****

**Assessment Report**

on

**“Customer Support Case Type Classification”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AIML)**

By

Name : Abhishek Kumar Singh

Roll Number : 202401100400009

Section: “A”

**Under the supervision of**

“BIKKI KUMAR”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

With the increasing number of online customer interactions, categorizing support cases efficiently has become essential for providing timely and relevant assistance. This project explores the application of supervised machine learning to **predict the type of support case** (e.g., technical, billing, or general) based on the length of the support message and the response time. Automating this classification process can help support teams prioritize and route cases more effectively.

**2. Problem Statement**

The goal of this project is to classify support cases into predefined categories—**technical**, **billing**, or **general**—based on their message\_length and response\_time. Accurate classification will enable better support case triaging and improve customer service response efficiency.

**3. Objectives**

* **Preprocess** the dataset to prepare it for modeling.
* **Encode** categorical labels for classification.
* **Train** a supervised learning model (Logistic Regression) to predict the case type.
* **Evaluate** the model’s performance using appropriate metrics.
* **Visualize** classification performance using a confusion matrix heatmap.

**4. Methodology**

* **Data Collection**: The dataset was provided in CSV format containing 100 entries with three features.
* **Data Preprocessing**:
  + Checked for missing values and verified data integrity.
  + Encoded the target categorical variable (case\_type) using Label Encoding.
  + Scaled the numerical features (message\_length and response\_time) using StandardScaler.
* **Model Building**:
  + Split the dataset into 80% training and 20% testing sets.
  + Trained a Logistic Regression classifier to predict the support case type.
* **Model Evaluation**:
  + Used metrics like accuracy, precision, recall, and F1-score.
  + Visualized the confusion matrix using a heatmap for better interpretability.

**5. Data Preprocessing**

* The dataset had **no missing values**.
* The case\_type column was converted into numeric labels using **Label Encoding**.
* The numerical features (message\_length, response\_time) were scaled using **StandardScaler**.
* The dataset was split into **training (80%)** and **testing (20%)** sets to evaluate generalization.

**6. Model Implementation**

Logistic Regression was chosen for its simplicity and effectiveness in multiclass classification. The model was trained on the preprocessed data and used to predict the case\_type on the test set. The model learns the relationship between message characteristics and the likely category of the case.

**7. Evaluation Metrics**

The following metrics were used to evaluate model performance:

* **Accuracy**: Overall correctness of predictions.
* **Precision**: The proportion of predicted labels that are correct for each class.
* **Recall**: The proportion of actual labels that were correctly identified.
* **F1 Score**: Harmonic mean of precision and recall.
* **Confusion Matrix**: A visual summary of prediction results for each class using a Seaborn heatmap.

**8. Results and Analysis**

* The model showed reasonable performance on the test data.
* The **confusion matrix heatmap** helped analyze how often each class was correctly predicted and where misclassifications occurred.
* Precision and recall scores varied slightly across different case types, indicating the need for potentially more complex models or more data for better generalization.

**9. Conclusion**

The Logistic Regression model successfully classified support cases based on message length and response time. While the results are promising, further improvements could involve:

* Using more complex models like Random Forest or SVM.
* Collecting more training data to improve model performance.
* Including additional features such as message sentiment or keywords.

This project highlights the feasibility of using machine learning to automate support case categorization, which can streamline operations and enhance customer experience.

