	20.75	0.64

Los resultados de la tabla 3 muestran la longitud de la raíz de tres variedades de lechuga (*Lactuca sativa* L.) donde el T3= (var. americana + arena+ H-NFT) con 29.86 cm, es superior a los otros tratamientos, seguido del T4= (var. americana +sustrato tradicional + H-NFT) con 26.61 cm; seguido del T1= (var. seda +arena + H-NFT) con 26.17 cm $\pm$ 0.45; seguido del T2= var seda + sustrato tradicional + H- NFT) con 25.80 cm; seguido del T5= (var. crespa + arena+ H-NFT) con 23.61 cm y finalmente el T6= (var. crespa +sustrato tradicional + H-NFT) con 21.54 cm.

3.1.1 Prueba de (ANOVA)

Con un nivel de confianza del 95%, los hallazgos se detallan en la tabla siguiente:

Tabla 4
Análisis de varianza longitud de raíz- cuarta evaluación

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	15889.800	9	1765.533	4766.476	0.000
Tratamiento	160.025	5	32.005	86.405	0.000
Bloque	2.670	3	0.890	2.403	0.108
Error	5.556	15	0.370		
Total	15895.356	24			

Modelo. En la tabla 4, es evidente que el valor p (sig = 0,000) está por debajo del umbral de significancia (< 0,05), por lo que es necesario rechazar la hipótesis nula. En consecuencia, se deduce que el modelo general presenta linealidad y que se cumplen los supuestos necesarios para un diseño de bloques completos aleatorios (DBCA).

Tratamientos. En la tabla 4, se observa que el valor p (sig = 0.000) cae por debajo del nivel de significancia (< 0.05), lo que resulta en el rechazo de la hipótesis nula. Por lo tanto, en la cuarta evaluación de tres variedades de lechuga (*Lactuca sativa* L), a saber, la seda, la americana y la crespa, se concluye que los dos tipos de sustrato utilizados en los cultivos hidroponicos con sirope y NFT tienen un efecto significativo sobre la longitud de las raíces.

Bloques. En la tabla 4, se observa que el valor p (sig = 0.108) supera el nivel de significancia (> 0.05), lo que lleva a la decisión de mantener la hipótesis nula (H_0). Por lo tanto, se afirma que los bloques no influyen en la longitud de las raíces durante la cuarta evaluación de las tres variedades de lechuga (*Lactuca sativa* L.), específicamente la seda, la americana y la crespa.

Para determinar cuál de los tratamientos: T1= (var. seda + arena + H-NFT), T2= var seda + sustrato tradicional + H- NFT), T3= (var. americana + arena+ H-NFT) T4= (var. americana +sustrato tradicional + H-NFT), T5= (var. crespa + arena+ H-NFT) y T6= (var. crespa +sustrato tradicional + H-NFT), producen los efectos más pronunciados en la longitud de las raíces. Durante la cuarta evaluación de las tres variedades de lechuga (*Lactuca sativa* L.), a saber, seda, americana y crespa. Se realizó una comparación múltiple de medias empleando la prueba de Tukey con un nivel de probabilidad del 95%, como se ilustra en la siguiente tabla.

3.1.2 Prueba de Tukey

Al 95%, los resultados se expresan en la siguiente tabla:

Tabla 5

Prueba de tukey, longitud de raíz – cuarta evaluación

Tratamiento	N	Subconjunto			
		1	2	3	4
T6=Var crespa +Sust Trad+ H-NFT	4	21.5400			
T5= Var crespa + arena+ H-NFT	4	23.6075			
T2= Var seda + Sust Trad+ H- NFT	4	25.8025			
T1= Var seda +arena + H-NFT	4	26.1725			
T4= Var americana +Sust Trad+ H- NFT	4	26.6075			
T3= Var americana + arena + H- NFT	4	29.8625			
Sig.		1.000	1.000	0.455	1.000

En la tabla 5, se observa en el primer sub conjunto homogéneo de las medias al

T6=Var crespa +Sust Trad+ H-NFT con 21.54 cm; en el segundo sub conjunto el T5= Var crespa + arena+ H-NFT con 23.60 cm; en el tercer sub conjunto el T2= Var seda + Sust Trad+ H- NFT con 25.80 cm, el T1= Var seda +arena + H-NFT con 26.17 cm y el T4= Var americana +Sust Trad+ H-NFT con 26.60 cm y en el cuarto sub conjunto, con un mayor promedio el T3= Var americana + arena + H-NFT, con 29.86 cm.

3.2 Altura de planta

La evaluación de campo de la altura de planta de tres variedades distintas de lechuga (*Lactuca sativa* L.) se llevó a cabo 60 días después de la siembra, mediante la hidroponía, Técnica de Película Nutriente (NFT)

Tabla 6

Estadístico descriptivo de la altura de planta (cm)

Tratamiento	Altura de planta (cm) cuarta evaluación			
	Media	Máximo	Mínimo	Desviación estándar

T1-Var seda +arena + H-NFT	21.93	22.11	21.81	0.14
T2-Var seda + Sust Trad+ H- NFT	22.60	22.70	22.52	0.07
T3-Var americana + arena + H- NFT	30.12	30.62	29.80	0.35
T4-Var americana +Sust Trad+ H- NFT	32.37	32.80	31.80	0.42
T5-Var crespa + arena+ H-NFT	22.72	23.04	22.43	0.25
T6-Var crespa +Sust Trad+ H- NFT	24.08	24.18	23.93	0.12

Los hallazgos indican la altura de planta de tres variedades distintas de lechuga (*Lactuca sativa* L.), en las que el T4= (var. americana + sustrato tradicional + H-NFT) mide 32,37 cm, superando a todos los demás tratamientos experimentales. Le sucede T3= (var. americana + arena + H-NFT) con 30,12 cm; seguido de T6= (var. crespa + sustrato tradicional + H-NFT) con 24,08 cm; posteriormente, T5= (var. crespa + arena + H-NFT) con 22,72 cm; seguido de T2= (var. americana + sustrato tradicional + H-NFT) con 22,60 cm y, finalmente, T1= (var. seda + arena + H-NFT) a 21,60 cm.

Tabla 7

Análisis de varianza altura de planta – cuarta evaluación

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	16169.513	9	1796.613	22522.487	0.000
Tratamiento	397.302	5	79.460	996.123	0.000
Bloque	0.020	3	0.007	0.085	0.967
Error	1.197	15	0.080		
Total	16170.709	24			

Modelo. En la tabla 7, es evidente que el valor p (sig = 0.000) es inferior al nivel de significancia predeterminado (< 0.05), por lo que es necesario rechazar la hipótesis nula. En consecuencia, comprobamos que el modelo general posee características lineales, por lo que cumple con los requisitos previos para un diseño de bloques completos aleatorios (DBCA).

Tratamientos. En la tabla 7, se observa que el valor p (sig = 0.000) es inferior al nivel de significancia establecido (< 0.05), lo que lleva al rechazo de la hipótesis nula. Por lo tanto, se concluye que existe un efecto significativo de los dos tipos de sustratos NFT hidropónicos y del suelo sobre la altura de las plantas, específicamente durante la cuarta evaluación en tres variedades de lechuga (*Lactuca sativa* L.): seda, americana y crespa.

Bloques. En la tabla 7, se observa que el valor p (sig = 0.967) supera el nivel de significancia (> 0.05), lo que nos guía hacia la decisión de aceptar la hipótesis nula (H_0). Por lo tanto, afirmamos que los bloques no influyen en la altura de las plantas durante la cuarta evaluación de las tres variedades de lechuga (*Lactuca sativa* L.): sedosa, americana y crespa.

Para determinar cuál de los tratamientos: T1= (var. seda + arena + H-NFT), T2= (var. seda + sustrato tradicional + H-NFT), T3= (var. americana + arena + H-NFT), T4= (var. americana + sustrato tradicional + H-NFT), T5= (var. crespa + arena + H-NFT) y T6= (var. crespa + el sustrato tradicional + H-NFT), ejercen un impacto más pronunciado en la altura de la planta. Durante la cuarta evaluación de las tres variedades de lechuga (*Lactuca sativa* L.): seda, americana y crespa. Se realizó una comparación

múltiple de medias utilizando la prueba de Tukey con un nivel de confianza del 95%, los resultados de que se presentan en la siguiente tabla.

Tabla 8

Prueba de tukey, altura de planta – cuarta evaluación

Tratamiento	N	Subconjunto				
		1	2	3	4	5
T1= Var seda +arena + H-NFT	4	21.9250				
T2= Var seda + Sust Trad+ H- NFT	4		22.6000			
T5= Var cressa + arena+ H-NFT	4		22.7225			
T6=Var cressa +Sust Trad+ H-NFT	4			24.0775		
T3= Var americana + arena + H-NFT	4				30.1200	
T4= Var americana +Sust Trad+ H- NFT	4					32.3675
Sig.		1.000	0.988	1.000	1.000	1.000

En la tabla 8, el subconjunto homogéneo inicial de los promedios en T1 comprende Var seda +arena + H-NFT con una medida de 21,92 cm; el subconjunto posterior T2 incluye Var seda + Sust Trad+ H- NFT con una medida de 22,60 cm y el T5 comprende Var cressa + arena+ H-NFT con una medida de 22,72 cm; el tercer subconjunto, T6, consiste en Var cressa +Sust Trad+ H-NFT con una medida de 24,07 cm; el cuarto subconjunto, el T3, abarca Var americana + arena + H-NFT, con una medida de 30,12 cm. En el quinto subconjunto, el T4 está formado Var americana +Sust Trad+ H-NFT con una medida de 32,36 cm.

3.3 N° de hojas

La evaluación empírica del recuento de las hojas de tres variedades distintas de lechuga (*Lactuca sativa* L) se llevó a cabo 60 días después de la siembra, empleando una metodología hidropónica conocida como Nutrient Film Technique (NFT).

Tabla 9

Estadístico descriptivo del número de hojas (Ud.)

Tratamiento	Número de hojas- cuarta evaluación			
	Media	Máximo	Mínimo	Desviación estándar
T1-Var seda +arena + H-NFT	24.93	25.10	24.70	0.17
T2-Var seda + Sust Trad+ H- NFT	19.85	20.10	19.70	0.17
T3-Var americana + arena + H-NFT	22.43	22.60	22.30	0.13
T4-Var americana +Sust Trad+ H- NFT	18.95	19.10	18.80	0.13
T5-Var cressa + arena+ H-NFT	16.13	16.30	15.90	0.17
T6-Var cressa +Sust Trad+ H- NFT	16.63	16.80	16.50	0.13

Los resultados del número de hojas de tres variedades de lechuga (*Lactuca sativa* L.) donde el T1= (var. seda +arena + H-NFT) con 24.93 cm, es superior a los otros tratamientos; seguido por el T3= (var. americana + arena+ H-NFT) con 22.43 cm; seguido por el T2= var seda + sustrato tradicional + H-NFT) con 19.85 cm; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 18.95 cm; seguido por el T6= (var. cressa +sustrato tradicional + H-NFT) con 16.63 cm y finalmente el T5= (var. cressa + arena+ H-NFT) con 16.13 cm.

Tabla 10

Análisis de varianza número de hojas – cuarta evaluación

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	9654,740	9	1072,749	47327,157	0,000
Tratamiento	229,863	5	45,973	2028,206	0,000
Bloque	0,070	3	0,023	1,029	0,408
Error	0,340	15	0,023		
Total	9655,080	24			

Modelo. En el cuadro 10, se observa que el valor p de significancia (sig) = 0.000 es inferior al umbral de significancia pre-determinado (< 0,05), por lo que es necesario rechazar la hipótesis nula, lo que lleva a concluir que el modelo general es lineal y que se cumple el requisito previo de un diseño de bloques completos aleatorios (DBCA).

Tratamientos. En la tabla 10, es evidente que el valor p de significancia (sig) = 0.000 está por debajo del nivel de significancia establecido (< 0.05), lo que resulta en el rechazo de la hipótesis nula, que permite afirmar que existe un efecto atribuible a los dos tipos de semillas y sustratos hidropónicos (NFT) en el recuento foliar (cuarta evaluación), en relación con tres variedades de lechuga (*Lactuca sativa* L): seda, americana y cressa.

Bloques. En la tabla 10, se observa que el valor p de significancia (sig) = 0,408 supera el umbral de significancia designado (> 0,05), por lo que se decide rechazar la hipótesis nula (Ho) y se afirma que los bloques no influyen en el recuento de las hojas (cuarta evaluación), en relación con tres variedades de lechuga (*Lactuca sativa* L): seda, americana y cressa.

Para determinar cuál de los tratamientos: T1 = (var. seda + arena + H-NFT), T2 = (var. seda + sustrato tradicional + H-NFT), T3 = (var. americana + arena + H-NFT), T4 = (var. americana + sustrato tradicional + H-NFT), T5 = (var. cressa+ arena + H-NFT) y T6 = (var. cressa + tradicional sustrato + H-NFT), ejercen un efecto más pronunciado en el recuento foliar: en la cuarta evaluación de tres variedades de lechuga (*Lactuca sativa* L): seda, americana y cressa, se realizó una comparación múltiple de medias utilizando la prueba de Tukey con un nivel de probabilidad del 95%, como se ilustra en el la tabla posterior.

Tabla 11

Prueba de tukey, número de hojas – 4ta evaluación

Tratamiento	N	Subconjunto					
		1	2	3	4	5	6
T5= Var cressa + arena+ H-NFT	4	16.1250					
T6=Var cressa +Sust Trad+ H- NFT	4		16.6250				
T4= Var americana +Sust Trad+ H- NFT	4			18.9500			
T2= Var seda + Sust Trad+ H- NFT	4				19.8500		
T3= Var americana + arena + H- NFT	4					22.4250	
T1= Var seda +arena + H-NFT	4						24.9250
Sig.		1.000	1.000	1.000	1.000	1.000	1.000

En la Tabla 11, el subconjunto homogéneo inicial de los promedios al T5= Var cressa + arena+ H-NFT con 16.12 unidades; en el segundo subconjunto el T6=Var cressa +Sust Trad+ H-NFT

con 16.62 unidades; en el tercer sub conjunto el T4= Var americana +Sust Trad+ H-NFT con 18.95 unidades; en el cuarto sub conjunto el T2= Var seda + Sust Trad+ H-NFT con 19.85 unidades y; en el quinto sub conjunto el T3= Var americana + arena + H-NFT con 22.42 unidades y el sexto sub conjunto el T1= Var seda arena + H-NFT con 24.92 unidades, con mayor numero.

3.4 Diámetro de la cabeza de lechuga

La evaluación de campo del diámetro de la cabeza de tres variedades distintas de lechuga (*Lactuca sativa* L) se llevó a cabo 60 días después de la siembra, mediante la hidroponía, Técnica de Película Nutriente (NFT).

Tabla 12

Estadístico descriptivo del diámetro de cabeza de lechuga(gr)

Tratamiento	Diámetro de la cabeza			
	Media	Máximo	Mínimo	Desviación estándar
T1-Var seda +arena + H-NFT	34.30	34.60	33.90	0.32
T2-Var seda + Sust Trad+ H- NFT	30.47	31.77	28.52	1.38
T3-Var americana + arena + H-NFT	21.80	22.30	21.25	0.46
T4-Var americana +Sust Trad+ H- NFT	22.43	23.42	21.60	0.78
T5-Var cressa + arena+ H-NFT	31.26	31.98	30.20	0.77
T6-Var cressa +Sust Trad+ H-NFT	28.60	28.90	28.10	0.36

Los hallazgos ilustran el diámetro de la cabeza de tres variedades distintas de lechuga (*Lactuca sativa* L.), donde T1= (var. seda +arena + H-NFT) con 34.30 cm, es superior a los otros tratamientos, seguido del T5= (var. cressa + arena+ H-NFT) con 31.26 cm, seguido del T2= var seda + sustrato tradicional + H- NFT) con 30.47 cm, seguido del T6= (var. cressa +sustrato tradicional + H-NFT) con 28.60 cm $\pm$ 0.36; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 22.43 cm y finalmente el T3= (var. americana + arena+ H-NFT) con 21.80 cm.

Tabla 13

Análisis de varianza diámetro de la cabeza de lechuga

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	19515.717	9	2168.413	4240.980	0.000
Tratamiento	504.719	5	100.944	197.426	0.000
Bloque	2.991	3	0.997	1.950	0.165
Error	7.670	15	0.511		
Total	19523.386	24			

Modelo. En la tabla 13, es evidente que el nivel de significancia del valor p de 0.000 es sustancialmente inferior al umbral de significancia predeterminado (< 0.05), por lo que es necesario rechazar la hipótesis nula; en consecuencia, se puede concluir que el modelo global es lineal y que se cumplen los requisitos previos para un diseño de bloques completos aleatorios (DBCA).

Tratamientos. En el cuadro 13, se observa que el nivel de significancia del valor p de 0.000 es significativamente inferior al umbral de significancia establecido (< 0.05), lo que lleva a rechazar la hipótesis nula, lo que indica que existe un efecto atribuible a los dos tipos distintos de semillas y sustratos hidropónicos (NFT) en el diámetro de la cabeza de tres variedades de lechuga (*Lactuca sativa* L), a saber, seda, americana y cressa.

Bloques. En el cuadro 13, se observa que el nivel de significancia del valor p de 0,165 supera el umbral de significancia designado ($> 0,05$), por lo que se decidió rechazar la hipótesis nula (Ho), afirmando así que los bloques no influyen en el diámetro de la cabeza de las tres variedades de lechuga (*Lactuca sativa* L), específicamente la seda, la americana y la cressa.

Para determinar cuál de los tratamientos: T1= (var. seda + arena + H-NFT), T2= (var. seda + sustrato tradicional + H-NFT), T3= (var. cressa + arena + H-NFT), T4= (var. americana + sustrato tradicional + H-NFT), T5= (var. cressa + arena + H-NFT) y T6= (var. cressa + sustrato tradicional + H-NFT), producen mayores efectos en el diámetro de la cabeza de las tres variedades de lechuga (*Lactuca sativa* L), a saber, seda, americana y cressa. Se realizó una comparación múltiple de medias utilizando la prueba de Tukey con un nivel de confianza del 95%, como se describe en la tabla siguiente.

Tabla 14

Prueba de tukey, diámetro de la cabeza de lechuga

Tratamiento	N	Subconjunto			
		1	2	3	4
T3= Var americana + arena + H-NFT	4	21.7950			
T4= Var americana +Sust Trad+ H- NFT	4	22.4300			
T6=Var cressa +Sust Trad+ H-NFT	4		28.6000		
T2= Var seda + Sust Trad+ H- NFT	4			30.4725	
T5= Var cressa + arena+ H-NFT	4			31.2575	
T1= Var seda +arena + H-NFT	4				34.3000
Sig.		0.803	1.000	0.639	1.000

En la tabla 14, se observa en el primer sub conjunto homogéneo de las medias al T3= Var americana + arena + H-NFT, con 21.79 cm y el T4= Var americana +Sust Trad+ H-NFT con 22.43 cm; en el segundo sub conjunto el T6=Var cressa +Sust Trad+ H-NFT con 28.60 cm; en el tercer sub conjunto el T2= Var seda + Sust Trad+ H-NFT con 30.47 cm y el T5= Var cressa + arena+ H-NFT con 31.25cm y en el cuarto sub conjunto con mayor diámetro el T1= Var seda +arena + H-NFT con 34.30cm.

4 DISCUSIONES

Longitud radicular. Los resultados se muestran en promedios correspondiente al T3= (var. americana + arena+ H-NFT) con 29.86 cm siendo superior a los otros tratamientos, seguido del T4= (var. americana +sustrato tradicional + H-NFT) con 26.61 cm seguido del T1= (var. seda +arena + H-NFT) con 26.17 cm $\pm$ 0.45; seguido del T2= var seda + sustrato tradicional + H- NFT) con 25.80 cm; seguido del T5= (var. cressa + arena+ H-NFT) con 23.61 cm $\pm$ 0.65 y finalmente el T6= (var. cressa +sustrato tradicional + H-NFT) con 21.54 cm. [9]Guapi, en sus resultados en cuanto a la longitud de raíz a los 45 días V1: Jade, presentó un promedio de 27.75 cm. y [10]Ceballos con respecto a la longitud de raíz obtuvo un promedio de 20.45 cm, aplicando la solución LM a la variedad Starfighter; en cambio Patagonia con la solución T obtuvieron un promedio de 13.68 cm.

Alra de la planta. Los resultados se presentan en promedios correspondiente al T4= (var. americana + sustrato tradicional + H-NFT) con 32.37 cm, es superior a los otros tratamientos, seguido del T3= (var. americana + arena+ H-NFT) con 30.12 cm; seguido del T6= (var. crespa + sustrato tradicional + H-NFT) con 24.08 cm; se-guido del T5= (var. crespa + arena+ H-NFT) con 22.72 cm; se-guido del T2= var seda + sustrato tradicional + H- NFT) con 22.60 cm y finalmente el T1= (var. seda +arena + H-NFT) con 21.60 cm. Según los hallazgos de [11]Tapia indica que al evaluar las variables de altura y recuento de hojas por planta a los 60 días, el tratamiento T1: (variado). Batavia + Guanu (solución nutritiva) produjo 27 hojas y logró una altura de planta de 39.84 cm, así mismo [9]Guapi, en los resultados en cuanto a la altura de la planta, las variedades V2: (Scarlet) y V3: (Starfighter) alcanzaron promedios de 17.56 y 16.43 cm respectivamente de igual [12]Guerra con sus resulta-dos de estudio en cuanto a la altura de planta de acuerdo al análisis estadístico a los 60 días (Marrón =28,97), (Romana=17,59), (Crespa=10,83)

Diámetro de la cabeza de lechuga. Los resultados se presentan en promedio de T1= (var. seda +arena + H-NFT) con 34.30 cm, es superior a los otros tratamientos, seguido del T5= (var. crespa + arena+ H-NFT) con 31.26 cm, seguido del T2= var seda + sustrato tradicional + H- NFT) con 30.47 cm, seguido del T6= (var. crespa +sustrato tradicional + H-NFT) con 28.60 cm; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 22.43 cm y finalmente el T3= (var. americana + arena+ H-NFT) con 21.80 cm. Según los hallazgos de [13]Robles indican que la Variedad Boston, obtuvo el mayor tamaño en el diámetro de la cabeza con un promedio 27.74 cm/cabeza, de igual forma [14]Criollo menciona que obtuvo resultados, con purín de cuy alcanzando un diámetro 21.51 cm según estadísticos.

Número de hojas. Los resultados se muestran en promedios T1= (var. seda +arena + H-NFT) con 24.93 cm, es superior a los otros tratamientos; seguido por el T3= (var. americana + arena+ H-NFT) con 22.43 cm; seguido por el T2= var seda + sustrato tradicional + H- NFT) con 19.85 cm; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 18.95 cm; seguido por el T6= (var. crespa +sustrato tradicional + H-NFT) con 16.63 cm y finalmente el T5= (var. crespa + arena+ H-NFT) con 16.13 cm. Según los hallazgos de [11]Tapia la altura y recuento de hojas por planta a los 60 días, el tratamiento T1: (variado). Batavia + Guanu (solución nutritiva) resultó tener 27 hojas en promedio según estadísticas, así mismo [15]Rojas según diferencias estadísticas entre los tratamientos (T1, T2 y T3) del sistema hidropónico. el T1 (purín de cuy) como solución nutritiva, generaron resultados: para Número de hojas 18.67 (unidades) por planta.

5 CONCLUSIONES

A los 60 días después de la siembra se registra diferencias significativas en la longitud radicular por planta siendo el mayor promedio la variedad americana + arena+ H-NFT con un promedio de 29.86 cm. Altura de planta, variedad americana+sustrato tradicional+H+NFT con un promedio de 32.37 cm. Diámetro de la cabeza de la lechuga variedad seda+arena+H+NFT, con un promedio de 34.30 cm y número

de hojas la variedad seda+arena+H+NFT con un promedio de 24.93 cm, llegando a una conclusión de la evaluación del efecto de dos tipos de sustrato de almágo más hidroponía NFT en las características morfológicas de tres variedades de lechuga (*Lactuca sativa* L.) en el INIA Andahuaylas Perú tuvo significancia, donde indica que los tratamientos que se utilizó sustratos de arena, junto con la hidroponía NFT, facilitan notablemente un crecimiento radicular y aéreo mejorado en comparación con los sustratos tradicional (tierra negra, arena y compost)

AGRADECIMIENTO

Se agradece al Instituto Nacional de Innovación Agraria (INIA) Andahuaylas-Apurímac, quienes brindaron su apoyo en la investigación, de igual forma a la Universidad Nacional Micaela Bastidas de Apurímac, por dar la oportunidad de investigar. Feria de Ciencia e Innovación Tecnológica - FERCYT UNAMBA 2022 - 2023

REFERENCIAS

- [1] C. Cuevas, "Diseño Y Construcción De Un Sistema Hidropónico Nutrient Film Technique;Como Alternativa Sustentable Decultivo, Produccióny Consumo De Hortalizas,Para El Beneficio De Los Habitantes Del Islote El Pardito En Baja California Sur.," 2021.
- [2] N. itzel Perez and A. Solorzano, "Eficiencia En La Produccion De Lechuga Hidroponica Caso Fresco Cultivos Hidroponicos,," p. 108, 2021.
- [3] V. Vargas, "cultivo hidropónico como alternativa sostenible a cultivos tradicionales," 2022.
- [4] Hydro Environmet., "Innovacion agricola," 2024. https://hydroenv.com.mx/catalogo/index.php?main_page=product_info&cPath=80&products_id=1635
- [5] X. Suclla, "Comparativo de tres frecuencias de riego en el crecimiento y desarrollo de lechuga hidropónica," 2023, 2023.
- [6] S. Molina and A. Mora, "UNIVERSIDAD DE GUAYAQUIL," 2020.
- [7] N. Cedeño, "Identificación del efecto de deficiencias nutricionales en el cultivo de lechuga (*Lactuca sativa* L.) bajo sistema hidropónico," Univ. Técnica Estatal Quevedo, 2022.
- [8] A. Gonzales, "universidad de guayaquil carrera de ingeniería agronómica previo a la obtención del título de : ingeniera agronómica tema : evaluación de soluciones nutritivas en lechuga (*lactuca sativa* l .) con la técnica hidropónica nft , provincia de santa elena . an," 2022.
- [9] A. GUAPI, K. ROBAYO, E. SALAZAR, and M. VIVAR, "Estrategias de cultivo : comparativa de tipologías de lechuga en sistemas hidropónicos," vol. 7, no. 3, pp. 68-82, 2024.
- [10] M. CEVALLOS, "Aplicación De Soluciones Nutritivas En Variedades De Lechuga En Cultivo Hidropónico Bajo El Sistema Nft,," Univ. Técnica Cotopaxi Fac., vol. 1, p. 101, 2020, [Online]. Available: <http://repositorio.utc.edu.ec/bitstream/27000/4501/1/PI-000727.pdf>
- [11] J. TAPIA, "Eficacia de soluciones nutritivas en la productividad de variedades de lechuga *lactuca sativa* l. mediante el sistema nft," 2022.
- [12] G. GUERRA and M. YUGSI, "Producción Hidropónico De Tres

Variedades De Lechuga (*Lactuca Sativa* L.) Bajo El Sistema Nft (Nutrien Film Technique)." 2022. [Online]. Available: <https://repositorio.utc.edu.ec/bitstream/27000/8980/1/UTC-PIM-000528.pdf>

- [13] F. ROBLES and L. VALLE, "T E S I S Efecto del sistema hidropónico recirculante con solución nutritiva en el rendimiento de seis variedades de lechugas (*Lactuca sativa* L) en la localidad de Yanacocha - Yanahuanca - Pasco. Autor:," vol. 21, no. 1, pp. 1-9, 2020.
- [14] G. CRIOLLO, "Análisis de los cambios morfológicos y nutricionales de las lechugas: *Lactuca sativa* L. variedades longifolia y capitata, debido a la influencia de la luz led, en un huerto urbano hidropónico en condiciones controladas," 2023. [Online]. Available: <https://dspace.ups.edu.ec/bitstream/123456789/26484/1/UPS-CT010995.pdf>
- [15] J. T. ROJAS, J. Bazán Hernández, M. P. Rodríguez Vasquez, A. C. Flores Quispe, and F. Del Carpio Delgado, "Efecto de la ceniza de bagazo de caña de azúcar en las propiedades mecánicas del concreto permeable," *Salud, Cienc. y Tecnol. - Ser. Conf.*, vol. 2, no. 21, p. 337, 2023, doi: 10.56294/sctconf2023337.

BIOGRAFÍAS

- Reyna Vargas-Peña, estudiante investigadora de la Escuela Académica Profesional de Ingeniería Agroecológica y Desarrollo Rural en la Universidad Nacional Micaela Bastida de Apurímac.
- Niki Franklin Flores Pacheco, Ingeniero Agrónomo, docente de la Universidad Nacional Micaela Bastidas de Apurímac, investigador en la línea de investigación: Agua, agricultura, silvicultura y pecuaria sostenible.

Rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) en dos tipos de sustrato en almacigo más hidroponía NFT en INIA Andahuaylas Perú-2024

Performance of three varieties of lettuce (*Lactuca sativa* L.) in two types of substrate in a seedbed plus NFT hydroponics in INIA Andahuaylas Perú-2024

Reyna Vargas-Peña ^A y Niki Franklin Flores-Pacheco ^B

Resumen— El objetivo de esta investigación fue evaluar el efecto de dos tipos de sustrato en almacigo más hidroponía NFT en el rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) en el INIA Andahuaylas. Se utilizó el diseño de bloques completamente al azar, con seis tratamientos; las variables evaluadas fueron: peso (g) y Rendimiento (Kg/1000 m²) de la lechuga. Los resultados evidenciaron diferencias significativas (sig. <0.05) Reportando el peso de la lechuga (g). T3= (var. americana + arena+ H-NFT) con 138.10 g, T1= (var. seda +arena + H-NFT) con 111.10 g; T4= (var. americana +sustrato tradicional + H-NFT) con 103.28 g; T5= (var. cressa + arena+ H-NFT) con 98.08 g; T6= (var. cressa +sustrato tradicional + H-NFT) con 95.15 g y T2= (var seda + sustrato tradicional + H- NFT) con 85.15 g. Rendimiento (kg/1000m².) T3= (var. americana + arena+ H-NFT) con 3452.50 kg/1000 m², T1= (var. seda +arena + H-NFT) con 2777.50 kg/1000 m²; T4= (var. americana +sustrato tradicional + H-NFT) con 2581.88 kg/1000 m²; T5= (var. cressa + arena+ H-NFT) con 2451.88 kg/1000 m²; T6= (var. cressa +sustrato tradicional + H-NFT) con 2378.75 kg/1000 m² y T2= var seda + sustrato tradicional + H- NFT). Concluimos que la variedad americana con sustrato de arena, tiene mejores rendimientos bajo el sistema hidroponía NFT.

Palabras clave: Almacigo, hidroponía, sustratos y rendimiento

Abstract— The objective of this research was to evaluate the effect of two types of substrate in a seedbed plus NFT hydroponics on the performance of three varieties of lettuce (*Lactuca sativa* L.) at the INIA Andahuaylas. The completely randomized block design is used, with six treatments; The variables evaluated were: weight (g) and Yield (Kg/1000 m²) of the lettuce. The results showed significant differences (sig. <0.05) Reporting the weight of the lettuce (g). T3= (American var. + sand + H-NFT) with 138.10 g, T1= (silk var. + sand + H-NFT) with 111.10 g; T4= (American var. + traditional substrate + H-NFT) with 103.28 g; T5= (var. cressa + sand+ H-NFT) with 98.08 g; T6= (var. crepe + traditional substrate + H-NFT) with 95.15 g and T2= (var. silk + traditional substrate + H-NFT) with 85.15 g. Yield (kg/1000m²) T3= (American variety + sand + H-NFT) with 3452.50 kg/1000 m², T1= (silk variety + sand + H-NFT) with 2777.50 kg/1000 m²; T4= (American var. + traditional substrate + H-NFT) with 2581.88 kg/1000 m²; T5= (var. cressa + sand+ H-NFT) with 2451.88 kg/1000 m²; T6= (var. crepe + traditional substrate + H-NFT) with 2378.75 kg/1000 m² and T2= var. silk + traditional substrate + H-NFT). We conclude that the American variety with a sand substrate has better yields under the NFT hydroponics system.

Keywords: Almacigo, hydroponics, substrates and performance

1 INTRODUCCIÓN

La lechuga (*Lactuca sativa* L.) se encuentra entre las verduras más consumidas a nivel mundial, debido a sus beneficios nutricionales y a su frecuente incorporación a la dieta diaria.[1][2].En el contexto peruano, su cultivo ha experimentado una expansión significativa, particularmente en sistemas de producción más avanzados desde el punto de vista tecnológico, como la hidroponía, que facilitan el cultivo sin suelo y mejoran la eficiencia de los recursos.[3] La hidroponía, con un énfasis particular en el sistema de la técnica de película de nutrientes (NFT), se ha convertido en una alternativa viable para aumentar la productividad de los cultivos hortícolas, ya que maximiza la absorción de nutrientes y agua a través de

un flujo continuo de solución nutritiva a través de los sistemas radiculares de las plantas.[4][5][6]

En este marco, se ha propuesto la exploración de técnicas innovadoras para mejorar la producción de lechuga en sistemas hidropónicos, lo que requiere la adaptación de las prácticas agrícolas para alinearlas con las condiciones locales. Un factor fundamental que influye en el éxito del cultivo en ambientes hidropónicos es la selección del sustrato empleado en la etapa de plantación. Los sustratos sirven para proporcionar soporte físico a las plántulas y ejercen un impacto directo en su crecimiento y desarrollo inicial, lo que influye en la productividad final de las plantas.



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 22/03/24 - Indexaciones

Número: doi.org/10.57166/riqchary/v6.n2.2024

Pág. 35 - 40 | Recibido 10/10/2024 ; Aceptado 05/11/2024



doi.org/10.57166/riqchary.v6.n2.2024.126

Autores:

A. ORCID iD <https://orcid.org/0009-0002-5638-3920>
Reyna Vargas Peña, Universidad Nacional Micaela Bastidas de Apurímac, 181677@unamba.edu.pe

B. ORCID iD <https://orcid.org/0000-0002-2772-0951>
Niki Franklin Flores Pacheco, Universidad Nacional Micaela Bastidas de Apurímac, nflores@unamba.edu.pe

Numerosos estudios han demostrado que la selección del sustrato durante la fase de almacenado puede afectar significativamente a la morfología, el crecimiento y el rendimiento general de las plantas.[7] En consecuencia, este estudio tiene como objetivo evaluar dos tipos de sustrato durante la etapa del almacenado junto con el sistema NFT en tres variedades de lechuga (*Lactuca sativa* L.) cultivadas en las instalaciones del INIA en Andahuaylas, Apurímac Perú.

La importancia de esta investigación se destaca por su potencial para mejorar la eficiencia de la producción de lechuga, sumado también a contribuir a la sostenibilidad agrícola en la región, maximizando así la utilización de recursos como el agua y los nutrientes. Además, los hallazgos derivados de este estudio pueden replicarse en otros cultivos dentro de los sistemas hidropónicos, fomentando así la innovación y la competitividad en el sector agrícola.

2 MÉTODO

El método adoptado fue experimental, ya que se observó el fenómeno formulado en la hipótesis, manipulando las variables independientes tres variedades de lechuga (*Lactuca Sativa* L) con dos tipos de sustratos (sustrato tradicional y arena) para evaluar sus efectos en las variables dependientes peso y rendimiento.

2.1 Tipo y nivel de investigación

El tipo de investigación. Por la duración en el tiempo es transversal (en un periodo agrícola). Por la naturaleza de los datos es cuantitativa (las variables evaluadas son cuantitativas). El nivel de investigación. Es descriptiva (se describen los resultados utilizando la estadística descriptiva). Es experimental (se manipularon las variables tres variedades de lechuga y dos tipos de sustratos) para medir el peso y rendimiento de la lechuga.

2.2 Diseño de investigación

El diseño adoptado es para un experimento, y se dio respuestas a las preguntas formuladas en la investigación y la validación de la hipótesis. Se utilizó el diseño de bloques completamente al azar (DBCA), con seis tratamientos que se describen a continuación:

T1= var seda +arena + hidroponia NFT

T2= var seda + sustrato tradicional + hidroponia NFT

T3= var americana + arena + hidroponia NFT

T4= var americana +sustrato tradicional + hidroponia NFT

T5= var cressa + arena+ hidroponia NFT

T6= var cressa +sustrato tradicional + hidroponia NFT

2.3 Población y muestra

La población lo conformaron 200 plantas de las tres variedades (seda, americana y cressa) de lechuga (*Lactuca sativa* L). La muestra, se determinó utilizando el método probabilístico a nivel de confianza al 95 %, obteniendo 132 unidades de plantas que fueron evaluadas.

2.4 Análisis estadístico

Los resultados de la investigación fueron recogidos en fichas de evaluación. Para el procesamiento de los datos se realizó las siguientes pruebas: Normalidad, Homogeneidad de varianzas. Para la prueba de hipótesis se realizó mediante ANOVA y la prueba de Tukey. Se procesó los datos dando respuesta al objetivo e hipótesis del estudio, utilizando la estadística descriptiva con gráficos y tablas.

Prueba de Normalidad

Para determinar si utilizamos la estadística paramétrica y no paramétrica. Se realizó la prueba de normalidad utilizando el estadístico Shapiro Wilk.

Tabla 1
Prueba de normalidad

Variable	Significancia de los tratamientos al 95%					
	T1= var seda +arena + H- NFT	T2=var seda + Sust Trad+ H-NFT	T3=var ameri- cana + arena + H- NFT	T4=var ameri- cana +Sust Trad+ H-NFT	T5=var cressa + arena+ H- NFT	T6=var cressa +Sust Trad+ H- NFT
Peso de la lechuga	0.915	0.722	0.469	0.976	0.161	0.650
Rendimiento	0.915	0.722	0.469	0.976	0.161	0.650

En la tabla 1, se observan los resultados de la prueba de normalidad y se analizó al nivel de significancia de 95%, se obtuvo valores mayores a 0.05 (Sig.>0.05). No rechazamos la hipótesis nula (H₀) y llegamos a la conclusión que los datos provienen de una población normal.

Prueba de homogeneidad de varianza

Para determinar la homogeneidad de varianza de las variables, utilizamos el estadístico de Levene.

Tabla 2
Prueba de homogeneidad de varianza

Variable	Estadístico de Levene	gl1	gl2	Sig.
Peso de la lechuga	0.949	5	18	0.474
Rendimiento	0.949	5	18	0.474

En la tabla 2, observamos que el valor de la significancia es mayor a 0.05 (Sig.>0.05), por lo que no se rechazó la hipótesis nula y se aceptó que las variables son homogéneas.

3 RESULTADOS

Rendimiento de las tres variedades de lechuga (*Lactuca sativa* L)

Se realizó las evaluaciones del peso de la planta y el rendimiento en 1000 m², de las tres variedades de lechuga (*Lactuca sativa* L), en la estación experimental INIA Chumbibamba, Andahuaylas Perú.

3.1 Peso de la planta

Evaluación en campo del peso de la planta de las tres variedades de lechuga (*Lactuca sativa* L) a los 60 días después de la siembra de almácigos bajo hidroponía NFT.

Tabla 3
Estadístico descriptivo del peso de la planta (g)

Tratamiento	Peso de la planta			
	Me- dia	Má- ximo	Mi- nimo	Desv. típica
T1= var seda +arena + H-NFT	111.10	112.50	109.40	1.31
T2= var seda + Sust Trad+ H- NFT	85.15	86.70	83.90	1.16
T3= var americana + arena + H-NFT	138.10	139.00	136.70	1.05
T4= var americana +Sust Trad+ H- NFT	103.28	104.70	102.00	1.13
T5= var cressa + arena+ H-NFT	98.08	98.30	97.90	0.21
T6=var cressa +Sust Trad+ H-NFT	95.15	96.30	94.30	0.84

Los resultados de la tabla 3, del peso de la planta de tres variedades de lechuga (*Lactuca sativa* L.). El T3= (var. americana + arena+ H-NFT) con 138.10 g, es superior a los otros tratamientos seguido por el T1= (var. seda +arena + H-NFT) con 111.10 g; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 103.28 g; seguido por el T5= (var. cressa + arena+ H-NFT) con 98.08 g; seguido por el T6= (var. cressa +sustrato tradicional + H-NFT) con 95.15 g y finalmente el T2= var seda + sustrato tradicional + H- NFT) con 85.15 g.

Prueba de ANOVA

Se realizó a un nivel de confianza del 95%, los resultados se expresan en tabla siguiente.

Tabla 4
Análisis de varianza peso de planta

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	272021.720 ^a	9	30224.636	45246.460	0.000
Tratamiento	6698.703	5	1339.741	2005.600	0.000
Bloque	8.535	3	2.845	4.259	0.230
Error	10.020	15	0.668		
Total	272031.740	24			

La tabla 4, se muestra los resultados del análisis de varianza, según:

El modelo. En la tabla 4, se observa que el p-valor sig = 0.000, es menor al nivel de significancia (< 0.05), motivo por el cual rechazamos la hipótesis nula, se concluye que el modelo general es lineal, y se cumple con el supuesto para un diseño de bloques completos al azar (DBCA).

Los tratamientos. En la tabla 4, se observa que el p-valor sig = 0.000, es menor al nivel de significancia (< 0.05), por lo cual se rechaza la hipótesis nula y se concluye que existe el efecto de los dos tipos de sustratos de almácigo e hidroponía NFT en el peso de tres variedades de lechuga (*Lactuca sativa* L) seda, americana y cressa.

Los bloques. En la tabla 4, se observa que el p-valor sig = 0.230, es mayor al nivel de significancia (< 0.05) que nos lleva a tomar la decisión de rechazar la hipótesis nula (H0) y manifestamos que los bloques no influyen en el peso de las tres variedades de lechuga (*Lactuca sativa* L) seda, americana y cressa.

Prueba de Tukey

Para determinar cuál de los tratamientos: T1= (var. seda +arena + H-NFT), T2= var seda + sustrato tradicional + H-NFT), T3= (var. americana + arena+ H-NFT), T4= (var. americana +sustrato tradicional + H-NFT), T5= (var. cressa + arena+ H-NFT) y T6= (var. cressa +sustrato tradicional + H-NFT), tienen mayores efectos en el peso de tres variedades de lechuga (*Lactuca sativa* L) seda, americana y cressa, se realizó la comparación múltiple de medias, a través de la prueba de Tukey al 95% de probabilidad, que se muestra en la tabla 5.

Tabla 5
Prueba de tukey, peso de planta

Tratamiento	N	Subconjunto					
		1	2	3	4	5	6
T2= var seda + Sust Trad+ H- NFT	4	85.15					
T6=var cressa +Sust Trad+ H-NFT	4		95.15				
T5= var cressa + arena+ H-NFT	4			98.07			
T4= var americana +Sust Trad+ H- NFT	4				103.27		
T1= var seda +arena + H-NFT	4					111.10	
T3= var americana + arena + H-NFT	4						138.10
Sig.		1.000	1.000	1.000	1.000	1.000	1.000

En la tabla 5, se observa en el primer sub conjunto homogéneo de las medias al T2= var seda + Sust Trad+ H-NFT con 85.15 g; en el segundo sub conjunto el T6=var cressa +Sust Trad+ H-NFT con 95.15 g; en el tercer sub conjunto el T5= var cressa + arena+ H-NFT con 98.07g; en el cuarto sub conjunto el T4= var americana +Sust Trad+ H-NFT con 103.27 g; en el quinto sub conjunto el T1= var seda +arena + H-NFT con 111.10 g y en el sexto sub conjunto el T3= var americana + arena + H-NFT, con 138.10 g, con mayor peso.

Analizando el resultado, los tratamientos más recomendables para obtener un mayor peso de lechuga es el T3= variedad americana + arena + hidroponía NFT.

3.2 Rendimiento (kg/1000 m²)

Evaluación en campo del rendimiento (kg/1000m²) de las tres variedades de lechuga (*Lactuca sativa* L.) a los 60 días después de la siembra de almácigos

Tabla 6
Estadístico descriptivo del rendimiento (kg/1000m²) de lechuga (*Lactuca Sativa* L)

Tratamiento	Rendimiento kg/1000 m ²			
	Media	Máximo	Mínimo	Desv. típica
T1= Var seda +arena + H-NFT	2777.50	2812.50	2735.00	32.72
T2= Var seda + Sust Trad+ H- NFT	2128.75	2167.50	2097.50	28.98
T3= Var americana + arena + H-NFT	3452.50	3475.00	3417.50	26.22
T4= Var americana +Sust Trad+ H- NFT	2581.88	2617.50	2550.00	28.31
T5= Var cressa + arena+ H-NFT	2451.88	2457.50	2447.50	5.15
T6=Var cressa +Sust Trad+ H-NFT	2378.75	2407.50	2357.50	20.97

Los resultados en la tabla 6, del rendimiento de tres variedades de lechuga (*Lactuca sativa* L.). El T3= (var. americana + arena+ H-NFT) con 3452.50 kg/1000 m², es superior a los otros tratamientos seguido por el T1= (var. seda +arena + H-NFT) con 2777.50 kg/1000 m²; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 2581.88 kg/1000 m²; seguido por el T5= (var. cressa + arena+ H-NFT) con 2451.88 kg/1000 m²; seguido por el T6= (var. cressa +sustrato tradicional + H-NFT) con 2378.75 kg/1000 m² y finalmente el T2= var seda + sustrato tradicional + H- NFT) con 2128.75 kg/1000m².

Prueba de ANOVA

Se realizó a un nivel de confianza del 95%, los resultados se expresan en tabla siguiente.

Tabla 7
Análisis de varianza- rendimiento de lechuga

Origen	Tipo III de suma de cuadrados	Gl	Media cuadrática	F	Sig.
Modelo	170013575.000a	9	18890397.222	45246.460	0.000
Tratamiento	4186689.583	5	837337.917	2005.600	0.000
Bloque	5334.375	3	1778.125	4.259	0.230
Error	6262.500	15	417.500		
Total	170019837.500	24			

La tabla 7, muestra los resultados del análisis de varianza, según:

El modelo. En la tabla 7, se observa que el p-valor sig = 0.000, es menor al nivel de significancia (< 0.05), motivo por el cual rechazamos la hipótesis nula, se concluye que el modelo general es lineal, y se cumple con el supuesto para un diseño de bloques completos al azar (DBCA).

Los tratamientos. En la tabla 7, se observa que el p-valor sig = 0.000, es menor al nivel de significancia (< 0.05), por lo cual se rechaza la hipótesis nula y se concluye que existe el efecto de los dos tipos de sustratos de almácigo e hidroponía NFT en el rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) seda, americana y cressa.

Los bloques. En la tabla 7, se observa que el p-valor sig = 0.230, es mayor al nivel de significancia (< 0.05) que nos lleva a tomar la decisión de rechazar la hipótesis nula (H₀) y manifestamos que los bloques no influyen en el rendimiento de las tres variedades de lechuga (*Lactuca sativa* L.) seda, americana y cressa.

Prueba de Tukey

Para determinar cuál de los tratamientos: T1= (var. seda +arena + H-NFT), T2= var seda + sustrato tradicional + H-NFT), T3= (var. americana + arena+ H-NFT), T4= (var. americana +sustrato tradicional + H-NFT), T5= (var. cressa + arena+ H-NFT) y T6= (var. cressa +sustrato tradicional + H-NFT), tienen mayores efectos en el rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) seda, americana y cressa, se realizó la comparación múltiple de medias, a través de la prueba de Tukey al 95% de probabilidad, que se muestra en la tabla siguiente.

Tabla 8
Prueba de tukey-Rendimiento

Tratamiento	N	Subconjunto					
		1	2	3	4	5	6
T2= var seda + Sust Trad+ H- NFT	4	2128.75					
T6=var cressa +Sust Trad+ H-NFT	4		2378.75				
T5= var cressa + arena+ H-NFT	4			2451.87			
T4= var americana +Sust Trad+ H-NFT	4				2581.87		
T1= var seda +arena + H-NFT	4					2777.50	
T3= var americana + arena + H-NFT	4						3452.50
Sig.		1.000	1.000	1.000	1.000	1.000	1.000

En la tabla 8, se observa en el primer sub conjunto homogéneo de las medias al el T2= var seda + Sust Trad+ H- NFT con 2128.75 kg/1000m²; en el segundo sub conjunto el T6=var cressa +Sust Trad+ H-NFT con 2378.75 kg/1000m²; en el tercer su conjunto el T5= var cressa + arena+ H-NFT con 2451.87 kg/1000m²; en el cuarto sub conjunto el T4= var americana +Sust Trad+ H-NFT con 2581.87 kg/1000m²; en el quinto sub conjunto el T1= Var seda +arena + H-NFT con 2777.50 kg/1000m² y en el sexto sub conjunto el T3= Var americana + arena + H-NFT, con 3452.50 kg/1000m², con mayor rendimiento.

Analizando el resultado, los tratamientos más recomendables para obtener un mayor rendimiento es el T3= Var americana + arena + H-NFT.

4 DISCUSIONES

Peso de la lechuga

Los resultados se presentan en promedios T3= (var. americana + arena+ H-NFT) con 138.10 g, es superior a los otros tratamientos seguido por el T1= (var. seda +arena + H-NFT) con 111.10 g; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 103.28 g; seguido por el T5= (var. cressa + arena+ H-NFT) con 98.08 g; seguido

por el T6= (var. cressa + sustrato tradicional + H-NFT) con 95.15 g y finalmente el T2= (var seda + sustrato tradicional + H- NFT) con 85.15 g. Nuestros resultados tienen relación con los hallazgos de [8] con las siguientes variedades marrón, romana y cressa en condiciones hidropónicas. Peso de planta (Marrón =173.44g), (Romana=107.32g), (Cressa=104.28 g), según los hallazgos de [9], peso de las lechugas 139.62 g y para el tamaño de la raíz fue de 15.00 cm.[10] registró pesos para cada cosecha que oscilaron entre 115,8 y 150,1 g por planta, con una media calculada de 139,0 g. [11] La variedad lisa presenta: 511,2 g, variedad carmín muestra 51,8 g. Las variedades mimosa y vera presentan 154,1g y 118g, respectivamente.

Rendimiento (kg/1000m².)

Los resultados se presentan en promedios de T3= (var. americana + arena+ H-NFT) con 3452.50 kg/1000 m², es superior a los otros tratamientos seguido por el T1= (var. seda +arena + H-NFT) con 2777.50 kg/1000 m²; seguido por el T4= (var. americana +sustrato tradicional + H-NFT) con 2581.88 kg/1000 m²; seguido por el T5= (var. cressa + arena+ H-NFT) con 2451.88 kg/1000 m²; seguido por el T6= (var. cressa +sustrato tradicional + H-NFT) con 2378.75 kg/1000 m² y finalmente el T2= var seda + sustrato tradicional + H- NFT) con una media de 2128.75 kg/1000m². Nuestros resultados tienen relación con los hallazgos de [12] producción de 3 tipologías de lechuga con 2 soluciones nutritivas bajo un sistema hidropónico en invernadero. El estudio consideró 3 variedades de lechuga variedad 1: jade, variedad 2: scarlet y variedad 3: starfighter y dos soluciones nutritivas, En términos de rendimiento se observó que la variedad 1 respondió de mejor manera con la solución 2 obteniendo una media de 37612,5 kg/ ha-1, la variedad 1 destacó por su longitud de raíz a los 45 días y su rendimiento en peso fresco, obteniendo resultados superiores en comparación con las otras variedades. [13] La variedad regina presenta un rendimiento promedio de 6,59 kg/m², mientras que la variedad rubinela muestra un rendimiento de 3,52 kg/m². [14] Rendimiento en campo: los tratamientos de 2 a 4 (2 litros por cada 20 litros de agua de biofermentación foliar EM y 4 toneladas por hectárea de compost EM) mostraron una eficacia superior, alcanzando el rango más alto con un rendimiento de 49.861,10 kg/ha-1.[15]La variedad white boston produjo un rendimiento de 174.608 g/planta (4.085 kg/m²), mientras que los cultivares grand rapids TBR y waldmann's green produjeron 158.525 g/planta (4.121 kg/m² en total) y 154.250 g/planta (4.010 kg/m² en total), respectivamente.

5 CONCLUSIONES

Existe el efecto de dos tipos de sustrato de almacigo más hidroponía NFT en el rendimiento de tres variedades de lechuga (*Lactuca sativa* L.) en el INIA Andahuaylas, Apurímac Perú; donde los datos revelan que las lechugas cultivadas en sustratos arena obtuvieron resultados superiores en comparación con las cultivadas en sustrato tradicional (tierra negra, arena y compost). Este resultado nos

lleva a concluir que es importante seleccionar los sustratos apropiados en la elaboración de los almacigos, para mejorar la producción en los sistemas hidropónicos; que puede influir sustancialmente en la viabilidad económica de la actividad agrícola.

AGRADECIMIENTOS

Los autores desean agradecer al director y personal que labora en el Instituto Nacional de Innovación Agraria - INIA, estación experimental de Chumbibamba- Andahuaylas. También, a las autoridades y directivos de la Universidad Nacional Micaela Bastidas de Apurímac, por el apoyo en la publicación de los resultados de la investigación.

REFERENCIAS

- [1] E. MORAN, "Efecto de sustancias minerales altamente diluidas (SMADs) en el cultivo hidropónico en plantas de lechuga (*Lactuca sativa* L.)," 2021. [Online]. Available: <https://repositorio.uteq.edu.ec/bitstream/43000/6501/1/T-UTEQ-311.pdf>
- [2] I. MEDINA, S. ARMEDO, J. GRIMPLET, and A. DIAS, "Validación de nuevos genes de referencia para estudios de expresión diferencial de genes involucrados en la síntesis de antocianinas en lechuga (*Lactuca sativa* L.) y especies silvestres relacionadas," 2022.
- [3] E. VELAZQUEZ, "Monitoreo, automatización y control de cultivos de lechuga en sistemas hidropónicos NFT," p. 6, 2021.
- [4] R. MEDRANO, "Efecto de abonos foliares en el cultivo de acelga (*Beta vulgaris* L. var. Cicla) en sistema hidropónico con la técnica de la película de nutrientes (NFT)," pp. 10–20, 2022, [Online]. Available: <https://hdl.handle.net/20.500.13080/7873>
- [5] M. RIOS, C. GARCIA, and M. TARAZONA, "Proyecto cultivo de lechugas ROSARIO," 2021.
- [6] C. CUEVAS, "Diseño Y Construcción De Un Sistema Hidropónico Nutrient Film Technique; Como Alternativa Sustentable Decultivo, Producción y Consumo De Hortalizas, Para El Beneficio De Los Habitantes Del Islote El Pardito En Baja California Sur.," 2021.
- [7] K. TEJADA, "Producción de Lechuga (*Lactuca sativa* L.) Y Acelga (*Beta vulgaris* L. var. cicla) Bajo un Sistema Jardines Verticales," 2022. [Online]. Available: <https://riunet.upv.es/bitstream/handle/10251/69118/LÓPEZ - CSA->
- [8] G. GUERRA and M. YUGSI, "Producción Hidropónica De Tres Variedades De Lechuga (*Lactuca Sativa* L.) Bajo El Sistema Nft (Nutrien Film Techniquel)." 2022. [Online]. Available: <https://repositorio.utc.edu.ec/bitstream/27000/8980/1/UTC-PIM-000528.pdf>
- [9] G. CRIOLLO, "Análisis de los cambios morfológicos y

- nutricionales de las lechugas: *Lactuca sativa* L. variedades longifolia y capitata, debido a la influencia de la luz led, en un huerto urbano hidropónico en condiciones controladas," 2023. [Online]. Available: <https://dspace.ups.edu.ec/bitstream/123456789/26484/1/UPS-CT010995.pdf>
- [10] R. PERTIERRA and J. QUISPE, "Análisis Económico De Lechugas Hidropónicas Bajo Economic Analysis of Hydroponic Lettuce Under Floating Root System in Semi - Arid Climate," vol. 31, no. 1, pp. 118–130, 2020.
- [11] J. CABRERA, "evaluación de cuatro cultivares de lechuga en parámetros agronomicos similares en la granja santa INES," 2021. [Online]. Available: <https://dspace.ucuenca.edu.ec/handle/123456789/36311>
- [12] A. GUAPI, K. ROBAYO, E. SALAZAR, and M. VIVAR, "Estrategias de cultivo: comparativa de tipologías de lechuga en sistemas hidropónicos," vol. 7, no. 3, pp. 68–82, 2024.
- [13] R. PALMA, "Comparativo del rendimiento de tres cultivares de lechuga (*Lactuca sativa* L.) empleando solución nutritiva y biol bajo sistema hidropónico nft en el fundo 'la banda' Huasacache, Arequipa 2017.," 2019.
- [14] A. FELIX, "efectos del biofermento foliar y compost con (em), en el rendimiento de lechuga (*Lactuca sativa* L. var. great lakes 659) en condiciones edafoclimaticas de la esperanza – amarilis," 2021.
- [15] N. RIVERA, "Evaluación de tres variedades de lechuga (*Lactuca sativa* L.) cultivadas con la técnica hidropónica de flujo laminar de nutrientes (NFT) en el centro experimental de cota cota-la paz," 2015. [Online]. Available: <file:///D:/EXCEL profesional/withi boston eso sin raiz.pdf>

BIOGRAFÍA

Reyna Vargas Peña, estudiante investigadora de la Escuela Académica Profesional de Ingeniería Agroecológica y Desarrollo Rural de la Universidad Nacional Micaela Bastidas de Apurímac.

Niki Franklin Flores Pacheco, Ingeniero Agrónomo, docente investigador de la Universidad Nacional Micaela Bastidas de Apurímac. Línea de investigación: Agua, agricultura, silvicultura y pecuaria sostenible.

Modelo predictivo de machine learning para el servicio de orientación vocacional hacia los estudiantes de secundaria en la Gerencia Regional de Trabajo y Promoción del Empleo Cusco 2024

Machine learning predictive model for vocational guidance service for high school students in the Regional Management of Labor and Employment Promotion Cusco 2024

Malu Beatriz Silva-Zarate ^A, Mario Aquino-Cruz ^B

Resumen— Este trabajo de investigación comparó tres modelos predictivos de Machine Learning: Árbol de Decisiones, Redes Neuronales y Regresión Logística, con el objetivo de identificar cuál fue el más efectivo para optimizar el procesamiento de los test vocacionales, los cuales tradicionalmente se analizaban y calificaban de forma manual, generando demoras y limitaciones en la orientación estudiantil. Esta situación afectaba la calidad del servicio de orientación vocacional, dificultando la emisión de los resultados a los estudiantes. Utilizando el test vocacional IEPPPO, se procesaron las respuestas aplicando la metodología KDD. Los modelos fueron desarrollados y evaluados en Python, utilizando indicadores de rendimiento como exactitud, precisión, recall y F1 score. Los resultados mostraron que el modelo de Árbol de Decisiones fue el más eficaz en la clasificación, alcanzando una exactitud de 86.00%, una precisión del 86.23%, un recall del 86.00% y un F1 score de 86.11%, superando a las Redes Neuronales y la Regresión Logística. Estas conclusiones evidencian el potencial del Árbol de Decisiones para automatizar la orientación vocacional, brindando un apoyo confiable a la Oficina de SOVIO al facilitar el procesamiento inmediato y automatizado de los resultados, permitiendo clasificar los tipos vocacionales según las respuestas de los estudiantes.

Palabras clave: Árbol de decisiones, Orientación vocacional, Redes neuronales, Regresión logística.

Abstract— This research paper compared three predictive machine learning models: Decision Tree, Neural Networks, and Logistic Regression, with the aim of identifying which was the most effective in optimizing the processing of vocational tests, which were traditionally analyzed and graded manually, generating delays and limitations in student guidance. This situation affected the quality of the vocational guidance service, making it difficult to issue results to students. Using the IEPPPO vocational test, the responses were processed applying the KDD methodology. The models were developed and evaluated in Python, using metrics such as accuracy, precision, recall, and F1 score. The results demonstrated that the Decision Tree model was the most effective in classification, achieving an accuracy of 86.00%, a precision of 86.23%, a recall of 86.00%, and an F1 score of 86.11%, outperforming Neural Networks and Logistic Regression. These findings demonstrate the potential of the Decision Tree to automate career guidance, providing reliable support to the SOVIO Office by facilitating immediate and automated processing of results and allowing for the classification of career types based on student responses.

Keywords: Decision tree, Vocational orientation, Neural networks, Logistic regression.



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 22/03/24 - [Indexaciones](#)

Número: doi.org/10.57166/riqchary/V6.n2.2024

Páginas: 41- 48 | Recibido 12/10/2024 ; Aceptado 04/11/2024

doi.org/10.57166/riqchary.v6.n2.2024.128

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0002-2456-150X>
Malu Beatriz Silva-Zarate, Bachiller de la Universidad Nacional Micaela Bastidas de Apurímac, malusilvaz2000@gmail.com.
- B. **ORCID iD** <https://orcid.org/0000-0002-2552-5669>
Mario Aquino-Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac, maquino@unamba.edu.pe

1 INTRODUCCIÓN

En los últimos años, la orientación vocacional ha adquirido una creciente relevancia dentro de los sistemas educativos a nivel global. Este proceso permite a los estudiantes explorar sus intereses, capacidades e intereses personales orientados a tomar decisiones fundamentadas sobre su trayectoria académica y laboral [1]. Sin embargo, persiste una notoria carencia de servicios de orientación vocacional que sean eficientes, accesibles y adaptados a las necesidades individuales de los estudiantes, lo cual repercute negativamente en su desarrollo académico y laboral. Diversos estudios resaltan que una orientación inadecuada puede derivar en una mala elección de carrera, desmotivación e incluso abandono escolar [2].

En el contexto latinoamericano, y específicamente en Perú, esta problemática se agrava por la limitada incorporación de herramientas tecnológicas en los programas de orientación vocacional. A pesar de que existen iniciativas destinadas a este fin, muchas de ellas no disponen de recursos técnicos adecuados que permitan optimizar su funcionamiento frente a una demanda estudiantil cada vez mayor [3]. En la región del Cusco, por ejemplo, el Servicio de Orientación Vocacional e Información Ocupacional (SOVIO), a cargo de la Gerencia Regional de Trabajo y Promoción del Empleo (GRTPE), sigue utilizando procesos manuales basados en papel, Microsoft Excel y Word, lo cual genera demoras significativas en el procesamiento, análisis y entrega de resultados [4].

Frente a esta situación, surge la necesidad de aplicar soluciones tecnológicas que permitan automatizar el procesamiento de los resultados de los test vocacionales y mejorar la precisión de las recomendaciones generadas. En este contexto, el Machine Learning (ML), como subdisciplina de la inteligencia artificial, ofrece un enfoque prometedor al permitir el análisis eficiente de grandes volúmenes de datos y la generación de recomendaciones personalizadas basadas en patrones identificados [5].

La presente investigación abordó esta problemática con el objetivo de evaluar y comparar tres modelos predictivos de Machine learning, Árbol de decisiones, Redes neuronales y Regresión logística, aplicados al análisis automatizado de datos del Inventario de Estilos Personales y Preferencias Ocupacionales (IEPPO). La comparación se realizará en función de métricas clave como la exactitud, la precisión, el recall y F1 Score, con el propósito de determinar cuál de estos modelos ofrece mejores resultados para apoyar el proceso de orientación vocacional en la GRTPE Cusco. La incorporación de estas tecnologías permitirá optimizar el procesamiento de datos y reducir tiempos de respuesta.

1.1 Trabajos Relacionados

En diversos estudios, se ha explorado la utilización de modelos predictivos y técnicas de Machine learning para optimizar la toma de decisiones académicas y vocacionales. Un estudio desarrolló un modelo predictivo para la elección de una carrera profesional para estudiantes de una universidad privada en Arequipa, en el cual, valió un modelo de regresión logística, como una herramienta de ayuda para, también revelando que factores como género y la edad de los estudiantes influyen en las respuestas y elección de su carrera [6]. Otro trabajo realizó la implementación de un sistema de predicción del rendimiento académico de estudiantes de ingeniería, en el cual utilizaron técnicas de minería de datos, utilizando la metodología KDD y algoritmos de clasificación como REPTree, Vote, SMOreg, RandomForest LWL, Bagging, Kstar, ZeroR, M5P, RandomTree e IBK, siendo el algoritmo Kstar el más eficiente [7].

Se evaluaron diversos modelos de machine learning para determinar su efectividad en la predicción de la deserción en el ámbito universitario, teniendo como métricas la medición de la precisión, la exactitud y exhaustividad, identificando que la Regresión Logística fue la técnica más efectiva en esta tarea [8]. Otro estudio ha sido desarrollado por un modelo de predicción para los resultados académicos de los estudiantes universitarios, utilizando métodos KDD y evaluando la precisión, sensibilidad y especificidad y aplicando el uso de algoritmos como SVM y KNN mostró que SVM fue el más eficaz, logrando altos niveles de precisión y mayor asertividad [9] [10]. Así también el estudio que desarrolló un modelo predictivo que apoye el seguimiento académico de estudiantes, realizó la comparación de diversos modelos de redes neuronales alcanzando una precisión de 98.97%, demostrando ser una herramienta eficaz para su objetivo [11].

A nivel internacional, un estudio desarrolló un agente inteligente basado en machine learning para predecir la elección de carrera de los estudiantes, los resultados mostraron que la técnica de k-vecinos cercanos fue la más eficaz para la predicción [12]. Además, en otro trabajo se compararon algoritmos de aprendizaje supervisado y no supervisado para sugerencias vocacionales, concluyendo que KNN y SVM fueron los más efectivos en la clasificación de vocaciones, alcanzando tasas de precisión del 90% [13]. Finalmente, en una investigación se implementó un sistema web utilizando técnicas de machine learning para mejorar la orientación vocacional de estudiantes de secundaria, utilizando la metodología KDD con resultados satisfactorios en la recomendación de carreras profesionales [14].

2 MATERIALES Y METODOLOGÍA

2.1 Materiales

- Datos de estudiantes: 3770 evaluaciones de estudiantes de 4° y 5° de secundaria de colegios públicos y privados en la Región Cusco, obtenidos mediante el Inventario de Estilos Personales y Preferencias Ocupacionales (IEPPO).
- Equipos de cómputo: Computadoras con sistema operativo Windows 64 bits, procesador Core i5 o superior, 8GB de RAM, y 512GB de almacenamiento.
- Software:
 - Lenguaje y librerías para análisis y entrenamiento: Python, empleando librerías como pandas y numpy para tratar los datos, matplotlib y seaborn para generar gráficos, y scikit-learn para construir y evaluar los modelos.
 - Entorno de desarrollo y pruebas: Google Colab, para la programación en la nube, entrenamiento de modelos y visualización de resultados. Microsoft Excel, para el procesamiento de datos.
 - Frameworks y herramientas para el desarrollo web: Frontend: React, con herramientas complementarias. Backend (servidores y procesamiento): NestJS y FastAPI. Base de datos: PostgreSQL.
- Conexión a internet: Para trabajar en la nube (Google Colab), desplegar servidores locales y consumir las APIs entre backend y frontend.
- Cuestionario IEPPO: Instrumento para obtener dato sobre las preferencias vocacionales de los estudiantes.

2.2 Tipo y nivel de investigación

Este estudio se clasifica como una investigación aplicada con nivel descriptivo, para Hernandez Sampieri y otros este se distingue por centrarse en la resolución de problemas concretos y aplicados, haciendo uso de conocimientos y teorías ya establecidas para generar resultados con aplicaciones directas en contextos reales, y descriptivo porque medirá y recogerá información de manera independiente o conjunta sobre los conceptos o las variables a las que se refieren [15].

2.3 Diseño de la investigación

El diseño es no experimental de corte transversal, observando fenómenos en su entorno natural sin manipular variables; en este estudio no se manipularon variables de forma directa, los datos se recopilaban tal como existen en la realidad, el enfoque fue analizar esos datos utilizando modelos predictivos, sin alterar el entorno ni aplicar intervenciones experimentales.

2.4 Población y muestra

La población objetivo incluye 3770 evaluaciones de estudiantes de cuarto y quinto año de secundaria de 38 colegios

tanto de instituciones públicas y privadas de la región Cusco, así también la muestra incluye el total de la población.

2.5 Proceso KDD

En la fig.1 se muestran las etapas del Proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD) que se utilizó para la investigación.

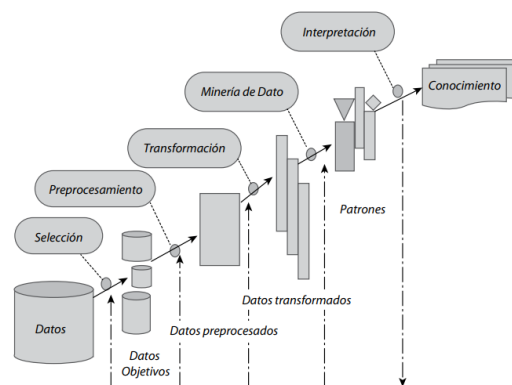


Fig. 1. Etapas del proceso KDD

2.6 Procedimiento

Primero se realizó la recopilación de datos de 3,770 estudiantes de 4° y 5° año de secundaria de instituciones públicas y privadas de región Cusco mediante la aplicación del test vocacional IEPPO, realizado entre agosto y diciembre de 2023, de forma presencial y virtual. La recolección fue realizada por la Gerencia Regional de Trabajo y Promoción del Empleo Cusco, a través de la Oficina SOVIO, conforme a la Directiva SOVIO N.° 001-2012-MTPE y la Resolución Ministerial N.° 116-2009-TR, que respaldan este servicio público gratuito dirigido a instituciones que no cuentan con profesionales en psicología. Los datos, de naturaleza sensible y privada, fueron almacenados en archivos Excel por colegio para su posterior clasificación en los tipos vocacionales.

Segundo, las respuestas de los test vocacionales IEPPO están divididas en tres partes: 33 preguntas en estilos personales, 47 preguntas en actividades de preferencia y 38 preguntas en percepción de habilidad, cada división fue respondida con respuesta de:

- No se parece a mí.
- Se parece a mí.
- Me interesa.
- Me interesa poco o nada.
- Soy hábil.
- No soy hábil.

Los datos fueron transformados y preparados para su análisis, lo que incluyó la consolidación y estructuración de las variables que fueron utilizadas.

Tercero, se realizó el procesamiento de las respuestas del test vocacional, obteniendo su clasificación según los 7 tipos vocacionales que son: liderazgo, técnico mecánico, social, organizado, artístico, emprendimiento e investigación. Se

realizó el proceso de selección y limpieza de la base de datos, garantizando la calidad; este proceso incluyó también la eliminación de datos duplicados, eliminación de test incompletos y corrección de errores, quedando un total de 2751 datos.

Cuarto, los datos finalmente obtenidos fueron estandarizadas para facilitar su uso en los modelos predictivos, se dividieron en tres segmentos, 70% para el entrenamiento, obteniendo un total de 1926 test vocacionales, 20% para la validación, obteniendo un total de 550 test vocacionales y 10% para el testeo, obteniendo 275 test vocacionales.

Quinto, se realizó el entrenamiento con los datos, utilizando la herramienta Python; se realizó en una computadora de 64 bit con sistema operativo Windows, se desarrolló con los tres modelos predictivos de machine learning (Árboles de Decisión, Redes Neuronales y Regresión Logística), los modelos fueron entrenados con el conjunto de entrenamiento y validados con conjunto de validación.

La arquitectura utilizada para el modelo de predicción basado en Redes neuronales se puede observar en la fig.2.

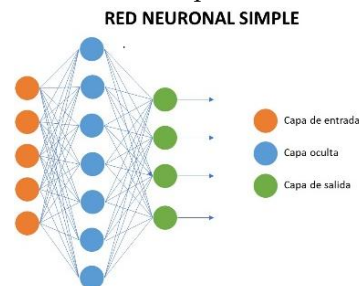


Fig. 2. Arquitectura utilizada para la Red Neuronal (Multilayer Perceptron, MLP) [16]

- Capa de Entrada (Input Layer): El número de neuronas en la capa de entrada se definió según la cantidad de características en el conjunto de datos, las cuales fueron 118, representando cada una de las preguntas respondidas por los estudiantes.
- Capa Oculta (Hidden Layer): Se configuró una capa oculta con 5 neuronas mediante el parámetro `hidden_layer_sizes`, buscando un modelo más simple que reduzca el riesgo de sobreajuste y optimice el tiempo de entrenamiento. Cada neurona aplica una transformación no lineal a los datos de entrada, permitiendo que la red capte relaciones complejas entre las variables.
- Capa de Salida (Output Layer): Esta capa contiene 7 neuronas, correspondiente a los 7 tipos vocacionales identificados. Cada neurona representa una clase, y el modelo predice la clase con la mayor probabilidad para cada estudiante.
- Se utilizó `max_iter` con valor de 1000 para asegurar que el algoritmo tenga suficientes iteraciones para converger, `Alpha` con valor de 1.0 como término de regularización para evitar el sobreajuste, y `random_state=42` para garantizar la reproducibilidad del entrenamiento. Estos valores fueron seleccionados para mantener un balance entre la precisión del modelo y el uso eficiente de los recursos computacionales.

En la fig.3 se muestra la arquitectura que se utilizó para el modelo de predicción de Árbol de decisiones.

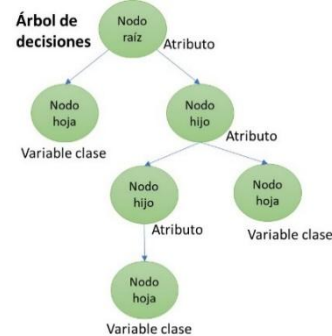


Fig. 3. Arquitectura utilizada para Árbol de decisiones

- Cada nodo interno representa una decisión basada en una característica de los datos (un nodo a una pregunta).
- Los nodos hoja representan las decisiones finales, que en este caso son los tipos vocacionales.
- Cada rama del árbol indica el resultado de una decisión, lo que lleva a otra pregunta.
- En la codificación realizada, el árbol creció automáticamente al entrenar el modelo, con el parámetro `random_state=42` para asegurar la reproducibilidad del resultado, se utilizó el criterio "gini" (impureza de Gini) para decidir las divisiones en el árbol.

En la fig.4 se muestra la arquitectura que se utilizó para el modelo de predicción de Regresión logística multiclase.

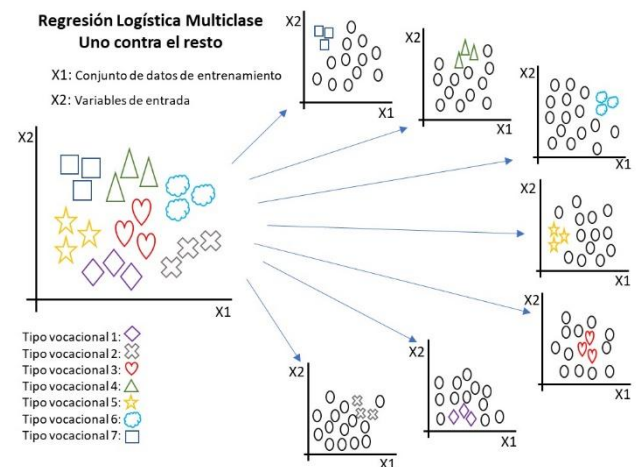


Fig. 4. Arquitectura utilizada para Regresión logística.

- Capa de entrada: Cada columna del test vocacional se convirtió en una característica para el modelo, los datos categóricos fueron convertidos en numéricos.
- Procesamiento - Uno contra el resto: Para cada tipo vocacional, se generó un clasificador que intentó predecir si una observación pertenece a esa clase o no; se realizó la aplicación de la función logística y ajuste de pesos.
- Capa de salida: Se seleccionó la clase final con la mayor

probabilidad de pertenecer a una de los 7 tipos vocacionales entre todas las predicciones de los clasificadores.

- d) Para la codificación se aseguró la convergencia y facilidad para reproducir los resultados, utilizando `max_iter`: incrementado a 1000 para permitir la convergencia del modelo teniendo en cuenta en tamaño del conjunto de datos y también el `random_state` configurado con un valor de 42 para la reproducibilidad.

2.7 Medición de desempeño

Para esta sección se ha utilizado indicadores de rendimiento para evaluar la eficacia de nuestros modelos predictivos, incluyendo la exactitud (accuracy), la precisión (precision), el recall (sensibilidad o recuperación) y F1 Score [17]. Las ecuaciones empleadas para calcular cada métrica son las siguientes:

- Exactitud (Accuracy): Esta métrica mide la proporción total de aciertos del modelo, incluye tanto las predicciones positivas correctas como las negativas correctas con respecto al número total de predicciones.

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precisión (Precision): La precisión refleja cuántas de las predicciones positivas realizadas por el modelo fueron realmente correctas.

$$Precision = \frac{TP}{TP + FP}$$

- Recall (Sensibilidad): Esta métrica refleja la habilidad del modelo para reconocer la totalidad de los casos positivos.

$$Recall = \frac{TP}{TP + FN}$$

En estas ecuaciones, TP, TN, FP y FN van a representar las cantidades de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

- F1 Score: Mide el balance entre la precisión y el recall, esta métrica utilizamos por contener datos desbalanceados.

$$F1\ Score = 2 * \frac{Precisión * Recall}{Precisión + Recall}$$

Estas métricas fueron fundamentales para la evaluación del rendimiento de los modelos predictivos como: Árbol de Decisión, Redes Neuronales y Regresión Logística, ya que cada uno de estos algoritmos tuvo un comportamiento diferente en términos de exactitud, precisión, recall y F1 score. El análisis conjunto de estas métricas nos permitieron seleccionar el modelo que mejor se ajusta a nuestros objetivos.

3 RESULTADOS

3.1 Evaluación de Modelos Predictivos

En este estudio se ha llevado a cabo un proceso de entrenamiento y evaluación de tres modelos predictivos: Árbol de Decisiones, Regresión Logística y Redes Neuronales, con el objetivo de determinar cuál es más efectivo para clasificar las respuestas de los estudiantes de secundaria en los 7 tipos vocacionales. Para ello, se utilizó una base de datos de 2751 evaluaciones, se fue procesando las respuestas textuales mediante codificación para convertirlas en datos numéricos, lo que permitió su posterior análisis con los modelos de clasificación; se codificaron tanto las respuestas a las preguntas del test como la variable objetivo (el tipo vocacional), utilizando `LabelEncoder` para convertir las respuestas en cadenas a valores numéricos. A continuación, se muestran los resultados obtenidos para cada uno de ellos.

3.1.1 Resultados del Árbol de Decisiones

El modelo de Árbol de Decisiones es un modelo basado en la creación de reglas lógicas [18], este modelo divide los datos en subconjuntos más pequeños, lo que le permitió manejar las relaciones complejas y no lineales. Los resultados obtenidos para la validación de este modelo fueron altos, los detalles de su desempeño se observan en la siguiente matriz de confusión en la fig.5 y su curva de aprendizaje del modelo se puede visualizar en la fig.6 en el cual se aprecia cómo evolucionaron las métricas de precisión tanto en el conjunto de entrenamiento como en de validación.

Matriz de Confusión - Árbol de Decisiones							
Verdadero	6	0	0	0	0	1	1
	0	14	0	1	0	0	0
	0	0	3	0	0	0	3
	0	0	1	71	4	3	3
	0	2	0	0	54	0	0
	0	0	0	1	3	36	7
	0	0	0	5	5	2	74
	Predicción						

Fig. 5. Matriz de confusión de modelo de predicción de Árbol de Decisiones.

- Exactitud: Con un resultado porcentual de 86.00%, el Árbol de Decisiones logró clasificar correctamente los casos en el conjunto de prueba, lo que significa que la mayoría de los estudiantes fueron clasificados a su tipo vocacional correcto. Este nivel de exactitud indica que el modelo pudo identificar patrones muy claros y consistentes entre las respuestas y los tipos vocacionales.
- Precisión: Obteniendo un resultado porcentual de 86.23%, esta métrica midió qué tan efectivo es el modelo al predecir cada tipo vocacional sin producir demasiados falsos positivos (errores de clasificación en los que el modelo asigna incorrectamente una clase).
- Recall: Con un resultado porcentual de 86.00%, esta métrica midió la capacidad del modelo para identificar correctamente a los estudiantes en su tipo vocacional, sin dejar de clasificar correctamente a los que realmente pertenecen a cada categoría.

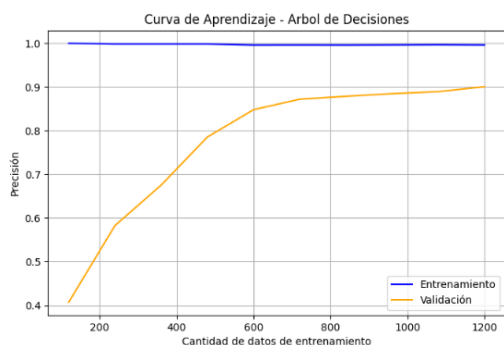


Fig. 6. Curva de aprendizaje del modelo de predicción Árbol de decisiones.

3.1.2 Resultados de la Regresión Logística

La Regresión Logística es un modelo lineal que intentó encontrar la mejor línea divisoria entre las clases en función de las características de entrada [19]. A pesar de ser un modelo sencillo y eficiente, los resultados de validación no alcanzaron el rendimiento del Árbol de Decisiones, su desempeño se observa en la siguiente matriz de confusión en la fig.7.

Matriz de Confusión - Regresión Logística

	1	0	1	3	0	0	3
1	6	0	0	2	0	0	0
0	0	13	0	0	1	0	1
0	0	0	3	0	0	3	0
0	0	0	0	48	9	9	16
0	0	0	0	3	37	2	14
0	0	0	0	10	1	28	8
0	0	0	0	15	5	8	58

Verdadero (eje Y) vs Predicción (eje X)

Fig. 7. Matriz de confusión de modelo de predicción de Regresión Logística.

- Exactitud: Se obtuvo un resultado porcentual de 64.33%, el modelo de Regresión Logística tuvo un desempeño significativamente inferior en comparación con el Árbol de Decisiones. Este resultado sugiere que las relaciones entre las respuestas del test vocacional y los tipos vocacionales no son completamente lineales, lo que dificulta a la Regresión Logística capturar esas relaciones.
- Precisión: Con un resultado porcentual de 65.43% muestra que el modelo produjo un número de falsos positivos, lo que indica que a menudo clasificó erróneamente a los estudiantes en sus tipos vocacionales incorrectos. Este valor también refleja la limitación del modelo para ajustar los datos, ya que no puede manejar de manera efectiva patrones no lineales.
- Recall: Se obtuvo un resultado porcentual de 64.33%, el cual nos indica que el modelo tuvo problemas para identificar correctamente todas las instancias de cada tipo vocacional. Esto se traduce en que muchas categorías no fueron completamente recuperadas, lo que significa que

el modelo no fue capaz de predecir correctamente una a los estudiantes para algunos tipos vocacionales.

3.1.3 Resultados de las Redes Neuronales

Las Redes Neuronales son modelos altamente flexibles [20], su función fue capturar relaciones no lineales y complejas entre los datos. En este caso, el rendimiento en el resultado de validación fue cercano al modelo de Regresión Logística, su matriz de confusión se puede ver en la fig.8.

Matriz de Confusión - Red Neuronal

	1	0	1	3	0	0	3
1	0	12	0	0	0	1	2
0	0	0	0	1	4	1	0
0	2	0	56	9	11	4	
0	0	0	2	40	4	10	
0	0	0	11	6	20	10	
0	1	0	11	3	6	65	

Verdadero (eje Y) vs Predicción (eje X)

Fig. 8. Matriz de confusión de modelo de predicción de Red Neuronal.

- Exactitud: Con un resultado porcentual de 64.67%, mejorando en un pequeño porcentaje a la Regresión Logística, pero aún lejos del Árbol de Decisiones. Esto sugiere que la Red Neuronal fue capaz de identificar algunas interacciones complejas en los datos, pero no de manera óptima.
- Precisión: Obteniendo un resultado porcentual de 64.04%, la Red Neuronal mostró una menor capacidad para reducir los falsos positivos en comparación con la Regresión Logística. Hubo errores en la clasificación, lo que implica que, sigue siendo menos eficiente que el Árbol de Decisiones.
- Recall: Con un resultado porcentual de 64.67%, indica que la Red Neuronal fue capaz de recuperar un número de instancias correctas en comparación con la Regresión Logística, pero tampoco logró alcanzar la perfección del Árbol de Decisiones.

3.1.4 Comparación de los Modelos

En la siguiente Tabla 1 se presentan los resultados de validación comparativos de los tres modelos en cuanto a exactitud, precisión, recall y F1 Score:

TABLA 1
Resultados comparativos de la validación en exactitud, precisión, recall, F1 Score para los tres modelos predictivos.

Modelo	Exactitud	Precisión	Recall	F1 Score
Árbol de Decisiones	0.86	0.86	0.86	0.861
Regresión Logística	0.64	0.65	0.64	0.649
Redes Neuronales	0.64	0.64	0.64	0.644

El Árbol de Decisiones mostró un rendimiento alto, lo que sugiere que las relaciones entre las respuestas del test vocacional y los tipos vocacionales son altamente discernibles por un modelo que utiliza reglas de decisión. Este resultado indica que el Árbol de Decisiones es una herramienta efectiva para este tipo de clasificación y, dado su rendimiento, podría ser directamente implementado en procesos de orientación vocacional con un buen nivel de confianza.

El desempeño de la Regresión Logística revela que este tipo de modelo no es el más adecuado para este problema, debido a la naturaleza compleja y no lineal de los datos. Aunque es un algoritmo fácil de implementar y eficiente computacionalmente, no fue capaz de capturar la riqueza de las relaciones entre las variables predictoras y la clasificación del tipo vocacional.

Las Redes Neuronales demostraron un rendimiento cercano a la Regresión Logística debido a su capacidad para capturar relaciones no lineales. Sin embargo, su desempeño aún fue considerablemente inferior al del Árbol de Decisiones.

En el caso del Árbol de Decisión, su alto rendimiento tanto en entrenamiento como en prueba cercano al 86% sugiere que el sobreajuste no fue significativo, puesto que se dividió en conjunto de prueba, de entrenamiento y testeo.

Se realizó una página web donde se integró el modelo de árbol de decisiones entrenado, permitiendo a los estudiantes ingresar sus respuestas como en la fig.9 y obtener su tipo vocacional de forma inmediata como se muestra en la fig.10. Esta solución tecnológica representa una mejora significativa frente al procesamiento tradicional de datos, que solía ser manual, lento y propenso a errores. Gracias a la automatización, se agiliza el proceso de orientación vocacional, se reduce el tiempo de análisis y se garantiza mayor precisión en los resultados, beneficiando así a la comunidad estudiantil con una herramienta moderna, eficiente y fácil de usar.

Fig. 9. Ingreso de respuestas del test IEPPO en la página web.

Fig. 10. Resultado de encuesta, clasificando su tipo vocacional.

4 CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

Como resultado de esta investigación, se concluyó que el modelo predictivo de Árbol de Decisiones fue el más efectivo para mejorar el servicio de orientación vocacional dirigido a los estudiantes de secundaria en la GRTPE Cusco, superando a las Redes Neuronales y a la Regresión Logística, lo que demuestra su capacidad para clasificar con mayor precisión los tipos vocacionales a partir de las respuestas del test IEPPO, optimizando así el procesamiento y análisis de los datos recolectados.

El análisis de la precisión mostró que el Árbol de decisiones es el modelo más preciso, con un valor porcentual en su validación de 86.23%. Esto implica que el modelo comete muy pocos errores al predecir el tipo vocacional de los estudiantes, reduciendo los falsos positivos y proporcionando predicciones confiables.

En términos de exactitud, que mide el porcentaje total de predicciones correctas, el Árbol de decisiones también superó a los demás modelos, con un valor porcentual en la validación de 86.00%; las Redes neuronales alcanzaron una exactitud de 64.67%, mientras que la Regresión logística obtuvo 64.33%.

En relación con el recall, el Árbol de Decisiones demostró nuevamente su superioridad con un valor porcentual en su validación de 86.00%, midiendo correctamente a los estudiantes que pertenecen a cada tipo vocacional.

4.2 Recomendaciones

Aunque el Árbol de Decisiones ha mostrado un rendimiento excelente en este estudio, se recomienda realizar un monitoreo continuo de su desempeño, puesto que los datos y las necesidades de los estudiantes pueden evolucionar, por lo que es importante que el modelo sea evaluado periódicamente para garantizar que continúe proporcionando resultados precisos y útiles.

Si bien el estudio se centró en tres modelos de predicción, resulta recomendable analizar otros enfoques de machine learning, como SVM y Random Forest, para determinar si ofrecen mayor precisión o estabilidad frente a diversas condiciones.

AGRADECIMIENTOS

Agradezco a la Gerencia Regional de Trabajo y Promoción del Empleo de Cusco (GRTPE Cusco) por su valioso apoyo en la recolección de datos a través de la aplicación de test vocacionales a los estudiantes de secundaria de la región Cusco, su colaboración fue fundamental para la conformación de la base de datos utilizada en el presente estudio, la cual permitió llevar a cabo el entrenamiento y evaluación de los tres modelos predictivos de Machine learning.

REFERENCIAS

- [1] T. De León Mendoza y R. Rodríguez Martínez, «El efecto de la orientación vocacional en la elección de carrera,» *Revista Mexicana de Orientación Educativa*, vol. V, n° 13, p. 7, 2008.
- [2] X. F. Erazo Guerra y E. d. R. Rosero Morales, «Orientación vocacional y su influencia en la deserción universitaria,» *Horizontes Revista de Investigación en Ciencias de la Educación*, vol. V, n° 18, p. 16, 2021. <https://doi.org/10.33996/revistahorizontes.v5i18.198>
- [3] OCDE, «Education at a Glance 2019 OECD Indicators,» Paris, 2019.
- [4] Ministerio de Trabajo y Promoción del Empleo Cusco, «Ministerio de Trabajo y Promoción del Empleo Cusco,» 21 Noviembre 2016. [En línea]. Available: <https://www2.trabajo.gob.pe/el-ministerio-2/sector-empleo/directorio-form-cap-lab/sobre-el-sovio/>. [Último acceso: 10 Octubre 2024].
- [5] C. OVALLE PAULINO, «Modelo predictivo basado en Machine Learning para la Cadena de Suministro y su influencia en la gestión logística de una empresa de venta de autos,» *Revista de la ACM*, p. 15, Abril 2022.
- [6] E. Franco Delgado y M. Polanco Valenzuela, «Elección de la carrera profesional: modelo predictivo en estudiantes de una universidad privada de Arequipa (Perú),» *Revista de Investigación en Psicología*, vol. XXVI, n° 2, p. 27, 2023. <https://doi.org/10.15381/rinvp.v26i2.25325>
- [7] N. Luna Reyes, «Implementación de un sistema de predicción del rendimiento académico de los estudiantes de Ingeniería de Sistemas de la Universidad Jose Maria Arguedas utilizando tecnicas de mineria de datos para la adecuada toma de decisiones,» *UNAJMA*, Andahuaylas, 2020.
- [8] A. E. Aco Tito, B. O. Hanco Condori y Y. Pérez Vera, «Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria,» *Revista Ibérica de Sistemas y Tecnologías de Información*, Arequipa, 2023.
- [9] J. D. García Dionisio, «Machine learning para predecir el rendimiento académico de los estudiantes universitarios,» Lima, 2021.
- [10] D. I. Candia Oviedo, «Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático,» Cusco, 2019.
- [11] H. Caselli Gismondi, «Modelo predictivo basado en Machine Learning como soporte para el seguimiento académico del estudiante universitario,» Chimbote, 2021.
- [12] D. Quimbayo, «Agente Inteligente para predecir la elección de carrera de estudiantes de bachiller,» Colombia, 2024.
- [13] B. E. Cisneros Bravo, «Algoritmo de Machine Learning y uso de propiedades semánticas para la identificación y sugerencia vocacional,» Estado de Mexico, 2023.
- [14] R. A. Leyva Osorio y K. A. Medina Arango, «Sistema de recomendación para vocacion profesional, aplicado a la carrera de ingeniería de sistemas ofrecida en la Universidad de Cundimarca, extensión facatativa,» Cundimarca, 2019.
- [15] R. Hernández Sampieri, C. Fernández Collado y M. d. P. Baptista Lucio, *Metodología de la Investigación*, vol. VI, Mexico: INTERAMERICANA EDITORES, 2014.
- [16] S. Savalia y V. Emamian, «Clasificación de arritmias cardíacas mediante perceptrones multicapa y redes neuronales convolucionales,» *Bioingeniería*, vol. V, n° 2, p. 12, 2018.
- [17] Y. Pérez Vera, A. E. Aco Tito y B. O. Hanco Condori, «Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria,» *Revista Ibérica de Sistemas y Tecnologías de Información*, n° 51, p. 15, 30 Setiembre 2023.
- [18] C. Arana, «MODELOS DE APRENDIZAJE AUTOMÁTICO MEDIANTE ÁRBOLES DE DECISIÓN,» Buenos Aires, 2021.
- [19] M. Domínguez, «Regresión Logística y Técnicas de Aprendizaje. Aplicaciones,» España, 2021.
- [20] C. Ruiz y M. Basualdo, «Redes Neuronales: Conceptos Básicos y Aplicaciones,» Buenos Aires, 2001.

BIOGRAFÍA

Malu Beatriz Silva Zarate, bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Mario Aquino Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac Perú, MSc. en Informática, investigador en las áreas de informática educativa, IoT, inteligencia artificial y ciberseguridad.

Determinación de la contaminación por metales pesados en los sedimentos superficiales de la subcuenca del río Antabamba en un período de mediano plazo

Determination of heavy metal contamination in the surface sediments of the Antabamba River sub-basin in a medium-term period

Darío Sánchez-Castillo^A, Pablo Zuloaga-Candia^B y Karina Gamarra-Peralta^C

Resumen— La investigación aborda el grave problema ambiental de la contaminación por metales pesados en los sedimentos superficiales de la subcuenca del río Antabamba, causado por actividades mineras, antropogénicas o procesos naturales. El objetivo principal es determinar la influencia de la contaminación por metales pesados en los sedimentos superficiales de la subcuenca del río Antabamba permite predecir la calidad ambiental durante un período de mediano plazo aplicando aprendizaje automático, analizando concentraciones de metales como del cromo, y considerando los estándares de calidad vigentes. Con un enfoque aplicativo y diseño experimental correlacional, se recolectaron y analizaron 28 muestras de sedimentos en 14 estaciones. El uso de algoritmos de aprendizaje automático permite predecir los límites permisibles de contaminación con el 75% de precisión, proporcionando datos clave para la gestión ambiental por meses. Los resultados subrayan la necesidad de implementar medidas de control y el cumplimiento normativo para reducir los impactos a corto y mediano plazo, y destacan la importancia de aplicar nuevos métodos para optimizar los pronósticos y futuras investigaciones.

Palabras clave: predicción, contaminación, metales pesados, sedimentos superficiales

Abstract— The research addresses the serious environmental problem of heavy metal contamination in the surface sediments of the Antabamba River sub-basin, caused by mining, anthropogenic activities or natural processes. The main objective is to determine the influence of heavy metal contamination in the surface sediments of the Antabamba River sub-basin to predict environmental quality over a medium-term period by applying machine learning, analyzing concentrations of metals such as chromium, and considering the standards. current quality standards. With an applicative approach and correlational experimental design, 28 sediment samples were collected and analyzed at 14 stations. The use of machine learning algorithms makes it possible to predict permissible pollution limits with 75% accuracy, providing key data for environmental management and mitigation strategies for months. The results underline the need to implement regulatory control and compliance to reduce short- and medium-term impacts, and highlight the importance of applying new methods to optimize forecasts and future research

Keywords: heavy metals, prediction, pollution, surface sediments.

1 INTRODUCCIÓN

La contaminación de los sedimentos superficiales por metales pesados (CSSMP), es un desafío crítico que impacta la calidad del agua, el suelo y la salud pública, así como el desarrollo socioeconómico de las comunidades afectadas [1]. Esta problemática es particularmente grave en los Sedimentos Superficiales (SS), que están contaminadas por actividades antropogénicas o por relaves mineros con altas concentraciones de metales pesados [2]. La complejidad de la contaminación por Metales Pesados (MP) radica en la dificultad de determinar sus concentracio-

nes exactas debido al comportamiento hidrofóbico [3] y la posibilidad de que estos contaminantes se vuelvan biodisponibles bajo ciertas condiciones [4];[1]. La escorrentía superficial durante las lluvias contribuye a la elevación de estos metales en los SS [5]. Esta situación es comparable a la crisis de CSSMP en países en desarrollo. Los MP, como mercurio (Hg), plomo (Pb) y cadmio (Cd), presentan riesgos significativos para la biota acuática y la salud humana, especialmente en áreas donde la minería es predominante [7]. Los SS, son sumideros de sustancias desechadas, donde los metales pesados tienden a bioacumularse, convirtiéndose en amenazas persistentes [8]. La falta de metodologías eficaces para evaluar la calidad de



Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x
Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 10/11/24 - Indexaciones
Número: doi.org/10.57166/riqchary/v6.n2.2024
Pág.: 49 - 54 | Recibido 28/10/2024 ; Aceptado 31/10/2024



doi.org/10.57166/riqchary.v6.n2.2024.127

Autores:

- A. ORCID iD <https://orcid.org/0000-0001-6654-6565>
Darío Dante Sánchez Castillo, trabaja en la Escuela Profesional de Ingeniería de Minas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, dsanchez@unamba.edu.pe
- B. ORCID iD <https://orcid.org/0000-0002-3678-8894>
Pablo Ruben Zuloaga Candia, trabaja en la Escuela Profesional de Ingeniería de Minas de la Universidad Nacional Micaela Bastidas de Apurímac - Perú, pzuaga@unamba.edu.pe
- C. ORCID iD <https://orcid.org/0000-0002-7702-9923>
Karina Gamarra Peralta, trabaja en la Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac - Perú, kgamarra@unamba.edu.pe

los SS, es un obstáculo en la identificación de contaminantes en ríos como el de Antabamba, donde la contaminación es creciente [9].

se requiere la implementación de modelos predictivos basados en el aprendizaje automático para mejorar la precisión en la evaluación de la calidad de los sedimentos [9]. Estos modelos son esenciales para identificar las áreas más susceptibles a la contaminación y optimizar los controles ambientales [3].

La Fiscalización Ambiental en Apurímac, reporta 343 casos en el 2022, donde los residuos sólidos contribuyen a la contaminación, seguidos por hidrocarburos y la minería [11]. Los datos indican que, la adecuada información prevendría muchos casos de contaminación. La escasez de datos específicos limita la evaluación precisa de la calidad de los sedimentos [12]; [13]. Por lo tanto, es crucial adoptar enfoques avanzados en la gestión sostenible de los recursos naturales para mitigar el impacto de la contaminación en la región [9].

En la región de Apurímac, las actividades mineras han comprometido la calidad del suelo, incrementando la concentración de estos MP y amenaza la biodiversidad [9]. Los ríos de esta región son vitales para la agricultura y el monitoreo es esencial para evitar la contaminación [2].

TABLA 1.

Coordenadas y Zona de muestreo

Referencia	Coordenadas de Muestreo			Altitud (msnm)
	Zona	Este	Norte	
Puente Huancaray	18 L	0713693	8423521	2797
Puente Pichirhua	18 L	0698026	8437463	2482



Fig.1. Río Antabamba y zonas de alta mineralización, extraído de Earth Google

2 METODOLOGÍA

2.1 Ámbito de estudio

El río Antabamba se encuentra en la región de Apurímac, Perú, y la subcuenca presenta una geografía diversa con varias zonas que tienen diferentes características ambientales que son influenciadas por actividades de origen antrópico, y naturales que ocurren en ambas márgenes [11], [14].

2.2 Tipo y nivel de investigación

La investigación es aplicada y correlacional [15], centrada en la relación entre la concentración de MP en los sedimentos del río Antabamba, comparando estos datos con estándares d El estudio adopta un enfoque retrospectivo y longitudinal. El

Los métodos tradicionales de evaluación son insuficientes, y calidad ambiental (ECA).

diseño de investigación es cuasiexperimental, con fase transversal y longitudinal [15]. La fase transversal involucra la recolección de datos en dos periodos estacionales mediante 28 muestras. La fase longitudinal aplica modelos de pronóstico en Python para proyectar la evolución de la contaminación por MP a mediano plazo.

2.3 Procedimiento

Para determinar la CSSMP del río Antabamba en un período de mediano plazo, se realizará lo siguiente:

- Determinar y analizar las concentraciones de los MP y el ECA de suelo, de acuerdo a la Normativa Nacional D. S. 011-2017-MINAM (Standard Methods, 2021; Perú, 2010), que establece los valores máximos permitidos para garantizar la conservación de la calidad ambiental [10]; [11], se analizan los datos clave que permiten caracterizar la calidad ambiental, realizar las correlaciones y se precisa la calidad ambiental.
- Precisar la calidad ambiental de la CSSMP del río Antabamba durante un periodo a mediano plazo aplicando algoritmos de aprendizaje automático.
- Precisar la calidad ambiental y el comportamiento de la CSSMP en el tiempo a mediano plazo, aplicando un modelo predictivo.

2.4 Población y muestra

La población, incluye todos los sedimentos superficiales de la subcuenca del río Antabamba, desde su origen hasta la confluencia con la subcuenca Chalhuanca.

La muestra, representativa de los SS en la subcuenca del río Antabamba es estratificada, se definieron 14 puntos de muestreo seleccionados estratégicamente a lo largo de 25 km. Considerando proximidad a proyectos mineros, diversidad geográfica y geológica, factores estacionales y accesibilidad para recolectar las muestras, con el fin de capturar la variabilidad espacial y temporal en la concentración de MP. Cada estrato representa una zona relevante del río en relación con fuentes de contaminación.

Precisar la calidad ambiental de la CSSMP del río Antabamba durante un periodo a mediano plazo aplicando algoritmos de aprendizaje automático.

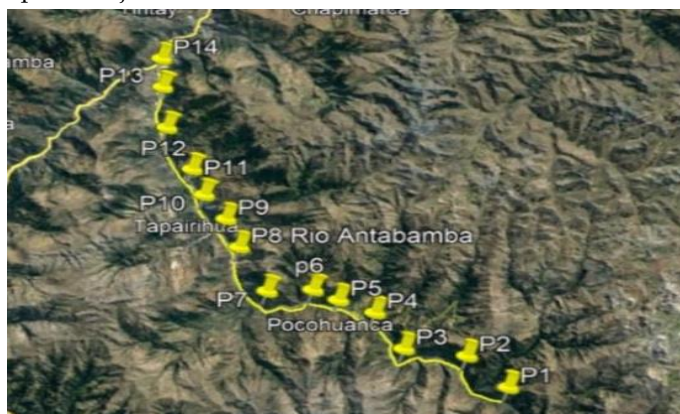


Fig. 2. Puntos de muestreo de sedimentos, extraído de Earth Google

La Fig. 2. Muestra los puntos de muestreo de sedimentosSe

establecieron modelos predictivos para precisar Iso ECAs y el comportamiento de la CSSMP en el tiempo, siendo estas recolectas en etpas de secas y lluvias (Ver Fig. 3).



Fig. 3. Recolección de muestras de sedimentos superficiales

2.5 Materiales e instrumentos

TABLA 2
Materiales y métodos

Pasos	Materiales-Equipos	Métodos
Recolección de muestras de SS.	Baldes, frascos de polietileno y Pyrex para SS, probetas y otros. GPS, balanza digital, malla N°20.	Muestreo estratificado en 14 puntos
Análisis de SS.	Espectrofotómetro de masas con plasma acoplado por inductividad (ICP-MS). Normas Method 3051 y 6020A.	Concentraciones de MP en laboratorio. Análisis)
	-Multiparámetro.	Normativa: ECAs
Predicción	-Laptop: Bases de datos con parámetros de ECAs, Software especializado (Python), algoritmos de ML, Herramientas de visualización y Técnicas de validación	Procesamiento y limpieza de datos Entrenamiento y validación de modelos. Predicción de ECAs durante un mediano plazo. Evaluación del rendimiento.

3 IMPLEMENTACIÓN

3.1 Origen de Datos y fuentes

Los SS, fueron analizados en el laboratorio de investigación de la Universidad Nacional de San Agustín de Arequipa. La técnica de análisis: ICP-MS. Los resultados del análisis presentaron diversas concentraciones de MP como son de: K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Br, Rb, Sr, Y, Zr, Nb y Ba.

Los factores fisicoquímicos analizados en laboratorio de la EPIM-UNAMBA, fue: Milivoltios de pH (MmvpH), Potencial de Hidrógeno (pH), Potencial Redox (Mvorp), Porcentaje de Oxígeno Disuelto (%DO), Partes Por Millón de Oxígeno Disuelto (PPmDo), Conductividad Eléctrica (uS/cm), Resistividad (MΩ.cm), Partes por millón de Sólidos Disueltos Totales (PPmTds), Salinidad (Psu), Temperatura en grados Celsius (T°C) y Presión (Psi). Además, se utilizaron investigaciones científicas publicadas en revistas académicas especializadas y la norma correspondiente al D. S. N.º 011-2017-MINAM [20], para limitar la investigación. Las fuentes permiten establecer una información sólida, confiable, para el análisis

y aplicación de nuevos métodos de investigación, centradas en mostrar el comportamiento de las concentraciones MP con ECAs. Los valores de concentración están expresados en partes por millón (ppm).

```
import pandas as pd

# Datos de metales pesados
data = {
    'Cr': [9.11, 3.26, 3.30, 9.59, 23.57, 4.79, 17.08, 9.01, 6.18],
    'As': [36.20, 63.44, 28.25, 52.58, 1947.63, 42.40, 114.77, 66],
    'Cd': [0.76, 0.31, 0.06, 0.82, 13.45, 0.50, 1.19, 1.30, 0.24],
    'Ba': [95.67, 258.11, 116.18, 253.61, 147.76, 61.17, 586.46],
    'Pb': [15.24, 56.11, 6.34, 23.43, 89.54, 25.93, 49.60, 34.29]
}

df = pd.DataFrame(data)
```

Fig. 4. DataSet de concentraciones de metales pesados

3.2. Algoritmos de aprendizaje automático

3.2.1 Bagging, o "Bootstrap Aggregating: Técnica que mejora la precisión de modelos al entrenar múltiples modelos en subconjuntos aleatorios de datos y promedia sus predicciones, reduciendo la varianza y el sobreajuste de entrenamiento. Ajusta pesos de la capa de salida, permite generalizar tareas de predicción, es rápido en tiempo real [16].

3.2.2 Extreme Gradient Boosting (XGBoost). Algoritmo que entrena múltiples árboles de decisión secuencialmente, con la técnica de boosting y cada árbol corrige los errores del anterior. Es eficiente con grandes volúmenes de datos y diferentes tipos de variables [17].

3.2.3 Voting Regresso. Modelo que integra modelos de regresión para mejorar la precisión de los aspectos a desarrollar [18].

4 RESULTADOS

4.1 Estadística Inferencial

TABLA 3
Normalidad de Metales Pesados y Ecas

Metal	Estadístico	p-valor	ECAs	Estadístico	p-valor
Cr*	0.9139	0.02462	Concentra-	210.73	1.737e-46
Cd*	0.3824	8.50e-10	ción_MP *		
As*	0.3612	5.53e-10	Eca_Agricola *	915.77	1.39e-199
Pb*	0.8918	0.00738	Eca_Residencial*	996.87	3.41e-217
Ba*	0.8043	0.00013	Eca_Industrial*	32.432	9.069e-08

*Los datos NO siguen una distribución normal.

4.2 Determinación límites permisibles de MP

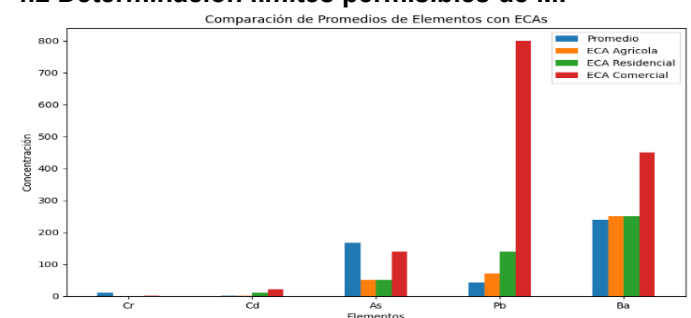


Fig. 5. Comparación de promedios de Metales Pesados con ECAs

La Fig. 5. indica que el Cr y As superan los límites permisibles de ECAs, en cambio los MP Ba, Pb y Cd se mantienen estables sin impactos negativos.

4.2 Matriz de correlación

Se obtuvo dos matrices de correlación de MP y de parámetros fisicoquímica siguientes:

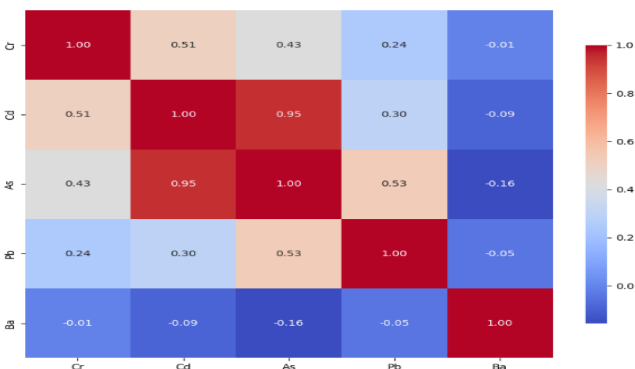


Fig. 6. Matriz de correlación de Spearman de metales pesados

La Fig. 6. Muestra la correlación de Spearman, de MP. Donde existe correlación:

- Alta y significativa entre Cr y Cd. Esto sugiere que a medida que aumenta la concentración de Cr, también tiende a aumentar la concentración de Cd, lo cual puede ser relevante en estudios de contaminación, gestión ambiental, predicción fiable y entender las relaciones clave entre MP y factores.
- Moderada y débil entre Cr y As y entre Pb y Ba, lo que indica que no es significativa. Indica que, aunque hay cierta relación, es menos confiable y podría ser influenciada por otros factores. Requiere un análisis cauteloso, por ser menos confiables.
- Sin correlación significativa entre Cr y Ba. No hay evidencia suficiente de una relación entre las concentraciones de MP y varían independientemente.
- Interdependencias complejas entre Cd y As, que pueden variar en diferentes condiciones ambientales. Necesitan métodos analíticos avanzados para entender cómo las variables interactúan en contextos más amplios.

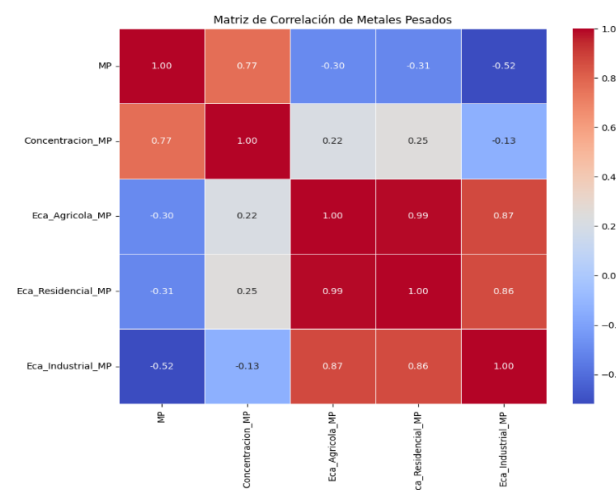


Fig. 7. Matriz de correlación de Spearman de factores fisicoquímicos

La Fig. 7, muestra la matriz de correlación de Spearman, de los MP y su concentración en distintos ECAs para usos agrícola, residencial e industrial. Esto indica que existe correlación:

- Alta y positiva entre de MP y Concentración_MP, lo que indica que a mayor concentración de MP, mayor presencia de contaminación general.
- Fuerte y positiva: de ECAs Agrícola, Residencial e Industrial entre sí, indica que los niveles de contaminación en un sector están fuertemente relacionados con los demás sectores.
- Débil y positiva de concentración de MP con los ECAs agrícola y residencial. Indica que, si aumenta la concentración de MP, los ECAs tienden a verse ligeramente afectados.
- Débil y Negativa de MP con el ECAs industrial indicando que, existe incremento en la concentración de MP y no está necesariamente asociado con un empeoramiento en la adecuación ambiental, quizás debido a más tolerancia a contaminantes en estas zonas. Así mismo, de MP y ECAs, donde a mayor presencia de MP tiende a reducir la calidad ambiental en términos de diferentes usos (agrícola, residencial e industrial).
- Se muestra una relación directa entre la presencia de MP y la disminución de la calidad del ambiente según los distintos ECAs, es decir muestra el incremento de la contaminación, frente a los permisibles.

4.3 Matriz de confusión de Algoritmos

- XGBoost: muestra la efectividad del modelo en clasificar como "Permisible" o "Contaminado" los metales pesados, mostrando precisión del 100% en todos los casos analizados. Los verdaderos positivos y negativos indican alta exactitud en la clasificación, validando la eficacia del modelo en contextos donde es crucial minimizar errores de clasificación.
- Bagging: muestra un rendimiento excelente (100% de precisión) para Cr, As, Pb y Ba, mientras que el Cd alcanza un 75% de precisión, indicando posibles mejoras. Los errores ocurren en la clasificación de casos "Contaminado" como "Permisible".
- Voting (XGBoost+Bagging), revela un rendimiento excelente (100% de precisión) para todos los MP. La matriz muestra un rendimiento alto (100% de precisión) para los MP Cr, Cd, As, Pb y Ba, no existe errores en la clasificación de casos "Contaminado" para Cr, como "Permisible" para todos los casos restantes de los MP (Ver Fig. 8).

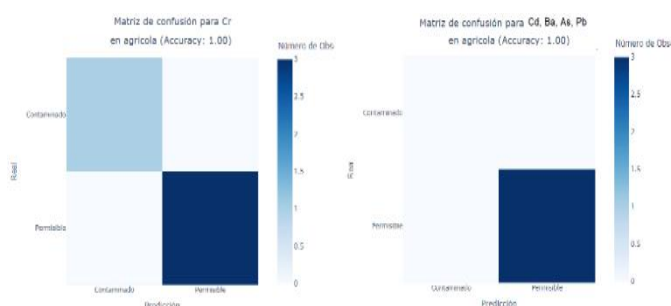


Fig. 8. Matriz de Confusión de Voting para los metales pesados con ECAs de suelo agrícola.

4.3.1 Validación Cruzada

Los algoritmos XGBoost (70%), Bagging (75%) y Voting, tienen una precisión media de 0.5375, 0.5375 y 0.5625, respectivamente, y una validez promedio de 75%, con Voting.

4.3 Predicción futura por meses

El comportamiento de la predicción por meses sugiere que, a corto y mediano plazo, ciertos metales pesados como el Cr representan un riesgo ambiental pronunciado, mientras que otros, como el Cd, As, Pb o Ba podrían convertirse en un problema a largo plazo debido a su crecimiento pausado en la zona de la microcuenca del río Antabamba.

4.3.1 Resultados de predicción de contaminación de sedimentos por metales pesados.

La determinación de contaminación a mediano plazo de 3 años con los algoritmos Xgboost, Bagging y Voting. Del cual, resalta la contaminación por el MP Cr: "Cromo". Por lo tanto, el suelo no cumple con los ECAs permisibles para uso 'agrícola' ni para uso 'industrial'. Se realizará una evaluación de contaminación agrícola, ya que las muestras analizadas del río Antabamba se utilizan para actividades agrícolas en la zona.

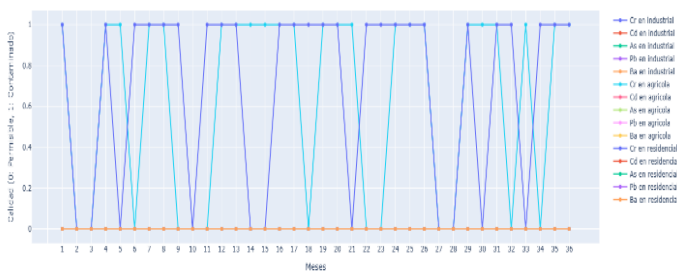


Fig. 10. Comparación de ECAs por mes con XGBoost

La Fig. 10, muestra resultados del algoritmo XGBoost, donde de 36 meses predichos de CSSMP por Cromo, 17 meses exceden los límites permitidos de ECAs, para suelo agrícola.

Los análisis de laboratorio finalizaron, en el mes de setiembre del 2023. Por lo cual, se considera agosto de 2023 como el 'mes 1' y a los meses siguientes se les asignan números consecutivos. Como sigue: 2, 3, 7, 8, 10, 12, 14, 15, 16, 19, 20, 21, 22, 24, 32, 33, 36.

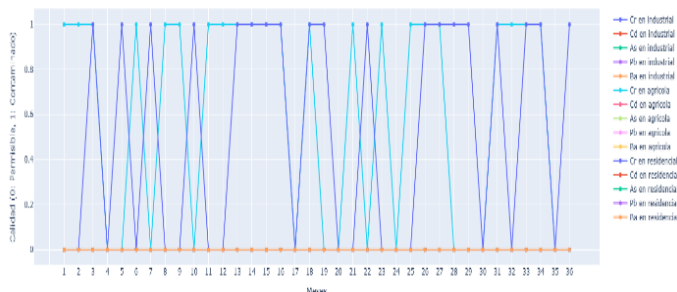
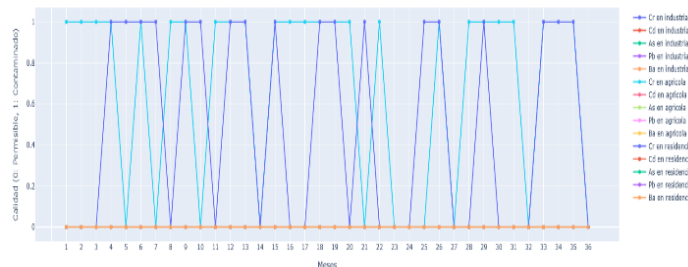


Fig. 11. Comparación de ECAs por Mes con Bagging

La Fig. 11, muestra resultados del algoritmo Bagging, donde de 36 meses predichos de contaminación por Cromo, muestra

24 meses que exceden los límites permitidos para suelo agrícola.

Se observa, la contaminación por Cr, en los meses: 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 15, 16, 18, 19, 20, 24, 25, 27, 29, 31, 32, 33, 34, 35. Fig. 12. Comparación de ECAs por Mes con Voting



La Fig. 12, muestra resultados, donde de 36 meses predichos de contaminación por Cromo, muestra 21 meses que exceden los límites permitidos para suelo agrícola.

Se observa, la contaminación por Cr, en los meses: 1, 2, 3, 4, 6, 8, 9, 11, 12, 15 al 20, 22, 26, 28 al 31. Por lo tanto, es necesario prever acciones de mitigación de contaminación específicas para esos meses.

4.3.2 Conclusiones de los resultados

Siendo el algoritmo Voting (XGBoost + Bagging), el algoritmo aplicado con una precisión del 75%, se determinó que la CSSMP será según la Fig. 13, en un rango de 36 meses, del mes de octubre del 2023 a setiembre del 2026, e indica que: exceden los límites permitidos de ECAs, para suelo agrícola por Cr, en los meses de: octubre, noviembre y diciembre del 2023; enero, marzo, mayo, junio, agosto, septiembre y diciembre del 2024; enero, febrero, marzo, abril, mayo, julio, noviembre del 2025 y enero, febrero, marzo y abril del 2026.

5 DISCUSIÓN

La CSSMP en el río Antabamba, basada en 28 muestras analizadas con ICP-MS, mejoró su precisión de predicción de 35% a 75% mediante el algoritmo Voting Classifier (Bagging + XGBoost) [17][6]. En un periodo a mediano plazo, el 58.3% de los meses superaron los límites permisibles de ECAs agrícolas para Cr. En el río Yellow, 200 muestras analizaron CSSMP con una precisión superior al 90% usando Bagging, RDT y SVM, donde más del 40% superan límites en ECAs para Pb y Zn [6]. Las predicciones adicionales en otros ríos alcanzaron más del 90% de precisión con XGBoost y RF, donde el 60% superan los límites de ECAs [17].

Por lo tanto, se observa la alta correlación positiva entre la concentración de MP y la contaminación [6] y la relación significativa entre los ECAs agrícolas, residenciales e industriales, siendo especialmente alta en los sectores agrícola y residencial [17]. La menor correlación en el sector industrial sugiere una mayor tolerancia a la CSSMP [6]. Los resultados validan la eficacia del uso de algoritmos y su relevancia en estudios de impacto ambiental de la CSSMP [17][6].

6 CONCLUSIONES

La investigación sobre la determinación de CSSMP en la subcuenca del río Antabamba revela contaminación que excede los límites permisibles de las ECAs agrícolas, destacando: que el Cr y As representan un riesgo ambiental significativo, mientras que otros MP (Cd, Ba, Pb) permanecen dentro de los ECAs. Las variables no siguen una distribución normal, se confirman correlaciones importantes entre Cr-As y Cr-Cd y los MP con la contaminación que generaran las Concentraciones de MP, por lo que influyen fuertemente en la CSSMP.

Se destaca el uso del algoritmo ML Voting para determinar la CSSMP en un tiempo a mediano plazo. El cual, alcanzó un 75% de precisión y evidencia que el 58% de los meses (21 de 36) superan los límites permisibles de ECAs agrícolas. Así mismo, el Cr impacta negativamente la calidad ambiental.

7 RECOMENDACIONES

Es esencial aplicar control y monitoreo continuo en meses críticos para mitigar riesgos de altos ECAs y enfermedades. Se recomienda el algoritmo Voting con modelos optimizados y otros factores ambientales.

AGRADECIMIENTOS

Los autores agradecen al Vicerrectorado de Investigación de la UNAMBA, por la subvención del presente trabajo. Así mismo, a la Facultad de Ingeniería Informática y de Sistemas por la publicación de presente trabajo.

REFERENCIAS

- [1] M. Zhang, P. He, G. Qiao, J. Huang, X. Yuan, and Q. Li, "Heavy metal contamination assessment of surface sediments of the Subei Shoal, China: Spatial distribution, source apportionment and ecological risk," *Chemosphere*, vol. 223, pp. 211–222, May 2019, doi: <https://doi.org/10.1016/j.chemosphere.2019.02.058>
- [2] M. Huaranga et al., "Contaminación por metales pesados en la Cuenca del Río Moche, 1980 – 2010, La Libertad – Perú," 2010. doi: <https://doi.org/10.17268/sci.agropecu.2012.03.05>
- [3] L. Chávez and A. Flores, "Contaminación por metales pesados en sedimentos del río Antabamba, Apurímac," *Revista Local de Ciencias Ambientales de Apurímac*, vol. 6, no. 1, pp. 25–40, 2022.
- [4] P. M. Bach, W. Rauch, P. S. Mikkelsen, D. T. McCarthy, and A. Deltic, "A critical review of integrated urban water modelling – Urban drainage and beyond," *Environmental Modelling & Software*, vol. 54, pp. 88–107, Apr. 2014, doi: <https://doi.org/10.1016/j.envsoft.2013.12.018>
- [5] M. E. Mohammad, N. Al-Ansari, and S. Knutsson, "Annual Runoff and Sediment in Duhok Reservoir Watershed Using SWAT and WEPP Models," *Engineering*, vol. 08, no. 07, pp. 410–422, 2016, doi: <https://doi.org/10.4236/eng.2016.87038>
- [6] J. Feng and H. Yang, "Predicting heavy metal pollution in sediments using bagging techniques," *Environmental Monitoring and Assessment*, vol. 195, no. 3, p. 576, 2023. doi: <https://doi.org/10.1007/s10661-023-10932-2>
- [7] M. Abraham, J. A. Silva, and E. Lobo, "Assessment of heavy metals contamination in surface sediments using multivariate statistical

techniques and artificial neural networks," *Environmental Monitoring and Assessment*, vol. 188, no. 3, p. 144, 2016. doi: <https://doi.org/10.1007/s10661-015-5165-8>

- [8] H. Ávila, E. Quintero, and N. Angulo, "Determinación de metales pesados en sedimentos superficiales costeros del sistema lago de Maracaibo, Venezuela," *Multiciencias*, vol. 14, no. 1, pp. 16–21, 2014. <https://www.redalyc.org/articulo.oa?id=90430816005>
- [9] R. Abarca and A. Vargas, "Contaminación por metales pesados en la región de Apurímac: Un análisis de la calidad del suelo," *Revista de Ciencias Ambientales*, vol. 45, no. 2, pp. 123–135, 2021.
- [10] Ministerio del Ambiente, "Informe sobre el estado de los suelos en Perú," Lima. Gobierno del Perú, 2020. <https://site.minam.gob.pe>
- [11] Oefa, "Oficina de Evaluación y Fiscalización Ambiental," 2022. [Online]. Available: <https://www.gob.pe/oefa>
- [12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: <https://doi.org/10.1023/A:1010933404324>
- [13] Vu Huy, "A Machine Learning Assessment to Predict the Sediment Transport Rate Under Oscillating Sheet Flow Conditions," *University of New Orleans*, 12–2019. <https://scholarworks.uno.edu/cgi/viewcontent.cgi?article=1134&context=honors-theses>
- [14] [SIAR, "Sistema de Información Ambiental Regional," [Online]. Available: <http://siar.regionsanmartin.gob.pe/>
- [15] M. González Mares, "Hernández-Sampieri, R. & Mendoza, C (2018). Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta," *Revista Universitaria Digital de Ciencias Sociales (RUDICS)*, vol. 10, no. 18, pp. 92–95, Jan. 2019, doi: <https://doi.org/10.22201/fesc.20072236e.2019.10.18.6>
- [16] M. R. Khan et al., "Ensemble learning methods for predicting heavy metal contamination in river sediments," *Environmental Monitoring and Assessment*, vol. 196, no. 2, p. 99, 2024. doi: <https://doi.org/10.22201/fesc.20072236e.2019.10.18.6>
- [17] Y. Wang et al., "Extreme Learning Machine for predicting sediment contamination by heavy metals," *Ecotoxicology and Environmental Safety*, vol. 238, p. 113550, 2024. doi: <https://doi.org/10.1016/j.ecoenv.2023.113550>
- [18] P. Zhang et al., "Fuzzy logic and machine learning integration for predicting heavy metal levels in sediment," *Journal of Cleaner Production*, vol. 405, p. 135450, 2023. doi: <https://doi.org/10.1016/j.jclepro.2023.135450>

BIOGRAFÍA



Darío Dante Sánchez Castillo. Egresado del Doctorado en Cs. Tecnología y Medio Ambiente de la UNAP, Maestro en Minería y Medio Ambiente. Título: Ing. Metalúrgico de la UNSAAC. Docente adscrito al departamento de Ingeniería de la UNAMBA, con experiencia en empresas mineras en instituciones estatales y privadas.



Pablo Ruben Zuloaga Candia. Doctor en Administración de la Educación, Magíster en Gestión de la Educación. Título: Ing. de Minas de la UNSAAC, con experiencia en obras mineras y civiles.



Karina Gamarra Peralta, Egresada del Doctorado de Ingeniería de Sistemas e Informática de UNMSM, M. Sc. de Informática con mención: en gestión de TIC. Título: Ing. de Sistemas. Docente adscrita al departamento de Ingeniería Informática y Sistemas de la UNAMBA.