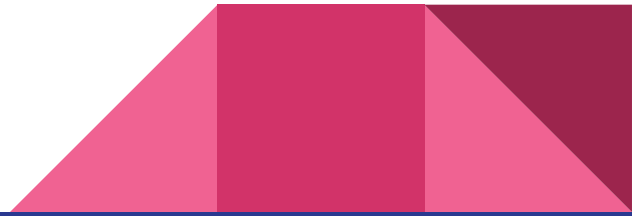
IT Incident Management Data Analytics

Udayangani Liyanaarachchi a.k.a. UL

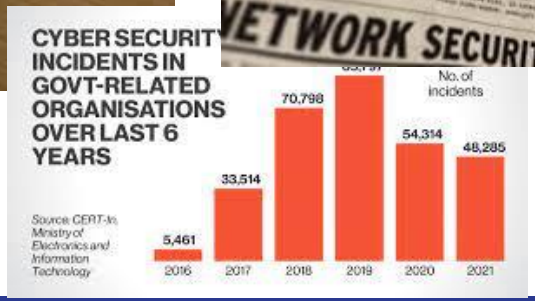
Contents

- What is an Incident
- What is Incident Management Data
- Background and Objective of the Project
- EDA & Cleansing
- Applying the models
- Model Evaluation
- Conclusion
- Next Steps



What is an Incident

An **Incident** is an unexpected occurrence of a disruption to a service or functionality of a product.



Why Manage Incidents?

Governments and Organizations today face large costs due to various incidents that occur which disrupts the service to the end users! (You, me and everyone), therefore it is important to manage incidents to;

- Better livelihood for people
- Maintain the product and service after sales service quality
- Customer satisfaction and retention
- Avoid costs to the organization due to outages
- Be competitive in the marketplace

Background and Objective of the Project

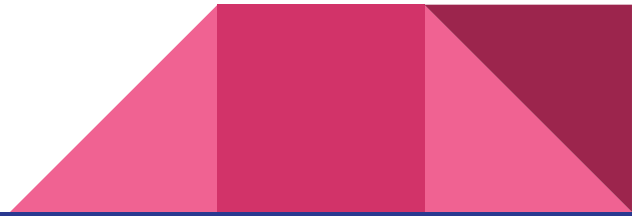
About the data set - incident management log from an IT company extracted from SNOW (ServiceNOW) combined with another relational database for enrichment

Source - https://www.kaggle.com/datasets/vipulshinde/incident-response-log

Expected Deliverables -

D1 - To create a predictive model to predict if a ticket will meet its SLA (Service Level Agreement)

D2 - To create a model to classify a ticket based on the incident features



EDA (Exploratory Data Analysis) and Data Cleansing

- 119998 Rows of data
- 36 Columns
- Most data are discrete and categorical, only 3 numeric features
- No null values found but could see '?' which concealed nulls

EDA (Exploratory Data Analysis) and Data Cleansing

- Replaced all the ‘?’ with numpy NaN
- Replaced NaN with zeros on selected columns
- Data values were strings hence removed the characters to make them numerical
- Changed the data type of the selected columns to integer and stored them as category type

EDA (Exploratory Data Analysis) and Data Cleansing

D1 - Predict SLA

Target - made_sla

features:

Active
U_priority_confirmation
sys_mod_count

D2 - Classify Incident

Target - assignment_group

features:

category
location
subcategory
u_symptoms

Applying the Models

Since the nature of data was discrete and categorical we need to use a classification model

- Logistic Regression
- SVM
- DecisionTreeClassifier
- RandomForrestClassifer

Model Evaluation

D1 - SLA Prediction -

Based on the models chosen, below results were seen and overall all the models were seen relatively accurate scoring over 90%

- Train/Test scores for all models were above 90%

- Comparison of confusion matrix

Model	TP	FP	FN	TN
LogisticRegression	1017	620	206	22157
DecisionTree	1615	450	507	27428
RandomForest	1615	450	507	27428

- Classification report shows ; 97% - 98% precision score for True and 76% - 83% for False across the above models respectively.
- Recall values between 98% - 99%

Model Evaluation

D2 - Classification based on features

Based on the 4 models used below are the scores achieved and other than logistic regression, others seem to be achieving better outcome

- Train/Test scores (when a data set with first 10000 rows were considered)

Model	Train	Test
Logistic Regression	28%	27%
SVM	68%	75%
Decision Tree	76%	69%
Random Forest	69%	76%

- Classification report was not seen as the best case for multi-class comparison however the precision scores were between 60-100% in all models.

Conclusion

- To Predict the probability of meeting SLA RandomForest Classifer was chosen as the best fitting model

- To classify the ticket to assignment group, svm was chosen as the best model

Next Steps

The above models resulted in reducing the number of IT incidents that occur by giving more insightful data and analytics around tickets being logged thereby the business operations were made smoother and better managed.

Next step will look at more other insights that can be inferred out of this data to improve the model, i.e. aging analysis, automatic chasers.

Application of model outcome to classify the tickets integrated to ServiceNow

Visualization of outcomes and predictions through reporting tools

Use Machine Learning/Generative AI to improve the model to be able to have efficient classification and issue identification using chatbot

Thank You