VOICE CALL QUALITY ANALYSIS FOR SUPERIOR CUSTOMER EXPERIENCE & RETENTION

[CAPSTONE PROJECT]

Abstract

The telecom industry struggles with customer retention due to poor voice call quality, leading to dissatisfaction and churn. The goal of this study is to develop predictive models that identify and address factors affecting call quality, aiming to improve service, enhance customer satisfaction, reduce churn, and optimize network resources

ANJALI S
anjaliposy@gmail.com.

Project Summary

Batch details	PGP-DSE June 2024 Bengaluru Batch
Domain of Project	Telecom and Churn
Proposed project title	Voice Call Quality Analysis for Superior Customer Experience & Retention
Mentor Name	Mr. Sourab Reddy

Date: 21st Jan'2025

Approved by Mentor

Mr. Sourab Reddy

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

Business Problem Understanding:

The telecom industry is highly competitive, with customer retention being a critical factor for success. Poor voice call quality is a significant driver of customer dissatisfaction and churn. The Telecom Regulatory Authority of India (TRAI) has collected extensive data on voice call quality and customer experience through the MyCall app, providing a rich dataset for analysis.

Business Objective:

The primary objective is to leverage the provided dataset to develop predictive models that identify and mitigate factors affecting voice call quality and customer experience. By doing so, the goal is to:

- Improve overall call quality.
- Enhance customer satisfaction.
- Reduce churn rates.
- Optimize network resources.

Approach:

- Data Preprocessing: Clean and normalize the dataset.
- Exploratory Data Analysis (EDA): Identify patterns and correlations.
- Model Development: Build and validate predictive models for call quality issues.
- Customer Feedback Correlation: Analyze the impact on customer satisfaction.
- Optimization Strategies: Use insights to optimize network resources and improve call quality.

Data Findings:

The telecommunications industry is highly competitive, with customer retention being a critical factor for success. Poor voice call quality is a major contributor to customer dissatisfaction and churn. To address this, the Telecom Regulatory Authority of India (TRAI) has collected extensive data through the MyCall app, providing valuable insights into call quality and customer experience.

This study leverages the provided dataset to develop predictive models that identify key factors affecting call quality and customer satisfaction. The approach includes data preprocessing, exploratory data analysis (EDA), and machine learning techniques to detect patterns and correlations. Additionally, customer feedback is analyzed to assess its impact on overall experience.

Findings reveal that factors such as network congestion, signal strength, and geographical variations significantly influence call quality. By addressing these issues through data-driven optimization strategies, telecom providers can enhance service reliability, improve customer satisfaction, and reduce churn. The insights from this study also aid in efficient network resource allocation, ensuring better quality of service (QoS) and regulatory compliance.

Conclusions:

By identifying and addressing the key factors impacting call quality, the company can significantly enhance overall call performance. Improved call quality translates to a better experience and greater satisfaction for customers. Consequently, satisfied customers are less likely to switch to other providers, resulting in lower churn rates and increased customer retention. Furthermore, efficiently allocating and optimizing network resources based on predictive insights will enhance service reliability and performance.

Industry Review - Current Practices

Call Quality Management Systems (CQMS):
Advanced CQMS systems integrate AI/ML models to predict and mitigate call quality issues in real time. Companies like Cisco, Google, and AWS are leveraging AI for real-time voice quality optimization.
They analyze live call streams and suggest adjustments such as bandwidth allocation or codec changes.

Speech Analytics:
Speech-to-text solutions combined with natural language processing (NLP) are employed to analyze customer-agent interactions.
Sentiment analysis is used to measure customer satisfaction indirectly from voice tone and keywords.

Location Optimization: Geospatial data helps optimize infrastructure and deploy microcells in areas with weak coverage, enhancing network reliability.

Customer Feedback Mechanisms:
Post-call surveys are commonly used to gather feedback on call quality.
Automated systems send satisfaction ratings and allow customers to report issues directly.

Cloud-Based Contact Centers:
Cloud services like Amazon Connect, Genesys Cloud CX, and Twilio Flex ensure scalable and reliable voice services with embedded analytics.
These platforms provide detailed insights into call quality and performance.

Technology Upgrades:
Transitioning to 4G/5G networks and implementing VoLTE and VoWiFi technologies improve voice quality and reduce latency.

Seasonal Adaptations:
Networks adjust to higher demand during peak times, using temporal data to forecast and plan for busy periods.

Background Research on Voice call quality

Understanding Voice Call Quality Metrics:

1. Mean Opinion Score (MOS): A subjective rating of voice quality on a scale from 1 to 5 (1 = poor quality, 5 = excellent).
2. Call Drop Rate (CDR): The percentage of calls that are unexpectedly dropped during a session.
3. Latency (Jitter and Delay): The time taken for data to travel from the sender to the receiver. Low latency is critical for good voice quality.
4. Packet Loss: The percentage of packets that are lost during transmission. High packet loss negatively impacts call quality.
5. Call Setup Time: The time it takes for a call to be established once initiated.
6. Signal Strength: Measures the strength of the signal during the call, which impacts call stability and quality.

Telecom Network Technologies Impacting Voice Quality:

1. 2G/3G Networks: Older technologies that often have lower quality (e.g., low data rates, high latency).
2. VoLTE (Voice over LTE): A technology that provides HD voice quality by using the 4G LTE network for voice calls.
3. VoNR (Voice over New Radio): Voice over 5G technology, which promises superior voice quality with ultra-low latency.
4. Wi-Fi Calling: Can provide better voice quality in areas with poor cellular coverage by using the user's Wi-Fi network.
5. Traditional PSTN (Public Switched Telephone Network) has largely been replaced by VoIP (Voice over Internet Protocol) systems due to cost efficiency and flexibility.

Factors Affecting Voice Call Quality:

1. Signal Strength and Coverage: Poor signal strength leads to dropped calls, jitter, and poor quality.
2. Network Congestion: High network traffic can lead to congestion, increasing latency and affecting voice clarity.
3. Environmental Factors: Geographical factors like urban vs. rural areas, physical obstructions (e.g., buildings, terrain), and user mobility (e.g., driving vs. stationary) impact call quality.
4. Device Characteristics: Different smartphones and network equipment may have varying capabilities, affecting voice quality (e.g., microphone, speaker, codec support).
5. Network Load and Interference: High interference from other electronic devices or poor backhaul connectivity can degrade call quality.

Impact on Customer Retention:
According to Gartner, a 5% increase in customer retention can boost profits by 25-95%.
High call quality ensures first-call resolution and reduces customer churn.

Integration with CRM Systems:
Call quality insights are integrated with CRM platforms like Salesforce and HubSpot to provide a 360-degree view of customer interactions.
This helps in personalizing customer experiences and understanding pain points.

Challenges:
High dependence on network infrastructure: Even minor disturbances in network performance can severely impact VoIP call quality.
Data privacy concerns: Monitoring calls often involves analyzing sensitive data, requiring compliance with GDPR, CCPA, etc.
Scalability: Analyzing millions of call records in real-time is computationally intensive.

Emerging Trends:
AI for predictive maintenance, crowdsourcing for real-time feedback, and the rollout of 5G are shaping the future of voice services.
Operators leverage feedback from customers via mobile apps or crowdsourced platforms (e.g., My Call app, Tutela).

In summary, telecom providers use a combination of data, technology, and customer insights to enhance voice call quality and meet customer expectations.

Key Focus Areas:

1. Customer Experience: Understanding customer perceptions through ratings and feedback.
2. Network Performance: Analyzing call drop patterns and network issues to optimize network coverage.
3. Geographical Insights: Identifying problem areas by correlating network performance with geographical data.
4. Technology Analysis: Comparing performance across different network types (e.g., 4G, 5G).

Literature Survey

Publications

The Paper **"Indian Voice Call Quality Datasets to Assess Customer Experience"** on IndiaAI.gov.in explains how the Telecom Regulatory Authority of India (TRAI) collects and makes voice call quality data publicly accessible. (Paper link)

1. Data Collection:

- TRAI gathers real-time user feedback on voice call quality through the MyCall app. Customers provide feedback on the quality of their voice calls, which helps TRAI collect valuable customer experience data along with network statistics.

2. Public Accessibility:

- The collected data is made publicly accessible to ensure transparency and to help researchers, telecom companies, and policymakers understand customer opinions better.

3. Coverage:

- The data encompasses feedback from various service providers across multiple locations, covering 3G, 4G, and 2G networks.

4. Purpose:

- The aim is to use this data to improve the overall quality of voice calls and enhance customer satisfaction in the telecom industry.

The Paper **"Determinants of Customer Satisfaction in Telecom Industry - A Study of Indian Telecom Industry**" by G. S. Popli and Dr. Madan .(Paper link)

1. Data Collection:

- A structured questionnaire was used to collect data from 150 respondents in the National Capital Region of Delhi, India.
- Descriptive statistics, correlation, and regression analysis were performed using SPSS software.

2. Coverage:

- The study focuses on the telecom industry within the National Capital Region of Delhi, India.

3. Purpose:

- The primary objective is to identify the determinants and factors impacting customer satisfaction in the telecom industry.
- By understanding and addressing these factors, telecom companies can enhance customer satisfaction and maximize the number of users.

Application

The Telecom Regulatory Authority of India (TRAI) uses a variety of applications and technologies to assess and monitor voice call quality in India. These tools help TRAI ensure that telecom service providers maintain high standards of service quality and comply with regulatory norms. One such application that we used in the project is from crowdsourced platform called "My Call App".

MyCall is a TRAI app designed to help users report their voice call quality experiences. The app allows mobile users to share feedback on call quality, including issues like call drops, call clarity, and network coverage.

The app collects crowdsourced data on voice quality and forwards it to TRAI. This data helps TRAI monitor real-time service quality across different telecom networks in India.

The app helps generate a QoS report based on user feedback and also informs telecom operators about problem areas in their network coverage.

About My Call Feedback App

People face a lot of issues due to low voice call quality and inability to express their opinion on a centralized platform. MyCall app provides a platform to all telecom subscribers in India to crowd source their opinion through a feedback rating process.

After completion of an incoming or outgoing call, a feedback pop-up will appear which is configurable under the settings tab. Users can also check the details of previous rated calls from the history tab as well as rate the unrated calls or rate multiple calls in one go.

The MyData tab gives a map based view of the rated calls and an option to filter the calls by specific time period. The settings tab provides configurables setting along with an option for users to select a rating style indicator from 3 available options per their preference.

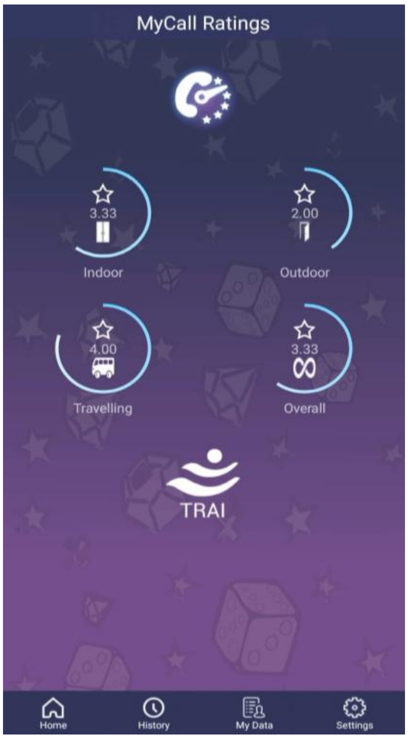
Challenges

- Lack of Awareness: Many users are unaware of the app.
- Alternative Channels: Users prefer other feedback methods.
- Trust Issues: Users doubt the impact of their feedback.

Recommendations

- Increase Awareness: Promote the app through targeted campaigns.
- Improve User Experience: Make the app more user-friendly.
- Incentivize Feedback: Offer rewards for regular feedback.
- Collaboration: Work with telecom providers to promote the app.

My Call App Overview and Voice Call Data Capture Process



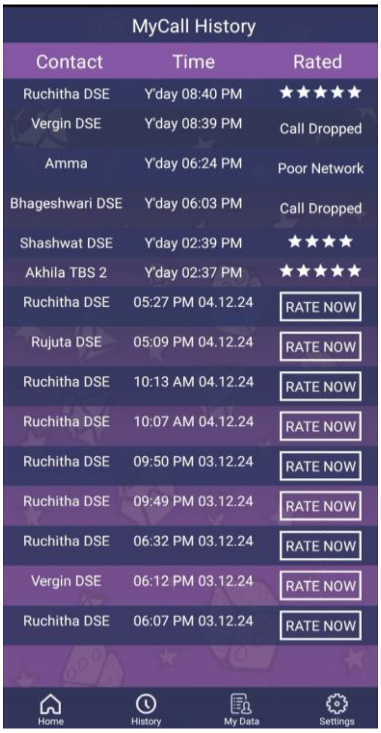
MyCall Ratings

Indoor: 3.33

Outdoor: 2.00

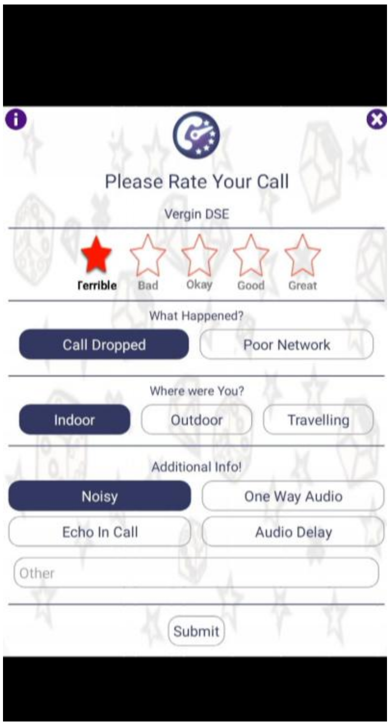
Travelling: 4.00

Overall: 3.33



MyCall History

Contact	Time	Rated
Ruchitha DSE	Y'day 08:40 PM	★★★★★
Vergin DSE	Y'day 08:39 PM	Call Dropped
Amma	Y'day 06:24 PM	Poor Network
Bhageshwari DSE	Y'day 06:03 PM	Call Dropped
Shashwat DSE	Y'day 02:39 PM	★★★★
Akhila TBS 2	Y'day 02:37 PM	★★★★★
Ruchitha DSE	05:27 PM 04.12.24	RATE NOW
Rujuta DSE	05:09 PM 04.12.24	RATE NOW
Ruchitha DSE	10:13 AM 04.12.24	RATE NOW
Ruchitha DSE	10:07 AM 04.12.24	RATE NOW
Ruchitha DSE	09:50 PM 03.12.24	RATE NOW
Ruchitha DSE	09:49 PM 03.12.24	RATE NOW
Ruchitha DSE	06:32 PM 03.12.24	RATE NOW
Vergin DSE	06:12 PM 03.12.24	RATE NOW
Ruchitha DSE	06:07 PM 03.12.24	RATE NOW



Please Rate Your Call

Vergin DSE

● Terrible ○ Bad ○ Okay ○ Good ○ Great

What Happened?

● Call Dropped ○ Poor Network

Where were You?

● Indoor ○ Outdoor ○ Travelling

Additional Info!

● Noisy ○ One Way Audio ○ Echo In Call ○ Audio Delay

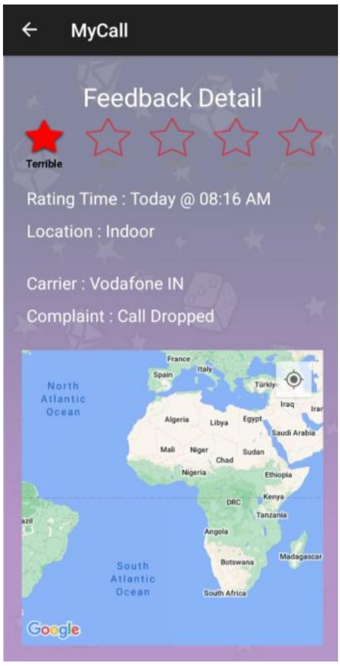
Other

Submit

Home screen that represents the average rating in different environments(Indoor ,Outdoor ,Travelling) and overall calls.

"MyCall History," which provides information about your calls, such as the contact name, call time, and rating status. If a call hasn't been rated, you can rate it.

Users rate call experience on a five-star scale (Terrible to Great) and select issues like Call Dropped, Poor Network. They can also specify the location where the call was made.



MyCall

Feedback Detail

★ Terrible

Rating Time : Today @ 08:16 AM

Location : Indoor

Carrier : Vodafone IN

Complaint : Call Dropped

This represents the feedback detail for each call which shows the ratings, date and time of the call, Operator that the customer is using, complaint that was reported , and the location type - indoor, outdoor and travelling and precise geographic coordinates (latitude and longitude) of the call location, enabling further spatial analysis.

This level of detail ensures a thorough understanding of call quality and user experiences.

All feedback provided is stored anonymously. This means that your ratings and comments are not linked to your personal information, ensuring your privacy while still contributing valuable data to improve services.

Fig 1 : User Interface of MyCall App

Past and Ongoing Research

Past Research on Voice Call Quality

Network Technology Impact (Pre-4G Era):
Early research on 2G networks (e.g., GSM) highlighted low bit rates, high call drop rates, and limited voice quality.
For 3G networks, studies focused on QoS, handover management, mobility, and real-time data impact on voice quality.

Voice over IP (VoIP) Quality:
Research optimized compression algorithms, jitter buffering, and error correction to enhance VoIP quality.
Key issues included packet loss, delay, echo cancellation, and codec efficiency (e.g., G.711, G.729).
QoE models were developed to quantify user satisfaction for VoIP and PSTN systems.

Voice Quality Evaluation:
Methods like Mean Opinion Score (MOS) were widely used for subjective assessments.
Objective metrics such as PESQ and POLQA were introduced to measure voice quality in line with human perception.

Impact of Network Congestion and Bandwidth:
Studies analyzed how network congestion and limited bandwidth caused issues like dropped calls, choppy voice, and echo.
Solutions focused on optimizing packet scheduling, bandwidth allocation, and traffic management for improved voice quality.

Ongoing Research on Voice Call Quality

- Ultra-low Latency and High-Definition Audio: Optimizing voice codecs and reducing latency for real-time communication.
- AI and ML Algorithms for Predictive Quality Management: Implementing intelligent systems to automatically adjust network parameters to improve voice quality.
- Seamless Integration Across Networks: Ensuring that voice calls can transition seamlessly between cellular, Wi-Fi, and other networks without degrading quality.
- Real-Time Monitoring and Analytics: Developing real-time quality monitoring tools to track network performance and user experience.
- Next-Gen Voice Features: Exploring advanced voice features such as voice biometrics, voice enhancement, and emotion recognition in calls.

Project Justification

Project Statement:
The telecom industry is highly competitive, with customer retention being a critical factor for success. Poor voice call quality is a significant driver of customer dissatisfaction. The Telecom Regulatory Authority of India (TRAI) has collected extensive data on voice call quality and customer experience through the MyCall app, providing a rich dataset for analysis.

Complexity Involved:

- Data Cleaning: Handling missing values as the location could be reported in any part of India, treating with mode or KNN imputation would lead to bias in the data, so we choose to go ahead with logical imputation.
- Imbalanced Dataset: Imbalanced data can lead to biased models, poor detection of the minority class, and misleading evaluation metrics. It may cause the model to overfit to the majority class, losing important insights and limiting generalization.
- There could have been more detailed features to explain why the call drop occurred.

Project Outcome:

The project outcome is to identify the dominant operators in each region based on customer ratings and determine which operators require service improvements to enhance customer satisfaction and competitiveness.

Commercial, Academic or Social value.

1. Commercial Value

Improved Customer Retention: By identifying areas where operators need to improve, companies can implement targeted strategies to enhance customer satisfaction.
Market Competitiveness: Insights into regional dominance allow operators to benchmark their performance against competitors and strategize accordingly.
Revenue Growth: Satisfied customers are more likely to continue using services and recommend them to others, directly impacting revenue.
Operational Efficiency: Data-driven recommendations help streamline processes, optimize agent performance, and reduce operational costs.

2. Academic Value

Advancing Research: The dataset offers opportunities for academic research in areas like natural language processing (NLP), sentiment analysis, and predictive modeling.
Educational Use: It can serve as a case study for students learning data analytics, machine learning, and customer service optimization.
Publication Opportunities: Researchers can publish findings in journals or present them at conferences, contributing to the knowledge base in customer relationship management and data science.

3. Social Value

Enhanced Customer Experience: By addressing pain points, operators can improve the quality of service, leading to higher customer satisfaction.
Equity in Service Delivery: Identifying underserved regions ensures that all customers, regardless of location, receive equitable and high-quality services.
Empowering Consumers: Insights can empower customers to make informed decisions when choosing operators based on regional performance and service quality.

Overview of the Final Process

Problem-Solving Methodology:

The project aimed to predict call quality ratings from the MyCall app dataset to identify key issues affecting telecom services and provide actionable insights for improvement. The approach focused on rigorous preprocessing, handling class imbalance, and leveraging advanced machine learning techniques.

Salient Features of the Data:

- Dataset Summary - 170,717 rows and 10 columns.
- The dataset, sourced from the MyCall app, included key features like Operator, In/Out/Travelling, Network Type, State Name, Region, Rating, and geographic coordinates.
- Ratings (1–5) represented customer satisfaction levels, with higher ratings being dominant, creating class imbalance.
- The dataset was diverse, covering different operators, environments, and regions across India.

Data Preprocessing:

Handling Missing Values:

- Non-standard placeholders like -1 in geographic data were replaced with NaN.
- Missing state names were populated using geographic coordinates via the geopy library.
- Records with unresolved inconsistencies were dropped.

Duplicates:

- Duplicates were preserved since each record reflected unique customer feedback.

Exploratory Data Analysis (EDA):

Univariate Analysis:

Target Variable (Rating):

Average rating: 3.46, skewed towards higher ratings (4 and 5).

Class imbalance was evident, with Ratings 1–3 underrepresented.

Categorical Variables:

RJio dominated among operators, followed by Airtel and Vodafone.

Most calls were rated in indoor environments, and 4G was the most used network type.

Numerical Variables:

Latitude and longitude covered diverse locations across India, enabling geographic analysis.

Bivariate Analysis:

- RJio had the highest ratings for 4G, with some localized issues.
- Indoor calls received slightly better ratings than outdoor or traveling calls.
- Regions like the West and South had better average ratings compared to the Northeast and Central regions.

Feature Engineering :

Feature Creation: Added a 'Region' feature by categorizing states into regional groups based on geographical locations.

Outlier Treatment: Kept longitude outliers as they represent valid geographic locations important for analysis.

Categorical Encoding: Applied Label Encoding to convert categorical variables into numerical format for model compatibility.

Scaling: Scaled numerical features (Latitude and Longitude) to ensure models perform optimally.

Feature Selection: Retained all features; Call Drop Category is the most important for predicting the target (Rating).

Dimensionality Reduction: Not needed due to the small number of features in the dataset.

Statistical Tests:

- Chi-Square Test (Categorical vs Categorical): The Chi-Square test was used to assess the relationship between categorical features and the target variable (Rating). The null hypothesis (independence between feature and target) is rejected for all categorical features, indicating they are statistically significant and relevant for model building.
- Kruskal-Wallis H Test (Categorical vs Numerical): For numerical features (Latitude and Longitude), the Kruskal-Wallis H test is applied to assess their relationship with the target variable (Rating). The null hypothesis (independence between feature and target) is rejected, suggesting a significant relationship, and these features will be retained for model building.

Techniques for Handling Class Imbalance :

1. Ensemble Methods: Use models like Random Forests, XGBoost, and Gradient Boosting, which handle class imbalance by adjusting weights or splitting criteria internally.
2. Cross-Validation: Ensures each fold maintains the original class distribution, preventing bias.
3. Class Weights: Assign higher weights to minority classes to give them more importance during model training.
4. SMOTE: Not used due to potential distortion; instead, ensemble methods and class weights are preferred for handling imbalance.

Modelling Approach:

The task is to predict call quality ratings (1-5) based on various features. The goal is to build a multi-class classification model. Multiple classifiers were tested, and performance metrics such as accuracy, precision, recall, and F1-score were used for evaluation.

Data Distribution Consistency

The Chi-Square and Kolmogorov-Smirnov tests confirmed that the distributions of both categorical and continuous variables are consistent between training and testing datasets, ensuring no data distribution shifts that could affect model performance.

Base Model Selection:

A Decision Tree was chosen as the baseline due to its ability to handle non-linear relationships, which were evident in the data based on correlation analysis with the target variable.

Logistic regression was avoided as the data lacked a clear linear relationship with the target variable, as confirmed through preliminary analysis.

Seven classifiers were evaluated, with XGBoost showing the best performance with low overfitting, competitive test accuracy (82.20%), and balanced precision and recall. Models like Decision Tree and AdaBoost showed higher overfitting with significant gaps between train and test accuracy.

Comparison to Benchmark:

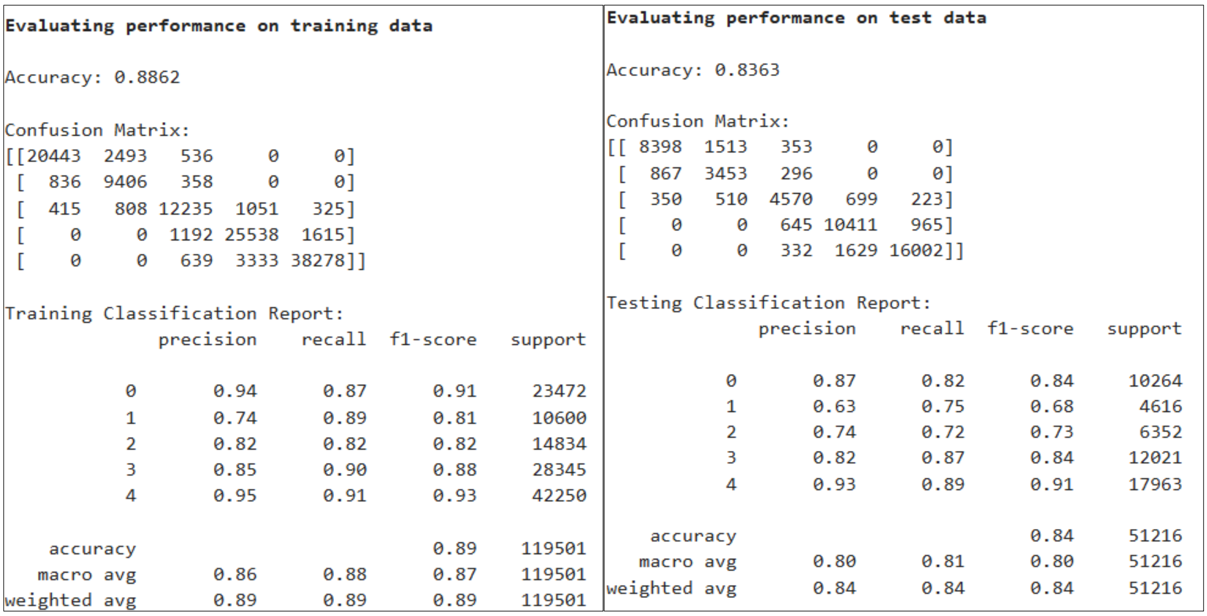
The final XGBoost model, with hyperparameter tuning and class weights, outperformed the benchmark in terms of accuracy (test accuracy 83.63%) and class balance. The benchmark models showed moderate overfitting and struggled with class imbalance, while XGBoost handled it better. Further optimization is required to reduce overfitting.

Final Model Assessment and Performance Metrics

XGBoost Classifier – Hyperparameter Tuning and Class Weights

Best Parameters :

{'subsample': 0.8, 'n_estimators': 300, 'min_child_weight': 3, 'max_depth': 10, 'learning_rate': 0.1, 'gamma': 0, 'colsample_bytree': 0.8}



```
Evaluating performance on training data

Accuracy: 0.8862

Confusion Matrix:
[[20443  2493   536     0     0]
 [  836  9406   358     0     0]
 [  415   808 12235  1051   325]
 [    0     0  1192 25538  1615]
 [    0     0   639  3333 38278]]

Training Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.87      0.91     23472
           1       0.74      0.89      0.81     10600
           2       0.82      0.82      0.82     14834
           3       0.85      0.90      0.88     28345
           4       0.95      0.91      0.93     42250

    accuracy                           0.89    119501
   macro avg       0.86      0.88      0.87    119501
weighted avg       0.89      0.89      0.89    119501
```

```
Evaluating performance on test data

Accuracy: 0.8363

Confusion Matrix:
[[ 8398  1513   353     0     0]
 [  867  3453   296     0     0]
 [  350   510  4570   699   223]
 [    0     0   645 10411   965]
 [    0     0   332  1629 16002]]

Testing Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.82      0.84     10264
           1       0.63      0.75      0.68      4616
           2       0.74      0.72      0.73      6352
           3       0.82      0.87      0.84     12021
           4       0.93      0.89      0.91     17963

    accuracy                           0.84     51216
   macro avg       0.80      0.81      0.80     51216
weighted avg       0.84      0.84      0.84     51216
```

Fig 2 : XGBoost Classification Report of Training and Test data

Step-by-step walk through of the solution

Step 1: Problem Definition and Data Understanding

Objective:

Identify key factors affecting call drops and predict call quality issues based on customer and network-related data.

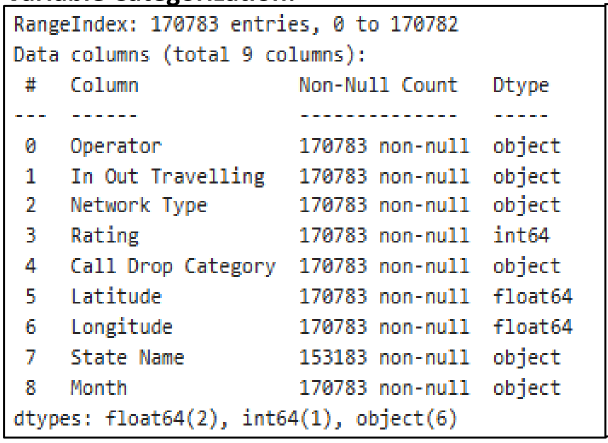
Data Overview:

The dataset contained features such as Operator, In Out Travelling, Network Type, Rating, Call Drop Category, Latitude, Longitude, State Name, Month, and Region.

Data Shape:

170,783 rows and 10 columns

Variable Categorization:



```
RangeIndex: 170783 entries, 0 to 170782
Data columns (total 9 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Operator            170783 non-null  object
 1   In Out Travelling   170783 non-null  object
 2   Network Type        170783 non-null  object
 3   Rating              170783 non-null  int64
 4   Call Drop Category  170783 non-null  object
 5   Latitude            170783 non-null  float64
 6   Longitude           170783 non-null  float64
 7   State Name          153183 non-null  object
 8   Month               170783 non-null  object
dtypes: float64(2), int64(1), object(6)
```

There are 6 categorical columns - 'Operator', 'IN Out Travelling', 'Network Type', 'Call Drop Category', 'State Name', 'Month'.

There are 3 Numerical columns - 'Rating' , 'Latitude' , 'Longitude'.

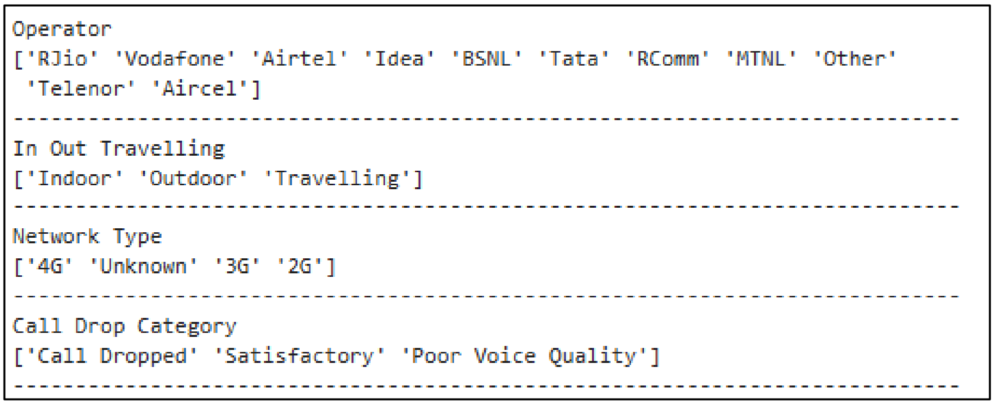
The "State Name" column contains null values. To ensure we account for non-standard missing values, we need to examine the unique values in each column.

Fig 3: Dataset Overview with Feature Types and Non-Null Counts

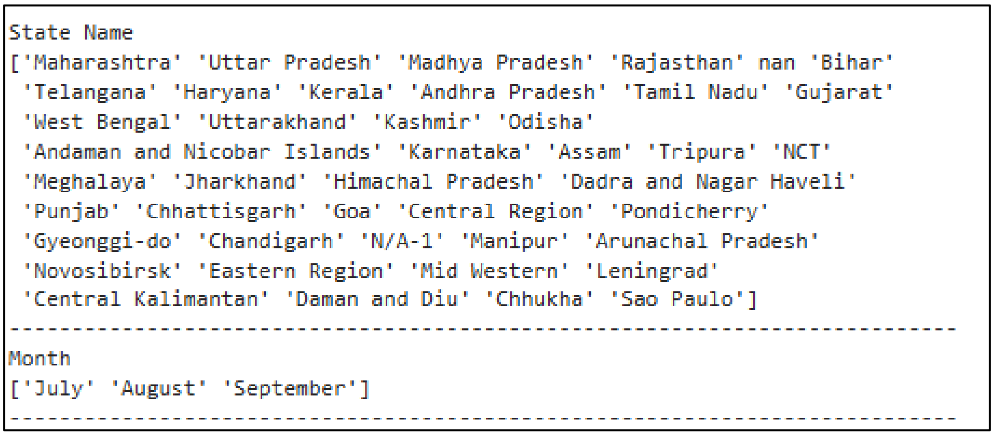
Step 2: Data Preprocessing

Handling Missing and Inconsistent Data

To identify non-standard missing values, we will review the unique values of each variable.



```
Operator
['RJio' 'Vodafone' 'Airtel' 'Idea' 'BSNL' 'Tata' 'RComm' 'MTNL' 'Other'
 'Telenor' 'Aircel']
----------------------------------------------------------------------------
In Out Travelling
['Indoor' 'Outdoor' 'Travelling']
----------------------------------------------------------------------------
Network Type
['4G' 'Unknown' '3G' '2G']
----------------------------------------------------------------------------
Call Drop Category
['Call Dropped' 'Satisfactory' 'Poor Voice Quality']
----------------------------------------------------------------------------
```



```
State Name
['Maharashtra' 'Uttar Pradesh' 'Madhya Pradesh' 'Rajasthan' nan 'Bihar'
 'Telangana' 'Haryana' 'Kerala' 'Andhra Pradesh' 'Tamil Nadu' 'Gujarat'
 'West Bengal' 'Uttarakhand' 'Kashmir' 'Odisha'
 'Andaman and Nicobar Islands' 'Karnataka' 'Assam' 'Tripura' 'NCT'
 'Meghalaya' 'Jharkhand' 'Himachal Pradesh' 'Dadra and Nagar Haveli'
 'Punjab' 'Chhattisgarh' 'Goa' 'Central Region' 'Pondicherry'
 'Gyeonggi-do' 'Chandigarh' 'N/A-1' 'Manipur' 'Arunachal Pradesh'
 'Novosibirsk' 'Eastern Region' 'Mid Western' 'Leningrad'
 'Central Kalimantan' 'Daman and Diu' 'Chhukha' 'Sao Paulo']
---------------------------------------------------------------------------
Month
['July' 'August' 'September']
---------------------------------------------------------------------------
```

Fig 4: Unique Entries for Categorical Features

1 .The 'Network Type' column contains "Unknown" as a missing value, which we will retain as is.
2. The 'State Name' column has "N/A-1" as a placeholder for missing values, which will be treated as such.
3. We will replace these non-standard missing values with np.nan to accurately count the total null values and handle them appropriately.

We will examine the value counts of each variable to detect any inconsistencies and address them accordingly.

Operator	**Counts**	**Proportions**
RJio	69581	40.742
Airtel	36522	21.385
Vodafone	32596	19.086
BSNL	17083	10.003
Idea	12857	7.528
MTNL	1413	0.827
Tata	608	0.356
Telenor	43	0.025
Other	33	0.019
RComm	26	0.015
Aircel	21	0.012
In Out Travelling	**Counts**	**Proportions**
Indoor	116842	68.415
Outdoor	32639	19.111
Travelling	21302	12.473
Network Type	**Counts**	**Proportions**
4G	97427	57.047
3G	36985	21.656
Unknown	27514	16.111
2G	8857	5.186

State Name	**Counts**	**Proportions**
Maharashtra	33428	24.205
West Bengal	10451	7.567
Rajasthan	10438	7.558
Uttar Pradesh	10210	7.393
Gujarat	9174	6.643
NCT	8823	6.389
Karnataka	8507	6.610
Tamil Nadu	8173	5.918
Telangana	7527	5.45
Haryana	4922	3.564
Kerala	4912	3.557
Odisha	4888	3.539
Bihar	3802	2.753
Madhya Pradesh	2982	2.159
Andhra Pradesh	2557	1.851
Uttarakhand	1628	1.179
Chattisgarh	1294	0.937
Jharkhand	901	0.652
Punjab	825	0.597
Meghalaya	594	0.43
Chandigarh	514	0.372
Assam	511	0.37
Goa	311	0.225

Call Drop Category	**Counts**	**Proportions**
Satisfactory	113482	66.448
Poor Voice Quality	43297	25.352
Call Dropped	14004	8.2
Month	**Counts**	**Proportions**
July	63689	37.292
September	62472	36.58
August	44622	26.128
Rating	**Counts**	**Proportions**
5	60215	35.258
4	40379	23.643
1	33753	19.764
3	21207	12.418
2	15229	8.917

Dadra and Nagar Haveli	236	0.171
Himachal Pradesh	186	0.135
Pondicherry	98	0.071
Kashmir	71	0.051
Tripura	67	0.049
Andaman & Nicobar Islands	43	0.031
Central Region	7	0.005
Central Kalimantan	7	0.005
Chukka	4	0.003
Sao Paulo	3	0.002
Manipur	2	0.001
Novosibirsk	2	0.001
Eastern Region	2	0.001
Gyeonggi-do	1	0.00005
Arunachal Pradesh	1	0.00005
Mid-Western	1	0.00005
Leningrad	1	0.00005
Daman and Diu	1	0.00005

Tab 1 : Value Counts and Proportions of Categories

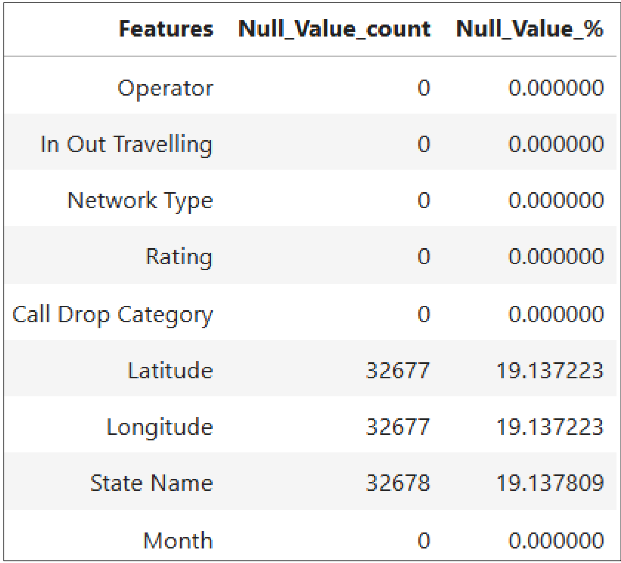
1.In the 'Operator' column, operators with a count below 100 will be grouped into a single category labelled 'Others'.
2.In the 'State Name' column, records with a count of 1 or 2 and all foreign countries will be removed, which has 32 records.
3.In the 'State Name' column, standardizing 'NCT' values by replacing them with 'Delhi', as it is the most commonly specified location.
4. As mentioned in data dictionary, Latitude and Longitude has nonstandard missing values : -1 , 0, which will replace these with np.nan to accurately count the total null values and handle them appropriately.

Duplicates in records

Removing duplicates in the voice call dataset may lead to the loss of valuable information, as multiple customers from the same location could have reported their call quality experiences independently. Each entry may represent a unique user experience, even if the location, operator, or other features appear identical.
Hence, instead of dropping duplicates, we keep all records to preserve the granularity of the data and ensure a more accurate analysis of voice call quality.

Missing value Treatment

Once non-standard missing values are handled and anomalies are corrected, we will verify the overall count of null values and check the shape of data.



Features	Null_Value_count	Null_Value_%
Operator	0	0.000000
In Out Travelling	0	0.000000
Network Type	0	0.000000
Rating	0	0.000000
Call Drop Category	0	0.000000
Latitude	32677	19.137223
Longitude	32677	19.137223
State Name	32678	19.137809
Month	0	0.000000

For records with a missing State Name but valid latitude and longitude, the 'geopy ' library will be used to retrieve and populate the State Name based on geographic coordinates.

For other records with null values, logical imputation will be performed by grouping the data and using the mode to fill the missing values.

Fig 5: Summary of Null Values: Count and Proportion

After the null value treatment , 34 records couldn't be logically imputed due to conflicting, or inconsistent information that couldn't be resolved through existing patterns while grouping.
So we will be dropping those 34 records – null values.

Step 3: Feature Engineering

Feature Creation

- Creating a new feature to categorize 'State Names' into regional groups such as North, South, East, West, Central and North-East based on geographical locations.
- This will involve generating a new column 'Region' that assigns each state to a specific region, which can be useful for modelling to capture regional trends or patterns in the data.

Presence of outliers and its treatment:

In the dataset, most features are categorical, so outliers are not a concern for those variables. However, the longitude feature contains some extreme values that might be considered outliers in a typical dataset. These outliers, however, represent valid geographic locations on the Earth's surface. Removing or adjusting these values could lead to the loss of important insights related to remote areas or regions with unique network behaviours. Since geographic data is crucial for understanding regional patterns in call quality, we choose not to treat these longitude outliers, as each point is valuable for the analysis.

Transformation of Categorical Variables for Machine Learning Models

In the dataset, categorical variables such as Operator, In/Out/Traveling, Network Type, Call Drop Category, State Name, Region, and Month need to be converted into numerical format to be compatible with machine learning models. Since most machine learning algorithms require numerical input, these categorical features must undergo transformation to ensure the models can process and interpret them effectively.

To achieve this, Label Encoding will be applied, where each unique category is assigned a numerical value. This technique is particularly suitable for both ordinal and nominal categorical data, allowing the model to appropriately handle and utilize the information contained within these variables. This transformation is crucial for enabling accurate model training and improving overall performance.

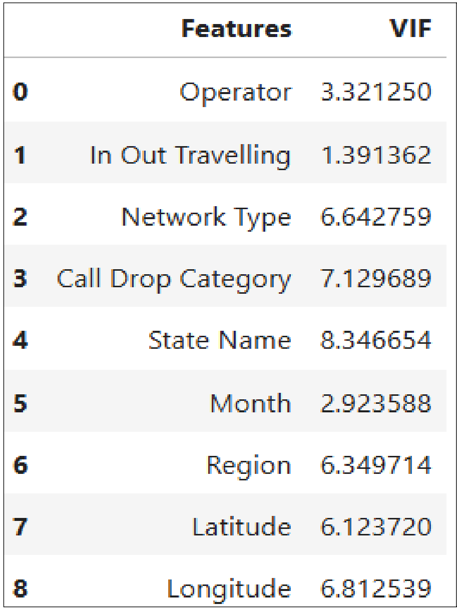
Scaling the data:

Since this dataset contains numerical features like Latitude and Longitude, scaling might be necessary. Models like k-NN, Logistic Regression can perform better when all features are on a similar scale.

Feature Selection:

When dealing with a classification problem, feature selection to address multicollinearity is crucial for ensuring that the model performs well and is not influenced by redundant features. Multicollinearity occurs when two or more predictor variables are highly correlated with each other, which can lead to issues in model interpretation and stability, particularly with linear models like Logistic Regression.

1. Checking Multi-collinearity using VIF

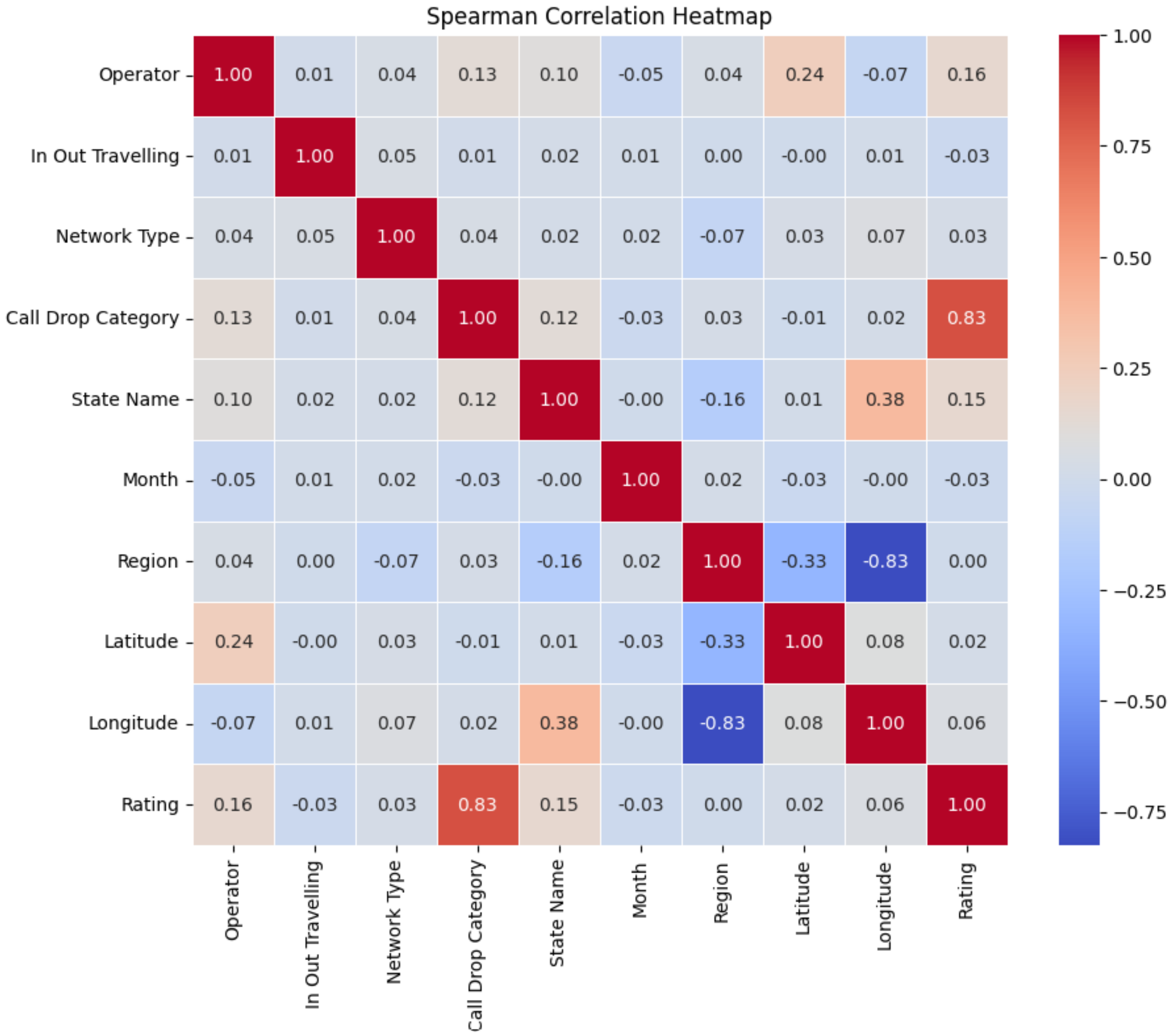


	Features	VIF
0	Operator	3.321250
1	In Out Travelling	1.391362
2	Network Type	6.642759
3	Call Drop Category	7.129689
4	State Name	8.346654
5	Month	2.923588
6	Region	6.349714
7	Latitude	6.123720
8	Longitude	6.812539

In the analysis of multicollinearity among the features, Variance Inflation Factor (VIF) was used to assess the presence of collinearity, especially among the categorical and numeric features. To perform this analysis, scaling was applied to the features, and VIF values were calculated. The results showed that all VIF values were below the threshold of 10, indicating that there is no significant multicollinearity among the features. This suggests that the features are independent of each other and will not cause instability in the model. As a result, all the features, including categorical ones, were retained and used for building the classification model, ensuring reliable and stable predictions.

Fig 6: Variance Inflation Factor (VIF) Analysis for Feature Multicollinearity

2. Checking Multi-collinearity using Correlation



	Operator	In Out Travelling	Network Type	Call Drop Category	State Name	Month	Region	Latitude	Longitude	Rating
Operator	1.00	0.01	0.04	0.13	0.10	-0.05	0.04	0.24	-0.07	0.16
In Out Travelling	0.01	1.00	0.05	0.01	0.02	0.01	0.00	-0.00	0.01	-0.03
Network Type	0.04	0.05	1.00	0.04	0.02	0.02	-0.07	0.03	0.07	0.03
Call Drop Category	0.13	0.01	0.04	1.00	0.12	-0.03	0.03	-0.01	0.02	0.83
State Name	0.10	0.02	0.02	0.12	1.00	-0.00	-0.16	0.01	0.38	0.15
Month	-0.05	0.01	0.02	-0.03	-0.00	1.00	0.02	-0.03	-0.00	-0.03
Region	0.04	0.00	-0.07	0.03	-0.16	0.02	1.00	-0.33	-0.83	0.00
Latitude	0.24	-0.00	0.03	-0.01	0.01	-0.03	-0.33	1.00	0.08	0.02
Longitude	-0.07	0.01	0.07	0.02	0.38	-0.00	-0.83	0.08	1.00	0.06
Rating	0.16	-0.03	0.03	0.83	0.15	-0.03	0.00	0.02	0.06	1.00

Fig 7 : Heatmap of Correlation Matrix for Feature Relationships

The correlation analysis of the features with the target variable - Rating, reveals that Call Drop Category has a strong positive correlation, making it the most important feature for prediction,while other show weak correlations, indicating minimal linear relationships with Rating.

Since there is no significant multicollinearity between independent variables, all features are retained. However, due to the lack of strong linear relationships between the features and the target, Logistic Regression is not the optimal choice for the base model. Instead, more flexible models like Decision Trees or Random Forests are better suited to capture the complexity of the data and improve predictive accuracy.

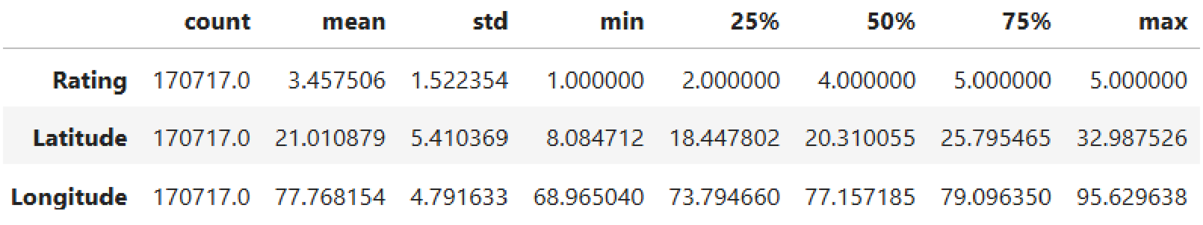
Dimensionality Reduction:

Since the dataset contains only a few features, there is no need for dimensionality reduction, as reducing dimensions could lead to the loss of important information. Therefore, dimensionality reduction will not be applied.

After completing all the preprocessing and feature engineering, the final shape of the dataset is:
No of Rows : 170717
No of Columns : 10

Step 4: Descriptive Analysis

Descriptive Statistics – Numerical variables



	count	mean	std	min	25%	50%	75%	max
Rating	170717.0	3.457506	1.522354	1.000000	2.000000	4.000000	5.000000	5.000000
Latitude	170717.0	21.010879	5.410369	8.084712	18.447802	20.310055	25.795465	32.987526
Longitude	170717.0	77.768154	4.791633	68.965040	73.794660	77.157185	79.096350	95.629638

Fig 8: Descriptive Statistics of Numerical Features

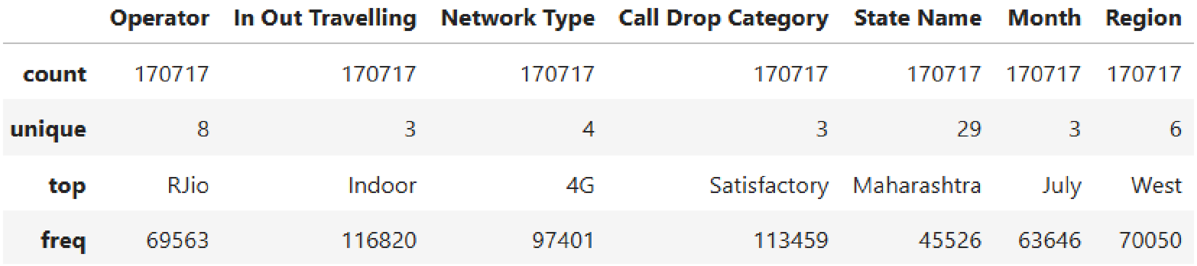
<u>**Inference**</u>

1. Target Variable (Rating):
   - The average rating is 3.46, which suggests a slight bias towards higher ratings.
   - Inference: The distribution of the Rating variable appears skewed toward higher values, suggesting potential class imbalance, where higher ratings (4 and 5) are dominant.

2. Latitude:
   - Min: 8.08 (close to southern latitudes of India).
   - Max: 32.99 (northern regions of India).
   - Inference: Latitude values gives the geographical data spans a wide range from southern to northern India.

3. Longitude:
   - Min: 68.96 (westernmost parts of India, e.g., Gujarat).
   - Max: 95.63 (easternmost regions of India, e.g., Arunachal Pradesh).
   - Inference: Longitude values confirm coverage from western to eastern India, indicating a geographically diverse dataset.

Descriptive Statistics – Categorical variables



	Operator	In Out Travelling	Network Type	Call Drop Category	State Name	Month	Region
count	170717	170717	170717	170717	170717	170717	170717
unique	8	3	4	3	29	3	6
top	RJio	Indoor	4G	Satisfactory	Maharashtra	July	West
freq	69563	116820	97401	113459	45526	63646	70050

Fig 9: Descriptive Statistics of Categorical Features

Inference

- RJio is the leading operator in this dataset, indicating it has a significant market share or user base.
- A majority of the call records are from indoor environments, which could be due to higher stability of indoor networks or user preference.
- The prevalence of 4G network usage indicates that it is the dominant technology among users in this dataset.
- Most call drops are categorized as "Satisfactory," implying that the service quality is generally acceptable but could still have some issues.
- Maharashtra has the highest number of call records, indicating a larger user base or more reporting from this state.
- The West region appears to have the highest number of records, which could correlate with population density or network infrastructure.

Step 5: Statistical Testing

Chi-Square (χ^2) Independence Test - Categorical vs Categorical

Statistical test for between categorical columns and target column

Null Hypothesis and Alternative Hypothesis

- Null Hypothesis (H_0): The target variable (Rating) is independent of the independent feature.
- Alternative Hypothesis (H_1): The target variable (Rating) is dependent on the independent feature.

When Null Hypothesis is Rejected:

- If the p-value < α (e.g., 0.05), the test suggests a statistically significant relationship between the feature and the target variable.This indicates that the independent feature could provide predictive power for target variable (Rating), making it relevant to include in the model.

When Null Hypothesis is Not Rejected:

- If the p-value ≥ α, the test does not provide evidence of a relationship between the feature and the target variable.The feature may not contribute meaningfully to predicting the target variable and can be considered for exclusion.

Feature	Chi-Squared Statistic	Degrees of Freedom	p-value	Conclusion
Operator	8114.759	28	0	Reject the null hypothesis. Ratings depend on Operator.
In Out Travelling	1614.001	8	0	Reject the null hypothesis. Ratings depend on In Out Travelling.
Network Type	2262.232	12	0	Reject the null hypothesis. Ratings depend on Network Type.
Call Drop Category	158618.636	8	0	Reject the null hypothesis. Ratings depend on Call Drop Category.
State Name	47363.651	112	0	Reject the null hypothesis. Ratings depend on State Name.
Month	310.745	8	2.12e-62	Reject the null hypothesis. Ratings depend on Month.
Region	8270.099	20	0	Reject the null hypothesis. Ratings depend on Region.

Tab 2 : Chi-Square Independence Test: p-value and Interpretation

Inference

The Chi-Square test results show that all categorical features have rejected the null hypothesis, indicating a significant association with the target variable. This suggests that each feature contributes meaningfully to the target and will be considered for model building.

Kruskal-Wallis H Test - Categorical vs Numerical

Statistical test for between numerical columns and target column

Since Latitude and Longitude doesn't follow a normal distribution, the Kruskal-Wallis H Test is used as a non-parametric test to compare the distributions of these numerical features across the different Rating categories. This is especially useful when the assumptions of parametric tests like ANOVA (normality and homogeneity of variance) may not hold.

Null Hypothesis and Alternative Hypothesis

- Null Hypothesis (H_0): The target variable (Rating) is independent of the independent feature (e.g., Latitude or Longitude). That is, the distribution of the feature does not differ across the different categories of the target variable.

- Alternative Hypothesis (H_1): The target variable (Rating) is dependent on the independent feature. That is, the distribution of the feature differs across at least one category of the target variable.

When Null Hypothesis is Rejected:

- Dependence Identified: If the p-value < α 0.05, the test suggests a statistically significant difference in the distribution of the feature (Latitude, Longitude, etc.) across the different categories of Rating. This indicates that the feature is related to the target variable and may contribute valuable information in predicting the target.
- Model Relevance: This suggests that the independent feature could be predictive of the target variable (Rating) and is therefore relevant to be included in the model for further analysis.

When Null Hypothesis is Not Rejected:

- No Evidence of Dependence: If the p-value ≥ α, the test does not provide enough evidence to suggest a relationship between the independent feature and the target variable. In this case, the feature may not show a statistically significant difference across the Rating categories.
- Possible Irrelevance: The feature may not provide meaningful information for predicting the target variable (Rating) and could be considered for exclusion from the model.

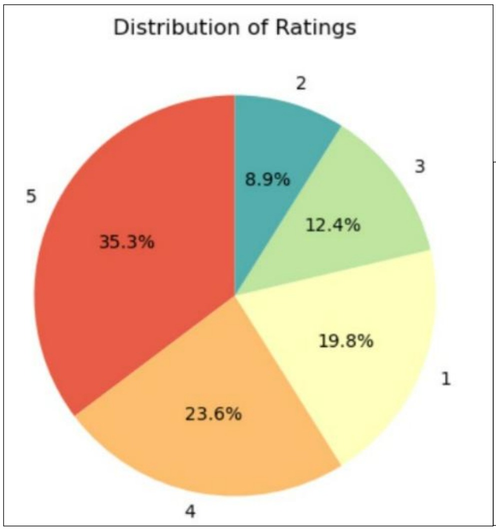
Feature	Kruskal-Wallis Statistic	p-value	Conclusion
Latitude	2013.26	0	Reject the null hypothesis. Latitude depends on Rating.
Longitude	2141.78	0	Reject the null hypothesis. Longitude depends on Rating.

Tab 3 : Krushkal-Wallis H Test : p-value and Interpretation

Inference

The Kruskal-Wallis H Test results indicate that all tested numerical features (e.g., Latitude, Longitude) have rejected the null hypothesis, suggesting a significant association between each feature and the target variable (Rating). This means that the distributions of these numerical features differ across the Rating categories, and they are likely to contribute valuable information in predicting the target. As a result, these features will be retained and considered for model building.

Step 6: Class imbalance and its treatment



The majority of the ratings are concentrated in the higher ratings (such as 5, with 35.3% of the total), while ratings like 2 and 3 have lower proportions (8.9% and 12.4%, respectively).

This imbalance could lead to challenges in model training, as the model may become biased towards predicting the majority classes (such as 5).

Fig 10: Visualization of Class Imbalance

Techniques for Handling Class Imbalance:

Ensemble Methods: Techniques like Random Forests, XGBoost, and Gradient Boosting are effective in handling imbalanced datasets as they adjust weights or splitting criteria internally, improving model learning for underrepresented classes without manual intervention.

Cross-Validation: Ensures that each fold in cross-validation reflects the original class distribution, preventing any bias during model evaluation. This method was applied during hyperparameter tuning.

Class Weights: By assigning higher weights to minority classes, the model places more importance on these underrepresented categories, enhancing performance without altering the natural data distribution.

SMOTE: While SMOTE generates synthetic samples for the minority class, it is not used in this case. SMOTE may lead to over-sampling of certain locations or features, especially if duplicates already represent the same rating from the same location, which could distort the data's natural distribution by merely creating copies rather than new, unseen observations.

Model Evaluation

The target column is rating of the call quality, which is a categorical variable with values ranging from 1 to 5. This makes the problem a multi-class classification task. The goal is to build predictive models that can accurately classify the call quality rating based on various input features, such as operator, network type, geographic location, and other related factors.

Several classification algorithms, including decision trees, random forests, will be considered for model building. The model's performance will be evaluated using appropriate metrics, such as accuracy, precision, recall, and F1-score, to ensure robust predictions and handle any class imbalance.

Data Distribution Consistency: Training vs. Testing

The dataset was split into training and testing sets with a 70-30 ratio for model building. To ensure the training and testing data were representative of the same underlying distribution, we performed a series of tests to check for any significant differences. This step was crucial to confirm that the model would generalize well and perform consistently on unseen data.

Tests Performed:

1. Chi-Square Test for Categorical Variables:

This test was applied to features such as Operator, In/Out/Travelling, Network Type, Call Drop Category, State Name, Month, and Region to assess the similarity in their distributions across the training and test sets.

The p-values from the Chi-Square test for all these features were greater than 0.05, indicating that there were no significant differences in the distributions of these categorical variables between the two datasets.

2. Kolmogorov-Smirnov (KS) Test for Continuous Variables:

The KS test was applied to Rating, Latitude, and Longitude to compare their distributions between the training and test sets.

For all these continuous variables, the p-values were also greater than 0.05, suggesting that the distributions in both datasets were similar.

Conclusion:

The results of both the Chi-Square test for categorical variables and the KS test for continuous variables show that the distributions in the train and test datasets are similar. With high p-values across all tests, there is no indication of significant discrepancies. This consistency ensures that the training and testing sets are representative of the same underlying data distribution. As a result, the model can be trained and tested confidently, with reduced risks of bias or performance issues due to data distribution shifts.

Baseline Model Performance: Evaluation with Default Parameters

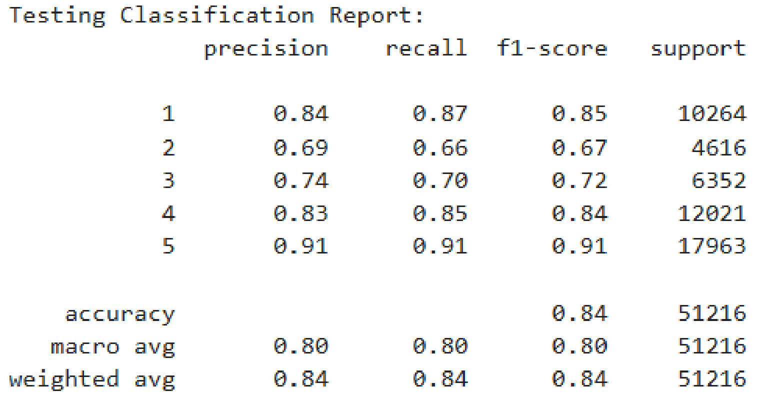
- Decision Tree Classifier – Base Model
- k-Nearest Neighbors (k-NN) Classifier
- Naïve Bayes Classifier
- Random Forest Classifier
- Gradient Boosting Classifier
- Adaptive Boosting (AdaBoost) Classifier
- Extreme Gradient Boosting (XGBoost) Classifier

Model Summary for each machine learning algorithm

Aspect	Details
Model Name	Decision Tree Classifier

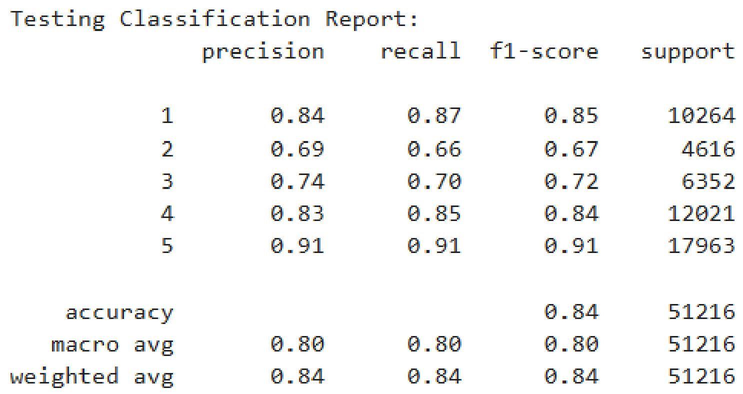
Default Parameters	criterion='gini'
	splitter='best'
	max_depth=None
	min_samples_split=2
	min_samples_leaf=1
	max_features=None
	max_leaf_nodes=None
	min_impurity_decrease=0.0
	class_weight=None

Reason for choosing	Non-linear relationships in data (ruled out logistic regression). Handles both categorical and numerical features. Simple and interpretable, making it a good baseline model.
Training Performance	Accuracy = 96.33% Weighted f1-score = 96%
Testing Performance	Accuracy = 83.76 % Weighted f1-score = 84%

Classification Report - Test	
	`Testing Classification Report:`
	`              precision    recall  f1-score   support`
	`           1       0.84      0.87      0.85     10264`
	`           2       0.69      0.66      0.67      4616`
	`           3       0.74      0.70      0.72      6352`
	`           4       0.83      0.85      0.84     12021`
	`           5       0.91      0.91      0.91     17963`
	`    accuracy                           0.84     51216`
	`   macro avg       0.80      0.80      0.80     51216`
	`weighted avg       0.84      0.84      0.84     51216`
	Fig 11: Classification report of test data for Decision Tree Classifier

Overfitting Evidence	High training accuracy (0.96) but significantly lower test accuracy (0.84), indicating overfitting.
Reasons for Poor Test Performance	Lack of Depth Control: Without limiting max_depth, the model grows too complex and memorizes training data rather than generalizing. Imbalance in Class Distribution: The model struggles with classifying ratings 1 and 2 due to insufficient emphasis on minority classes. Greedy Splitting: Decision Trees split on the best feature at each node but may fail to find globally optimal splits.
Conclusion	Implement hyperparameter tuning (e.g., limit max_depth, increase min_samples_leaf) to reduce overfitting. Use class weights to prioritize ratings 1 and 2. Consider ensemble methods (e.g., Random Forest, Gradient Boosting) for better generalization.

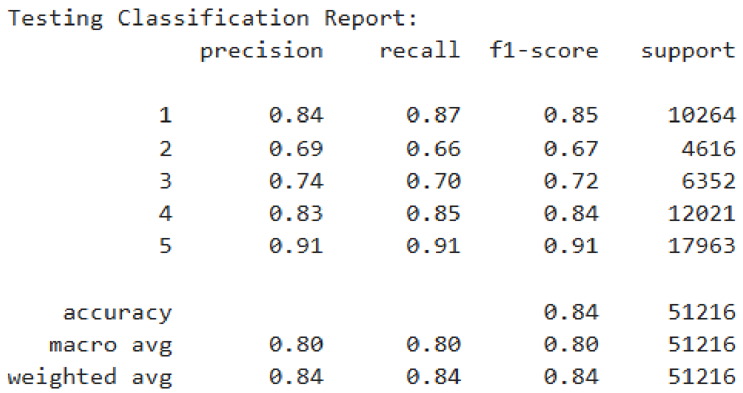
Tab 4: Decision Tree Model Performance: Summary and Insights

<table>
<tr><th>Aspect</th><th>Details</th></tr>
<tr><td>Model Name</td><td>k-Nearest Neighbors (k-NN) Classifier</td></tr>
<tr><td>Default Parameters</td><td>n_neighbors=5
weights='uniform'
algorithm='auto'
leaf_size=30
p=2
metric='minkowski’</td></tr>
<tr><td>Reason for choosing</td><td>k-NN’s non-parametric nature allows for quick evaluation without making assumptions about data distribution.
It also enables fast prototyping and computational efficiency, making it useful for understanding the data and its characteristics early in the modelling process.
Even if not used in the final model, k-NN helps establish baseline metrics for further comparison.</td></tr>
<tr><td>Training Performance</td><td>Accuracy = 87%
Weighted f1-score = 87%</td></tr>
<tr><td>Testing Performance</td><td>Accuracy = 83 %
Weighted f1-score = 83%</td></tr>
<tr><td>Classification Report - Test</td><td>
<pre>
Testing Classification Report:
              precision   recall f1-score   support

           1   0.84   0.87   0.85   10264
           2   0.69   0.66   0.67   4616
           3   0.74   0.70   0.72   6352
           4   0.83   0.85   0.84   12021
           5   0.91   0.91   0.91   17963

    accuracy   0.84   51216
   macro avg   0.80   0.80   0.80   51216
weighted avg   0.84   0.84   0.84   51216
</pre>
Fig 12: Classification report of test data for k-NN Classifier</td></tr>
<tr><td>Limitations of k-NN for This Dataset</td><td>Categorical Features: k-NN is not well-suited for datasets with many categorical features like operator, in/out/travel, and network type. Distance metrics used in k-NN are more effective for numeric and continuous data, leading to reduced model efficacy.
Lazy Learning: k-NN stores all training data and makes predictions at runtime, making it computationally expensive and slower on large datasets. This is impractical for real-time churn prediction scenarios.
Fixed k Value: The default k=5 may not be optimal for this dataset, as it may not balance bias and variance effectively.</td></tr>
<tr><td>Conclusion</td><td>Given the high proportion of categorical features, the inefficiency of k-NN’s lazy learning approach, and its reliance on distance-based metrics that are less effective for mixed-type data, k-NN is not a suitable model for this classification problem. It will not be considered for further optimization.</td></tr>
</table>

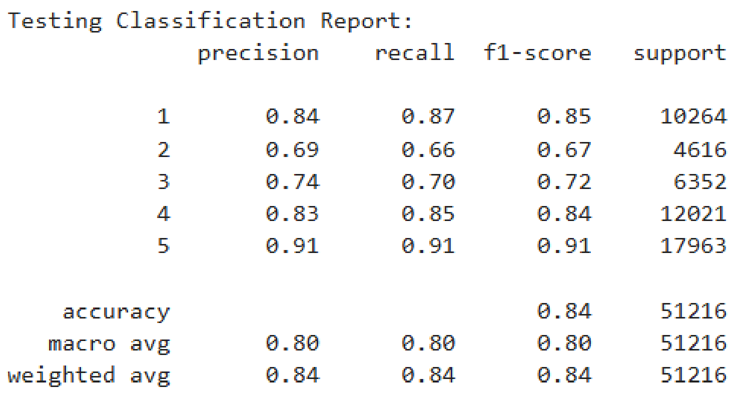
Tab 5: k-NN Model Performance: Summary and Insights

Aspect	**Details**
Model Name	Naïve Bayes Classifier
Reason for choosing	Naive Bayes was selected to quickly check model performance with simpler assumptions (Gaussian distribution for features) and a relatively straightforward approach.
Training Performance	Accuracy = 57% Weighted f1-score = 53%
Testing Performance	Accuracy = 57% Weighted f1-score = 52%
Classification Report - Test	 Testing Classification Report: precision recall f1-score support 1 0.84 0.87 0.85 10264 2 0.69 0.66 0.67 4616 3 0.74 0.70 0.72 6352 4 0.83 0.85 0.84 12021 5 0.91 0.91 0.91 17963 accuracy 0.84 51216 macro avg 0.80 0.80 0.80 51216 weighted avg 0.84 0.84 0.84 51216 Fig 13: Classification report of test data for Naïve Bayes Classifier
Reasons for Poor Test Performance	The model's performance is significantly low, especially for class 3, which had very poor precision and recall. This suggests that Gaussian Naive Bayes struggles with the data's feature distributions and class imbalance. Additionally, its assumptions about feature independence and Gaussian distributions do not align well with the underlying complexity of the data, leading to poor overall classification results. The model does not generalize well to unseen data as seen by the very similar performance on both training and test sets. There is no significant overfitting, but the model's predictive power is weak overall.
Conclusion	It should be excluded from further consideration for this dataset, as more complex models like XGBoost and Gradient Boosting can show better performance.

Tab 6: Naïve Bayes Model Performance: Summary and Insights

Aspect	**Details**
Model Name	Random Forest Classifier

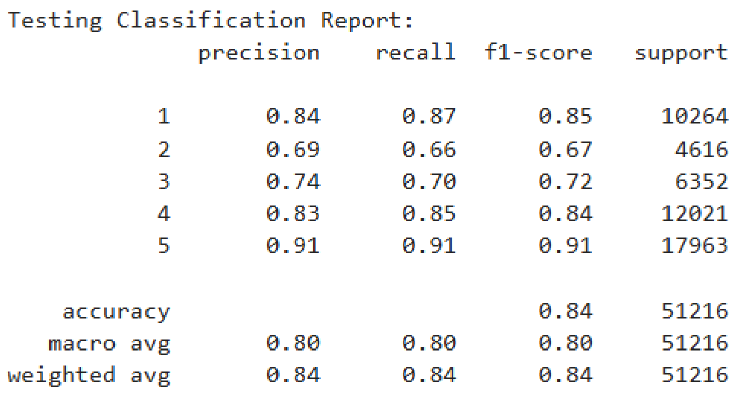
Default Parameters	n_estimators=100
	criterion='gini'
	max_depth=None
	min_samples_split=2
	min_samples_leaf=1
	max_features='sqrt'
	max_leaf_nodes=None
	min_impurity_decrease=0.0
	bootstrap=True
	oob_score=False
	class_weight=None
	max_samples=None

Reason for choosing	Random Forest is chosen due to its ability to handle high-dimensional data, capture complex relationships, and provide good generalization capabilities for classification tasks. Ensemble method reduces overfitting compared to a single decision tree, offering better performance overall. Random Forest is less sensitive to noise and irrelevant features compared to simpler models like decision trees.
Training Performance	Accuracy = 96% Weighted f1-score = 96%
Testing Performance	Accuracy = 85 % Weighted f1-score = 85%
Classification Report - Test	 Testing Classification Report: precision recall f1-score support 1 0.84 0.87 0.85 10264 2 0.69 0.66 0.67 4616 3 0.74 0.70 0.72 6352 4 0.83 0.85 0.84 12021 5 0.91 0.91 0.91 17963 accuracy 0.84 51216 macro avg 0.80 0.80 0.80 51216 weighted avg 0.84 0.84 0.84 51216 Fig 14: Classification report of test data for Random Forest Classifier
Reasons for Poor Test Performance	Overfitting: Random Forest tends to perform very well on the training set but can overfit if the trees are too deep or the model is too complex. Imbalance in Class Distribution: The model struggles with minority classes (ratings 1 and 2) as they are underrepresented in the dataset.
Conclusion	While Random Forest provides strong performance on training data, it shows overfitting when applied to the test set. The model may need hyperparameter tuning (e.g., limiting depth, using class weights) to improve generalization and performance on minority classes.

Tab 7: Random Forest Model Performance: Summary and Insights

Aspect	Details
Model Name	Gradient Boosting Classifier

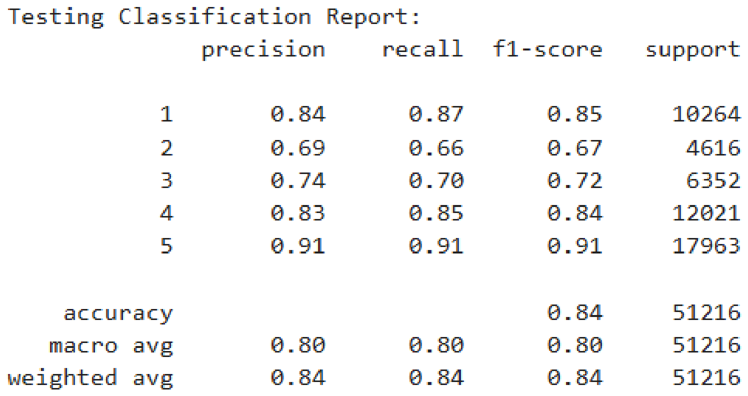
Default Parameters	loss='log_loss'
	learning_rate=0.1,
	n_estimators=10
	subsample=1.0
	criterion='friedman_mse'
	min_samples_split=2
	min_samples_leaf=1
	max_depth=3
	min_impurity_decrease=0.0
	max_features=None

Reason for choosing	Gradient Boosting was chosen because it is a powerful ensemble method known for handling complex data patterns and interactions. It typically performs better than single decision trees by combining weak learners and is less prone to overfitting compared to a single model.
Training Performance	Accuracy = 73 % Weighted f1-score = 72 %
Testing Performance	Accuracy = 73 % Weighted f1-score = 72%
Classification Report - Test	 Testing Classification Report: precision recall f1-score support 1 0.84 0.87 0.85 10264 2 0.69 0.66 0.67 4616 3 0.74 0.70 0.72 6352 4 0.83 0.85 0.84 12021 5 0.91 0.91 0.91 17963 accuracy 0.84 51216 macro avg 0.80 0.80 0.80 51216 weighted avg 0.84 0.84 0.84 51216 Fig 15: Classification report of test data for Gradient Boosting Classifier
Reasons for Poor Test Performance	Inadequate recall for minority classes: The model struggles with ratings 2 and 3 due to their lower occurrence in the dataset, which lowers the model's performance for these ratings. Class Imbalance: The model might underperform on classes with lower support (ratings 1 and 2), as it does not prioritize those classes well. Training Data Distribution: Gradient Boosting focuses on optimizing the errors made by previous estimators, which can sometimes lead to a bias against less frequent classes or classes with higher misclassification errors.
Conclusion	Gradient Boosting performs reasonably well but struggles with class imbalance and lower-represented classes, like ratings 1 and 2, impacting its overall performance. Hyperparameter tuning (e.g., adjusting the number of estimators, max depth, and class weights) could improve results.

Tab 8: Gradient Boosting Model Performance: Summary and Insights

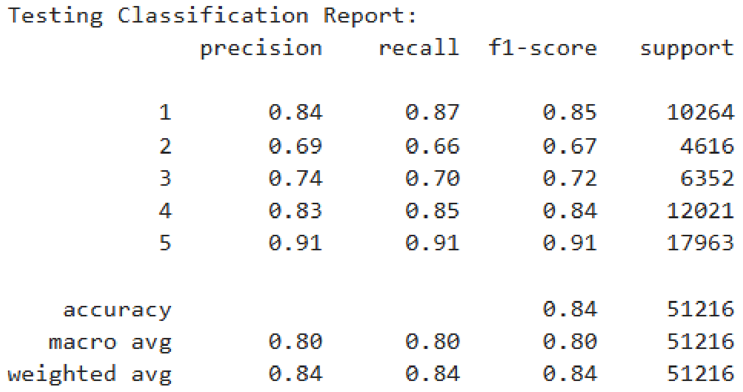
Aspect	**Details**
Model Name	Adaptive Boosting (AdaBoost) Classifier

Default Parameters	stimator=None
	n_estimators=50
	learning_rate=1.0
	algorithm='SAMME.R'

Reason for choosing	AdaBoost was chosen for its ability to boost weak classifiers and improve performance iteratively by focusing on errors made by previous models. It is effective for classification tasks with complex relationships between features and the target variable.
Training Performance	Accuracy = 96% Weighted f1-score = 96%
Testing Performance	Accuracy = 84 % Weighted f1-score = 84%
Classification Report - Test	Testing Classification Report: precision recall f1-score support 1 0.84 0.87 0.85 10264 2 0.69 0.66 0.67 4616 3 0.74 0.70 0.72 6352 4 0.83 0.85 0.84 12021 5 0.91 0.91 0.91 17963 accuracy 0.84 51216 macro avg 0.80 0.80 0.80 51216 weighted avg 0.84 0.84 0.84 51216 Fig 16: Classification report of test data for AdaBoost Classifier
Reasons for Poor Test Performance	High Complexity of the Model: AdaBoost is sensitive to noise and outliers, and may overfit on training data if not properly tuned. Inadequate Regularization: Default parameters do not include sufficient regularization to prevent overfitting, especially with a large number of estimators.
Conclusion	Despite the training accuracy being high, the model struggles to generalize well to the test data, indicating overfitting. Given this issue, and the fact that other models such as Gradient Boosting and XGBoost have provided better performance in terms of accuracy and generalization, this model is not suitable for further tuning or deployment.

Tab 9: Ada Boost Model Performance: Summary and Insights

<table>
<tr><th>Aspect</th><th>Details</th></tr>
<tr><td>Model Name</td><td>Extreme Gradient Boosting (XGBoost) Classifier</td></tr>
<tr><td>Default Parameters</td><td>n_estimators: 100
learning_rate: 0.1
max_depth: 6</td></tr>
<tr><td>Reason for choosing</td><td>XGBoost was chosen for its strong performance with imbalanced datasets, ability to handle large datasets, and flexibility to tune various parameters for improvement.
XGBoost has performed well in classification tasks with complex patterns.</td></tr>
<tr><td>Training Performance</td><td>Accuracy = 84%
Weighted f1-score = 84%</td></tr>
<tr><td>Testing Performance</td><td>Accuracy = 84%
Weighted f1-score = 84%</td></tr>
<tr><td>Classification Report - Test</td><td>



```
Testing Classification Report:
              precision    recall  f1-score   support

           1       0.84      0.87      0.85     10264
           2       0.69      0.66      0.67      4616
           3       0.74      0.70      0.72      6352
           4       0.83      0.85      0.84     12021
           5       0.91      0.91      0.91     17963

    accuracy                           0.84     51216
   macro avg       0.80      0.80      0.80     51216
weighted avg       0.84      0.84      0.84     51216
```

Fig 17: Classification report of test data for XGBoost Classifier</td></tr>
<tr><td>Reasons for Performance</td><td>The model is performing well, with a good balance between precision and recall for most categories.
Accuracy Difference: The small difference between training (0.8427) and test accuracy (0.822) indicates that the model is not overfitting and is generalizing reasonably well.
Minor Drop in Accuracy: The drop in accuracy is likely due to the inherent difficulty of classifying the more imbalanced categories (low ratings 1 and 2), but the model still performs better than simpler models like AdaBoost.
Overall, the precision, recall, and F1-score for the XGBoost model are better than the other models tested.The model shows a good balance between precision and recall, especially for the low and high ratings, while avoiding overfitting, as evidenced by the minimal drop in accuracy between training and test data. This indicates strong generalization performance on unseen data.</td></tr>
<tr><td>Conclusion</td><td>XGBoost is the chosen model due to its better generalization ability, higher overall accuracy, and strong performance in handling complex data relationships. This model will be a good candidate for further optimization through hyperparameter tuning and class weights.</td></tr>
</table>

Tab 10: XGBoost Model Performance: Summary and Insights

Overall – Model Summary of all the ML algorithms – with default parameters

Model	Train Accuracy	Test Accuracy	Accuracy Difference (%)	Inference
Decision Tree Classifier	96.33%	83.76%	12.57%	High overfitting, poor generalization.
k-NN Classifier	87.24%	82.79%	4.45%	Moderate gap, better generalization than some models.
Naive Bayes Classifier	56.63%	56.62%	0.01%	Almost no overfitting; poor performance overall.
Random Forest Classifier	96.33%	85.41%	10.92%	High overfitting,
Gradient Boosting Classifier	73.98%	73.36%	0.62%	Small gap, but low accuracy
AdaBoost Classifier	96.33%	83.99%	12.34%	High overfitting, similar to Decision Tree.
XGBoost Classifier	84.27%	82.20%	2.07%	Small gap, better performance compared to others.

Tab 11: Overall Model Performance: Summary and Insights – with Default Parameters

Inference

XGBoost appears to be the best overall model here, as its test accuracy is competitive, and it maintains a low difference between train and test accuracy (2.07%). Additionally, XGBoost shows solid overall precision, recall, and F1-score, which makes it suitable for the dataset at hand.

Hyperparameter Tuning & Overfitting:

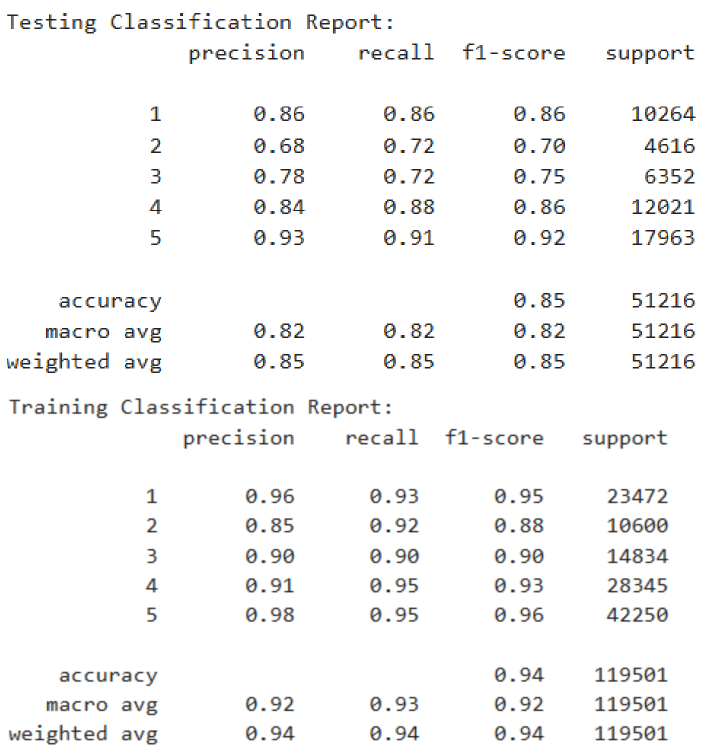
Models with a significant gap between training and test accuracy, such as Decision Tree and AdaBoost, may be overfitting to the training data. These models should be avoided in this case, as they are prone to poor generalization on unseen data despite yielding strong performance on the training set.

A difference less than 10% is generally a good indicator that the model isn't overfitting much. If the train accuracy is much higher than test accuracy, there might still be room for model regularization (for example, adjusting the learning rate, increasing or decreasing the number of estimators in Random Forest, Gradient Boosting, or XGBoost).

Since the XGBoost classifier has a low gap between train and test accuracy, it appears stable and could benefit from hyperparameter tuning without significant overfitting. Key areas to focus on during hyperparameter tuning would include:

Learning rate, Number of estimators, Maximum depth of trees, Regularization parameters (e.g., gamma, lambda, and alpha).

Model Summary After Hyperparameter Tuning and Class Weights

<table>
<tr><th>Aspect</th><th>Details</th></tr>
<tr><td>Model Name</td><td>Random Forest Classifier</td></tr>
<tr><td>Best Parameters</td><td>n_estimators: 300
min_samples_split: 5
min_samples_leaf: 1
max_features: 'log2'
max_depth: 20
class_weight= 'balanced'</td></tr>
<tr><td>Training Performance</td><td>Accuracy = 94%</td></tr>
<tr><td>Testing Performance</td><td>Accuracy = 85%</td></tr>
<tr><td>Classification Report – Train & Test</td><td>
<pre>
Testing Classification Report:
              precision   recall f1-score   support

           1   0.86   0.86   0.86   10264
           2   0.68   0.72   0.70   4616
           3   0.78   0.72   0.75   6352
           4   0.84   0.88   0.86   12021
           5   0.93   0.91   0.92   17963

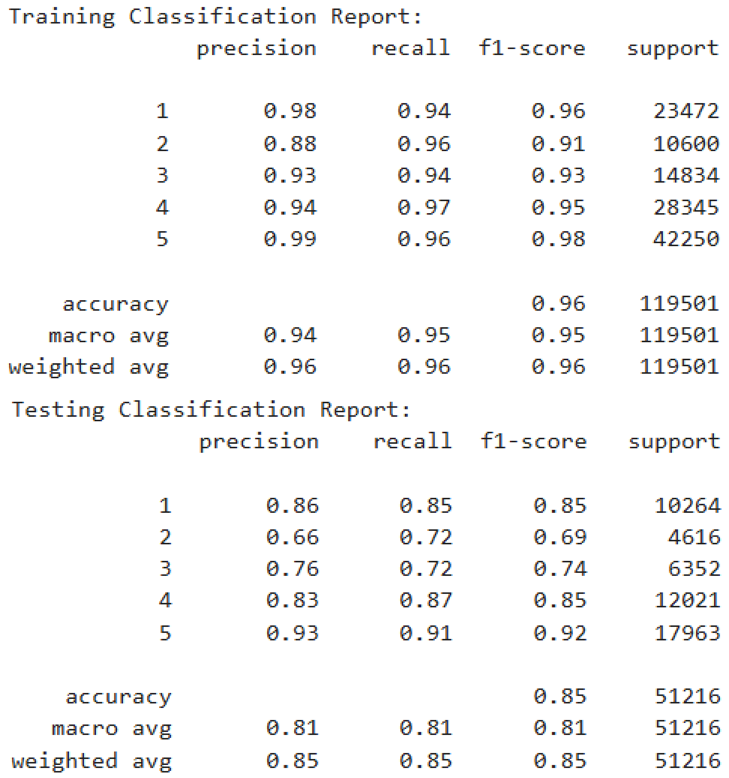
    accuracy   0.85   51216
   macro avg   0.82   0.82   0.82   51216
weighted avg   0.85   0.85   0.85   51216

Training Classification Report:
              precision   recall f1-score   support

           1   0.96   0.93   0.95   23472
           2   0.85   0.92   0.88   10600
           3   0.90   0.90   0.90   14834
           4   0.91   0.95   0.93   28345
           5   0.98   0.95   0.96   42250

    accuracy   0.94   119501
   macro avg   0.92   0.93   0.92   119501
weighted avg   0.94   0.94   0.94   119501
</pre>
Fig 18: Classification report of train & test data for Random Forest Classifier with Hyper parameter tuning & class weights</td></tr>
<tr><td>Key Inferences</td><td>Potential for Class 2 and 3 Improvement: Lower precision and F1-score for these classes indicate the need for better discrimination through additional feature engineering or different algorithms.
The tuned Random Forest Classifier with class weights reduced training accuracy by ~2% compared to the default parameters, indicating a slight decrease in overfitting.
The reduction in the training-test accuracy gap (~2%) suggests improved model robustness, though it did not significantly boost real-world predictive performance.
This indicates that the default parameters already provided a strong baseline, and further improvements may require alternative models rather than relying solely on hyperparameter tuning.</td></tr>
</table>

Tab 12: Random Forest Model Performance: Summary and Insights – with Hyperparameter & class weights

<table>
<tr><th>Aspect</th><th>Details</th></tr>
<tr><td>Model Name</td><td>Gradient Boosting Classifier</td></tr>
<tr><td>Best Parameters</td><td>'subsample': 1.0
'n_estimators': 150
'min_samples_split': 10
'min_samples_leaf': 2
'max_depth': 10
'learning_rate': 0.5</td></tr>
<tr><td>Training Performance</td><td>Accuracy = 95%</td></tr>
<tr><td>Testing Performance</td><td>Accuracy = 84%</td></tr>
<tr><td>Classification Report – Train & Test</td><td>
<pre>
Training Classification Report:
              precision   recall f1-score   support

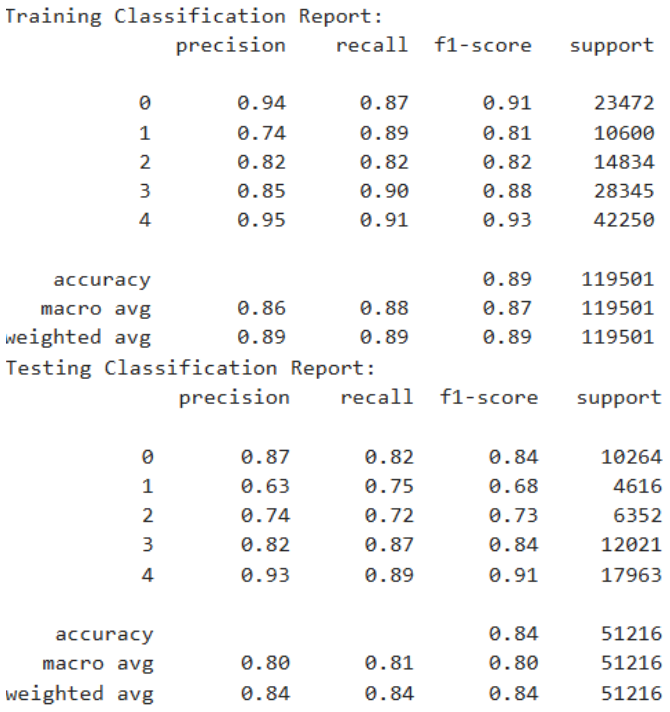
           1   0.98   0.94   0.96   23472
           2   0.88   0.96   0.91   10600
           3   0.93   0.94   0.93   14834
           4   0.94   0.97   0.95   28345
           5   0.99   0.96   0.98   42250

    accuracy   0.96   119501
   macro avg   0.94   0.95   0.95   119501
weighted avg   0.96   0.96   0.96   119501
Testing Classification Report:
              precision   recall f1-score   support

           1   0.86   0.85   0.85   10264
           2   0.66   0.72   0.69   4616
           3   0.76   0.72   0.74   6352
           4   0.83   0.87   0.85   12021
           5   0.93   0.91   0.92   17963

    accuracy   0.85   51216
   macro avg   0.81   0.81   0.81   51216
weighted avg   0.85   0.85   0.85   51216
</pre>
Fig 19: Classification report of train & test data for Gradient Boosting Classifier with Hyper parameter tuning & class weights</td></tr>
<tr><td>Key Inferences</td><td>The performance of the model before applying hyperparameter tuning and class weights showed both training and test accuracy at approximately 73%, indicating no overfitting but significant underfitting, as the model was not learning complex patterns effectively. After tuning and introducing class weights, training accuracy increased dramatically to 95.63%, and test accuracy improved to 84.65%. However, this introduced a <b>10% gap between training and test accuracy</b>, suggesting the model is now overfitting by learning the training data too well at the cost of generalization to unseen data.
Although the overall test accuracy is better than before, the large disparity highlights a need for further refinement. Techniques such as more aggressive regularization, using a different algorithm like XGBoost should be considered to achieve a better balance between bias and variance.</td></tr>
</table>

Tab 13: Gradient Boosting Model Performance: Summary and Insights – with Hyperparameter & class weights

<table>
<tr><th>Aspect</th><th>Details</th></tr>
<tr><td>Model Name</td><td>Extreme Gradient Boosting (XGBoost) Classifier</td></tr>
<tr><td>Best Parameters</td><td>'subsample': 0.8
'n_estimators': 300
'min_child_weight': 3
'max_depth': 10
'learning_rate': 0.1
'gamma': 0
'colsample_bytree': 0.8</td></tr>
<tr><td>Training Performance</td><td>Accuracy = 89%</td></tr>
<tr><td>Testing Performance</td><td>Accuracy = 84%</td></tr>
<tr><td>Classification Report
– Train & Test</td><td>
<pre>
Training Classification Report:
              precision   recall f1-score   support

           0   0.94   0.87   0.91   23472
           1   0.74   0.89   0.81   10600
           2   0.82   0.82   0.82   14834
           3   0.85   0.90   0.88   28345
           4   0.95   0.91   0.93   42250

    accuracy   0.89   119501
   macro avg   0.86   0.88   0.87   119501
weighted avg   0.89   0.89   0.89   119501
Testing Classification Report:
              precision   recall f1-score   support

           0   0.87   0.82   0.84   10264
           1   0.63   0.75   0.68   4616
           2   0.74   0.72   0.73   6352
           3   0.82   0.87   0.84   12021
           4   0.93   0.89   0.91   17963

    accuracy   0.84   51216
   macro avg   0.80   0.81   0.80   51216
weighted avg   0.84   0.84   0.84   51216
</pre>
Fig 20: Classification report of train & test data for XGBoost Classifier with Hyper parameter tuning & class weights</td></tr>
<tr><td>Key Inferences</td><td>After applying hyperparameter tuning and class weights to the XGBoost model, significant improvements were observed in both training and test performance.
The training accuracy increased to 89%, while test accuracy reached 84%. This represents a slight increase in the overfitting gap compared to the default model, which previously showed only a 2% difference between training and test accuracy.
However, the overall precision, recall, and accuracy for the tuned model improved, especially in terms of generalization, with better handling of class imbalances.
Although the model did show some overfitting, it performed better overall, making it a more reliable choice for prediction than other models like Gradient Boosting or Random Forest. Further tuning and optimizations could enhance performance even more, particularly for underperforming classes.</td></tr>
</table>

Tab 14: XGBoost Model Performance: Summary and Insights – with Hyperparameter & class weights

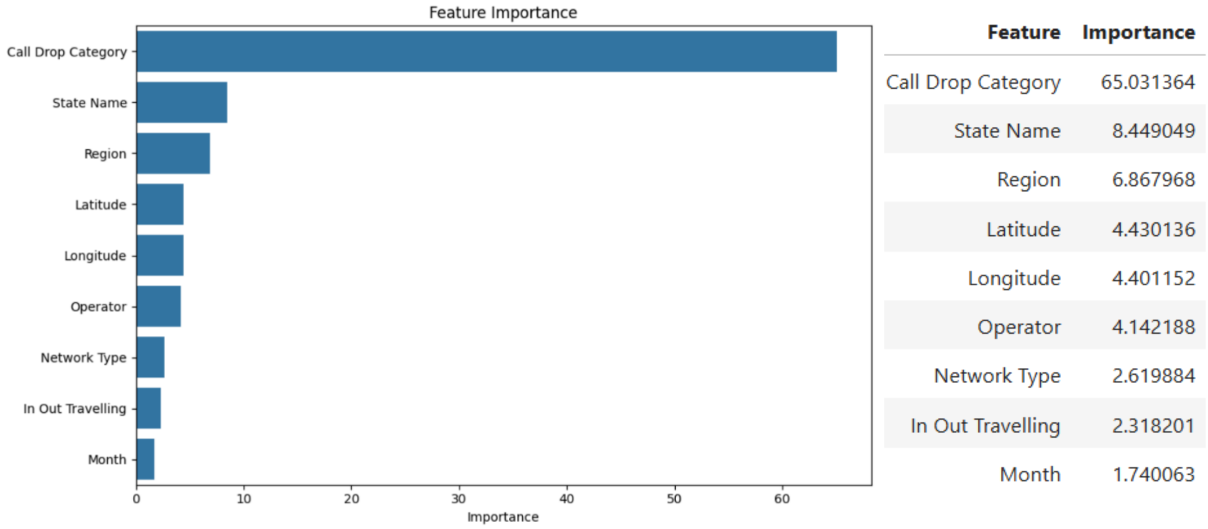
Overall Model Summary Before and After Hyperparameter Tuning

Hyperparameter tuning with class weights was performed for three classification models—Random Forest Classifier, Gradient Boosting Classifier, and XGBoost Classifier. RandomizedSearchCV was employed to identify the optimal hyperparameters for each model, ensuring better handling of class imbalance through the use of class weights.

Model	Train Accuracy	Test Accuracy	Accuracy Gap (%)	Overfitting/ Generalization	Recommended Model
Random Forest Classifier	0.9633	0.8541	10.92%	Gap difference = -2.66% High overfitting; performance drops after tuning Not recommended due to overfitting.	
Random Forest Classifier **– HT & CW**	0.9364	0.8538	8.26%		
GradientBoosting Classifier	0.7398	0.7336	0.62%	Gap difference = +10.36% Significant overfitting, but some improvement in generalization when compared to Random Forest. Not recommended due to overfitting.	
Gradient Boosting Classifier **-HT & CW**	0.9563	0.8465	10.98%		
XGBoost Classifier	0.8427	0.822	2.07%	Gap difference = +3.92% Balanced overfitting, better performance and generalization. Recommended for balanced performance	
XGBoost Classifier **-HT & CW**	0.8862	0.8363	5.99%		

Tab 15: Overall Model Performance: Summary and Insights – with Hyperparameter & class weights

Feature Importance



Feature	Importance
Call Drop Category	65.031364
State Name	8.449049
Region	6.867968
Latitude	4.430136
Longitude	4.401152
Operator	4.142188
Network Type	2.619884
In Out Travelling	2.318201
Month	1.740063

Fig 21 : Bar Chart – Feature Importance of Final Model

Inference : The model is heavily driven by Call Drop Category, followed by State Name and Region. Geographical features (latitude, longitude, and region) and telecom-related features (operator and network type) also play significant roles, while temporal factors like Month are less important. This suggests that the model is sensitive to network performance and geographical factors, with limited influence from time-related trends.

Comparison to Benchmark

When comparing the final solution to the initial benchmark, there are notable improvements in terms of both accuracy and handling class imbalance. The benchmark set out the goal of achieving strong performance across various classification models while addressing issues of overfitting and class imbalance.

- Final Solution (XGBoost with Hyperparameter Tuning and Class Weights): The model achieved test accuracy of 83.63% and train accuracy of 88.62%, outperforming the initial benchmark in terms of overall performance metrics. While a slight overfitting gap was observed, the precision and recall for individual classes improved, especially for underrepresented classes, addressing the class imbalance issue more effectively than the benchmark model.

- Benchmark Model: Initially, the models used (Gradient Boosting and Random Forest) showed decent accuracy but suffered from moderate overfitting and underperformance on certain classes. The benchmark accuracy ranged between 73% for a simpler model and 80-85% for others, with class imbalance not fully accounted for.

Conclusion

Yes, the final solution improved on the benchmark, primarily due to the implementation of hyperparameter tuning and class weights, which helped enhance both model accuracy and class balance handling. The improvements in performance were mainly seen in XGBoost, which proved to be the most robust solution in terms of generalization and class handling. However, despite the improvements, the final model still displayed some overfitting, and further optimization is needed for even better generalization.

Data Visualization – Univariate Analysis

Operator

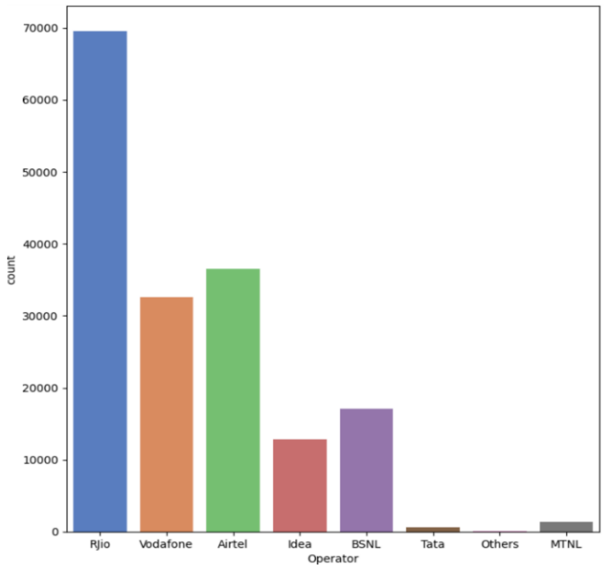


Fig 22 : Bar Chart - Frequency of Observations by Operator

Distribution: The majority of MY Call App users are on the RJio network, which holds the largest share, followed by Airtel and then Vodafone. Since Idea merged with Vodafone during this period, the dataset reflects separate operators for the two, capturing their distinct performance before the merger was completed.
Operators like MTNL, Tata, and Others have few users, indicating a niche market or underrepresentation in 2018.

RJio dominates the user base, reflecting its market leadership, driven by widespread 4G availability and affordable plans.

In/Out/Travelling

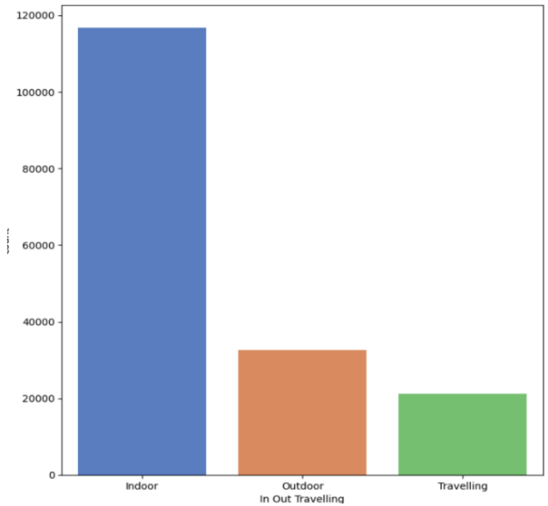


Fig 23 : Bar Chart - Frequency of Observations of In/Out/Travel

Distribution: The majority of users are either Indoor followed by Outdoor, with relatively fewer users traveling.

Inference: The app is primarily used when users are indoors or outdoors, indicating more local use.

Network Type

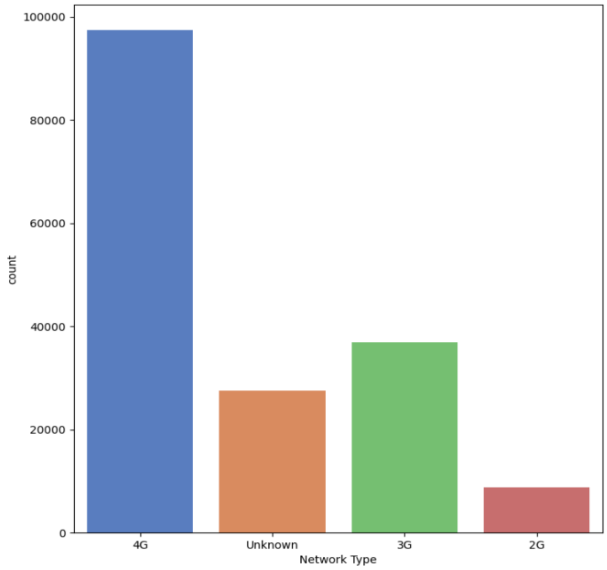


Fig 24 : Bar Chart - Frequency of Observations of Network Type

Distribution: 4G is the most commonly used network type, followed by 3G and 2G.

Inference: The dataset indicates a strong preference for 4G, which aligns with current trends in mobile usage in India. 3G is still widely used, but 2G is far less common.

Call Drop Category

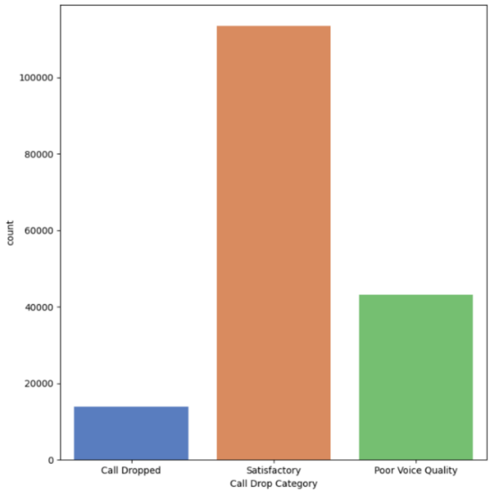


Fig 25:Bar Chart-Frequency of Observations by Call drop category

Inference: A large proportion of users experience satisfactory call quality, suggesting that network coverage is generally good. However, the number of users reporting poor voice quality and dropped calls indicates room for improvement in network performance.

Other Insights: If this data reflects user complaints or app usage, addressing the "Poor Voice Quality" category could be an opportunity for operators to improve customer experience.

Rating

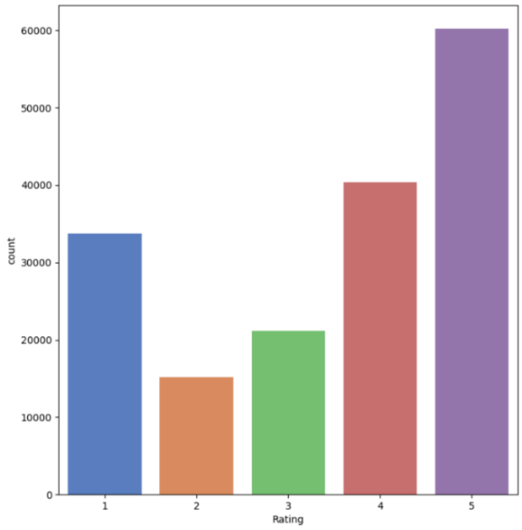


Fig 26 : Bar Chart - Frequency of Observations by Rating

Inference: The app receives mostly positive feedback, with 5 and 4 being the dominant ratings. This suggests that users are generally satisfied with the app or service. However, the presence of many 1, 2, and 3 ratings indicates areas where users are not fully satisfied. These users may have encountered issues that could require investigation (e.g., call drops, network qual

Region

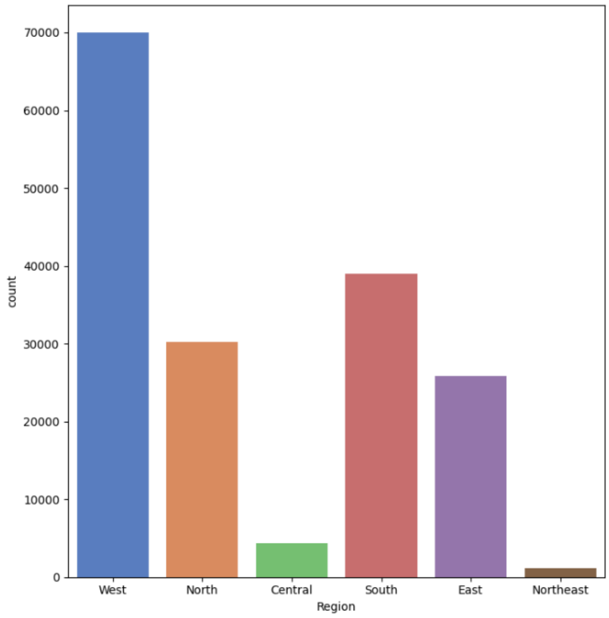


Fig 27 : Bar Chart - Frequency of Observations by Region

The West region dominates, likely due to Marketing for the app was initially focused here.

This dominance should be considered when analyzing user behaviour by region. The Northeast and Central regions show lower engagement, which could point to opportunities for targeted marketing on app or improved network infrastructure.

State Name

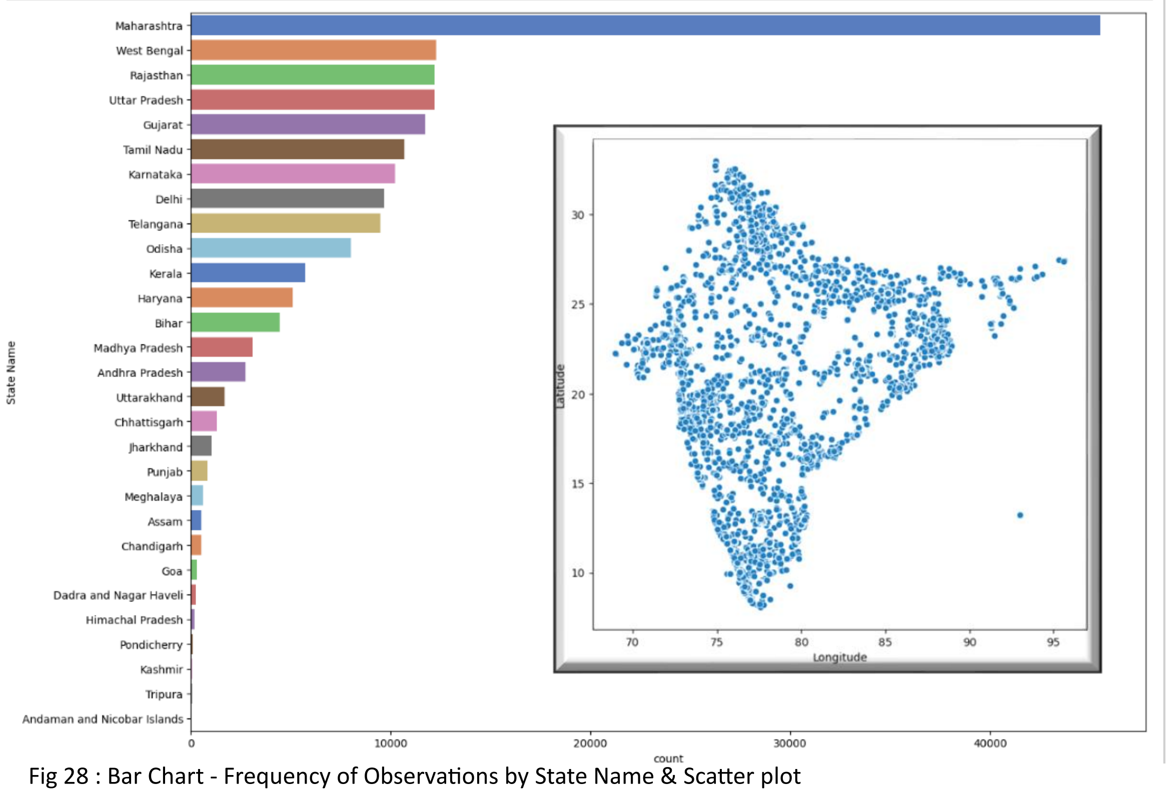


Fig 28 : Bar Chart - Frequency of Observations by State Name & Scatter plot

- Inference from Distribution: There is a higher concentration of data points in Western, Southern, and Northern India, particularly in states like Maharashtra, West Bengal, Rajasthan, Tamil Nadu, Uttar Pradesh, and Gujarat. This trend aligns with the app's reported dominance in Maharashtra, where initial marketing and launch efforts, including pilot testing and awareness campaigns, were focused to evaluate user engagement and app performance before expanding to other states. The concentration in these regions suggests better network infrastructure, higher smartphone penetration, and more active app users, contributing to the app’s adoption.

- The Northeast and Central parts of India (such as Madhya Pradesh, Chhattisgarh, and Bihar) show fewer data points. This could be indicative of lower app adoption in these areas, potentially due to limited network coverage, lower smartphone usage, or lack of targeted marketing. Similarly, areas like Jammu & Kashmir and the Andaman and Nicobar Islands show minimal user presence, likely reflecting both geographic isolation and lower adoption rates.

- The higher density in metropolitan areas such as Delhi, Mumbai, Chennai, and Bangalore is expected, as urban regions tend to have better mobile network infrastructure and higher smartphone usage, leading to increased app downloads and engagement. The distribution suggests that more urban users are adopting the MYCall app compared to rural areas, where network issues or lack of awareness may hinder adoption.

- Targeted Marketing: Focus on Northeast and Central India with localized campaigns and telecom partnerships to boost adoption. Network Infrastructure: Improve coverage and offer promotions in regions with fewer users to drive engagement.
- Urban Expansion: Strengthen presence in urban areas with new features and promotions to grow the user base. These insights can refine marketing and app development strategies.

Data Visualization – Bivariate Analysis – Operator vs Network Type

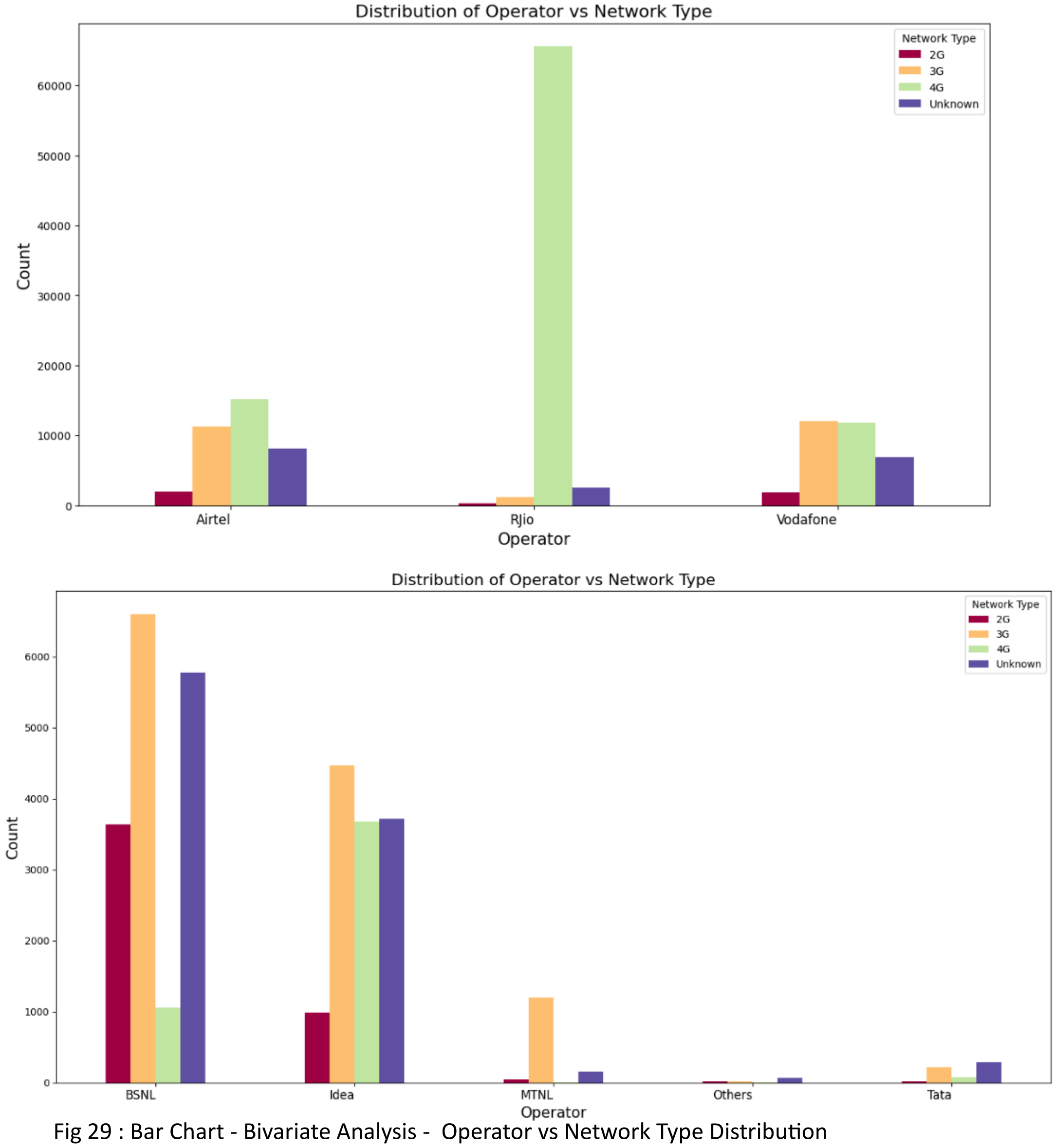


Fig 29 : Bar Chart - Bivariate Analysis - Operator vs Network Type Distribution

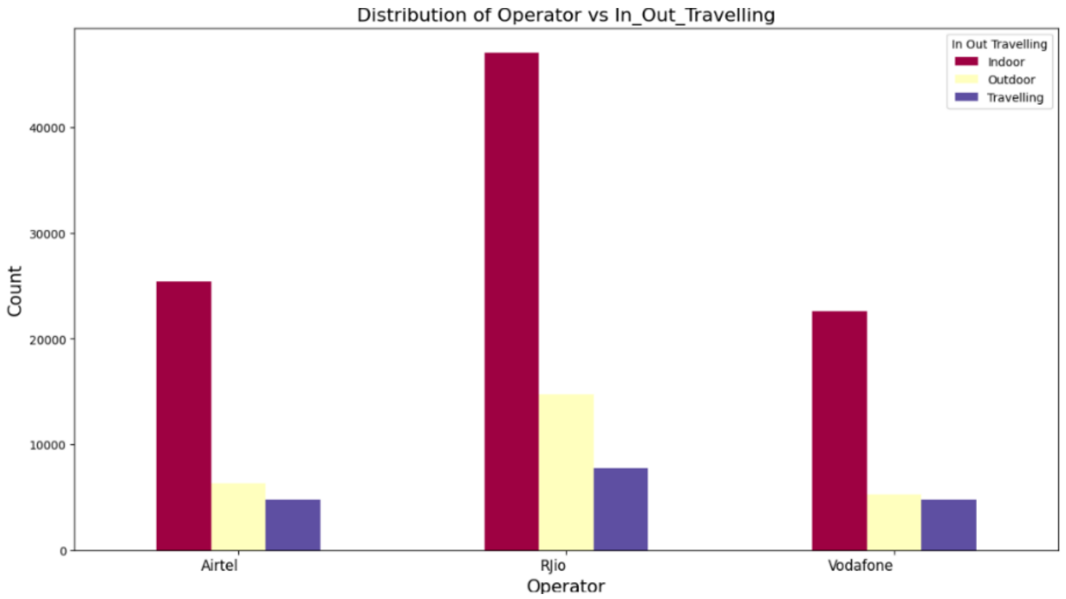
Insights:

- RJio leads with the highest number of 4G users, a result of its aggressive roll-out of 4G infrastructure and disruptive pricing. The company's focus on offering affordable 4G data and free voice calls during its early years enabled it to quickly gain a large market share.

- While Airtel and Vodafone have a significant 4G user base, they still maintain a large 3G user base. This reflects the slower pace at which they transitioned to 4G technology, perhaps due to a larger pre-existing customer base reliant on 3G.

- BSNL, Idea, and MTNL still had a substantial number of 3G users, indicating that these operators were slower to transition to 4G. This delay can be attributed to factors such as insufficient 4G spectrum, financial constraints, or limited investment in upgrading their infrastructure. Their focus remained on retaining 3G customers, while 4G adoption was relatively slower.

Data Visualization – Bivariate Analysis – Operator vs In Out Travelling



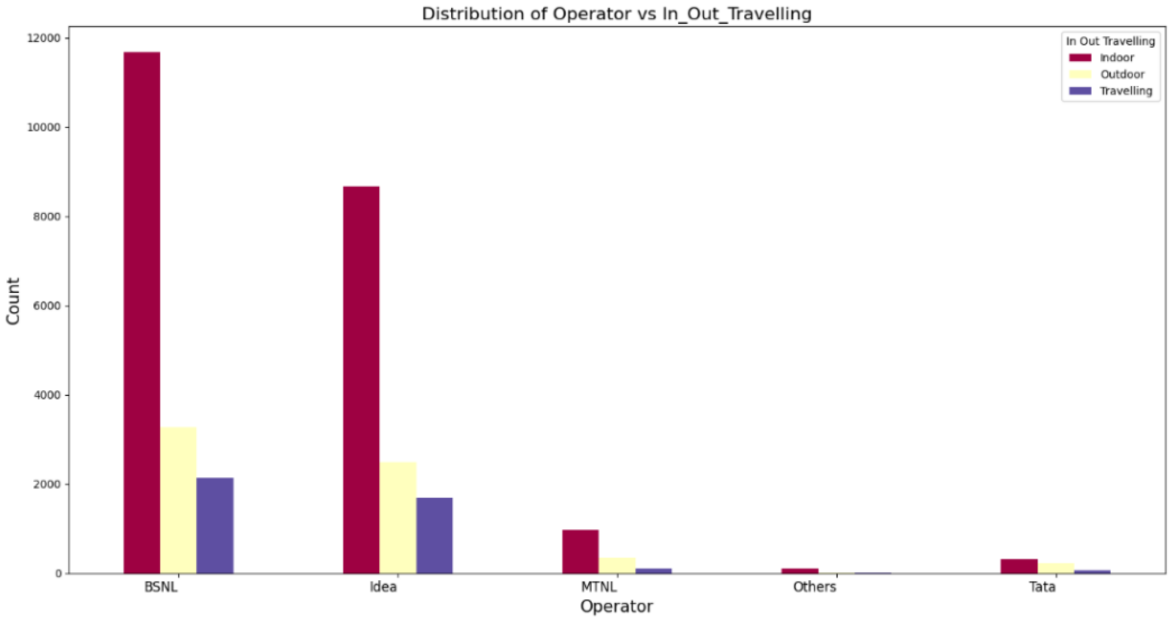


Fig 30 : Bar Chart - Bivariate Analysis - Operator vs In Out Travelling

Insights :

Across all operators, indoor usage significantly outweighs outdoor and travelling usage, suggesting better connectivity or higher activity in indoor settings.

Outdoor and travelling categories are led by RJio, Airtel, and Vodafone, likely due to their broader and more modern network infrastructure.

Operators like BSNL, Idea, MTNL, Tata, and Others have smaller user bases in all activity categories, highlighting their limited coverage or user adoption.

Data Visualization – Bivariate Analysis – Operator vs Rating

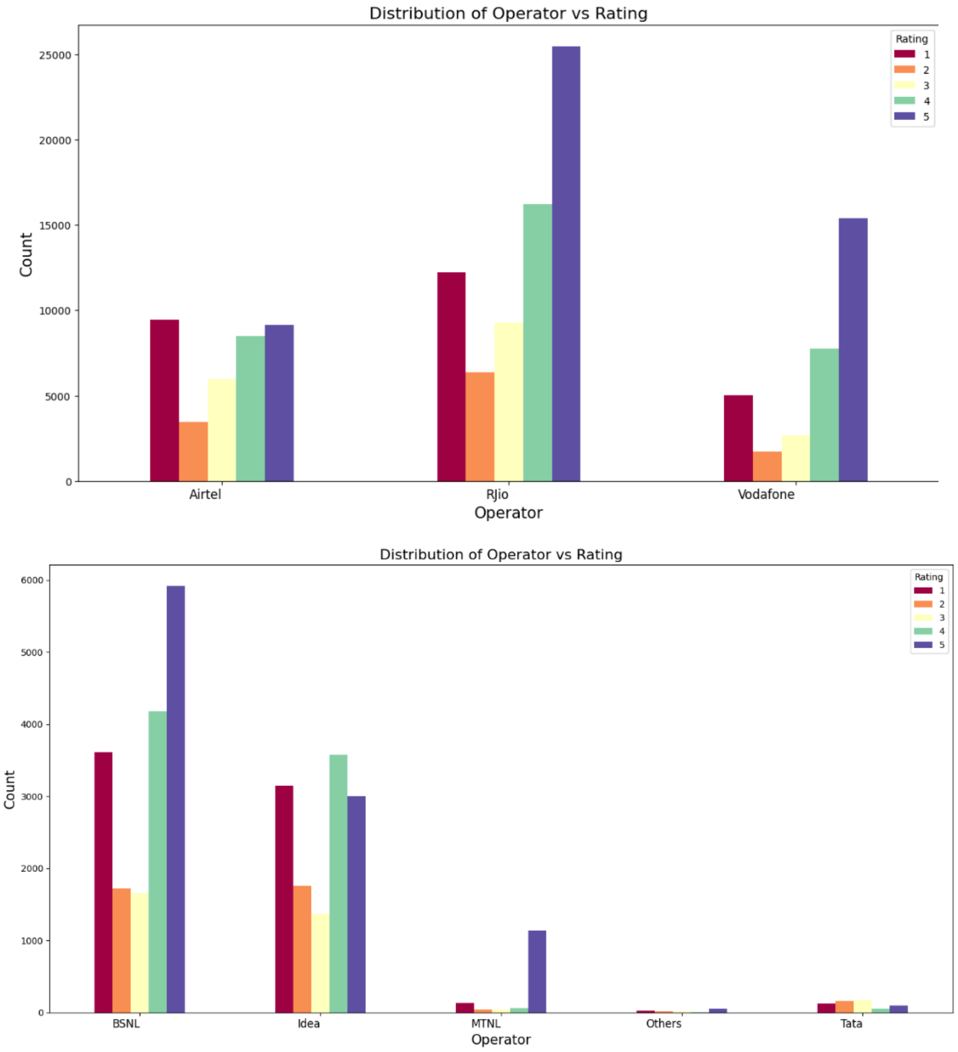


Fig 31 : Bar Chart - Bivariate Analysis - Operator vs Rating

Insights :

RJio dominates Ratings 4 and 5, offering generally good call quality, but the high Rating 1 count suggests localized issues or congestion that need attention.

Vodafone has a moderate Rating 1 count but a high Rating 5 count, indicating overall good call quality with occasional regional issues.

Airtel shows near-equal Ratings 1, 4, and 5, indicating significant variability in call quality and the need for improvements in regions with Rating 1.

Idea mirrors Airtel’s pattern, reflecting inconsistent performance and the need for better network stability and coverage.

BSNL has a high Rating 1 count, but its Rating 5 count is still noticeable, suggesting inconsistent network quality.

MTNL has low Rating 1 and high Rating 5 counts, indicating a niche but satisfied user base with limited overall market presence.

Data Visualization – Bivariate Analysis – Operator vs Average Rating

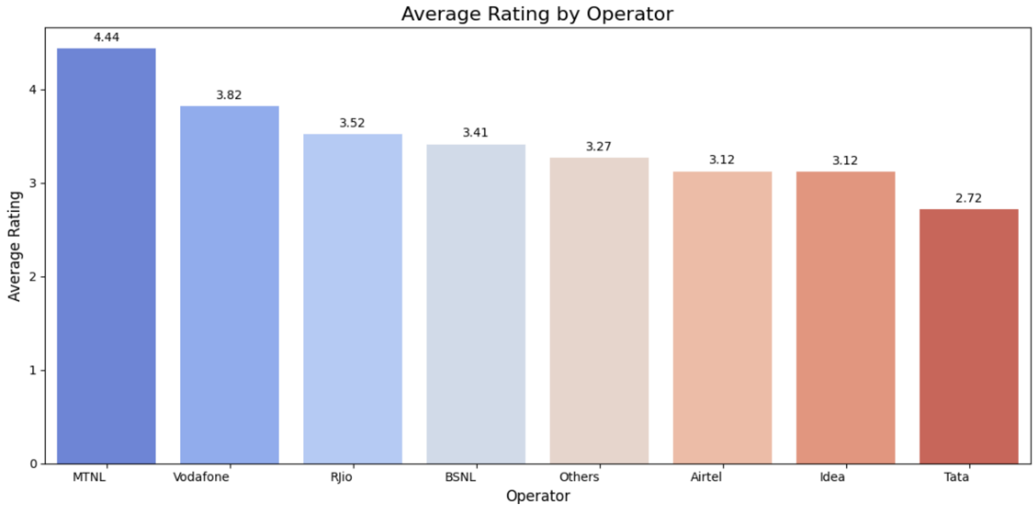


Fig 32 : Bar Chart - Bivariate Analysis - Operator vs Average Rating

Insights:

MTNL stands out with the highest average rating and a satisfied niche user base.

Vodafone and RJio show good overall quality but with some localized issues, while BSNL exhibits inconsistent service across regions.

Airtel and Idea need to address significant variability in call quality, and Tata faces the most challenges with the poorest service.

Data Visualization – Bivariate Analysis – Rating vs Call Drop Category

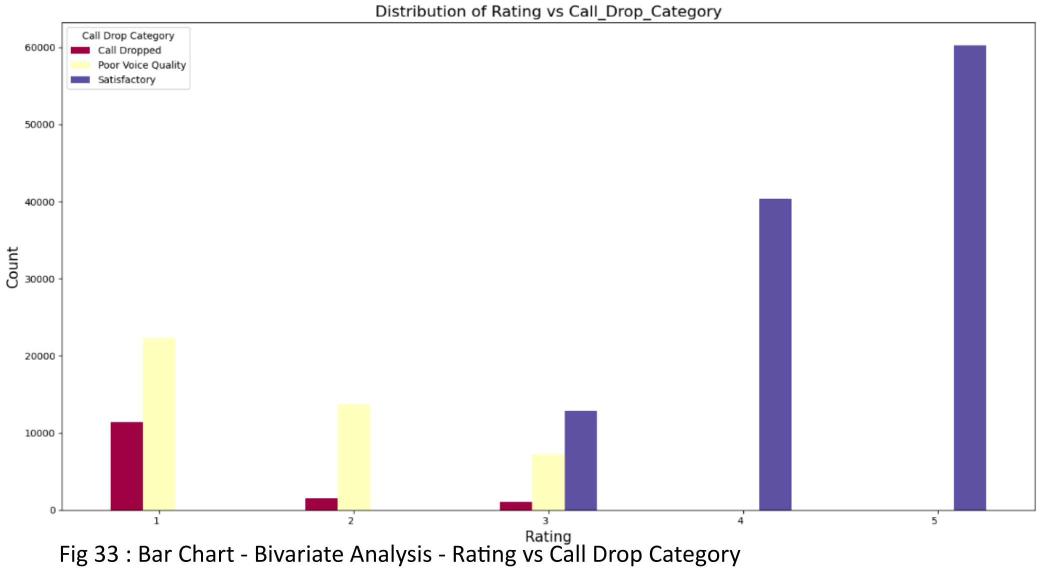


Fig 33 : Bar Chart - Bivariate Analysis - Rating vs Call Drop Category

Insights:

Ratings 1 and 2 indicate poor network quality with high instances of call drops and poor voice quality.

Rating 3 reflects moderate service with occasional issues but also satisfactory experiences for some users.

Ratings 4 and 5 show high customer satisfaction, with no significant issues, reflecting strong network performance.

Data Visualization – Bivariate Analysis – Network Type vs Rating

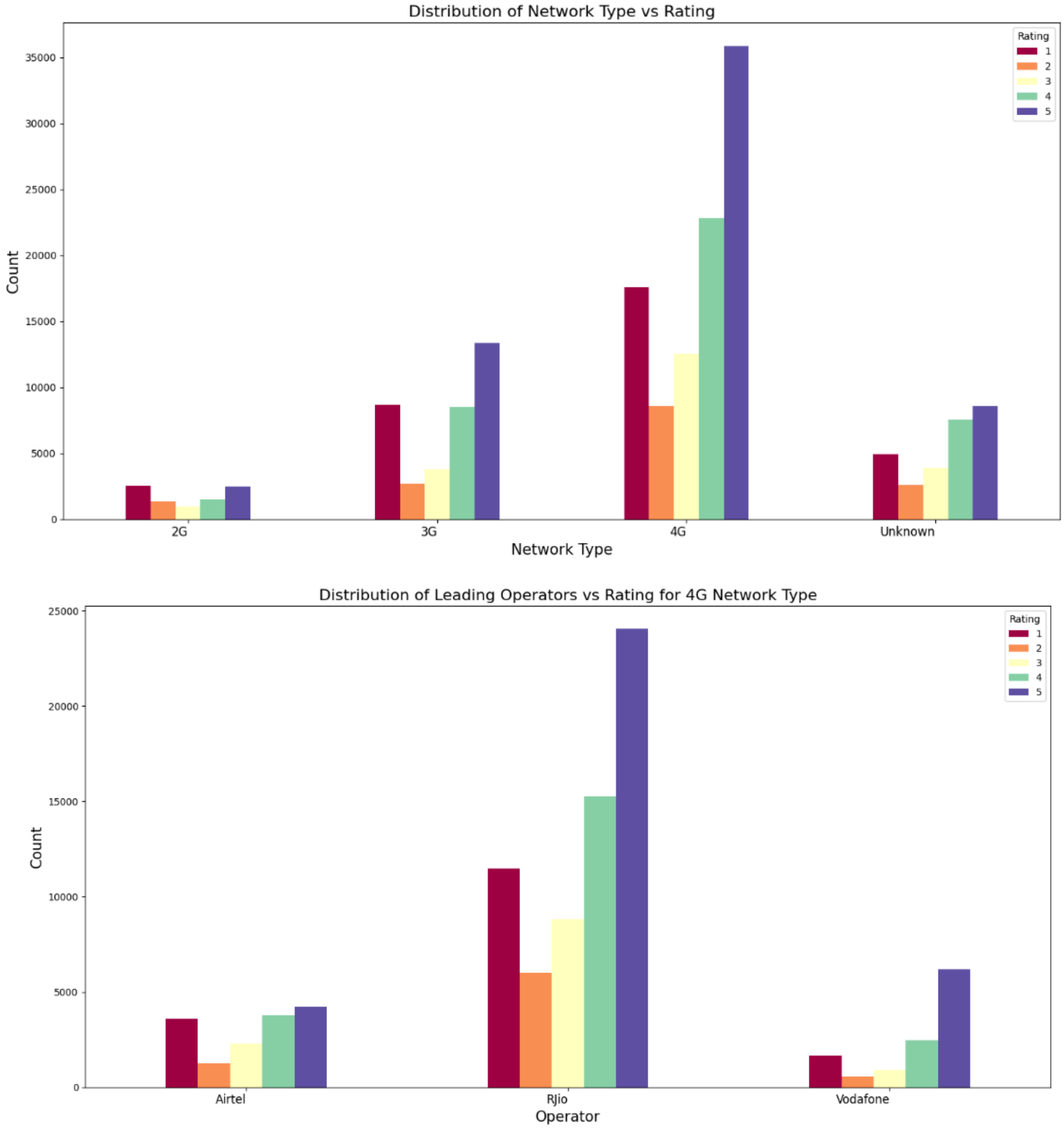


Fig 34 : Bar Chart - Bivariate Analysis – Network Type vs Rating

Insights :

2G shows the highest proportion of poor and average experiences (Ratings 1-3), indicating the network's limitations. However, some users do report satisfactory experiences.

3G provides a better balance, with a larger number of users reporting good to excellent quality (Ratings 4 and 5), though still has significant issues in some areas.

The higher number of 4G users indicates that it provides the most excellent experiences (Rating 5), but some localized issues affect a segment of users. These issues could stem from network congestion or coverage gaps in less-developed or rural regions.

To better understand the 4G pattern among leading operators, we will analyze the data to identify which operator is performing poorly in 4G.

Despite RJio's large user base, Vodafone offers the most consistent 4G experience with fewer poor ratings. Airtel needs improvements in certain areas, while RJio's high Rating 5 count is offset by a significant number of Rating 1 users, highlighting areas for network improvement.

Data Visualization – Multi-variate Analysis – Leading operators & Network vs Avg. Rating

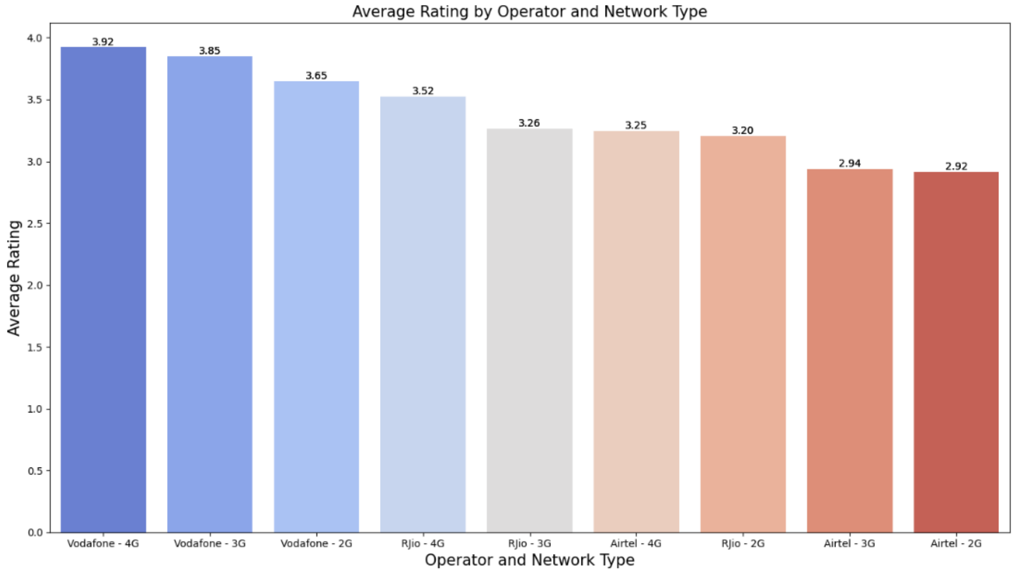


Fig 35 : Bar Chart - Multivariate Analysis – Leading Operators & Network type vs Avg. Rating

Insights:

Vodafone’s 4G network has the highest average rating, suggesting it provides the best user experience among the three operators.

RJio’s networks are rated consistently across 4G, but lower than Vodafone’s equivalent network types.

Airtel's network ratings are the lowest across the board, indicating a need for improvement in user experience.

Data Visualization – Multi-variate Analysis – Leading operators & InOutTravel vs Avg.Rating

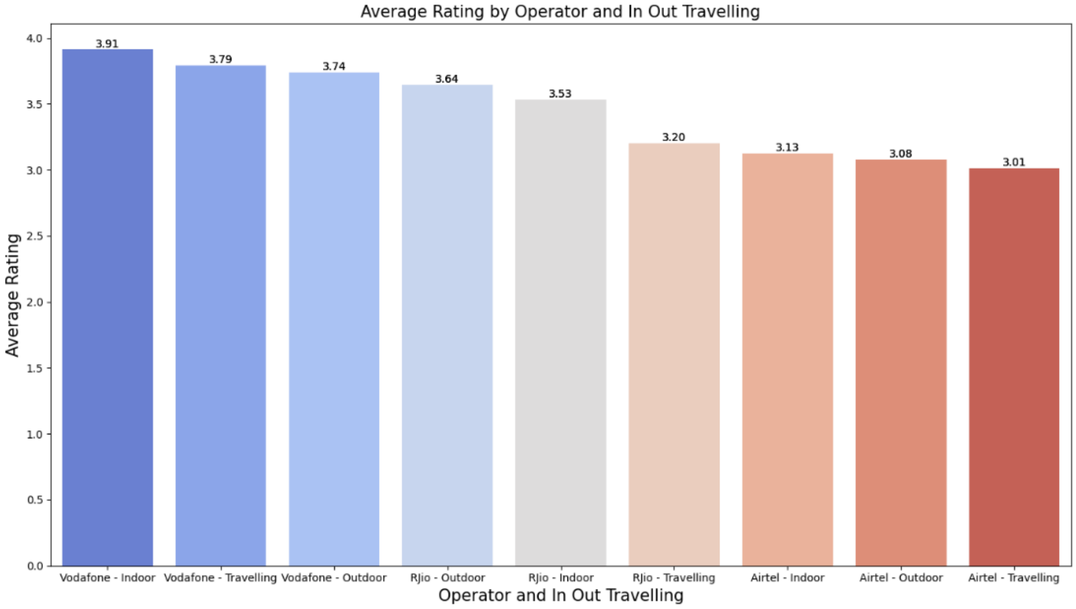


Fig 36 : Bar Chart - Multivariate Analysis – Leading Operators & In_Out_Travel vs Avg. Rating

Insights:

Vodafone is the top performer across all categories, providing the best user experience in indoor, outdoor, and traveling environments.

RJio performs well outdoors and indoors but struggles with consistency while traveling, likely due to signal variability or environmental factors.

Airtel maintains similar performance across all environments but at a lower overall level compared to Vodafone and RJio.

Data Visualization – Multi-variate Analysis – Leading operators & Region vs Avg. Rating

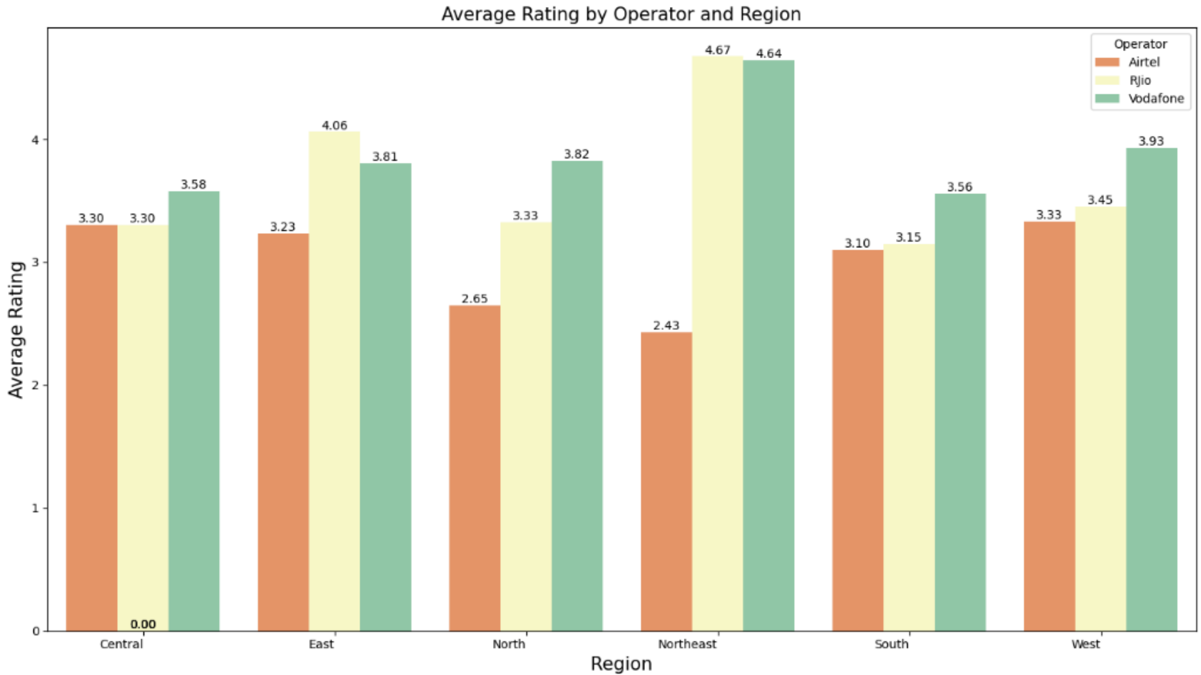


Fig 37 : Bar Chart - Multivariate Analysis – Leading Operators & Region vs Avg. Rating

Insights:

Vodafone: Consistently high ratings across all regions, particularly strong in the North East region. This suggests that Vodafone has robust network performance and high user satisfaction.

RJio: Highest ratings in the East and Northeast regions, indicating excellent performance there. RJio also performs well in the West region but has room for improvement in other regions.

Airtel: Lowest ratings in most regions, especially in the North and Northeast. This indicates a need for significant improvement in network performance and user satisfaction.

The varying ratings across regions might reflect differences in network infrastructure, service quality, and user experience, which could be influenced by factors like coverage, network speed, and customer support. The higher ratings for RJio and Vodafone in certain regions could also indicate stronger market penetration and greater customer satisfaction in those areas.

Data Visualization – Multi-variate Analysis – Leading operators & Geocoordinates – Low Rating (1,2)

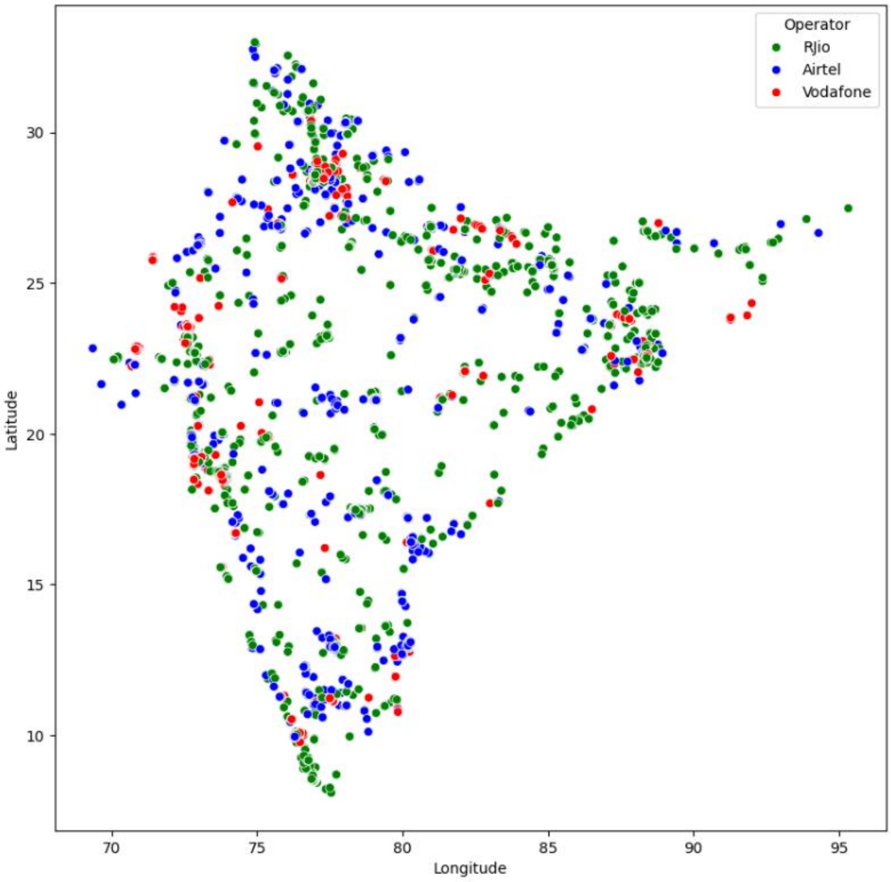


Fig 38 :Scatter Plot - Multivariate Analysis – Leading Operators & Geo-Coordinates vs Low Rating

Regions with Overall Poor Performance (All Networks): The North Indian region (specifically around state borders of Punjab, Haryana, Himachal Pradesh, and Uttarakhand) and some parts of West Bengal and Maharashtra are showing dense clusters where all three networks (RJio, Airtel, and Vodafone) perform poorly. These regions could be geographically challenging areas (e.g., hilly terrain, border regions) or underserved due to infrastructure limitations. Collaborative or independent infrastructure development is necessary in these areas to improve connectivity for all operators.

RJio shows high-density clusters of low ratings in Kerala, Gujarat, Telangana, and Odisha. These regions might be experiencing significant network congestion or coverage gaps despite being highly active areas for RJio users. Immediate measures should be taken to analyze and optimize network performance in these regions due to their dense user base. Apart from the dense clusters, RJio's network issues appear sporadically across the rest of the country, with lower density. This suggests that overall, RJio is doing better in most regions, but isolated pockets of issues remain.

Airtel's low performance is scattered across the country rather than being concentrated in specific regions. This pattern indicates widespread but less severe network issues, possibly related to capacity constraints or inconsistent service quality. Airtel should prioritize improving its network stability and capacity across diverse geographies instead of focusing on specific regions.

Vodafone's low ratings are concentrated on the periphery of the country rather than the central regions. This pattern suggests that Vodafone may have weaker infrastructure or coverage in border areas and coastal regions, potentially due to lower investment or geographical challenges.

Model Deployment (https://voice-call-quality-analysis.streamlit.app/)

The final model was deployed as a web application using Streamlit, providing an interactive platform for predicting telecom call quality ratings. Users can input relevant features like operator, call environment, network type, and geographical location to receive a predicted customer rating (1–5) for voice call quality.

Key Features:

User-friendly Interface: Streamlit offers a simple and interactive interface for both technical and non-technical users.
Users can easily adjust inputs to observe real-time predictions and insights.

Real-time Predictions: The application processes user inputs and delivers instant predictions based on the trained model.
Useful for telecom providers to quickly identify areas with potential call quality issues.

Scalability and Accessibility: Hosted as a web-based tool, the application can be accessed from anywhere, ensuring broader usability across teams and stakeholders.
Supports decision-making for network optimization and customer experience improvement.

Potential Use Cases: Identifying low-rated geographical regions to focus on service enhancement.
Providing real-time feedback on potential network issues to operators.

The deployment bridges the gap between predictive analytics and practical application, making the model accessible for real-world decision-making and actionable insights.

Feedback Loop for Rating Prediction

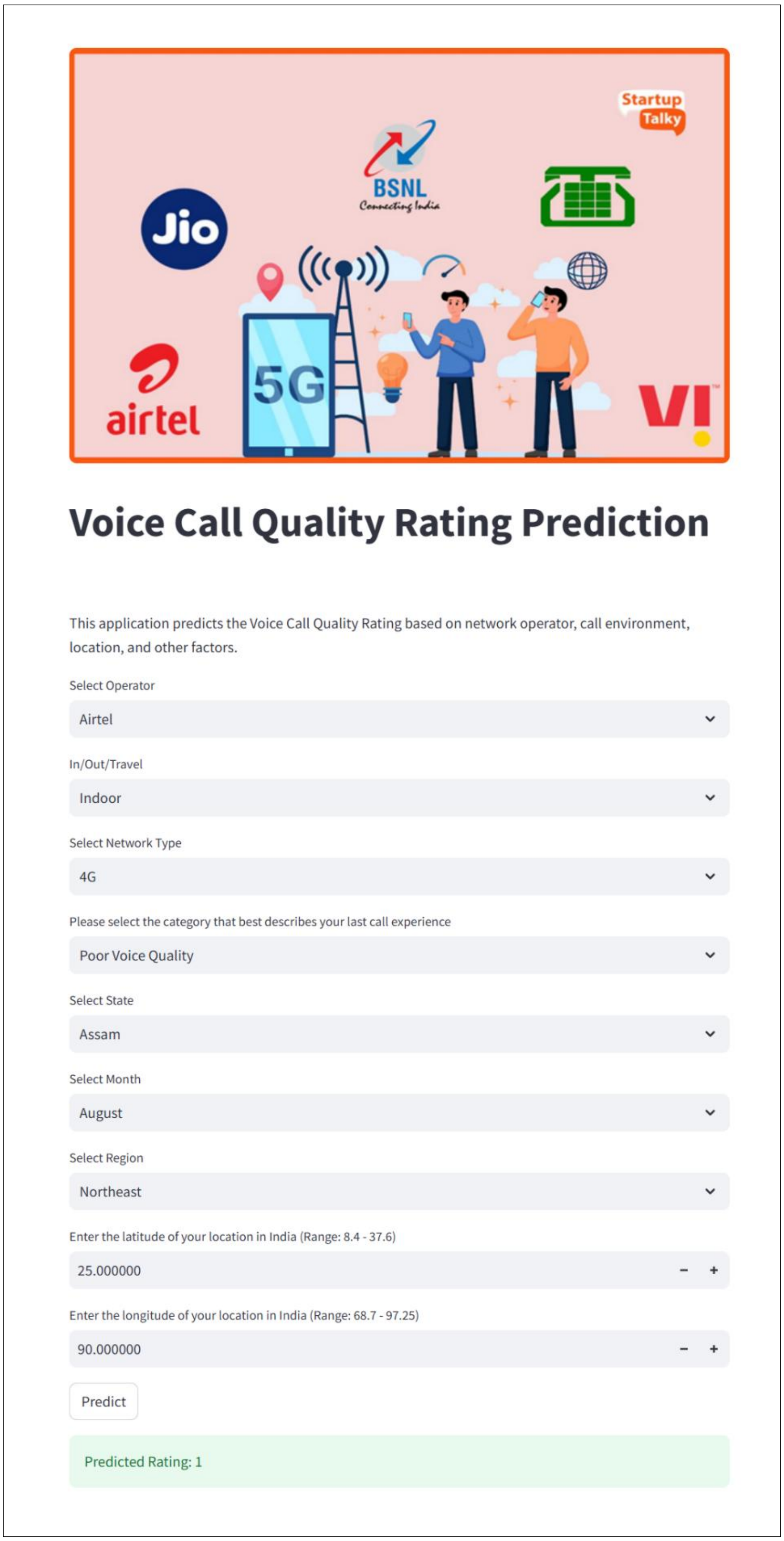
Prediction of Ratings:
The model predicts customer ratings (1–5), identifying regions, operators, or scenarios with low ratings.
These insights guide telecom providers to prioritize network upgrades or resolve specific issues.

Actionable Insights and Interventions:
Operators act on predictions to improve service quality in low-rated regions, such as deploying infrastructure upgrades or optimizing configurations.

New Data Reflects Improvements:
As corrective measures are implemented, updated customer feedback captures the effectiveness of these actions.
Improved ratings in these areas form new data patterns.

Retraining for Enhanced Predictions:
The model is retrained with updated data to reflect the latest trends, ensuring improved accuracy and relevance in future predictions.

Continuous Improvement Cycle:
This iterative loop—predictions driving actions, followed by retraining—enables ongoing service enhancements and more accurate forecasting of customer satisfaction.



Voice Call Quality Rating Prediction

This application predicts the Voice Call Quality Rating based on network operator, call environment, location, and other factors.

Select Operator
Airtel

In/Out/Travel
Indoor

Select Network Type
4G

Please select the category that best describes your last call experience
Poor Voice Quality

Select State
Assam

Select Month
August

Select Region
Northeast

Enter the latitude of your location in India (Range: 8.4 - 37.6)
25.000000

Enter the longitude of your location in India (Range: 68.7 - 97.25)
90.000000

Predict

Predicted Rating: 1

Fig 39 : Deployment Screenshot: Streamlit App

Implications

Better Customer Retention: Enhanced call quality directly reduces frustrations caused by dropped calls, poor signal strength, or low audio clarity, which are major drivers of customer churn in the telecom industry.

Satisfied customers are less likely to switch to competitors, leading to increased customer loyalty and higher lifetime value per customer. Improved retention ultimately translates to stable and growing revenue streams for telecom operators.

Targeted Improvements: The solution identifies the root causes of poor call quality (e.g., network congestion, weak signal strength, or high latency) and locates problem areas, such as regions with poor ratings.

Telecom providers can then focus investments in these specific problem areas, ensuring that resources are directed where they will have the most significant impact, improving overall network efficiency.

Customer-Centric Approach: By leveraging customer feedback as a core part of the analysis, the solution aligns network upgrades with actual user concerns.

This approach demonstrates a commitment to listening to customer needs, building trust and enhancing the brand's reputation for reliability and service quality.

Efficient Resource Allocation: Telecom operators often face budget constraints for infrastructure upgrades. This solution enables them to allocate resources strategically, reducing unnecessary spending on regions that already perform well and concentrating on underperforming areas.

Improved Industry Standing: By maintaining high call quality consistently, operators can outperform competitors and position themselves as leaders in service quality.

Meeting or exceeding benchmarks set by regulatory bodies such as TRAI enhances the company’s credibility and improves its market reputation.

Regulatory Compliance: Addressing call quality issues proactively ensures compliance with telecom regulations and quality standards.

This minimizes the risk of fines, penalties, or reputational damage while fostering a positive relationship with regulatory authorities.

Recommendations

1. Prioritize Low-Rating Regions:

Focus efforts on regions with the highest density of low ratings, such as North India and parts of the Northeast, to address issues related to poor signal strength, high call drop rates, and network congestion.

Strengthen infrastructure in these regions by upgrading to 4G/5G networks, installing additional cell towers, and using advanced technologies like VoLTE (Voice over LTE) and VoWiFi (Voice over Wi-Fi).

2. Real-Time Monitoring:

Deploy AI/ML-based systems that can monitor network performance in real time, identifying potential issues like congestion or outages before they affect users.

Automatically adjust network parameters (e.g., bandwidth allocation or traffic rerouting) to maintain high service quality.

3. Incentivize Participation – Encourage Feedback:

Encourage users to share feedback on call quality through apps like MyCall by offering incentives, such as:

- Free data packs or discounts on monthly bills for frequent feedback.

- Entry into rewards programs or contests for detailed feedback.

- Gamify the process by introducing badges or leaderboards for users who consistently provide feedback.

4. Continuously Validate and Update Models:

- Regularly test and update the predictive models to incorporate new data and account for changing user behaviors or technological advancements.

Limitations

Dynamic Network Conditions: External factors such as weather, sudden network outages, or unexpected traffic spikes may disrupt service in ways the model cannot predict.

Enhancement: Incorporate real-time data streams and anomaly detection systems that can adapt to sudden changes in network conditions.

Scalability Challenges: Real-time prediction and monitoring for millions of users across a large telecom network may require significant computational resources and infrastructure.

Enhancement: Use distributed computing and cloud-based architectures to scale predictions and monitoring systems efficiently.

Privacy and Regulatory Compliance: Collecting and analyzing user data, including geolocation and call details, requires strict adherence to privacy laws like GDPR, CCPA, or local regulations in India.

Enhancement: Use anonymization techniques and secure data storage systems to comply with privacy regulations and gain user trust.

Model Limitations - Lack of Contextual Understanding:

The model relies heavily on numerical and categorical data but doesn’t account for unstructured data like customer comments or sentiment, which may provide deeper insights into dissatisfaction.

Enhancement: Integrate natural language processing (NLP) techniques to analyze unstructured customer feedback for a more comprehensive view of issues.

Black-Box Nature of Advanced Models:

Some machine learning models, like XGBoost or neural networks, can be difficult to interpret, making it challenging to explain decisions to stakeholders or regulators.

Enhancement: Use explainable AI (XAI) techniques to make model predictions more transparent and understandable.

External and Operational Limitations - Adoption of Feedback Mechanisms:

Many users are unaware of feedback apps like MyCall, leading to underrepresentation in certain regions or demographics.

Enhancement: Increase awareness through marketing campaigns, partnerships with telecom providers, or in-app incentives.

Competitor and Environmental Variables:

Competitor strategies, changes in consumer preferences, or external economic factors may influence results in ways the model does not account for.

Enhancement: Continuously monitor competitor data and external trends, and adjust predictions to align with market dynamics.

Closing Reflections

Insights Gained:

Importance of Data Quality: The success of the model was closely tied to the quality and representativeness of the data. A well-curated dataset, with diverse feedback from various regions and user groups, provided a more accurate understanding of network performance, leading to more reliable predictions.

Balancing Performance and Overfitting: The project emphasized the need to balance model performance with the risk of overfitting. While achieving high accuracy on training data is important, ensuring the model generalizes well to unseen data was equally crucial. Fine-tuning hyperparameters and utilizing regularization techniques helped address this challenge.

Real-World Application of Predictive Modeling: The project highlighted how predictive models can directly address real-world challenges, such as improving telecom network performance and customer experience. Aligning model outputs with actionable business insights was key to achieving tangible results.

What Could Be Done Differently:

Use of NLP (Natural Language Processing): Incorporating NLP techniques into the analysis of customer feedback (such as text data from surveys or complaints) provided deeper insights into network issues. Text classification and sentiment analysis helped identify common customer concerns and complaints related to call quality, which were crucial for improving predictions and recommendations.

Expand Data Collection Efforts: Greater emphasis on obtaining feedback from underrepresented regions and ensuring a balanced dataset from the outset would improve model performance. Collaborations with telecom providers or community organizations could help increase participation.

Integrate More Real-Time Data: Integrating real-time network performance data, such as bandwidth usage, latency, and traffic volume, would complement customer feedback and provide a more dynamic response to network conditions.

Collaborate with Domain Experts: Closer engagement with telecom network engineers and business stakeholders would lead to a better understanding of real-time network management challenges. This collaboration would help align model predictions with business objectives more effectively.

Monitor and Update the Model Continuously: Establishing a systematic process for continuous model monitoring and retraining would ensure that predictions remain accurate as the network infrastructure evolves and user needs change.

In conclusion, the project provided valuable insights into model development, data preparation, and problem-solving. Future work should focus on comprehensive data collection, the integration of real-time data, and ongoing collaboration with domain experts to improve the overall solution. Furthermore, expanding the use of NLP to process and analyze textual feedback could further enhance model accuracy and customer satisfaction.

APPENDIX - Dataset and Domain - Data Dictionary

Column Name	**Definition**	**Data Type**	**Example**
Operator	This refers to the telecom company providing the service. It helps in identifying which operator's network is being assessed.	Category	“RJio” , “Airtel”
In Out Travelling	Indoor: Calls made or received inside buildings. Outdoor: Calls made or received outside buildings. Travelling: Calls made or received while moving, such as in a vehicle.	Category	“Indoor” , “Outdoor”
Network Type	Indicates the type of mobile network used during the call .	Category	“2G”,”3G”,”4G”
Rating	Customer ratings for call quality, usually on a scale (e.g., 1-5 stars). This provides insights into user satisfaction and perceived call quality.	Numeric	(e.g., 1-5 stars)
Call Drop Category	Satisfactory: Calls rated positively by customers. Call Dropped: Calls that were unexpectedly disconnected. Poor Voice Quality: Calls with unsatisfactory sound quality.	Category	Satisfactory, Call Dropped, Poor Voice Quality
Longitude, Latitude	Geolocation coordinates indicating the exact location where the call was made or received. This helps in mapping call quality issues to specific geographic regions. If Latitude and Longitude is 0,0 or -1,-1 it means user Mobile has disable the functionality to let Android Api to provide that details to My CALL APP	Numeric	19.064262, 72.996913
State Name	The state in which the call occurred. This allows for regional analysis of call quality and identifying areas with significant issues.	Category	“Karnataka”, ” Maharashtra”
Month	Created this column to combine the three months dataset.	Category	“July”, “August”, “September”
Region	Added a 'Region' column to group 'State Names' into North, South, East, West, and North-East for regional trend analysis and improved modeling.	Category	“North”, “South”,” West”, “East”

REFERENCES

Raw data collected from OGD(Open Government Data) Platform India

https://www.data.gov.in/catalog/voice-call-quality-customer-experience

This catalog provides Customers Feedback Captured using TRAI MyCAll App. Customers rate their perceived experience about voice call quality in real time and help TRAI gather customer experience data along with other network data.

July Call Quality

August Call Quality

September Call Quality

Publications – Paper Link

https://indiaai.gov.in/article/indian-voice-call-quality-datasets-to-assess-customer-experience?form=MG0AV3

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2277570

TRAI MYCall App

https://play.google.com/store/apps/details?id=com.trai.mycall&hl=en_IN