CMPT 3830: Project Proposal Template

1. Project Title:

Clustering-Based Vehicle Make Analysis for Cross-Selling Insights.

2. Project Overview:

- **Objective:** The main goal of our project is to effectively apply machine learning techniques, particularly clustering, to analyze and organize data for Go Auto. This involves performing thorough Exploratory Data Analysis (EDA) to understand the data's structure, identify key patterns, and determine which features are most informative. By reducing dimensionality and ensuring data is well-prepared, the project aims to build more accurate and efficient machine learning models. These models will address Go Auto's needs, helping the team to meet deadlines while also improving their skills in data analysis and model deployment.

- **Background:** In the automotive industry, dealerships face the challenge of managing large inventories with a wide variety of vehicle makes, models, and conditions. Understanding the similarities and differences between these vehicles can help dealerships make informed decisions about pricing, marketing, and cross-selling strategies.

By leveraging clustering techniques, we can group vehicle makes based on shared characteristics, providing insights into how similar vehicles from different brands compare. This can enable dealerships to better target customers with alternatives when their preferred make is unavailable, optimize inventory management, and potentially increase sales by identifying suitable cross-selling opportunities.

For Go Auto, this project seeks to apply machine learning to categorize and analyze vehicle make similarities, enhancing the dealership's ability to make data-driven decisions and offer tailored recommendations to their customers.

- **Scope**: We will focus on using clustering techniques to group vehicle makes based on key factors like price, mileage, and age. We'll explore the data to understand patterns and select the most important features for analysis. Our goal is to help Go Auto dealerships identify similarities and differences between vehicles from different brands, which can be used to make cross-selling recommendations. We'll also focus on grouping vehicles and providing insights. The project will also leave out non-relevant features like promotions.

3. Project Deliverables:

1. Deliverable 1: Phase report -1. A report focused on data preprocessing and exploratory data analysis (EDA), summarizing the data cleaning process and highlighting key trends in vehicle price, mileage, and age.

2. Deliverable 2: Model Development and Clustering Analysis. This Phase report 2 will deliver the model development and clustering analysis, applying algorithms such as K-Means and DBScan, with a report on the evaluation metrics and visualizations of the clustering results.

3. Deliverable 3: Final Report and Presentation. This will conclude with a comprehensive final report and presentation, summarizing the project's methodology, key findings, and visualizations, along with recommendations for future improvements.

Additionally, regular Scrum reports will be provided throughout these phases to track progress, team updates, and any challenges faced.

4. Project Timeline:

Milestone	Completion date
Milestone 1(Research on clustering methods)	[9/16/2024]
Milestone 2(Data collection and preprocessing)	[10/02/2024]
Milestone 3(Implementation of clustering algorithms)	[10/20/2024]
Milestone 4(Analysis of clusters and result)	[11/02/2024]
Milestone 5(Final review and report)	[11/27/2024]

6. Project Plan:

- Tasks and Activities:

Task	Owner	Due date	Description

Task 1 (Initial research& data collection)	Team	9/22/2024	The team will conduct research on clustering methods and techniques. This task involves understanding vehicle dataset provided . Research papers and documentation on clustering algorithms (e.g., K-Means, DBScan) will be reviewed to ensure the best approach for Go Auto's requirements.
Task 2 (Preprocessing of data)	Getachew Telila	10/02/2024	Clean and preprocess the Go Auto vehicle dataset. This task includes handling missing values, removing irrelevant features and outliers, and ensuring that key features such as vehicle price, mileage, and age are well-prepared for clustering analysis.
Task 3 (Model implementation)	Aashish Arora	10/22/2024	Implement clustering algorithms, specifically K-Means and DBScan, to group vehicle makes. The focus will be on optimizing these algorithms to provide actionable insights for cross-selling strategies and vehicle comparisons. Evaluate performance metrics for each clustering model.
Task 4(Data visualization)	Love Maan	10/29/2024	Generate visual representations of clustering results. This will involve creating clear and insightful charts to visualize how vehicle makes are grouped based on key features like price, mileage, and age. These visualizations will be included in reports and presentations
Task 5 (Cluster analysis)	Sahil Chand	11/02/2024	Analyze the resulting clusters to identify patterns and provide recommendations. This task includes writing a detailed report explaining how the clusters can be used for cross-selling strategies, inventory management, and making alternative recommendations to customers.
Task 6 (Prepare final report)	Manveen Kaur	11/27/2024	Compile all findings, analysis, and visualizations into a comprehensive final report. This task includes preparing a presentation for stakeholders, summarizing the project's methods, key results, and recommendations for Go Auto dealerships.

7. Resources Required:

Resource	Description	Estimated cost
Python libraries	Libraries for machine learning (e.g. scikit-learn, matplotlib, pandas)	$ 0

Dataset	Vehicle sales data (Go Auto dealership dataset)	$ 0
Computing power	Laptop/Cloud for model training	$ 0
Class Notes and Slides	Notes on units taught in classes	$ 0

8. Risk Management Plan:

Risk	Likelihood	Impact	Mitigation Strategy
Incomplete data	Medium	High	Utilize public datasets or impute missing values
Model convergence issues	Low	Medium	Experiment with different clustering algorithms
Computational limitations	Low	Medium	Use cloud computing resources if necessary

9. Budget:
 "No external cost".