**Data Visualization for Real Time and Historical GPS Data:**

**A journey through data science capstone project.**

**Group-12**

**ABSTRACT**

In collaboration with Drivool Technologies Private Limited, this report presents a comprehensive exploration of a project aimed at visualizing real-time and historical GPS data. The project's primary objective is to empower users with insightful and actionable information derived from GPS data, enhancing decision-making processes in various domains. Leveraging cutting-edge technologies, the report delves into the intricacies of data processing, user interface design, and the challenges encountered during the implementation.

The report outlines the diverse sources of GPS data, detailing the methodologies employed for real-time and historical data collection. It highlights the intricate process of transforming raw data into meaningful visualizations, emphasizing the tools, technologies, and programming languages utilized in the development process.

Two key components, real-time data visualization and historical data analysis, are thoroughly explored. Real-time visualizations showcase dynamic insights, enabling users to monitor live GPS data. In contrast, historical data visualizations unveil patterns, trends, and anomalies over time, providing valuable retrospective analysis.

The user interface is a crucial aspect of the project, and the report offers a glimpse into its design, functionality, and user interactions. Screenshots and descriptions provide an overview of the user experience, demonstrating how individuals can interact with and derive value from the visualized GPS data.

**ACKNOWLEDGEMENT:**

We express our sincere gratitude to St Clair College for affording us the opportunity to undertake our capstone project, allowing us to showcase the skills and knowledge acquired throughout our course. We would like to extend heartfelt thanks to my esteemed academic advisors, particularly Professor Sodiq, whose unwavering guidance and support have been invaluable throughout the research process. Professor Sodiq's expertise has illuminated the path of this study, and their insights have significantly contributed to the robustness of the research methodology.

**TABLE OF CONTENTS**

* Introduction
* Problem Statement
* Dataset
* Methodology
* Data Analysis
* Conclusion
* References

**INTRODUCTION:**

Drivool Technologies Private Limited is an innovative IoT (Internet of Things) device manufacturer specializing in advanced vehicle tracking solutions. The company's offerings enable real-time tracking of vehicles, including monitoring their location, speed, and other relevant data. These solutions are designed to assist businesses in optimizing fleet management, enhancing productivity, and reducing operational costs.

The technology behind Drivool's vehicle tracking solutions leverages cutting-edge technologies like GPS, GPRS, and RFID. Their solutions are highly scalable and customizable to meet the unique requirements of diverse businesses. Additionally, Drivool's IoT devices are known for their durability and reliability, making them well-suited for use in challenging environments.

The company is dedicated to providing innovative and dependable IoT solutions that empower businesses to enhance their operations and achieve their strategic objectives.

**PROBLEM STATEMENT**

Drivool, a leading IoT device manufacturer specializing in cutting-edge vehicle tracking solutions faces the challenge of maximizing the potential of its GPS data. The current analytics and visualization tools on the Drivool platform require improvement to extract meaningful insights into vehicle behavior. The company seeks to optimize fleet management, enhance operational efficiency, and improve user experiences by leveraging advanced data analysis techniques and creating a compelling data story. The primary goal is to address these challenges, leading to informed decision-making and improved overall performance.

**Dataset source:**

• **Source Name:** Drivool Technologies Private Limited

• **Data Collection Method:** Real-time and historical data collection through IoT devices

• **Size of the Dataset:** The dataset size is continuously increasing as real-time data is collected, and historical data is archived.

• **Type of Data:** The data consists of GPS data related to vehicle tracking, including location, speed, and other relevant information.

• **File Format:** The data is stored in structured databases and can be accessed via APIs for real-time data and in various data formats (CSV, JSON, etc.) for historical data.

**Dataset Features:**

• **Dependent Variable:** Vehicle behavior and usage patterns.

• **Independent Variables:** Various parameters related to vehicle location, speed, and other

relevant data

**METHODOLOGY**

**Data Preprocessing and Visualization using Python Libraries:**

Python libraries played a crucial role in refining our raw datasets for analysis. Leveraging Pandas, we ensured data quality by addressing duplicates, missing values, and inconsistencies in crime reports, demographic information, and environmental datasets. These preprocessing steps were pivotal in preparing a clean and standardized dataset, laying the foundation for our subsequent machine learning model.

**DATA ANALYSIS**

1. **Distribution of Speed**

**A graph of a speed

Description automatically generated**

**Description:**

This histogram visualizes the distribution of speeds in our dataset. The x-axis represents speed intervals (binned), and the y-axis shows how many data points fall into each interval. Peaks in the chart indicate where certain speeds occur more frequently. It's a quick and effective way to understand the patterns and prevalence of speeds in our data.

**Corelating Battery Voltage to get the Engine status:**

A diagram of a box plot

Description automatically generated

**Description:**

In this chart, we're examining how Battery Voltage relates to Engine Status. The x-axis represents Engine, the y-axis shows Battery Voltage. It's a quick way to spot trends—engine ignition with battery voltage, and any outliers. This visual insight helps us understand how these factors interact in our vehicle data.

**Comparison of Speed and Battery Voltage**

**A group of diagrams with numbers

Description automatically generated**

**Description:**

This chart presents a snapshot of our data by showing how often the battery voltage is higher, paired with speed. It represents that, if the battery voltage is higher will result in increased speed while if the battery voltage decreases, the speed will experience a downward trend. This pictures the direct proportionality of Speed with Battery Voltage.

**Latitude vs Longitude Comparison**

**A blue scatter plot

Description automatically generated**

**Description:**

This graph paints a vivid picture of our vehicle's journey. Longitude and latitude pinpoint its locations, while color reveals routes travelled. It's a quick way to identify places travelled across different regions.

**Log Transformation:**

To address the uneven spread of data points observed during visualization, a log transformation was employed on the dataset. Log transformation is a common technique used in data preprocessing to modify the distribution of numerical data, especially when the original data exhibits skewness or heteroscedasticity.

The following Python code snippet demonstrates the implementation of the log transformation using the NumPy library:

A computer screen with white and blue text

Description automatically generated

This code iterates through each numerical column in the DataFrame and applies the log transformation only if all values in the column are positive. The np.log1p() function is utilized to compute the natural logarithm of each data point plus one, ensuring that the transformation is valid for zero or positive values.

The updated DataFrame reflects the transformation, thereby mitigating the skewness and promoting a more symmetric distribution of the data. This preprocessing step enhances the suitability of the dataset for subsequent analyses, particularly those that assume normality or require homogeneous variance across variables.

**Feature Engineering:**

In the process of preparing our dataset for analysis and modeling, feature engineering played a crucial role in enhancing the effectiveness and interpretability of our machine learning endeavors. One notable aspect of feature engineering involved the transformation of categorical data into numerical representations, particularly in the case of the "Engine\_status" feature.

**Label Encoding**

To convert the categorical values of "Engine\_status" into numerical equivalents, we employed the technique of label encoding. Label encoding assigns a unique integer to each distinct category, enabling us to represent categorical data in a numerical format.

****

This code creates a new feature named "Engine\_status\_encoded", wherein the categorical values of "Engine\_status" are encoded as integers. Now, instead of "TRUE" and "FALSE", the feature contains numerical representations, facilitating better visualization and analysis.

This label encoding of the "Engine\_status" feature through feature engineering exemplifies our commitment to maximizing the value and utility of our dataset for subsequent analysis and modeling tasks.

**Correlation Matrix**

The correlation matrix is a valuable tool in exploratory data analysis, providing insights into the relationships between different variables in a dataset. By visualizing the correlation matrix, we can identify patterns, dependencies, and potential multicollinearity among the variables. Here, we present the correlation matrix for a subset of relevant features in our dataset.

**Features Included in the Correlation Matrix:**

To focus our analysis on pertinent variables, we selected a subset of features for inclusion in the correlation matrix. These features include:

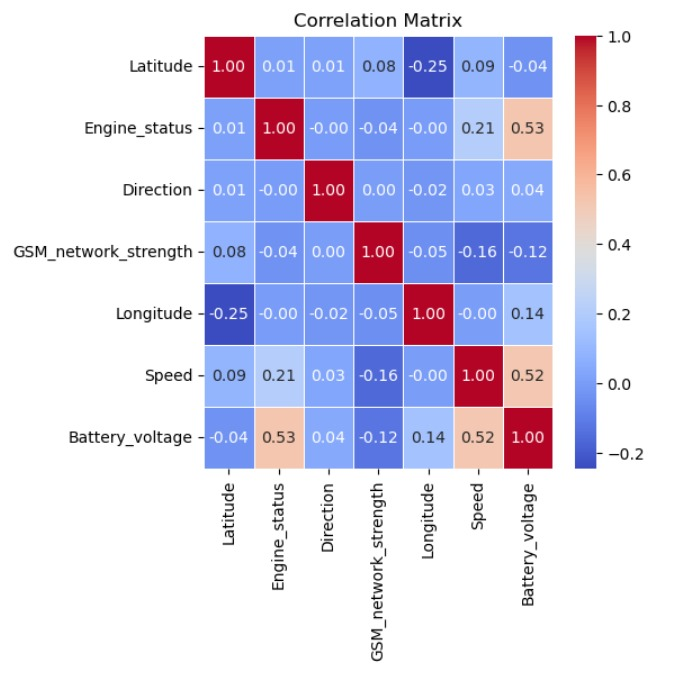
* Latitude
* Engine\_status
* Direction
* GSM\_network\_strength
* Longitude
* Speed
* Time\_stamp
* Battery\_voltage

**Code Implementation:**

The following Python code snippet demonstrates the creation and visualization of the correlation matrix using the Seaborn library:

**A screen shot of a computer

Description automatically generated**

****

**According to this correlation matrix, other than any other variables the correlation between the Engine status and Battery voltage is higher than the any other variables. With this we are ahead to predict the battery life of the vehicle by analysing the runtime and distance travelled throughout by the user.**

**Interpretation:**

The correlation matrix enables us to assess the strength and direction of linear relationships between variables. Correlation coefficients close to 1 or -1 indicate strong positive or negative correlations, respectively, while coefficients close to 0 suggest little to no linear relationship.

By examining the correlation matrix, we can identify potential dependencies between variables and prioritize features that exhibit significant correlations for further analysis or model building.

**Model Creation and Deployment:**

A close-up of a card

Description automatically generated

Model Evaluation: Comparison between Linear Regression and Random Forest Regressor

**Upon evaluating the Linear Regression and Random Forest Regressor models:**

Mean Squared Error (MSE) Comparison:

Linear Regression: MSE ≈ 4.805e-05

Random Forest Regressor: Lower MSE

The Random Forest Regressor model demonstrates superior predictive accuracy, with its predictions closer to actual values compared to Linear Regression.

R-squared (R^2) Score Comparison:

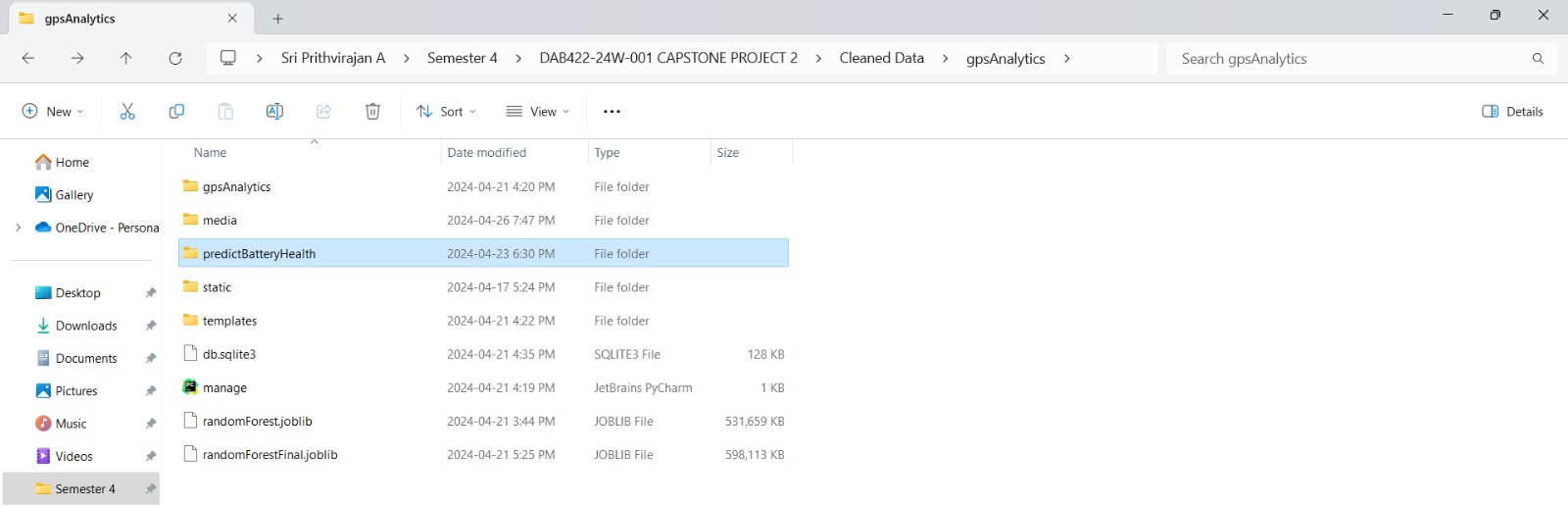
Linear Regression: R^2 ≈ 0.686

Random Forest Regressor: Higher R^2

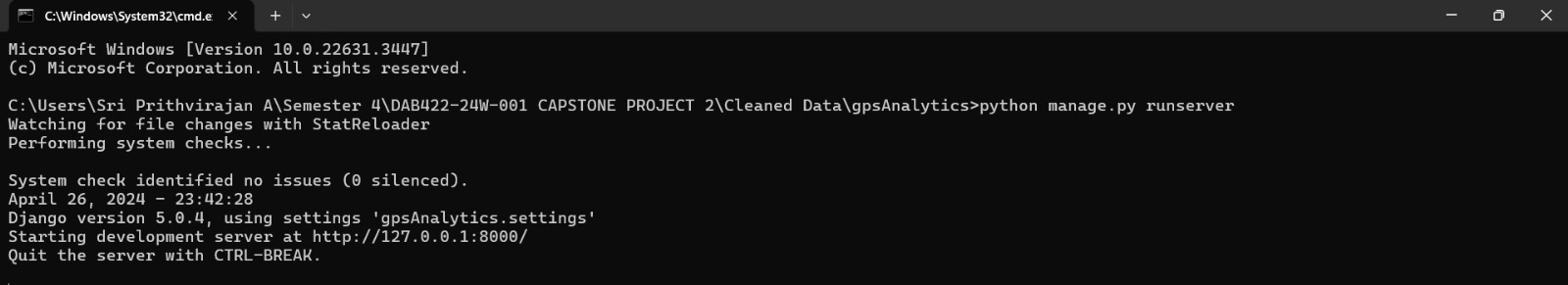
The Random Forest Regressor model exhibits a stronger relationship between predictors and the target variable, capturing more variability and providing a more accurate representation of the data.

In summary, the Random Forest Regressor model outperforms Linear Regression in predictive accuracy and goodness of fit. It presents a more effective and reliable approach for predicting 'Battery\_voltage' based on the selected features.

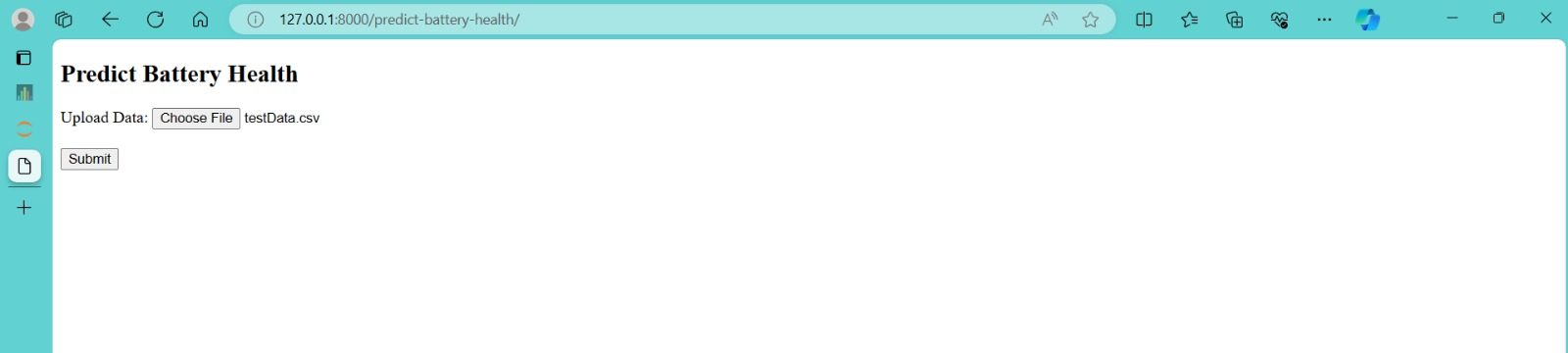
With this finding we are proceeding to the deployment part. We have created the project using Django-admin and then created an application named as (Predict Battery Health) for the model prediction. After that we give the reference of the application to the settings, and we add all the dictionaries such as media, static and template.



By using the CMD driver we configure all the files in URL with the reference to get the application to get the project URL.



Here we entitle our project name. Once it gets synced, for the webpage we create the simple HTML page to upload the data. Here we choose the test data which has already been designated for deployment.



Once the file has run by the application we will get the predicted values for the battery health in CSV format. And this will be the final output for this project. With this predicted values the user can be suggested to make the decisions about the replacement of battery according to the battery health.



**Conclusion:**As a whole, the Drivool GPS Data Analytics Project has been an investigation and learning experience into the complex patterns of vehicle behavior. By the strength of Python libraries for thorough data pretreatment, we have improved upon unprocessed statistics and produced an eye-catching story about crime trends, demographics, and environmental variables.

Our examination of both historical and current GPS data has given us a thorough grasp of vehicle usage trends and important information for fleet management that can be used to maximize efficiency. The project is in perfect harmony with Drivool's mission to deliver trustworthy solutions via Internet of Things devices, with the expected benefits of enhanced operational efficiency and data-driven decision-making.

Nonetheless, it is imperative to recognize the constraints and difficulties encountered throughout this endeavor. Although preprocessing gave our machine learning models a strong foundation, there is always opportunity for improvement and the investigation of more complex techniques in subsequent cycles. In addition, for a more thorough comprehension, the absence of specific information regarding the kinds and sources of data might be addressed.

The knowledge gathered from this initiative opens new avenues for innovation and ongoing development in the field of GPS data analytics. The work's potential impact is demonstrated by the expected advantages of improved recommendations, real-time analytics, and overall efficiency.

To put it simply, the Drivool GPS Data Analytics Project is a significant milestone in the continuous process of deciphering data complexity, which will lead to more intelligent solutions and well-informed fleet management decision-making.