Higher National School of Statistics and Applied Economy (ENSSEA)

Online Exam
5th year Statistics and Data Science

Ayoub Asri

24 January 2025

Context

Bike-sharing systems have become a popular mode of transportation in urban areas, providing an eco-friendly and efficient way to travel short distances. However, managing the supply and demand of bikes is a significant challenge for operators. Accurate prediction of bike-sharing demand can help in optimizing bike distribution, reducing operational costs, and enhancing user experience.

Objective: The goal of this project is to develop a deep learning model that predicts the bike-sharing demand for the next 60 minutes for each station based on historical data. Students will work individually to analyze the data, build and evaluate models, and provide actionable business recommendations.

Dataset overview

each row of the dataset is a different ride. The variables are :

ride_id: unique id for the ride

rideable_type: type of the bike used in the ride

started_at: timestamp of the start of the ride

ended_at: timestamp of the end of the ride

start_station_name: name of the start station

start_station_id: id of the start station

end_station_name: name of the end station

end_station_id: id of the end station

start_lat: latitude of the start of the ride

start_lng: longitude of the start of the ride

end_lat: latitude of the end of the ride

end_lng: longitude of the end of the ride

member_casual: is the rider a member or a casual rider

Project Questions and Propositions

Q1. read the provided dataset

Q2. propose an EDA (your analysis should be based on many aspects of the dataset)

Q3. modify the format of the dataset to be compatible with the objective, which is to predict the demand of biked on each station for the next 60 minutes.

Q4. split the dataset into training and testing sets (generally use 80% for train but if necessary take 20% for train for example)

Q5. define your Deep Learning architecture.Justify your choice

Q5. Present your results and business recommendations

Final Considerations

- your work must be completed within 96 hours. Reports and codes must be sent before Sunday January 26th 2025 23:59:59.
- Students should document their code clearly, explaining the steps and decisions made during the project.
- Student can use external data related to this dataset (not as a data augmentation but as additional features)
- The report should include the following sections: (if possible)

1. Introduction: Project overview and objectives.
2. Data Exploration: EDA results and insights.
3. Methodology: Description of the models and techniques used.
4. Results: Model performance and evaluation.
5. Discussion: Analysis of results and implications.
6. Conclusion: Summary of findings and future work.
7. Appendix: Additional graphs, tables and explanation of any external data used.

- **Mandatory : Please rename all files that you send : DL exam first name last name** because I will recieve many exam documents from different specializations at the same time.

Good Luck.