ANALYSIS AND VISUAL EXPLORATION OF PREDICTION ALGORITHMS FOR PUBLIC BICYCLE SHARING SYSTEMS

Alexandra Cortez Ordoñez and Pere-Pau Vázquez
ViRVIG, Universitat Politècnica de Catalunya
Carrer Llorens I Artigas, 4-6, 08028 Barcelona, Spain

ABSTRACT
Public bicycle sharing systems have become an increasingly popular means of transportation in many cities around the world. However, the information shown in mobile apps or websites is commonly limited to the system’s current status and is of little use for both citizens and responsible planning entities. The vast amount of data produced by these managing systems makes it feasible to elaborate and present predictive models that may help its users in the decision-making process. For example, if a user finds a station empty, the application could provide an estimation of when a new bicycle would be available. In this paper, we explore the suitability of several prediction algorithms applied to this case of bicycle availability, and we present a web-based tool to visually explore their prediction errors under different time frames. Even though a quick quantitative analysis may initially suggest that Random Forest yields a lower error, our visual exploration interface allows us to perform a more thorough analysis and detect subtle but relevant differences between algorithms depending on variables such as the station’s behavior, hourly intervals, days, or types of days (weekdays and weekends). This case illustrates the potential of visual representation together with quantitative metrics to compare prediction algorithms with a higher level of detail, which can, in turn, assist application designers and decision-makers to dynamically adjust the best model for their specific scenarios.

KEYWORDS
Visualization Systems and Tools, Visual Analytics, Bike Sharing Systems, Forecasting Algorithms

1. INTRODUCTION

During the last decades, the popularity of public bicycle sharing systems (BSS) has been significantly increased in many medium and large cities. Not only have these transportation systems brought about a noticeable reduction in traffic congestion, noise, and pollution but also have had a considerable positive impact on user’s health (Ricci, 2015). Despite their substantial benefits, local governments often delegate the management of their operations to contractors through public tenders. These operators are compensated independently of the number of users, thus having little incentive to promote, improve, or increase its use. As a result, BSS management and end-user applications are not quite sophisticated and provide a very limited amount of information, sometimes even erroneous, harming user’s enthusiasm towards this public service (Chen et al., 2020). The advance of Information Technologies (IT) and the Internet of Things (IoT) has enabled the collection of richer data to a greater extent which is unfortunately not used today and could be leveraged to include modern features such as predictions with manifold benefits.

In our previous paper (Cortez & Vázquez, 2021) we created a web-based visualization system for the exploratory analysis of Barcelona’s BSS. Our system offers novel features not existing in others such as the prediction of slots or bicycles of each type (mechanical or electrical), analysis of docking stations suffering outages, or a prioritized list of stations close to a point of interest. Its development was based on the requirements and input from regular users and experts in policy-making, system operation, and urban planning, collected through several semi-structured interviews. The visualization system, showed in Figure 1, allows the first group to obtain precise information on the future availability of bicycles or drop-off slots to assist them in deciding whether to wait for a bicycle, walk to a nearby station or use other means of transportation. Besides, the system helps the second group understand patterns of usage and get deeper
insights. For instance, policymakers could take decisions regarding resizing, changing the placement of docking stations, or reinforcing other public transport systems. Managers, on their side, could use the information to improve re-balancing operations. After the evaluation of the system, we found that both groups of users considered highly useful the availability prediction of mechanical and electrical bikes and free slots in the following two hours. This has been the initial motivation that drives us to investigate various prediction models and the best way to evaluate them easily and efficiently.

In this paper, we seek to close the gap in extant research by analyzing the suitability of different prediction algorithms and their most influential variables using Barcelona’s BSS as a use case. Our innovative interactive application is useful for bike-sharing visualization systems developers. It facilitates the rapid comparison and understanding of the strengths and weaknesses of different prediction algorithms by visually presenting the information. This tool was initially conceived to compare the performance of four forecasting models: Random Forest, Linear Regression, ARIMA, and Prophet. However, the application was developed in a modular way to ensure transferability to other similar data sources.

Figure 1. The initial web-based visualization system was applied to Barcelona’s Bicing. The map shows the availability of bicycles in stations (red colors indicate low availability). The right view displays the prediction for the next two hours for the selected station and the closest ones prioritized by difficulty to access (altitude + distance)

2. RELATED WORK

Large studies for comparing different prediction models have been developed in the last years in a variety of domains, for instance on economics (Alon et al., 2001), urban traffic (Vázquez et al., 2020), health (Lim et al., 2000) or more recently COVID propagation (Appadu et al., 2021). Other investigations try to focus on a more general domain such as time series (Bontempi et al., 2012; Sharda & Patil, 1992) or classification problems (Caruana & Niculescu-Mizil, 2006; King et al., 1995). We have found few extensive studies for the general comparison of prediction algorithms (Ahmed et al., 2010; Fischer et al., 2018; Nguyen & Park, 2020). These investigations mainly concentrate on the analysis of novel machine learning algorithms vs traditional forecasting methods. Their results are presented using tables or the default plots provided by the software employed. We share with them the interest in the comparison of traditional and novel forecasting techniques; however, we also focus on a specific domain, BSS, and provide bike-sharing visualization developers with a novel visual interface to compare the outcomes of the selected models easily and dynamically.

The analysis of BSSs usage has been addressed in multiple ways. One common approach is to study the effect of urban configuration on bicycle flows and destination preferences (Faghih-Imani et al., 2014; Faghih-Imani & Eluru, 2015). Other investigations analyze how the elevation of stations may affect the trips
(Frade & Ribeiro, 2014; I. Kim et al., 2020). The influence of important events (Xie & Wang, 2018), and the correlation to weather data (Gehbart & Noland, 2014) and other variables, such as the calendar events (El-Assi et al., 2017; K. Kim, 2018; Younes et al., 2020) have been also studied. Therefore, building on a common focus to characterize bicycles’ availability and use, this study presents suitable prediction algorithms and thorough comparison.

Other studies evaluate the pickup and drop-off events applying regression to model the counts of bicycles (Holmgren et al., 2017, 2018). However, they only provide numerical measurements from which is difficult to get deeper insights such as the algorithms performance on weekdays and weekends, or how the amount of activity may affect them. We also evaluated regression and compared it with other prediction models. More sophisticated algorithms have been applied to BSS scenarios, such as Graph Convolutional Neural Networks (Lin et al., 2018), hierarchical prediction model and Gradient Boosting Regression Tree (Li et al., 2015), among others. These papers propose a single model to analyze the bikes’ demand. In contrast, we are interested in bike availability and we evaluate four different prediction algorithms that are commonly used, easy to understand, and available through various libraries in R or Python. The comparison between different forecasting algorithms in the BSS context has also been investigated (Feng & Wang, 2017; Hulot et al., 2018a; Wang & Kim, 2018; Xu et al., 2018), but these studies mainly illustrate their findings with tables and static charts, and no visual tool is developed to help users to drill down the data and gather a deeper understanding of the prediction algorithms’ behavior.

Finally, this work is a continuation of a previous paper (Cortez & Vázquez, 2021) where we created a visualization system for the exploratory analysis of Barcelona’s BSS. From our interviews with users, system operators, and policymakers, we found that even if the prediction of bike availability for the following two hours was considered highly useful, it was not present in official applications, mainly focused on usage patterns analysis (Dai et al., 2020; Froehlich et al., 2009; Shi et al., 2019) and availability status (Meddin et al., 2020; Oliveira et al., 2016; Oppermann et al., 2018). These interviews motivated our interest to further investigate and compare different prediction algorithms that can be applied to BSS scenarios for which we developed a visual interface that facilitates a quick and effective evaluation of their performance.

3. METHODS AND RESULTS

In this section, we present the process followed to collect the data and select the features, a description of the models used, and the results. The forecasting and initial data processing are performed using R. The visual interface is built using D3 and Flask, and data management is carried out in Python.

The first step we need to perform is the data gathering and processing. Given that the information needed for this type of analysis differs among public BSS, we concentrate on Barcelona’s BSS named Bicing. Its data is available through the Open Data BCN website which is updated monthly. For each month there are more than 3.5 million records since data collection occurs approximately every 5 minutes. For this study, only information from 2019 is used to train and evaluate the models since 2020 had an abnormal behavior due to the lockdown and several mobility restrictions. Moreover, the company that provides the bike service in Barcelona was gradually changed during 2019, which caused a lack of continuity in the information. Thus, we focus on the last months of 2019 from September to November.

The Barcelona BSS has 424 docking stations, however, after data cleaning only 409 stations are kept. Data treatment includes deleting information of docking stations, that even when appearing in the set, are still not used, as well as others that were not working properly. Besides, garbled information such as dates outside the analyzed range or negative data availability is removed. From these data, the additional information necessary to train the models is computed: day of the week (and month), total docks available in each station, and the number of pickups and drop-offs (absolute and relative) per hour. Considering that we want to get further insights into the behavior of the algorithms, we also calculate usage ratios by hour.

Usage Ratios are derived from the frequency of pickups and drop-offs and help us differentiate between static and dynamic docking stations. It is calculated as follows:

\[ \text{use}_{ijk} = \sum \text{arrivals}_{ijk} + \sum \text{departures}_{ijk}, \text{where } i = \text{StationID}, j = \text{day}, k = \text{hour} \]
To better interpret this ratio and have a common scale we re-scale these values between 0 and 1:

\[
\text{Usage Ratio}_{ijk} = \frac{(\text{use}_{ijk} - \text{MinUse})}{\text{MaxUse} - \text{MinUse}}, \quad \text{where } i = \text{StationID}, j = \text{day}, k = \text{hour}
\]

Once our data is clean, we proceed with the selection of features to train the prediction algorithms. It is based on the information available in Barcelona’s BSS and previous studies (Lin et al., 2018; Wang & Kim, 2018; Xu et al., 2018) which have contributed with valuable knowledge to understand the factors that might affect the bike-sharing demand. These attributes can be classified as station-related or weather features. The first group is composed by day of the week (from Monday to Sunday), time slot, number of available bikes (mechanical and electrical), and station’s capacity. Weather features are temperature, wind, humidity, atmospheric pressure, and overall weather classified in four categories: sunny, cloudy, light rain, and heavy rain. The features day of the week, time slot, and weather are categorical variables while the others are numerical.

Then, we analyze the performance of different prediction algorithms. Several models to predict bike sharing demand have been tested by different authors (Lin et al., 2018; Wang & Kim, 2018; Xu et al., 2018), some of them are quite sophisticated and developed specifically for this scenario while others are extensively used in many domains. For this study, we focus on prediction algorithms that are popular among data scientists because of their ease of use and interpretation, their availability through libraries in R and Python and their low computational resources requirements. Besides, similar to several studies (Alon et al., 2001; Sharda & Patil, 1992), we test the performance of traditional methods (Linear Regression, and Autoregressive integrated moving average) and new machine learning techniques (Random Forest, and Prophet). We have considered Linear regression as a base model because of its common use in predictive analysis. The Autoregressive integrated moving average is widely used to predict events that happen over time. Given that interpretability of the models and their parameters is not a concern for this study, Random Forest is chosen due to its ability to avoid overfitting. Previous works (Hulot et al., 2018b; Wang & Kim, 2018; Xu et al., 2018) show how these models have been broadly used in the BSS field. Finally, we also test Prophet as an example of novel methods developed by private companies which are gaining more relevance in recent years. The description of the models applied is presented as follows:

**Linear Model** is one of the most popular and studied forecasting methods. It can be defined as an approach to modeling the relationship between two or more variables by fitting a linear equation to observed data (Darlington & Hayes, 2017). One variable is known as the dependent variable and the others are considered as explanatory or independent variables. The relationship between them is constructed using linear predictor functions and a disturbance term or error that adds "noise" to this relationship. Linear models are widely used since their unknown parameters and their statistical properties are easier to estimate from the data itself.

**Autoregressive integrated moving average** or Arima is one of the most widely used approaches to time series forecasting. It can be considered as a combination of the differenced autoregressive model with the moving average model (Hillmer & Tiao, 1982). Its components are: i) Autoregression (AR) which shows that the time series variable is regressed on its own lagged or past values. ii) Integrated (I) part represents the differenced values of d order necessary for the time series to become stationary, it means raw values are replaced by the difference between data and the previous values. iii) Moving average (MA) indicates the dependency between an observation and its residual error, i.e., the forecast error can be represented as a linear combination of past errors. Each component is represented as a parameter with a standard notation: i) p: the lag order, ii) d: the degree of difference, iii) q: the order of the moving average.

**Random Forest** is a supervised learning algorithm that consists of randomly generating multiple decision trees and training them using the bagging method (Feng & Wang, 2017; Ho, 1995). As individual decision trees mainly suffer from overfitting, Random Forest trains decision trees with different parts of the training set by creating random subsets of the features and building smaller trees using those subsets. Then, the subtrees are combined reducing the variance. One of the advantages of this method is to increase the overall result as the combination of many learning models will outperform any individual model. Its main limitation is its slowness and the high need of computational resources for many trees.
Prophet is open-source software released by Facebook in 2017 and implemented in R and Python (Taylor & Letham, 2017). It is specifically developed for time series and based on an additive model where non-linear trends are fitted with different seasonality (yearly, weekly, and daily), and holiday effects. Lately, Prophet has gained popularity due to its simplicity as it allows users to adjust parameters intuitively without knowing the details of the underlying model. Moreover, it is robust to outliers, missing data, and dramatic trend shifts.

For each station, all models are trained using September and October data and both mechanical and electrical bikes’ availability are predicted for November 2019. Furthermore, the natural logarithm of the predicted variable (availability) is also computed to check if results could be improved, but it did not improve the performance of the original variables. We also evaluate several error measures to compare the models. Our primary choice is the root-mean-square error (RMSE) as others (e.g., MAE, MAPE) were strongly correlated and did not add any valuable insights.

The error measures comparison for the four models described is shown in Table 1. Results suggest that electrical bikes’ availability prediction has the smallest levels of error as a result of low variance. In general, using non-transformed data in both mechanical and electrical forecasting, Random Forest is giving the lowest RMSE levels using an optimal amount of 500 trees. Similarly, the Multiple Linear Regression model also exhibits good data fitting, while requiring less computational resources. In contrast, Arima and Prophet do not adequately capture the characteristics of bicycle trips and cannot make an accurate prediction. A possible explanation is their better suitability for time series data with strong seasonal effects and several seasons of historical data, which is not our case for the reasons abovementioned. Table 1 also displays the RMSE for the time slot from 7:00 to 21:00, when the Bicing usage ratio is higher compared with other periods when it is close to zero. Moreover, data is divided into weekdays and weekends since their usage patterns are quite different. The analysis using different time frames allows us to discover interesting insights. For instance, the error is smaller when we focus on the period with the highest usage ratio (from 7:00 to 21:00). Likewise, the error on weekends tends to be higher than on weekdays. Another interesting result is the similar performance of some algorithms during weekends.

Table 1. RMSE results. Random Forest seems to be more suitable for capturing trips characteristics. However, a deeper analysis shows some interesting outcomes.

<table>
<thead>
<tr>
<th></th>
<th>Mechanical / Electrical</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE weekdays 7:00-21:00</td>
<td>RMSE weekdays 7:00-21:00</td>
<td>RMSE weekends 7:00-21:00</td>
<td>RMSE weekends 7:00-21:00</td>
<td></td>
</tr>
<tr>
<td>Random Forest</td>
<td>5.11 / 0.60</td>
<td>4.99 / 0.57</td>
<td>5.92 / 0.64</td>
<td>5.64 / 0.61</td>
<td></td>
</tr>
<tr>
<td>Linear Model</td>
<td>5.44 / 0.66</td>
<td>5.34 / 0.63</td>
<td>5.53 / 0.63</td>
<td>5.10 / 0.59</td>
<td></td>
</tr>
<tr>
<td>Prophet</td>
<td>5.98 / 0.65</td>
<td>5.86 / 0.62</td>
<td>5.85 / 0.62</td>
<td>5.63 / 0.58</td>
<td></td>
</tr>
<tr>
<td>Arima</td>
<td>6.48 / 0.69</td>
<td>6.65 / 0.67</td>
<td>6.70 / 0.67</td>
<td>6.84 / 0.65</td>
<td></td>
</tr>
</tbody>
</table>

4. VISUAL TOOL

Obtaining numerical performance measures for prediction algorithms is not enough if we want to fully understand the complete behavior of the system. It is well known that descriptive statistics such as the average or the median, overly simplify information. In our context, users may be interested in investigating whether the prediction algorithms have larger errors at certain moments of the day, or whether they are influenced seasonally or by the day of the week. These values could be difficult to analyze in tables, therefore, it is very useful to visualize the output of the prediction algorithms against each other, as well as with different configurations.

Overview. We have designed a web-based multiple-view application that facilitates the analysis of these algorithms through a set of heatmaps as shown in . There are three linked views. The control panel (1) is used to select the parameters to analyze. The larger portion of the screen (3) is devoted to four different heatmaps that encode the RMSE results for each prediction algorithm. Details can be accessed using the details-on-demand view (2) named Maximum errors. It provides contextual information about maximum errors, the id of the station where it is produced, and when these values are found.
Filtering and interaction techniques. To facilitate the exploration, we provide several filtering modes. The default mode shows the Hour view where all stations are arranged along the X-axis and hours in the Y-axis. Through the control panel, users can switch between different days of the month, group information by time ranges, or hide early hours which usually have usage ratios close to zero. This configuration can also be changed to the Day view, where users can visualize the summary of RMSE per station and the day of the week (from Monday to Sunday). It also allows users to display data for the whole week, filtered by weekdays or grouped by weekend/weekday.

Given that the Bicing system has both electrical and mechanical bikes, and their specific availability is a strong decision factor (Cortez & Vázquez, 2021) (e.g. uphill trips), it becomes interesting to analyze whether their prediction algorithms errors exhibit different behavior. Therefore, we also let the user toggle between electrical or mechanical bikes.

Figure 2. Visual Interface. The control panel (1) lets the user control what variables are shown in the heatmaps (3). (2) shows the maximum errors in all algorithms, as well as the stations and when these errors are produced.

As we explain in the next section, a smaller number of bicycles picked and dropped off might result in higher errors in the prediction. Hence, each error heatmap can be transformed into an activity heatmap by toggling its top-right checkbox labeled Show Activity.

We have implemented some cross-highlighting methods to facilitate the analysis and allow users to identify patterns that cannot be easily detected through tables or general numerical measures:

- Similar behavior between prediction algorithms.
- Larger prediction errors on certain days or hours.
- Relationship between errors and usage ratios.
- Stations that exhibit larger errors for all algorithms.

This way, instead of inspecting a single algorithm, users can analyze the four of them at once and check if they behave similarly for distinct parameters such as day of the month, time of the day, etc. Upon hovering on any heatmap, the station under the mouse is highlighted in all the heatmaps and users can compare easily the RMSE performance. It is also possible to select a particular station by looking for its ID using the Search button or emphasize a set of stations using the Filter stations button. As the number of stations is high (more than 400), just labeling them would pose difficulties to the user to relate the same position in different heatmaps.

5. USE CASES

This section presents two of the more significant use cases that we have found.
Use case 1. Our visual tool allows us to easily identify different patterns that can be useful to select or discard a model, or to dynamically choose the best algorithm for different time frames. Using the heatmaps it is possible to identify a similar behavior between Random Forest and Multiple Linear Regression algorithms, as shown in . Even though the error is slightly larger in the case of regression, both models show similar results. Switching between the activity and the RMSE also reveals interesting discoveries. For instance, during weekdays, Arima tends to be less accurate between 10 a.m. and 6 p.m. when the activity is high (-top right). From 12 a.m. to 6 a.m., the activity is very low in all stations, thus, the application offers an option to remove this time range and focus only on the hours with the highest activity and, therefore, the greatest interest to analyze. Moreover, it can be inferred that stations with lower usage ratios usually have higher prediction errors and Prophet is especially sensitive to this situation as shown in Figure 3.

![Prophet](image1)

![Prophet](image2)

Figure 3. The Show Activity checkbox allows the user to toggle between error data (top) and activity data (bottom). Stations with higher usage ratios are predicted better than the stations with a lower amount of activity.

Use case 2. The need for different prediction models for the same station in different time frames can be rapidly identified using our tool. For example, when aggregating RSME results and sorting them by day of the week (Day view), both Random Forest and Prophet show that station 171 presents the biggest prediction error for mechanical bicycles on November 30 (Figure 4-top right). If we only consider working days, station 171 continues to be the one with the greatest error on November 22. If the Hour view is selected and station 171 is highlighted, the behavior of the RMSE for each algorithm and day of the month can be inspected.
In fact, this station presents higher levels of RMSE error during several days. On November 22, the visual analysis and the exploration of the tables presented on the details on-demand view (Figure 4-top left), allows us to identify that each model performs better in a different time frame, which may suggest choosing a different algorithm depending on the range of hours to improve the prediction.

![Figure 4](image)

**Figure 4.** The details-on-demand view shows the maximum errors for each algorithm. On the left, the Hour view displays the hours with the largest errors for the selected day. The Day view, on the right, shows the day with the largest RMSE.

### 6. CONCLUSION

In this study, a novel approach for the selection of prediction algorithms through visual exploration has been presented. Several widely used prediction algorithms have been analyzed in the context of public Bicycle Sharing Systems (BSS) using Barcelona's open data. For this purpose, an interactive full web-based visualization system was developed to represent in an easy and intuitive manner the forecasting error (RMSE) obtained from the application of the following four prediction algorithms to our dataset: Multiple Linear Regression, Arima, Random Forest, and Facebook's Prophet. A quick quantitative analysis suggests that Random Forest performs better and gives the smallest RMSE to predict the availability of both mechanical and electrical bikes, which often display distinct patterns. Nonetheless, a single metric does not usually represent the whole picture in such dynamic systems and can lead to sub-optimal decisions. For example, we observe that algorithms perform differently depending on the station's behavior and other variables such as hourly intervals, days, or types of days (weekdays and weekends). In addition, a relationship between the station's usage ratio (activity) and the prediction error has been found. Therefore, the proposed type of visual representation clearly allows for a more thorough analysis which can assist BSS application designers and decision-makers to select the best model for their specific scenarios.

Among the research limitations, the irruption of the COVID pandemic and a change of operator in Barcelona's BSS were the most restricting factors. As a result, the data used was time-constrained to those months without a strict lockdown, leaving not enough data for more advanced prediction models or the exploitation of seasonality. Besides, Barcelona's Open Data website provides only monthly updates for its BSS, thus the impossibility to perform real-time analysis. Hence, we believe that the real potential of this visualization comparison method can be further explored in a more optimal situation. This work has been partially motivated by the interviews with a government official responsible for monitoring the BSS performance. In the future, we want to evaluate the system together with various BSS stakeholders so as to improve its user-friendliness and the comprehension of the represented information.
ACKNOWLEDGEMENT

Partially supported by project TIN2017-88515-C2-1-R(GEN3DLIVE), from the Spanish Ministerio de Economía y Competitividad, by 839 FEDER (EU) funds.

REFERENCES


69


