

UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN

Escuela de Posgrado

**MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA -
ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN**

**MODELO DE PREDICCIÓN A PARTIR DE LA MINERÍA DE DATOS
BASADO EN CASOS, VINCULADOS AL ESTUDIO DE LA
RADIACIÓN ULTRAVIOLETA UV-B AREQUIPA 2017**

TESIS

PRESENTADA POR:

WILSON FERRIOL MAMANI YANAPA

Para optar el Grado Académico de:

**MAESTRO EN CIENCIAS (*MAGISTER SCIENTIEAE*) CON MENCIÓN EN
INGENIERÍA DE SISTEMAS E INFORMÁTICA - ADMINISTRACIÓN
DE TECNOLOGÍAS DE INFORMACIÓN**

TACNA-PERÚ


2021

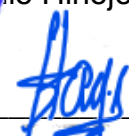
UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN-TACNA
Escuela de Posgrado

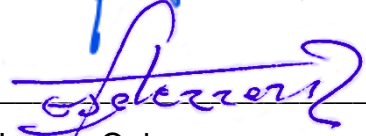
**MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA-
ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN**


**MODELO DE PREDICCIÓN A PARTIR DE LA MINERÍA DE DATOS
BASADO EN CASOS, VINCULADOS AL ESTUDIO DE LA
RADIACIÓN ULTRAVIOLETA UV-B AREQUIPA 2017**

Tesis sustentada y aprobada el 21 de abril del 2021; estando el jurado calificador integrado por:

PRESIDENTE : 
Dr. Edwin Antonio Hinojosa Ramos

SECRETARIO : 
M. Sc. Edgar Aurelio Taya Acosta

MIEMBRO : 
Dr. José Herrera Quispe

ASESOR : 
Dr. José Herrera Quispe

DEDICATORIA

El presente trabajo de investigación está dedicado a la Gran Nación Aymara, a la cual pertenezco ¡Jallalla Aymaramarka! ¡Jallalla Pachamama!

De otra parte, a todas las personas que creen en el conocimiento como fuente de libertad.

AGRADECIMIENTOS

A mi familia, mis padres y hermanos por todo el apoyo brindado.

A mi asesor de tesis.

CONTENIDO

ÍNDICE DE FIGURAS ..	vii
ÍNDICE DE TABLAS ..	xi
RESUMEN ..	xiii
ABSTRACT ..	xiv
INTRODUCCIÓN	
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	
1.1 DESCRIPCIÓN DEL PROBLEMA ..	3
1.2 FORMULACIÓN DEL PROBLEMA ..	4
1.3 JUSTIFICACIÓN DE LA INVESTIGACIÓN ..	4
1.4 OBJETIVOS ..	5
1.4.1 Objetivo General ..	5
1.4.2 Objetivos Específicos ..	5
1.5 HIPÓTESIS ..	5
CAPÍTULO II: MARCO TEÓRICO	
2.1 ANTECEDENTES DE ESTUDIO ..	6
2.2 BASES TEÓRICAS ..	12
2.2.1 Descubrimiento Conocimiento En Base De Datos (KDD) ..	12
2.2.2 Proceso KDD ..	12
2.2.3 Minería de Datos ..	12
2.2.4 Técnicas de Minería De Datos ..	12
2.2.5 Modelos de Predicción ..	14
2.3 DEFINICIÓN DE TÉRMINOS ..	20
2.3.1 El Sol ..	20
2.3.2 La Radiación Solar ..	20
2.3.3 Radiación Ultravioleta ..	20
2.3.4 Índice Ultravioleta ..	21
2.3.5 Métricas de Medición ..	21
CAPÍTULO III: MARCO METODOLÓGICO	
3.1 TIPO Y DISEÑO DE LA INVESTIGACIÓN ..	24

3.2 POBLACIÓN Y/O MUESTRA	24
3.3 OPERACIONALIZACIÓN DE VARIABLES	24
3.4 TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS	25
3.5 PROCESAMIENTO Y ANÁLISIS DE DATOS	25
CAPÍTULO IV: MARCO FILOSÓFICO	
CAPÍTULO V: RESULTADOS	
5.1 RESULTADOS: %RMSE, RMSE, MAPE, MAE, PRECISIÓN, CORRELACIÓN.....	31
5.2 RESULTADOS ANUALES 2011-2017	35
5.3 RESULTADOS MENSUALES AÑO 2017	38
5.4 RESULTADOS MÍNIMOS 2011-2017	41
5.5 RESULTADOS MÁXIMOS 2011-2017	44
5.6 AUTOREGRESIÓN DE MEDIAS MÓVILES (ARMA)	47
5.7 RAZONAMIENTO BASADO EN CASOS (RBC)	54
5.8 REDES NEURONALES CON MEMORIA DE LARGO Y CORTO PLAZO (LSTM)	60
5.9 COMPARACIÓN DE MÉTRICAS DE MODELOS ARMA, RBC, LSTM	66
CAPÍTULO VI: ANÁLISIS Y DISCUSIÓN	
CONCLUSIONES	70
RECOMENDACIONES.....	72
REFERENCIAS BIBLIOGRÁFICAS.....	73
ANEXOS	78
ANEXO 1: MATRIZ DE CONSISTENCIA	78
ANEXO 2: MODELO AUTOREGRESIVO DE MEDIAS MÓVILES (ARMA)	79
ANEXO 3: MODELO RAZONAMIENTO BASADO EN CASOS (RBC)	84
ANEXO 4: MODELO REDES NEURONALES CON MEMORIA DE LARGO Y CORTO PLAZO (LSTM).....	87

ÍNDICE DE FIGURAS

Figura 1 Proceso del Descubrimiento de Conocimiento en Base de Datos	13
Figura 2 Clasificación de Técnicas de Minería de Datos	14
Figura 3 Rasgos predictores y Rasgos Objetivos	15
Figura 4 Diagrama de células LSTM de redes neuronales	16
Figura 5 Diagrama de Flujo del modelo ARMA	17
Figura 6 Diagrama de Flujo del modelo RBC	18
Figura 7 Diagrama de Flujo del modelo LSTM	19
Figura 8 Componentes de Radiación Solar	20
Figura 9 Escala de Índices de Radiación Ultravioleta	21
Figura 10 Box-Plot para datos de los años 2011-2017	28
Figura 11 Box-Plot para datos de los meses enero-diciembre del año 2017 ..	29
Figura 12 Box-Plot para datos de los meses junio (mínimos) de los años 2011- 2017	29
Figura 13 Box-Plot para datos de los meses diciembre (máximos) de los años 2011-2017	30
Figura 14 % RMSE de los Modelos ARMA, RBC, LSTM para los años 2011- 2017	35
Figura 15 RMSE de los Modelos ARMA, RBC, LSTM para los años 2011-2017	35
Figura 16 MAPE de los Modelos ARMA, RBC, LSTM para los años 2011-2017	36
Figura 17 MAE de los Modelos ARMA, RBC, LSTM para los años 2011-2017	36
Figura 18 Precisión de los Modelos ARMA, RBC, LSTM para los años 2011- 2017	37
Figura 19 Correlación de los Modelos ARMA, RBC, LSTM para los años 2011- 2017	37
Figura 20 % RMSE de los Modelos ARMA, RBC, LSTM para los meses del año 2017	38

Figura 21 RMSE de los Modelos ARMA, RBC, LSTM para los meses del año 2017.....	38
Figura 22 MAPE de los Modelos ARMA, RBC, LSTM para los meses del año 2017.....	39
Figura 23 MAE de los Modelos ARMA, RBC, LSTM para los meses del año 2017.....	39
Figura 24 Precisión de los Modelos ARMA, RBC, LSTM para los meses del año 2017.....	40
Figura 25 Correlación de los Modelos ARMA, RBC, LSTM para los meses del año 2017.....	40
Figura 26 % RMSE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	41
Figura 27 RMSE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	41
Figura 28 MAPE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	42
Figura 29 MAE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	42
Figura 30 Precisión de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	43
Figura 31 Correlación de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.....	43
Figura 32 % RMSE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	44
Figura 33 RMSE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	44
Figura 34 MAPE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	45
Figura 35 MAE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	45

Figura 36 Precisión de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	46
Figura 37 Correlación de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	46
Figura 38 Test AIC de Parámetros P, D, Q del modelo ARMA.	48
Figura 39 RMSE del periodo 2011-2017 para el modelo ARMA.	49
Figura 40 Coeficiente de Correlación del periodo 2011-2017 para el modelo ARMA.....	49
Figura 41 RMSE del periodo enero-diciembre 2017 para el modelo ARMA	51
Figura 42 Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo ARMA.	51
Figura 43 Predicción de Índices de Radiación Ultravioleta enero 2017 para el modelo ARMA.....	53
Figura 44 Dispersión de los Índice de Radiación Ultravioleta del año 2017 para el modelo ARMA.	53
Figura 45 RMSE de Parámetros K del modelo RBC.	54
Figura 46 RMSE del periodo 2011-2017 para el modelo RBC.....	55
Figura 47 Coeficiente de Correlación del periodo 2011-2017 para el modelo RBC.	56
Figura 48 RMSE del periodo enero-diciembre 2017 para el modelo RBC. ____	56
Figura 49 Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo RBC.....	57
Figura 50 Predicción de Índices de Radiación Ultravioleta enero 2017 para el modelo RBC.....	58
Figura 51 Dispersión de los Índice de Radiación Ultravioleta del año 2017 para el modelo RBC.	58
Figura 52 RMSE del Número de Células LSTM del modelo LSTM.	60
Figura 53 RMSE del periodo 2011-2017 para el modelo LSTM.....	61
Figura 54 Coeficiente de Correlación del periodo 2011-2017 para el modelo LSTM.	62
Figura 55 RMSE del periodo enero-diciembre 2017 para el modelo LSTM	63

Figura 56 Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo LSTM.....	63
Figura 57 Predicción de Índices de Radiación Ultravioleta enero 2017 para el modelo LSTM.....	64
Figura 58 Dispersión de los Índice de Radiación Ultravioleta del año 2017 para el modelo LSTM.	64
Figura 59 Índice de Radiación Ultravioleta y predicciones del año 2017 para el modelo ARMA, RBC, LSTM.....	66

ÍNDICE DE TABLAS

Tabla 1 Métricas de medición de Modelos ARMA, RBC, LSTM para los años 2011-2017.	31
Tabla 2 Métricas de medición de Modelos ARMA, RBC, LSTM para los meses del 2017.	32
Tabla 3 Métricas de medición de Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017.	33
Tabla 4 Métricas de medición de Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017.....	34
Tabla 5 Test AIC de Parámetros P, D, Q del modelo ARMA.....	47
Tabla 6 RSME y Coeficiente de Correlación para el periodo 2011-2017.....	48
Tabla 7 RSME y Coeficiente de Correlación para el periodo enero-diciembre del año 2017.	50
Tabla 8 Índice de Radiación Ultravioleta y Predicción para el mes enero 2017.	52
Tabla 9 RMSE de Parámetros K del modelo RBC.....	54
Tabla 10 RSME y Coeficiente de Correlación del periodo 2011-2017 para el modelo RBC.....	55
Tabla 11 RSME y Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo RBC	57
Tabla 12 Índice de Radiación Ultravioleta y Predicción para el mes enero 2017 para el modelo RBC	59
Tabla 13 RMSE del Número de Células LSTM del modelo LSTM.....	60
Tabla 14 RSME y Coeficiente de Correlación del periodo 2011-2017 para el modelo LSTM.....	61
Tabla 15 RSME y Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo LSTM.....	62
Tabla 16 Índice de Radiación Ultravioleta y Predicción para el mes enero 2017 para el modelo LSTM.....	65
Tabla 17 Métricas para los modelos ARMA, RBC, LSTM del año 2017.....	66

Tabla 18 Tabla Comparativa del Índice de Radiación Ultravioleta y Predicción
para los modelos ARMA, RBC, LSTM.....67

RESUMEN

En el presente estudio se ha propuesto un Modelo de Predicción a partir de la Minería de Datos mediante el Modelo Razonamiento Basado en Casos, vinculado al estudio de la radiación UV-B en la ciudad de Arequipa, para el año 2017. El presente estudio es de tipo Cuasi-Experimental, donde la población de estudio corresponde a datos recolectados y proporcionados por el Ministerio de Salud (MINSA) sede Arequipa, y la muestra, son los datos meteorológicos de índices de radiación ultravioleta de los años 2010 al 2017. Para el procesamiento y análisis de datos, se utilizó software Octave y Python. Se dividieron los datos en dos partes: El entrenamiento se realizó con los datos de los años 2010 hasta 2016 y las pruebas de predicción se realizaron con los datos del año 2017 que representa aproximadamente un 10 % de los datos. Durante el comparativo de tres modelos de minería de datos, el modelo Autoregresivo de Medias Móviles (ARMA) presentó el coeficiente de correlación $R = 0.022$ para el año 2017 de 365 datos, mientras que el modelo Razonamiento Basado en Casos (RBC) se obtuvo un coeficiente de correlación de $R = 0.695$ y Raíz del Error Medio Cuadrático Porcentual (% RMSE) de 32 %. El modelo de redes neuronales con Memoria de Largo y Corto Plazo (LSTM) se obtuvo un % RMSE de 28 % y coeficiente de correlación de $R = 0.728$. Ha sido posible desarrollar un modelo de predicción mediante un Razonamiento Basado en Casos (RBC) con % RMSE de 32 % y 68 % de precisión para la predicción del año 2017.

Palabras Clave: Minería de Datos, Proceso KDD, Razonamiento Basado en Casos, RBC, ARMA, LSTMA, Memoria de Largo y Corto Plazo.

ABSTRACT

In the present study, a prediction model has been proposed from Data Mining and Case Based Reasoning (CBR) model, linked to the estimation of UV-B radiation in Arequipa city. The population studied corresponded to the data collected and provided by the Ministry of Health (MINSa) based in Arequipa and the sample was meteorological data on Ultraviolet Radiation Indexes for the years 2010 to 2017. Software was used for data processing and analysis Octave and Python. The data has been divided into two parts: The training was carried out with the data for the years 2010 to 2016 and the prediction tests were performed with the data for the year 2017, which represents approximately 10 % of the data. During the comparison of results of 3 data mining models, the Autoregressive Average Model (ARMA) presented the correlation coefficient $R = 0,022$ for the year 2017 of 365 data, while the Case Based Reasoning (CBR) model had a correlation coefficient of $R = 0,695$ and Root of Normalized Mean Square Error (% RMSE) of 32 %. The Long-Short Term Memory (LSTM) model obtained the RMSE of 281 which corresponds to a % RMSE of 28 % and a correlation coefficient of $R = 0,728$. It has been possible to develop a prediction model using Case Based Reasoning (RBC) with a % RMSE of 32 % and 68 % accuracy for the prediction of the year 2017.

Key Words: DATA MINING, KDD PROCESS, CASE BASED REASONING, CBR, ARMA, LSTM, LONG-SHORT TERM MEMORY.

INTRODUCCIÓN

En los años 80 se inicia la aparición del agujero capa de ozono directamente relacionado a emisiones de gases freón, provenientes de la actividad humana creciente (Chubachi et al, 1984). Los agujeros en la capa de ozono son generados debido a la baja concentración de ozono que en consecuencia permite el ingreso de gran cantidad de radiación ultravioleta de longitudes de onda corta (UV-B). Los bordes se sitúan principalmente en el sur de los continentes americano y Oceanía (Longstreth et al, 1998).

Se sabe también que, se han implementado campañas de salud en los últimos años a fin de dar a conocer los riesgos de la sobreexposición a la radiación solar y sus efectos en la piel. Se hace imperativo la difusión del correcto uso de implementos de protección como sombreros, vestimenta de mangas largas con el objetivo de disminuir daños en la piel que es de mayor importancia en diferentes grupos de riesgo (Pacientes con xerodema pigmentosum, síndrome nevus de células básicas o receptores de trasplante de órganos) quienes presentan mayores riesgos de adquirir cáncer de piel.

De otra parte, la radiación ultravioleta (UV-B) en altas dosis provoca daños en la salud humana (induce Eridema o quemadura de sol, daños en el ADN, cáncer de piel y cataratas. Sin embargo, en pequeñas cantidades son beneficiosas para la salud a través de la producción de vitamina D (Rosso Fernández, 2014).

Así también, produce cambios en plantas y ecosistemas terrestres (Heisler et al, 2003); en muchas especies de plantas se incrementa la susceptibilidad a enfermedades debido a fotosíntesis modificadas (Caldwell et al, 1998); existen evidencias de daño en cultivo y desarrollo de cosechas, se disminuye la productividad de fitoplancton, existen modificaciones en comunidades microbiológicas e invertebradas (Kakani et al, 2003). En los humanos también se

han demostrado la relación entre el desarrollo de melanomas malignos y las exposiciones de Radiación Ultravioleta para la población infantil (Østerlind et al, 1988).

Existen diversas organizaciones mundiales como la Organización Meteorológica Mundial (OMM), la Organización Mundial de la Salud (OMS) y el Programa de Naciones Unidas para el Medio Ambiente (PNUMA) que han propuesto una métrica simple, índice ultravioleta (I-UV) para informar a la población del riesgo durante las horas del día. El índice ultravioleta (I-UV) se describe en: Bajo (1 a 2), Moderado (3 a 4), Alto (5 a 7), Muy Alto (8 a 10) y Extremadamente Alto (11 a más) (OMS, 2003).

En el Perú, durante el monitoreo de los efectos de la radiación solar, el Ministerio de Salud del Perú informa que cada año se diagnostican más de cuatro mil casos de cáncer de piel en relación a la sobre-exposición solar en presencia de elevados índices de radiación ultravioleta; Así también, el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), realiza monitoreo del índice de radiación ultravioleta en diferentes ciudades peruanas.

Por ello, es de vital importancia la implementación de modelos computacionales capaces de predecir los índices de radiación ultravioleta, a fin de estudiar el comportamiento y realizar acciones de prevención en la población nacional y local.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 DESCRIPCIÓN DEL PROBLEMA

En presencia del agujero de la capa de ozono, se incrementan diversas enfermedades y el cáncer de piel en todo el mundo, que principalmente tiene su origen en la exposición a los elevados índices de radiación ultravioleta (Sordo y Gutiérrez, 2013).

De otra parte, en Perú el cáncer de piel tiene la frecuencia de ser el tercer más frecuente en hombres y cuarto en mujeres (Gutiérrez et al, 2009) que fueron estudiados en el primer semestre del año 2012 (Ramos, 2012).

Por tanto, el gobierno peruano, a través de iniciativa del congreso, ha promulgado la ley 30102 del 2013, bajo el título "*Ley que dispone medidas preventivas contra los efectos nocivos para la salud por la exposición prolongada a la radiación solar*" (de amat-loza, 2014), donde, en su artículo 1, enfatiza en la prevención de daños ocasionados por la exposición prolongada a los rayos ultravioleta en trabajadores, estudiantes y niños, en aplicación del decreto ley, en diversas instituciones del Gobierno Peruano, como Municipalidades, Oficinas Públicas, Gobiernos Regionales, los trabajadores deben de portar elementos de protección a fin de evitar una sobre-exposición a radiación ultravioleta UV-B durante el día.

Aún no se presentan estudios sobre predicción de radiación solar, que tienen consecuencias directas para la salud de la población y efectos al medio ambiente; de igual modo, no se ha escrito sobre simulaciones basadas en los datos registrados anteriormente en instituciones como el Ministerio de Salud

(MINSA) y el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI).

1.2 FORMULACIÓN DEL PROBLEMA

¿Será posible desarrollar un Modelo de Predicción a partir de la Minería de Datos Basado en Casos, vinculados al estudio de la radiación ultravioleta UV-B en la ciudad de Arequipa?

1.3 JUSTIFICACIÓN DE LA INVESTIGACIÓN

Existe una constante preocupación desde el Gobierno Central, a través del Ministerio de Salud y el Congreso de la República, en la implementación de diversas líneas de acción para combatir o mitigar los efectos que pueda tener en la población vulnerable.

Nuestra propuesta de estudio busca cumplir esta inquietud de previsión y ser una fuente para futuros estudios de la radiación ultravioleta (UV-B).

La minería de datos ha tenido buenos resultados a nivel nacional en otros campos de estudio como las finanzas (*fraude electrónico en entidades bancarias*) (Naupas Caraza, 2016), *deserción académica* (Sulla Torres, 2015), medicina (*hipertensión arterial*) (Díaz Avendaño, 2016).

Es también un tema de estudio relativamente novedoso y actual en el ámbito peruano según las nuevas disposiciones gubernamentales.

1.4 OBJETIVOS

1.4.1 Objetivo General

Proponer un modelo de alta precisión en la predicción de los índices de radiación ultravioleta en la ciudad de Arequipa para el año 2017.

1.4.2 Objetivos Específicos

Tenemos los siguientes objetivos específicos durante el estudio:

- Determinar el error en la Predicción del modelo Autoregresivo de Medias Móviles (ARMA).
- Determinar el error en la Predicción del modelo de Razonamiento Basado en Casos (RBC).
- Determinar el error en la Predicción del modelo de Redes Neuronales con Memoria de Largo y Corto Plazo (LSTM).
- Realizar el análisis comparativo de los modelos ARMA, RBC y LSTM.

1.5 Hipótesis

El modelo Razonamiento Basado en Casos (RBC), es el modelo con menor error en la predicción de los índices de radiación ultravioleta en la ciudad de Arequipa para el año 2017.

CAPÍTULO II

MARCO TEÓRICO

2.1 ANTECEDENTES DE ESTUDIO

Se ha realizado búsquedas de bibliografía reciente relacionada al estudio de la minería de datos, orientada al pronóstico de radiación UV, efectos para la salud de la población peruana. Así se ha consultado en bibliotecas, hemerotecas, bases de datos científicas: Engineering Village, Web of sciences, ACM Digital Library, IET (IEEE Electronic Library). De igual modo, en buscadores de publicaciones científicas como Scopus y Scholar Google.

ANÁLISIS DE SERIES DE TIEMPO UNIVARIANTE APLICANDO METODOLOGÍA DE BOX-JENKINS PARA LA PREDICCIÓN DE OZONO EN LA CIUDAD DE CALI, COLOMBIA.

Mauricio Jarmillo Ayerbe, Daniel Enrique González Gómez, María Eugenia Núñez Cabrera, Gloria Esperanza Portilla, Jesús Heriberto Lucio García (Lahura et al, 2003).

Resumen

El autor aplica el modelo ARMA en la predicción a corto plazo de la concentración del ozono troposférico de la ciudad de Cali, Colombia a través de un análisis de una sola variable de series de tiempo. Durante los meses de abril a julio del 2003, 2496 datos concentración de ozono fueron obtenidos de estaciones de la Red de Monitoreo de Calidad del Aire (RMCA) de Cali. Los datos son divididos en dos partes de 93 días y 11 días, donde los primeros se utilizaron en la estimación del modelo y los restantes para su validación. (Lahura et al, 2003).

Conclusión

El autor indica que el modelo que mayor proximidad, en el proceso generador de la serie, corresponde a un ARMA (24,4), del cual se desprende el coeficiente de información ajustado R^2 de 0.72, que se interpreta como un buen nivel de correspondencia entre el modelo y los datos estudiados.

El modelo propuesto por el autor satisface la predicción hasta por un periodo de 8 horas en adelante. De otra parte, se establece la posible utilidad de la metodología Box-Jenkins como complemento al análisis de datos de monitoreo de la calidad del aire.

Las medidas preventivas deberían incluir restricciones vehiculares y advertencias de exposición en grupos vulnerables como ancianos y niños, así evitar zonas con altas concentraciones de ozono en exteriores. El análisis de una sola variable, presentado en el estudio, constituye una herramienta básica si se compara con otras técnicas de variables múltiples de predicción, pero con estimaciones similares.

APROXIMACIÓN A LA ESTIMACIÓN PARA EL POTENCIAL DE RADIACIÓN SOLAR CON COMPLEJOS PATRONES ESPACIALES VÍA TÉCNICAS DE MACHINE LEARNING.

A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine learning techniques Choongwan Koo, Wenzhuo Li, Seung Hyun Cha, Shaojie Zhang (Koo et al, 2019).

Resumen

La radiación solar global sobre la superficie horizontal en un específico lugar, debería ser investigado en avance, así el sistema de energía solar podría ser implementado correctamente y eficientemente. Así entonces, el *promedio mensual de la radiación solar diaria* (MADSR) en China, tiene complejos patrones espaciales y sus estaciones de observación que aún carecen debido al alto costo de equipos. Para dirigir estos retos, este estudio tiene por objetivo desarrollar una aproximación a la estimación para el MADSR, con su complejo patrón espacial sobre una vasta área en China, a través de técnicas de Machine Learning (Por ejemplo, método de clustering (k-means) y un modelo Avanzado de Razonamiento Basado en Casos (Ha-CBR). El MADSR y la relevante información, fueron colectados en 97 ciudades de China para 10 años (Desde 2006 hasta 2015). La **precisión** del promedio de predicción de la aproximación propuesta fue determinada al 93.23 %.

Conclusión

Este estudio intenta desarrollar una nueva aproximación sobre la estimación para el *promedio mensual de la radiación solar diaria* MADSR, con su complejo patrón espacial sobre una extensa área de China, en aplicación de técnicas de machine learning (Por ejemplo: Cluster k-means, RBC, MRA, ANN y GA). Un total de 11640 datos (97 ciudades x 10 años x 12 meses) fueron utilizados en desarrollar una nueva aproximación durante la estimación.

Primeramente, el método Cluster K-Means fue usado para clasificar las cinco zonas de radiación solar en China, analizando el promedio de los 10 años de MADSR (Desde 2006 hasta 2015) en 97 ciudades.

En segundo lugar, el modelo A-CBR fue desarrollado para estimar el MADSR para cinco zonas de radiación solar. La precisión del promedio de predicción fue determinada al 93.23 % mostrando una prometedora forma de estimar la radiación solar global en la superficie horizontal en algún sitio específico.

PREDICCIÓN DE SERIES DE TIEMPO DE RADIACIÓN SOLAR DIARIA PREPROCESADA UTILIZANDO REDES NEURONALES.

Forecasting of preprocessed daily solar radiation time series using neural networks, Christophe Paoli, Cyril Voyant, Marc Muselli, Marie-Laure Nivet (Paoli et al, 2010).

Resumen

Se presenta una aplicación de Redes Neuronales Artificiales (ANNs) de las energías renovables. Se ve particularmente la red perceptron multicapa (MLP) la cual ha sido la más utilizada en las arquitecturas ANNs, ambas en las energías renovables y la predicción de series temporales. Se ha utilizado un MLP y un pre-procesamiento ad-hoc para desarrollar una metodología para predicciones diarias de radiación solar global en una superficie horizontal. Los primeros resultados son prometedores con raíz del error medio cuadrático porcentual (%RMSE) aproximado al 21 % y RMSE aproximado a 359MJ/m². El MLP optimizado presenta predicciones similares o mejores que el modelo convencional y referencia métodos tales a ARIMA, Inferencia Bayesiana, Cadenas de Markov y K-Vecinos Cercanos. El simulador propuesto ha sido obtenido utilizando 19 años de datos válidos provenientes de la estación de Ajaccio (Isla de Corsica, Francia, 41°55'N, 8°44'E, 4m sobre el nivel del mar). La metodología de predicción ha sido validada en malla de paneles fotovoltaicos Mono-Si con potencia de 1.175KWc. Se utilizaron seis métodos de predicción (ANN, clear sky model, combinaciones); que permiten predecir la mejor potencia de producción DC PV en un horizonte h + 1 La energía acumulada DC PV en un periodo de 6 meses muestra una alta correlación entre lo simulado y lo observado ($R^2 > 0.99$ y $nRMSE < 2\%$)

Conclusión

El desarrollo de una red neuronal artificial de predicción para determinar la irradiación global en horizonte diaria (d+1), el cual puede ayudar en la

administración eléctrica con una red- PV sistema de potencia conectado. Este modelo de predicción ha sido comparado con otros métodos predictores (AR, ARMA, K-NN, MARKOV CHAIN, ETC). Se ha utilizado una serie de tiempo pre-procesada ad-hoc y una serie de tiempo diseñada MLP; Sin embargo, la ubicación fue muy específica con la proximidad al mar y la montaña que puede tener gran impacto en la nebulosidad, se obtuvieron resultados relevantes.

Sin Preprocesamiento, el modelo autoregresivo AR (8) y Redes Neuronales, presentaron los mejores RMSE con 36MJ/m² (998 Wh-m²) y nRMSE de 21 %, aproximadamente, comparado a la cadena de Markov, Bayes, y K-NN, donde el nRMSE fue de 25-26 %, aproximadamente. Así entonces, el pre-procesamiento anual del método ANN basado en índices de claridad e índices de cielo claro, reduce los errores de pronóstico del 5-6 % aproximadamente, (nRMSE 20 %) comparado a clásicos predictores como cadenas de Markov.

Estas herramientas de simulación han sido satisfactoriamente validadas en la predicción de energía de corriente directa de 1.175 KWp en sistemas de potencia de paneles fotovoltaicos (PV) de silicio monocristalino conectado a la red. En un punto vista estacional, ANN con pre-procesamiento de cielo claro en meses de invierno (enero, febrero) representa una adecuada solución (nRMSE 37 %, aproximadamente). Para meses de verano ANN sin Pre-procesamiento genera los mejores resultados (nRMSE 15 %, aproximadamente). La producción PV acumulada, simulada y medida con muy buenos resultados ($R^2 > 0.99$) validando todo el proceso de predicción (nRMSE < 2 % para el periodo de 6 meses).

2.2 BASES TEÓRICAS

2.2.1 Descubrimiento de Conocimiento en Base de Datos (KDD)

El descubrimiento de Conocimiento en Base de Datos (KDD) se desarrolla desde un enfoque de investigación que involucra diversas áreas como bases de datos, aprendizaje automático, reconocimiento de patrones, estadística, teoría de la información, inteligencia artificial, visualización de datos (Riquelme Santos et al, 2006).

2.2.2 Proceso KDD

En el conjunto del proceso KDD, la Minería de Datos es la parte central del conjunto de métodos y procesos progresivos (López, 2007), el cual se visualiza en la figura 1.

2.2.3 Minería de Datos

Son el conjunto de técnicas en la búsqueda de nuevas e importantes relaciones, patrones y tendencia para un grupo de data organizada. Actualmente, los avances en el desarrollo poder computacional la minería de datos tiende a realizarse de forma automática (López, 2007).

2.2.4 Técnicas de Minería de Datos

Existen tres métodos o técnicas principales de minería de datos: técnicas predictivas, técnicas descriptivas y técnicas auxiliares (López, 2007) que se grafican en la figura 2.

Figura 1

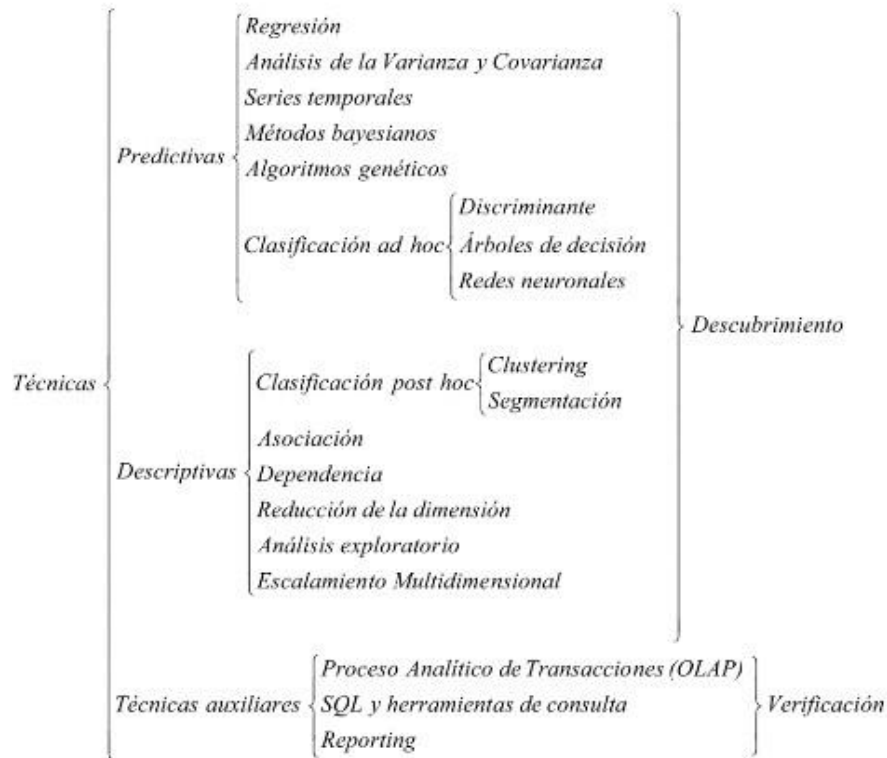
Proceso del Descubrimiento de Conocimiento en Base de Datos



Fuente: (López, 2007)

Figura 2

Clasificación de Técnicas de Minería de Datos



Fuente: (López, 2007)

2.2.5 Modelos de Predicción

Autoregresión de Medias Móviles (ARMA)

La palabra ARMA significa Modelos Autorregresivos de Medias Móviles, se define como un modelo Autorregresivo, donde toda Y_t se puede expresar como una ecuación lineal de sus propios valores pasados (parte sistemática) y un término de error (innovación). El orden del modelo expresa el número de observaciones retrasadas de la serie temporal analizada que intervienen en la ecuación. (De Arce and Mahia, 2003).

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t$$

Razonamiento Basado en Casos (RBC)

El razonamiento basado en casos, es un modelo utilizado para resolver problemas actuales adaptando soluciones de problemas anteriores (Jiménez Builes, 2006).

El Razonamiento Basado en Casos para la toma de Decisiones, es un Sistema de Decisión $S=(U,X^uY)$ y se define en términos de un par (U,X^uY) donde U es un conjunto finito no vacío de objetos, eventos llamados casos; de otra parte, se tienen los rasgos predictores (X) y rasgos objetivos (Y) que se consideran atributos o propiedades. (Rodríguez, S. C. et al., 2011).

Donde cada rasgo predictor $x_i \in X$ puede ser considerado una función, que mapea elementos de U en el conjunto que mapea elementos de U en el conjunto $M_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$, el cual se denomina conjunto de valores del rasgo predictor $X_i: U \rightarrow M_i$.

De manera análoga, cada rasgo objetivo $y_j \in Y$ puede ser considerado una función que mapea elementos de U en el conjunto $N_j = \{y_{j1}, y_{j2}, \dots, y_{jn_j}\}$ que se denomina conjunto de valores del rasgo predictor $Y_j: U \rightarrow N_j$.

Figura 3

Rasgos predictores y rasgos objetivos

	<i>Rasgos predictores</i>			<i>Rasgo objetivo</i>
Casos	x_1	...	x_n	y_1
O_1	$x_1(O_1)$...	$x_n(O_1)$	$y_1(O_1)$
O_2	$x_1(O_2)$...	$x_n(O_2)$	$y_1(O_2)$
...
O_m	$x_1(O_m)$...	$x_n(O_m)$	$y_1(O_m)$

Fuente:(Rodríguez, S. C. et al., 2011)

Memoria de Largo y Corto Plazo (LSTM)

Las redes neuronales LSTM son un tipo de redes neuronales recurrentes. Son una evolución de los perceptrones de las redes neuronales elementales en las que hay ciclos en las reconexiones entre las neuronas de cada capa. Son, además, un bloque de construcción que se puede usar de forma recursiva y que automáticamente permite la asignación y lectura de valores por parte de la red. (Guridi Mateos, 2017).

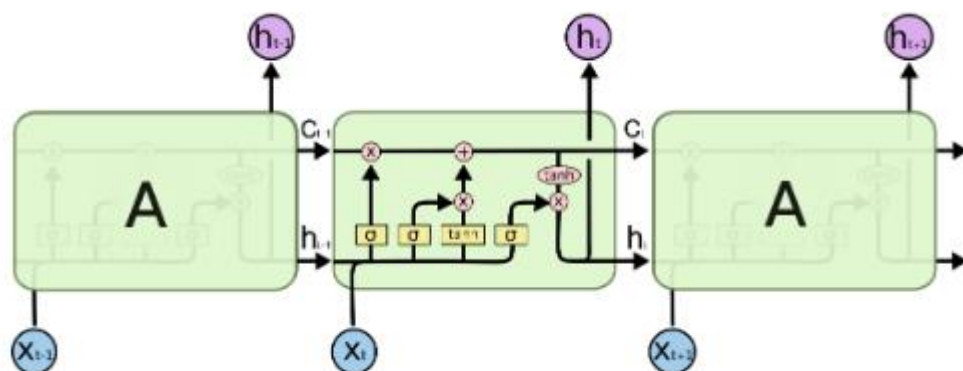
En esta arquitectura vemos que la función de activación de estas neuronas es la función identidad ($f(x) = x$), junto con el peso de la puerta de recordar. Esta es la aproximación a las funciones de actualización de una neurona LSTM:

$$C_t = C_{t-1} \cdot \text{puerta1}(h_{t-1}, x_t) + \text{puerta2}(h_{t-1}, x_t) \cdot W(h_{t-1}, x_t)$$

$$h_t = C_t \cdot \text{puerta3}(h_{t-1}, x_t)$$

Figura 4

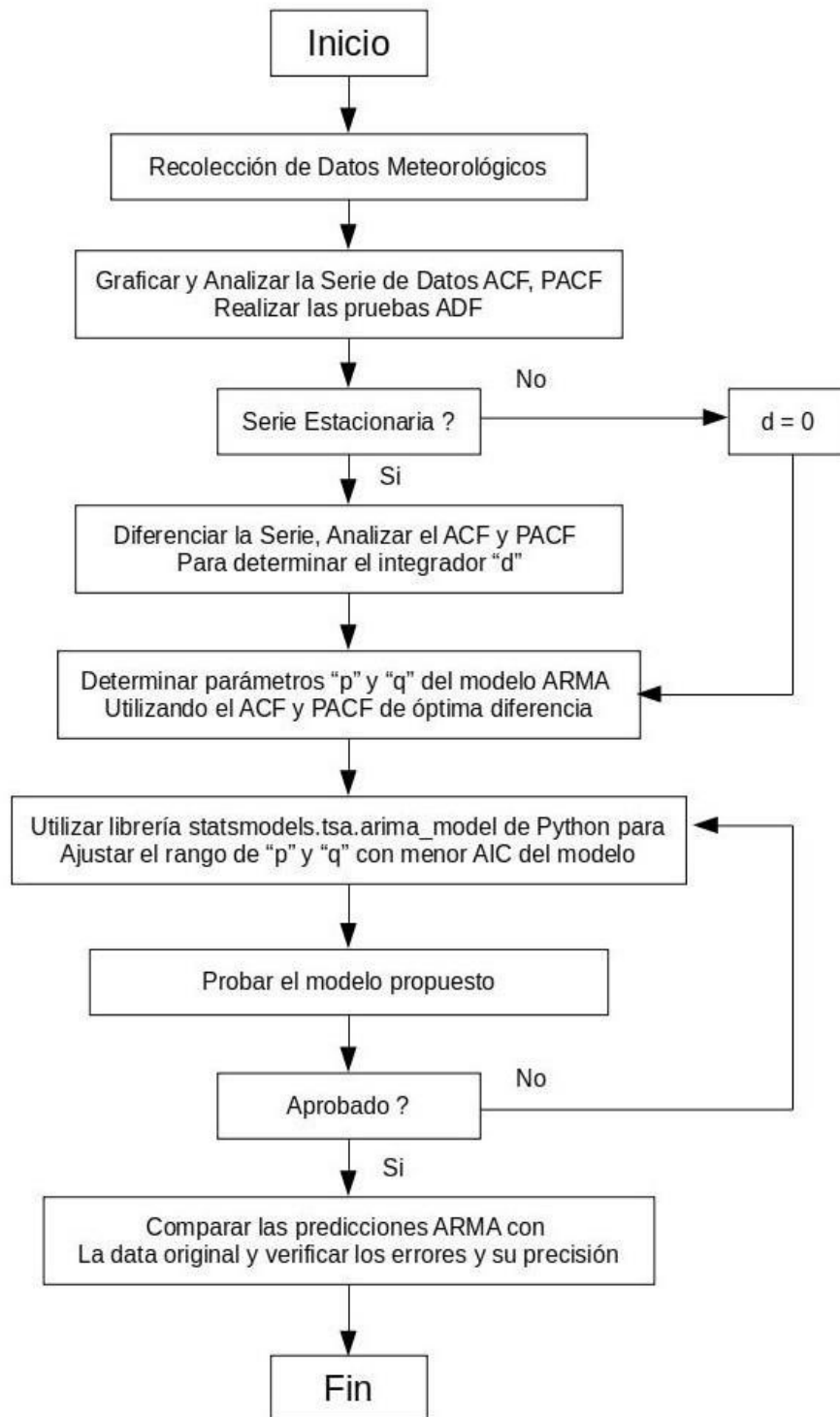
Diagrama de células LSTM de redes neuronales



Fuente: (Guridi Mateos, 2017).

Figura 5

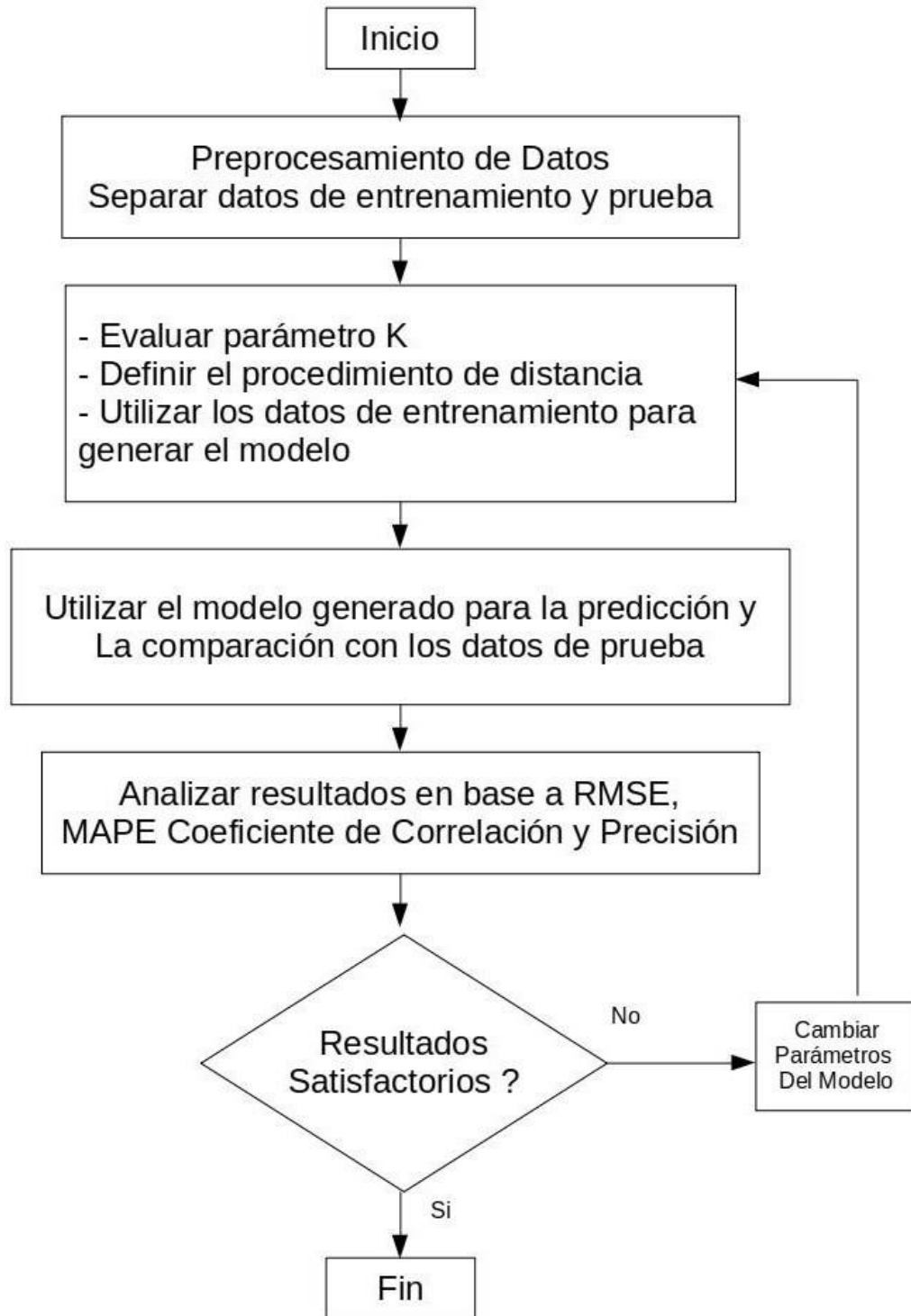
Diagrama de Flujo del modelo ARMA



Fuente: (Machekposhti et al, 2018)

Figura 6

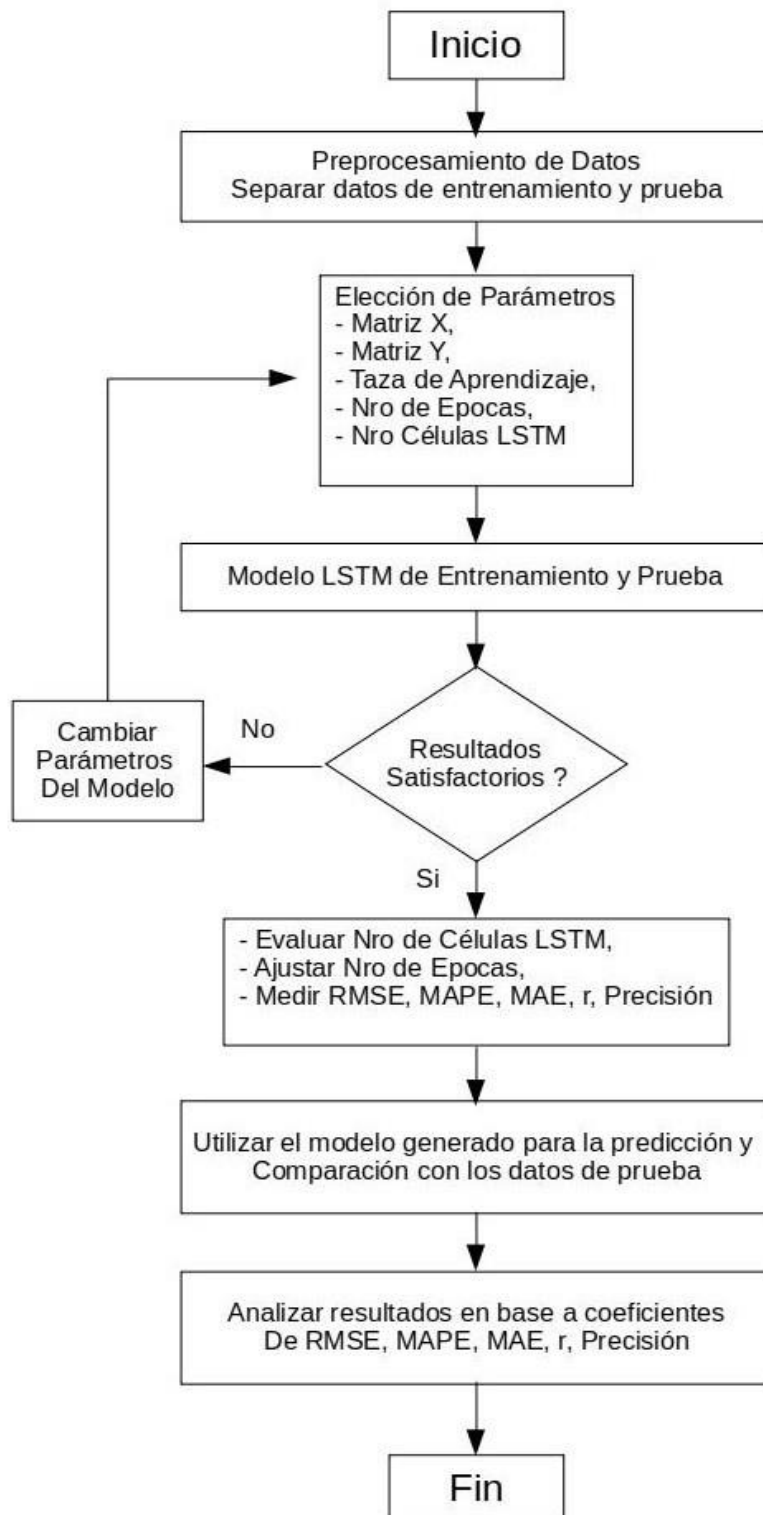
Diagrama de Flujo del modelo RBC



Fuente: (Garg et al, 2020)

Figura 7

Diagrama de Flujo del modelo LSTM



Fuente: (Zhang, 2017)

2.3 DEFINICIÓN DE TÉRMINOS

2.3.1 El Sol

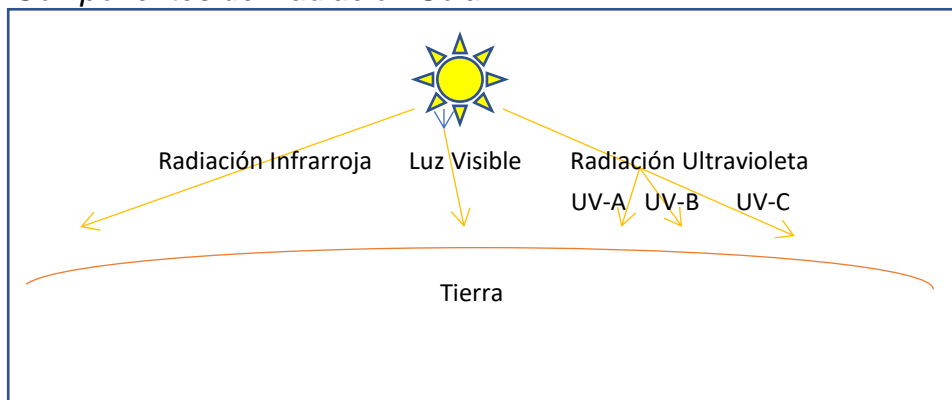
El sol es la estrella más cercana al planeta Tierra con un diámetro de $1,39 \times 10^9 m$ y de $1,5 \times 10^{11} m$ de distancia. La gran cantidad de energía generada durante su vida útil hace su recorrido y alcanza nuestro planeta en forma de radiación (Duffie, 2013).

2.3.2 La Radiación Solar

La gran cantidad de energía generada por el sol, tiene su composición de diversas longitudes de onda, sin embargo, el mayor porcentaje (99 %) corresponde a la luz ultravioleta, visible e infrarroja (Cabrera-Silva, 2005).

Figura 8

Componentes de Radiación Solar



Fuente: Propio

2.3.3 Radiación Ultravioleta

La radiación ultravioleta es la radiación electromagnética con longitud de onda inferiores a la luz visible, es decir, menores a 400nm (Cabrera-Silva, 2005)

Se puede subdividir en los siguientes tipos según su longitud de onda:

- UV-A (315 a 400nm).
- UV-B (280 a 315nm).

- UV-C (200 a 280nm).
- Vacuum, que se propaga solo en condiciones de vacío.

2.3.4 Índice Ultravioleta

El índice ultravioleta (IUV) representa la intensidad de la radiación ultravioleta en la superficie de la Tierra. El índice se expresa en números enteros superiores a cero, a valores más altos, mayor probabilidad de lesiones en la piel y ojos (PNUD, 2003).

Figura 9

Escala de Índice de Radiación Ultravioleta



Fuente: (PNUD, 2003)

2.3.5 Métricas de Medición

Raíz del Error Medio Cuadrático Porcentual (%RMSE)

La raíz del error medio cuadrático porcentual se expresa mediante la ecuación siguiente (Bitri et al, 2014), donde: y_t : Valor Actual, \hat{y}_t : Valor de Predicción, T : Número de Datos.

$$\% RMSE = \frac{100 \cdot T}{\sum_{t=1}^T y_t} \times \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}$$

Raíz del Error Medio Cuadrático (RMSE)

Es la raíz cuadrada del promedio de errores al cuadrado. Esta métrica de medición es sensible a valores atípicos (Pontius et al, 2008) dónde: y_t : Valor Actual, \hat{y}_t : Valor de Predicción, T : Numero de Datos.

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}$$

Error Medio Absoluto Porcentual (MAPE)

El valor promedio absoluto del error normalizado en porcentaje (Koo et al, 2019) donde: A_t : Valor Actual, F_t : Valor de Predicción, n : Numero de Datos.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Error Medio Absoluto (MAE)

Corresponde al valor promedio absoluto del error $|e| = x_i - y_i$ (Willmott & Matsuura, 2005) donde: x_i : Valor Actual, y_i : Valor de Predicción, n : Número de Datos.

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n}$$

Precisión

El valor absoluto del error normalizado en porcentaje para obtener la precisión (Koo et al, 2019) donde: A_t : Valor Actual, F_t : Valor de Predicción, n : Número de Datos.

$$PA = 100 - \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Coefficiente de Correlación (r)

Corresponde la relación lineal existente entre dos conjuntos de variables, sus valores varían entre -1 y 1, cuando no existe correlación su valor cae en 0 (Lahura et al, 2003) donde: x_i, y_i : Valores Individuales de Datos, n :

Número de Datos.

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

CAPÍTULO III

MARCO METODOLÓGICO

3.1 TIPO Y DISEÑO DE LA INVESTIGACIÓN

El presente estudio es de tipo Cuasi-Experimental.

3.2 POBLACIÓN Y/O MUESTRA

La población de estudio corresponde a los datos proporcionados por el Ministerio de Salud (MINSa) y la muestra son los datos meteorológicos de índices de radiación ultravioleta (I-UV) de los años 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017.

3.3 OPERACIONALIZACIÓN DE VARIABLES

Variable Independiente: Tiempo

Variable Dependiente: Índice de Radiación UV

3.3.1 Indicadores:

Tiempo: año (2017)

Índice de Radiación UV

- Desempeño en la predicción del modelo RBC
 - %RMSE, RMSE, MAPE, MAE, Coeficiente de correlaciones
- Desempeño en la predicción del modelo ARMA
 - %RMSE, RMSE, MAPE, MAE, Coeficiente de correlaciones
- Desempeño en la predicción del modelo LSTM
 - %RMSE, RMSE, MAPE, MAE, Coeficiente de correlaciones

3.4 TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS

Los datos son recolectados mediante instrumentación perteneciente al MINSA sede Arequipa.

3.5 PROCESAMIENTO Y ANÁLISIS DE DATOS

Se utilizó una computadora con conexión a internet y plataforma Google-Colab con software Python versión 3.6 que se emplea durante la implementación del algoritmo de predicción y el procesamiento de datos en los modelos ARMA, RBC, LSTM. El software Octave versión 4.22 se emplea durante el análisis de los resultados obtenidos.

CAPÍTULO IV

MARCO FILOSÓFICO

El enfoque principal del presente estudio es la salud y la ciencia, de donde se desprende que la salud es hoy día cosa pública, objetivada como “bienestar” (Mainetti, 1990) lo cual se proporciona a la población por sus respectivos estados a fin de obtener el bienestar pleno. Es difícil comentar del crecimiento personal del individuo si no se atiende la salud también. De este modo, es que la constitución de la Organización Mundial de la Salud reza: “La salud de todas las personas es fundamental para alcanzar la paz y seguridad, además, depende de la cooperación individual y de los estados” (Mainetti, 1990).

La salud moderna se basa en la medicina que se presenta como la nueva caja de pandora de la sociedad industrial, revestida de todos los dones y, a la vez, fuente de muchos males; ella alimenta la esperanza de la humanidad en la ambrosía, el pan de la salud y la mortalidad, que se transforma en el pan de la enfermedad y la locura (Mainetti, 1990). Así entonces, se tiene que en la medicina y la salud siempre hay en el acto asistencial, por más restricta y pobremente trabajada que esté, una inexorable dimensión hermenéutica, la necesidad de saber cómo determinados saberes generales pueden ser aplicados a un paciente concreto (Ayres, 2008). Es posible indicar que la salud del individuo precisa atención desde un punto de vista más científico.

Es en este sentido, que la ciencia aplica sus métodos en el estudio de las causas, tratamiento o consecuencias de los estados de salud del individuo; cuando se propone un modelo predictivo que estudie el comportamiento de algún componente de la radiación ultravioleta (UV-B) en la salud de la población, se da un pequeño aporte en mejorar la calidad de vida de la población.

Para este fin, el estudio de la ciencia y su metodología, han tenido transformaciones durante la historia por parte del ser humano, siempre con las mejores intenciones de preservar la vida. En ese proceso, las ciencias fueron abandonando el positivismo ingenuo que sustentaba la metafísica causalista y empirista, en dirección a una concepción más perspectivista y constructorista de las evidencias que producen (Ayres, 2008).

Para Rodríguez, la concepción filosófica de origen platónico, según la cual el ideal de la ciencia consiste en forjar una teoría que reduce la totalidad de la diversidad sensible a la unidad de un principio inteligible que posibilite deducir la totalidad de lo real, tanto del pasado como del futuro, a partir de la comprensión del presente (Rodríguez, 2008).

La presente investigación se sostiene en la corriente filosófica del positivismo, del cual se desprenden hechos objetivos y verificables con resultados reales y aplicables en algún sentido estrictamente científico, a fin de contribuir al progreso material y moral de los seres humanos. Para desarrollar un conocimiento sistemático es necesario el uso de un lenguaje lógico, riguroso, preciso y formal (Díaz Tello, 2014).

Para finalizar, se analiza que la salud necesita permanentemente de la ciencia, como una tarea constante a lo largo de los años, donde se resolverán antiguos paradigmas científicos y se crearán nuevos, según Poincaré: La teoría y la ciencia más actual, es siempre una tarea inacabada. (Rodríguez, 2008).

CAPÍTULO V

RESULTADOS

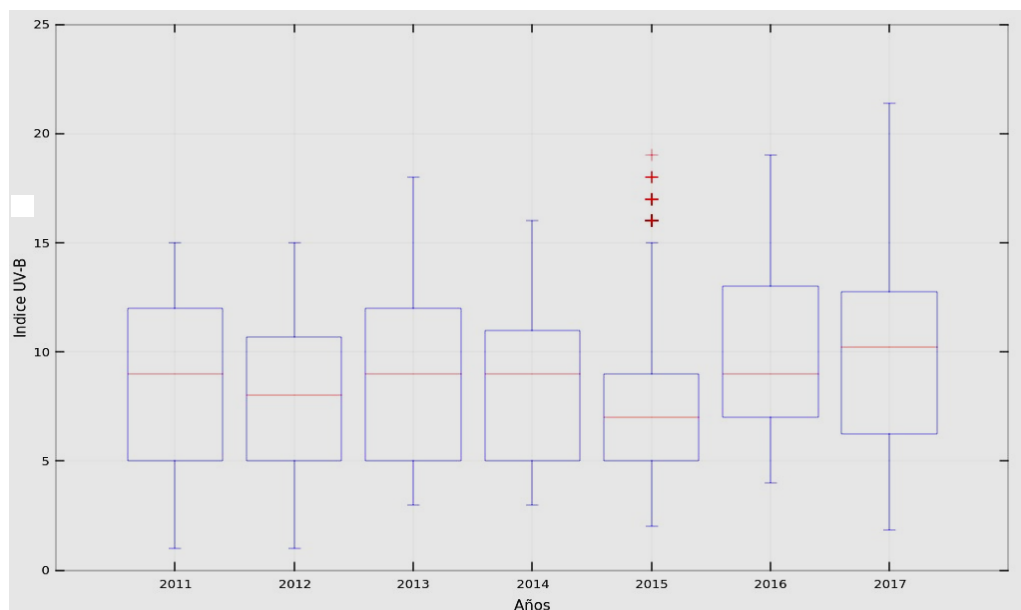
Durante el presente trabajo de investigación sobre Minería de Datos, se ha utilizado tres modelos, en especial el modelo Razonamiento Basado en Casos, a fin de comparar sus resultados y permitir confirmar la hipótesis planteada.

Los modelos a emplear son: Autoregresión de Medias Móviles (ARMA), Razonamiento Basado en Casos (RBC) y Redes Neuronales con Memoria de Largo y Corto (LSTM).

En el estudio de los modelos predictivos ARMA, RBC, LSTM se ha dividido los datos en Entrenamiento del periodo 2010-2016 y datos de Prueba el año 2017, que se grafican en las figuras 7, 8, 9 y 10.

Figura 10

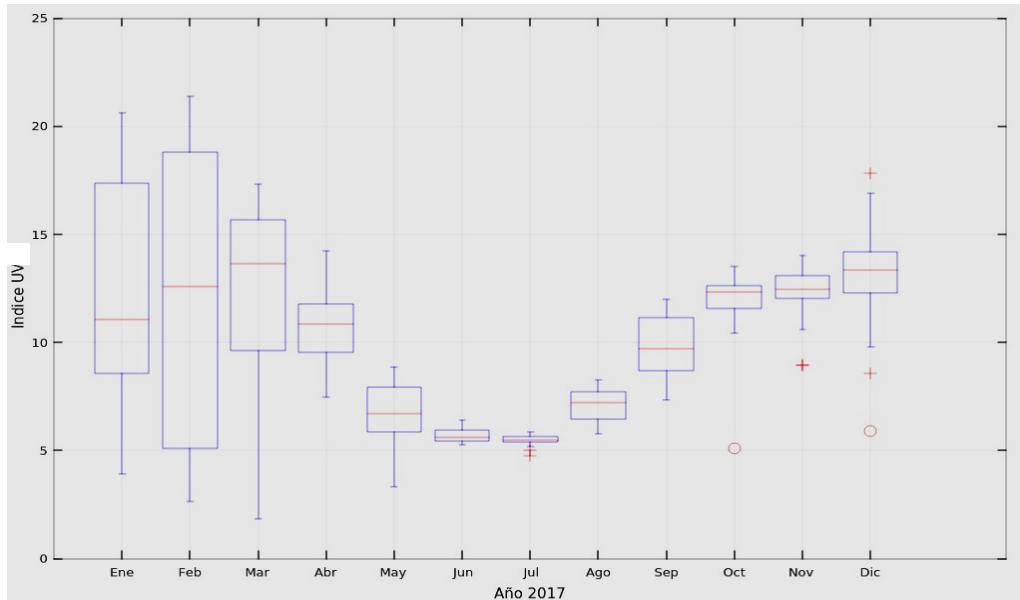
Box-Plot para datos de los años 2011-2017



Fuente: Propia

Figura 11

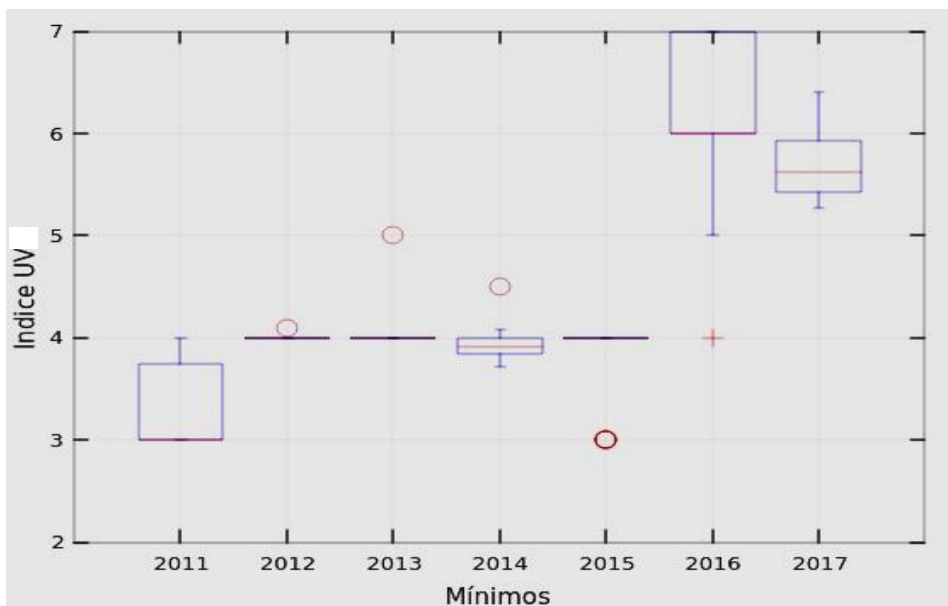
Box-Plot para datos de los meses enero-diciembre del año 2017



Fuente: Propia

Figura 12

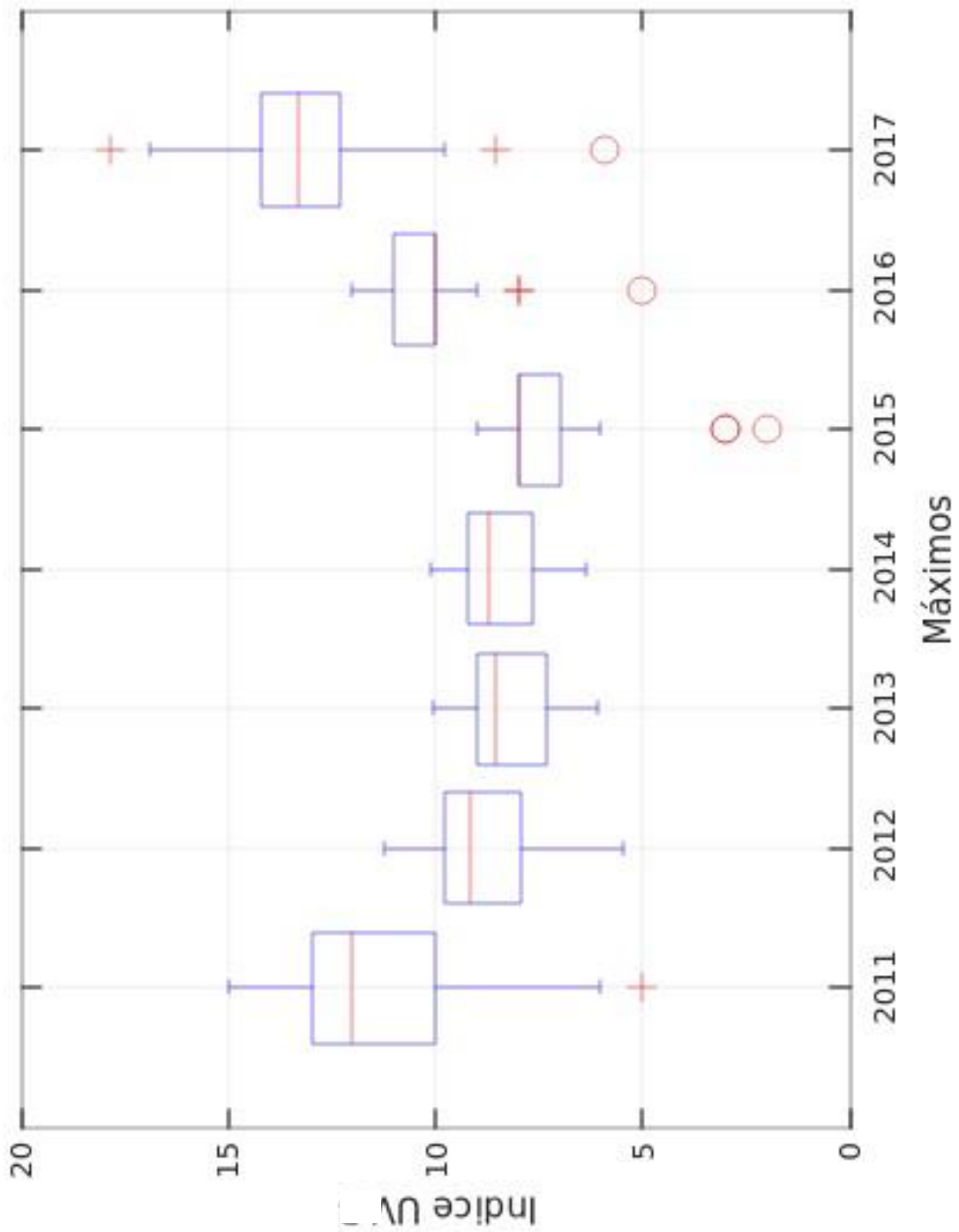
Box-Plot para datos de los meses junio (Mínimos) de los años 2011-2017



Fuente: Propia

Figura 13

Box-Plot para datos de los meses diciembre (Máximos) de los años 2011-2017



Fuente: Propia

5.1 RESULTADOS: % RMSE, RMSE, MAPE, MAE, PRECISIÓN, CORRELACIÓN

En la presente sección se tienen tablas con los resultados de indicadores obtenidos aplicando los algoritmos ARMA, RBC, LSTM. Tablas 1, 2, 3 y 4.

Tabla 1

Métricas de medición de Modelos ARMA, RBC, LSTM para los años 2011-2017

		2011	2012	2013	2014	2015	2016	2017	Prom.
%RMSE	ARMA	49.41	40.32	50.38	45.91	51.68	53.96	40.38	47.43
	RBC	19.59	23.13	22.94	21.28	27.28	15.11	31.71	23.01
	LSTM	19.18	20.83	21.95	19.41	27.56	16.14	28.01	21.87
RMSE	ARMA	4.32	3.3	4.49	3.94	3.89	5.46	4.05	4.21
	RBC	1.71	1.85	2.02	1.76	2.05	1.53	3.02	1.99
	LSTM	1.68	1.71	1.96	1.67	2.08	1.64	2.81	1.93
MAPE	ARMA	41.44	34.92	41.67	39.57	41.78	44.58	34.26	39.75
	RBC	10.63	13.88	13.07	11.21	13.83	1.02	18.90	13.10
	LSTM	11.76	12.21	13.36	1.14	15.57	1.14	16.06	13.11
MAE	ARMA	3.63	2.86	3.71	3.40	3.15	4.51	3.44	3.53
	RBC	0.93	1.11	1.15	0.93	1.04	1.03	1.80	1.14
	LSTM	1.03	1	1.19	0.98	1.17	1.16	1.61	1.16
Precisión	ARMA	58.56	65.08	58.33	60.43	58.22	55.42	65.74	60.25
	RBC	80.41	76.87	77.06	78.72	72.72	84.89	68.29	76.99
	LSTM	88.24	87.79	86.64	88.60	84.43	88.60	83.94	86.89
Correlación	ARMA	0.24	0.04	0.08	0.03	0.59	0.38	0.02	0.17
	RBC	0.89	0.84	0.85	0.87	0.82	0.90	0.69	0.84
	LSTM	0.89	0.85	0.85	0.87	0.81	0.90	0.73	0.84

Fuente: Propia

Tabla 2

Métricas de medición de Modelos ARMA, RBC, LSTM para los meses del año 2017

	MES	1	2	3	4	5	6	7	8	9	10	11	12	Prom
%RMSE	ARMA	47.01	54.51	37.92	17.35	51.28	72.73	79.68	39.3	14.2	22.01	23.69	32.19	40.99
	RBC	46.37	45.01	42.88	18.91	18.82	14.57	14	10.21	11.65	12.73	16.07	26.38	23.13
	LSTM	43.8	42.87	38.18	11.11	12.96	5.34	4.78	2.8	8.69	13.2	9.31	21.65	17.89
RMSE	ARMA	5.73	7.06	4.65	1.9	3.46	4.17	4.37	2.79	1.41	2.61	2.95	4.2	3.77
	RBC	5.43	5.56	5.04	1.97	1.18	0.75	0.7	0.67	1.1	1.44	1.92	3.3	2.42
	LSTM	5.34	5.55	4.68	1.21	0.87	0.31	0.26	0.2	0.86	1.57	1.16	2.82	2.07
MAPE	ARMA	39.01	49.12	34.06	13.73	47.35	72.52	79.56	38.02	12.51	20.54	22.66	28.45	38.13
	RBC	37.47	36.99	32.42	12.82	14.95	10.97	9.74	6.86	8.51	8	9.75	18.56	17.25
	LSTM	36.9	34.37	29.63	7.54	9.34	3.72	3.44	2.34	6.95	8.55	7.54	17.21	13.96
MAE	ARMA	4.75	6.36	4.18	1.5	3.19	4.16	4.36	2.7	1.24	2.44	2.82	3.71	3.45
	RBC	4.39	4.57	3.81	1.33	0.94	0.57	0.48	0.45	0.8	0.9	1.17	2.32	1.81
	LSTM	4.5	4.45	3.64	0.82	0.63	0.21	0.19	0.17	0.69	1.02	0.94	2.25	1.62
Precisión	ARMA	60.99	50.88	65.94	86.27	52.65	27.48	20.44	61.98	87.49	79.46	77.34	71.55	61.87
	RBC	62.53	63.01	67.58	87.18	85.05	89.03	90.26	93.14	91.49	92	90.25	81.44	82.75
	LSTM	63.1	65.63	70.37	92.46	90.66	96.28	96.56	97.66	93.05	91.45	92.46	82.79	86.04
Correlación	ARMA	0.12	0.33	0.24	0.81	0.86	0.77	0.54	0.97	0.87	0.11	0.24	0.3	0.01
	RBC	0.33	0.53	0.07	0.42	0.6	0.13	0.17	0.85	0.67	0.46	0.06	0.03	0.35
	LSTM	0.36	0.55	0.07	0.69	0.79	0.56	0.27	0.97	0.81	0.35	0.22	0.08	0.48

Fuente: Propia

Tabla 3

Métricas de medición de Modelos ARMA, RBC, LSTM para los meses junio
(Mínimos) de los años 2011-2017

		2011	2012	2013	2014	2015	2016	2017	Prom.
%RMSE	ARMA	230.9	104.2	63.44	66.37	127.9	85.16	72.73	107.2
	RBC	9.68	4.58	6.4	18.73	10.74	13.79	14.57	11.21
	LSTM	20.24	8.61	10.45	6.7	12.48	13.19	5.34	11.
RMSE	ARMA	7.54	4.17	2.56	2.61	4.86	5.17	4.17	4.44
	RBC	0.32	0.18	0.26	0.61	0.41	0.84	0.75	0.48
	LSTM	0.66	0.34	0.42	0.26	0.47	0.80	0.31	0.47
MAPE	ARMA	230.5	104.1	63.28	66.3	127.4	84.49	72.52	1075.8
	RBC	3.06	0.92	1.65	11.34	4.39	8.24	10.97	8.9
	LSTM	18.93	7.37	9.26	4.89	9.33	9.36	3.72	8
MAE	ARMA	7.53	4.17	2.55	2.61	4.84	5.13	4.16	4.43
	RBC	0.1	0.04	0.07	0.37	0.17	0.50	0.57	0.26
	LSTM	0.62	0.29	0.37	0.19	0.35	0.57	0.21	0.37
Precisión	ARMA	13.05	4.14	36.72	33.7	27.41	15.51	27.48	6.95
	RBC	96.94	99.08	98.35	88.66	95.61	91.76	89.03	94.2
	LSTM	81.07	92.63	90.74	95.12	90.67	90.64	96.28	91.02
Correlación	ARMA	0.77	0.03	0.16	0.74	0.55	0.61	0.77	0.35
	RBC	0.84	0.03	0.03	0.54	0.55	0.46	0.13	0.35
	LSTM	0.87	0.05	0.07	0.51	0.52	0.51	0.56	0.41

Fuente: Propia

Tabla 4

Métricas de medición de Modelos ARMA, RBC, LSTM para los meses diciembre (Máximos) de los años 2011-2017

		2011	2012	2013	2014	2015	2016	2017	Prom.
%RMSE	ARMA	24.74	26.03	35.00	34.23	32.50	34.23	32.19	31.27
	RBC	26.83	23.23	17.13	15.23	25.11	14.84	26.38	21.25
	LSTM	23.88	17.13	13.46	11.73	24.96	15.88	21.65	18.38
RMSE	ARMA	2.72	2.29	2.89	2.90	2.37	3.47	4.20	2.98
	RBC	2.95	1.93	1.33	1.22	1.82	1.50	3.30	2.01
	LSTM	2.63	1.51	1.11	1.00	1.82	1.61	2.82	1.78
MAPE	ARMA	22.00	22.75	32.38	32.07	22.75	32.84	28.45	27.61
	RBC	20.53	18.29	13.69	10.48	14.67	10.19	18.56	15.20
	LSTM	17.83	13.07	10.22	8.76	16.06	10.76	17.21	13.41
MAE	ARMA	2.42	2.00	2.67	2.72	1.66	3.33	3.71	2.64
	RBC	2.26	1.52	1.07	0.84	1.07	1.03	2.32	1.44
	LSTM	1.96	1.15	0.84	0.74	1.17	1.09	2.25	1.31
Precisión	ARMA	75.26	73.97	65.00	65.77	67.50	65.77	67.81	68.73
	RBC	73.17	76.77	82.87	84.77	74.89	85.16	73.62	78.75
	LSTM	76.12	82.87	86.54	88.27	75.04	84.12	78.35	81.62
Correlación	ARMA	0.41	0.43	0.44	0.44	0.40	0.02	0.30	0.14
	RBC	0.12	0.33	0.21	0.20	0.35	0.25	0.03	0.20
	LSTM	0.24	0.30	0.35	0.40	0.35	0.01	0.08	0.25

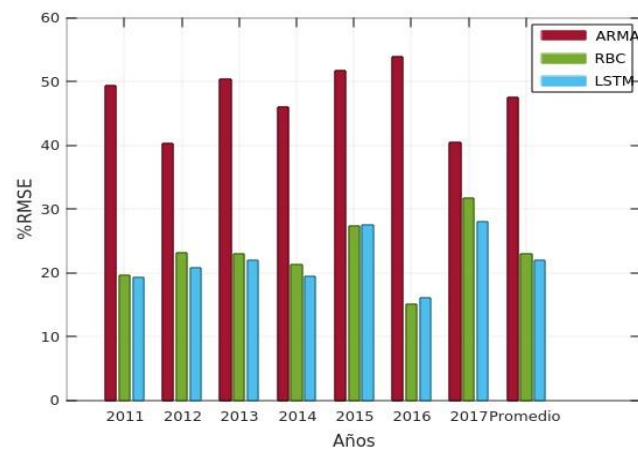
Fuente: Propia

5.2 RESULTADOS ANUALES 2011-2017

En esta sección se grafican los resultados ANUALES de los indicadores % RMSE, RMSE, MAPE, MAE, Precisión, Correlación para los modelos ARMA, RBC, LSTM.

Figura 14

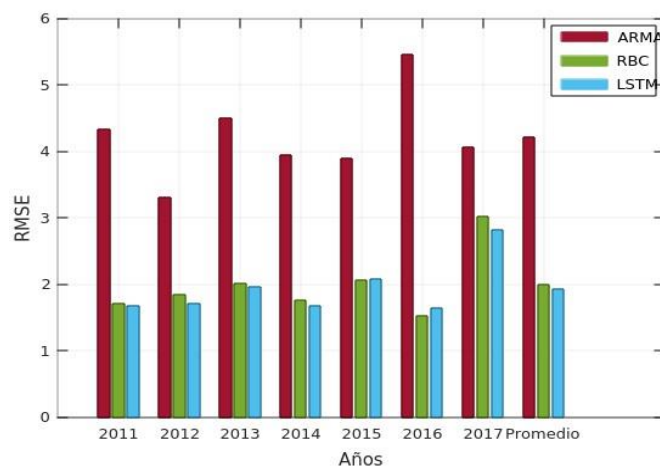
% RMSE de los Modelos ARMA, RBC, LSTM para los años 2011-2017



Fuente: Propia

Figura 15

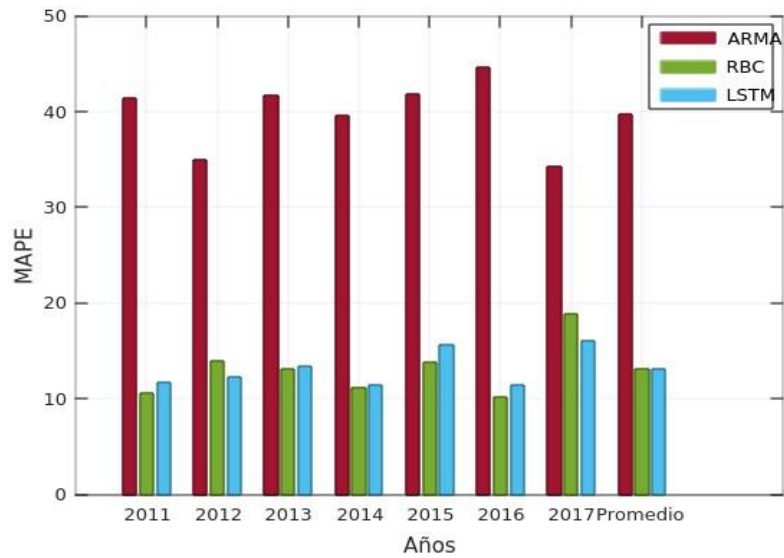
RMSE de los Modelos ARMA, RBC, LSTM para los años 2011-2017



Fuente: Propia

Figura 16

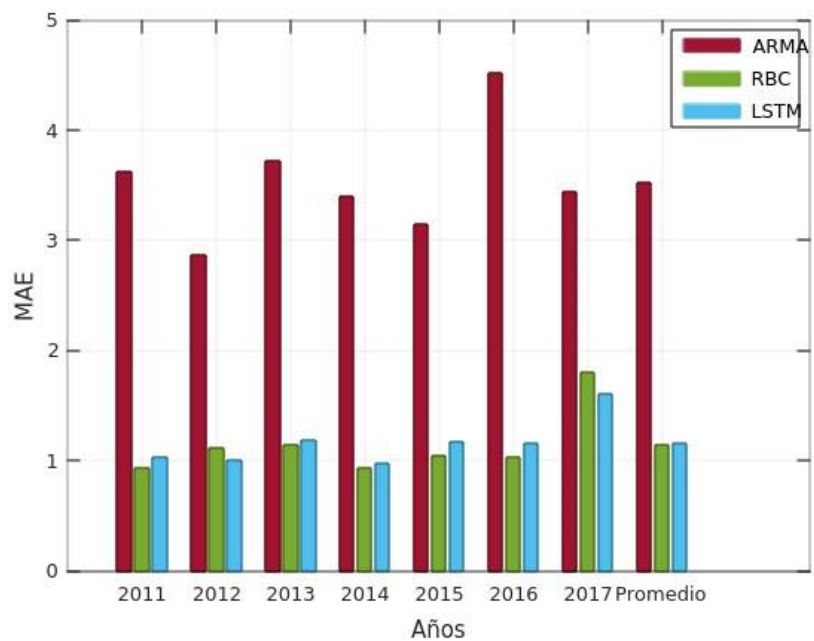
MAPE de los Modelos ARMA, RBC, LSTM para los años 2011-2017



Fuente: Propia

Figura 17

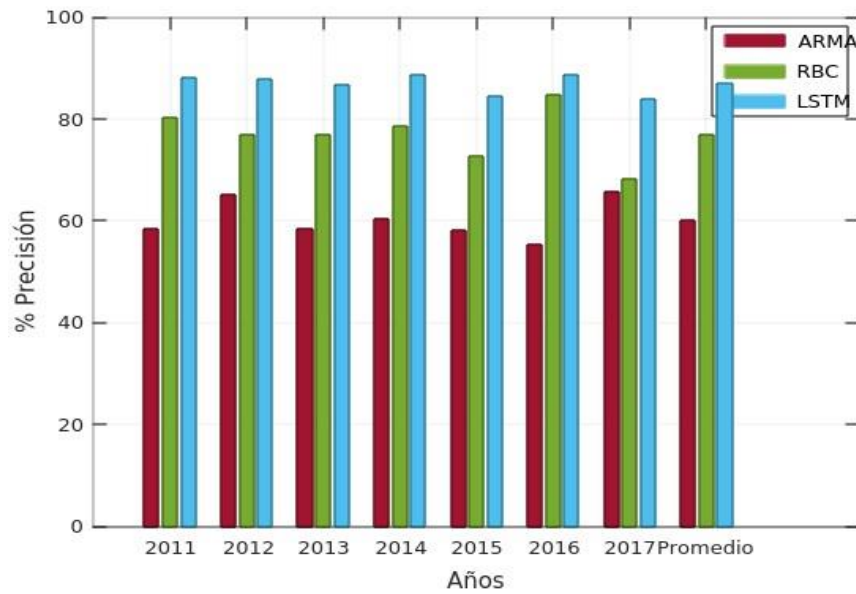
MAE de los Modelos ARMA, RBC, LSTM para los años 2011-2017



Fuente: Propia

Figura 18

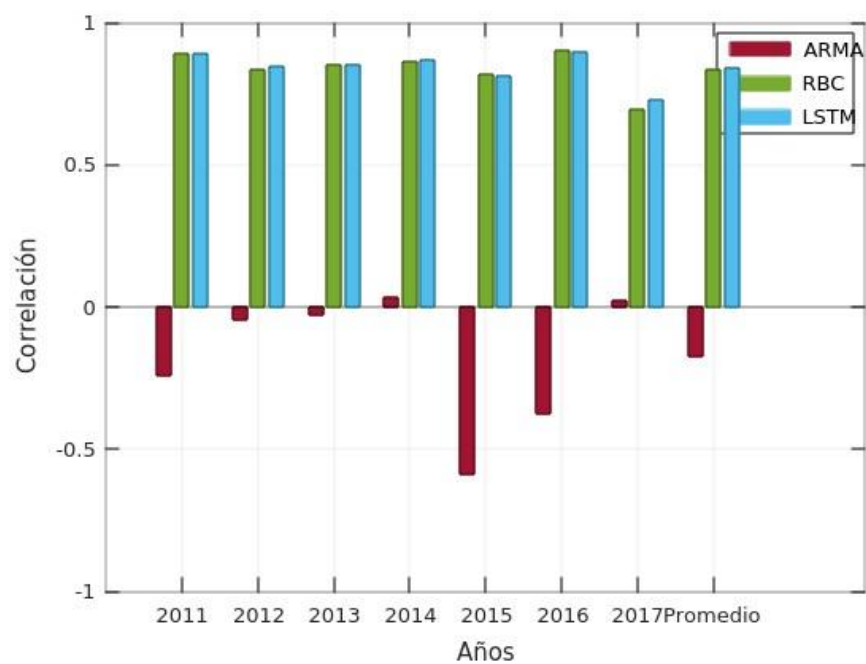
Precisión de los Modelos ARMA, RBC, LSTM para los años 2011-2017



Fuente: Propia

Figura 19

Correlación de los Modelos ARMA, RBC, LSTM para los años 2011-2017



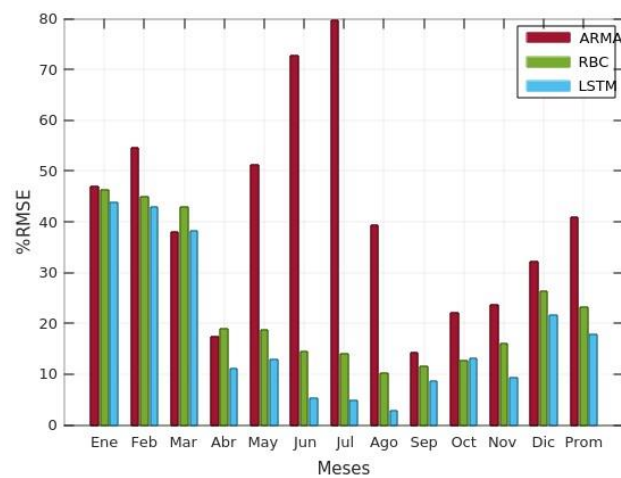
Fuente: Propia

5.3 RESULTADOS MENSUALES AÑO 2017

En esta sección se grafican los resultados MENSUALES para el año 2017, de los indicadores % RMSE, RMSE, MAPE, MAE, Precisión, Correlación para los modelos ARMA, RBC, LSTM.

Figura 20

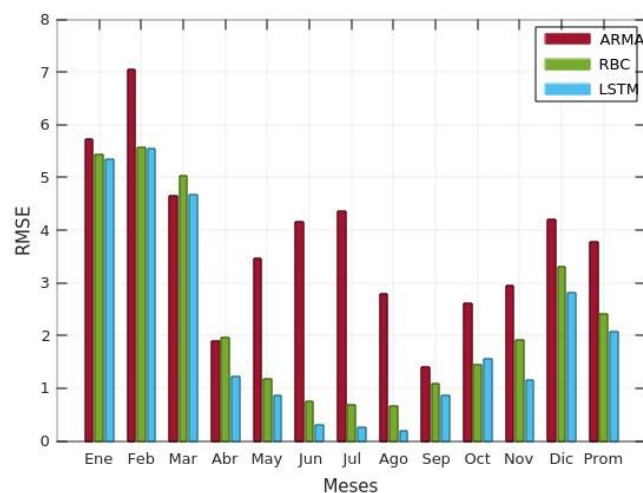
% RMSE de los Modelos ARMA, RBC, LSTM para los meses del año 2017



Fuente: Propia

Figura 21

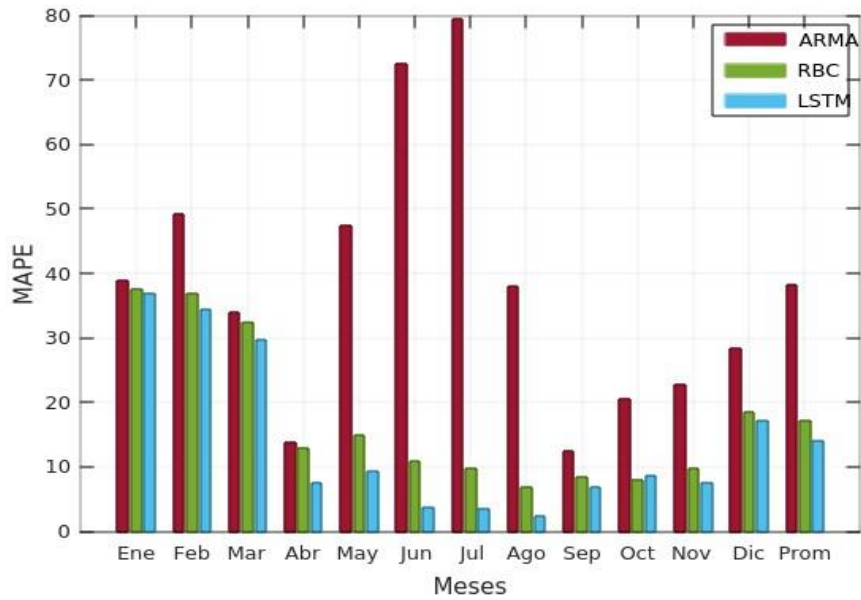
RMSE de los Modelos ARMA, RBC, LSTM para los meses del año 2017



Fuente: Propia

Figura 22

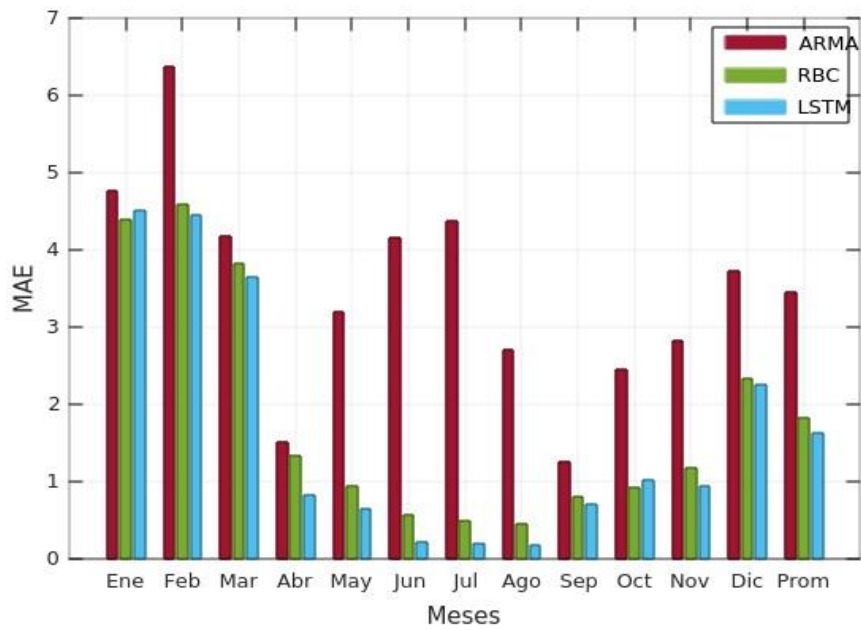
MAPE de los Modelos ARMA, RBC, LSTM para los meses del año 2017



Fuente: Propia

Figura 23

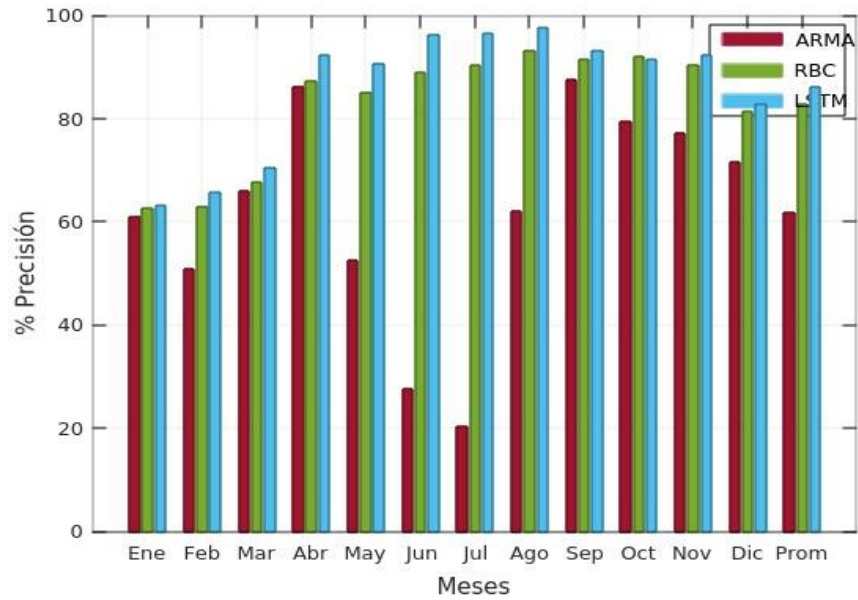
MAE de los Modelos ARMA, RBC, LSTM para los meses del año 2017



Fuente: Propia

Figura 24

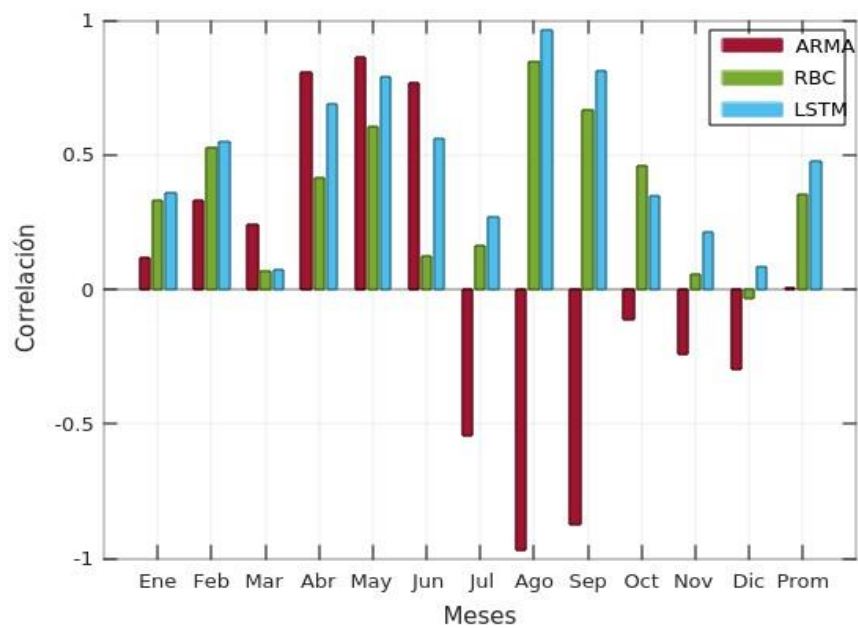
Precisión de los Modelos ARMA, RBC, LSTM para los meses del año 2017



Fuente: Propia

Figura 25

Correlación de los Modelos ARMA, RBC, LSTM para los meses del año 2017



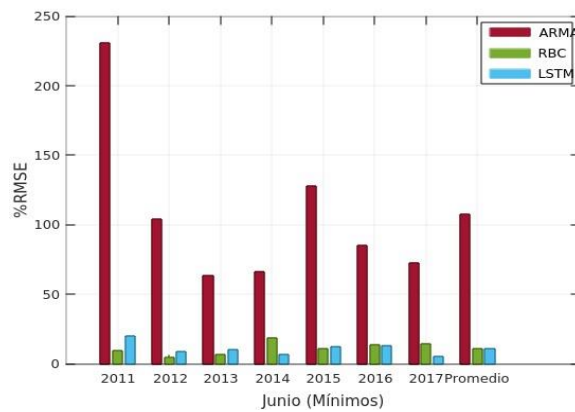
Fuente: Propia

5.4 RESULTADOS MÍNIMOS 2011-2017

En esta sección se grafican los resultados de meses Junio (MINIMOS) para los años 2011-2017 de los indicadores % RMSE, RMSE, MAPE, MAE, Precisión, Correlación para los modelos ARMA, RBC, LSTM.

Figura 26

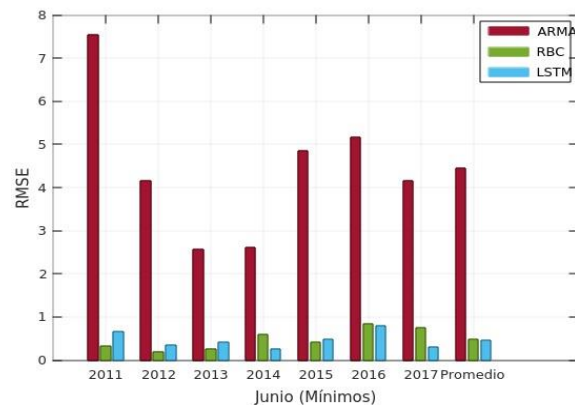
% RMSE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



Fuente: Propia

Figura 27

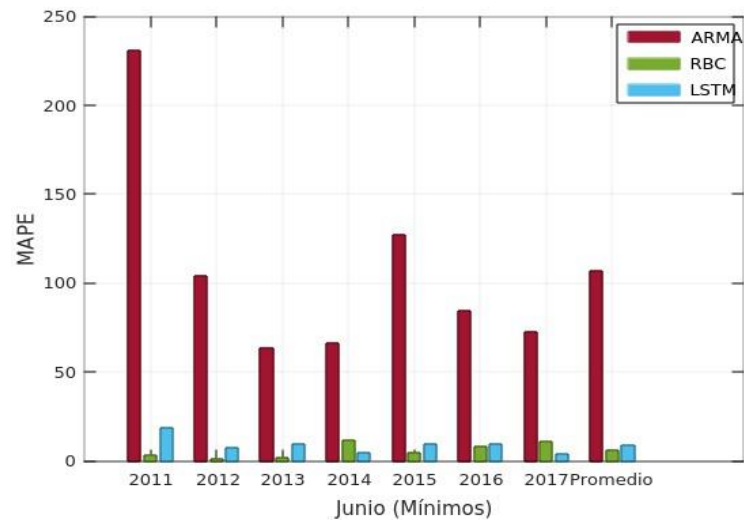
RMSE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



Fuente: Propia

Figura 28

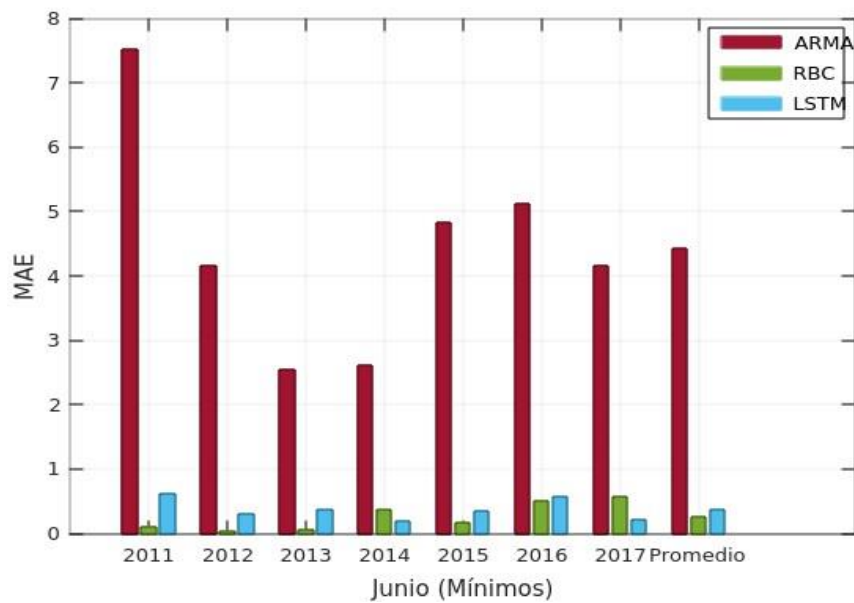
MAPE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



Fuente: Propia

Figura 29

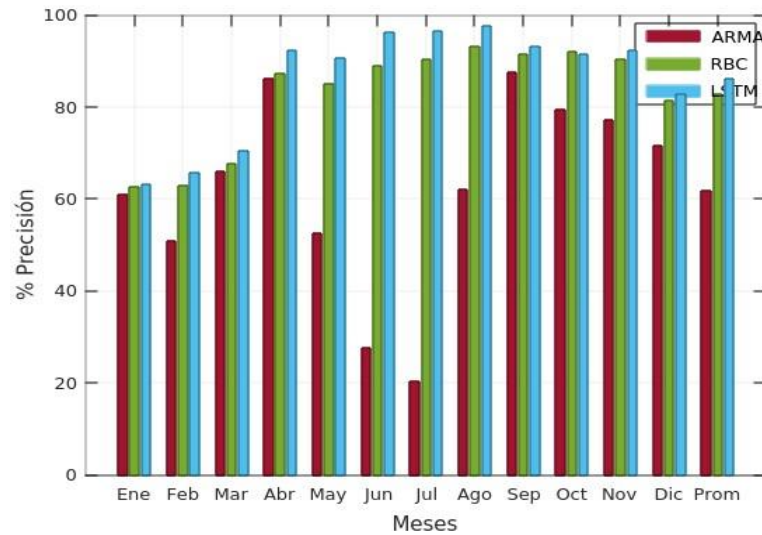
MAE de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



Fuente: Propia

Figura 30

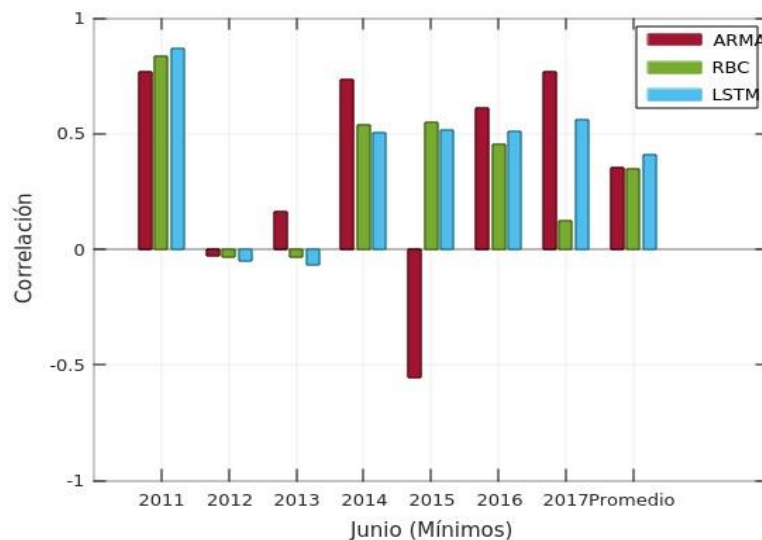
Precisión de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



Fuente: Propia

Figura 31

Correlación de los Modelos ARMA, RBC, LSTM para los meses junio (mínimos) de los años 2011-2017



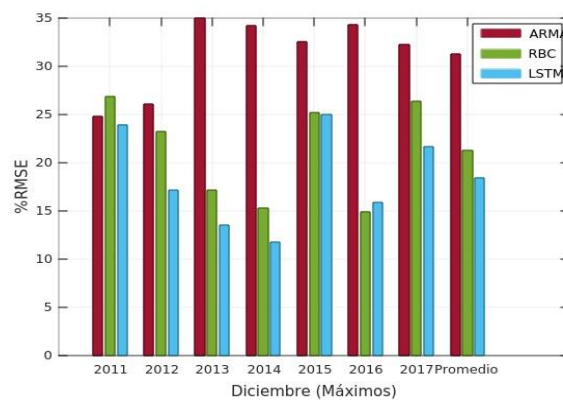
Fuente: Propia

5.5 RESULTADOS MÁXIMOS 2011-2017

En esta sección se grafican los resultados de meses Diciembre (MÁXIMOS) para los años 2011-2017 de los indicadores % RMSE, RMSE, MAPE, MAE, Precisión, Correlación para los modelos ARMA, RBC, LSTM.

Figura 32

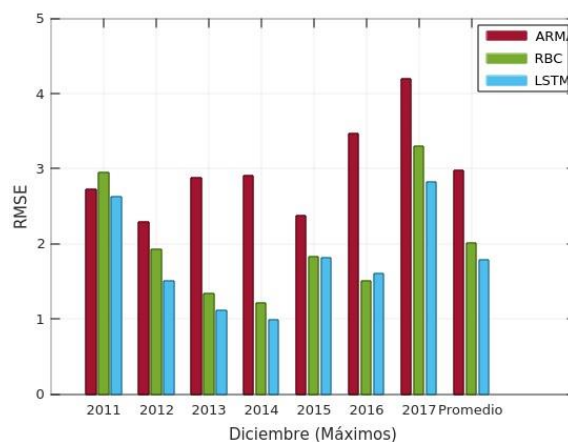
% RMSE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

Figura 33

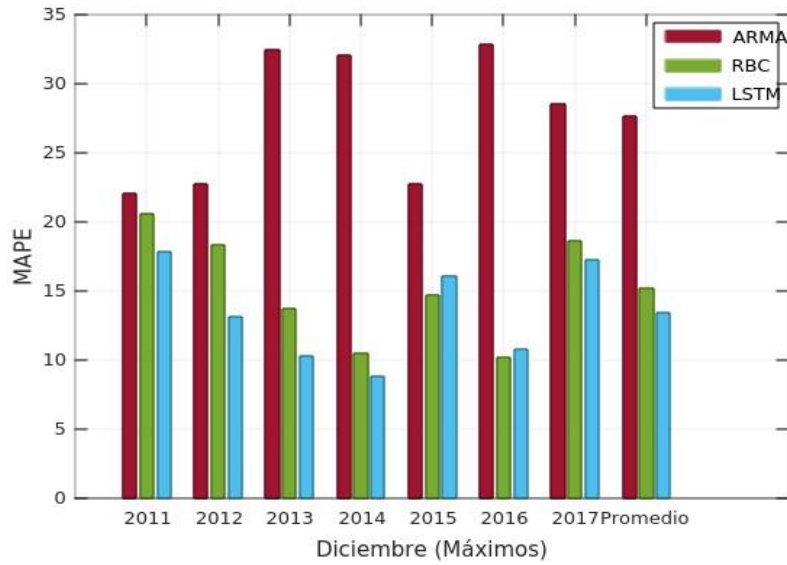
RMSE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

Figura 34

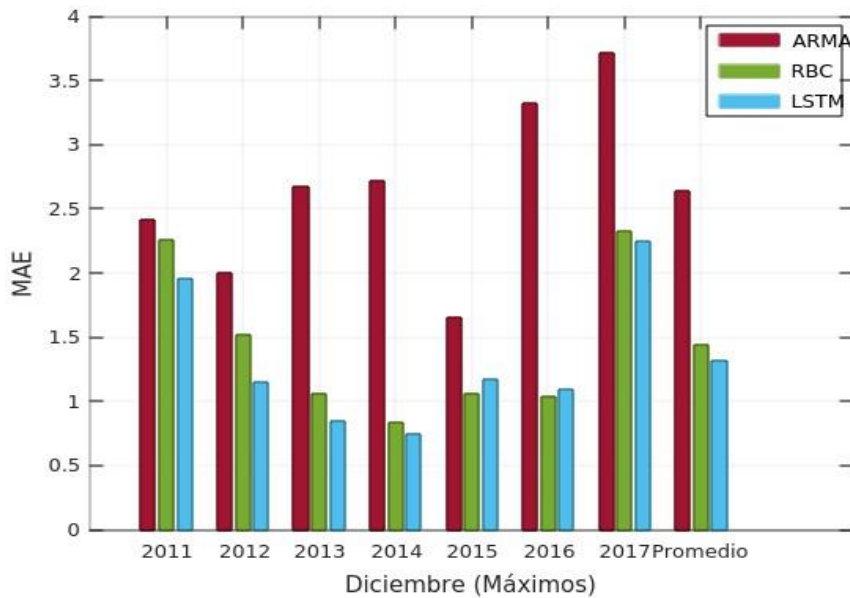
MAPE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

Figura 35

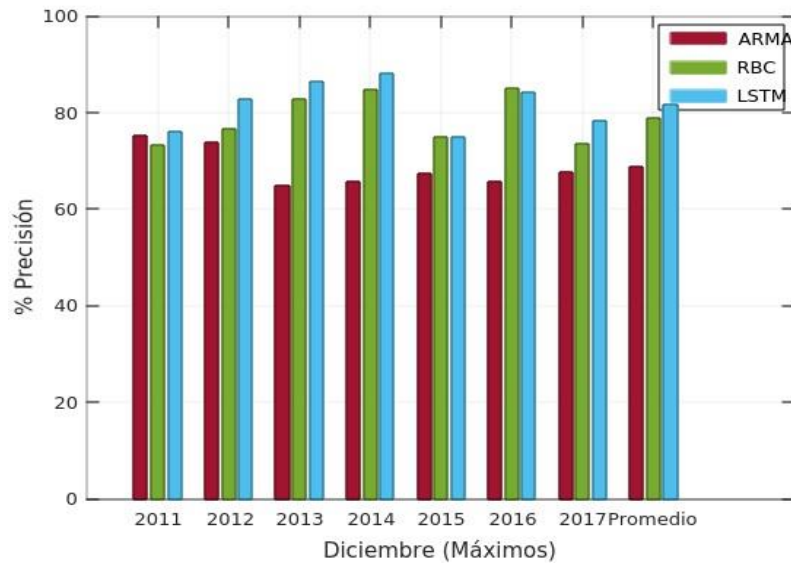
MAE de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

Figura 36

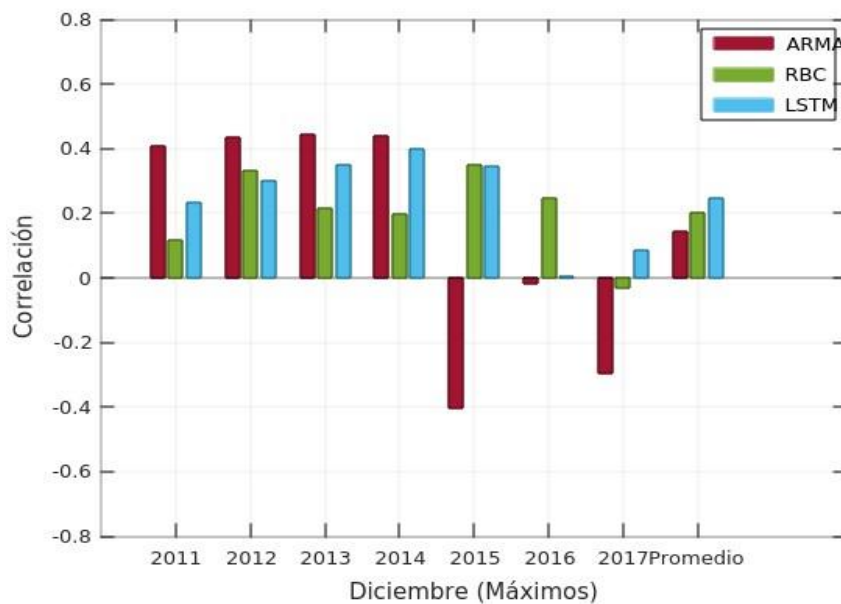
Precisión de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

Figura 37

Correlación de los Modelos ARMA, RBC, LSTM para los meses diciembre (máximos) de los años 2011-2017



Fuente: Propia

5.6 AUTOREGRESIÓN DE MEDIAS MÓVILES (ARMA)

Se grafica los valores de test AIC según diversos parámetros (p, q) que sirven para definir modelo ARMA.

Tabla 5

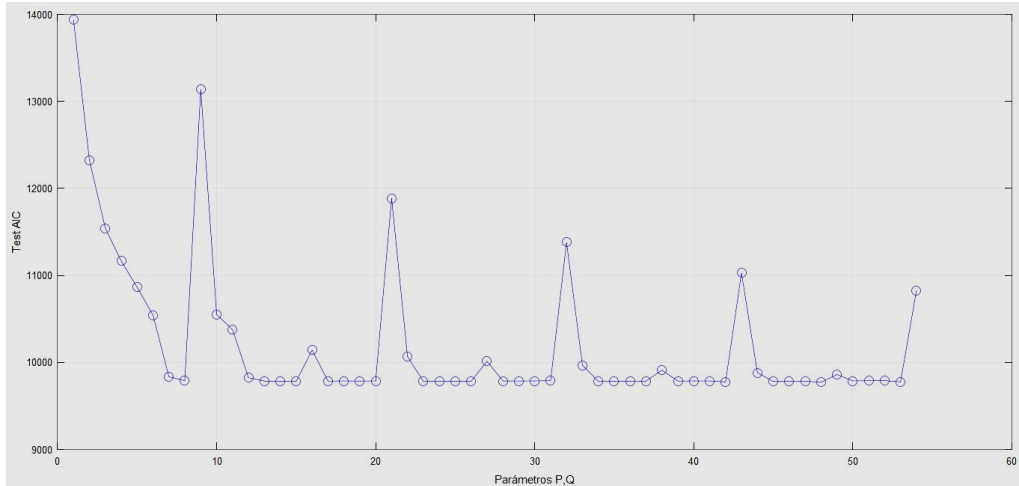
Test AIC de Parámetros P, D, Q del modelo ARMA

n	P	D	q	aic	n	p	d	q	aic	n	p	d	q	aic
1	0	0	0	13937.32	21	1	2	0	11885.98	41	3	1	3	9787.12
2	0	0	1	12325.27	22	2	0	0	10067.61	42	3	1	4	9775.38
3	0	0	2	11543.08	23	2	0	1	9781.40	43	3	2	0	11028.74
4	0	0	3	11165.06	24	2	0	2	9783.15	44	4	0	0	9881.88
5	0	0	4	10863.13	25	2	0	3	9781.53	45	4	0	1	9780.95
6	0	1	0	10543.88	26	2	0	4	9783.34	46	4	0	2	9783.11
7	0	1	1	9834.94	27	2	1	0	10013.10	47	4	0	3	9784.34
8	0	1	2	9790.73	28	2	1	1	9787.05	48	4	0	4	9770.70
9	0	2	0	13136.94	29	2	1	2	9786.39	49	4	1	0	9861.64
10	0	2	1	10550.58	30	2	1	3	9787.60	50	4	1	1	9787.58
11	1	0	0	10376.33	31	2	1	4	9789.57	51	4	1	2	9789.14
12	1	0	1	9825.05	32	2	2	0	11379.58	52	4	1	3	9790.89
13	1	0	2	9785.61	33	3	0	0	9967.07	53	4	1	4	9776.41
14	1	0	3	9780.93	34	3	0	1	9783.02	54	4	2	0	10825.39
15	1	0	4	9781.86	35	3	0	2	9782.39					
16	1	1	0	10142.27	36	3	0	3	9783.36					
17	1	1	1	9785.65	37	3	0	4	9785.34					
18	1	1	2	9787.23	38	3	1	0	9910.77					
19	1	1	3	9785.95	39	3	1	1	9785.59					
20	1	1	4	9787.58	40	3	1	2	9787.45					

Fuente: Propia

Figura 38

Test AIC de Parámetros P, D, Q del modelo ARMA



Fuente: Propia

Durante el modelamiento Auto-Regresivo de Medias Móviles (ARMA), se han obtenido resultados de predicción desde el año 2011 al 2017, de acuerdo a la raíz del error medio cuadrático (RMSE) y coeficiente de regresión Lineal (r) que se muestra en la tabla 6 y se visualiza en la figura 39 y figura 40.

Tabla 6

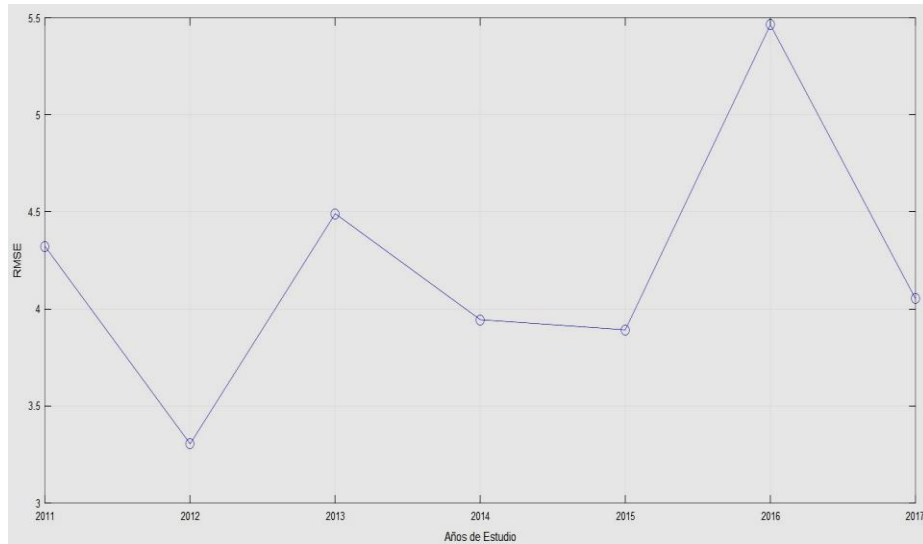
RSME y Coeficiente de Correlación para el periodo 2011-2017

AÑO	RMSE	r
2017	4.053	0.022
2016	5.464	-0.375
2015	3.892	-0.588
2014	3.944	0.032
2013	4.491	-0.027
2012	3.305	-0.042
2011	4.322	-0.238
PROMEDIO	4.210	-0.174

Fuente: Propia

Figura 39

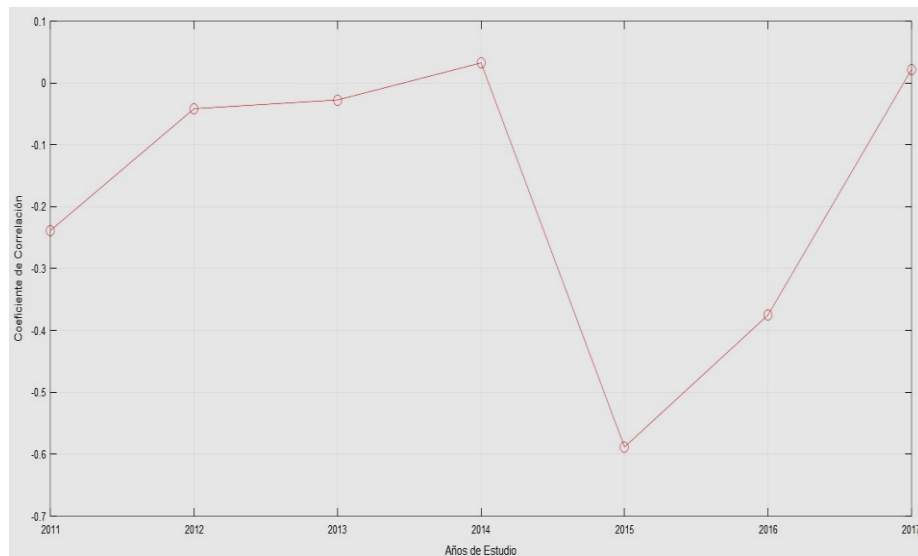
RMSE del periodo 2011-2017 para el modelo ARMA



Fuente: Propia

Figura 40

Coefficiente de Correlación del periodo 2011-2017 para el modelo ARMA



Fuente: Propia

Enseguida se tienen la Raíz de Error Medio Cuadrático (RMSE) y Coeficiente de Regresión Lineal (r) para los distintos meses del año 2017 que se muestra en la tabla 7 y se visualiza en las figuras 41 y 42.

Tabla 7

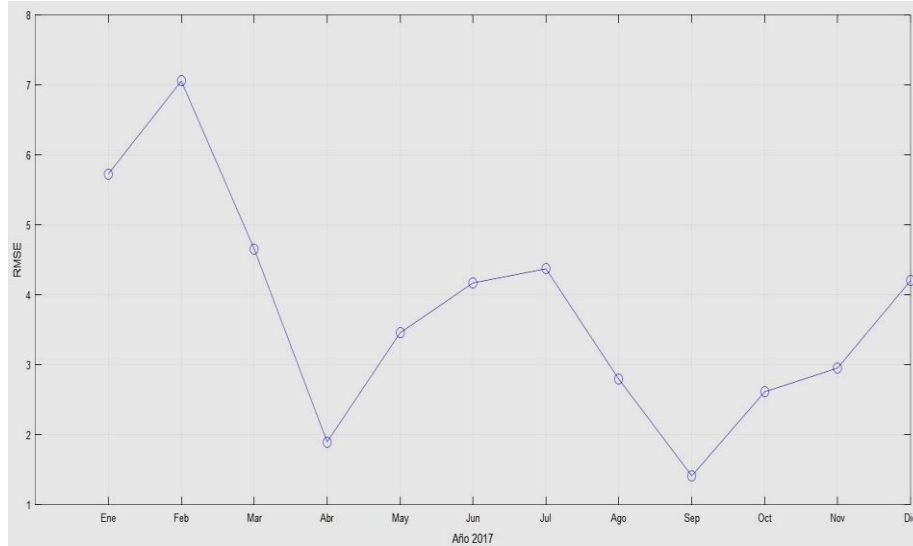
RSME y Coeficiente de Correlación para el periodo enero-diciembre del año 2017

MES	RMSE	R
ENERO	5.728	0.116
FEBRERO	7.055	0.329
MARZO	4.652	0.239
ABRIL	1.896	0.808
MAYO	3.455	0.861
JUNIO	4.169	0.771
JULIO	4.370	-0.545
AGOSTO	2.792	-0.968
SEPTIEMBRE	1.410	-0.874
OCTUBRE	2.612	-0.114
NOVIEMBRE	2.952	-0.238
DICIEMBRE	4.200	-0.297
PROMEDIO	3.774	0.007

Fuente: Propia

Figura 41

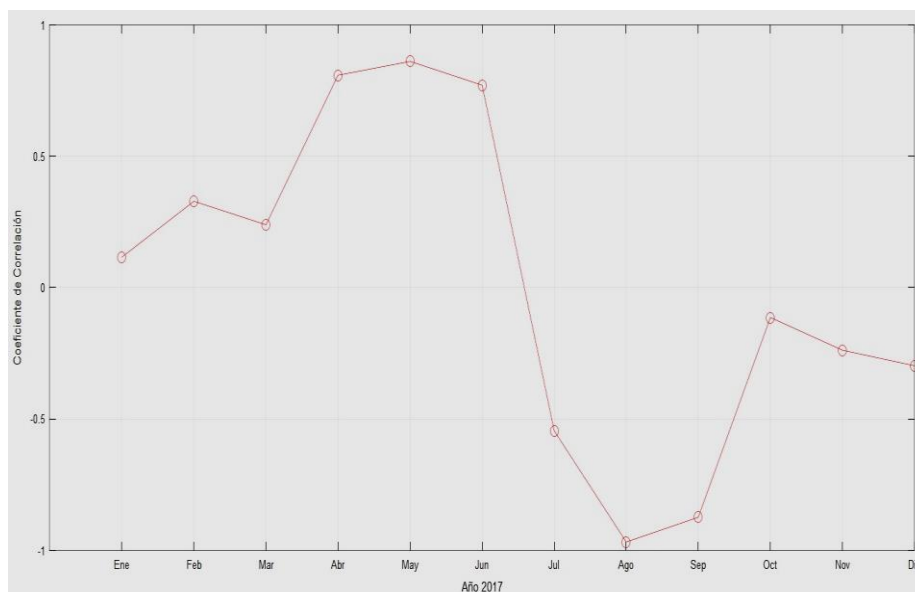
RMSE del periodo enero-diciembre 2017 para el modelo ARMA



Fuente: Propia

Figura 42

Coefficiente de Correlación del periodo enero-diciembre 2017 para el modelo ARMA



Fuente: Propia

Finalmente, se presenta los resultados de predicción los índices de prueba e índices de predicción para los días del mes de enero del año 2017, a partir de datos del Ministerio de Salud – Arequipa, que se muestra en la tabla 8 y se visualiza en la figura 43 y la dispersión del año 2017 en la figura 44.

Tabla 8

Índice de Radiación Ultravioleta y Predicción para el mes enero2017

DÍA	Índice UV	Predicción	DÍA	Índice UV	Predicción
1	12.240	10.617	17	11.002	10.110
2	10.152	10.265	18	4.608	10.185
3	13.104	9.972	19	11.074	10.152
4	16.531	9.914	20	5.170	10.053
5	14.026	10.081	21	4.810	9.993
6	10.742	10.264	22	4.608	10.026
7	20.390	10.269	23	10.454	10.109
8	17.237	10.106	24	14.962	10.152
9	14.256	9.952	25	9.533	10.115
10	4.277	9.958	26	18.878	10.042
11	18.461	10.101	27	17.482	10.008
12	7.646	10.222	28	19.685	10.042
13	20.635	10.202	29	18.302	10.102
14	10.267	10.073	30	10.930	10.124
15	3.902	9.974	31	17.770	10.088
16	4.651	10.000			

Fuente: Propia

5.7 RAZONAMIENTO BASADO EN CASOS (RBC)

Se grafica los valores RMSE de k Vecinos Cercanos, que sirve para definir modelo RBC.

Tabla 9

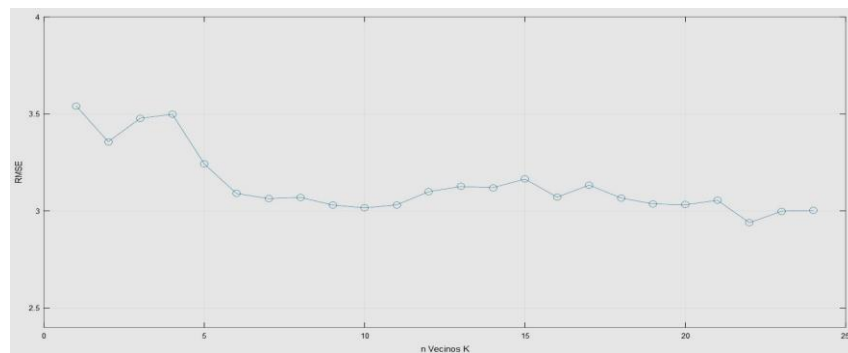
RMSE de Parámetros K del modelo RBC

K	RMSE	K	RMSE
1	3.54	13	3.13
2	3.35	14	3.12
3	3.47	15	3.16
4	3.49	16	3.07
5	3.24	17	3.13
6	3.09	18	3.06
7	3.06	19	3.03
8	3.07	20	3.03
9	3.03	21	3.05
10	3.01	22	2.93
11	3.03	23	2.99
12	3.09	24	3.00

Fuente: Propia

Figura 45

RMSE de Parámetros K del modelo RBC



Fuente: Propia

Durante el modelamiento Razonamiento Basado en Casos (RBC) se han obtenido resultados de predicción desde el año 2011 al 2017, de acuerdo a la Raíz de Error Medio Cuadrático (RMSE) y Coeficiente de Regresión Lineal (r) que se muestra en la tabla 10 y se visualiza en las figuras 45 y 47.

Tabla 10

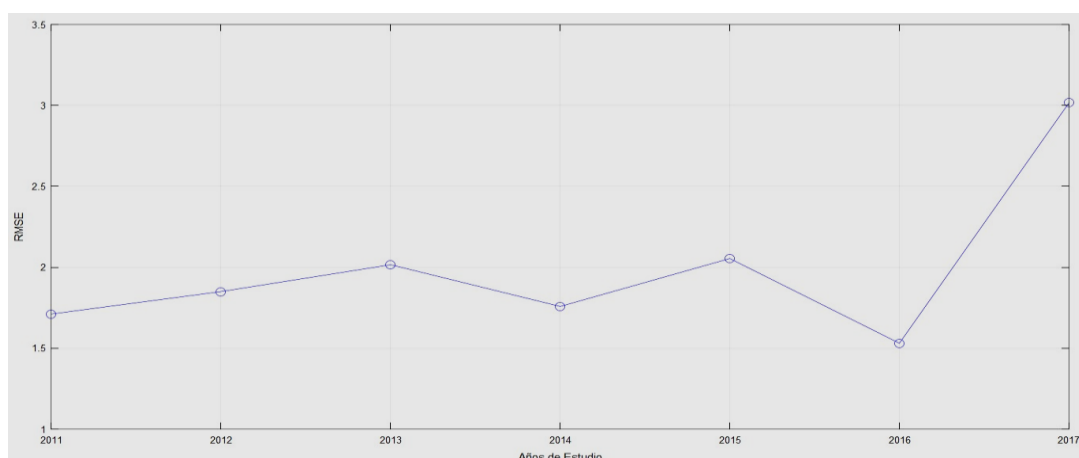
RMSE y Coeficiente de Correlación del periodo 2011-2017 para el modelo RBC

AÑO	RMSE	R
2017	3.016	0.695
2016	1.531	0.904
2015	2.054	0.820
2014	1.758	0.865
2013	2.016	0.850
2012	1.850	0.837
2011	1.711	0.890
PROMEDIO	1.991	0.837

Fuente: Propia

Figura 46

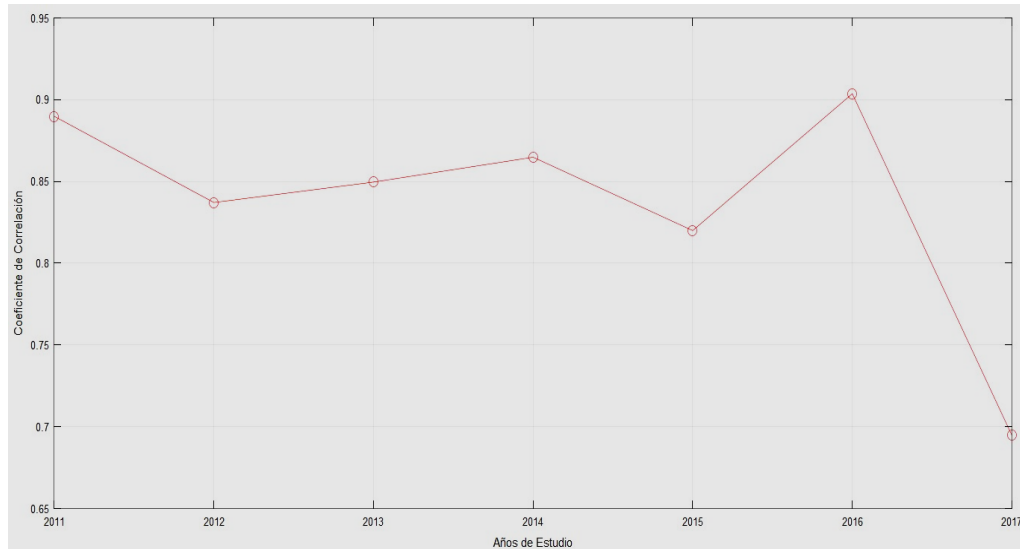
RMSE del periodo 2011-2017 para el modelo RBC



Fuente: Propia

Figura 47

Coefficiente de Correlación del periodo 2011-2017 para el modelo. RBC

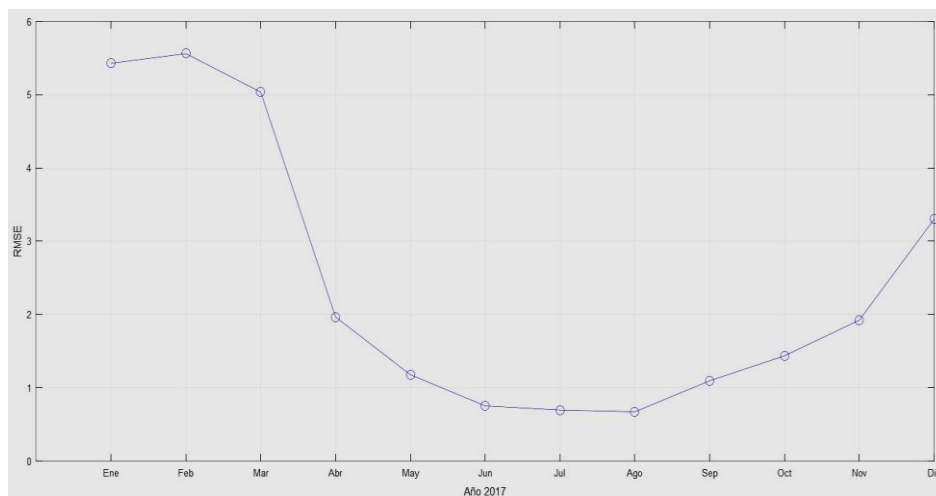


Fuente: Propia

Enseguida, se tienen la raíz de error medio cuadrático (RMSE) y coeficiente de regresión lineal (r), para los distintos meses del año 2017 que se muestra en la tabla 11 y se visualiza en las figuras 48 y 49.

Figura 48

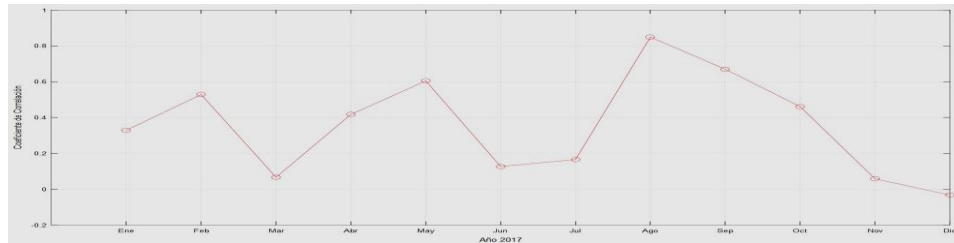
RMSE del periodo enero-diciembre 2017 para el modelo RBC



Fuente: Propia

Figura 49

Coeficiente de correlación del periodo enero-diciembre 2017 para el modelo RBC



Fuente: Propia

Tabla 11

RSME y Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo RBC

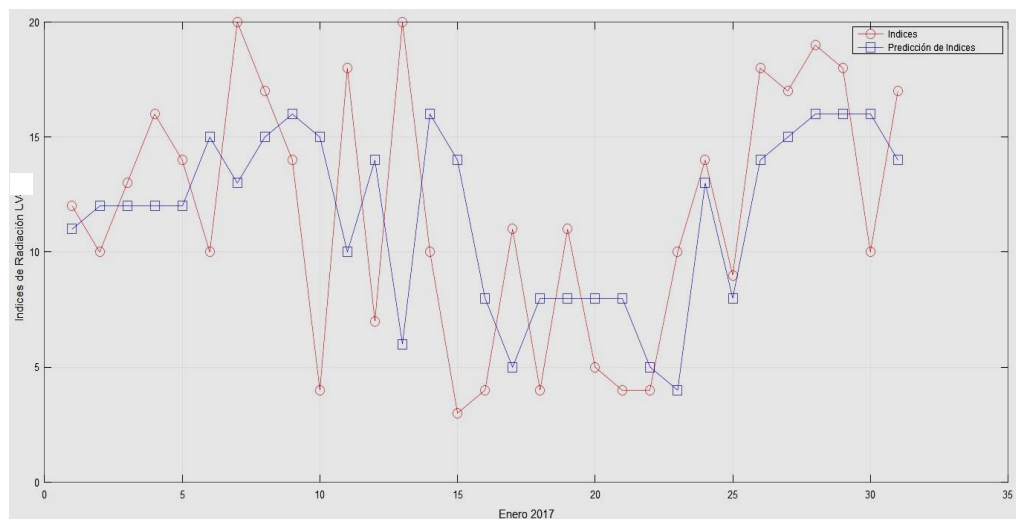
MES	RMSE	r
ENERO	5.430	0.329
FEBRERO	5.561	0.528
MARZO	5.035	0.066
ABRIL	1.966	0.417
MAYO	1.178	0.605
JUNIO	0.753	0.126
JULIO	0.696	0.166
AGOSTO	0.672	0.849
SEPTIEMBRE	1.095	0.670
OCTUBRE	1.437	0.461
NOVIEMBRE	1.924	0.058
DICIEMBRE	3.302	-0.033
PROMEDIO	2.421	0.354

Fuente: Propia

Finalmente, se presenta los resultados de Predicción los índices de Prueba e índices de Predicción, para los días del mes de enero del año 2017 a partir de datos del Ministerio de Salud – Arequipa, que se muestra en la tabla 12 y se visualiza en la figura 50; y la dispersión del año 2017 en la figura 51.

Figura 50

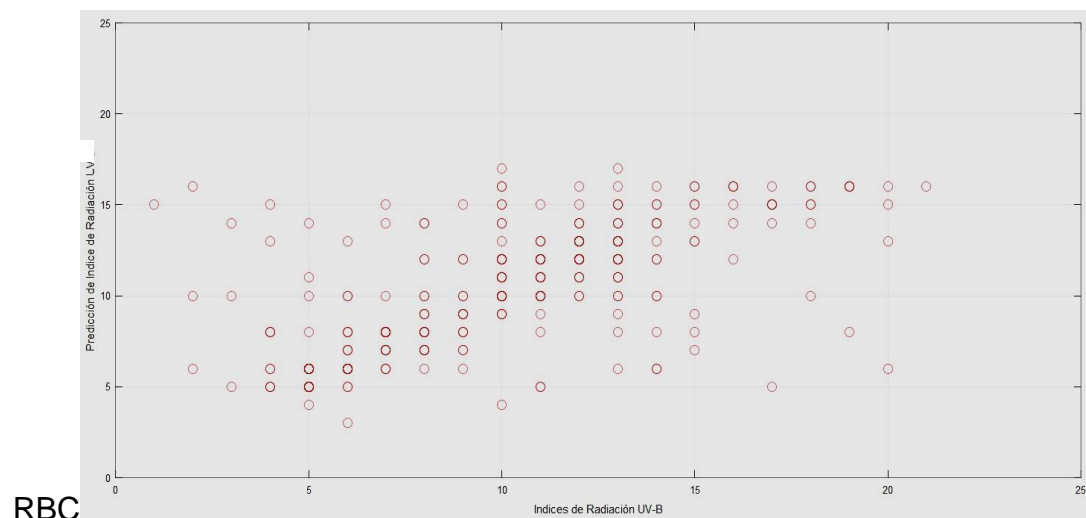
Predicción de Índices de Radiación Ultravioleta enero 2017 para el modelo RBC



Fuente: Propia

Figura 51

Dispersión de los Índice de Radiación Ultravioleta del año 2017 para el modelo



RBC

Fuente: Propia

Tabla 12

Índice de Radiación Ultravioleta y Predicción para el mes enero 2017 para el modelo RBC

DÍA	Índice UV	Predicción	DÍA	Índice UV	Predicción
1	12.240	11	17	11.002	5
2	10.152	12	18	4.608	8
3	13.104	12	19	11.074	8
4	16.531	12	20	5.170	8
5	14.026	12	21	4.810	8
6	10.742	15	22	4.608	5
7	20.390	13	23	10.454	4
8	17.237	15	24	14.962	13
9	14.256	16	25	9.533	8
10	4.277	15	26	18.878	14
11	18.461	10	27	17.482	15
12	7.646	14	28	19.685	16
13	20.635	6	29	18.302	16
14	10.267	16	30	10.930	16
15	3.902	14	31	17.770	14
16	4.651	8			

Fuente: Propia

5.8 REDES NEURONALES CON MEMORIA DE LARGO Y CORTO PLAZO (LSTM)

Se muestran los valores RMSE del número de Células LSTM que sirven para definir modelo LSTM.

Tabla 13

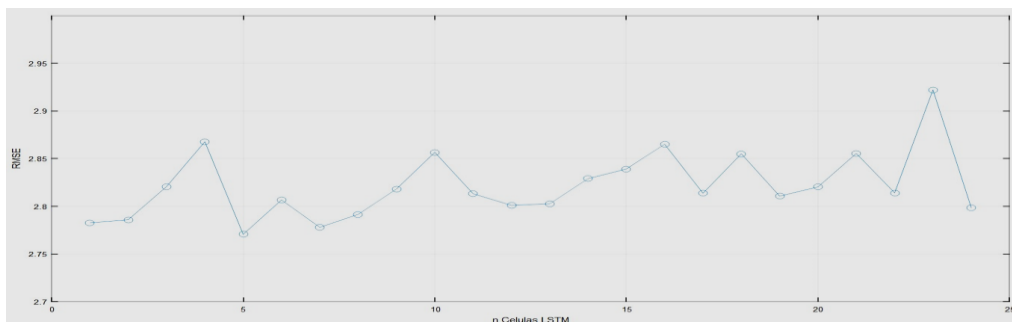
RMSE del Número de Células LSTM del modelo LSTM

n	RMSE	n	RMSE
1	2.782	13	2.802
2	2.780	14	2.829
3	2.820	15	2.839
4	2.867	16	2.865
5	2.770	17	2.813
6	2.806	18	2.855
7	2.778	19	2.810
8	2.791	20	2.820
9	2.818	21	2.855
10	2.856	22	2.814
11	2.813	23	2.922
12	2.801	24	2.798

Fuente: Propia

Figura 52

RMSE del Número de Células LSTM del modelo LSTM



Fuente: Propia

Durante el modelamiento de Redes Neuronales con Memoria de Largo y Corto Plazo (LSTM), se han obtenido resultados de predicción desde el año 2011 al 2017, de acuerdo a la Raíz de Error Medio Cuadrático (RMSE) y Coeficiente de Regresión Lineal (r) que se muestra en la tabla 14 y se visualiza en las figuras 53 y 54.

Tabla 14

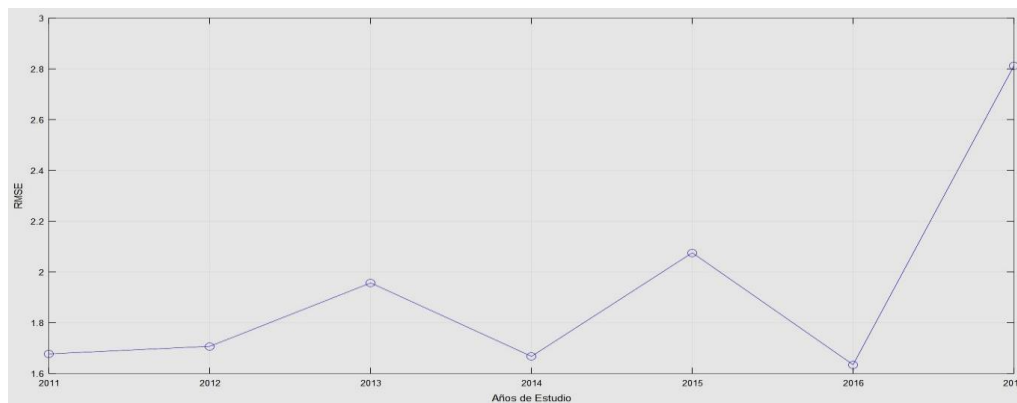
RSME y Coeficiente de Correlación del periodo 2011-2017 para el modelo LSTM

Año	RMSE	r
2017	2.812	0.728
2016	1.635	0.899
2015	2.076	0.810
2014	1.667	0.868
2013	1.957	0.854
2012	1.707	0.848
2011	1.677	0.891
PROMEDIO	1.933	0.843

Fuente: Propia

Figura 53

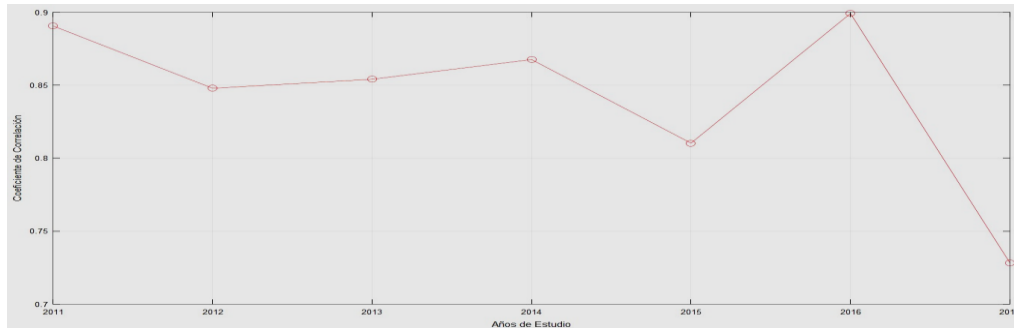
RMSE del periodo 2011-2017 para el modelo LSTM



Fuente: Propia

Figura 54

Coefficiente de Correlación del periodo 2011-2017 para el modelo LSTM



Fuente: Propia

Enseguida, se tienen la Raíz de Error Medio Cuadrático (RMSE) y Coeficiente de Regresión Lineal (r), para los distintos meses del año 2017 que se muestra en la tabla 15 y se visualiza en las figuras 55 y 56.

Tabla 15

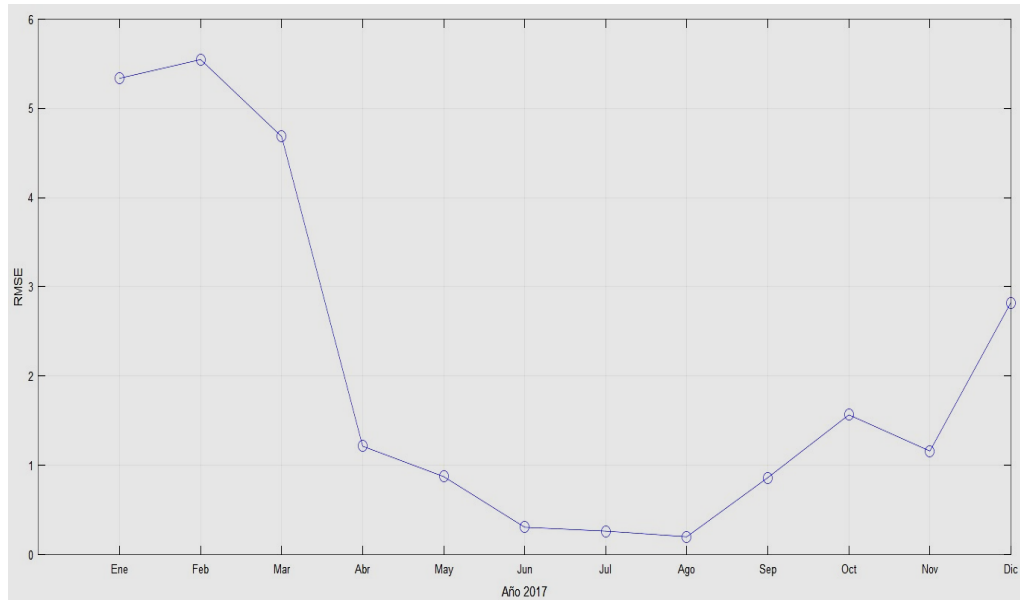
RSME y Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo LSTM

Mes	RMSE	R
ENERO	5.338	0.359
FEBRERO	5.548	0.553
MARZO	4.684	0.074
ABRIL	1.214	0.690
MAYO	0.873	0.789
JUNIO	0.306	0.561
JULIO	0.262	0.270
AGOSTO	0.199	0.967
SEPTIEMBRE	0.863	0.811
OCTUBRE	1.566	0.348
NOVIEMBRE	1.160	0.215
DICIEMBRE	2.824	0.084
PROMEDIO	2.070	0.477

Fuente: Propia

Figura 55

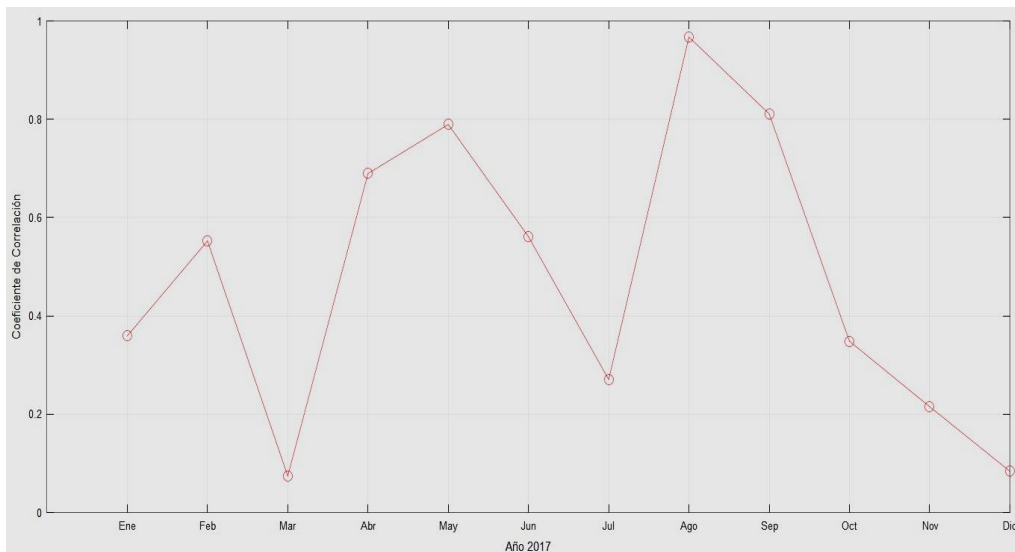
RMSE del periodo enero-diciembre 2017 para el modelo LSTM



Fuente: Propia

Figura 56

Coeficiente de Correlación del periodo enero-diciembre 2017 para el modelo LSTM

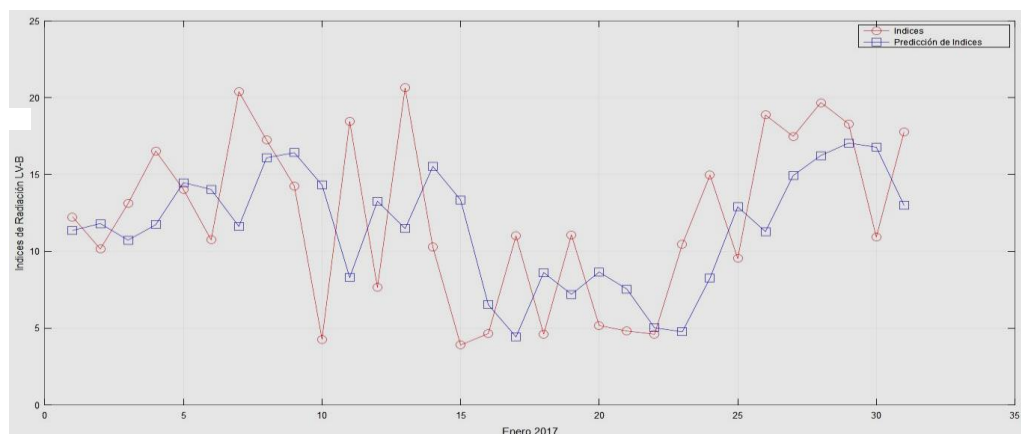


Fuente: Propia

Finalmente, se presenta los resultados de predicción los índices de prueba e índices de predicción para los días del mes de enero del año 2017, a partir de datos del Ministerio de Salud – Arequipa, que se muestra en la tabla 16 y se visualiza en la figura 57; y la dispersión de los índices de radiación ultravioleta del año 2017, en la figura 58.

Figura 57

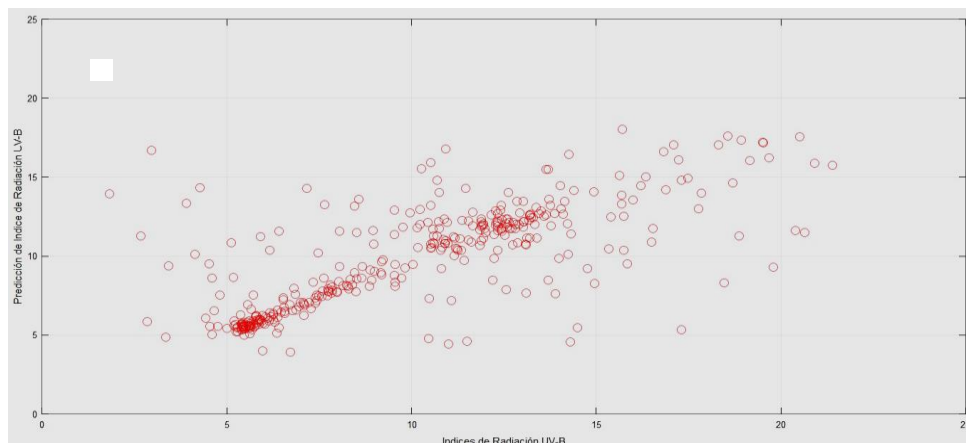
Predicción de Índices de Radiación Ultravioleta enero 2017 para el modelo LSTM



Fuente: Propia

Figura 58

Dispersión de los Índice de Radiación Ultravioleta del año 2017 para el modelo LSTM



Fuente: Propia

Tabla 16

Índice de Radiación Ultravioleta y Predicción para el mes enero 2017 para el modelo LSTM

DÍA	Índice UV	Predicción	DÍA	Índice UV	Predicción
1	12.240	11.363	17	11.002	4.431
2	10.152	11.809	18	4.608	8.619
3	13.104	10.728	19	11.074	7.204
4	16.531	11.731	20	5.170	8.649
5	14.026	14.475	21	4.810	7.543
6	10.742	14.026	22	4.608	5.020
7	20.390	11.623	23	10.454	4.761
8	17.237	16.090	24	14.962	8.258
9	14.256	16.422	25	9.533	12.893
10	4.277	14.318	26	18.878	11.285
11	18.461	8.291	27	17.482	14.930
12	7.646	13.269	28	19.685	16.241
13	20.635	11.481	29	18.302	17.047
14	10.267	15.525	30	10.930	16.781
15	3.902	13.349	31	17.770	12.995
16	4.651	6.523			

Fuente: Propia

5.9 COMPARACIÓN DE MÉTRICAS DE MODELOS ARMA, RBC, LSTM

En la presente sección, se muestra la tabla 17 de resumen de resultados % RMSE, RMSE, MAPE, MAE, precisión y coeficiente de correlación para el año 2017.

Tabla 17

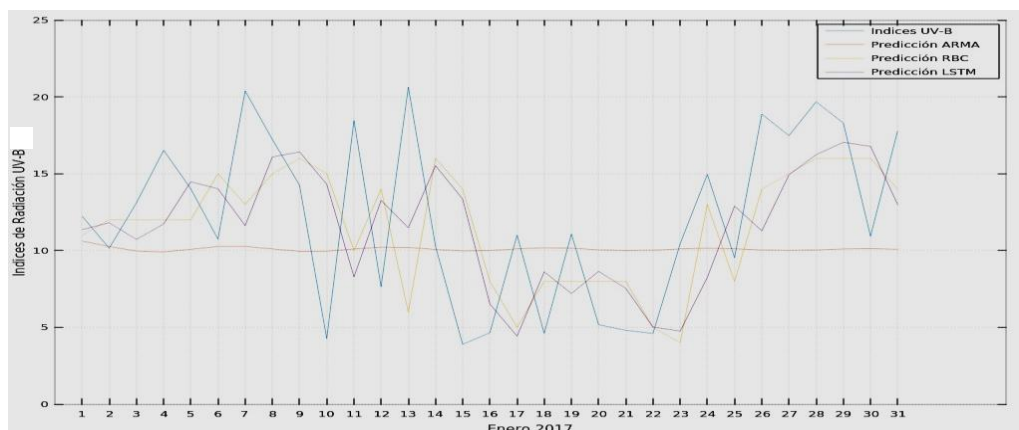
Métricas para los modelos ARMA, RBC, LSTM del año 2017

	ARMA	RBC	LSTM
%RMSE	40.38	31.71	28.01
RMSE	4.05	3.02	2.81
MAPE	34.26	18.9	16.06
MAE	3.44	1.8	1.61
PRECISIÓN	65.74	68.29	83.94
COEF. DE CORRELACIÓN	0.02	0.69	0.73

Fuente: Propia

Figura 59

Índice de Radiación Ultravioleta y predicciones del año 2017 para el modelo ARMA, RBC, LSTM



Fuente: Propia

Tabla 18

Tabla Comparativa del Índice de Radiación Ultravioleta y predicción para los modelos ARMA, RBC, LSTM

DÍA	Índice UV	ARMA	RBC	LSTM
1	12.24	10.62	11	11.36
2	10.15	10.27	12	11.81
3	13.10	9.97	12	10.73
4	16.53	9.91	12	11.73
5	14.03	10.08	12	14.48
6	10.74	10.26	15	14.03
7	20.39	10.27	13	11.62
8	17.24	10.11	15	16.09
9	14.26	9.95	16	16.42
10	4.28	9.96	15	14.32
11	18.46	10.10	10	8.29
12	7.65	10.22	14	13.27
13	20.64	10.20	6	11.48
14	10.26	10.07	16	15.53
15	3.90	9.97	14	13.35
16	4.65	10	8	6.52
17	11.00	10.11	5	4.43
18	4.61	10.19	8	8.62
19	11.07	10.15	8	7.20
20	5.17	10.05	8	8.65
21	4.81	9.99	8	7.54
22	4.61	10.03	5	5.02
23	10.45	10.11	4	4.76
24	14.96	10.15	13	8.26
25	9.53	10.12	8	12.89
26	18.88	10.04	14	11.29
27	17.48	10.01	15	14.93
28	19.69	10.04	16	16.24
29	18.30	10.10	16	17.05
30	10.93	10.12	16	16.78
31	17.77	10.09	14	12.99

Fuente: Propia

CAPÍTULO VI

ANÁLISIS Y DISCUSIÓN

Se ha dividido los datos en dos partes: una de entrenamiento y otra de prueba. El entrenamiento se realiza con los datos de los años 2010 hasta 2016 y las pruebas de predicción se realizan con los datos del año 2017 que representa aproximadamente un 10 % de los datos obtenidos del ministerio de salud (MINSa).

La metodología Autoregresión de Medias Móviles (ARMA), también es utilizada en investigación realizada por (Lahura et al, 2003), en el trabajo de investigación titulado ANÁLISIS DE SERIES DE TIEMPO UNIVARIANTE APLICANDO METODOLOGÍA DE BOX-JENKINS PARA LA PREDICCIÓN DE OZONO EN LA CIUDAD DE CALI, COLOMBIA”; los siguientes resultados de coeficiente de determinación $R^2 = 0,7242$ que corresponde a un coeficiente de correlación de $R = 0,8508$ aproximadamente, correspondiente a 104 días durante el periodo abril a julio del año 2003.

El modelo ARMA implementado en la presente investigación, destaca el coeficiente de correlación $R = 0,022$ para el año 2017, de 365 datos registrados por el ministerio de salud (MINSa).

La metodología de Razonamiento Basado en Casos (RBC), también es utilizada en investigación realizada por (Koo et al, 2019) en la tesis denominada “APROXIMACIÓN A LA ESTIMACIÓN PARA EL POTENCIAL DE RADIACIÓN SOLAR CON COMPLEJOS PATRONES ESPACIALES VÍA TÉCNICAS DE MACHINE LEARNING”. En el trabajo de investigación se implementan los modelos Razonamiento Basado en Casos (RBC) con 90.351 % de Precisión, Redes Neuronales (ANN) con 91.928 % de Precisión y Análisis de Regresión Múltiple (MRA), con 91.143 % de precisión que corresponde a un % RMSE de

9.65 %, 8.072 % y 8.86 %, respectivamente. En el presente trabajo de investigación el modelo RBC se tiene un % RMSE de 32 % y coeficiente de correlación de $R = 0,695$.

La metodología de Redes Neuronales con Memoria de Largo y Corto Plazo (LSTM), se emplea en la investigación realizada por (Paoli et al, 2010) en el trabajo de investigación titulado "*Predicción de Series de Tiempo De Radiación Solar Diaria Pre-procesada Utilizando Redes Neuronales*"; los siguientes resultados de % RMSE 21 % y un coeficiente de determinación $R^2 = 0,99$ que corresponde a un coeficiente de correlación de $R = 0,995$. El modelo LSTM implementado en el presente trabajo de obtiene el RMSE de 2812 que corresponde a un % RMSE de 28 % y coeficiente de correlación de $R = 0,728$.

Finalmente, se analiza el comparativo de la predicción para los tres modelos en estudio, que el modelo LSTM presenta el menor % RMSE de 28 % y el modelo ARMA presenta % RMSE de 40 % y coeficiente de correlación $R = 0,022$ del año 2017 para los índices de radiación ultravioleta de la ciudad de Arequipa que se muestra en la tabla 17.

El modelo RBC tiene resultados intermedios con % RMSE de 32 % para y Coeficiente de Correlación de los datos observados de $R = 0,695$ en la predicción de los Índices de Radiación Ultravioleta para el año 2017.

CONCLUSIONES

1. Considerando las herramientas informáticas y los datos meteorológicos disponibles y aplicando las tres técnicas de minería de datos; ha sido posible proponer un modelo de predicción a partir de la minería de datos basado en casos, vinculado al estudio de la radiación ultravioleta UVB en la ciudad de Arequipa para el año 2017.
2. El modelo Autoregresión de Medias Móviles (ARMA) tiene % RMSE de 40 %, es decir, la menor precisión de los tres modelos comparados, sin embargo, es de muy fácil implementación y costo computacional, se adapta fácilmente a los tipos de datos y, sin embargo, se sacrifican los resultados con baja precisión que dependen principalmente de los parámetros ($p = 4$, $q = 4$).
3. Ha sido posible desarrollar un modelo de predicción mediante un Modelo de Razonamiento Basado en Casos (RBC) con el indicador Raíz de Error Medio Cuadrático Porcentual (%RMSE) de 32 % y 68 % de precisión para la predicción de índices de radiación ultravioleta (I-UV) del año 2017. El resultado depende de la cantidad de casos disponibles para su comparación. De los datos solicitados al MINSA, se tienen desde el año 2010 al 2017, sin embargo, con mayor cantidad de años la precisión podría mejorar; De otra parte, es importante mencionar que también depende del número de Rasgos Predictores ($r_p = 4$) y Vecinos Cercanos que para el presente estudio se determinó en $k = 21$.
4. Según los resultados mostrados en el capítulo anterior, se verifica que el modelo de Redes Neuronales con Memoria de Largo y Corto Plazo (LSTM) tiene un % RMSE de 28 %, es decir, la mejor precisión de predicción con 72 % que ha sido posible gracias a la facilidad de

implementar algoritmos de redes neuronales bajo lenguaje Python. Las variables principales fueron el número de células LSTM ($n = 5$) así como también el número de épocas (épocas = 10).

5. En análisis de los tres modelos del presente estudio, se tiene que el mejor modelo mediante su indicador %, RMSE es el modelo LSTM con 28 %, en segundo lugar, está el modelo RBC con 32 % y en último lugar, el modelo ARMA con 40 %.

RECOMENDACIONES

1. Sería recomendable implementar la minería de datos con el modelo Razonamiento Basado en Casos (RBC), en la predicción de otras variables meteorológicas como las precipitaciones, las temperaturas o la humedad relativa, relacionados al estudio del fenómeno del niño.
2. El modelo ARMA presenta deficiencias en la predicción del índice de radiación ultravioleta debido a la alta volatilidad de sus valores.
3. El modelo Razonamiento Basado en Casos puede ser replicado de manera muy accesible y bajo costo, sin embargo, es importante la cantidad necesaria de data para su óptimo funcionamiento.
4. El modelo LSTM de Redes Neuronales puede ser un complemento al modelo Razonamiento Basado en Casos, debido a que presenta mejor precisión en resultados.
5. Los modelos que han tenido mucha demanda en el ámbito nacional son, el modelo ARMA y RBC; sin embargo, ante el avance tecnológico del país y el abaratamiento del costo computacional en la nube, es posible implementar el modelo LSTM.

REFERENCIAS BIBLIOGRÁFICAS

- Ayres, J R d C M (2008) Para comprender el sentido práctico de las acciones de salud: contribuciones de la hermenéutica filosófica *Salud colectiva*, 4:159–172.
- Bitri, M, Grazhdani, S, and Ahmeti, A (2014) Validation of the aquacrop model for full and deficit irrigated potato production in environmental condition of korç,a zone, south-eastern Albania.
- Cabrera-Silva, S (2005) *Radiación ultravioleta y salud* Editorial Universitaria.
- Caldwell, M M, Bjorn, L O, Bornman, J F, Flint, S D, Kulandaivelu, G, Teramura, A H, and Tevini, M (1998) Effects of increased solar ultraviolet radiation on terrestrial ecosystems *Journal of Photochemistry and Photobiology B: Biology*, 46(1-3):40–52.
- Chubachi, S et al (1984) Preliminary result of ozone observations at syowa station from february 1982 to january 1983.
- De Amat-Loza, F (2014) Arequipa y la radiación Ultravioleta *dermatología peruana*, 24(1):5.
- De Arce, R and Mahia, R (2003) Modelos Arima *Programa CITUS: Técnicas de Variables Financieras*.
- Díaz Avendaño, A A (2016) Técnicas de Minería de datos para predicción del diagnóstico de hipertensión arterial.

- Díaz Tello, S (2014) Sistema integral bajo el enfoque de minería de datos y redes neuronales para la predicción y control de la contaminación atmosférica por pm10 en la ciudad de Chimbote.
- Duffie, J A, B W A (2013) *Solar engineering of thermal processes* John Wiley Sons.
- Garg, A, Huang, H, Kushvaha, V, Madhushri, P, Kamchoom, V, Wani, I, Koshy, N, and Zhu, H-H (2020) Mechanism of biochar soil pore–gas–water interaction: gas properties of biochar-amended sandy soil at different degrees of compaction using knn modeling *Acta Geophysica*, 68(1):207–217.
- Gutiérrez, E, Chía, H, Galarza, C, Ramos, W, Mendoza, M, Macetas, R, Gómez, A, Tello, M, Ronceros, G, Uribe, M, et al (2009) Enfermedades dermatológicas en un servicio académico asistencial universitario de lima-Perú, 2004-2007 *Dermatol Perú*, pages 306–313.
- Guridi Mateos, G. (2017). Modelos de redes neuronales recurrentes en clasificación de patentes.
- Heisler, G M, Grant, R H, and Gao, W (2003) Ultraviolet radiation and its impacts on agriculture and forests *Agricultural and Forest Meteorology* 120: 3-7, 120.
- Índice, U (2003) solar mundial: guía práctica recomendación conjunta de: Organización mundial de la salud, organización meteorológica mundial, programa de las naciones unidas para el medio ambiente, comisión internacional de protección contra la radiación no ionizante.

- Jiménez Builes, J A (2006) Un modelo de planificación instruccional usando razonamiento basado en casos en sistemas multi-agente para entornos integrados de sistemas tutoriales inteligentes y ambientes colaborativos de aprendizaje *Escuela de Sistemas*.
- Kakani, V, Reddy, K, Zhao, D, and Sailaja, K (2003) Field crop responses to ultraviolet-b radiation: a review *Agricultural and forest meteorology*, 120(1-4):191–218.
- Koo, C, Li, W, Cha, S H, and Zhang, S (2019) A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine-learning techniques *Renewable energy*, 133:575–592.
- Lahura, E et al (2003) *El coeficiente de correlación y correlaciones espúreas*, volume 218 Pontificia Universidad Católica del Perú, Departamento de Economía.
- Longstreth, J, De Gruijl, F, Kripke, M, Abseck, S, Arnold, F, Slaper, H, Velders, G, Takizawa, Y, and Van der Leun, J (1998) Health risks *Journal of Photochemistry and Photobiology B: Biology*, 46(1-3):20–39.
- López, C P (2007) *Minería de datos: técnicas y herramientas* Editorial Paraninfo.
- Machekposhti, K H, Sedghi, H, Telvari, A, and Babazadeh, H (2018) Modeling climate variables of rivers basin using time series analysis (case study: Karkheh river basin at iran) *Civ Eng J*, 4:78.
- Mainetti, J A (1990) Bioética: una nueva filosofía de la salud *Boletín de la Oficina Sanitaria Panamericana (OSP)*; 108 (5-6), mayo-jun 1990.

- Naupas Caraza, C M (2016) Minería de datos aplicada a la detección de fraude electrónico en entidades bancarias.
- Østerlind, A, Tucker, M, Stone, B, and Jensen, O (1988) The danish case-control study of cutaneous malignant melanoma ii importance of uv-light exposure *International Journal of Cancer*, 42(3):319–324.
- Paoli, C, Voyant, C, Muselli, M, and Nivet, M-L (2010) Forecasting of preprocessed daily solar radiation time series using neural networks *Solar Energy*, 84(12):2146–2160.
- Pontius, R G, Thontteh, O, and Chen, H (2008) Components of information for multiple resolution comparison between maps that share a real variable *Environmental and Ecological Statistics*, 15(2):111–142.
- Ramos, W (2012) Vigilancia de cáncer al I semestre 2012 Ministerio de salud de Perú. Dirección general de epidemiología, red nacional de epidemiología. *Boletín Epidemiológico (Lima)*, 21(27):435–7.
- Riquelme Santos, J C, Ruiz, R, and Gilbert, K (2006) Minería de datos: Conceptos y tendencias *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, 10 (29), 11-18.
- Rodríguez, A M (2008) Filosofía y ciencia *Revista de Filosofía de la Universidad de Costa Rica*.
- Rodríguez, S. C., Rodríguez, E. F. G., Hernández, H. C., Carvajal, Y. L., Cardoso, G. C., & Martínez, I. G. (2011). Sistema experto basado en casos para el diagnóstico de la hipertensión arterial. *Revista Facultad de Ingeniería Universidad de Antioquia*, (60), 202-213.

Rosso Fernández, C (2014) Comportamientos de protección solar y percepción de riesgo de cáncer de piel en la población española.

Sordo, C and Gutiérrez, C (2013) Cáncer de piel y radiación solar: experiencia peruana en la prevención y detección temprana del cáncer de piel y melanoma *Revista Peruana de Medicina Experimental y Salud Publica*, 30:113–117.

Sulla Torres, J (2015) *Aplicación de técnicas supervisadas de minería de datos para determinar la predicción de deserción académica* Tesis, UNIVERSIDAD CATÓLICA DE SANTA MARÍA.

Willmott, C J and Matsuura, K (2005) Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance *Climate research*, 30(1):79–82.

Zhang, Y Mlstudio: Machine learning made easy

ANEXOS

ANEXO 1: MATRIZ DE CONSISTENCIA:

MODELO DE PREDICCIÓN A PARTIR DE LA MINERÍA DE DATOS BASADO EN CASOS, VINCULADAS AL ESTUDIO DE LA RADIACIÓN ULTRAVIOLETA UV-B AREQUIPA 2017			
PROBLEMA GENERAL	OBJETIVO GENERAL	HIPOTESIS	VARIABLES
<p>¿Será posible desarrollar un Modelo de Predicción a partir de la Minería de Datos Basado en Casos vinculado al estudio de la radiación ultravioleta UV-B en la ciudad de Arequipa, 2017?</p>	<p>Proponer un modelo de alta precisión en la predicción de los índices de radiación ultravioleta en la ciudad de Arequipa para el año 2017.</p>	<p>El modelo Razonamiento Basado en Casos (RBC) es el modelo con menor error en la predicción de los índices de radiación ultravioleta en la ciudad de Arequipa para el año 2017.</p>	<p>TIPO DE INVESTIGACION: Aplicada</p> <p>NIVEL DE INVESTIGACION Predictivo</p>
	<p>OBJETIVOS ESPECÍFICOS:</p> <ul style="list-style-type: none"> ◦ Determinar el error en la Predicción del modelo lineal ARMA ◦ Determinar el error en la Predicción del modelo Razonamiento Basado en Casos (RBC). ◦ Determinar el error en la Predicción del Modelo Memoria de Largo y Corto Plazo (LSTM). ◦ Realizar el análisis comparativo de los Modelos .ARMA, RBC, LSTM. 		<p>INDICADORES:</p> <ul style="list-style-type: none"> • Desempeño en la predicción del modelo RBC • Desempeño en la predicción del modelo ARMA • Desempeño en la predicción del modelo LSTM. <p>MÉTRICAS:</p> <ul style="list-style-type: none"> • %RMSE • RMSE • MAPE • MAE • Precisión • Coeficiente de Correlación
			<p>DISEÑO DE INVESTIGACION Cuasi-Experimental</p> <p>POBLACION Y MUESTRA Índices de radiación ultravioleta 2010-2017</p> <p>TRATAMIENTO DE DATOS Técnicas Matemáticas</p>

ANEXO 2: Modelo Autoregresivo de Medias Móviles (ARMA)

Código Python:

```
import numpy as np
import pandas as pd
from pandas import read_csv
from matplotlib import pyplot as plt
from sklearn.metrics import mean_squared_error
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
import math
from datetime import datetime

def difference(dataset, interval=1):
    diff = list()
    for i in range(interval, len(dataset)):
        value = dataset[i] - dataset[i - interval]
        diff.append(value)
    return np.array(diff)

dataframe = pd.read_csv('/content/gdrive/My Drive/Dm/2010_2017_v1.csv', usecols =
[2])
dataset = dataframe.values
```

```

dataset = dataset.astype('float32')
print(datetime. now())
# Original Series
fig, (axes1,axes2,axes3) = plt.subplots(3, 1, sharex=True)
axes1.plot(dataset); axes1.set_title('Original Series')
axes2.plot(difference(dataset)); axes2.set_title('1st Order Differencing')
axes3.plot(difference(difference(dataset))); axes3.set_title('2nd Order Differencing')

fig, (axes1, axes2, axes3, axes4) = plt.subplots(4, 1, sharex=True)
fig, plot_acf(dataset ,lags=25, ax=axes1)
fig, plot_pacf(dataset,lags=25, ax=axes2)
fig, plot_acf(difference(dataset),lags=25, ax=axes3)
fig, plot_acf(difference(difference(dataset)),lags=25, ax=axes4)

result = adfuller(dataframe['in'])
print('ADF Statistic: {}'.format(result[0]))
print('p-value: {}'.format(result[1]))
print('Critical Values:')
for key, value in result[4].items():
    print("\t{}: {}".format(key, value))

t=365
train = dataset[:len(dataset)-t]
test = dataset[len(dataset)-t:]

import itertools
p=d=q=range(0,5)
pdq = list(itertools.product(p,d,q))

parametros = []
model_arima_aic = []

```

```

rRMSE_arima=[]

import warnings
warnings.filterwarnings('ignore')
for param in pdq:
    try:
        model_arima = ARIMA(train,order=param)
        model_arima_fit = model_arima.fit()
        # Forecast
        yp= model_arima_fit.forecast(steps=t)[0]
        rRMSE = math.sqrt(mean_squared_error(test, yp))
        parametros.append(param)
        model_arima_aic.append(model_arima_fit.aic)
        rRMSE_arima.append(rRMSE)
        print(param,model_arima_fit.aic,rRMSE)
    except:
        continue

plt.plot(model_arima_aic)

future_forecast = stepwise_model.predict(n_periods = t)
plt.plot(future_forecast[:31])
plt.plot(test[:31])

import csv
myFile = open('/content/gdrive/My Drive/Dm/parametros_arima.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],parametros))

print("Writing complete")

```

```

print(datetime. now())

import csv
myFile = open('/content/gdrive/My Drive/Dm/aic_arima.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],model_arima_aic))

print("Writing complete")
print(datetime. now())

model_arima = ARIMA(train,order=(4,0,4))
model_arima_fit = model_arima.fit()
print(model_arima_fit.summary())

yp= model_arima_fit.forecast(steps=t)[0]
rRMSE = math.sqrt(mean_squared_error(test, yp))
ypp = yp.reshape(-1,1)

plt.plot(test[:15])
plt.plot(yp[:15],color='red')

import csv
myFile = open('/content/gdrive/My Drive/Dm/yp_arima2017.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],yp))

print("Writing complete")

```

```
print(datetime. now())
```

```
import csv
```

```
myFile = open('/content/gdrive/My Drive/Dm/y_arima2017.csv', 'w')
```

```
with myFile:
```

```
    writer = csv.writer(myFile)
```

```
    writer.writerows(map(lambda x: [x],test))
```

```
print("Writing complete")
```

```
print(datetime. now())
```

```
\end{verbatim}
```

ANEXO 3: Modelo Razonamiento Basado en Casos (RBC)

Código Python:

```
import math
import numpy as np
import pandas as pd
from pandas import read_csv
import matplotlib.pyplot as plt
import sklearn
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split

dataframe = read_csv('/content/gdrive/My Drive/Dm/2010_2017_v1.csv', usecols = [2] )
dataset = dataframe.values
dataset = dataset.astype('float32')
dataset = np.append(dataset,[[1]],axis=0)

def create_dataset(dataset, look_back=1):
    dataX, dataY = [], []
    for i in range(len(dataset) - look_back - 1):
        a = dataset[i:(i + look_back), 0]
        dataX.append(a)
        dataY.append(dataset[i + look_back, 0])
    return np.array(dataX), np.array(dataY)

look_back = 2

t=365+3
```

```

train = dataset[:len(dataset)-t]
test = dataset[len(dataset)-t:]

X_train, y_train = create_dataset(train, look_back)
X_test, y_test = create_dataset(test, look_back)
y_train = y_train.astype('int32')
y_test = y_test.astype('int32')

k_range = range(1, 25)
rRMSE = []

for k in k_range:
    knn = KNeighborsClassifier(n_neighbors = k)
    knn.fit(X_train, y_train)
    y_pred = knn.predict(X_test)
    rm = math.sqrt(mean_squared_error(y_pred, y_test))
    rRMSE.append(rm)
plt.plot(k_range, rRMSE, markersize = 10)
plt.xlabel('Value of K for KNN')
plt.ylabel('rmse')
rRMSE = np.array(rRMSE)
rRMSE = rRMSE.astype('float32')

import csv
myFile = open('/content/gdrive/My Drive/Dm/rRMSE_KNN.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],rRMSE))

print("Writing complete")

```

```
knn = KNeighborsClassifier(n_neighbors=21)
knn.fit(X_train, y_train)
yp = knn.predict(X_test)

import csv
myFile = open('/content/gdrive/My Drive/Dm/yp_KNN2017.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],yp))

print("Writing complete")

import csv
myFile = open('/content/gdrive/My Drive/Dm/y_KNN2017.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],y_test))

print("Writing complete")
```

ANEXO 4: Modelo Redes Neuronales con Memoria de Largo y Corto Plazo (LSTM)

Codigo Python:

```
import math
import numpy
from pandas import read_csv
import matplotlib.pyplot as plt
from keras.layers import Dense
from keras.layers import LSTM
from keras.models import Sequential
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error

def create_dataset(dataset, look_back=1):
    dataX, dataY = [], []
    for i in range(len(dataset)-look_back-1):
        a = dataset[i:(i+look_back), 0]
        dataX.append(a)
        dataY.append(dataset[i + look_back, 0])
    return numpy.array(dataX), numpy.array(dataY)

numpy.random.seed(7)
dataframe = read_csv('/content/gdrive/My Drive/Dm/2010_2017_v1.csv', usecols = [2] )
dataset = dataframe.values
dataset = dataset.astype('float32')
dataset = numpy.append(dataset,[[1]],axis=0)

scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
```

```

t=365+3
train = dataset[:len(dataset)-t]
test = dataset[len(dataset)-t:]

look_back = 2
trainX, trainY = create_dataset(train, look_back)
testX, testY = create_dataset(test, look_back)

trainX = numpy.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = numpy.reshape(testX, (testX.shape[0], 1, testX.shape[1]))

testY = scaler.inverse_transform([testY])

k_range = range(1, 25)
rRMSE_lstm = []

for k in k_range:
    model = Sequential()
    model.add(LSTM(k, input_shape=(1, look_back)))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer='adam')
    model.fit(trainX, trainY, epochs=5, batch_size=1, verbose=2)

    testPredict = model.predict(testX)
    testPredict = scaler.inverse_transform(testPredict)
    rRMSE = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
    rRMSE_lstm.append(rRMSE)

plt.plot(k_range, rRMSE_lstm, markersize = 10)

import csv

```

```

myFile = open('/content/gdrive/My Drive/Dm/rRMSE_lstm.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],rRMSE_lstm))

```

```
print("Writing complete")
```

```

model = Sequential()
model.add(LSTM(5, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=10, batch_size=1, verbose=2)

```

```

testPredict = model.predict(testX)
testPredict = scaler.inverse_transform(testPredict)
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
ypp = testY.reshape(-1,1)
plt.plot(testPredict)
plt.plot(ypp)

```

```
print("Test Score: %.2f RMSE" % (testScore))
```

```

import csv
myFile = open('/content/gdrive/My Drive/Dm/y_LSTM2017.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],ypp))

```

```
print("Writing complete")
```

```
import csv
```

```
myFile = open('/content/gdrive/My Drive/Dm/yp_LSTM2017.csv', 'w')
with myFile:
    writer = csv.writer(myFile)
    writer.writerows(map(lambda x: [x],testPredict))

print("Writing complete")
```