UNIVERSIDAD DEL AZUAY

DEPARTAMENTO DE POSGRADOS

TÍTULO: DISEÑO DE UN MODELO DE PREDICCIÓN DE LA DEMANDA DE VIAJES ENTRE ESTACIONES DEL SISTEMA DE BICICLETA PÚBLICA EN LA CIUDAD DE CUENCA

Trabajo de graduación previo a la obtención del título de:

Magister en Matemática Aplicada

Autor:

Nombres y Apellidos: Mgtr. María Fernanda Samaniego Larriva

Director:

Nombres y Apellidos: PhD. Iván Andrés Mendoza Vázquez

Cuenca – Ecuador

2023

AGRADECIMIENTOS

Agradezco a Dios por sus bendiciones y por brindarme buena salud. A mi familia, Juan, José y Julita, ha sido mi mayor apoyo en cada proyecto y decisión que he tomado. También agradezco al Ing. Iván Mendoza por su liderazgo como director de tesis, al Ing. Julio Mosquera y al Ing. Jonathan Avilés por su orientación en el tribunal. Mi gratitud se extiende al Ing. Mateo Coello por su intermediación con EMOV EP y al Ing. Omar Delgado por su ayuda con la información de la red de sensores remotos del IERSE. Además, agradezco al Ing. Cristian Rojas, Ing. Ana Ochoa e Ing. Santiago Núñez por permitirme utilizar sus equipos de computación. Sin todos ustedes, este logro no habría sido posible.

DEDICATORIA

Dedico este trabajo de investigación a mi amada familia: Juan, Susana, José y Julita. Siempre han sido mi pilar y mi fuente inagotable de apoyo. En cada paso de este camino, ustedes han estado ahí, brindándome su amor incondicional y su aliento constante.

INDICE

INDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE ANEXOS

DISEÑO DE UN MODELO DE PREDICCIÓN DE LA DEMANDA DE VIAJES ENTRE ESTACIONES DEL SISTEMA DE BICICLETA PÚBLICA EN LA CIUDAD DE CUENCA

María Fernanda Samaniego Larriva mafersamaniego@uazuay.edu.ec - Iván Andrés Mendoza Vázquez imendoza@uazuay.edu.ec

RESUMEN

Se desarrolló un modelo para predecir la demanda de viajes entre estaciones del sistema de bicicleta pública de la ciudad de Cuenca. Para construir este modelo, se recopiló información de diversas fuentes. Mediante un análisis de correlación entre las variables, se seleccionaron las más relevantes para las predicciones, como la estación de bicicleta, temperatura, humedad, ruido, fin de semana y hora del día. El análisis se realizó con los datos desde el 1 de abril del 2.020 al 31 de diciembre del 2.022. Inicialmente se intentó realizar predicciones a nivel horario, pero debido al bajo uso por estación, se obtuvieron mejores resultados al agrupar los datos en intervalos de mañana, tarde y noche. Para generar el modelo de predicción, se utilizó el algoritmo de random forest, el cual mostró un mayor coeficiente de determinación en las predicciones tanto para las llegadas como para las salidas.

Palabras clave.

Modelo de predicción, Minería de datos, Sistema de bicicleta pública, Análisis de correlación, *Random forest*.

ABSTRACT

In this study, a model was developed to predict the demand of trips between stations of the public bicycle system in Cuenca. To build this model, information was collected from various sources. Through a correlation analysis between variables, the most relevant variables for predictions were selected, such as bicycle station, temperature, humidity, noise, weekend and time of day. The analysis was performed with data from April 1, 2020 to December 31, 2002. Initially, predictions were attempted at the hourly level, but due to low usage per station, better results were obtained by grouping the data into morning, afternoon, and evening intervals. To generate the prediction model, the random forest algorithm was used, which showed a higher coefficient of determination in the predictions for both arrivals and departures.

Keywords

Prediction model, Data mining, Public bike system, Correlation analysis, Random forest.

1 INTRODUCCIÓN

Los sistemas de bicicleta compartida (BSS) son una alternativa de movilidad activa, sostenible no contaminante. Su uso tiene muchos beneficios, entre ellos mejora la calidad de vida, reducción de costos de movilización comparada con la alternativa de transporte motorizado privado, es asequible porque no requiere una inversión inicial, se reducen los riesgos de robo, entre otros.

Los sistemas de bicicletas compartidas generalmente son automatizados y generan registros de información básicas del viaje: como lugar origen, lugar destino, fecha, tiempo de uso, cada vez que los usuarios tienen contacto con los dispositivos asociados con este sistema (Yang y otros, 2018) (BiciCuenca, s.f.), el proceso, variables generadas y almacenamiento de la información depende de cada operador.

El uso de BSS está condicionado a varios factores que se revisarán más adelante, algunos de los aspectos analizados son la hora del día, si es un día normal de trabajo, aspectos climáticos, perfiles socio demográficos del entorno, características del uso del suelo, etc. A los registros de estos datos los llamaremos variables y la repetición de un conjunto de comportamientos pueden ser reconocidos como patrones.

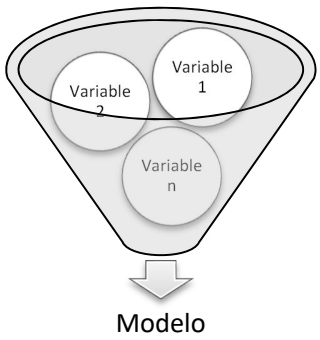


Ilustración 1 Esquema de un modelo sencillo. Elaboración propia.

En la Ilustración 1, se pueden identificar a las variables como insumos para generar el modelo. El embudo simboliza el conjunto de procesos y técnicas que fueron usadas para encontrar patrones de similitud del comportamiento, este proceso se llama minería de datos. Generalmente, para generar un modelo a partir de una simulación computacional, se requieren dos conjuntos de datos: uno para entrenarse, es decir para extraer patrones y otro para probar su exactitud y eficacia.

Los modelos generados, reconocen las características de las variables con las que fueron entrenados y replican los resultados.

En la ciudad de Cuenca, la empresa que opera el Sistema de bicicletas compartidas es BiciCuenca, cuenta con una flota de bicicletas distribuidas en 20 estaciones de la ciudad con capacidades de entre 12 y 28 bicicletas cada una. Según (EMOV EP., 2021) en la ciudad de Cuenca, la alcaldía realizó una inversión en el mes de junio del año 2021 de 8´081.380 millones de dólares: para "la construcción de 13.5km de red de ciclovías nominada "Cuenca Unida en Bici", la cual forma parte de los 125km de infraestructura ciclística planificada en la actual administración". En la Ilustración 2 se puede ver las estaciones de bicicletas y las ciclovías de la ciudad.

Los datos con los que se cuenta en el presente proyecto son los registros de uso de la bicicleta pública BiciCuenca desde abril 2019 hasta diciembre 2.022 (EMOV EP., 2023), la información meteorológica se obtuvo a través del IERSE (IERSE, 2023) para ese mismo lapso de tiempo. Se espera que la información climática pueda utilizarse para complementar la información de uso de bicicletas y modelarla consiguiendo una mayor precisión en la predicción.

Al momento no se dispone de un sistema que prediga la demanda de bicicletas, en las estaciones puede haber escasez o saturación del sistema. En el presente trabajo de investigación se formulará un modelo que prediga la demanda del uso de las bicicletas públicas cada hora durante el día en cada estación del sistema BiciCuenca en la ciudad de Cuenca, tomando una porción de datos para entrenamiento y otra porción para verificar su funcionamiento.

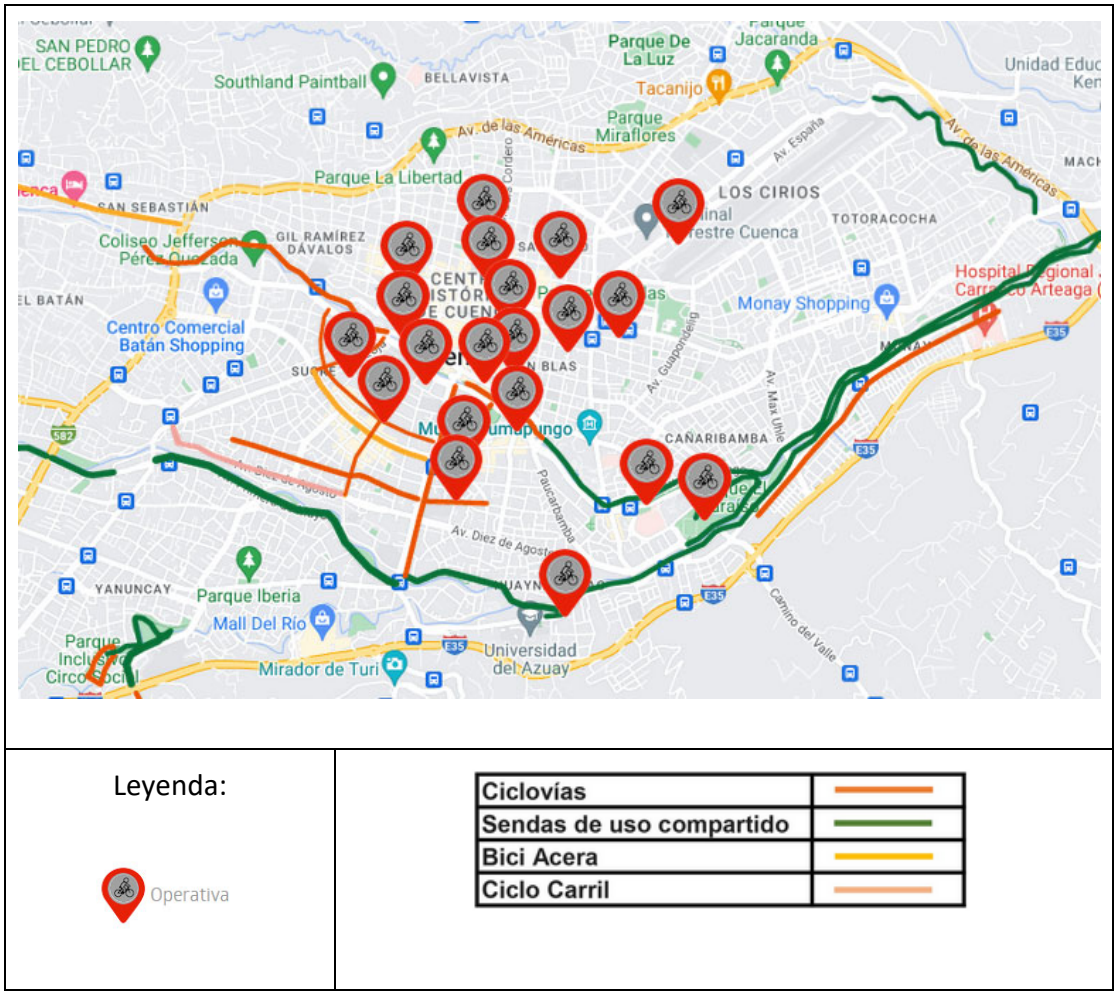


Ilustración 2 Mapa de Estaciones Bicicletas Públicas. Tomado de (BiciCuenca, s.f.)

2 REVISIÓN BIBLIOGRÁFICA

Los siguientes trabajos tratan el problema de la predicción de demanda de bicicletas públicas en diferentes contextos.

Antes de ahondar en las predicciones, de manera general: se llaman variables de entrada a las características que se conocen, a partir de una o más variables de entrada, se obtendrá una predicción o una variable de salida. Es por esta razón que es importante elegir variables de entrada que sean relevantes y relacionadas con la variable a encontrar.

Para las predicciones de demanda del uso de los sistemas de bicicletas compartidas (Yang y otros, 2016) agrega al análisis de información del clima,

estacionalidad, características especiales de la fecha: si es feriado, características específicas de las estaciones, si están cerradas, distancia entre estaciones, capacidad, tiempo mínimo y máximo entre estaciones, si el usuario es usuario regular. (Gao & Lee, 2019) información del clima, feriados o huelgas. (Li y otros, 2015) información del clima, locaciones de estaciones. (Bustamante y otros, 2022) utiliza número de viajes entre vecindarios, agrega al análisis información del clima, fin de semana, feriado, información demográfica, uso de suelo del entorno de las estaciones, distancias, altitud, transporte público, infraestructura del entorno amigable para bicicletas, ciclovías, o vías compartidas, disponibilidad de bicicletas.

Hay muchas maneras de realizar las predicciones, entre ellas tenemos modelos cuantitativos que se basan en tendencias para predecir la demanda y aprendizaje de máquinas que mediante varias técnicas como árboles de decisiones, algoritmos genéticos o redes neuronales.

(Kaltenbrunner y otros, 2010) **,** (Li y otros, 2015), (Yang y otros, 2016) analizan el uso de series de tiempo ARMA *(Autoregressive Moving Average*) para replicar tendencias en función de datos históricos incorporando información o variables del entorno.

Para la calibración del modelo (Gao & Lee, 2019) utiliza red neuronal de propagación hacia atrás, Algoritmo de C-medias difusas, Algoritmo genético híbrido. (Yang y otros, 2016) utiliza el algoritmo de *random forest*, (Yang y otros, 2018) utiliza aprendizaje profundo (DL) utilizando las redes neuronales convolucionales (CNNs); (Guidon y otros, 2020) utiliza regresiones espaciales y el algoritmo de *random forest*.

En la tabla que se presenta a continuación, se muestra a detalle las características de cada uno de los artículos revisados con detalles del lugar de análisis, modelos utilizados, en algunos casos se muestran los modelos que fueron utilizados para comparar resultados. También se muestra la variable o resultados que predicen y las variables independientes que fueron utilizadas ya sea para un análisis inicial o implementos para generar la predicción.

#	Autores	Año	Título	Lugar de análisis	Modelos Utilizados	¿Qué predice?	Variables Independientes
1	Andreas Kaltenbrunner, Rodrigo Meza, Jens Grivolla, Joan Codina, Rafael Banchs (Kaltenbrunner y otros, 2010)	2.010	*Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system*	Sistema de bicicleta compartida Bicing de Barcelona	Series de tiempo: ARMA (*Auto-regressive Moving Average*). Prueba usando predictor básico: Predicción basada en gradiente para una estación.	Disponibilidad de bicicletas.	Fecha y hora, Número de bicicletas salidas de estación, Disponibilidad de bicicletas, Espacios disponibles.
2	Yexin Li, Yu Zheng, Huichu Zhang, Lei Chen (Li y otros, 2015)	2.015	*Traffic Prediction in a Bike-Sharing System*	Ciudades de Nueva York y Washington D.C.	Predicción jerárquica (HP) basada en la agrupación de estaciones (BC) e inferencia basada en múltiples similitudes (MSI); predicción de llegadas basada en la agrupación de estaciones, transición entre grupos y duración del viaje (P-TD). También utiliza modelos HA para valores promedio de períodos de tiempo y días de la semana; ARMA para predecir valores futuros en una serie temporal entrada / salida, árboles de regresión GBRT para predecir patrones; HP-	Número de bicicletas que serán rentadas desde y hacia cada estación en un período futuro, para que la reasignación pueda ser ejecutada con anticipación.	Información georeferenciada, Fecha y hora, Tiempo de trayectoria, Número de bicicletas salidas de estación, Número de bicicletas llegadas de estación, Fin de semana, Temperatura aire, Valores climáticos específicos, Velocidad viento.

#	Autores	Año	Título	Lugar de análisis	Modelos Utilizados	¿Qué predice?	Variables Independientes
					KNN método jerárquico para predecir a nivel general y luego a cada grupo; y GC para dividir la ciudad en rejillas y agruparlas estaciones dentro de la misma rejilla.		
3	Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, Thomas Moscibroda (Yang y otros, 2016)	2.016	*Mobility Modeling and Prediction in Bike-Sharing Systems*	Sistema de bicicletas públicas compartidas Hangzhou *Public Bicycle Transport Service Development Co.*	*Random Forest* se compara con ARMA, HA, HP-MSI. Para la predicción se utiliza sólo datos con las características buscadas.	Tráfico temporal entre estaciones.	Capacidad de Estaciones, Fecha y hora, Distancias entre estaciones, Tiempo de trayectoria, Número de bicicletas salidas de estación, Número de bicicletas llegadas de estación, Disponibilidad de bicicletas, Espacios disponibles, Fin de semana, Día trabajo, Temperatura aire, Humedad, Valores climáticos específicos, Velocidad viento.
4	Hong Yang, Kun Xie, Kaan Ozbay, Yifang Ma, Zhenyu Wang	2.018	*Use of Deep Learning to Predict Daily Usage of Bike Sharing Systems*	Sistema de Bicicleta Citi Bike en New York City (NYC)	Utiliza aprendizaje profundo (DL) y redes neuronales convolucionales (CNNs); para la comparación	Uso diario de bicicletas compartidas a nivel de	Origen y destino del viaje, Información geo referenciada, Fecha y hora, Número de bicicletas salidas de

#	Autores	Año	Título	Lugar de análisis	Modelos Utilizados	¿Qué predice?	Variables Independientes
	(Yang y otros, 2018)				utilizó NN y series temporales ARIMA.	ciudad y de estaciones.	estación, Fin de semana, Día trabajo, Temperatura aire, Valores climáticos específicos, Precipitación, Caracterización Usuario.
5	Xuehong Gao, Gyu M. Lee (Gao & Lee, 2019)	2.019	*Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning*	Washington D.C., USA	Red de propagación hacia atrás previo una clasificación mediante un algoritmo genético híbrido de C-medias difusas.	Demanda del uso del sistema de bicicleta compartida.	Fecha y hora, Fin de semana, Día trabajo, Temperatura aire, Humedad, Valores climáticos específicos, Velocidad viento.
6	Sergio Guidon, Daniel J. Reck, Kay Axhausen (Guidon y otros, 2020)	2.020	*Expanding a(n) (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests*	Sistema de Bicicletas compartidas "Smide" en Suiza.	Utiliza regresiones espaciales y el algoritmo de *random forest*.	Número de reservas por zona, analiza demanda para predicción en una nueva estación de bicicleta compartida.	Origen y destino del viaje, Información geo referenciada, Número de bicicletas salidas de estación, Información relacionada con entorno geográfico, Cercanía con lugares de interés, Infraestructura Ciclovías, Cercanía medios de transporte masivo, Densidad poblacional.

#	Autores	Año	Título	Lugar de análisis	Modelos Utilizados	¿Qué predice?	Variables Independientes
7	Yi Cao y Yixiao Wang (Cao & Wang, 2022)	2.022	*Shared Cycling Demand Prediction during COVID-19 Combined with Urban Computing and Spatiotemporal Residual Network*	Gobierno Municipal de Shenzhen	Computación urbana y red residual de atención espacio-temporal, USTARN.	Predicción demanda de uso sistema de bicicleta compartida.	Origen y destino del viaje, Información geo referenciada, Fecha y hora, Tiempo de trayectoria, Fin de semana, Día trabajo, Información específica de COVID, Temperatura aire, Valores climáticos específicos, Velocidad viento.
8	Bustamante, X., Federo, R., & Fernández-i-Marin, X. (Bustamante y otros, 2022)	2.022	*Riding the wave: Predicting the use of the bike-sharing system in Barcelona before and during COVID-19*	Sistema de bicicletas compartido Bicing de Barcelona	Aprendizaje automático probabilístico en una investigación cuasi-experimental.	Uso del sistema bicicletas compartidas antes y durante el COVID.	Información geo referenciada, Fecha y hora, Número de bicicletas salidas de estación, Número de bicicletas llegadas de estación, Disponibilidad de bicicletas, Espacios disponibles, Fin de semana, Día trabajo, Temperatura aire, Humedad, Velocidad viento, Precipitación, Información relacionada con entorno geográfico, Infraestructura Ciclovías.

Tabla 1 Tabla comparativa de bibliografía consultada para el uso de modelos de predicción de BSS. Elaboración propia.

3 METODOLOGÍA

Esta investigación es un estudio comparativo con base cuantitativo longitudinal, en la que se añadieron a los datos variables externas como clima, feriado y fechas con restricciones para determinar si son relevantes o influyen en la predicción de tráfico y uso de bicicletas públicas. Como se puede ver en la Ilustración 3, de manera general, para la selección del modelo de predicción se realizaron los siguientes procesos: Se recopilaron y limpiaron los datos, se analizaron datos de distintas fuentes; se realizó la limpieza de datos para eliminar datos fuera de rango aceptable; se generaron archivos con los datos ya procesados; se analizaron los datos para determinar características y correlaciones; se seleccionaron las variables para aplicar las predicciones mediante varios modelos. Por lo que se generaron al menos dos archivos uno para entrenamiento y otro para validación de acuerdo para cada modelo y tipo de predicción. Se generaron, analizaron los resultados de los distintos modelos y se seleccionaron los modelos que se acercan mejor a predecir los datos de uso de las estaciones como salidas y como llegadas.

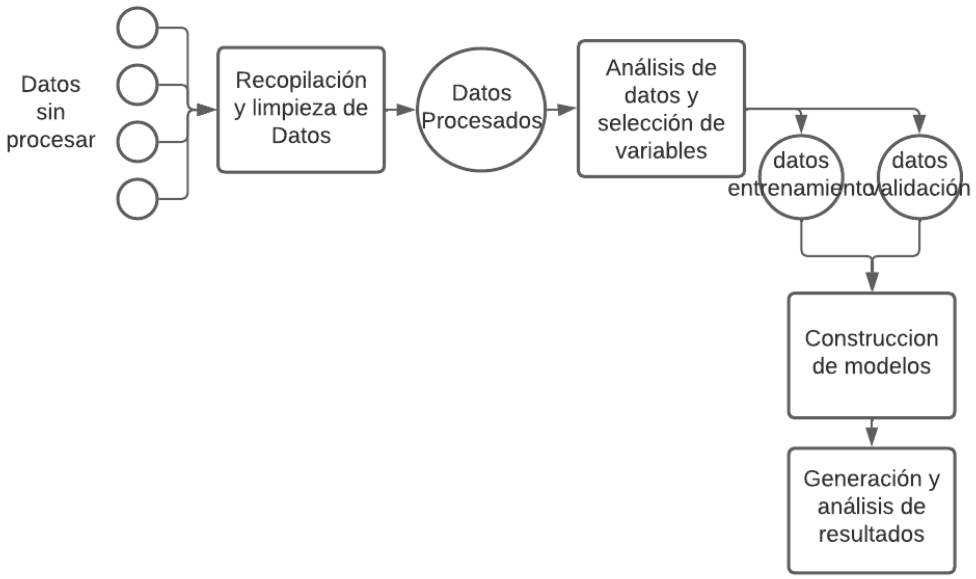


Ilustración 3 Esquema del proceso de selección de modelo de predicción. Elaboración propia.

3.1 Recopilación y limpieza de datos

Los datos principales con los que se desarrolló el presente proyecto son los registros de uso de la bicicleta pública BiciCuenca desde 1 de abril 2.019 hasta 31 de diciembre 2.022 (EMOV EP., 2023). Se recibieron 76.952 registros, de los que se eliminaron registros con estaciones de prueba, otros que no contenían origen y/o destino, quedando para un análisis inicial 76.519 registros. En el Anexo 1 se muestra una tabla en la que se describen los datos, ahí se puede identificar 2.582 registros con estado "Cancelado" y también los que tenían un uso fuera de los rangos normales, por lo que, se realizó una winsorización que consiste en eliminar los datos extremos, el 5% superior y el 5% inferior, ya que según (García-Santana & Pérez-Canto, 2021) reduce su influencia en las estimaciones de los parámetros en el análisis de datos, lo que puede mejorar la precisión y la fiabilidad de los datos.

Para agregar las características climáticas a las estaciones, se añadió información que se encuentra en el intervalo de estudio: 1 de abril 2.019 hasta 31 de Diciembre 2.022 (IERSE, 2023), en la que se recibieron 8.631.265 registros climáticos, entre ellos se encontraban registros no válidos, se unificaron los registros de los sensores válidos en 1.229.056 registros con las características que se muestran en el Anexo 2. De estos registros, se filtraron e unificaron por fecha los datos de humedad relativa, presión atmosférica, temperatura ambiental y ruido de los datos; se establecieron rangos de datos válidos y se eliminaron los valores fuera de estos; se agregaron los datos por estación de sensores obteniendo las medianas de los valores con intervalos horarios; se interpolaron los valores faltantes de los sensores climáticos y se creó un listado con sus ubicaciones, calculándose las distancias a las estaciones del sistema BSS para seleccionar la más cercana.

Al analizar los datos de bicicletas se detectaron rangos de fechas en los que no hay registros, coincidiendo con feriados, restricciones de movilidad por pandemia y por paros nacionales. En el Anexo 3 se muestra el detalle de dichas fechas. Para este análisis, se creó un listado de fechas con feriados (C.A. EL UNIVERSO, 2018) (C.A. EL UNIVERSO, 2019) (C.A. EL UNIVERSO, 2021), paros nacionales (Pérez Torres, 2019)

(Colaboradores de Wikipedia, 2023) y restricciones por pandemia (Colaboradores de Wikipedia, 2023).

Para consolidar los datos, se creó una serie de registros periódicos horarios durante el intervalo de estudio para todas las estaciones de bicicletas. Se agregaron los datos de fecha: fecha, fecha transformada a horas, año, mes, día, día de la semana y hora del día. Luego se unieron las características de la estación de bicicleta, incluyendo su identificador, nombre, longitud, latitud y altura. Se agregaron características especiales de las fechas, como restricciones de movilidad por pandemia, paros y días feriados. Las características del clima se seleccionaron y agruparon las lecturas del sensor más cercano a la estación de bicicleta: humedad relativa, presión atmosférica, temperatura ambiental y ruido ambiental. Por último, se calculó la cantidad de bicicletas cuyo origen y destino fueron la estación.

En resumen, se crearon registros horarios para las veinte estaciones de bicicletas, generándose así 31.519 registros para cada estación dando un total de 630.380 registros con las cantidades de uso para origen y destino, es decir las llegadas y salidas por estación.

3.2 Análisis de datos y selección de variables

Una vez realizada la limpieza de los datos, se crearon ilustraciones para mostrar gráficamente la relación entre las variables estudiadas, para determinar las relaciones entre pares de variables se realizó una matriz de correlación. Se seleccionaron para el estudio las variables con mayor valor de correlación. La matriz de correlación $\mathrm{Corr(X)} = \mathrm{SD}^{-1} * \mathrm{Cov(X)} * \mathrm{SD}^{-1}$, donde SD^{-1} es la inversa de la matriz de desviaciones estándar y Cov(X) es la matriz de covarianza.

3.3 Construcción de modelos

Se estableció para todos los modelos dos predicciones para llegadas y salidas, en las que se requirieron dos grupos de datos: para el entrenamiento y para validación, se analizaron las combinaciones para fecha de inicio del entrenamiento y fecha inicio de la validación: para el inicio del entrenamiento se seleccionaron dos fechas hitos: al inicio el

servicio de BSS, y el 18 mayo de 2.020, fecha en la que se reanudó el servicio después de las restricciones por pandemia; para el inicio de la validación: enero 2.022 y otro para diciembre 2.022. Otro parámetro que se evaluó es si se utilizaron los datos de todas las estaciones para el entrenamiento y validación ó si se realizó paralelamente el análisis por cada estación y luego se juntaron los resultados. Adicionalmente para la predicción de las llegadas, se incluyó en algunos casos la variable salidas que fue previamente predicha.

Para el modelo con serie de tiempo se utilizó únicamente la fecha y la cantidad de llegadas y salidas para cada predicción. En la aplicación del modelo neuronal básico, se utilizaron dentro de las entradas, valores anteriores a la fecha llamados *lags*, para que la predicción considere la sucesión de datos en el comportamiento diario de cada estación. Para otros predictores se eligieron los datos que tienen mayor correlación absoluta con las llegadas y salidas.

3.4 Generación y análisis de resultados

Para utilizar cada modelo, se calibraron con los datos de entrenamiento. Luego de obtenidos los resultados de predicciones, se tabularon y compararon estos resultados con los datos de validación mediante la utilización del "*Squared Correlation*" o coeficiente de determinación $r^2 = \sum(y\_pred\_i - \bar{y})^2 / \sum(y\_i - \bar{y})^2$, donde: y_pred_i es el valor predicho de la variable dependiente para la i-ésima muestra; $\bar{y}$ es la media de las observaciones; y_i: es el valor real de la variable dependiente para la i-ésima muestra. Otra medida con la que se midieron los datos es el "*Root Mean Squared Error*" raíz del error cuadrático medio: RMSE = $\sqrt{(1/n) * \sum(y\_i - y\_pred\_i)^2}$, también se incluirá el error absoluto $Error\ absoluto = \left(\frac{1}{n}\right)\sum(|y\_i - y\_pred\_i|)$ y el error relativo $Error\ relativo = \left(\frac{1}{n}\right)\sum\left(\left|\frac{y_i - y_{pred_i}}{y_i}\right|\right) * 100$, donde: n: es el número total de muestras u observaciones; y_i: es el valor real de la variable dependiente para la i-ésima muestra; y_pred_i: es el valor predicho de la variable dependiente para la i-ésima muestra.

En base a los resultados del coeficiente de determinación, se realizaron gráficas comparativas tanto para salidas, así como para llegadas en cada escenario de grupos de variables seleccionadas, dónde se eligieron los modelos que tienen el mayor valor de

coeficiente de determinación.

4 RESULTADOS

4.1 Recopilación y limpieza de datos

Una vez creados los intervalos de tiempo con periodicidad horaria entre el 1 de abril 2.019 hasta 31 de diciembre 2.022, con la información para cada estación, del clima, fechas de feriados, fechas de restricciones, cantidad de bicicletas que llegaron y salieron de cada en estación. Con respecto a los datos climáticos hay escenarios faltantes de ruido ambiental 27,45%, humedad relativa 0,47%, presión atmosférica 1,46% y temperatura ambiental 2,33%: por lo que se realizaron interpolaciones para rellenar espacios vacíos.

4.2 Análisis de datos y selección de variables

Para ilustrar el comportamiento de la frecuencia de uso del BSS BiciCuenca, se generó una gráfica con la cantidad de uso diario durante el intervalo de estudio desde el 1 de abril 2.019 hasta 31 de diciembre de 2.022.

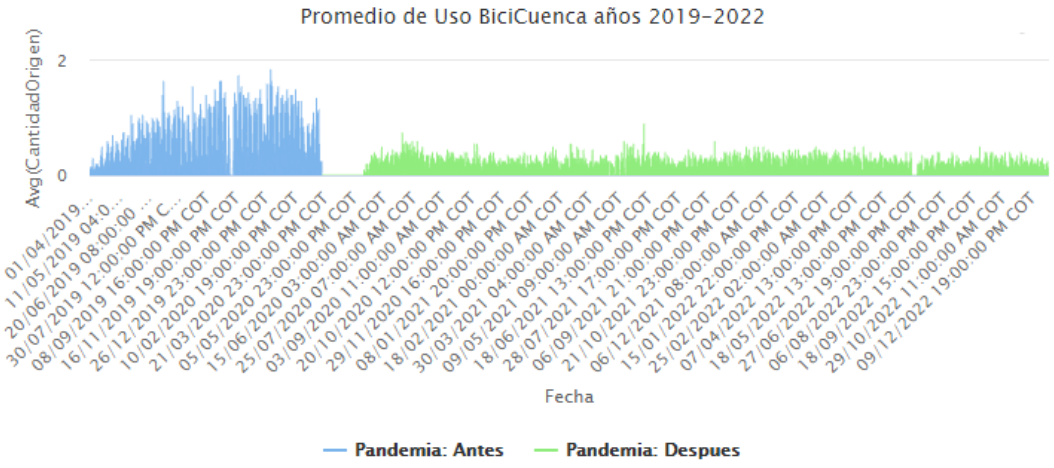


Ilustración 4 Gráfica promedio de uso del BSS Bicicuenca entre los años 2.019-2.022. Fuente de Datos: (EMOV EP., 2023). Elaboración propia.

En la Ilustración 4 se pueden distinguir 2 períodos distintos: antes y después de la pandemia COVID-19, en el Anexo 4 se muestra la cantidad de usuarios por hora en estos dos intervalos de tiempo.

Como se puede observar en Anexo 4 y Anexo 5: la cantidad de usuarios por hora del BSS BiciCuenca ha bajado a partir del inicio de la pandemia.

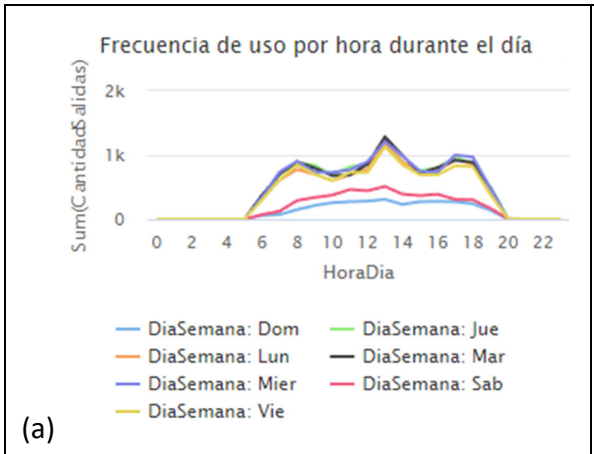
En el Anexo 5 el comportamiento de la frecuencia tanto para las salidas, así como para las llegadas en su mayoría es cero, esto se debe a que se consideró todas las horas del día como base.

En la Ilustración 4 se pueden distinguir espacios en cero o vacíos, en el Anexo 3 podemos ver a detalle los intervalos y los motivos aparentes: los paros nacionales registrados en los años 2.019 (Colaboradores de Wikipedia, 2022) y 2.022 (Colaboradores de Wikipedia, 2023) así como restricciones de movilidad ocasionadas por la pandemia COVID-19 el año 2.020 (Colaboradores de Wikipedia, 2023) son las posibles causas en las que posiblemente se suspendió el servicio. Otro caso, en el que no hay datos, es el primero de enero de todos los años de estudio y el 24 de mayo del 2.020 que pudo ser causada por el feriado de la Batalla de Pichincha (C.A. EL UNIVERSO, 2019), así como también a las restricciones de movilidad causadas por la pandemia COVID-19 registradas en la provincia (Colaboradores de Wikipedia, 2023). Lo que demuestra que las restricciones de movilidad inciden en el uso del BSS.

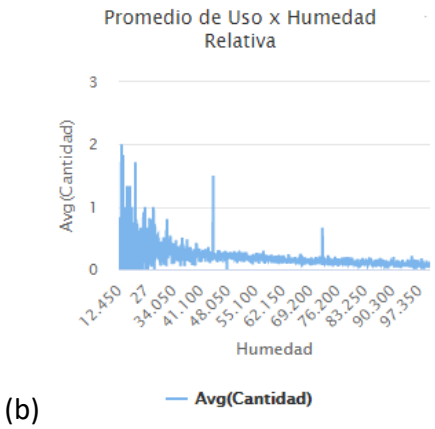
En la Ilustración 5 se muestran las gráficas de las relaciones de frecuencia de uso entre variables hora del día y día de la semana; el promedio de la cantidad por hora contrastadas con las variables de humedad relativa, ruido ambiental, temperatura ambiental y presión atmosférica. Para contrastar las imágenes, se realizó una matriz de correlación que se encuentra en el Anexo 6, en la que se determinaron relaciones entre pares de variables.

Como se ve en la Ilustración 5 (a), en las horas de inicio y fin del servicio, se registra un uso menor del sistema. A las 8H00 y 17H00 existe un valor máximo relativo y un máximo absoluto alrededor de las 13H00. El uso durante la semana se puede identificar la diferencia de uso entre semana y fines de semana. Siendo el uso entre semana mayor: registrándose los martes, miércoles y jueves una mayor cantidad de uso. Al analizar al comportamiento del uso durante la semana, se puede distinguir una disminución de uso en los fines de semana, existe una correlación negativa, lo que

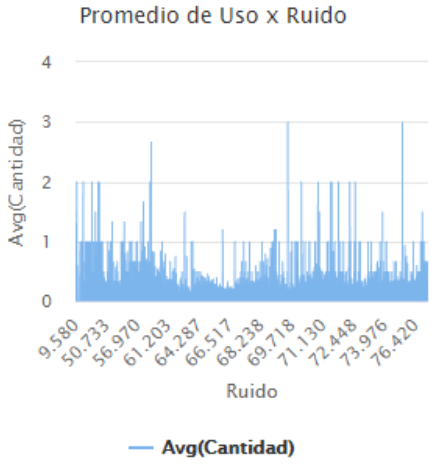
confirma que en los fines de semana se reduce el uso del sistema BSS. Con respecto a la humedad relativa se puede identificar que en el inicio se encuentra muy dispersa y lecturas con amplias diferencias. Cuando se incrementa el porcentaje de humedad, el promedio tiene oscilaciones menos amplias y se reduce. Existe una correlación negativa, mientras se registra mayor humedad, es menor el uso tanto para salidas como llegadas. Existe una correlación positiva con el ruido de 0,064, se puede interpretar como que en los lugares de mayor ruido se registra una mayor cantidad de uso del BSS.



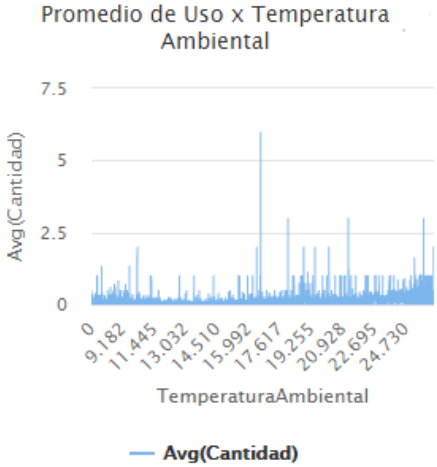
(a)



(b)



(c)



(d)

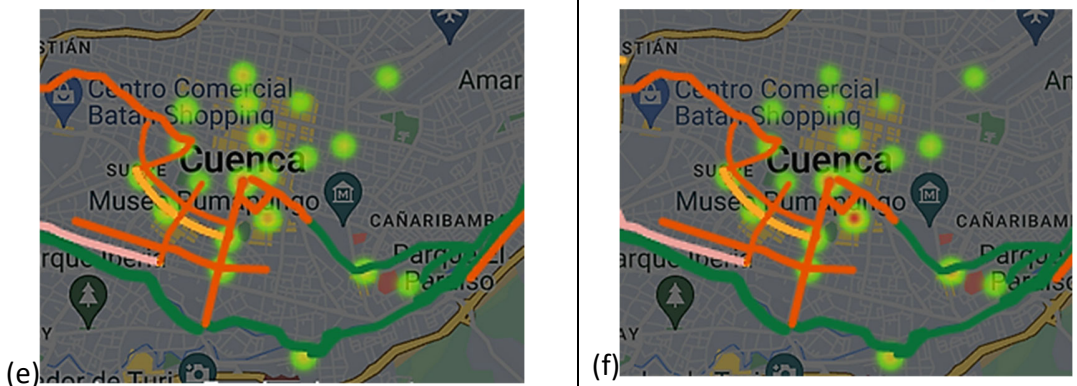


Ilustración 5 Análisis descriptivo del BSS Cuenca entre 18 Mayo 2.020 – 31 diciembre 2.022. Fuentes de Datos: (EMOV EP., 2023), (BiciCuenca, s.f.). Elaboración propia: (a) Frecuencia de uso por hora durante el día; (b) Promedio de uso x Humedad Relativa; (c) Promedio de uso x Ruido Ambiental; (d) Promedio de uso x Temperatura Ambiental; (e) Mapa de calor para salidas de estaciones; (f) Mapa de calor para llegadas de estaciones.

Si hacemos referencia al Anexo 6: revisando la correlación entre las llegadas y salidas con la temperatura ambiental, es positiva, en la Ilustración 5 (d) se distingue que: a mayor temperatura, mayor uso del BSS. Un dato importante a resaltar es que la correlación de la altura de la estación con la cantidad de salidas es positiva y con la cantidad de llegadas es negativa, por lo que se podría interpretar que las estaciones de origen a una altura mayor son más frecuentes lugares de salida y los lugares a menor altura son más frecuentes para destinos. En la Ilustración 5 (e) se puede ver un mapa de calor para salidas, en la que se pueden identificar que las estaciones de salidas más utilizadas son las que se encuentran a mayor altura, mientras que en la Ilustración 5 (f) se evidencia un mayor uso de las estaciones con menor altura. Con respecto a la relación positiva entre la cantidad de salidas y llegadas, se encontró una correlación alta, la máxima de entre las variables. A mayor cantidad de salidas, mayor cantidad de llegadas.

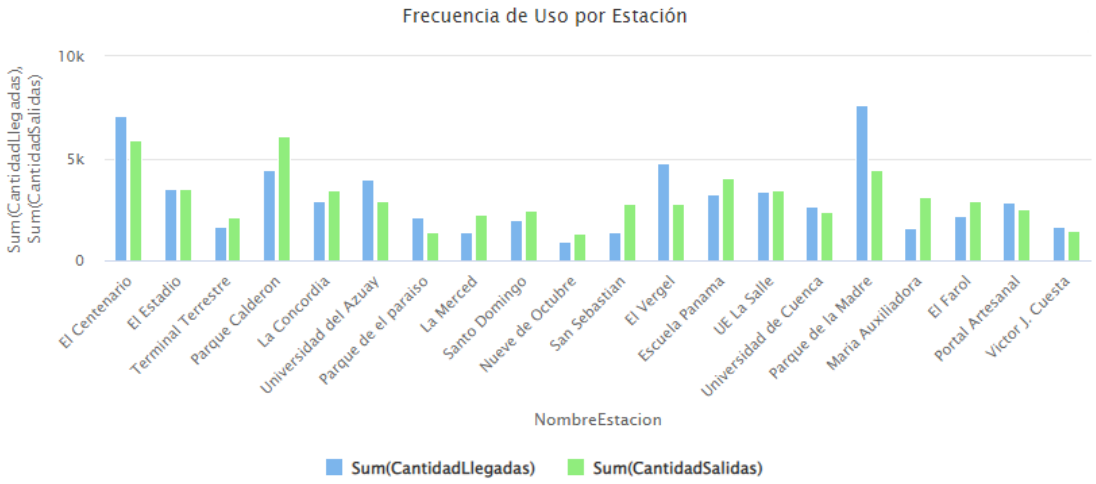


Ilustración 6 Cantidad de salidas y llegadas por estación entre 18 Mayo 2.020 – 31 diciembre 2.022. Fuentes de Datos: (EMOV EP., 2023). Elaboración propia.

En la Ilustración 6, se puede identificar que los lugares desde los que se registra una mayor cantidad de salidas son El Centenario, Parque Calderón y Escuela Panamá; mientras que la mayor cantidad de llegadas se registra en El Centenario, Parque de la Madre y El Vergel.

De la matriz de correlación se eligieron las variables con mayor correlación absoluta: humedad, temperatura ambiental, ruido, estación de bicicleta, fin de semana, fecha y hora del día donde el coeficiente de correlación absoluto fue alto.

En la Tabla 2 se pueden observar los detalles de los datos: registros horarios para las veinte estaciones de bicicletas, generándose así 31.519 registros para cada estación dando un total de 630.380 registros con las cantidades de uso para origen y destino.

Variable	Descripción de Variable	Unidad de Medida	Tipo	Estadísticas		
CantidadOrigen	Cantidad de registros de bicicleta que salieron de una estación por hora.	Unidades	Entero	**Mínimo**	**Máximo**	**Media**
				0	14	0,097
CantidadDestino	Cantidad de registros de bicicleta que llegaron a una estación por hora.	Unidades	Entero	**Mínimo**	**Máximo**	**Media**
				0	11	0,097
Fecha	Fecha del registro.		Fecha Hora	**Mínimo**	**Máximo**	**Duración**
				1/4/2.019 00:00	31/12/2.022 18:00	1.370 días, 18 horas, 0 minutos, 0 segundos
id_estacion_bicicleta	Identificador de la estación de bicicletas.		Entero	**Mínimo**	**Máximo**	**Media**
				1	23	11,450
HoraDia	Hora del día en la que se realizó el registro.		Entero	**Mínimo**	**Máximo**	**Media**
				0	23	11,450
FinSemana	1 si es sábado o domingo; 0 días entre semana.		Entero	**Mínimo**	**Máximo**	**Media**
				0	1	0,284
AlturaEstacion	Altura de la estación.	Metros	Real	**Mínimo**	**Máximo**	**Media**
				2.487,943	2.556,916	2.529,906
Humedad	Porcentaje de humedad registrada.	Porcentaje de Humedad	Real	**Mínimo**	**Máximo**	**Media**
				12,450	100	76,183
Ruido	Ruido registrado.	Decibelios	Real	**Mínimo**	**Máximo**	**Media**
				9,580	90,295	65,458
TemperaturaAmbiental	Temperatura ambiental registrada.	Grados Centígrados	Real	**Mínimo**	**Máximo**	**Media**
				0	31,340	15,883

Tabla 2 Descripción de los datos seleccionados para las predicciones. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

4.3 Construcción de modelos

Se definieron 4 escenarios principales como se puede ver en la Tabla 3 para realizar las predicciones, los que se ejecutaron utilizando todos datos de manera general y parcialmente, de cada estación a la vez, tanto para predecir las salidas, así como para predecir las llegadas. En las llegadas se realizaron pruebas incluyendo la variable predicha de las salidas.

Escenarios de datos para las predicciones				
Nomenclatura	Fecha intervalo de entrenamiento	Cantidad de datos para entrenamiento	Fecha Intervalo para validación	Cantidad de datos para validación
E1	1/abril/2.019 – 1/enero/2.022	57.480	1/enero/2.022 – 31/diciembre/2.022	21.580
E2	1/abril/2.019 – 1/diciembre/2.022	77.220	1/diciembre/2.022 – 31/diciembre/2.022	1.840
E3	18/mayo/2.020 – 1/enero/2.022	34.660	1/enero/2.022 – 31/diciembre/2.022	21.580
E4	18/mayo/2.020 – 1/diciembre/2.022	54.400	1/diciembre/2.022 – 31/diciembre/2.022	1.840

Tabla 3 Escenarios de datos para las predicciones. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

4.4 Generación y análisis de resultados

Al aplicar el modelo de *random forest* se obtuvo el coeficiente de determinación de 0,057 que es demasiado bajo, pues prácticamente no predice, ya que la mayoría de los datos son cero que puede deberse, a que según la distribución del Anexo 5 la presencia de valores cero es mayoritaria; también se realizó una clasificación en la que se utilizaron los métodos de árboles de decisión, *random forest, deep learning*: los mejores resultados realizados se lograron con deep learning, obteniendo una exactitud entre 91,21% y 92,5%, pero la cantidad predicha es mayoritariamente cero y en el mejor de los casos predijo 75

de las 2.177 veces que se encuentra el valor uno; valores mayores no acierta en la predicción. Al utilizar otros métodos, estos predijeron únicamente valores de cero o cercanos a cero, pues representan casi el 93,83% del total de los datos, por lo que se decidió establecer intervalos temporales mayores para obtener resultados que puedan ser relevantes y aporten al estudio. Dado que el uso del sistema es reducido, se establecieron los siguientes intervalos: fuera de horario desde 22H00 a 5H00, mañana hasta las 12H00, tarde hasta 18H00 y noche hasta 22H00. En el Anexo 7, se puede observar la matriz de correlación con la clasificación de intervalos de horas, las correlaciones con las cantidades de llegadas y salidas se acentuaron, siendo más representativas las variables de hora del día: en la mañana y tarde tienden a subir, en la noche y madrugada la relación es negativa entonces la cantidad baja; además de una mayor representatividad las variables de temperatura ambiental, ruido ambiental, humedad relativa.

Al implementar el método de ARIMA con los datos de la estación más utilizada que es el Centenario, utilizando únicamente valores salidas con *lag* de quince para que se incluyan las 15 horas en las que está el estudio y sin *lag* el RMSE fue 5,1633E+137 +/- 8,1965E+138. representando un error excesivamente alto, por lo que se descartó el análisis con esta metodología. Esta decisión se respaldó con la función de autocorrelación ACF, Anexo 10, y la función de autocorrelación parcial PACF, en las que se encontraron que los p-valores no son significativos, es decir, los datos no son estacionales.

Para la predicción de las salidas se realizaron dos técnicas: análisis general, en el que se utilizaron las variables seleccionadas con la información de todas las estaciones; el análisis paralelo, se analizó individualmente la información de cada estación, se juntaron los resultados de predicciones. En la Tabla 3 se encuentra una descripción de los escenarios generados y en el Anexo 8 se encuentran los resultados de las predicciones de salida. En la Ilustración 7 se puede distinguir que, entre los modelos utilizados, el mejor modelo para predecir los datos de salida es el de *random forest* porque tiene los más altos valores de coeficiente de determinación, entre ellos el más alto es 0,225; Con respecto a los valores de los errores RMSE, error absoluto y error relativo no se distingue por poseer ningún valor extremo.

Para la predicción de las llegadas se utilizó el análisis general con las variables seleccionadas, como se presenta en la Tabla 3 y en el Anexo 9. Una variación del análisis

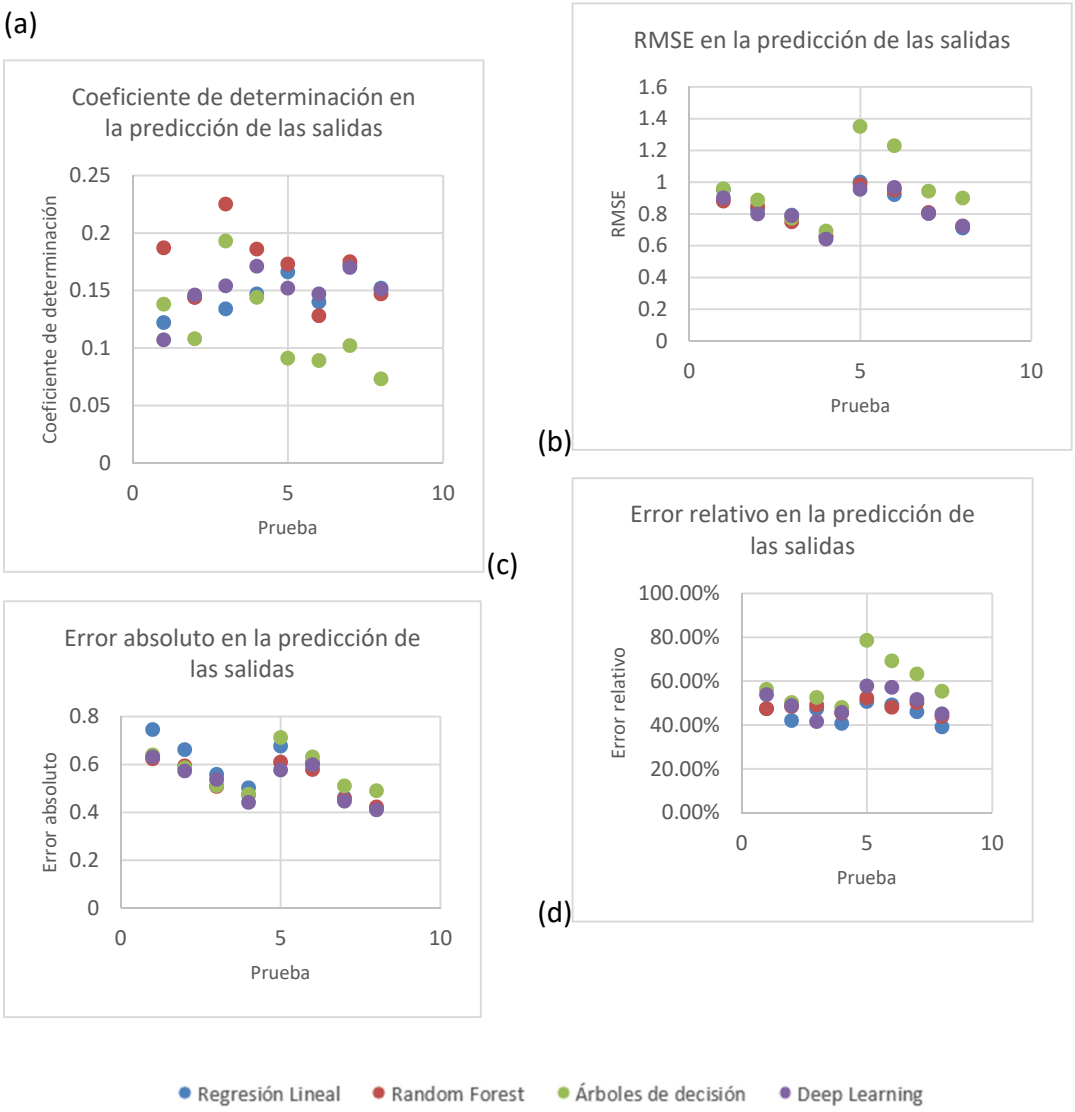


Ilustración 7 Gráficas de los resultados de pruebas para predicciones de salida: (a) del coeficiente de determinación; (b) RMSE; (c) error absoluto; (d) error relativos. En la Tabla 3 se encuentran los escenarios aplicados y en el Anexo 8 los resultados obtenidos. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

general y análisis paralelo en la que se incluyeron los valores de la cantidad de salidas predicha: es decir, la cantidad de salidas para cada una de las 20 estaciones.

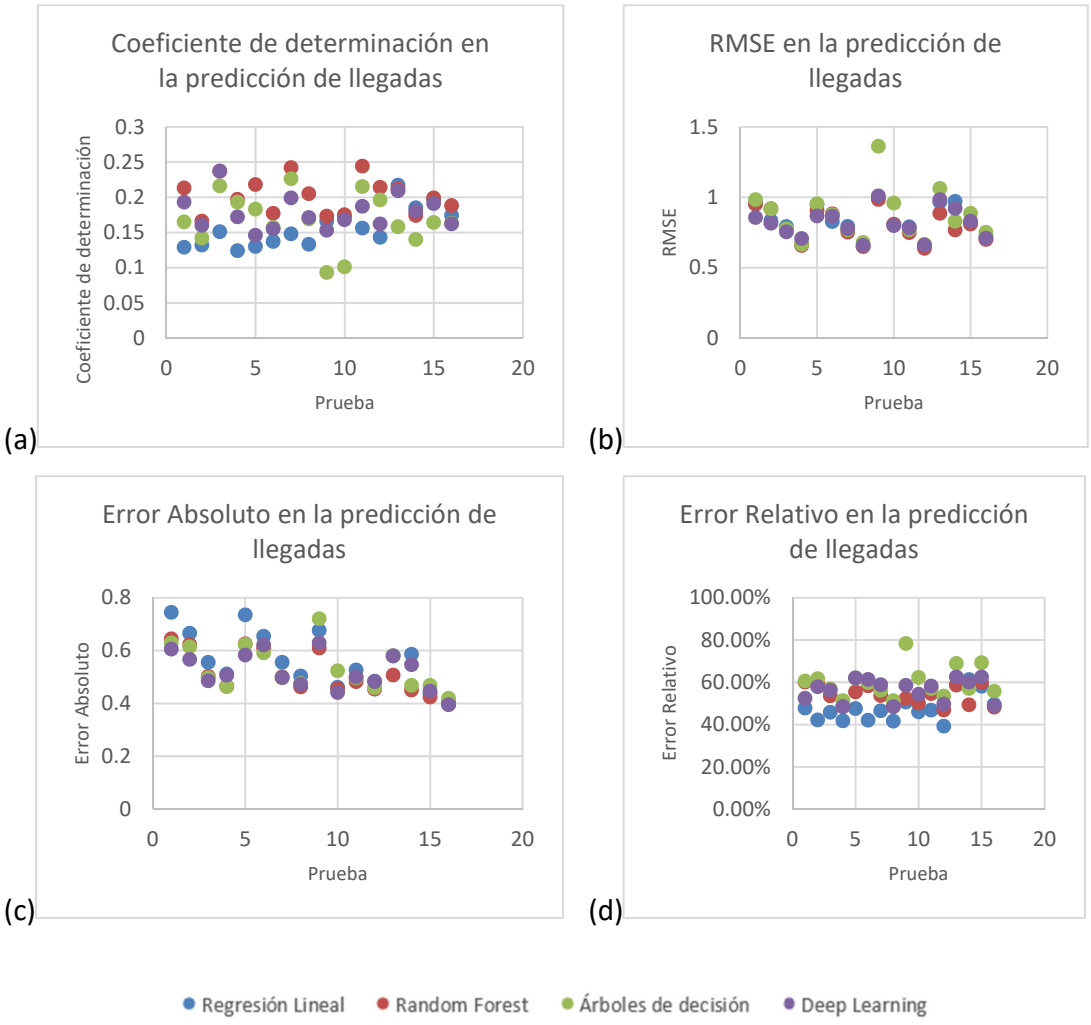


Ilustración 8 Gráficas de los resultados de pruebas para predicciones de llegadas: (a) del coeficiente de determinación; (b) RMSE; (c) error absoluto; (d) error relativos. En la Tabla 3 se encuentran los escenarios aplicados y en el Anexo 9 los resultados obtenidos. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

En la Ilustración 8 se muestran los resultados de las predicciones de las llegadas. Para las predicciones que utilizaron como entrada la variable de la cantidad de salida, se utilizaron las predicciones obtenidas por el método de *random forest* que fue generado utilizando el análisis general con predicciones de salidas con un coeficiente de determinación de 0,225.

Como se puede ver en el Anexo 8 y Anexo 9 los escenarios que tienen mejores resultados, es decir, mayor coeficiente de determinación y menores errores absoluto, relativo y RMSE es el escenario 3: Datos de entrenamiento entre 18/mayo/2.020 y

1/enero/2.022; y datos de validación entre 1/enero/2.022 y 1/diciembre/2.022.

Modelo	Predicción Salidas		Predicción Llegadas			
			Análisis General		Análisis Paralelo por estación	
	Análisis General	Análisis Paralelo por estación	No incluye salidas	Incluye salidas	No incluye salidas	Incluye salidas
Error Absoluto						
Regresión Lineal	0,558	0,461	0,555	0,555	0,526	0,437
Random Forest	0,507	0,455	0,502	0,5	0,482	0,424
Árboles de decisión	0,511	0,51	0,497	0,496	0,496	0,468
Deep Learning	0,537	0,446	0,485	0,498	0,5	0,446
Error Relativo						
Regresión Lineal	47,27%	45,96%	45,73%	46,45%	46,79%	58,03%
Random Forest	48,96%	50,11%	53,58%	53,71%	54,56%	59,73%
Árboles de decisión	52,46%	63,21%	56,86%	56,13%	57,01%	69,24%
Deep Learning	41,46%	51,57%	56,15%	58,76%	58,19%	62,45%
RMSE						
Regresión Lineal	0,792	0,803	0,791	0,792	0,789	0,815
Random Forest	0,75	0,808	0,76	0,752	0,748	0,809
Árboles de decisión	0,772	0,943	0,775	0,768	0,769	0,885
Deep Learning	0,788	0,803	0,754	0,776	0,781	0,83
Coeficiente de Determinación						
Regresión Lineal	0,134	0,172	0,151	0,148	0,156	0,197
Random Forest	0,225	0,175	0,237	0,242	0,244	0,199
Árboles de decisión	0,193	0,102	0,216	0,226	0,215	0,164
Deep Learning	0,154	0,17	0,237	0,199	0,187	0,191

Tabla 4 Resultados del mejor escenario: Datos de entrenamiento entre 18/mayo/2.020 y 1/enero/2.022; y datos de validación entre 1/enero/2.022 y 1/diciembre/2.022. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

Como se puede ver en la Tabla 4, el mejor escenario está con *random forest*, que tiene el coeficiente de determinación más alto: para las salidas es 0,225, y para las llegadas es 0,244 también el RMSE, error absoluto y error relativo es el menor.

Para la predicción de salidas el mejor escenario fue el que utilizó datos de entrenamiento a partir del 18 de mayo del 2.020 y 1 de enero del 2.022; y datos de validación entre 1 de enero del 2.022 y 1 de diciembre del 2.022 utilizando un análisis general.

Para la predicción de llegadas el mejor escenario fue el que utilizó datos de entrenamiento a partir del 18 de mayo del 2.020 y 1 de enero del 2.022; y datos de validación entre 1 de enero del 2.022 y 1 de diciembre del 2.022, utilizando un análisis paralelo, es decir, se tomó la información de cada estación independientemente sin utilizar para el entrenamiento valores de salidas predichas previamente.

En la Tabla 5 se presenta una muestra de los valores obtenidos mediante las predicciones de entradas y salidas. En las predicciones de salida dado que el entrenamiento fue en general: en el horario 05H00 - 12H00 hay desaciertos en la predicción de ceros en estaciones de bajo uso; en el horario entre 12H00 - 18H00 se podría decir que hay mayor cantidad de valores acertados; en el horario entre 18H00 - 22H00, todas las predicciones son cero. En las predicciones de llegadas, dado que el entrenamiento fue en paralelo, en el horario 05H00 - 12H00 hay mayor cantidad de precisión en los aciertos, en el horario entre 18H00 - 22H00 todas las predicciones son cero.

Variable	Salidas						Llegadas					
Horario	05H00 - 12H00		12H00 - 18H00		18H00 - 22H00		05H00 - 12H00		12H00 - 18H00		18H00 - 22H00	
Estación	Valor	Predicción	Valor	Predicción	Valor	Predicción	Valor	Predicción	Valor	Predicción	Valor	Predicción
El Centenario	0	2	4	3	0	0	2	3	1	2	0	0
Parque Calderón	1	1	3	4	0	0	1	1	2	0	0	0
La Concordia	4	2	0	1	1	0	1	1	2	1	0	0
Universidad del Azuay	1	2	1	1	0	0	2	1	3	1	1	0
Parque de el paraíso	0	0	0	1	0	0	0	0	1	1	0	0
La Merced	0	1	2	1	1	0	2	1	1	0	0	0
Santo Domingo	0	1	1	1	0	0	0	0	2	0	0	0
Nueve de Octubre	0	1	0	1	0	0	0	0	1	0	0	0
San Sebastián	0	1	1	1	0	0	0	0	0	0	0	0
El Vergel	0	1	1	1	0	0	0	1	1	1	2	0
Escuela Panamá	3	2	2	2	0	0	1	2	2	1	0	0
UE La Salle	1	2	1	1	0	0	1	1	3	1	0	0
Universidad de Cuenca	1	1	0	3	1	0	0	0	0	0	1	0
Parque de la Madre	1	1	1	1	1	0	1	2	3	1	1	0
María Auxiliadora	1	1	2	1	0	0	0	0	0	0	0	0
El Farol	0	1	4	1	0	0	0	2	1	1	0	0
Portal Artesanal	0	1	0	1	0	0	0	0	0	1	0	0
Víctor J. Cuesta	0	0	1	1	0	0	0	0	0	0	0	0
El Estadio	0	1	0	1	0	0	2	1	1	1	0	0
Terminal Terrestre	0	1	1	1	0	0	0	0	0	0	0	0

Tabla 5 Muestra de predicciones de salidas y llegadas para la fecha 18 de mayo 2.022: Datos de entrenamiento entre 18/mayo/2.019 y 1/enero/2.022; y datos de validación entre 1/enero/2.022 y 1/diciembre/2.022. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

5 DISCUSIÓN

El uso del BSS BiciCuenca disminuyó a partir de la pandemia, en el 2.021, se construyeron nuevas ciclovías para fomentar el uso de bicicletas. Sin embargo, como la fecha del estudio finaliza en diciembre de 2.022, no se pudo evidenciar un aumento en el uso. Para futuras investigaciones, se recomienda analizar variables que incluyan aspectos sociales del uso del BSS, como la cantidad de afiliados o personas que tienen activa la suscripción para acceder a este servicio, edad, género e historial de uso.

Hubo fechas en las que no hay información de uso: 1 de enero, restricciones de pandemia, restricciones de movilidad por huelga. Después de la pandemia, a partir del 18 de mayo del 2.020, el comportamiento de uso del BSS cambió. A pesar de que las variables de restricciones tienen una clara influencia en el uso, no se consideraron para la predicción, ya que son características únicas.

Con respecto a las variables climáticas, hubo rangos de fechas sin datos y se realizaron interpolaciones, lo que podría haber afectado el análisis y la relación entre variables para las predicciones, ya que varían según la hora del día, el mes del año y la ubicación. Para futuros trabajos, se recomienda incluir la variable de lluvia y la cantidad de lluvia, así como reemplazar los datos faltantes generándolos para cada estación de sensor para simular el comportamiento con un análisis estacional y espacial.

Al extender el intervalo de tiempo de horas a mañana, tarde y noche, se disiparon los valores de humedad, ruido, presión y temperatura, y se acentuó la correlación entre estas variables con las de salida y llegada como se puede observar en el Anexo 6 y Anexo 7. La desventaja es que se perdió la calidad de las variables de entrada y se redujeron los datos para el entrenamiento y la creación de los modelos.

Los p-valores no significativos en el análisis de la función de autocorrelación ACF, presentes en el Anexo 10, y la función de autocorrelación parcial PACF: sugieren que no hay correlación significativa entre las observaciones en el conjunto de datos de

salidas o llegadas. No hay patrones de dependencia lineal significativos entre las variables a lo largo del tiempo, es decir, no existe estacionalidad o patrones repetitivos. Esto puede deberse a la falta de cantidad de datos o ausencia de factores significativos que afecten a las variables en el tiempo.

Para futuros trabajos se recomienda dividir en sub-escenarios, es decir, hacer predicciones de acuerdo a características específicas, como predicciones por hora, por estación, así como en eventos específicos, como paros, feriados, semanas laborales y fines de semana. Dado que el análisis paralelo de cada estación por individual, tuvo buenos resultados que se ajustan a cada estación. Se recomienda recolectar datos por estación en períodos mayores de tiempo para mejorar precisión en las predicciones.

6 CONCLUSIÓN

Se utilizaron los datos disponibles de uso del sistema de bicicletas públicas BiciCuenca en intervalos horarios para la ciudad de Cuenca. Sin embargo, dado que la cantidad de uso por hora era muy baja, el 93,83% de los datos estaban representados por ceros. Por lo tanto, se decidió aumentar los intervalos temporales para obtener resultados más significativos y relevantes para el estudio. Para ello, se agruparon los intervalos temporales para la mañana de 05H00 a 12H00, tarde de 12H00 a 18H00 y noche de 18H00 a 22H00 en todas las estaciones.

Para predecir la demanda del sistema de bicicletas públicas en Cuenca, se seleccionaron las variables con mayor correlación absoluta con la cantidad de salidas y cantidad de llegadas, entre las que se incluyen la estación de bicicleta, temperatura ambiental, humedad, ruido, fin de semana, altura de la estación y hora del día.

El mejor escenario para las predicciones de salidas y llegadas se obtuvo utilizando datos de entrenamiento desde el 18 de mayo del 2.020 hasta antes del 1 de enero del 2.022; y datos de validación desde el 1 de enero del 2.022 hasta el 31 de diciembre del 2.022. El modelo con un mayor coeficiente de determinación fue el de *random forest*, frente a modelos de regresión lineal, árboles de decisión y *deep learning*: para las salidas obtuvo 0,225, y para las llegadas es 0,244. Además, se obtuvieron valores menores de

RMSE, error absoluto y error relativo. A pesar de que estos modelos fueron los mejores, aún se necesitan mejoras para la predicción de la demanda del BSS BiciCuenca.

7 REFERENCIAS

Bachand-Marleau, J., Lee, B. H., & El-Geneidy, A. M. (2012). Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use. *Transportation Research Record, 2314*(1), 66-71.

BiciCuenca. (s.f.). *Acerca del Proyecto*. Retrieved 31 de Enero de 2022, from https://www.bicicuenca.com/sobre.aspx

BiciCuenca. (s.f.). *Mapa de las Estaciones*. Retrieved 28 de Marzo de 2023, from Bici Pública Cuenca: https://www.bicicuenca.com/mapaestacao.aspx

Bustamante, X., Federo, R., & Fernández-i-Marin, X. (2022). Riding the wave: Predicting the use of the bike-sharing system in Barcelona before and during COVID-19. *Sustainable Cities and Society*, 83, 103929.

C.A. EL UNIVERSO. (29 de Octubre de 2018). Calendario de feriados del 2019 en Ecuador. *El Universo*. https://www.eluniverso.com/noticias/2018/10/25/nota/7015913/feriados-ecuador-2019/

C.A. EL UNIVERSO. (22 de Noviembre de 2019). Feriados en Ecuador durante el 2020. *El Universo*. https://www.eluniverso.com/noticias/2019/11/22/nota/7615766/feriados-ecuador-2020/

C.A. EL UNIVERSO. (10 de Marzo de 2021). Ecuador: Estos son los días de feriado que aún quedan en el 2021. *El Universo*. https://www.eluniverso.com/noticias/ecuador/ecuador-calendario-de-los-dias-de-feriado-que-aun-quedan-en-el-2021-nota/

Cano Sanmartín, C. F. (2022). Implicaciones del SARS-CoV-2 en el uso del sistema de bicicleta pública de Ecuador entre los años 2019-2021. Cuenca, Ecuador: (Bachelor's thesis, Universidad del Azuay).

Cao, Y., & Wang, Y. (2022). Shared Cycling Demand Prediction during COVID-19 Combined with Urban Computing and Spatiotemporal Residual Network. *Sustainability, 14*(16), 9888.

Colaboradores de Wikipedia. (23 de Diciembre de 2022). *Manifestaciones en Ecuador de octubre de 2019*. (L. e. Wikipedia, Editor) Retrieved 13 de Marzo de 2023, from https://es.wikipedia.org/w/index.php?title=Manifestaciones_en_Ecuador_de_octubre_de_2019&oldid=148131942

Colaboradores de Wikipedia. (13 de Marzo de 2023). *Pandemia de COVID-19 en Ecuador*. (L. e. Wikipedia, Editor) Retrieved 25 de Marzo de 2023, from https://es.wikipedia.org/wiki/Pandemia_de_COVID-19_en_Ecuador

Colaboradores de Wikipedia. (18 de Marzo de 2023). *Paro Nacional de Ecuador de 2022*. (L. e. Wikipedia, Editor) Retrieved 25 de Marzo de 2023, from https://es.wikipedia.org/w/index.php?title=Paro_Nacional_de_Ecuador_de_2022&oldid=149971847

Consejo de Participación Ciudadana y Control Social de Ecuador. (10 de Marzo de 2023). *Veeduría ciudadana recomienda revisar el proyecto de Bici Pública en Cuenca*. Retrieved 13 de Marzo de 2020.

Diario El Mercurio. (7 de Abril de 2022). El feriado por las fiestas de fundación de Cuenca se traslada al lunes 11 de abril. *El Mercurio*. https://elmercurio.com.ec/2022/04/07/el-feriado-por-las-fiestas-de-fundacion-de-cuenca-se-traslada-al-lunes-11-de-abril/

EMOV EP. (Marzo de 2020).

EMOV EP. (Marzo de 2021). Base de Datos de detalle Uso de Bicicleta Pública, Bicicuenca.

EMOV EP. (16 de Junio de 2021). *Inicia la construcción de 13.5km de la red de ciclovía*. Retrieved 31 de Enero de 2022, from EMOV EP.: https://www.emov.gob.ec/inicia-la-construccion-de-13-5km-de-la-red-de-

ciclovia/

EMOV EP. (27 de Febrero de 2023). Base de Datos de detalle Uso de Bicicleta Pública, Bicicuenca, desde 1 Abril 2019 hasta 31 Diciembre 2022. Cuenca, Ecuador.

Gao, X., & Lee, G. (2019). Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. *Computers & Industrial Engineering*, 128, 60-69.

García-Santana, M., & Pérez-Canto, S. (2021). *Winsorización: una técnica estadística para reducir la influencia de valores extremos en el análisis de datos económicos, 29(1)*, 43-56. Revista de Economía Aplicada. https://doi.org/10.1016/j.revecap.2020.04.003

Guidon, S., Reck, D. J., & Axhausen, K. (2020). Expanding a (n)(electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests. *Journal of Transport Geography*, 84, 102692.

IERSE. (1 de Marzo de 2023). Datos pertenecientes a la red de sensores remotos entre 1 Abril 2019 hasta 31 Diciembre 2022. Cuenca, Ecuador.

INAMHI. (s.f.). *Información Estaciones INAMHI.* Retrieved 31 de Enero de 2022, from http://sedc.fonag.org.ec/reportes/inamhi

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4), 455-466.

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (Noviembre de 2015). Traffic prediction in a bike-sharing system. *In Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 1-10.

Ministerio de Turismo de Ecuador. (2 de Enero de 2019). *CALENDARIO DE FERIADOS ECUADOR 2019 ENERO*. Retrieved 9 de Marzo de 2023, from https://www.turismo.gob.ec/wp-content/uploads/2019/01/FERIADOS-2019.pdf

Ministerio de Turismo de Ecuador. (9 de Marzo de 2020). *Calendario de Feriados*. https://www.turismo.gob.ec/wp-content/uploads/2020/03/CALENDARIO-DE-FERIADOS.pdf

Ministerio de Turismo de Ecuador. (Noviembre de 2021). *Calendario de Feriados Nacionales 2022*. https://drive.google.com/file/d/19dCPwi2S0fhGkIAUSBAy1QOY2_GoKd3w/view

Öztaş, C. C., & Ölmez, M. (4 de Enero de 2022). *Using Cycling as an Indicator for Urban Quality of Life*. TheCityFix: https://thecityfix.com/blog/using-cycling-as-an-indicator-for-urban-quality-of-life/

Pérez Torres, L. (28 de Noviembre de 2019). Cronología del paro en Ecuador, y lo que vino después. *Deutsche Welle*.

RapidMiner Inc. (8 de Marzo de 2023). RapidMiner: https://rapidminer.com/

Romero, J., & Palomeque, C. (2022). Análisis de percepción y parámetros de recorrido de transmisión mecánica y eléctricas en el contexto del sistema de bicicleta pública en la ciudad de Cuenca. Cuenca, Ecuador: (Bachelor's thesis, Universidad del Azuay).

Sinche Solis, D. P., & Zhinin Auquilla, D. F. (2020). Análisis de aceptación del sistema de transporte bicicleta pública en la ciudad de Cuenca. Cuenca, Ecuador: (Bachelor's thesis, Universidad Politécnica Salesiana Sede Cuenca).

Weatherspark. (25 de Marzo de 2023). *Tiempo promedio en abril en Cuenca, Ecuador*. https://es.weatherspark.com/m/19348/4/Tiempo-promedio-en-abril-en-Cuenca-Ecuador

Yang, H., Xie, K., Ozbay, K., Ma, Y., & Wang, Z. (2018). Use of deep learning to predict daily usage of bike sharing systems. *Transportation research record*, 2672(36), 92-102.

Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda. (2016). Mobility modeling and prediction in bike-sharing systems. *In Proceedings of the 14th annual international conference on mobile systems, applications, and services*, (pp. 165-178).

8 ANEXOS

Anexo 1 Descripción variables del dataset de uso del Sistema de bicicleta pública compartida BiciCuenca (EMOV EP., 2023). Elaboración propia.

<table>
<tr><th>Variable</th><th>Unidad de Medida</th><th>Tipo</th><th colspan="3">Estadísticas</th></tr>
<tr><td rowspan="2">Fecha</td><td rowspan="2"></td><td rowspan="2">Fecha Hora</td><td>Mínimo</td><td>Máximo</td><td>Duración</td></tr>
<tr><td>1-abr-19 9:02</td><td>31-dic-22 19:34</td><td>1.370 días, 10 horas, 40 minutos, 29 segundos</td></tr>
<tr><td rowspan="2">Estado</td><td rowspan="2"></td><td rowspan="2">Nominal</td><td>Más pequeño</td><td>Más grande</td><td>Datos</td></tr>
<tr><td>Cancelada (2.582)</td><td>Finalizada (73.937)</td><td><table><tr><td>Estado</td><td>Frecuencia</td></tr><tr><td>Finalizada</td><td>73.937</td></tr><tr><td>Cancelada</td><td>2.582</td></tr></table></td></tr>
<tr><td rowspan="2">Código Bicicleta</td><td rowspan="2">Unidades</td><td rowspan="2">Entero</td><td>Mínimo</td><td>Máximo</td><td>Media</td><td>Desviación Estándar</td></tr>
<tr><td>1.210</td><td>104.4291</td><td>23.692,73</td><td>41.316,4</td></tr>
<tr><td rowspan="2">Medio</td><td rowspan="2"></td><td rowspan="2">Nominal</td><td>Más pequeño</td><td>Más grande</td><td colspan="2">Datos</td></tr>
<tr><td>APP IPHONE (8.825)</td><td>LECTOR (54.924)</td><td colspan="2"><table><tr><th>Medio</th><th>Frecuencia</th></tr><tr><td>APP IPHONE</td><td>8.825</td></tr><tr><td>APP ANDROID</td><td>12.770</td></tr><tr><td>LECTOR</td><td>54.924</td></tr></table></td></tr>
<tr><td rowspan="2">Tiempo de Uso</td><td rowspan="2">Segundos</td><td rowspan="2">Entero</td><td>Mínimo</td><td>Máximo</td><td colspan="2">Media</td></tr>
<tr><td>0</td><td>243.771</td><td colspan="2">770.658</td></tr>
<tr><td rowspan="2">Estación Origen</td><td rowspan="2"></td><td rowspan="2">Nominal</td><td>Más pequeño</td><td>Más grande</td><td colspan="2">Datos</td></tr>
<tr><td>Nueve de Octubre (1.834)</td><td>El Centenario (7.617)</td><td colspan="2"><table><tr><th>Origen</th><th>Frecuencia</th></tr><tr><td>El Centenario</td><td>7.617</td></tr><tr><td>Parque Calderón</td><td>7.249</td></tr><tr><td>Parque de la Madre</td><td>5.633</td></tr><tr><td>Escuela Panamá</td><td>4.837</td></tr><tr><td>La Concordia</td><td>4.390</td></tr></table></td></tr>
</table>

<table>
<tr><th>Variable</th><th>Unidad de Medida</th><th>Tipo</th><th colspan="4">Estadísticas</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td colspan="2"><table>
<tr><td>El Estadio</td><td>4.149</td></tr>
<tr><td>UE La Salle</td><td>4.134</td></tr>
<tr><td>María Auxiliadora</td><td>3.782</td></tr>
<tr><td>El Farol</td><td>3.660</td></tr>
<tr><td>Universidad del Azuay</td><td>3.618</td></tr>
<tr><td>San Sebastián</td><td>3.599</td></tr>
<tr><td>El Vergel</td><td>3.315</td></tr>
<tr><td>Portal Artesanal</td><td>3.153</td></tr>
<tr><td>Universidad de Cuenca</td><td>3.023</td></tr>
<tr><td>Santo Domingo</td><td>3.005</td></tr>
<tr><td>La Merced</td><td>2.817</td></tr>
<tr><td>Terminal Terrestre</td><td>2.762</td></tr>
<tr><td>Parque de el paraíso</td><td>1.986</td></tr>
<tr><td>Víctor J. Cuesta</td><td>1.956</td></tr>
<tr><td>Nueve de Octubre</td><td>1.834</td></tr>
</table></td></tr>
<tr><td rowspan="3">Estación Destino</td><td rowspan="3"></td><td rowspan="3">Nominal</td><td>Más pequeño</td><td>Más grande</td><td colspan="2">Datos</td></tr>
<tr><td>Nueve de Octubre (1.355)</td><td>Parque de la Madre (9.220)</td><td>Destino</td><td>Frecuencia</td></tr>
<tr><td colspan="2"><table>
<tr><td>Parque de la Madre</td><td>9.220</td></tr>
<tr><td>El Centenario</td><td>8.756</td></tr>
<tr><td>El Vergel</td><td>5.514</td></tr>
<tr><td>Parque Calderón</td><td>5.369</td></tr>
<tr><td>Universidad del Azuay</td><td>4.716</td></tr>
<tr><td>El Estadio</td><td>4.173</td></tr>
</table></td></tr>
</table>

Variable	Unidad de Medida	Tipo	Estadísticas			
					UE La Salle	4.049
					Escuela Panamá	3.946
					La Concordia	3.918
					Portal Artesanal	3.523
					Universidad de Cuenca	3.389
					El Farol	2.899
					Parque de el paraíso	2.880
					Santo Domingo	2.520
					Terminal Terrestre	2.233
					San Sebastián	2.129
					Víctor J. Cuesta	2.109
					María Auxiliadora	1.997
Costo	Dólares	Flotante	**Mínimo**	**Máximo**	**Media**	**Desviación Estándar**
			0	9,99	0,030996	0,19945753

Anexo 2 Descripción variables del dataset de datos pertenecientes a la red de sensores remotos de Cuenca (IERSE, 2023). Elaboración propia.

<table>
<tr><th>Variable</th><th>Unidad Medida</th><th>Tipo</th><th colspan="3">Estadísticas</th></tr>
<tr><td rowspan="2">Fecha</td><td rowspan="2"></td><td rowspan="2">Fecha Hora</td><td>Mínimo</td><td>Máximo</td><td>Duración</td></tr>
<tr><td>1-abr-19 0:00</td><td>31-dic-22 11:58</td><td>1.370 días, 23 horas, 57 minutos, 43 segundos</td></tr>
<tr><td rowspan="2">Estación Climática</td><td rowspan="2"></td><td rowspan="2">Nominal</td><td>Más pequeño</td><td>Más grande</td><td>Datos</td></tr>
<tr><td>SCP88 (10)</td><td>SCP06 (810.680)</td><td>
<table>
<tr><th>Estación</th><th>Frecuencia</th></tr>
<tr><td>SCP06</td><td>810.680</td></tr>
<tr><td>SCP04</td><td>771.000</td></tr>
<tr><td>SCP09</td><td>654.970</td></tr>
<tr><td>SCP07</td><td>621.265</td></tr>
<tr><td>SCP03</td><td>620.320</td></tr>
<tr><td>SCP08</td><td>616.245</td></tr>
<tr><td>SCP10</td><td>598.670</td></tr>
<tr><td>SCP11</td><td>537.835</td></tr>
<tr><td>SCP05</td><td>526.880</td></tr>
<tr><td>SCP15</td><td>495.470</td></tr>
<tr><td>SCP16</td><td>486.380</td></tr>
<tr><td>SCP01</td><td>467.685</td></tr>
<tr><td>SCP13</td><td>424.290</td></tr>
<tr><td>SCP02</td><td>423.220</td></tr>
<tr><td>SCP17</td><td>341.040</td></tr>
<tr><td>SCP14</td><td>96.270</td></tr>
<tr><td>SCP12</td><td>67.295</td></tr>
<tr><td>SCP18</td><td>44.880</td></tr>
<tr><td>SCP19</td><td>26.860</td></tr>
<tr><td>SCP88</td><td>10</td></tr>
</table>
</td></tr>
<tr><td rowspan="2">Humedad Relativa</td><td rowspan="2">Porcentaje de Humedad</td><td rowspan="2">Real</td><td>Mínimo</td><td>Máximo</td><td>Media</td></tr>
<tr><td>-1.000,333</td><td>100</td><td>69,902</td></tr>
<tr><td rowspan="2">Ruido</td><td rowspan="2">Decibelios</td><td rowspan="2">Real</td><td>Mínimo</td><td>Máximo</td><td>Media</td></tr>
<tr><td>-5,970</td><td>96,170</td><td>60,430</td></tr>
<tr><td rowspan="2">Presión Atmosférica</td><td rowspan="2">Pascales</td><td rowspan="2">Real</td><td>Mínimo</td><td>Máximo</td><td>Media</td><td>Desviación Estándar</td></tr>
<tr><td>-1001</td><td>7.685.887,500</td><td>75.608,171</td><td>10.057,756</td></tr>
<tr><td rowspan="2">Temperatura del Aire</td><td rowspan="2">Grados Centígrados</td><td rowspan="2">Real</td><td>Mínimo</td><td>Máximo</td><td>Media</td></tr>
<tr><td>-1001</td><td>191,090</td><td>13,986</td></tr>
</table>

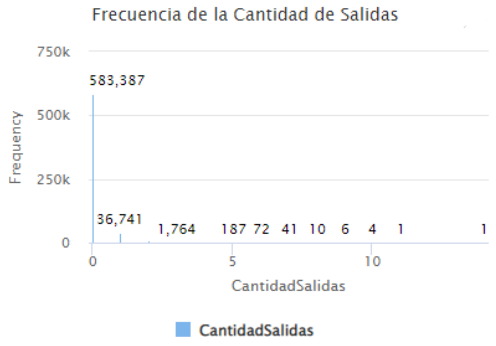
Anexo 3 Rangos de Fechas en las que no existen registros y eventos relacionados. Fuentes de Datos: (EMOV EP., 2023), (C.A. EL UNIVERSO, 2018), (C.A. EL UNIVERSO, 2019), (Pérez Torres, 2019), (Colaboradores de Wikipedia, 2023), (Colaboradores de Wikipedia, 2023). Elaboración propia.

Año	**Rango de Fechas que no existen registros**	**Eventos**
2019	7 Abril.	Se desconoce.
2019	9 al 13 de Octubre.	Paro Nacional.
2.020	1 de Enero.	Feriado primero de enero.
2.020	17 de Marzo – 17 de Mayo.	Restricciones de movilidad por pandemia.
2.020	24 de Mayo.	Feriado Batalla de Pichincha, restricciones de movilidad por pandemia.
2.021	24 y 25 de Abril. 1, 2, 8, 9, 15, 16 de Mayo.	Restricciones de movilidad por pandemia.
2.022	1 de Enero.	Feriado primero de enero.
2.022	23 al 30 de Junio.	Paro Nacional.

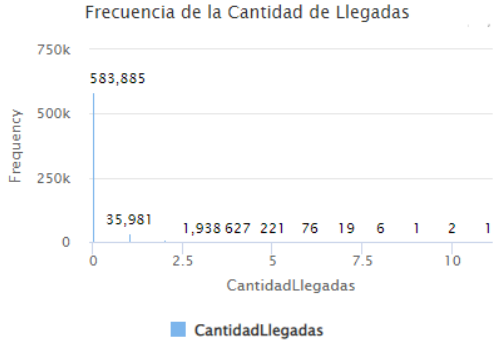
Anexo 4 Valores descriptivos de la cantidad de salidas de usuarios por hora antes y después de la Pandemia. Fuente de Datos: (EMOV EP., 2023). Elaboración propia.

Pandemia	Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo	Desviación Estándar	Varianza
Antes	0,000	0,000	0,000	0,326	1,000	14,000	0,638	0,406
Después	0,000	0,000	0,000	0,085	0,000	6,000	0,308	0,095

Anexo 5 Frecuencia de la cantidad de llegadas y salidas. Fuente de Datos: (EMOV EP., 2020). Elaboración propia.

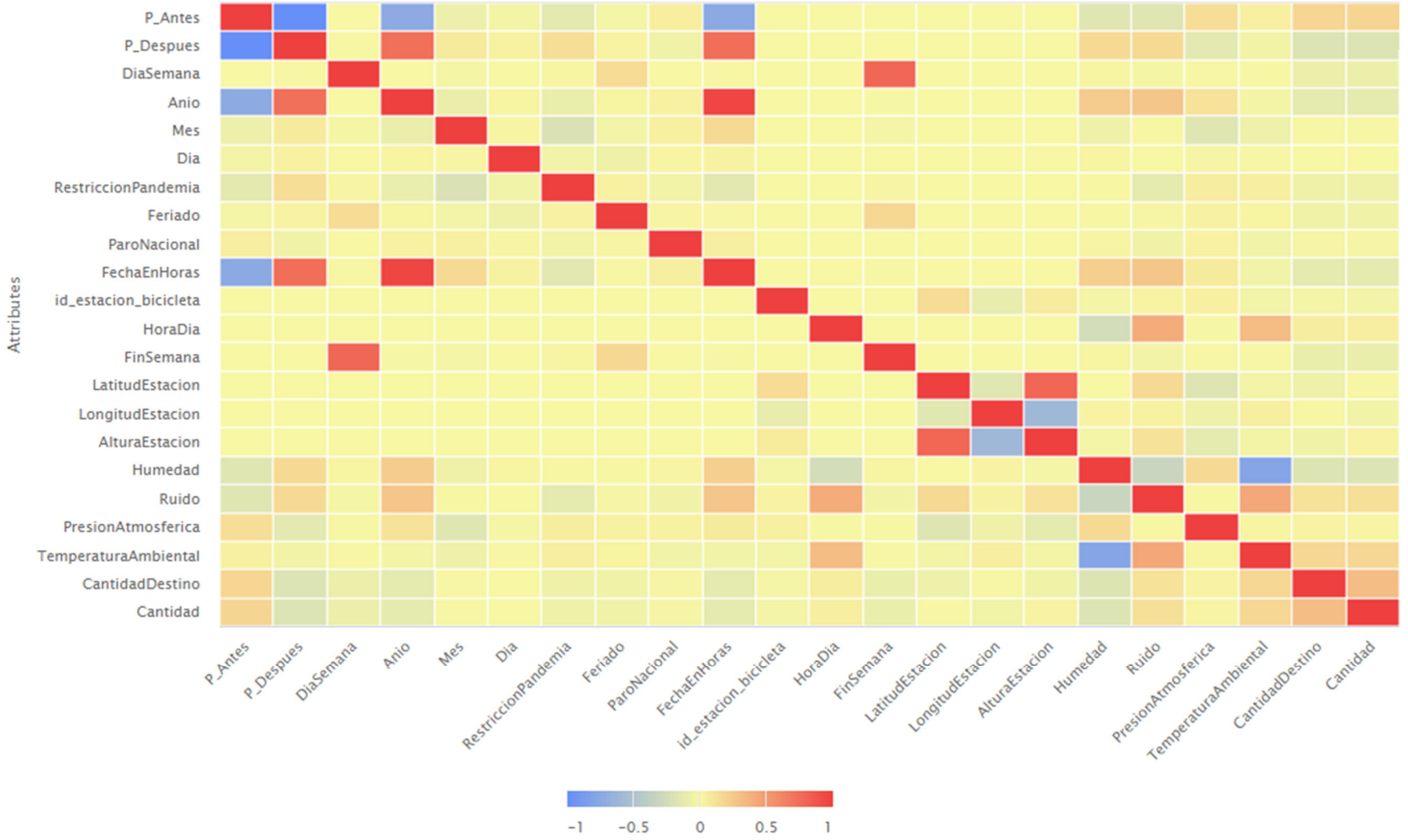


(a) Frecuencia de la cantidad de salidas.

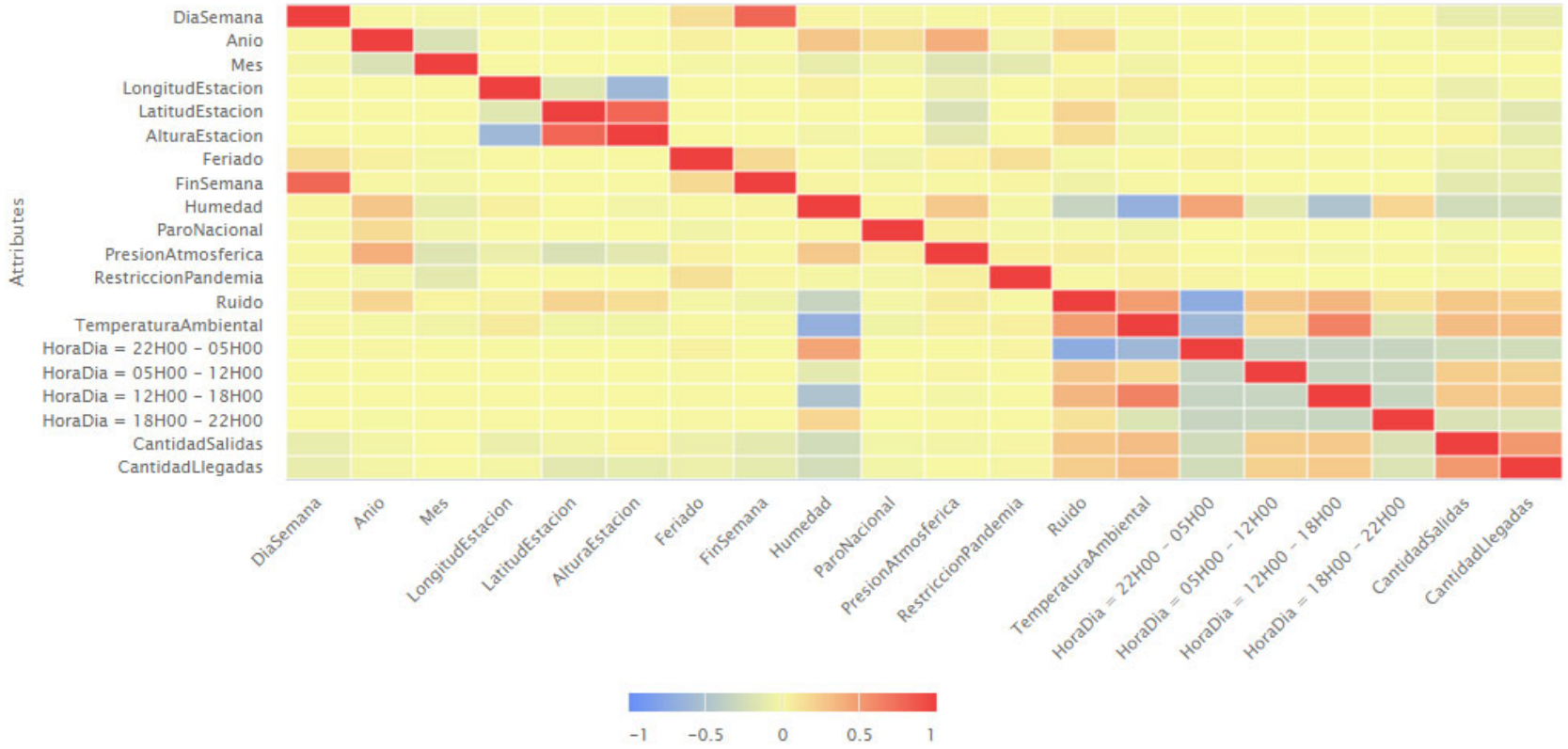


(b) Frecuencia de la cantidad de llegadas.

Anexo 6 Gráfica de la Matriz de correlación entre las variables del estudio con intervalos horarios. Elaboración Propia.



Anexo 7 Gráfica de la Matriz de correlación entre las variables de estudio con intervalos de tiempo personalizado. Elaboración Propia



Anexo 8 Resultados de la aplicación de los modelos de predicción para las salidas. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

Resultados de Predicciones de Salidas								
	G: Análisis General				P: Análisis Paralelo por estación			
Modelo	E1	E2	E3	E4	E1	E2	E3	E4
Error Absoluto								
Regresión Lineal	0,745	0,661	0,558	0,502	0,676	0,607	0,461	0,417
Random Forest	0,623	0,594	0,507	0,472	0,609	0,578	0,455	0,422
Árboles de decisión	0,639	0,585	0,511	0,475	0,712	0,631	0,51	0,49
Deep Learning	0,631	0,572	0,537	0,441	0,576	0,598	0,446	0,41
Error Relativo								
Regresión Lineal	47,32%	41,97%	47,27%	40,58%	50,61%	49,06%	45,96%	39,07%
Random Forest	47,36%	48,37%	48,96%	45,32%	52,25%	47,98%	50,11%	43,90%
Árboles de decisión	56,12%	50,19%	52,46%	47,89%	78,53%	69,12%	63,21%	55,32%
Deep Learning	53,71%	48,74%	41,46%	45,67%	57,70%	57,05%	51,57%	45,03%
RMSE								
Regresión Lineal	0,952	0,83	0,792	0,658	1	0,921	0,803	0,71
Random Forest	0,88	0,847	0,75	0,657	0,984	0,95	0,808	0,723
Árboles de decisión	0,957	0,887	0,772	0,692	1,35	1,229	0,943	0,9
Deep Learning	0,901	0,799	0,788	0,641	0,954	0,966	0,803	0,722
Coeficiente de Determinación								
Regresión Lineal	0,122	0,144	0,134	0,147	0,166	0,14	0,172	0,152
Random Forest	0,187	0,144	0,225	0,186	0,173	0,128	0,175	0,147
Árboles de decisión	0,138	0,108	0,193	0,144	0,091	0,089	0,102	0,073
Deep Learning	0,107	0,146	0,154	0,171	0,152	0,147	0,17	0,151

Anexo 9 Resultados de la aplicación de los modelos de predicción para las llegadas. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

Resultados de Predicciones de Llegadas

	G: Análisis General								P:Análisis Paralelo por estación							
	A: No incluye salidas				B: Incluye salidas				A: No incluye salidas				B: Incluye salidas			
Modelo	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
Error Absoluto																
Regresión Lineal	0,744	0,665	0,555	0,511	0,735	0,654	0,555	0,503	0,676	0,461	0,526	0,484	0,58	0,586	0,437	0,405
Random Forest	0,644	0,623	0,502	0,465	0,625	0,609	0,5	0,462	0,609	0,455	0,482	0,454	0,507	0,45	0,424	0,396
Árboles de decisión	0,629	0,614	0,497	0,463	0,623	0,591	0,496	0,475	0,72	0,523	0,496	0,459	0,58	0,468	0,468	0,419
Deep Learning	0,605	0,566	0,485	0,506	0,583	0,621	0,498	0,472	0,629	0,441	0,5	0,482	0,579	0,546	0,446	0,395
Error Relativo																
Regresión Lineal	47,73%	42,16%	45,73%	41,70%	47,48%	41,98%	46,45%	41,55%	50,61%	45,96%	46,79%	39,22%	58,76%	61,22%	58,03%	49,41%
Random Forest	60,02%	59,35%	53,58%	48,46%	55,41%	58,28%	53,71%	48,57%	52,25%	50,11%	54,56%	46,88%	58,67%	49,31%	59,73%	48,21%
Árboles de decisión	60,52%	61,62%	56,86%	51,28%	61,90%	59,74%	56,13%	51,26%	78,29%	62,21%	57,01%	53,44%	68,89%	56,99%	69,24%	55,66%
Deep Learning	52,37%	57,79%	56,15%	48,50%	62,03%	61,25%	58,76%	48,38%	58,46%	54,31%	58,19%	49,62%	62,48%	60,14%	62,45%	48,88%
RMSE																
Regresión Lineal	0,954	0,833	0,791	0,664	0,947	0,827	0,792	0,656	1	0,803	0,789	0,659	0,966	0,97	0,815	0,719
Random Forest	0,950	0,917	0,76	0,657	0,905	0,882	0,752	0,649	0,984	0,808	0,748	0,637	0,885	0,767	0,809	0,7
Árboles de decisión	0,983	0,917	0,775	0,665	0,953	0,875	0,768	0,679	1,361	0,958	0,769	0,664	1,062	0,829	0,885	0,751
Deep Learning	0,856	0,815	0,754	0,706	0,867	0,864	0,776	0,658	1,009	0,798	0,781	0,66	0,983	0,921	0,83	0,709
Coeficiente de Determinación																
Regresión Lineal	0,129	0,132	0,151	0,124	0,13	0,137	0,148	0,133	0,166	0,172	0,156	0,143	0,217	0,185	0,197	0,174
Random Forest	0,213	0,166	0,237	0,197	0,218	0,177	0,242	0,205	0,173	0,175	0,244	0,214	0,212	0,174	0,199	0,188
Árboles de decisión	0,165	0,142	0,216	0,193	0,183	0,157	0,226	0,169	0,093	0,101	0,215	0,196	0,158	0,14	0,164	0,163
Deep Learning	0,193	0,16	0,237	0,172	0,146	0,155	0,199	0,171	0,153	0,168	0,187	0,162	0,209	0,18	0,191	0,162

Anexo 10 Resultados de los p valores obtenidos al aplicar la función de autocorrelación ACF de los datos de salidas. Fuentes de Datos: (EMOV EP., 2023; IERSE, 2023). Elaboración propia.

ACF <dbl>	p.valor <dbl>
1.00000000	1.992887
-0.47745646	1.996604
-0.28448568	1.997976
0.53388656	1.996202
-0.24487328	1.998258
-0.24810991	1.998235
0.44467428	1.996837
-0.21291488	1.998486
-0.21949204	1.998439
0.41495738	1.997048