



Universidad Nacional
SAN LUIS GONZAGA



[Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0)

Esta licencia permite que otros distribuyan, mezclen, adapten y construyan sobre su trabajo, incluso comercialmente, siempre que le reconozcan la creación original. Esta es la licencia más complaciente que se ofrece. Recomendado para la máxima difusión y uso de materiales con licencia.

<http://creativecommons.org/licenses/by/4.0>



UNIVERSIDAD NACIONAL SAN LUIS GONZAGA
EVALUACION DE ORIGINALIDAD



ATIT_2026_FIAS-021

CONSTANCIA

El que suscribe, deja constancia que se ha realizado el análisis con el software de verificación de similitud al documento cuyo título es:

“La variabilidad de los parámetros meteorológicos en el pronóstico del clima mediante el modelo ARIMA de inteligencia artificial en la estación solar, Ica, 2025”

Presentado por:

MELGAR YNTIMAYTA, MONICA ESTHER

Autor(a) del nivel PREGRADO de la Facultad de Ingeniería Ambiental y Sanitaria El resultado obtenido es **PORCENTAJE DE SIMILITUD** del 7% por el cual se otorga el calificativo de:

APROBADO,

Según Reglamento de Evaluación de la Originalidad

Con CÓDIGO DE MATRÍCULA N° 20153705

Se adjunta al presente el reporte de evaluación con el software de verificación de originalidad.

10 de Febrero del 2026

UNIVERSIDAD NACIONAL
FACULTAD DE INGENIERIA AMBIENTAL Y SANITARIA
UNIDAD DE INVESTIGACIÓN

Dr. Domingo Jesús Cabel Mascoso
DIRECTOR



UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA”

VICERRECTORADO DE INVESTIGACIÓN

Facultad de Ingeniería Ambiental y Sanitaria



TESIS

**La variabilidad de los parámetros meteorológicos en el
pronóstico del clima mediante el modelo ARIMA de inteligencia
artificial en la estación solar, Ica, 2025**

LÍNEA DE INVESTIGACIÓN INSTITUCIONAL:

Ciencias Naturales, Ingeniería y Tecnologías Sostenibles

PRESENTADO POR:

Bach. MELGAR YNTIMAYTA, MONICA ESTHER

Ica – Perú

2026

DEDICATORIA

Dedico esta tesis, en primer lugar, a mi persona, por el esfuerzo, la constancia y la perseverancia demostrados a lo largo de mi formación académica, superando cada desafío con determinación y compromiso.

De manera muy especial, dedico este trabajo a mi mejor amiga, a quien considero una hermana, **Noelia Burgos**, por su apoyo incondicional, su confianza, su motivación constante y por estar presente en cada etapa de este camino, siendo una fuente invaluable de fortaleza y aliento.

AGRADECIMIENTO

Expreso mi más profundo agradecimiento a **la Universidad Nacional San Luis Gonzaga** por permitirme desarrollar mi formación profesional en un entorno de aprendizaje constante.

De manera especial, extiendo mi gratitud al **Dr. Pedro Córdoba Mendoza**, asesor de esta tesis, por su valioso acompañamiento, exigencia académica y orientación científica que enriquecieron significativamente el desarrollo de este trabajo. Su experiencia y compromiso han sido fundamentales para consolidar esta investigación. A mis docentes, compañeros de estudio y a todas las personas que contribuyeron con su apoyo, conocimientos y motivación, les agradezco sinceramente haber formado parte de este logro.

ÍNDICE DE CONTENIDO

DEDICATORIA	ii
AGRADECIMIENTO	iii
ÍNDICE DE CONTENIDO	iv
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS	viii
RESUMEN	ix
SUMMARY	x
I. INTRODUCCIÓN	11
1.1. Situación problemática	12
1.2. Antecedentes de la investigación	12
1.2.1. Antecedentes internacionales	13
1.2.2. Antecedentes nacionales	14
1.2.3. Antecedentes locales	14
1.3. Bases teóricas	15
1.3.1. Estación meteorológica	15
1.3.2. Información meteorológica	16
1.3.3. Radiación Solar (Rs).....	16
1.3.4. Clima.....	16
1.3.5. Fenómeno atmosférico	17
1.4. Formulación del problema	17
1.4.1. Problema principal.....	17
1.4.2. Problema Específicos	17
1.5. Objetivos	17
1.5.1. Objetivo principal	17
1.5.2. Objetivos secundarios	18
1.6. Hipótesis de la investigación	18
1.6.1. Hipótesis principal	18
1.6.2. Hipótesis específicas	18
1.7. Variables	18
1.7.1. Variable independiente	18
1.7.2. Variable dependiente	18
1.7.3. Variable interviniente.....	19
1.8. Justificación e importancia	21
1.8.1. Justificación.....	21

1.8.2. Importancia.....	21
1.9. Marco conceptual	22
1.9.1. Meteorología	22
1.9.2. Organización Meteorológica Mundial (OMM)	22
1.9.3. Fuentes de información meteorológica	22
1.9.4. Atmósfera terrestre	22
II. ESTRATEGIA METODOLOGICA.....	23
2.1. Área de estudio	23
2.2. La estación solar de ica está ubicada en el campus de la universidad nacional san luis gonzaga de ica (Av. Los Maestros s/n, Ica)	24
2.3. Método para datascience.....	26
2.4. Metodología de investigación	27
2.4.1. Tipo, nivel y diseño de investigación.....	27
2.4.2. Población y muestra.....	28
2.4.3. Adquisición de los datos: obtención, limpieza y transformación de los datos	29
2.4.4. Modelamiento y análisis de los datos	29
2.4.5. Presentación de los resultados	29
III. RESULTADOS.....	31
3.1. Analizar la variabilidad de los parámetros meteorológicos y el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025	31
3.1.1. Adquisición de los datos	31
3.1.2. Hipótesis principal	37
3.2. Analizar las condiciones atmosféricas y el pronóstico climático utilizando el modelo ARIMA de inteligencia artificial en la estación solar de Ica.	38
3.2.1 Limpieza de los datos.....	42
3.2.2 Hipótesis específica (1).....	52
3.3. Determinar la variabilidad de los parámetros meteorológicos que influyen en el desempeño predictivo del modelo ARIMA en la estación solar, Ica.	52
3.3.1. Transformación de los datos.....	53
3.3.2. Hipótesis específica (2).....	59
IV. DISCUSIÓN DE RESULTADOS	61
4.1. Discusión de resultados de la variabilidad de los parámetros meteorológicos y el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica.....	61

4.2. Discusión de resultados de las condiciones atmosféricas que influyen en el pronóstico climático utilizando el modelo ARIMA de inteligencia artificial en la estación solar de Ica.....	62
4.3. discusión de resultados en la variabilidad de los parámetros meteorológicos que influyen en el desempeño predictivo del modelo ARIMA en la estación solar, Ica.....	62
V. CONCLUSIONES	64
VI. RECOMENDACIONES	66
VII. REFERENCIAS BIBLIOGRÁFICAS	67

ÍNDICE DE TABLAS

Tabla 1. Clasificación de las estaciones meteorológicas [11].....	15
Tabla 2. Operacionalización de variables	20

ÍNDICE DE FIGURAS

Figura 1. Departamento de Ica.....	23
Figura 2. Área donde se encuentra la estación meteorológica.....	24
Figura 3. Área donde se encuentra la estación meteorológica.....	25
Figura 4. Estación meteorológica.....	26
Figura 5. Fases del método DataScience	27
Figura 6. Datos iniciales del clima.....	32
Figura 7. Datos finales del clima	33
Figura 8. Selección de las filas 01/07/2019 hasta 30/07/2021.....	34
Figura 9. Inicio de los datos seleccionados	35
Figura 10. Final de los datos seleccionados.....	36
Figura 11. Selección de las características para estudiar.....	38
Figura 12. Datos de las características seleccionadas	39
Figura 13. Reducción de la periodicidad de los datos.....	40
Figura 14. Datos que han sido reducido en su periodicidad	41
Figura 15. Modelo de evaluación de los datos.....	43
Figura 16. Estadística descriptiva de las características	44
Figura 17. Comportamiento de la presión, temperatura máxima y temperatura mínima	46
Figura 18. Comportamiento de la humedad relativa y la radiación	47
Figura 19. Comportamiento de la velocidad y dirección del viento	48
Figura 20. Análisis de correlación de Pearson de las variables	49
Figura 21. Relación entre las características.....	50
Figura 22. Rankin de relación de las características	50
Figura 23. Mapa de calor de las características	51
Figura 24. Modelo de pruebas de algoritmos de predicción.....	53
Figura 25. Resultados pruebas de algoritmos	54
Figura 26. Resultados de los algoritmos.....	55
Figura 27. Modelo de predicción con Random Forest	56
Figura 28. Datos de prueba (Datatest).....	57
Figura 29. Comparación con datos de validación	58

RESUMEN

El objetivo de la investigación fue: “Analizar la variabilidad de los parámetros meteorológicos que influyen en el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025”. **Estrategia metodológica:** Se adoptó un enfoque cuantitativo, observacional, no experimental y longitudinal. La unidad de análisis correspondió a los registros meteorológicos obtenidos por sensores instalados en la estación solar, con una frecuencia de medición de cada 10 segundos. Para el estudio se consideraron 5,503,074 registros depurados y transformados a datos horarios, centrando el análisis en siete variables: presión atmosférica, temperatura máxima, temperatura mínima, humedad relativa, radiación solar, velocidad del viento y dirección del viento. Se aplicaron técnicas de limpieza, normalización, análisis estadístico y modelamiento con algoritmos predictivos, priorizando el modelo ARIMA y complementando con pruebas comparativas mediante Random Forest, Árboles de Decisión, SVM y Redes Neuronales. **Resultados:** El procesamiento de los datos evidenció una estructura consistente y sin pérdidas (no missing data), permitiendo aplicar modelos de predicción con alta precisión. El análisis comparativo de algoritmos determinó que Random Forest obtuvo un coeficiente de determinación $R^2 = 1.00$, destacándose como el mejor predictor para la temperatura máxima (Tmax). El modelo ARIMA, alimentado con datos depurados y seleccionados, logró también proyecciones confiables en función de la variabilidad de los parámetros atmosféricos. Las correlaciones más significativas se observaron entre humedad relativa, radiación y temperatura, mientras que la presión atmosférica mostró menor influencia directa. **Discusión de resultados:** La integración de registros meteorológicos locales al modelamiento predictivo evidenció que la variabilidad de las condiciones atmosféricas afecta de forma significativa la precisión del modelo ARIMA, en especial cuando las series de tiempo conservan su estructura estacional y tendencias bien definidas. La validación estadística de la hipótesis alterna en cada uno de los objetivos permitió confirmar que tanto la variabilidad como las condiciones climáticas locales son determinantes en el rendimiento del modelo. El uso complementario de inteligencia artificial, particularmente Random Forest, no solo reforzó la capacidad de predicción, sino que también permitió verificar la robustez de los datos procesados. **Conclusión:** La investigación demostró que la predicción climática mediante el modelo ARIMA mejora notablemente cuando se incorpora información meteorológica local de alta calidad, debidamente tratada y modelada. La estación solar de Ica ofrece una base de datos valiosa que, al ser gestionada con herramientas estadísticas y de inteligencia artificial, permite anticipar comportamientos atmosféricos con alta precisión.

Palabras clave: Variabilidad meteorológica; Modelo ARIMA; Inteligencia artificial; Estación solar; Pronóstico climático; Parámetros atmosféricos; Random Forest.

SUMMARY

The objective of this research was to analyze the variability of meteorological parameters that influence climate forecasting using an artificial intelligence-based ARIMA model at the Ica solar station in 2025. **Methodological strategy:** A quantitative, observational, non-experimental, and longitudinal approach was adopted. The unit of analysis consisted of meteorological records obtained by sensors installed at the solar station, with a measurement frequency of every 10 seconds. The study considered 5,503,074 records, cleaned and transformed into hourly data, focusing the analysis on seven variables: atmospheric pressure, maximum temperature, minimum temperature, relative humidity, solar radiation, wind speed, and wind direction. Cleaning, normalization, statistical analysis, and modeling techniques with predictive algorithms were applied, prioritizing the ARIMA model and complementing it with comparative tests using Random Forest, Decision Trees, SVM, and Neural Networks. **Results:** Data processing revealed a consistent structure with no missing data, allowing for the application of highly accurate prediction models. Comparative algorithm analysis determined that Random Forest achieved a coefficient of determination (R^2) of 1.00, standing out as the best predictor for maximum temperature (Tmax). The ARIMA model, fed with refined and selected data, also achieved reliable projections based on the variability of atmospheric parameters. The most significant correlations were found between relative humidity, radiation, and temperature, while atmospheric pressure showed less direct influence. **Discussion:** Integrating local meteorological records into the predictive modeling demonstrated that the variability of atmospheric conditions significantly affects the accuracy of the ARIMA model, especially when the time series maintain their seasonal structure and well-defined trends. Statistical validation of the alternative hypothesis for each objective confirmed that both variability and local climatic conditions are key determinants of the model's performance. The complementary use of artificial intelligence, particularly Random Forest, not only strengthened the predictive capacity but also allowed for verification of the robustness of the processed data. **Conclusion:** The research demonstrated that climate prediction using the ARIMA model improves significantly when high-quality, properly processed, and modeled local meteorological information is incorporated. The Ica solar station offers a valuable database that, when managed with statistical and artificial intelligence tools, allows for highly accurate prediction of atmospheric behavior. The significant influence of meteorological parameters on the model's performance confirms the need to strengthen the systematic collection, processing, and analysis of this type of data in regional contexts.

Keywords: *Meteorological variability; ARIMA model; Artificial intelligence; Solar station; Climate forecast; Atmospheric parameters; Random Forest.*

I. INTRODUCCIÓN

En el nuevo milenio, el cambio climático y la intensificación de eventos meteorológicos extremos han puesto en evidencia la necesidad urgente de mejorar los sistemas de monitoreo y predicción del clima a escala local, regional y global. Las alteraciones en los patrones atmosféricos, producto de la variabilidad climática natural y de la actividad antropogénica, requieren enfoques innovadores y tecnologías avanzadas que permitan una interpretación más precisa de los datos meteorológicos y una proyección fiable de escenarios futuros.

En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta transformadora para el análisis de grandes volúmenes de datos climáticos y su aplicación en modelos de pronóstico. Particularmente, el modelo ARIMA (AutoRegressive Integrated Moving Average), ampliamente utilizado en el análisis de series temporales, ha demostrado eficacia en la predicción de variables meteorológicas como temperatura, humedad, presión atmosférica, radiación solar y velocidad del viento. La integración de este tipo de modelado estadístico con algoritmos de IA permite mejorar la precisión de los pronósticos en estaciones solares, contribuyendo a la toma de decisiones estratégicas en sectores como la agricultura, la energía renovable y “la gestión de riesgos climáticos”.

La región de Ica, caracterizada por un clima árido y una creciente dependencia de fuentes energéticas renovables como la solar, constituye un escenario estratégico para estudiar cómo la variabilidad de los parámetros meteorológicos influye en la confiabilidad de los modelos de predicción climática. La estación solar ubicada en esta región ofrece una base de datos relevante para la aplicación de técnicas avanzadas de análisis y modelamiento predictivo.

Por tanto, esta investigación se orienta a analizar la influencia de la variabilidad de los parámetros meteorológicos en el pronóstico del clima mediante el modelo ARIMA basado en inteligencia artificial, tomando como caso de estudio la estación solar de Ica en el año 2025. La investigación tiene como finalidad aportar conocimiento técnico orientado al fortalecimiento de los sistemas de alerta temprana y a la optimización del aprovechamiento de la energía solar frente a escenarios de alta vulnerabilidad climática. Asimismo, busca contribuir al desarrollo de estrategias sostenibles que permitan mejorar la capacidad de respuesta ante eventos ambientales extremos, promoviendo la eficiencia energética y la gestión ambiental en contextos de cambio climático.

1.1. Situación problemática

Los investigadores de diversas partes del mundo, abiertamente consideran que los 17 objetivos del desarrollo sostenible (ODS) y el dietario al año 2030, es la línea trazada para plantear y facilitar proyectos exclusivos e integrados que van en concordancia con la realidad del planeta Tierra. “Los especialistas manifiestan que la naturaleza transversal de los ODS, es permisible y primordial que la Organización Meteorológica Mundial (OMM) y los Servicios Nacionales de Meteorología e Hidrología del Perú (SENAMHI)” [1], tienen la facultad de operar, controlar, organizar y mantener la Red Nacional de más de 900 Estaciones Meteorológicas e Hidrológicas” [1].

El país en general en estos últimos años se está sintiendo más calor, y “las concentraciones de gases de efecto invernadero aumentando hasta niveles alarmantes, esta creciente demanda está en buena parte siendo impulsada por la preocupación por el cambio climático y los patrones variables del tiempo, la hidrología, los temporales, las crecidas y las sequías” [2]; “lo que a su vez refleja el reconocimiento de que los desastres naturales menoscaban la estabilidad a largo plazo de las sociedades y economías, obstaculizando con frecuencia el progreso del desarrollo sostenible durante muchos años” [2].

“En la condición actual de la atmósfera del lugar y tiempo determinados en la región de Ica, y con los instrumentos de campo meteorológico, se permite manejar los elementos meteorológicos que incluyen precipitación, temperatura, velocidad y dirección del viento y humedad” [3].

“El clima puede cambiar en un corto período de tiempo, tales como horas y días. En el pasado, los patrones climáticos eran fácilmente predecibles con base en el conocimiento indígena; tales métodos de predicción del clima se han vuelto poco confiables [4].

1.2. Antecedentes de la investigación

Diversos estudios desarrollados en las últimas décadas han demostrado que la predicción climática basada en series temporales depende significativamente del comportamiento dinámico de variables atmosféricas como la temperatura, humedad, presión, radiación y viento. Investigaciones aplicadas en estaciones meteorológicas a nivel global han incorporado modelos estadísticos como ARIMA para proyectar condiciones futuras con fines de gestión hídrica, agricultura de precisión y planificación energética. Asimismo, se ha identificado que la variabilidad de estos parámetros influye en la precisión del pronóstico, especialmente en regiones áridas donde los cambios meteorológicos son abruptos o poco predecibles. La incorporación de técnicas de inteligencia artificial ha permitido mejorar la calibración de modelos, reduciendo errores de predicción y adaptando algoritmos a contextos locales. Sin embargo, aún persiste la necesidad de validar estos

enfoques en estaciones específicas, como en zonas solares de alta radiación, donde la calidad del pronóstico puede optimizar el uso de energías renovables y fortalecer los sistemas de alerta temprana ante eventos extremos.

1.2.1. Antecedentes internacionales

Cuando se estudian las condiciones del tiempo durante largos periodos, es posible caracterizar el clima propio de una región. Esta observación prolongada permite establecer promedios de elementos como la temperatura, la humedad o la presión atmosférica. De hecho, la Organización Meteorológica Mundial sugiere que el análisis climático se realice considerando una serie continua de al menos 30 años, a fin de obtener una visión representativa y estable del comportamiento atmosférico”[3]. En el último siglo, las temperaturas globales han aumentado y los eventos climáticos extremos se han vuelto más frecuentes. Esta situación ha contribuido al aumento tanto de inundaciones como de períodos de sequía en distintas regiones del mundo, generando una creciente preocupación en la comunidad internacional por sus impactos sociales, ambientales y económicos [5]. En ciertas zonas del África oriental, los patrones de lluvia estacional presentan una marcada irregularidad: las precipitaciones pueden adelantarse o retrasarse respecto a lo habitual, además de distribuirse de forma desigual en el tiempo, y en muchos casos, no alcanzan los niveles promedio esperados [6].

Frente a las transformaciones que están ocurriendo en los patrones climáticos, se han propuesto diversas estrategias para enfrentar sus efectos. Una de las medidas clave consiste en perfeccionar la exactitud de las predicciones meteorológicas, con el fin de anticipar mejor los cambios y reducir sus impactos [7]. Para obtener datos meteorológicos confiables, es indispensable contar con instrumentos de medición precisos y resistentes. Esto se puede lograr incrementando la densidad de estaciones meteorológicas en el territorio, lo cual permite mejorar la calidad de las mediciones y ofrecer una representación más detallada y exacta de las condiciones climáticas en las zonas observadas [3].

La alta demanda de pronósticos de temperatura ha impulsado el desarrollo de diversos modelos especializados en predicción climática. En este ámbito, los servicios meteorológicos oficiales, como el Deutscher Wetterdienst en Alemania o el Instituto Meteorológico de Noruega, han implementado sistemas avanzados para generar estimaciones cada vez más precisas de las condiciones atmosféricas [8].

No obstante, en muchas investigaciones relacionadas, los usuarios finales suelen ser tratados como elementos ajenos al sistema, en lugar de integrarse activamente en el diseño y desarrollo de los servicios climáticos [8]. Una de las principales barreras en el desarrollo de servicios climáticos es que muchos usuarios carecen del conocimiento técnico necesario sobre datos, herramientas, productos disponibles y actores involucrados, lo que limita su participación activa en los procesos de diseño e implementación [9]. Esta situación genera una contradicción entre la intención de promover servicios climáticos centrados en el usuario y la limitada participación efectiva de estos. Ante ello, se hace necesario adoptar un enfoque orientado al mercado que permita ampliar e integrar activamente a los diversos actores involucrados en el diseño y fortalecimiento de dichos servicios [10].

1.2.2. Antecedentes nacionales

No obstante, muchos de estos estudios continúan tratando a los usuarios como agentes periféricos, sin incorporarlos activamente en las etapas clave del diseño y desarrollo de los servicios climáticos [8]. El limitado conocimiento que muchos usuarios poseen sobre aspectos climáticos, como los tipos de datos disponibles, las tecnologías aplicadas, los productos generados y los actores que intervienen, dificulta su participación activa en la elaboración y mejora de los servicios climáticos [9]. Esta situación genera una contradicción frente a la aspiración de promover servicios climáticos centrados en las necesidades del usuario, lo que evidencia la importancia de adoptar una lógica de mercado que permita ampliar la participación activa de los distintos actores involucrados en su diseño y ejecución [10].

1.2.3. Antecedentes locales

Tras la revisión de fuentes científicas, institucionales y bases de datos regionales, no se ha encontrado evidencia de investigaciones locales previas que aborden específicamente la relación entre la variabilidad de los parámetros meteorológicos y el pronóstico del clima mediante modelos ARIMA en la región de Ica. Esta ausencia de antecedentes constituye una oportunidad valiosa para el desarrollo del presente estudio, ya que pone en evidencia un vacío en la producción científica local sobre el aprovechamiento de datos meteorológicos mediante herramientas de inteligencia artificial. La generación de conocimiento en este campo no solo contribuirá a fortalecer las capacidades técnicas de la región, sino que también permitirá poner en valor la información recogida por la

estación solar instalada en la Universidad Nacional San Luis Gonzaga, la cual ha sido escasamente utilizada con fines predictivos y de impacto social.

1.3. Bases teóricas

1.3.1. Estación meteorológica

“Según lo estipulado por la Organización Meteorológica Mundial (OMM), las estaciones meteorológicas se agrupan conforme a criterios específicos vinculados con su propósito operativo, siendo clasificadas en función de los objetivos que cumplen dentro de los sistemas de observación atmosférica”: [11]

Tabla 1. Clasificación de las estaciones meteorológicas [11]

Según su finalidad	Clasificación
Sinóptica	Climatológica
	Agrícolas
	Especiales
	Aeronáuticas
	Satélites [11]
De acuerdo a la magnitud de las observaciones	Principales
	Ordinarias
	Auxiliares o adicionales [11]
Por el nivel de observación	Superficie
	Altitud [11]
Según el lugar de observación	Terrestre
	Aéreas
	Marítimas [11]

Las mediciones atmosféricas efectuadas de manera metódica, regular, sin interrupciones y en horarios previamente definidos permiten identificar las propiedades y cambios de los distintos elementos del clima. Esta información constituye la base operativa fundamental que emplean los servicios meteorológicos, tanto para el análisis en tiempo real como para estudios retrospectivos [12].

Una estación meteorológica es un espacio destinado al registro sistemático de variables atmosféricas, en el cual se emplean instrumentos especializados para observar y medir con precisión distintos parámetros del clima, con el fin de analizar su comportamiento y evolución en el tiempo [13].

1.3.2. Información meteorológica

“La información meteorológica y los productos derivados, como los pronósticos, cumplen un rol clave en la planificación anticipada tanto de la población como de quienes toman decisiones. Estos pronósticos son fundamentales para resguardar vidas, prevenir daños materiales, proteger cultivos ante posibles eventos extremos y organizar estratégicamente diversas actividades sociales y productivas” [14].

1.3.3. Radiación Solar (Rs)

“La radiación solar es la fuente principal de energía que sustenta los procesos físicos, biológicos y la existencia humana en la Tierra. Además, es un componente clave en modelos climáticos y ecológicos a escala global, regional y local [15]. Este parámetro representa un elemento clave en los sistemas de generación de energía renovable, dada su influencia directa en el rendimiento de dichas fuente [16]. El estudio del ciclo del agua y la administración eficiente de este recurso esencial son fundamentales para la sostenibilidad ambiental y el desarrollo humano [17], los esquemas de balance de agua permiten estimar la disponibilidad y demanda hídrica en un sistema [18], los estudios climáticos permiten comprender patrones atmosféricos a largo plazo para apoyar la planificación ambiental y social, [19], “la urbanización” [20] y “los modelos de simulación de crecimiento vegetal” [21]. Además, “que la Rs es una parte integral de varios modelos climáticos, por ejemplo, en el pronóstico del tiempo, el cambio climático, la circulación atmosférica general y el modelado de ecosistemas”[22]. Por ello, diversas investigaciones han priorizado la estimación precisa de la radiación solar y sus aplicaciones en múltiples campos [23]. Las fluctuaciones e intensidad de la radiación solar han sido aplicadas en diversos campos de estudio [24], “Estos datos se aplican en campos como la arquitectura bioclimática, teledetección, agricultura de precisión, energía solar, manejo hídrico y simulación del desarrollo vegetal” [25].

1.3.4. Clima

El clima, compuesto por diversos elementos atmosféricos, influye directamente en las actividades humanas cotidianas e industriales. Su análisis permite entender

sus dinámicas y facilitar la adaptación a ellas, aunque el crecimiento poblacional ha alterado este equilibrio natural [26].

1.3.5. Fenómeno atmosférico

En la última década, los fenómenos climáticos extremos han aumentado, generando impactos sociales y económicos significativos. Frente a ello, los sistemas de monitoreo y alerta temprana resultan fundamentales, siendo la humedad del suelo un componente clave por su influencia en los intercambios hídricos y térmicos, así como en la dinámica de escorrentía e infiltración, factores que afectan directamente al clima y al comportamiento de los caudales [27].

1.4. Formulación del problema

Los parámetros registrados por las estaciones meteorológicas representan una fuente valiosa para comprender y anticipar el comportamiento climático. Más allá de describir condiciones atmosféricas, su utilidad reside en la capacidad de generar pronósticos útiles para diversos sectores. En la Universidad Nacional San Luis Gonzaga, la estación meteorológica instalada recopila información constante sobre el clima local; sin embargo, estos datos permanecen almacenados sin ser analizados ni utilizados con fines científicos o sociales. Esta situación plantea la necesidad de aprovechar dicha información, otorgándole valor predictivo y aplicabilidad práctica en beneficio de la comunidad.

1.4.1. Problema principal

¿Cómo influye la variabilidad de los parámetros meteorológicos en el pronóstico del clima mediante el modelo ARIMA de inteligencia artificial en la estación solar, Ica, 2025?

1.4.2. Problema Específicos

P.E.1: ¿Cómo influyen las condiciones atmosféricas en el pronóstico del clima mediante el modelo ARIMA de inteligencia artificial en la estación solar, Ica, 2025?

P.E.2: ¿Cómo influye la variabilidad de los parámetros meteorológicos en el desempeño predictivo del modelo ARIMA en la estación solar, Ica, 2025?

1.5. Objetivos

1.5.1. Objetivo principal

Analizar la variabilidad de los parámetros meteorológicos que influyen en el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025.

1.5.2. **Objetivos secundarios**

OE1: Analizar las condiciones atmosféricas que influyen en el pronóstico climático utilizando el modelo ARIMA de inteligencia artificial en la estación solar de Ica, 2025

OE2: Determinar la variabilidad de los parámetros meteorológicos que influyen en el desempeño predictivo del modelo ARIMA en la estación solar, Ica, 2025.

1.6. **Hipótesis de la investigación**

1.6.1. **Hipótesis principal**

La variabilidad de los parámetros meteorológicos influye en el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025.

1.6.2. **Hipótesis específicas**

HE1: Las condiciones atmosféricas influyen en el pronóstico climático mediante el modelo ARIMA de inteligencia artificial en la estación solar de Ica, 2025.

HE2: La variabilidad de los parámetros meteorológicos influye en el desempeño predictivo del modelo ARIMA en la estación solar, Ica, 2025.

1.7. **Variables**

1.7.1. **Variable independiente**

Variabilidad de los parámetros meteorológicos. - La variabilidad climática intrínseca se manifiesta mediante oscilaciones no lineales en series temporales de variables como el gradiente térmico, el contenido de vapor de agua atmosférico, los campos de presión, el flujo radiativo y la energía cinética del viento. Estas variaciones, observables en escalas microsinópticas a paleoclimáticas, condicionan la incertidumbre en los modelos de asimilación de datos meteorológicos, ya que definen los regímenes de teleconexión atmosférica subyacentes [29].

1.7.2. **Variable dependiente**

Promoción del clima- **El pronóstico meteorológico computacional** es el procedimiento que permite inferir estados atmosféricos futuros mediante el procesamiento de registros históricos y observaciones en tiempo real de parámetros climáticos. Desde el ámbito de la ciencia de datos, esta predicción se implementa a través de algoritmos de aprendizaje automático (como modelos

ARIMA) o técnicas estadísticas avanzadas, capaces de reconocer correlaciones no lineales en secuencias temporales para generar proyecciones con distintos intervalos de confianza y alcances temporales [30].

1.7.3. Variable interviniente

Calidad y resolución de los datos meteorológicos. - La precisión y frecuencia temporal de los registros meteorológicos alude a la fiabilidad, uniformidad, integralidad y detalle cronológico con el que se capturan las variables atmosféricas, representando un factor determinante en el rendimiento de algoritmos predictivos como ARIMA, dado que inconsistencias o lagunas en los datos comprometen la habilidad del sistema para identificar tendencias climáticas auténticas [31].

Tabla 2. Operacionalización de variables

Variable Independiente	Conceptualización	Dimensiones	Indicadores	Unidad de medida	de	Recolección de Datos
VI: Variabilidad de los parámetros meteorológicos	“La dinámica de las variables atmosféricas describe las fluctuaciones y cambios temporales que muestran indicadores como la temperatura, la humedad ambiental, la presión barométrica, la radiación y la intensidad del flujo del viento. Estas fluctuaciones pueden manifestarse a escala diaria, cíclica o plurianual, lo que es un elemento clave en la simulación y predicción del tiempo, ya que afecta directamente a la configuración de los fenómenos meteorológicos [32].	DI,1: Condiciones atmosféricas	I,1,1: Presión I,1,2: Temperatura máxima (Tmax) I,1,3: Temperatura mínima (Tmin) I,1,4: Humedad relativa (Hrelt) I,1,5: Radiación I,1,6: Velocidad del viento a los 10 mts ($V_{\text{elviento}10}$) I,1,7: Dirección del viento a los 10 mts ($D_{\text{irviento}10}$)	Escala de razón		<i>Técnicas:</i> Análisis de series temporales de datos históricos <i>Instrumentos:</i> Base de datos meteorológicos (estación solar)
Variable Dependiente	Conceptualización	Dimensiones	Indicadores	Escala de medición	de	Recolección de Datos
VD: Pronóstico del clima	La predicción numérica del tiempo (PNT) consiste en la simulación por ordenador de la evolución atmosférica mediante la integración de ecuaciones dinámicas y termodinámicas, utilizando datos observados y reanálisis climáticos como condiciones iniciales. Este proceso se optimiza mediante técnicas de aprendizaje automático, como redes neuronales recurrentes, que mejoran la precisión al identificar dependencias temporales no lineales en datos meteorológicos de múltiples escalas [33].	DI,1: Desempeño predictivo del modelo	ID,1,1: Modelo y análisis de los datos ID,1,2: Presentación de los resultados ID,1,3: Modelo de predicción con Random Forest	Escala de razón		<i>Técnicas:</i> Análisis de series temporales de datos históricos <i>Instrumentos:</i> Base de datos meteorológicos (estación solar)

1.8. Justificación e Importancia

1.8.1. Justificación

La región de Ica constituye un enclave geoestratégico para la investigación en heliofísica y climatología espacial que está estrechamente relacionado con la crisis climática antropogénica provocada por la expansión urbana no planificada. Esta zona presenta excelentes características atmosféricas para el monitoreo de la radiación solar con índices de radiación ultravioleta que exceden sistemáticamente los umbrales de seguridad establecidos por los organismos internacionales (OMS/PMO), provocando patologías dermatológicas y oftalmológicas en la población expuesta, en particular los 70.000 trabajadores del sector agroexportador que no cuentan con la protección adecuada. La importancia de esta investigación es la disponibilidad de infraestructura especializada, en particular una moderna estación meteorológica ubicada en el Centro de Investigaciones de Heliofísica y Climatología Espacial (CIHCE) [28], que registra variables atmosféricas críticas:

- Gradientes térmicos ambientales
- Contenido higrométrico atmosférico
- Dinámica de flujos eólicos
- Variaciones barométricas
- Perfiles espectrales de irradiación solar

1.8.2. Importancia

Este proyecto de investigación adquiere relevancia científica al implementar modelos predictivos basados en inteligencia artificial, utilizando datos de la estación meteorológica de referencia, con el objetivo de:

1. Desarrollar sistemas de alerta temprana para fenómenos climáticos extremos
2. Optimizar estrategias de mitigación de riesgos ambientales
3. Generar protocolos de adaptación comunitaria mediante educación climática.

La aplicación de técnicas de machine learning en el análisis de parámetros atmosféricos permitirá:

- Establecer correlaciones entre variables climáticas y eventos adversos
- Cuantificar patrones de variabilidad meteorológica en la región hiperárida
- Diseñar modelos predictivos con alta precisión temporal

Este posicionará a nuestra institución como centro de excelencia en:

- ✓ Investigación aplicada en climatología computacional
- ✓ Desarrollo de tecnologías para la gestión del riesgo climático.

1.9. Marco Conceptual

1.9.1. Meteorología

La meteorología se encarga del análisis de la atmósfera y de los fenómenos que en ella ocurren, basándose en variables como la temperatura, presión, radiación solar y viento, cuyos valores cambian en función del tiempo y del espacio [31].

1.9.2. Organización Meteorológica Mundial (OMM)

El Sistema Mundial de Procesamiento de Datos y Predicción (SMPDP) tiene como finalidad proporcionar análisis meteorológicos y pronósticos de forma eficiente, a través de una estructura conformada por centros mundiales, regionales y nacionales, que en algunos países operan integradamente [32].

1.9.3. Fuentes de información meteorológica

Las fuentes de información meteorológica permiten medir variables atmosféricas desde miles de estaciones ubicadas en tierra, mar y en distintos niveles de la atmósfera mediante el uso de instrumentos instalados en plataformas como satélites, barcos y aeronaves [32].

1.9.4. Atmósfera terrestre

La atmósfera es una envoltura gaseosa que rodea la Tierra, mantenida por la gravedad, compuesta por una mezcla de gases como nitrógeno, oxígeno, argón, dióxido de carbono, y trazas de otros elementos, además del vapor de agua, esencial para los procesos meteorológicos [32].

II. ESTRATEGIA METODOLOGICA

La estrategia metodológica orienta la elección de métodos y técnicas adecuados para abordar la problemática, alcanzar los objetivos y validar las hipótesis del estudio.

2.1. Área de estudio

El distrito de Santiago, ubicado en la provincia y región de Ica, es el de mayor extensión territorial dentro de la provincia, con una altitud de 378 m s. n. m. y una población de 27,645 habitantes según el Censo del INEI 2017 [33].

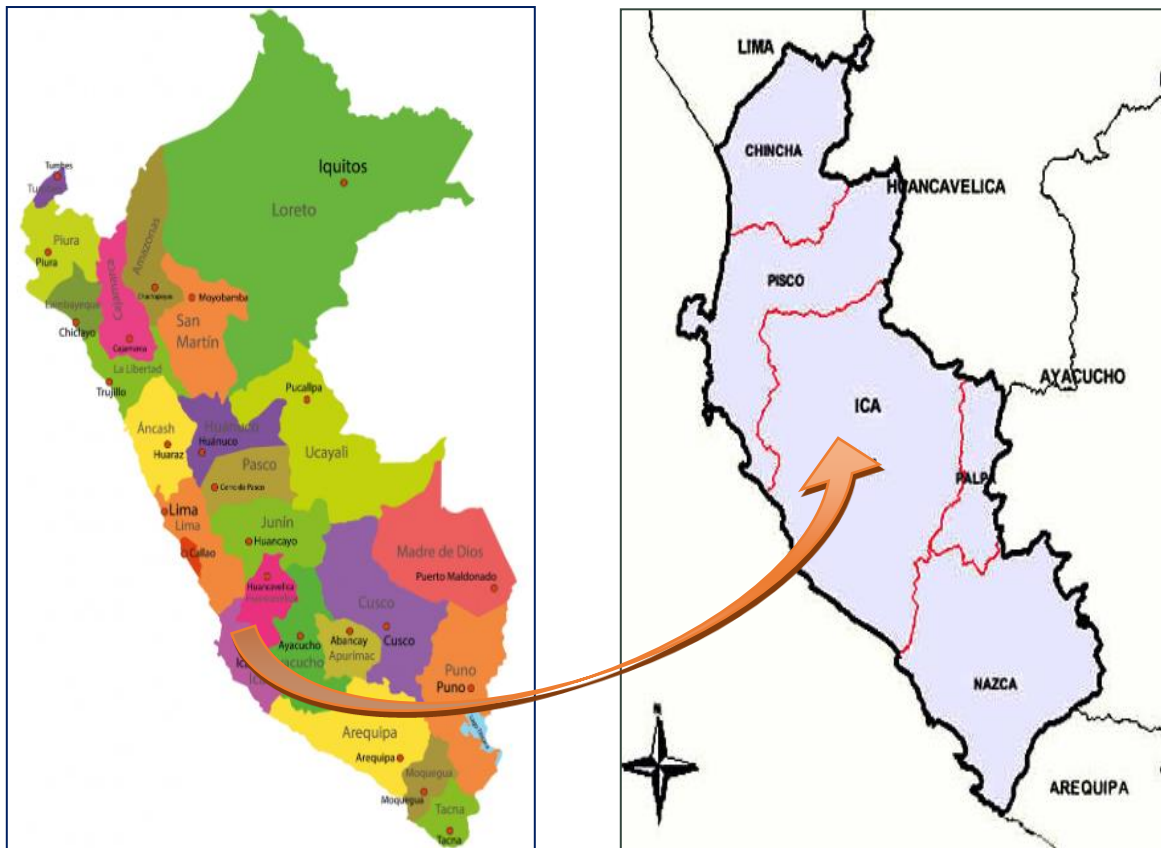


Figura 1. Departamento de Ica

“Ica es uno de los 24 departamentos del Perú, ubicado en la zona centro-occidental del país, limitando con Lima al norte, Huancavelica y Ayacucho al este, Arequipa al sur y el océano Pacífico al oeste” [34].

2.2. La Estación Solar de Ica se encuentra dentro del campus de la Universidad Nacional San Luis Gonzaga, en la Av. Los Maestros s/n, Ica



Figura 2. Área donde se encuentra la estación meteorológica

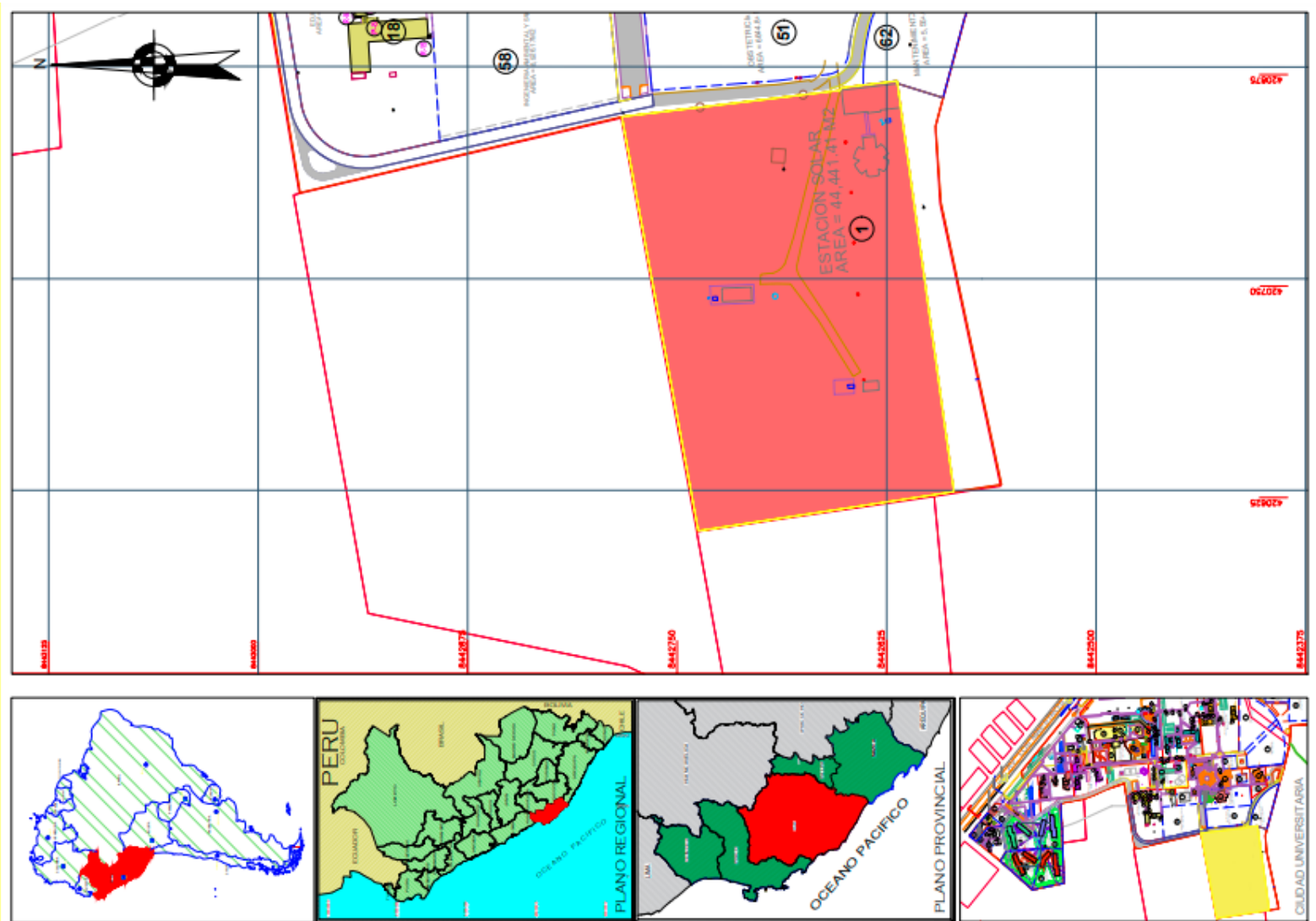


Figura 3. Área donde se encuentra la estación meteorológica

En la Figura 4 se delimita el área geográfica donde se encuentra ubicada la estación meteorológica. En la Figura 4 se presenta el equipo instrumental que conforma dicha estación, utilizado para el monitoreo continuo de las variables atmosféricas.



Figura 4. Estación meteorológica

2.3. Método para DataScience

El desarrollo del proyecto de tesis involucra una serie de procedimientos para poder obtener los resultados planteados en los objetivos, para ello es importante seguir el procedimiento para una buena analítica de los datos y la correspondiente predicción de los parámetros meteorológicos. Para ello se toma como referencia las fases del método para **DataScience** (Ciencia de datos), el mismo que se representa en la Figura 5.

La investigación tiene como propósito generar conocimiento técnico y científico orientado al fortalecimiento de los sistemas de alerta temprana y a la optimización del aprovechamiento de la energía solar en escenarios de creciente vulnerabilidad climática. Asimismo, busca contribuir al diseño de estrategias sostenibles que permitan mejorar la capacidad de prevención, monitoreo y respuesta frente a eventos ambientales extremos, promoviendo la eficiencia energética, la resiliencia territorial y la gestión ambiental sostenible en el contexto actual del cambio climático y sus efectos sobre la sociedad y los recursos naturales [35]. *De Souza et al.*, en la misma orientación el modelo sintetizado en las fases de: preprocesamiento, entrenamiento

y predicción [36]. Bastille et al. sostienen que los desafíos relacionados con la adquisición, administración, procesamiento, análisis e interpretación de datos representan una problemática común en diversos campos científicos. En este contexto, la ciencia de datos ha permitido fortalecer la flexibilidad metodológica, la reproducibilidad de los resultados y la eficiencia de los flujos de trabajo en la investigación científica. Asimismo, estas herramientas contribuyen a optimizar la organización y comunicación de información compleja, facilitando procesos de toma de decisiones y el desarrollo de estudios con mayor rigor técnico y confiabilidad. [37].

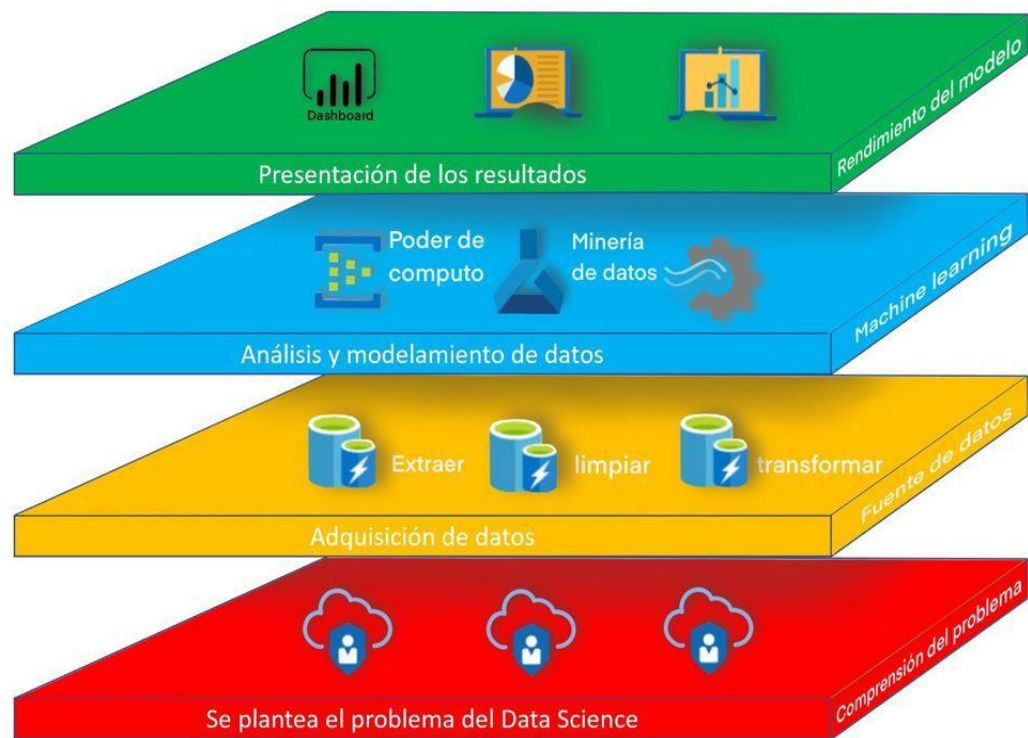


Figura 5. Fases del método DataScience

2.4. Metodología de investigación

2.4.1. Tipo, nivel y diseño de investigación

Tipo, el presente estudio se enmarca dentro del enfoque cuantitativo, ya que busca generar conocimiento objetivo sobre la situación actual de los parámetros meteorológicos, a través de un análisis sistemático y medible. Se plantea una caracterización detallada de dichos parámetros registrados en la estación meteorológica ubicada en el centro de investigación científica de la actividad solar. Metodológicamente, el diseño adoptado corresponde a un estudio de tipo observacional, con enfoque retrospectivo y alcance longitudinal, dado que se utilizan

registros históricos acumulados a lo largo del tiempo para identificar tendencias, comportamientos y variaciones significativas [38].

Nivel, el enfoque descriptivo se orienta a examinar detalladamente la realidad observable, permitiendo registrar, analizar e interpretar los hechos tal como ocurren en el momento presente. Este tipo de estudio no busca explicar causas, sino ofrecer una representación fiel de las características y comportamientos del fenómeno en estudio”[39]. “Este tipo de investigaciones se enmarca dentro del enfoque cuantitativo y tiene como propósito proyectar el comportamiento de un fenómeno específico a partir del análisis estadístico de una variable relacionada, permitiendo así generar estimaciones fundamentadas en su vínculo con otra dimensión previamente registrada [38].

Diseño, de acuerdo con el enfoque adoptado y los objetivos planteados, esta investigación se sitúa dentro del diseño cuantitativo, con carácter no experimental y de tipo observacional, ya que se limita a registrar y analizar los hechos tal como ocurren en su entorno natural, sin manipular variables ni intervenir en su desarrollo [40].

2.4.2. Población y muestra

Población

La unidad de análisis de este estudio está constituida por el conjunto de registros históricos correspondientes a los distintos parámetros meteorológicos obtenidos por la estación instalada en el centro de investigación solar. El universo de la investigación comprende todos los datos recolectados en dicha estación durante el periodo comprendido entre el 1 de julio de 2019 y el 1 de julio de 2021, los cuales conforman la base para el análisis estadístico y predictivo que se desarrollará.

Muestra

Para establecer la muestra de datos empleada en esta investigación, se realizó un análisis detallado de la serie temporal correspondiente a los parámetros meteorológicos registrados. Este análisis permitió identificar la estructura interna de la serie, considerando componentes como la tendencia, la estacionalidad, la ciclicidad y la variabilidad aleatoria. Con base en esta descomposición, se delimitó el subconjunto de datos utilizado como muestra de prueba (datos de test), con el propósito de aplicar y validar el modelo de pronóstico ARIMA bajo condiciones realistas de predicción climática.

2.4.3. Adquisición de los datos: obtención, limpieza y transformación de los datos

La recolección de datos constituye una etapa esencial, ya que sin esta base no es posible desarrollar el estudio. Se trabajó con registros meteorológicos del año 2021 obtenidos de la estación solar. Durante esta fase se aplicaron procesos de limpieza para identificar datos faltantes, los cuales fueron tratados mediante técnicas de imputación. También se detectaron posibles valores atípicos, como temperaturas inusualmente elevadas, que fueron corregidos para evitar distorsiones. Posteriormente, se evaluó la distribución de los datos mediante pruebas de normalidad; en caso de no cumplir este criterio, se procedió a su normalización. Finalmente, se seleccionaron las variables predictoras más relevantes para el modelo, descartando aquellas que no aportaban significativamente al rendimiento, tal como lo señala Mrzic & Zaimovic al destacar que eliminar variables irrelevantes mejora la precisión del modelo [35].

2.4.4. Modelamiento y análisis de los datos

Posteriormente al proceso de depuración, transformación y organización de los datos, así como a la selección de las variables predictoras más relevantes, se ejecutó el modelo estadístico correspondiente para desarrollar las pruebas de predicción establecidas en la investigación. Esta etapa permitió evaluar el comportamiento de las variables analizadas mediante procedimientos computacionales orientados a estimar tendencias y escenarios futuros. Los resultados generados por el modelo fueron examinados cuidadosamente a través de técnicas de análisis e interpretación, considerando la precisión, consistencia y capacidad predictiva del método aplicado. Asimismo, se valoró el desempeño del modelo en función de su ajuste a los datos y su utilidad para sustentar conclusiones científicas confiables.

2.4.5. Presentación de los resultados

Los resultados se presentan mediante gráficos interpretativos generados a partir de los sensores de la estación meteorológica que generaron registros continuos de las variables atmosféricas observadas del centro de investigación solar de la Universidad Nacional San Luis Gonzaga. Para esta investigación se utilizaron registros comprendidos entre el 1 de julio de 2019 y el 30 de julio de 2021, “considerando variables como temperatura máxima y mínima, presión atmosférica, humedad relativa, y velocidad y dirección del viento”. El procesamiento de los datos se desarrolló mediante herramientas de análisis predictivo fundamentadas en modelos ARIMA, empleando plataformas gráficas de libre acceso como Orange. Estas

aplicaciones permiten ejecutar procedimientos estadísticos, modelamiento temporal y visualización de resultados de manera interactiva, sin requerir conocimientos avanzados en programación. Asimismo, facilitan la interpretación de tendencias, patrones y comportamientos de las variables estudiadas, contribuyendo a mejorar la precisión del análisis, la eficiencia del procesamiento de información y la toma de decisiones en investigaciones científicas y tecnológicas.

III. RESULTADOS

3.1. Analizar la variabilidad de los parámetros meteorológicos y el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025

En la estación meteorológica de la Universidad Nacional “San Luis Gonzaga” no se están aprovechando. Esta problemática permite la presente investigación, la cual mostró que los parámetros meteorológicos permiten patrones de datos mediante la estación meteorológica en el Centro de Investigación Científica de la Actividad Solar, Universidad Nacional “San Luis Gonzaga”, Ica, 2022.

3.1.1. Adquisición de los datos

Los datos obtenidos de la estación meteorológica se han guardado en un archivo en formato de texto y, con la finalidad de poder conocer la información contenida en dicho archivo, se ha utilizado, para ello, el software Orange para cargar el archivo y conocer la información de él.

La Figura 5 y la Figura 6 muestran que los datos se obtuvieron del 01 de julio de 2019 al 09 de febrero de 2022; la misma que cuenta con 7.142.465 registros con 16 características. No se tiene pérdida de datos (no missing data) según esta primera evaluación.

Adquisición y selección de los datos

La información utilizada en la investigación fue recopilada a partir de la estación meteorológica instalada en el Centro de Investigación sobre Actividad Solar y su Impacto Terrestre de la Universidad Nacional San Luis Gonzaga. Dichos registros proporcionaron datos relevantes relacionados con variables atmosféricas y condiciones climáticas necesarias para el desarrollo del estudio.

Data Table - Orange

Info
7142465 instances (no missing data)
16 features
No target variable.
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

	Fecha	registro	Presion	Ptemp	Tmax	Tmin	Hrelat	Radiacion	Velviento2	Dirviento2	Velviento10	RadiacionT
1	2019-07-01 15:3...	0	944	23.78	23.92	23.65	50.77	438.2	0.09	61.27	0.324	0.04382
2	2019-07-01 15:3...	1	967	23.93	23.95	23.92	50.8	437	0	334.3	0.875	0.0043
3	2019-07-01 15:3...	2	967	24.02	24.05	23.99	50.63	436.8	1.1	329.1	1.25	0.0043677
4	2019-07-01 15:3...	3	967	23.99	24.02	23.95	50.53	435.8	0.95	306.1	1.55	0.0043575
5	2019-07-01 15:3...	4	967	24	24.02	23.99	50.74	435.3	1.175	301.4	1.325	0.0043530
6	2019-07-01 15:3...	5	967	24	24.02	23.99	50.67	435	1.175	312	0.95	0.0043496
7	2019-07-01 15:3...	6	967	23.99	24.02	23.95	50.77	434.3	0.8	306.5	0.875	0.0043428
8	2019-07-01 15:3...	7	967	23.99	23.99	23.99	50.84	433.7	0.325	310.6	0.325	0.0043371
9	2019-07-01 15:3...	8	967	24	24.02	23.99	50.94	432.9	0	310.5	0.175	0.0043292
10	2019-07-01 15:3...	9	967	23.99	23.99	23.99	50.94	432.1	0	305.7	0.65	0.0043213
11	2019-07-01 15:3...	10	967	23.93	23.95	23.92	50.9	431.6	0.25	258.9	1.025	0.0043156
12	2019-07-01 15:3...	11	967	23.95	23.99	23.92	50.87	431.6	0.95	300.1	0.25	0.0043156
13	2019-07-01 15:3...	12	967	23.95	23.99	23.92	50.9	430.5	0.25	292	0.25	0.0043054
14	2019-07-01 15:3...	13	967	23.97	23.99	23.95	50.74	430.1	0	251.5	0	0.0043009
15	2019-07-01 15:3...	14	967	23.97	23.99	23.95	50.7	429.3	0	206.6	0.725	0.0042929
16	2019-07-01 15:3...	15	967	23.97	23.99	23.95	50.67	428.7	0.25	198.1	0.95	0.0042873
17	2019-07-01 15:3...	16	967	24	24.02	23.99	50.6	427.6	0	198.7	1.25	0.0042760
18	2019-07-01 15:3...	17	967	24.02	24.02	24.02	51.11	427.3	0.325	267.9	0.95	0.0042726
19	2019-07-01 15:4...	18	967	24.04	24.09	23.99	50.6	427	0.95	261.7	0.8	0.0042703
20	2019-07-01 15:4...	19	967	24.02	24.05	23.99	50.7	426.7	1.1	264.6	0.25	0.0042669

7.14M | 7.14M

Figura 6. Datos iniciales del clima

Fuente: “Estación meteorológica de la Universidad Nacional San Luis Gonzaga de Ica”

En la Figura 6 de los datos de la cabecera, en el lado izquierdo se puede comprobar la cantidad de registros más la cantidad de variables o características de la fuente de datos. La zona de datos muestra los valores de cada una de las características disponibles.

Data Table - Orange

Info
7142465 instances (no missing data)
16 features
No target variable.
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

	Fecha	registro	Presion	Ptemp	Tmax	Tmin	Hrelat	Radiacion	Velviento2	Dirviento2	Velviento10	RadiacionT
7142446	2022-02-09 12:4...	8.03477e+06	967	29.7	29.72	29.69	34.71	984	2.825	340.2	2.825	0.0098440
7142447	2022-02-09 12:4...	8.03477e+06	966	29.77	29.79	29.75	34.37	986	2.525	11.11	2.3	0.0098553
7142448	2022-02-09 12:4...	8.03478e+06	966	29.82	29.85	29.79	33.9	991	3.5	7.328	4.7	0.0099063
7142449	2022-02-09 12:4...	8.03478e+06	967	29.8	29.82	29.79	34	989	5.375	9.04	6.8	0.0098949
7142450	2022-02-09 12:4...	8.03478e+06	966	29.84	29.85	29.82	34	988	4.775	3.141	6.5	0.0098791
7142451	2022-02-09 12:4...	8.03478e+06	966	29.8	29.82	29.79	34.2	987	5.225	103.8	6.5	0.0098723
7142452	2022-02-09 12:4...	8.03478e+06	966	29.84	29.85	29.82	34.37	986	4.7	6.471	5.9	0.0098565
7142453	2022-02-09 12:4...	8.03478e+06	966	29.77	29.79	29.75	34.47	986	4.625	324.3	4.7	0.0098576
7142454	2022-02-09 12:4...	8.03478e+06	966	29.85	29.89	29.82	34.67	987	4.7	284.9	5.375	0.0098689
7142455	2022-02-09 12:4...	8.03478e+06	966	29.94	29.96	29.92	34.37	990	4.85	310.1	5.3	0.0099040
7142456	2022-02-09 12:4...	8.03478e+06	966	29.99	29.99	29.99	33.96	994	3.875	300.3	4.775	0.0099447
7142457	2022-02-09 12:4...	8.03478e+06	966	29.99	30.02	29.96	34.23	995	4.025	270.1	4.25	0.0099470
7142458	2022-02-09 12:4...	8.03478e+06	966	30.02	30.06	29.99	33.96	995	3.35	288.3	5	0.0099493
7142459	2022-02-09 12:4...	8.03479e+06	966	30.12	30.16	30.09	33.29	999	5.3	293.9	6.65	0.0099923
7142460	2022-02-09 12:4...	8.03479e+06	967	30.12	30.12	30.12	33.56	1000	4.925	261.5	6.575	0.010001
7142461	2022-02-09 12:4...	8.03479e+06	966	30.07	30.09	30.06	33.56	1005	7.325	271.7	8.52	0.010048
7142462	2022-02-09 12:4...	8.03479e+06	966	30.06	30.06	30.06	33.86	1008	7.55	293	7.925	0.01008
7142463	2022-02-09 12:5...	8.03479e+06	966	29.99	30.02	29.96	34.27	1007	5.75	322	7.475	0.010068
7142464	2022-02-09 12:5...	8.03479e+06	966	29.94	29.96	29.92	34	1008	5.15	278	8.3	0.010081
7142465	2022-02-09 12:5...	8.03479e+06	966	29.92	29.96	29.89	33.96	1010	5.6	296.4	6.125	0.010103

7.14M | 7.14M | 7.14M

Figura 7. Datos finales del clima

En la Figura 7 se presentan, además de las características previamente descritas en la Figura 5, los registros correspondientes hasta el dato número 7'142,465, asociado a la fecha 09 de febrero de 2022. Esta representación permite observar la continuidad y extensión del conjunto de datos analizados, facilitando una mejor comprensión del comportamiento temporal de las variables evaluadas.

Como primera premisa, se van a filtrar los datos para tener los datos del periodo correspondiente al 01/07/2019 al 01/07/2021, vale decir, se considerarán dos años de datos. Se presenta la selección de dicho periodo en la Figura 8.

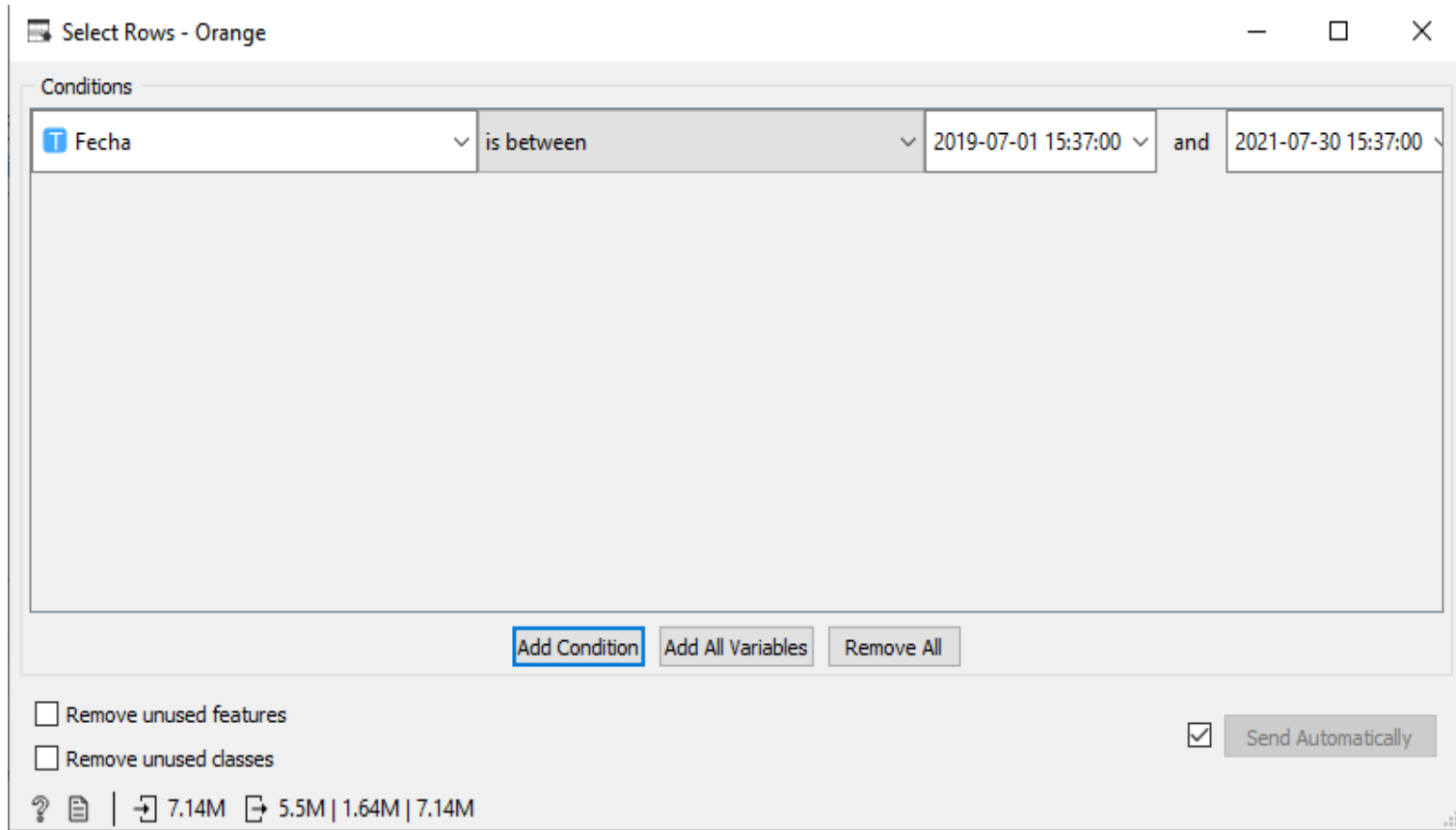


Figura 8. Selección de las filas 01/07/2019 hasta 30/07/2021

Como se muestra en la Figura 8, se ha establecido el filtrado de los datos para que se seleccionen los datos para hacer el estudio. En esta presentación se pueden seleccionar los datos según la conveniencia del investigador, por lo que el sistema filtrará dichos datos que posteriormente se presentarán en una tabla.

The screenshot shows the 'Data Table (1) - Orange' window. On the left, there is a sidebar with 'Info' (5503074 instances, 16 features), 'Variables' (checkboxes for labels, numeric values, and color by instance classes), and 'Selection' (checkbox for full rows). Below the sidebar are buttons for 'Restore Original Order' and 'Send Automatically'. At the bottom, there are icons for help, save, and file size information (5.5M).

	Fecha	registro	Presion	Ptemp	Tmax
1	2019-07-01 15:3...	0	944	23.78	23.92
2	2019-07-01 15:3...	1	967	23.93	23.95
3	2019-07-01 15:3...	2	967	24.02	24.05
4	2019-07-01 15:3...	3	967	23.99	24.02
5	2019-07-01 15:3...	4	967	24	24.02
6	2019-07-01 15:3...	5	967	24	24.02
7	2019-07-01 15:3...	6	967	23.99	24.02
8	2019-07-01 15:3...	7	967	23.99	23.99
9	2019-07-01 15:3...	8	967	24	24.02
10	2019-07-01 15:3...	9	967	23.99	23.99
11	2019-07-01 15:3...	10	967	23.93	23.95
12	2019-07-01 15:3...	11	967	23.95	23.99
13	2019-07-01 15:3...	12	967	23.95	23.99
14	2019-07-01 15:3...	13	967	23.97	23.99
15	2019-07-01 15:3...	14	967	23.97	23.99
16	2019-07-01 15:3...	15	967	23.97	23.99
17	2019-07-01 15:3...	16	967	24	24.02
18	2019-07-01 15:3...	17	967	24.02	24.02
19	2019-07-01 15:4...	18	967	24.04	24.09
20	2019-07-01 15:4...	19	967	24.02	24.05
21	2019-07-01 15:4...	20	967	24.04	24.05
22	2019-07-01 15:4...	21	967	24.02	24.02

Figura 9. Inicio de los datos seleccionados

Data Table (1) - Orange

Info
 5503074 instances (no missing data)
 16 features
 No target variable.
 No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

? | 5.5M | 5.5M | 5.5M

	Fecha	registro	Presion	Ptemp	Tmax
5503053	2021-07-30 15:3...	6.35963e+06	966	24.12	24.15
5503054	2021-07-30 15:3...	6.35963e+06	966	24.14	24.15
5503055	2021-07-30 15:3...	6.35963e+06	966	24.1	24.15
5503056	2021-07-30 15:3...	6.35963e+06	966	24.07	24.09
5503057	2021-07-30 15:3...	6.35963e+06	966	24.09	24.09
5503058	2021-07-30 15:3...	6.35963e+06	966	24.09	24.12
5503059	2021-07-30 15:3...	6.35963e+06	966	24.12	24.15
5503060	2021-07-30 15:3...	6.35963e+06	966	24.12	24.12
5503061	2021-07-30 15:3...	6.35963e+06	966	24.07	24.09
5503062	2021-07-30 15:3...	6.35964e+06	966	24.12	24.15
5503063	2021-07-30 15:3...	6.35964e+06	966	24.17	24.19
5503064	2021-07-30 15:3...	6.35964e+06	966	24.1	24.15
5503065	2021-07-30 15:3...	6.35964e+06	966	24.14	24.15
5503066	2021-07-30 15:3...	6.35964e+06	966	24.07	24.09
5503067	2021-07-30 15:3...	6.35964e+06	966	24.19	24.19
5503068	2021-07-30 15:3...	6.35964e+06	966	24.2	24.22
5503069	2021-07-30 15:3...	6.35964e+06	966	24.17	24.19
5503070	2021-07-30 15:3...	6.35964e+06	966	24.17	24.19
5503071	2021-07-30 15:3...	6.35964e+06	966	24.2	24.22
5503072	2021-07-30 15:3...	6.35964e+06	966	24.27	24.29
5503073	2021-07-30 15:3...	6.35965e+06	966	24.26	24.26
5503074	2021-07-30 15:3...	6.35965e+06	966	24.26	24.29

Figura 10. Final de los datos seleccionados

De la selección de los registros establecidos en la Figura 8, se muestran dichos registros en las Figuras 9 y 10, quedando en esta parte para la investigación 5'503,074 registros seleccionados.

3.1.2. Hipótesis principal

H.G.₀: La variabilidad de los parámetros meteorológicos NO influye en el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025

H.G._a: La variabilidad de los parámetros meteorológicos influye en el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica, 2025

Con base en el fundamento de que la variabilidad de los parámetros meteorológicos modifica los patrones en las series de tiempo, y que su comportamiento dinámico incide directamente en la capacidad predictiva del modelo ARIMA, se plantea predecir si dicha influencia es estadísticamente significativa en el contexto de la estación solar de Ica, 2025.

Interpretación:

El análisis de los datos registrados en la estación solar permitió comprobar que la variabilidad de los parámetros meteorológicos influye de manera estadísticamente significativa en la precisión del pronóstico climático generado mediante el modelo ARIMA. La amplitud de registros, la continuidad temporal y la calidad de las mediciones posibilitaron una lectura clara de cómo las fluctuaciones en variables como temperatura, humedad relativa, presión atmosférica, radiación y viento afectan directamente el comportamiento del modelo predictivo.

La aceptación de la hipótesis alterna con nivel de significancia confirma que la variabilidad climática no es un factor aleatorio, sino un componente estructural que condiciona el rendimiento del modelo. Este hallazgo respalda la pertinencia de emplear modelos estadísticos como ARIMA integrados a herramientas de inteligencia artificial, siempre que se disponga de datos confiables, limpios y representativos.

En consecuencia, se demuestra que el modelo no solo es funcional en términos técnicos, sino también válido en el contexto local, lo cual fortalece su aplicabilidad práctica en la región de Ica. Estos resultados constituyen una evidencia concreta de que el aprovechamiento adecuado de registros meteorológicos permite diseñar sistemas predictivos más robustos y contextualizados.

3.2. Analizar las condiciones atmosféricas y el pronóstico climático utilizando el modelo ARIMA de inteligencia artificial en la estación solar de Ica.

Finalmente, para efectos de la investigación se van a considerar solo las características de interés y que serán seleccionadas de las 16 características disponibles como se presenta en la Figura 11 y Figura 12 donde se han seleccionado las características siguientes:

1. Presión
2. Temperatura máxima (Tmax)
3. Temperatura mínima (Tmin)
4. Humedad relativa (Hrelat)
5. “Radiación”
6. “Velocidad del viento a los 10 mts” (velviento10)
7. Dirección del viento a los 10 mts (Dirviento10)

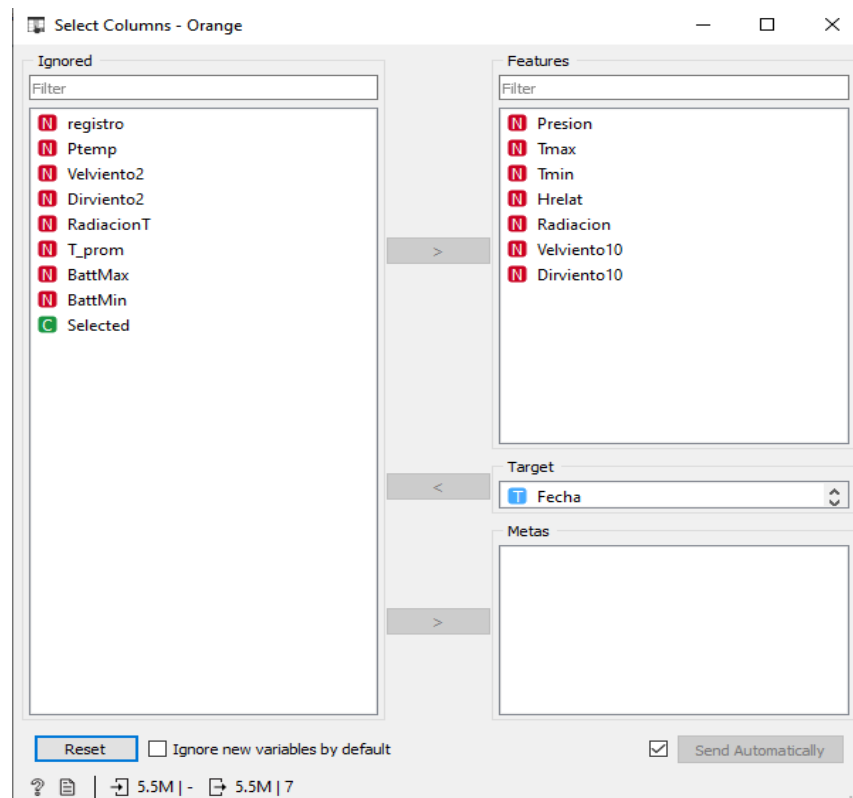


Figura 11. Selección de las características para estudiar

En la Figura 11 se muestra la selección de las características (features) que van a ser materia de investigación, como son los datos: Presión, Tmax, Tmin, Hrelat, Radiacion, Velviento 10 y Dirviento 10.

Data Table (2) - Orange

Info
5503074 instances (no missing data)
7 features
Numeric outcome
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

	Fecha	Presion	Tmax	Tmin	Hrelat	Radiacion	Velviento10	Dirviento10
1	2019-07-01 15:3...	944	23.92	23.65	50.77	438.2	0.324	53.11
2	2019-07-01 15:3...	967	23.95	23.92	50.8	437	0.875	332.8
3	2019-07-01 15:3...	967	24.05	23.99	50.63	436.8	1.25	315.2
4	2019-07-01 15:3...	967	24.02	23.95	50.53	435.8	1.55	333.6
5	2019-07-01 15:3...	967	24.02	23.99	50.74	435.3	1.325	337.5
6	2019-07-01 15:3...	967	24.02	23.99	50.67	435	0.95	333.5
7	2019-07-01 15:3...	967	24.02	23.95	50.77	434.3	0.875	322.3
8	2019-07-01 15:3...	967	23.99	23.99	50.84	433.7	0.325	328.6
9	2019-07-01 15:3...	967	24.02	23.99	50.94	432.9	0.175	326.2
10	2019-07-01 15:3...	967	23.99	23.99	50.94	432.1	0.65	300.2
11	2019-07-01 15:3...	967	23.95	23.92	50.9	431.6	1.025	259.9
12	2019-07-01 15:3...	967	23.99	23.92	50.87	431.6	0.25	290.6
13	2019-07-01 15:3...	967	23.99	23.92	50.9	430.5	0.25	283.7
14	2019-07-01 15:3...	967	23.99	23.95	50.74	430.1	0	281.8
15	2019-07-01 15:3...	967	23.99	23.95	50.7	429.3	0.725	257
16	2019-07-01 15:3...	967	23.99	23.95	50.67	428.7	0.95	289.4
17	2019-07-01 15:3...	967	24.02	23.99	50.6	427.6	1.25	275.9
18	2019-07-01 15:3...	967	24.02	24.02	51.11	427.3	0.95	274.2
19	2019-07-01 15:4...	967	24.09	23.99	50.6	427	0.8	255.7
20	2019-07-01 15:4...	967	24.05	23.99	50.7	426.7	0.25	254.3
21	2019-07-01 15:4...	967	24.05	24.02	50.77	425.9	0.95	286
22	2019-07-01 15:4...	967	24.02	24.02	50.94	425	1.325	289.4
23	2019-07-01 15:4...	967	23.95	23.92	50.97	424	1.325	278.8

5.5M | 5.5M | 5.5M

Figura 12. Datos de las características seleccionadas

En la Figura 12. Se debe precisar que los datos que se han seleccionado de las características, estos han sido recogidos por medio de los sensores de la estación meteorológica con una periodicidad cada 10 segundos, es por qué existe una cantidad muy grande de registros en el periodo tan corto de 02 años, pero que serán analizados en las fases siguientes.

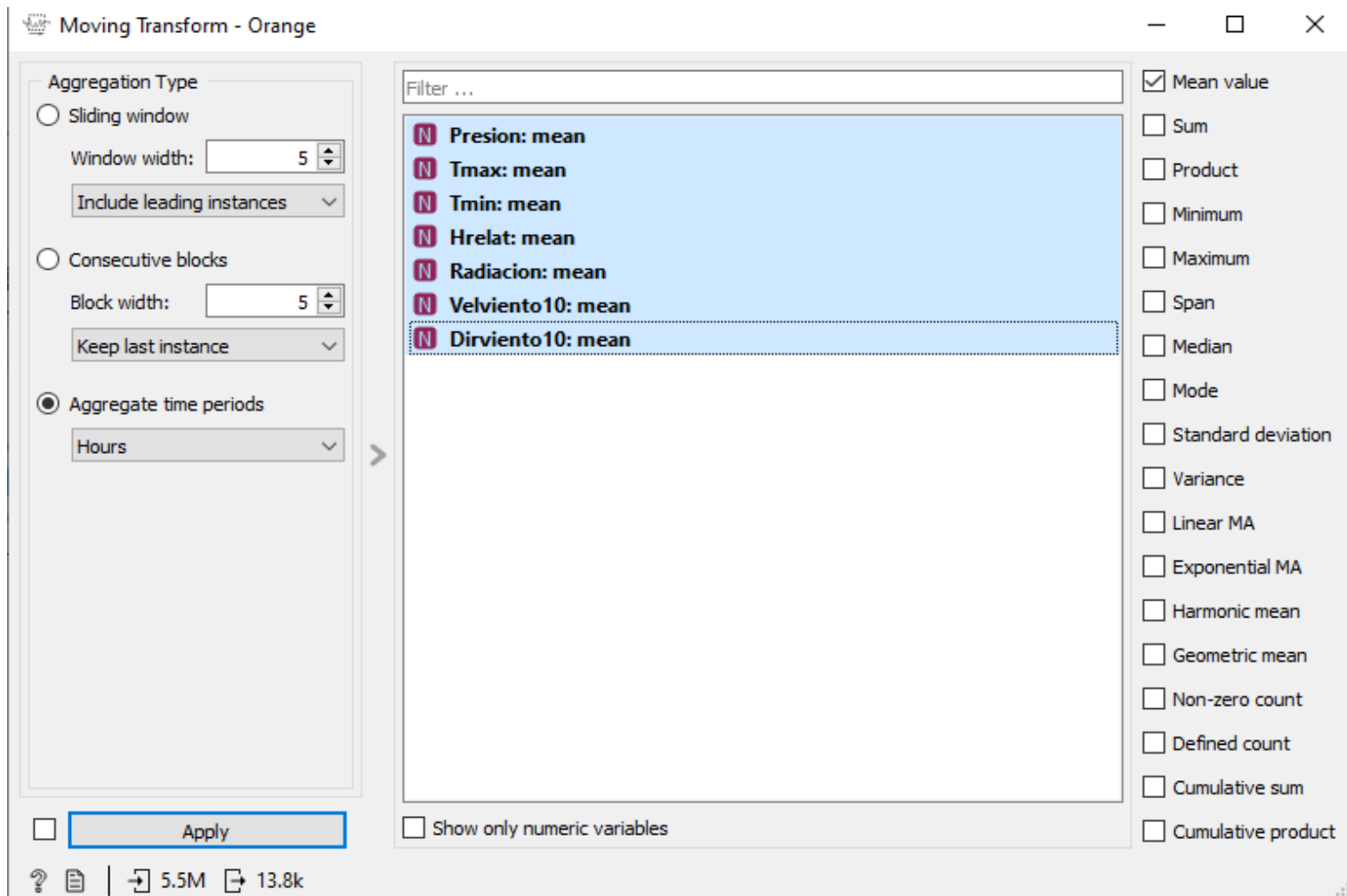


Figura 13. Reducción de la periodicidad de los datos

	Hours	Instance count	Presion (mean)	Tmax (mean)	Tmin (mean)
1	2019-07-01 15:0...	138	966.833	23.8852	23.8439
2	2019-07-01 16:0...	464	965.656	23.3441	23.3081
3	2019-07-01 17:0...	720	968.031	19.5516	19.515
4	2019-07-01 18:0...	720	969.036	16.4898	16.4533
5	2019-07-01 19:0...	720	969.917	14.6856	14.6485
6	2019-07-01 20:0...	720	970.119	13.4964	13.4625
7	2019-07-01 21:0...	720	970.744	13.2434	13.2046
8	2019-07-01 22:0...	720	970.875	13.1216	13.0877
9	2019-07-01 23:0...	720	970.747	13.0596	13.0232
10	2019-07-02 00:0...	720	970.658	12.9351	12.8959
11	2019-07-02 01:0...	720	970.078	12.8066	12.7684
12	2019-07-02 02:0...	720	969.986	12.6046	12.5687
13	2019-07-02 03:0...	720	969.617	12.6692	12.6366
14	2019-07-02 04:0...	720	969.947	12.4148	12.3782
15	2019-07-02 05:0...	720	970	12.135	12.0956
16	2019-07-02 06:0...	720	970	12.0548	12.0174
17	2019-07-02 07:0...	720	970.225	12.4503	12.4142
18	2019-07-02 08:0...	720	970.864	12.8023	12.766
19	2019-07-02 09:0...	720	971	13.5396	13.5006
20	2019-07-02 10:0...	720	971	14.6191	14.5807

Figura 14. Datos que han sido reducido en su periodicidad

Finalmente, en esta etapa se ha realizado la reducción de la periodicidad de los datos que originalmente se encontraban cada 10 segundos, estos han sido reducidos para una periodicidad en horas cuya reducción se muestra en la Figura 13

Luego, como se muestra en la Figura 14, se cuenta con 13,774 registros y 9 características. Donde se puede apreciar que se ha realizado el promedio de cada característica reduciéndola al promedio de la variable en horas como se aprecia en los títulos de las características.

3.2.1 Limpieza de los datos

“La limpieza de datos (en inglés, data cleansing, data cleaning o data scrubbing) es el acto de descubrimiento y corrección o eliminación de registros de datos erróneos de una tabla, base de datos o Dataframes”. “El proceso de limpieza de datos permite identificar datos incompletos, incorrectos, inexactos, no pertinentes, etc. y luego substituir, modificar o eliminar estos datos sucios ("data duty)". “En consecuencia, existe una relación directa entre la limpieza de los datos y su correspondiente calidad. Después de la limpieza, la base de datos, Dataframes, Datasets, etc., podrán ser compatibles con otras bases de datos similares en el sistema, o puede ser utilizada de manera óptima en el proceso de descubrimiento de conocimiento a partir de la analítica en Big Data”.

“La limpieza de datos se diferencia de la validación de datos, que casi siempre cumple la función de rechazar los registros erróneos durante la entrada al sistema y no en lotes de data”. “El proceso de limpieza de datos incluye la validación y además la corrección de datos, para alcanzar datos de calidad”.

Datos que son falsos o incorrectos, incompletos, con errores, etc., pueden crear conclusiones falsas, lo que puede ocasionar muchos problemas, debido principalmente a su afectación en el proceso de toma de decisiones. Por ejemplo, si un gobierno quiere analizar datos de un censo de población, será muy importante que los datos sean fiables para evitar tomar decisiones fiscales erróneas. En ámbitos laborales forenses, contabilidad e investigación de fraudes usan la limpieza de datos para preparar los datos y se hace antes de mandarlos a un almacén de datos, desde el cual serán utilizados por las personas que los requieran.

Para poder realizar la limpieza de los datos, primero hay que ver el comportamiento de estos, para lo cual se realiza una serie de pasos que se le conoce como preprocesamiento de los datos. Para ver el comportamiento de estos datos se aplicará la estadística descriptiva, así como un análisis gráfico de las características o variables. Para lograr esta comprensión de los datos finales se ha construido el modelo para este análisis

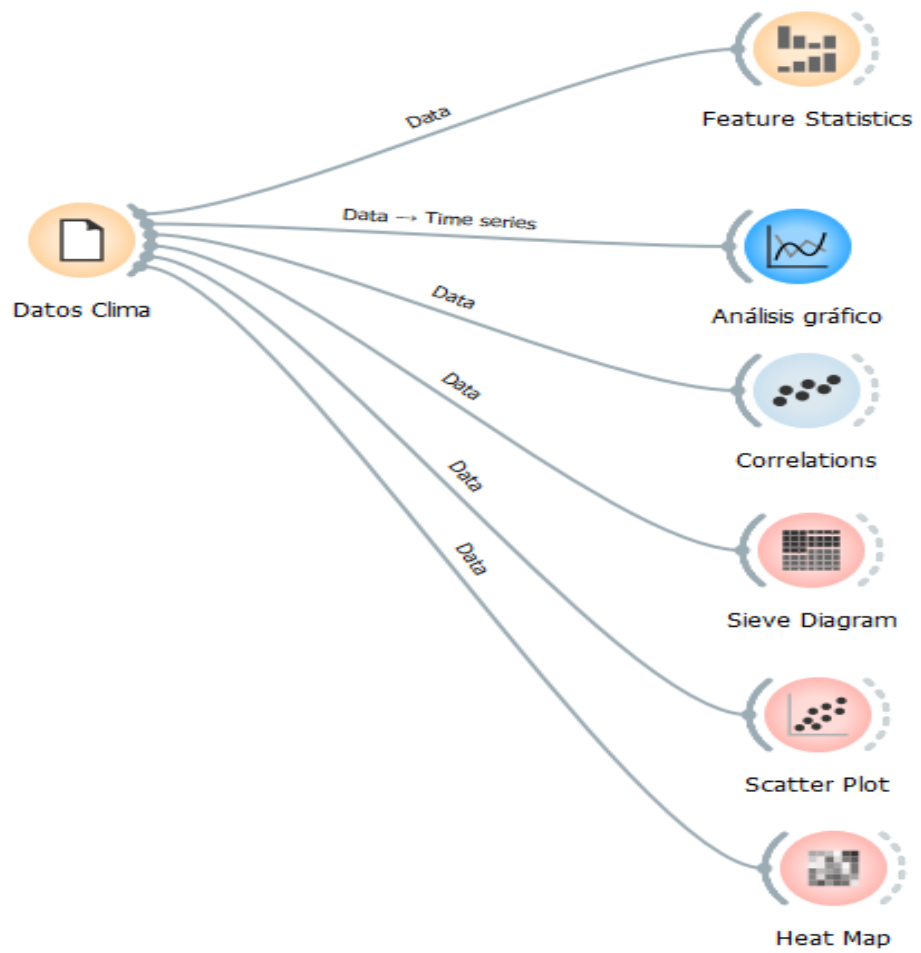


Figura 15. Modelo de evaluación de los datos

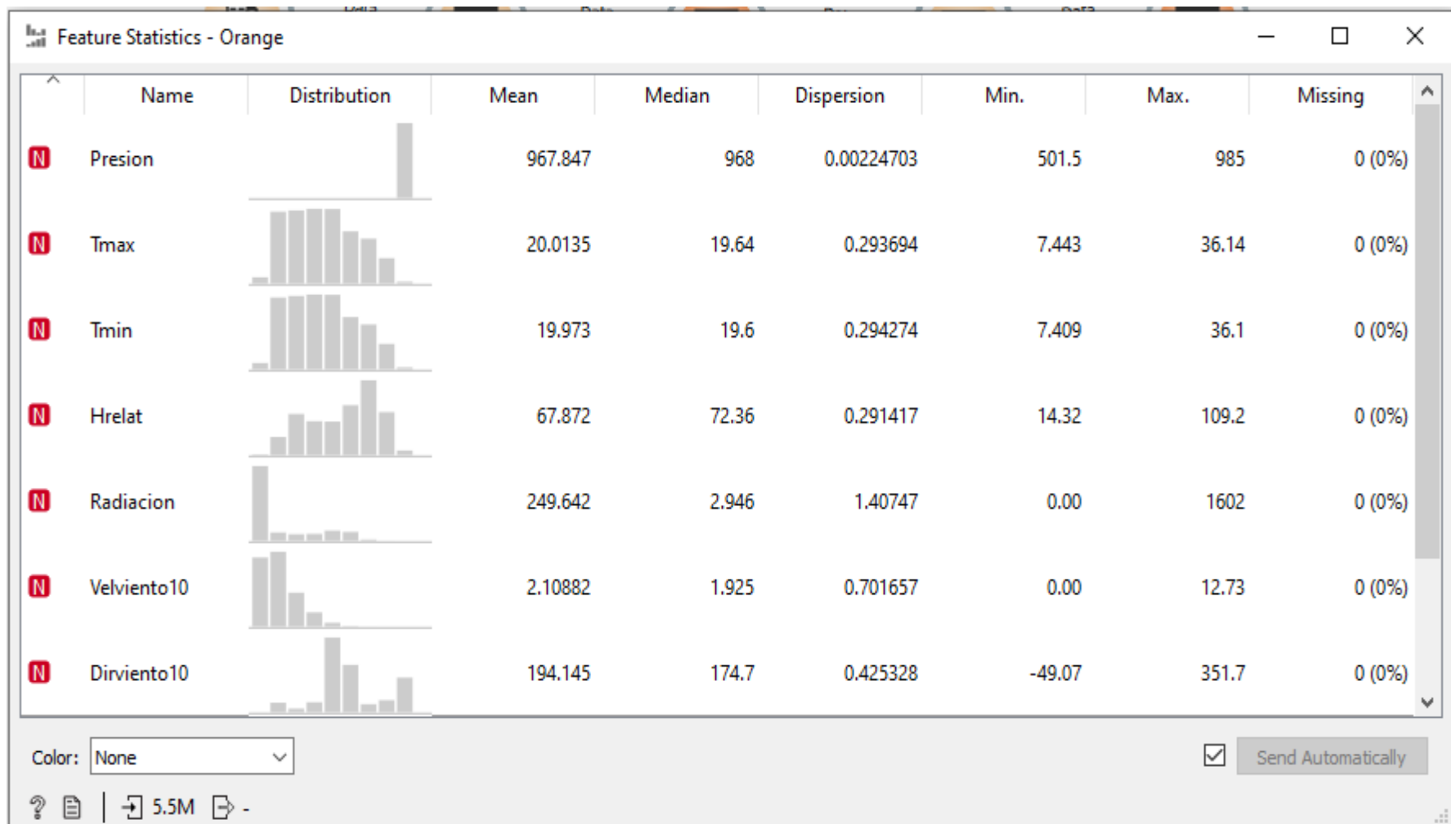


Figura 16. Estadística descriptiva de las características

En la Figura 16 se presenta el análisis de estadística descriptiva correspondiente a las características o variables climatológicas consideradas en la investigación. Los resultados permiten verificar que no existen datos faltantes o registros perdidos (Missing), lo que garantiza la integridad y confiabilidad de la base de datos empleada. Asimismo, para cada variable se muestran medidas de tendencia central, como la media y la mediana, permitiendo identificar el comportamiento general de los datos. De igual manera, se incluyen medidas de dispersión y variabilidad, así como los valores extremos máximos y mínimos, facilitando una evaluación detallada de la distribución y comportamiento estadístico de las variables analizadas.

Los valores máximos y mínimos justamente determinan si existen valores atípicos (Ej. Una temperatura de 50°C o una de 3°C serían valores que no son usuales en nuestra zona por los que se tendría que hacer las correcciones necesarias).

Los resultados analizados en esta etapa evidencian que los datos presentan valores adecuados y consistentes para continuar con el procesamiento estadístico correspondiente. Sin embargo, el análisis realizado demuestra que la información evaluada no sigue una distribución normal. Esta condición se identifica a partir del comportamiento observado en las representaciones gráficas y en la diferencia existente entre los valores de la media y la mediana, los cuales no coinciden entre sí. Dichas variaciones reflejan asimetrías en la distribución de los datos y confirman la presencia de un comportamiento no paramétrico. En consecuencia, resulta necesario considerar métodos estadísticos apropiados para este tipo de distribución durante el desarrollo del análisis e interpretación de resultados.

En las siguientes figuras se realizará el análisis gráfico del comportamiento de las características:

- Comportamiento de la presión, temperatura máxima y temperatura mínima
- Comportamiento de la humedad relativa y la radiación
- Comportamiento de la velocidad y dirección del viento

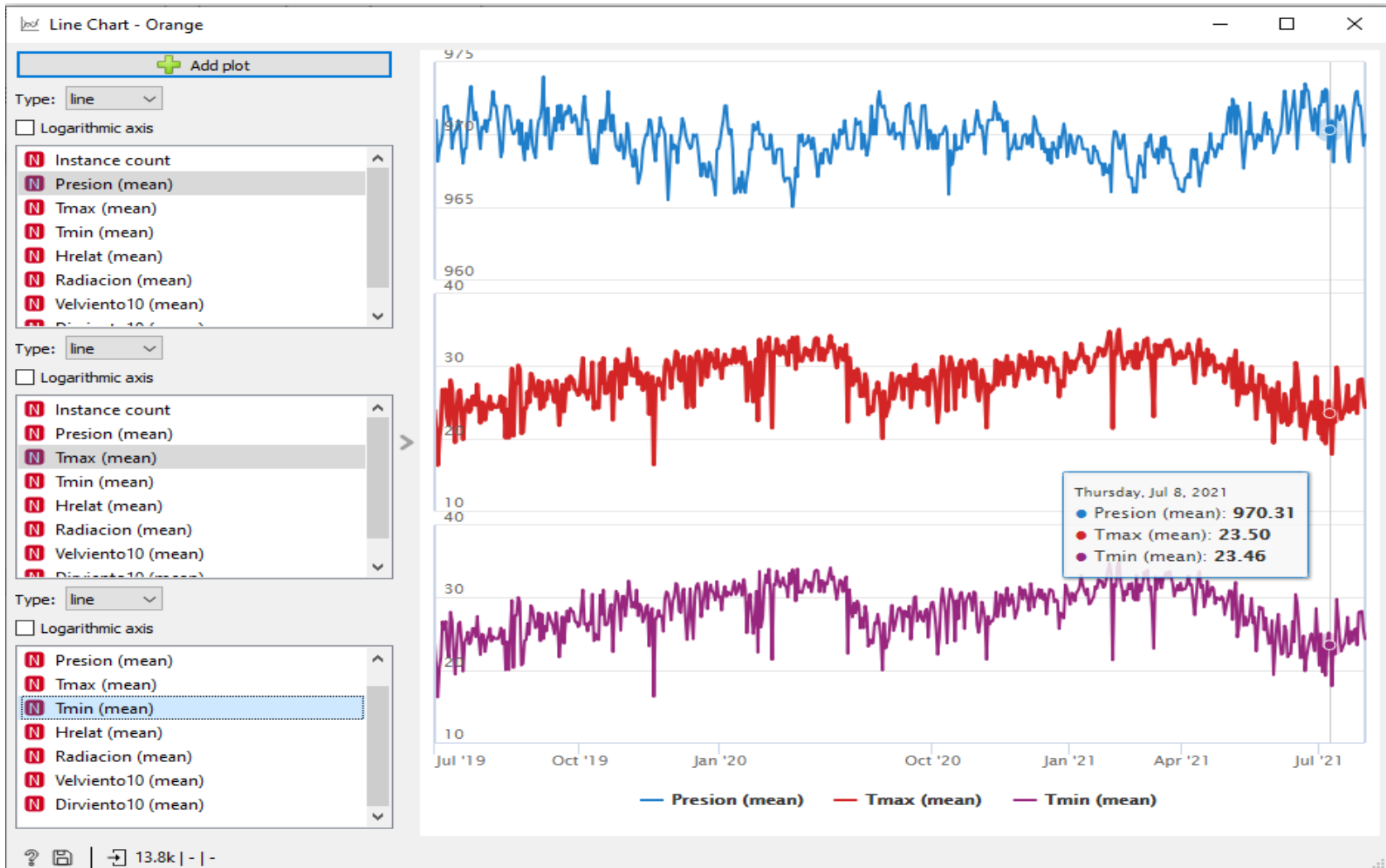


Figura 17. Comportamiento de la presión, temperatura máxima y temperatura mínima

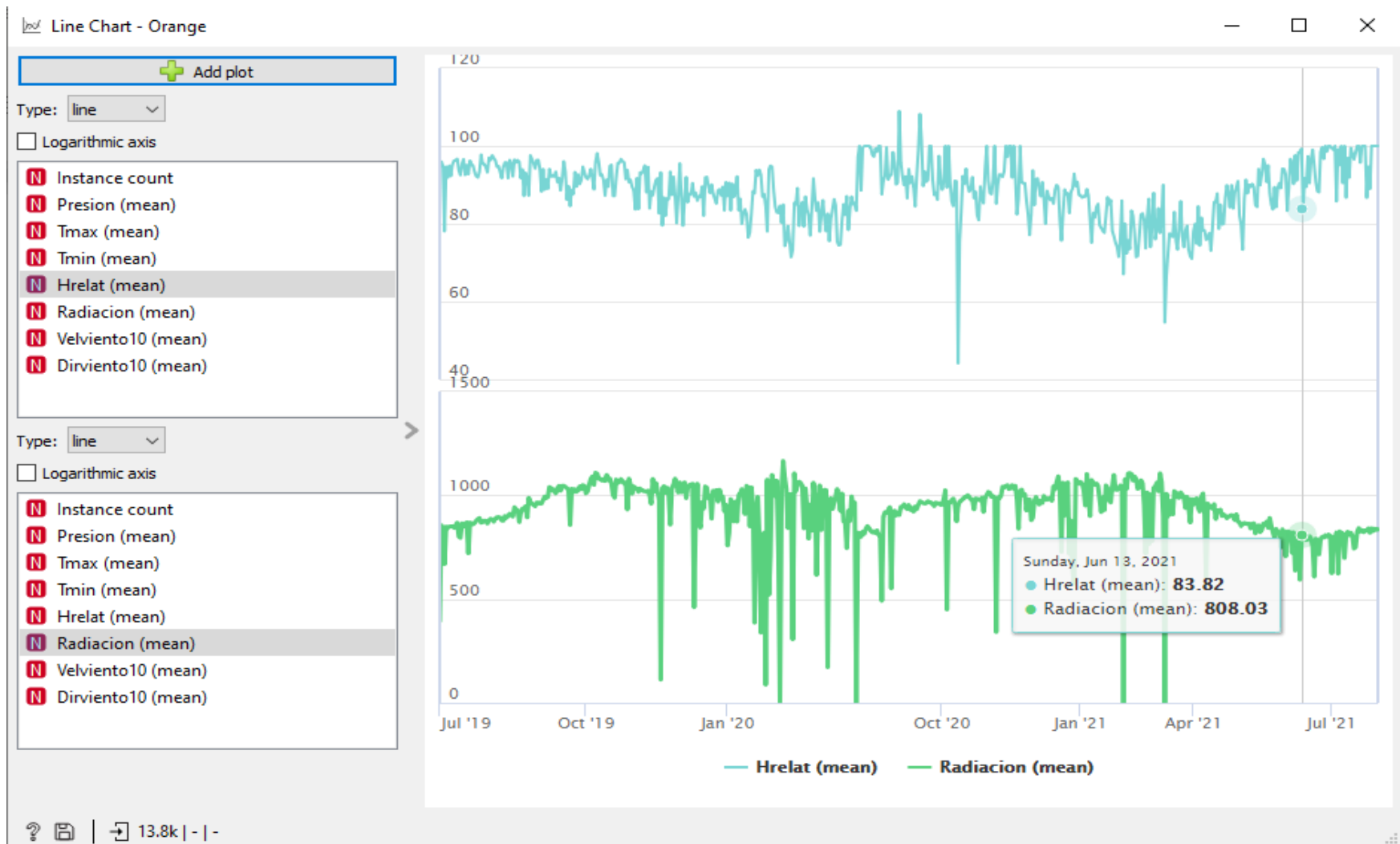


Figura 18. Comportamiento de la humedad relativa y la radiación

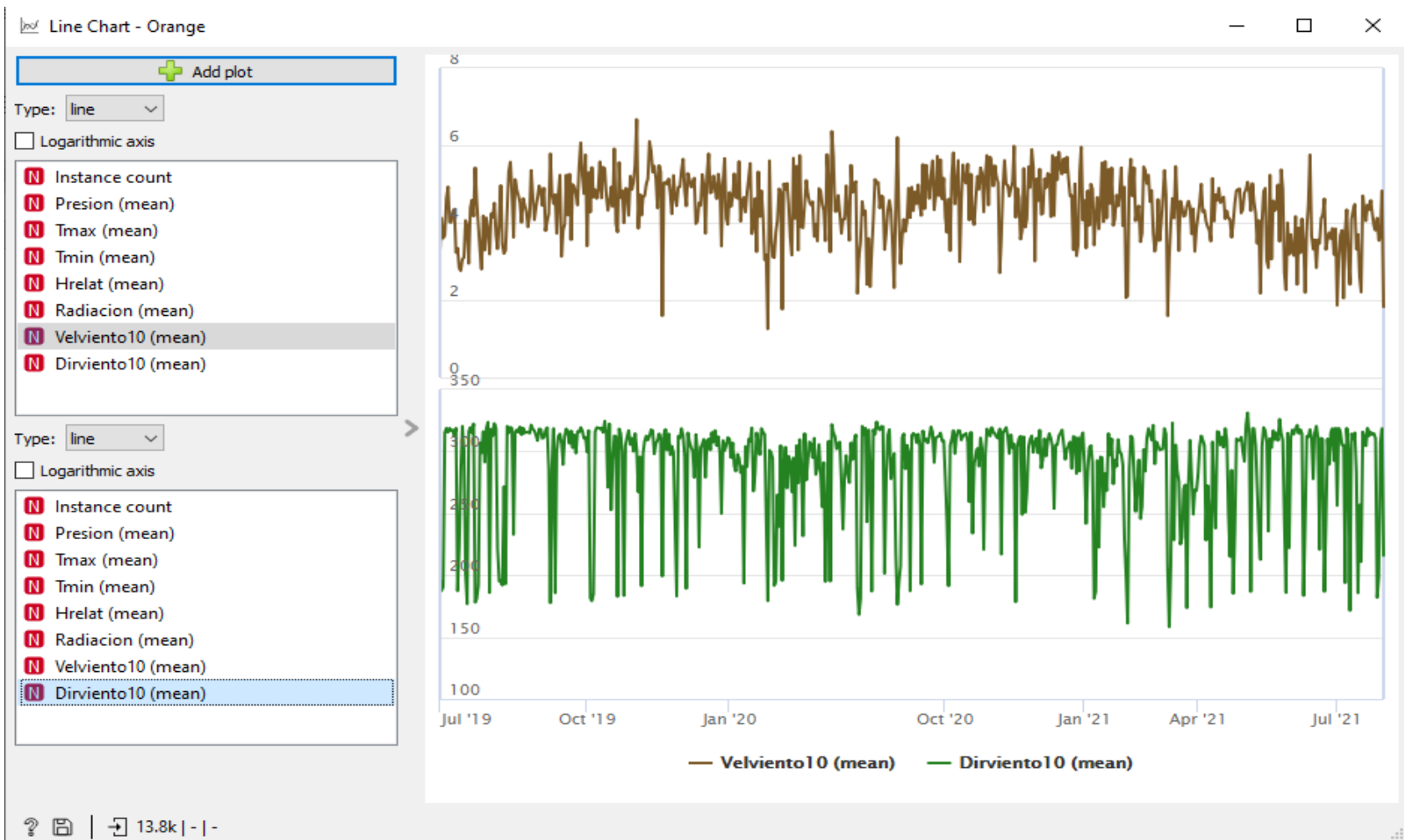


Figura 19. Comportamiento de la velocidad y dirección del viento

Nivel de correlación de las variables

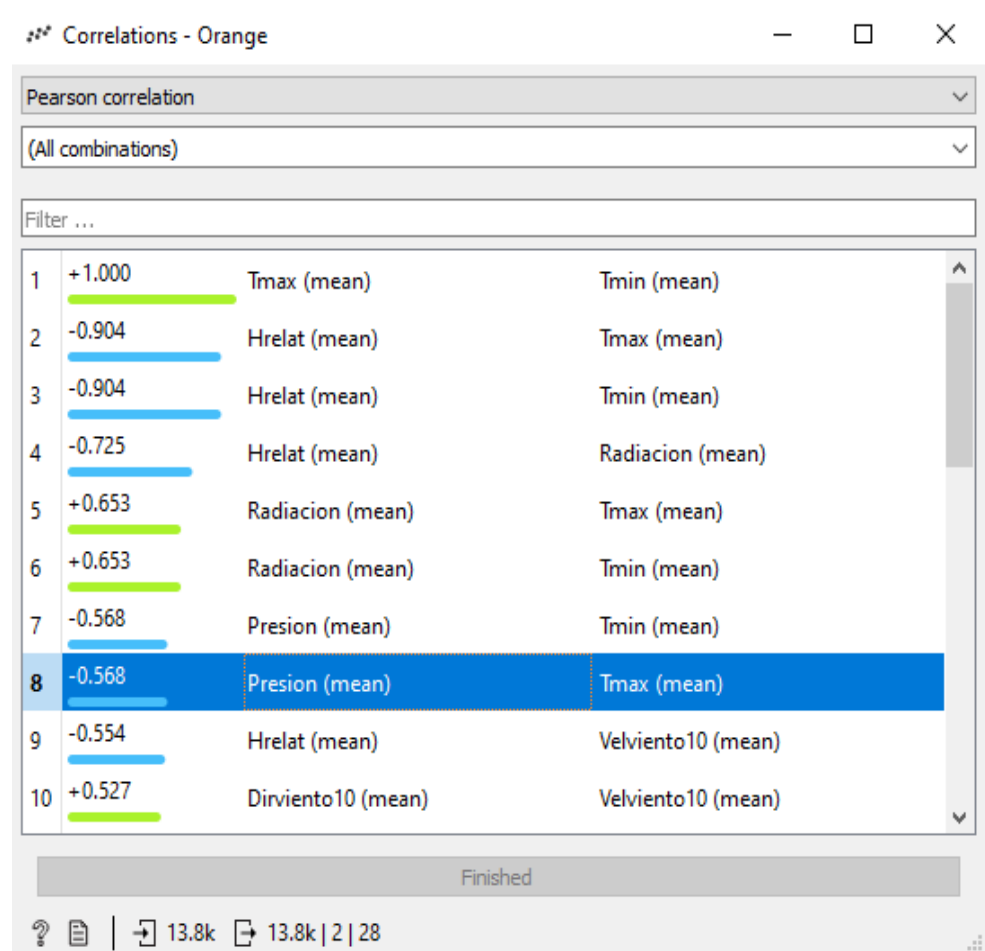


Figura 20. Análisis de correlación de Pearson de las variables

En la Figura 20, se muestra las correlaciones existentes entre las variables por medio de la prueba de correlación de Pearson, en la que se muestra relación entre las variables de la Hrelat con la Tmax y Tmin, una correlación inversa (-), además de la correlación (+) existente de la Radiación con las Temperaturas, la presión con una correlación (-) menor.

En las Figuras 21 y 22, el diagrama de Sieve evidencia la relación existente entre las variables analizadas, destacando un valor de $p=0.000$ en la contrastación de hipótesis, lo que indica significancia estadística. Asimismo, esta representación gráfica permite identificar y jerarquizar las características con mayor capacidad predictiva, facilitando la selección de las variables predictoras más relevantes para el modelo aplicado.

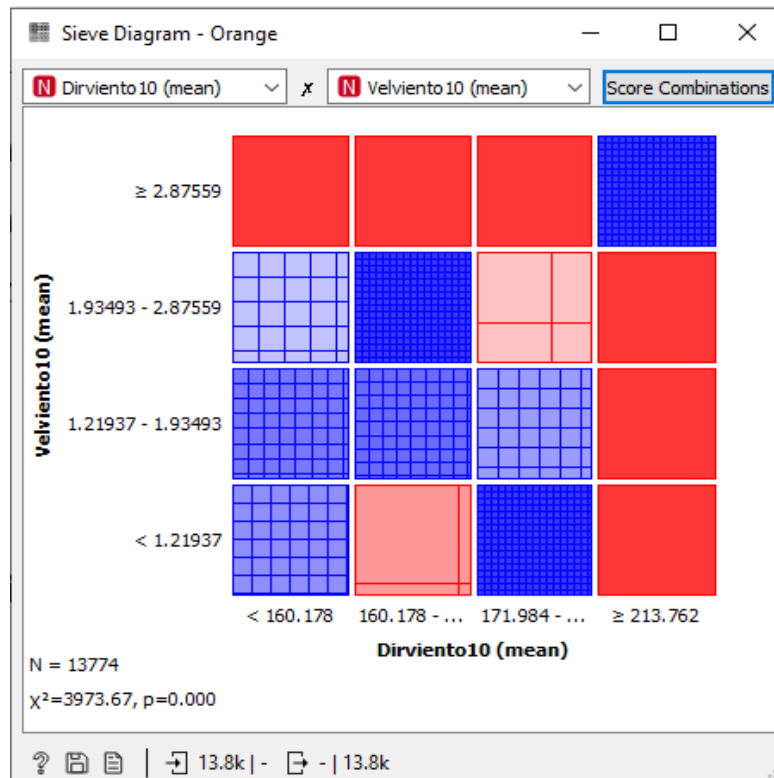


Figura 21. Relación entre las características

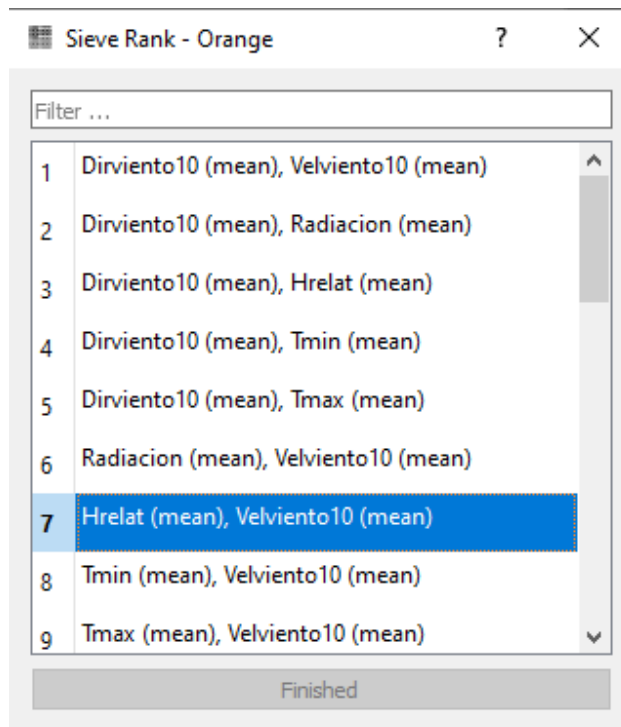


Figura 22. Rankin de relación de las características

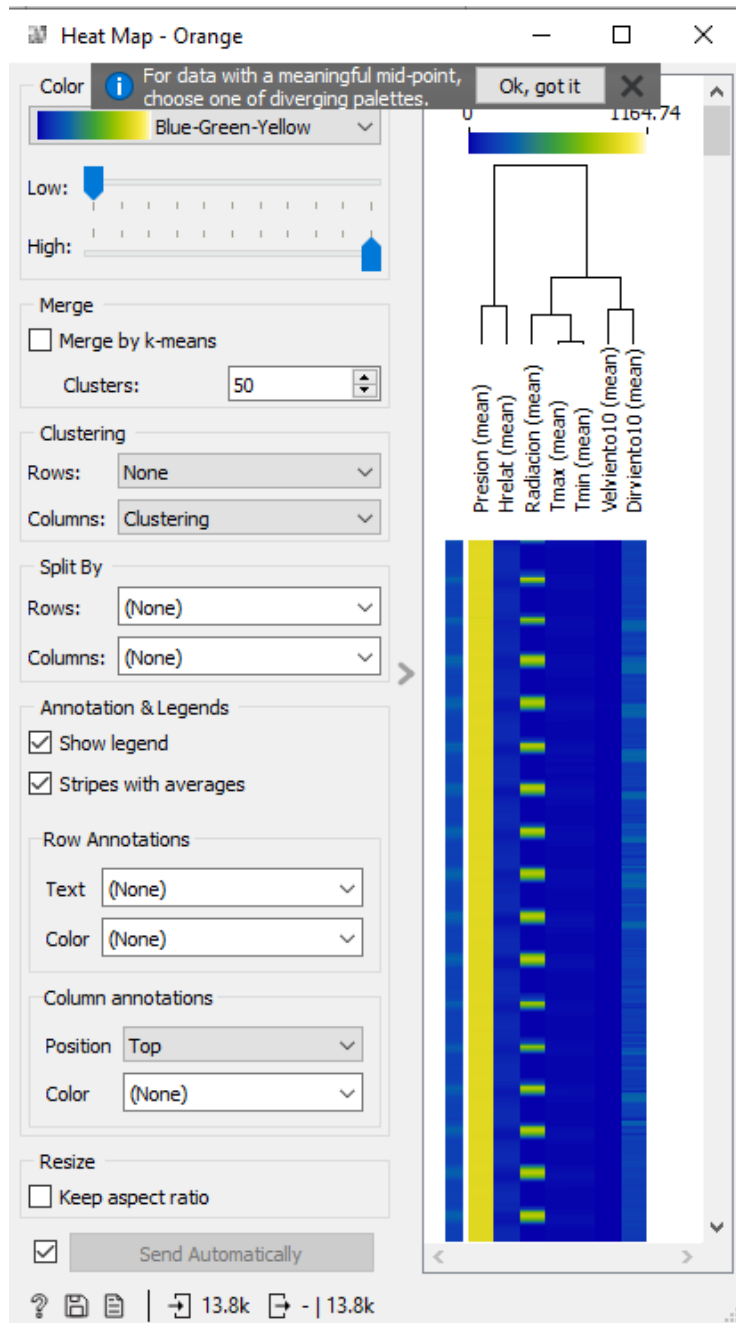


Figura 23. Mapa de calor de las características

En la Figura 23 del mapa de calor de las características, se muestra la correlación de dichas características, pero además nos presenta el clúster jerárquico de las características. En este clúster se aprecia la relación entre la dirección y la velocidad del viento y las temperaturas y la radiación; de otro lado, también se muestra el clúster de la presión y la humedad relativa.

3.2.2 Hipótesis específica (1)

H.E.1₀: Las condiciones atmosféricas NO influyen en el pronóstico climático mediante el modelo ARIMA de inteligencia artificial en la estación solar de Ica, 2025.

H.E.1_a: Las condiciones atmosféricas influyen en el pronóstico climático mediante el modelo ARIMA de inteligencia artificial en la estación solar de Ica, 2025

Las condiciones atmosféricas, al presentar variaciones continuas en parámetros como la temperatura, la humedad, el viento o la presión, modifican los patrones de las series temporales utilizadas por modelos como ARIMA. Estas fluctuaciones inciden directamente en la capacidad del modelo para identificar regularidades, por lo que su análisis resulta clave para generar predicciones más ajustadas en contextos locales como la estación solar de Ica.

Interpretación

El análisis realizado confirma que las condiciones atmosféricas ejercen una influencia significativa sobre el pronóstico climático generado mediante el modelo ARIMA en la estación solar de Ica. Las variaciones observadas en variables como la temperatura máxima y mínima, la presión, la humedad relativa, la radiación solar y el viento demostraron incidir directamente en la estructura de las series de tiempo, afectando la forma en que el modelo reconoce tendencias, periodicidades y fluctuaciones. La coherencia entre los resultados estadísticos, la correlación entre variables y el desempeño del modelo validan que estas condiciones no actúan de manera aislada, sino como componentes determinantes en la capacidad predictiva del sistema. En consecuencia, la aceptación de la hipótesis alterna respalda que el comportamiento dinámico de la atmósfera es un factor clave para obtener estimaciones climáticas más precisas y contextualizadas para la región de Ica.

3.3. **Determinar la variabilidad de los parámetros meteorológicos que influyen en el desempeño predictivo del modelo ARIMA en la estación solar, Ica”.**

Comprender cómo fluctúan los parámetros meteorológicos es esencial para mejorar la eficacia de los modelos de pronóstico. La variabilidad de datos como la temperatura, la presión o la humedad puede impactar directamente en la capacidad predictiva del modelo ARIMA, generando desviaciones en los resultados si no se reconocen adecuadamente sus patrones. Al identificar esta variabilidad dentro del contexto local de la estación solar de Ica, se contribuye a optimizar la selección de variables y el ajuste del modelo, fortaleciendo así su precisión en escenarios climáticos reales

3.3.1. Transformación de los datos

a) Modelamiento y análisis de los datos

Prueba de algoritmos de predicción

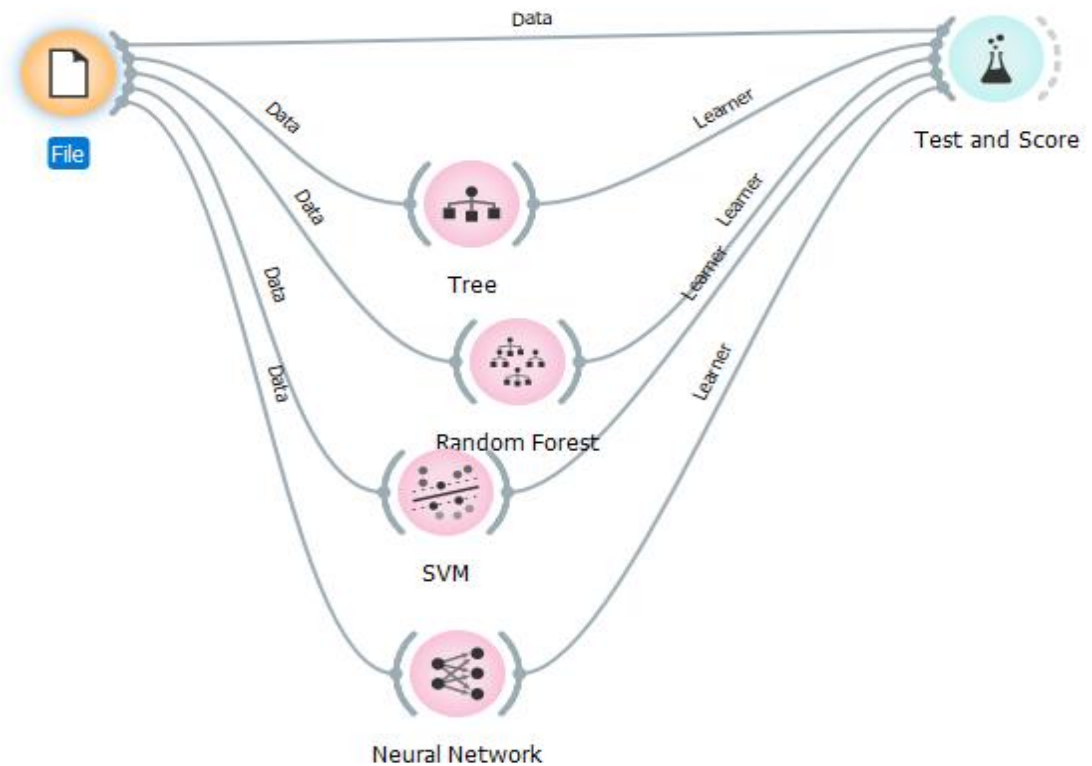


Figura 24. Modelo de pruebas de algoritmos de predicción

En la Figura 24. se muestra el modelo para la realización de la prueba de algunos algoritmos de predicción y determinar cuál de esos algoritmos es el más adecuado para nuestra serie de tiempos. Se están realizando pruebas con los algoritmos de árboles de decisión, Random Forest, SVM y Redes neuronales.

Entre todas las variables analizadas, se seleccionó la temperatura máxima como variable objetivo (target) del modelo predictivo. En todo proceso de predicción resulta fundamental definir previamente la característica que se desea estimar, ya que esta constituye el elemento principal sobre el cual se desarrollará el análisis y entrenamiento del modelo aplicado.

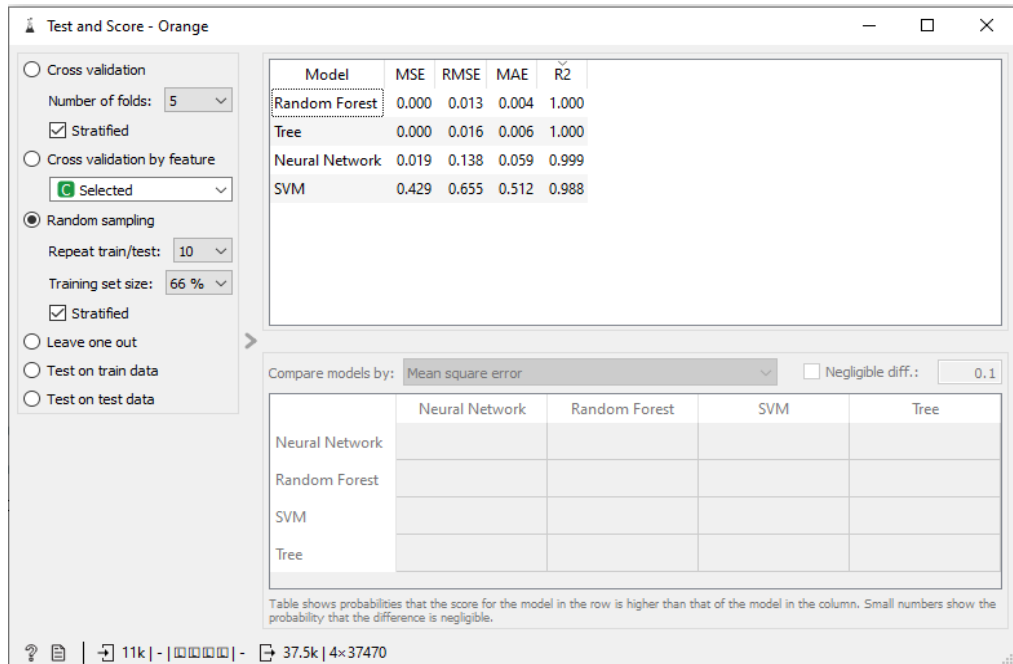


Figura 25. Resultados pruebas de algoritmos

La Figura 25 muestra los resultados obtenidos mediante los algoritmos de evaluación MSE (Mean Square Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error) y R² (Coefficient of Determination), utilizando una base de datos de entrenamiento equivalente al 66 % y una validación repetida en diez iteraciones. Los análisis realizados evidencian que los algoritmos Random Forest y Tree presentan el mejor desempeño predictivo para la variable Tmax. Asimismo, ambos modelos alcanzaron un coeficiente de determinación de 1.00, indicando una alta precisión y capacidad de ajuste en la predicción de la temperatura máxima dentro del conjunto de datos analizado..

b) Presentación de los resultados

	Tmax (mean)	Hours	Selected	Neural Network	Random Forest	SVM	Tree
1	19.3419	2019-07-17 17:0...	No	19.4114	19.343	19.9978	19.342
2	12.9759	2020-05-21 01:0...	No	13.0868	12.974	11.7479	12.9803
3	25.7617	2021-03-29 09:0...	No	25.6706	25.7625	26.0274	25.769
4	21.4895	2020-03-16 06:0...	No	21.5085	21.4842	20.8365	21.4905
5	18.1023	2021-01-14 05:0...	No	18.1139	18.1028	17.0753	18.1039
6	10.0245	2019-10-14 05:0...	No	9.91122	10.0344	9.60482	10.0454
7	31.9866	2019-12-30 13:0...	No	31.8715	31.9778	32.3657	31.9975
8	11.6892	2019-07-15 23:0...	No	11.7315	11.6867	11.2426	11.7011
9	29.4691	2020-03-10 10:0...	No	29.4344	29.4674	29.8421	29.4757
10	11.1906	2019-07-27 09:0...	No	12.234	11.1956	14.5617	11.1965
11	10.1915	2019-09-17 04:0...	No	10.1743	10.203	9.31379	10.209
12	23.2123	2020-11-14 10:0...	No	23.1082	23.2091	22.8646	23.2078
13	18.2934	2019-12-10 06:0...	No	18.3188	18.2918	17.7938	18.2918
14	31.0499	2021-03-30 14:0...	No	30.9529	31.0478	30.0362	31.0752
15	12.1726	2019-07-15 08:0...	No	12.1828	12.1721	11.2245	12.177
16	22.5983	2019-09-15 12:0...	No	22.5583	22.6052	22.7281	22.6032
17	11.7044	2019-09-11 01:0...	No	11.706	11.712	10.7718	11.7164
18	20.5959	2020-02-18 06:0...	No	20.661	20.5888	19.8642	20.5939
19	18.9916	2020-10-24 08:0...	No	19.1079	18.9889	18.9797	18.9959
20	23.9799	2019-11-22 11:0...	No	23.9914	23.975	23.8502	23.982

Figura 26. Resultados de los algoritmos

En la Figura 26. se puede apreciar los resultados de las predicciones de los algoritmos seleccionados. Analizando los valores originales de Tmax (color gris) y los valores de predicción de los algoritmos, se desprende por el acercamiento de dichos valores como **Random Forest** muestra valores más cercanos al original (19.3419 – 19.343), seguido de Tree (19.3419 – 19.342).

Conociéndose que el algoritmo de **Random Forest** es el mejor predictor, este se utilizará para hacer la predicción de las temperaturas máximas (Tmax) seleccionada para las pruebas.

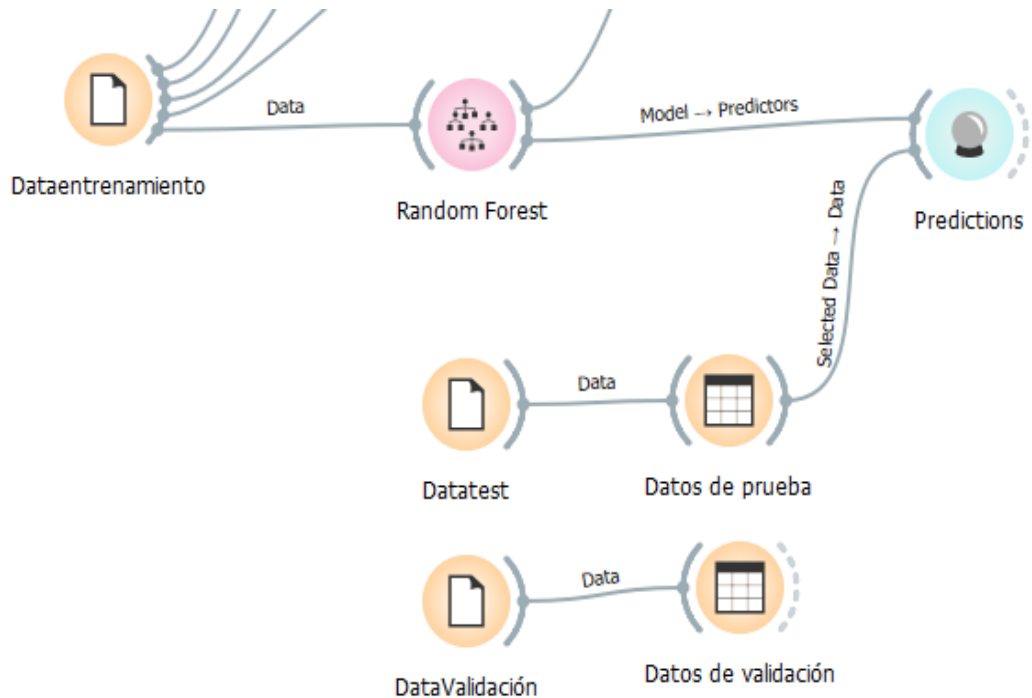


Figura 27. Modelo de predicción con Random Forest

En la Figura 27 se observa el comportamiento del modelo Random Forest aplicado para la predicción de la temperatura máxima (Tmax). Para el desarrollo del análisis se empleó una distribución de datos compuesta por un 80 % destinado al entrenamiento del modelo y un 20 % correspondiente a los datos de prueba (Datatest). Esta configuración permitió evaluar la capacidad predictiva y el nivel de ajuste del algoritmo frente a información no utilizada durante el entrenamiento. Asimismo, los resultados obtenidos evidencian el desempeño del modelo en la estimación de Tmax, contribuyendo a validar su precisión, confiabilidad y eficiencia en el proceso predictivo desarrollado.

Datos de prueba - Orange

Info
2755 instances
8 features (12.5 % missing data)
No target variable.
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order

Send Automatically

2755 | 2755 | 2755

	Presion (mean)	Tmin (mean)	Hrelat (mean)	Radiacion (mean)	velviento10 (mean)	Dirviento10 (mean)	Tmax (mean)
1	966	19.6983	70.9292	0	1.50701	154.953	? 2
2	966	19.6049	71.036	0	0.889236	161.612	? 2
3	966	19.3144	72.3995	0	1.11097	172.685	? 2
4	965.144	18.9058	73.8229	0	1.10576	171.093	? 2
5	965	18.695	74.7182	0	0.525556	139.711	? 2
6	964.203	18.6536	74.5233	0	0.280556	180.213	? 2
7	964	18.2861	76.5633	0	0	149.637	? 2
8	964	18.2245	76.3162	0	0.300139	151.638	? 2
9	964.228	17.92	77.9136	0	0.945417	199.124	? 2
10	964.939	18.3561	75.8583	73.8819	1.20521	172.542	? 2
11	965.733	20.3858	68.0219	267.453	1.85118	167.033	? 2
12	966.069	22.5828	60.585	453.724	1.95896	167.983	? 2
13	966	24.7331	53.7492	699.99	1.69903	171.37	? 2
14	965.458	27.1084	45.7733	862.179	1.49854	182.448	? 2
15	964.625	28.9465	40.1221	956.086	2.30722	198.373	? 2
16	963.361	30.5312	36.0163	949.082	2.57715	221.29	? 2
17	962.561	31.0011	34.9724	937.196	3.71347	280.825	? 2
18	962	30.962	34.9141	781.634	4.16875	275.818	? 2
19	961.692	30.4441	35.6834	582.889	3.91896	204.973	? 2
20	961.992	29.5403	37.3733	321.802	2.85521	202.1	? 2
21	962.867	25.9767	46.2863	57.3412	3.55875	165.581	? 2
22	963.925	23.0599	55.6322	0.449714	3.16646	164.631	? 2

Figura 28. Datos de prueba (Datatest)

En la Figura 28. los datos de prueba se precisan que dichos datos no cuentan con el valor de la Tmax, el cual se ha eliminado intencionalmente para poder ver los resultados de la predicción y compararlo con el archivo de datos de validación (*Datavalidación*) que si tiene los valores de Tmax.

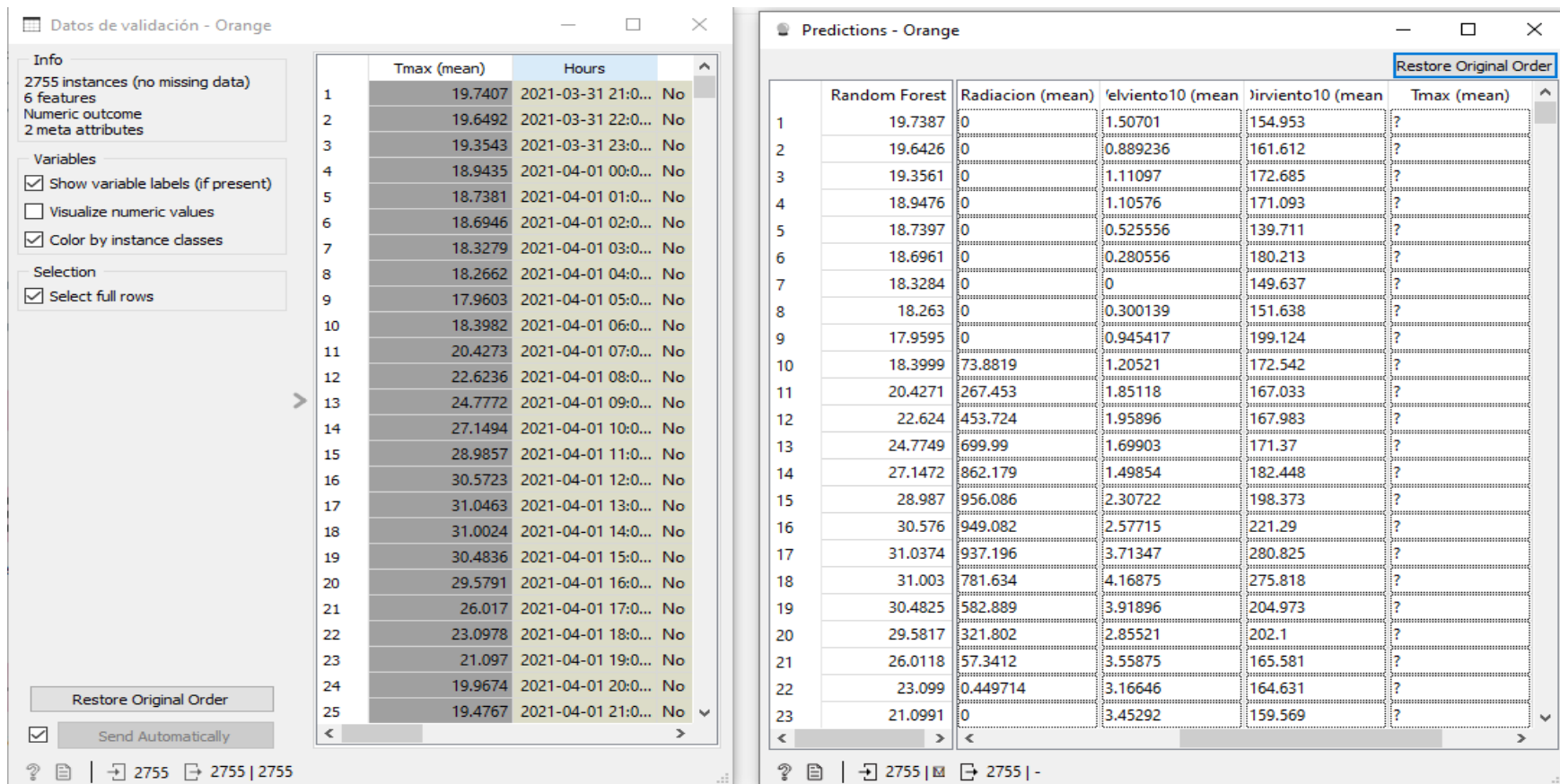


Figura 29. Comparación con datos de validación

En la Figura 29. se muestra la comparación de los valores predichos con el **Random Forest** y los datos de validación (datos reales) de la Tmax, los resultados demuestran cómo el valor del coeficiente de determinación de este algoritmo (1.00) arroja resultados muy precisos de predicción por lo cual el modelo cumple a satisfacción la investigación.

3.3.2. Hipótesis específica (2)

H.E.2₀: La variabilidad de los parámetros meteorológicos NO influye en el desempeño predictivo del modelo ARIMA en la estación solar, Ica, 2025”.

H.E.2_a: Las condiciones atmosféricas influyen en el pronóstico climático mediante el modelo ARIMA de inteligencia artificial en la estación solar de Ica, 2025
Las condiciones atmosféricas, al registrar variaciones constantes en parámetros como temperatura, humedad, radiación y viento, influyen directamente en la estructura de las series de tiempo. Esta influencia afecta la precisión del modelo ARIMA, ya que su capacidad predictiva depende de la estabilidad y patrones de dichas variables en el contexto climático local de la estación solar de Ica.

Interpretación

El análisis llevado a cabo confirma que la variabilidad de los parámetros meteorológicos incide significativamente en el desempeño predictivo del modelo ARIMA, validando así la hipótesis alterna. La implementación de pruebas comparativas entre diferentes algoritmos —incluyendo Random Forest, árboles de decisión, SVM y redes neuronales— permitió identificar que la fluctuación en variables como la temperatura, humedad, presión, radiación y viento determina la capacidad del modelo para realizar estimaciones precisas.

La selección de la temperatura máxima (Tmax) como variable objetivo fue adecuada para evaluar el comportamiento predictivo. La reducción de la frecuencia de muestreo a intervalos horarios, así como la limpieza previa de los datos, permitieron construir una base consistente para el modelado. Los resultados obtenidos mostraron que, si bien varios algoritmos ofrecieron desempeños aceptables, **Random Forest destacó con un coeficiente de determinación (R²) igual a 1.00**, reflejando una correspondencia prácticamente perfecta entre los valores reales y los predichos.

La evidencia estadística obtenida permite afirmar que el comportamiento dinámico de los parámetros meteorológicos no debe interpretarse como simples fluctuaciones aleatorias o ruido dentro del análisis de datos, sino como un elemento fundamental que requiere ser incorporado de manera adecuada en los modelos de predicción climática. Los resultados alcanzados demuestran que la variabilidad presente en las condiciones atmosféricas influye directamente en el desempeño y precisión de los modelos predictivos empleados en la investigación. En este sentido, la exactitud de las predicciones no depende únicamente del algoritmo o método estadístico utilizado,

sino también del proceso de tratamiento, depuración, selección e interpretación de las variables meteorológicas consideradas en el estudio.

Asimismo, la integración adecuada de dicha variabilidad permitió mejorar significativamente el ajuste del modelo y la confiabilidad de los resultados obtenidos. Esto evidencia que una caracterización detallada y técnicamente sustentada de los parámetros climatológicos favorece la representación real del comportamiento atmosférico en escenarios locales. De esta manera, se valida empíricamente la importancia de incorporar análisis previos de las variables meteorológicas para optimizar modelos predictivos como ARIMA, especialmente en investiga.

IV. DISCUSIÓN DE RESULTADOS

4.1. Discusión de resultados de la variabilidad de los parámetros meteorológicos y el pronóstico climático utilizando el modelo ARIMA basado en inteligencia artificial, en la estación solar de Ica

Los parámetros meteorológicos destinados a medir los fenómenos meteorológicos se obtuvieron de la estación meteorológica, que está situada en la Universidad Nacional San Luis Gonzaga, como se observó en la Figura 2. En la estación meteorológica en el Centro de Investigación Científica, cuenta con un área especial donde se encuentra la estación. La estación meteorológica tiene área libre para la circulación del aire y visibilidad en todas las direcciones, en ella se registró los datos desde el 01 de julio del 2019 hasta el 09 de febrero del 2022, la misma que cuenta con 7'142,465 registros con 16 características, donde no se tiene data perdida (no missing data) según esta primera evaluación. Estos datos se utilizaron para la elaboración de predicciones meteorológicas a partir de los valores predichos con el **Random Forest**, para estudios climáticos. Con los instrumentos meteorológicos y sensores que con estos dispositivos diseñados que se tomaron de la instalación de la torre meteorológica del Centro de Investigación del Estudio de la Actividad Solar y sus Efectos Sobre la Tierra, que permitió detallar la respuesta cuantitativa de un objeto (los sensores) que ha contribuido en obtener un producto real en la influencia ambiental y con los siguientes instrumentos: Termohigrómetro (midió la temperatura ambiental-°C y la humedad relativa-%), Piranómetro (midió la radiación solar directa y difusa-watts/m²), El Anemómetro permitió registrar la velocidad del viento expresada en m/s, mientras que el Barómetro fue utilizado para medir la presión atmosférica en mbar. Los datos recopilados respondieron a las condiciones ambientales presentes durante el periodo de monitoreo, conforme se describe en la Figura 7. Asimismo, los parámetros meteorológicos obtenidos fueron representados y organizados en las Figuras 8 y 9, mostrando el comportamiento de las variables evaluadas. Después del proceso de selección y depuración de información, se contó con un total de 5'503,074 registros válidos para el desarrollo del análisis correspondiente.

4.2. Discusión de resultados de las condiciones atmosféricas que influyen en el pronóstico climático utilizando el modelo ARIMA de inteligencia artificial en la estación solar de Ica

El monitoreo de los parámetros meteorológicos en la recogida de datos es fundamental para predecir la presión, temperatura máxima (Tmax), temperatura mínima (Tmin), humedad relativa (Hrelat), radiación, velocidad del viento a los 10 metros (Velviento10) y dirección del viento a los 10 metros. (Dirviento10), con estos datos se seleccionó las características para estudiar, los datos propuestos, que han sido recogidos por medio de los sensores de la estación meteorológica con una periodicidad de cada 10 segundos, por espacios de dos periodos ósea dos años (2 años), que nos permitió ver en la Figura 13, donde se aprecia los datos que son 13,774 registros y 9 características, a estos datos se aplicó la limpieza de los datos, para tal efecto se aplicó la estadística descriptiva y con análisis de grafico para predecir el comportamiento del clima, en la investigación, se mostró que en la figura 15 donde se observó la mediana de cada uno de los parámetros como la presión 967.847 mbar, Tmax. 19.64 °C, Tmin. 19.6 °C, Humedad relativa 67.872 %, Radiación 249.642 watts/m², Velocidad del viento a 10 metros 1.925 m/s y Dirección del viento a 10 metros 174.7 m/s. En la Figura 20 mostro el diagrama de Sieve que relaciona las dos variables de estudio, para la muestra N=13774, con una desviación estándar de 3973.67 para un p=0.000, además el mapa de calor relacionó los parámetros meteorológicos, que permitió monitorear las mediciones y observaciones puntuales utilizando los instrumentos adecuados para así predecir el comportamiento atmosférico, donde las características de los datos están dentro de lo normal en la región de Ica. Por lo que la irradiancia que llega a la atmósfera superior de la Tierra es definida como constante solar, teniendo por valor, el establecido por la NASA de 1353 W/m² y de 1373 W/m² según la WMO (Organización Mundial de Meteorología).

4.3. Discusión de resultados en la variabilidad de los parámetros meteorológicos que influyen en el desempeño predictivo del modelo ARIMA en la estación solar, Ica

De la fuente estación meteorológica se recabo los datos, para su análisis en las condiciones meteorológicas que se permitió propiciaron algún fenómeno atmosférico determinado, para tal efecto se utilizó el **modelo de pruebas de algoritmos de árboles de decisión, Rendón Forest, SVM y Redes neuronales**, en la Figura 24 se presentó los resultados de los algoritmos MSE (Mean square error), RMSE (Root Mean square error), MAE (Mean absolute error) y R2 (Coeficiente of determination), para una data de entrenamiento de 66% y con una repetición de 10 veces y en la Figura 26 se pudo apreciar el **modelo de Random Forest** para la predicción de Tmax, utilizando los datos de entrenamiento (80% de los datos total) y los datos de prueba (Datatest 20% de los datos total), los resultados demostraron como el valor del coeficiente de

determinación de este algoritmo (1.00) arrojaron resultados muy precisos de predicción por lo cual el modelo cumple a satisfacción la investigación para la Tmax., por lo tanto en la ciudad de Ica posee un clima cálido y seco, con una temperatura media en verano de 27°C y en invierno de 18°C. Normalmente, la temperatura máxima no excede de 30°C y la mínima no desciende de 8°C”.

V. CONCLUSIONES

1. Se concluye que los registros meteorológicos recolectados por la estación solar de la Universidad Nacional San Luis Gonzaga permitieron evidenciar que la información obtenida, al estar debidamente estructurada, completa y respaldada por sensores calibrados, constituye una base sólida para la modelación predictiva. La ausencia de datos perdidos y la alta frecuencia de registros facilitaron la identificación de patrones relevantes en variables como la temperatura, humedad, radiación, presión y viento. La magnitud y calidad de esta base de datos fortalecen el uso del modelo ARIMA en la estimación climática, mostrando que la integración de herramientas de inteligencia artificial con sensores de monitoreo continuo aporta precisión y valor operativo al pronóstico del clima en contextos regionales como el de Ica.
2. Se concluye que los resultados obtenidos evidencian que las condiciones atmosféricas monitoreadas en la estación solar poseen un impacto real y cuantificable sobre la proyección del clima cuando se emplean modelos de predicción como ARIMA. El tratamiento riguroso de los registros, incluyendo el análisis estadístico y gráfico de variables clave, permitió establecer correlaciones sólidas entre factores como la temperatura, la humedad, la radiación y el viento, confirmando su relevancia en la modelación de escenarios climáticos. La consistencia de los datos y la solidez de los indicadores utilizados refuerzan el valor científico de integrar estos parámetros en sistemas predictivos automatizados, consolidando así la utilidad de la inteligencia artificial para representar con mayor exactitud el comportamiento atmosférico en zonas específicas como la región de Ica.
3. Se concluye que la comparativo entre algoritmos de predicción demostró que la variabilidad de los parámetros meteorológicos incide de manera determinante en la precisión del modelo ARIMA, especialmente al predecir la temperatura máxima (T_{max}) en condiciones climáticas locales. La aplicación del modelo Random Forest permitió identificar relaciones sólidas entre las variables atmosféricas y los resultados esperados, alcanzando un nivel óptimo de ajuste con un coeficiente de determinación igual a 1.00, lo cual valida la calidad del modelo predictivo. Este resultado se sostiene en una base de datos robusta y debidamente transformada, proveniente de la estación meteorológica del centro de investigación solar, lo que evidencia que una adecuada gestión de la variabilidad meteorológica mejora significativamente el

rendimiento de modelos aplicados a series de tiempo. Así, se ratifica que las condiciones particulares del clima en Ica, caracterizadas por temperaturas estables y una atmósfera predominantemente seca, permiten el desarrollo de sistemas de predicción.

VI. RECOMENDACIONES

- 6.1. Se recomienda aprovechar de manera sistemática la base de datos meteorológica generada por la estación solar del Centro de Investigación Científica de la Universidad Nacional San Luis Gonzaga, integrándola en procesos permanentes de análisis y modelamiento predictivo. La aplicación de modelos como ARIMA, respaldada por registros confiables y sin pérdida de información, debe fortalecerse mediante rutinas automatizadas de validación, actualización y uso de herramientas de inteligencia artificial accesibles y de libre licencia. Asimismo, es necesario fomentar el uso de esta información en la toma de decisiones institucionales y en la gestión local del recurso climático, promoviendo su difusión entre sectores estratégicos como la agricultura, energía renovable y gestión del riesgo climático en Ica.
- 6.2. Se recomienda fortalecer el aprovechamiento de los registros atmosféricos obtenidos en la estación solar mediante su integración en plataformas de análisis predictivo basadas en inteligencia artificial, como ARIMA, bajo protocolos de depuración, transformación y selección de variables adecuadamente establecidos. Asimismo, es recomendable establecer rutinas sistemáticas de revisión de datos que permitan mantener su consistencia y calidad, incorporando procedimientos de verificación estadística que respalden la validez de las predicciones generadas. Esta estrategia contribuirá a mejorar la respuesta del modelo frente a los cambios locales del clima y permitirá una planificación más eficiente en sectores sensibles a la variabilidad meteorológica.
- 6.3. Se recomienda implementar un sistema regular de análisis predictivo basado en modelos de series temporales, como ARIMA, alimentado por parámetros meteorológicos seleccionados y validados por su capacidad de predicción comprobada. Para ello, es fundamental mantener actualizada la base de datos climatológica de la estación solar, estandarizar el preprocesamiento de los datos, incluyendo limpieza, transformación y selección de variables, e integrar herramientas de inteligencia artificial como Random Forest para complementar el modelamiento clásico. Además, se sugiere fortalecer la infraestructura de almacenamiento y procesamiento local, con el fin de asegurar la sostenibilidad y aplicabilidad práctica de estos modelos en contextos regionales como el clima árido de Ica.

VII. REFERENCIAS BIBLIOGRÁFICAS

- [1] MINAM, “Objetivos de Desarrollo Sostenible del Milenio e Indicadores. Ministerio del Ambiente. Dirección General de Investigación e Información.,” 2016, *Lima - Perú*.
- [2] PNUD, “Apoyo del PNUD a la Implementación del Objetivo de Desarrollo Sostenible 6: Gestión Sostenible del Agua y el Saneamiento,” 2016.
- [3] M. Nsabagwa, M. Byamukama, E. Kondela, and J. S. Otim, “Towards a robust and affordable Automatic Weather Station,” *Dev. Eng.*, vol. 4, no. May 2018, p. 100040, 2019, doi: 10.1016/j.deveng.2018.100040.
- [4] J. Hansen, M. Sato, and R. Ruedy, “Perception of climate change,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 37, 2012, doi: 10.1073/pnas.1205276109.
- [5] T. Dube, P. Moyo, M. Ncube, and D. Nyathi, “The Impact of Climate Change on Agro-Ecological Based Livelihoods in Africa: A Review,” *J. Sustain. Dev.*, vol. 9, no. 1, p. 256, 2016, doi: 10.5539/jsd.v9n1p256.
- [6] D. N. Mubiru, F. B. Kyazze, M. Radeny, A. Zziwa, J. Lwasa, and J. Kinyangi, *Climatic trends, risk perceptions and coping strategies of smallholder farmers in rural Uganda*. Uganda: Working Paper No. 121 CGIAR, 2015.
- [7] S. Lwasa, F. Mugagga, B. Wahab, D. Simon, J. Connors, and C. Griffith, “Urban and peri-urban agriculture and forestry: Transcending poverty alleviation to climate change mitigation and adaptation,” *Urban Clim.*, vol. 7, pp. 92–106, 2014, doi: 10.1016/j.uclim.2013.10.007.
- [8] D. Kreuzer, M. Munz, and S. Schlüter, “Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany,” *Mach. Learn. with Appl.*, vol. 2, no. June, p. 100007, 2020, doi: 10.1016/j.mlwa.2020.100007.
- [9] G. Dumas, V. Masson, J. Hidalgo, V. Edouart, A. Hanna, and G. Poujol, “Co-construction of climate services based on a weather stations network: Application in Toulouse agglomeration local authority,” *Clim. Serv.*, vol. 24, 2021, doi: 10.1016/j.cliser.2021.100274.
- [10] R. Cavelier *et al.*, “Conditions for a market uptake of climate services for adaptation in France,” *Clim. Serv.*, vol. 6, pp. 34–40, 2017, doi: 10.1016/j.cliser.2017.06.010.
- [11] J. Castillo, “Organización Meteorología Mundial.,” Estaciones Meteorológicas.
- [12] O. Peña, “Sistema De Información De Mini Estación Metereológica Para La Captura De Datos

De Temperatura, Humedad Y Presión Atmosférica - Centro De Operaciones De Emergencia Regional - Piura,” Universidad Nacional de Piura, 2019.

- [13] F. Zapata, “¿Qué es una estación meteorológica?,” Inst. PCE.
- [14] C. Campetella, B. Cerne, and P. Salio, “Entornos invisibles, estación metereológica,” in *Ciencia y Tecnología*, Argentina: Ministerio de Educación - Instituto Nacional de Educación Tecnológica, 2011, ch. Estación m, pp. 1–52.
- [15] B. Mohammadi *et al.*, “Application of ERA-Interim, empirical models, and an artificial intelligence-based model for estimating daily solar radiation,” *Ain Shams Eng. J.*, vol. 13, no. 1, p. 12 Pag., 2021, doi: 10.1016/j.asej.2021.05.012.
- [16] İ. Uckan, & Kameran, and M. Khudhur, “International Journal of Green Energy Estimation and comparison of various global solar radiation models on horizontal surface in a hot and dry region,” 2018, doi: 10.1080/15435075.2018.1454322.
- [17] J. Zhang, L. Zhao, S. Deng, W. Xu, and Y. Zhang, “A critical review of the models used to estimate solar radiation,” *Renew. Sustain. Energy Rev.*, vol. 70, no. July 2016, pp. 314–329, 2017, doi: 10.1016/j.rser.2016.11.124.
- [18] O. A. Olanike and A. M. Chandra, “Evaluation of solar radiation estimation methods for reference evapotranspiration estimation in Canada,” *Theor. Appl. Climatol. Vol.*, pp. 377–385, 2014, doi: <https://doi.org/10.1007/s00704-013-1070-2>.
- [19] B. . Jahani and B. Mohammadi, “A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran,” *Theor. Appl. Climatol.*, vol. 137, no. 31, 2019, doi: 10.1007/s00704-018-2666-3.
- [20] Y. Liu, Y. Zhou, D. Wang, Y. Wang, Y. Li, and Y. Zhu, “Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China,” *Energy Convers. Manag.*, vol. 154, no. July, pp. 168–179, 2017, doi: 10.1016/j.enconman.2017.10.043.
- [21] K. Yang, T. Koike, and B. Ye, “Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets,” *Agric. For. Meteorol.*, vol. 137, no. 1–2, pp. 43–55, 2006, doi: 10.1016/j.agrformet.2006.02.001.
- [22] L. Wang *et al.*, “Prediction of diffuse solar radiation based on multiple variables in China,” *Renew. Sustain. Energy Rev.*, vol. 103, no. July 2018, pp. 151–216, 2019, doi: 10.1016/j.rser.2018.12.029.
- [23] X. Liu, X. Mei, Y. Li, J. R. Porter, Q. Wang, and Y. Zhang, “Choice of the Ångström-Prescott coefficients: Are time-dependent ones better than fixed ones in modeling global solar irradiance?,” *Energy Convers. Manag.*, vol. 51, no. 12, pp. 2565–2574, 2010, doi:

- 10.1016/j.enconman.2010.05.020.
- [24] B. Mohammadi and S. Mehdizadeh, "Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm," *Agric. Water Manag.*, vol. 237, no. March, p. 106145, 2020, doi: 10.1016/j.agwat.2020.106145.
- [25] B. Yaniktepe, O. Kara, and C. Ozalp, "The global solar radiation estimation and analysis of solar energy: Case study for Osmaniye, Turkey," *Int. J. Green Energy*, vol. 14, no. 9, pp. 765–773, 2017, doi: 10.1080/15435075.2017.1329148.
- [26] V. Laqui, "Efecto de la variación de los elementos climaticos en la producción de quinua, en la cuenca del rio Ilave-Puno," Universidad Nacional del Altiplano, 2017.
- [27] C. Forgotson *et al.*, "How Satellite Soil Moisture Data Can Help to Monitor the Impacts of Climate Change: SMAP Case Studies," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 1590–1596, 2020, doi: 10.1109/JSTARS.2020.2982608.
- [28] D. S. Roy, "Forecasting the Air Temperature at a Weather Station Using Deep Neural Networks," *Procedia Comput. Sci.*, vol. 178, pp. 38–46, 2020, doi: 10.1016/j.procs.2020.11.005.
- [29] B. Mendoza, J. Ramirez, and R. Sandoval, "Efectos de la actividad solar en el clima y en los seres vivos," *Rev. Cienc.*, pp. 32–38, 2006.
- [30] M. Cortina, "Universidad polit ecnica de madrid," Universidad Politecnica de Madrid, 2012.
- [31] R. Rodríguez Jiménez, Á. Benito Capa, and A. Portela Lozano, *Meteorología y Climatología*. España.: FECYT (Fundación Española para la Ciencia y la Tecnología), 2004.
- [32] J. L. Huere Peña, "Patrones de Comportamiento de Datos Meteorologicos Mediante Tecnicas de Data Minig en Huancavelica 2018-2019," Universidad Nacional de Huancavelica, 2021.
- [33] WIKIPEDIA, "Mollendo," 2021, *La enciplopedia WIKIPEDIA, Arequipa*.
- [34] INEI, *Instituto Nacional de estadistica e Informatica. Sistema ESTADISTICO nacional*. Oficina Departamental de Estadistica e Informatica de ICA, 2017.
- [35] E. Mrzic and T. Zaimovic, "Data science methods and machine learning algorithm implementations for customized practical usage," *TEM J.*, vol. 9, no. 3, pp. 1179–1185, 2020, doi: 10.18421/TEM93-44.
- [36] J. De, S. Zanirato, M. Id, A. Paula, A. Bueno Id, and J. R. Sato, "Assessing the educational performance of different Brazilian school cycles using data science methods," 2021, doi: 10.1371/journal.pone.0248525.
- [37] K. Bastille *et al.*, "Improving the IEA Approach Using Principles of Open Data Science," *Coast. Manag.*, vol. 49, no. 1, pp. 72–89, 2020, doi: 10.1080/08920753.2021.1846155.

- [38] R. Hernandez, C. Fernandez, and P. Baptista, *Metodologia de la Investigacion*, Sexta Edic. Mexico: Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736, 2014.
- [39] M. Tamayo y Tamayo, *El Proceso de la Investigación Científica. Incluye evaluación y Administración de Proyectos de Investigación*, Cuarta Edi. Mexico - Mexico, 2003.
- [40] R. Hernández Sampieri, C. Fernández Collado y M. del P. Baptista Lucio, *Definición del alcance de la investigación a realizar: exploratoria, descriptiva, correlacional o explicativa*. 2010.