

UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE INGENIERÍA AMBIENTAL

Trabajo de titulación previo a la obtención del título de:

INGENIERO AMBIENTAL

TRABAJO EXPERIMENTAL:

MODELAMIENTO DE LA PRECIPITACIÓN EN LA ZONA URBANA DE

LA CIUDAD DE CUENCA.

AUTORES:

MARÍA EUGENIA QUINTEROS ESPINOZA

OMAR EMMANUEL RAMIREZ PARRA

TUTOR:

DIEGO ROMAN CABRERA MENDIETA, Ph. D

Cuenca-Ecuador

2018

CESIÓN DE DERECHOS DE AUTOR

Nosotros, María Eugenia Quinteros Espinoza, con documento de identificación N° 0302087689 y Omar Emmanuel Ramírez Parra con documento de identificación N° 0302332465, manifestamos nuestra voluntad y cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales, en virtud de que somos autores del Trabajo de Titulación denominado: “MODELAMIENTO DE LA PRECIPITACIÓN EN LA ZONA URBANA DE LA CIUDAD DE CUENCA” mismo que ha sido desarrollado para optar por el título de Ingeniero Ambiental en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación al determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, julio de 2018.

María Eugenia Quinteros Espinoza

0302087689

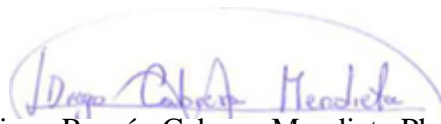
Omar Emmanuel Ramírez Parra

0302332465

CERTIFICACIÓN

Yo declaro que bajo mi tutoría fue desarrollado el trabajo de titulación:
“MODELAMIENTO DE LA PRECIPITACIÓN EN LA ZONA URBANA DE LA
CIUDAD DE CUENCA” realizado por los autores María Eugenia Quinteros Espinoza y
Omar Emmanuel Ramírez Parra, obteniendo el Trabajo Experimental que cumple con
todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, julio de 2018.



Diego Román Cabrera Mendieta Ph.D.

0104028303

DECLARATORIA DE RESPONSABILIDAD

Nosotros, María Eugenia Quinteros Espinoza, con documento de identificación N° 0302087689 y Omar Emmanuel Ramírez Parra con documento de identificación N° 0302332465, autores del trabajo de Titulación “MODELAMIENTO DE LA PRECIPITACIÓN EN LA ZONA URBANA DE LA CIUDAD DE CUENCA” certificamos que el total contenido de esta investigación es de nuestra exclusiva responsabilidad y autoría



María Eugenia Quinteros Espinoza

0302087689



Omar Emmanuel Ramírez Parra

0302332465

DEDICATORIA

El presente trabajo le dedico primeramente a Dios por la fuerza y sabiduría brindada durante esta etapa de mi vida, a mi madre Rosita y mi padre Omar por su apoyo constante, de igual manera a mis abuelos Rhino y Bertha que han sido mi pilar fundamental para seguir adelante, los cuales me han llenado e inculcado de valores, lo que ha sido importante para cumplir esta meta propuesta. A mi hermano Josué que han sido motivación para seguir adelante; a mis amigos y demás familiares que han sido importantes en toda mi etapa de vida.

Omar Emmanuel

El presente trabajo le dedico a Dios por la fortaleza y sabiduría brindada para cumplir esta meta en mi vida, a mi padre Olmedo por su apoyo incondicional, inculcándome experiencias y valores los cuales han sido importantes para llegar a cumplir esta meta propuesta. A mis hermanas Sofía y Judith que han sido mi pilar y principal motivación para seguir adelante; a mis abuelitos Dilio y Libia, mi tía Margoth que han sido mi ejemplo y han guiado mi camino desde niña, proporcionándome experiencias y cariño; a mis amigos y demás familiares que han sido muy importantes en esta etapa de mi vida compartiendo buenos y malos momentos.

María Eugenia

AGRADECIMIENTO

Queremos expresar nuestro más sincero agradecimiento a nuestro tutor Diego Cabrera Mendieta Ph.D, quien nos brindó y compartió experiencias y conocimientos adquiridos durante su vida profesional

A todos los miembros del Grupo de Investigación y Desarrollo en Tecnologías Industriales (GIDTEC) por el apoyo y ayuda brindada para realizar la presente investigación.

A los docentes de la Carrera de Ingeniería Ambiental por sus aportes académicos importantes para nuestra formación profesional.

A ETAPA EP, a través del Ing. Mario Gualpa, quien nos colaboró con información y experiencia para poder ejecutar la presente investigación.

RESUMEN

De este proyecto de investigación radica en la necesidad de conocer con anticipación los fenómenos hidrológicos extremos que se pueden dar en la ciudad de Cuenca, los cuales pueden afectar de manera directa o indirecta el bienestar de la ciudadanía. Mediante la generación de un modelo de pronóstico de precipitación, se pueden tomar decisiones para reducir recursos destinados para la mitigación de los daños y poder realizar un manejo adecuado del agua para la satisfacción de la demanda del mismo recurso dentro de la ciudad.

El objetivo principal es desarrollar un modelo para el pronóstico de precipitación a partir de datos meteorológicos del período mayo de 2014 hasta junio de 2017 en la zona urbana de Cuenca.

El desarrollo del modelo de pronóstico se realizó en 7 etapas: I. Delimitación del área, II. Obtención de los datos, III. Preprocesamiento, IV. Análisis estadístico, V. Modelamiento, VI. Evaluación, VII. Presentación en un SIG.

Los datos se preprocesaron en 3 etapas. El análisis estadístico evaluó el comportamiento de las variables meteorológicas. Para el entrenamiento del modelo se escogió las redes neuronales recurrentes de tipo LSTM. En la evaluación del modelo se usaron métricas de rendimiento (matriz de confusión, recall, precisión, f1-score, accuracy). Finalmente, las probabilidades obtenidas desde cada uno de los modelos fueron publicadas por un servidor socket desarrollado dentro de Quantum GIS con la herramienta Python. La mayor cantidad de horas que se logró pronosticar sin rebasar la frontera de 50% de exactitud de un clasificador trivial fue de 14 horas, y la menor cantidad de horas fue de 3 horas. Concluyendo que las RNN de tipo LSTM son eficientes para aplicaciones de pronóstico de variables meteorológicas, trabajando con series de tiempo.

ABSTRACT

This research project lies in the need to know in advance the extreme hydrological phenomena that can occur in the city of Cuenca, which can directly or indirectly affect the well-being of citizens. Through the generation of a precipitation forecasting model, decisions can be made to reduce resources destined for the mitigation of damages and to be able to carry out an adequate water management to satisfy the demand of the same resource within the city.

The main objective is to develop a model for the precipitation forecast based on meteorological data from May 2014 to June 2017 in the urban area of Cuenca.

The development of the forecast model was carried out in 7 stages: I. Delimitation of the area, II. Obtaining the data, III. Preprocessing, IV. Statistical analysis, V. Modeling, VI. Evaluation, VII. Presentation in a GIS.

The data was preprocessed in 3 stages. The statistical analysis evaluated the behavior of the meteorological variables. For the training of the model, recurrent neural networks of LSTM type were chosen. In the evaluation of the model, performance metrics (confusion matrix, recall, precision, f1-score, accuracy) were used. Finally, the probabilities obtained from each of the models were published by a socket server developed within Quantum GIS with the Python tool. The largest number of hours that could be predicted without exceeding the border of 50% accuracy of a trivial classifier was 14 hours, and the lowest number of hours was 3 hours. Concluding that the LSTM type RNNs are efficient for forecasting meteorological variables, working with time series.

INDICE GENERAL

1. INTRODUCCIÓN	1
1.1. Antecedentes	1
1.2. Problema	3
1.3. Objetivos	5
1.3.1. Objetivo General	5
1.3.2. Objetivo Especifico	5
2. FUNDAMENTO TEÓRICO	6
2.1. Meteorología	6
2.1.1. Atmósfera	6
2.1.2. Variables meteorológicas	7
- Precipitación	7
- Temperatura	8
- Humedad Relativa	8
- Presión atmosférica	8
- Viento	8
2.2. Machine Learning	9
2.2.1. Conjunto de datos	9
2.2.2. Tipos de aprendizaje	10
- Aprendizaje supervisado	11
- Aprendizaje no supervisado	12
2.2.3. Evaluación de desempeño	12
1. Hold-out	12
2. Validación cruzada con K-folds	13
2.2.4. Métricas	14
1. Matriz de confusión	15
2. Precisión	16
3. Recall	16
4. F1-score	16
5. Exactitud (accuracy)	16
2.3. Modelos Neuronales	17
2.3.1. Redes Neuronales Clásicas	17
2.3.2. Redes Neuronales Recurrentes (RNN)	18
- Función de costo	19
- Función de activación	20
2.3.3. Long Short Term Memory	22

2.4.	Lenguaje de programación.....	25
2.4.1	Lenguaje Python.....	26
2.4.2	Tensorflow.....	27
2.5.	Sistemas de Información Geográfico.....	27
2.5.1	Quantum GIS.....	27
-	Interpolación.....	28
2.6.	Análisis estadístico descriptivo.....	30
2.6.1	Población y muestra	30
2.6.2	Medida de posición o localización	31
-	Media	31
2.6.3	Medidas de dispersión o variabilidad.....	31
-	Cuartiles	31
-	Varianza y desviación estándar	32
-	Curtosis y asimetría	32
3.	METODOLOGIA	33
3.1	Delimitación del área de estudio	34
3.2	Obtención de los datos	35
3.3	Preprocesamiento de los datos	37
1.	Eliminación de mediciones incompletas:.....	37
2.	Disminución de la tasa de muestreo:	38
3.	Discretización de la precipitación	39
3.4	Análisis estadístico.....	41
3.5	Creación del modelo de pronóstico.....	42
3.6	Evaluación del modelo de pronóstico.	44
3.7	Presentación del modelo en un sistema de información geográfico.....	45
4.	RESULTADOS Y ANALISIS	45
4.1	Preprocesamiento de series de tiempo.....	45
4.2	Comportamiento de las variables meteorológicas	48
1.	Precipitación	48
2.	Temperatura	71
3.	Humedad relativa.....	74
4.	Presión atmosférica	78
5.	Velocidad del viento	81
6.	Dirección del viento.....	85
4.3	Modelamiento predictivo basado en redes neuronales.	93
4.4	Validación de desempeño.....	93
4.5	Presentación del modelo en SIG	94

5. CONCLUSIONES Y RECOMENDACIONES	95
5.1. Conclusiones	95
5.2. Recomendaciones	98
5.3. Trabajos a futuro	99
6. BIBLIOGRAFIA	100
Anexo A: Parámetros de los mejores modelos de pronóstico	108
Anexo B: Métricas de rendimiento de los mejores modelos de pronóstico	110

INDICE DE GRÁFICAS

Gráfica 1 Gráfica del conjunto de datos considerando su evolución temporal.	10
Gráfica 2: Ejemplo de Hold-out aplicado a un conjunto de datos para una partición de 70% para entrenamiento, 15% para validación y 15% para prueba con estratificación.....	13
Gráfica 3: Ejemplo de validación cruzada con k-folds aplicado a un conjunto de datos para una partición de 75% para entrenamiento, 25% para validación.	14
Gráfica 4: Estructura de una red neuronal clásica.....	18
Gráfica 5: Función de activación ReLU.....	21
Gráfica 6: Esquema de la configuración de LSTM.....	23
Gráfica 7: Forget gate.....	24
Gráfica 8: Input gate.....	24
Gráfica 9: Actualización de la memoria.....	25
Gráfica 10: Salida.....	25
Gráfica 11: Esquema de la metodología del proyecto.....	34
Gráfica 12: Mapa del área de estudio.....	35
Gráfica 13: Distribución espacial de las estaciones meteorológicas.	37
Gráfica 14: Diagrama de flujo de eliminación de mediciones incompletas.....	38
Gráfica 15: Diagrama de flujo de disminución de la tasa de muestreo.....	39
Gráfica 16: Diagrama de flujo de discretización de la precipitación.....	40
Gráfica 17: Etapas del entrenamiento.....	43
Gráfica 18: Gráfica de barras de pérdida de precipitación.....	47
Gráfica 19: Gráfica de barras de precipitación acumulada en el mes de febrero de los años 2015, 2016, 2017 de las nueve estaciones de estudio.	49
Gráfica 20: Gráfica de barras de precipitación acumulada en el mes de marzo de los años 2015, 2016, 2017 de las nueve estaciones de estudio.	50
Gráfica 21: Gráfica de barras de precipitación acumulada en el mes de junio de los años 2014, 2015, 2016, 2017 de las nueve estaciones de estudio.	50
Gráfica 22: Gráfica de barras de precipitación acumulada en el mes de diciembre de los años 2014, 2015, 2016 de las nueve estaciones de estudio.	51
Gráfica 23: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2014 de las 9 estaciones.	52
Gráfica 24: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2015 de las 9 estaciones.	53
Gráfica 25: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2016 de las 9 estaciones.	54
Gráfica 26: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2017 de las 9 estaciones.	55
Gráfica 27: Gráfica de la máxima precipitación diaria del 24 de marzo del 2015.....	56
Gráfica 28: Gráfica de la máxima precipitación diaria del 10 de marzo del 2017.....	57
Gráfica 29: Gráfica de la máxima precipitación diaria del 18 de noviembre del 2015.....	57
Gráfica 30: Gráfica de la máxima precipitación diaria del 18 de noviembre del 2015.....	58
Gráfica 31: Gráfica de la máxima precipitación diaria del 3 de mayo de 2015.....	58
Gráfica 32: Gráfica de la máxima precipitación diaria del 13 de enero del 2016.....	59
Gráfica 33: Gráfica de la máxima precipitación diaria del 13 de marzo del 2016.....	59
Gráfica 34: Gráfica de la máxima precipitación diaria del 29 de marzo del 2017.....	60
Gráfica 35: Gráfico de cajas y bigotes del mes de enero correspondiente a la precipitación media diaria en los años 2015 y 2016 en las 9 estaciones de estudio.	61
Gráfica 36: Gráfico de cajas y bigotes del mes de marzo correspondiente a la precipitación media diaria en los años 2015, 2016 y 2017 en las 9 estaciones de estudio.....	62

Gráfica 37: Gráfico de cajas y bigotes del mes de mayo correspondiente a la precipitación media diaria en los años 2014, 2015, 2016 y 2017 en las 9 estaciones de estudio.	62
Gráfica 38: Gráfico de cajas y bigotes del mes de julio correspondiente a la precipitación media diaria en los años 2014, 2015 y 2016 en las 9 estaciones de estudio.	63
Gráfica 39: Histograma de precipitación de marzo del 2017 en la estación 3	64
Gráfica 40: Histograma de precipitación de abril del 2016 en la estación 4.....	64
Gráfica 41: Histograma de precipitación de agosto del 2014 en la estación 1	65
Gráfica 42: Histograma de precipitación de septiembre del 2015 en la estación 2.....	65
Gráfica 43: Comparación de medias mensuales y anual en el año 2014.....	66
Gráfica 44: Comparación de medias mensuales y anual en el año 2015.....	67
Gráfica 45: Comparación de medias mensuales y anual en el año 2016.....	69
Gráfica 46: Comparación de medias mensuales y anual en el año 2017.....	70
Gráfica 47: Medias mensuales de temperatura en el año 2014	72
Gráfica 48: Medias mensuales de temperatura en el año 2015	73
Gráfica 49: Medias mensuales de temperatura en el año 2016	73
Gráfica 50: Medias mensuales de temperatura en el año 2016	74
Gráfica 51: Medias mensuales de humedad relativa en el año 2014.....	75
Gráfica 52: Medias mensuales de humedad relativa en el año 2015.....	76
Gráfica 53: Medias mensuales de humedad relativa en el año 2016.....	77
Gráfica 54: Medias mensuales de humedad relativa en el año 2017.....	78
Gráfica 55: Medias mensuales de presión atmosférica en el año 2014.....	78
Gráfica 56: Medias mensuales de presión atmosférica en el año 2015	79
Gráfica 57: Medias mensuales de presión atmosférica en el año 2016.....	80
Gráfica 58: Medias mensuales de presión atmosférica en el año 2017	81
Gráfica 59: Medias mensuales de velocidad del viento en el año 2014.....	82
Gráfica 60: Medias mensuales de velocidad del viento en el año 2015	83
Gráfica 61: Medias mensuales de velocidad del viento en el año 2016.....	84
Gráfica 62: Medias mensuales de velocidad del viento en el año 2017.....	85
Gráfica 63: Dirección del viento en el mes de enero de la estación 7.....	86
Gráfica 64: Dirección del viento en el mes de marzo de la estación 8.....	87
Gráfica 65: Dirección del viento en el mes de febrero de la estación 9.	88
Gráfica 66: Dirección del viento en el mes de mayo de la estación 7.....	89
Gráfica 67: Dirección del viento en el mes de junio de la estación 8.	90
Gráfica 68: Dirección del viento en el mes de septiembre de la estación 9	91
Gráfica 69: Gráfica de interpolación de la media anual 2015.	94
Gráfica 70: Gráfica de interpolación de la media anual 2015.	95

INDICE DE TABLAS

Tabla 1: Representación de un conjunto de datos D	10
Tabla 2: Matriz de confusión	16
Tabla 3: Ficha de información de las estaciones meteorológicas empleadas en el modelo.	36
Tabla 4: Hiperparámetros variantes para el entrenamiento del modelo.	44
Tabla 5: Número total de datos brindados por ETAPA EP.....	46
Tabla 6: Fecha de datos completos.	48
Tabla 7: Resumen de precipitación acumulada en el año 2014.....	52
Tabla 8: Resumen de precipitación acumulada en el año 2015.....	54
Tabla 9: Resumen de precipitación acumulada en el año 2016.....	55
Tabla 10: Resumen de precipitación acumulada en el año 2017.....	56
Tabla 11: Media mensual vs media anual en el año 2014.....	66
Tabla 12: Media mensual vs media anual en el año 2015.....	68
Tabla 13: Media mensual vs media anual en el año 2016.....	70
Tabla 14: Media mensual vs media anual en el año 2017.....	71
Tabla 15: Ejemplo de prueba de parámetros en la estación 1 a 1 hora de pronóstico.....	93

1. INTRODUCCIÓN

1.1. Antecedentes

La cordillera de Los Andes es particular por su topografía debido a que sufre una caída altitudinal extrema, conocida como la Depresión de Huancabamba (Samaniego, Eguiguren, Maita, & Aguirre, 2015). Esta depresión es consecuencia de la formación de una barrera climática que tiene incidencia sobre tres de cuatro regiones naturales del Ecuador, siendo estas: la región Costa o Litoral, Andina o Sierra y Trásandina o Amazónica, por lo que tienen índices de temperatura y lluvia muy variados. De igual manera, por la ubicación geográfica (1°20'N y 5°S) del Ecuador se presentan diversos tipos de climas y microclimas, lo que genera altos niveles de biodiversidad. (Samaniego et al., 2015).

Los avances científicos en el tiempo han sido muy beneficiosos para la sociedad. Gracias al desarrollo de diversos instrumentos se puede estimar con mayor precisión varios factores climáticos que anteriormente no se lograba medir. Entre los instrumentos que se han desarrollado se encuentran barómetros, termómetros y pluviómetros, los cuales están instalados en casi todos los lugares de la superficie de la Tierra. Las mediciones recolectadas por estos instrumentos permiten realizar modelos de predicción, hacer comparaciones, buscar clasificaciones, etc.

La atmósfera es una capa gaseosa que rodea la Tierra, sin ella no existiría la vida. En este sentido es importante medir el estado de la atmósfera mediante sus variables: temperatura, presión, dirección y velocidad del viento, humedad y la precipitación (Campetella, Claudia; Cerne, 2011). El pronóstico meteorológico es un procedimiento que permite especular una situación futura a partir de condiciones actuales y pasadas.

En 1904 el noruego Bjerknes propuso que el estado del tiempo se puede pronosticar cuantitativamente a partir de ecuaciones hidrodinámicas y termodinámicas. Al final de la Segunda Guerra Mundial, se desarrolló la primera computadora electrónica, que gracias a

sus futuros avances, en 1950 se logró realizar el primer pronóstico numérico de tiempo a 24h (Campetella, Claudia; Cerne, 2011).

En los últimos años, se han realizado varios estudios en el desarrollo de pronósticos de precipitaciones utilizando varias técnicas de aprendizaje automático que generan resultados satisfactorios (Huang, Lin, Huang, & Xing, 2017). Marzaban & Stumpf, (1998) investigaron sobre una red neuronal (NN) que permita diagnosticar que circulaciones de viento detectadas por el Laboratorio Nacional de Tormentas Severas produce vientos dañinos, obteniendo como resultado la existencia o no de los mismos. Hall, Brooks, & Doswell (1999) estudiaron el mismo método para la probabilidad y pronóstico de precipitación cuantitativa verificando los pronósticos de dos años contra una red con 36 pluviómetros en Dallas, Texas. Maqsood et al. (2002) investigaron el desarrollo de una técnica de neurocomputación para pronosticar el clima pico usando datos de un año de temperatura máxima diaria y velocidad del viento en Vancouver, Canadá. French, Krajewski, & Cuykendall (1992) lograron pronosticar los campos de intensidad de lluvia en el plazo de 1 h mediante el algoritmo de retropropagación. Navone & Ceccatto (1994) lograron predecir las lluvias monzónicas de verano sobre la India mediante el uso de NN obteniendo resultados mejorados con respecto a métodos convencionales. Freiwan & Cigizoglu (2005) lograron predecir la cantidad de precipitación mensual mediante NN usando datos reales de la estación meteorológica del aeropuerto de Amman. Valverde Ramírez, De Campos Velho, & Ferreira (2005) generaron pronósticos cuantitativos específicos del sitio de precipitación diaria mediante una cartografía no lineal entre variables meteorológicas y datos pluviométricos durante el verano e invierno del periodo 1997-2002 en Sao Paulo, Brasil usando NN. Ingsrisawang, Ingsrisawang, & Somchit (2008) utilizaron la máquina de soporte vectorial, NN y el árbol de decisión para pronosticar lluvias a corto plazo desarrollando modelos de clasificación y predicción,

demostrando como se usa la selección de características para identificar la relación entre ocurrencias de lluvia y otras condiciones climáticas. Hong (2008) implementó una máquina de soporte vectorial con el objetivo de pronosticar valores de profundidad de lluvia al Norte de Taiwán.

1.2.Problema

El problema planteado en el tema de investigación radica en la necesidad de conocer con anticipación los fenómenos hidrológicos extremos (inundaciones o sequías) que se pueden dar en la ciudad de Cuenca los cuales pueden afectar de manera directa o indirecta el bienestar de la ciudadanía. Mediante la generación de un modelo de pronóstico de precipitación, se pueden tomar decisiones para reducir recursos destinados para la mitigación de los daños que pueden causar estos fenómenos y poder realizar un manejo adecuado de este recurso para la satisfacción de la demanda del mismo dentro de la ciudad.

El agua es considerada como un recurso potencial y el más importante en la tierra, al administrar este recurso racionalmente puede ser una fuente inagotable, siendo este estudio indispensable para el ser humano.

La predicción meteorológica es cada día más importante en la sociedad. Con la meteorología se puede conocer cualquier factor climático y su impacto sobre un ecosistema específico; por ejemplo, se pueden conocer los gradientes de presión, temperatura, densidad del aire, con los cuales es posible establecer la variación del clima (Castillo González, 2015). Un pronóstico meteorológico claro y confiable es fundamental para la gestión de riesgos, agricultura, industria, energías renovables, diseños arquitectónicos urbanos, gestión de transportes, salud pública y seguridad (Huang et al., 2017; Maqsood et al., 2005; National Research Council, 2012). Actualmente debido a varios factores, existen condiciones meteorológicas extremas que afectan varias actividades (productivas, industriales, agrícolas) lo que conlleva a pérdidas económicas, por la gran inversión para

contrarrestar los efectos de invierno y verano. La precipitación es una variable climática de gran importancia para los sistemas hidrológico, agrícola, industrial y energético. El entendimiento de su comportamiento temporal y espacial es de sumo interés, especialmente en los estudios de riesgos climáticos, donde la disponibilidad de información de alta resolución y de buena calidad es esencial (Gabino, 2016).

“Ecuador es un país con elevado índice de vulnerabilidad ante factores naturales y antrópicos. Desde 1990 a 2009 el país registro 65 desastres de gran magnitud; el 60% de estos fueron provocados por fenómenos hidrometeorológicos (sequías, inundaciones, deslizamientos) y el 40% por eventos geofísicos (sismos, erupciones volcánicas)” (CEPAL, 2003; SENPLADES, 2013, p. 143).

Según la Comisión Económica para América Latina y el Caribe (2003) el 33% de las pérdidas directas e indirectas (vidas humanas, infraestructura y el área productiva) en la región fueron causados por eventos naturales. El fenómeno de El Niño durante los años de 1982-1983 y 1997-1998 causó pérdidas económicas de aproximadamente USD 3900 millones (CEPAL, 2003; SENPLADES, 2013). “En 2012 la estación invernal afectó a las provincias del Azuay, El Oro, Esmeraldas, Guayas, Loja, Los Ríos, Manabí, que se declararon en estado de excepción, se vieron afectados alrededor de 195.147 hectáreas, generando costos de USD 237.9 millones” (SENPLADES, 2013, p. 143).

Según el Plan Nacional de Desarrollo Toda una Vida dentro del objetivo 2 nos habla del derecho a vivir en un ambiente sano y ecológicamente equilibrado lo cual es indispensable para el desarrollo humano (SENPLADES, 2017). Por lo cual, el estudio meteorológico de la ciudad de Cuenca puede ser un ejemplo para el desarrollo de proyectos similares en el país, lo cual ayudará a disminuir riesgos en pérdidas humanas y materiales.

La ciudad de Cuenca en los años 2010 y 2016 ha pasado por épocas de sequias fuertes en los meses de Enero y Diciembre, siendo la ciudadanía los principales afectados debido

al racionamiento del agua por parte de la empresa ETAPA (Astudillo, 2016). De igual forma causa un inconveniente en la economía de los agricultores por la falta de producción de alimentos de consumo diario, los comerciantes aumentan su costos y los productores pierden su inversión (F. Machado, 2016). Otro problema que se presenta en estas épocas son los incendios forestales perdiéndose grandes hectáreas y poniendo en riesgos a la población (J. Machado, 2016).

Concretamente, Cuenca tiene un clima privilegiado por ubicarse dentro de un extenso valle en medio de la columna andina con una ubicación geográfica $2^{\circ}54'08''S$ $79^{\circ}00'19''O$ con temperaturas que varían entre 7 a $15^{\circ}C$ en invierno y 12 a $25^{\circ}C$ en verano además en esta ciudad existen 4 ríos que son Tomebamba, Tarqui, Yanuncay y Machángara. Debido a su ubicación la ciudad puede experimentar en algunas ocasiones varios cambios de clima y temperatura durante el transcurso del día. Debido a las razones anteriores, el presente trabajo pretende realizar el análisis del comportamiento meteorológico y desarrollar un modelo de pronóstico de precipitación para la ciudad de Cuenca, que luego pueda servir de ayuda a los organismos competentes para poder establecer medidas preventivas que nos permita contrarrestar los daños que pueden causar estos factores climatológicos dentro de su zona urbana.

1.3.Objetivos

1.3.1. Objetivo General

Desarrollar un modelo para el pronóstico de precipitación a partir de datos meteorológicos del período mayo de 2014 hasta junio de 2017 en la zona urbana de Cuenca.

1.3.2. Objetivo Especifico

- I. Analizar los factores que influyen en el proceso de precipitación con la información obtenida en el estado del arte, con la finalidad de definir las variables del modelo.
- II. Preprocesar las series de tiempo mediante el lenguaje de programación Python.

- III. Evaluar los datos proporcionados por ETAPA EP, mediante un análisis estadístico, que permita estudiar el comportamiento intraanual e intradiario de las variables meteorológicas, variabilidad temporal y espacial de la intensidad de la lluvia.
- IV. Implementar modelos basados en datos mediante el uso de redes neuronales recurrentes con diferentes configuraciones para el pronóstico de precipitación entre 1-24 horas.
- V. Validar el desempeño de los modelos anteriores con datos de prueba, para la selección del mejor modelo, mediante métricas clásicas usadas en tareas de aprendizaje automático.
- VI. Presentar los resultados del modelo óptimo mediante un Sistema de Información Geográfico para su visualización.

2. FUNDAMENTO TEÓRICO

2.1.Meteorología

La Meteorología es la ciencia que estudia la atmósfera, sus propiedades y los fenómenos que se presentan dentro de ella, los cuales son denominados meteoros. El estudio de la atmósfera se basa en el análisis de variables meteorológicas, como la temperatura, la precipitación, radiación solar, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo (Rodríguez, Rosa Maria; Benito, Agueda; Portela, 2004).

2.1.1. Atmósfera

La atmósfera es la única barrera entre el universo y la tierra. En esta se distingue 5 capas: tropósfera, estratósfera, mesósfera, termosfera y exosfera, cada una a diferentes alturas y con temperaturas que definen prácticamente las distintas capas. El 78% de la atmósfera es nitrógeno, 21% es oxígeno y el 1% son constituyentes menores, esto en proporciones constantes hasta unos 100 km (Baume, 2013).

2.1.2. Variables meteorológicas

Las variables meteorológicas son parte de la atmosfera, el conjunto de variables establece una condición climática. A continuación, se describirá las siguientes variables meteorológicas.

- **Precipitación**

Es una variable meteorológica importante en el ciclo hidrológico del agua y se define como el retorno del agua de la atmósfera a la superficie terrestre. Se puede presentar de diferentes maneras, ya sea en estado líquido como lluvia o estado sólido como nieve, lo cual depende de la temperatura de las masas de aire (Sucozhañay Calle, 2015). La distribución geográfica, intensidad y frecuencia de la precipitación en la superficie terrestre es muy compleja de definir porque depende de la altitud, distribución continental, océanos y del relieve (Galarza, 2016).

Para que se origine la precipitación se deben cumplir las siguientes condiciones (Breña & Jacobo, 2006):

- Presencia de núcleos de condensación.
- Temperaturas cercanas a la del punto de rocío.
- Abasto continuo de vapor de agua.
- Incremento del tamaño de las gotas a través de colisiones.

Según su origen se conocen tres tipos de precipitación:

- Ciclónicas son las provocadas por los frentes asociados a una borrasca o ciclón.
- Convección se produce por el ascenso de bolsas de aire caliente (tormentas de verano).
- Orográficas se presentan cuando las masas de aire húmedo son obligadas a ascender al encontrar una barrera montañosa.

- **Temperatura**

La temperatura es uno de los parámetros físicos más sensibles del clima, caracterizado por el calor o transferencia de energía entre sistemas (Sierra, 2006). Es una magnitud relacionada con la rapidez del movimiento de las partículas que constituyen la materia, cuanto mayor agitación de las partículas se presenta mayor es la temperatura (Rodríguez, Rosa Maria; Benito, Agueda; Portela, 2004).

- **Humedad Relativa**

La humedad relativa es una variable meteorológica que expresa la relación entre la cantidad de humedad del aire y la cantidad del aire contenido a la misma temperatura y presión si estuviese saturado; se expresa en %. (Sierra, 2006). Existen 3 tipos de humedad:

- Humedad relativa: Masa de vapor de agua, en gramos, contenida en 1m³ de aire seco
- Humedad específica: Masa de vapor de agua, en gramos, contenida en 1 kg de aire.
- Razón de mezcla: Masa de vapor de agua, en gramos, que hay en 1 kg de aire seco.

- **Presión atmosférica**

La atmósfera contiene gran cantidad de moléculas de gas que son atraídas hacia el planeta por la fuerza de gravedad. La medida del peso de la columna de aire sobre unidad de área que se encuentra en un sitio determinado es conocido como presión atmosférica (Galarza, 2016).

- **Viento**

Es una variable meteorológica que hace referencia al movimiento horizontal de masas de aire sobre la superficie de la tierra, causadas por la diferencia de presiones atmosféricas

atribuidas a la variación de temperaturas (Sierra, 2006). La circulación de las masas del viento se origina por: el déficit de temperatura (la masa de aire frío ocupa el espacio libre que dejó el ascenso del aire caliente) o déficit de presión (el aire se direcciona, desde las zonas de alta presión, hacia zonas de baja presión) (Galarza, 2016).

2.2. Machine Learning

Machine Learning conocido como aprendizaje automático, es una rama de la inteligencia artificial, tiene como objetivo desarrollar sistemas que aprendan a partir de la experiencia, para lo cual utiliza algoritmos capaces de generalizar patrones complejos en una gran cantidad de datos, adaptándose a cambios y mejorando el rendimiento con la experiencia (Arcila-Calderón, Barbosa-Caro, & Cabezuelo-Lorenzo, 2016) .

Uno de los objetivos del Aprendizaje Automático es extraer información de un objeto que no ha sido observado, basándose en los datos que son conocidos de objetos similares. Lo anterior puede ser extrapolado a información temporal, es decir, se puede predecir el comportamiento a futuro a partir de lo que ha ocurrido en el pasado.

A continuación, se hablará sobre algunos conceptos necesarios del Aprendizaje Automático.

2.2.1. Conjunto de datos

Conjunto de datos es un grupo de instancias similar a las muestras en el análisis estadístico. Normalmente está compuesto por n instancias con características específicas y comunes. Las características permiten distinguir a cada instancia como similar o diferente del resto, y así encontrar las relaciones entre los elementos de un conjunto de datos (Cabrera Mendieta, 2017)

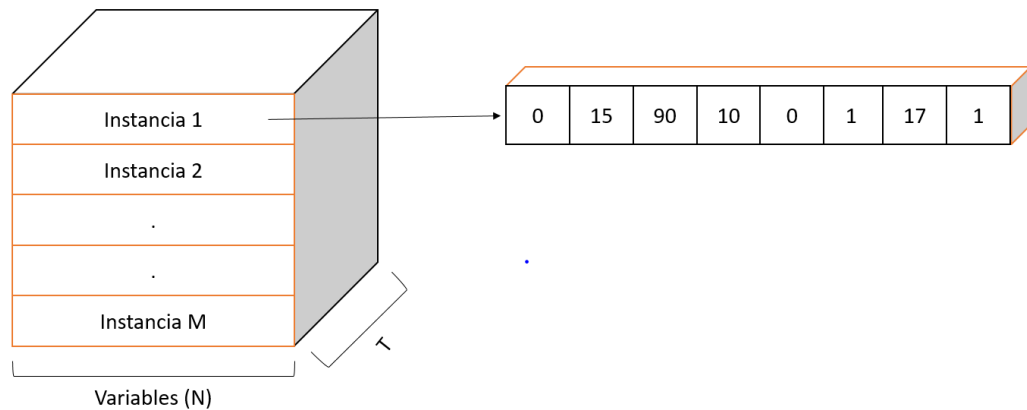
La manera de representar un conjunto de datos es de forma tabular, donde la fila, a , representa una instancia, Da , y la columna, b , representa la característica Fb . Se puede

apreciar un ejemplo en la Tabla 1, donde se presenta un conjunto de datos D con M instancias y cada una de ellas con N características.

Tabla 1: Representación de un conjunto de datos D

Instancia	$F1$	$F2$	Fb	F_N
$D1$	V_1^1	V_1^2	V_1^b	V_1^N
$D2$	V_2^1	V_2^2	V_2^b	V_2^N
....
D_a	V_a^1	V_a^2	V_a^b	V_a^N
....
D_M	V_M^1	V_M^2	V_M^b	V_M^N

El conjunto de datos para series temporales, tiene como característica una tercera dimensión que se dará en cada uno de los instantes de tiempo t , es decir para cada instancia, D , se tienen ciertas características, Fb , además estas características son adquiridas a lo largo de un rango temporal T (Gráfica 1).



Gráfica 1 Gráfica del conjunto de datos considerando su evolución temporal.

2.2.2. Tipos de aprendizaje

El uso del aprendizaje automático en la actualidad es muy diverso, una de las maneras de encontrar la mejor técnica de aprendizaje para una aplicación es determinando el tipo de aprendizaje. Los tipos de aprendizaje de Machine Learning son tres, aprendizaje

supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. A continuación, se explicarán los dos primeros tipos.

- **Aprendizaje supervisado**

El aprendizaje supervisado es un tipo de aprendizaje donde se desea estimar una función desconocida con datos o características denominadas variables de entrada, que permite obtener una variable de salida.

Formalizando esta idea, podríamos escribir:

$$\hat{y} = f(x) \quad (1)$$

Donde x es el vector de variables de entrada, f es la función que se consigue aprender y, por lo tanto \hat{y} es el pronóstico que devuelve f y que debe ser lo más cercano a un valor verdadero, y , para todas las instancias x del conjunto de datos D . Por lo tanto el aprendizaje supervisado se caracteriza por tener en D pares (x, y) con valores de y conocidos para cada una de las instancias.

Dependiendo del tipo de variable de salida, el aprendizaje supervisado se clasifica a su vez en dos tipos, clasificación y regresión.

- **Clasificación**

Es una sub-categoría del aprendizaje supervisado, cuyo objetivo es predecir clases. Cuando el número de clases posibles es 2 se conoce como Clasificador Binario. Por el contrario, si posee más de dos clases se le denomina Clasificador Multi-clase (Cabrera Mendieta, 2017).

Para el presente trabajo se utilizará el clasificador binario debido a que se disponen de dos clases (lluvia o no lluvia) para una posible variable aleatoria de salida (precipitación).

- **Regresión**

Es una sub-categoría del aprendizaje supervisado, el cual predice variables de salida continuas.

- **Aprendizaje no supervisado**

El aprendizaje no supervisado es un tipo de aprendizaje donde se desea estimar una función desconocida, a partir de conjuntos de datos con datos de entrada sin respuestas etiquetadas (datos de salida). Busca patrones que permite agrupar los datos. Un ejemplo de esta búsqueda de los patrones se da mediante clustering.

- **Clustering**

Es la técnica más común dentro del aprendizaje no supervisado, es utilizado en el análisis de datos exploratorios, para encontrar patrones ocultos o agrupaciones en los datos.

2.2.3. Evaluación de desempeño

Las técnicas más importantes dentro de la evaluación de desempeño son: hold-out y validación cruzada por k-fold. A continuación, describiremos cada una de ellas:

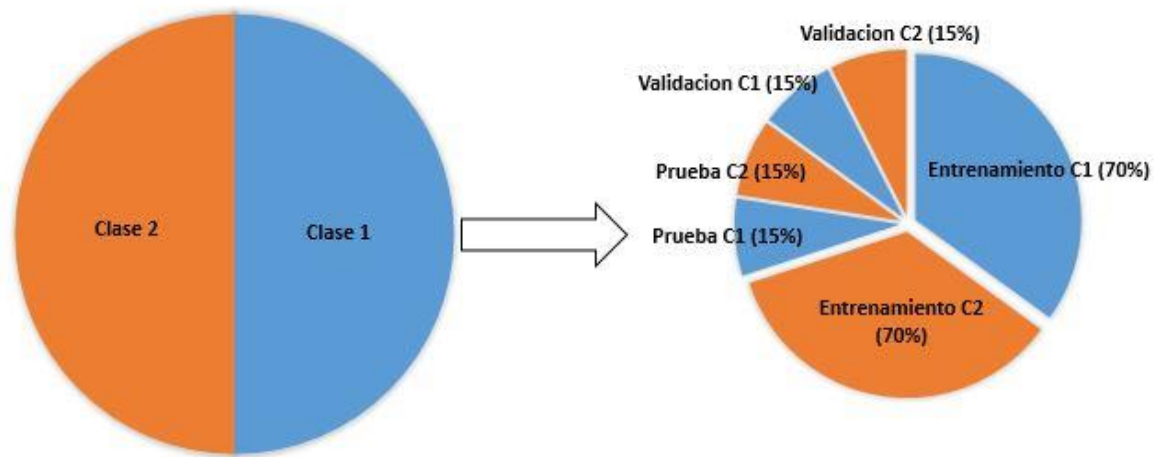
1. Hold-out

Es una técnica que divide el conjunto de datos en 3 grupos denominados entrenamiento, validación y prueba. El modelo empieza con el entrenamiento donde una parte del conjunto de datos ingresa al modelo, por lo general el 70% de datos son usados en la primera etapa. La segunda etapa se realiza periódicamente para evaluar el modelo durante el entrenamiento con el 15% del conjunto de datos. El entrenamiento termina cuando el rendimiento en la validación es aceptable o deja de mejorar. En la tercera etapa el modelo obtenido es evaluado con el resto de datos los cuales son nuevos para el modelo.

La ventaja que presenta este modelo es que su costo computacional es bajo, debido a que el modelo es entrenado una sola vez para llegar a un resultado final.

Junto con la partición de datos se realiza un proceso de estratificación cuyo objetivo es mantener la misma proporción de todas las clases en cada uno de los conjuntos.

Dentro del proceso de Hold-out se tiene dos casos. En el primero se realiza la división en dos conjuntos, uno de entrenamiento y otro de prueba. En el segundo caso se realiza la división en 3 partes una de entrenamiento, otra de validación y otra de prueba. A continuación, en la gráfica 2 se presentará un ejemplo de la segunda división que posteriormente será usada en el presente trabajo.



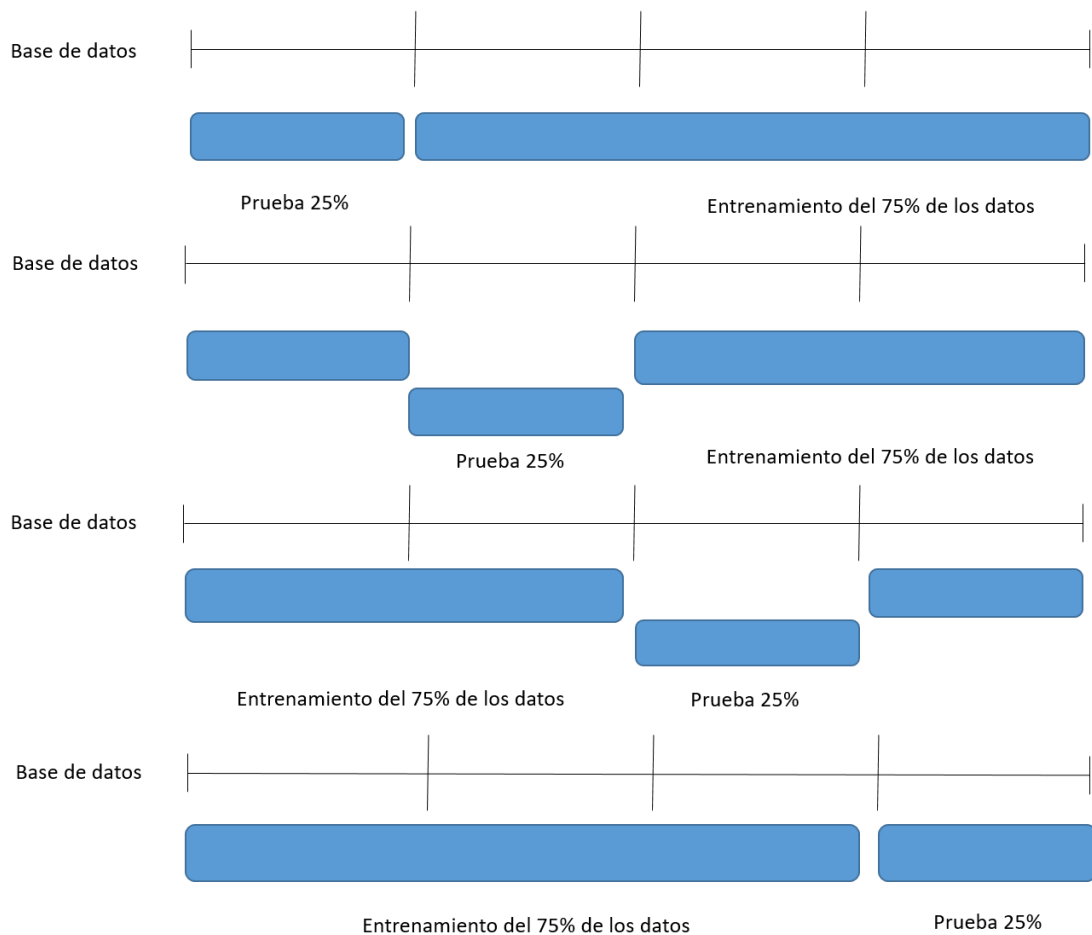
Gráfica 2: Ejemplo de Hold-out aplicado a un conjunto de datos para una partición de 70% para entrenamiento, 15% para validación y 15% para prueba con estratificación

2. Validación cruzada con K-folds

El conjunto de datos se divide en q partes iguales por la estratificación, en cada partición obtenida se entrena el modelo con las partes restantes $q - 1$ y es validada con la partición sobrante. (Reitermanová, 2010). De esta forma mediante esta técnica se realiza todas las combinaciones posibles, para obtener la estimación del error. El método se aplica cuando la base de datos es pequeña y no es posible realizar una validación de tipo Hold-out.

Una vez realizada todas las pruebas se procede a obtener la métrica de evaluación seleccionada para el problema de cada una de ellas. Luego se puede calcular una métrica general de desempeño del modelo, a través de un promedio de todos los resultados de la evaluación de cada partición. De igual manera se obtiene la varianza estimada en el proceso de evaluación, el cual disminuirá a medida que el número de particiones incremente. La desventaja que presenta esta técnica es el alto coste computacional al tener que generar tantos modelos como particiones elegidas.

A continuación, en la gráfica 3 se puede ver un ejemplo de validación cruzada con k-



fold con $q = 4$ particiones.

Gráfica 3: Ejemplo de validación cruzada con k-folds aplicado a un conjunto de datos para una partición de 75% para entrenamiento, 25% para validación.

2.2.4. Métricas

Las métricas son utilizadas para la evaluación de un modelo las cuales son capaces de definir el desempeño de este, por lo cual en esta sección vamos hablar de las métricas usadas para el modelo de pronóstico de precipitación.

1. Matriz de confusión

La matriz de confusión también puede ser conocida como matriz de contingencia. Es una tabla que permite visualizar de manera organizada los tipos de errores que un algoritmo de clasificación puede presentar. La matriz de confusión está formado por 2 columnas que contienen los valores predichos (0-1) y 2 filas que representan los ejemplos reales (0-1) (Staudemeyer & Omlin, 2013). Los elementos que podemos encontrar en la matriz son:

- **Verdadero positivo (VP).** – Son los elementos de una clase que verdaderamente fueron clasificados en esa clase.

$$VP(c) = M(c, c) \quad (2)$$

- **Verdadero negativo (VN).** –Son los elementos que no pertenecen a esa clase y que realmente el modelo predijo que no pertenecen a esa clase.

$$VN(c) = \sum_{i=1, i \neq c}^n M(i, i) \quad (3)$$

- **Falso positivo (FP).** - Es el número de elementos que de forma errónea fueron clasificadas como pertenecientes a esa clase.

$$FN(c) = \sum_{i=1, i \neq c}^n M(i, c) \quad (4)$$

- **Falso negativo (FN).** - Es el número de elementos de la clase que fueron clasificadas de forma incorrecta en el resto de clases

$$VN(c) = \sum_{i=1, i \neq c}^n M(c, i) \quad (5)$$

En la tabla 2 se presenta la matriz de confusión para un caso de clasificación binaria.

Tabla 2: Matriz de confusión

		Predicciones	
		Cond. P. Positiva	Cond. P. Negativa
Ejemplos reales	Cond. Positiva	VP	FN
	Cond. Negativa	FP	VN

2. Precisión

Esta métrica indica la capacidad del clasificador de no etiquetar como positiva una instancia que es negativa (Scikit-Learn, s/f).

$$Precisión = \frac{VP}{VP + FP} \quad (6)$$

3. Recall

Es la capacidad del clasificador para encontrar todas las muestras positivas (Scikit-Learn, s/f)

$$Recall = \frac{VP}{VP + FN} \quad (7)$$

4. F1-score

Esta métrica puede interpretarse como una media de la precisión y recall. Expresada por la siguiente fórmula:

$$F1 - score = 2 * \frac{Precisión * Recall}{Precisión + Recall} \quad (8)$$

5. Exactitud (accuracy)

Esta métrica mide la cercanía que tiene una cantidad medida con respecto al valor real de dicha cantidad (Joint Committee For Guides In Metrology, 2012). La ecuación 9 indica la expresión para calcular esta métrica:

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (9)$$

2.3. Modelos Neuronales

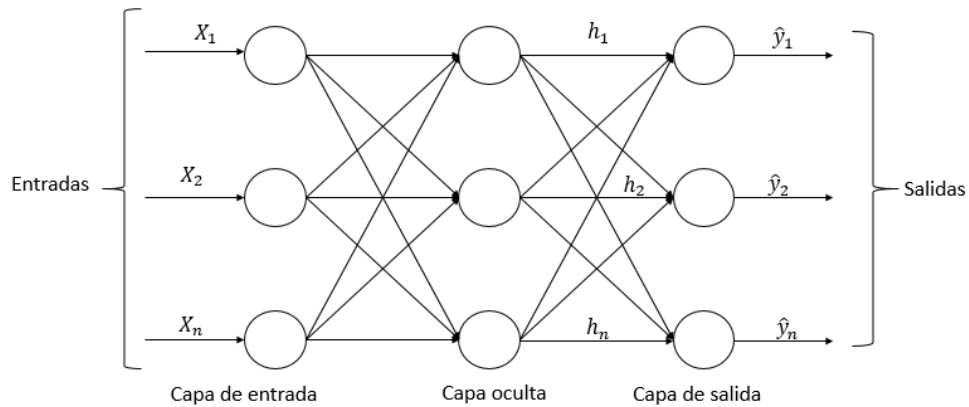
En la actualidad existen cientos de modelos neuronales disponibles para ser usados en entornos de aprendizaje automático como el perceptrón (Minsky & Papert, 1988), perceptrón multicapa (Rosenblatt & Van Der, 1986), redes neuronales clásicas (Werbos, 1974), redes neuronales recurrentes (Hornik, 1991), ADALINE (Widrow & Hoff, 1960), entre otros. A continuación, se hablará de las redes neuronales clásicas como referencia y de las redes neuronales recurrentes que serán usadas en el presente trabajo.

2.3.1 Redes Neuronales Clásicas

Las redes neuronales clásicas están compuestas por 3 capas (entrada, oculta y salida), cada una de estas se encuentran conformadas por unidades denominadas neuronas. Las capas de entrada son los valores iniciales que ingresan al modelo, la capa oculta forma la estructura interna de la red, y la de salida contiene los valores que se pretende obtener a la salida de la red.

Las capas de entrada contienen valores que son transformados mediante funciones de activación, cada entrada de las capas se encuentra multiplicada por pesos, los cuales son aprendidos siguiendo un criterio de minimización del error (retropropagación). La retropropagación tiene como objetivo calcular los gradientes del error con respecto a cada parámetro (pesos de la red), aplicando la regla de la cadena para encontrar los gradientes de las capas anteriores.

Posteriormente se conectan las salidas de la capa oculta con las entradas de la capa de salida usando otros pesos, siguiendo el mismo criterio de retropropagación. A continuación, en la gráfica 4 se puede apreciar la estructura de una red neuronal clásica.



Gráfica 4: Estructura de una red neuronal clásica.

2.3.2 Redes Neuronales Recurrentes (RNN)

Son un tipo de red neuronal que tienen caminos de retroalimentación entre todos los elementos que las conforman. Cada neurona se encuentra conectada con las neuronas posteriores en las siguientes capas, las capas anteriores y hacia ellos mismo a través de pesos (parámetros), los cuales se van actualizando y cambiando con el algoritmo de entrenamiento, con la finalidad de alcanzar los parámetros o metas de operación. Estas redes son más eficaces en la solución de problemas con no-linealidades temporales significativas, son útiles en el reconocimiento de patrones en series temporales (Barradas, 2009).

Las entradas a la red reciben un vector $x(t)$ de tamaño N_{in} , el cual va a ser igual al número de neuronas presentes en la capa de entrada, cada elemento se representara en notación matricial:

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_{N_{in}}(t) \end{bmatrix} \quad (10)$$

El vector de la capa oculta $h(t)$ de tamaño N_o procesa los valores de entrada y esta será igual al número de neuronas en la capa oculta:

$$h(t) = \begin{bmatrix} h_1(t) \\ h_2(t) \\ \cdot \\ \cdot \\ h_{N_o}(t) \end{bmatrix} \quad (11)$$

El vector anterior es procesado en la capa de salida para obtener un vector $\hat{y}(t)$, de tamaño N_{out} :

$$\hat{y}(t) = \begin{bmatrix} \hat{y}_1(t) \\ \hat{y}_2(t) \\ \cdot \\ \cdot \\ \hat{y}_{N_{out}}(t) \end{bmatrix} \quad (12)$$

Las RNN están diseñadas para capturar información de secuencias o series de tiempo. La operación se define como sigue:

$$h(t) = f(W_{in}x(t) + Wh(t-1)) \quad (13)$$

Dentro de la terminología de las RNN el argumento $h(t)$ es referido como las señales de activación en la capa oculta, f la función de activación no lineal, W_{in} es el peso entre la capa de entrada y la capa oculta, $x(t)$ se conoce como los valores de las variables de entrada en el instante t , W es el peso de la conexión de las variables dentro de la capa oculta y $h(t-1)$ salida de la capa oculta en el instante anterior.

Para obtener el valor de salida $\hat{y}(t)$ de la RNN se usó la ecuación (14) en donde g es la función de activación de la salida, W_{out} el peso de la capa final y $h(t)$ es la salida de la capa oculta obtenido con la ecuación (13), la salida se define como sigue:

$$\hat{y}(t) = g(W_{out}h(t)) \quad (14)$$

- Función de costo

La función de costo permite cuantificar que tan bien se encuentran las predicciones de la red con respecto a los datos reales. Primero la secuencia de valores de salida se encuentra definida en un intervalo $(0 \dots T)$ de tiempo, representada en forma de vector y como:

$$y = \begin{bmatrix} y(0) \\ y(1) \\ \cdot \\ \cdot \\ y(t) \\ \cdot \\ \cdot \\ y(T) \end{bmatrix} \quad (15)$$

El vector de y cuenta con su secuencia de pronóstico correspondiente, que se obtiene como resultado de las RNN, y se encuentra definido como \hat{y} :

$$\hat{y} = \begin{bmatrix} \hat{y}(0) \\ \hat{y}(1) \\ \cdot \\ \cdot \\ \hat{y}(t) \\ \cdot \\ \cdot \\ \hat{y}(T) \end{bmatrix} \quad (16)$$

Cada elemento de y representa el vector de salida real del conjunto de datos D y \hat{y} viene dada por la red para cada instante de tiempo en la secuencia.

La función de costo cuantifica el error entre estas dos secuencias. El costo total E se puede presentar como la suma de las funciones de costo parciales $E_t(y(t), \hat{y}(t))$, para cada instante t y se define como:

$$E = \sum_{t=0}^T E_t \quad (17)$$

- Función de activación

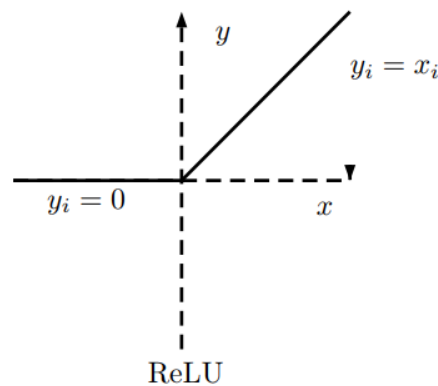
La función de activación calcula el estado de actividad de una neurona la cual transforma la entrada de los datos en un valor de activación, cuyo rango va normalmente de 0 a 1 o de -1 a 1 (Matich, 2001). Las funciones más comunes son: función lineal,

función sigmoide, función tangente hiperbólica, softmax y ReLU. En este trabajo se escogieron las dos últimas funciones mencionadas y se detallaran a continuación:

La función ReLU (rectified linear unit) permite que todos los valores positivos pasen positivos sin hacer cambios y asigna a todos los valores negativos el valor 0, esto se define de la siguiente manera:

$$y_i = \begin{cases} x_i & \text{si } x_i \geq 0 \\ 0 & \text{si } x_i < 0 \end{cases} \quad (18)$$

Expresada de la siguiente manera:



Gráfica 5: Función de activación ReLU

Fuente: (Xu, Wang, Chen, & Li, 2015)

La función de activación softmax es usada para la salida de la red. Esta es una generalización de la función sigmoide, la cual permite a la capa de salida tener diferentes probabilidades en cada nodo, en total estas deben sumar 1. La salida de esta función es equivalente a una distribución de probabilidad categórica.

$$Softmax(\hat{y}) = \frac{e^{y_i}}{\sum_{j=0}^{c-1} e^{y_j}}, i = 0 \dots T \quad (19)$$

La ventaja que presentan las RNN es que son aproximadores universales de sistemas dinámicos (Cabrera et al., 2017; Cabrera et al. 2018) lo que quiere decir que cualquier sistema que evoluciona en el tiempo puede ser estimado por este tipo de redes (Hornik, 1991). Por lo contrario, el problema que generan las RNN son dos, cuando los pesos en la matriz son pequeños (menores que 1), puede conducir a una situación que se conoce como

gradiente evanescente. La señal del gradiente puede ser tan pequeña que el aprendizaje se vuelve lento, lo que dificulta el aprendizaje de los datos en plazos largos de tiempo. El otro caso es llamado explosión del gradiente y se presenta cuando los pesos son grandes (mayores que 1), lo cual hace que el aprendizaje no llegue a converger (Montesdeoca, 2016) y el sistema de RNN sea inestable. Como solución a este inconveniente se aplica el tipo de RNN conocido como LSTM (Long Short Term Memory), el cual ha demostrado ser la solución al problema de las redes recurrentes. Este tipo de red cuenta con información adicional dentro de la neurona la cual es llamada memoria.

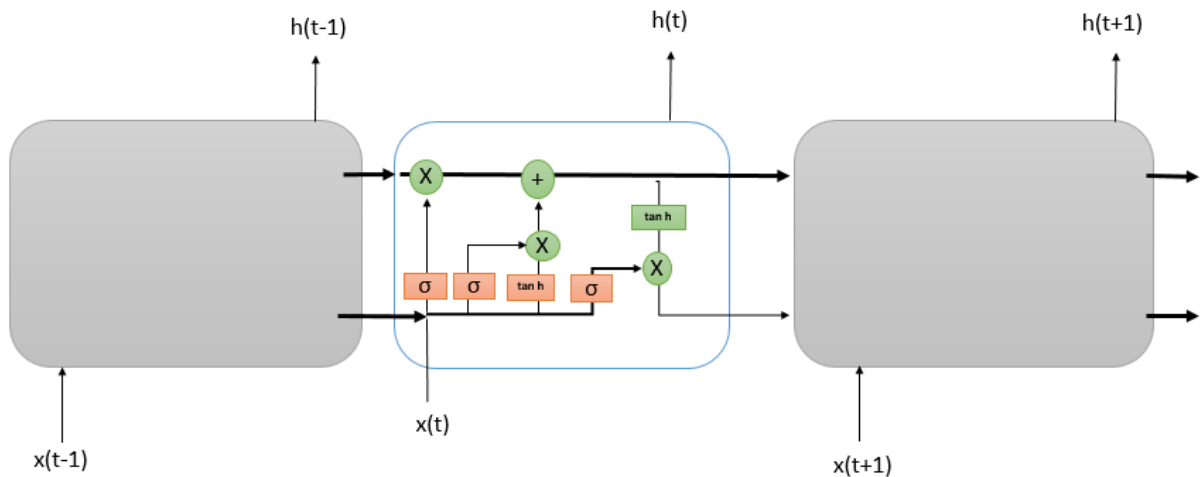
2.3.3 Long Short Term Memory

Son un tipo de RNN, conocidas por tener memoria a largo plazo, capaces de clasificar, procesar y predecir series de tiempo. La celda de memoria de un LSTM contiene tres puertas que son: input gate (puerta de entrada), output gate (puerta de salida) y forget gate (puerta del olvido) (Beaufays, Sak, & Senior, 2014; Jones, 2017; Yao et al., 2014).

Las RNN de tipo LSTM contiene los siguientes componentes:

- Forget gate “ f ” (sigmoide)
- Candidatos “ \tilde{C} ” (tanh)
- Input gate “ i ” (sigmoide)
- Output gate “ o ” (sigmoide)
- Estado oculto “ h ” (vector)
- Estado de memoria “ C ” (vector)

En la gráfica 6 se indica el diagrama de LSTM para cada paso de tiempo (t).



Gráfica 6: Esquema de la configuración de LSTM

Las puertas del modelo LSTM se encuentran formadas por una función de activación sigmoide (σ) junto con operaciones aritméticas como suma o multiplicación (Montesdeoca, 2016). Las puertas que contienen la función sigmoide están conformadas por valores entre cero y uno, los cuales permiten decidir cuanta información pasa a la siguiente etapa.

LSTM está representado por las siguientes formulas:

$$f_t = \sigma(W_f h(t - 1), x(t)) + b_f \quad (20)$$

$$\tilde{C}_t = \tanh(W_c h(t - 1), x(t)) + b_c \quad (21)$$

$$i_t = \sigma(W_i h(t - 1), x_i) + b_i \quad (22)$$

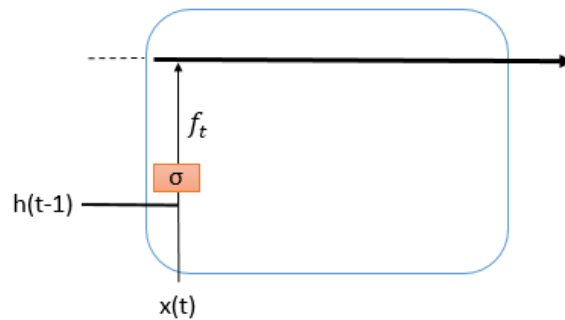
$$o_t = \sigma(W_o h(t - 1), x_i) + b_o \quad (23)$$

$$C_t = (f_t C_{t-1} + i_t \tilde{C}_t) \quad (24)$$

$$h_t = (o_t \tanh(C_t)) \quad (25)$$

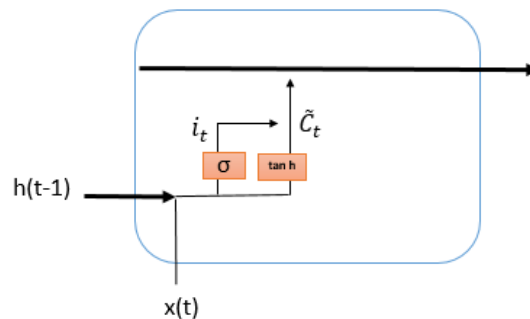
Primeramente, las estructuras LSTM deciden qué información va a olvidar de la memoria, esta decisión es tomada por la puerta conocida como “Forget gate” (f_t) que se

calcula por la concatenación de los valores de entrada $x(t)$ y los valores de salida de la neurona anterior $h(t - 1)$.



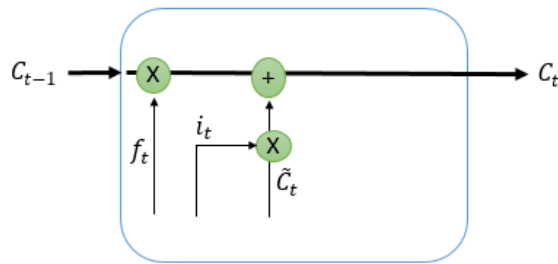
Gráfica 7: Forget gate

Luego decide qué información nueva será almacenada en la memoria mediante dos etapas, en la primera etapa el input gate (i_t) decide que valores se van a actualizar. Posteriormente se crea un vector de valores nuevos denominados candidatos (\tilde{C}_t). Al finalizar las dos etapas se concatenan los resultados anteriores para crear una actualización de estado.



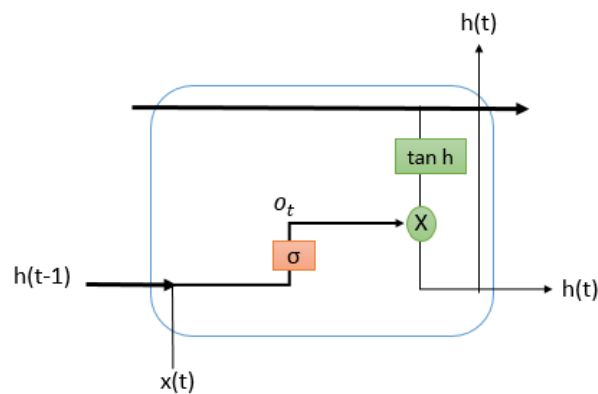
Gráfica 8: Input gate

En esta etapa se procede actualizar el estado de la memoria cambiando C_{t-1} por C_t , aplicando las operaciones de cada puerta (multiplicación, suma). Primeramente, se multiplica ($C_{t-1} f_t$) y a esto se suma ($i_t \tilde{C}_t$).



Gráfica 9: Actualización de la memoria

La salida de la red (o_t) se calcula en función de los valores de entrada $x(t)$ y $h(t-1)$ multiplicados a una función sigmoide (σ). Luego se multiplica los valores del estado de memoria (C_t) calculados anteriormente por la \tanh (permite que los valores tomen rangos entre -1 y 1). Finalmente se multiplica los dos valores anteriores para dar salida a la neurona h_t .



Gráfica 10: Salida

2.4. Lenguaje de programación

Lenguaje de programación es el elemento para la comunicación entre un programador y una computadora, siendo esta comunicación unidireccional, es decir las órdenes son emitidas por el programador y ejecutadas por la computadora. A continuación, se va hablar sobre el lenguaje de comunicación Python y sus librerías que se utilizó para el presente trabajo.

2.4.1 Lenguaje Python

Python es un lenguaje de programación sencillo, pero con un gran alcance, que cuenta con extensas librerías (pandas, numpy, matplotlib, seaborn, etc.). La cual le da un entorno poderoso para la informática científica (van Rossum, 2009).

Las librerías utilizadas en el presente trabajo se describen a continuación:

1. Pandas es una librería de código abierto para el análisis de datos tabulares y series de tiempo que proporciona herramientas y estructuras para el análisis de datos en Python. Pandas ofrece las siguientes estructuras (series, DataFrame, etc.) (Bloice & Holzinger, 2016).
 - Las series son arreglos unidimensionales con indexación, que permite almacenar varios tipos de datos. Este tipo de estructuras pueden ser generadas desde diccionarios o listas.
 - Los DataFrame son estructuras de datos similar a las tablas de bases de datos relacionales como SQL. También pueden ser utilizados como matrices bidimensionales (Krishna, 2017)
2. Numpy es el paquete fundamental para la informática científica de Python con arreglos multidimensionales de alta eficiencia y diseñados para cálculo científico. Este paquete contiene: tipos de datos, clases, funciones y módulos que posibilitan la creación y manejo de arreglos n dimensionales. Los arreglos de Numpy son alternativas a las listas de Python, siendo más rápidos, fáciles de trabajar y permiten realizar cálculos a través de arreglos completos. Un arreglo debe mantener las mismas dimensiones en todos sus elementos (NumPy community, 2010).
3. Seaborn es una librería que proporciona una interfaz de alto nivel para dibujar gráficos estadísticos basada en Matplotlib (Bloice & Holzinger, 2016)

4. Matplotlib es una librería 2D de trazados matriciales, el cual produce figuras de calidad en una variedad de formatos impresos. Esta librería está escrita en Python y hace un uso intensivo de Numpy. Matplotlib puede generar gráficos de barras, histogramas, diagramas de dispersión, entre otros, con tan solo pocas líneas de código (Hunter, Dale, Firing, & Droettboom, 2018).

2.4.2 Tensorflow

Creado por el grupo de investigación de inteligencia de Google, cuenta con un sólido respaldo para el aprendizaje profundo y el aprendizaje automático. Es una librería de código abierto que representa cálculos en forma de grafos (Tensorflow, s/f), almacena datos internamente en una estructura de grafo, facilitando la visualización de las dependencias entre operaciones y su asignación a diferentes dispositivos como los procesadores gráficos (Nuñez, 2016). La arquitectura de este permite una fácil implementación de computación en varias plataformas CPU, GPU, TPU).

2.5. Sistemas de Información Geográfico

Un Sistema de Información Geográfica (SIG) es una integración entre software, hardware y datos geográficos, llegando a ser una herramienta utilizada para la toma de decisiones. Lo que permite al usuario decidir cómo manejar el territorio analizado por medio del almacenamiento, interpolación, procesamiento y análisis de la información. Existen algunos tipos de SIG los más usados en la actualidad es el ArcGIS, GvGIS, Kosmo, Grass GIS y Quantum GIS (Santovenia Díaz, Javier Tarragó Montalvo & Cañedo, 2009)

En este trabajo se ha elegido Quantum GIS por ser un software gratuito que trabaja en cualquier sistema operativo.

2.5.1 Quantum GIS

QGIS es un Sistema de Información Geográfica de código abierto, el cual representa la información geográfica, donde se puede visualizar, editar, gestionar, analizar datos y diseñar mapas imprimibles. Posee una arquitectura extensible basada en un sistema de plugins que pueden ser utilizados o creados mediante lenguajes de programación como Python (Mendez, 2011).

Dentro del software Quantum Gis se realizó la interpolación que sirvió para presentar de manera gráfica los resultados obtenidos.

- **Interpolación**

La interpolación de datos a partir de un mapa de puntos (X, Y), permite la estimación de una variable Z, siendo esta una variable continua convertida en isoclinas, siendo ese su principal objetivo (García Gonzales, s/f).

La interpolación en un SIG, suele utilizarse para obtener capas raster siendo esta la variable a interpolar, por lo tanto, cada celda es un punto a interpolar (García Gonzales, s/f).

El método de estimación más usado para la interpolación es el Kriging, en donde a partir de información de una variable (precipitación), medida en ciertas áreas, permite predecir el valor de dicha variable en ciertas ubicaciones donde existe carencia de muestreo o no existe información (Peña, 2016).

Las estimaciones resultantes de este proceso se consideran óptimas ya que el error es mínimo.

La palabra Kriging proviene del nombre de D. G. Krige, geólogo sudafricano que desarrolló varios trabajos de suma importancia en la predicción de reservas de oro. El método de Kriging se basa en un conjunto de técnicas y métodos usados para la predicción espacial, su fundamento es el error medio cuadrático de predicción, esto puede llevarse a cabo expresando el valor de la variable aleatoria evaluada en cierta posición, $Z(x_0)$, que se

encuentra dentro del dominio de un sistema D , tal que $x \in D$, como una combinación lineal de los datos $Z(x_1), Z(x_2), Z(x_3), \dots, Z(x_n)$ (Giraldo Henao, 2002 como se cita en (Montalvo, 2017; Peña, 2016)):

$$Z^*(X_0) = \lambda_1 Z(X_1) + \lambda_2 Z(X_2) + \lambda_3 Z(X_3) + \dots + \lambda_n Z(X_n) = \sum_{i=1}^n \lambda_i Z(X_i) \quad (26)$$

Donde λ_i representa los pesos de los valores originales que se calculan en función de la distancia entre los puntos muestreados y el punto en donde se realizará la predicción correspondiente. El peso óptimo de λ_i es calculado de tal manera que la suma de los cuadrados del error sea mínimo al estimar $Z^*(x_0)$ por $Z(x_0)$ (Abad, 2014).

Ventajas

- Permite cartografiar la distribución de la población.
- Mayor número de datos repartidos en un área permite mayor nivel de detalle.
- Nos permite obtener información de áreas no exploradas (Garcia Gonzales, s/f)

El método de interpolación Kriging presenta 3 clases (Kriging Ordinario, Kriging Universal y Kriging Residual), para nuestra área de estudio se utilizó el Kriging Ordinario, el cual será descrito posteriormente.

Kriging Ordinario (OK)

Kriging es un método para predecir valores de una variable la cual no posee información, a través de una combinación lineal de valores medios, cuyos pesos dependen de la correlación espacial entre ellos, la suma de dichos pesos debe ser uno (Peña, 2016).

La desventaja que presenta este método al analizar fenómenos meteorológicos (precipitación) es que la información de entrada no cambiara con el tiempo o espacio, debido a que la media y la varianza permanece constantes, y los fenómenos meteorológicos no son

estacionarios (Peña, 2016).

$$E[Z(x+h) - Z(x)] = 0 \quad (27)$$

Donde E es el valor esperado o media, (x) y $(x+h)$ pertenecen al dominio de los datos, y h es la distancia entre dos puntos. También se asume que la varianza de la diferencia entre dos valores se encuentra en función solamente de la distancia (h), y no de su ubicación espacial x. la varianza se expresa como:

$$Var[Z(x+h) - Z(x)] = 2\gamma(h) \quad (28)$$

Donde $\gamma(h)$ es el semivariograma. Basado en esto, el predictor para el método Kriging Ordinario se expresa así:

$$\gamma(h, \alpha) = \frac{1}{2N(h, \alpha)} \sum_{i=0}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (29)$$

Donde:

$\gamma(h, \alpha)$ = Semivarianza como una función de la magnitud de la distancia entre puntos o vectores de separación (h) y su dirección (α).

$N(h, \alpha)$ = número de pares de observación separado por la distancia (h) y su dirección.

$Z(x_i)$ = variable aleatoria en la posición (x_i) (Abad, 2014).

2.6. Análisis estadístico descriptivo

La utilidad de la estadística es de proporcionar un conjunto de normas que permiten cuantificar determinados fenómenos, cada fenómeno está formado por un conjunto de personas o cosas a las que llamamos población (Orellana, 2001).

2.6.1 Población y muestra

Se llama población, universo o colectivo al conjunto de elementos que poseen una o más características observables en un ambiente geográfico determinado y en un lapso de tiempo. Muestra es un subconjunto representativo de la población elegido en términos de

representatividad, cuando hablamos de esto nos referimos a que debe ser lo más parecido posible a la población desde un punto de vista cuantitativo y cualitativo. (Montero, 2007,p.2)

2.6.2 Medida de posición o localización

La medida de posición, es aquel número que tiene la intención de indicar el centro de la distribución de un conjunto de datos (Orellana, 2001). Por consiguiente, se mencionará acerca de la media.

- Media

Es la medida de posición más frecuentemente usada. Para calcular la media aritmética de un conjunto de observaciones se suman todos los valores y se divide por el número total de observaciones (Mendenhall, Beaver, Beaver, Sánchez, & Patiño, 2002; Orellana, 2001).

$$\bar{X} = \frac{X1 + X2 + \dots + Xn}{n} \quad (30)$$

2.6.3 Medidas de dispersión o variabilidad

Las medidas de dispersión nos indican que tan separados están los datos entre sí y con respecto al valor central (Instituto Guatemalteco de Educación Radiofónica, 2016).

A continuación, se explicarán ciertas medidas de dispersión que han sido usadas en el presente trabajo.

- Cuartiles

Los cuartiles son los 3 valores que dividen a la muestra en cuatro partes iguales, el primer cuartil $Q_{1/4}$ será medida tal que el 25% de datos sean inferiores a su valor y el 75% de datos sean los superiores, el segundo cuartil $Q_{1/2}$ es la mediana y $Q_{3/4}$ será mayor a las $3/4$

partes de los cuartiles anteriores (Gorgas Garcia, Cardiel Lopez, & Zamorano Calvo, 2011).

$$Q_{1/4} = 25(n + 1) \quad (31)$$

$$Q_{3/4} = 75(n + 1) \quad (32)$$

Donde

n: número de población.

- Varianza y desviación estándar

La varianza de una muestra de n mediciones es la suma de las desviaciones cuadradas de las mediciones alrededor la media dividida entre n-1 (Mendenhall et al., 2002). Las observaciones que se encuentran a la derecha de la media tendrán distancias positivas, en tanto que las observaciones menores tendrán distancias negativas y la suma de las distancias a la media serán igual a cero (Orellana, 2001).

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (33)$$

La desviación estándar mide cuán lejos se encuentran los datos de la media muestral, se presenta como la raíz cuadrada positiva de la varianza (Orellana, 2001).

$$s = \sqrt{S^2} \quad (34)$$

- Curtosis y asimetría

La curtosis indica cómo está el agrupamiento de datos de acuerdo al valor central pudiendo tomar 3 formas: leptocúrtica si los valores están alrededor del valor central, platicúrtica si están alejados, y mesocúrtica si los valores poseen una distribución normal (Gorgas Garcia et al., 2011).

El coeficiente de curtosis se define como el cociente entre el momento de cuarto orden respecto a las medias y la cuarta potencia de la desviación típica (Gorgas Garcia et al., 2011).

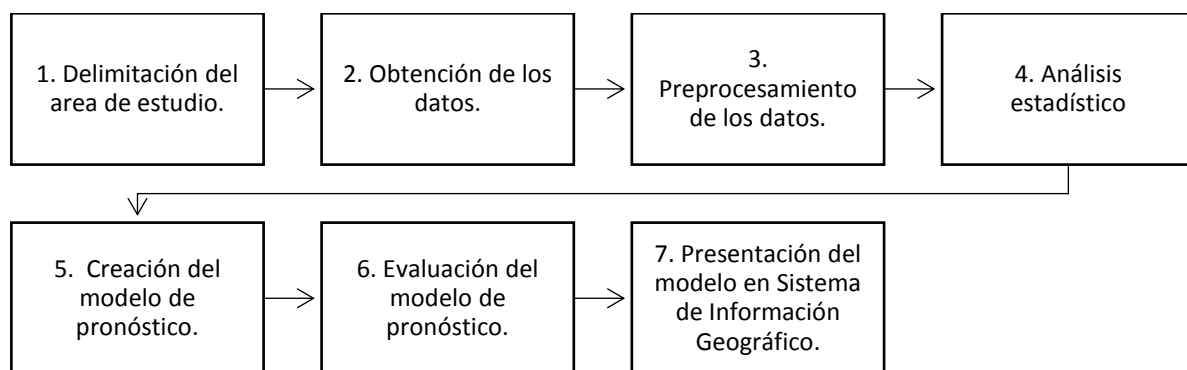
$$g_2 = \frac{m_4}{s^4} \text{ donde } m_4 = \frac{\sum_i^k = 1 (x_i - \bar{X})^4 n_i}{N} \quad (35)$$

Este coeficiente es adimensional, mientras más puntiaguda sea la distribución alcanzara valores mayores. Una distribución de medidas es simétrica cuando los valores de la variable equidistante a uno y otro lado del valor central tiene igual frecuencia. Es decir, un histograma simétrico (Gorgas Garcia et al., 2011).

3. METODOLOGIA

La generación de un modelo de pronóstico de precipitación, nos permite tomar decisiones para la mitigación de los daños que pueden causar los fenómenos tales como inundaciones o sequias y poder realizar un manejo adecuado de este recurso garantizando que la ciudadanía tenga un ambiente sano y ecológicamente equilibrado como estipula el Plan Nacional del Buen Vivir.

En la gráfica 11 se presentan las etapas usadas en el presente trabajo para la obtención un modelo de pronóstico de precipitación:



Gráfica 11: Esquema de la metodología del proyecto

3.1 Delimitación del área de estudio

La zona urbana del cantón Cuenca, se ubica en la región centro sur del Ecuador, entre las coordenadas 2°30' - 3°10' de latitud sur y 78°51' - 79°40' de longitud oeste. Pertenece a la región Sierra, provincia del Azuay. La variación altitudinal de su territorio administrativo varía entre 20 y 4560 msnm, mientras que Cuenca (zona urbana) tiene una cota media de 2550 msnm. Limita al norte con la provincia del Cañar, y al oeste con la Provincia del Guayas. Dentro de la provincia del Azuay, limita al sur con los cantones Camilo Ponce Enríquez, San Fernando, Santa Isabel y Girón, hacia el este con los cantones Paute, Gualaceo y Sigsig.

El cantón Cuenca tiene una extensión de 3665.33 km², de los cuales 73.01 km² (2%) corresponde a la zona urbana y 3592.31 km² (98% a la zona rural). Según la División Política Administrativa (DPA), Cuenca se compone por 15 parroquias urbanas y 21 parroquias rurales.

Hidrológicamente la ciudad se encuentra ubicada en la parte alta de la cuenca del río Paute, cuyo caudal es un afluente del río Namangoza, este a su vez es afluente del río Santiago que desemboca en el río Marañón, el cual termina en el Amazonas que desemboca en el Océano Atlántico. La ciudad de Cuenca se encuentra atravesada por 4 ríos

que provienen de las subcuencas del Tarqui, Tomebamba, Machángara y Yanuncay, siendo una ciudad rica en recursos hídricos.

El clima en la ciudad está influenciado por el régimen de la costa del Pacífico por el oeste y por masas de aire continentales y tropicales del Atlántico por el este. La ciudad de Cuenca se encuentra localizada en una zona donde la precipitación presenta una estructura bimodal, es decir presenta dos picos de épocas húmedas. La estación seca ocurre entre los meses de Junio a Agosto con una media anual de 660 a 1100 mm (Sucozhañay Calle, 2015).



Gráfica 12: Mapa del área de estudio.

3.2 Obtención de los datos

Se recolectaron datos de la Empresa Pública Municipal de Telecomunicaciones, Agua Potable, Alcantarillado y Saneamiento de Cuenca (ETAPA EP) de 3 estaciones meteorológicas y 6 pluviométricas dentro de la zona urbana de la ciudad de Cuenca.

Población: Variables meteorológicas en la zona urbana de Cuenca.

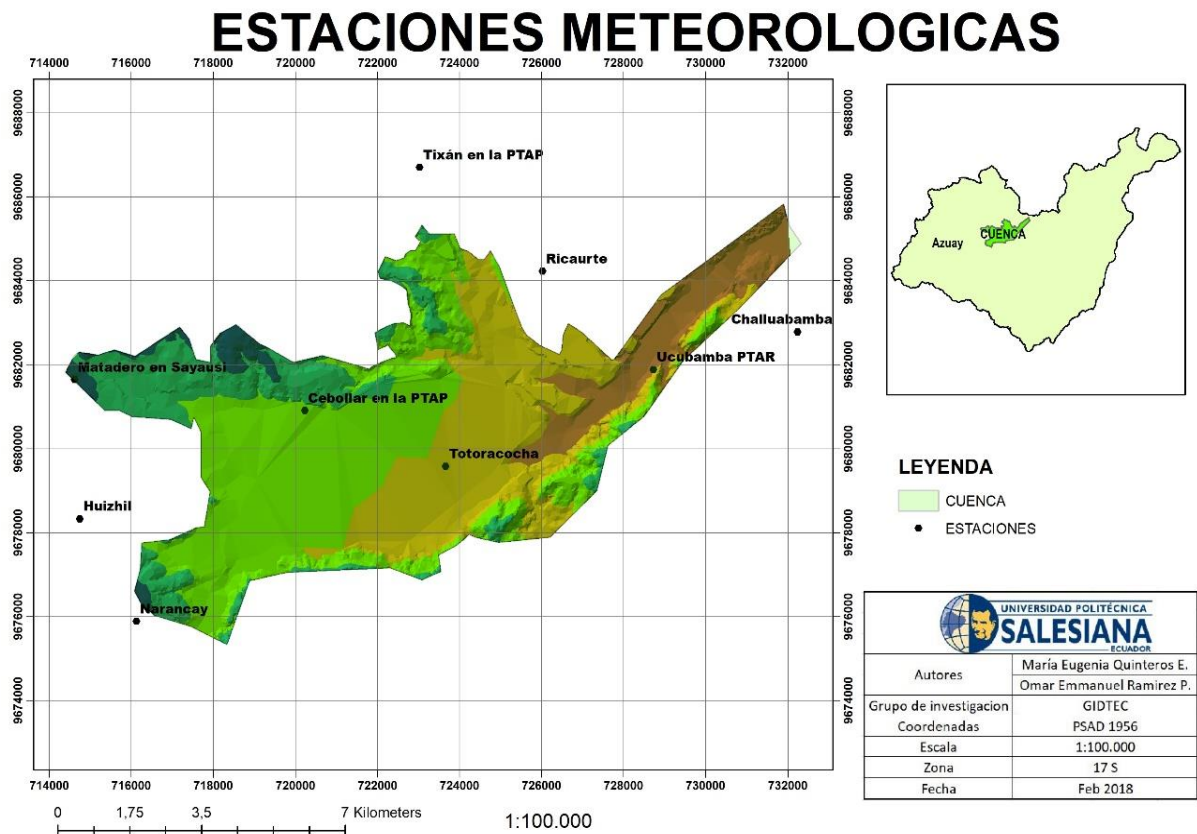
Muestra: Mediciones de las variables meteorológicas cada 5 minutos durante los meses de mayo 2014 a junio 2017 con su ubicación descrita en la Tabla 3, en un periodo aproximado de 4 años que es un rango temporal suficiente para generar modelos de pronóstico a partir de datos como se indica en Maqsood & Abraham, (2007).

Los registros de precipitación fueron proporcionados por ETAPA EP durante el periodo del 13 de mayo del 2014 a las 6h15 hasta el 12 de junio del 2017 a las 16h50, estos datos fueron tomados de las 9 estaciones correspondientes a Challuabamba, Narancay, Huizhil, Matadero de Sayausi, Totoracocha, Ricaurte, Ucubamba, Cebollar y Tixán. La precipitación es la única variable de estudio que se encuentra en todas las estaciones. Las variables de temperatura, humedad, presión atmosférica, velocidad y dirección del viento se encuentran en las estaciones correspondientes a Ucubamba, Cebollar y Tixán.

En la gráfica 13 se aprecia la localización de las estaciones en la ciudad de Cuenca.

Tabla 3: Ficha de información de las estaciones meteorológicas empleadas en el modelo.

Estación	Tipo	X	Y	Z	Código
Challuabamba	Pluviógrafo	732240	9682775	2521	01
Narancay	Pluviógrafo	716123	9675876	2728	02
Huizhil	Pluviógrafo	714747	9678318	2773	03
Matadero en Sayausi	Limni+Pluvi	714618	9681633	2693	04
Totoracocha	Pluviógrafo	723662	9679572	2516	05
Ricaurte	Pluviógrafo	726027	9684215	2600	06
Ucubamba PTAR	Meteorológica	728722	9681876	2425	07
Cebollar en la PTAP	Meteorológica	720230	9680898	2652	08
Tixán en la PTAP	Meteorológica	723026	9686691	2708	09



Gráfica 13: Distribución espacial de las estaciones meteorológicas.

3.3 Preprocesamiento de los datos

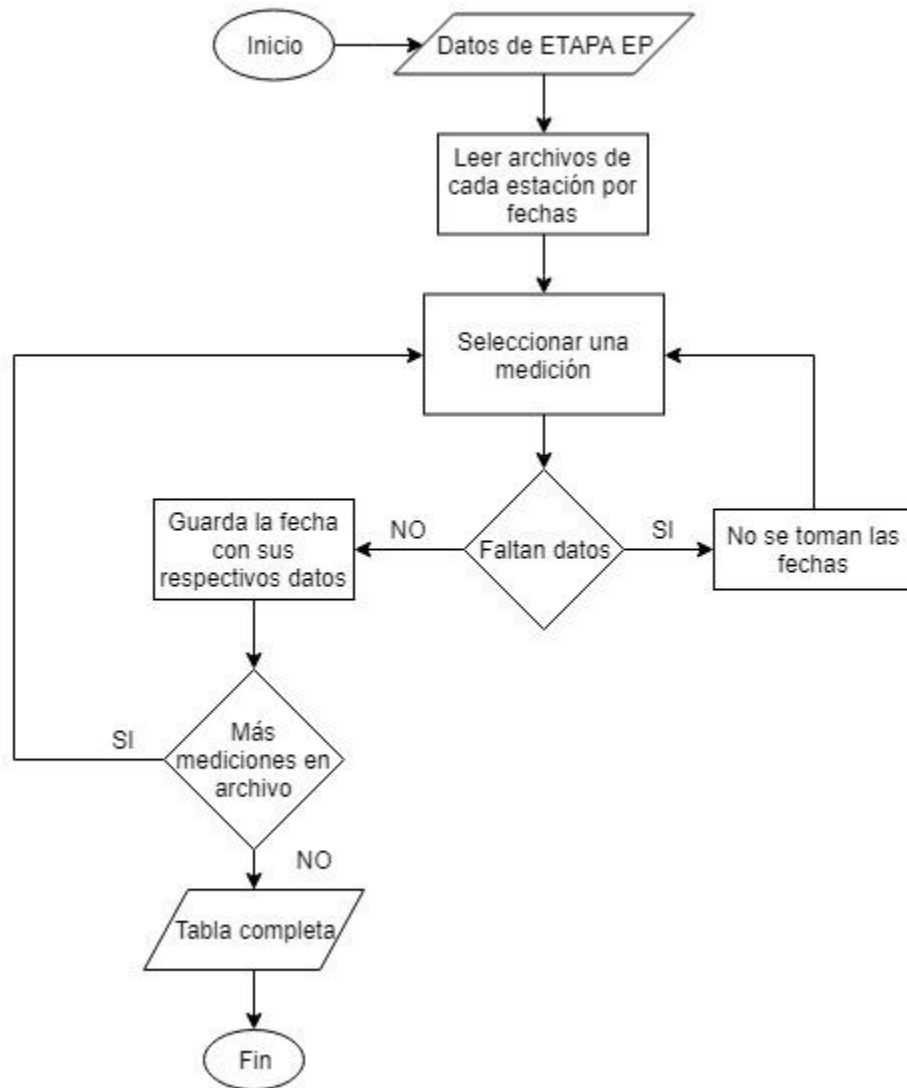
En el preprocesamiento se realizó la limpieza de los datos brutos para transformarlos en datos con formato para poder ser usados luego en el entrenamiento de la red.

El preprocesamiento de datos se realizó en 4 etapas:

1. Eliminación de mediciones incompletas:

En esta etapa se revisó todas las mediciones de las variables meteorológicas en las 9 estaciones, escogiendo las que se encontraban completas y eliminando mediciones incompletas.

En la gráfica 14 se indica el procedimiento para el preproceso de datos mediante un diagrama de flujo:

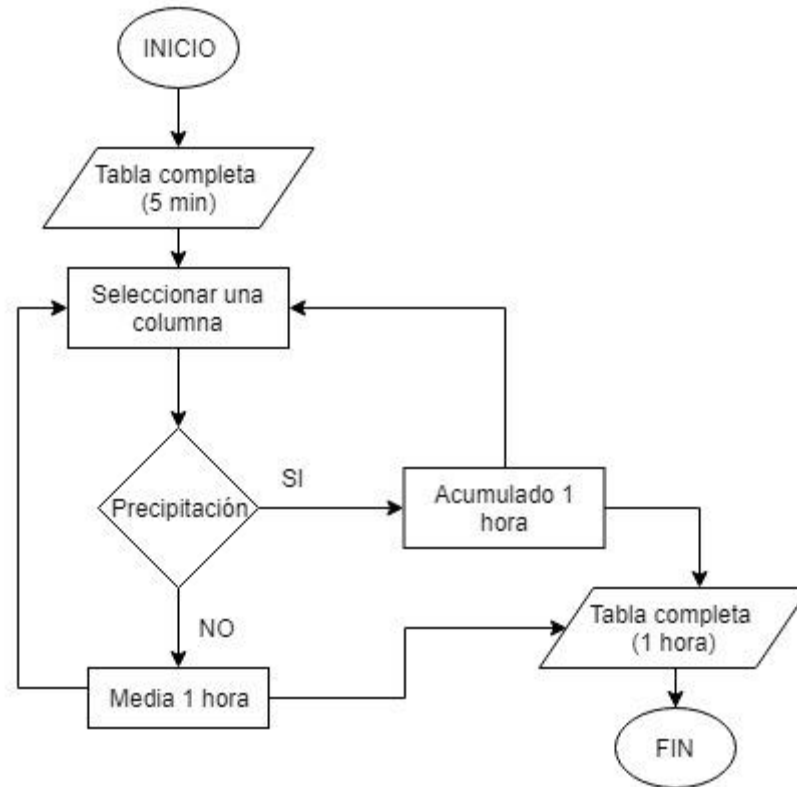


Gráfica 14: Diagrama de flujo de eliminación de mediciones incompletas

2. Disminución de la tasa de muestreo:

La tasa de muestreo inicial fue de 5 minutos. No se recomienda una tasa tan grande para cumplir con el objetivo de estimar si existirá precipitación en un intervalo de 1-24 h, por lo que se procedió a calcular el acumulado de las variables de precipitación y la media de las variables restantes (temperatura, humedad, presión atmosférica, velocidad y dirección del viento) obteniendo un submuestreo de las mediciones cada hora.

En la gráfica 15 se aprecia el procedimiento realizado para la disminución de la tasa de muestreo mediante el diagrama de flujo:

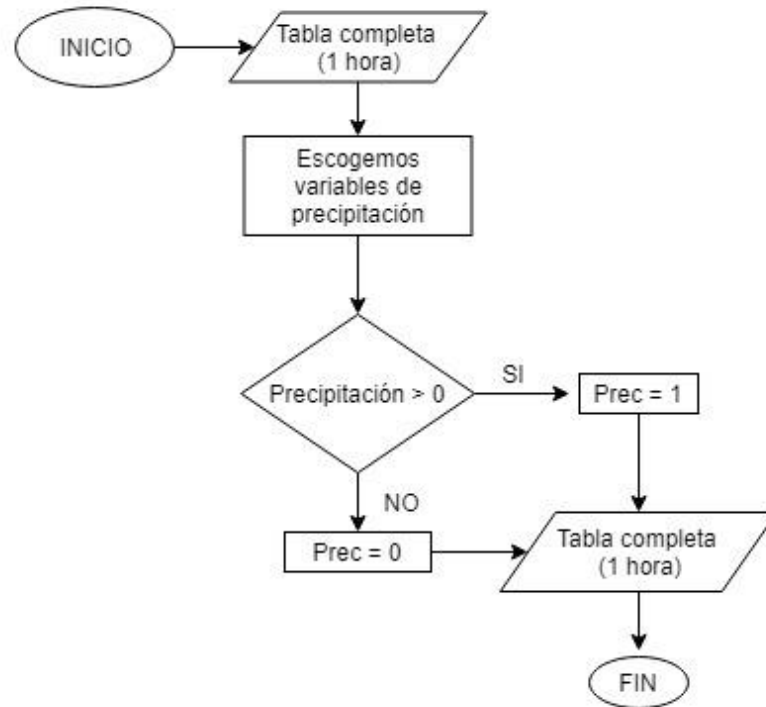


Gráfica 15: Diagrama de flujo de disminución de la tasa de muestreo

3. Discretización de la precipitación

Una vez submuestreados los datos en un intervalo de 1 hora se discretizó las variables de precipitación. Todos los valores de precipitación acumulados que son mayores a cero son considerados como lluvia a los cuales se asignó el valor de (1), y los valores igual a 0 como no lluvia (0), obteniendo dos posibles valores para la variable aleatoria de precipitación.

En la gráfica 16 se explica cómo se realiza la discretización de la precipitación:



Gráfica 16: Diagrama de flujo de discretización de la precipitación

Se aplicó la distribución de probabilidad de Bernoulli, la cual indica una probabilidad de éxito (1), y una probabilidad de error (0), es decir trabaja con dos valores.

$$P(x) = p^x(1 - p)^{1-x} \quad x = 0,1 \quad (36)$$

Dónde:

p = probabilidad del éxito

$(1 - p)$ = Fracaso

La distribución de probabilidad de la variable aleatoria de precipitación L , para un tiempo $t + k$, conocidos los datos de medición hasta un tiempo t queda expresada como:

$$P(L_{(t+k)} \mid \frac{(Temp_{7-9}, Hum_{7-9}, VV_{7-9}, DV_{7-9}, Pa_{7-9}, L_{1-9})}{\leq t}) \quad (37)$$

Donde:

$Temp$ = Temperatura

Hum = Humedad

VV = Velocidad del viento

Pa = Presión atmosférica

L = Precipitación

DV = Dirección del viento

t = Tiempo

k = Valor de salida deseado

3.4 Análisis estadístico

Se realizó el análisis estadístico de los datos para evaluar el comportamiento de las variables meteorológicas.

Con el fin de determinar la cantidad máxima y mínima de precipitación por mes en los diferentes años de estudio se realizará una gráfica de barras con los valores acumulados de precipitación en las 9 estaciones meteorológicas.

De igual manera para determinar la distribución y la simetría de la variable de precipitación se utilizará gráficos de cajas y bigotes (boxplot) en los datos mensuales de cada estación.

Con el propósito de cuantificar los días que presentan mayor cantidad de precipitación se realizará gráficas de histogramas mensuales de cada estación en los diferentes años de estudio, así se podrá observar que meses tienen mayores días de precipitación y la cantidad de precipitación presente.

Para analizar meses húmedos y secos en los diferentes años de estudio se realizará la comparación de las medias mensuales de cada estación con la media anual respectiva, esto permitirá tener un enfoque en que tan disperso se encuentra la precipitación mensual con respecto a la media anual.

Para las variables de temperatura, humedad, presión atmosférica y velocidad del viento se realizará gráficas de barras de las medias de cada una de las estaciones, cabe recalcar que estas variables se encuentran presentes solo en las estaciones 7,8,9.

Para la variable dirección del viento se trabaja con gráficos de rosa de los vientos para determinar hacia qué dirección sopla más el viento, este se realizará para cada mes de cada estación, al igual que las variables indicadas anteriormente esta se encuentra presente solo en las estaciones 7,8,9.

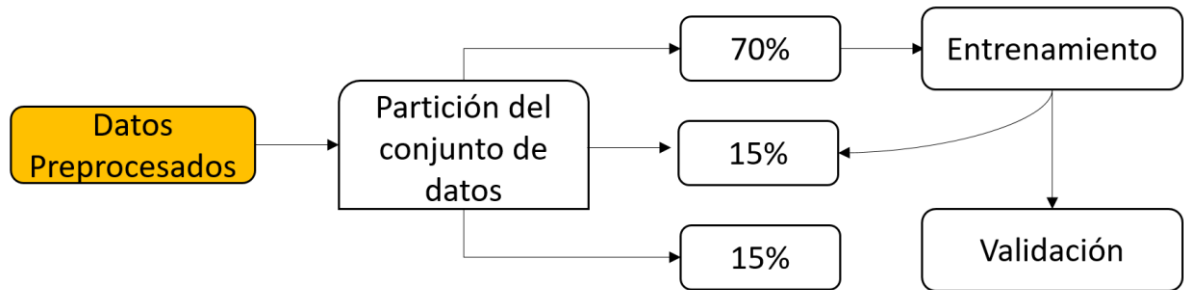
3.5 Creación del modelo de pronóstico.

El conjunto de datos que se utiliza para entrenar el modelo está conformado por 6 variables meteorológicas (temperatura, humedad, precipitación, presión atmosférica, velocidad y dirección del viento) situadas en 3 estaciones meteorológicas y 6 pluviométricas.

El modelo a usarse en el presente trabajo son las Redes Neuronales Recurrentes (RNN), puesto que permiten trabajar con series de tiempo. Se usó el tipo Long Short Term Memory (LSTM), el cual contiene una celda de memoria que permite recordar información de una gran cantidad de datos, por otro lado, este tipo de red soluciona los problemas del gradiente que presentan las redes recurrentes.

En el modelo de RNN de tipo LSTM se usó dos tipos de funciones de activación la función ReLU dentro de las capas ocultas y softmax para las capas de salida. ReLU permite que todos los valores positivos pasen positivos sin hacer cambios y asigna a todos los valores negativos el valor 0, mediante esta función se obtiene dos casos 0 o 1 donde 0 indica no lluvia y 1 lluvia. La función softmax codifica los parámetros de una función de Bernoulli, en la cual se obtiene dos valores a la salida del modelo, 0 que indica no lluvia y 1 que indica lluvia.

A continuación, en la gráfica 17, se puede apreciar el proceso que se sigue para el entrenamiento de la red LSTM.



Gráfica 17: Etapas del entrenamiento

Al finalizar la etapa de preprocesamiento indicada anteriormente, como resultado se obtiene una base de datos con un tensor tridimensional (gráfica 1), formado por la cantidad de datos s (varía según la estación meteorológica), número de variables N (24), y los pasos de tiempo (t) , además, cuenta con un vector de salida deseado (k) que es específico para cada una de las estaciones meteorológicas. Se creará un modelo para cada una de las estaciones meteorológicas es por eso que el vector nos viene individualizado para cada una de las estaciones.

Luego el conjunto de datos es particionado utilizando la técnica de Hold-out en la que se dividió el conjunto de datos en 70% entrenamiento, 15% validación y 15% prueba, utilizando la técnica con estratificación. La partición del conjunto de datos se realizó siguiendo un orden de tiempo; es decir para el 70% de datos utilizados de entrenamiento se tomó el 70% de los datos del inicio, luego el 15% de validación fue el 15% siguiente en tiempo y por último el 15% restantes para prueba. Se realizó en orden de tiempo para evitar mezclar datos pasados con futuros, puesto que se desea estimar a futuro.

Una vez con los datos particionados se utiliza el conjunto de entrenamiento para ajustar los hiperparámetros de la red neuronal, en el cual se usa diferentes configuraciones variando los hiperparámetros indicados en la tabla 4.

Una vez ajustados los hiperparámetros el modelo empieza a ser entrenado, se establece como parámetro fijo 100 iteraciones para el entrenamiento. Cada 5 iteraciones el conjunto

de validación es probado en el modelo, donde se obtiene métricas para evaluar que el modelo vaya mejorando a medida que se va entrenando. En cada parte de la evaluación del modelo si este ha mejorado con respecto a los modelos anteriores se almacena en una memoria el mejor modelo, de lo contrario se mantiene el mejor modelo generado hasta el momento. Una vez que se cumplan estas condiciones se ha terminado la fase de entrenamiento.

Tabla 4: Hiperparámetros variantes para el entrenamiento del modelo.

	1
NÚMERO DE CAPAS	2
	3
	0,1
TASA DE APRENDIZAJE	0,01
	0,001
	5000
TAMAÑO DEL BATCH	10000
	15000
	5
NÚMERO DE NEURONA EN CADA CAPA OCULTA	10
	15

3.6 Evaluación del modelo de pronóstico.

Para evaluar el rendimiento final y asegurar que el modelo es capaz de clasificar los datos que no han sido vistos aun, se usó el conjunto de prueba, el mismo que constituye el 15% de los datos. Al finalizar el entrenamiento, los datos ingresan a la red y evalúan el desempeño del modelo según las métricas (vistas en la sección 2.2.4).

En el caso que se disponga de nuevos datos de las diferentes estaciones meteorológicas el proceso para realizar inferencias sobre estos datos se deberá empezar desde la parte del preprocesamiento, los datos ingresaran a cada uno de los modelos según la estación meteorológica que corresponda y se obtendrá como salida la probabilidad de lluvia para esa estación.

3.7 Presentación del modelo en un sistema de información geográfico

Las probabilidades obtenidas desde cada uno de los modelos son publicadas por un servidor socket, donde un script cliente desarrollado dentro de Quantum GIS con la herramienta Python toma esta información para desplegarse dentro de la interfaz del programa. El despliegue de la interfaz realiza una interpolación entre cada una de las estaciones para obtener un mapa de distribución de probabilidades.

Los datos obtenidos luego de un proceso de inferencia son enviados mediante un servidor Socket (servidor-cliente).

4. RESULTADOS Y ANALISIS

4.1 Preprocesamiento de series de tiempo

En esta sección se puede apreciar el análisis de las series de tiempo de los datos correspondientes a las variables meteorológicas recopiladas por la empresa ETAPA EP. Se realizó un examen preliminar de los datos, para determinar la calidad de información.

Los datos analizados en las 9 estaciones (meteorológicas y pluviométricas), fueron obtenidos desde el mes de mayo del 2014 hasta junio del 2017 con una frecuencia de 5 minutos.

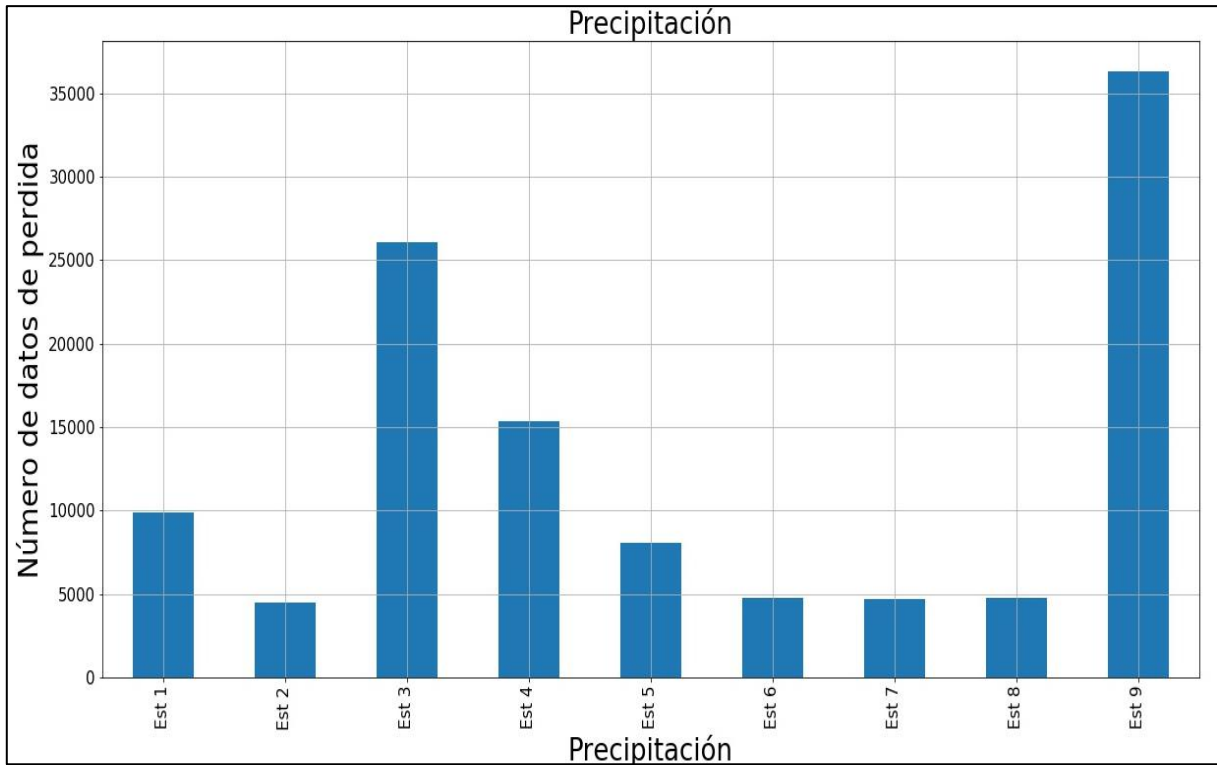
El total de datos brindados inicialmente por la empresa ETAPA EP fue de 7443014, el número de datos faltantes fue de 342970. La mayor parte pertenece a la estación 9 con un total de 217818 datos faltantes, esta estación posee el rango de fechas más amplio de datos faltantes que va desde el 13 de mayo del 2014 a las 6:15 hasta el 31 de agosto del 2014 a las 18:15. La variable que cuenta con la mayor cantidad de datos es precipitación de la estación 2 con 399900 datos.

En la tabla 5 se puede apreciar la cantidad de datos totales y datos faltantes de cada variable respecto a su estación.

Tabla 5: Número total de datos brindados por ETAPA EP

Estación	Variables	Datos Totales	Datos Faltantes
1	Precipitación	314555	9861
2	Precipitación	319900	4516
3	Precipitación	298355	26061
4	Precipitación	309064	15352
5	Precipitación	316376	8040
6	Precipitación	319652	4764
7	Temperatura	319741	4675
	Humedad Relativa	319741	4675
	Presión atmosférica	319741	4675
	Precipitación	319741	4675
	Velocidad del viento	319740	4676
	Dirección del viento	319740	4676
8	Temperatura	319665	4751
	Humedad Relativa	319665	4751
	Presión atmosférica	319665	4751
	Precipitación	319665	4751
	Velocidad del viento	319665	4751
	Dirección del viento	319665	4751
9	Temperatura	288113	36303
	Humedad Relativa	288113	36303
	Presión atmosférica	288113	36303
	Precipitación	288113	36303
	Velocidad del viento	288113	36303
	Dirección del viento	288113	36303

En la gráfica 18 se aprecia el número de pérdida de datos para la variable de precipitación, donde se aprecia la estación 9 como la estación con mayor pérdida de datos, y la estación 2 la variable con menor pérdida de datos con un total de 4516 datos.



Gráfica 18: Gráfica de barras de pérdida de precipitación.

Luego del análisis de los datos se procedió a la primera etapa de procesamiento en la que se eliminaron los datos incompletos, obteniendo como resultado la tabla 6 con los rangos de datos completos. Los cuales fueron usados posteriormente para el entrenamiento del modelo.

Tabla 6: Fecha de datos completos.

Fecha de inicio		Fecha de finalización		Fecha de inicio		Fecha de finalización	
31/8/2014	18:20	5/1/2015	14:40	31/5/2016	11:00	11/7/2016	14:50
6/1/2015	9:20	6/1/2015	18:55	11/7/2016	15:00	2/8/2016	13:25
10/1/2015	11:00	13/1/2015	9:50	2/8/2016	14:30	9/8/2016	13:50
13/1/2015	12:15	9/4/2015	12:35	23/8/2016	9:50	15/9/2016	11:20
9/4/2015	15:00	10/4/2015	10:45	15/9/2016	11:30	28/9/2016	14:30
10/4/2015	11:50	6/5/2015	18:55	28/9/2016	16:05	10/10/2016	15:25
10/5/2015	0:05	13/8/2015	23:00	10/10/2016	15:50	7/11/2016	20:00
14/8/2015	21:05	1/11/2015	16:50	25/11/2016	11:55	29/11/2016	10:50
18/11/2015	10:05	24/11/2015	12:35	29/11/2016	11:00	2/12/2016	14:20
17/12/2015	14:00	31/12/2015	14:45	2/12/2016	14:40	6/12/2016	14:20
7/1/2016	11:40	11/1/2016	8:00	6/12/2016	14:45	5/1/2017	8:45
11/1/2016	16:40	12/1/2016	4:20	17/1/2017	16:10	26/1/2017	15:20
12/1/2016	13:00	13/1/2016	12:50	26/1/2017	15:30	14/3/2017	11:15
13/1/2016	13:00	14/1/2016	23:00	14/3/2017	11:25	2/5/2017	15:10
14/1/2016	23:10	18/1/2016	17:00	2/5/2017	15:20	11/5/2017	14:25
19/1/2016	1:40	20/1/2016	16:45	12/5/2017	17:00	15/5/2017	10:15
20/1/2016	16:55	21/1/2016	23:30	15/5/2017	17:50	18/5/2017	11:45
21/1/2016	23:40	26/1/2016	21:40	18/5/2017	12:45	19/5/2017	15:20
27/1/2016	6:25	24/5/2016	9:45	19/5/2017	16:20	8/6/2017	15:00
24/5/2016	9:55	28/5/2016	3:00				

Para entrenar el modelo se obtuvo un total de 260982 datos completos.

4.2 Comportamiento de las variables meteorológicas

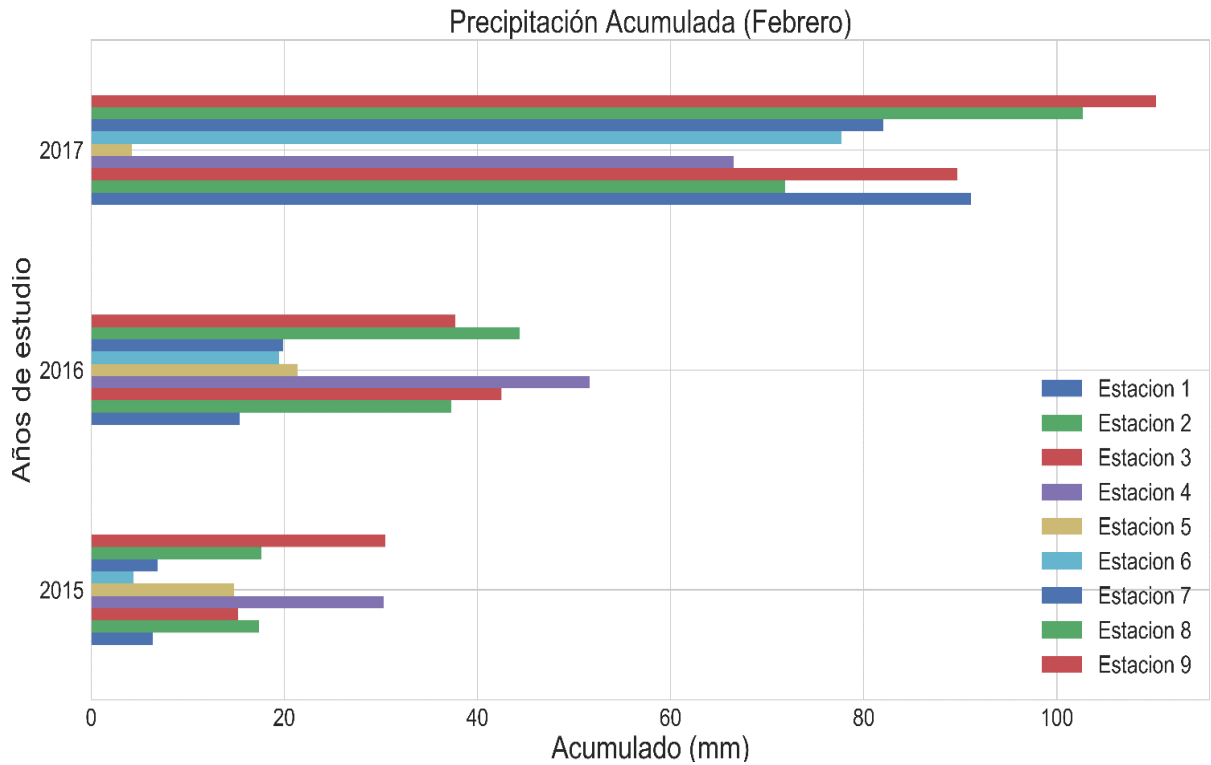
En el siguiente apartado se muestra los gráficos más relevantes obtenidos del análisis estadístico de las variables meteorológicas.

1. Precipitación

Se realizó el análisis de precipitación acumulada por mes dentro de los años de estudio (2014-2017), donde se obtuvieron los siguientes resultados.

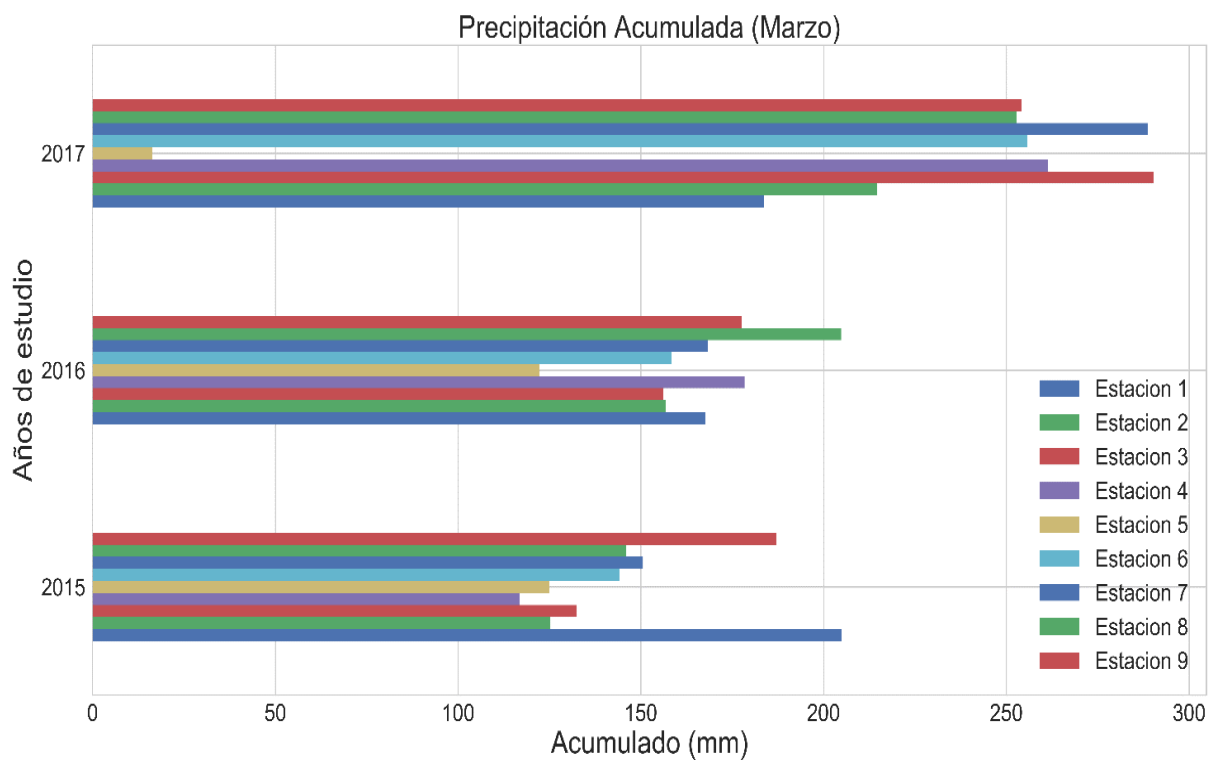
Los meses de febrero, marzo y junio, fueron elegidos para ser analizados puesto que son los meses más representativos en comparación de los otros meses, además presentan un comportamiento muy distinto en cada año.

En la gráfica 19 se analizó la precipitación acumulada (mm) en el mes de febrero en los 3 años de estudio en donde se pudo observar que en el año 2017 presentó mayor cantidad de precipitación y el año 2015 presentó menor cantidad de precipitación.



Gráfica 19: Gráfica de barras de precipitación acumulada en el mes de febrero de los años 2015, 2016, 2017 de las nueve estaciones de estudio.

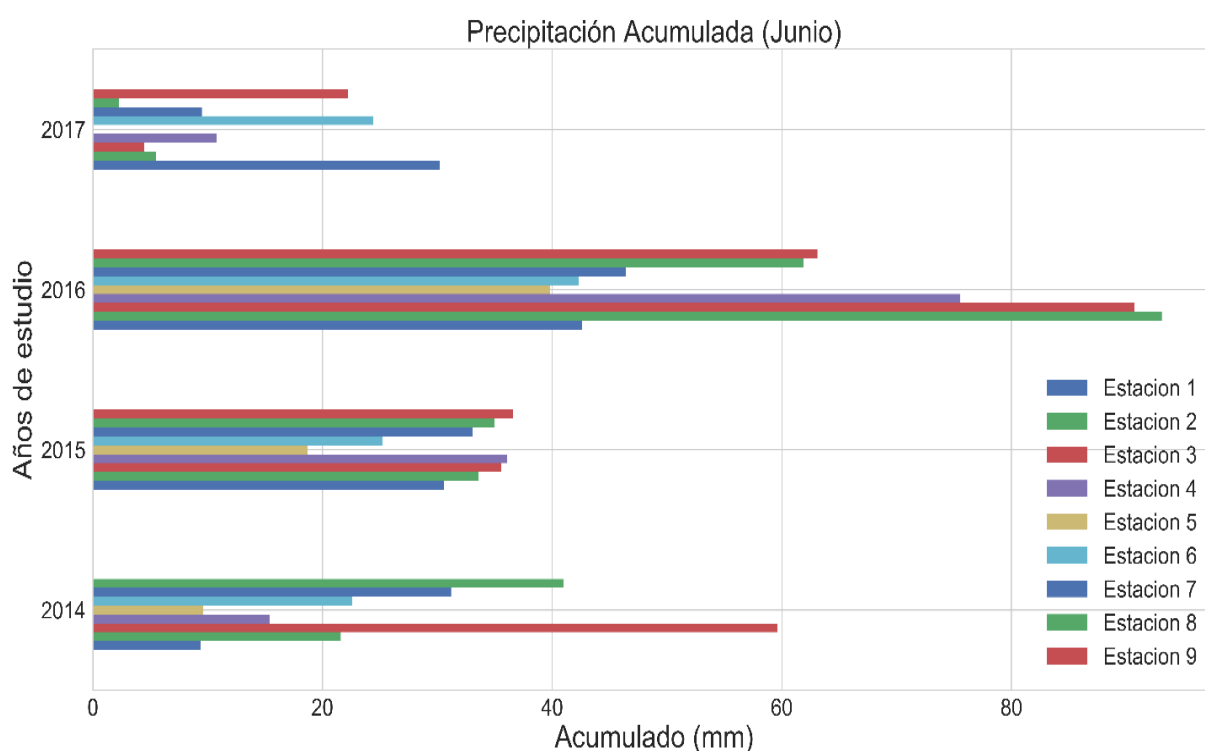
En la gráfica 20 se presenta los valores acumulados de precipitación en el mes de marzo durante el periodo de 2014-2017, se observa que en el año 2017 existe mayor



cantidad de precipitación, por lo contrario, los años 2015-2016 presentan datos similares.

Gráfica 20: Gráfica de barras de precipitación acumulada en el mes de marzo de los años 2015, 2016, 2017 de las nueve estaciones de estudio.

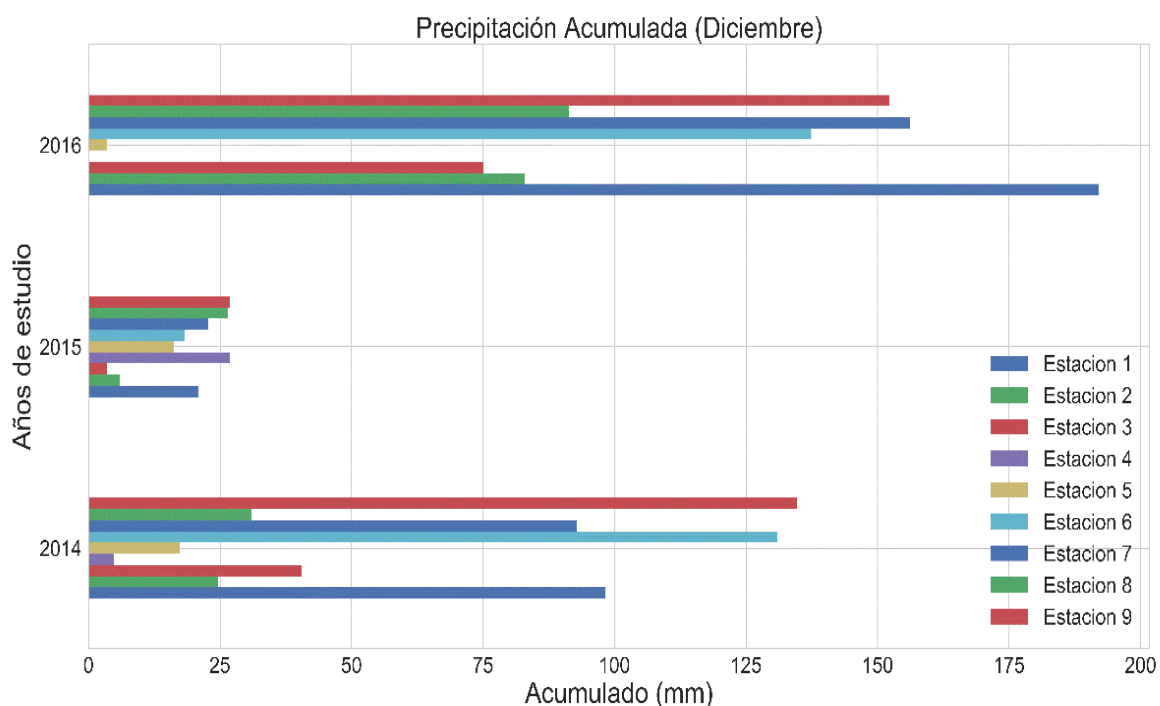
De igual manera se analizó el mes de junio en la gráfica 21 donde se observó que el año 2017 presenta menor cantidad de precipitación acumulada, por lo contrario, el año 2016 presenta mayor cantidad de precipitación.



Gráfica 21: Gráfica de barras de precipitación acumulada en el mes de junio de los años 2014, 2015, 2016, 2017 de las nueve estaciones de estudio.

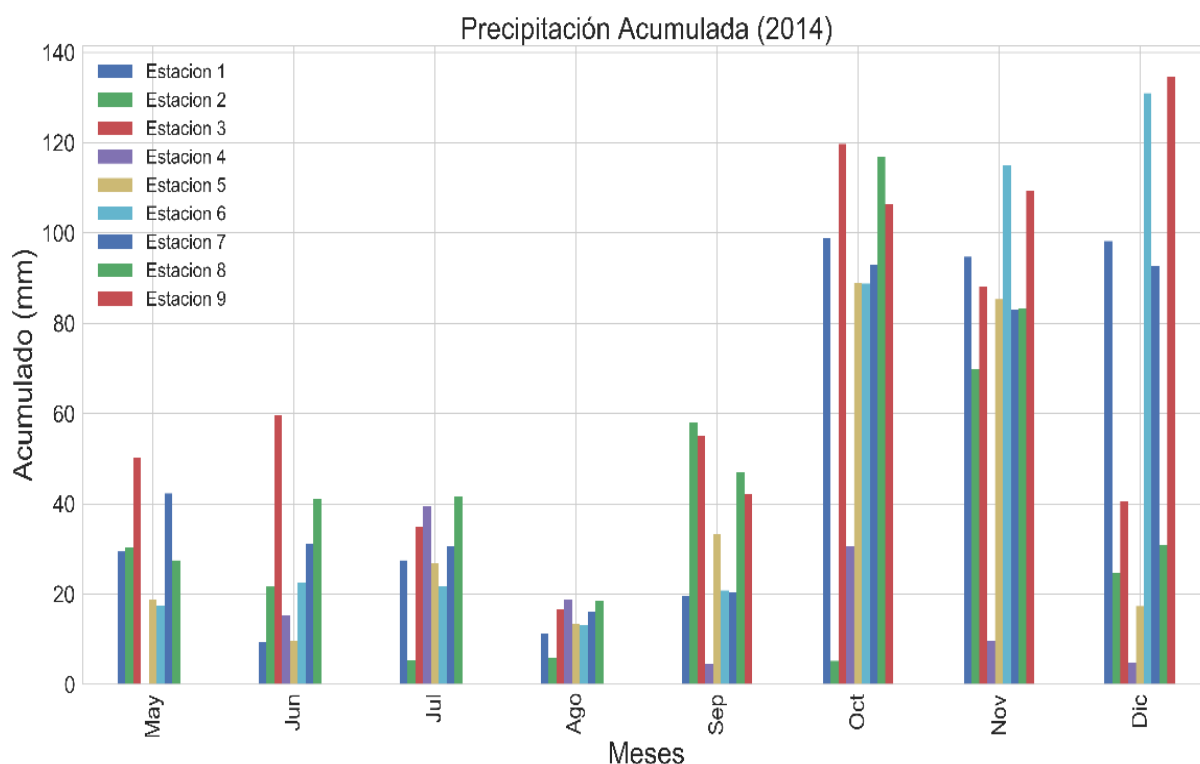
La gráfica 22 muestra los valores acumulados de precipitación (mm) de las 9 estaciones meteorológicas en el mes de diciembre. En el gráfico de barras se observa que el

2016 es el año con más precipitación acumulada y el año 2015 presenta menor cantidad de precipitación en ciertas estaciones.



Gráfica 22: Gráfica de barras de precipitación acumulada en el mes de diciembre de los años 2014, 2015, 2016 de las nueve estaciones de estudio.

De igual manera se realizó el análisis mensual de los datos por cada estación, representada por diferentes colores en los diferentes años de estudio. La gráfica 23 indica las precipitaciones acumuladas de cada estación en el año 2014.



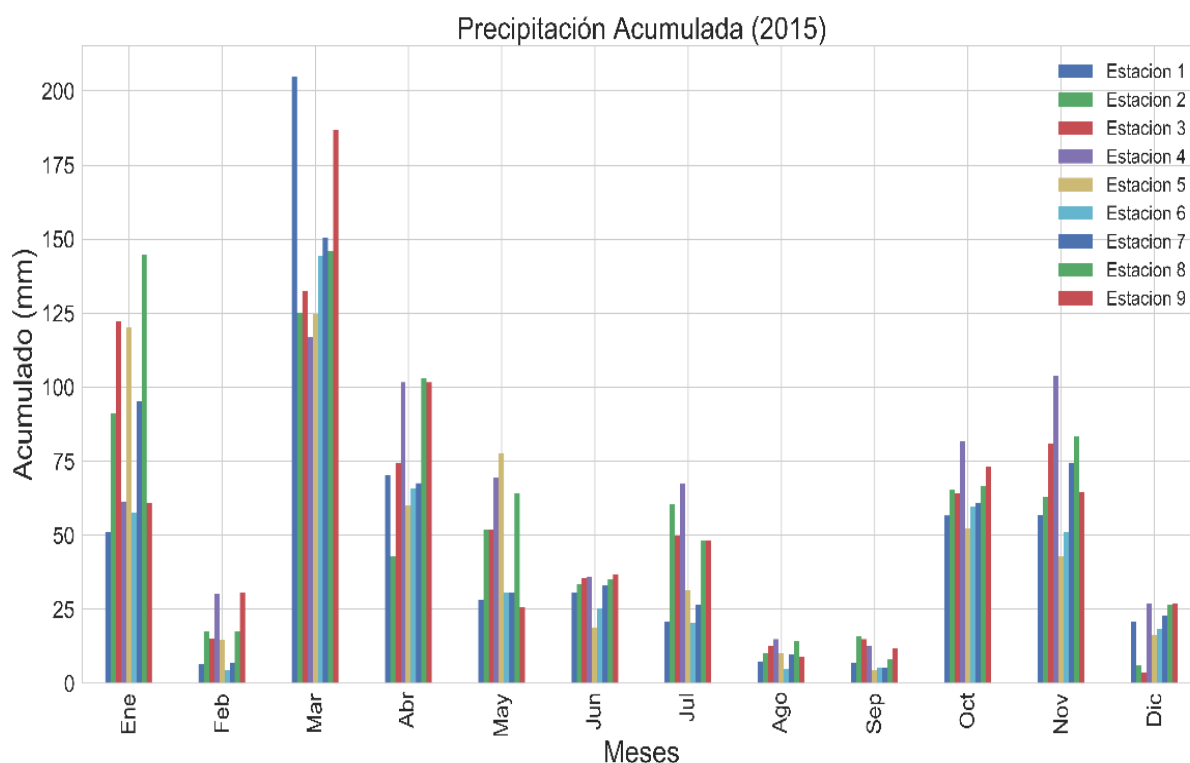
Gráfica 23: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2014 de las 9 estaciones.

En el año 2014 la estación 3 (Huizhil) posee los valores máximos de precipitación durante los meses de mayo, junio y octubre. Por lo contrario, la estación 4 (Sayausi) posee los valores mínimos durante los meses de septiembre, noviembre y diciembre (Tabla 7).

Tabla 7: Resumen de precipitación acumulada en el año 2014

2014				
MES	MAX (mm)	ESTACIÓN	MIN (mm)	ESTACIÓN
MAYO	50,3	3	15,5	6
JUNIO	59,59	3	9,4	1
JULIO	41,5	8	5,3	2
AGOSTO	18,7	4	0	9
SEPTIEMBRE	58	2	4,6	4
OCTUBRE	119,69	3	5,2	2
NOVIEMBRE	115	6	9,6	4
DICIEMBRE	134,7	9	4,9	4

La gráfica 24 indica los valores del año 2015 mensuales por estación.



Gráfica 24: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2015 de las 9 estaciones.

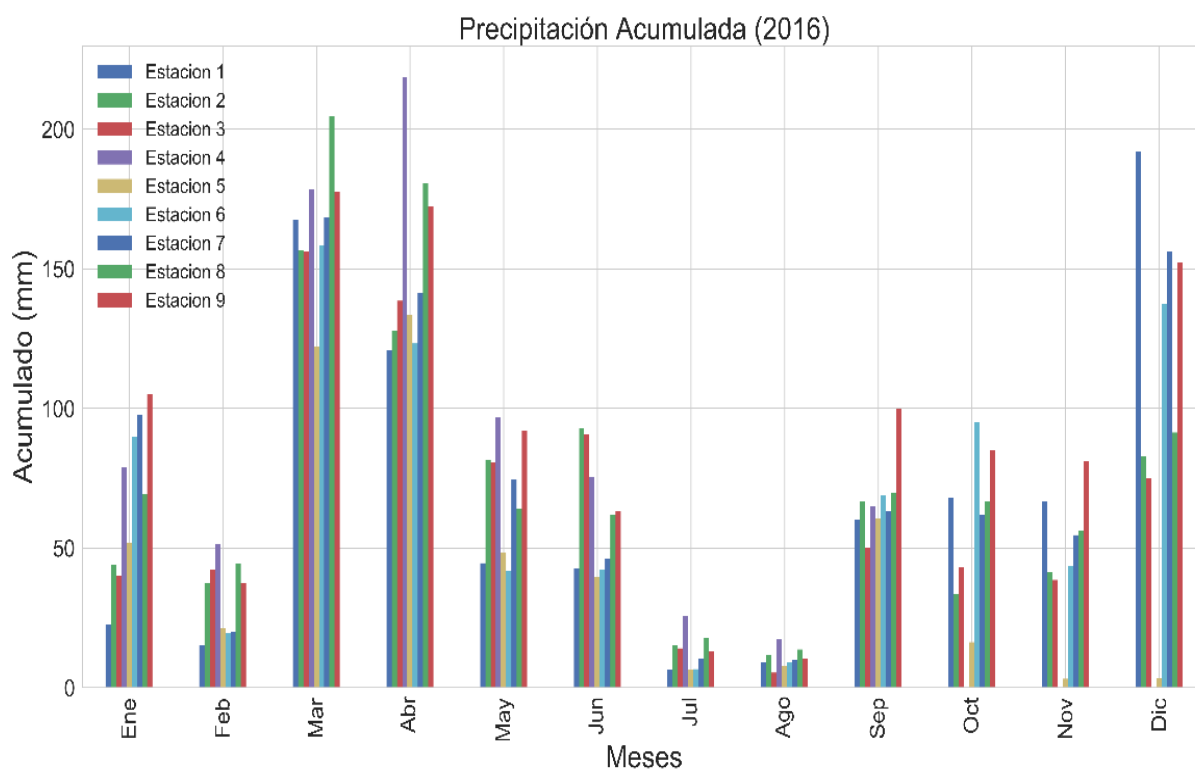
Los meses más secos en el año 2015 son febrero, agosto, septiembre y diciembre.

Los meses más húmedos encontrados fueron en enero, marzo, abril, noviembre. La estación 4 presentó los máximos durante más meses en el año y la estación 5 presentó valores mínimos (Tabla 8).

Tabla 8: Resumen de precipitación acumulada en el año 2015

2015				
MES	MAX (mm)	ESTACIÓN	MIN (mm)	ESTACIÓN
ENERO	144,8	8	51,2	1
FEBRERO	30,5	9	4,4	9
MARZO	204,9	1	116,9	4
ABRIL	103,1	8	42,8	2
MAYO	77,5	5	25,8	9
JUNIO	36,6	9	18,7	5
JULIO	58,1	4	26,2	6
AGOSTO	14,9	4	4,7	6
SEPTIEMBRE	16	2	4,7	6
OCTUBRE	73,3	9	52,2	5
NOVIEMBRE	103,6	4	42,9	5
DICIEMBRE	26,9	4	3,6	3

La gráfica 25 indica los valores del año 2016 mensuales por estación.



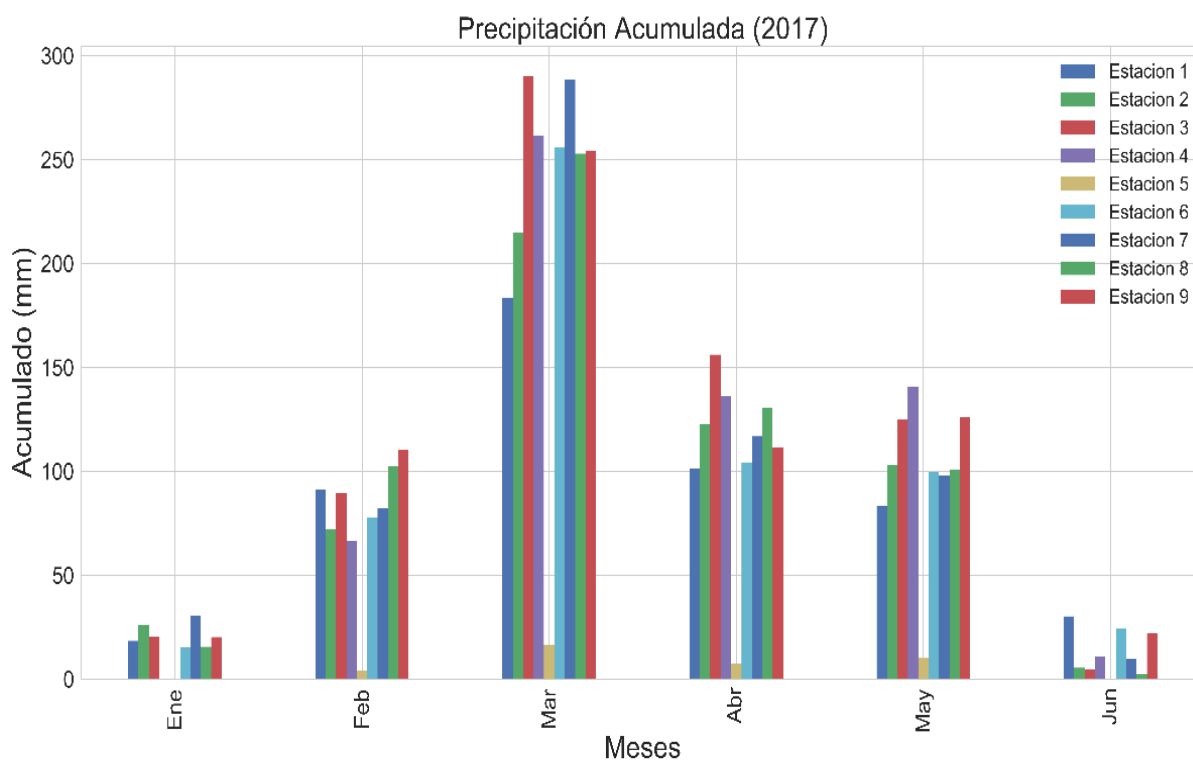
Gráfica 25: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2016 de las 9 estaciones.

En el año 2016 la estación 4 presenta mayor cantidad de máximos durante el año al igual que el año 2015 y la estación 1 presente mayor cantidad de mínimos (Tabla 9).

Tabla 9: Resumen de precipitación acumulada en el año 2016

2016				
MES	MAX (mm)	ESTACIÓN	MIN (mm)	ESTACIÓN
ENERO	105,1	9	22,7	1
FEBRERO	51,6	4	15,4	1
MARZO	204,8	8	122,3	5
ABRIL	218,79	4	120,9	1
MAYO	96,89	4	41,99	6
JUNIO	93,1	2	42,3	6
JULIO	25,7	4	6,3	1
AGOSTO	17,2	4	5,4	3
SEPTIEMBRE	100	9	63,2	7
OCTUBRE	95,29	6	0	4
NOVIEMBRE	81	9	0	4
DICIEMBRE	192,09	1	0	4

La gráfica 26 indica los valores máximos y mínimos del año 2017 mensuales por estación.



Gráfica 26: Gráfica de barras mensuales acumuladas correspondientes a la precipitación del año 2017 de las 9 estaciones.

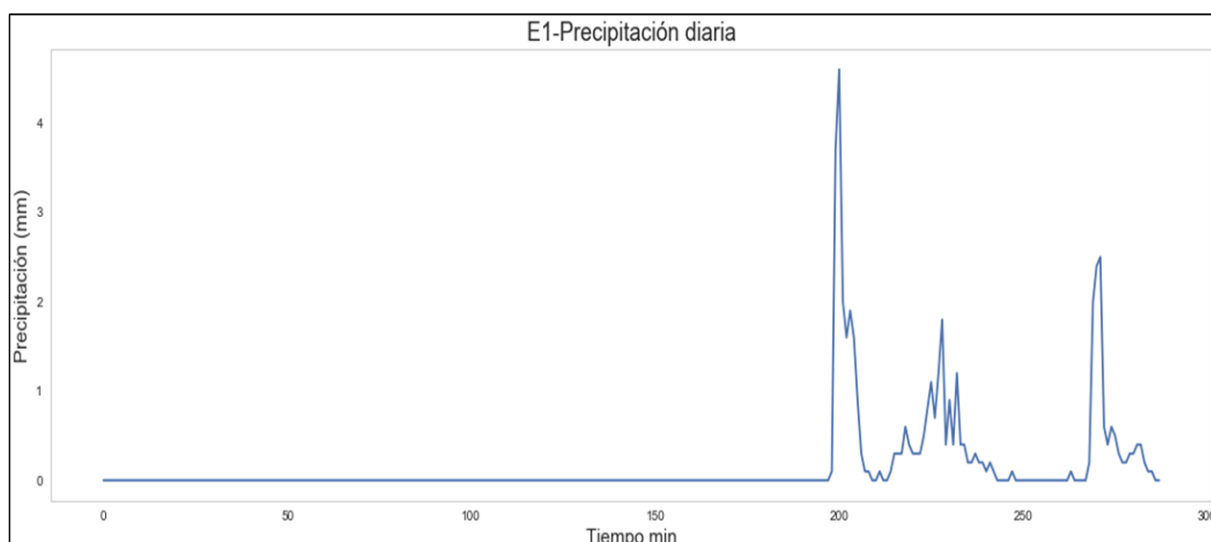
Todos los meses en el año 2017 la estación 5 han presentado valores mínimos. Entre marzo y abril la estación 3 presenta los valores máximos (Tabla 10).

Tabla 10: Resumen de precipitación acumulada en el año 2017

2017				
MES	MAX (mm)	ESTACIÓN	MIN (mm)	ESTACIÓN
ENERO	30,7	7	0	4-5
FEBRERO	110,3	9	4,2	5
MARZO	290,19	3	16,3	5
ABRIL	156,1	3	7,7	5
MAYO	140,6	4	10	5
JUNIO	30,2	1	0	5

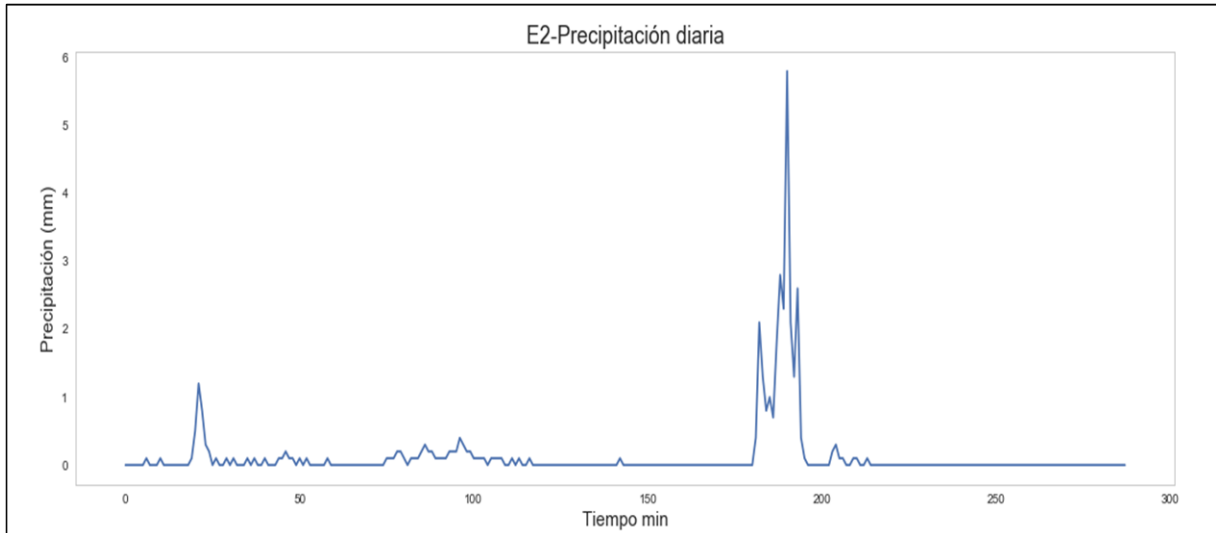
Se realizó el análisis diario de precipitación en las diferentes estaciones, donde a partir de las medias diarias de precipitación se escogieron los valores máximos, y se graficó su comportamiento diario cada 5 minutos.

En la estación 1 se encontró el máximo diario el 24 de marzo de 2015, a continuación, la gráfica del comportamiento diario. La mayor parte del día la precipitación es de 0 mm, esta incrementa aproximadamente a las 16h30 donde existe precipitación hasta de 5mm. (Gráfica 27).



Gráfica 27: Gráfica de la máxima precipitación diaria del 24 de marzo del 2015.

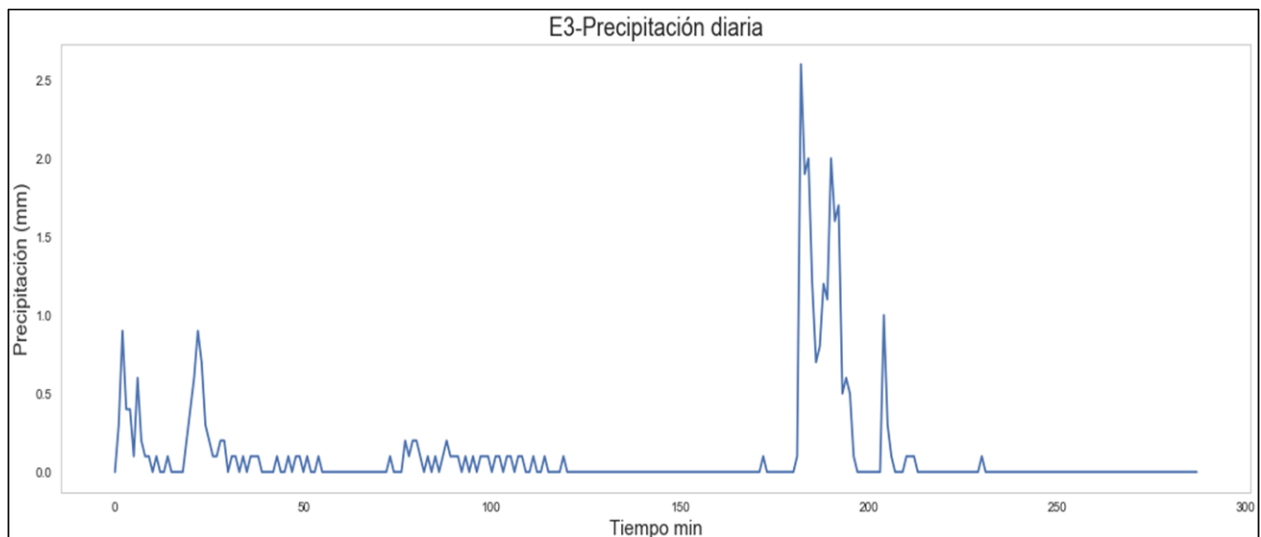
En la estación 2 el máximo diario fue 10 de marzo del 2017, en la gráfica se puede observar que la precipitación alcanza aproximadamente los 5.8 mm, aproximadamente a las 15h10 empieza a existir una variación en la precipitación, alcanzando valores altos.



(Gráfica 28)

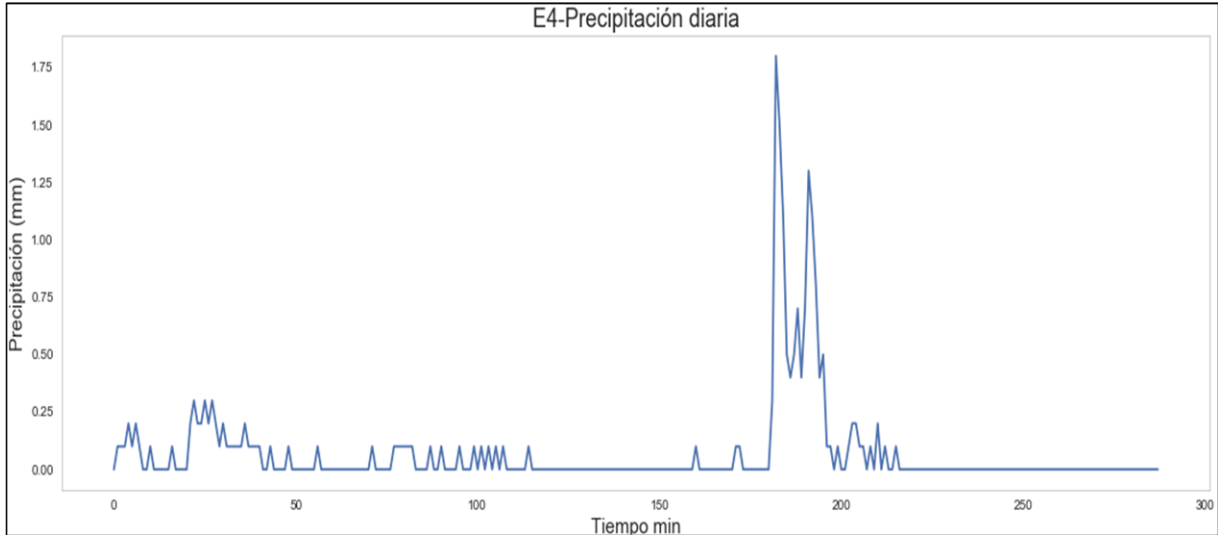
Gráfica 28: Gráfica de la máxima precipitación diaria del 10 de marzo del 2017.

La precipitación máxima registrada en la estación 3 fue el 18 de noviembre del 2015, existiendo lluvia a lo largo del día, existiendo pocos instantes sin precipitación. Alcanzando una cantidad máxima de 6.5 mm en el día. (Gráfica 29).



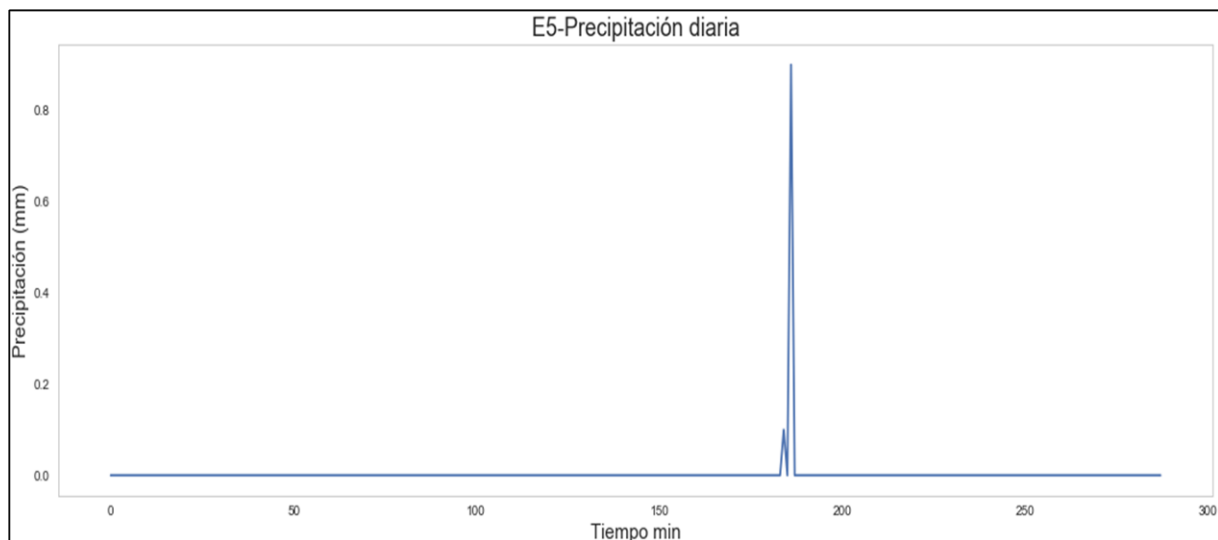
Gráfica 29: Gráfica de la máxima precipitación diaria del 18 de noviembre del 2015.

La estación 4 registro su precipitación máxima el día 23 de enero del 2015. En esta estación no se notó cambios tan fuertes en la cantidad de precipitación diaria, ya que la máxima registrada en el día fue de 2 mm. (Gráfica 30).



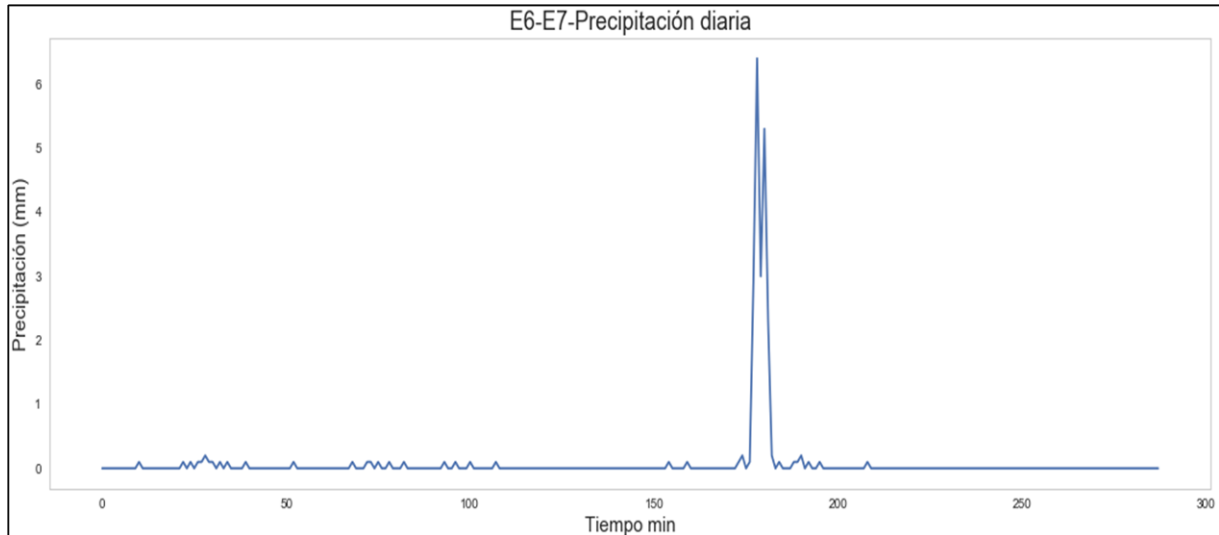
Gráfica 30: Gráfica de la máxima precipitación diaria del 18 de noviembre del 2015.

La estación 5 registro la máxima precipitación el 3 de mayo de 2015. Se puede apreciar que la estación 5 en ese día registro poca precipitación existiendo un solo pico que alcanza 1 mm. (Gráfica 31).



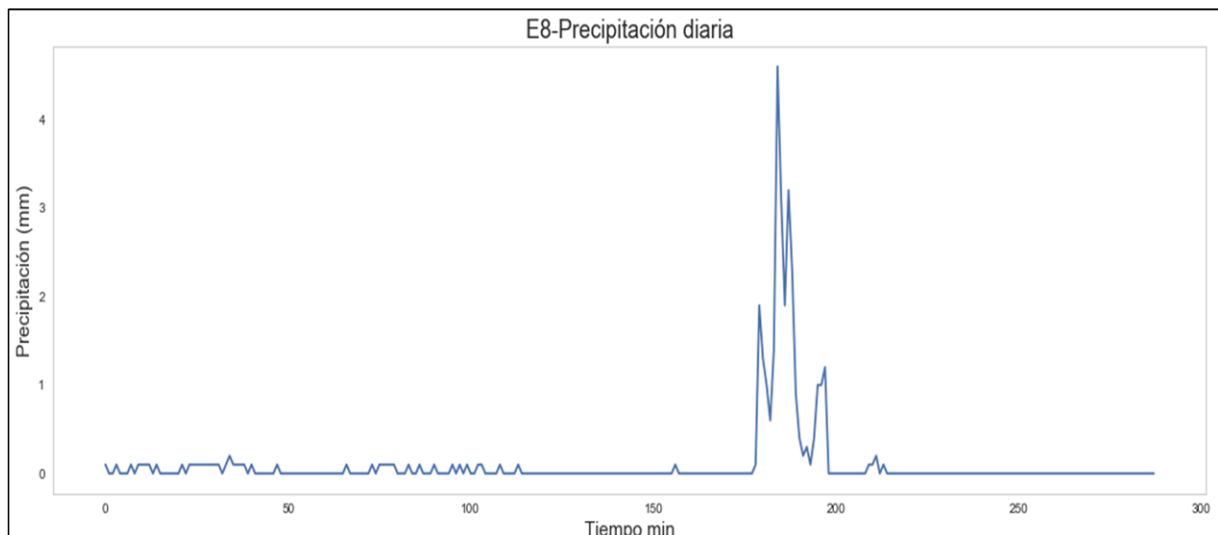
Gráfica 31: Gráfica de la máxima precipitación diaria del 3 de mayo de 2015.

La estación 6 y 7 registraron el día 13 de enero del 2016. En la siguiente grafica se puede observar que existe pequeñas cantidades de precipitación durante el día, se presenta un pico que avanza 7 mm y luego disminuye la precipitación. (Gráfica 32).



Gráfica 32: Gráfica de la máxima precipitación diaria del 13 de enero del 2016.

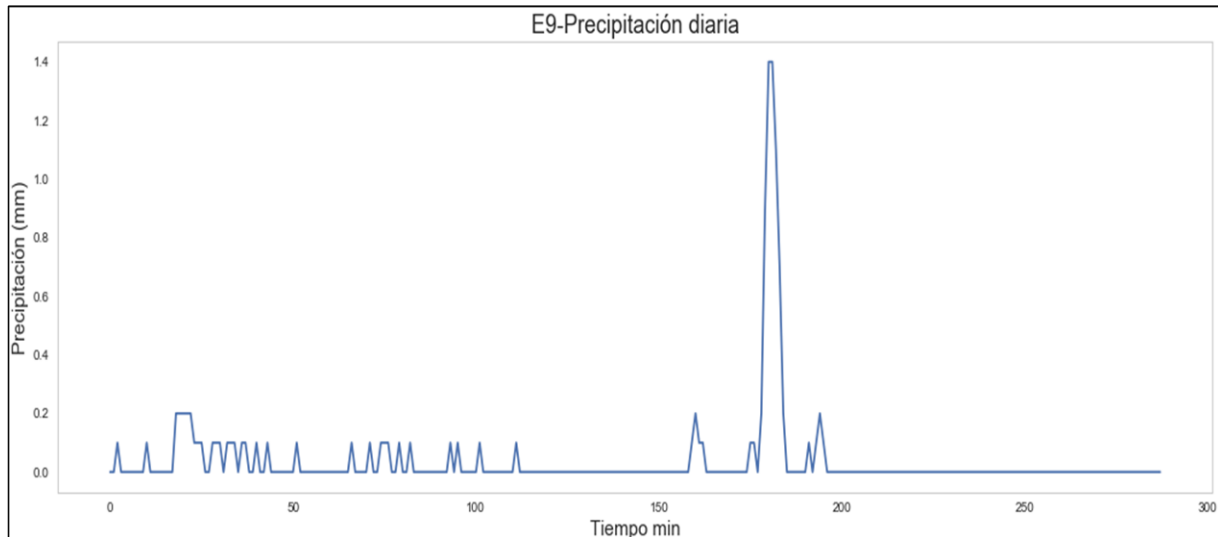
La estacion 8 representa el dia 13 de marzo de 2016 en la siguiente grafica, este dia presenta pequeñas cantidades de precipitacion durante el dia, en horas de la tarde presenta un pico que alcanza aporximadamente los 5 mm y luego disminuye la precipitacion.



(Gráfica 33).

Gráfica 33: Gráfica de la máxima precipitación diaria del 13 de marzo del 2016.

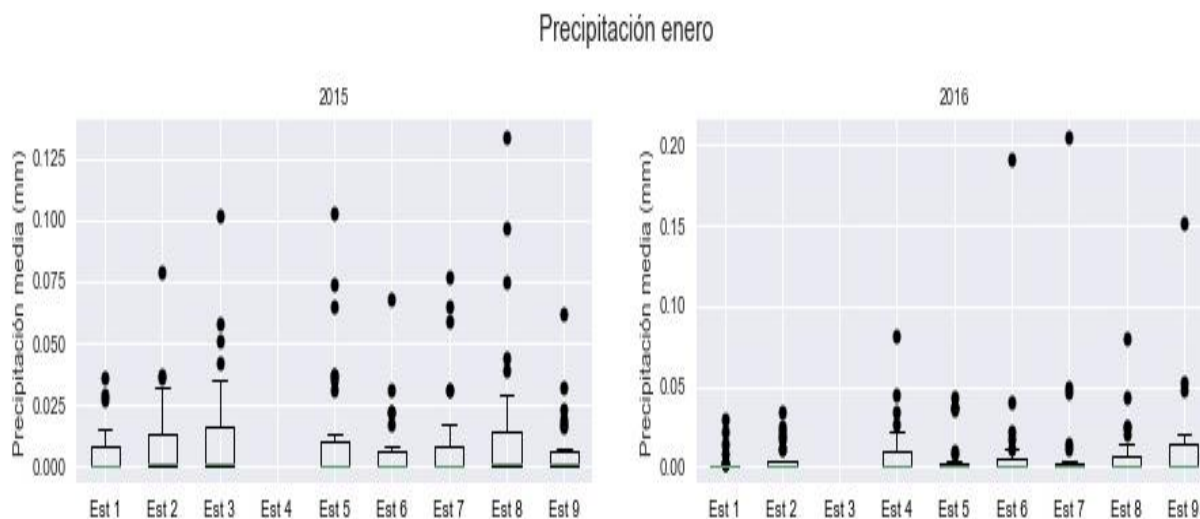
Para la estación 9 se graficó el día 29 de marzo del 2017, en este gráfico se puede analizar que existe pequeñas cantidades de precipitación durante la mañana que no sobrepasan los 0.3 mm, en la tarde se aprecia que existe un pico que alcanza los 1.4 mm, para que luego la precipitación disminuya hasta alcanzar los 0 mm. (Gráfica 34)



Gráfica 34: Gráfica de la máxima precipitación diaria del 29 de marzo del 2017

Se realizó gráficas de cajas y bigotes con los datos mensuales en todas las estaciones, presentando los resultados de los meses más relevantes. Al analizar los datos de precipitación mediante cajas se puede apreciar en la gráfica 35 que los 2 años de estudio presentan datos asimétricos, con una distribución no normal sesgada a la derecha (positiva), en el conjunto de datos existe poca variabilidad con una curtosis leptocúrtica y los datos muy similares.

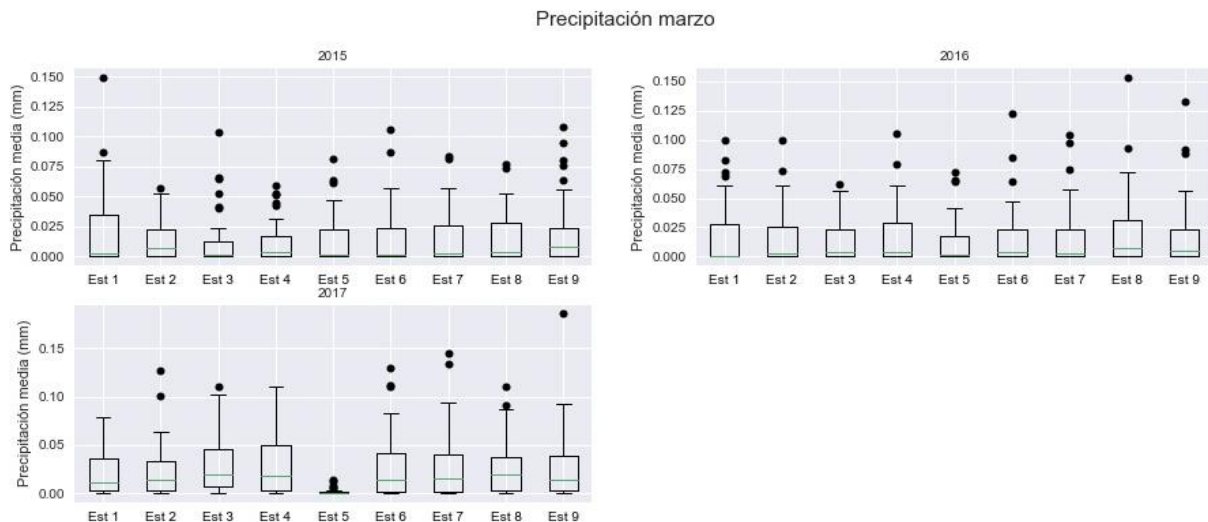
Se aprecia que en el año 2015 existió mayor variabilidad de los datos con respecto al 2016, a excepción de la estación 4 en el 2015 y la estación 3 en el 2016 que no existen datos.



Gráfica 35: Gráfico de cajas y bigotes del mes de enero correspondiente a la precipitación media diaria en los años 2015 y 2016 en las 9 estaciones de estudio.

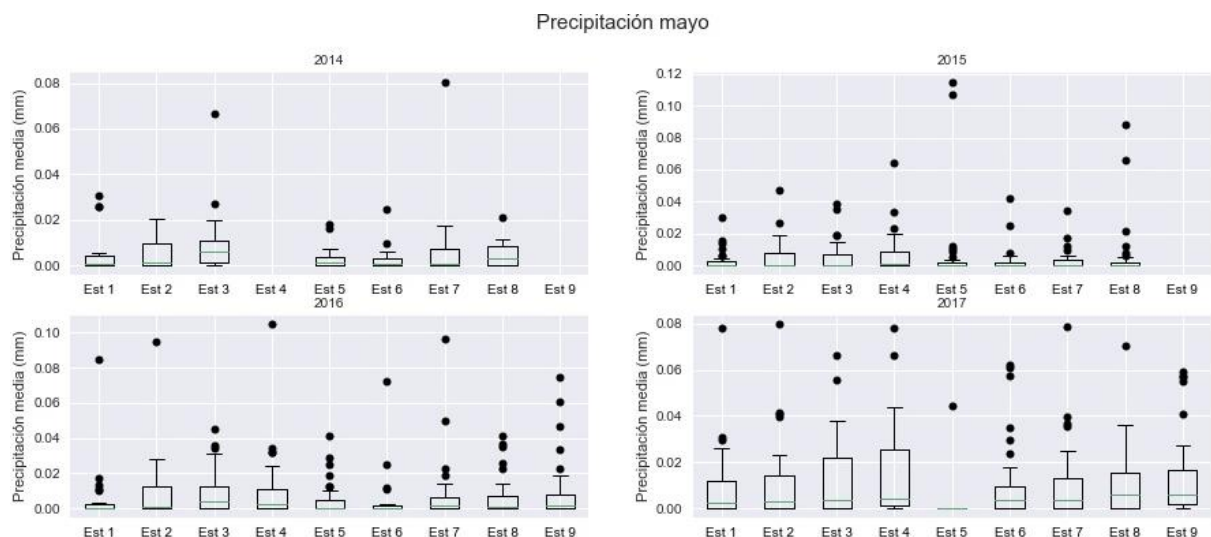
En la gráfica 36 se aprecia los datos del mes de marzo en los años 2015, 2016, 2017 donde se observa que en este mes los datos se encuentran dispersos con mayor variabilidad, una distribución no normal sesgada a la derecha y una curtosis platicúrtica. La estación 5 en el año 2017 presenta varios datos atípicos y poca variabilidad de datos.

En el mes de marzo se aprecian datos altos de precipitación en todos los años de estudio. La estación 5 en el año 2017 presentan valores mínimos, donde no se aprecia precipitación.



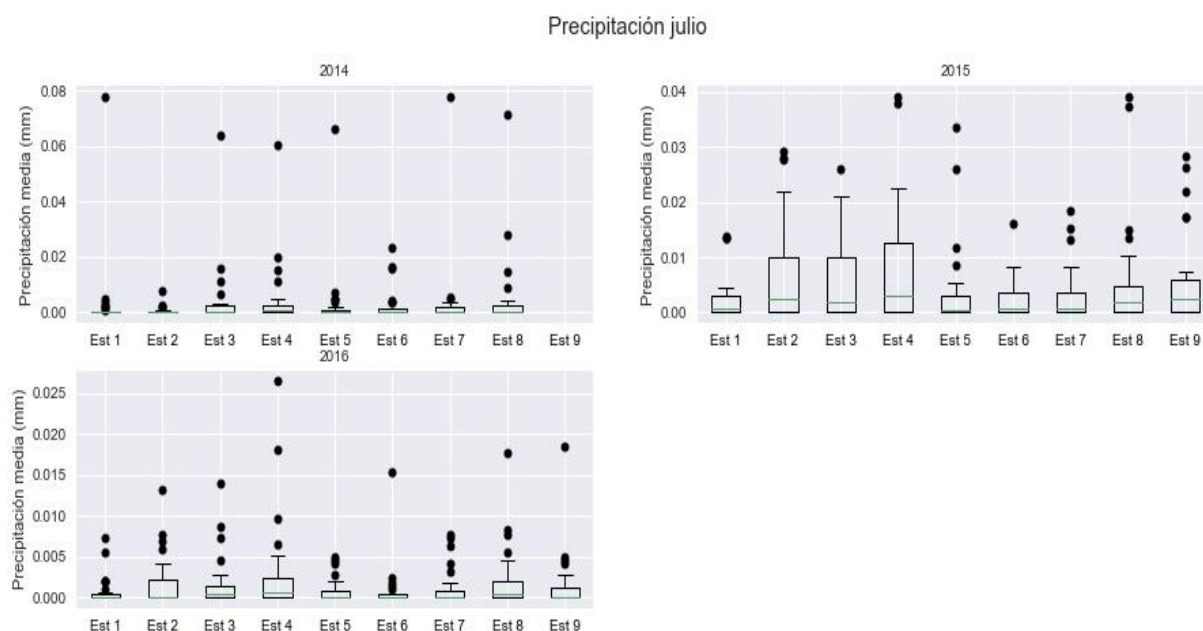
Gráfica 36: Gráfico de cajas y bigotes del mes de marzo correspondiente a la precipitación media diaria en los años 2015, 2016 y 2017 en las 9 estaciones de estudio.

En la gráfica 37 se aprecia el grafico de caja y bigotes correspondiente al mes de mayo donde se observó que existe mayor variabilidad y una distribución platicúrtica en el año 2017. La estación 5 presentan valores de precipitación mínimos. Se presenta mayor cantidad de datos anómalos en los años 2016 y 2017. En todos los años se presenta una distribución no normal sesgada a la derecha. Los años 2014, 2015 y 2016 presentan una curtosis leptocúrtica. En el año 2014 y 2015 no existen datos en la estación 9 y en el 2014 en la estación 4 en el mes de mayo.



Gráfica 37: Gráfico de cajas y bigotes del mes de mayo correspondiente a la precipitación media diaria en los años 2014, 2015, 2016 y 2017 en las 9 estaciones de estudio.

En la gráfica 38 se muestra el grafico de caja y bigotes para el mes de julio, donde se observa que en el año 2015 existe mayor variabilidad y dispersión de los datos con una curtosis de tipo platicúrtica, indicando mayor precipitación en ese año. Todos los años presentan una distribución no normal sesgada a la derecha. El año 2014 y 2016 presenta



una curtosis de tipo leptocúrtica.

Gráfica 38: Gráfico de cajas y bigotes del mes de julio correspondiente a la precipitación media diaria en los años 2014, 2015 y 2016 en las 9 estaciones de estudio.

Se realizó histogramas con los valores diarios acumulados de precipitación para analizar los días de lluvia en cada mes en las 9 estaciones de los años 2014, 2015, 2016 y 2017.

En la gráfica 39 y 40 se aprecia que los meses de marzo y abril respectivamente son los periodos más lluviosos. Las gráficas nos indican una distribución no normal de los datos con una asimetría hacia la derecha (positiva). En el mes de marzo se contabilizó 6 días con valores de 0 a 1 mm.



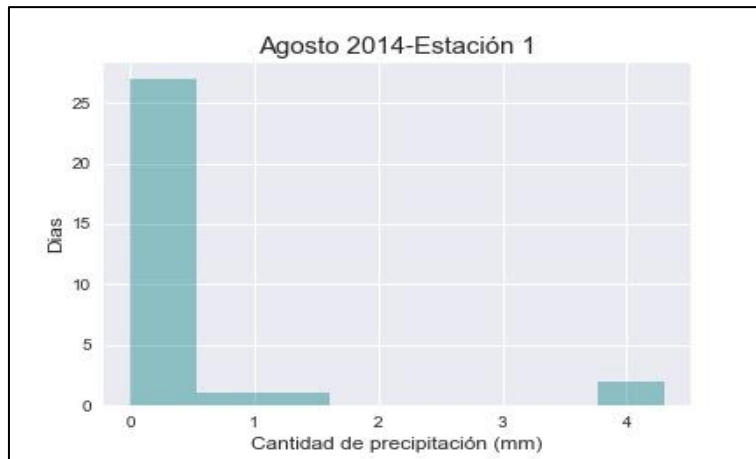
Gráfica 39: Histograma de precipitación de marzo del 2017 en la estación 3

En la estación 4 en el mes de abril se contabilizó, 9 días de precipitación entre 0 a 1 mm.



Gráfica 40: Histograma de precipitación de abril del 2016 en la estación 4

En la gráfica 41 y 42 se aprecia que los meses agosto y septiembre respectivamente son los más secos. Al igual que los histogramas anteriores la distribución de los datos no son normales con una asimetría hacia la derecha. En el mes de agosto se contabilizó 27 días.



Gráfica 41: Histograma de precipitación de agosto del 2014 en la estación 1

Al analizar las precipitaciones en un rango de 0 a 1 mm se identificaron 24 días.



Gráfica 42: Histograma de precipitación de septiembre del 2015 en la estación 2

De acuerdo a los resultados obtenidos en las gráficas de histogramas los meses de marzo y abril son meses lluviosos en el cual se encuentran en la categoría de lluvias fuertes en un rango de 20-70 mm diarios. Para los meses de agosto y setiembre se puede establecer que son meses secos debido a que se encuentran en un rango de 0-5 mm diarios.

Se realizó el análisis de la media anual con respecto a la media mensual de los 4 años de estudio obteniendo la siguiente información.

En la gráfica 43 se observan los resultados del año 2014, donde se observa que las medias mensuales en el año 2014 tienen un comportamiento diferente en todas las estaciones.



Gráfica 43: Comparación de medias mensuales y anual en el año 2014

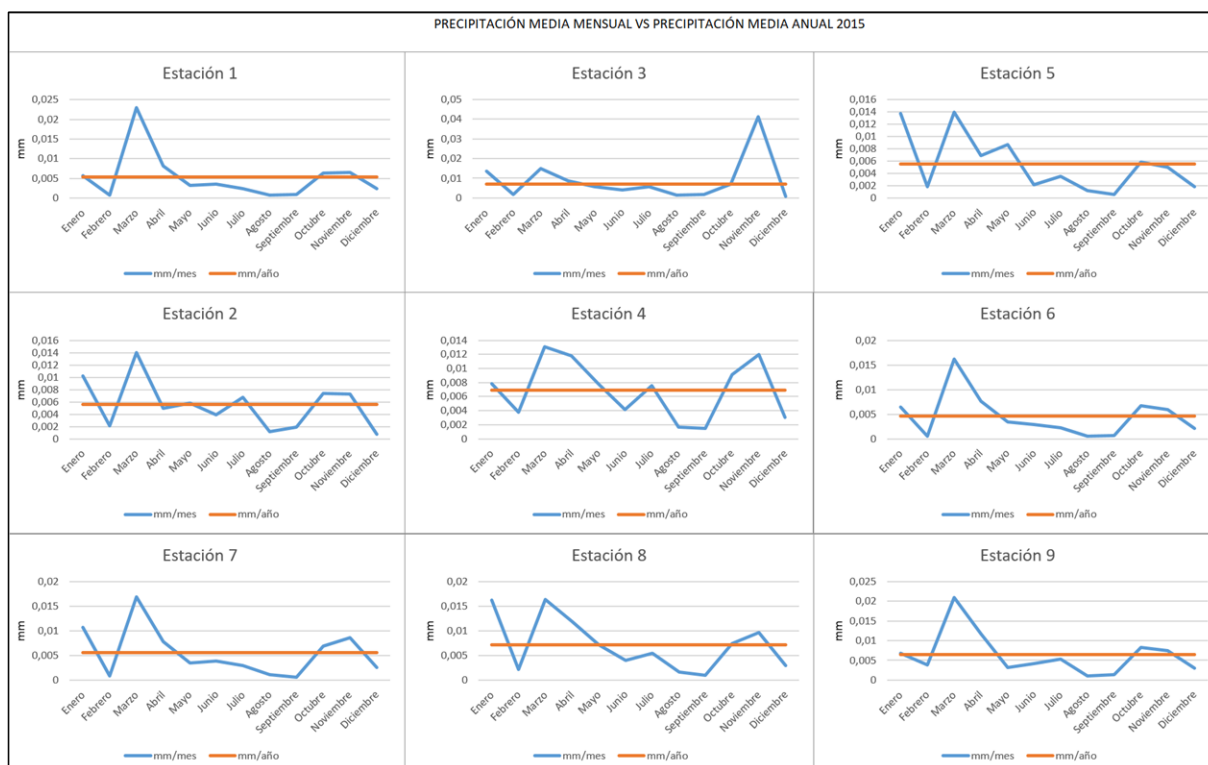
Se ha señalado con **negrita** los valores de la media anual de cada estación y se han comparado con los valores de las medias mensuales, se analizó que en el 2014 existen 27 datos que se encuentran sobre la media anual, siendo estos meses húmedos (octubre y noviembre) los cuales se encuentran en **negrita**. En el mes de agosto no se registró ningún valor sobre la media por lo que se dice que es un mes seco (Tabla 11)

Tabla 11: Media mensual vs media anual en el año 2014

ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN	ESTACIÓN
1	2	3	4	5	6	7	8	9

Media anual	<i>0,006393</i>	<i>0,003297</i>	<i>0,006936</i>	<i>0,002187</i>	<i>0,004648</i>	<i>0,006426</i>	<i>0,006115</i>	<i>0,006068</i>	<i>0,011</i>
MAYO	0,0059	<i>0,0056</i>	<i>0,0093</i>		0,0037	0,003245	<i>0,007841</i>	0,0050595	
JUNIO	0,001853	0,0025	0,006899	<i>0,004672</i>	0,003895	0,002616	0,003622	0,004747	
JULIO	0,003	0,000594	0,003909	<i>0,0044</i>	0,003	0,002442	0,003439	0,004694	
AGOSTO	0,0016	0,000672	0,001871	0,002095	0,0015	0,001478	0,001815	0,002083	0
SEPTIEMBRE	0,0022	<i>0,006713</i>	0,006377	0,000532	0,0038	0,002407	0,002361	0,005428	0,00487
OCTUBRE	<i>0,011</i>	0,000582	<i>0,13407</i>	<i>0,0034</i>	<i>0,0099</i>	<i>0,0099</i>	<i>0,0104</i>	<i>0,012082</i>	<i>0,011</i>
NOVIEMBRE	<i>0,0109</i>	<i>0,008</i>	<i>0,0101</i>	0,0011	<i>0,0098</i>	<i>0,01331</i>	<i>0,009618</i>	<i>0,009653</i>	<i>0,012</i>
DICIEMBRE	<i>0,0109</i>	0,0027	0,004536	0,000549	0,00193	<i>0,01467</i>	<i>0,0103</i>	0,003472	<i>0,015</i>

En el año 2015 se visualiza que los picos del grafico se encuentran en el mismo mes, las gráficas de las estaciones 1, 6, 7 y 9 presentan formas muy similares (Gráfica 44).



Gráfica 44: Comparación de medias mensuales y anual en el año 2015

Los valores de la media anual de cada estación se han señalado con negrita y cursiva en la tabla 12 y se han comparado con los valores de las medias mensuales, se analizó que en el 2015 existen 47 datos que se encuentran sobre la media anual, considerados como meses húmedos (enero, marzo, abril, octubre y noviembre) los cuales se encuentran en

negrita. Los meses de febrero, junio, agosto y septiembre no se registró ningún dato sobre la media por lo que se han considerado como meses secos.

	ESTACIÓN 1	ESTACIÓN 2	ESTACIÓN 3	ESTACIÓN 4	ESTACIÓN 5	ESTACIÓN 6	ESTACIÓN 7	ESTACIÓN 8	ESTACIÓN 9
Media anual	0,005338	0,005591	0,007038	0,006948	0,005482	0,0046	0,005556	0,007205	0,0064
ENERO	0,0057	0,01	0,01369	0,0078	0,01379	0,006471	0,010685	0,016226	0,0068
FEBRERO	0,000794	0,0021	0,001885	0,003757	0,001835	0,000546	0,000856	0,002183	0,00378
MARZO	0,0229	0,14	0,0148	0,013	0,01399	0,01615	0,0168	0,016342	0,0209
ABRIL	0,0081	0,0049	0,0086	0,011759	0,006944	0,0076	0,0078	0,011933	0,011
MAYO	0,00317	0,0058	0,0058	0,007796	0,008681	0,0034	0,00342	0,00718	0,0032
JUNIO	0,003542	0,0038	0,0041	0,004178	0,002164	0,0029	0,003831	0,004	0,0042
JULIO	0,002352	0,0067	0,0056	0,007549	0,003528	0,0022	0,002957	0,00538	0,00539
AGOSTO	0,000806	0,0011	0,00141	0,001669	0,001177	0,000526	0,001	0,0016	0,001
SEPTIEMBRE	0,00081	0,0018	0,001725	0,001447	0,000532	0,000613	0,000625	0,000926	0,0013
OCTUBRE	0,00634	0,0073	0,0072	0,009151	0,005847	0,006676	0,006832	0,007471	0,0082
NOVIEMBRE	0,006562	0,0072	0,04133	0,012	0,004965	0,0059	0,0086	0,00964	0,00747
DICIEMBRE	0,002341	0,000744	0,000891	0,003	0,001815	0,00205	0,002554	0,002968	0,003

Tabla 12: Media mensual vs media anual en el año 2015

En el año 2016 se visualiza que los picos del grafico se encuentran en el mismo mes, las gráficas de las estaciones 4 y 5 presentan formas muy similares al igual que las estaciones 2 y 3 (Gráfica 45).



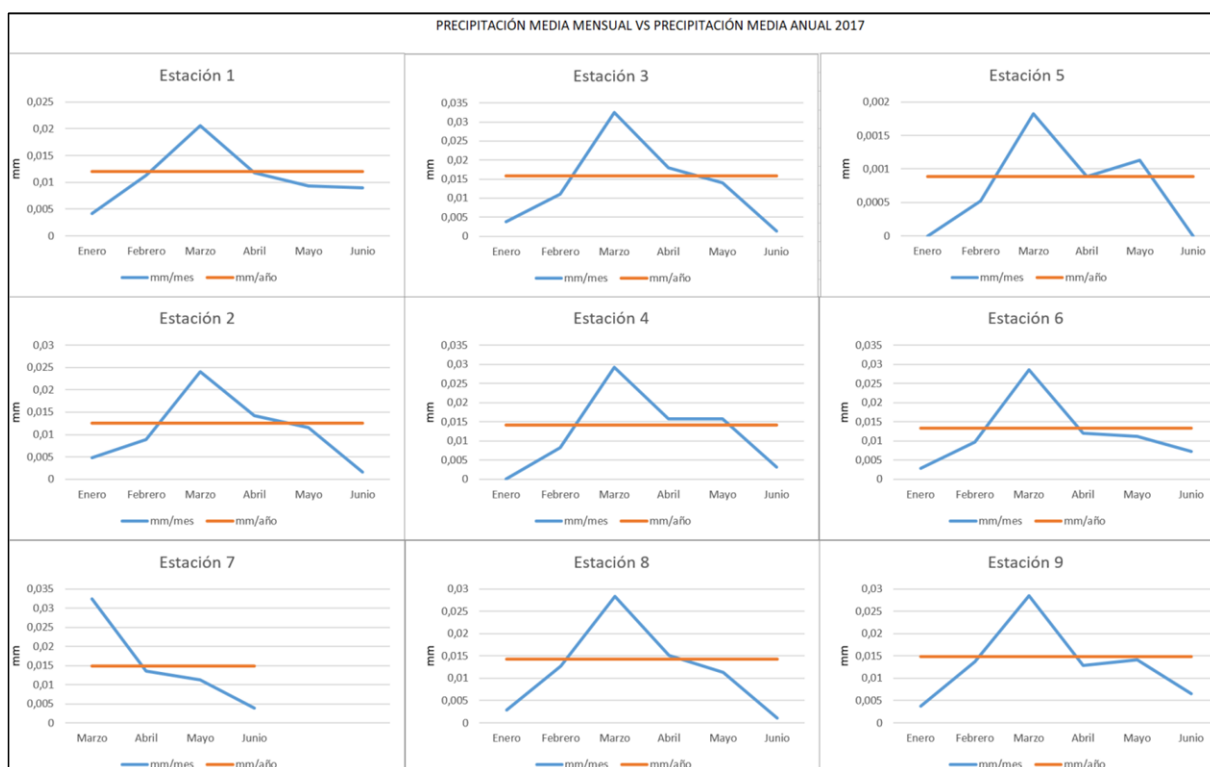
Gráfica 45: Comparación de medias mensuales y anual en el año 2016

Los valores de la media anual de cada estación se han señalado con **negrita y cursiva** en la tabla 13 y se han comparado con los valores de las medias mensuales, se analizó que en el 2016 existen 44 valores que sobrepasan la media anual, considerados como meses húmedos (marzo y abril) los cuales se encuentran en **negrita**. Los meses de febrero, julio y agosto no se registró ningún dato sobre la media por lo que se han considerado como meses secos.

Tabla 13: Media mensual vs media anual en el año 2016

	ESTACIÓN 1	ESTACIÓN 2	ESTACIÓN 3	ESTACIÓN 4	ESTACIÓN 5	ESTACIÓN 6	ESTACIÓN 7	ESTACIÓN 8	ESTACIÓN 9
Media anual	0,007753	0,0075	0,0082	0,007663	0,00489	0,008023	0,008582	0,008925	0,01
ENERO	0,00257	0,005	0,005652	0,008826	0,0058	0,010297	0,01096	0,00777	0,0117
FEBRERO	0,001844	0,00044	0,005089	0,0061	0,0025	0,002335	0,0023	0,0053	0,0045
MARZO	0,0187	0,017	0,017	0,019	0,013698	0,01771	0,0188	0,0229	0,0198
ABRIL	0,014	0,014	0,016	0,025	0,01544	0,0142	0,016	0,0209	0,0199
MAYO	0,005	0,009151	0,009	0,01	0,0054	0,0052	0,0083	0,0071	0,0103
JUNIO	0,0049	0,01	0,01	0,0087	0,0046	0,0048	0,0053	0,0071	0,007
JULIO	0,0007	0,0017	0,0015	0,0028	0,000739	0,00075	0,001	0,002	0,00145
AGOSTO	0,001	0,00131	0,001	0,0019	0,000862	0,001	0,0011	0,0015	0,00116
SEPTIEMBRE	0,0069	0,007721	0,0058	0,0075	0,007025	0,0079	0,0073	0,00805	0,0115
OCTUBRE	0,0076	0,0037	0,0048	0	0,001827	0,0106	0,006922	0,00746	0,0095
NOVIEMBRE	0,007	0,0048	0,01	0	0,00037	0,005	0,006331	0,0065	0,0093
DICIEMBRE	0,0215	0,009297	0,0084	0	0,000403	0,015	0,0174	0,010237	0,017

En el año 2017 se observa que los picos del grafico se encuentran en el mismo mes, las gráficas de las estaciones 2,3,4 y 8 presentan formas muy similares al igual que las



estaciones 6 y 7, y las estaciones 5 y 9 (Gráfica 46).

Gráfica 46: Comparación de medias mensuales y anual en el año 2017

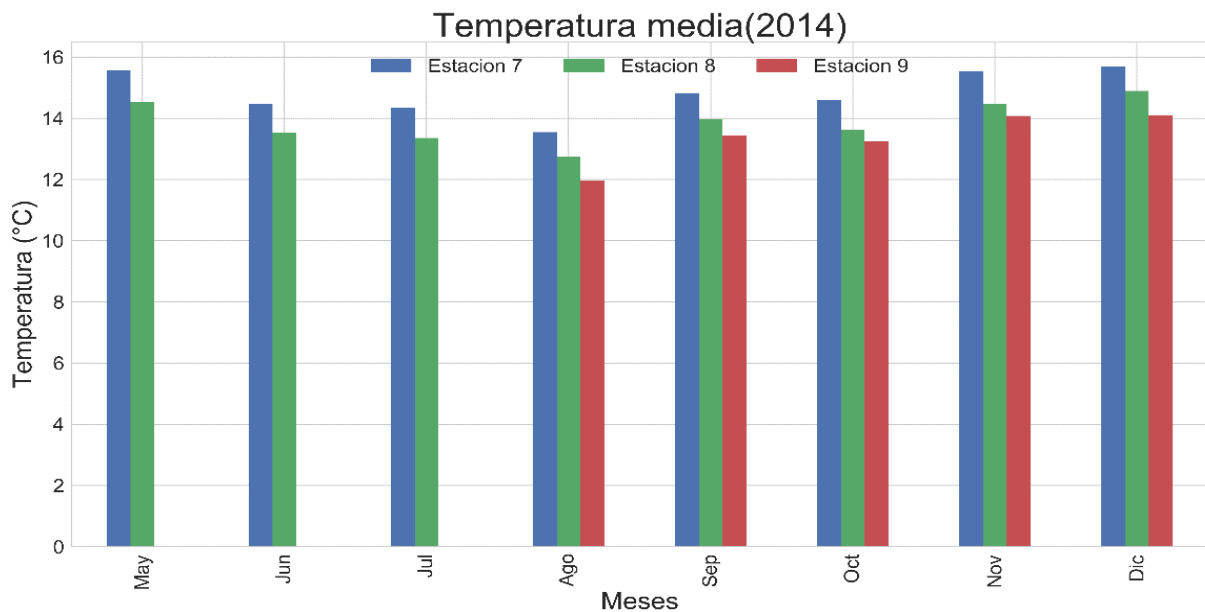
Los valores de la media anual de cada estación se han señalado con negrita y cursiva en la tabla 14 y se han comparado con los valores de las medias mensuales, se analizó que en el 2017 existen 21 valores que sobrepasan la media anual, considerados como meses húmedos (marzo) los cuales se han señalado en negrita. Los meses de enero, febrero, y junio no se registró ningún dato sobre la media por lo que se han considerado como meses secos.

Tabla 14: Media mensual vs media anual en el año 2017

	ESTACIÓ N 1	ESTACIÓ N 2	ESTACIÓ N 3	ESTACIÓ 4	ESTACIÓ N 5	ESTACIÓ N 6	ESTACIÓ N 7	ESTACIÓ 8	ESTACIÓ 9
Media anual	<i>0,011735</i>	<i>0,01254</i>	<i>0,01584</i>	<i>0,014209</i>	<i>0,000883</i>	<i>0,0133336</i>	<i>0,01482</i>	<i>0,014353</i>	<i>0,0148</i>
ENERO	0,0034	0,0038	0,0038	0	0	0,002875	0,005696	0,00289	0,0037
FEBRERO	0,011297	0,01	0,01	0,0082	0,000521	0,009635	0,0101	0,0127	0,0136
MARZO	0,02	0,032	0,032	0,029	0,001826	0,02864	0,032	0,028	0,02846
ABRIL	0,011	0,018	0,018	0,015	0,000891	0,012049	0,0135	0,015	0,01291
MAYO	0,0093	0,014	0,014	0,015	0,001128	0,011229	0,01123	0,011	0,01414
JUNIO	0,0089	0,00133	0,00133	0,0032	0	0,007243	0,003832	0,00105	0,006589

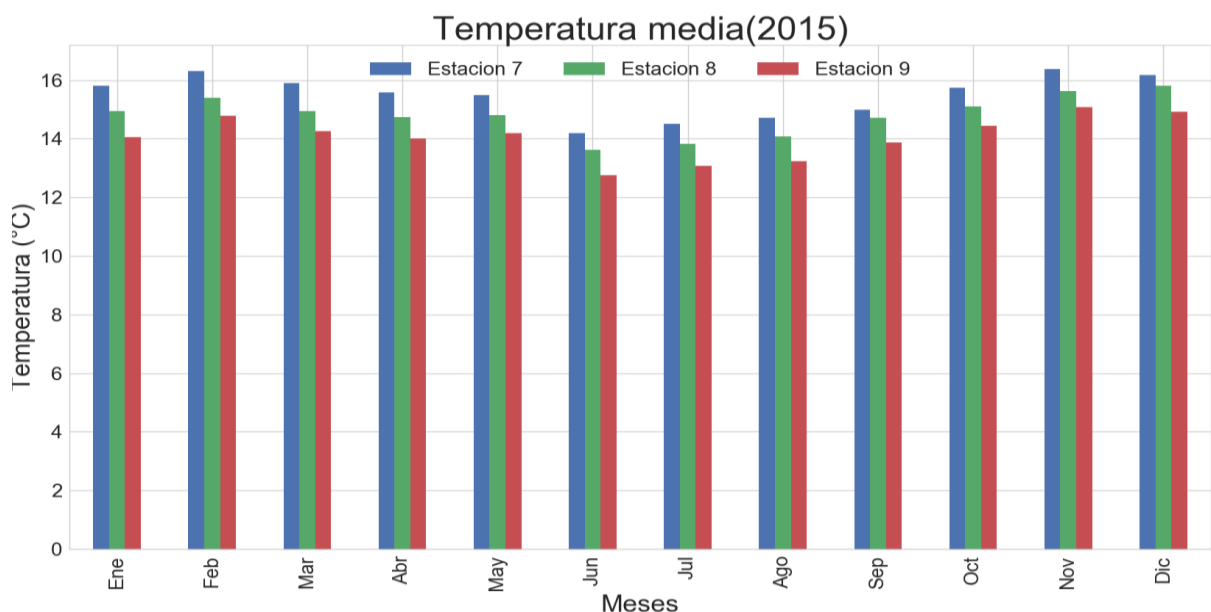
2. Temperatura

En la gráfica 47 se aprecia los valores de temperatura registrados en las 3 estaciones meteorológicas durante el año 2014, existe una carencia de datos en la estación 9 en los meses de mayo, junio y julio. Las temperaturas más altas se registraron en el mes de diciembre en las 3 estaciones con valores de 15.70°C en la estación 7, 14.89°C en la estación 8 y 14.10°C en la estación 9. Las temperaturas mínimas se registran en el mes de agosto con un total de 13.54°C en la estación 7, 12.74°C en la estación 8 y para la estación 9 un total de 11.95°C.



Gráfica 47: Medias mensuales de temperatura en el año 2014

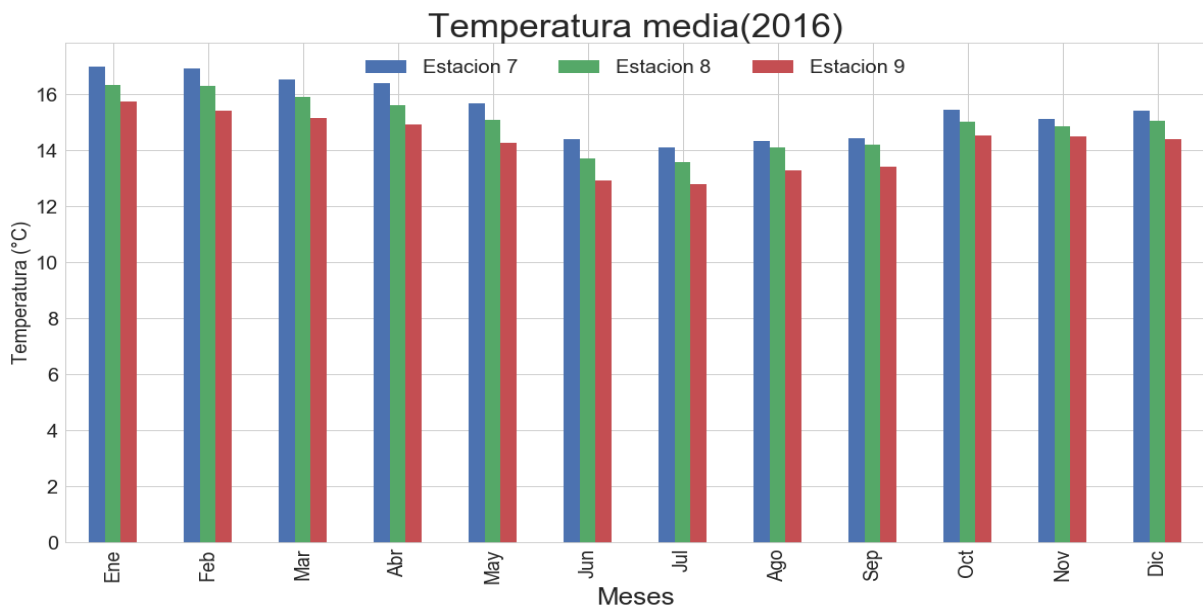
En la gráfica 46 se aprecia valores de temperatura registradas en las 3 estaciones meteorológicas durante el año 2015, donde la temperatura máxima es la estación 7 y la



temperatura mínima es la estación 9 las temperaturas mínimas. El mes de noviembre presenta la temperatura máxima con un total de 16.37°C, en el mes de diciembre se registra la máxima temperatura en la estación 8 con un total de 15.81°C y en la estación 9 en noviembre con 15.08°C. Las temperaturas mínimas en las estaciones 7,8 y 9 se registran en el mes de junio con un total de 14.19°C, 13.61°C, 12.76°C respectivamente.

Gráfica 48: Medias mensuales de temperatura en el año 2015

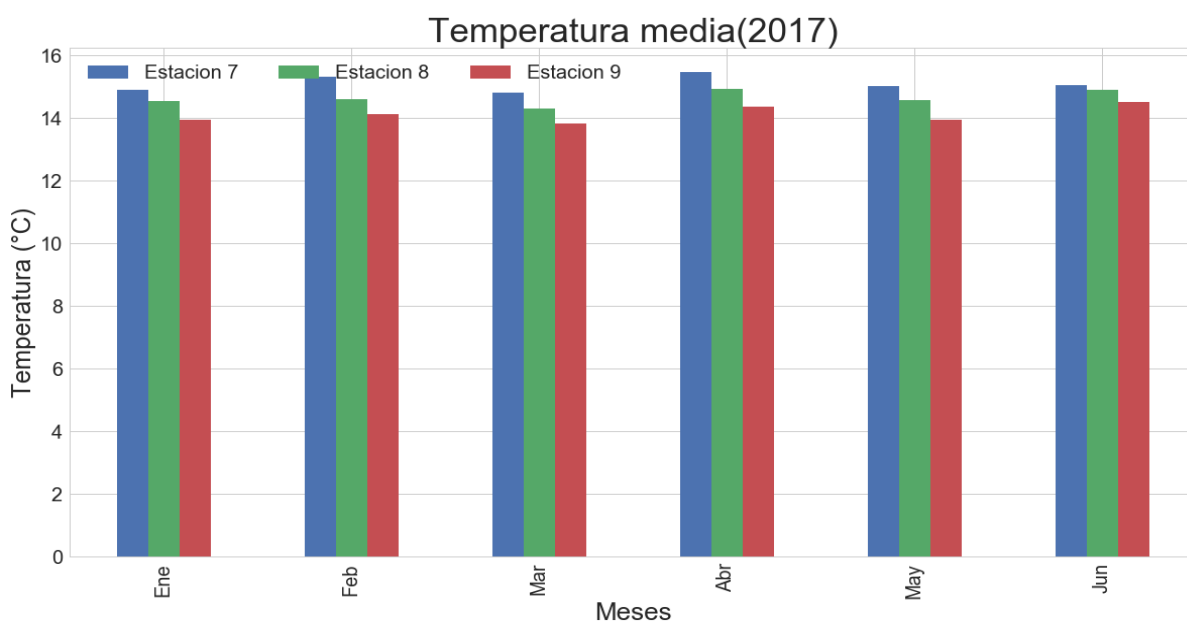
En la gráfica 49 se aparecía valores de temperatura registradas en las 3 estaciones meteorológicas durante el año 2016, las temperaturas máximas que se reflejan son en el mes de mayo con valores de 16.98°C, 16.33°C, 15.74°C para las estaciones 7,8 y 9 respectivamente. Los valores mínimos se aprecian en el mes de julio con un total de 14.11°C para la estación 7, 16.33°C en la estación 8 y 12.79°C en la estación 9.



Gráfica 49: Medias mensuales de temperatura en el año 2016

En la gráfica 50 se aprecian los valores medios de temperatura en el año 2017, en el mes de abril se registran los valores máximos en el mes de abril en la estación 7 y 8 (15.46°C y 14.93°C) y en la estación 9 en el mes de junio con un total de 14.52°C. los

valores mínimos se registran en el mes de marzo en las estaciones 7,8 y 9 (14.82°C, 14.31°C, 13.83°C). En este año solo se posee registros desde enero hasta junio.

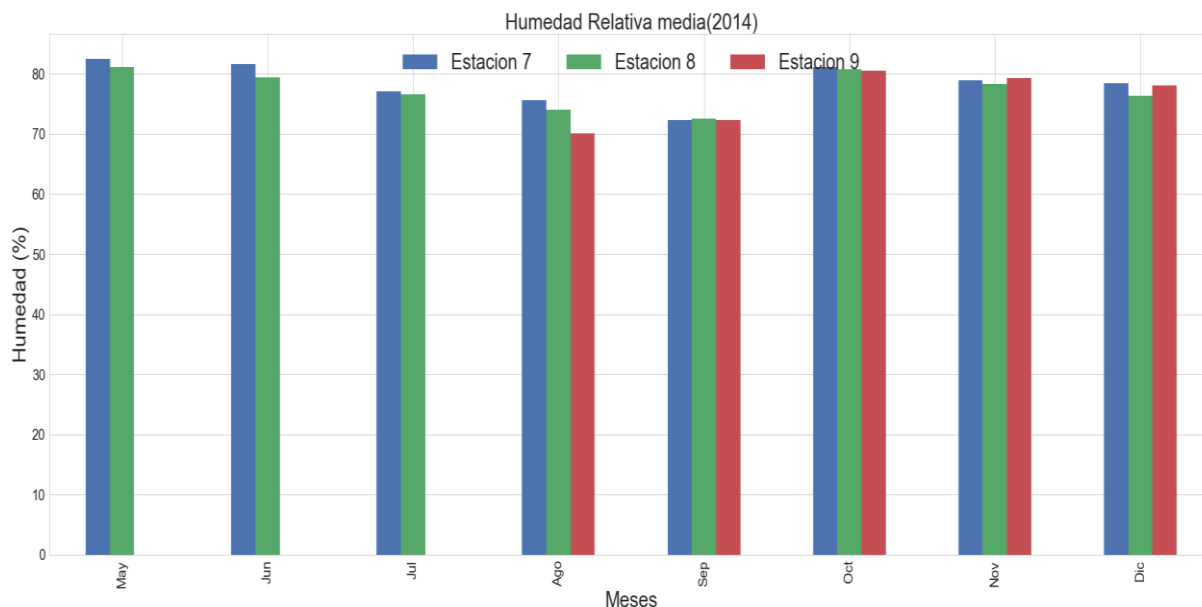


Gráfica 50: Medias mensuales de temperatura en el año 2016

3. Humedad relativa

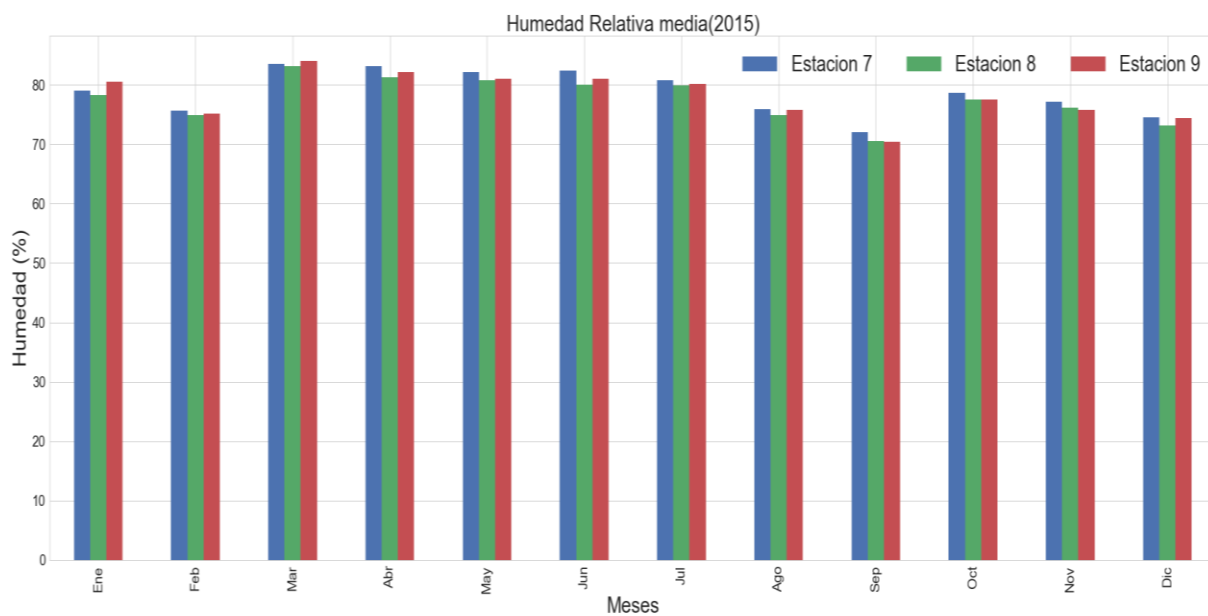
En la gráfica 51 la humedad relativa registrada en el año 2014 indica que los valores máximos se encuentran en el mes de mayo en las estaciones 7 y 8 (82.48%, 81.08%), en la estación 9 se registró un total de 80.52% en el mes de octubre. Los valores mínimos

registrados para las estaciones 7 y 8 se dan en el mes de septiembre (72.26%, 72.51%) y en el mes de agosto para la estación 9 con un valor de 70.15%. Cabe recalcar que no se cuentan con registros de los meses de mayo a julio de la estación 9.



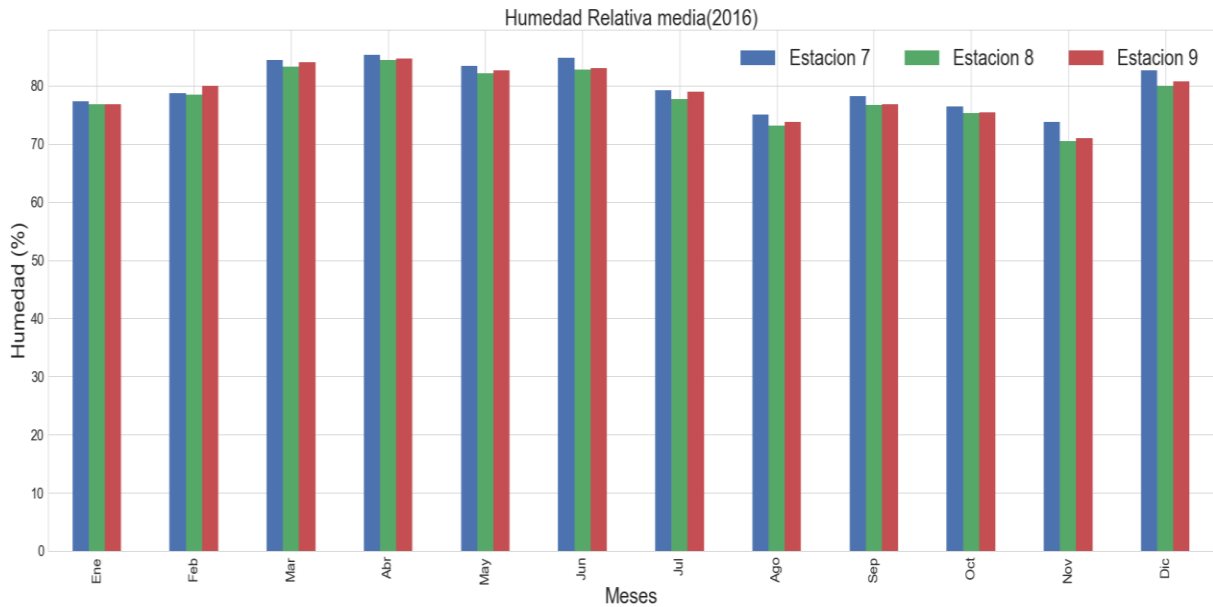
Gráfica 51: Medias mensuales de humedad relativa en el año 2014

En la gráfica 52 la humedad relativa en el año 2015 presenta los valores máximos en el mes de marzo con registros de 83.56%, 83.10%, 84.01% para las estaciones 7,8 y 9 respectivamente. Los valores mínimos se registraron en el mes de septiembre con un total de 72.06%, 70.47% y 70.45% respectivamente para las estaciones 7,8 y 9.



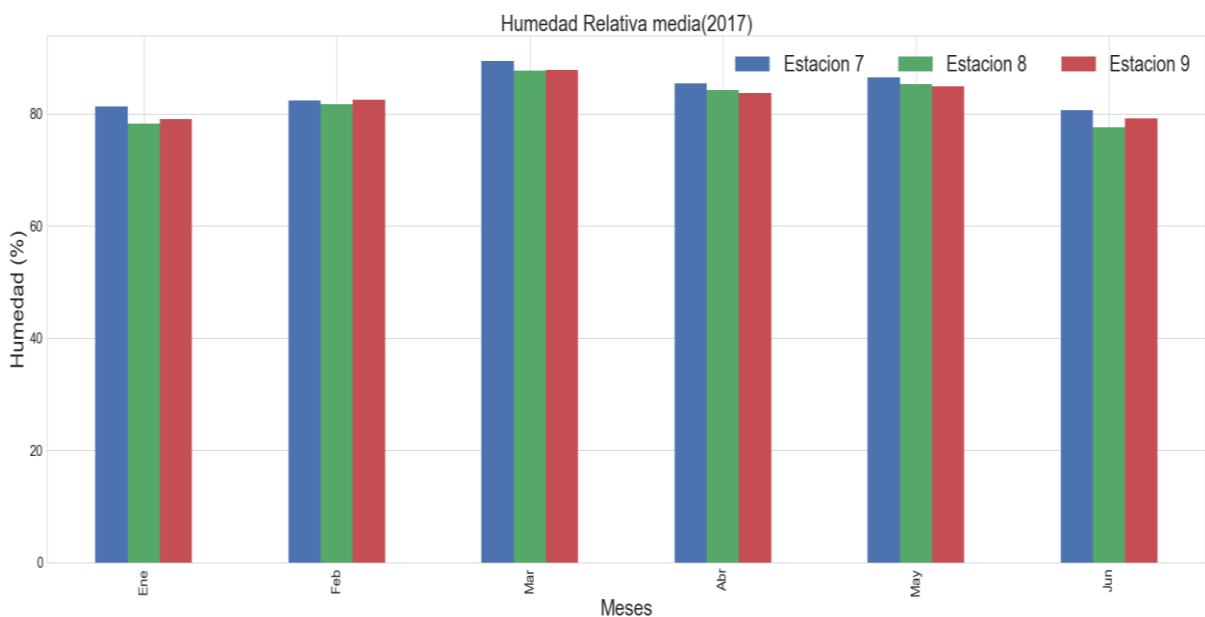
Gráfica 52: Medias mensuales de humedad relativa en el año 2015

En la gráfica 53 se observa la que la humedad relativa máxima en el año 2016 se dio en el mes de abril para la estación 7 con 85.31, estación 8 con 84.45% y la estación 9 con un total de 84.73%. La humedad relativa mínima registrada fue de 73.74%, 70.45% y 71.03% en las estaciones 7,8 y 9 respectivamente en el mes de noviembre.



Gráfica 53: Medias mensuales de humedad relativa en el año 2016

En la gráfica 54 se observa que la humedad relativa máxima del año 2017 se registró en el mes marzo con un total de 89.40%, 87.68% y 87.80% para la estación 7,8 y 9 respectivamente. Los valores mínimos se registraron en la estación 7 y 8 en el mes de junio con un total de 80.67% y 77.59% en la estación 9 se registró un total de 79.01% en el mes



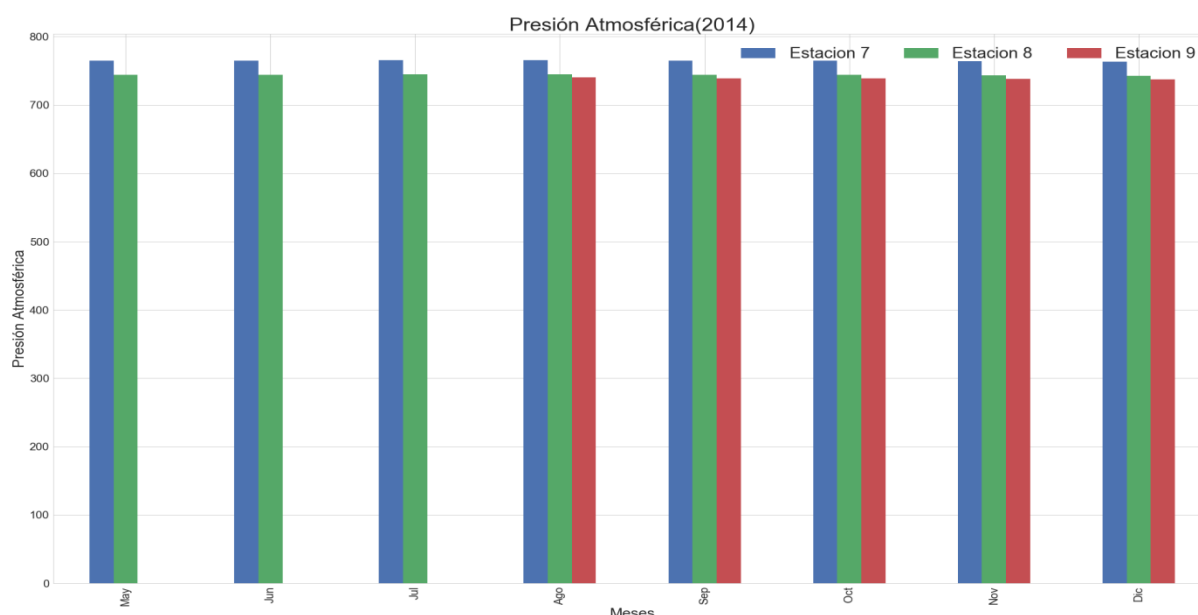
de enero. Cabe recalcar que se analizó desde el mes de enero a junio.

Gráfica 54: Medias mensuales de humedad relativa en el año 2017

Al analizar la humedad relativa se puede notar que los valores máximos se registran en el primer semestre del año a excepción del año 2014 en la estación 9, los valores mínimos se registran en el segundo semestre del año a excepción de 2017 en la estación 9.

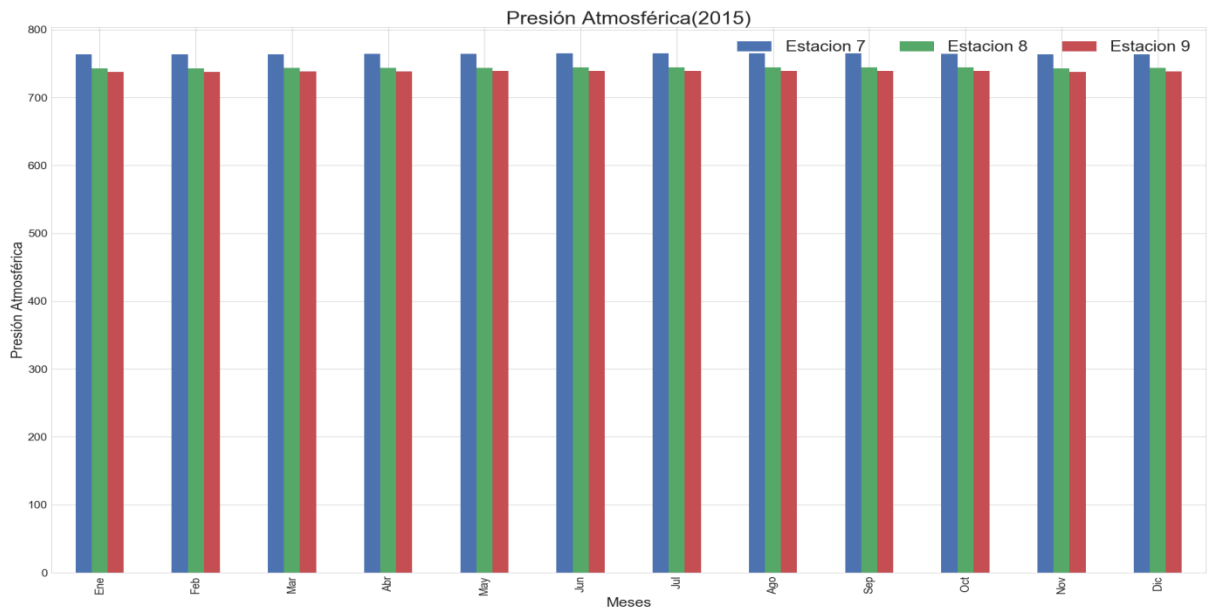
4. Presión atmosférica

En la gráfica 55 en el año 2014 la presión atmosférica máxima se registra en el mes de julio en las estaciones 7 y 8 (765.85 y 745.10), en la estación 9 se registra en el mes de agosto un total de 740.27, los valores mínimos se registran para las tres estaciones en el mes de diciembre (763.58, 742.96, 737.77) respectivamente para las estaciones 7,8 y 9.



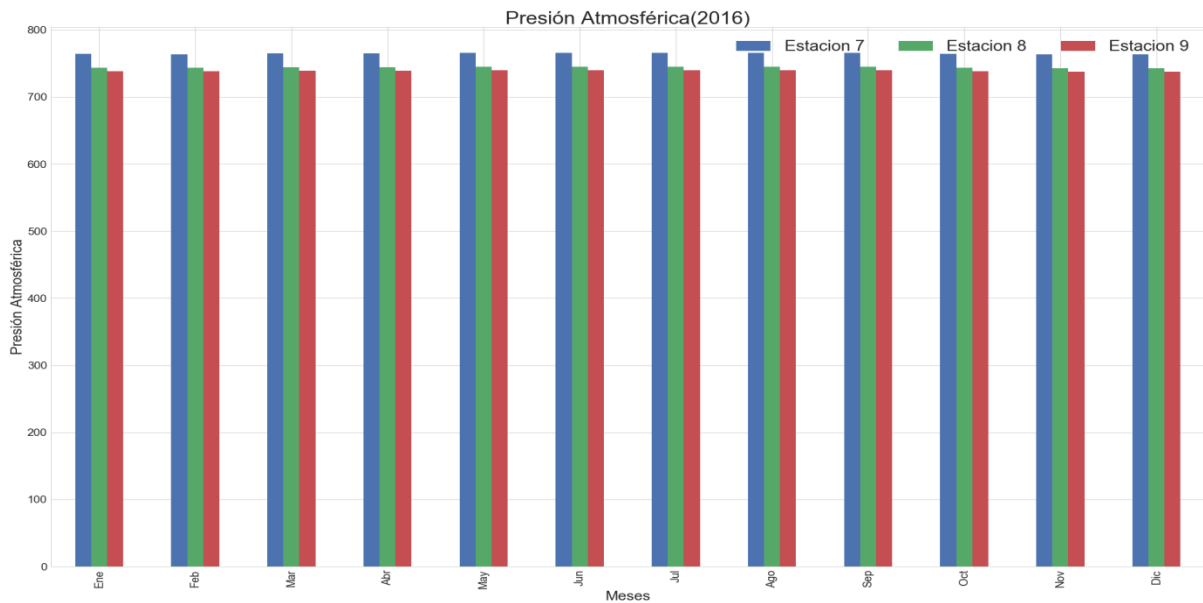
Gráfica 55: Medias mensuales de presión atmosférica en el año 2014

En la gráfica 56 en el año 2015 las tres estaciones se comportan de manera similar en sus valores máximos y mínimos. En el mes de junio se registran los valores máximos con un total de 765.55, 744.76 y 739.61. Los valores mínimos se registraron en el mes de noviembre con valores de 763.52, 742.94 y 737.78.



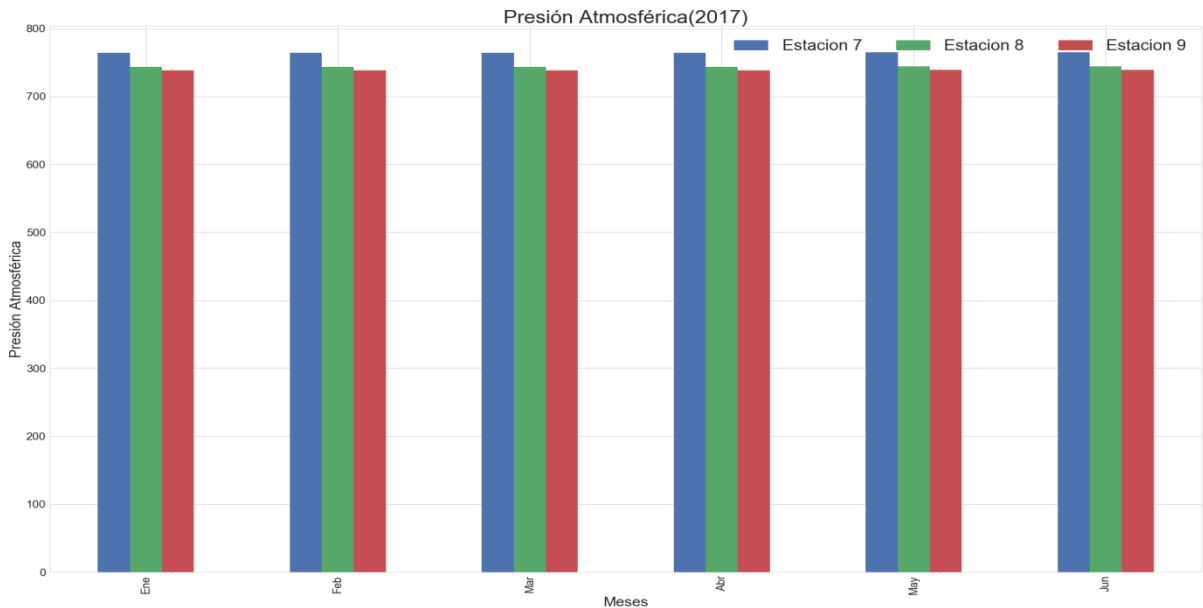
Gráfica 56: Medias mensuales de presión atmosférica en el año 2015

En la gráfica 56 en el año 2016 las estaciones se comportan de manera muy similar, las tres registran su máximo en el mes de junio con un total de 766.15, 745.35 y 748.15. En el mes de diciembre registran sus valores mínimos con 763.68 en la estación 7, 743.08 en la estación 8 y 737.86 en la estación 9.



Gráfica 57: Medias mensuales de presión atmosférica en el año 2016

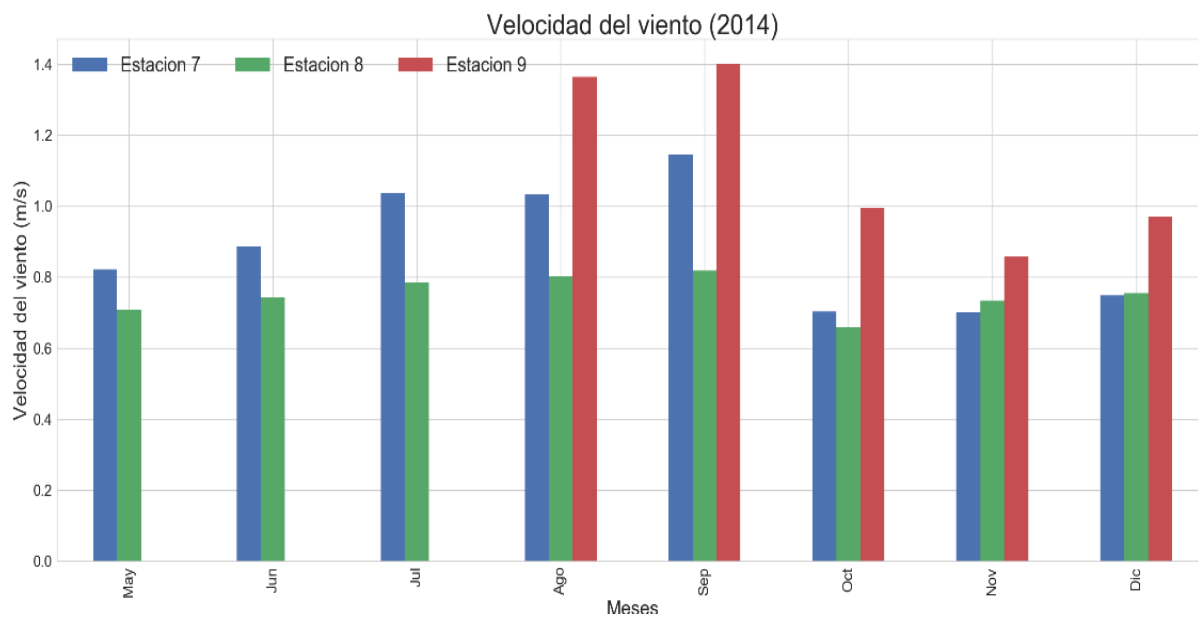
En la gráfica 58 en el año 2017 los valores máximos se registran en el mes de enero en la estación 7 y 8 con índices de 0.702 m/s y 0.742 m/s, en la estación 9 el valor máximo se registra en junio con un total de 1.036 m/s. Los valores mínimos se registran en el mes de junio en la estación 7 con valores de 0.006 m/s, en la estación 8 en el mes de mayo con un total de 0.506 m/s y en la estación 9 un total de 0.6603 en el mes de marzo.



Gráfica 58: Medias mensuales de presión atmosférica en el año 2017

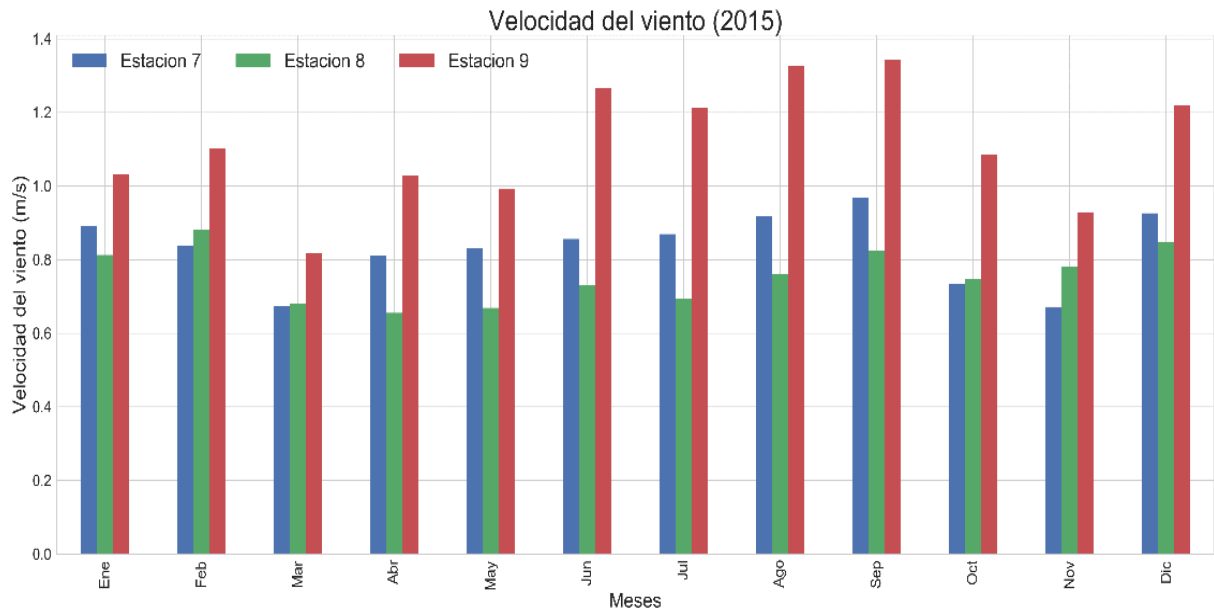
5. Velocidad del viento

En la gráfica 59 se analizó la velocidad de viento en el año 2014, no existen registros de datos de la estación 9 en los meses de mayo a junio. Se observó que los valores máximos se registraron en el mes de septiembre en las tres estaciones con valores de 1.14 m/s en la estación 7, 0.81 m/s en la estación 8 y 1.40 m/s en la estación 9. Los valores mínimos se registraron en el mes de noviembre para la estación 7 y 9 (0.701 m/s y 0.857 m/s) respectivamente en la estación 8 se registró en el mes de octubre un total de 0.658 m/s.



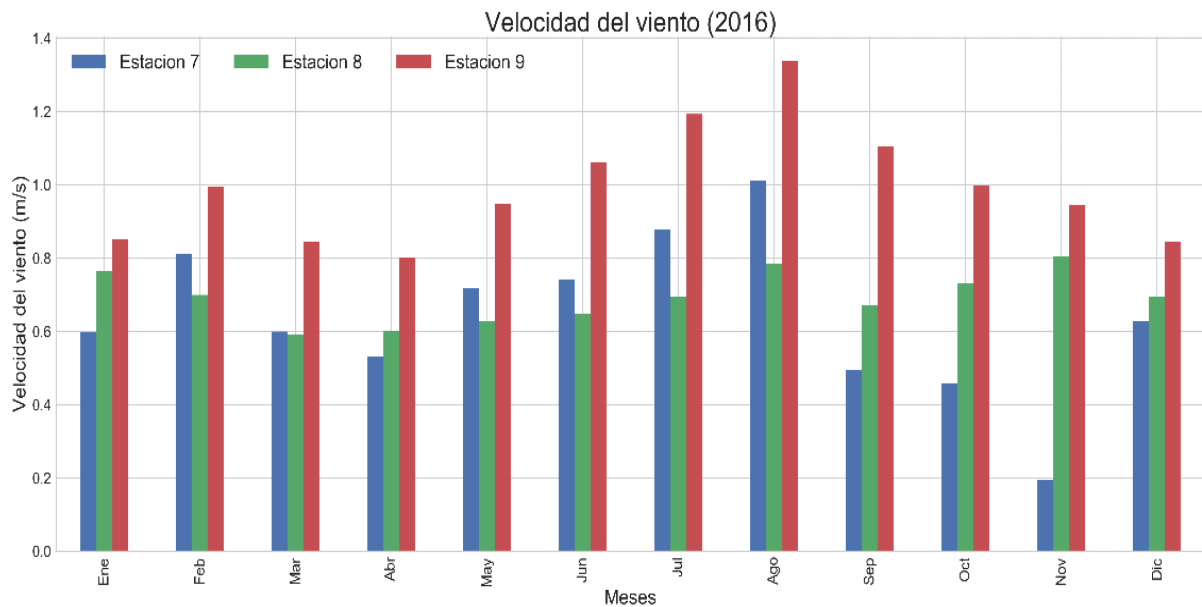
Gráfica 59: Medias mensuales de velocidad del viento en el año 2014

En la gráfica 60 en el año 2015 los valores máximos se obtuvieron en el mes de septiembre en las estaciones 7 y 9 (0.967 m/s y 1.342 m/s) en el mes de febrero se registra la velocidad media máxima en la estación 8 con un valor de 0.881 m/s. Los valores mínimos se registran en el mes de noviembre en la estación 7 (0.671 m/s), en abril en la estación 8 (0.654 m/s) en el mes de marzo un total de 0.817 m/s en la estación 9.



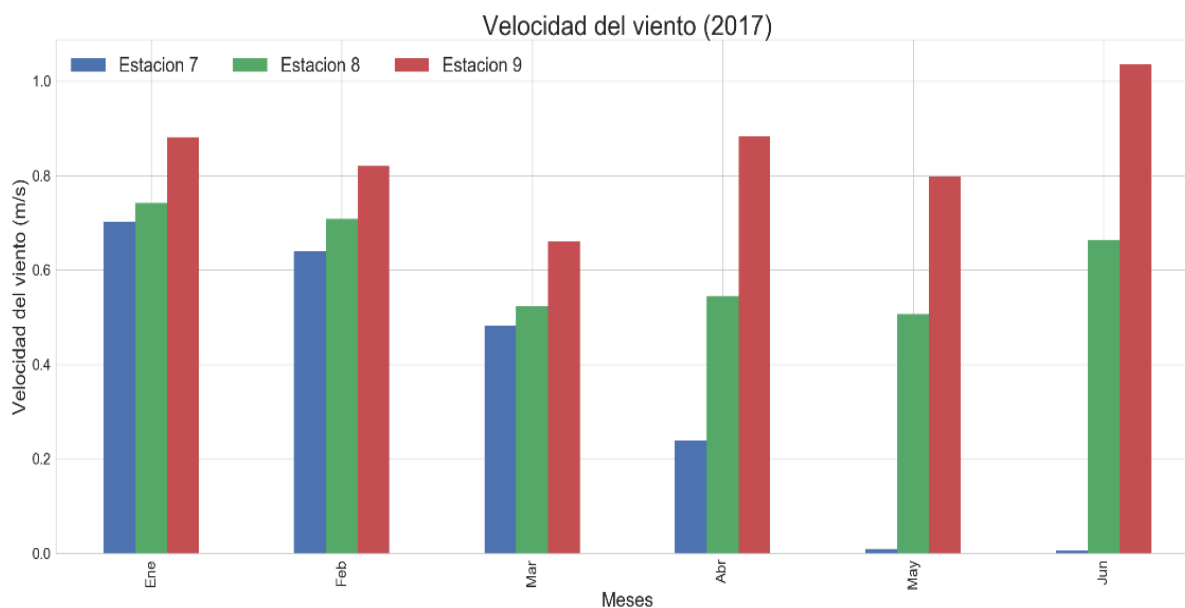
Gráfica 60: Medias mensuales de velocidad del viento en el año 2015

En la gráfica 61 en el año 2016 los valores máximos se registran en el mes de agosto en la estación 7 y 9 con valores de 1.012 m/s y 1.338 m/s respectivamente, en el mes de noviembre se registra en la estación 8 un total de 0.805 m/s. los valores mínimos corresponden a el mes de noviembre en la estación 7 con un total de 0.192 m/s, en marzo en la estación 8 con 0.590 m/s y en abril en la estación 9 con 0.802 m/s.



Gráfica 61: Medias mensuales de velocidad del viento en el año 2016

En la gráfica 62 en el año 2017 los valores máximos se registran en el mes de enero en la estación 7 y 8 con índices de 0.702 m/s y 0.742 m/s, en la estación 9 el valor máximo se registra en junio con un total de 1.036 m/s. Los valores mínimos se registran en el mes de junio en la estación 7 con valores de 0.006 m/s, en la estación 8 en el mes de mayo con un total de 0.506 m/s y en la estación 9 un total de 0.6603 en el mes de marzo.



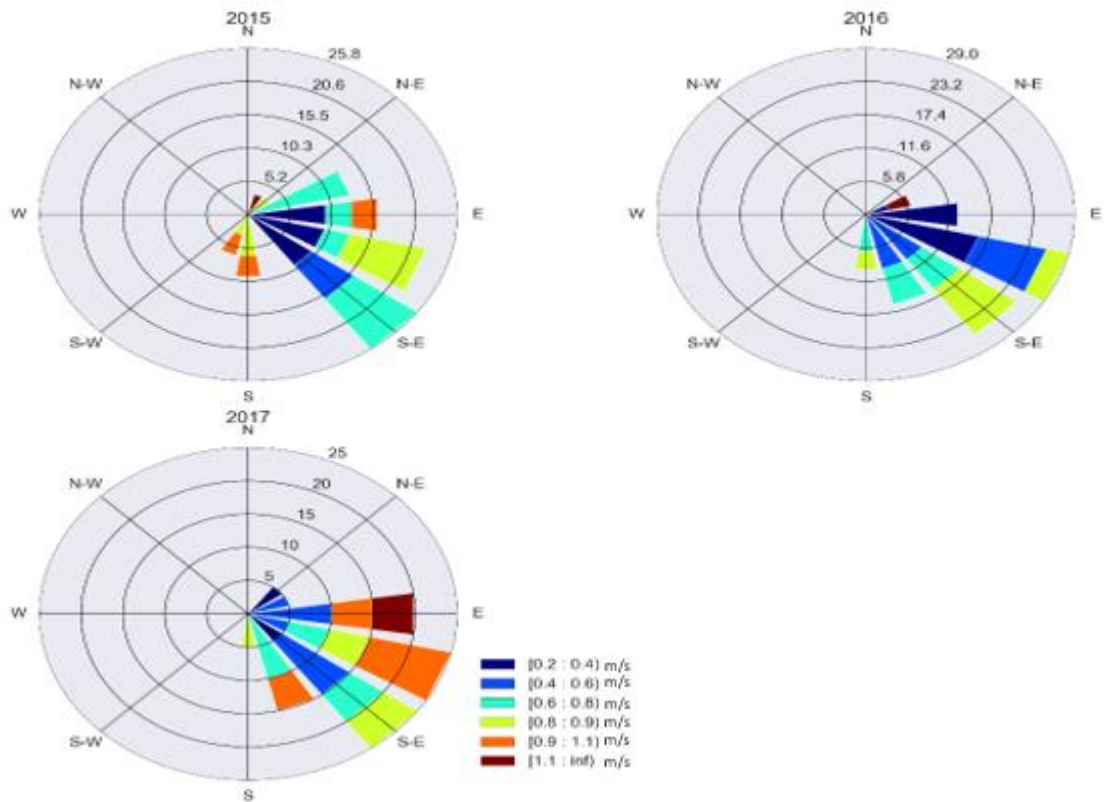
Gráfica 62: Medias mensuales de velocidad del viento en el año 2017.

6. Dirección del viento

En los siguientes gráficos se aprecia la dirección del viento mensual en las distintas estaciones.

En la gráfica 63 la rosa de los vientos de la estación 7 en el mes de enero nos indica que el viento predominante en el año 2015 es hacia el S y E-S-E con unos vientos intermedios y hacia el E con menor frecuencia unos vientos fuertes, en el año 2016 hacia el E, E-S-E y S-E predominan los vientos intermedios, en el año 2017 predomina los vientos fuertes hacia el E, E-S-E con unos vientos fuertes y hacia el S-E con unos vientos intermedios.

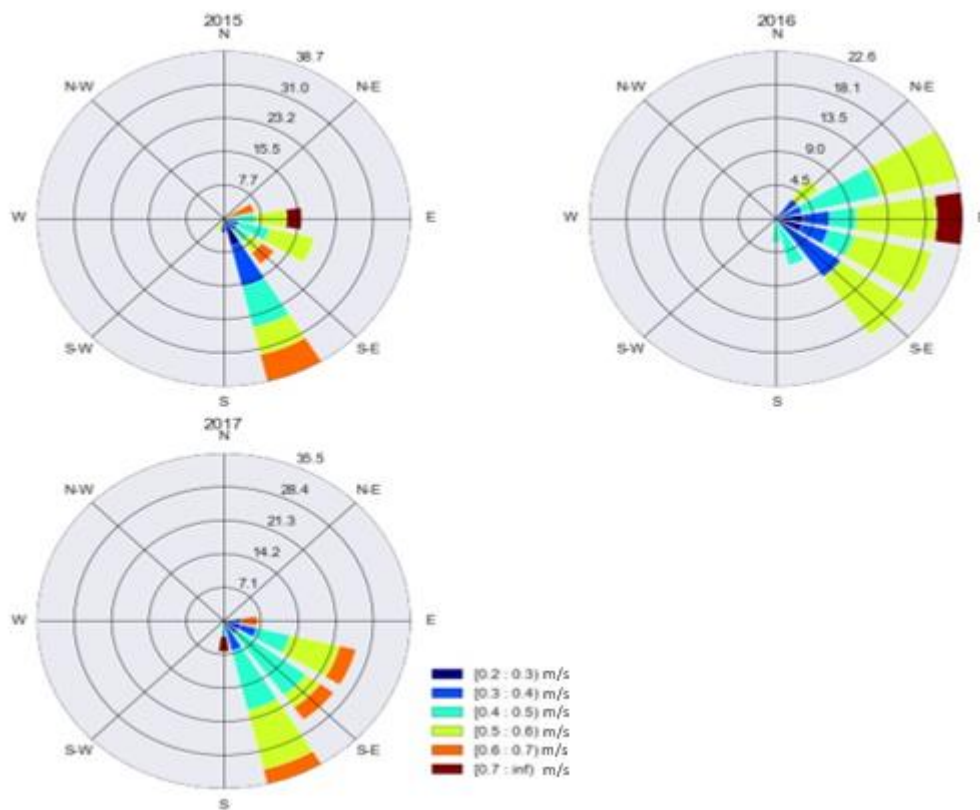
Dirección del viento Enero (Estacion 7)



Gráfica 63: Dirección del viento en el mes de enero de la estación 7.

En la gráfica 64 la rosa de los vientos de la estación 8 en el mes de Marzo nos indica que los vientos más predominantes en el año 2015 es hacia el S-S-E con unos vientos fuertes e intermedios, y en menor frecuencia pero con vientos más fuertes hacia el E, en el año 2016 hacia el E-N-E, E, E-S-E y S-E predominan los más vientos fuertes y los intermedios en menor escala, en el año 2017 predominan los vientos fuertes e intermedios hacia el E-S-E, S-E y S-S-E y con menor frecuencia pero con vientos más fuertes predomina hacia el S.

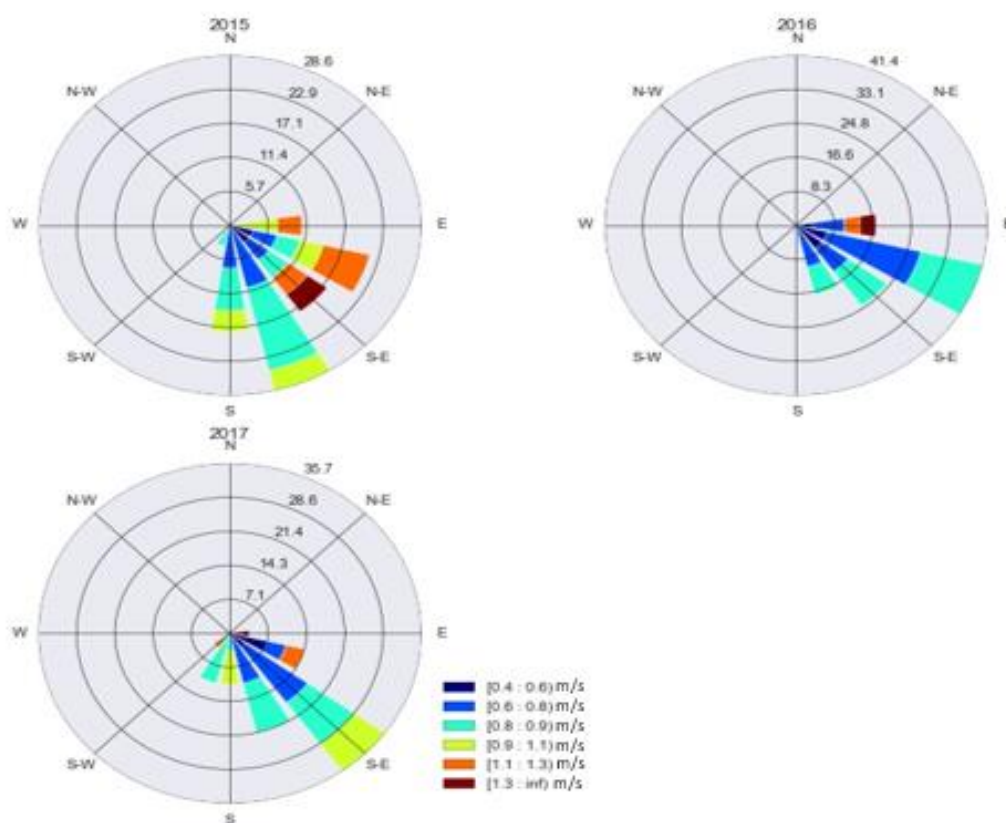
Dirección del viento Marzo (Estacion 8)



Gráfica 64: Dirección del viento en el mes de marzo de la estación 8.

En la gráfica 65 la rosa de los vientos de la estación 9 en el mes de febrero nos indica que el viento más predominante en el año 2015 es hacia el E, E-S-E, S-E y S-S-E con unos vientos fuertes e intermedios, en el año 2016 hacia el E-S-E predominan los vientos intermedios y con menor frecuencia, pero con vientos más fuertes predomina hacia el E, en el año 2017 predominan los vientos intermedios hacia el Sureste y con menor frecuencia hacia el E-S-E predominan los vientos fuertes.

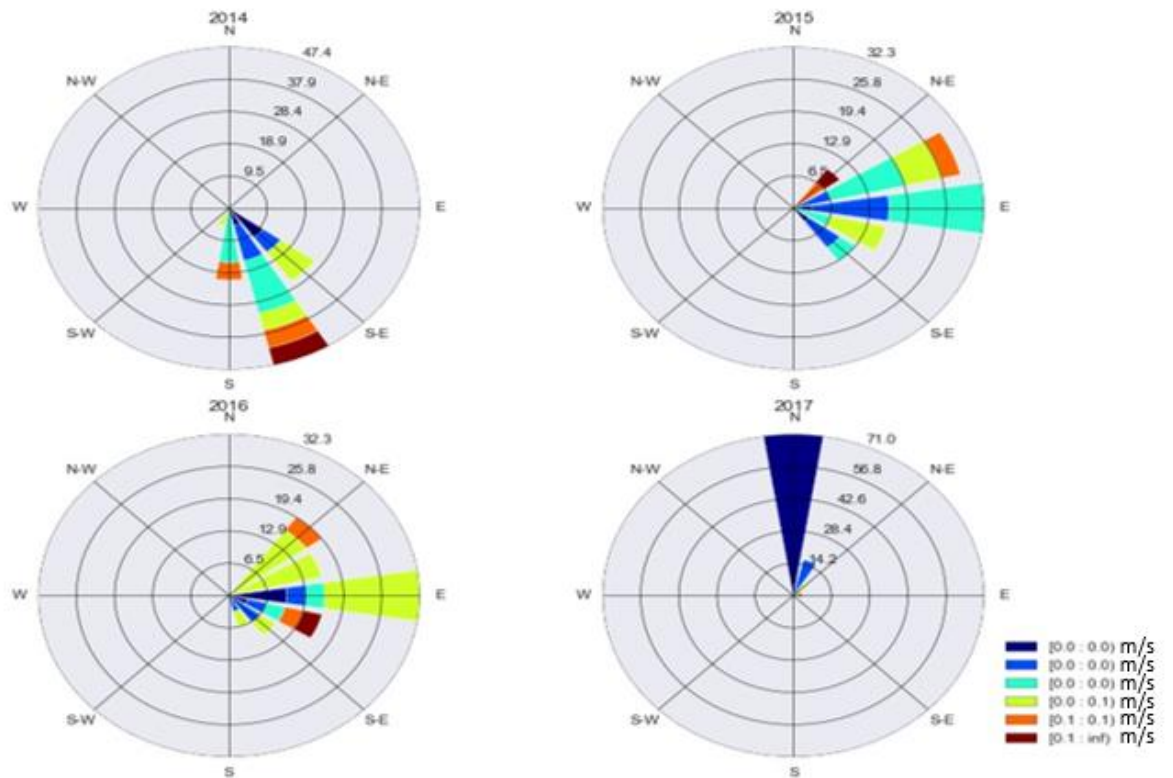
Dirección del viento Febrero (Estacion 9)



Gráfica 65: Dirección del viento en el mes de febrero de la estación 9.

En la gráfica 66 la rosa de los vientos de la estación 7 en el mes de mayo nos indica que los vientos más predominantes en el año 2014 predominan al S-S-E, en el año 2015 predomina al E-N-E los vientos fuertes y con mayor frecuencia hacia el E con vientos intermedios, en el año 2016 predomina los vientos fuertes e intermedios hacia el N-E y E y E-S-E, en el año 2017 se posee vientos suaves hacia el N.

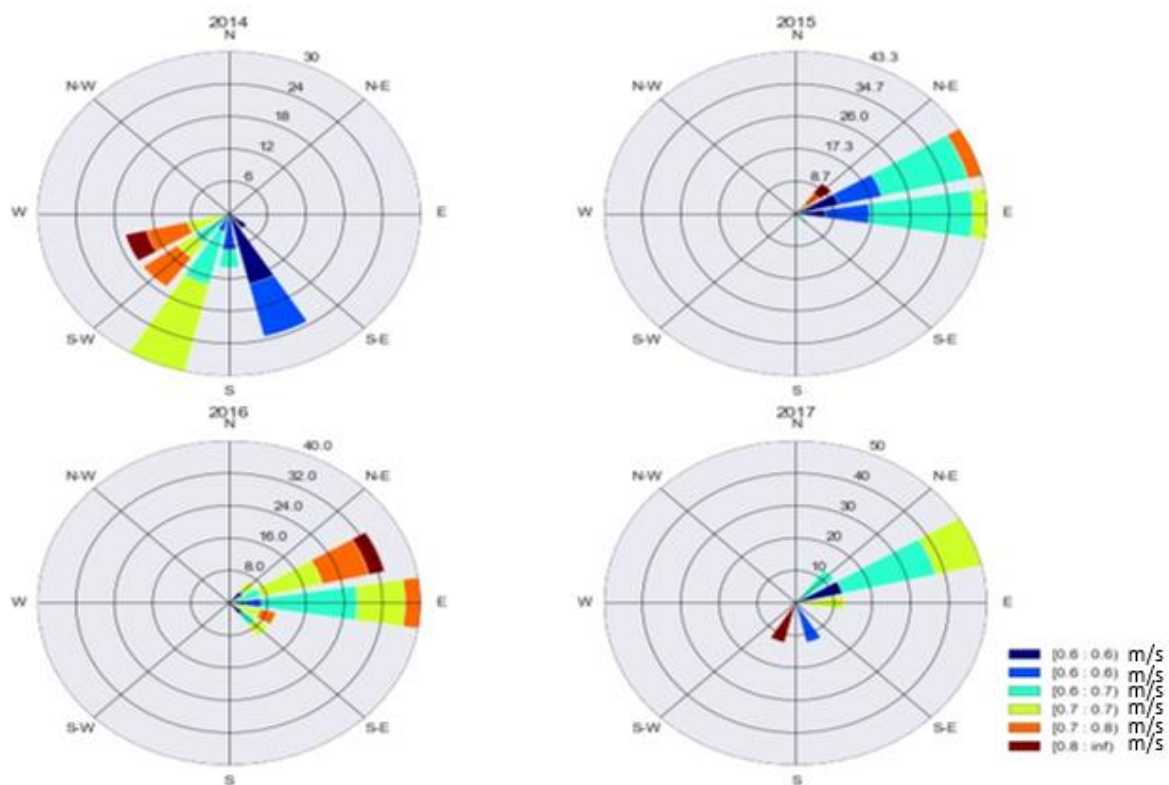
Dirección del viento Mayo (Estacion 7)



Gráfica 66: Dirección del viento en el mes de mayo de la estación 7.

En la gráfica 66 la estación 8 en el mes de junio nos indica que los vientos predominantes en el año 2014 con unos vientos fuertes hacia el S-O y O-S-O y hacia el S-S-E existe vientos débiles, en el año 2015 predomina al Este y E-N-E los vientos intermedios, en el año 2016 predomina los vientos fuertes hacia el E y E-N-E, en el año 2017 predomina los vientos leves hacia Noreste y hacia el Sureste existe vientos fuertes, pero con menor frecuencia en comparación con los otros.

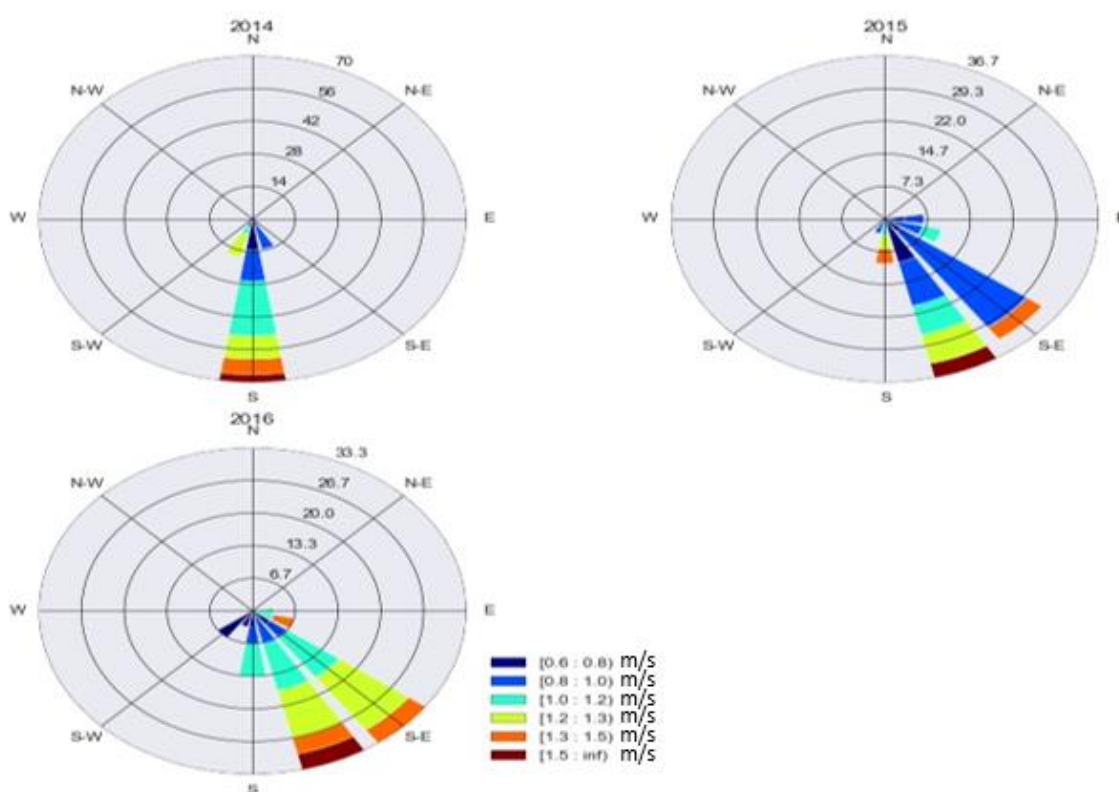
Dirección del viento Junio (Estacion 8)



Gráfica 67: Dirección del viento en el mes de junio de la estación 8.

En la gráfica 67 la rosa de los vientos de la estación 9 en el mes de septiembre nos indica que el viento más predominante en el año 2014 es hacia el S con unos vientos fuertes, en el año 2015 es hacia el S-E y S-S-E con vientos Fuertes, y en el año 2016 predomina los vientos fuertes y con hacia el hacia el S-E y al S-S-E.

Dirección del viento Septiembre (Estacion 9)



Gráfica 68: Dirección del viento en el mes de septiembre de la estación 9

Análisis

De acuerdo a los resultados obtenidos de las gráficas de precipitación se pudo observar lo siguiente:

La estación 3 Huizhil presenta las máximas precipitaciones en los años 2014 y 2017, esto puede deberse a que cuenta con grandes áreas boscosas y en esa zona se encuentra presente el río Tarqui, las presencias de grandes fuentes de agua alteran patrones al aumentar la cantidad de agua que se evapora generando mayor precipitación.

La estación 4, Matadero de Sayausi presenta las máximas precipitaciones en los años 2015 y 2016, esto se debe a que en la mayoría de su territorio se encuentra ecosistemas cubiertos de paramo, el cual se encuentra dentro Área Protegida del Parque Nacional el

Cajas. Como se sabe dentro de esta área, el Cajas, existe extensas lagunas las cuales alteran patrones al aumentar la cantidad de agua que se evapora generando mayor precipitación.

Por lo contrario, las estaciones que presentan menor cantidad de precipitación son 1,5 y 6, esto puede deberse a la presencia de mayor población en el área urbana, edificaciones altas, casas, la presencia de asfalto en las calles, el promedio de vehículos por familia, falta de áreas verdes. La presencia de estos aspectos físicos no permite que se dé el efecto albedo por completo. El efecto albedo es el rebote de energía de la superficie terrestre a la atmósfera, al presentarse una superficie oscura (asfalto) y edificaciones la mayor parte de radiación queda retenida en la superficie, generando microclimas. La generación de microclimas incrementa la temperatura en $0,06^{\circ}\text{C}$ cada década. Al incrementar la temperatura, la precipitación disminuye, a estos factores puede deberse que en estas dos zonas no se presenten precipitaciones fuertes.

La presencia de estos aspectos físicos dentro de la zona urbana da lugar al efecto invernadero, para este caso determinados gases que son componentes del planeta retienen parte de la energía que el suelo emite al ser calentado por la radiación solar. Este proceso evita que la energía regrese inmediatamente al espacio incrementando la temperatura y generando más microclimas.

La estación 2 de Narancay, presenta una precipitación media en comparación a las demás estaciones y esto podría deberse a la presencia de áreas boscosas y de igual manera a la presencia del río Tarqui.

Para las estaciones 7,8 y 9 se realizó un análisis con las gráficas de temperatura debido a que solo estas 3 estaciones son meteorológicas. De acuerdo al análisis efectuado se pudo observar que la estación 9 Tixán, que presenta mayor cantidad de precipitación es la que menor temperatura posee, debido a que estas dos variables son inversas, seguida de la estación 8 Cebollar, que posee una temperatura intermedia y por último la estación 7

Ucubamba posee poca precipitación con respecto a las anteriores, pero posee mayor temperatura.

4.3 Modelamiento predictivo basado en redes neuronales.

Como se indicó en la metodología (sección 3.5) se modificaron los hiperparámetros (número de capas, tasa de aprendizaje, tamaño del batch y número de neuronas en cada capa oculta) en los entrenamientos para cada tiempo de pronóstico (k) en todas las estaciones meteorológicas. Los modelos con mejor desempeño en el conjunto de validación fueron almacenados para su uso posterior en la etapa de inferencia.

En la tabla 15 se puede apreciar un ejemplo de los parámetros que fueron modificados en el entrenamiento de cada modelo.

Tabla 15: Ejemplo de prueba de parámetros en la estación 1 a 1 hora de pronóstico.

Tiempo de pronóstico (k)	Estación	Parámetro	Tamaño del batch	Número de capas	Número de neuronas en cada capa oculta	Tasa de aprendizaje
1	1	Prueba				
		1	10000	1	5	0,001
		2	10000	2	5	0,001
		3	10000	3	5	0,001
		4	10000	1	5	0,01
		5	10000	1	5	0,1
		6	15000	1	5	0,1
		7	5000	1	5	0,1
		8	10000	1	10	0,1
9	10000	1	15	0,1		

4.4 Validación de desempeño

Al culminar cada entrenamiento se evaluaron los mejores modelos con los datos de prueba, donde se obtuvieron métricas de rendimiento para cada uno de ellos.

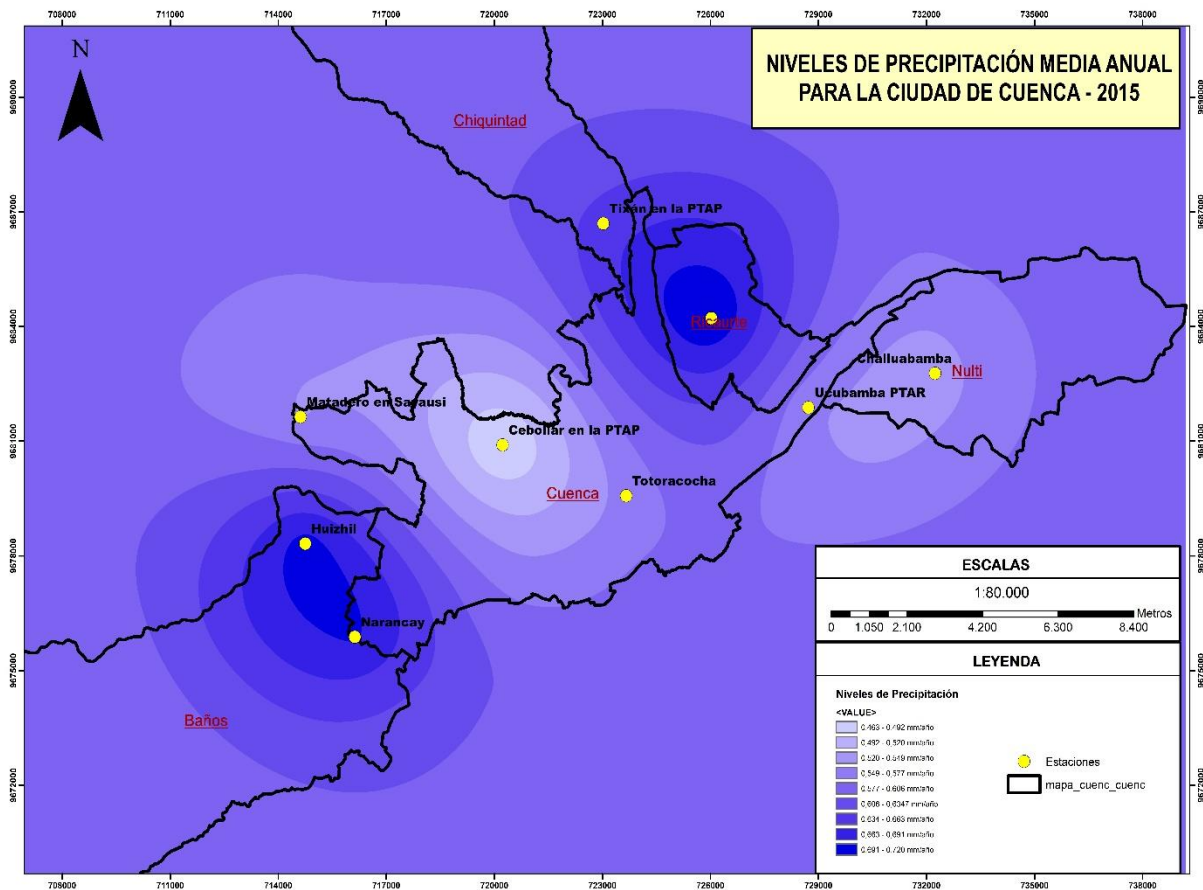
Según las métricas para cada estación y valor de salida deseado (1-24 horas) se obtuvieron los modelos definitivos que se muestran en el Anexo A.

Los resultados de las métricas por clase de los mejores modelos son expuestos en el anexo B donde se calcula matriz de confusión, precisión, recall, F1-score y exactitud.

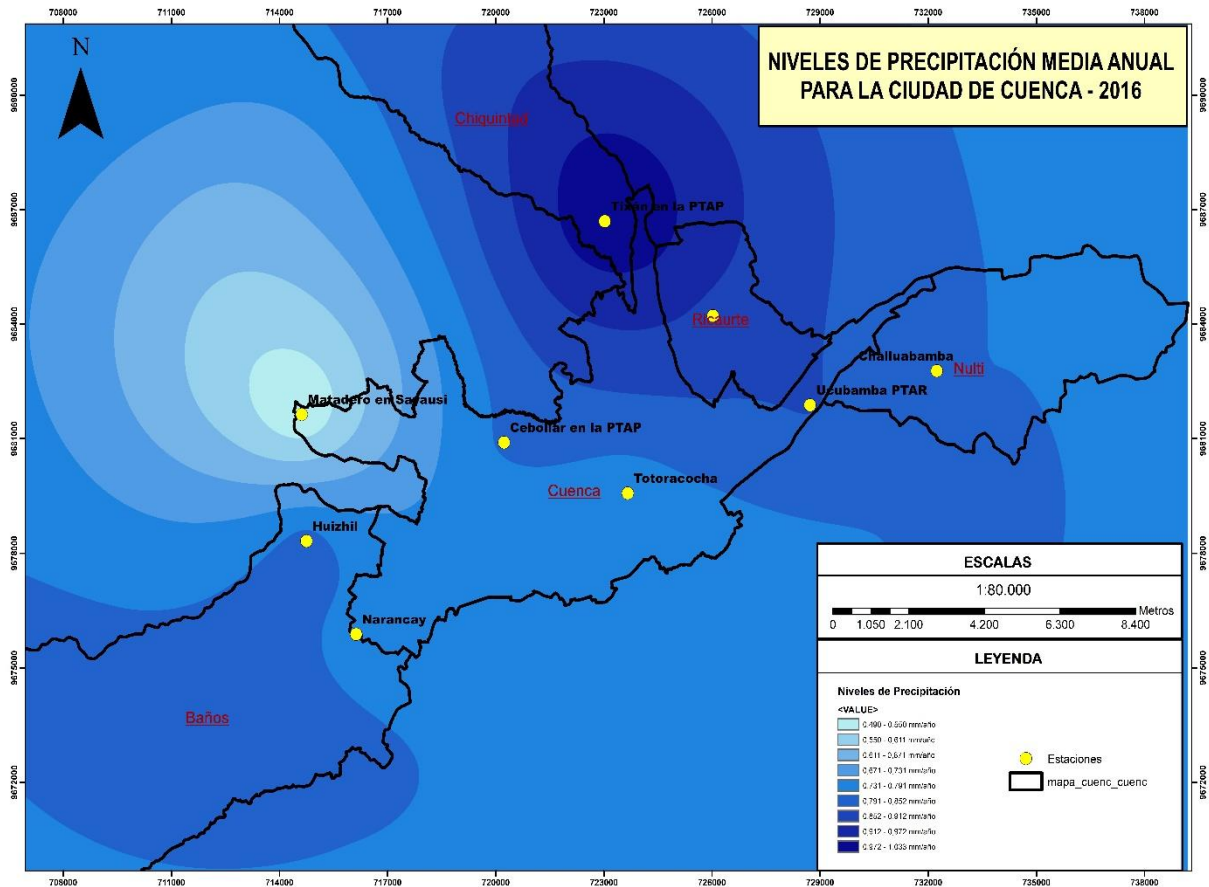
En las 9 estaciones se obtuvieron diferentes modelos a diferentes horas. La mayor cantidad de horas que se pudo pronosticar sin rebasar la frontera de 50% de exactitud de un clasificador trivial (lanzamiento de una moneda) fue de 14 horas en la estación 6 y la menor cantidad de horas fue en la estación 3 con 3 horas (Anexo B).

4.5 Presentación del modelo en SIG

Al culminar la validación del modelo, los valores de probabilidad de lluvia obtenidos son enviados al Software Q Gis a través de Sockets con una arquitectura de comunicación cliente-servidor. Las probabilidades obtenidas de cada modelo son interpoladas entre cada estación para obtener un mapa de distribuciones de probabilidades (Gráfica 69, 70).



Gráfica 69: Gráfica de interpolación de la media anual 2015.



Gráfica 70: Gráfica de interpolación de la media anual 2015.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

La presente investigación tuvo como objetivo principal, obtener un modelo para el pronóstico de precipitación a partir de los datos meteorológicos brindados por ETAPA EP del período mayo de 2014 hasta junio de 2017 en la zona urbana de Cuenca.

Primeramente, se realizó la revisión del estado del arte como fundamento teórico, para luego desarrollar la parte práctica.

Los datos fueron entregados por la empresa ETAPA EP; inicialmente se obtuvo un total de 603952 datos, estos fueron preprocesados mediante el lenguaje de programación Python, con el propósito de eliminar datos anómalos y obtener datos limpios. Al concluir con el Preprocesamiento, se obtuvo un total de 260982 datos completos en las 9 estaciones

meteorológicas para entrenar el modelo de pronóstico. En total existió la pérdida de 342970 datos.

El análisis estadístico realizado fue descriptivo, puesto que nuestro propósito fue estudiar el comportamiento mensual y anual de las variables meteorológicas, así como la variabilidad temporal y espacial de la precipitación. Al realizar el análisis estadístico descriptivo se pudo definir los meses más húmedos (febrero, marzo, abril) y los más secos del año (agosto, septiembre, octubre). Los meses de enero, mayo, junio, julio, noviembre y diciembre presentan precipitaciones, pero estas son leves en comparación de los meses lluviosos. Desde la perspectiva ambiental se encontró una relación entre el comportamiento de las variables meteorológicas y la ubicación de las mismas. La estación 3 Huizhil, presenta máximas precipitaciones en los años 2014, la estación 4 Matadero de Sayausi, presentan las máximas precipitaciones en los años 2015. En estas dos zonas (Sayausi y Huizhil) podría existir mayor precipitación por la ubicación geográfica, debido a que Huizhil cuenta con grandes áreas boscosas y la presencia del río Tarqui, mientras que en la estación 4 de Matadero de Sayausi mayor parte de territorio, se encuentra conformado por ecosistemas cubiertos de paramo, esta estación se encuentra dentro Área Protegida del Parque Nacional el Cajas. Por otro lado, se pudo realizar un análisis de las zonas más secas. Las estaciones que presentaron menor cantidad de precipitación son Challuabamba, Totoracocha y Ricaurte (estaciones 1, 5 y 6) respectivamente, puede deberse a la presencia de mayor población en el área urbana, edificaciones altas, casas, la presencia de asfalto en las calles, el promedio de vehículos por familia, falta de áreas verdes. La presencia de todos los aspectos mencionados anteriormente permite la generación de microclimas, lo cual incrementa la temperatura en esa zona.

Para las estaciones meteorológicas Ucubamba, Cebollar y Tixán (estaciones 7, 8 y 9) se realizó un análisis entre las variables meteorológicas de precipitación y temperatura,

algunos autores indican que existe una relación inversa entre estas dos variables. De acuerdo al análisis efectuado se observó que la estación 9 Tixán, presenta mayor cantidad de precipitación y menor temperatura, mientras que la estación 7 Ucubamba posee poca precipitación con respecto a las anteriores, pero posee mayor temperatura.

Los meses que mayores precipitaciones presentan son febrero, marzo, y abril, de acuerdo a la información revisada, estas fechas son periodos de invierno.

Una vez realizado el análisis estadístico, con una idea de forma general sobre el comportamiento de la variable de precipitación, se entrenó un modelo de pronóstico de precipitación mediante una Red Neuronal Recurrente de tipo LSTM. Para el entrenamiento de la red se ajustan hiperparámetros que varían en cada entrenamiento, hasta encontrar la mejor configuración del modelo a medida que se entrena el mismo. De acuerdo a todas las pruebas realizadas se pudo observar que la mayor parte de modelos se ajusta a un número de 2 a 3 capas, manteniendo una tasa de aprendizaje casi constante de entre 0,01 y 0,001, de igual manera el tamaño del batch varía entre 75000 a 100000 y por último el número de neuronas en la capa oculta varía entre 10 a 15. Se estableció como parámetro fijo 100 iteraciones para el entrenamiento del modelo. Cada 5 iteraciones el conjunto de validación es probado en el modelo, donde se obtiene métricas para evaluar que el modelo vaya mejorando a medida que se va entrenando. En cada parte de la evaluación del modelo si este ha mejorado con respecto a los modelos anteriores se almacena en una memoria el mejor modelo, de lo contrario se mantiene el mejor modelo generado hasta el momento. Una vez que se cumplan estas condiciones se ha terminado la fase de entrenamiento.

No se puede generar un modelo genérico capaz de predecir en todas en las estaciones de la ciudad de Cuenca, debido a que cada una de estaciones tiene un comportamiento diferente, por ejemplo, ante las mismas condiciones de temperatura, presión atmosférica; adicionalmente todas las variables que se han detallado en la entrada en unos sectores

llueve y en otros no. Por lo tanto, no se ha podido generar un modelo genérico capaz de realizar esta tarea, por eso se ha tenido que realizar un modelo para cada estación, donde se combinó la información de todo el resto de estaciones que influye en el comportamiento de cada uno de los lugares.

Se pudo observar que el comportamiento de las métricas clásicas disminuye su exactitud a lo largo del tiempo, lo cual tiene lógica puesto que es más difícil pronosticar a un largo periodo

Con todas las mejores configuraciones se logró realizar el pronóstico de cada estación, generando 9 modelos. La mayor cantidad de horas que se pudo pronosticar sin rebasar la frontera de 50% de exactitud de un clasificador trivial (lanzamiento de una moneda) fue de 14 horas en la estación 6 y la menor cantidad de horas fue en la estación 3 con 3 horas.

Con esto se puede concluir que las RNN de tipo LSTM son eficientes para aplicaciones de pronóstico de variables meteorológicas, trabajando con series de tiempo. La eficiencia del modelo siempre dependerá de la calidad de datos que se maneje.

Los valores de precipitación no son iguales en un mismo instante de tiempo, es por ello que se realizó un modelo para cada estación.

Luego del planteamiento de la metodología la generación de un modelo, no representa un coste computacional alto por lo cual se realizó un modelo por estación.

Los resultados obtenidos de probabilidad de lluvia fueron publicados en un mapa a través del programa Quantum GIS con la herramienta Python, el despliegue de esta interfaz realiza una interpolación entre cada estación para obtener un mapa de distribución de probabilidades.

5.2.Recomendaciones

Para realizar este tipo de proyectos se recomienda el uso de un servidor de computo que cuente con al menos una tarjeta gráfica de alto desempeño para realizar el

entrenamiento del modelo, ya que para una gran cantidad de datos se necesita una gran cantidad de memoria RAM y gran cantidad de capacidad de procesamiento. Para las pruebas iniciales se puede usar un computador normal.

Se recomienda como indispensable revisar la calidad de los datos antes de plantearse el desarrollo de un modelo.

Se debe tener en consideración las características propias del proceso que se desea modelar, por ejemplo, en la ciudad de Cuenca se desea modelar la precipitación y se ha visto muchas veces que la precipitación varía de un lugar a otro dentro de la misma ciudad. Por ello luego del planteamiento de la metodología se recomienda que se realice modelos computacionales para diferentes zonas dependiendo de sus características.

Se recomienda realizar un tipo de prueba post hoc, entre cada estación para dar un mayor peso al análisis estadístico.

Realizar un análisis estadístico de cajas y bigotes con los datos acumulados de precipitación.

5.3.Trabajos a futuro

- Se propone la detección de las zonas con microclimas en la zona urbana de Cuenca, puesto que de manera visual se pudo detectar variabilidades en distintas zonas, usando pruebas basadas en datos como clustering.
- Realizar modelos de pronóstico mediante otras técnicas de aprendizaje automático como redes neuronales convolucionales.
- Realizar un modelo de pronóstico que sea capaz de pronosticar niveles de precipitación.
- Se propone la aplicación de esta metodología para otros sistemas dinámicos dentro del área ambiental con otras variables.

6. BIBLIOGRAFIA

- Abad, O. (2014). *Estimacion de precipitacion espacial mediante correlacion con variables secundarias y la ayuda de tecnologias de Informacion Geografica*. Universidad de Cuenca. Recuperado a partir de <http://tesis.uson.mx/digital/tesis/docs/21897/Capitulo2.pdf>
- Arcila-Calderón, C., Barbosa-Caro, E., & Cabezuelo-Lorenzo, F. (2016). Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de la Información*, 25(4), 623. <https://doi.org/10.3145/epi.2016.jul.12>
- Astudillo, G. (2016, noviembre 21). El río Tomebamba es el reflejo de la sequía en Cuenca. *El Comercio*. Recuperado a partir de <http://www.elcomercio.com/actualidad/clima-riotomebamba-sequia-cuenca-ecuador.html>
- Barradas, A. (2009). “*Detección y prevención de intrusiones usando redes neuronales recurrentes*”. Instituto Politécnico Nacional.
- Baume. (2013). La Atmósfera terrestre. En *Astronomía Observacional*. Recuperado a partir de http://fcaglp.unlp.edu.ar/~observacional/teoricas/notas_2016/c04a-atmosfera.pdf
- Beaufays, F., Sak, H., & Senior, A. (2014). Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. *Interspeech*, (September), 338–342. <https://doi.org/arXiv:1402.1128>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Bloice, M., & Holzinger, A. (2016). A Tutorial on Machine Learning and Data Science Tools with Python, 9605(November). <https://doi.org/10.1007/978-3-319-50478-0>
- Cabrera Mendieta, D. R. (2017). *Modelado de Sistemas Dinámicos con Machine Learning*.

Universidad de Sevilla.

Cabrera, D., Sancho, F., Cerrada, M., Sanchez, R., & Tobar, F. (2018). *Echo state network and variational autoencoder for efficient one-class learning on dynamical systems*.

Journal of Intelligent & Fuzzy Systems

Cabrera, D., Sancho, F., Li, C., Cerrada, M., Sanchez, R., Pacheco, F., & Valente, J.

(2017). *Automatic feature extraction of time-series applied to fault severity assesment of helical gearbox in stationary and non-stationary speed operation*. Applied Soft

Computing

Campetella, Claudia; Cerne, B. P. (2011). *Entornos invisibles*. Ministerio de Educación.

Instituto Nacional de Educación Tecnológica. Buenos Aires. Recuperado a partir de

http://www.inet.edu.ar/wp-content/uploads/2012/11/C1_Parque_de_diversionesR.pdf

Castillo González, N. V. (2015). *Técnicas de Machine Learning para el Post-Proceso de la predicción de la Irradiancia*. Universidad de Granada.

CEPAL. (2003). Com. Recuperado a partir de <http://www.cepal.org/es>

Freiwan, M., & Cigizoglu, K. (2005). Prediction of total monthly rainfall in Jordan using

feed forward backpropagation method. *Fresenius Environmental Bulletin*, 14, 142–151.

French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, 137(1–4), 1–31.

[https://doi.org/10.1016/0022-1694\(92\)90046-X](https://doi.org/10.1016/0022-1694(92)90046-X)

Gabino, A. (2016). *Predicción de la precipitación pluvial en función de elementos*

hidrometeorológicos en la cuenca del Río Ramis. Universidad Nacional del Altiplano Puno “.

Galarza, D. Priscila. (2016). *Evaluación meteorológica de la cuenca del río Machangara*.

Universidad de Cuenca.

García Gonzales, J. C. F. (s/f). La interpolación como método de representación

cartográfica para la distribución de la población: Aplicación a la provincia de Albacete.

Giraldo Henao, R. (2002). *Introducción a la geoestadística*.

Gorgas Garcia, J., Cardiel Lopez, N., & Zamorano Calvo, J. (2011). *Estadística Básica Para Estudiantes de Ciencias. Universidad Complutense de Madrid*.

Hall, T., Brooks, H., & Doswell, C. (1999). Precipitation forecasting using a neural network. *Weather and Forecasting*, 14(3), 338–345.

[https://doi.org/http://dx.doi.org/10.1175/1520-0434\(1999\)014<0338:PFUANN>2.0.CO;2](https://doi.org/http://dx.doi.org/10.1175/1520-0434(1999)014<0338:PFUANN>2.0.CO;2)

Hong, W. C. (2008). Rainfall forecasting by technological machine learning models. *Applied Mathematics and Computation*, 200(1), 41–57.

<https://doi.org/10.1016/j.amc.2007.10.046>

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)

Huang, M., Lin, R., Huang, S., & Xing, T. (2017). A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. *Advanced Engineering Informatics*, 33, 89–95. <https://doi.org/10.1016/j.aei.2017.05.003>

Hunter, J., Dale, D., Firing, E., & Droettboom, M. (2018). Matplotlib. Recuperado a partir de <https://matplotlib.org/Matplotlib.pdf>

Ingsrisawang, L., Ingsrisawang, S., & Somchit, S. (2008). Machine Learning Techniques for Short-Term Rain Forecasting System in the Northeastern Part of Thailand. *World Acad Sci*, 41, 248–253.

Instituto Guatemalteco de Educación Radiofónica. (2016). *Estadística Descriptiva*. Guatemala.

Joint Committee For Guides In Metrology. (2012). *Vocabulaire international de*

métrologie. VIM3: International Vocabulary of Metrology (Vol. 3).

[https://doi.org/10.1016/0263-2241\(85\)90006-5](https://doi.org/10.1016/0263-2241(85)90006-5)

- Jones, T. (2017). Arquitecturas de aprendizaje profundo. Recuperado a partir de <https://www.ibm.com/developerworks/ssa/library/cc-machine-learning-deep-learning-architectures/index.html>
- Krishna, M. (2017). Pandas Guide. Recuperado a partir de <https://media.readthedocs.org/pdf/pandasguide/latest/pandasguide.pdf>
- Machado, F. (2016, enero 8). La sequía causa problemas en Cuenca, Latacunga, Manta y la frontera. *El telégrafo*. Recuperado a partir de <http://www.eltelegrafo.com.ec/noticias/ecuador/3/la-sequia-causa-problemas-en-cuenca-latacunga-manta-y-en-la-frontera>
- Machado, J. (2016, noviembre 21). Alarma en Cuenca por incendios y sequía. *El Tiempo*. Recuperado a partir de <http://www.eltiempo.com.ec/noticias/sucesos/9/402829/alarma-en-cuenca-por-incendios-y-sequia>
- Maqsood, I., & Abraham, A. (2007). Weather analysis using ensemble of connectionist learning paradigms. *Applied Soft Computing Journal*, 7(3), 995–1004. <https://doi.org/10.1016/j.asoc.2006.06.005>
- Maqsood, I., Khan, M. R., Huang, G. H., & Abdalla, R. (2005). Application of soft computing models to hourly weather analysis in southern Saskatchewan, Canada. *Engineering Applications of Artificial Intelligence*, 18(1), 115–125. <https://doi.org/10.1016/j.engappai.2004.08.019>
- Maqsood, I., Muhammad, R., & Abraham, A. (2002). Neural-computing based Canadian weather analysis. *The 2nd International Workshop on Intelligent Systems Design and Applications*, 39–44.

- Marzaban, C., & Stumpf, G. (1998). A neural network for damaging wind prediction. *Weather and Forecasting*, 151–163.
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*.
- Mendenhall, W., Beaver, R., Beaver, B. M., Sánchez, F., & Patiño, A. (2002). *Introducción a la probabilidad y estadística*.
- Mendez, J. (2011). Quantum Gis (QGIS): Un Sistema De Información Geográfica Basado en Software Libre. Recuperado a partir de <https://gkudos.com/blog/2011/02/15/quantum-gis-qgis-un-sistema-de-informacion-geografica-basado-en-software-libre/>
- Minsky, M., & Papert, S. (1988). *Perceptrons an introduction to computational geometry* (Expanded e). Cambridge, Mass MIT Press 1988.
- Montalvo, M. (2017). *Modelado Poisson en la variación geoestadística de ventas*. Universidad Central del Ecuador.
- Montero, J. M. (2007). *Estadística Descriptiva*. (C. M. De la Fuente, Ed.). España. Recuperado a partir de https://books.google.com.ec/books?id=D6sj2d0xTgUC&printsec=frontcover&dq=estadística+descriptiva&hl=es-419&sa=X&redir_esc=y#v=onepage&q=estadística+descriptiva&f=false
- Montesdeoca, B. (2016). *Estudios de predicción en series temporales de datos meteorológicos utilizando redes neuronales recurrentes*. Universidad de las Palmas de Gran Canaria.
- National Research Council. (2012). Urban Meteorology: Forecasting, Monitoring and meeting User's Needs. Washington DC. <https://doi.org/10.17226/13328>.
- Navone, H. D., & Ceccatto, H. A. (1994). Predicting Indian monsoon rainfall: a neural network approach. *Climate Dynamics*, 10(6–7), 305–312.

- <https://doi.org/10.1007/BF00228029>
- NumPy community. (2010). NumPy User Guide. Recuperado a partir de <https://docs.scipy.org/doc/numpy-1.5.x/numpy-user.pdf>
- Nuñez, F. J. (2016). *Diseño de un sistema de reconocimiento automático de matrículas de vehículos mediante una red neuronal convolucional*. Universitat Oberta de Catalunya. Recuperado a partir de <http://openaccess.uoc.edu/webapps/o2/handle/10609/52222%0Ahttp://openaccess.uoc.edu/webapps/o2/bitstream/10609/52222/6/fnunezsTFM0616memòria.pdf>
- Orellana, L. (2001). *Estadística descriptiva*.
- Peña, P. (2016). *Integración de imágenes satelitales de precipitación TRMM con información de pluviómetros para los años 2001 y 2002 en Ecuador*. Universidad de Cuenca.
- Reitermanová, Z. (2010). Data Splitting. *Week of Doctoral Students 2010 -- Proceedings of Contributed Papers*, 31–36. Recuperado a partir de https://www.mff.cuni.cz/veda/konference/wds/proc/pdf10/WDS10_105_i1_Reitermanova.pdf
- Rodriguez, Rosa Maria; Benito, Agueda; Portela, A. (2004). *Meteorología y Climatología* (FECYT). España. <https://doi.org/M-XXXXXX-2004>
- Rosenblatt, F., & Van Der, M. (1986). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. *Brain Theory*, 245–248. https://doi.org/10.1007/978-3-642-70911-1_20
- Samaniego, N., Eguiguren, P., Maita, J., & Aguirre, N. (2015). Clima de la Región Sur del Ecuador: historia y tendencias. *Cambio climático y biodiversidad: Estudio de caso de los páramos del Parque Nacional Podocarpus, Ecuador*, (December), 43–63.
- Santovenia Díaz, Javier Tarragó Montalvo, C., & Cañedo, R. (2009). Sistemas de

- información geográfica para la gestión de la información. *ACIMED*. Recuperado a partir de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352009001100007
- Scikit-Learn. (s/f). User guide. 3.3 Model evaluation: quantifying the quality of predictions. Recuperado a partir de http://scikit-learn.org/stable/modules/model_evaluation.html#precision-recall-f-measure-metrics
- SENPLADES. (2013). *Plan Nacional del Buen Vivir. Obtenido de Constituyendo un Estado Plurinacional e Intercultural: http://issuu.com/publisenplades/docs/pnbv_2009-2013* (Secretaria). Recuperado a partir de <http://biblioteca.gestionderiesgos.gob.ec/items/show/55>
- SENPLADES. (2017). *Pla Nacional de Desarrollo Toda una vida 2017-2021*. Quito.
- Sierra, M. (2006). *Establecer La Asociación Existente Entre Las Variables Meteorológicas Temperatura, Velocidad Del Viento Y Precipitación Y Las Concentraciones De Pm10 Registradas En La Red De Calidad Del Aire De Bogota D.C.*
- Staudemeyer, R., & Omlin, C. (2013). Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data.
- Sucozhañay Calle, A. E. (2015). *Análisis de la variabilidad espacio temporal de la precipitación en la ciudad de Cuenca durante el periodo febrero-julio de 2014*. Universidad de Cuenca.
- Tensorflow. (s/f). TensorFlow. Recuperado a partir de <https://www.tensorflow.org/>
- Valverde Ramírez, M. C., De Campos Velho, H. F., & Ferreira, N. J. (2005). Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *Journal of Hydrology*, 301(1–4), 146–162. <https://doi.org/10.1016/j.jhydrol.2004.06.028>
- van Rossum, G. (2009). *El tutorial de Python*. (F. J. Drake, Ed.), *Python* (Vol. 1). Buenos

Aires.

Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Harvard University.

Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. *1960 IRE WESCON Convention Record*, (4), 96–104.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical Evaluation of Rectified Activations in Convolutional Network, 2. Recuperado a partir de <http://arxiv.org/abs/1505.00853>

Yao, K., Peng, B., Zhang, Y., Yu, D., Zweig, G., & Shi, Y. (2014). Spoken Language Understanding Using Long-Term Memory Neural Networks. *Spoken Language Technology*, 189–194. <https://doi.org/10.1109/SLT.2014.7078572>

7. ANEXOS

Anexo A: Parámetros de los mejores modelos de pronóstico.

A continuación, se presenta la tabla con los parámetros obtenidos para los mejores modelos. Con la siguiente leyenda para cada elemento.

Leyenda	
E	Estación
k	Valor de salida deseado
BS	Tamaño del batch
L	Número de capas
H_S	Número de neuronas en cada capa oculta
L_R	Tasa de aprendizaje

PARÁMETROS				
CODIGO	BS	L	H_S	L_R
E1 k1	10000	1	15	0,1
E1 k2	5000	2	5	0,01
E1 k3	5000	1	10	0,01
E1 k4	5000	1	5	0,01
E1 k5	15000	1	10	0,001
E1 k6	10000	1	5	0,01
E1 k7	20000	1	5	0,001
E1 k8	10000	1	5	0,001
E1 k9	10000	2	10	0,001
E1 k10	10000	1	10	0,01
E1 k11	10000	2	5	0,001
E2 k1	10000	1	5	0,01
E2 k2	15000	1	5	0,01
E2 k3	10000	2	10	0,001
E2 k4	15000	1	5	0,01
E3 k1	10000	1	5	0,01
E3 k2	10000	1	5	0,01
E3 k3	10000	1	5	0,001
E4 k1	20000	1	5	0,01
E4 k2	10000	1	10	0,01
E4 k3	15000	1	5	0,1
E4 k4	15000	1	5	0,001
E4 k5	10000	1	10	0,1
E4 k6	15000	1	5	0,001
E4 k7	10000	3	5	0,001
E4 k8	10000	1	5	0,01

CODIGO	BS	L	H_S	L_R
E5 k 1	10000	1	10	0,01
E5 k 2	10000	1	10	0,001
E5 k 3	10000	2	5	0,01
E5 k 4	10000	1	5	0,001
E5 k 5	20000	2	10	0,001
E5 k 6	10000	1	5	0,01
E5 k 7	10000	1	10	0,001
E5 k 8	15000	1	5	0,01
E5 k 9	10000	2	5	0,001
E5 k 10	10000	1	5	0,01
E5 k 11	10000	2	5	0,001
E6 k1	10000	1	10	0,01
E6 k2	10000	2	10	0,1
E6 k3	10000	1	15	0,1
E6 k4	10000	1	5	0,001
E6 k5	10000	1	5	0,01
E6 k6	10000	1	10	0,01
E6 k7	10000	1	10	0,1
E6 k8	10000	1	5	0,01
E6 k9	15000	1	5	0,001
E6 k10	10000	2	5	0,001
E6 k11	15000	2	5	0,001
E6 k12	10000	1	5	0,001
E6 k13	10000	1	5	0,1
E6 k14	15000	1	10	0,001
E7 k1	10000	1	10	0,01
E7 k2	20000	2	5	0,01
E7 k3	10000	1	5	0,01
E7 k4	10000	1	5	0,01
E7 k5	10000	1	5	0,001
E7 k6	10000	1	5	0,001
E8 k1	10000	1	5	0,01
E8 k2	10000	2	5	0,01
E8 k3	15000	1	5	0,01
E8 k4	10000	1	5	0,001
E8 k5	15000	1	10	0,001
E8 k6	15000	1	5	0,1
E8 k7	10000	1	5	0,01
E9 k1	15000	1	5	0,1
E9 k2	10000	1	10	0,01
E9 k3	20000	1	5	0,01
E9 k4	10000	1	10	0,001
E9 k5	15000	1	5	0,001

Anexo B: Métricas de rendimiento de los mejores modelos de pronóstico.

A continuación, se presentan los valores obtenidos en las métricas de los mejores modelos encontrados, se presenta la leyenda de cada elemento.

Leyenda	
VP	Verdadero positivo
VN	Verdadero negativo
REC_0	Recall clase 0
REC_1	Recall clase 1
F1_SCO_0	F1-score clase 0
F1_SCO_1	F1-score clase 1
PRECI_0	Precisión clase 0
PRECI_1	Precisión clase 1
EXAC	Exactitud

MÉTRICAS									
CODIGO	VP	VN	REC_0	REC_1	F1_SCO_0	F1_SCO_1	PRECI_0	PRECI_1	EXAC
E1 k1	0,8	0,8387	0,8	0,8387	0,816	0,8229	0,8322	0,8078	0,8195
E1 k2	0,7129	0,8337	0,7129	0,8337	0,7587	0,7862	0,8108	0,7438	0,7733
E1 k3	0,6368	0,7915	0,6368	0,7915	0,6902	0,7347	0,7534	0,6855	0,7142
E1 k4	0,6158	0,7717	0,6158	0,7717	0,6679	0,7159	0,7295	0,6676	0,6938
E1 k5	0,6143	0,7642	0,6143	0,7642	0,6641	0,7109	0,7227	0,6646	0,6893
E1 k6	0,6298	0,6327	0,6298	0,6327	0,6307	0,6318	0,6316	0,6309	0,6313
E1 k7	0,5552	0,7444	0,5552	0,7444	0,6132	0,68	0,6847	0,626	0,6498
E1 k8	0,5215	0,7562	0,5215	0,7562	0,5908	0,6768	0,6814	0,6124	0,6162
E1 k9	0,5912	0,591	0,5912	0,591	0,5911	0,591	0,5911	0,5912	0,5911
E1 k10	0,5375	0,6633	0,5375	0,6633	0,5736	0,6241	0,6149	0,5892	0,6004
E1 k11	0,6686	0,5586	0,6686	0,5586	0,6367	0,5911	0,6023	0,6276	0,6136
E2 k1	0,7477	0,8608	0,7477	0,8608	0,7925	0,8147	0,8431	0,7733	0,8043
E2 k2	0,6648	0,8413	0,6648	0,8413	0,7291	0,773	0,8072	0,7151	0,753
E2 k3	0,6369	0,8195	0,6369	0,8195	0,7	0,7509	0,7792	0,693	0,7282
E2 k4	0,5447	0,8474	0,5447	0,8474	0,6419	0,736	0,7812	0,6505	0,6961
E3 k1	0,7964	0,8355	0,7964	0,8355	0,8123	0,8195	0,8288	0,8041	0,816
E3 k2	0,6498	0,847	0,6498	0,847	0,7209	0,771	0,8094	0,7075	0,7484
E3 k3	0,6128	0,7992	0,6128	0,7992	0,6758	0,731	0,7532	0,6733	0,706
E4 k1	0,7904	0,843	0,7904	0,843	0,8118	0,8214	0,8343	0,8	0,8167
E4 k2	0,6377	0,8571	0,6377	0,8571	0,7163	0,77224	0,8169	0,7029	0,7474
E4 k3	0,6487	0,7556	0,6487	0,7556	0,6853	0,7172	0,7263	0,6826	0,7021
E4 k4	0,5949	0,7085	0,5949	0,7085	0,6307	0,6704	0,6711	0,6362	0,6517

CODIGO	VP	VN	REC_0	REC_1	F1 SCO_0	F1 SCO_1	PRECI_0	PRECI_1	EXAC
E4 k5	0,576	0,7376	0,576	0,7376	0,6266	0,6825	0,687	0,635	0,6568
E4 k6	0,5389	0,7488	0,5389	0,7488	0,6021	0,6777	0,6821	0,6189	0,6439
E4 k7	0,5845	0,695	0,5845	0,695	0,4401	0,707	0,8082	0,5709	0,6153
E4 k8	0,6042	0,6224	0,6042	0,6224	0,6098	0,6168	0,6154	0,6113	0,6133
E5 k1	0,72885	0,9523	0,72885	0,9523	0,8203	0,8565	0,9386	0,7782	0,8404
E5 k2	0,6306	0,8571	0,6306	0,8571	0,7112	0,7699	0,8153	0,6988	0,767
E5 k3	0,648	0,7677	0,648	0,7677	0,6892	0,7243	0,7361	0,6856	0,7338
E5 k4	0,6322	0,9047	0,6322	0,9047	0,732	0,7962	0,8609	0,711	0,7685
E5 k5	0,5646	0,7481	0,5646	0,7481	0,6217	0,6852	0,6915	0,6321	0,6564
E5 k6	0,525	0,7628	0,525	0,7628	0,5958	0,6817	0,6888	0,6162	0,6439
E5 k7	0,6017	0,6495	0,6017	0,6495	0,6164	0,6343	0,6319	0,6199	0,6256
E5 k8	0,5993	0,6798	0,5993	0,6798	0,6244	0,6535	0,6517	0,6291	0,6395
E5 k9	0,5553	0,7512	0,5553	0,7512	0,6156	0,6842	0,6906	0,6281	0,6532
E5 k10	0,5364	0,716	0,5364	0,716	0,5893	0,657	0,6539	0,607	0,6262
E5 k11	0,5039	0,7735	0,5039	0,7735	0,583	0,6926	0,6916	0,6098	0,6396
E6 k1	0,796	0,856	0,796	0,856	0,8207	0,8311	0,8469	0,8076	0,826
E6 k2	0,6788	0,8317	0,6788	0,8317	0,7351	0,7727	0,8013	0,7215	0,7553
E6 k3	0,636	0,7995	0,636	0,7995	0,6926	0,7391	0,7603	0,6871	0,7177
E6 k4	0,6	0,7726	0,6	0,7726	0,6569	0,7113	0,7252	0,659	0,6864
E6 k5	0,5693	0,7677	0,5693	0,7677	0,632	0,6984	0,7102	0,6406	0,6685
E6 k6	0,596	0,731	0,596	0,731	0,6391	0,6848	0,689	0,644	0,6635
E6 k7	0,5617	0,7475	0,5617	0,7475	0,6192	0,684	0,6899	0,6304	0,6464
E6 k8	0,5764	0,6896	0,5764	0,6896	0,611	0,6527	0,65	0,6195	0,633
E6 k9	0,5163	0,7438	0,5163	0,7438	0,5826	0,6678	0,6684	0,6059	0,6301
E6 k10	0,5087	0,7959	0,5087	0,7959	0,5929	0,694	0,7104	0,6173	0,6506
E6 k11	0,5418	0,7679	0,5418	0,7679	0,7679	0,6899	0,7001	0,6263	0,6548
E6 k12	0,5636	0,6888	0,5636	0,6888	0,6013	0,6482	0,6443	0,6122	0,6262
E6 k13	0,5696	0,7037	0,5696	0,7037	0,7037	0,6594	0,6578	0,6205	0,6366
E6 k14	0,5294	0,7827	0,5294	0,7827	0,6062	0,6947	0,709	0,6345	0,656
E7 k1	0,8033	0,8422	0,8033	0,8422	0,8192	0,8261	0,8358	0,8106	0,8227
E7 k2	0,7055	0,7981	0,7055	0,7981	0,7397	0,7628	0,7775	0,7304	0,7518
E7 k3	0,663	0,7627	0,663	0,7627	0,6978	0,7265	0,7365	0,6936	0,7129
E7 k4	0,5935	0,7674	0,5935	0,7674	0,65	0,706	0,7184	0,6537	0,6804
E7 k5	0,6075	0,7279	0,6075	0,7279	0,6464	0,6865	0,6906	0,6497	0,6677
E7 k6	0,5919	0,7	0,5919	0,7	0,6257	0,6641	0,6636	0,6317	0,6459
E8 k1	0,7939	0,8625	0,7939	0,8625	0,8221	0,8339	0,8524	0,8071	0,8282
E8 k2	0,6296	0,8314	0,6296	0,8314	0,7002	0,7552	0,7888	0,6918	0,7305
E8 k3	0,5978	0,8203	0,5978	0,8203	0,6727	0,7382	0,7689	0,671	0,7091

CODIGO	VP	VN	REC_0	REC_1	F1 SCO_0	F1 SCO_1	PRECI_0	PRECI_1	EXAC
E8 k4	0,5426	0,7849	0,5426	0,7849	0,6174	0,7001	0,7161	0,6318	0,6637
E8 k5	0,544	0,7583	0,544	0,7583	0,6093	0,6849	0,6924	0,6245	0,6512
E8 k6	0,5364	0,7228	0,5364	0,7228	0,5916	0,6612	0,6593	0,6092	0,6292
E8 k7	0,5485	0,7533	0,5485	0,7533	0,6111	0,6833	0,6898	0,6252	0,6509
E9 k1	0,7555	0,8399	0,7555	0,8399	0,7888	0,8059	0,8251	0,7745	0,7977
E9 k2	0,6454	0,8109	0,6454	0,8109	0,7037	0,749	0,7735	0,6958	0,7282
E9 k3	0,5432	0,8105	0,5432	0,8105	0,627	0,7149	0,7414	0,6395	0,6768
E9 k4	0,5414	0,7748	0,5414	0,7748	0,613	0,6938	0,7063	0,6282	0,6581
E9 k5	0,5181	0,7505	0,5181	0,7505	0,5862	0,6724	0,675	0,609	0,6343