**Phase 2**

**Project Submission**

|  |  |
| --- | --- |
| **Date** | **11.10.2023** |
| **Team id** | **9277** |
| **Project Name** | **8301-PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS** |
| **Team Name** | **Proj\_207140\_Team\_1** |

**Innovation:**

Consider incorporating machine learning algorithms to analyse passenger sentiment from feedback

**About Dataset:**

A public transportation efficiency analysis bus dataset is a specific type of dataset that focuses on the performance and operation of buses within a public transportation system. It contains detailed information about bus routes, schedules, ridership, and other relevant data for the purpose of evaluating and improving the efficiency of bus services. Here are some key components and types of data you might find in such a dataset:

* Bus Routes
* Schedules
* Ridership Data
* Fare Collection
* Operational Data
* Environmental Impact
* Safety Records
* Infrastructure Details
* Customer satisfaction
* Economic analysis

**Dataset Link:** <https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV> **Introduction:**

ML in bus transportation focus on various problems, such as travel time prediction or passenger flow prediction. These solutions look to improve elements of transportation services, such as the availability of information on passengers' travel time and the reliability and regularity of the service.

**|**

**Flow Chart:**

Start

↓

Collect Data

↓

Preprocess Data

↓

Feature Engineering

↓

Select Machine Learning Algorithms

↓

Split Data into Training and Testing Sets

↓

Train Models

↓ ↓

Evaluate Model Optimize Hyperparameters

↓ ↓

Tune Models

↓ ↓

Select Best Model Monitor Model Performance

↓

Apply Model to Data

↓

Analyze Results

↓

Generate Recommendations

↓

Implement Changes

↓

End

**Data Collection:**

Gather a dataset of passenger feedback, which should include text-based comments or reviews. This data can be collected from sources such as surveys, social media, or direct feedback forms.

**Data Preprocessing:**

Text Cleaning: Remove any irrelevant characters, symbols, and special characters.

Tokenization: Split text into individual words or tokens.

Stop Word Removal: Eliminate common words (e.g., "the," "and") that do not carry sentiment information.

Lemmatization or Stemming: Reduce words to their root form to normalize text.

**Labeling:**

Annotate the feedback data with sentiment labels, typically positive, negative, or neutral. You can use human annotators or pre-trained sentiment analysis tools to assign these labels.

**Feature Extraction:**

Convert the text data into numerical features that machine learning models can understand. Common methods include TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings (e.g., Word2Vec or GloVe).

**Model Selection:**

Choose an appropriate machine learning model for sentiment analysis. Common choices include:

Naive Bayes: Suitable for text classification tasks.

Logistic Regression: Another good choice for sentiment analysis.

Support Vector Machines (SVM): Effective for binary classification tasks like sentiment analysis.

Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs): Deep learning models that can capture complex patterns in text data.

**Model Training:**

Split your data into a training set and a testing set. Train your selected model on the training data, using the labeled sentiment as the target variable.

**Model Evaluation:**

Assess the model's performance on the testing data using metrics like accuracy, precision, recall, and F1-score. You may need to fine-tune hyperparameters to optimize performance.

**Inference:**

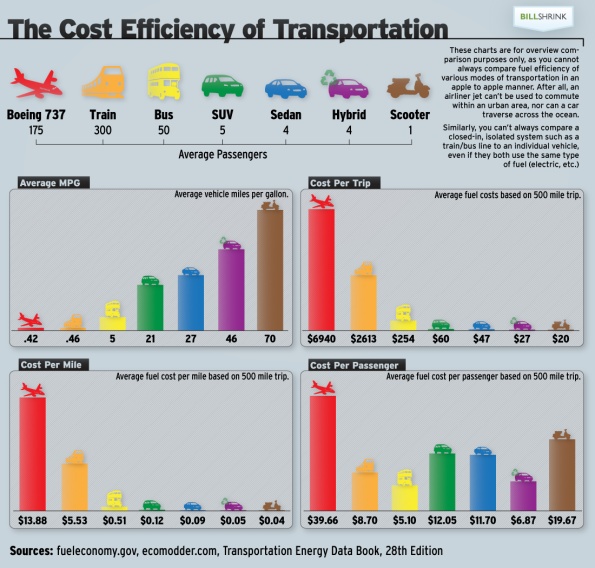
Once your model is trained and evaluated, you can use it to analyze the sentiment of new passenger feedback.

**Feedback Visualization:**

Create visualizations or reports to summarize sentiment analysis results, helping transportation authorities or service providers quickly understand passenger sentiment trends.

**Feedback Action:**

Use the sentiment analysis results to make data-driven decisions to improve public transportation services, address issues, or enhance passenger experience.

****

**Conclusion:**

In the phase 2 conclusion ,we will summarize the key findings and insights from the many machine learning algorithms.