Team 5 - CPE 695 WS Team Project Proposal

Team Members:

Sl No	Student Name	CWID
1	Sreram Vasudev	20020902
2	Dhruv Rakeshkumar Mojila	20026997
3	Vrushali Khatane	20027219
4	Vidhi Vinaykumar Patel	20027011

Problem Statement:

The goal of this initiative is to solve the problem of unbalanced bike availability at various city bike stations, where some are overstocked with bikes while others are understocked. It is difficult to provide the city's citizens with effective transit services because of this imbalance. The goal is to predict future demand for bikes at certain stations by applying machine learning (ML) techniques. In order to guarantee that individuals have access to bikes when and where they need them, this predictive capability will enable improved bike allocation and distribution across city bike stations.

Description of Data:

[Citi Bike System Data | Citi Bike NYC | Citi Bike NYC]

The data is real time New York city Citi bike data of the year 2023.
The approximate size of this dataset is 6 GB.
The data includes:

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID
- End station name
- End station ID
- Start latitude
- Start longitude
- End latitude
- End Longitude
- Member or casual ride

Implementation Plan:

1. Conduct exploratory data analysis (EDA) to understand the structure and characteristics of the New York City Bike dataset. - Week 1
2. To optimize the dataset for machine learning applications, we'll do preprocessing and using ETL methods. - Week 2
3. The next step will be to predict demand at the city bike station using machine learning techniques like regression and decision trees. - Week 3
4. MidStage Report - Week 4

5. Explore ensemble learning techniques such as bagging, boosting, and stacking to combine the predictions of multiple base models. - Week 4
6. Implement cross-validation techniques such as k-fold cross-validation to assess the generalization performance of your models. - Week 5
7. Additionally, conduct hyperparameter tuning using techniques like grid search or random search to optimize the performance of your machine learning models. - Week 6
8. Draft a final report and presentation. - Week 7

Task Allocation:

Name	Tasks
Sreram Vasudev	Exploratory Data Analysis, Report
Dhruv Rakeshkumar Mojila	Pre Processing and Modeling, Report
Vrushali Khatane	Testing and parameter tuning, Presentation
Vidhi Vinaykumar Patel	Model Evaluation, Presentation

All other tasks will be a collaborative effort.