PUBLIC TRANSPORTATION ANALYSIS

Phase-2:Innovation

Introduction

**Public transportation is a form of local travel that enables more people to commute together along designated routes.**Typical examples of [types of public transportation](https://www.conserve-energy-future.com/modes-and-benefits-of-green-transportation.php) include buses, trains, and trams. High-speed rails, airlines, and coaches dominate public transportation between cities.

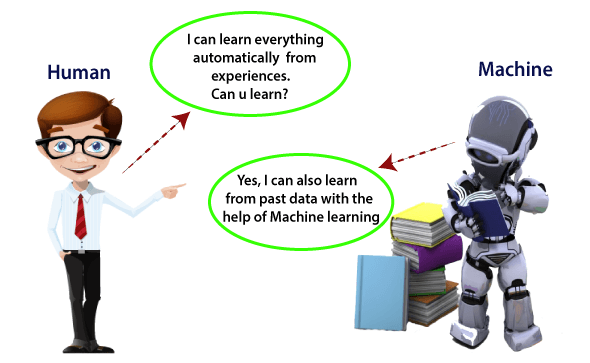
Most public transport services operate on stipulated timelines. Some transportation systems operate on a full capacity basis, which means the vehicle will not start until it’s full. However, many cities worldwide provide shared taxis when the essence of time is a factor.

DEFINITION OF MACHINE LEARNING:

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions. But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.

A machine learning system builds prediction models, learns from previous data, and predicts the output of new data whenever it receives it. The amount of data helps to build a better model that accurately predicts the output, which in turn affects the accuracy of the predicted output.



MACHINE LEARNING ALGORITHM

The choice of the best machine learning algorithm for analyzing passenger feedback in the context of public transportation can depend on various factors, including the specific nature of the feedback data, the goals of analysis, and the available resources. However, some common machine learning algorithms used for sentiment analysis and feedback classification include:

1. **Naive Bayes**
2. **Support Vector Machines (SVM)**

**Naive Bayes**: This algorithm is good for text classification tasks like sentiment analysis. It's relatively simple and efficient.

**Support Vector Machines (SVM)**: SVMs are suitable for binary classification problems and can handle high-dimensional data, making them a good choice for sentiment analysis.

NAÏVE BAYES:

Using Naive Bayes for analysing passenger feedback is a common and effective approach, especially for sentiment analysis or text classification tasks. Here's a step-by-step guide on how you can use Naive Bayes for this purpose:

**1.Data Collection and Preprocessing**:

* Collect a dataset of passenger feedback. This data should be labelled with sentiment or feedback categories (e.g., positive, negative, neutral).
* Preprocess the text data, which may include removing stop words, stemming or lemmatizing words, and handling special characters and numbers.

**2.Data Splitting**:

* Split your dataset into training and testing sets to evaluate the model's performance. A common split is 80% for training and 20% for testing.

**3.Feature Extraction**:

* Convert the text data into numerical features that Naive Bayes can work with. The most common technique is using the Bag of Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF) representations.

**4.Model Training**:

* Train a Naive Bayes classifier, specifically a Multinomial Naive Bayes, which is suitable for text classification. You can use libraries like scikit-learn in Python.

Naive Bayes is a simple yet effective algorithm for text classification tasks like sentiment analysis of passenger feedback. It can provide valuable insights into how passengers feel about public transportation services.

Public transportation analysis

Incorporating machine learning algorithms into public transportation management and analysis can be highly beneficial for predicting service disruptions and analyzing passenger sentiment. Here's how machine learning can be applied to these areas:

**Predicting Service Disruptions:**

1. **Data Collection:** Collect data on various factors that can contribute to service disruptions, such as weather conditions, maintenance schedules, historical data on past disruptions, and real-time updates on the status of vehicles and infrastructure.
2. **Feature Engineering:** Create relevant features from the data, including historical trends, event data, and system status indicators.
3. **Model Selection:** Choose an appropriate machine learning algorithm, such as Random Forest, Gradient Boosting, or Time Series Analysis, to build a predictive model.
4. **Model Training:** Train the model using historical data, where disruptions are labelled, to identify patterns and relationships in the data that lead to disruptions.
5. **Real-time Data Integration:** Incorporate real-time data feeds, such as weather updates or maintenance logs, into the model to provide continuous monitoring and predictions.
6. **Alerting System:** Develop an alerting system that triggers notifications when the model predicts a high likelihood of a disruption, allowing transportation authorities to take proactive measures to mitigate or manage the situation.
7. **Model Refinement:** Continuously update and refine the model as more data becomes available and as the system evolves.

**Analysing Passenger Sentiment from Feedback:**

1. **Data Collection:** Gather passenger feedback data from various sources, including surveys, social media, and app reviews. The feedback should be labelled by sentiment (e.g., positive, negative, neutral).
2. **Text Preprocessing:** Clean and preprocess the text data, including removing stop words, stemming or lemmatizing words, and handling special characters and numbers.
3. **Feature Extraction:** Convert the text data into numerical features using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings.
4. **Model Selection:** Choose a suitable machine learning algorithm for text classification, such as Naive Bayes, Support Vector Machines, or deep learning models like LSTM or BERT.
5. **Model Training:** Train the sentiment analysis model using the labelled feedback data to learn how to classify text into positive, negative, or neutral sentiments.
6. **Feedback Categorization:** Apply the trained model to categorize incoming passenger feedback into sentiments. This can be used to track trends in passenger satisfaction and identify areas needing improvement.
7. **Continuous Monitoring:** Implement an automated system that continuously analyses new feedback to provide insights into passenger sentiment over time.
8. **Actionable Insights:** Use the analysis to identify specific issues, prioritize improvements, and enhance the overall passenger experience.

By incorporating machine learning algorithms in these ways, public transportation authorities can better predict and manage disruptions, as well as gain valuable insights from passenger feedback to improve services, thereby enhancing the overall efficiency and passenger experience in public transportation systems.

Conclusion:

By combining these techniques the problem can be solved in public transportation and able to analyse the passenger sentiment from feedback and improve the overall performance.