

Automatic Ticket Assignment

Interim Report

# 

# Index

[Team Details 2](#_Toc37546120)

[Index 3](#_Toc37546121)

[Summary of the problem statement, Data and findings 4](#_Toc37546122)

[Problem Statement 4](#_Toc37546123)

[Data & Findings 4](#_Toc37546124)

[Summary of the approach to EDA and Pre-Processing 5](#_Toc37546125)

[Visualization 5](#_Toc37546126)

[Word Cloud 5](#_Toc37546127)

[Charts 5](#_Toc37546128)

[Decide Model and Model building 6](#_Toc37546129)

[Model performance - Approaches to improve model 7](#_Toc37546130)

[Code Snippet 8](#_Toc37546131)

[Finalized results 9](#_Toc37546132)

[Link to code and references 10](#_Toc37546133)

# Team Details

|  |  |
| --- | --- |
| Charan  Harsha V Chunduri  Jayapavan P Deshpande  Richa  Soumendra Bagh  Visakh Vijayaprakash | [Charantej588@gmail.com](mailto:Charantej588@gmail.com)  [harshu244@gmail.com](mailto:harshu244@gmail.com)  [jayapavanpd@gmail.com](mailto:jayapavanpd@gmail.com)  [soumendra.bagh@gmail.com](mailto:soumendra.bagh@gmail.com)  visakhvijayaprakash@gmail.com |

# **Summary of the Problem Statement, Data and Findings**

## 

## Problem Statement

Manual assignment of incidents is time consuming and requires human efforts. There may be mistakes due to human errors and resource consumption is carried out ineffectively because of the misaddressing. On the other hand, manual assignment increases the response and resolution times which result in user satisfaction deterioration / poor customer service.

## Abstract

Applying traditional machine learning and neural network-based NLP to automatically classify tickets and assign them to the right owner in a timely manner to save effort, increase user satisfaction and improve throughput in the ticketing pipeline of an organization.

## Data & Findings

1. The dataset comprises of **8500 rows** and **4 columns**
2. All columns are of type object containing textual information.
3. There are **8 null/missing values** present in the Short description and **1 null/missing values** present in the description column
4. **Password reset** is one of the most occuring tickets which reflects in the Short description column.
5. The top occuring Description in the dataset is only the text **'the'**, which absolutely doesn't make any sense. hence by looking at the Short description of such rows reveals that these are also a category of Password reset.

### Data provided in format

CSV

### Total Records

8500

### Data Fields

|  |  |
| --- | --- |
| Short description | A summary of the issue faced by the user |
| Description | Detailed description of the issue |
| Assignment group | GRP\_0 ~ GRP\_73 (total 74 classes of Assignment group) |

### Sample data

| **Short description** | **Description** | **Assignment group** |
| --- | --- | --- |
| login issue | -verified user details.(employee# & manager na... | GRP\_0 |
| outlook | \r\n\r\nreceived from: hmjdrvpb.komuaywn@gmail... | GRP\_0 |
| cant log in to vpn | \r\n\r\nreceived from: eylqgodm.ybqkwiam@gmail... | GRP\_0 |

### Distribution of classes and Observation

1. High imbalance seen in data for target column in our dataset with GRP\_0 having highest percent of representation
2. Many classes with very little representation.
3. Null values in Data:
   1. Short description **8**
   2. Description **1**
   3. Assignment group **0**
4. Observed few Ticket having Non-English ticket descriptions

## Summary of the approach to EDA and Pre-Processing

### Cleaning processes applied

1. Removal of trailing spaces
2. Removal of line breaks and tabs (\r\n\t)
3. Removal of special characters
4. Removal of extra spaces
5. Convert the data having garbled text such as mojibake using ftfy Library.

## Visualization

# Decide Model and Model building

# Model performance - Approaches to improve model

# Code Snippet

# Finalized results

# Link to code and references