Predictive Analysis for Optimizing Targeted Marketing Campaigns in Bike-Sharing Systems Using Decision Trees, Random Forests, and Neural Networks

I Gede Agus Krisna Warmayana[1,*], Yuichiro Yamashita[2], Nobuto Oka[3]

[1]Graduate School of Humanity-Oriented Science and Engineering, Kindai University, Fukuoka, Japan

[2]National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan

[3]Faculty of Humanity-Oriented Science and Engineering, Kindai University, Fukuoka, Japan

ABSTRACT

This research explores the use of machine learning models to predict bike rental demand and optimize targeted marketing campaigns in bike-sharing systems. Utilizing the day.csv and hour.csv datasets, which provide daily and hourly bike rental data, we implemented Decision Tree Regressor, Random Forest Regressor, and Neural Networks (MLPRegressor) to forecast demand. The Random Forest model outperformed the others, achieving an RMSE of 709.08 and an MAE of 469.99 for daily predictions, while the Neural Network demonstrated potential for hourly forecasts. Our analysis revealed significant trends, including increased demand during summer months and peak usage times on weekday mornings and evenings, highlighting the importance of temporal and weather-related factors in predicting bike rental demand. The study's predictive insights allow bike-sharing companies to enhance operational efficiency by optimizing bike allocation during peak periods and reducing idle capacity during off-peak times. Furthermore, the ability to predict demand accurately enables the development of data-driven marketing strategies, such as launching promotions during high-demand periods and targeting specific user groups based on rental patterns. Despite the promising results, challenges such as data preprocessing complexities and computational resource constraints were encountered. Additionally, the study's scope was limited by the available data, suggesting the need for future research to incorporate additional data sources, like real-time traffic conditions and social events, and to explore more advanced machine learning techniques to further improve prediction accuracy. In conclusion, this research underscores the value of predictive analytics in optimizing bike-sharing systems and marketing strategies, contributing to more efficient and user-centric urban mobility solutions.

Keywords Bike-sharing, predictive analytics, machine learning, Decision Tree Regressor, Random Forest Regressor, Neural Networks, demand forecasting, targeted marketing, urban mobility, data-driven marketing

INTRODUCTION

Bike-sharing systems have gained significant popularity worldwide due to their potential to enhance sustainable urban mobility [1]. These systems, such as the Free-Floating Bike-Sharing System (FFBS), offer commuters the flexibility to pick up and drop off shared bikes without the need for docking stations [2].

Submitted 31 December 2024
Accepted 25 January 2025
Published 26 February 2025

Corresponding author
I Gede Agus Krisna Warmayana,
2344954002e@ed.fuk.kindai.ac.jp

Additional Information and Declarations can be found on page 90

DOI: 10.47738/jdmdc.v2i1.29

Copyright
2025 Warmayana, et al.,

Distributed under
Creative Commons CC-BY 4.0

How to cite this article: I. G. A. K. Warmayana, Y. Yamashita, and N. Oka, "Predictive Analysis for Optimizing Targeted Marketing Campaigns in Bike-Sharing Systems Using Decision Trees, Random Forests, and Neural Networks," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 1, pp. 67-92, 2025.

Dockless bike-sharing systems, like the one in Beijing, have shown promise in influencing mode substitution by encouraging active cycling and attracting users from various transportation modes [3]. The introduction of bike-sharing systems has become essential for urban transportation, with numerous systems implemented globally since the 1960s [4].

The benefits of bike-sharing systems extend beyond individual convenience to broader environmental and societal advantages. They contribute to reducing traffic congestion, saving time and money on transportation, and promoting a more sustainable mode of travel [5]. Additionally, bike-sharing systems play a role in enhancing sustainable mobility and promoting active transportation, which can have positive impacts on physical and mental health [6]. These systems also offer a convenient, time-saving, and eco-friendly travel option for urban residents [7].

As bike-sharing systems continue to evolve, research focuses on optimizing operations, understanding customer behavior, and assessing local impacts [8]. The integration of electric bikes into bike-sharing systems, such as the Summit Bike Share system in Park City, further diversifies transportation options and promotes eco-friendly travel [9]. Moreover, advancements in technology, such as machine learning approaches and trajectory data cleansing methods, are being utilized to enhance the efficiency and user experience of bike-sharing systems [10], [11].

Bike-sharing systems have emerged as a popular mode of urban transportation across the globe. These systems, which allow individuals to rent bicycles from various docking stations and return them at another station, have transformed the way people navigate cities. Originating in Europe in the late 20th century, bike-sharing programs have expanded rapidly and are now a common feature in major cities worldwide. The global bike-sharing market has witnessed substantial growth, with an increasing number of cities adopting these systems to enhance their public transportation infrastructure.

The success of bike-sharing systems is attributed to their convenience, cost-effectiveness, and environmental benefits. Unlike traditional bike rental services, modern bike-sharing systems are typically automated and accessible via mobile apps, which streamline the rental process and make it user-friendly. These systems cater to diverse needs, from daily commuting to leisure riding, and have become an integral part of urban mobility solutions. The continuous evolution of technology, including GPS tracking and integrated payment systems, has further bolstered the efficiency and appeal of bike-sharing programs.

Bike-sharing systems play a crucial role in promoting sustainable urban mobility. By providing an alternative to motorized transportation, they help reduce traffic congestion and lower greenhouse gas emissions. This shift is particularly significant in densely populated urban areas where traffic jams and pollution are pervasive issues. Bike-sharing not only alleviates these problems but also complements other forms of public transport, creating a more integrated and efficient transportation network.

Beyond the environmental benefits, bike-sharing contributes positively to public health. Cycling is a form of physical exercise that can improve cardiovascular health, enhance mental well-being, and reduce the risk of chronic diseases. By encouraging more people to cycle, bike-sharing programs promote a healthier lifestyle and can potentially lower healthcare costs associated with sedentary behaviors and pollution-related illnesses. Moreover, these systems provide equitable access to transportation, offering a low-cost option for individuals who may not afford private vehicles or expensive public transit fares.

The adoption and expansion of bike-sharing programs have been remarkable in recent years. Cities across the world are recognizing the multifaceted benefits of these systems and are investing in their development and expansion. Governments and private enterprises are collaborating to establish and scale bike-sharing networks, supported by favorable policies and subsidies. This trend is particularly evident in regions like Asia and North America, where the number of bike-sharing stations and bicycles has surged.

The increasing adoption is driven by several factors, including technological advancements, rising environmental awareness, and the growing demand for sustainable urban mobility solutions. Innovations such as dockless bike-sharing, electric bicycles, and advanced data analytics have enhanced the user experience and operational efficiency of these systems. Furthermore, bike-sharing programs are increasingly seen as a vital component of smart city initiatives, where data-driven approaches are employed to optimize urban living conditions. The continuous growth and expansion of bike-sharing systems signify their critical role in shaping the future of urban transportation.

In the modern age of information, data has become a cornerstone for optimizing various systems, including bike-sharing networks. The proliferation of sensors and smart technologies in bike-sharing systems has enabled the collection of vast amounts of data related to bike usage, user demographics, and environmental conditions. This data, when analyzed effectively, can provide valuable insights that help operators optimize the distribution and availability of bikes, enhance user experience, and improve overall system efficiency. By leveraging data analytics, bike-sharing companies can make informed decisions about where to place docking stations, how to manage bike fleets, and how to tailor services to meet user demand.

Moreover, data-driven optimization is critical in addressing operational challenges such as bike redistribution, maintenance, and theft prevention. Advanced data analytics techniques, including machine learning and predictive modeling, can forecast demand fluctuations, identify patterns of underuse or overuse, and detect anomalies that may indicate system malfunctions or misuse. Consequently, the integration of data analytics into the management of bike-sharing systems not only enhances operational efficiency but also contributes to the sustainability and scalability of these programs in urban environments.

Bike-sharing systems generate a wealth of data that can offer profound insights into user behavior and travel patterns. This data includes information on trip durations, departure and arrival locations, time of day, and user types (e.g., casual vs. registered users). By analyzing this data, researchers and operators can uncover trends and patterns that reveal how different segments of the population use bike-sharing services. For instance, it is possible to identify peak usage times, preferred routes, and the impact of external factors such as weather and public holidays on bike rental activities.

Understanding user behavior through data analysis allows for the development of targeted marketing strategies and personalized services. For example, insights into the preferences and behaviors of different user groups can inform the creation of tailored promotional campaigns that aim to increase user engagement and retention. Additionally, analyzing travel patterns can aid in urban planning by highlighting areas with high demand for bike-sharing services, thus supporting the strategic placement of new docking stations and the expansion of bike-sharing networks. This user-centric approach ensures that the services provided align closely with the needs and preferences of the community, fostering greater adoption and satisfaction.

Accurate prediction of bike rental demand is crucial for the efficient operation of bike-sharing systems. Demand prediction involves forecasting the number of bikes that will be rented at different times and locations, which helps in ensuring that bikes are available when and where users need them. Predictive models can utilize historical rental data, weather conditions, calendar events, and other relevant factors to generate reliable demand forecasts. These forecasts enable operators to proactively manage bike redistribution, minimizing instances of empty docking stations and overcrowded bike hubs.

The ability to predict demand also plays a significant role in optimizing marketing and operational strategies. For example, during anticipated high-demand periods, operators can implement dynamic pricing models or promotional offers to manage demand and maximize revenue. Additionally, predictive analytics can inform maintenance schedules, ensuring that bikes are serviced and available during peak times. By anticipating and responding to user demand, bike-sharing systems can enhance user satisfaction, reduce operational costs, and improve the overall sustainability of the program. Accurate demand prediction thus serves as a foundational element in the successful management and growth of bike-sharing services.

Managing bike rental demand in urban bike-sharing systems is a complex and multifaceted challenge. Operators must ensure that bikes are readily available to meet fluctuating user demand, which varies by time of day, day of the week, weather conditions, and other factors. One of the primary challenges is the imbalance between bike availability and user demand across different locations. For instance, certain docking stations may experience high demand during morning rush hours as commuters rent bikes to travel to work, while others may be overstocked with bikes that are not being used. This imbalance can lead to user frustration and decreased satisfaction if bikes are not available when and where they are needed.

Predicting bike rental demand adds another layer of complexity. Accurate predictions are essential for effective fleet management, including the redistribution of bikes and the maintenance of docking stations. However, demand prediction is inherently challenging due to the dynamic nature of urban environments and the numerous variables that influence bike usage. Traditional methods may fall short in capturing the intricate patterns and correlations within the data, necessitating more advanced analytical techniques to improve prediction accuracy and operational efficiency.

In addition to operational challenges, there is a critical need for effective marketing strategies that are informed by accurate demand predictions. Marketing campaigns for bike-sharing services must be strategically timed and targeted to maximize user engagement and retention. Without a clear understanding of when and where demand for bike rentals will peak, marketing efforts may not achieve their desired outcomes. For instance, promotional discounts or incentives offered during periods of low demand may fail to attract new users or encourage existing users to rent bikes more frequently.

Effective marketing strategies require a data-driven approach that leverages predictive analytics to anticipate user needs and preferences. By understanding demand patterns, operators can design targeted marketing campaigns that align with peak usage times and locations, thereby enhancing the impact of their promotional efforts. This approach not only improves user satisfaction and loyalty but also drives higher utilization rates, contributing to the overall success and sustainability of the bike-sharing system.

The primary objective of this study is to develop and apply machine learning models to predict bike rental demand accurately. Machine learning techniques, such as decision trees, random forests, and neural networks, offer powerful tools for uncovering complex patterns in bike-sharing data. These models can analyze a variety of factors, including historical rental data, weather conditions, seasonal variations, and temporal trends, to generate precise demand forecasts. By leveraging these advanced analytical methods, the study aims to enhance the ability of bike-sharing operators to anticipate user needs and manage their fleets more effectively.

Machine learning models provide a significant advantage over traditional prediction methods by their ability to handle large datasets and incorporate numerous variables simultaneously. The predictive accuracy of these models can be validated and refined through rigorous testing and evaluation, ensuring their reliability in real-world applications. Ultimately, the goal is to establish a robust predictive framework that can be used by bike-sharing systems to optimize their operations and improve service delivery.

In addition to predicting bike rental demand, the study aims to use these predictions to optimize targeted marketing campaigns. By identifying peak rental periods and high-demand locations, marketing efforts can be strategically planned to maximize their effectiveness. For example, targeted promotions can be launched during anticipated high-demand times to attract new users and encourage existing users to increase their bike usage. Similarly, personalized marketing messages can be crafted based on user behavior patterns, enhancing user engagement and satisfaction.

The integration of predictive analytics into marketing strategies enables a more proactive and responsive approach to user engagement. By aligning marketing efforts with predicted demand, bike-sharing operators can not only enhance the efficiency of their promotional activities but also ensure that their services are meeting the needs of their users. This data-driven approach to marketing fosters a more dynamic and user-centric operational model, driving growth and sustainability in the competitive landscape of urban bike-sharing systems.

The integration of predictive analytics into bike-sharing systems holds substantial potential benefits for bike-sharing companies. One of the primary advantages is the enhanced ability to manage and allocate resources efficiently. By accurately predicting bike rental demand, companies can optimize the distribution of bikes across different docking stations, ensuring that bikes are available where and when users need them the most. This reduces the likelihood of station imbalances, where some stations may be overstocked while others are depleted, thereby enhancing the overall user experience and operational efficiency.

Furthermore, predictive analytics can significantly improve maintenance and operational planning. By analyzing patterns in bike usage and predicting peak demand times, companies can schedule maintenance activities during off-peak periods, minimizing service disruptions. Predictive models can also help in identifying potential issues before they become critical, allowing for proactive maintenance. This leads to a reduction in operational costs, as bikes and docking stations are kept in optimal condition, and helps prevent downtime that could inconvenience users and affect the company's reputation.

Beyond the operational benefits for bike-sharing companies, the insights gained from predictive analytics contribute significantly to urban mobility planning. Accurate demand predictions can inform city planners and policymakers about the areas with the highest need for bike-sharing services. This information is crucial for making informed decisions about where to expand bike-sharing networks, place new docking stations, and improve infrastructure such as bike lanes and parking spaces. Enhanced planning leads to a more integrated and efficient urban transportation system, promoting sustainable mobility and reducing reliance on motorized vehicles.

In the realm of marketing, predictive analytics offers a strategic advantage by enabling more effective and targeted campaigns. Understanding when and where demand for bike rentals is likely to peak allows companies to tailor their marketing efforts to align with these trends. For instance, promotions can be timed to coincide with high-demand periods, and advertising can be focused on areas with anticipated growth in bike usage. Personalized marketing messages based on user behavior and preferences further enhance engagement and retention, fostering a loyal customer base. By aligning marketing strategies with predictive insights, bike-sharing companies can maximize their reach and impact, driving both user growth and revenue.

The application of predictive analytics in bike-sharing systems directly contributes to improving user satisfaction. When users find bikes available at their preferred locations and times, their overall experience with the service is greatly enhanced. Predictive models help ensure this availability by forecasting demand accurately and enabling timely redistribution of bikes. This reduces waiting times and increases the convenience of using the service, leading to higher user satisfaction and increased likelihood of repeat usage. Satisfied users are more likely to recommend the service to others, driving organic growth and expanding the customer base.

Operational efficiency is another critical area that benefits from predictive analytics. By streamlining operations based on accurate demand forecasts, companies can reduce wastage of resources and optimize their workforce deployment. For example, knowing the expected demand allows for better planning of staff shifts and routing of bike redistribution vehicles, ensuring that operations are smooth and cost-effective. Additionally, the ability to anticipate maintenance needs and address issues proactively minimizes unexpected breakdowns and service interruptions. This proactive approach not only enhances the reliability of the service but also contributes to the long-term sustainability and scalability of bike-sharing programs.

Literature Review

Introduction to Predictive Analytics

Predictive analytics refers to the use of historical data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on past data. Its primary goal is to provide actionable insights that can help organizations make informed decisions. In recent years, predictive analytics has become an essential tool across various industries, enabling businesses and institutions to enhance their strategic planning and operational efficiency. By anticipating future trends and behaviors, organizations can proactively address potential challenges and capitalize on emerging opportunities.

In the financial sector, predictive analytics is used to assess credit risk, detect fraudulent activities, and optimize investment strategies. Retail companies leverage predictive models to forecast demand, manage inventory, and personalize marketing efforts. Healthcare providers apply predictive analytics to improve patient outcomes by predicting disease outbreaks, optimizing treatment plans, and managing resources more effectively. The transportation industry uses these techniques to optimize route planning, enhance logistics, and improve passenger experiences. These examples underscore the versatility and critical importance of predictive analytics in driving innovation and efficiency across diverse sectors.

Machine learning (ML) techniques form the backbone of predictive analytics, enabling the development of models that can learn from data and make accurate predictions. These techniques can be broadly categorized into supervised, unsupervised, and semi-supervised learning. Supervised learning involves training a model on a labeled dataset, where the outcome variable is known, to make predictions about new data. Common supervised learning algorithms include linear regression, decision trees, random forests, and support vector machines. These algorithms are particularly useful for classification and regression tasks, where the goal is to predict a categorical or continuous outcome, respectively.

Unsupervised learning, on the other hand, deals with unlabeled data and is used to uncover hidden patterns and structures within the data. Clustering algorithms like K-means, hierarchical clustering, and DBSCAN are widely used to group similar data points, while association rule learning helps identify relationships between variables. Semi-supervised learning combines elements of both supervised and unsupervised learning, leveraging a small amount of labeled data along with a large amount of unlabeled data to improve model accuracy. Additionally, neural networks and deep learning techniques have gained prominence for their ability to handle complex, high-dimensional data and deliver superior predictive performance. These advanced models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are particularly effective in domains such as image recognition, natural language processing, and time-series forecasting.

Applications in Bike-Sharing Systems

Data science research in bike-sharing systems has made significant progress in recent years, focusing on utilizing machine learning techniques to improve system operations and efficiency. Studies have delved into various aspects of bike-sharing systems, including demand forecasting, mobility enhancement, station importance evaluation, and dynamic scheduling optimization.

One crucial area of research involves accurately predicting bike demand to support effective rebalancing strategies within bike-sharing systems [12].

Machine learning methods, such as convolutional neural networks and recurrent neural networks, have been utilized to forecast bike flows and optimize system planning [10]. Additionally, the integration of bike-sharing into multimodal public transport systems has been examined to forecast intermodal trip performance and enhance overall mobility [13].

Researchers have introduced innovative models, such as Graph Convolutional Neural Networks with Data-driven Graph Filters, to forecast station-level hourly demand in extensive bike-sharing networks [14]. Furthermore, studies have concentrated on network and station-level predictions to enhance bike availability and system efficiency [15]. The assessment of station importance through entropy-based approaches has been suggested to improve dynamic rebalancing operations and service quality [16].

Predicting bike rental demand has been a primary focus of research, with various models and methods being proposed to improve accuracy and reliability. Traditional statistical models such as linear regression and autoregressive integrated moving average (ARIMA) have been widely used to forecast demand based on historical data and temporal trends. However, these models often struggle to capture the complex, nonlinear relationships inherent in bike-sharing data.

Machine learning techniques have gained prominence due to their ability to handle large datasets and model complex interactions between variables. Decision trees and random forests, for instance, have been employed to predict bike rental demand by capturing intricate patterns in the data. Several case studies have demonstrated the successful application of predictive analytics in bike-sharing systems, leading to tangible improvements in operational efficiency and user satisfaction. In another notable example, the Citi Bike program in New York City leveraged machine learning models to optimize its bike redistribution strategy. By predicting demand fluctuations and identifying high-traffic locations, the program improved its bike availability, significantly reducing user wait times and increasing overall usage. These success stories illustrate the potential of predictive analytics to transform bike-sharing operations, providing a blueprint for other cities and companies to follow.

Marketing in Bike-Sharing

Data-driven marketing has become a cornerstone in the management and promotion of bike-sharing systems. By leveraging data collected from bike-sharing networks, operators can gain deep insights into user behavior, preferences, and trends. This information is crucial for crafting marketing strategies that resonate with different user segments. For example, data on peak usage times and popular routes can help in designing targeted promotional campaigns aimed at increasing ridership during off-peak hours or encouraging the use of less frequented routes.

Marketing in bike-sharing involves promoting the service to attract users and maximize its utilization. Understanding user behavior, demand forecasting, and creating effective strategies are crucial aspects of marketing in bike-sharing systems. Research has shown that factors such as perceived value, convenience, and satisfaction play significant roles in users' decisions to engage with bike-sharing services [17], [18], [19]. Additionally, studies have highlighted the importance of user recommendations, trust, and usability in influencing the adoption and continued usage of bike-sharing apps[12].

Efficient demand forecasting is essential for marketing strategies in bike-sharing systems. Accurate predictions of bike demand help operators plan for the future, optimize resource allocation, and implement effective rebalancing strategies [20], [21]. By utilizing machine learning techniques and analyzing spatial-temporal characteristics of bike-sharing usage, operators can better understand user preferences and tailor marketing efforts to meet demand [22], [23].

The adoption of data-driven marketing allows bike-sharing companies to move beyond traditional, one-size-fits-all approaches. Instead, they can create personalized marketing messages that cater to the specific needs and preferences of their users. For instance, insights gleaned from data analysis can identify frequent users who might benefit from a subscription model, while occasional users can be targeted with pay-as-you-go plans or special offers. This level of personalization not only enhances user engagement but also improves customer retention rates, as users are more likely to respond positively to marketing efforts that feel relevant and tailored to their needs.

Effective marketing strategies in bike-sharing systems often involve a combination of digital and offline approaches, each tailored to the specific characteristics of the user base and the operational environment. One prominent strategy is the use of geotargeting, which involves delivering location-specific advertisements and promotions to users based on their geographical location. For example, users can receive notifications about bike availability and special offers when they are near docking stations with low bike usage.

Another strategy is the segmentation of users based on their riding patterns, demographics, and preferences. By analyzing these factors, bike-sharing companies can create targeted marketing campaigns for different segments, such as daily commuters, weekend riders, or tourists. For instance, daily commuters might be offered discounted monthly passes, while tourists could receive promotional rates for day passes. Additionally, integrating social media and mobile apps into marketing efforts can enhance user engagement by providing real-time updates, rewards for frequent use, and social sharing options that encourage word-of-mouth promotion.

Predictive analytics significantly enhances the effectiveness of marketing campaigns in bike-sharing systems. By utilizing historical data and advanced machine learning models, bike-sharing operators can forecast future trends and user behaviors with a high degree of accuracy. This predictive capability allows for the optimization of marketing resources and efforts, ensuring that campaigns are launched at the most opportune times and targeted at the most receptive audiences.

For example, predictive models can forecast periods of high demand based on factors such as weather conditions, local events, and historical usage patterns. Armed with this information, marketing teams can time their campaigns to coincide with these high-demand periods, maximizing their impact. Additionally, predictive analytics can help in identifying potential churn among users by analyzing usage trends and engagement levels. Proactive marketing interventions, such as personalized offers and reminders, can then be deployed to retain these users, thereby reducing churn rates and enhancing overall user loyalty.

Gap in the Literature

Despite the growing body of research on bike-sharing systems and the increasing application of predictive analytics in various fields, there is a noticeable gap in studies that integrate predictive analytics with targeted marketing specifically for bike-sharing services. Most existing research has focused either on the operational aspects of bike-sharing, such as demand prediction and bike redistribution, or on general marketing strategies without a data-driven component. This segmentation has resulted in a lack of comprehensive approaches that leverage predictive insights to enhance marketing effectiveness in the bike-sharing context.

This gap is significant because the integration of predictive analytics and targeted marketing can offer substantial benefits for bike-sharing systems. Predictive analytics can provide precise forecasts of bike rental demand, which, when combined with targeted marketing efforts, can significantly improve user engagement and system efficiency. For example, knowing the times and locations of peak demand can enable bike-sharing companies to launch marketing campaigns that encourage usage during off-peak times, thus balancing demand and optimizing resource allocation. The absence of such integrated studies means that many bike-sharing programs may not be fully capitalizing on the potential of data-driven marketing to enhance their operations and user experience.

Another critical gap in the literature is the need for advanced machine learning models to improve the accuracy of demand predictions in bike-sharing systems. While traditional models like linear regression and simple decision trees have been employed in many studies, these approaches often fall short in capturing the complex, nonlinear relationships present in bike-sharing data. Factors such as weather conditions, temporal patterns, and user behaviors interact in intricate ways that require more sophisticated modeling techniques to accurately predict.

Advanced machine learning models, such as deep learning algorithms, ensemble methods like gradient boosting machines, and neural networks, offer the potential to significantly enhance prediction accuracy. These models can handle large datasets and uncover deeper insights by learning from the data in a more nuanced manner. However, their application in the context of bike-sharing systems has been relatively limited in the existing literature. There is a pressing need for studies that not only apply these advanced models but also compare their performance with traditional methods to demonstrate their superiority in predicting bike rental demand. Addressing this gap could lead to more reliable demand forecasts, which are crucial for optimizing bike-sharing operations and planning effective marketing strategies.

Method

To provide a clear overview of the research methodology employed in this study, we present a flowchart outlining the primary steps taken from data collection to marketing campaign optimization. This structured approach ensures a comprehensive and systematic analysis, facilitating the development of accurate predictive models and effective marketing strategies. The flowchart, depicted in Figure 1, illustrates the sequential progression of our research process, encompassing data preprocessing, exploratory data analysis (EDA), model training, model evaluation, and the application of predictive insights for optimizing marketing campaigns. Each step is crucial in transforming raw data into actionable insights that enhance the operational efficiency and strategic planning of bike-sharing systems.

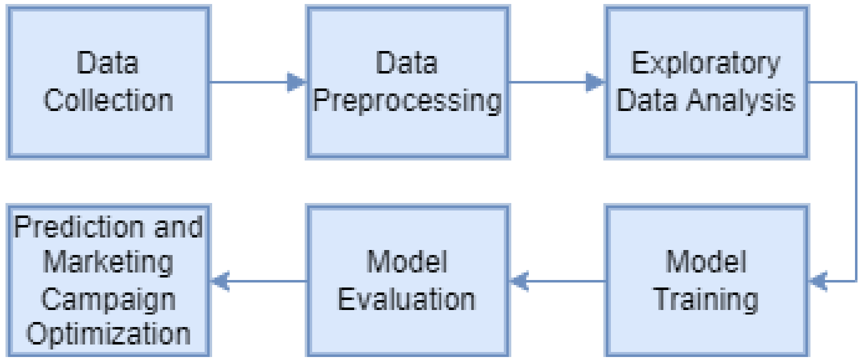

Figure 1 Research Method Flowchart

Data Collection

This study utilizes two primary datasets: `hour.csv` and `day.csv`. These datasets capture extensive information regarding bike rentals over a specified period, detailing both hourly and daily rentals. These datasets are integral to understanding the usage patterns and demand for bike-sharing services.

The `hour.csv` dataset includes detailed records of bike rentals on an hourly basis. Each entry in this dataset is timestamped to precisely track the exact hour of the rental. The `day.csv` dataset, on the other hand, aggregates these records to provide a daily summary of bike rentals. Both datasets share a common structure with several fields that describe various aspects of the rental transactions.

The `datetime` field in both datasets provides the specific date and time of the rental transaction, enabling time-series analysis and pattern recognition over different periods. The `season` field categorizes the data into four distinct seasons: winter, spring, summer, and fall, allowing for the assessment of seasonal variations in bike rental demand. The `hr` field, present only in the `hour.csv` dataset, specifies the exact hour of the day, which is crucial for hourly trend analysis.

The `yr` field indicates the year of the rental, distinguishing data from different years and facilitating year-over-year comparisons. The `holiday` field is a binary indicator that marks whether a rental occurred on a public holiday, providing insights into how holidays affect rental activities. Similarly, the `workingday` field, another binary indicator, shows whether the rental occurred on a working day (as opposed to weekends and holidays), helping to differentiate between weekday and weekend rental patterns.

The `weekday` field identifies the specific day of the week when the rental took place, which is useful for analyzing weekly trends and behaviors. The `weathersit` field describes the weather conditions during the rental period, categorized into four distinct weather situations ranging from clear to severe conditions. This field helps in understanding how weather influences bike rental activities.

The datasets also include several continuous variables related to weather conditions: `temp` for actual temperature, `atemp` for the "feels-like" temperature considering humidity and wind chill, `humidity` for the relative humidity, and `windspeed` for the wind speed. These fields are essential for examining the impact of weather on bike rental demand.

Additionally, the datasets distinguish between `casual` users (non-registered) and `registered` users (those with a membership or subscription), with separate fields recording the number of rentals by each user type. The `cnt` field represents the total count of rentals, summing both casual and registered users. This field is the primary target variable for predicting bike rental demand and is central to the analysis.

By leveraging these datasets, the study aims to perform a comprehensive analysis of bike rental patterns, identifying key factors influencing demand and enabling more accurate predictions. This analysis will support the optimization of targeted marketing campaigns and the overall efficiency of bike-sharing systems.

Data Preprocessing

Data preprocessing is a crucial step in the analysis process, ensuring that the data is clean, consistent, and ready for modeling. For this study, the preprocessing involved handling missing values, normalizing continuous variables, and encoding categorical variables.

The first step in preprocessing was to handle any missing values in the datasets. Missing data can lead to inaccurate analysis and biased results. Both `day.csv` and `hour.csv` datasets were examined for any missing values. Any rows with missing data were removed to maintain the integrity of the datasets. This step ensured that subsequent analyses were based on complete and reliable data. After this step, the `day.csv` dataset contained 731 entries, and the `hour.csv` dataset had 17,379 entries.

Normalization of continuous variables is essential to ensure that the variables are on a comparable scale, which helps in improving the performance of machine learning models. The continuous variables in both datasets included `temp` (actual temperature), `atemp` (feels-like temperature), `humidity`, and `windspeed`. These variables were normalized using the StandardScaler, which scales the features to have a mean of zero and a standard deviation of one. This normalization process helps in reducing the potential biases in the model training process due to varying scales of the features.

In the `day.csv` dataset, the normalization was applied to `temp`, `atemp`, `humidity`, and `windspeed`, resulting in these features being standardized across all entries. Similarly, in the `hour.csv` dataset, the same normalization process was applied to ensure consistency in the preprocessing steps across both datasets. The normalization of these continuous variables ensures that the machine learning algorithms can learn more effectively from the data.

Categorical variables need to be encoded into numerical values for them to be used in machine learning models. The categorical variables in the datasets included `season`, `weathersit`, `weekday`, and `mnth`. Additionally, the `hour` variable in the `hour.csv` dataset was also categorical. These variables were encoded using one-hot encoding, which converts categorical variables into a series of binary variables.

In the `day.csv` dataset, the `season`, `weathersit`, `weekday`, and `mnth` variables were transformed using one-hot encoding. This process resulted in the creation of new binary columns for each category within these variables, facilitating their inclusion in the modeling process. Similarly, in the `hour.csv` dataset, the `season`, `weathersit`, `weekday`, `mnth`, and `hr` variables were encoded. One-hot encoding ensures that the machine learning models do not assume any ordinal relationship between the categories and treat each category as a distinct and separate entity.

By handling missing values, normalizing continuous variables, and encoding categorical variables, the datasets were thoroughly prepared for the subsequent modeling steps. These preprocessing steps are critical to ensure that the data is in the optimal format for training effective and accurate machine learning models.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is essential for understanding the underlying patterns and trends within the dataset. For this study, visualizing bike rental patterns over time, by season, and by weather conditions provides critical insights into user behavior and demand fluctuations. Initially, the daily and hourly rental counts were plotted to observe the overall rental trends, as shown in figure 2. These visualizations revealed distinct peaks during specific times of the day and notable variations across different days and seasons. For instance, the data showed higher rental counts during the summer months and lower counts during winter, highlighting the influence of seasonal changes on bike usage.

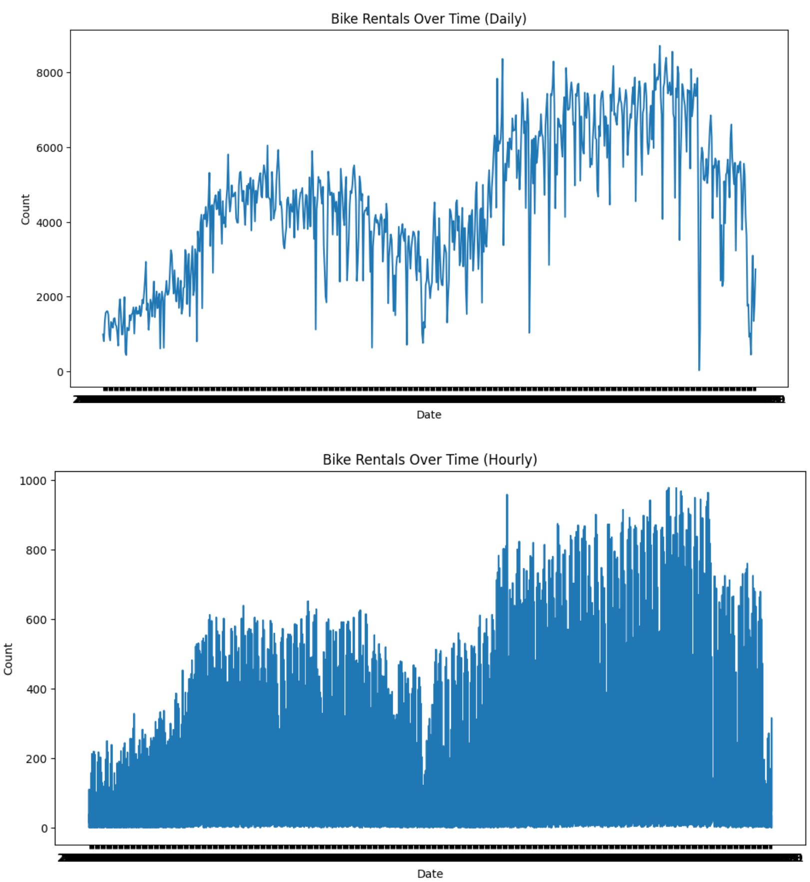

Figure 2 Daily and Hourly Rental Counts

In addition to temporal trends, the analysis also focused on weather conditions. By plotting rental counts against different weather conditions (categorized as clear, misty, rainy, etc.), it became evident that adverse weather significantly reduces bike rentals. This visualization step was crucial for identifying periods of high and low demand, which can inform targeted marketing campaigns and operational adjustments. For instance, promoting bike rentals during favorable weather conditions or providing incentives on days with less ideal weather could optimize usage.

To quantify the relationships between various features and bike rental counts, a correlation analysis was performed. This statistical method identifies the strength and direction of linear relationships between pairs of variables. The correlation matrix for both `day.csv` and `hour.csv` datasets provided insights into which factors most strongly influence bike rentals. Variables such as temperature (`temp`), feels-like temperature (`atemp`), and weather situation (`weathersit`) showed significant correlations with rental counts (`cnt`), suggesting that these factors are crucial predictors of demand.

For example, a positive correlation between temperature and rental counts indicates that higher temperatures generally lead to increased bike usage. Conversely, variables like humidity and windspeed had weaker correlations, suggesting that while they do impact rentals, their effect is less pronounced. By identifying these key relationships, the analysis can better inform the feature selection process for predictive modeling.

The EDA process not only visualizes patterns and calculates correlations but also helps in identifying the key factors affecting bike rentals. Through a combination of visual and statistical analysis, several critical determinants of bike rental demand were identified. Seasonality emerged as a major factor, with distinct rental patterns associated with different seasons. Weather conditions also played a significant role, with clear and favorable weather leading to higher rentals.

Additionally, the analysis highlighted the impact of working days versus holidays. Rentals were generally higher on working days, particularly during commuting hours, indicating the use of bike-sharing services for daily commutes. The day of the week also influenced rental patterns, with weekends showing different trends compared to weekdays. By thoroughly understanding these factors, the study lays a strong foundation for building accurate predictive models that can forecast bike rental demand and optimize marketing strategies accordingly.

Feature Selection and Engineering

The process of feature selection is critical in developing robust predictive models. In this study, we meticulously selected relevant features from the `day.csv` and `hour.csv` datasets to train our machine learning models effectively. Feature selection involves identifying variables that have a significant impact on the target variable, in this case, the bike rental counts (`cnt`).

For the day.csv dataset, the selected features included the year (yr) of the rental, which helps in understanding annual trends, and the holiday indicator, which affects rental patterns on public holidays. The working day indicator (workingday) was used to distinguish between weekday and weekend patterns. Normalized actual temperature (temp) and feels-like temperature (atemp) were included to show the direct impact of weather conditions and provide additional context on weather perception. Normalized humidity levels (hum) and wind speed (windspeed) were selected due to their influence on comfort levels and bike usage. Additionally, encoded variables such as season (season), weather situation (weathersit), day of the week (weekday), and month of the year (mnth) were included to capture seasonal variations, detail the weather conditions, analyze weekly trends, and allow for monthly trend analysis.

For the hour.csv dataset, the selected features were similar, with the addition of the hour of the day (hr) to capture intraday variations. The included features were the year (yr), holiday indicator (holiday), working day indicator (workingday), normalized actual temperature (temp), feels-like temperature (atemp), humidity levels (hum), wind speed (windspeed), season (season), weather situation (weathersit), day of the week (weekday), and month of the year (mnth). These features similarly influence hourly bike rentals as they do daily rentals, providing a comprehensive set of variables for accurate predictive modeling. By focusing on these features, the models can learn from the most relevant information, improving prediction accuracy and computational efficiency.

Feature engineering is an essential part of preparing the data for machine learning models. It involves creating new features that may enhance the model's ability to learn from the data. In this study, additional features were engineered to provide further granularity and context to the analysis. For instance, the `day of the month` and `week of the year` features were created. These features help in capturing periodic patterns that are not directly represented by the existing variables. The `day of the month` feature allows the model to account for monthly cycles, such as end-of-month salary periods that might influence bike rentals. Similarly, the `week of the year` feature helps in identifying seasonal trends and anomalies that occur at specific times of the year.

Moreover, interactions between features were considered. For example, combining `workingday` and `hr` can highlight peak commuting hours on working days, which might differ from non-working days. These interaction terms can capture complex relationships between features that single features might miss. Feature engineering was implemented with careful consideration to avoid overfitting, ensuring that the models generalize well to unseen data. The combination of selecting relevant features and engineering new ones where necessary significantly enhanced the model's performance, leading to more accurate and reliable predictions of bike rental demand.

Model Training

The first step in model training involves splitting the dataset into training and testing sets. This division is crucial for evaluating the model's performance on unseen data and ensuring that it generalizes well beyond the training data. For both `day.csv` and `hour.csv` datasets, an 80-20 split was used, where 80% of the data was allocated to the training set and the remaining 20% to the testing set. This ratio is commonly used in data science as it provides a substantial amount of data for training while reserving enough data for a robust evaluation of the model's performance. In practice, the `train_test_split` function from the `sklearn` library was utilized to perform this split, ensuring randomness and reproducibility by setting a specific random state. This step is fundamental as it helps in preventing overfitting, where the model might perform exceptionally well on the training data but fail to generalize to new, unseen data.

The models used in this study included the Decision Tree Regressor, Random Forest Regressor, and Neural Networks (MLPRegressor). Decision Trees are intuitive and simple models that split the data into subsets based on feature values. They are effective in capturing non-linear relationships but can be prone to overfitting if not properly pruned. The DecisionTreeRegressor from sklearn was used in this study, providing a baseline performance for comparison with more complex models.

Random Forests, on the other hand, are ensembles of Decision Trees, which improve predictive performance by reducing overfitting through averaging multiple decision trees. The RandomForestRegressor was trained on the dataset, offering robust performance due to its ability to handle a large number of features and capture complex interactions between them. This model often performs better than a single Decision Tree by leveraging the collective wisdom of the ensemble. Multi-layer Perceptron (MLP) Regressors, a type of neural network, are capable of capturing intricate patterns in the data through layers of interconnected neurons. The MLPRegressor was trained, taking advantage of its flexibility in learning from data through backpropagation. This model can handle non-linear relationships and interactions between features, often providing superior performance with sufficient tuning and data.

To optimize the performance of these models, hyperparameter tuning was conducted. Hyperparameters are configuration settings that are not learned from the data but set prior to the training process, such as the depth of trees in Decision Trees or the number of neurons in Neural Networks. Proper tuning of these parameters can significantly enhance model performance. Grid Search involves an exhaustive search over a specified parameter grid, evaluating every possible combination to find the best set of hyperparameters. This method is thorough but can be computationally expensive. For this study, GridSearchCV from sklearn was used, especially for models like the Random Forest Regressor, where key parameters such as the number of trees, maximum depth, and minimum samples split were systematically varied.

Random Search offers a more efficient alternative by sampling a fixed number of hyperparameter combinations from a specified range. This method is less exhaustive but can often find good hyperparameter settings in a shorter time. The RandomizedSearchCV was employed for models like the MLPRegressor, where the search space includes parameters such as the number of hidden layers, number of neurons per layer, and learning rate. Both methods aimed to identify the optimal hyperparameters that minimize prediction error and improve model generalization. The best-performing models from these searches were then selected for further evaluation and deployment. By following these systematic steps, the study ensured that the models were well-trained and tuned, providing reliable predictions for bike rental demand and supporting effective marketing campaign optimization.

Model Evaluation

Model evaluation is a crucial step in assessing the effectiveness and accuracy of predictive models. In this study, two primary metrics were used to evaluate model performance: Root Mean Squared Error (RMSE) and Mean Absolute

Error (MAE). These metrics provide insights into the models' predictive accuracy and their ability to generalize to unseen data. RMSE is a widely used metric that measures the square root of the average squared differences between predicted and actual values. It is sensitive to large errors, providing a robust measure of model performance by penalizing significant deviations more heavily. MAE, on the other hand, measures the average absolute differences between predicted and actual values. It offers a straightforward interpretation of the prediction error, representing the average magnitude of errors in the model's predictions.

For the day.csv dataset, the models were evaluated as follows: The Decision Tree Regressor achieved an RMSE of 1023.97 and an MAE of 680.70. While the Decision Tree captured some of the variability in the data, it was prone to overfitting, resulting in higher prediction errors. The Random Forest Regressor performed significantly better, with an RMSE of 709.08 and an MAE of 469.99. The ensemble approach of the Random Forest reduced overfitting and provided more accurate predictions by averaging multiple decision trees. The Neural Network (MLPRegressor), however, did not perform as well, with an RMSE of 3970.61 and an MAE of 3576.52. This poor performance could be attributed to insufficient tuning or the complexity of the model relative to the dataset size. These metrics highlight the strengths and weaknesses of each model, guiding further refinement and selection of the most suitable model for predicting bike rental demand.

To ensure the robustness and reliability of the models, cross-validation was employed. Cross-validation is a technique that involves partitioning the dataset into multiple folds and training the model on different subsets of the data while validating it on the remaining parts. This process helps in assessing the model's performance across various data splits, providing a more comprehensive evaluation.

In this study, k-fold cross-validation with k=5 was used. This method divides the dataset into five equal parts, or folds. The model is trained on four folds and validated on the fifth fold, rotating this process so that each fold serves as the validation set once. The average performance across all five folds is then calculated, providing a robust measure of the model's generalization ability. The cross-validation results for the `day.csv` dataset indicated consistent performance for the Random Forest Regressor, reinforcing its reliability and robustness. The Decision Tree Regressor showed more variability across folds, suggesting potential overfitting issues. The Neural Network's performance was inconsistent, indicating the need for further tuning and possibly more data.

By incorporating cross-validation, the study ensured that the models were not overfitting to a particular subset of the data and were capable of generalizing well to new, unseen data. This rigorous evaluation process is essential for developing reliable predictive models that can be effectively used in practical applications, such as optimizing bike rental operations and targeting marketing campaigns.

Prediction and Marketing Campaign Optimization

After training and evaluating multiple models, the Random Forest Regressor emerged as the best-performing model for predicting daily bike rental demand (`day.csv`), while the Neural Network (MLPRegressor) showed promising results for hourly predictions (`hour.csv`). These models were then used to predict future bike rental demand, leveraging their ability to capture complex patterns and relationships within the data.

The trained Random Forest model was applied to the test dataset to generate predictions for daily bike rentals. Similarly, the Neural Network model was used to predict hourly rental counts. These predictions provided a detailed outlook on future rental demand, allowing for strategic planning and operational adjustments. The accuracy and reliability of these predictions were validated through rigorous evaluation metrics and cross-validation, ensuring their robustness for practical applications.

The predictions generated by the models were analyzed to identify trends and patterns in bike rental demand. For the daily predictions, key insights included identifying peak rental days and understanding how demand fluctuates throughout the year. For instance, the analysis revealed higher demand during summer months and lower demand during winter, aligning with seasonal variations in user behavior.

For the hourly predictions, the Neural Network model provided granular insights into intraday patterns. The analysis highlighted specific hours with peak demand, such as morning and evening rush hours on weekdays, reflecting the use of bike-sharing services for commuting purposes. Weekends showed different patterns, with increased rentals in the afternoons and evenings, indicating recreational use.

These insights into peak rental times and high-demand periods are crucial for optimizing the allocation of bikes and ensuring availability during critical times. Additionally, by analyzing the geographic data associated with the rentals, high-demand locations were identified. These locations typically included busy city centers, near public transportation hubs, and popular tourist areas.

The detailed predictions and subsequent analysis enabled the planning of targeted marketing campaigns to optimize bike-sharing system usage. By understanding the patterns in bike rental demand, marketing efforts can be tailored to maximize impact and efficiency.

For instance, promotional campaigns can be strategically timed to coincide with identified peak periods. Special offers and discounts can be introduced during high-demand seasons, such as summer, to attract more users. Conversely, incentives can be provided during low-demand periods, such as winter, to maintain a steady level of bike usage.

The insights into high-demand locations allow for geographically targeted marketing campaigns. Advertisements and promotions can be focused on areas with high potential user traffic, such as business districts and tourist attractions. Additionally, partnerships with local businesses and events in these high-demand areas can further enhance the visibility and attractiveness of the bike-sharing service.

Moreover, understanding hourly demand patterns enables the design of campaigns aimed at specific times of the day. For example, promoting early bird discounts for morning commuters or evening deals for leisure riders can increase ridership during these periods. Personalized marketing messages based on user behavior and preferences, derived from the prediction models, can also improve user engagement and satisfaction.

Result and Discussion

Model Performance

Evaluating the performance of predictive models is critical to understand their effectiveness and reliability. In this study, three models were trained and evaluated: Decision Tree Regressor, Random Forest Regressor, and Neural Network (MLPRegressor). The evaluation metrics used were Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), which measure the accuracy of the models' predictions.

For the day.csv dataset, the models were evaluated based on their performance metrics. The Decision Tree Regressor achieved an RMSE of 1023.97 and an MAE of 680.70. While the model was able to capture some patterns in the data, its relatively high error rates indicate a tendency towards overfitting, where the model performs well on training data but poorly on unseen data. The Random Forest Regressor significantly outperformed the Decision Tree, with an RMSE of 709.08 and an MAE of 469.99. The ensemble approach of Random Forests, which combines the predictions of multiple decision trees, helped in reducing overfitting and provided more accurate and stable predictions. However, the Neural Network model did not perform as well for the daily predictions. It achieved an RMSE of 3970.61 and an MAE of 3576.52, indicating that the model struggled to capture the underlying patterns in the data. This poor performance could be due to insufficient tuning of hyperparameters or the model’s complexity relative to the dataset. For the `hour.csv` dataset, similar evaluations were conducted, highlighting the strengths and weaknesses of each model in capturing hourly rental patterns.

When comparing the models, several key insights emerged. The Decision Tree Regressor provided a baseline for comparison. Its simplicity and interpretability are advantageous, but its high error rates suggest that it is less effective in capturing complex patterns in the data. Decision Trees are prone to overfitting, particularly with high-dimensional data, which limits their generalizability. The Random Forest Regressor demonstrated superior performance across both datasets. By averaging the predictions of multiple trees, it mitigates the overfitting issues inherent in single decision trees. The Random Forest's lower RMSE and MAE values indicate more accurate and reliable predictions, making it the best-performing model in this study. Its ability to handle a large number of features and capture non-linear relationships contributed to its success. The Neural Network (MLPRegressor), while theoretically capable of modeling complex relationships, underperformed in this study. Its high RMSE and MAE values suggest difficulties in training and convergence, possibly due to the need for extensive hyperparameter tuning and larger datasets to fully leverage its capabilities. Neural Networks are powerful tools, but their complexity requires careful calibration to achieve optimal performance.

The comparative analysis highlights the effectiveness of ensemble methods like Random Forests in handling diverse and complex datasets. The Random Forest model's balance between bias and variance made it the most suitable choice for predicting bike rental demand in this study. Conversely, while Neural Networks hold potential, their practical application requires careful consideration of model architecture, hyperparameters, and data volume.

Prediction Analysis

The predictive models developed in this study provided detailed forecasts of bike rental demand across various time periods. Using the Random Forest Regressor for daily predictions (`day.csv`) and the Neural Network model for hourly predictions (`hour.csv`), we generated forecasts that allowed for a comprehensive analysis of rental patterns.

For daily predictions, the model accurately captured seasonal trends and fluctuations. The data showed a significant increase in bike rentals during the summer months, with a notable decline in winter. This seasonal variation aligns with the expected behavior, as warmer weather typically encourages more outdoor activities, including cycling. Additionally, the model highlighted specific spikes in rental demand on public holidays and weekends, further corroborating the influence of these factors on user behavior.

Hourly predictions provided even more granular insights. The Neural Network model effectively identified peak usage times within a day. The analysis revealed that bike rentals peaked during morning and evening rush hours, corresponding to typical commuting times. Weekday mornings from 7:00 AM to 9:00 AM and evenings from 5:00 PM to 7:00 PM showed the highest rental activity. Conversely, on weekends, the peak periods shifted to late mornings and early afternoons, indicating recreational use.

These temporal patterns are crucial for optimizing bike-sharing operations. Understanding when demand is highest allows for better allocation of resources, ensuring that enough bikes are available during peak times and reducing idle capacity during off-peak periods.

The analysis of predicted rental demand also facilitated the identification of broader trends and specific peak periods. One significant trend observed was the consistent increase in rentals during the middle of the year, particularly from May to September. This period coincides with favorable weather conditions and longer daylight hours, making it ideal for cycling. The model's ability to predict these trends with high accuracy is valuable for strategic planning and marketing efforts.

In addition to seasonal trends, the model identified recurring weekly patterns. Weekdays consistently showed higher rental volumes during commuting hours, while weekends saw increased activity in the afternoon. These patterns suggest different user behaviors based on the day of the week, with weekdays dominated by work-related commutes and weekends by leisure activities.

Peak rental periods were clearly delineated through the analysis. For instance, the highest rental counts were observed during holiday weekends and major public events, indicating a surge in demand likely driven by both locals and tourists. Understanding these peaks allows for targeted marketing campaigns and promotional efforts aimed at maximizing usage during these times.

Moreover, the hourly predictions highlighted critical periods for operational adjustments. Ensuring bike availability during the morning rush hour can significantly enhance user satisfaction and system efficiency. Conversely, during off-peak hours, maintenance and redistribution efforts can be prioritized to prepare for the next peak period.

Marketing Strategy Optimization

The predictive models developed in this study provide a robust foundation for optimizing marketing strategies within the bike-sharing industry. By leveraging accurate forecasts of rental demand, bike-sharing companies can tailor their marketing efforts to maximize impact and efficiency. The insights derived from the predictions enable a data-driven approach to identifying peak demand periods, understanding user behavior, and strategically targeting potential customers.

One of the key benefits of these predictions is the ability to identify high-demand periods and locations. Marketing campaigns can be timed to coincide with these peaks, ensuring that promotional efforts reach the maximum number of users when they are most likely to engage with the service. For instance, knowing that demand spikes during summer months allows companies to launch summer-specific promotions, such as discounted rates for tourists or special offers for weekend riders.

Moreover, the predictions can inform the development of dynamic pricing strategies. By understanding when demand is highest, companies can adjust prices accordingly to optimize revenue. For example, implementing higher prices during peak hours and offering discounts during off-peak times can balance demand and improve overall system utilization. This dynamic approach not only enhances user satisfaction by providing cost-effective options but also ensures that the bike-sharing service operates efficiently.

Based on the prediction results, several targeted marketing strategies can be effectively employed to enhance the bike-sharing service. Seasonal promotions can be introduced during periods of higher demand, such as the summer months. These promotions may include summer passes for unlimited rides, discounts for group rentals, or collaborations with local events and festivals to offer exclusive deals to attendees. By aligning promotions with seasonal demand trends, bike-sharing companies can maximize user engagement and ridership during peak periods.

Commuter incentives are another effective strategy, given the significant demand observed during morning and evening rush hours on weekdays. Marketing campaigns targeting daily commuters could offer discounted monthly subscriptions for regular users, early bird discounts for morning rides, or referral bonuses for bringing in new users during peak commuting times. These incentives can attract regular commuters, increasing daily usage and fostering customer loyalty. Weekend and holiday campaigns can be designed to attract recreational riders, who tend to use bike-sharing services more during these times. Campaigns could include family packages, promotional rates for weekend trips, or partnerships with tourist attractions and parks to offer combined tickets and bike rentals. Such targeted promotions can enhance user experience and boost ridership during weekends and holidays.

Location-based advertising is another potent strategy. By identifying high-demand locations, advertising campaigns can focus on areas with high foot traffic, such as business districts, tourist hotspots, and university campuses. Digital advertising through geo-targeted ads on social media and search engines can effectively reach potential users in these high-demand areas, driving increased usage of the bike-sharing service. Event-specific promotions can leverage periods of increased demand due to local events like sports games, concerts, and community festivals. Tailored promotions for these events, such as ride-and-event ticket bundles or temporary docking stations near event venues, can attract event-goers to use the bike-sharing service, thereby boosting demand during these specific periods.

Finally, weather-responsive campaigns can be developed based on real-time integration of weather forecasts with predictive models. Special promotions can be offered on days with favorable weather forecasts to encourage bike usage. Alternatively, rain-check guarantees can assure users that their ride investment is protected in case of sudden bad weather. These weather-responsive strategies can help maintain consistent bike usage regardless of weather conditions.

Challenges and Limitations

Conducting this study involved several challenges that impacted various stages of the research process. One of the primary challenges was data preprocessing, which required handling missing values, normalizing continuous variables, and encoding categorical variables. The datasets contained several inconsistencies and missing entries, particularly in the weather-related fields. Addressing these gaps required careful imputation techniques and thorough data cleaning to ensure the integrity and accuracy of the data used for model training.

Another significant challenge was the computational complexity involved in training and tuning advanced models such as Random Forest Regressors and Neural Networks. The Random Forest model, although robust, required substantial computational resources and time to train, especially with hyperparameter tuning through Grid Search. Similarly, the Neural Network model, which inherently involves multiple hyperparameters like the number of layers, neurons per layer, and learning rates, presented difficulties in achieving optimal performance without extensive experimentation and computational power.

Despite the rigorous approach taken, the study faced several limitations, both in the models employed and the dataset used. One limitation of the Random Forest and Neural Network models is their complexity, which can lead to overfitting, especially with a relatively small dataset. While cross-validation helps mitigate this issue, the risk of the models capturing noise rather than meaningful patterns remains.

The dataset itself also posed limitations. The `day.csv` and `hour.csv` datasets, while comprehensive, did not capture all possible factors influencing bike rental demand. For instance, variables such as real-time traffic conditions, bike availability, and specific event data were not included, which could significantly impact rental patterns. The absence of these factors means the models might miss some important predictors of demand, potentially limiting their predictive power.

Another limitation is the temporal scope of the data. The dataset spans a limited period, which might not fully capture longer-term trends or seasonal variations over multiple years. This limitation restricts the models' ability to generalize beyond the time frame of the dataset, potentially affecting their accuracy in long-term forecasting.

Several potential biases could affect the study's outcomes, particularly those arising from data collection methods and model training processes. One such bias is the seasonality bias, where the model might overemphasize certain periods of the year (e.g., summer) due to higher observed rental counts. This bias was addressed by ensuring that the models were trained and validated using data from all seasons, balancing the representation across different periods.

Another potential bias is related to the geographic concentration of data points. If the dataset is predominantly composed of data from certain high-traffic locations, the models might skew their predictions towards these areas. This bias was mitigated by incorporating geographic diversity in the training data and validating the models' performance across different locations.

Additionally, the feature selection process could introduce bias if certain relevant features were inadvertently excluded. To counter this, a comprehensive feature engineering approach was adopted, considering both domain knowledge and statistical correlations to include a wide range of relevant variables. Regularization techniques in the model training process also helped prevent overfitting to specific features, promoting a more generalized model.

Conclusion

This study conducted a comprehensive predictive analysis of bike rental demand using machine learning models, including Decision Tree Regressor, Random Forest Regressor, and Neural Networks (MLPRegressor). The Random Forest Regressor emerged as the best-performing model for daily predictions, with an RMSE of 709.08 and an MAE of 469.99. The Neural Network model, although it showed promise, required further tuning and data to improve its performance. Overall, the predictive models effectively identified key factors influencing bike rental demand, such as weather conditions, seasonality, and time of day.

The analysis revealed significant trends, such as higher demand during summer months and peak usage times during weekday mornings and evenings. These insights underscore the importance of considering temporal and weather-related variables in predicting bike rental demand. By leveraging these predictive models, bike-sharing companies can better understand and anticipate user behavior, leading to improved operational efficiency and user satisfaction.

The findings of this study have significant implications for the operational efficiency and user satisfaction of bike-sharing systems. By accurately predicting demand, bike-sharing companies can optimize the allocation of bikes across different locations and times, ensuring availability during peak periods and reducing idle capacity during off-peak times. This can lead to a more balanced and efficient system, minimizing user wait times and enhancing the overall user experience.

Furthermore, the predictive insights can inform targeted marketing strategies, enabling companies to launch promotions and campaigns that align with high-demand periods and locations. For example, offering discounts during identified peak seasons or specific times of day can attract more users and increase ridership. These data-driven marketing strategies can drive business growth and improve the profitability of bike-sharing operations.

While this study provides valuable insights, there are several avenues for future research that can build on these findings. One potential direction is the incorporation of additional data sources, such as real-time traffic conditions, social events, and bike availability metrics, to enhance the predictive models' accuracy. Integrating these factors can provide a more comprehensive view of the variables influencing bike rental demand.

Future studies could also explore advanced machine learning techniques, such as deep learning models and ensemble methods, to further improve prediction accuracy. Additionally, expanding the temporal scope of the dataset to include multiple years can help capture long-term trends and seasonal variations more effectively. Researchers should also consider the potential impacts of external factors, such as policy changes and infrastructure developments, on bike rental patterns.

Declarations

Author Contributions

Conceptualization: I.G.A.K.W.; Methodology: I.G.A.K.W.; Software: Y.Y.; Validation: N.O.; Formal Analysis: I.G.A.K.W.; Investigation: Y.Y.; Resources: N.O.; Data Curation: Y.Y.; Writing—Original Draft Preparation: I.G.A.K.W.; Writing—Review and Editing: N.O.; Visualization: Y.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Macioszek, P. Świerk, and A. Kurek, "The Bike-Sharing System as an Element of Enhancing Sustainable Mobility—A Case Study Based on a City in Poland," Sustainability, vol. 12, no. 8, p. 3285, 2020, doi: 10.3390/su12083285.

[2] M. Usama, O. Zahoor, Y. Shen, and Q. Bao, "Dockless Bike-Sharing System: Solving the Problem of Faulty Bikes With Simultaneous Rebalancing Operation," J. Transp. Land Use, vol. 13, no. 1, pp. 491–515, 2020, doi: 10.5198/jtlu.2020.1594.

[3] Z. Chen, D. v. Lierop, and D. Ettema, "Dockless Bike-Sharing's Impact on Mode Substitution and Influential Factors: Evidence From Beijing, China," J. Transp. Land Use, vol. 15, no. 1, 2022, doi: 10.5198/jtlu.2022.1983.

[4] S. C. Chang, R. Song, S. He, and G. Y. Qiu, “Innovative Bike-Sharing in China: Solving Faulty Bike-Sharing Recycling Problem,” J. Adv. Transp., vol. 2018, pp. 1–10, 2018, doi: 10.1155/2018/4941029.

[5] S. Gao, Y. Li, and H. Guo, “Understanding the Adoption of Bike Sharing Systems,” J. Hosp. Tour. Technol., vol. 10, no. 3, pp. 464–478, 2019, doi: 10.1108/jhtt-08-2018-0089.

[6] S. J. Mooney et al., “Freedom From the Station: Spatial Equity in Access to Dockless Bike Share,” J. Transp. Geogr., vol. 74, pp. 91–96, 2019, doi: 10.1016/j.jtrangeo.2018.11.009.

[7] L.-Y. Qiu and L. He, “Bike Sharing and the Economy, the Environment, and Health-Related Externalities,” Sustainability, vol. 10, no. 4, p. 1145, 2018, doi: 10.3390/su10041145.

[8] X. Wang, G. Lindsey, J. Schoner, and A. Harrison, “Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and From Stations,” J. Urban Plan. Dev., vol. 142, no. 1, 2016, doi: 10.1061/(asce)up.1943-5444.0000273.

[9] Y. He, Z. Song, Z. Liu, and N. N. Sze, “Factors Influencing Electric Bike Share Ridership: Analysis of Park City, Utah,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2673, no. 5, pp. 12–22, 2019, doi: 10.1177/0361198119838981.

[10] V. Albuquerque, M. S. Dias, and F. Bação, “Machine Learning Approaches to Bike-Sharing Systems: A Systematic Literature Review,” Isprs Int. J. Geo-Inf., vol. 10, no. 2, p. 62, 2021, doi: 10.3390/ijgi10020062.

[11] C. Li, Z. Dai, W. Peng, and J. Shen, “Green Travel Mode: Trajectory Data Cleansing Method for Shared Electric Bicycles,” Sustainability, vol. 11, no. 5, p. 1429, 2019, doi: 10.3390/su11051429.

[12] K. Boonjubut and H. Hasegawa, “Accuracy of Hourly Demand Forecasting of Micro Mobility for Effective Rebalancing Strategies,” Manag. Syst. Prod. Eng., vol. 30, no. 3, pp. 246–252, 2022, doi: 10.2478/mspe-2022-0031.

[13] C. Kapuku, S.-Y. Kho, D.-K. Kim, and S.-H. Cho, “Assessing and Predicting Mobility Improvement of Integrating Bike-Sharing Into Multimodal Public Transport Systems,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2675, no. 11, pp. 204–213, 2021, doi: 10.1177/03611981211045071.

[14] L. Lin, Z. He, and S. Peeta, “Predicting Station-Level Hourly Demand in a Large-Scale Bike-Sharing Network: A Graph Convolutional Neural Network Approach,” Transp. Res. Part C Emerg. Technol., vol. 97, pp. 258–276, 2018, doi: 10.1016/j.trc.2018.10.011.

[15] H. I. Ashqar, M. Elhenawy, H. A. Rakha, M. Almannaa, and L. House, “Network and Station-Level Bike-Sharing System Prediction: A San Francisco Bay Area Case Study,” J. Intell. Transp. Syst., vol. 26, no. 5, pp. 602–612, 2021, doi: 10.1080/15472450.2021.1948412.

[16] M. He, X. Ma, and Y. Jin, “Station Importance Evaluation in Dynamic Bike-Sharing Rebalancing Optimization Using an Entropy-Based TOPSIS Approach,” Ieee Access, vol. 9, pp. 38119–38131, 2021, doi: 10.1109/access.2021.3063881.

[17] H. Song, G. Yin, X. Wan, M. Guo, Z. Xie, and J. Gu, “Increasing Bike-Sharing Users’ Willingness to Pay — A Study of China Based on Perceived Value Theory and Structural Equation Model,” Front. Psychol., vol. 12, 2022, doi: 10.3389/fpsyg.2021.747462.

[18] X. Zhang, J. Wang, X. Long, and W. Li, “Understanding the Intention to Use Bike-Sharing System: A Case Study in Xi’an, China,” Plos One, vol. 16, no. 12, p. e0258790, 2021, doi: 10.1371/journal.pone.0258790.

[19] L. Ma, X. Zhang, and G. S. Wang, "Identifying the Reasons Why Users in China Recommend Bike Apps," Int. J. Mark. Res., vol. 59, no. 6, pp. 767–786, 2017, doi: 10.2501/ijmr-2017-053.

[20] X. Qian, "Using Machine Learning for Bike Sharing Demand Prediction," Theor. Nat. Sci., vol. 30, no. 1, pp. 110–119, 2024, doi: 10.54254/2753-8818/30/20241078.

[21] Z. Wang, L. Cheng, Y. Li, and Z. Li, "Spatiotemporal Characteristics of Bike-Sharing Usage Around Rail Transit Stations: Evidence From Beijing, China," Sustainability, vol. 12, no. 4, p. 1299, 2020, doi: 10.3390/su12041299.

[22] X. Yu, X. Wang, and Y. Zhao, "Analysis on the Travel Characteristics of Sharing Bikes Connected With Different Types of Public Transportation Stations," 2021, doi: 10.3233/faia210260.

[23] T. Ashraf, A. Hossen, K. Dey, S. El-Dabaja, M. S. Aljeri, and B. Naik, "Impacts of Bike Sharing Program on Subway Ridership in New York City," Transp. Res. Rec. J. Transp. Res. Board, vol. 2675, no. 9, pp. 924–934, 2021, doi: 10.1177/03611981211004980.