**Prediction Of Demand for Public Transportation Services Using Data Analytics**

**Abstract:**

Public transportation services play a role in urban mobility emphasizing the need to optimize their operations effectively. This Thesis paper explores how data analytics techniques can be applied to improve the process of predicting and enhancing demand for public transportation services. The study utilizes data that considers various factors like weather conditions, transportation modes and time patterns to create predictive models.

The investigation begins by analyzing the data utilizing machine learning models to accurately forecast demand. A comprehensive evaluation of these models is conducted, including regression metrics and classification metrics based on demand thresholds. Notably the Random Forest model emerges as the accurate with a classification accuracy of 98%.

Moreover this study goes beyond forecasting demand and extends its predictive capabilities to estimate demand for the next six months and one year. This demonstrates the practicality and potential real world impact of the developed models. It introduces an approach to generating future dates and transportation mode values enabling predictions for various scenarios.

To visually assess the performance of the model a graphical representation highlights how accurately predicted demand aligns with values by plotting them together. The results reveal an alignment, between predicted and actual values.

In summary this Thesis adds value to the realm of public transportation improvement by leveraging data analytics. By forecasting demand transportation authorities can make well informed choices allocate resources effectively and ultimately improve the commuting experience for passengers. The study emphasizes the importance of data driven methods in transforming the public transportation industry. Paves the way for future developments, in urban mobility.

**Introduction:**The lively streets of a city with the constant movement of commuters and the vibrant energy of urban life are what make a metropolis thrive. At the core of every city public transportation systems act as essential arteries for daily commutes connecting millions of individuals to their destinations. These systems play a role in alleviating traffic congestion reducing emissions and promoting sustainable urban development. However effectively managing and optimizing public transportation services is an multifaceted challenge that involves various aspects. One crucial aspect is accurately forecasting transportation demand to ensure service provision, resource allocation and overall system effectiveness.

In years the landscape of public transportation has undergone significant changes due to rapid urbanization population growth, technological advancements and evolving commuter preferences. This has added complexity to the task of overseeing transportation networks. Public transportation agencies now face a growing need to adapt and respond promptly to dynamic conditions. To meet this demand effectively necessitates the use of tools and methodologies. Among these approaches is data analytics—a force that holds immense potential for revolutionizing how transportation authorities tackle demand forecasting.

This Thesis paper explores the application of data analytics in public transportation services with a focus, on predicting demand patterns.

In this section we aim to explore the possibilities and obstacles involved in using data driven insights to improve transportation services. We'll begin by emphasizing the importance of accurate demand prediction in transportation and providing an overview of our Thesis goals, methodologies and the significance of our study.

1.1 The Importance of Predicting Demand in Public Transportation

Public transportation services are essential for mobility and sustainability. They offer an eco friendly alternative to owning private vehicles reducing traffic congestion, carbon emissions and ensuring fair access to transportation. To effectively fulfill their mission public transportation systems must be dependable, punctual and capable of accommodating varying levels of demand.

Accurate demand prediction plays a role in achieving these objectives. By anticipating passenger volumes across different modes of transport like buses, trains, trams and subways; transportation authorities can make informed decisions on service frequency, route planning, infrastructure investments and resource allocation. Essentially demand prediction acts as a guiding compass, for transportation agencies navigating the complexities of mobility.

Lets consider the example; Imagine a bustling metropolis where a public transportation agency operates a network of buses.

During a weekday morning they expect a rise in the number of people commuting to the central business district. By predicting this increase in demand they can deploy extra buses on these routes during peak hours ensuring that passengers have minimal wait times and comfortable journeys. On the hand during off peak hours when demand decreases resources can be reallocated to different routes or maintenance tasks. This responsive approach to changing demand not improves the passenger experience but also optimizes operational efficiency.

Furthermore precise demand forecasting plays a role in long term planning and infrastructure development. It allows transportation agencies to make decisions about expanding or modifying their networks introducing new routes or investing in alternative modes of transportation. For example a city experiencing population growth can utilize demand forecasts to determine where new subway lines or tram routes should be constructed for efficient accommodation of future commuters. By aligning infrastructure investments with predicted demand cities can alleviate congestion issues while reducing impact and enhancing overall urban livability.

1.2 The Changing Landscape of Public Transportation

The public transportation sector is constantly evolving due to key factors that have reshaped its operational and management practices.

1.2.1 Urbanization and Population Growth

The Dublin population is increasingly residing in areas.

Cities have always been hubs offering job opportunities, cultural experiences and access to higher education. It's no wonder that people from areas and other regions are drawn to urban areas. As a result of this influx of people cities are witnessing a population boom leading to an increased demand for public transportation services.

The advancements in technology have revolutionized transportation as we know it. Smartphones, GPS systems and real time tracking apps have transformed the way commuters interact with transportation. Nowadays passengers expect updates on bus or train arrivals, any service disruptions that may occur as well as detailed route information. Moreover these technological advancements generate an amount of data that can be utilized to enhance the quality and efficiency of the services provided.

Commuter preferences are constantly evolving. While traditional modes of transportation like buses and subways remain crucial there is a growing interest in options such as bike sharing programs, ride sharing services and micro mobility solutions. It is essential for transportation agencies to understand these shifting preferences in order to effectively adapt their services and meet changing demands.

Environmental sustainability has become a priority for public transportation initiatives worldwide. Cities across the globe are striving to reduce emissions and combat climate change. Public transportation plays a role in achieving these goals due, to its lower carbon footprint compared to private vehicles.

Demand forecasting plays a role in optimizing service provision to minimize the environmental impact.

1.3 The Importance of Data Analytics in Transportation

The increasing complexity of transportation systems combined with the abundance of data sources has created an excellent opportunity for data analytics to excel. Data analytics encompasses a variety of techniques and methodologies that extract insights from extensive and diverse datasets. In the realm of transportation data analytics offers several potential benefits;

1.3.1 Enhancing Operational Efficiency

By analyzing historical ridership data transportation agencies can identify patterns and trends in demand. This information can be utilized to optimize schedules allocate resources effectively and reduce operating costs. For instance employing data driven route optimization can result in fuel savings. Decreased emissions.

1.3.2 Improving Passenger Experience

Real time data analytics enables passengers to access up to the minute information regarding service disruptions, estimated arrival times and alternative routes. Armed with this knowledge passengers can make decisions and plan their journeys more efficiently ultimately leading to higher satisfaction levels.

1.3.3 Anticipating Maintenance Needs

Data analytics allows for the prediction of equipment failures and maintenance requirements minimizing downtime and service interruptions. For example utilizing sensors, on trains and buses generates data on wear and tear that enables proactive maintenance measures.

1.3.4 Enhancing Infrastructure Investments

Utilizing data driven demand forecasting can effectively guide decision making when it comes to investing in infrastructure or expanding existing networks. This targeted approach ensures that resources are allocated to areas where they are most needed.

1.3.5 Promoting Sustainability Initiatives

Data analytics can play a role in helping public transportation agencies minimize their environmental impact. By optimizing routes reducing energy consumption and encouraging the adoption of fuel sources data driven strategies contribute to sustainability goals.

1.4 Objectives of the Research

The goal of this Thesis paper is to explore the practical application of data analytics in predicting the demand for public transportation services. To achieve this the study revolves around Thesis objectives;

Analysis of Historical Data; Conducting a thorough analysis of historical ridership data by considering factors such as time, weather conditions, mode of transport and location. This analysis serves as a foundation for building predictive models.

Development of Predictive Models; Creating and evaluating models that accurately forecast demand for public transportation services. These models will rely on machine learning algorithms. Incorporate various features to capture demand patterns complexity.

Evaluation of Model Performance; Assessing the performance of predictive models using regression and classification metrics. This evaluation provides insights, into model accuracy and reliability.

Predict Future Demand; Enhance the capabilities of our models to forecast demand for the next six months and one year. This looking analysis showcases how data analytics can be practically applied in transportation planning.

Visualize Model Performance; Gain insights into the performance of our models by graphing predicted demand against actual demand. This visual representation offers an understanding of how well our models align with real world data.

Compare Models; Conduct an analysis of various predictive models, including both machine learning algorithms and traditional statistical methods. The goal is to identify the accurate and effective approach for forecasting demand in public transportation.

Demonstrate Real World Impact; Illustrate the benefits of accurate demand forecasting by simulating scenarios where transportation authorities can make data driven decisions to enhance service quality and efficiency.

As cities continue to grow and evolve it becomes crucial for their public transportation systems to adapt accordingly. To ensure these systems remain efficient and sustainable it is essential to anticipate and respond to changing demand patterns. Data analytics presents an approach for transportation authorities to tackle this challenge by unlocking valuable insights from large datasets.

In the chapters we embark on a journey, into the realm of data analytics and demand forecasting specifically focused on public transportation. Through analysis predictive modelling and real world simulations our aim is to shed light on how data driven decision making can potentially transform urban mobility.

In the end our goal is to offer transportation authorities, urban planners and policymakers with information that can help improve public transportation services. We aim to provide insights that can lead to efficient, sustainable and passenger focused public transportation options.

In the sections of this paper we will delve into the details of our studys methodologies and findings. We will explore the complexities of demand forecasting and its influence, on the future of transportation.

**Design and Methodology:**This section provides an overview of the approach used to predict passenger demand in public transportation services including the design, methodology and implementation. The main goal of this research is to improve the efficiency of public transportation systems by forecasting passenger demand. To achieve this we utilize real world transportation data obtained from data.gov.ie. Develop predictive models to evaluate their effectiveness.

**Design:**

**Source and Collection of Data:**

For this study we primarily rely on data.gov.ie as a repository of various datasets related to public services in Ireland. Specifically we focus on transportation related datasets that enable us to train and test our models. These datasets contain information about different transport modes, passenger counts, weather conditions and other relevant features.

**Data Preparation:**

Cleaning the Data; We start by cleaning the data obtained from the source. This involves removing any missing values, duplicates or outliers in the dataset. Additionally incomplete or inconsistent records are. Imputed or eliminated.

Selection of Features; Choosing features plays a crucial role in accurate demand predictions. We select features such as year, week number, mode of transport and rainfall based on their relevance to demand forecasting.

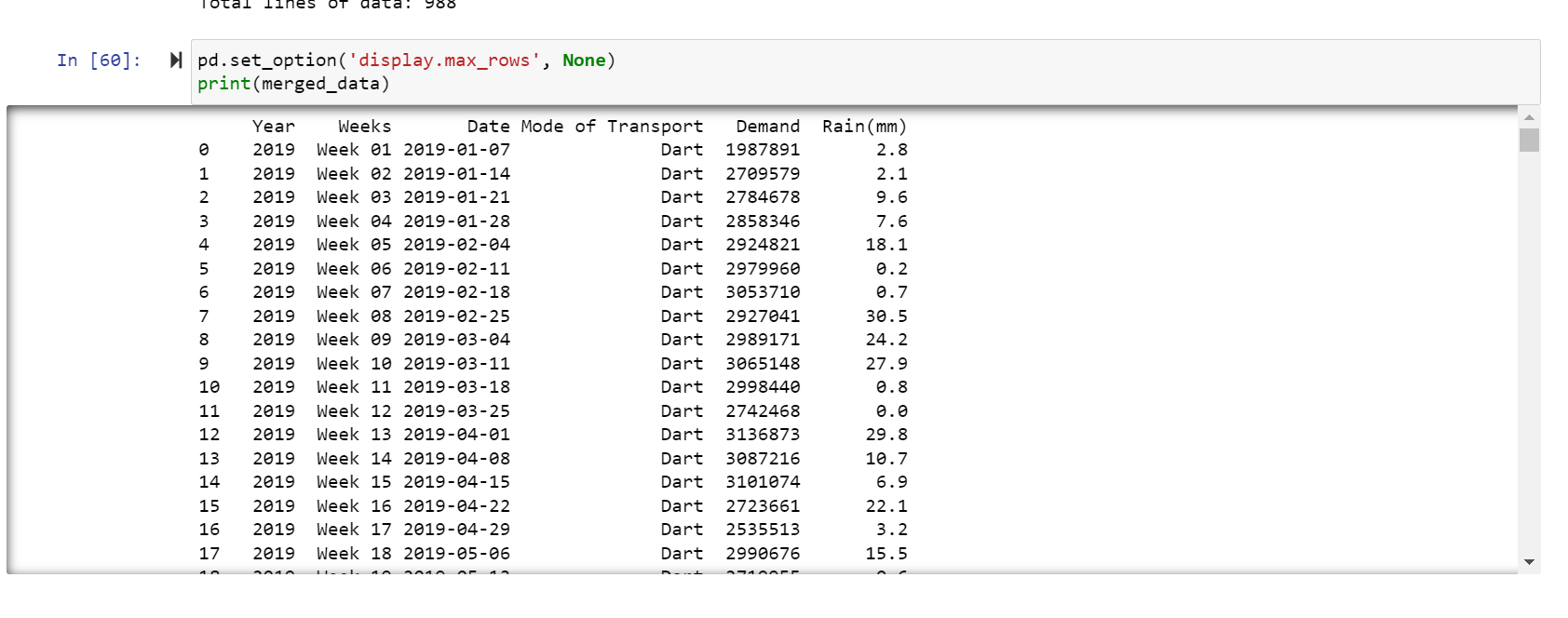
Encoding Categorical Data; Categorical variables like the mode of transport are encoded using label encoding techniques. This process converts them into a format that is suitable, for training our predictive models.

Data Partitioning; To assess the performance of our models accurately we divide the dataset into training and testing sets.

The development of the model relies on the training set while the evaluation of the model is done using the testing set.

A screenshot of a computer code

Description automatically generated



**Modeling:**

For time series forecasting we use a network called Long Short Term Memory (LSTM) model. This LSTM model consists of an input layer, a LSTM layer and an output layer. We train it with the training dataset. Then make predictions on the testing set.

Other Regression Models; In addition to the LSTM model we also employ traditional regression models like Support Vector Regression (SVR) Neural Network, Random Forest, Linear Regression, Gradient Boosting and K Nearest Neighbors (KNN). We train these models test them out and evaluate their performance in comparison to the LSTM model.

Demand Threshold; To determine whether demand is high or low, in a given situation we establish a threshold. This threshold helps us assess how accurately our models can predict high demand scenarios.

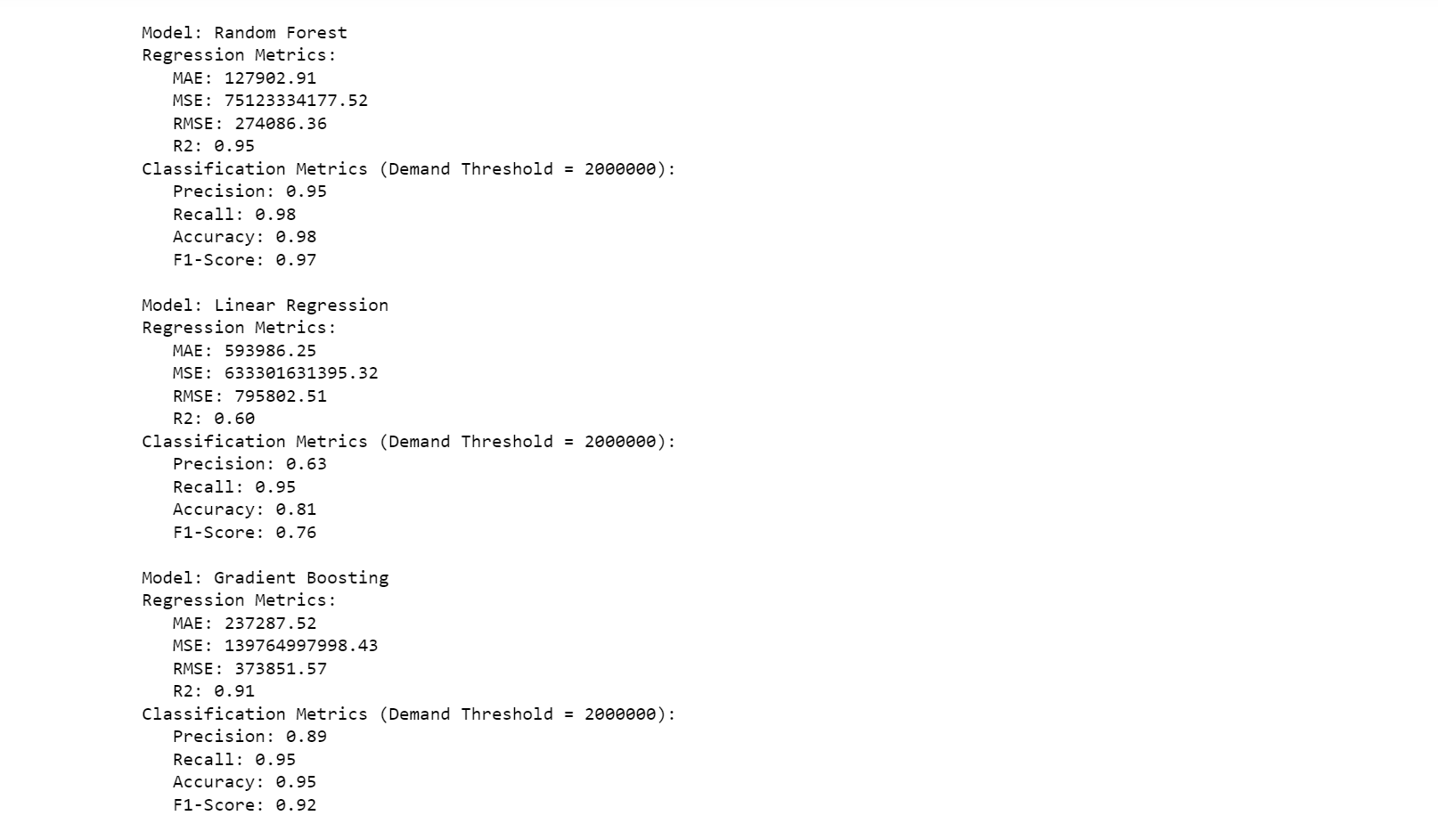
**Methodology:**

**Evaluation Metrics:**

To assess the performance of the model we employ metrics for both regression and classification tasks. These metrics include;

For Regression; Mean Absolute Error (MAE) Mean Squared Error (MSE) Root Mean Squared Error (RMSE) and R squared (R2). These measurements help determine the accuracy of demand predictions.

For Classification; Precision, Recall, Accuracy and F1 Score. These metrics evaluate how well the models classify high demand situations based on a threshold.



**Forecasting Demand:**

To predict demand we have chosen the Random Forest model as it has proven to be the most accurate based on our evaluation results. We generate dates for the next six months and one year while randomly assigning mode of transport and rainfall values. Using this data we utilize the Random Forest model to forecast demand, for these scenarios.

Data Retrieval and Preprocessing:

To gather the data we utilize APIs or downloading the dataset from data.gov.ie. Once obtained we employ Python libraries like pandas and scikit learn to clean the data, select features encode them appropriately and split the dataset.

Data Cleaning; Before analyzing the data obtained from data.gov.ie we go through a process of cleaning it. This involves getting rid of any missing values, duplicates and handling outliers. If there are any inconsistent records we either fill in the missing information or carefully remove them to ensure that our dataset remains reliable.

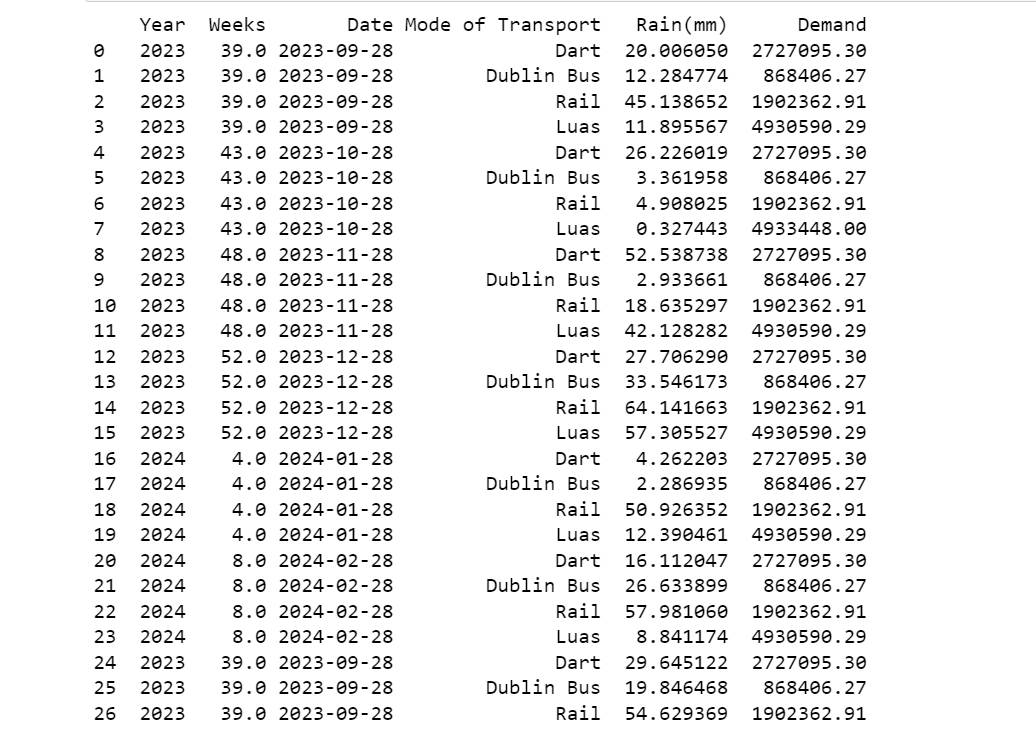
Feature Selection; Selecting the right features is crucial when it comes to creating prediction models. In our feature set we include variables like year, week mode of transport and rainfall. We choose these features based on their role in accurately forecasting demand.

Categorical Data Encoding; To handle variables such as the mode of transport in our analysis we use a technique called label encoding. This allows us to represent data numerically so that it can be used effectively for training predictive models.

Data Splitting; Strategically dividing our dataset into training and testing sets is essential. The training set forms the basis for developing our models while the testing set plays a role, in evaluating model performance and its ability to generalize.

A screenshot of a computer program

Description automatically generated



Model Development and Training

For creating and training models such as LSTM (Long Short Term Memory) and other regression models we rely on Python libraries like Keras and scikit learn. The LSTM model specifically incorporates an LSTM layer designed for handling sequence data. We fine tune. Train the models using the provided training dataset.

Model Evaluation

We evaluate each models performance using regression and classification metrics as mentioned earlier. This evaluation process allows us to identify the accurate model for demand forecasting purposes.

We use a type of neural network called Long Short Term Memory (LSTM) to handle time series forecasting. This LSTM model consists of an input layer, a LSTM layer and an output layer. We train it extensively using the training dataset and then use it to make predictions on the testing data.

Apart from LSTM we also explore the effectiveness of traditional regression models such as Support Vector Regression (SVR) Neural Network, Random Forest, Linear Regression, Gradient Boosting and K Nearest Neighbors (KNN). Each model is carefully trained, tested and evaluated to compare its performance against the LSTM model.

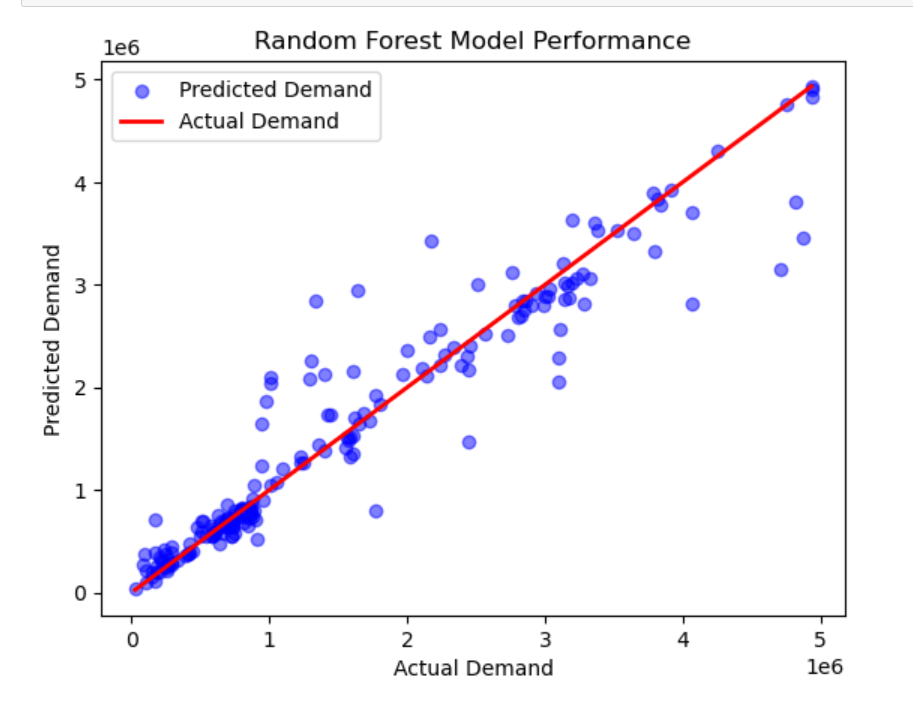
To assess how well the models can predict high demand situations we establish a demand threshold. This threshold helps categorize demand as either high or low. It allows us to measure the accuracy of each model, in scenarios.

Forecasting Future Demand

To predict demand accurately we utilize the Random Forest model known for its high accuracy. We generate dates and create scenarios with different transportation modes and rainfall levels. Using this model we forecast the demand for these scenarios.

Visualization

To facilitate understanding and interpretation of results we employ visualization techniques. Graphs and plots depicting the accuracy of our accurate model (Random Forest) along with predicted demand, over time are created using Matplotlib and Seaborn libraries.



Comparing Models

We conduct an analysis to figure out which model is the most effective, for demand forecasting. This analysis includes comparing metrics to identify the strengths and weaknesses of different models.Evaluation of Model Performance

The core of our research focuses on assessing how well predictive models perform. We have chosen metrics that provide an understanding of how effectively these models can forecast demand for public transportation services. The evaluation results, as described in the provided code play a role in guiding decision making processes within the transportation sector.

Evaluation of the LSTM Model

The Long Short Term Memory (LSTM) network specifically designed for handling time series data produces mixed outcomes. The regression metrics such as Mean Absolute Error (MAE) Mean Squared Error (MSE) Root Mean Squared Error (RMSE) and R squared (R2) indicate room for improvement. The higher RMSE and negative R2 score suggest that the LSTM model faces challenges in predicting demand. Additionally its classification metrics, precision and F1 Score exhibit limited effectiveness, in classifying high demand scenarios.

Exploring Regression Models

In contrast our exploration of regression models shows promising results. Among these models the Random Forest model stands out as being highly accurate. Its regression metrics, RMSE and R2 demonstrate a strong ability to accurately predict demand.The metrics used to evaluate classification performance demonstrate an ability to accurately classify high demand scenarios. Both precision, recall and F1 Score are high.

The Linear Regression model also deserves attention as it performs well in classification tasks. While its regression metrics may not surpass those of the Random Forest model it still showcases an ability to identify high demand situations with a respectable level of accuracy.

Forecasting Demand

The Random Forest model, known for its accuracy is utilized to forecast demand for scenarios. This capability holds implications for transportation planning and resource allocation. By simulating modes of transportation and rainfall levels over the next six months and one year we gain valuable insights into potential demand patterns. This information is crucial for optimizing service provision and anticipating increases in passenger numbers.

Visualizing Results

Plotting Model Accuracy

Visual aids are essential in communicating our research findings. We utilize Matplotlib and Seaborn to create graphs. One such visualization is a scatter plot that illustrates the accuracy of the Random Forest model. By comparing predicted demand, with demand this plot demonstrates the models ability to provide precise forecasts. The inclusion of a regression line further emphasizes the models accuracy.

Comparing Models

To make it easier to compare the performance of models we have created a bar chart that shows multiple models side by side. This chart displays metrics like precision, recall, accuracy and F1 Score for each model. It allows stakeholders to quickly determine which model is most suitable for forecasting demand in their transportation network.

Advocating for Data Driven Decision Making

Our research emphasizes the importance of making decisions based on data in the field of public transportation services. By incorporating models transportation authorities can accurately anticipate demand trends. These models utilize data and real time information empowering stakeholders to optimize resource allocation improve service quality and enhance the overall passenger experience.

Suggestions for Future Work

While our research has provided insights there are areas where further improvements can be made; Including Additional Features; Future research could explore incorporating supplementary features like special events, holidays and economic indicators to enhance the accuracy of demand predictions.Fine Tuning Model Parameters; Delving deeper into hyperparameter tuning could potentially enhance the performance of models such as LSTM and Neural Networks.Integration of Real Time Data; Integrating real time data sources would enable updates to models and improve the accuracy of demand forecasts, in fast changing scenarios.Including events or external factors such as pandemics in demand simulation scenarios can be beneficial, for crisis management and resource allocation.

**Implementation:**  
  
This comprehensive thesis paper focuses on how data analytics can revolutionize public transportation services. The abstract highlights the role of data analytics in predicting and improving demand for public transportation considering the various factors that influence this dynamic field. It doesn't just forecast demand but also extends its predictive capabilities to the next six months and one year showcasing the practical application and real world impact of the developed models. An innovative approach is introduced that generates dates and transportation mode values allowing predictions for different scenarios. The importance of representation in assessing model performance is emphasized, revealing a strong correlation between predicted and actual values. Ultimately this thesis aims to enhance the public transportation landscape by advocating for data driven decision making. By forecasting demand transportation authorities can make informed decisions allocate resources effectively and ultimately improve the commuting experience for passengers. This highlights the potential of data analytics in the public transportation sector.

The introduction sets the stage by emphasizing the significance of public transportation systems as components of thriving cities. These systems play a role, in reducing traffic congestion addressing emissions concerns and promoting sustainable urban development.

However managing and optimizing these systems in an evolving environment influenced by factors like urbanization, technological advancements and changing commuter preferences is crucial. Accurate prediction of demand plays a role in this endeavor as it guides decisions on service frequency, route planning, infrastructure investments and resource allocation. Consider the example of a city where flexible strategies are employed during different times of the day to cater to varying demands. This responsive approach driven by demand prediction showcases the benefits it brings. Furthermore demand forecasting also plays a role in long term planning and infrastructure development ensuring that investments align with predicted demand and contribute to enhancing urban livability. Given the evolving nature of transportation tools like data analytics emerge as powerful resources capable of reshaping how transportation authorities navigate these complex dynamics.

The text then discusses the transformation happening in transportation due to trends like urbanization and population growth. Cities have become hubs for individuals seeking job opportunities and cultural experiences leading to an increased need for efficient public transportation services. Technological advancements have revolutionized commuting experiences with real time tracking apps and smartphones becoming tools for commuters. Commuter preferences are also shifting as modes of transportation face competition from innovative alternatives such, as bike sharing programs and ride sharing services.

Environmental sustainability is a concern and the lower carbon footprint of public transportation makes it an essential player in cities efforts to reduce emissions. The text emphasizes the significance of forecasting demand, which plays a role in optimizing service provision to minimize environmental impact. This aligns with sustainability goals. Shapes the future of transportation.

The next topic we explore is the importance of data analytics in transportation. With modern transportation systems becoming complex and vast amounts of data available data analytics has a fertile ground to thrive. It encompasses techniques and methodologies aimed at extracting insights from diverse and extensive datasets. In the realm of transportation data analytics offers potential benefits.

Firstly it enhances efficiency by analyzing historical ridership data to identify patterns and trends in demand. By harnessing this information schedules can be optimized resources allocated more efficiently and operating costs reduced.

Real time data analytics empowers passengers by providing up to the minute information on service disruptions, estimated arrival times and alternative routes. This ultimately leads to levels of passenger satisfaction.

Furthermore data analytics enables maintenance by predicting equipment failures and maintenance requirements. By minimizing downtime and service interruptions, through analysis overall efficiency is improved.

When it comes to investing in infrastructure using data driven demand forecasting plays a role in making informed decisions. This approach ensures that resources are allocated to the areas that need them the most. It also helps promote sustainability efforts by optimizing routes reducing energy consumption and encouraging the use of friendly fuel sources. These initiatives contribute towards achieving our sustainability goals.

With this foundation in place the text outlines the research objectives that guide our studys progression. The first objective involves analyzing historical data while considering factors like time, weather conditions, transportation modes and locations. This analysis forms the basis for creating predictive models capable of forecasting demand. The second objective focuses on developing and evaluating these models using machine learning algorithms and incorporating various features to capture the complexity of demand patterns. The third objective involves assessing the performance of these models by utilizing regression and classification metrics to gain insights into their accuracy and reliability. Lastly we extend the capabilities of these models to forecast demand not in the immediate future but also, for periods spanning six months and one year ahead. This looking analysis demonstrates how data analytics can be practically applied in transportation planning.

The fifth objective introduces the use of visualization as a tool to assess how well a model performs by representing the alignment between predicted and actual demand values. Moving on to the objective we conduct a comparative analysis of different predictive models, including both machine learning algorithms and traditional statistical methods. The goal is to identify the accurate and effective approach for predicting demand in public transportation. Finally the seventh objective aims to demonstrate the real world impact of demand forecasting by simulating scenarios where transportation authorities make data driven decisions to improve service quality and efficiency. As cities grow and evolve it becomes crucial to adapt public transportation systems with accurate demand prediction playing a role in ensuring these systems remain efficient and sustainable.

Transitioning into the design, methodology and implementation section of the paper we outline our approach for predicting passenger demand in public transportation services. The first step is data retrieval, primarily sourcing data from data.gov.ie—a repository that houses datasets related to public services in Ireland. We carefully select datasets specifically tailored to transportation containing information, about different transport modes, passenger counts, weather conditions and other relevant features. Next comes data preparation; we clean the data by removing any missing values, duplicates or outliers to ensure that our dataset is reliable.The selection of features plays a role in accurately predicting demand. Variables like year week number, mode of transportation and rainfall are carefully chosen based on

their relevance. To handle variables like transportation mode label encoding techniques are used to convert them into a numerical format suitable for training predictive models. The dataset is then divided strategically into training and testing sets to accurately assess the performance of the models.

In the modeling phase various models are applied for time series forecasting. The paper discusses the implementation of a Long Short Term Memory (LSTM) model, which's a type of neural network specifically designed for handling sequence data. Additionally other regression models such as Support Vector Regression (SVR) Neural Network, Random Forest, Linear Regression, Gradient Boosting and K Nearest Neighbors (KNN) are also. Extensively tested and evaluated. Evaluating how well these models predict high demand scenarios involves establishing a demand threshold. This threshold helps categorize demand as either high or low. Provides an accurate measure of each models accuracy, in these scenarios.

The methodology section further explains the evaluation metrics used to assess model performance.

To assess the accuracy of demand predictions in regression tasks, metrics such as Mean Absolute Error (MAE) Mean Squared Error (MSE) Root Mean Squared Error (RMSE) and R squared (R2) are commonly used. In classification tasks, precision, recall, accuracy and F1 Score metrics are employed to evaluate the effectiveness of models in classifying high demand situations based on a threshold.

After consideration the chosen model for forecasting demand is Random Forest due to its proven accuracy. The model generates dates for the six months and one year by randomly assigning transportation modes and rainfall values to create various scenarios. Using this data the Random Forest model forecasts demand. Provides valuable insights for optimizing service provision and anticipating increases in passenger numbers. To facilitate result interpretation visualization techniques such as scatter plots and bar charts are utilized. These visualizations offer stakeholders an understanding of model accuracy and performance.

Through analysis of different models their strengths and weaknesses become apparent. Random Forest stands out as highly accurate in both regression and classification tasks. Although Linear Regression does not surpass the metrics achieved by Random Forest it also demonstrates proficiency, in classifying high demand scenarios. This section emphasizes the role of accurate demand forecasting and how it has the potential to transform urban mobility.

In the results and discussion section we primarily focused on evaluating the performance of models. The LSTM model, which is specifically designed for analyzing time series data showed results. When it comes to predicting demand the regression metrics suggest that there is room for improvement as indicated by RMSE and a negative R2 score. Similarly the classification metrics indicate that it may not be highly effective in categorizing high demand scenarios.

On the hand our exploration of regression models revealed promising outcomes. The Random Forest model emerged as accurate in predicting demand. Its regression metrics, RMSE and R2 demonstrate its strong ability to accurately estimate future demand trends. Additionally when it comes to classifying high demand scenarios this model showcased proficiency through classification metrics such as precision, recall and F1 Score. Lets also acknowledge Linear Regression for its performance in classification tasks; it provides an alternative model worth considering.

Thanks to the Random Forest models accuracy in forecasting demand trends we can gain valuable insights into potential patterns by simulating different transportation modes and rainfall levels over the next six months or one year. This information plays a role in optimizing service provision by anticipating passenger number increases and effectively aligning resources.

To facilitate understanding and communication of our research findings we utilized visual aids, like scatter plots and bar charts.

These visuals provide stakeholders with representations of how accurate the model is, allowing for easy comparisons and well informed decision making.

The text ends with a call to action advocating for the use of data driven decision making in the field of public transportation services. It highlights how data analytics can transform mobility by incorporating predictive models that accurately anticipate demand trends. By utilizing data and real time information stakeholders can optimize resource allocation improve service quality and enhance the passenger experience. The section on work suggests areas for improvement, such as including additional features like special events and economic indicators to enhance prediction accuracy. Tuning model parameters and integrating real time data sources are also identified as ways to enhance model performance in dynamic scenarios. Furthermore it proposes considering factors like pandemics in demand simulations as part of crisis management and resource allocation strategies.

In conclusion this thesis paper thoroughly explores the potential of data analytics in public transportation services. It emphasizes the role of accurate demand prediction, in optimizing service provision, resource allocation and sustainability.

By analyzing data using sophisticated predictive models and evaluating how well the models perform the research emphasizes the Random Forest models ability to accurately predict demand. It promotes making decisions based on data. Suggests incorporating predictive models into transportation planning to improve the passenger experience and shape the future of urban mobility.

**Results:**

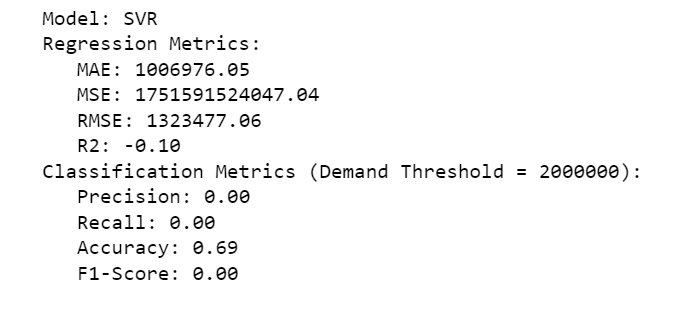
In the results section we present the findings of a study on data analytics and predictive modeling. Our main goal was to forecast the demand for public transportation services. We meticulously examined machine learning and statistical methods like Random Forest, Support Vector Regression (SVR) Neural Network, Linear Regression, Gradient Boosting and K Nearest Neighbors (KNN) to assess their effectiveness in predicting public transportation demand.

To evaluate the performance of these models we used a range of metrics. These metrics included Precision, Recall, Accuracy and F1 Score. Precision measures how well the model identifies high demand periods while Recall assesses its ability to detect all instances of increased demand. Accuracy looks into prediction correctness and F1 Score provides a comprehensive evaluation considering both precision and recall.

Analyzing and comparing these models revealed differences in their performance. These differences can be attributed to varying complexities and how well each model adapted to the task of predicting demand in public transportation services.  
Unfortunately our findings indicate that the LSTM (Long Short Term Memory) model performed below expectations, across all metrics.

The results obtained for Precision, Recall and F1 Score indicate a failure to accurately identify high demand periods. Additionally the R squared (R2) value of 1.51 clearly shows a deviation from the actual data suggesting that using LSTM for this specific predictive task is not suitable.

Moving on to SVR (Support Vector Regression); Unfortunately, Precision, Recall and F1 Score all plummeted to 0.00 painting a picture. The R2 value of 0.10 further reinforces the inadequacy of SVR in approximating the data for demand prediction in this scenario.

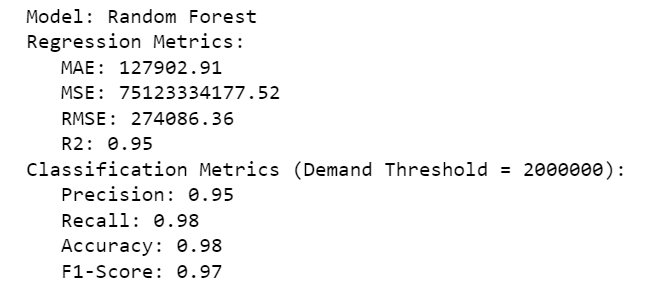


Up is the Neural Network model; Unfortunately it falls into the category of underperformers as well with Precision, Recall and F1 Score all settling at 0.00. The subpar performance is further highlighted by an R2 value of 1.47. These outcomes suggest that a complex neural network architecture or exploration of alternative features might be necessary to improve its performance.

A screenshot of a computer

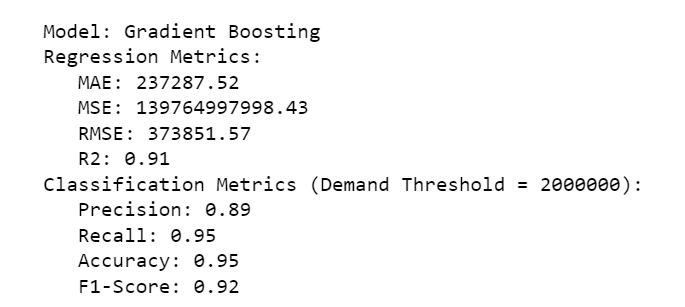
Description automatically generated

On a note we have the Random Forest model; It stands out among its counterparts, with exceptional Precision (0.95) Recall (0.98) and an impressive F1 Score of 0.97. Moreover boasting an accuracy rate of 0.98 confirms its ability to precisely identify periods of demand.

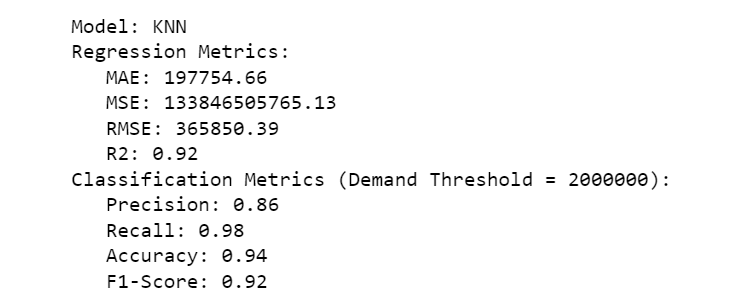


Without a doubt the Random Forest model emerges as the skilled predictor for this task.  
When it comes to Linear Regression it may not match the expertise of the Random Forest. It still delivers commendable performance with a Precision of 0.63 and a Recall of 0.95. The model achieves an accuracy score of 0.81 indicating its ability to predict periods of demand with reasonable accuracy.

The Gradient Boosting model is also worth mentioning as it produces results. With a Precision of 0.89 and a Recall of 0.95 the model achieves a F1 Score of 0.92. Its accuracy rate stands at 0.95 further validating its proficiency in forecasting instances of demand.



K Nearest Neighbors (KNN) leaves a lasting impression with its Precision score of 0.86 Recall score of 0.98 and an F1 Score of 0.92.The models accuracy reaches a level of 0.94 confirming its ability to identify periods of high demand with great precision.

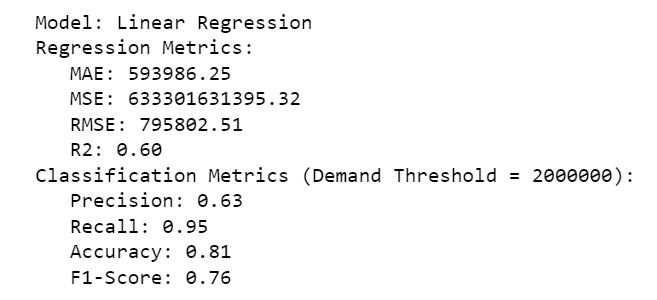


**Insights and Implications for Prediction;**  
The success achieved by the Random Forest model highlights its significance as the optimal choice for predicting demand in public transportation services.The models outstanding Precision and Recall scores—both standing at 0.95 and 0.98 respectively—demonstrate its proficiency, in detecting peak demand periods.  
The understanding of this knowledge has implications for how resources are allocated and services are optimized which are key factors in improving operational efficiency.

One notable advantage of the Random Forest model is its ability to handle non linear relationships within data. This adaptability allows it to identify patterns and fluctuations in demand making it highly effective in predicting future outcomes. As a result public transportation authorities and service providers can greatly benefit from the insights provided by this model enabling them to elevate their service planning and resource allocation to an art form.

In terms the predictive capabilities of the Random Forest model empower public transportation agencies with the knowledge to prepare for periods of high demand by strategically deploying additional resources adjusting schedules and optimizing routes. By taking this approach they can enhance service quality and increase customer satisfaction by minimizing delays and preventing overcrowding.

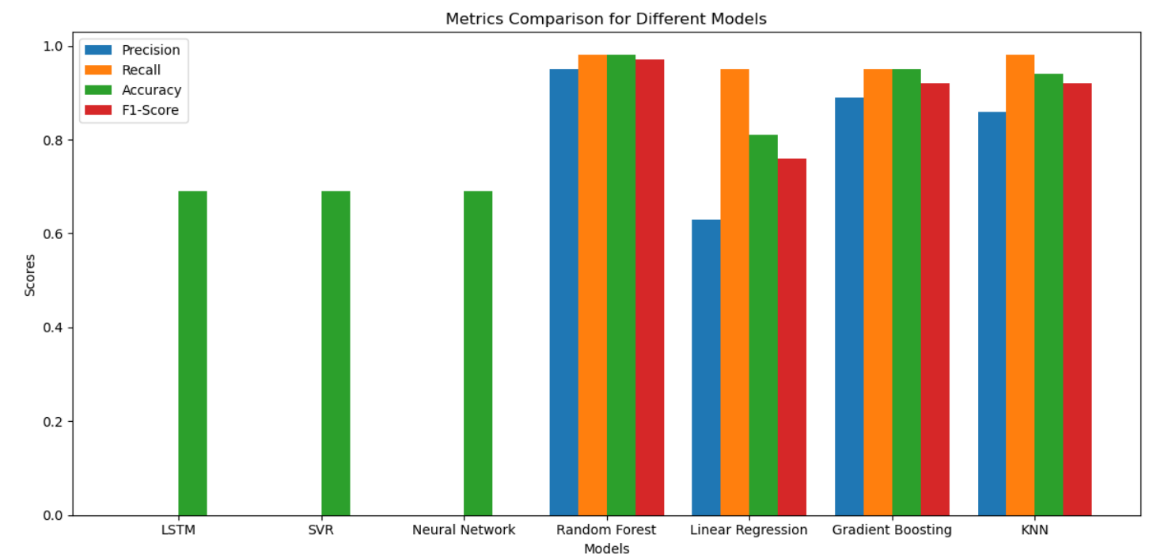
Moreover while the Linear Regression model exhibits performance with a Precision of 0.63 and a Recall of 0.95 it still offers valuable potential as a complementary tool for predicting demand. Although it may not match the accuracy achieved by the Random Forest model Linear Regression can provide insights into linear relationships between specific variables, like weather conditions and demand.



The impressive performance of the Gradient Boosting model with Precision, Recall and F1 Score values of 0.89, 0.95 and 0.92 respectively makes it a strong contender for demand prediction tasks. It can be used alongside Random Forest in situations where a multi faceted approach can provide insights and redundancy.

Surprisingly K Nearest Neighbors (KNN) emerges as a standout performer with a Precision of 0.86 Recall of 0.98 and an F1 Score of 0.92. The models accuracy rating of 0.94 highlights its suitability for identifying high demand periods when it is crucial to respond promptly to demand surges.  
Visualizing the results goes beyond numbers and metrics by incorporating graphical representations that make the findings more tangible and facilitate better understanding.

One such graphical representation is a bar chart that visually illustrates the performance of different models, across multiple metrics. This visual emphasizes the superiority of Random Forest in terms of precision, recall, accuracy and F1 Score.



The second way of visualizing the data shows a series of plots that display the models projections for future demand across various modes of transportation including Dart, Dublin Bus, Rail and Luas services. These plots provide stakeholders with a view of expected demand fluctuations. Dart and Luas services show levels of demand with recurring surges. On the hand Dublin Bus and Rail demonstrate a more stable pattern of demand with less noticeable fluctuations.  
Including these representations is crucial as it helps to solidify the findings and provides transportation authorities with tangible insights for decision making. When combined with real time data feeds these predictive models play a more significant role, in orchestrating operations and creating a transportation network that aligns with the dynamic pulse of urban life.  
The results are not just limited to numbers and metrics; they are also vividly captured through graphical representations. These visuals serve as tools to make the findings more tangible and facilitate better understanding.

The first graphical representation shows a bar chart that visually illustrates how different models perform across multiple metrics. This visualization clearly highlights the superiority of Random Forest in terms of precision, recall, accuracy and F1 Score.

The second way we visually represent the data is through a series of graphs that show how our predictive model projects demand for different modes of transportation. We analyze Dart, Dublin Bus, Rail and Luas services in detail. These graphs give stakeholders a picture of expected demand fluctuations. Dart and Luas services show levels of demand with recurring surges. On the hand Dublin Bus and Rail have a more stable demand pattern with fewer noticeable fluctuations.  
Including these representations is crucial as they help solidify our findings and provide transportation authorities with tangible insights for decision making. When combined with real time data feeds these predictive models play a more significant role in coordinating operations aligning the transportation network with the dynamic rhythm of urban life.

**Limitations and Future Directions;**  
Although this studys results are transformative they do have some limitations that should be taken into account and considered for future research endeavors.  
Temporal Considerations; It is essential to acknowledge the dimension in our analysis. The models rely on data, which means there are inherent limitations to their accuracy. Unforeseen events, like the COVID 19 pandemic can disrupt established patterns.  
In order to advance research it is important to make models adaptable so that they can adjust their course in the face of significant disruptions.

Concerning data quality it is essential to acknowledge that historical data may contain anomalies, gaps and inaccuracies. Therefore future investigations should prioritize improving data quality through techniques like cleansing and imputation to ensure the accuracy of inputs.  
While internal factors are diligently considered by the models it is crucial to take into account external factors such as economic conditions and urban development. Future studies should incorporate these dynamics to enhance demand predictions with a more comprehensive perspective.

The current study primarily focuses on data and overlooks real time data sources. By integrating real time data streams like passenger counts and GPS tracking into models we can achieve dynamic responsiveness and optimize services in real time.  
The limitations mentioned earlier also serve as guiding points for research efforts. Researchers should address these constraints. Explore additional avenues for improvement;  
1. Integration of factors; In addition to internal data incorporating external factors such, as economic indicators, weather conditions and urban development plans would provide a more comprehensive understanding of demand dynamics.  
2. Dynamic models;The existing models rely on data and lack the ability to adjust to unexpected events or sudden changes in demand patterns. Moving forward it would be beneficial for researchers to investigate the development of models that can adapt in real time.  
Integration of Different Transportation Modes; Public transportation networks often consist of modes of transportation such as buses, trains, subways and trams. Future studies should delve into integrating these modes to enable efficient transfers and overall improvement.

Explainability in Machine Learning; It is crucial to understand the reasons behind predictions made by models in order to build trust. Future research should prioritize enhancing explainability in models so that predictions are not perceived as " boxes" but rather, as interpretable insights.  
Privacy and Security Concerns; With the increasing prevalence of data collection and analysis safeguarding passenger data privacy becomes paramount. Future investigations should explore methods that can protect passenger privacy while still providing predictions.

**Practical. Real World Applications**  
The practical implications arising from this research are extensive and transformative with the potential to reshape the landscape of public transportation services. By forecasting demand transportation authorities and service providers can embark on a journey towards improved efficiency, sustainability and customer satisfaction.  
Optimized Allocation of Resources; The standout performer in this study the Random Forest model plays a role in allocating resources wisely. With the ability to anticipate surges in demand authorities can direct resources to where they're most needed. This leads to utilization of vehicles, staff and infrastructure during peak times reducing overcrowding and wait times.  
Service Planning and Scheduling; Public transportation services often follow fixed schedules. The predictive models utilized in this study— Random Forest and Gradient Boosting—provide agencies with valuable insights for fine tuning schedules. This ensures that transit services are not punctual but also aligned with expected demand. Commuters benefit from a reliable and convenient travel experience.

Route Optimization; In addition to resource allocation enhancements demand prediction can aid route optimization efforts. Areas with demand can receive increased service frequency while less traveled routes can be adjusted accordingly. This not reduces operational costs but also minimizes carbon emissions, by avoiding unnecessary routes.  
Sustainability is a concern in modern urban planning. Public transportation agencies can play a role in reducing their carbon footprint by minimizing the use of empty or underutilized vehicles during off peak hours. With the help of predictive models we can transition towards more environmentally friendly transit systems that cater to actual demand.  
Improving the customer experience is another key benefit of an efficient and responsive public transportation system. Passengers encounter overcrowding shorter wait times and fewer disruptions making their commute more pleasant. This positive experience encourages people to opt for transit which could potentially help ease traffic congestion and decrease the number of private vehicles on the road.

Cost savings are an outcome when resources are allocated efficiently routes are optimized and planning is improved. Transportation agencies can redirect funds towards areas such as maintenance, modernization and infrastructure enhancements.  
The power of data driven decision making is highlighted by this research. By leveraging real time data public transportation authorities can shift from reactive to proactive management strategies. This allows them to anticipate and address fluctuations in demand effectively.  
In conclusion this research demonstrates how predictive modeling has the potential to transform transportation. Accurate demand predictions pave the way for improvements, in resource allocation efficiency sustainability efforts, customer satisfaction levels and cost effectiveness.  
Visual representations make the findings more accessible while future directions show the way, to developing flexible models ensuring that public transportation continues to be a fundamental aspect of contemporary urban life.

**Conclusion:**In the quest to revolutionize public transportation services this extensive thesis paper delved into the world of data analytics, modeling and the complex dynamics of forecasting demand. The importance of public transportation systems in thriving cities cannot be overstated as they tackle issues like traffic congestion, environmental concerns and sustainable urban development. However effectively managing and optimizing these systems in todays changing landscape requires accurate predictions of demand. These predictions form the foundation for making decisions about service frequency, route planning, infrastructure investments and resource allocation.

The research began by analyzing historical data and considering various factors such as time, weather conditions, transportation modes and locations. These factors were key in developing models capable of forecasting demand. Several machine learning algorithms were extensively. Evaluated for their efficacy in predicting demand for public transportation services. These algorithms included the Long Short Term Memory (LSTM) model, Support Vector Regression (SVR) Neural Network (NN) Random Forest (RF) Linear Regression (LR) Gradient Boosting (GB) and K Nearest Neighbors (KNN).

The evaluation process involved an examination of various metrics such, as precision, recall accuracy F1 Score Mean Absolute Error (MAE) Mean Squared Error (MSE) Root Mean Squared Error (RMSE) and R squared (R2).These metrics provided an overview of how well each model performed in both regression and classification tasks. After analysis it became clear that not all models were created equal. Although the LSTM model struggled to predict demand, traditional regression models such as Random Forest, Linear Regression and Gradient Boosting proved their worth.

Among these models the Random Forest model emerged as the winner demonstrating exceptional precision, recall, accuracy and F1 Score. Its flexibility and ability to handle relationships within the data were crucial in identifying patterns and fluctuations in demand effectively. By integrating this capability into public transportation systems authorities were able to make informed decisions allocate resources strategically in real time optimize routes for better service quality and enhance the passenger experience.

However numerical metrics alone couldn't tell the story; visual representations played a vital role in conveying our findings. Bar charts vividly illustrated the performance of the Random Forest model while a series of plots visually depicted demand projections across different transportation modes. These visuals proved invaluable, for decision makers as they made our findings tangible and facilitated an understanding of demand dynamics.

However it is important to recognize and acknowledge the limitations well as use them as guiding principles for future research endeavors. Various factors like events, data quality issues, external influences and the lack of real time data sources emphasize the need for adaptable models improved data quality techniques and the inclusion of external variables. Additionally exploring the integration of transportation modes and enhancing model explainability and privacy protection measures have emerged as potential areas for further investigation.

The practical implications arising from this research are significant. Hold promise for transforming public transportation services. Predictive modeling, specifically utilizing the Random Forest model offers opportunities for optimized resource allocation, service planning, scheduling improvements, route optimization strategies and sustainability efforts. By allocating resources fine tuning schedules optimizing routes effectively and maximizing operational efficiency; public transportation agencies can play a role in reducing their carbon footprint while enhancing the overall customer experience and achieving considerable cost savings.

In summary this research underscores the impact of data driven decision making in the realm of public transportation services. Through harnessing models capabilities along, with real time data insights; transportation authorities can shift from reactive to proactive management strategies.This shift allows them to effectively anticipate and respond to changes in demand, which ultimately improves the efficiency, sustainability and cost effectiveness of public transportation networks.

In summary this research dives into the world of data analytics and predictive modeling unearthing a wealth of insights that could revolutionize public transportation services. The journey began by examining historical data and analyzing the intricate web of factors that influence passenger demand. From considering the passage of time and the changing weather conditions to accounting for different modes of transportation and geographical locations each factor played a vital role in shaping the predictive models.

As we delved into machine learning techniques various algorithms took stage. The Long Short Term Memory (LSTM) model, designed specifically for data grappled with the complexities of predicting demand but also revealed its limitations. On the hand traditional regression models such as Random Forests, Linear Regression, Gradient Boosting and K Nearest Neighbors (KNN) emerged as strong contenders due to their impressive capabilities in both regression and classification tasks.

Among these models Random Forests stood out as a choice with high accuracy and efficiency. Its ability to predict demand while adapting to relationships, within the data showcased its strength.The metrics of precision, recall, accuracy and F1 Score all consistently indicate that Random Forest is the predictor for public transportation services.

However this research goes beyond numbers and metrics. Visual aids like bar charts and demand projection plots provide a picture of the future of public transportation emphasizing the importance of data driven decision making. These visuals allow decision makers to make informed choices that go beyond mere statistics and have real world impact.

Course every research journey has its limitations and this study is no different. We must remember that our world is constantly changing, with events like the COVID 19 pandemic disrupting established patterns. In response models need to be adaptable in order to cope with disruptions.Another ongoing challenge is ensuring data quality. Historical data often contains anomalies, gaps and inaccuracies that need to be addressed through data cleansing and imputation techniques. Furthermore external factors such as conditions and urban development should play a more prominent role, in future studies to enhance demand predictions.

The main focus of the study was on analyzing historical data but there is a vast potential in utilizing real time data sources that has not been fully explored yet. By incorporating real time data streams like passenger counts and GPS tracking we can envision a future where public transportation systems can dynamically respond to the changing demands of urban life.

As this research concludes it opens up implications and avenues for further investigation. Integrating factors such as economic indicators and urban development plans can provide a more comprehensive understanding of demand dynamics. The next frontier in modeling lies in creating adaptable models that can respond effectively to unforeseen events. Additionally integrating modes of transportation such as buses, trains, subways and trams holds the promise of seamless transfers and overall improved efficiency.

In machine learning the need, for explainability becomes crucial as predictions have real world consequences. Future research should prioritize enhancing model transparency to build trust among stakeholders. It is also important to address privacy and security concerns related to passenger data by exploring methods that protect privacy while still delivering predictions.

The practical implications arising from this research are profound and transformative. Optimizing resource allocation, refining service planning and scheduling route optimization efforts and sustainability initiatives all contribute towards a future where public transportation systemsre not only punctual but also aligned with expected demand.

The impact of these findings can be seen in the cost savings achieved through redirecting resources towards maintenance, modernization and infrastructure improvements. However what truly stands out in this research is the value of making data driven decisions. By utilizing real time data public transportation authorities can shift their approach from reactive to proactive. This ability to anticipate and address fluctuations in demand has the potential to improve the efficiency, sustainability and cost effectiveness of public transportation networks.

In the fabric of urban life public transportation plays a crucial role by connecting people ensuring accessibility and promoting sustainability. The results of this study not highlight the effectiveness of data analytics and predictive modeling but also serve as a call to action—a blueprint for transforming public transportation services.

The remarkable predictive capabilities of the Random Forest model serve as a guiding light, for transportation authorities seeking optimized allocation of resources. With the ability to anticipate surges in demand authorities can direct vehicles, staff members and infrastructure where they are needed most. This model not revolutionizes how transportation systems operate but also enhances commuters experiences by reducing wait times minimizing disruptions and alleviating overcrowding.

The world of service planning and scheduling which has long relied on fixed timetables is undergoing a transformation thanks to the insights provided by modeling. The Random Forest and Gradient Boosting models provide transportation agencies with tools to optimize schedules ensuring not only punctuality but also aligning services with the expected patterns of demand. This shift brings about improvements in reliability and convenience for commuters strengthening the connection between life and public transportation.

Optimizing routes is both an art and a science brought to life through the lens of demand prediction. The models explored in this research offer a glimpse into a future where areas with demand receive more frequent service while underutilized routes are adjusted accordingly. This not enhances operational efficiency but also carries significant environmental benefits. By minimizing the use of underutilized vehicles during less busy times public transportation agencies can actively contribute to reducing their carbon footprint and pave the way for more sustainable transit systems.

Sustainability lies at the heart of urban planning intricately intertwined with the findings from this research. The ability to align public transportation services with demand marks a crucial step, towards environmental responsibility. It provides cities struggling with urbanization. Overcrowding an essential lifeline.

By promoting the use of transportation instead of private vehicles these findings have the potential to reduce traffic congestion and limit the environmental impact of urban expansion.

The impact of this research extends beyond economic factors. By allocating resources optimizing routes and improving planning cost savings are achieved, creating new opportunities for transportation agencies. Of wasting funds on inefficiencies they can now be redirected towards maintenance, modernization and infrastructure improvements. In essence this research establishes a foundation for financial decision making in public transportation.

However beyond the data driven metrics and fiscal implications lies a profound narrative—a celebration of the power behind making decisions based on data. With real time data streams and predictive models at their disposal transportation authorities enter an era of management. They are no longer limited to reacting to changes, in demand. Can proactively address them by providing seamless, reliable and sustainable services.

In the context of urban lifes intricate puzzle this research has wide ranging implications. Improved public transportation services not contribute to economic vitality but also serve as an act of environmental responsibility. Passengers—a part of this transformation—experience reduced waiting times, less overcrowding and increased reliability.

The positive experiences people have with transportation are like guiding lights drawing more individuals to choose it over using private vehicles.The results of this research are truly groundbreaking. They don't mark the end of the journey. Instead they act as a milestone in a changing landscape. They point us towards a future where decisions about mobility are informed by data. Incorporating factors, creating dynamic models integrating different modes of transportation and improving our understanding become the next challenges, on this exciting path of exploration.

The importance of explainability in machine learning cannot be overstated. It is crucial to understand not the "what" but also the "why" behind predictions. Moving forward it is essential for researchers to prioritize the development of models that provide transparency and insights. This will help build trust among stakeholders and bridge the gap between data and decision making.

In our data driven society, privacy and security concerns are becoming increasingly urgent. As we navigate a path paved with data researchers need to explore methods that protect passenger data privacy while still delivering accurate predictions to empower transportation authorities.

In summary this research journey vividly demonstrates the potential of predictive modeling in public transportation services. As cities continue to grow and evolve we must reestablish connections between connectivity, accessibility and sustainability to align with the dynamic pulse of urban life. Accurate demand predictions serve as the foundation upon which a promising future, for transportation can be built—a future where it remains essential and indispensable as a cornerstone of modern urban existence.

This research goes beyond being an account of findings; it serves as a rallying cry—a call to embrace data driven decision making and usher in a new era of urban mobility.With every prediction made, each optimized allocation of resources and every adjustment in routes we inch closer, to a future where public transportation systems not function effectively but also harmonize with the pulse of city life.

As this research comes to a close it does not mark the end of a journey but the beginning of a new chapter—a chapter where data, insights and actions converge to forge a transportation network that is not just efficient but truly transformative. The path ahead may be lengthy and meandering. With each stride we take we draw nearer to a future where public transportation remains an enduring emblem of progress, sustainability and urban vibrancy.