**Leveraging Data in Last-Mile Delivery: 2021 Amazon Last Mile Routing Research Challenge Analysis**

CIV1506: Freight Transportation and ITS Applications

Final Term Project

Monday March 15, 2024

by:

Rome Thaniratt (1005841608)

&

Michelle Yip (1004097416)

Department of Civil & Mineral Engineering

University of Toronto

**Table of Contents:**

**1.0 Introduction**.............................................................................................................................3

**1.1Project Objective**..........................................................................................................3

**2.0 Literature Review.**..................................................................................................................3

**2.1 Vehicle Routing Problem (VRP).**...............................................................................4

**2.2 Route-Based Approach.**..............................................................................................4

**2.3 Instance-Based Approach.**.........................................................................................4

**3.0 Methodology**............................................................................................................................4

**4.0 Exploratory Data Analysis**.....................................................................................................6

**4.1 Data Overview**.............................................................................................................6

**4.2 Feature Extraction**......................................................................................................6

**4.2.1 Route-Specific Features**...............................................................................6

**4.2.2 Package-Specific Features**...........................................................................7

**4.2.3 Stop-Specific Features**.................................................................................7

**4.3 Feature Engineering**...................................................................................................7

**4.4 Feature Analysis**..........................................................................................................8

**4.4.1 Route Score**...................................................................................................8

**4.4.2 Station Code**.................................................................................................8

**4.4.3 Weekday**......................................................................................................10

**4.4.4 Departure Hour**..........................................................................................10

**4.4.5 Vehicle Capacity**.........................................................................................10

**4.4.6 Number of Specified Windows**.................................................................11

**4.4.7 Total Service Time**.....................................................................................11

**4.4.8 Package Scan Status**...................................................................................12

**4.4.9 Number of Stops**.........................................................................................12

**4.4.10 Total Travel Time**.....................................................................................13

**4.5 Route-ID Visualization**..............................................................................................13

**4.6 Feature Correlations**..................................................................................................14

**4.6.1 Numerical Features**.....................................................................................14

**4.6.2 Categorical Features**...................................................................................15

**5.0 Results and Discussion**...........................................................................................................16

**5.1 Clustering Analysis**....................................................................................................16

**5.2 Gradient Boosted Classifier**......................................................................................18

**5.3 Random Forest Classifier**.........................................................................................19

**5.4 Discussion**...................................................................................................................19

**6.0 Conclusion**..............................................................................................................................20

**7.0 References**...............................................................................................................................21

**1.0 Introduction**

In the dynamic and ever-evolving landscape of logistics and supply chain management, optimizing last-mile delivery operations emerges as an indispensable endeavor, essential for both operational efficiency and customer satisfaction. The last mile, the final leg of the delivery process, represents a critical point where the efficiency and effectiveness of logistics operations are put to the ultimate test. Its intricate nature, characterized by diverse challenges such as variable traffic conditions, complex urban environments, and tight delivery windows, underscores the significance of employing innovative approaches to enhance service quality and streamline operations.

Against this backdrop, our project embarks on a data-driven exploration of last-mile delivery optimization, leveraging the nuanced insights encapsulated within the 2021 Amazon last mile delivery dataset. While other methodologies predominantly focus on sequence analysis of delivery routes, our approach diverges by prioritizing the utilization of instance-based features. These features, drawn both from the inherent attributes of the dataset and meticulously engineered through advanced analytics techniques, serve as the cornerstone of our methodology. By harnessing the power of sophisticated machine learning algorithms, particularly the Random Forest Classifier and Gradient Boosted Classifier, we aim to develop predictive models capable of discerning and categorizing route efficiency levels with a high degree of accuracy.

**1.1 Project Objective**

At the heart of our project lies a dual-fold objective aimed at shedding light on the intricacies of last-mile delivery efficiency:

Firstly, we seek to classify route efficiency levels based on an expansive array of instance-based features extracted from the dataset. Unlike conventional studies that rely solely on route sequencing algorithms, our methodology embraces a holistic approach, encompassing a diverse spectrum of factors influencing delivery performance. By dissecting and analyzing these features in-depth, we aim to develop robust predictive models capable of accurately gauging route efficiency levels across various operational contexts.

Secondly, our project aims to elucidate the relative importance of these instance-based features in the final classifiers. Through meticulous experimentation and feature importance analysis, we endeavor to unravel the underlying drivers of route efficiency, discerning the pivotal factors that dictate delivery performance. This endeavor not only enhances our understanding of last-mile logistics dynamics but also equips logistics operators with actionable insights to optimize their delivery operations effectively and adaptively.

**2.0 Literature Review**

The term “last-mile delivery” often refers to the logistical operations involved in transporting goods during the final segment of their journey to reach the customer (Florio, A, Costa, P, & Ozarik, S, 2021). With the significant rise in e-commence orders, logistics service providers encounter many challenges. The challenges associated with last-mile delivery primarily arise from the need to distribute individual orders to multiple addresses. Despite the increasing research on optimization methods, there’s still a lack of focus on how these methods can be practically applied to make last-mile deliveries more efficient. Thus, one way to solve and improve efficiency in the last-mile delivery is Vehicle Routing Problem (VRP).

It is important to note that there are other research papers focusing on route-based approaches for the 2021 Amazon Last Mile Routing Research Challenge. However, this paper focuses on instance-based approaches, creating features for each route, although sequencing was not a primary focus.

**2.1 Vehicle Routing Problem (VRP)**

The Vehicle Routing Problem (VRP) mainly emphasizing the efficient distribution of goods through the utilization of a fleet of vehicles (Wenyi, L). VRP was based off from the Traveling Salesman Problem (TSP) and was formally introduced by Dantzig and Ramser in 1959, by finding the optimal shortest route then returning to departure with no repetitive visits (Singh, A, & Wenyi, L). Though VRP in transportation can present it challenges, it shows real life logistics by exemplifying significant savings of costs, efficiency, time constraints, and customer satisfaction (Wenyi, L).

**2.2 Route-Based Approach**

A “route-based approach” typically refers to a method in logistics in transportation that determines the most optimal route for delivery drivers. This approach's objective is to optimize and enhance design vehicle paths to efficiently serve multiple locations, such as warehouses/ distribution centers, while considering factors like time constraints, traffic, minimizing total distance, etc (Woods, C, 2023). There are many benefits of route-based approach, such as improving efficiency like fuel and cost, increasing customer satisfaction by receiving service on time, and enhancing productivity from removing detours and delays (Woods, C, 2023).

**2.3 Instance-Based Approach**

A “instance-based approach” typically involves making decisions or solving problems based on the specific instance, rather than following predetermined rules or models. This approach uses more complex symbolic presentations but relies on comparing and analyzing instances to make recommendations and predictions (Keogh, E, 2011). Thus, this paper will generalize the dataset from Amazon and make new predictions.

**3.0 Methodology**

Our methodology encompasses an end-to-end approach to last-mile delivery optimization. The schematic representation of our pipeline, outlined in Figure 1, delineates each stage of our methodology, from data acquisition to model evaluation.

* Data Wrangling:

The first phase of our methodology involves data wrangling, where we gather, assess, and preprocess the 2021 Amazon last mile delivery dataset. This crucial step entails addressing data inconsistencies between each data file within the compressed file and ensuring data integrity to lay a solid foundation for subsequent analyses.

* Exploratory Data Analysis (EDA):

Following data wrangling, we conduct exploratory data analysis (EDA) to gain comprehensive insights into the underlying patterns and distributions within the dataset. Through visualizations, statistical summaries, and correlation analyses, we uncover key trends, outliers, and potential areas of interest that inform subsequent feature engineering efforts. The main objective of conducting an EDA is to unveil potential patterns that can influence the classification of route efficiency scores.

* Feature Engineering:

Armed with insights gleaned from EDA, we embark on feature engineering, a pivotal stage where we transform and augment the dataset to extract meaningful predictors for our predictive models. This process involves creating new features, transforming existing ones, and encoding categorical variables to enhance our models' predictive power.

* Train/Test Split:

With the feature-engineered dataset in hand, we partition it into training and testing sets using a predefined ratio of 70:30. The training set serves as the data on which our models are trained, while the testing set remains unseen during training and is used to evaluate the models' performance on unseen data.

* Pre-processing:

Before model training commences, we apply necessary pre-processing steps to standardize or normalize features and address any potential data imbalances. This ensures that our models are robust and capable of generalizing well to unseen data. In the case of our model, we decided to keep all unique target variables under route scores in hopes to better understand what contributes to ‘low’ route scores despite having drastic imbalance within the target variables.

* Model Training and Evaluation:

Finally, we train our machine learning models, specifically the Random Forest Classifier and Gradient Boosted Classifier, on the pre-processed training data. Leveraging the wealth of features engineered earlier, these models learn to classify route efficiency levels based on instance-based features.

Following model training, we evaluate their performance using a range of evaluation metrics, including accuracy, precision, recall, and F1-score. This comprehensive assessment lets us gauge our models' efficacy in accurately classifying route efficiency levels, identify potential improvement areas, and understand the importances of each feature within the feature space.

A diagram of a data flow

Description automatically generated**Figure 1: End-to-End Pipeline Schematic**

**4.0 Exploratory Data Analysis**

**4.1 Data Overview:**

The compressed file provided for analysis comprises six distinct files: package\_data, route\_data, actual\_sequence, travel\_time\_data, and invalid\_sequence\_scores. Each file captures essential information relevant to the last-mile delivery process, offering insights into various aspects of route planning, package handling, and logistical operations. The dataset provided comprises 6112 distinct delivery routes, underscoring the complexity of the dataset presenting more than enough opportunity for in-depth analysis and modeling.

**4.2 Feature Extraction:**

From these files, we defined three primary categories of features that serve as the foundation of our analysis: route-specific, package-specific, and stop-specific features. Each category captures unique attributes that collectively contribute to the overall understanding of last-mile delivery dynamics. Before extraction, it is important to note that the provided files are in json format and needed appropriate measures to wrangle accordingly. The codes used can be found in the attached python notebook.

**4.2.1 Route-Specific Features:**

* Route ID: Serving as a unique identifier, the Route ID distinguishes each delivery route within the dataset.
* Station Code: Comprising a unique depot ID represented by a combination of three letters and a numerical identifier (e.g., DLA7 for Depot Los Angeles #7).
* Date and Departure Time: The Date and Departure Time attributes denote the departure date and time of each delivery route, offering temporal context crucial for analyzing route scheduling.
* Vehicle Capacity: Reflecting the capacity of the vehicle utilized for each delivery route, the Vehicle Capacity feature provides insights into resource allocation and operational efficiency.
* Route Scores: Ranging from High to Low, Route Scores represent an assessment of each route's efficiency, serving as a key performance indicator for route optimization efforts.

**4.2.2 Package-Specific Features:**

* Time Window: Defined by start and end times, the Time Window feature specifies the designated delivery window for each package, guiding route planning and scheduling decisions.
* Planned Service Time: Reflecting the anticipated service duration for each package, the Planned Service Time attribute aids in estimating route completion times and resource allocation.
* Package Dimensions: Comprising the height, width, and depth of each package in cubic centimeters, Package Dimensions provide insights into spatial constraints and packaging requirements.
* Scan Status: Categorized as delivered, delivery attempted, or rejected, the Scan Status feature tracks the status of package delivery attempts.

**4.2.3 Stop-Specific Features**:

* Stops: Represented as a dictionary item, the Stops feature encompasses unique stop identifiers along with their corresponding latitude, longitude, and zone ID.

**4.3 Feature Engineering**

The table below summarizes the engineered features from those preliminarily extracted and listed above.

|  |  |  |
| --- | --- | --- |
| **Route-Specific:** | **Package-Specific:** | **Stop-Specific:** |
| * Station group * Delivery weekday * Departure hour * Total travel time * Delivery time | * Package count * Number of specified time windows * Total package volume * Used vehicle capacity * Total counts of each scan status | * Number of stops * Zone prefixes * Zone number 1 * Zone number 2 * Zone suffixes * Stop sequence |

**Table 2: Engineered Features by Different Sources**

**4.4 Feature Analysis**

In this section, we delve into an examination of the various features included in our dataset, comprising both original attributes and engineered predictors. Our objective is twofold: to extract fundamental statistics and information pertaining to each feature and to explore potential correlations with our target variable, route score. Through comprehensive analysis, we aim to uncover insights into the predictive power and relevance of individual features in determining route efficiency levels. By scrutinizing feature distributions, and measures of central tendency, we seek to determine patterns that may inform subsequent modeling efforts and guide strategic decision-making in last-mile delivery optimization.

**4.4.1 Route Score**

This attribute serves as a crucial determinant of route efficiency levels. Upon examination, we observe that the distribution of route scores within our dataset reveals a notable imbalance: there are 2718 instances classified as 'High' efficiency, 3292 instances categorized as 'Medium', and 102 instances labeled as 'Low'. This inherent class imbalance underscores the complexity of our classification task and emphasizes the importance of robust modeling techniques capable of effectively handling skewed class distributions. Despite this imbalance, we made the deliberate decision to retain all unique labels for route scores, recognizing the significance of understanding factors contributing to low-scored routes in addition to optimizing classification accuracy.

**4.4.2 Station Code**

A graph of a distribution of stations

Description automatically generated

**Figure 2: Distribution of Stations by Route Score**

The distribution of delivery routes by their originating stations provides valuable insights into geographic variations in last-mile delivery operations. Among the stations analyzed, DLA (Los Angeles) stands out as the most frequent origin, followed by DSE (Seattle), DCH (Chicago), DBO (Boston), and DAU (Austin). Interestingly, a closer examination of route scores reveals intriguing patterns across different stations. While routes originating from DLA display a relatively balanced distribution between 'High' and 'Medium' route scores, stations such as DCH and DBO exhibit a notable disparity. Here, 'Medium' route scores occur almost twice as frequently as 'High' scores, indicating potential inefficiencies or challenges specific to these regions.

**4.4.3 Weekday**

Examining the distribution of delivery routes by the day of the week reveals interesting patterns. Deliveries are predominantly concentrated on weekdays, with Monday through Friday seeing higher volumes compared to weekends. Interestingly, while weekdays are dominated by 'Medium' route scores, weekends exhibit a higher proportion of 'High' scores. This suggests potential efficiency improvements or optimized scheduling strategies during weekend deliveries, contributing to enhanced route performance and customer satisfaction.

A graph of a distribution of weeks

Description automatically generated

**Figure 3: Distribution of Weekdays by Route Score**

**4.4.4 Departure Hour**

A graph of a number of different colored bars

Description automatically generated

**Figure 4: Distribution of Departure Times (Hour) by Route Score**

Analyzing the "Departure Hour" for each delivery route provides insights into temporal patterns and their potential impact on route efficiency. The included bar plot, segmented by different route scores, illustrates the distribution of departure hours across various efficiency levels. Surprisingly, the analysis reveals a consistent distribution of route scores across all departure hours. Regardless of when routes commence, the proportions of 'High', 'Medium', and 'Low' route scores remain stable. This suggests that the hour of departure may not be a reliable indicator of route efficiency.

**4.4.5 Vehicle Capacity**

The dataset lists three unique vehicle capacities: 4175 vehicles with a capacity of 3313071 cm3, 1926 vehicles with a capacity of 4247527 cm3, and only 11 vehicles with a capacity of 3114853.25 cm3, suggesting a potential outdated or less commonly used version.

The accompanying bar plot shows the distribution of route scores across different vehicle capacities. The distribution of route scores appears consistent across all sizes, indicating that vehicle capacity may not reliably predict route efficiency. These observations prompt a deeper exploration of alternative factors influencing route scores to enhance predictive modeling accuracy and operational optimization strategies.

A graph of a distribution of vehicle capacity

Description automatically generated

**Figure 5: Distribution of Vehicle Size by Route Score**

**4.4.6 Number of Specified Windows**

A graph of a distribution of delivery

Description automatically generatedA screenshot of a computer screen

Description automatically generated

**Figure 6: Distribution of Number of Deliveries with Specified Time Windows by Route Score**

Within the dataset, some packages are assigned specified time windows for delivery, while others are not. High-scored routes tend to have more specified time windows, while low-scored routes have fewer. This suggests that specified time windows act as constraints within the delivery process. The accompanying figure illustrates this distribution, emphasizing the impact of scheduling constraints on route efficiency levels. High-scored routes benefit from adherence to specified time windows, while optimization opportunities may exist for low-scored routes with fewer constraints.

**4.4.7 Total Service Time**

The section investigates the "Total Service Time" for each delivery route, calculated by summing the service times for all packages on the route. The accompanying figure illustrates these observations, highlighting the differences in total service time distribution between high and medium scored routes.  High-scored routes exhibit a greater spread, lower distribution peak, and higher mean compared to medium routes. This suggests that high-scored routes may involve a wider range of service durations and potentially more variability in delivery tasks.

A graph of a service

Description automatically generatedA screenshot of a computer screen

Description automatically generated

**Figure 7: Distribution of Total Service Time (min) by Route Score**

**4.4.8 Package Scan Status**

In this section, we examine the "Package Scan Status," which indicates whether packages are successfully delivered or rejected along the delivery routes. For each route, we have compiled the total number of rejected and successfully delivered packages. Rejected deliveries are rare occurrences. However, among the instances where rejections do occur, low-scored routes appear to have the highest frequency. This observation suggests a potential correlation between route efficiency and the occurrence of rejected deliveries. Furthermore, our analysis reveals that high-scored routes tend to have a higher average number of delivered packages compared to medium-scored routes.

**4.4.9 Number of Stops**

The distribution of stops, segmented by route score, reveals intriguing patterns. High route scores exhibit a right-skewed distribution with a wider curve, suggesting variability in the number of stops. Conversely, medium route scores display a narrower distribution with a higher peak, indicating more consistent stop counts. Medium route scores have a higher average number of stops compared to high route scores. This observation underscores potential differences in route complexity and operational requirements between high and medium scored routes.

A graph of a distribution of stops

Description automatically generatedA screenshot of a data

Description automatically generated

**Figure 8: Distribution of Number of Stops by Route Score**

**4.4.10 Total Travel Time**

This section examines the "Total Travel Time" aggregated from sequence data for each delivery route, representing the cumulative time spent traveling between stops. The distribution of total travel time, segmented by route scores, is also presented. Observations reveal distinct differences between medium and high scored routes. Medium routes exhibit higher peaks in the distribution but also have a higher average total travel time compared to high scored routes. This discrepancy is expected, as high-quality routes typically involve shorter travel times between stops.

A graph of a distribution of travel time

Description automatically generatedA screenshot of a computer

Description automatically generated

**Figure 9: Distribution of Total Travel Time (min) by Route Score**

**4.5 Route-ID Visualization**

This section unveils a visual exploration of the zone IDs present in our dataset, representing the distinct zones encountered along each delivery route. These zone IDs encode valuable information about how Amazon segments delivery destinations into zones and sub-zones, providing a structured framework for logistical operations.

Each zone ID comprises a prefix (e.g., 'D'), first number, second number, and suffix (e.g., 'H'), delineating specific geographical areas within the delivery network. This structured nomenclature aids in organizing and managing delivery routes efficiently.

The accompanying map visualization employs color-coding to represent different zone prefixes for routes traversing the Los Angeles (LA) region. For instance, zones prefixed with 'A' are depicted in red, while transitions to zones with prefix 'H' are denoted by magenta hues.

Additionally, a complementary figure enhances our understanding by visualizing the interplay between zone suffixes and second zone numbers. This detailed visualization offers a deeper level of granularity, by examining how zone suffixes and second zone numbers interact, we gain insights into the specific regions and sub-regions traversed during the delivery process. These segmented zone id components can then be used as features in our classification model.

A map with many points

Description automatically generated  A map of a city with points

Description automatically generated

**Figure 10: Route Visualization Based on Prefix and First Zone Number**

A map with points on it

Description automatically generated  A map of a city with points

Description automatically generated

**Figure 11: Route Visualization Based on Suffix and Second Zone Number**

**4.6 Feature Correlations**

**4.6.1 Numerical Features**

In this section, a correlation heatmap is presented, showcasing all pre-existing and generated numerical features derived from the dataset. Pearson's correlation coefficient was utilized to quantify the relationships between these features. These correlations serve as a sanity check, validating intuitive relationships between features that are expected to be related.

Key correlations observed include:

* Package count and number of stops: A logical relationship suggesting that routes with more packages may entail more stops.
* Package count and total package volume: An expected correlation indicating that more packages would likely result in a higher total package volume.
* Package count and number of packages delivered: Intuitively, routes with more packages should lead to more deliveries.
* Number of stops and total travel time (between stops): A sensible correlation, implying routes with more stops may require more time for travel between them.
* Number of stops and total delivery time (travel + service): A logical connection suggesting that routes with more stops may incur longer overall delivery times due to both travel and service durations.

A screen shot of a chart

Description automatically generated

**Figure 12: Correlation Heatmap for Numerical Features**

**4.6.2 Categorical Features**

In this section, we extend the correlation analysis to categorical features, utilizing Cramer's V coefficient to quantify the associations between these features. Like Pearson's correlation coefficient, Cramer's V gives insight into the strength of association between categorical variables.

Two notable correlations observed are:

* Departure hour and station group: This correlation explores the relationship between the time of departure and the grouping of stations. Understanding this association can offer insights into the scheduling and operational dynamics of last-mile delivery routes.
* Vehicle capacity and station code: Examining the relationship between vehicle capacity and station codes provides valuable insights into the allocation of resources and logistical operations at different depot locations.

A screenshot of a computer screen

Description automatically generated

**Figure 13: Correlation Heatmap for Categorical Features**

**5.0 Results and Discussion**

**5.1 Clustering Analysis**

In this subsection, we explore the results of the clustering analysis conducted on the defined features using KMeans clustering. Optimal clustering with three clusters was determined using the elbow method, followed by dimensionality reduction to two dimensions using Principal Component Analysis (PCA) for visualization.

The Elbow method plot (Figure 14) confirms the selection of three clusters, while Figure 15 illustrates the clusters in a PC1 vs PC2 plot.

Results for Each Cluster:

* Cluster 1 is predominantly characterized by high-scored routes (60%) with a notable presence of medium-scored routes (39%). These routes exhibit the least mean number of stops but have the highest mean service time and number of specified time windows. They also demonstrate the least mean travel time and the greatest number of packages per stop. Typically, these routes involve shorter distances and prioritize time-sensitive deliveries.
* Cluster 0 consists primarily of medium-scored routes (65%), with a significant proportion of high-scored routes (33%). These routes feature a low number of specified time windows and a moderate number of stops. They display the least number of delivered items, with the most travel time and the least service time. Cluster 0 routes typically cover longer distances and involve less time-sensitive deliveries.
* Cluster 2 comprises a mix of medium-scored (56%) and high-scored routes (40%). Similar to Cluster 0, these routes exhibit fewer specified time windows and a moderate number of stops. However, they differ in having more delivered packages and package counts.

A graph with a line

Description automatically generated

**Figure 14: Elbow Method to Determine Optimal Number of Clusters**

A diagram of a cluster of yellow and purple dots

Description automatically generated

**Figure 15: Cluster Analysis of Features**

**5.2 Gradient Boosted Classifier**

Based on the results of the gradient boosted classifier, it seems that the model performs well in classifying routes with a medium efficiency level, achieving a precision of 0.65 and a recall of 0.79. This indicates that the model effectively identifies and correctly predicts a significant portion of medium-scored routes. However, the precision and recall for high-scored routes are comparatively lower, with values of 0.66 and 0.50, respectively. This suggests that while the model performs reasonably well in identifying high-scored routes, there is room for improvement in correctly classifying them.

It's concerning to note that the classifier fails to predict any low-scored routes correctly, as indicated by the precision and recall values of 0.00. This indicates a significant limitation of the model, particularly in identifying routes with low efficiency levels which can be attributed to the data imbalance.

The model's overall accuracy is 0.65, which suggests it performs moderately well in correctly classifying routes across all efficiency levels. However, it's essential to consider the imbalanced nature of the dataset, particularly the limited representation of low-scored routes, which may impact the model's performance.

Reviewing the confusion matrix, it's evident that the model struggles the most with correctly identifying low-scored routes, as indicated by the large number of false negatives (10 out of 12 actual low-scored routes are misclassified as high-scored). There are also notable false positives for high-scored routes, with 413 instances incorrectly predicted as medium-scored.

A diagram of a chart

Description automatically generated with medium confidence**A screenshot of a computer screen

Description automatically generated**

**Figure 16: Confusion Matrix and Metrics for Test Set**

A graph with blue squares

Description automatically generated

**Figure 17: Gradient Boosting Classifier Feature Importance**

**5.3 Random Forest Classifier**

A blue and white diagram

Description automatically generated**A screenshot of a graph

Description automatically generated**

**Figure 18: Confusion Matrix and Metrics for Test Set**

A graph of a number of blue bars

Description automatically generated

**Figure 19: Feature Importance Plot of Random Forest Classifier**

Comparing the results of the random forest classifier with the gradient boosted classifier, we observe similar performance metrics. The precision, recall, and F1-score for each efficiency level are comparable between the two models, with both achieving an overall accuracy of 0.65.

However, like the gradient boosted classifier, the random forest classifier also struggles with correctly identifying low-scored routes, with a precision and recall of 0.00. This indicates a common limitation across both models in effectively classifying routes with low efficiency levels.

Examining the confusion matrix, we see a similar pattern of misclassifications, particularly with false positives for high-scored routes and challenges in correctly identifying low-scored routes.

Overall, while both models demonstrate moderate performance in classifying medium-scored routes, there is room for improvement, especially in accurately identifying low-scored routes. Further refinement and potentially addressing the class imbalance issue could lead to enhanced performance in last-mile delivery route efficiency classification.

**5.4 Discussion**

The insights gained from the feature importance analysis of the gradient boosted and random forest classifiers offer valuable guidance for optimizing delivery route planning and decision-making processes.

For the gradient boosted classifier, the most influential features include Total Travel Time, Number of Specified Time Windows, and Used Capacity. This underscores the significance of factors related to time efficiency and resource utilization in determining route scores. These insights highlight the importance of prioritizing efficient travel time management and optimizing capacity utilization to enhance operational efficiency and customer satisfaction in last-mile logistics.

In contrast, the random forest classifier assigns relatively equal importance to all features, with Total Travel Time, Travel/Delivery Time, and Total Delivery Time being the most prominent. While this model lacks the distinct prioritization observed in the gradient boosted classifier, it still emphasizes the critical role of time-related factors in route efficiency classification.

During exploratory data analysis (EDA), key findings underscored the importance of understanding variations in total travel time and the number of stops for optimizing route planning and resource allocation. Additionally, insights into the relationship between package scan status and route scores highlight the need for effective strategies to minimize rejected deliveries and optimize operational performance.

Furthermore, understanding the factors influencing total service time and their impact on route scores enables the refinement of predictive models and the development of targeted strategies to enhance operational performance and customer satisfaction.

Overall, these insights emphasize the importance of leveraging data-driven approaches to gain a deeper understanding of route efficiency determinants and inform decision-making processes in last-mile logistics. By prioritizing features that contribute significantly to route scores and addressing operational inefficiencies identified through EDA, logistics operators can optimize delivery routes, improve service quality, and enhance customer satisfaction.

**6.0 Conclusion**

In conclusion, our data-driven exploration of last-mile delivery optimization has provided valuable insights into the factors influencing route efficiency. Through the utilization of machine learning techniques such as gradient boosted and random forest classifiers, we have uncovered key features that significantly impact route scores, including Total Travel Time, Number of Specified Time Windows, and Used Capacity.

The findings from our exploratory data analysis (EDA) underscore the importance of understanding variations in total travel time, the number of stops, and package scan status for effective route planning and resource allocation. These insights offer actionable strategies for optimizing operational efficiency and enhancing customer satisfaction in last-mile logistics.

By leveraging data-driven approaches, logistics operators can make informed decisions to refine predictive models, optimize delivery routes, and improve overall service quality. Moving forward, continued collaboration between data scientists and logistics experts will be essential for driving innovation and achieving excellence in last-mile delivery operations.

**7.0 References**

Daniel Merchán, Jatin Arora, Julian Pachon, Karthik Konduri, Matthias Winkenbach, Steven Parks, Joseph Noszek (2024) 2021 Amazon Last Mile Routing Research Challenge: Data Set. Transportation Science 58(1):8-11. https://doi.org/10.1287/ trsc.2022.1173

Keogh, E. (2011). Instance-Based Learning. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164- 8\_409

Singh, Ankur. (2023). Understanding Last-Mile Delivery: An Analysis of the Amazon Last Mile Routing Dataset. 10.13140/RG.2.2.14352.99841.

Wenyi, L. (2023). Research on vehicle routing problem and application scenarios. *Handbook of Mobility Data Mining*, 63–88. <https://doi.org/10.1016/b978-0-323-95892-9.00006-1>

Woods, C. (2023, June 14). *What is route planning?*. Locate2u. <https://www.locate2u.com/route-planning/what-is-route-planning/#:~:text=Route%20planning%20refers%20to%20the,deliveries%20faster%20and%20more%20efficiently>.