Customer Satisfaction Prediction – Project Report

1. Introduction

In this project, we developed a machine learning model to predict **Customer Satisfaction Rating** based on historical customer support ticket data. The dataset includes customer demographic information, ticket types, resolution times, and more. Our goal is to understand what influences customer satisfaction and use this information to improve customer support operations.

2. Tools and Technologies Used

- **Programming Language**: Python

- **Data Handling & Analysis**: pandas, numpy

- **Data Visualization**: matplotlib, seaborn

- **Machine Learning**: scikit-learn

- **Model Used**: RandomForestClassifier

- **Encoding**: LabelEncoder for categorical features

- **IDE/Platform**: Google Colab / Jupyter Notebook

3. Dataset Overview

The dataset contains customer support tickets, with the following important features:

- **Customer Age, Gender**

- **Product Purchased, Date of Purchase**

- **Ticket Type, Ticket Priority, Ticket Status**

- **Ticket Channel** (Email, Phone, Chat, Social Media)

- **First Response Time**

- **Time to Resolution**

- **Customer Satisfaction Rating** (Target variable)

4. Steps Performed

A. Data Import and Inspection

- Imported the dataset using pandas

- Displayed head() and info() to understand structure and data types

B. Missing Value Handling

- Found missing values in Resolution, First Response Time, Time to Resolution, and Customer Satisfaction Rating

- Dropped rows where Customer Satisfaction Rating was missing (since it's the target)

C. Data Cleaning

- Removed non-essential features like Ticket ID, Customer Email, and Ticket Description

- Encoded categorical columns (e.g., Gender, Product, Ticket Type, Priority) using LabelEncoder

D. Exploratory Data Analysis (EDA)

- Visualized correlations using heatmaps

- Used bar plots to show distribution of ticket types, priorities, and channels

- Explored satisfaction level distributions across categories

E. Model Building

- Split dataset into training and testing sets (80-20 split)

- Trained a **RandomForestClassifier** to predict satisfaction

- Evaluated model using accuracy, confusion matrix, and classification report

---

5. Model Performance

- Achieved a good level of accuracy on the test set

- **Top Features** influencing customer satisfaction were:

  - **Ticket Priority**

  - **Time to Resolution**

  - **First Response Time**

   o **Ticket Type**

---

6. Conclusion

The machine learning model built in this project can successfully predict customer satisfaction ratings using support ticket data. The insights from EDA and model explainability help customer service teams prioritize cases and respond more effectively.

---

7. Future Scope

- Apply **sentiment analysis** using NLP on ticket descriptions

- Try **advanced ML models** like XGBoost or LightGBM

- Implement **real-time satisfaction prediction dashboards**

- Automate ticket routing based on predicted **satisfaction prediction dashboards**

- Automate ticket routing based on predicted satisfaction level

- Integrate findings into **CRM systems** for customer support optimization