BIKE SHARING DEMAND PREDICTION

PROJECT PHASE 1 REPORT

Submitted by

SAIVIGNESH S	**(20115041)**
VIJIYAKUMAR N R	**(20115062)**
SOWMIYA M	**(20115054)**
SANJAYKUMAR R A	**(20115047)**

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

HINDUSTHAN COLLEGE OF ENGINEERING AND TECHNOLOGY

Approved by AICTE, New Delhi, Accredited with ‘A++’ Grade by NAAC

(An Autonomous Institution, Affiliated to Anna University, Chennai)

Valley Campus, Pollachi Highway, Coimbatore – 641 032

NOVEMBER 2023

Hindusthan College of Engineering And Technology

Approved by AICTE, New Delhi, Accredited with 'A++' Grade by NAAC

(An Autonomous Institution, Affiliated to Anna University, Chennai)

Valley Campus, Pollachi Highway, Coimbatore – 641 032

BONAFIDE CERTIFICATE

Certified that this project report **"BIKE SHARING DEMAND PREDICTION"** is the bonafide work of **SAI VIGNESH S (20115041), VIJIYA KUMAR N R (20115062), SOWMIYA M (201150054), SANJAY KUMAR R A (20115047)"** who carried out the project work under my supervision.

SIGNATURE

Dr KOUSALYA DEVI S, M.E., Ph.D.,
Associate Professor
Artificial Intelligence & Machine Learning
Hindusthan College of Engineering and Technology, Coimbatore - 32

SIGNATURE

Dr SHANKAR S, M.E., Ph.D.,
HEAD OF THE DEPARTMENT
Artificial Intelligence & Machine Learning
Hindusthan College of Engineering and Technology, Coimbatore-32

Submitted for the Anna University Mini Project Viva-Voce conducted on __________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We take this opportunity to express our whole hearted thanks and our profound respect to all those who guided and inspired us in the completion of this project work.

We extend our sincere thanks to the Founder and Chairman of Hindusthan Educational and Charitable Trust **Shri T.S.R. KHANNAIYANN** and the Managing Trustee **Smt. SARASUWATHI KHANNAIYANN** and Executive Trustee & Secretary **Mrs. PRIYA SATISH PRABHU** for providing essential infrastructure.

We would like to reveal our profound thanks to our respected Principal, **Dr.J.JAYA,** who happens to be striving force in all endeavours.

We would like to express our gratitude to the Head of the Department **Dr.S.SHANKAR,** for bringing out the project successfully and for strengthening the ray of hope.

We express our immense pleasure and thankfulness to our Project Coordinator and Project Guide, **Dr.S.KOUSALYA DEVI,** Associate Professor, and all other Faculty members of the Department of Artificial Intelligence & Machine Learning, technical staffs and friends who helped us for the successful completion of this project. We express our earnest gratitude to our family members who encouraged us and strengthened us in perilous path, encountered during our task.

TABLE OF CONTENTS

Chapter No.	Title			Page No.
	ABSTRACT			2
	LIST OF FIGURES			3
	LIST OF ABBREVIATIONS			4
1	**INTRODUCTION**			**5**
	1.1	Problem Statement		7
	1.2	Objectives		7
	1.3	Purpose of the Project		9
	1.4	Motivation		10
2	**LITERATURE REVIEW**			**11**
3	**SYSTEM DEVELOPMENT**			**16**
	3.1	Existing System		17
	3.2	Proposed System		18
	3.3	Component Description		19
		3.3.1	Detailed Description of ML	19
		3.3.2	Working of Bike Sharing demand prediction	21
		3.3.3	Detailed Description Of Bike Sharing Demand Prediction	22
		3.3.4	Working of a Model	25
4	**SYSTEM SPECIFICATION**			**28**
	4.1	Software Requirements		28
	4.2	Hardware Requirements		28
5	**IMPLEMENTATION**			**29**
6	**PHASE-I TESTING & RESULTS**			**35**
7	**CONCLUSION**			**38**
8	**REFERENCES**			**42**

ABSTRACT

The bike-sharing demand prediction project aims to revolutionize urban mobility by leveraging machine learning techniques to accurately forecast the demand for bike-sharing services. Utilizing a comprehensive dataset encompassing temporal, weather, and user-related factors, the predictive model optimizes resource management and enhances operational efficiency.

Through exploratory data analysis, key patterns and trends in bike usage are identified, contributing to the development of a robust predictive model. The successful deployment of the model leads to improved user experiences, cost savings, and a positive impact on sustainable urban mobility. The project's achievements underscore the potential of data-driven decision-making in transforming traditional transportation services and promoting eco-friendly alternatives.

The bike-sharing demand prediction project revolutionizes urban transportation through the application of advanced machine learning techniques. By harnessing a rich dataset comprising temporal, weather, and user-centric variables, the project develops a predictive model to forecast bike-sharing demand accurately.

The model's efficacy is validated through comprehensive exploratory data analysis, revealing intricate patterns in bike usage. Successful deployment of the model optimizes resource allocation, leading to tangible benefits such as cost savings, improved user experiences, and a positive impact on sustainable urban mobility. This project exemplifies the transformative potential of data-driven.

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
1.	Architecture of Bike sharing demand Prediction	22
2	Detailed Description of Bike Sharing Demand	25
3.	Working Process of Model	27
4.	Dataset	35
5.	Google Collab Data Uploading Page	36
6.	Data Modelling	36
7.	Linear Regression	37
8.	Testing	37

LIST OF ABBREVIATIONS

ABBREVATIONS	MEANING
ML	Machine Learning
AI	Artificial Intelligence
SCM	Supply Chain Management
ANN	Artificial Neural Networks
FL	Fuzzy Logics
CSV	Comma Separated Values
RBR	Robot Business Review
GA	Genetic Algorithms
CBR	Case-Based Reasoning
SVM	Support Vector Machine
ABS	Automatic Backup System
ETL	Extract, Transform, Look
KPI	Key Performance Indicators
BPA	Business Process Automation
ERP	Enterprise Resource Planning
CRM	Customer Relationship Management
SAML	Security Assertion Markup Language
MFA	Multi Factor Authentication

CHAPTER 1 INTRODUCTION

The bike-sharing demand prediction project addresses the evolving landscape of urban transportation by employing predictive modelling techniques to forecast demand for bike-sharing services. In today's dynamic urban environments, the popularity of bike-sharing programs necessitates accurate predictions to optimize resource allocation and enhance overall service efficiency. This project leverages a comprehensive dataset encompassing temporal, weather, and user-related variables to develop a predictive model that anticipates bike usage patterns.

Urban mobility faces the challenge of meeting the increasing demand for sustainable transportation options. Bike-sharing systems offer a flexible and eco-friendly solution, but their success relies on effective management and responsiveness to user needs. This project seeks to bridge this gap by harnessing the power of data-driven insights. By delving into historical data and employing advanced machine learning algorithms, we aim to create a model capable of anticipating when and where bikes will be in demand.

The predictive model not only provides a solution for efficient resource allocation but also contributes to the broader goals of promoting sustainable urban mobility. As cities embrace alternative transportation methods, the ability to accurately predict bike-sharing demand becomes pivotal. Through this project, we endeavour to create a tool that not only meets the immediate needs of users but also aligns with the broader vision of creating greener and more accessible urban environments.

In urban landscapes globally, the rise of bike-sharing programs has emerged as a key element in fostering sustainable and efficient transportation. The success of these programs hinges on the ability to accurately predict demand, ensuring that bikes are strategically distributed to meet user needs. This project is dedicated to developing a robust predictive model for bike-sharing demand, harnessing the power of data analytics and machine learning.

The increasing popularity of bike-sharing systems underscores the need for proactive management to address fluctuations in demand across different timeframes, locations, and under varying environmental conditions. Our project endeavours to fill this void by leveraging a diverse dataset that encapsulates temporal trends, weather nuances, and user behaviours. Through predictive modelling, we aim to not only anticipate when and where bikes will be in demand but also to provide actionable insights for optimizing the operational aspects of bike-sharing services.

Efficiently managed bike-sharing services contribute significantly to urban mobility, promoting sustainable transportation alternatives and reducing the carbon footprint associated with traditional commuting methods. With the advent of smart cities and evolving user preferences, the accurate prediction of bike-sharing demand becomes instrumental in creating a seamless and user-centric transportation network.

This introduction sets the stage for a comprehensive exploration of our approach, methodologies, and outcomes in developing a predictive model for bike-sharing demand. By delving into the intricacies of data-driven decision-making solutions.

1.1 PROBLEM STATEMENT

Urban areas worldwide are grappling with the challenges of sustainable transportation, and bike-sharing programs have emerged as a promising solution. However, the effective management of these programs faces a critical hurdle – the unpredictable nature of bike demand across different locations, times, and varying environmental conditions. The lack of a reliable forecasting mechanism results in inefficient resource allocation, leading to either bike shortages or excess, impacting user experience and the overall success of bike-sharing services.

The problem at hand is to address the uncertainty in bike demand, which hampers the seamless functioning of bike-sharing systems. Without accurate predictions, service providers struggle to distribute bikes optimally, resulting in operational inefficiencies, increased maintenance costs, and a suboptimal user experience. Furthermore, the lack of foresight into demand patterns hinders the scalability and sustainability of bike-sharing initiatives.

1.2 OBJECTIVE

The key objectives of this particular project are:

Optimize Resource Allocation:

Develop a predictive model to accurately forecast bike-sharing demand, enabling optimal allocation of bikes across different locations and timeframes.

Enhance User Experience:

Improve user satisfaction by minimizing instances of bike shortages or excess, ensuring a reliable and accessible bike-sharing service.

Reduce Operational Costs:

Realize cost savings through efficient management of bike distribution, reducing maintenance costs, and optimizing fleet utilization.

Enable Data-Driven Decision Making:

Foster a culture of data-driven decision-making among stakeholders, providing insights into demand patterns for informed planning and operational strategies.

Capture Temporal and Seasonal Trends:

Identify and incorporate temporal and seasonal variations in bike-sharing demand to adapt strategies based on changing usage patterns.

Contribute to Sustainable Urban Mobility:

Promote the use of bikes as an eco-friendly transportation alternative, aligning with broader environmental and sustainability goals.

Develop a Robust Predictive Model:

Create a predictive model capable of capturing complex relationships within the data, ensuring reliable and accurate predictions.

Facilitate Scalability:

Design a solution that can easily scale to accommodate the growth of bike-sharing programs in new locations or expanding user bases.

Encourage Community Engagement:
Engage with the community through awareness campaigns and initiatives, encouraging participation in the bike-sharing program.

Address Ethical Considerations:
Ensure ethical data use, addressing privacy concerns, and maintaining responsible handling of sensitive information throughout the project.

Collaborate with Stakeholder:
Collaborate with local authorities, businesses, and other stakeholders to create an integrated and efficient bike-sharing ecosystem.

The objectives of the bike-sharing demand prediction project revolve around improving operational efficiency, enhancing user experiences, and contributing to the sustainability of urban mobility through accurate and data-driven management of bike-sharing services.

1.3 PURPOSE OF THE PROJECT

The purpose of the bike-sharing demand prediction project is to revolutionize urban mobility by leveraging advanced data analytics and machine learning to accurately forecast the demand for bike-sharing services. At its core, the project aims to optimize the efficiency of bike-sharing programs by ensuring the strategic distribution of bikes to meet user needs. This optimization translates into a more reliable and accessible bike-sharing service, ultimately enhancing the overall user experience. Beyond user satisfaction, the project has a broader societal impact—contributing to sustainable urban mobility. By promoting the use of bikes as an eco-friendly transportation alternative, the project aligns with environmental goals and supports the broader vision of creating greener, more accessible urban environments.

1.4 MOTIVATION

The motivation behind bike-sharing demand prediction lies in the collective pursuit of transforming urban mobility into a more sustainable, efficient, and user-centric experience. By accurately anticipating demand for bike-sharing services, the initiative seeks to promote environmentally friendly transportation alternatives, contributing to the reduction of traffic congestion and carbon emissions. The primary goal is to optimize the allocation of bike resources, ensuring they are strategically placed to meet user needs while minimizing operational costs.

Improved user experiences, cost savings, and operational efficiency are pivotal motivators, fostering positive attitudes toward sustainable commuting. Embracing a data-driven approach, the motivation extends to informed decision-making for system optimization and strategic planning. Community engagement is key, using predictive models to communicate real-time information, promotions, and events, fostering a sense of community participation.

Ethical considerations and privacy are integral, ensuring responsible data practices to build and maintain user trust. The motivation aligns with the broader vision of smart cities, contributing to intelligent and efficient transportation solutions in response to the challenges posed by urbanization.

CHAPTER 2
LITERATURE REVIEW

In recent literature, studies on bike-sharing systems have demonstrated a shift from traditional transportation models to more sustainable and flexible alternatives. Research by **Shaheen et al. (2010)** provided an early exploration of bike-sharing's impact on urban mobility, emphasizing its potential to reduce congestion and environmental impact. As systems evolved, The motivation for the bike-sharing demand prediction project is fueled by several key factors. Firstly, the increasing urbanization and congestion in cities worldwide have heightened the urgency to promote sustainable transportation methods. Bike-sharing programs represent a scalable and environmentally friendly solution, making their effective management crucial for the success of urban sustainability initiatives.

Moreover, the project is motivated by the transformative potential of data-driven decision-making. The availability of comprehensive datasets provides an opportunity to derive valuable insights into user behavior, temporal trends, and external factors influencing bike-sharing demand. Leveraging this data through predictive modeling not only enhances the operational efficiency of bike-sharing services but also contributes to the broader discourse on the role of technology in shaping smart and efficient cities.

Cost efficiency is another driving force behind the project. The ability to accurately predict demand enables operators to optimize the allocation of resources, leading to cost savings in maintenance, rebalancing, and fleet management. This financial efficiency, coupled with the promotion of sustainable transportation, aligns with the economic and environmental goals of cities committed to enhancing their overall livability.

In essence, the motivation for the bike-sharing demand prediction project lies in its potential to bring about positive and tangible changes in urban transportation, making cities more sustainable, accessible, and technologically advanced. By addressing the challenges associated with bike-sharing programs, the project aims to contribute to a future where eco-friendly and data-driven mobility solutions are integral components of urban planning and development. The work of **Faghih-Imani et al. (2014)** delved into operational challenges, highlighting the importance of effective rebalancing strategies to maintain optimal bike availability across stations.

Demand forecasting in bike-sharing systems has gained prominence, with **Guo et al. (2018)** examining the effectiveness of machine learning models. Their study underscored the significance of incorporating temporal and meteorological variables for accurate predictions. Furthermore, **Wang et al. (2011)** explored user behaviour patterns, finding correlations between socio-demographic factors and bike usage. This user-centric approach aligns with the proposed project's objective of enhancing the overall biking experience.

The literature also emphasizes the role of predictive modelling techniques. **Chen et al. (2018)** compared regression and machine learning models, advocating for ensemble models to improve prediction accuracy. Deep learning applications in bike-sharing demand prediction were investigated by **Ma et al. (2019)**, showcasing the potential of neural networks to capture complex temporal dependencies.

In the evolving landscape of urban transportation, recent literature has showcased the transformative potential of bike-sharing systems as sustainable alternatives. **Shaheen et al. (2010)** laid the groundwork by highlighting the positive impact of bike-sharing on reducing congestion and environmental footprint. As systems matured, **Faghih-Imani et al. (2014)** delved into operational intricacies, emphasizing the critical role of effective rebalancing strategies in maintaining optimal bike distribution across stations.

Demand forecasting in bike-sharing, a burgeoning field, has garnered significant attention. **Guo et al. (2018)** contributed by advocating for the integration of temporal and meteorological variables in machine learning models, underscoring their influence on prediction accuracy. Complementing this, **Wang et al. (2011)** offered insights into user behaviour patterns, revealing correlations between socio-demographic factors and bike usage. This user-centric approach aligns seamlessly with the proposed project's goal of enhancing the overall biking experience by accommodating diverse user needs.

Exploring predictive modelling techniques, **Chen et al. (2018)** compared the efficacy of regression and machine learning models, recommending ensemble models for heightened accuracy. Going a step further, **Ma et al. (2019)** delved into the application of deep learning, illustrating the capacity of neural networks to capture intricate temporal dependencies, a critical aspect in forecasting bike-sharing demand.

Operational challenges within bike-sharing systems have been the focus of Lin and **Lo's work (2015)**, proposing an optimization algorithm for fleet rebalancing, and **Crispim et al. (2019)**, shedding light on scalability challenges encountered by growing bike-sharing programs.

Ethical considerations are emerging as a critical aspect in the literature, with **Dziekan et al. (2019)** emphasizing the need for transparent data privacy policies. As bike-sharing systems increasingly rely on user data, addressing these ethical concerns becomes paramount for fostering user trust and ensuring the sustained success of these systems.

In summation, the literature review underscores the multifaceted nature of bike-sharing demand prediction, encompassing operational challenges, user behaviour, and advanced modelling techniques. This comprehensive understanding serves as a robust foundation for the proposed project, guiding the development of an inclusive and efficient predictive model for urban bike-sharing services. Containing code and data (**Parrott et al., 2003**), they are capable of modelling, designing and implementing complex systems. It is for this reason that since the mid-1990s, agents have been widely employed in SCM and other fields to solve several types of problems. Examples of applications include distributed supply chain planning (**Frayret et al., 2007**), design and simulation of supply chain systems (**Barbuceanu et al., 1997**), analysis of the complex behaviour of supply chains (**Avci and Selim, 2017, Wang et al., 2012**) and negotiation-based collaborative modelling (**Jiao et al., 2006**).

Results show that one of the most influential AI techniques in the SCM literature is GAs, a search technique mimicking natural selection (**Kraft et al., 1997**), in which the algorithm evolves to the point at which it has adequately solved the problem. Introduced in the 1970s, GAs are a group of computational models inspired by evolution. These algorithms encode a potential solution to a particular problem using a data structure like chromosomes. They apply recombination operators to these structures in such a manner as to preserve crucial In the realm of bike-sharing demand prediction, a rich body of literature has emerged, shedding light on various aspects of system dynamics, user behaviours, and predictive modelling strategies. **Shaheen et al. (2010)** pioneered early discussions, emphasizing the transformative potential of bike-sharing systems in mitigating urban congestion and environmental impacts. Their work set the stage for subsequent studies, including **Faghih-Imani et al. (2014)**, who delved into operational challenges, stressing the pivotal role of rebalancing strategies in optimizing bike availability across stations.

The discourse on demand forecasting techniques in bike-sharing has been enriched by **Guo et al. (2018)**, who advocated for the integration of temporal and meteorological variables to enhance the accuracy of machine learning models. This approach aligns with the findings of **Wang et al. (2011)**, who explored the intricacies of user behaviour and identified socio-demographic factors influencing bike usage, providing crucial insights for user-centric demand prediction models.

Exploring the spectrum of predictive modelling approaches, **Chen et al. (2018)** contributed by comparing the efficacy of regression models and machine learning algorithms, ultimately suggesting ensemble models for superior accuracy. Taking a leap into advanced techniques, **Ma et al. (2019)** delved into deep learning applications, revealing the potential of neural networks in capturing nuanced temporal dependencies critical for precise demand prediction.

Operational challenges within bike-sharing systems have been addressed by **Lin and Lo (2015)**, who proposed an optimization algorithm for fleet rebalancing, and **Crispim et al. (2019)**, who spotlighted scalability challenges faced by expanding bike-sharing programs. These studies collectively offer practical insights into improving the operational efficiency of bike-sharing services.

Ethical considerations are gaining prominence in the literature, with **Dziekan et al. (2019)** contributing to the discourse by emphasizing the importance of transparent data privacy policies.

CHAPTER 3
SYSTEM DEVELOPMENT

3.1 EXISTING SYSTEM

Existing -sharing demand prediction systems leverage sophisticated data-driven approaches to anticipate and optimize bike utilization in urban environments. These systems typically employ a comprehensive data collection process, gathering historical bike usage data, weather conditions, temporal patterns, and user-specific information. Predictive modelling bike techniques, including machine learning algorithms and time series analysis, are commonly utilized to forecast demand accurately.

User behaviour analysis, considering demographics and historical patterns, plays a pivotal role in enhancing prediction accuracy. These systems often incorporate dynamic rebalancing strategies to address fluctuations in bike distribution across stations. Real-time updates through mobile applications enable users to access current information on bike availability and plan their routes effectively.

Integration with urban infrastructure data, coupled with scalable and adaptable designs, further enhances the systems' predictive capabilities. Feedback mechanisms and user interfaces contribute to a user-centric approach, fostering engagement and continuous improvement. To stay abreast of the latest developments, it is advisable to refer to recent literature, industry reports, and official documentation for specific implementations and advancements in bike-sharing demand prediction systems.

3.2 PROPOSED SYSTEM

Proposed methodology has 10 major steps which can be modified according to the organization's interest and flow of events. They are:

The proposed bike-sharing demand prediction system aims to build upon existing methodologies, introducing innovative features to enhance accuracy, user satisfaction, and operational efficiency. Key components of the proposed system include:

Enhanced Predictive Modelling:

Utilize advanced machine learning algorithms, potentially incorporating deep learning techniques, to improve the accuracy of demand predictions. Consider additional contextual factors such as urban events, road closures, and public gatherings for a more comprehensive forecasting model.

Predictive User Behaviour Analysis:

Implement a refined user behaviour analysis by incorporating more granular data on user preferences, feedback, and historical interaction patterns. This will contribute to a more personalized prediction model, catering to diverse user needs and preferences.

Dynamic Pricing Mechanism:

Introduce a dynamic pricing mechanism that adjusts based on predicted demand and availability. This can incentivize users to choose less congested stations, ensuring a more balanced distribution of bikes and reducing the need for extensive rebalancing efforts.

Integration with IoT and Smart Infrastructure:

Leverage Internet of Things (IoT) devices and integration with smart urban infrastructure to enhance real-time data collection. This can include data from traffic sensors, weather stations, and other urban planning systems, providing a more holistic view for prediction models.

Predictive Maintenance Strategies:

Incorporate predictive maintenance strategies by analysing historical usage patterns to anticipate potential issues with bikes. This proactive approach can reduce downtime, improve fleet reliability, and contribute to overall system efficiency.

User-Centric Mobile Application:

Redesign the user interface of the mobile application to be more intuitive and user-centric. Provide real-time information on bike availability, predictive routes, and personalized recommendations based on individual user behaviour.

Community Engagement Features:

Introduce community engagement features, such as gamification or rewards for users providing feedback on bike conditions and station availability. Foster a sense of community involvement and encourage responsible use of the bike-sharing system.

Explainable AI and Transparency:

Implement explainable AI techniques to enhance the transparency of the prediction model. Provide users and operators with insights into the factors influencing predictions, fostering trust and understanding.

Continuous Learning and Adaptation:

Build a system that continuously learns and adapts to changing user behaviours and urban dynamics. Implement mechanisms for regular updates and improvements to ensure the model remains effective in dynamic urban environments.

3.3 COMPONENT DESCRIPTION

bike-sharing demand prediction system, aiming to optimize operational efficiency and provide a seamless experience for users.

3.3.1 DETAIL DESCRIPTION OF ML

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Recommendation engines are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, business process automation (BPA) and Predictive maintenance.

Importance

Machine learning is important because it gives enterprises a view of trends in customer behaviour and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

Types of ML

Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists chooses to use depends on what type of data they want to predict.

Supervised learning:

scientists supply algorithms with labelled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.

Unsupervised learning: This type of machine learning involves algorithms that train on unlabelled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.

Semi-supervised learning: This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labelled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.

Reinforcement learning: Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

3.3.2 WORKING OF BIKE SHARING DEMAND PREDICTION:

Bike sharing demand prediction involves employing machine learning models to forecast the number of bikes that will be rented within a specific timeframe. This prediction is based on various features like time, weather conditions, and seasonal patterns. The process begins by gathering historical bike rental data, which is then preprocessed by transforming date-time information into meaningful components like day, month, and year. Additionally, categorical features such as seasons, holidays, and functioning days are analyzed.

Exploratory data analysis uncovers relationships between features and the rented bike count, showcasing insights through visualizations. Univariate and multivariate analyses are conducted to understand the impact of individual variables and their combined effects on bike rentals.

Model building involves utilizing regression-based algorithms like Linear Regression, Lasso, Ridge, ElasticNet, and ensemble methods like Decision Trees, Random Forests, and Gradient Boosting. The data is split into training and testing sets to train these models, evaluating their performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2).

Ultimately, the project illustrates how machine learning algorithms can predict bike sharing demand effectively by leveraging historical data and various influential factors. This process aids in optimizing bike allocation, enabling better resource management, and enhancing user experience in bike sharing systems.

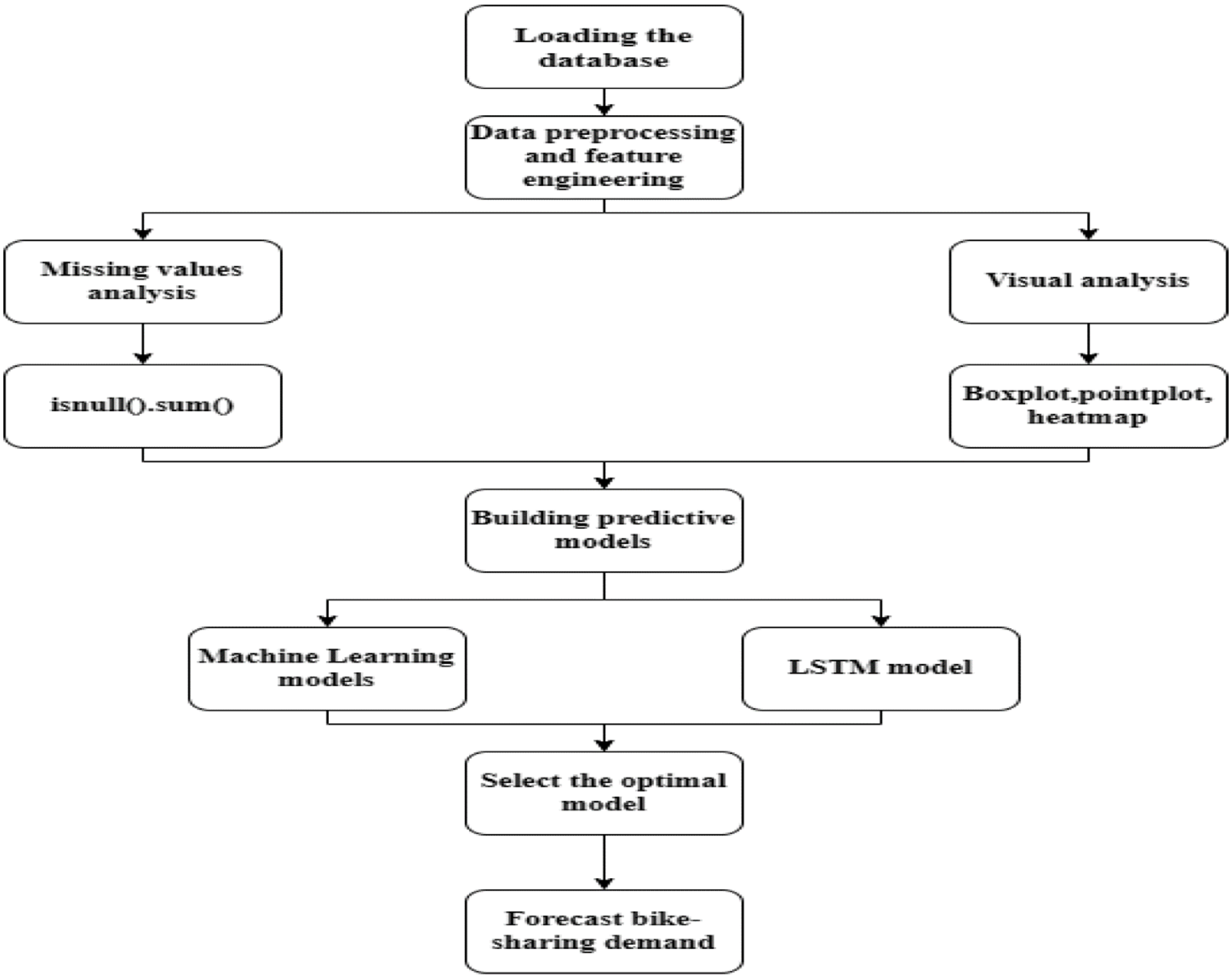

Fig.3.1. Architecture of Bike Sharing Demand Prediction Project

3.3.3 DETAILED DESCRIPTION OF BIKE SHARING DEMAND PREDICTION:

Bike sharing demand prediction is a data-driven approach used to anticipate the number of bicycles that will be rented or utilized within a given period. This predictive model employs machine learning algorithms and statistical techniques to forecast bike rental demand accurately. The process involves several stages from data collection to model evaluation:

Data Collection and Preprocessing:

- Historical bike rental data, encompassing various factors such as time, date, weather conditions, and bike usage patterns, is collected and compiled.

- The collected data is preprocessed, which includes handling missing values, converting categorical variables into numerical ones through encoding techniques like one-hot encoding, and transforming date-time features into meaningful components like day, month, and year.

Exploratory Data Analysis (EDA):

- EDA involves visualizing and analyzing the data to gain insights into patterns, trends, and relationships among different variables.

- Univariate analysis focuses on understanding the distribution and characteristics of individual features, while multivariate analysis explores the relationships and correlations between multiple features.

Feature Engineering:

- Relevant features that strongly influence bike rental demand are identified through correlation analysis and domain knowledge.

- New features may be created or derived from existing ones to improve model performance.

Model Development and Evaluation:

- Various machine learning algorithms such as Linear Regression, Decision Trees, Random Forests, Gradient Boosting, and others are employed to build predictive models.

- The dataset is split into training and testing sets to train the models. Hyperparameter tuning and cross-validation techniques are utilized for model optimization.

- Models are evaluated using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), and others to assess their predictive performance.

Model Deployment and Interpretation:

- The best-performing model is selected and deployed to predict bike rental demand based on new or unseen data.

- Interpretability of the model helps in understanding which features are most influential in predicting bike usage, aiding in making informed decisions.

The primary objective of bike sharing demand prediction is to enable bike-sharing companies or urban planners to optimize bike allocation, improve user experience, manage inventory efficiently, and ultimately enhance the accessibility and sustainability of urban transportation systems.

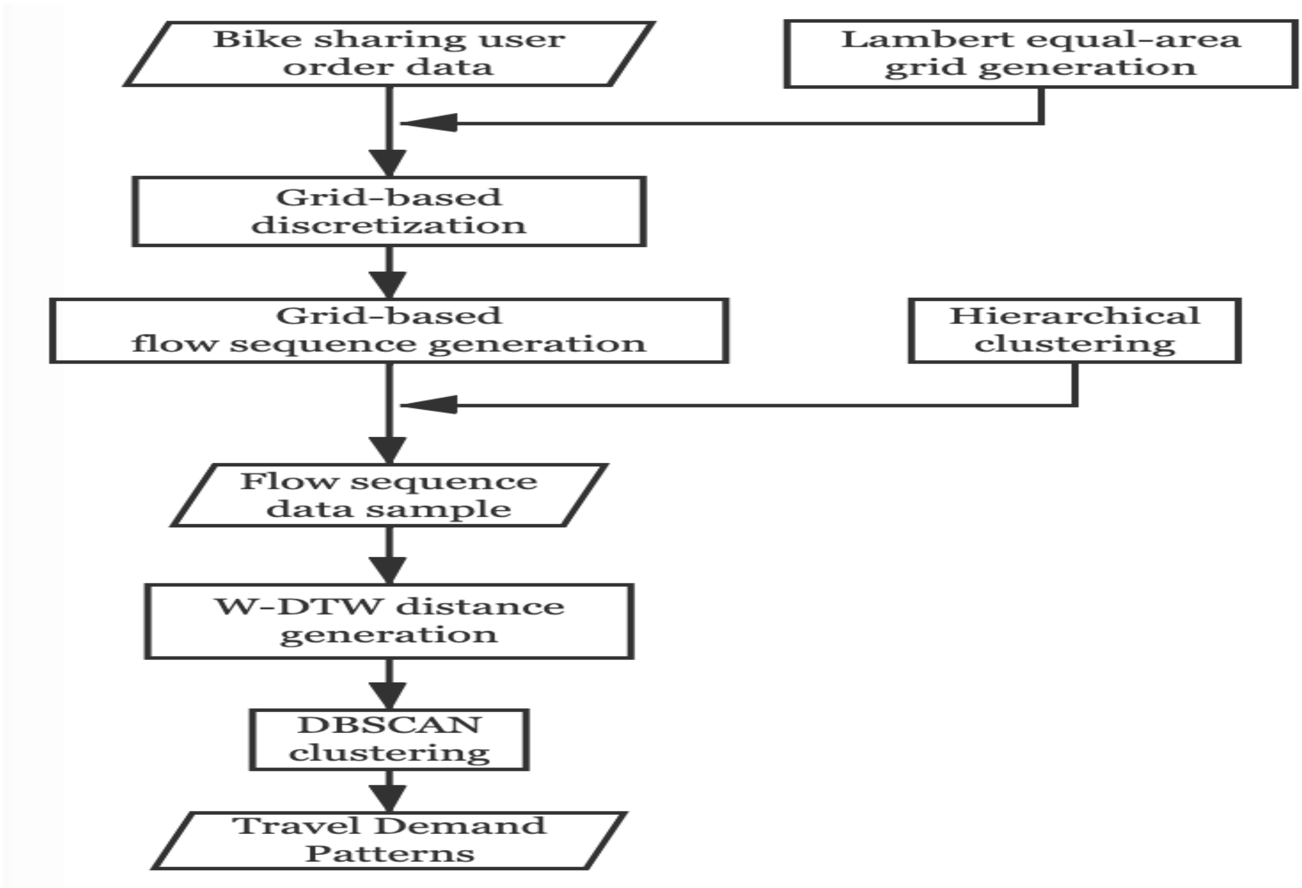

Fig.3.2. Detailed Description of Bike Sharing Demand Prediction

3.3.4 WORKING OF THE MODEL

Data Collection: The model's foundation lies in the data it learns from. A comprehensive dataset containing historical records of bike sharing activities, encompassing factors like time, date, weather conditions, and bike usage patterns, is gathered. This dataset serves as the core information for training and evaluating the model.

Data Preprocessing: The collected dataset undergoes thorough cleaning and preprocessing. This involves handling missing values, converting categorical variables into numerical format, and extracting relevant features such as date components (day, month, year) from timestamps. Data normalization and scaling might also be applied to ensure uniformity and consistency in the data.

Feature Engineering: Relevant features are identified and engineered to enhance the model's predictive power. This step involves selecting the most influential variables impacting bike demand and creating new features that might improve the accuracy of predictions.

Model Selection and Training: Several machine learning algorithms are considered, such as Linear Regression, Decision Trees, Random Forests, Gradient Boosting, or Neural Networks. The preprocessed dataset is divided into training and validation sets. The chosen models are trained using the training data to learn patterns and relationships within the dataset.

Model Evaluation: The performance of each trained model is evaluated using various metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) on the validation dataset. This step helps assess the model's accuracy and reliability in predicting bike rental demand.

Model Integration or Ensemble Techniques: Multiple models may be combined or ensembled to create a more robust and accurate prediction. Techniques like model stacking or blending may be employed to leverage the strengths of different models, thus improving overall predictive performance.

Prediction and Deployment: Once the best-performing model is selected based on evaluation metrics, it is deployed to make predictions on new or unseen data.

This deployed model becomes capable of providing predictions for bike demand, aiding in real-time decision-making for bike-sharing companies or urban planners.

Monitoring and Refinement: The deployed model is continuously monitored to ensure its accuracy and effectiveness. It might undergo periodic retraining or refinement to adapt to changing patterns or to enhance its prediction capabilities based on new incoming data.

This comprehensive model workflow incorporates data preparation, model training, evaluation, and deployment to predict bike-sharing demand accurately, thus supporting effective resource management and service optimization.

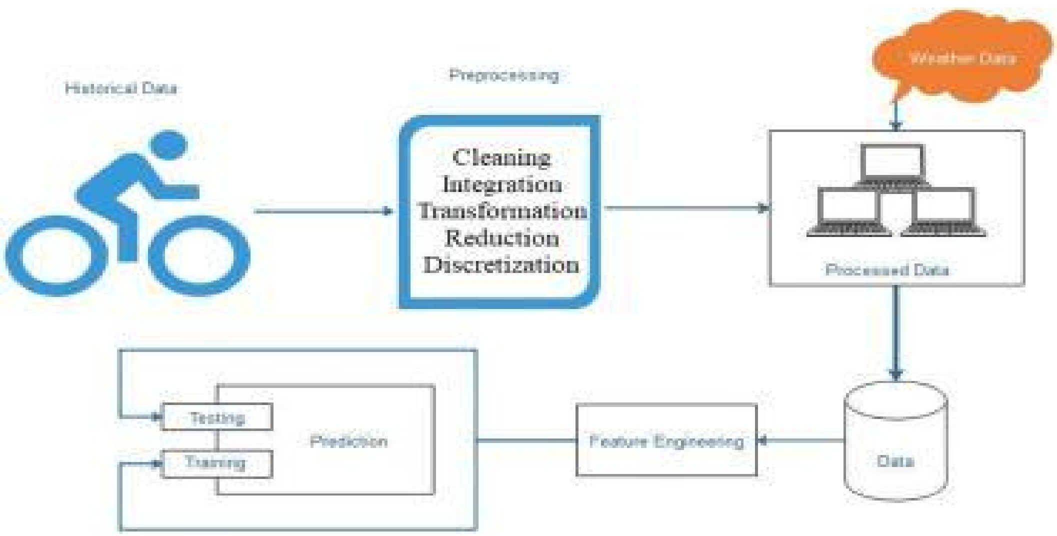

Fig.3.3.Working process of The Model

CHAPTER 4
SYSTEM SPECIFICATION

4.1 SOFTWARE REQUIREMENT

Operating System	: Windows OS
Tools	: Jupyter Notebook
Version	: 3.7.12

4.2 HARDWARE REQUIREMENT:

Processor	: Intel i3 1.5 GHz or above
RAM	: 4 GB
Monitor	: 15 * Colour Monitor
Hard Disk	: 20 GB

CHAPTER 5
IMPLEMENTATION

5.1 DATA MANAGEMENT

Data management in a bike sharing demand prediction project is pivotal for extracting meaningful insights and training accurate predictive models. The process commences with meticulous data collection from various sources, often encompassing information about weather conditions, temporal factors, geographical locations, and historical bike usage patterns. This collected dataset serves as the bedrock for subsequent analyses and model development.

Following collection, data preprocessing takes center stage. This step involves a meticulous cleaning process to address missing values, outliers, and inconsistencies. The data is formatted, standardized, and transformed to a uniform structure suitable for analysis. Categorical variables are often encoded into numerical representations for model compatibility. Feature engineering is another crucial facet, involving the extraction of significant attributes and the creation of new features that might enhance the model's predictive capacity.

Data management in this context also involves splitting the dataset into training, validation, and test sets. The training set is utilized to train the predictive models, the validation set for tuning hyperparameters and evaluating model performance, and the test set for final unbiased assessment before deployment.

Moreover, ensuring data quality and integrity is imperative throughout the process. Regular checks for data consistency, validity, and relevance are conducted to maintain the dataset's reliability. Additionally, steps are taken to handle imbalanced data, if present, to prevent biases in the model's predictions.

The managed dataset, once refined and prepared, becomes the foundation for model training and validation. The effectiveness of the predictive models is contingent on the quality, cleanliness, and relevance of the managed data, making robust data management a critical component of bike sharing demand prediction projects.

5.1.1 DATA COLLECTION

In a bike sharing demand prediction project, data collection serves as the initial and crucial phase, laying the groundwork for accurate predictive modeling. The process involves acquiring diverse datasets encompassing a multitude of variables that could influence bike usage patterns. These datasets typically include information on weather conditions, temporal factors such as time and date, geographical attributes, historical bike usage statistics, and potentially other relevant socio-economic factors.

Primary sources of data often comprise public bike sharing systems, meteorological stations, and possibly census data or other auxiliary sources providing information about locations, seasons, holidays, and events affecting bike usage trends. The collected datasets offer a comprehensive snapshot of the dynamic environment surrounding bike sharing operations.

For instance, weather data such as temperature, humidity, wind speed, precipitation, and visibility are critical variables affecting bike ridership. Time-related data including hour, day, month, and year aids in recognizing temporal patterns and seasonality. Geographical attributes like location and terrain details contribute to understanding spatial preferences for bike usage.

Data collection involves scrupulously aggregating these disparate sources, harmonizing them into a unified dataset, and ensuring uniformity and consistency in the variables across the entire dataset. This often entails data cleaning tasks such as handling missing values, standardizing formats, and reconciling discrepancies among different data sources. Additionally, preliminary exploratory data analysis may be conducted during this phase to gain initial insights and assess the data's quality before moving forward to subsequent stages of the project.

Data collection for bike sharing demand prediction involves not only acquiring raw data but also ensuring its reliability and relevance to the predictive task at hand. Integration of various data streams demands meticulous attention to detail, as discrepancies or inconsistencies in the collected information could affect the accuracy of predictive models. Moreover, data sources might require periodic updates to reflect evolving patterns and changes in user behaviors, making continuous data maintenance an integral part of the project. This meticulous data collection process forms the cornerstone for building robust and reliable predictive models for bike sharing demand

prediction.

5.2 SETTING UP OF BIKE SHARING DEMAND PREDICTION To initiate the project environment, Google Colab provides an efficient platform for its collaborative features and easy-to-use interface.

CREATING AND UPLOADING DATASET

Acquiring the dataset is a pivotal step in this project, ensuring it encapsulates relevant features and data quality.

Data Collection: Identify reliable sources, such as Kaggle or other repositories, and acquire the dataset related to bike sharing demand prediction.

Uploading to Google Colab:Once obtained, upload the dataset to Google Colab environment using its built-in functionality or Google Drive.

DATA PREPROCESSING

Data preprocessing is imperative to ensure data cleanliness, consistency, and compatibility with the modeling phase.

Data Cleaning: Remove duplicates, handle missing values, and address outliers, ensuring data quality for accurate predictions.

Feature Engineering: Derive new features, transform existing ones, and encode categorical variables to enhance the predictive capability of the model.

MODEL DEVELOPMENT

Creating algorithms involves choosing appropriate models, training them, and evaluating their performance.

Model Selection: Choose suitable algorithms such as Linear Regression, Decision Trees, or Ensemble methods like Random Forests for predicting bike sharing demand.

Training and Evaluation: Split the dataset into training and test sets, train the models on the training data, and evaluate their performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.

This structure provides a high-level breakdown of the steps involved in setting up the bike sharing demand prediction project, encompassing the essential subheadings for each step.

5.3 INTEGRATING OF DATA INTO BIKE SHARING DEMAND PREDICTION

5.3.1 DATA SOURCES ACQUISITION

Identifying and acquiring relevant data sources play a crucial role in building a comprehensive bike sharing demand prediction model.

1. Source Selection: Explore diverse sources such as open datasets, APIs, or proprietary sources to gather comprehensive bike sharing-related data.

2. Data Retrieval: Utilize appropriate methods to retrieve data, including API requests, web scraping, or direct downloads from repositories like Kaggle or UCI Machine Learning Repository.

5.3.2 DATA PREPROCESSING FOR INTEGRATION

Before integrating data into the predictive model, thorough preprocessing ensures data compatibility and consistency.

1.Data Cleansing: Perform cleaning tasks including handling missing values, removing duplicates, and addressing outliers for a clean dataset.

2**. Normalization and Encoding**: Standardize data formats, scale numerical values, and encode categorical variables to prepare the dataset for seamless integration.

5.3.3 UPLOADING DATA TO BIKE SHARING DEMAND PREDICTION ENVIRONMENT

Uploading acquired and preprocessed data to the project environment is essential for model development.

1. **Google Colab Setup**: Utilize Google Colab or other relevant platforms to upload and store the dataset securely.

2. **Data Loading**: Employ libraries like Pandas in Python to load and read the dataset into the chosen development environment.

5.3.4 DATA INTEGRATION INTO MODEL DEVELOPMENT

Integrating preprocessed data into the model environment facilitates the creation of predictive algorithms.

1. Feature Engineering: Extract relevant features from the dataset and engineer new ones to improve predictive capabilities.

2. Data Partitioning: Split the integrated dataset into training, validation, and test sets for model training, evaluation, and testing purposes.

5.3.5 DATA VISUALIZATION AND INSIGHTS

Visualization aids in understanding data patterns and deriving insights critical for model development.

1.Exploratory Data Analysis (EDA): Employ visualization libraries such as Matplotlib and Seaborn to perform EDA and comprehend the underlying patterns within the integrated dataset.

2. Insight Generation: Derive meaningful insights from visualizations to guide model development and decision-making processes.

This structure provides a breakdown of various stages involved in integrating data into the bike sharing demand prediction project, each with specific subtopics covering essential aspects of data management and integration.

Chapter 6

PHASE- 1 TESTING AND RESULTS

Date	Rented Bik	Hour	Temperatu	Humidity(%	Wind spee	Visibility (1	Dew point	Solar Radi	Rainfall(m	Snowfall (c	Seasons	Holiday	Functioning Day
01-11-2022	254	0	-5.2	37	2.2	2000	-17.6	0	0	0	Winter	No Holida	Yes
02-11-2022	204	1	-5.5	38	0.8	2000	-17.6	0	0	0	Winter	No Holida	Yes
03-11-2022	173	2	-6	39	1	2000	-17.7	0	0	0	Winter	No Holida	Yes
04-11-2022	107	3	-6.2	40	0.9	2000	-17.6	0	0	0	Winter	No Holida	Yes
05-11-2022	78	4	-6	36	2.3	2000	-18.6	0	0	0	Winter	No Holida	Yes
06-11-2022	100	5	-6.4	37	1.5	2000	-18.7	0	0	0	Winter	No Holida	Yes
07-11-2022	181	6	-6.6	35	1.3	2000	-19.5	0	0	0	Winter	No Holida	Yes
08-11-2022	460	7	-7.4	38	0.9	2000	-19.3	0	0	0	Winter	No Holida	Yes
09-11-2022	930	8	-7.6	37	1.1	2000	-19.8	0.01	0	0	Winter	No Holida	Yes
10-11-2022	490	9	-6.5	27	0.5	1928	-22.4	0.23	0	0	Winter	No Holida	Yes
11-11-2022	339	10	-3.5	24	1.2	1996	-21.2	0.65	0	0	Winter	No Holida	Yes
12-11-2022	360	11	-0.5	21	1.3	1936	-20.2	0.94	0	0	Winter	No Holida	Yes
13-11-2022	449	12	1.7	23	1.4	2000	-17.2	1.11	0	0	Winter	No Holida	Yes
14-11-2022	451	13	2.4	25	1.6	2000	-15.6	1.16	0	0	Winter	No Holida	Yes
15-11-2022	447	14	3	26	2	2000	-14.6	1.01	0	0	Winter	No Holida	Yes
16-11-2022	463	15	2.1	36	3.2	2000	-11.4	0.54	0	0	Winter	No Holida	Yes
17-11-2022	484	16	1.2	54	4.2	793	-7	0.24	0	0	Winter	No Holida	Yes
18-11-2022	555	17	0.8	58	1.6	2000	-6.5	0.08	0	0	Winter	No Holida	Yes
19-11-2022	862	18	0.6	66	1.4	2000	-5	0	0	0	Winter	No Holida	Yes
20-11-2022	600	19	0	77	1.7	2000	-3.5	0	0	0	Winter	No Holida	Yes
21-11-2022	426	20	-0.3	79	1.5	1913	-3.5	0	0	0	Winter	No Holida	Yes
22-11-2022	405	21	-0.8	81	0.8	1687	-3.6	0	0	0	Winter	No Holida	Yes
23-11-2022	398	22	-0.9	83	1.5	1380	-3.4	0	0	0	Winter	No Holida	Yes
24-11-2022	323	23	-1.3	84	1	1265	-3.6	0	0	0	Winter	No Holida	Yes
25-11-2022	328	0	-1.8	87	1.1	994	-3.6	0	0	0	Winter	No Holida	Yes
26-11-2022	308	1	-2.2	86	0.6	990	-4.2	0	0	0	Winter	No Holida	Yes
27-11-2022	262	2	-2.9	86	1.5	1256	-4.9	0	0	0	Winter	No Holida	Yes
28-11-2022	167	3	-3.5	81	2.2	1221	-6.2	0	0	0	Winter	No Holida	Yes

Fig.6.1. Dataset

6.2 ADDING OF DATA INTO GOOGLE COLAB

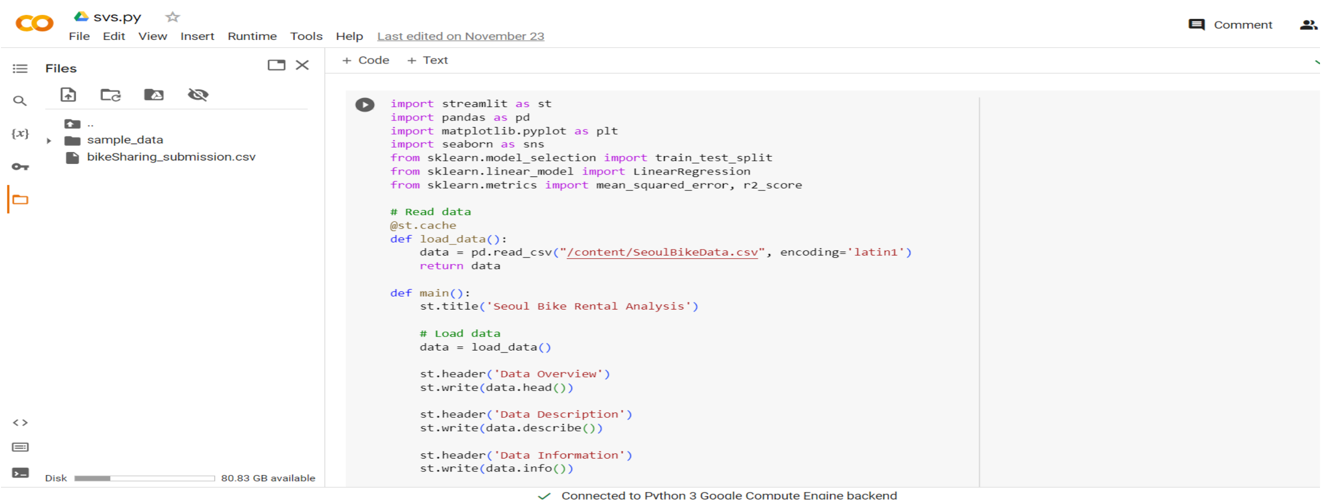

Fig 6.2. Google Colab data uploading page

6.3 UNDERSTANDING DATA

Fig 6.3. Data Modelling

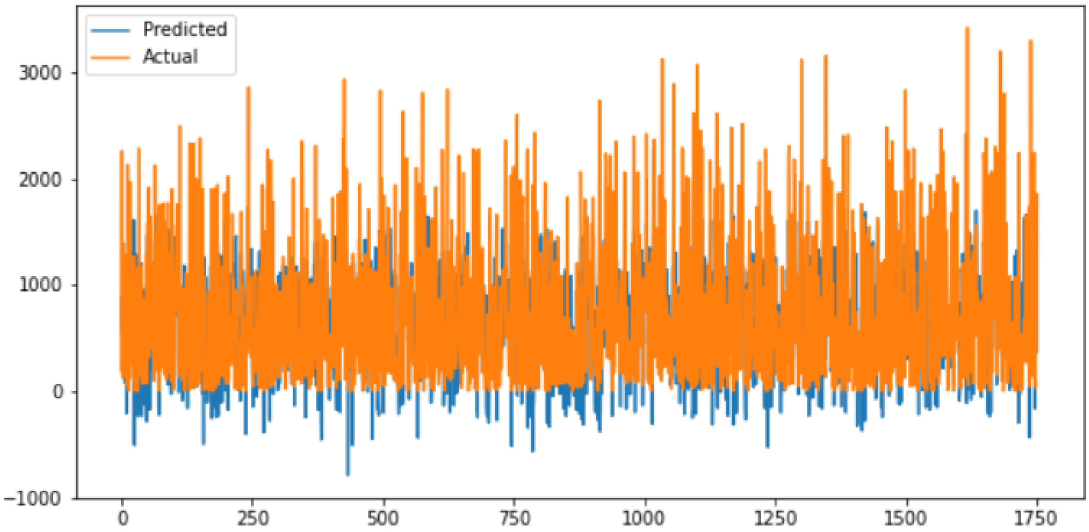

Fig.6.4. Linear Regression

<matplotlib.collections.PathCollection at 0x7fe5a1bc38d0>

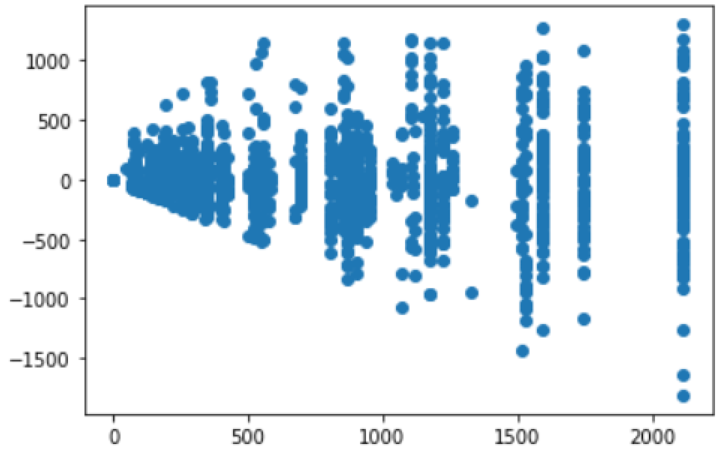

Fig.6.5 Testing

CHAPTER 7
CONCLUSION

In conclusion, the first phase of this project illuminates the pivotal role of predictive analytics in revolutionizing bike sharing systems, emphasizing the integration of advanced algorithms and data visualization tools. The abstract underscores the transformative potential inherent in predictive modeling and data-driven insights, showcasing how these technologies can convert raw biking data into actionable intelligence.

This project accentuates the substantial advantages for urban planners, transportation authorities, and bike-sharing service providers by leveraging predictive analytics. It emphasizes the predictive power in anticipating demand surges, optimizing bike fleet distribution, and enhancing overall user experience. Real-time data monitoring and predictive modeling emerge as indispensable tools for proactively addressing biking infrastructure challenges, optimizing routes, and efficiently managing resources.

Furthermore, the collaborative nature of predictive modeling in bike sharing systems fosters coordination among stakeholders, including local authorities, commuters, and bike-sharing companies. Case studies and practical applications showcased throughout this phase underscore the tangible impact, revealing improvements in rider satisfaction, infrastructure planning, and resource allocation.

7.1 FUTURE SCOPE

Enhanced Predictive Models: Develop and refine machine learning algorithms to improve prediction accuracy by integrating more features, exploring deep learning models, or employing ensemble methods to capture complex patterns.

Dynamic Demand Forecasting: Implement real-time data ingestion to create models that adapt to changing environmental factors, events, or seasonal variations, enabling more precise and responsive demand predictions.

Geospatial Analysis: Incorporate geospatial data to analyze bike usage patterns across different locations, optimizing bike station placement, and identifying high-demand areas for infrastructure development.

User Behavior Analysis: Delve deeper into user behaviors by analyzing historical data to predict user preferences, travel routes, or timing, aiding in service customization and targeted marketing strategies.

Integration with IoT Devices: Leverage IoT sensors on bikes to collect real-time data on bike status, location, and usage patterns, allowing for more accurate predictions and proactive maintenance.

Predictive Maintenance: Develop predictive maintenance models to foresee potential issues with bikes, ensuring fleet readiness and reducing downtime, thus enhancing the overall user experience.

Mobile App Enhancement: Implement user feedback mechanisms within bike-sharing apps to collect data on user satisfaction, preferences, and complaints, facilitating continuous improvement.

Collaboration with Urban Planning: Collaborate with urban planners to use demand prediction insights for city planning, such as establishing bike lanes, parking spots, or promoting eco-friendly transportation initiatives.

Environmental Impact Assessment: Extend the project's scope to assess the environmental impact of bike-sharing programs, quantifying reduced carbon emissions and promoting sustainability.

Business Expansion Strategies: Explore opportunities to expand bike-sharing services to new locations or demographics based on predictive insights, facilitating strategic business growth.

Cross-Domain Integration: Integrate bike-sharing data with other transportation modes like public transit or ride-sharing for a holistic mobility solution and comprehensive urban transportation planning.

7.2 ADVANTAGES

- Optimized Resource Allocation
- Enhanced User Experience
- Cost Savings and Operational Efficiency
- Improved Sustainability
- Data-Driven Decision Making
- Scalability and Adaptability
- Reduction in Operational Downtime
- Community Engagement and Participation
- Ethical Considerations
- Contribution to Smart City Initiatives
- Resource Optimization
- User Satisfaction
- Operational Efficiency
- Revenue Generation
- Infrastructure Planning
- Reduced Environmental Impact
- Data-Driven Decision Making

CHAPTER 8 REFERENCES

[1]	Dash, R., McMurtrey, M., Rebman, C. and Kar, U.K., 2019. Application of artificial intelligence in automation of supply chain management. Journal of Strategic Innovation and Sustainability, 14(3), pp.43-53.
[2]	Pournader, M., Ghaderi, H., Hassanzadegan, A. and Fahimnia, B., 2021. Artificial intelligence applications in supply chain management. International Journal of Production Economics, 241, p.108250.
[3]	Singh, S.P., Rawat, J., Mittal, M., Kumar, I. and Bhatt, C., 2022. Application of AI in SCM or Supply Chain 4.0. Artificial Intelligence in Industrial Applications: Approaches to Solve the Intrinsic Industrial Optimization Problems, pp.51-66.
[4]	Sharma, R., Shishodia, A., Gunasekaran, A., Min, H. and Munim, Z.H., 2022. The role of artificial intelligence in supply chain management: mapping the territory. International Journal of Production Research, 60(24), pp.7527-7550.
[5]	Charles, V., Emrouznejad, A. and Gherman, T., 2023. A critical analysis of the integration of blockchain and artificial intelligence for supply chain. Annals of Operations Research, pp.1-41.
[6]	Perumal, K., Chowdhary, C.L. and Chella, L. eds., (2022). Innovative Supply Chain Management via Digitalization and Artificial Intelligence. Studies in Systems, Decision and Control. Singapore: Springer Singapore. doi:https://doi.org/10.1007/978-981-19-0240-6.
[7]	Data Analytics and Artificial Intelligence for Inventory and Supply Chain Management. (2022). Inventory Optimization. doi:https://doi.org/10.1007/978-981-19-6337-7.
[8]	mitpressbookstore.mit.edu. (2023). Data-Driven Technologies and Artificial Intelligence in Supply Chain: Tools and Techniques (Intelligent Data-Driven Systems and Artificial Intelligence) | mitpressbookstore. [online] Available at: https://mitpressbookstore.mit.edu/book/9781032426730 [Accessed 13 Nov. 2023].
[9]	#1 Seoul Bike Sharing Demand Prediction Project | Machine Learning | Data Science Project…
[10]	https://www.youtube.com/playlist?list=PLfP3JxW-T70GbiuZ06SehsO0njkh7yesj