Capstone Project Report

Predictive Analytics for Optimizing Emergency Department Operations

Jean Applys Cherizol & Junior Zephir

Github: https://github.com/jeancheri/EDPredictiveEfficiency

Project Statement

Have you ever experienced a long wait time at an emergency department? If so, we suspect you might face inevitable frustration. Emergency departments know this is an ongoing problem and have optimized their resources to help patients cope with long waiting and visit times. The objective of this project is to use predictive analytics on wait times (WAITTIME) and length of visits (LOV) to optimize Emergency Department operations.

For this project, we used emergency service data provided by the National Ambulatory Medical Care Survey (NAMCS). Our approach consists of creating four models using supervised learning techniques. We propose to answer the following questions:

1. Can we effectively predict WAITTIME in emergency rooms to better manage patient flow and optimize the allocation of resources?
2. Can we predict the Length of Visit (LOV) in the emergency department?

We integrated regression and classification methods for each question to provide comprehensive insights. For WAITTIME, we developed a regression model to predict WAITTIME duration in minutes and a classification model to determine whether WAITTIME is considered normal (<=30 minutes) or high (> 30 minutes) based on a predefined threshold.

Likewise, for the Length of Visit (LOV), we developed a regression model to estimate the LOV in minutes. We also defined a classification model to predict whether the LOV is normal (<=120 minutes) or high (>120 minutes) based on a specific threshold for classification.

Broader Impact Statement

Developing and applying these four prediction models aims to increase emergency services' operational effectiveness. These methods have the potential for long-term impact on numerous benefactors such as healthcare administrators, physicians, emergency response coordinators, legislators, government health agencies, academics, and students in medicine and health informatics.

Through the implementation and integration of this project, Emergency Departments (ED) may be able to use pertinent information to optimize the provision of emergency services. When fully implemented, emergency departments can use fewer resources to predict WAITTIME and LOV. This provides patient awareness of the estimated wait times and duration of stay in the emergency room.

Ethical Concern

We have attempted to ensure the reliability and fairness of our forecasts in optimizing emergency services (ES) operations through carefully selected and verified features to avoid introducing biases linked to bias-sensitive variables, such as demographic indicators.

We could not establish a continuous feedback mechanism involving emergency service personnel to capture real-time information on the model's practical effectiveness and fairness. Incorporating feedback mechanisms could allow us to adapt to new models and changing patient demographics through regular updates and retraining cycles. The Centers for Disease Control and Prevention (CDC) and Health Insurance Portability and Accountability Act (HIPAA), where we obtain information, have already complied with Data Privacy and Compliance.

Methodology

Data cleaning and preprocessing: We cleaned the data using different strategies for dealing with null values, NaN, including the distribution of features and outliers, among other strategies. We applied different engineering techniques, like transforming some features and creating new ones. We then transformed our data by creating processing tools that can be used separately by different data splits. The data split occurs first, before normalizing or applying any imputation, such as filling NaN with mean, to prevent data leakage. We created and used a configuration file, config.py, which manages the relatively different paths in our project. The config.py also contains different global variables, like random_seeds, to ensure reproductivity while using the data and the models.

Model selection: We tried different regression and classification models, capitalizing on the strengths of every technique given the character of our independent variable and the prediction targets. Ultimately, the CatBoostRegressor and LGBMRegressor emerged as models with higher performance for the metrics used at some stage in the validation phases.

Features selection for WAITTIME: After training and selecting the best models to predict the “WAITTIME,” we chose the twenty most essential features using Shapley Additive exPlanations (SHAP). SHAP is a method from game theory that assigns each feature's contribution to the model's output, enhancing interpretability by quantifying how significantly each feature affects the predictions. This precise attribution helps refine the model by focusing on essential features, thereby improving the effectiveness of some features in conducting predictive analytics in emergency department operations.

Features selection for “LOV”: We used best_model.feature_importances_ from our best-trained model. The process extracts the important features from the best model and indicates how much each feature contributes to the model's predictions. These features are then normalized by dividing by their sum to ensure they sum within zero to one, making them easier to compare on a consistent scale. A threshold is applied to filter out features with relatively low importance (in this case, greater or equal to 0.01), focusing on those most significantly influencing the model's output.

Evaluation Strategy

We measured the effectiveness of the regression models using R-squared (R^2) and the Mean Absolute Error (MAE) for each WAITTIME and LOV. For the classification Models, we created new target variables WAITTIME_BINARY and LOV_BINARY to classify WAITTIME and LOV as `normal' or 'high' primarily based on predefined thresholds. We measured the effectiveness of the classification models with precision, recall, F1-score, and ROC.

To evaluate the models' capability to predict unseen data without any data leakage, after cleaning the data and applying the same features engineering, we then split our data into three categories:

1. train data (70%)
2. validation data (15%),
3. test data (15%)

After splitting the data, we applied normalization techniques and filled in some missing data with the mean. We then utilized the transformed data in the training process, by training different models and picking the most performant ones using cross-validation techniques, such as.

After selecting the best models, we performed hyperparameter tuning. We then evaluated the trained best model with its best parameters. Lastly, we used the test data—other unseen data—to confirm the performance of our model and its ability to predict and work well on unseen data.

WAITTIME and LOV Prediction: The model's success depends on its ability to predict whether the wait time will be normal or high and for how long. Metrics include R-squared (R^2), the Mean Absolute Error (MAE), precision, recall, F1-score, and universal accuracy.

Data Collection and Cleaning

The dataset used for this project was collected from the 2014 National Hospital Ambulatory Medical Care Survey (NHAMCS). This data set focuses on Emergency Department (ED) visits across various hospitals in the United States. NAMCS is a national probability sample survey conducted by the Division of Health Care Statistics, National Center for Health Statistics (NCHS). The dataset and specifications are a public use file and can be found and downloaded here. The files are in SAS format (sas7bdat).

The original dataset contains 23,844 records, and more than 1012 features capturing various aspects of ED related to each visit, such as patient demographics, visit characteristics, diagnostic tests, procedures, medications, and hospital-level information.

In this project, we implemented a comprehensive data-cleaning process to ensure the reliability and validity of our analysis. This process was executed using a custom Python class, DataCleaner, designed to handle various aspects of data preprocessing required for our dataset.

Below is a detailed explanation of each step taken during the data-cleaning session:

- **Conversion of Mixed-Type Columns**: We standardized columns containing mixed data types to prevent mismatch errors during analysis. This step ensures consistency across our dataset and allows for more accurate data manipulation and analysis.
- **Removal of Trailing Zeros**: Numeric columns were evaluated, and trailing zeros were removed. This is to convert floats to integers wherever applicable and simplify data representation without losing precision.
- **Cleaning Byte-Type Columns:** Columns stored as byte types due to encoding issues were converted to strings, ensuring all textual data is consistently handled and analyzed as string objects.

- **Handling Negative Values**: Negative values in numeric columns, which could represent missing or erroneous data, were replaced with NaN to indicate missing information, ensuring the integrity of our statistical analysis.
- **Outlier Removal**: Outliers for the critical target variable were identified using the interquartile range (IQR) method and removed to prevent skewed analysis results. This step is crucial for maintaining the robustness of our predictive models.
- **Dropping Unnecessary Columns**: Columns that only contained missing values were removed from the dataset, as they provide no value to our analysis. Also, columns matching specific patterns deemed irrelevant to our study (e.g., specific medication identifiers) were dropped.
- **Combining Medication Columns:** Medication columns were consolidated into a single RX_combined column, aggregating all medication-related information into a unified field. This transformation simplifies the dataset and focuses the analysis on relevant medication data.
- **Filling Missing Values:** For categorical variables, missing values were filled with the most frequent categorical values, and infinite values in the dataset were replaced with NaN to standardize missing data representation. We used the median metric for numerical features to fill in the missing values.
- **Dropping Columns by Pattern:** Specific patterns in column names (e.g., COMSTAT\d+$ or GPMED(27|28|29|30)$) were identified, and columns matching these patterns were dropped from the dataset to eliminate unnecessary complexity and focus on the most relevant features.

This structured approach to data cleaning addressed several challenges inherent in working with real-world healthcare data, including handling mixed data types, managing missing values, and consolidating medication information. By carefully preparing our dataset through these steps, we ensured that the subsequent analysis could proceed smoothly and accurately, providing meaningful insights into optimizing emergency department operations.

Exploratory Data Analysis

To explore our dataset, we conducted appropriate univariate (focused on understanding individual variables) and bivariate (examined the relationships between two variables) analyses on each attribute to grasp the dataset's structure and its relationship with hospital length of visit (LOV) and the wait time (WAITTIME). For categorical data, we counted how often each category occurred, while for numerical data, we looked at measures of spread (like variance, standard deviation, and interquartile range) and central tendency (such as mean, median, and mode).

	LOV	WAITTIME
count	17959.00	17959.00
mean	170.44	36.52

std	104.77	45.48
min	0.00	0.00
25%	89.00	9.00
50%	148.00	21.00
75%	232.00	46.00
max	504.00	474.00

Table 1: Descriptive Statistical results of LOV and wait time

Figure 1 provides a snapshot of these exploratory findings for the distribution of LOV and WAIT TIME from the raw dataset. The points form an “S" shape, indicating that LOV and WAITTIME distributions have heavier tails than a normal distribution. This means there are more extreme values (low and high) than expected if both were normally distributed.

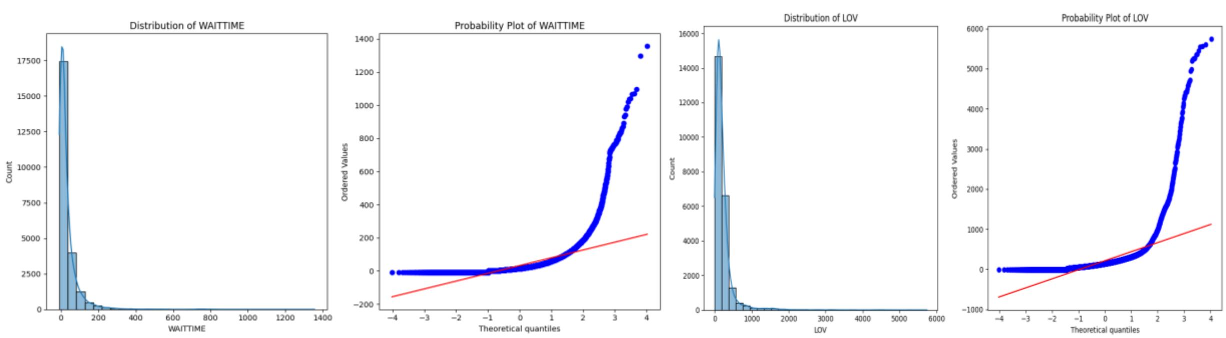


Figure 1: Distribution of LOV and WAIT TIME with outliers

Figure 1a shows the distribution similarity for length of visit and wait time between male and female patients. This might suggest that the triage process and subsequent treatment are standardized and not heavily influenced by the patient’s sex.

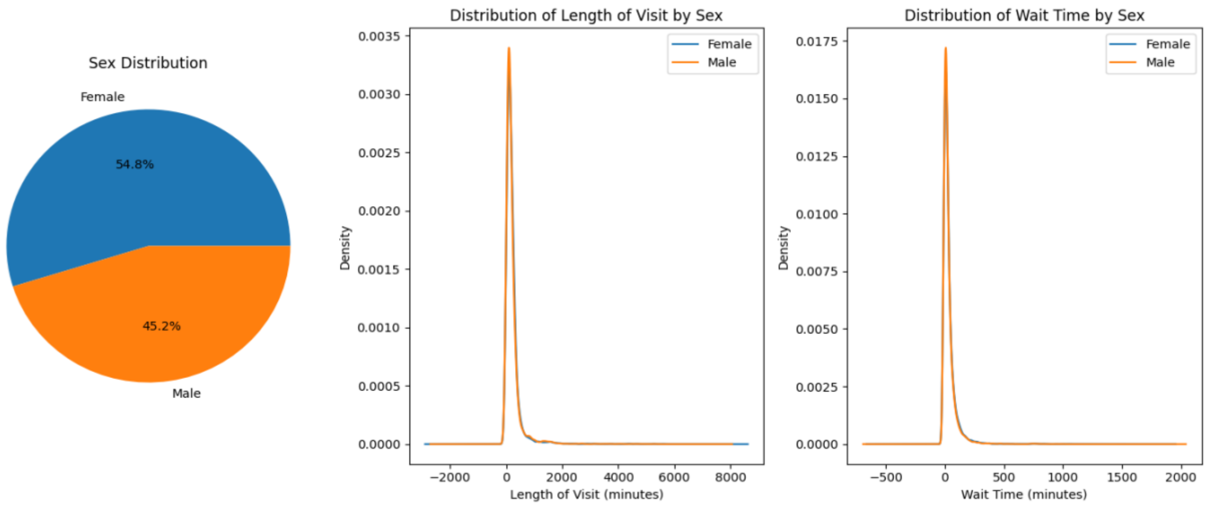


Figure 1a: Distribution of LOV and WAIT TIME by sex

Before going into our analysis, we will examine the distribution of Length of Visit (LOV) after filtering out any outliers to ensure a clear understanding of the data

Figure 1b shows the distribution of Length of Visit (LOV) in hours after removing outliers. It is right-skewed, with the majority of visits being shorter. Most LOV values cluster on the lower end of the scale, with a mean LOV of 2.84 hours, as indicated by the dashed red line. The frequency decreases as the LOV increases, showing that longer visits are less common.

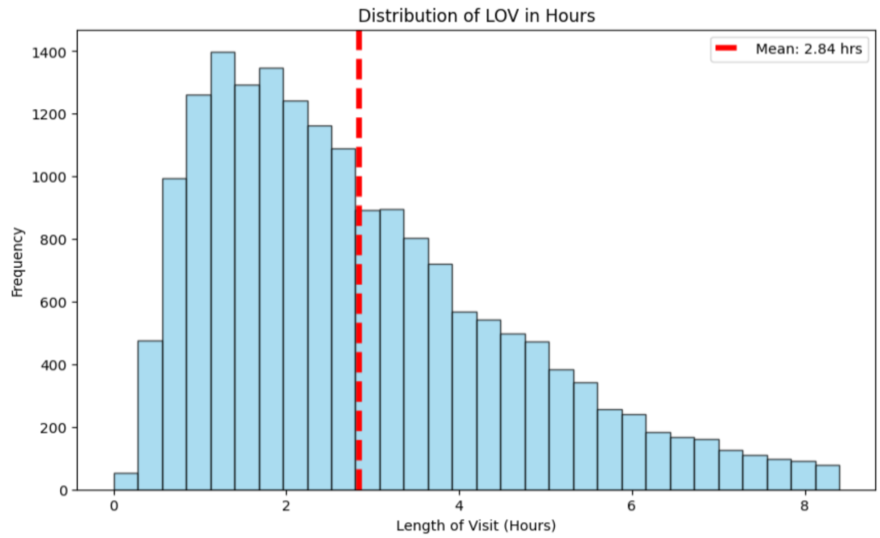


Figure 1b: Distribution of LOV values in hours without outliers

∉ In Figure 2, the data shows that both waiting time and length of visit at ED facilities are longer for older and younger patients. The oldest 75+ age group experiences the most extended length of visit visits. The figure also indicates that the waiting time generally varies with different age categories. The 35 to 54-year-old age group has the longest average wait time, 35 minutes.

- The Under 1 and 1-14 age groups have the shortest wait times, estimated at 25 minutes.
- The length of the visit increases with age. The 75+ age group has the most extended average visit length of over 200 minutes.
- Younger age groups, such as those under 1 through 15-24, have shorter average visit lengths, between 100 and 150 minutes.

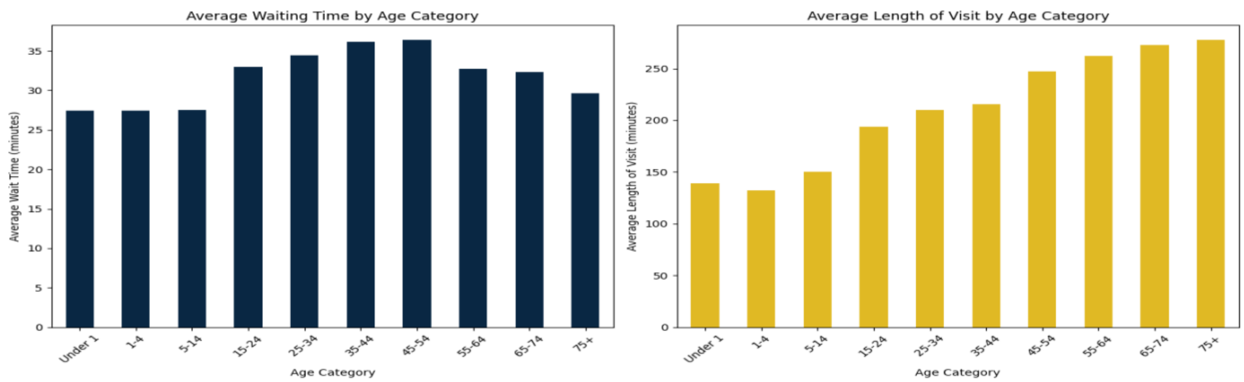


Figure 2: Average waiting time and average length of visit by age category

Figure 3 demonstrates the Time Series Distribution of LOV over Months, tracking the average Length of Visit (LOV) to an emergency department during 2014.

As we can see, the shortest wait time occurred in July while the longest wait time occurred in August.

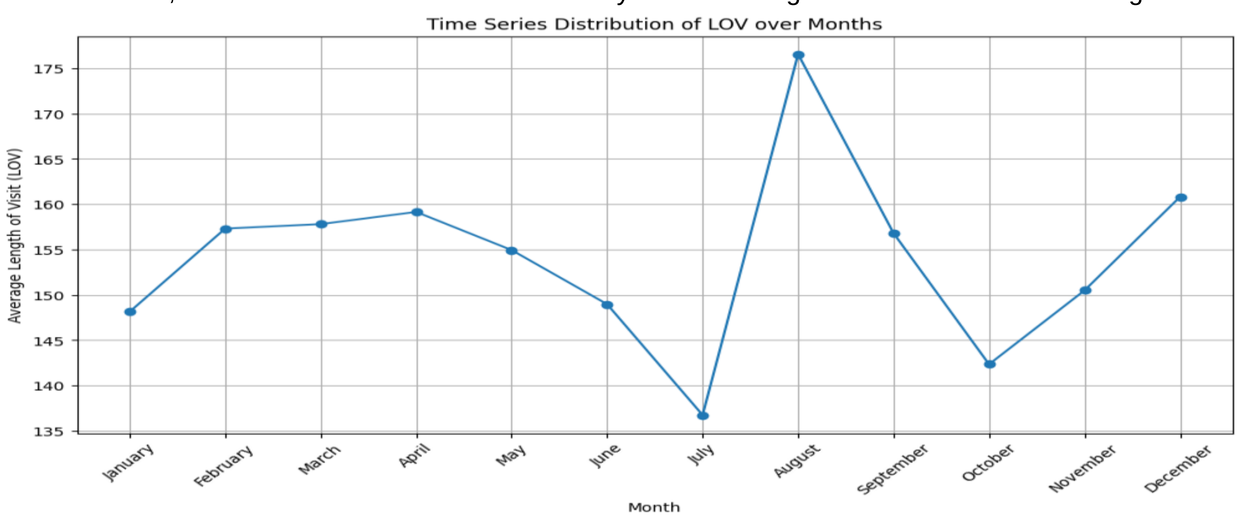


Figure 3: Time Series Distribution of LOV over Months for the year 2014

Technical depth - WAITTIME

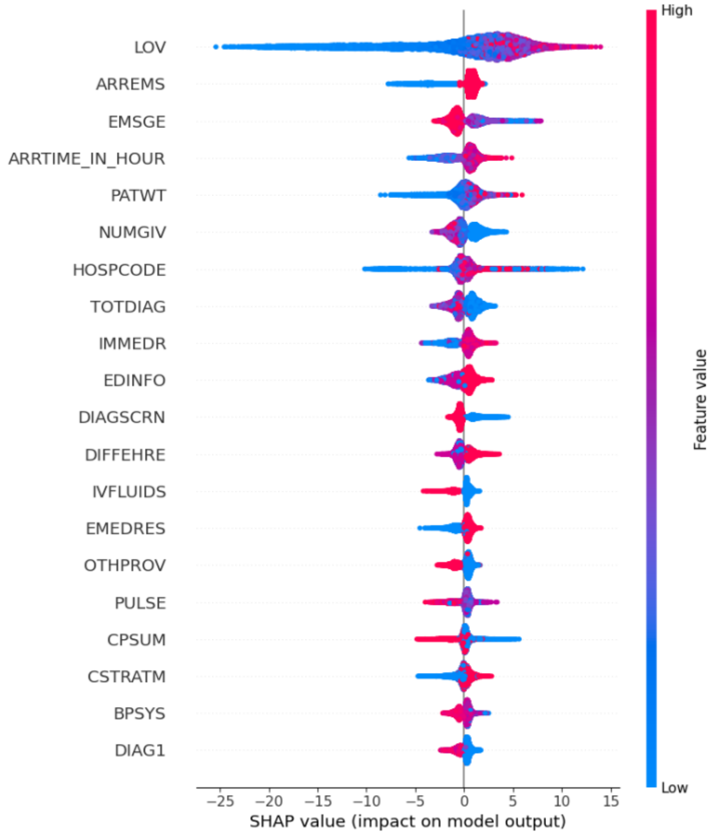


Figure 4: Top features selection

Summary for WAITTIME Prediction

For WAITTIME prediction, LGBMRegressor is the most powerful, acceptable, and predictable model. LGBM's mean Absolute Error (MAE) is 15.13. It is moderate because it is more than half the mean WAITTIME (26.47) and very close to the median WAITTIME. The model's regression score has an R-squared (R2) of 0.29, which means it explains 29% of the variability in WAITTIME. This is due to the need to include background knowledge when dealing with ED features, which could be improved by using more sophisticated modeling techniques to make finer distinctions between artworks.

We tried different *n* values to determine the number of features (n) of the most influential features as shown in Figure 5 below. MAE and R2 do not increase above and below 50. As the highlighted plot shows, the performance was optimal when n equals 50. We did hyperparameter tuning on the best model via cross-validation, which optimized the performance of both MAE and R2.

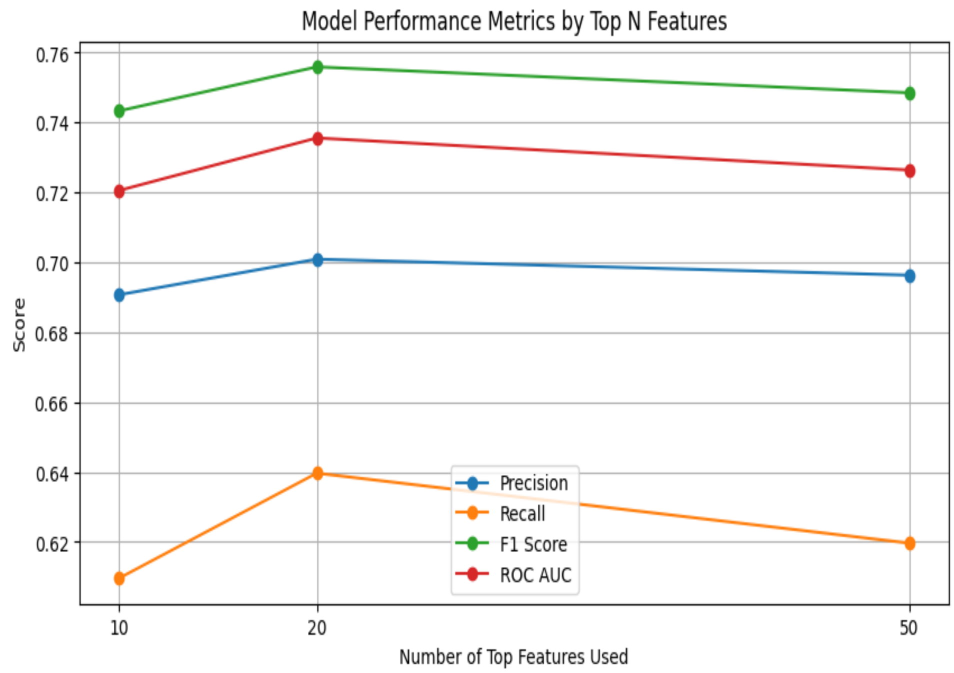


Figure 5: Selection of top 20 features

Summary for classification of WAITTIME as normal or high

Predicting whether WAITTIME is normal or high shows that the most effective model in classifying WAITTIME_BINARY was the CatBoostClassifier. Indeed, when the value of WAITTIME is less than or equal to 30 min, WAITTIME is considered normal(class 0), and when it is greater than 30 min, it is considered high(class 1). The model differentiates subtly using the cutoff point.

There is a precision vs. Recall trade-off in CatBoostClassifier - for class 0 (Normal WAITTIME), the precision is 0.83, indicating that the model is right most of the time when it predicts normal WAITTIME. RECALL is 0.70, indicating that the model missed 30% of normal WAITTIME. For Class 1 (High WAITTIME), Precision now falls to 0.60, indicating that the model needs to be more accurate at predicting high WAITTIME. Still, the recall goes up to 0.75, indicating that the model is good at picking up true cases of high WAITTIME.

Confusion Matrix—Based on the confusion matrix, the model shows high true positive and negative rates. The graph indicates that the model can classify WAITTIME_BINARY. In addition, the graph indicates that the model may sometimes misclassify WAITTIME.

Heatmap: In the given heatmap, the classification report of the model is represented in a graphical format; classification "0" represents (Normal WAITTIME) which has high precision and F1-score. However, class 1 (High WAITTIME) has relatively low precision and F1-score. There may be a model bias toward favoring the class of Normal.

Summary for LOV Prediction

The NHAMCS dataset contains LOV values for each patient, allowing us to use regression and classification techniques to build prediction models. For the regression technique, we assessed various models and identified the CatBoost Regressor as the most accurate. This model provides a Mean Absolute Error (MAE) of about 51 minutes, corresponding to about 30% of the average visit length, positioning it as a relatively minor error given the broad range of data. Moreover, 55.6% explains the variability in LOV (as indicated by the R-squared value of 0.556), suggesting it captures more than half of the underlying patterns, which is commendable though not exhaustive. This performance suggests that the CatBoost Regressor can be a valuable tool for operational planning in emergency department (ED) settings, particularly for managing resources and scheduling in emergency departments. However, we recognize our predictive model's limitations and need to consider additional enhancements to improve its reliability and application in critical decision-making processes.

Figure 6 visually represents our model's predictive accuracy by comparing the actual versus predicted values. This comparison is key to evaluating the model's performance and understanding its predictive capabilities in practical scenarios.

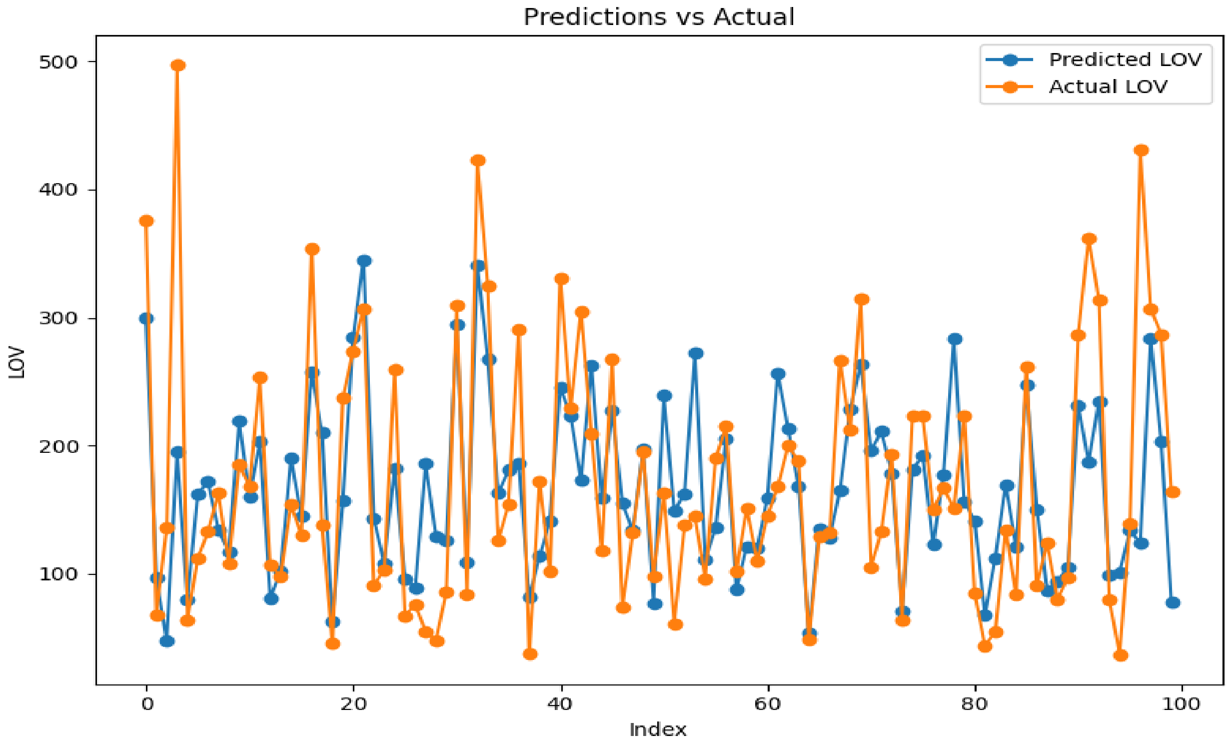


Figure 6: Predict versus actual LOV values

Our preliminary analyses indicated that classification techniques performed better than regression-based approaches based on our dataset, and classifiers may also be more useful for ED workflows.

Prediction for classification tasks

There are many machine learning models commonly used for binary classification tasks. Some of these models have simpler forms, while others are more complex. For instance, the logistic regression model relies on the logistic (sigmoid) function and ensembles such as the Decision Tree.

Prediction Model Results

We used the Scikit-learn Python library in our analysis, where we evaluated the performance of our model by using precision, recall AUC, and F1-score metrics, which can be defined as:

$$precision = \frac{TP}{TP + FP} \qquad recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad F1 - score = 2 * \frac{precision.recall}{precision + recall}$$

TP, TN, FP, and FN refer to the number of true positive, true negative, false positive, and false negative classifications, respectively. Note that we use weighted F1- 11 scores, the average F1 Score calculated for each label is obtained by considering the proportion of each label in the dataset. Precision-recall and ROC curves can be used to understand the trade-offs in model performances towards particular classes.

Model Name	**Accuracy**	**F1 Weighted**	**Precision**	**Recall**	**ROC AUC**	Is best model
CatBoostClassifier	0.8245	0.8241	0.8223	0.8245	0.8994	Yes
LGBMClassifier	0.8225	0.8222	0.8216	0.8225	0.8961	No
XGBClassifier	0.8129	0.8087	0.8088	0.8111	0.8752	No
XGBClassifier (Updated)	0.8163	0.8159	0.8157	0.8163	0.8906	No
Gradient Boosting Classifier	0.8111	0.8089	0.8090	0.8111	0.8760	No
RandomForestClassifier	0.8185	0.8018	0.8179	0.8185	0.8913	No
Random Forest Classifier	0.8142	0.8136	0.8133	0.8142	0.8892	No
MLPClassifier	0.7886	0.7909	0.7905	0.7846	0.8607	No

Table 2: Models’ metrics comparison for LOV

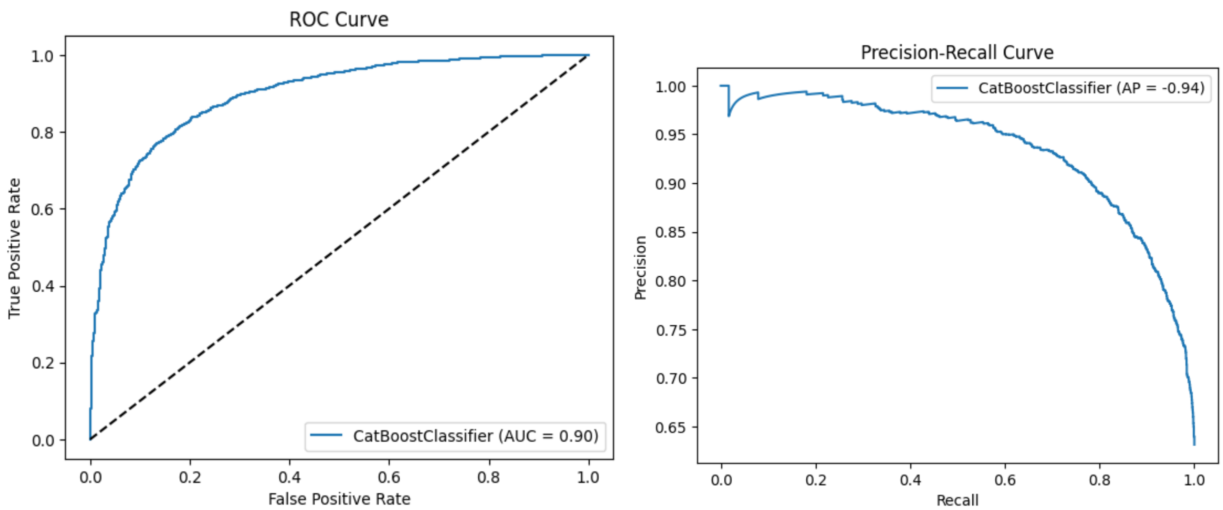
CatBoostClassifier is the best model based on evaluation and performs better compared to other models based on several factors.

The CatBoostClassifier, shown in Figure 7, paints a picture of a high-performing model adept at LOV binary classification. With an AUC on the ROC curve at 0.90, the model distinguishes between classes with high accuracy. This near-perfect AUC score indicates the model's strength in prediction accuracy.

Furthermore, the Precision-Recall curve, with an average precision score of 0.94, reflects the model's prediction accuracy across a range of thresholds. This score indicates a model that not only returns relevant results but also does so with a low rate of false positives and negatives.

Supporting metrics reinforce the model's solid performance: accuracy of 81.67%, precision of 88.49%, and an F1 Score of 84.90%. The model demonstrates a strong balance in identifying true positives and negatives.

The CatBoostClassifier's performance is robust across various metrics, highlighting its utility and accuracy in LOV binary classification tasks and making it a valuable analytics tool for predicting the LOV for the ED.



(a): ROC Curve

(b): Precision-Recall Curve

Figure 7: ROC and Precision-recall curves (plotted using 5-fold cross-validation)

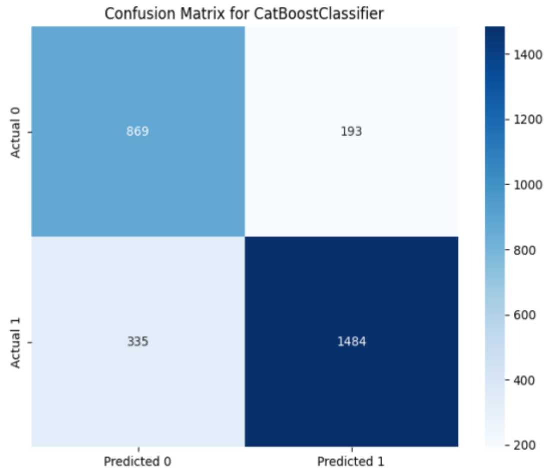


Figure 8a: Confusion Matrix for LOV

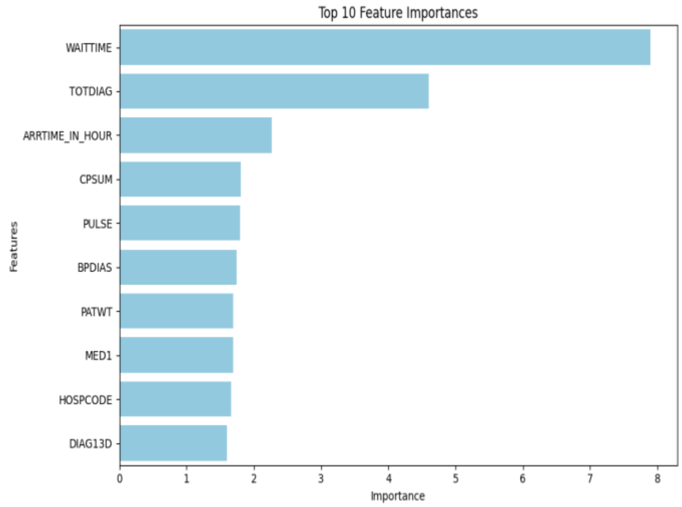


Figure 8b: Top 10 features importance for LOV

LOV values for normal (<120 min) patients have a very wide range, demonstrating the difficulty of discriminating between normal and high lengths of Visit within the dataset. On the other hand, the mean LOV value for patients with a falsely predicted length of Visit (i.e, FN) was still substantially shorter than patients with true normal visits predictions (mean LOV 276 vs 262.8 hours), suggesting that models offer meaningful information even when predictions are “incorrect.”.

Outcome for LOV

Our study provides insights on several aspects of using machine learning-based predictive modeling for LOV prediction. We used a confusion matrix to evaluate the performance of our best model (CatboostClassifier) on a classification task. Here are the insights from the graph:

- The confusion matrix displays the number of true positives, true negatives, false positives, and false negatives.
- A high number of true positives (1,484) and true negatives (869) suggests that the model has a strong classification ability for both classes.
- The relatively lower number of false positives (193) (type I error) and false negatives (335) (type II error) indicates the model's efficiency in minimizing incorrect classifications.
- The model is more conservative with the positive class, with fewer false positives than false negatives, suggesting a potential focus on precision over recall.

While the model demonstrates excellent predictive performance with a strong ability to recognize true classifications and a high importance placed on 'WAITTIME' and “TOTDIAG” in determining outcomes, there is still room for improvement, especially in reducing the number of false positives and false negatives. This could be addressed by further model tuning, class rebalancing techniques, or gathering more representative data.

Translation of WAITTIME outcome into Insights

1. **MAE and R2 for WAITTIME**: The LGBMRegressor model is the most accurate tool for predicting waiting times (WAITTIME) within the Emergency Department (ED). With a Mean Absolute Error (MAE) of 15.13 minutes, the model's predictions are typically within 15 minutes of actual waiting times. This level of accuracy can be attributed to the model's ability to consider multiple factors simultaneously, such as a patient's understanding, arrival time, and other top predictors.

   Furthermore, the LGBMRegressor R2 value of 0.29 indicates that the model can explain 29% of the variability in WAITTIME. This means the model can identify significant factors influencing waiting times, such as staffing levels, patient volumes, and other operational factors. By understanding these factors, ED administrators can utilize the LGBMRegressor's predictions to tailor staffing levels during anticipated peak and off-peak periods. For example, if the model predicts longer wait times, administrators can effectively deploy additional staff to manage the increased patient load. Similarly, staffing levels can be scaled back during shorter predicted periods, potentially reducing labor costs.

2. **Addressing False Positives with the Current Model for WAITTIME:** Our wait time prediction system is designed to provide accurate estimates to our customers. Our model has shown a precision rate of 0.83 for Class 0, representing normal wait times. This means that our system has an accuracy rate of 83% in predicting normal wait times. We can optimally minimize false alerts and allocate resources by predicting high wait times. We take pride in providing our customers with the best possible experience, and our wait time prediction system is crucial.

3. **Addressing False Negatives with the Current Model for WAITTIME:** Regarding the prediction of high wait times (Class 1), the model exhibits a respectable recall rate of 0.75, effectively identifying the high wait times and mitigating the possibility of understaffing during crucial scenarios. Nevertheless, there is still an opportunity for enhancement, as there remains a probability of overlooking certain high wait times.

Translation of LOV outcome into Insights

The Mean Absolute Error provides an average of how much the predicted LOV deviates from the actual LOV. Using the Mean Absolute Error (MAE) and R-squared values, we gauge the precision of our CatBoostRegressor model to aid resource optimization in the Emergency Department (ED). Our model's MAE of 51.58 signifies that, on average, the predicted length of visit (LOV) can deviate by approximately 51 minutes from the actual visit duration. While this gives us a strong baseline for understanding patient flow, the R-squared value of 0.55 indicates that there's still substantial variability in LOV that the model has yet to capture. By understanding this margin of error, the ED can adjust staffing schedules to ensure sufficient coverage during times when longer visits are predicted.

Addressing False Positives for LOV with the Current Model:

Our model has an 88% precision rate for predicting standard LOV (Class 0). This means that when our model predicts a normal range LOV, it is correct 88% of the time, which is essential for managing ED operations and patient expectations efficiently. This high precision helps us effectively reduce the occurrence of false positive instances where long lengths of visits are predicted but do not materialize ensuring judicious use of our ED resources.

Addressing False Negatives with the Current Model:

Our model also has a recall rate of 81% for predicting extended LOV (Class 1), showing it successfully identifies most instances when patients might face long visits. By focusing on minimizing false negative situations where we fail to anticipate a long LOV we can better manage periods of high patient volume and prevent understaffing, crucial for patient care and satisfaction.

Benefits for Stakeholders – ED Staff/Administration:

Accurate predictions of LOV are vital for avoiding over resources during periods of normal patient flow and ensuring adequate resources when longer waits are predicted. By prioritizing patients accurately and reducing the risk of unanticipated long LOVs, our model contributes to enhanced patient care, improved operational efficiency, and an optimized patient experience, especially for those with extended LOV. The insights gained from our predictive model are instrumental in the strategic allocation of staff and resources, ultimately benefiting the entire ED ecosystem.

Translation of WAITIME and LOV outcome into Insights

1. **Implement real-time predictive analysis:** Based on the previous results from our analysis, LGBMRegressor and CatBoostClassifier have some predictive power for WAITTIME; Additionally, we found that CatBoostRegressor and CatBoostClassifier have predictive power for LOV. Therefore, these predictive analytics should be integrated into existing ED workflows after improvement. The integration of this solution should include user interface designs, real-time data integration, and feedback loops from Emergency room staff.

2. **Continuous improvement:** Continuously update predictive models with real-time data for better predictability. Update models regularly to reflect current trends in Emergency room operations.

3. **Dynamic Resource Allocation:** This helps increase staffing levels during peak periods and quickly process non-emergency visits.

4. **Patient Flow Management:** ED can use predictive analytics to manage patient flow by ranking them based on predicted wait times, which can help reduce congestion during critical periods.

5. **Operational adjustments:** ARRTIME_IN_HOUR is a highly predictive feature that can help ED adjust staffing schedules for busy hours. However, patient characteristics and reason for visit are more important and should be used for resource allocation and triage.

6. **Triage Process:** Monitoring the significant functionalities outlined in the APPENDIX is crucial. If the foremost two characteristics can accurately predict patients in need of surgery, it may be

advantageous to incorporate them into triage systems. Doing so could facilitate the prioritization of patients, resulting in reduced wait times for those requiring surgery urgently.

7. **Expectation Setting—Improving Communication:** ED can use predictive values to communicate more effectively with staff and patients about the expected variation in wait times. This can help reduce negative experiences caused by the variation.

8. **Benefits for Stakeholders ED Staff/Administration:** Precisely forecasting extended WAITTIME and LOV is crucial to avoid unnecessary staffing and enhance cost-effectiveness. Prioritizing patients with utmost accuracy reduces the risk of false negatives, leading to better overall care and experience, particularly for those facing extended wait times.

Feedback from Mentor

Our mentor suggested creating a summary notebook for each main prediction question. This would summarize different scripts' training, evaluation, and testing parts.

Critical Evaluation and Limitations

Performance Evaluation: Upon close examination of the means, the regression and class models provide practical insights to some extent. Specifically, the classification models effectively differentiate between normal and high WAITTIME and LOV. Conversely, while the regression models display accurate predictions of WAITTIME and LOV duration, as evident by the MAE scores, their R^2 score is weak, indicating their incapability to account for the variance in the data.

Achievements and Challenges: Although the models showed significant improvements, they still faced difficulties handling highly diverse emergency room situations. However, the models resulted in valuable insights, despite numerous parameters and variables that fall beyond their influence due to the unpredictable nature of the emergency room environment. To achieve this, feedback and expertise from subject matter experts working in emergency rooms are essential. By leveraging the insights of these experts, we can better integrate predictive analytics into the operations of Emergency Departments, leading to improved efficiency and patient care.

Conclusion

To summarize, this project proposes four predictive analytics models that can predict WAITTIME and LOV in emergency departments (ED). These models can help with staffing adjustments, process optimization, resource allocation, and improving communication with patients. Implementing these models can enhance care efficiency, increasing patient, family, and staff satisfaction. It is important to routinely review and adjust the threshold to minimize false predictions. The application of predictive analytics in healthcare can lead to a more efficient and effective healthcare system.

Statement of work

Name	Task
Jean Applys Cherizol	All analysis and work-related work to WAITTIME
Junior Zephir	All analysis and work related to LOV

References

Reference 1 - Title: Machine Learning to improve prediction for frequent emergency department use: a retrospective cohort study. retrieved from
Link: https://www.nature.com/articles/s41598-023-27568-6

Reference 2 - Title: Remodeling Emergency Department Utilization via Predictive Analytics (ECG Management Consultants Blog). retrieved from
Link: https://www.ecgmc.com/insights/blog/2161/shaping-emergency-department-utilization-through-predictive-analytics

Reference 3 - Title: Predicting hospital admission at emergency department triage using machine learning. retrieved from
Link: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016

Reference 4 - We got help from Grammarly by using these AI prompts:
Prompts created by Grammarly
- "Shorten it"
- "Improve it"
- "Clean up notes"

Reference 5 - We got help from Chat-GPT:
- "Improve comments in our code"
- "Clean up code"

APPENDIX- Top 20 most important features

1. **LOV (length of visit):** The relative severity (or complexity) of the visit and the corresponding urgency of the care needed determine how long the stay will last. 1.
2. **TOTDIAG (Total Diagnoses):** This represents the number of diagnoses of each patient, which might affect the length of the visit: if a patient has multiple diagnoses, their visit might need more time to be completed.
3. **ARREMS (Arrival by EMS):** If people arrive in cars sent by emergency medical services, this often means greater criticality and a more extended, demanding visit.

4. **ARRTIME_IN_HOUR (Arrival Time in Hours):** The hour of the day a patient arrives could correlate with peak operation hours or busier times, thus impacting wait times.
5. **NUMGIV (Number Given):** concerns the number of interventions or medications offered—the more that have been added to a case, the more complex the treatment might be.
6. **CSTRATM (Chest Trauma)**: This applies only to patients who present with chest trauma and are felt to be at higher risk of more extended evaluation and treatment times.
7. **IMMEDR (Immediate Rooming):** Patients needing to be roomed immediately may have more pressing needs than the others and may drive up WAITTIME.
8. **DIAGSCRN (Diagnostic Screening)**: Y/N = if yes, diagnostic screenings were performed as needed before proceeding with definitive surgery, which can delay WAITTIME.
9. **PATWT (Patient Weight)**: This could be important for dose calculation, or the type of equipment required and thus affect WAITTIME.
10. **EMEDRES (Emergency Medicine Resident)**: Due to time constraints, the level of medical experience could differ from treatment idea to treatment idea.
11. **WIRELESS (wireless technology use)**: This refers to wireless communications and monitoring technology that could simplify specific underlying processes and impact WAITTIME.
12. **EMSGE (EMS place)**: WAITTIME might depend on the geography or origin of the EMS (e.g., the distance, direction, or traffic situation), implying an association or dependence between each.
13. **HOSPCODE (Hospital Code):** One hospital may treat more patients, have more dedicated personnel, or use different procedures than another; thus, HOSPCODE should be considered when assessing AVERAGE WAITTIME.
14. **FASTTRAK (Fast Track)**: Some ED have a fast-track system to help patients with less severe problems avoid WAITTIME.
15. **IVFLUIDS (Intravenous Fluids Administered):** IV fluids often correlate with disease severity and thus imply LONGER WAITTIME.
16. **CPSUM (Composite Score Summary)**: This composite score for patient acuity or triage category drives WAITTIME priority.
17. **EMRED (Emergency Readiness)**: Like EMEDRES, it may represent staffing or readiness levels affecting WAITTIME.
18. **EMSGER (EMS Geriatrics):** If this applies to geriatric patients from EMS, it means that WAITTIME will be longer due to special needs.
19. **ANYIMAGE - (Any Imaging Performed):** Imaging tests typically add to WAITTIME since there is startup time to do and read the images.
20. **RES (RESPR = Respiration Rate or Status)**: Many respiratory problems require faster interventions, so if the triage system is in place, the RES status needs to be considered when assigning a WAITTIME and determining whether the patient goes into the 'immediate' group.