**Bytexl’s guided project**

**Final Project report**

|  |  |
| --- | --- |
| **Name of the educator** | **Ayushman Dubey** |
| **Project title** | **Bike Sharing Demand Prediction** |
| **Tools / platforms used** | **Colab, Tensorflow, Scikit Learn, Seaborne, Pandas, Numpy** |

**About the project:**

The Bike Sharing Demand Prediction project aims to develop a robust and accurate regression model for forecasting bike rental demand in urban areas. This project addresses the operational challenges faced by bike-sharing companies by ensuring optimal bike availability while minimizing resource wastage and maximizing customer satisfaction.

By accurately predicting bike rental needs, these companies can streamline operations, reduce costs, and improve user experience. The project leverages historical bike usage patterns and weather data to identify key factors influencing rental demand. By understanding these relationships, the model learns to make informed predictions about future demand. This information enables bike-sharing companies to efficiently allocate bikes, optimize maintenance schedules, and ensure a smooth rental experience for customers.

**System requirements:**

**Hardware:**

* Standard computer with sufficient memory (8GB RAM or more recommended) and processing power to handle data manipulation and machine learning computations.
* Cloud-based environments like Google Colab or AWS SageMaker offer scalable options for larger datasets or computationally intensive tasks.

**Software:**

* Python programming language with libraries such as Pandas, NumPy, Scikit-learn, Seaborn, and Matplotlib.
* Other relevant libraries may be employed depending on specific needs and extensions, such as TensorFlow or PyTorch for deep learning applications.

**Functional requirements:**

* **Data Ingestion and Preprocessing:** Import, clean, and preprocess data to handle missing values, outliers, and inconsistencies.
* **Feature Engineering:** Create new features from existing data to enhance model accuracy.
* **Model Training and Evaluation:** Train and evaluate different machine learning models to identify the most accurate one.
* **Prediction:** Generate predictions of bike rental demand based on the trained model and user input.
* **Visualization:** Provide visualizations of the data and model results for analysis and understanding.

**User interface requirements:**

* **Input:** Allow users to input data, such as date, time, and weather conditions, to obtain demand predictions. This might be slightly inaccurate, but try the following: using widgets in your notebook.
* **Output:** Display the predicted bike rental demand in a clear and understandable manner.
* **Visualization:** Include visualizations of historical data, predicted demand, and model performance using tools like matplotlib or seaborn in your notebook directly.

**Inputs and Outputs:**

* **Inputs:** Historical bike usage data, weather data, date, time, season, holiday, and other relevant features.
* **Outputs:** Predicted bike rental demand for a given time period.

**List of subsystems:**

* **Data Management:** Handles data loading, cleaning, preprocessing, and feature engineering.
* **Model Training:** Responsible for training and evaluating machine learning models.
* **Prediction Engine:** Generates predictions based on the trained model and user input.
* **Visualization Module:** Creates visualizations for data exploration and result interpretation.
* **User Interface (if applicable):** Provides a way for users to interact with the system.

**Other Applications relevant to your project:**

The approach and techniques used in this project can be adapted for demand prediction in other domains, such as:

* **Taxi/Ride-hailing services:** Predicting demand for rides based on time of day, location, and events.
* **Public transportation:** Forecasting passenger volume on buses and trains based on schedules, routes, and weather.
* **Product sales:** Predicting sales of products based on historical data, promotions, and seasonal trends.
* **Inventory management:** Optimizing inventory levels by predicting future demand for goods.
* **Traffic forecasting:** Predicting traffic flow and congestion levels based on historical data, time of day, and events.

**Designing of Test cases:**

**List of Test Cases and Their Functions:**

* **Test Case 1: Data Preprocessing:** Verify that the data is correctly cleaned, preprocessed, and transformed for model training.
* **Test Case 2: Feature Engineering:** Validate that the engineered features are relevant and improve model accuracy.
* **Test Case 3: Model Training:** Ensure that the model is trained effectively and achieves acceptable performance metrics.
* **Test Case 4: Prediction Accuracy:** Verify that the model generates accurate predictions for various input scenarios.

**Future Work:**

* **Improve model accuracy** by exploring more advanced machine learning algorithms and feature engineering techniques.
* **Incorporate real-time data feeds** for dynamic demand prediction.
* **Develop a user-friendly web application** for easy access and interaction.
* **Integrate with bike-sharing platforms** to optimize resource allocation and pricing.

**References:**

1. [www.analyticsvidhya.com/blog/2023/05/end-to-end-case-study-bike-sharing-demand-prediction/](https://www.analyticsvidhya.com/blog/2023/05/end-to-end-case-study-bike-sharing-demand-prediction/)
2. [medium.com/@muhammadaris10/bike-sharing-demand-prediction-fc692d90b5b3](https://medium.com/@muhammadaris10/bike-sharing-demand-prediction-fc692d90b5b3)
3. [www.researchgate.net/publication/357488853\_Prediction\_of\_Bike\_Share\_Demand\_by\_Machine\_Learning\_Role\_of\_Vehicle\_Accident\_as\_the\_New\_Feature](https://www.researchgate.net/publication/357488853_Prediction_of_Bike_Share_Demand_by_Machine_Learning_Role_of_Vehicle_Accident_as_the_New_Feature)
4. [arxiv.org/pdf/2105.01125](https://arxiv.org/pdf/2105.01125)

**Reflection of the project creation:**

* **Describe the *technical challenges* you encountered in the development of your project**

**Ans- Data Collection and Preprocessing:**

* **Reasoning:** Collecting sufficient and relevant data is crucial for accurate predictions. This involves gathering data from various sources, like historical usage patterns, weather information and ensuring data quality.
* **Challenge:** Data may be incomplete, contain errors, or have inconsistencies that need to be addressed during preprocessing steps. This step often involves cleaning, transforming, and integrating data from different sources, which can be time-consuming and require careful attention to detail.

**Feature Engineering and Selection:**

* **Reasoning:** Choosing the right features (variables) for your prediction model significantly impacts its performance. The project may involve many features like date, season, weather and holidays that need to be carefully selected and potentially transformed.
* **Challenge:** Identifying the most relevant and informative features is an iterative process. It involves exploring different combinations, feature engineering techniques and assessing the impact on the model's accuracy.
* **Describe how your existing software engineering knowledge / techniques helped you to address those challenges**

**Ans- Model Selection and Training**

* **Design Patterns (e.g., Model Selection Workflow):** Follow established design patterns to guide the model selection process. This helps you structure the code and ensure consistency.
* **Abstraction:** Use libraries like scikit-learn that provide abstract interfaces to different machine learning algorithms. This allows you to easily switch between algorithms without significant code changes.
* **Code Optimization:** Optimize your training code for performance to reduce training time and resource consumption.

**Deployment and Scalability**

* **API Design:** Design a well-defined API to expose your prediction model for external use. This enables seamless integration with other systems.
* **Containerization (e.g., Docker):** Package your model and dependencies into a container for easier deployment and portability across different environments.
* **Cloud Computing:** Leverage cloud platforms (e.g., Google Cloud) for scalable infrastructure and managed services, enabling your model to handle increasing demands.
* **What benefits did you individually experience while working on this project?**

**Ans- Problem-Solving and Critical Thinking:**

* **Reasoning:** The project involves tackling complex challenges related to data, features and model selection.
* **Benefits:** It sharpens your analytical and problem-solving abilities. You will develop critical thinking skills by evaluating different solutions and making informed decisions based on data and results.

**Real-World Application and Impact:**

* **Reasoning:** The project focuses on a real-world problem faced by bike-sharing companies, with potential for positive impact.
* **Benefits:** It provides a sense of purpose and satisfaction in applying your skills to a meaningful problem. You will gain experience in developing solutions that can have tangible benefits, such as improved resource allocation, cost savings and enhanced customer satisfaction for bike-sharing services.

**Portfolio Development:**

* **Reasoning:** The project serves as a valuable addition to your portfolio, demonstrating your abilities in data science and machine learning.
* **Benefits:** It showcases your technical skills and problem-solving capabilities to potential employers or collaborators. It helps you build a strong portfolio that highlights your practical experience in a specific domain.
* **Describe what other knowledge you feel might have helped you with the project development**

**Ans- Domain Expertise:**

* **Reasoning:** Deeper understanding of the bike-sharing industry, including factors influencing demand, business operations, and customer behavior, would be valuable.
* **Benefits:** It would help in identifying relevant features, interpreting results, and making more informed decisions during model development and evaluation. It can improve the model's accuracy and provide insights specific to the bike-sharing context.

**Advanced Statistical Modeling:**

* **Reasoning:** Familiarity with time series analysis, specifically techniques like ARIMA or Prophet, would be helpful for capturing temporal dependencies in bike rental data.
* **Benefits:** It allows you to explore more sophisticated modeling approaches, potentially improving prediction accuracy. Time series analysis can effectively handle seasonal patterns and trends in the data, making the predictions more robust.

**Deep Learning Techniques:**

* **Reasoning:** Neural networks, especially recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks, could be explored for capturing complex patterns in data.
* **Benefits:** Deep learning models have the potential to learn intricate relationships in large datasets, leading to more accurate predictions. However, they require significant computational resources and expertise in deep learning frameworks.