UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

MASTER THESIS

Neural network model, based on time series, to forecast availability in the bike-shared systems

---

A thesis submitted in fulfillment of the requirements for the University Master's Degree in Artificial Intelligence Research

in the

Departament of Artificial Intelligence

Author: Bosco Gutiérrez Gómez

Supervisor: José Luis Aznarte Mellado, PhD

Co-supervisor: Rodrígo de Medrano López

September 19, 2021

Madrid

Abstract

The rental of the public system of shared bicycles is a service aimed at all citizens of the city of Madrid and Barcelona, as an alternative element of clean transport that contributes to a more sustainable mobility model and the promotion of more balanced transport habits and healthy. Mobility services are increasingly based on technology and data collection, not only directly related to mobility flows, but also to other variables that affect it to a greater extent such as meteorology, pollution, strikes and temporary events. Knowing where, when and how people move is key to matching supply with demand. A better understanding of their behavior will allow us to better adapt these transport systems and optimize resources. It is necessary to predict how the system will behave to anticipate movements. Deep learning techniques have shown significant improvements in prediction over traditional models, but some difficulties and open questions remain regarding their applicability, accuracy, and ability to provide practical information. Our approach in this paper is based on comparing different models capable of predicting at least 6 hours in advance which stations are likely to be full or empty.

Acknowledgements

Me gustaría agradecer especialmente al Dr. José Luis Aznarte la confianza depositada en mí al aceptarme para la realización de este máster, así como su tutela. A Rodrígo de Medrano por su orientación y apoyo en el día a día. Y por supuesto a mi familia, Belén, Adriana y Marina, que no solo han aportado su apoyo, sino que han cedido gran cantidad de su tiempo juntos para que yo pudiera dedicarme a esta tarea.

Contents

List of Figures

List of Tables

1 Introduction

The purpose of this first chapter is to give an overview of the problem and a description of the contents that are developed in this work. The objectives are presented, as well as the analysis of the data, the methodology used and the results obtained.

1.1 Presentation

The city organizes its mobility through different transport systems that act individually, but which together form a mobility ecosystem. This is an ecosystem with great imbalances by giving priority to some transport over others. A better understanding of their behavior will allow us to better adapt these transport systems and optimize resources. The city's ecosystem is supported by a physical infrastructure such as roads, bridges or tunnels. This infrastructure conditions movement flows within cities and these investments in infrastructure are high, so design errors represent a great economic and functional cost for citizens. In addition to the road, there are the stops and routes made by public transport, which is the basis of mobility in any developed city. This is more or less flexible depending on its type. For example, the route of the bus and its stops is much more flexible than the subway, whose underground infrastructure represents a technical challenge and a large investment, so any errors are difficult to correct or cannot be adapted to changing trends within the city it faces. The technological infrastructure is also important, such as mobile phone towers, identification cameras for access control of certain areas, toll arches, access barriers or communication of these systems.

The public bicycle-shared system rental is a service aimed at all citizens of the city of Madrid and Barcelona, as an alternative element of clean transport that contributes to a more sustainable mobility model and the promotion of more balanced and healthy transport habits.

Mobility services are increasingly based on technology and data collection, not only directly related to mobility flows, but also to other variables that affect it to a greater extent such as weather, pollution, strikes, specific events or the day of the week and the month. Knowing where, when and how people move is key to matching supply with demand. For this, it will be necessary to work with different data sources and explore others, which are not so obvious at first, but which can give us the keys to these behaviors.

The study focuses on the comparison of the cities of Madrid and Barcelona, which together have more than 9,000 bicycles, 700 stations and 300,000 users. Based on usage data and profiles from the public bicycle network in Madrid (BiciMad) and Barcelona (Bicing), we developed a usage prediction model based on time, journeys and climate conditions. The stations have a limited number of anchorages, around 20, where you can park your bicycles. Within the user experience, one of the biggest frustrations is not being able to leave the bikes because the station is completely full or not being able to have a bike because the station is completely empty. The municipalities, which are responsible for the service, apply financial penalties to the operators in the contracts every time these situations occur. To solve these problems, the operator must balance the distribution of bicycles in the stations, moving vehicles from one station to another. The constant flows of bicycles are conditioned by different variables such as working hours, the weather, the day of the week or events, which makes it necessary to predict how the system will behave in order to anticipate movements.

1.2 Precedents

Although the use of shared bicycles has been a reality for some time in different cities around the world, it is in recent years and especially after the global pandemic caused by Covid 19, when it has gained a relevant role. "Recent analysis on the micro-mobility industry finds that, globally, it is set to become one of the most attractive forms of sustainable mobility thanks to its convenience, affordability and safety. Comprising bike-sharing, kick scooter sharing, and scooter sharing, the global micro-mobility market is estimated to register 31.9 million vehicles by 2025, up from 20.5 million in 2020" [1]. Different changes in trends have favored cities to become aware of the importance of mobility in cities. These trends are framed in different areas such as social, economic, infrastructure or cultural. With the new generations, habits are being changed from the ownership of assets to the use and enjoyment of services. As this is a new trend, it is not very clear if it is driven by the greater difficulty of access to different goods such as motor vehicles and / or a temporary behavior conditioned by the lack of responsibilities and obligations that will grow with the increase of members in the family. The growth trend of cities in the face of rural depopulation or the development of technology that makes the use and payment of new forms of transport more accessible in its different modes such as shared transport. What seems like a fact rather than a trend is that governments, administrations and those responsible for public transport and mobility in cities have incorporated the shared bicycle service as one more mode of transport that helps solve the increasingly growing problem congestion in cities. The concentration of the population in cities requires different solutions to help mitigate the problems caused by displacement and its high cost, both economically and in terms of time and health caused by high rates of pollution. Shared bicycle systems are here to stay and now we have to understand the habits of use and the parameters that will allow us to locate the stations in an efficient way and balance the distribution of bicycles in the stations. In addition, we need to know what other variables influence usage patterns in order to develop predictive models that allow us to optimize and improve the service.

1.3 State of the Art

Many studies have been conducted using available data to analyze bike sharing systems [12, 14, 18]. The main problem is that aggregate data was available in general, while it was not common to access individual data that reflected origin-destination journeys, availability of vehicles at stations or type of user (sporadic, regular or intensive). The difficulty of accessing these data [13] has conditioned the type of studies on mere descriptive analyzes until recent years where techniques with a predictive approach based mainly on time series and crossing data that could influence behaviors such as the weather [21, 23, 26], the level of congestion in the city, pollution [15] or other sociodemographic conditions [19, 20]. Bagus and Purnama published in 2015 one of the first articles [2] that attempts to develop a predictive model of use favored by availability of origin-destination vector data in a disaggregated form, contrary to what had been available until now. Once they have defined the typology of users, spatial analysis and route estimation, they apply a Markov-based model to predict the next destination station for the user. The results of this study are promising. Seguí and

Mateu [3] asses three components: The evolution of the modal distribution and its relationship with bicycle traffic, the territorial coverage of the system and its operation and, finally, the characteristics, profile and the motivations of the users. This is a very descriptive work that, as a handicap, uses the origin-destination matrices in an aggregate way, which makes it difficult to assign the route individually. In any case, this does not prevent us from analyzing the growth and movement flows that make up the bicycle ecosystem in the city of Palma.

Finally, Ali and Mahmood [4] review 26 articles from two perspectives: the first, the use of deep learning architecture and the second, all the data collected by various devices such as sensors, IoT or GPS, revealing that the LSTM (Long Short-Term Memory) of neural networks is the most used architecture for the prediction of traffic flows. Regarding Spatio-temporal forecasting, an open research field whose interest is growing exponentially.

Rodrigo de Medrano and José L. Aznarte in their work [5] focus on creating a complex deep neural framework for spatio-temporal traffic forecasting. His proposal is based on an interpretable attention-based neural network in which several modules are combined in order to capture key spatio-temporal time series components. This article serves as the basis for the present study and the performance of time series by neural networks.

1.4 Objectives

The objective of this master´s thesis is the comparison of different predictive models of neural networks on the data of two cities (Madrid and Barcelona) according to their usage habits in a time series format and adding climatological factors such as temperature, rain and wind. It seeks to understand what conditions determine the mode and intensity of use and if they can be extrapolated to other cities. For this, our approach is based on building a model capable of predicting at least 6 hours in advance which stations are likely to fill or empty.

1.5 Thesis overview

After an introduction where we locate the scope of this work, background, state of art and objectives, we proceed to describe and analyze the sources and quality of the data that we are going to have. These are the key to the development of this work. We continue with the methodology, breaking down how the different variables become part of the process and which are the models that will allow us to develop our predictive algorithm. Once the model is established, we will comment on what is the social impact and the use that should be given and how it improves the current methods used. The expected results and how to present it are also discussed.

2 Data analysis and description

As we have previously commented, the challenge we face is to demonstrate which model performs best when predicting station availability and whether there are other factors that influence user habits. For this, we will use a time series based on the use data of the shared bicycle systems of Madrid and Barcelona. In addition, as exogenous factors, we will add climatological variables such as rain, temperature and wind. The distinction of the nature of the two types of data is important because, while usage data is the basis for model performance, the true influence of climatological data on forecasting models and the impact on improving the model are an unknown that we hope to be able to solve throughout this study.

2.1 Data presentation

The time window chosen for this study is from March 2019 to February 2020, trying to avoid the impact that the COVID 19 pandemic produced on the use of these systems. We consider that the stoppage of the service during the following months and the subsequent irregular use caused by the different restrictions could translate into erratic behaviors within the time series. Although a study period of several years would be desirable to obtain optimal results, we are conditioned by the obtaining of the data and the computational capacity, so we consider that the established space of a full year is sufficient for the purpose of this study.

The magnitude of the data used also recommends the containment of temporal space. The size of the databases used are made up of more than 15 million time steps (one for each movement) in Barcelona and more than 5 million time steps for the city of Madrid.

Next, we proceed to describe the different databases and their sources:

- Bicycle-shared usage data in Madrid: Provided by *Portal de datos abiertos del Ayuntamiento de Madrid*.[1] The data about the BiciMAD service is offered in JSON format, a lightweight text format for data exchange that is widely used today and widely adopted as an alternative to XML. The information presented is relative to the movements of the bicycles. Being understood as movement, the transfer of a bicycle from a station of origin to a station of destination. This information includes data related to the movement itself (departure station or destination station) and data related to the user who performs the movement, duly anonymized to allow statistical studies, without being able to know any specific data about the user. Also included are the movements of bicycles carried out by company personnel, for example, to transport bicycles to stations where a higher demand is expected or to carry out a repair. That is why we speak of "movements" and not "trips".

This data set presents the day-by-day disaggregated information on:

- Users: New daily registrations in the system in its two modalities (annual and occasional subscriptions).

- 03/31/2017 id: Movement identifier.

- user day code: User code. For the same date, all the movements of the same user will have the same code, in order to be able to carry out statistical studies of the daily trends of the users.

- idunplug station: Number of the station from which the bike is unhooked.

- idunplug base: Number of the base from which the bike is unhooked.

- idplug station: Number of the station where the bike is hooked up.

- idplug base: Number of the base where the bike is attached.

- unplug hourTime: Time zone in which the bicycle is unhooked. For reasons of

[1] https://datos.madrid.es/portal/site/egob/

anonymity, the start time of the movement is provided, without the information of minutes and seconds. All movements started during the same time will have the same starting data.

- travel time: Total time in seconds, between unhooking and hitching the bike.

- track: Detail of the route made by the bicycle between the starting station and the destination station, in GeoJSON format. If it exists, it will contain a "FeatureCollection" element which in turn will contain a "Features" element of type list.

- Bicycle-shared usage data in Barcelona: Provided by *Open Data BCN*.[2]
The dataset Bicing service use of the city of Barcelona is available in Json format or CSV and it includes:

- Users: New daily registrations in the system in its two modalities (annual and occasional subscriptions).

- 03/31/2017 id: Movement identifier.

- user day code: User code. For the same date, all the movements of the same user will have the same code, in order to be able to carry out statistical studies of the daily trends of the users.

- idunplug station: Number of the station from which the bike is unhooked.

- idunplug base: Number of the base from which the bike is unhooked.

- idplug station: Number of the station where the bike is hooked up.

- idplug base: Number of the base where the bike is attached.

- unplug hourTime: Time zone in which the bicycle is unhooked. For reasons of anonymity, the start time of the movement is provided, without the information of minutes and seconds. All movements started during the same time will have the same starting data.

- travel time: Total time in seconds, between unhooking and hitching the bike.

- geometry: Indicates the position of the bike and contains:

- type: Type of the position.

- coordinates: Longitude and latitude coordinates of the position.

- type: type of element.

- Weather data in Madrid and Barcelona: Provided by *Agencia Estatal de Meteo-*

[2] https://opendata-ajuntament.barcelona.cat/es/us-servei-bicing

rología - AEMET.[3]

Weather observations consist of daily temperature in Celsius degrees, wind speed measured in km/h, daily rainfall in mm. Weather information is taken from a single station.

2.2 Data cleaning

Next, we are going to describe the cleaning work carried out in the different databases:
- Databases of the BiciMad and Bicing bicycle-shared systems: Once the databases have been downloaded from the different sources in monthly files in Json and csv format, we integrate them into a single annual database. This gave us two databases of more than 15 million time steps in Barcelona and more than 5 million in Madrid with more than 20 variables each. For the description of the habits of use we handle more variables than for the temporary matrix of availability of stations, such as the distance traveled and the time used, frequency of use according to the day of the week, monthly and yearly in Barcelona. For Madrid we will be able to discriminate by time of use and the frequency of use according to the day of the week, monthly and yearly.

2.3 Data analysis

In this section we are going to make a description of the habits of use of the citizens of the shared bicycle systems in Madrid and Barcelona. As we can see is that the use in the two cities decreases on weekends compared to the use during the working days of the week. On the other hand, the busiest day of the week in Barcelona is Thursday as it is shown in figure 1, while in Madrid it is Monday as it is shown in figure 2.

[3]http://www.aemet.es/es/datos abiertos/AEMET OpenData

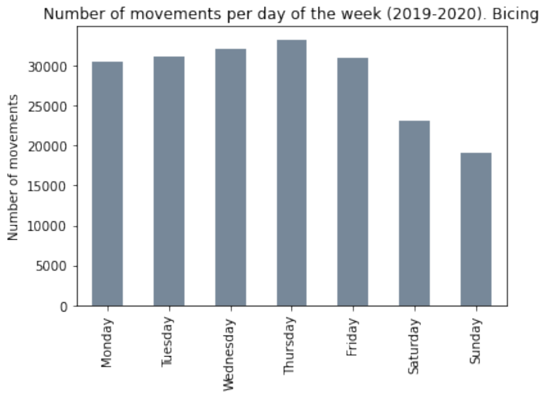

Figure 1: Use by day of the week. Bicing

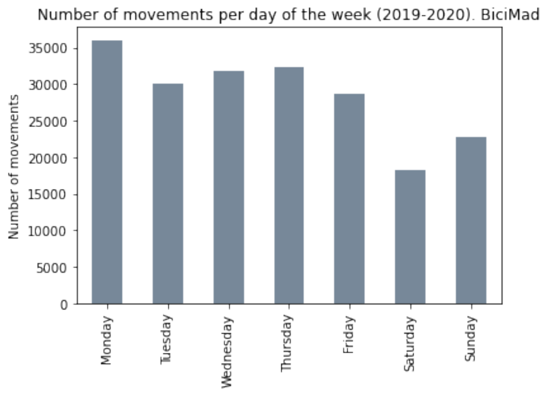

Figure 2: Use by day of the week. BiciMad

Likewise, the peak hours of the day in the two cities are between 6:00 a.m. and 9:00 a.m. and between 5:00 p.m. and 7:00 p.m., which correspond to the entry and exit of working hours (Showing figure 3 and figure 4). This gives us indications of the reason for the majority use is for commuting to work.

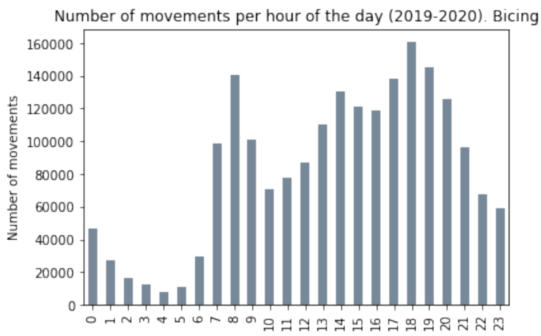

Figure 3: Use by time of day. Bicing

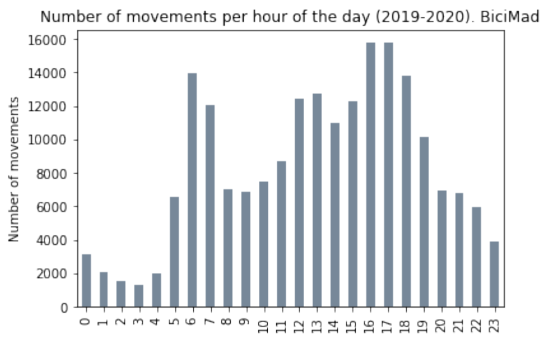

Figure 4: Use by time of day. BiciMad

In the case of Barcelona, we see how use decreases in the months of March, August and December, which are the months that coincide with holiday periods (Showing figure 5 and figure 6).

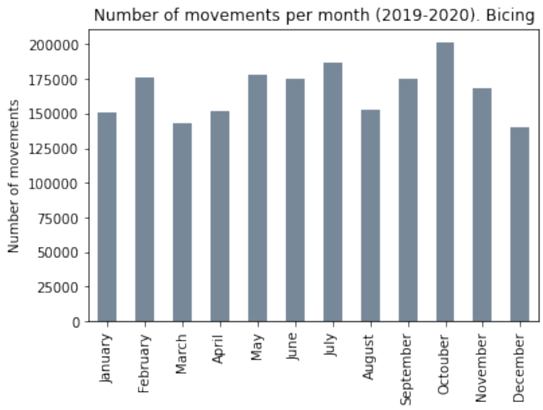

Figure 5: Use according to month. Bicing

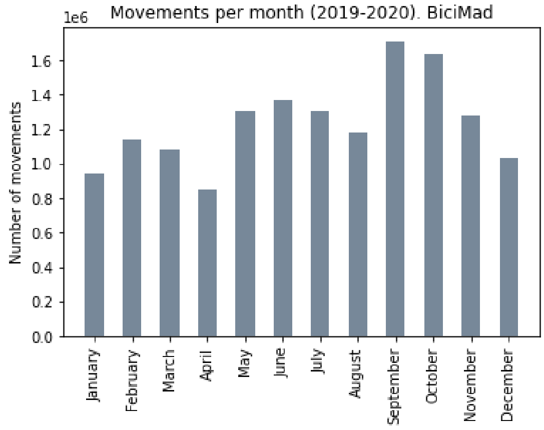

Figure 6: Use according to month. BiciMad

The distribution of the duration of the trips in both cities are similar and are around a range of 400 and 600 seconds (Showing figure 7 and figure 8).

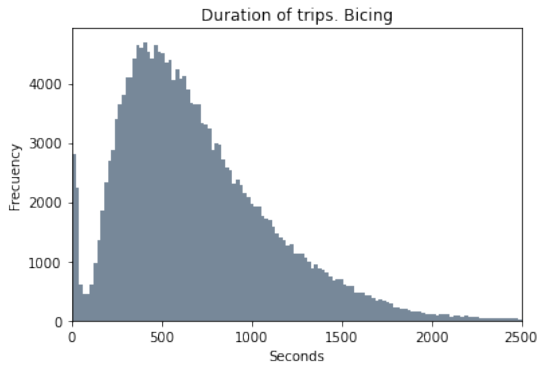

Figure 7: Duration of trips. Bicing

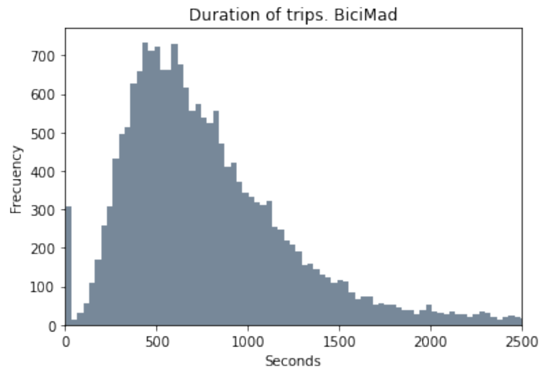

Figure 8: Duration of trips. BiciMad

Most trips in Barcelona cover a distance of between 1 and 3 kilometers as shown in figure 9.

2.4 Weather analysis

Of the multiple climatological variables provided by the AEMET stations, we are going to focus on temperature (Showing figure 10 and figure 11), rainfall (Showing figure 12 and figure 13) and wind speed (Showing figure 14 and figure 15). All of them are variables that, with some logic, should influence [10] the behavior of users when

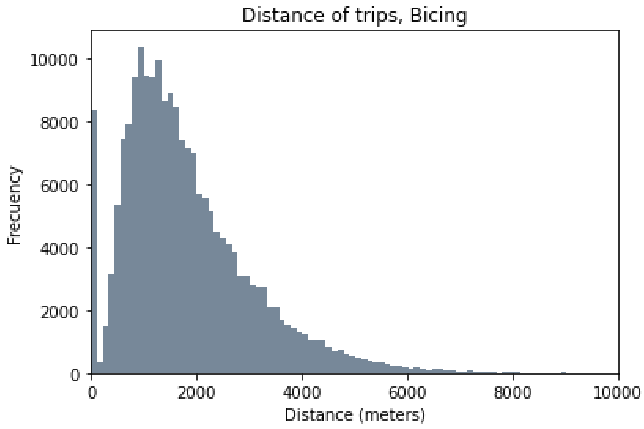

Figure 9: Distance of trips Bicing

making the decision to use the bicycle or not.

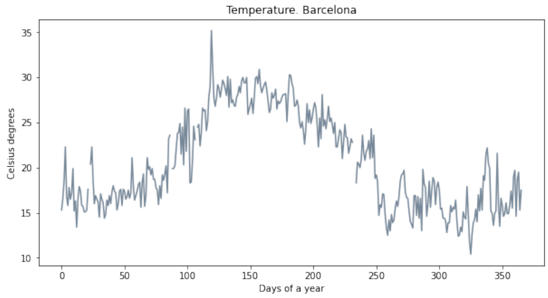

Figure 10: Temperature. Barcelona

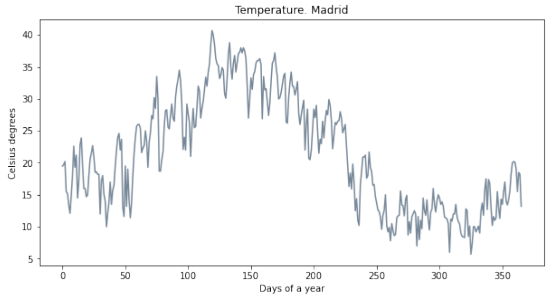

Figure 11: Temperature. Madrid

As is well known, the weather is seasonal, repeating its parameters annually and its characteristic behavior can be seen in each season.

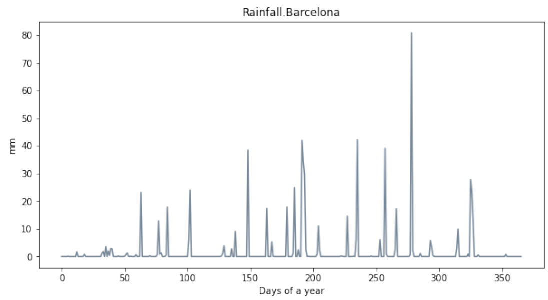

Figure 12: Rainfall. Barcelona

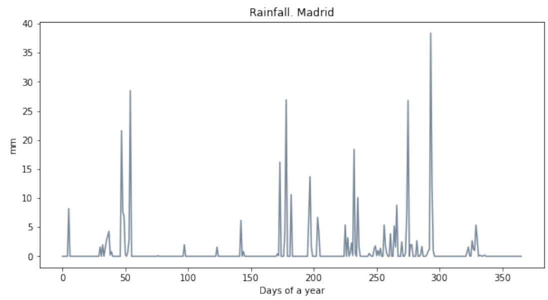

Figure 13: Rainfall. Madrid

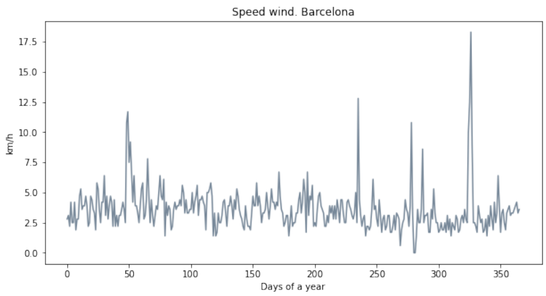

Figure 14: Speed wind. Barcelona

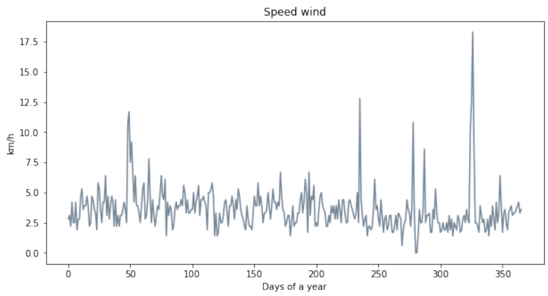

Figure 15: Speed wind. Madrid

2.5 Availability matrices

When constructing our matrices, we do not take into account their spatial situation [11] because we do not consider it relevant for obtaining our goals.
For the temporary availability matrix we use 4 variables that are station id of departure, time of departure, station id of arrival and time of arrival. In addition, we group the movements for each hour of the day, in such a way that each station will give us a net integer number that indicates the number of bicycles that have left during that hour minus the number of bicycles that have arrived. We group in an hour because we understand that it is the appropriate interval to predict the behavior of each station and to foresee if it is necessary for an operator to come to take or collect bicycles. For the Bicing matrix we used 8766 timesteps x 439 stations, presented in table 1, and for the BiciMad matrix we used 8766 timesteps x 169 stations.

3 Forecasting models

3.1 Introduction

The first challenge we face is choosing the models that we are going to confront in order to understand which ones work best for this specific problem [17, 24]. A first premise is to understand that each model has a series of advantages and disadvantages that will indicate the suitability of that model for that specific database and certain objectives. Choosing the right model is one of the skills that every data scientist must

Table 1: Movements matrix

Stations Timesteps	1	2	3	...	495	496	prec	tmax	velmean
01/03/2019 0:00	3	0	0	...	0	0	0.0	15.3	2.8
01/03/2019 1:00	0	-3	0	...	0	0	0.0	15.3	2.8
01/03/2019 2:00	4	1	-1	...	0	0	0.0	15.3	2.8
01/03/2019 3:00	0	-1	0	...	0	0	0.0	15.3	2.8
01/03/2019 4:00	0	0	-1	...	0	0	0.0	15.3	2.8
	...	...	...	...	...	...	...	...	...
28/02/2020 19:00	-5	1	0	...	-1	-7	0.0	17.5	3.6
28/02/2020 20:00	1	3	0	...	1	-1	0.0	17.5	3.6
28/02/2020 21:00	2	0	1	...	-1	4	0.0	17.5	3.6
28/02/2020 22:00	-5	4	-3	...	-2	-5	0.0	17.5	3.6
28/02/2020 23:00	3	-5	5	...	-2	8	0.0	17.5	3.6

develop in order to truly benefit from a database.

ANN (artificial neural networks) are inspired by the biological neural networks of the human brain. To do this, they try to replicate the elements and behavior of the brain and neurons. These behaviors include learning through experience, generalizing from previous examples to new examples and abstracting the main characteristics of a series of data that apparently do not present common or relative aspects. Another is the adaptation of their behavior depending on the environment. They are shown a set of inputs and adjusted to produce consistent outputs.

In the neural network, the process element acts as a neuron and is a simple processing unit that receives and combines signals from and to other neurons. Showing in figure 16.

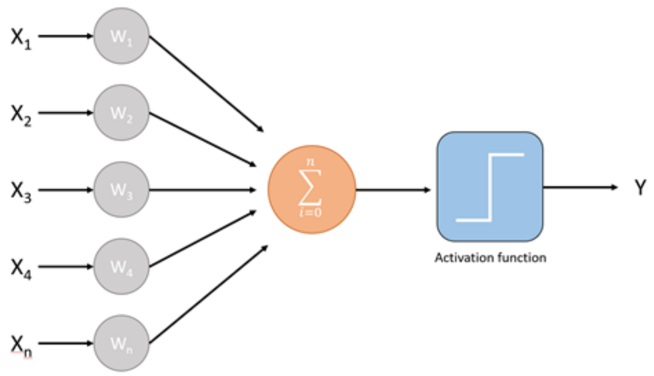

Figure 16: Artificial neuron model

The vector representation of an artificial neuron is: NET = X * W where NET is the output, X the input vector and W the vector of weights. The process element has multiple inputs that it combines as a sum. Each input has a W_i weight that is adjusted automatically as the neural network learns. This is modified by an activation function, which keeps the set of output values in a certain range, usually (0,1) or (-1,1) and giving an output value. This output can be connected in turn to the inputs of other artificial neurons, forming what we call a neural network. These process elements are organized in groups called levels or layers forming a sequence. There are two layers that connect with the outside, which are the input and the output and the rest of the intermediate layers are called hidden.

The artificial neural network as well as biological networks learn by repetition, that is why better results are achieved the larger the database to train and the more times we train it. Showing figure 17.

- Forward propagation: It is the process that goes forward, starting with neurons in the input layer and ending in neurons in the output layer. In this process, each neuron performs a weighted sum of all the inputs according to some weights, passes the result through an activation function and generates the result, which is passed on to the next layer. This process is repeated until the last layer of neurons that returns us the result of the network.

- Backpropagation: From the result of the cost function, the backpropagation algorithm tries to determine how much fault each of the neurons has on an error committed iteratively from the last layer to the first.

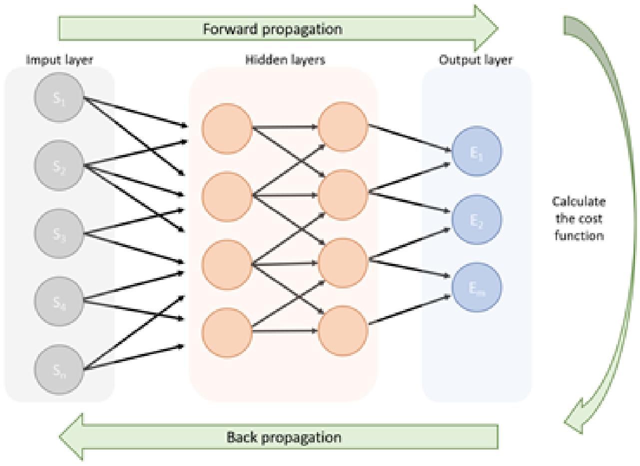

Figure 17: Learning process of a neural network

3.2 Dense Neural Network (DNN)

A DNN is actually a fully connected Artificial Neural Network (ANN). The dense layer is a neural network layer that is deeply connected, which means that each neuron in the dense layer receives information from all the neurons in its previous layer. In the background, the dense layer performs matrix-vector multiplication. The values used in the array are actually parameters that can be trained and updated with the help of backpropagation. The output generated by the dense layer is a dimensional vector "m". Therefore, the dense layer is basically used to change the dimensions of the vector. Dense layers also apply operations like rotation, scaling, translation on the vector. With respect to the learning process, DNNs use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. They can learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners [7].

The size of a DNN must be adjusted according to the characteristics of each problem, due to the limited computational power available. In this study, we used fully connected, dense neural layers where the output of one layer serves as the input to the next.

Although DNNs are based on simple mathematical operations, models that are obtained by combining neurons with non-linear activations in several hidden layers cannot be evaluated with a mathematical formula. This property of the DNN is called lack of transparency by Roscher [6]. We understand by transparency if the processes that extract the parameters of the model from the training data and generate labels from the test data can be clearly described and motivated by the designer of the approach.

3.3 Convolutional Neural Network (CNN)

CNN is a type of Artificial Neural Network with supervised learning that processes its layers to identify different characteristics in the inputs that make it able to identify objects. Convolutions are really good at detecting simple structures in an image, and then putting those simple functions together to build even more complex functions. In a convolutional network, this process occurs over a series of many layers, each of which performs a convolution on the result of the previous layer. For this, CNN contains several specialized hidden layers with a hierarchy: this means that the first layers can detect lines, curves and are specialized until they reach deeper layers that recognize complex shapes such as a face or the silhouette of an animal, as well as patterns of behavior of a time series such as our case in this study.
Convolutional neural networks are distinguished from other neural networks by their superior performance with input from image, voice, or audio signals. They have three main types of layers, which are (Showing figure 18.[4]):
- Convolutional layer.
- Pooling layer.
- Fully-connected (FC) layer.

[4]https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

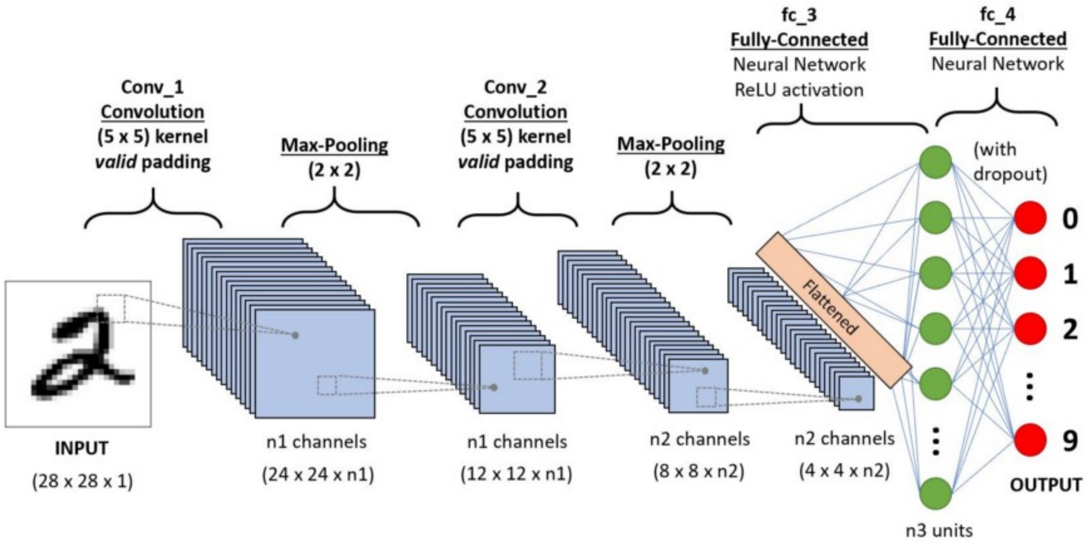

Figure 18: Process of a Convolutional Neural Network

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or grouped layers, the fully connected layer is the final layer. With each layer, CNN increases its complexity, identifying larger portions of the image. The previous layers focus on simple features, such as colors and borders. As the image data progresses through the CNN layers, it begins to recognize larger elements or shapes of the object until it finally identifies the desired object or pattern.

3.4 Long Short Term Memory (LSTM)

Recurrent neural networks, RNN, are a type of network capable of recognizing and predicting data sequences over time, such as texts, genomes, spoken speech or numerical series. The definition of recurrent is based on loops that allow the exit of the network or a part of it at a given moment to serve as an entrance to the network itself at the next moment. The operation can be described as a multilayer perceptron with a single hidden layer where the output of the perceptron is used as input in the following evaluation. This recurrence in the network architecture is the property that allows the network to "remember" information over time. As we add layers to it, its modeling capacity will grow so that it will be able to recognize larger sequences with less and less error.

LSTMs are a special type of recurring networks [16, 22]. While standard recurring networks can model short-term dependencies (close relationships in the time series), LSTMs can learn long dependencies, so it could be said that they have a longer-term "memory". LSTMs were introduced by Hochreiter and Schmidhuber (1997) [8], and were later refined and popularized.

The LSTM contains its information in memory in the sense that a neuron in an LSTM can read, write, and erase information from its memory. The neuron decides whether to store or eliminate information inside (opening the door or not to store), based on the importance it assigns to the information it is receiving. The importance assignment is decided through the weights, which are also learned by the algorithm. That is, you learn over time what information is relevant and what is not (Showing figure 19.[5])

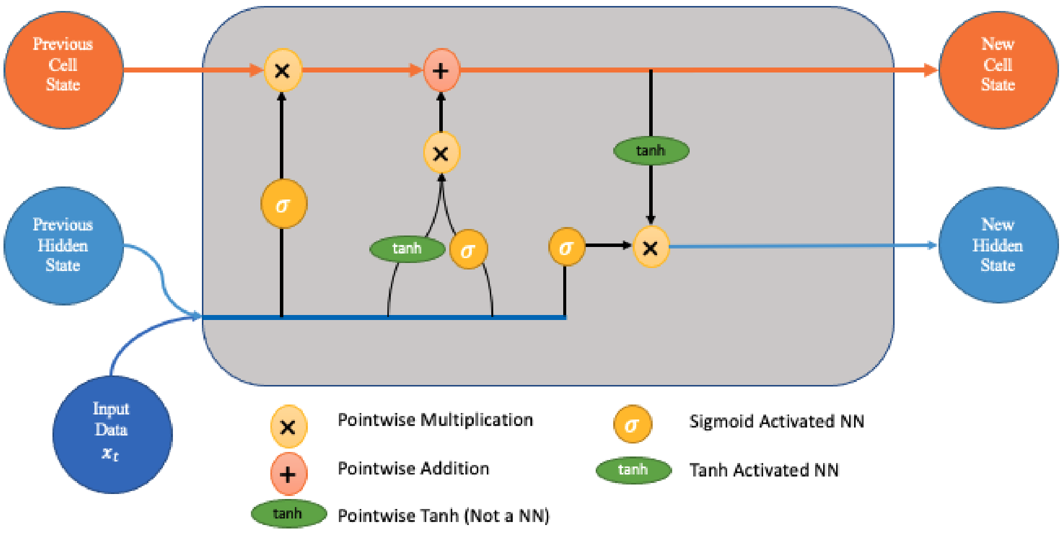

Figure 19: Process of a LSTM

In an LSTM neuron there are three gates to these information "cells": input gate, forget gate, and output gate. These doors are in charge of deciding whether or not a new entry is allowed, the information is eliminated or it is allowed to affect the exit in the current time step. The gates in an LSTM are analogous to a sigmoid shape, which means that they go from 0 to 1. The fact that they are analogous to a sigmoid activation

[5]https://https://ichi.pro/es/lstm-networks-una-explicacion-detallada/

function allows them to be incorporated into the Backpropagation process. Vanishing Gradients problems are solved through LSTM because it keeps the gradients steep enough and therefore the training is relatively short and the precision high.

3.5 Gated Recurrent Unit (GRU)

GRU layers appeared in 2014 by Kyunghyun Cho [9] et al and use the same principle as LSTM, but are simplified so that their performance is on par with LSTM but is more computationally efficient and relatively new in comparison. Like LSTM, GRU uses gates to control the flow of information. The GRU architecture offers some improvements over LSTM. Another interesting aspect of GRU is that, unlike LSTM, it does not have a separate cell state (Ct). It only has a hidden state (Ht). Due to the simpler architecture, GRUs are faster to train.(Showing figure 20).

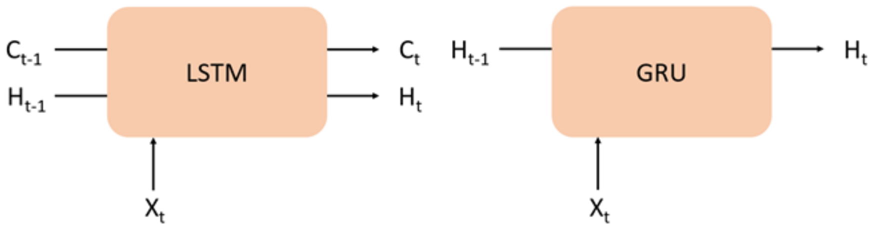

Figure 20: LSTM and GRU comparison

3.6 Persistence Model

The naive forecast or persistence algorithm is the most cost-effective prediction model and provides a benchmark against which to compare more sophisticated models. This prediction method is suitable for time series. The naive forecast method uses the value at the previous time point (t - 1) to predict the expected result at the next time point (t + 1).

$$\hat{y}_{t+1} = y_t$$

If the time series under study shows seasonality, the seasonal approach of the naive method may be more appropriate, thus being the predictions the same as the previous season.

$$\hat{y}_{t+1} = y_{t-m}$$

The persistence model does not present statistical, mathematical or computational difficulties, however, given the non-linear nature of some time series such as the use of the bicycle service, it is necessary to complement it with other more robust prediction models.

3.7 Autoregressive Model

This method models [25] the next instant in a time series or sequence as a linear function of the observations in the previous time instants. The model notation needs to specify the order of the model p as a parameter in the autoregression function AR(p). For example, AR(1) is a first-order autoregression model.

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ... + \Phi_p y_{t-p} + \varepsilon_t$$

where ε_t is the white noise, Φ_i are coefficients and c a constant.

4 Experimental settings

4.1 Experimental design

In order to choose the model that best suits our interests, we must be able to use the tools and their metrics that allow us to compare these models in the same order of magnitude.
The reasons why we want to evaluate the predictive capacity of a model are:
- We want to estimate, in a generalized way, the predictive capacity of our model for future data.
- Increase the predictive capacity by adjusting the learning algorithm and selecting the best model given a starting hypothesis.
- Identify the best ML (Machine learning) algorithm that best fits our case, through the comparison of several algorithms.
Estimating model behavior is one of the biggest challenges in ML.

4.2 Validation

After choosing and generating the models and verifying the fit method used, we must evaluate the ability of the models to predict the target variable. For the evaluation, you can use some method called resampling, among which are cross-validation and Bootstrap. These methods allow us to fit the models as many times as necessary using subsets of the training data. For this study, the method chosen is cross validation.
There are, within the model, two types of error that we must differentiate, such as the training error rate, which is the average error of the model when predicting based on the subset of data dedicated to training. By using already manipulated data, you overestimate the measure of the prediction. And the test error rate that gives us the average of the prediction on the subset of the data dedicated to the test. This is more reliable since it predicts new observations using known predictors but not the response variable.
The problem is being able to access new data sets with which to quantify the error. If they are quantitative variables, the error is measured by mean square error (MSE). If they are qualitative variables, for classification, the measurement is the proportion of incorrect variables over the total set of predictions. For this study and the evaluation of the models, we will focus on Mean Absolute Error(MAE), defined as the number of all correct predictions divided by the number of examples in the database.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i|$$

The K-Fold Cross-Validation method is an iterative process. It consists of dividing the data into k groups of approximately the same size. k-1 groups are used to train the model and one of the groups is used as a test, this process is repeated k times using a different group as a test in each iteration. The process generates k estimates of the test error, the average of which is used as the final estimate. In this case we use the version called “rolling origin” according to which the forecasting origin is updated successively and the forecast are produced from each origin, showing figure 21.

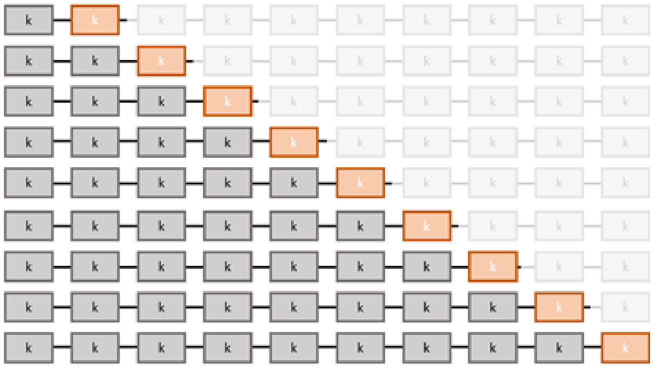

Figure 21: K-fold Cross-Validation

Two advantages of the K-Fold Cross-Validation method are considered, which are necessary less computational requirements and a better balance between bias and variance.

4.3 Hyper-parametrization and parameter tuning

To achieve the best possible results in our model, and also in our baselines, we look for hyperparameters in each model. Therefore, each hyperparameter is selected using the cross-validation approach presented above. After evaluating different values in different parameters, we present the chosen values. In table 2 and 3 we detail below the tested values for each hyperparameter.

Table 2: Characteristics of experimentation with Bicing

Model	**Time/epochs**	**Layers**	**Units**	**Epochs**	**Batch**	**Kernel**	**Params**
DENSE	3.53	5	100	100	64	-	523.839
CNN	49.23	3	100	100	64	(5,5)	154.468
LSTM	5.05	2	100	100	64	-	342.242
GRU	8.30	5	100	100	64	-	388.439

4.4 Model evaluation

To assess whether one method is a better fit than another, the variances associated with the estimates of the predictability obtained through the validation metrics must be

Table 3: Characteristics of experimentation with BiciMad

Model	Time/epochs	Layers	Units	Epochs	Batch	Kernel	Params
DENSE	2.82	5	100	100	64	-	226.569
CNN	26.64	3	100	100	64	(5,5)	124.498
LSTM	5.42	2	100	100	64	-	205.469
GRU	11.26	5	100	100	64	-	280.169

taken into account and whether there is sufficient evidence of superiority. They must be models trained on the same data, same partitions and same order.

5 Results and discussion

In this chapter we describe the decisions taken and experiments conducted in order to evaluate the performance of the several approaches introduced in chapter 3. Before explaining the results, we must establish what questions we want to answer. And they would be these:
- What models are more accurate to predict the state of the stations?
- Do the models improve by adding meteorological variables?
- Is there a correlation between cycling and weather conditions?
- Which models predict best with an erratic time series due to exceptional conditions like COVID 19?
Through these questions we will try to evaluate the results of the experimentation.

5.1 General results and feature importance

What can be seen when observing the results obtained is that, for this specific case, models based on neural networks work better. Within the four models, it is the convolutional model and the Dense model that have slightly better performance than the recurring models (GRU and LSTM). This can be seen more clearly in the Bicing models in Barcelona presented in figure 22 and figure 23 and its comparison in table 4

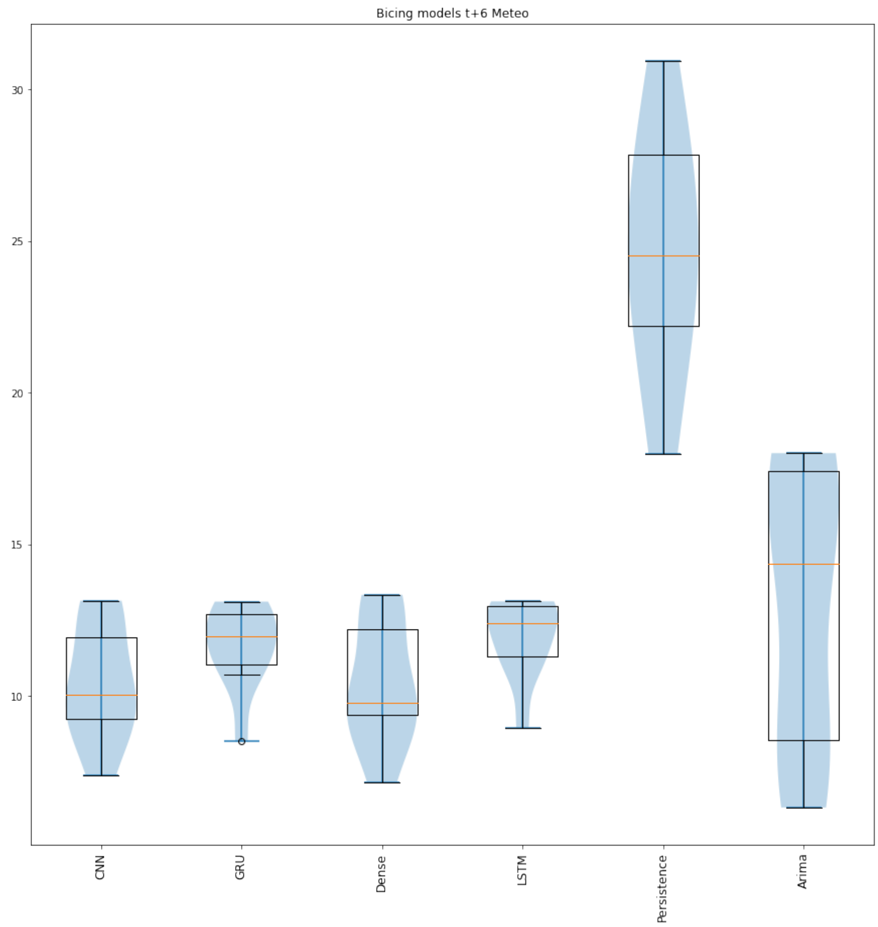

Figure 22: Bicing Models

Table 4: MAE Bicing and BiciMad

Meteo t + 6	**Bicing**		**BiciMad**	
Model	**Median MAE**	**Mean MAE**	**Median MAE**	**Mean MAE**
CNN	10.03	10.44	9.09	9.66
GRU	11.97	11.66	9.31	9.98
DENSE	9.77	10.42	9.19	9.8
LSTM	12.41	11.93	9.43	10.05
PERSISTENCE	24.51	24.60	10.96	10.61
ARIMA	14.36	13.04	9.91	14.31

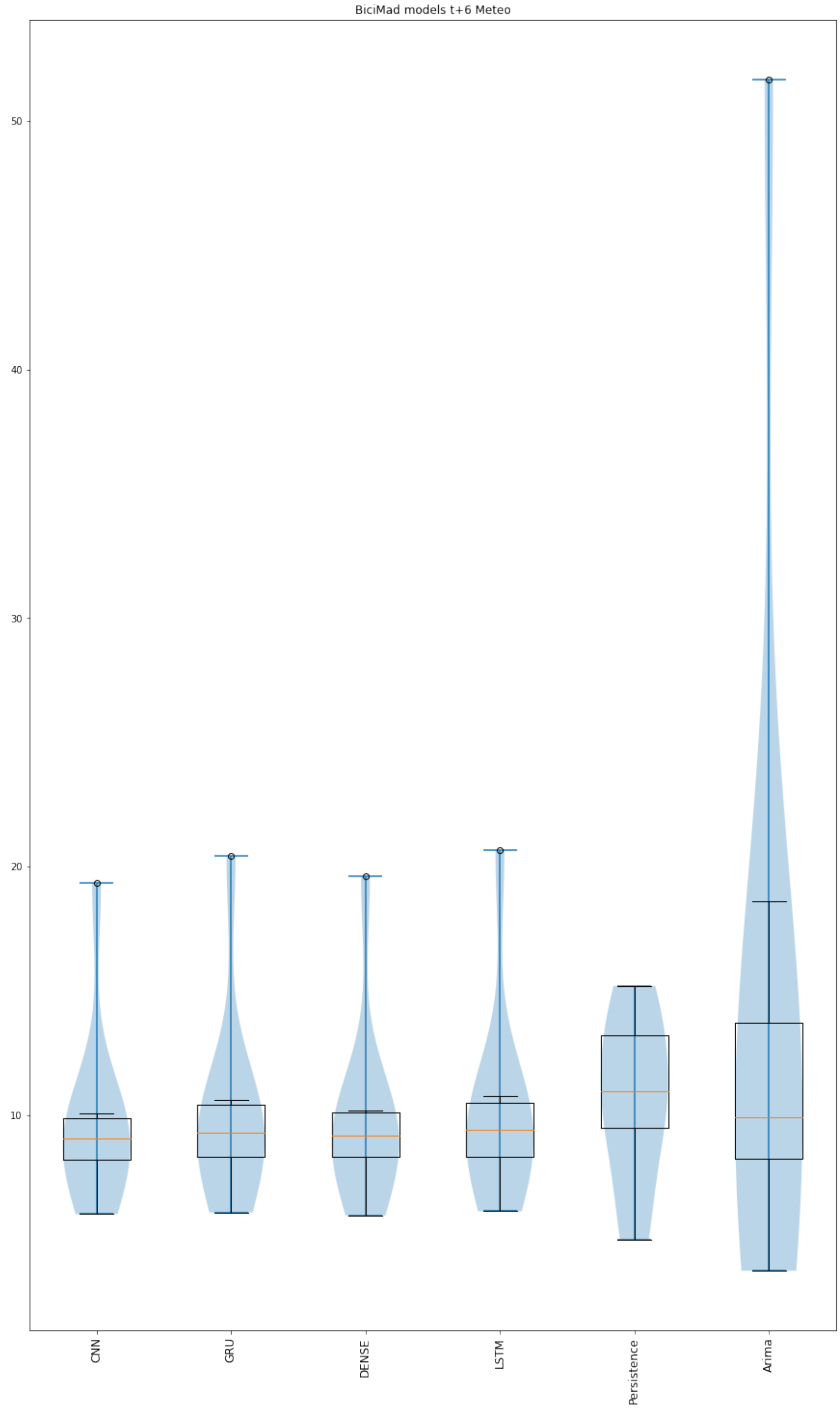

Figure 23: BiciMad Models

The Persistence and Arima models perform significantly worse. However, as we can see in the tables, despite having a good performance, the convolutional network takes a much longer calculation time than the rest of the models. From a performance point of view, we should go for any of the others. This leads us to wonder how to achieve better performance of the convolutional network, so we tried a new experiment that consists of comparing the two best networks (Dense and CNN) and with a similar behavior so far, in the 12-month situation before the Covid Pandemic (March 2019 to February 2020) with the case in which we include the 6 months after (March 2019 to August 2020). With this, what we intend is to stress the model since the time series behaves erratically with a month without movements (since March 14, 2020) and a subsequent recovery of services in a staggered manner. It is in this case when we can appreciate the virtues of the neural network in all its dimension, improving performance compared to Dense. (Showing figure 24 and table 5)

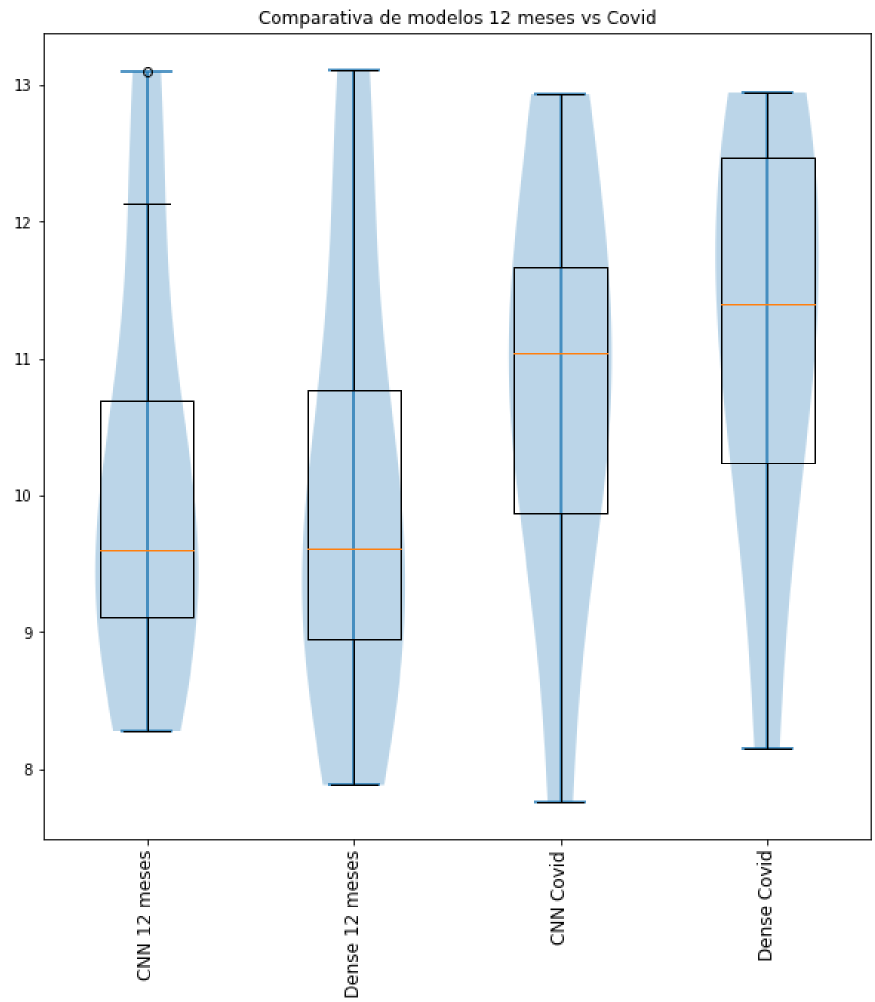

Figure 24: Comparison of models precovid vs postcovid

Table 5: MAE Bicing. Covid vs 12 months

Bicing	**Model**	**Median MAE**	**Mean MAE**
Covid	CNN	11.04	10.84
	DENSE	11.40	11.22
12 months	CNN	9.61	10.11
	DENSE	9.62	10.04

But following this reasoning, we think that we should compare the matrices with the meteorological variables versus the matrices without them. Logically, the meteorological variables should give better results. (Showing figure 25, figure 26 and table 6).

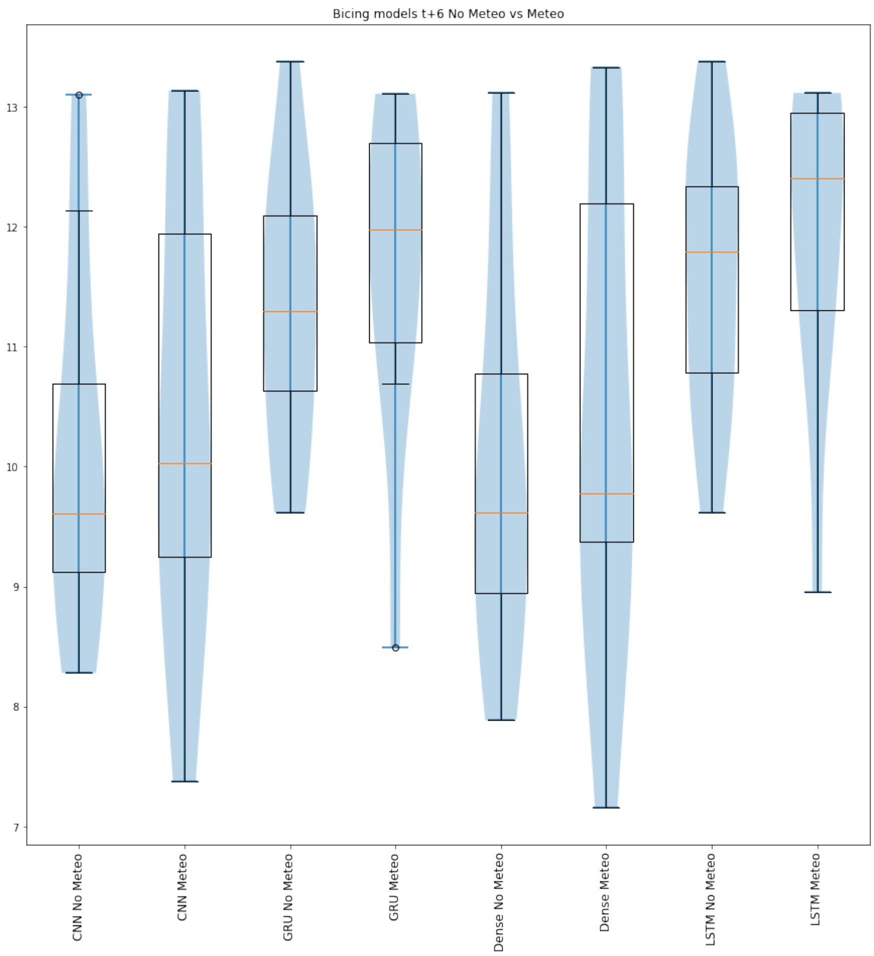

Figure 25: Comparison of models Bicing meteo vs no meteo

Table 6: MAE Bicing and BiciMad 12 months

12 months	Bicing		BiciMad	
Model	**Median MAE**	**Mean MAE**	**Median MAE**	**Mean MAE**
CNN	9.61	10.11	8.84	8.80
GRU	11.29	11.31	8.96	9.03
DENSE	9.62	10.04	8.81	8.83
LSTM	11.79	11.59	9.03	9.16

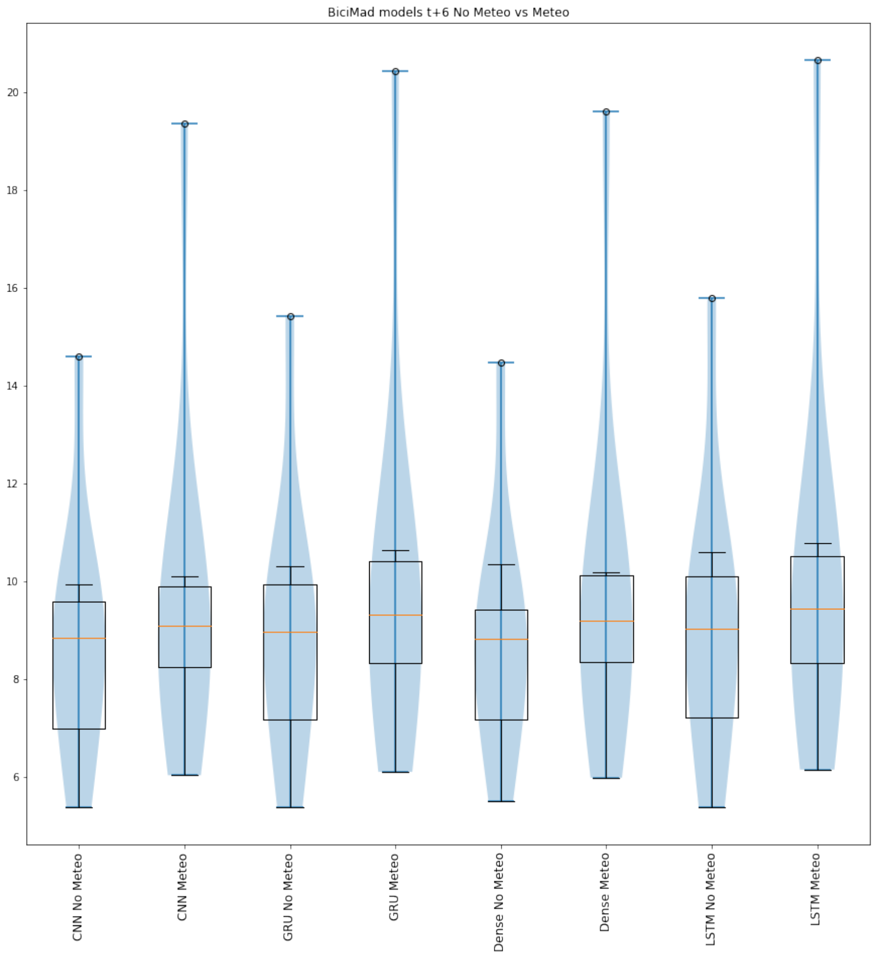

Figure 26: Comparison of models Bicing meteo vs no meteo

What may seem surprising to us, which is that the matrices work better without the meteorological variables, finds its logic in that having daily meteorological data and repeating it 24 hours a day (as if the meteorological conditions of wind, rain or temperature remained unchanged at throughout the day), not only is it not an advantage, but quite the opposite, since it gives erroneous information when it comes to correlating uses with weather conditions.

5.2 Conclusions

In this work we set ourselves as objectives:
- The comparison of different predictive models of neural networks on the data of two cities (Madrid and Barcelona) according to their usage habits in a time series format and adding climatological factors such as temperature, rain and wind.
- Identify the best ML algorithm that best fits our case, through the comparison of several algorithms.
- To build a model capable of predicting at least 6 hours in advance which stations are likely to fill or empty and evaluate the predictive capacity of the model.

We can consider that the objectives have been met and that the lessons learned are:
- The hyperparameterization of the models to be able to compare them in a fair way is essential for the data obtained to be reliable and to avoid any biases or a directing of our research.
- Neural networks have obtained better results but have required greater computational capacity. Depending on the precision we need, it may be more efficient to use linear regressions that lose precision but gain speed and save computational capacity, as we can see in our ARIMA model.
- The use of meteorological variables should help our model, the wind and rain being more decisive in extreme situations, but for this it is necessary to have the complete series of data.
- Although a priori, we could think that recurring networks should behave better than convolutional ones, the fact that CNN is able to compete with Dense with almost three times fewer parameters (and therefore less biased to overtraining and better capacity to generalization) requires us to open our minds to predetermined ideas.

5.3 Further developments

For future developments we should be able to obtain conclusive results on the influence of meteorological variables, as well as find other types of variables that can condition the behavior of cyclists such as pollution, points of interest such as restaurants, museums and shows and the temporary events that occur in the city such as sporting events.
We also propose to use a database with a longer period of time, between 3 and 5 years, where the models can test their performance with greater reliability.
The stations of origin or destination of the vehicles that operated in each station were not within the scope of the study, but it would be very interesting to add these data to the matrices to analyze the behavior of the models and check if it gives greater robustness to the resulting forecasts. In addition, it would provide information on the flow of movements of cyclists and their behavior within the city.

References

[1] Melissa Tan: *"Micro-mobility Sharing Underpinning City Mobility Networks"*, Frost and Sullivan's. (2021)

[2] Ida Bagus Irawan Purnama, Neil Bergmann, Raja Jurdak y Kun Zhao: *"Characterising and predicting urban Mobility Dynamics by mining bike sharing system data"*, Proceedings of the 12th IEEE International Conference on Ubiquitous Intelligence and Computing (UIC 2015). IEEE, United States of America, pp. 159-167.

[3] Seguí Pons, Joana Maria; Mateu Lladó, Jaume; Ruiz Pérez, Maurici: *"Los sistemas de bicicleta pública y la movilidad urbana sostenible. Un análisis en la ciudad de Palma (Mallorca, Islas Baleares)"*, Boletín de la Asociación de Geógrafos Españoles, Núm. 71 (2016), p. 227-245. DOI 10.21138/bage.2281.

[4] Ali, Usman, and Mahmood, Tariq: *"Using Deep Learning to Predict Short Term Traffic Flow: A Systematic Literature Review."*, Intelligent Transport Systems – From Research and Development to the Market Uptake 222 (2018): 90-101. Web.

[5] Rodrigo de Medrano, José L. Aznarte: *"A spatio-temporal attention-based spot-forecasting framework for urban traffic prediction"*, Applied Soft Computing, Volume 96, 2020, 106615, ISSN 1568-4946

[6] Ribana Roscher, Bastian Bohn, Marco F. Duarte, Jochen Garcke: *"Explainable Machine Learning for Scientific Insights and Discoveries"*, Institute of Electrical and Electronics Engineers (IEEE), Volume 8, ISSN 2169-3536

[7] Moustakidis, S., Papandrianos, N.I., Christodolou, E: *"Dense neural networks in knee osteoarthritis classification: a study on accuracy and fairness"*, Neural Comput and Applic (2020)

[8] Hochreiter, S. and Schmidhuber, J.: *"Long short-term memory"*, Neural computation, 9(8):1735–1780. (1997)

[9] Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua: *"Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation"*, Computer Science, Mathematics. ArXiv. (2014)

[10] Pengfei Lin, Jiancheng Weng, Quan Liang, Dimitros Alivanistos, Siyong Ma: *"Impact of whearher conditions and built environment on public bikesharing trips in Beijing"*, Springer Science. (2019)

[11] Medrano, Rodrigo De, and Aznarte, José L: *"On the Inclusion of Spatial Information for Spatio-temporal Neural Networks."*, Neural Computing and Applications (2021): Neural Computing and Applications, 2021. Web.

[12] Chen, Po-Chuan, Hsieh, He-Yen, Su, Kuan-Wu, Sigalingging, Xanno Kharis, Chen, Yan-Ru, and Leu, Jenq-Shiou: *"Predicting Station Level Demand in a Bike-sharing System Using Recurrent Neural Networks."*, IET Intelligent Transport Systems 14.6 (2020): 554-61. Web.

[13] Wang, Senzhang, Cao, Jiannong, and Yu, Philip: *"Deep Learning for Spatio-Temporal Data Mining: A Survey."*, IEEE Transactions on Knowledge and Data Engineering (2020): 1. Web.

[14] Xiao, Guangnian, Wang, Ruinan, Zhang, Chunqin, and Ni, Anning: *"Demand Prediction for a Public Bike Sharing Program Based on Spatio-temporal Graph Convolutional Networks."*, Multimedia Tools and Applications 80.15 (2020): 22907-2925. Web.

[15] Prybutok, V. R., Yi, J., and Mitchell, D: *"Comparison of neural network models with arima and regression models for prediction of houston's daily maximum ozone concentrations."*, European Journal of Operational Research, (2000) 122(1):31–40.

[16] Hochreiter, Sepp, and Schmidhuber, Jürgen: *"Long Short-Term Memory."*, Neural Computation 9.8 (1997): 1735-780. Web.

[17] Gamboa, John Cristian Borges: *"Deep Learning for Time-Series Analysis"*, (2017). Web.

[18] P. Borgnat, C. Robardet, J-B. Rouquier, P. Abry, E. Fleury, and P. Flandrin: *"Shared bicycles in a city: A signal processing and data analysis perspective."*, Advances in Complex Systems, 14 (3): 415-438, 2010.

[19] R. Beecham, J. Wood, and A. Bowerman: *"Studying commuting behaviours using collaborative visual analytics."*, Computer, Environment and Urban System, 47:5-15, 2014.

[20] A. Asahara, A. Sato, K. Maruyama, and K. Seto: *"Pedestrian-movement Prediction based on Mixed Markov-chain Model."*, In Proceedings of the 19th International Conference on Advances in Geographic Information Systems, pages 25-33, USA, 2011.

[21] Romeu, P., Zamora-Martínez, F., Botella-Rocamora, P., Pardo, J: *"Time-series forecasting of indoor temperature using pre-trained deep neural networks."*, Artificial Neural Networks and Machine Learning ICANN 2013, pp. 451458. Springer (2013)

[22] Gers, F. A., Eck, D., Schmidhuber, J.: *"Applying lstm to time series predictable through time-window approaches."*, Neural Nets WIRN Vietri-01. Springer, pp. 193–200. (2002)

[23] Grover, A., Kapoor, A., Horvitz, E.: *“A deep hybrid model for weather forecasting.”*, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. (2015)

[24] Hyndman, R. J., Athanasopoulos, G.: *“Forecasting: principles and practice.”*, OTexts. (2018)

[25] Khashei, M., Bijari, M: *“A novel hybridization of artificial neural networks and arima models for time series forecasting.”*, Applied Soft Computing 11(2), 2664–2675. (2011)

[26] Liu, J. N., Hu, Y., He, Y., Chan, P. W., Lai, L.: *“Deep neural network modeling for big data weather forecasting.”*, Information Granularity, Big Data, and Computational Intelligence. Springer, pp. 389–408. (2015)