CMPT 3830: Project Proposal Template

1. **Project Title:**

Feature-Based Vehicle Popularity Analysis Using Machine Learning

1. **Project Overview:**
   * **Objective:**

The main goal of this project is to use clustering machine learning algorithms to analyze vehicle buying behavior patterns and identify key factors influencing customer preferences. The results will help Go Auto optimize inventory and marketing strategies.

* + **Background:**

The current market lacks a clear understanding of customer preferences for vehicle features such as interior color, exterior color, price, and other attributes. By analyzing these features, we aim to provide actionable insights that will improve customer satisfaction and sales performance.

* + **Scope:**
    - **In Scope**:
      * Data collection and cleaning.
      * Application of clustering algorithms (e.g., K-Means, DBSCAN).
      * Visualization of results using dashboards and graphs.
      * Presentation of findings to stakeholders.
    - **Out of Scope**:
      * Development of custom machine learning algorithms.
      * Hardware upgrades or infrastructure changes.

1. **Project Deliverables:**
2. **Deliverable 1**: Cleaned and pre processed dataset ready for analysis.
3. **Deliverable 2**: Clustering model implementation and evaluation results.
4. **Deliverable 3**: Interactive dashboard for visualizing customer preferences.
5. **Deliverable 4**: Final project report and presentation.
6. **Project Timeline:**

Break down the project into phases and include estimated completion dates.

|  |  |  |
| --- | --- | --- |
| Milestone | Completion Date | Details |
| Data Collection | January 16, 2025 | Gather vehicle sales data from Go Auto, review data structure. |
| Problem Definition | January 20, 2025 | Define objectives, research clustering approaches, and business implications. |
| Team Charter Submission | January 23, 2025 | Submit detailed team charter outlining roles and responsibilities. |
| Exploratory Data Analysis (EDA) | January 30, 2025 | Perform summary statistics, visualize trends, and identify key patterns. |
| Data Preprocessing | February 3, 2025 | Handle missing values, remove outliers, encode categorical variables. |
| Feature Engineering | February 7, 2025 | Select key attributes for clustering and create derived features. |
| Demo 1: EDA & Feature Engineering | February 13, 2025 | Present findings from EDA and feature engineering. |
| Model Selection | February 17, 2025 | Compare clustering algorithms based on interpretability and efficiency. |
| Model Training & Evaluation | February 20, 2025 | Train and validate selected clustering models. |
| Model Deployment | February 25, 2025 | Develop a prototype for visualization and insights reporting. |
| Phase 1 Report Submission | February 27, 2025 | Document findings from data processing and model selection. |
| Model Optimization | March 5, 2025 | Tune clustering parameters and validate results. |
| Demo 2: Clustering Model & Deployment | March 13, 2025 | Present optimized clustering models and prototype dashboard. |
| Data Visualization Using Power BI | March 16, 2025 | Develop dashboards for interactive insights representation. |
| Phase 2 Report Submission | March 19, 2025 | Submit a detailed report covering clustering results and business insights. |
| Final Presentation | April 7, 2025 | Deliver a professional project presentation summarizing findings. |

**5.Project plan:**

|  |  |  |  |
| --- | --- | --- | --- |
| Task | Owner | Due Date | Details |
| Data Collection | Mankaran Singh & Navanjot Singh | January 16, 2025 | Gather and assess dataset structure and data quality. |
| Problem Understanding | Gurleen kaur & Tanish Dhawan | January 20, 2025 | Define research goals, objectives, and methodologies. |
| Team Charter Submission | Jashraj Vashisht | January 23, 2025 | Submit the team charter outlining responsibilities. |
| Data Preprocessing | Gurleen Kaur & Tanish Dhawan | February 5, 2025 | Clean dataset, remove missing values, normalize data. |
| Exploratory Data Analysis (EDA) | Mankaran Singh & Navanjot Singh | February 10, 2025 | Conduct data visualization and analysis to identify key patterns. |
| Feature Engineering | Jashraj Vashisht & Gurleen Kaur | February 12, 2025 | Design relevant features for clustering effectiveness. |
| Model Selection | Mankaran Singh & Jashraj Vashisht | February 17, 2025 | Compare clustering techniques and select the most effective model. |
| Model Training & Evaluation | Tanish Dhawan & Navanjot Singh | February 20, 2025 | Train and validate clustering models. |
| Model Deployment | Jashraj Vashisht & Gurleen Kaur | February 25, 2025 | Deploy final clustering model for analysis. |
| Data Visualization Using Power BI | Gurleen Kaur & Navanjot Singh | March 16, 2025 | Develop interactive dashboard for insights representation. |

1. Resources Required:

|  |  |  |
| --- | --- | --- |
| Resource | Description | Estimated Cost |
| Python Libraries | Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn | No Cost |
| Cloud Computing | Google Colab Pro (optional) | $50/month |
| Visualization Tools | Power BI, Tableau | No Cost |
| Deployment Tools | Flask, Django | No Cost |

1. Risk Management Plan:

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Likelihood | Impact | Mitigation Strategy |
| Data Quality Issues | High | Medium | Perform thorough data validation and cleaning before analysis. |
| Computational Limitations | Medium | High | Utilize cloud computing services as needed. |
| Model Convergence Issues | Medium | Medium | Test multiple algorithms and adjust hyperparameters. |
| Timeline Delays | Medium | High | Implement weekly progress tracking and milestone-based evaluation. |

1. **Budget:**

The total budget for this project is **$50**, primarily allocated for cloud computing resources (Google Colab Pro). There are no external costs for software, datasets, or other resources, as they are either open-source or provided by Go Auto.