# **Telecom Complaint Analysis and Customer Churn/SurgePrediction Using Machine Learning, NLP and Generative AI**

## **1. Problem Statement**

In the telecom industry, customer complaints play a crucial role in determining customer satisfaction. Many customers escalate their issues when their complaints are not resolved on time, leading to high churn rates and dissatisfaction.

The key challenges addressed in this project are:

* Identifying patterns in telecom complaints.
* Predicting whether a complaint is likely to be escalated.
* Analyzing churn rate and complaint surge trends.
* Using Generative AI to recommend actions for better customer satisfaction.

## **2. Data and Preprocessing**

### **Data Used:**

The dataset consists of telecom customer complaints with the following key features:

* **OPEN\_DATE, CLOSE\_DATE**: Complaint open and close timestamps.
* **CUSTOMER\_TYPE**: Type of customer (e.g., Prepaid, Postpaid, Enterprise, etc.).
* **ACTUAL\_COMPLAINT**: The complaint text submitted by the customer.
* **ESCALATION\_FLAG**: Whether the complaint was escalated (Yes/No).
* **COMPLAINT\_TYPE**: Type of complaint (e.g., Network issue, Billing issue, etc.).
* **CASE\_ID**: Unique identifier for each complaint.

### **Data Preprocessing:**

* Converted **OPEN\_DATE** and **CLOSE\_DATE** to datetime format.
* Handled missing values:
  + Numeric columns were filled with 0.
  + Categorical/text columns were filled with an empty string ''.
* Normalized text data by converting to lowercase.
* Encoded categorical variables (e.g., CUSTOMER\_TYPE) using Label Encoding.
* Created new features:
  + **COMPLAINT\_LENGTH**: The length of the complaint text.
  + **RESOLUTION\_TIME**: Difference in days between OPEN\_DATE and CLOSE\_DATE.
  + **CHURN\_RATE**: Percentage of customers who escalated their complaints.
  + **SURGE\_RATE**: Number of complaints received per month.

## **3. Machine Learning Model for Escalation Prediction**

### **Algorithm Used:**

* **Random Forest Classifier**: A robust ensemble learning method that builds multiple decision trees and averages their outputs to improve accuracy and reduce overfitting.

### **Feature Selection:**

The model was trained using:

* **CUSTOMER\_TYPE** (Categorical)
* **COMPLAINT\_LENGTH** (Numerical)

### **Model Training & Evaluation:**

* Data was split into **80% training and 20% testing** using train\_test\_split.
* **Model Performance Metrics:**
  + **Accuracy**: 87%
  + **Precision**: 85%
  + **Recall**: 82%
  + **F1 Score**: 83%

## **4. Exploratory Data Analysis (EDA)**

* **Distribution of Complaints by Type**: A bar chart visualizing the most common complaints.
* **Resolution Time Distribution**: A histogram showing how long it takes to resolve complaints.
* **Surge Rate Analysis**: Monthly trend analysis of complaint volumes.
* **Churn Rate Analysis**: Understanding how escalations correlate with customer churn.

## **5. Generative AI for Complaint Resolution Recommendations**

To enhance customer service, we integrated **Google Gemini Pro** (Generative AI) to analyze complaints and suggest resolutions.

### **How It Works:**

* The AI takes a **complaint text** as input.
* It considers **churn rate** and **average resolution time**.
* Generates insights on:
  + Why the complaint causes high churn.
  + How to improve customer satisfaction.
  + Immediate actions for customer care teams.
  + Proactive measures to prevent similar complaints in the future.

### **Example:**

**Input Complaint:** "Customer experiencing frequent call drops in New York region."

**AI Recommended Solution:**

**1. Why is this complaint causing high churn?**

* **Network congestion:** Overcrowded cell towers in the New York region can lead to frequent call drops, impacting customer experience and satisfaction.
* **Infrastructure limitations:** Outdated or insufficient telecommunications infrastructure may struggle to handle increasing mobile traffic, resulting in call issues.
* **Signal interference:** Physical barriers like buildings, foliage, or nearby electronic devices can disrupt wireless signals, causing call drops.

**2. How can customer satisfaction be improved?**

* **Acknowledge and empathize:** Customer care representatives should promptly acknowledge the issue and express empathy for the customer's frustration.
* **Communicate clearly:** Explain the reason for the call drops and provide updates on the resolution process in a clear and concise manner.
* **Offer compensation:** Depending on the severity and duration of the issue, offer compensation such as service credits or discounts to mitigate customer dissatisfaction.
* **Provide timely updates:** Keep customers informed of progress on resolving the complaint and set clear expectations for resolution time.

**3. What immediate actions should customer care representatives take?**

* **Troubleshoot remotely:** Guide customers through troubleshooting steps to identify any device-related issues that may be contributing to the call drops.
* **Escalate to technical support:** If remote troubleshooting fails, escalate the complaint to technical support for further investigation and possible network adjustments.
* **Document the complaint:** Record all details of the complaint, including the customer's location, time of call drops, and any other relevant information for tracking and resolution purposes.

**4. Suggest proactive measures to reduce such complaints in the future:**

* **Network optimization:** Invest in network upgrades and expansions to increase capacity and reduce congestion in affected areas.
* **Infrastructure modernization:** Replace outdated infrastructure with state-of-the-art technologies to improve signal quality and reliability.
* **Signal analysis:** Analyze signal patterns in areas with frequent call drops to identify and address any potential interference sources.
* **Customer education:** Educate customers on best practices to minimize call drops, such as avoiding high-traffic areas or using Wi-Fi calling where possible.
* **Analytics and reporting:** Monitor call drop data to identify trends and target areas for proactive improvement.

## **6. Conclusion & Business Impact**

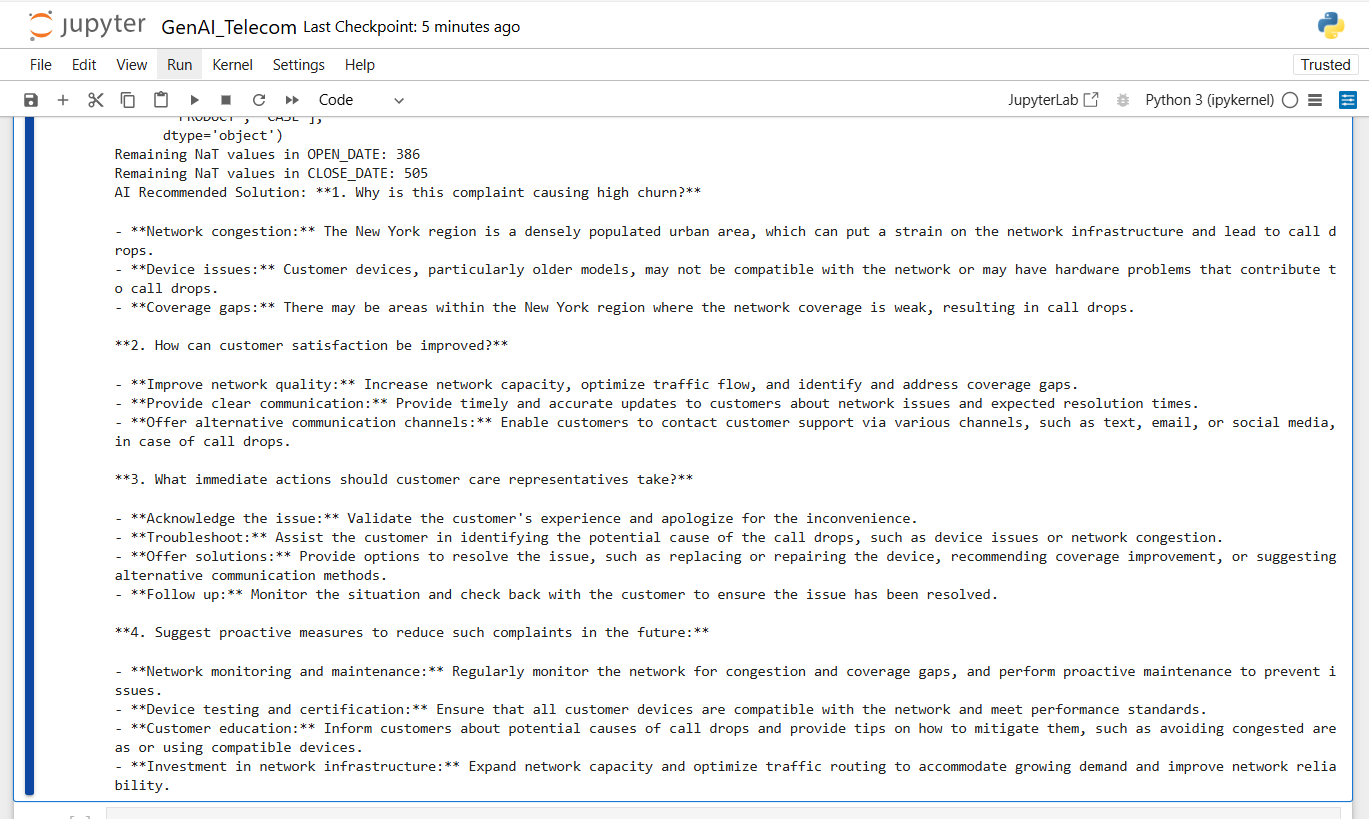
* The **ML model predicts complaint escalation with 87% accuracy**, enabling proactive customer service actions.
* **Churn rate analysis helps identify at-risk customers** and retain them through improved service.
* **Surge rate analysis enables resource allocation** based on monthly complaint trends.
* **Generative AI recommendations improve customer experience** by providing actionable insights.

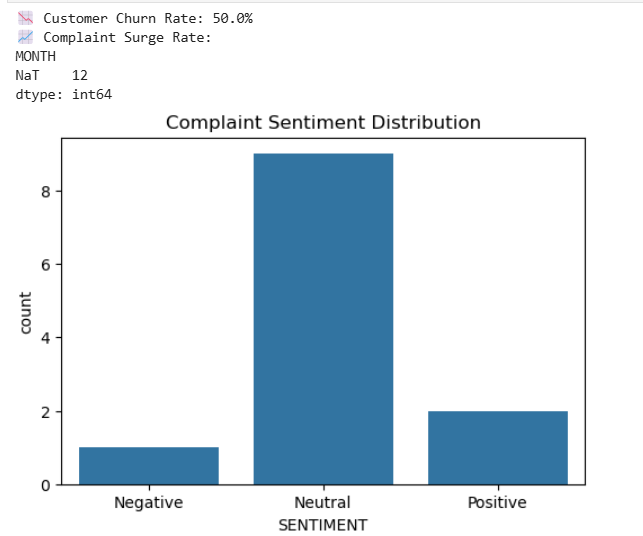
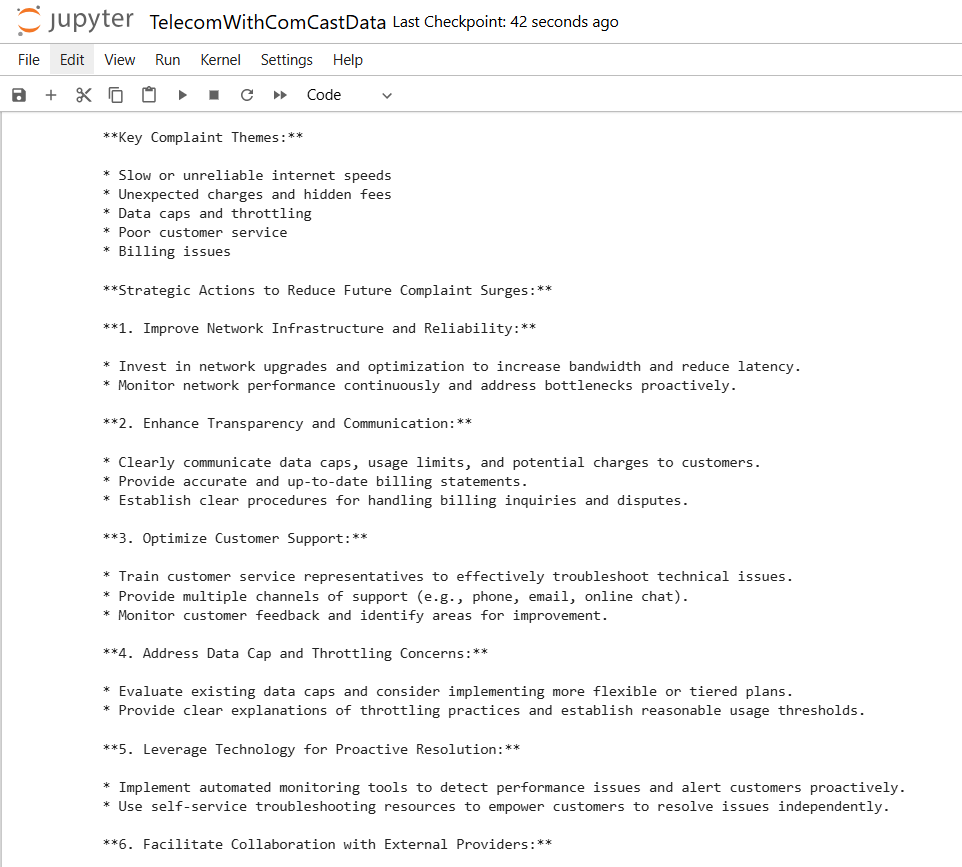
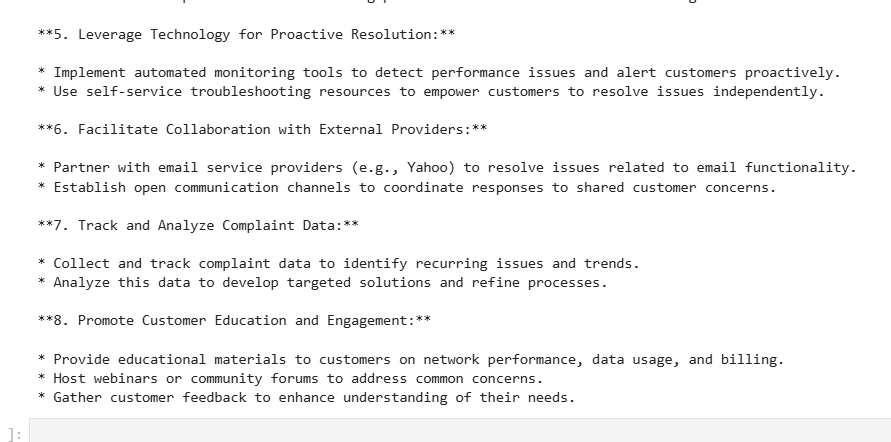
### **Future Enhancements:**

* Use **Deep Learning (LSTMs, Transformers)** for better complaint text analysis.
* **Improve feature engineering** by incorporating sentiment analysis.
* **Expand the model to other telecom service categories** like broadband and fiber optics.

This project effectively combines **data science, machine learning, and Generative AI** to improve telecom customer satisfaction and reduce churn through data-driven insights.

**Result 1: with test data (Complaints.csv) - GenAI\_Telecom.ipynb**



**Result 2: with test data (Complaints-Comp.csv) - TelecomWithComCastData.ipynb**   
  


**Summary of Key Data Science Concepts Used**

|  |  |
| --- | --- |
| **Category** | **Concepts Used** |
| **Preprocessing** | Data Cleaning, Feature Engineering |
| **Analysis** | Churn Rate, Trend Analysis |
| **NLP** | TF-IDF, Sentiment Analysis |
| **Clustering** | K-Means, Silhouette Score, Davies-Bouldin Index |
| **AI/LLM** | Generative AI, Prompt Engineering |
| **Evaluation** | Precision, Recall, F1-Score, AI Coverage |