UTA THE UNIVERSITY OF TEXAS AT ARLINGTON

Predictive Analytics for Incident Resolution Time

Group 4
Rushika Badri Prasad
Trupti Shailendra Gandhi
Girish Sunkadakatte Chandrappa

UTA

Problem Statement

- **Service Level Agreement (SLA) are** critical in IT Service management, which defines the required timeframes for resolving incidents based on their urgency, impact, and priority levels. Failing to meet these SLA can result in **financial penalties and weaken client trust.** Accurately predicting resolution time is challenging due to the diverse and dynamic nature of incident attributes.

Objective

- Our project aims to analyze historical incident data and **build a predictive model that classifies** incidents **resolution time required into different categories**. By analyzing the factors affecting the resolution time. Our model enables IT teams to identify high-risk tickets early, allocate resources proactively, and optimize workflows. These insights will support data driven SLA discussions with clients and help teams to improve service quality and operational efficiency.

Dataset Overview

Dataset comprises of **141,712 observations and 37 features (4 numerical and 33 categorical features).** Key predictive features influencing resolution time include operational and organizational metrics, time based and incident metadata, SLA risk indicators, detailed incident category information. These attributes play a critical role in modeling how quickly IT service incidents are resolved.

Category	Features
Operational Metrics	reassignment_count, reopen_count, sys_mod_count, time_taken
Organizational Attributes	assignment_group, assigned_to, caller_id, opened_by, resolved_by, sys_created_by, sys_updated_by
Time Based Metrics	opened_at, resolved_at, closed_at, sys_created_at, sys_updated_at, closed_code
Incident Metadata	incident_state, active, made_sla, knowledge, u_priority_confirmation, notify, location, cmdb_ci, vendor
SLA Risk Indicators	impact, urgency, priority
Identifier	number
Target	time_group
Issue details	category, subcategory, u_symptom, contact_type, problem_id, rfc, caused_by

Exploratory Data Analysis

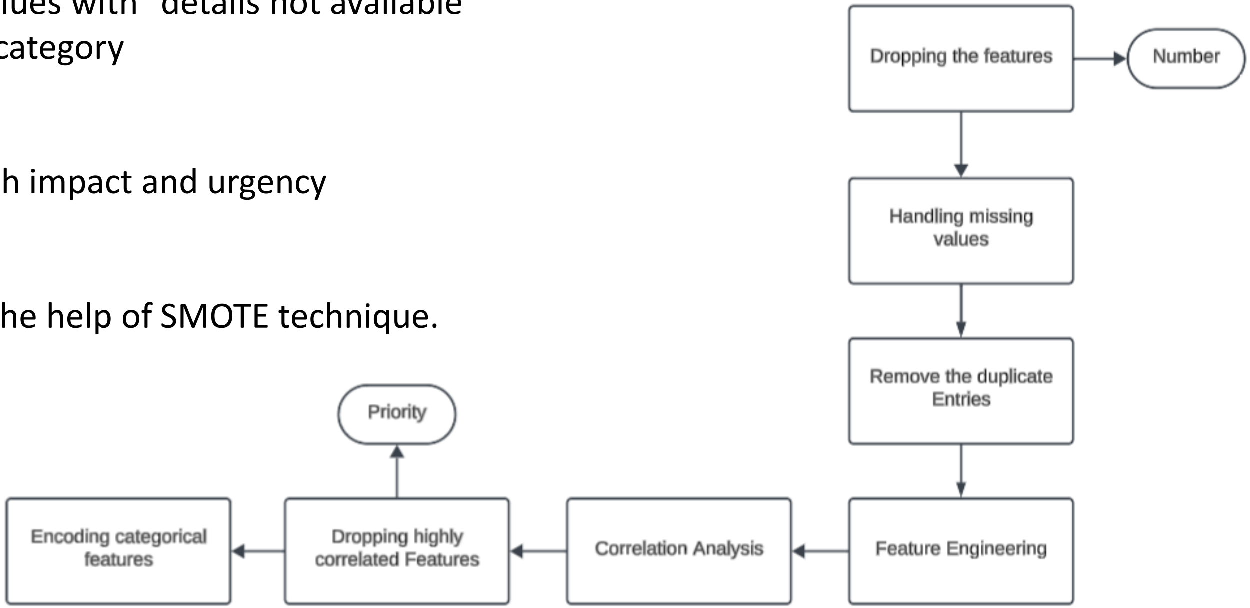
http://powerbi_dash_board.com

Data Preprocessing

- Handling missing values
  - Location based on caller id
  - Category based on subcategory
  - U_symptoms updated missing values with “details not available”
  - Assignment group based on sub category

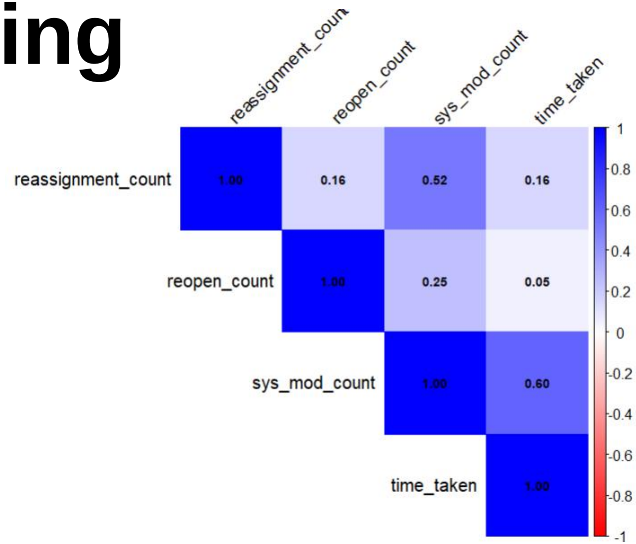
- Correlation analysis
  - Priority was highly correlated with impact and urgency

- Data Imbalance
  - Handled imbalance issue with the help of SMOTE technique.

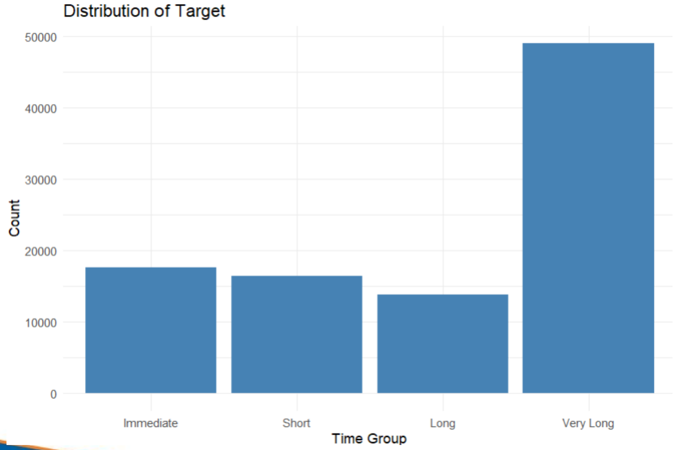


Data Preprocessing

Correlation Heat map

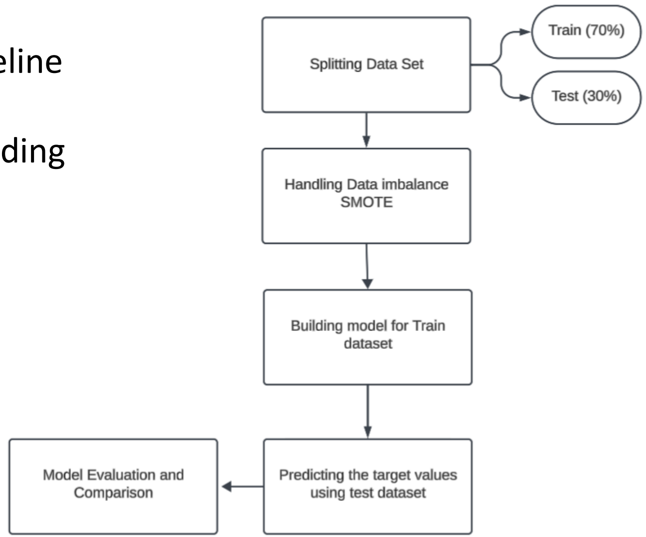


Target variable distribution



Models and Evaluation

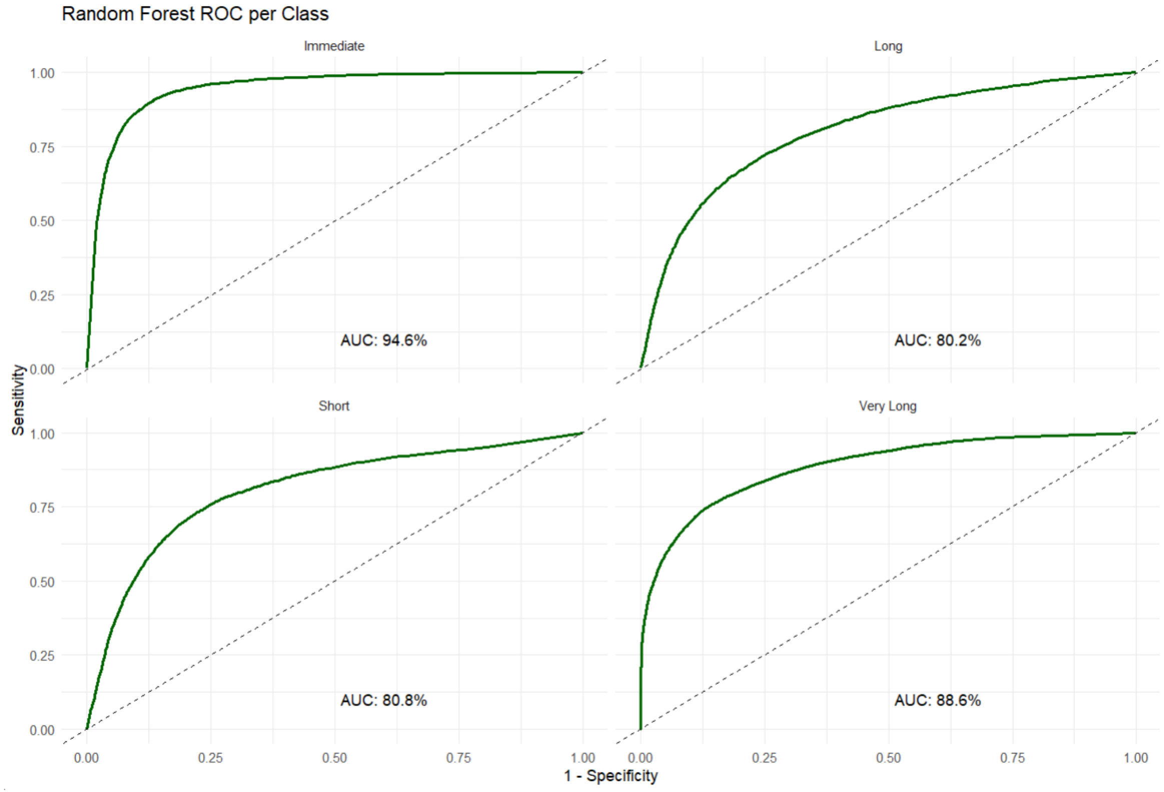
- Random Forest and XGBoost both performed significantly better than the baseline models (Decision Tree and C5.0).
- Random Forest outperformed XGBoost across all key evaluation metrics, including accuracy, precision, recall, and F1-score.
- Higher Precision and F1 Score indicates better classification and minimizing misclassification errors when compared to XGBoost model.



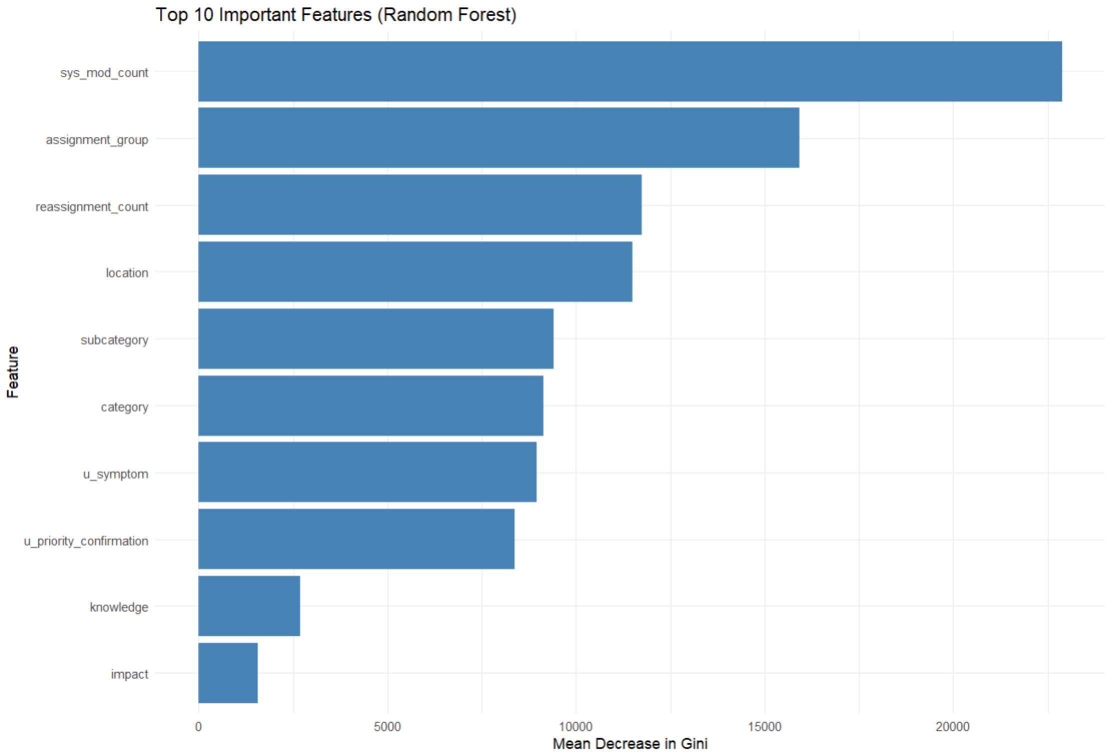
Models	Accuracy	Precision	Recall	F1 Score
Random Forest	69.90%	62.80%	61.80%	61.70%
XG Boost	64.80%	54.80%	54.20%	54.50%

UTA

ROC Curve


Random Forest ROC per Class
Immediate
Long
Short
Very Long
AUC: 94.6%
AUC: 80.2%
AUC: 80.8%
AUC: 88.6%
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00

Feature Importance



Summary and Conclusion

- Post comparing multiple classification models, **Random Forest** was the top performer across all performance metrics.
- Model demonstrated a good performance, with an overall accuracy of 69.9% and well balance between precision and recall.
- Features like **sys_mod_count, assignment group and reassignment count** are the top features having highest impact on the ticket resolution time.
- ROC AUC Scores for class in the target feature shows highest discriminative power.
- Model helps by improving SLA performance by predicting delays early, using resources more efficiently and keeping clients satisfied through data-driven decisions.

Future Scope

- **Improved Data Quality:** Complete and consistent dataset, as the current one had a high volume of missing values.
- **Quarterly Model Updates:** Continuously updating and retrain the model to reflect recent trends and maintain prediction accuracy.
- **Advanced Feature Engineering**: Use of feature selection and transformation techniques.
- **Real-Time SLA Monitoring Integration:** Integrate the model into real time incident dashboards.
- **Client-Specific SLA Reporting:** Tailor SLA risk predictions to individual client profiles.

Thank  You