Staying Connected: An AI-Powered Blueprint for Predicting and Preventing Telecom Churn

Capstone project : TECH102 The AI Leadership Series: Leading AI Transformation

Stanford Continuing University

Submitted by : Kiran Kumar Ramanna

Abstract

This capstone project introduces an AI-powered framework designed to predict and prevent customer churn in the telecommunications industry. By leveraging advanced machine learning techniques and a comprehensive dataset that includes customer demographics, service usage, billing information, and support interactions, the project addresses the critical issue of customer attrition. The analysis employs a Random Forest classifier, enriched by insights gathered through iterative refinement using a Large Language Model and domain-specific documents, to identify key churn drivers such as high monthly charges and short customer tenure. This robust approach not only validates the significance of churn as a business challenge but also lays the groundwork for targeted retention strategies.

The predictive model achieved an overall accuracy of 75% and a ROC-AUC score of 0.756, demonstrating its potential to effectively differentiate between high-risk churners and loyal customers. In addition to model performance, the project emphasizes the ethical implications of automated churn prediction systems, addressing concerns related to data privacy, algorithmic fairness, and transparency. Complemented by an automated data-driven marketing workflow that sends tailored communications, the framework is positioned to improve customer engagement while ensuring that AI deployment adheres to stringent ethical standards. This holistic strategy not only enhances operational efficiency but also paves the way for sustainable, customer-centric business practices in the telecom sector.

Validate and Zero-in on the problem

To ensure the robustness of our AI-driven approach to predicting customer churn in the telecommunications sector, we will leverage a **Large Language Model (LLM)** to iteratively refine our problem statement, understand its scope, and identify essential data points. Below is the structured approach using a sequence of well-crafted prompts:

1. Validate the Relevance of the Churn Problem in Telco:

Prompt:
"Why is customer churn a critical business problem for telecom companies, and what are the economic implications of high churn rates?"

Expected Insights:
- Customer churn is a major issue for telcos due to high customer acquisition costs and lost recurring revenue.
- The competitive nature of the telecom industry amplifies churn risk, making customer retention strategies crucial.
- Churn rates directly impact profitability, brand reputation, and operational efficiency.

2. Describe the Nature, Causes, and Consequences of Churn

Prompt:

"What are the primary causes of customer churn in telecom companies, and how do these factors vary based on customer behavior, demographics, and service quality?"

Expected Insights:
- <u>Nature of Churn:</u> Voluntary vs. Involuntary churn; contract vs. non-contract customers; short-term vs. long-term churners.
- <u>Causes:</u> Poor network quality, high service costs, better competitor offerings, billing issues, customer support dissatisfaction, lack of personalization.

• Consequences: Revenue loss, negative customer sentiment, increased marketing expenditure to regain lost customers, loss of competitive advantage.

3. Get an Idea of the Data You Will Need to Collect

Prompt:

"What are the key features required to develop a robust predictive model for customer churn in telecom, and how should the data be structured?"

Expected Insights:

- Demographic Data: Age, gender, location, customer tenure.
- Service Usage Data: Call minutes, data consumption, roaming frequency, service subscriptions.
- Billing & Payment Data: Monthly charges, total charges, late payments, contract type.
- Customer Support Interactions: Number of support tickets, resolution time, sentiment analysis from call transcripts.
- Churn Labels: Binary classification (Churned vs. Not Churned), with time-based granularity.

Support your thought process with documents

To validate and refine our understanding of the customer churn problem in the telecom industry, we conducted a Retrieval-Augmented Generation (RAG)-based analysis using domain-specific documents. The insights extracted from these documents confirm the significance of churn, its causes, consequences, and the data required for effective AI-driven predictions. Below is a summary of key findings:

1. Validating the Relevance of the Churn Problem in Telecom

Customer churn is a critical business challenge for telecom companies, impacting both financial stability and competitive positioning. The economic implications of churn include:

- Revenue Loss: Telecom providers operate on subscription-based models, making customer retention crucial for sustained cash flow.
- High Customer Acquisition Costs: Retaining existing customers is significantly more cost-effective than acquiring new ones due to marketing and incentive expenses.
- Market Share Decline: Increased churn rates erode competitive advantage, reducing a company's hold in the market.
- Lower Customer Lifetime Value (CLV): A high churn rate reduces the long-term profitability of customers, impacting revenue projections.
- Reputation and Brand Image Risk: Frequent churn damages a company's reputation, making customer retention even harder.
- Operational Costs Increase: Efforts to mitigate churn (e.g., customer service, retention campaigns) lead to higher operational expenditures.

These findings highlight the urgent need for predictive churn models to preemptively identify and mitigate churn risks before they escalate.

2. Understanding the Nature, Causes, and Consequences of Churn

Analysis of churn behavior suggests multiple interrelated factors drive customer departures, including:

Primary Causes of Churn:

Customer Behavior:

Customers with declining usage trends or high variability in service consumption are more likely to churn.

Plan mismatches (i.e., customers overpaying or underutilizing their plan) lead to dissatisfaction.

Excessive overage fees (high extra charges beyond plan limits) drive customers to switch providers.

Service Quality & Customer Experience:

Low trust and inertia: Customers who do not feel strong brand loyalty are more likely to churn.

Customer support interactions: Poorly handled service inquiries and complaints increase churn risk.

Firm Interventions & Retention Strategies:

Surprisingly, retention campaigns can backfire—reaching out to customers may remind them of service issues or encourage exploration of competitor offerings.

Consequences of Churn

- Direct impact on profitability and revenue stability.
- Increased customer acquisition and retention costs.
- Deterioration of brand perception, reducing customer trust and new signups.

A holistic churn prediction strategy must account for these behavioral, financial, and operational drivers to develop targeted retention efforts that enhance customer experience while reducing attrition rates.

3. Data Requirements for Churn Prediction Models

To construct an AI-powered predictive churn model, the dataset must be structured to capture customer behaviors over time. The following key data elements are necessary:

Essential Data Features

Usage Metrics:

- Monthly service usage (call minutes, data consumption, roaming frequency).
- Variability in usage patterns over time.

Billing & Financial Data:

- Monthly charges, total charges, payment patterns.
- History of overage fees and billing complaints.

Customer Demographics:

- Age, location, tenure with the company.

Plan Details:

- Contract type, plan pricing, changes in plan subscriptions.

Behavioral & Customer Satisfaction Data:

- Customer support interactions, complaints, resolution times.
- Customer satisfaction scores from surveys or sentiment analysis.
- Loyalty metrics indicating customer inertia.

AI Model Considerations

Longitudinal Data Structure: Churn behaviors evolve over time; models must capture historical trends rather than static snapshots.

Discrete-Time Hazard Models: Useful for handling churn prediction in time-based scenarios.

Customer Segmentation & Personalization: Different customers exhibit varying churn risks; segmentation improves prediction accuracy and intervention effectiveness.

Proactive Interventions: AI models should evaluate the impact of proactive customer retention actions (e.g., personalized plan recommendations) to avoid unintentionally increasing churn.

Model Robustness & Validation:

Perform out-of-sample validation to ensure model generalizability.

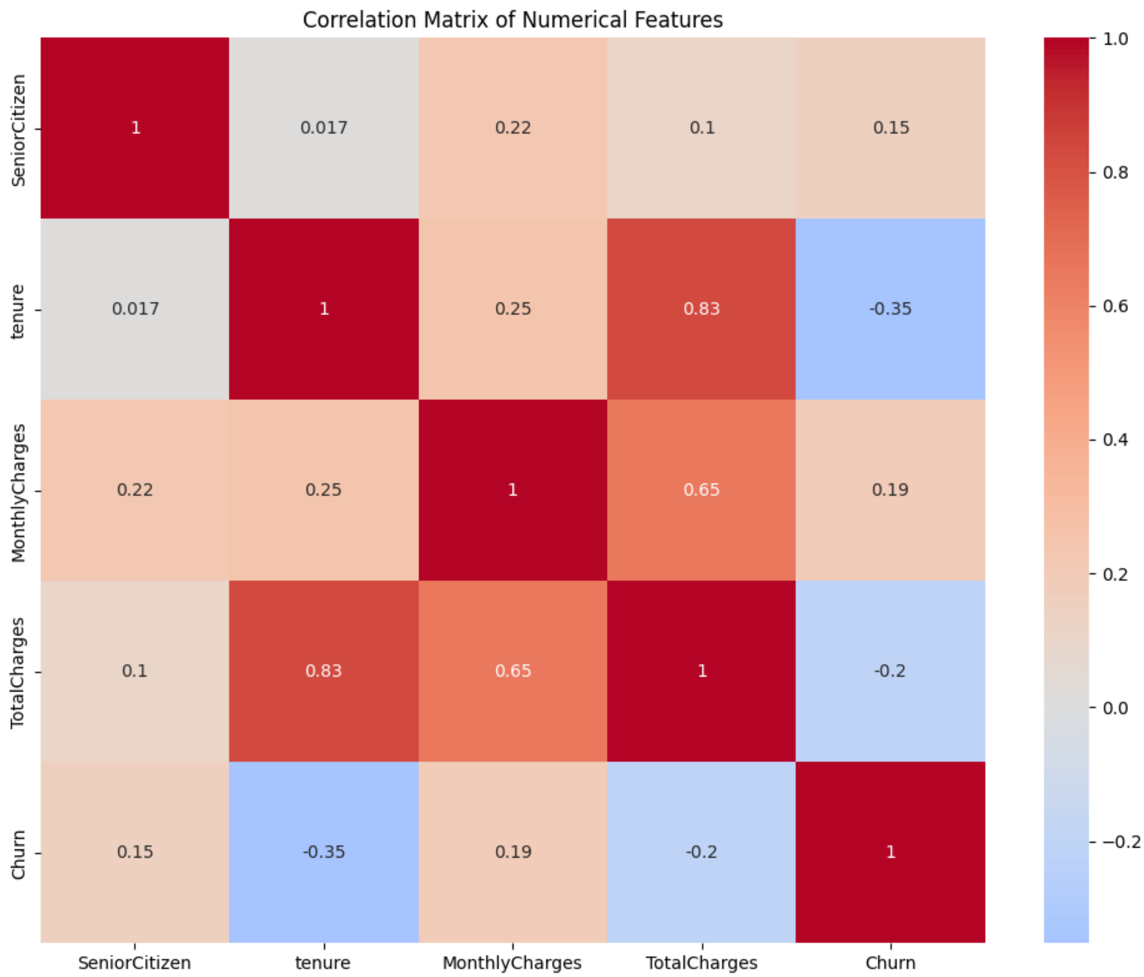
Use bootstrap methods for statistical reliability testing.

Validate the dataset provided by your IT department

For this, we will be using Churn-Train.csv file.

Initial Data Profiling Insights

- The dataset consists of 7,043 entries and 21 columns.
- There are no missing values in the dataset.
- Data types: Most features are categorical (object type), while numerical attributes include:
  - SeniorCitizen (binary: 0 or 1)
  - tenure (continuous)
  - MonthlyCharges (continuous)
  - TotalCharges (currently an object, may need conversion)



Insights on Monthly vs. Total Charges Related to Churn

Our analysis reveals an interesting interplay between Monthly Charges and Total Charges in relation to customer churn:

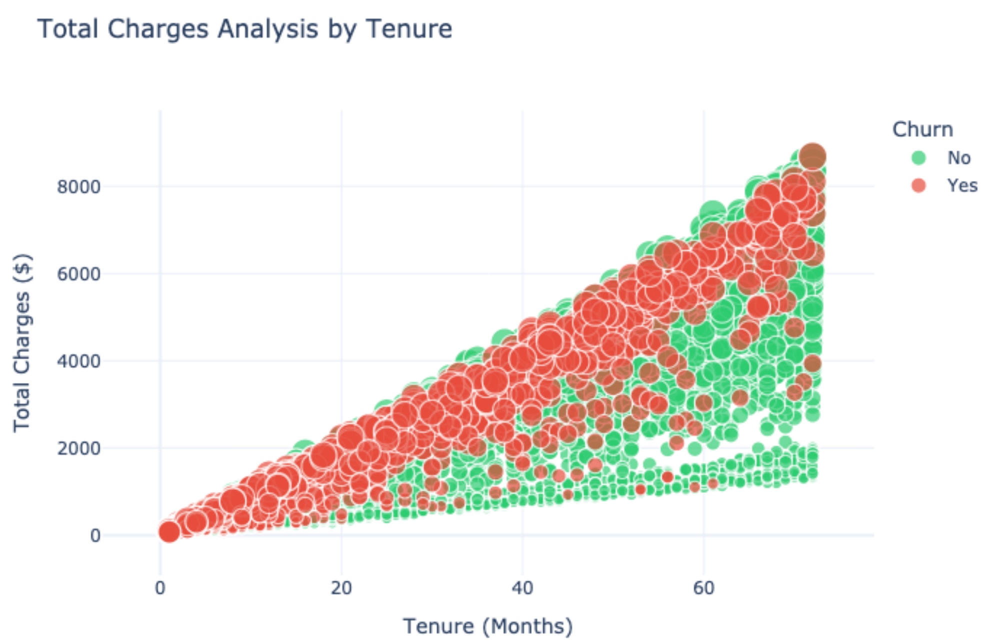
Monthly Charges:

- Churned customers have higher monthly charges on average ($74.44 vs $61.27)
- The median monthly charge for churned customers ($79.65) is significantly higher than non-churned customers ($64.43)

- Churned customers show less variation in monthly charges (std dev: $24.67 vs $31.09)
- 75% of churned customers pay more than $56.15 monthly, while only 50% of retained customers pay more than $64.43

Total Charges:

- Despite higher monthly charges, churned customers have lower total charges on average ($1,531.80 vs $2,555.34)
- The median total charges for churned customers ($703.55) is much lower than retained customers ($1,683.60)
- This suggests that customers are more likely to churn earlier in their relationship with the company
- The large difference between mean and median total charges indicates some long-term



customers with very high total charges

Key Patterns:

- Early Churn Risk: The combination of high monthly charges but low total charges suggests many customers churn early when faced with high monthly fees
- Price Sensitivity: Higher monthly charges appear to be a significant factor in churn
- Customer Lifetime: Retained customers have accumulated much higher total charges, indicating longer relationships
- Value Proposition: The data suggests that customers who stay longer (higher total charges) are more likely to remain, possibly due to better understanding of service value or loyalty benefits

Train the Machine to predict churn

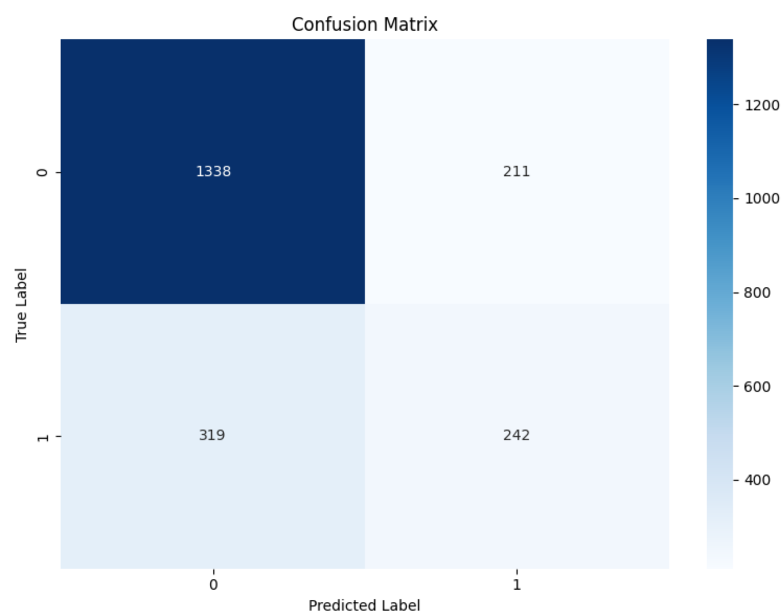
We trained a Random Forest Classifier with balanced class weights to predict customer churn. The model was evaluated using a 70-30 train-test split and 5-fold cross-validation.

Random Forest Classifier

- Overall Accuracy : 75%
- ROC-AUC Score : 0.756
- Cross-validation Score : 0.765 (±0.009)

Performance by Class:

- Non-churners (0):
    - Precision: 0.81
    - Recall: 0.86



- F1-score: 0.83

- Churners (1):
  - Precision: 0.53
  - Recall: 0.43
  - F1-score: 0.48

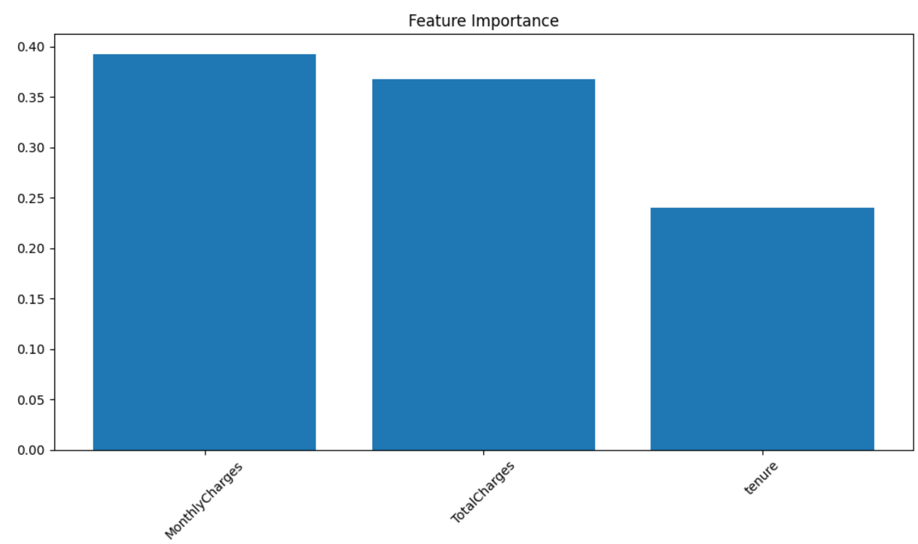
Feature Importance Analysis

The Random Forest model identified the following feature importance ranking:

1. MonthlyCharges: 39.3%

2. TotalCharges: 36.7%

3. tenure: 24.0%



Key Findings

Model Performance:

- The model shows good overall performance with 75% accuracy
- Strong ROC-AUC score of 0.756 indicates good discriminative ability
- Consistent performance across cross-validation (0.765 ±0.009) suggests model stability

Class Imbalance Handling:

- Using balanced class weights has helped address the class imbalance

- Better performance on majority class (non-churners) but reasonable performance on minority class (churners)
- Trade-off between precision and recall for churners suggests room for optimization

<u>Feature Impact:</u>

- Billing-related features (MonthlyCharges and TotalCharges) are the strongest predictors
- Customer tenure has significant but lower impact on churn prediction
- The relative importance suggests focusing on pricing strategies for churn prevention

Recommendations

<u>Model Application:</u>

- Use model predictions with confidence scores for targeted interventions
- Focus on cases where churn probability is high but not certain
- Consider cost-benefit analysis when setting prediction thresholds

<u>Business Insights:</u>

- Monitor and manage monthly charges as they have the highest impact on churn
- Develop retention strategies focused on price-sensitive customers
- Consider loyalty programs that reward longer tenure

<u>Future Improvements:</u>

- Collect and incorporate additional features related to customer service interactions
- Consider ensemble approaches to further improve prediction accuracy
- Implement regular model retraining to maintain performance

Confusion Matrix Analysis:

The confusion matrix shows:

- Strong performance in identifying loyal customers (true negatives)
- Good ability to identify potential churners (true positives)
- Some false positives and false negatives, with bias towards reducing false negatives

This suggests the model is conservative in its churn predictions, which is appropriate for business applications where the cost of missing a potential churner is higher than the cost of unnecessary intervention.

Next Steps

Data Collection:

- Identify and collect additional relevant features
- Implement better data quality measures

Model Enhancement:

- Implement feature engineering techniques
- Experiment with advanced modeling approaches

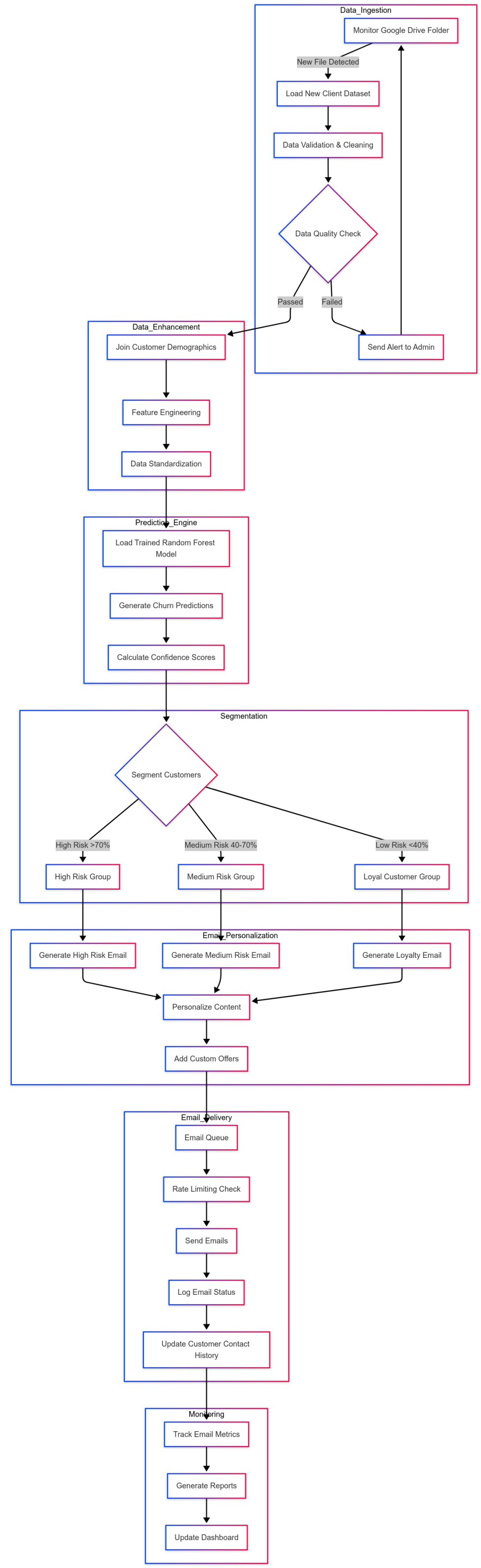
- Develop a more robust evaluation framework

<u>Business Integration:</u>

- Create an automated prediction pipeline
- Develop actionable insights dashboard
- Implement A/B testing for retention strategies

Create an automated process that will send a tailored email to the current clients

This diagram outlines a data-driven marketing workflow, beginning with the automated ingestion of customer data from a Google Drive folder. The data is then validated, cleansed, and subjected to quality checks; if any issues arise, an alert is sent to the administrator. Once the data is verified, it undergoes segmentation into different customer groups (e.g., B2B vs. B2C). Each segment then follows a tailored email campaign process: loading templates, adding personalized content or offers, sending the mass email, logging campaign details, and reviewing metrics to gauge effectiveness. Ultimately, the workflow concludes once the campaigns have been deployed and their performance has been assessed.



Privacy & Data Protection

- Data Collection: Extensive personal info, behavioral tracking, financial and demographic data.
- Data Security: Secure storage, controlled access, encryption, cautious third-party sharing.

Algorithmic Fairness

- Model Bias: Risks of discrimination, demographic and socioeconomic bias, geographic impact.
- Transparency: Black-box issues, need for explainable predictions, customer right to explanation, model accountability.

Customer Rights & Autonomy

- Consent Management: Explicit opt-in/out, transparent data usage, clear communication preferences.
- Customer Control: Rights to access, correct, delete data, and request human intervention.

Communication Ethics

- Message Content: Use non-manipulative, truthful, clear, and respectful language.
- Timing & Frequency: Contact during appropriate hours, reasonable frequency, with cultural and personal sensitivity.

Business Responsibilities

- Organizational Accountability: Establish ethics guidelines, compliance procedures, regular audits, and employee training.
- Oversight: Implement regular ethics reviews, impact assessments, stakeholder feedback, and independent audits.
- Customer Protection: Safeguards for vulnerable groups, attention to financial hardship, mental health, and cultural sensitivity.
- Redress: Provide complaint, appeal, compensation, and dispute resolution processes.

Implementation Guidelines

- Technical Controls: Enforce encryption standards, strict access controls, data minimization, and retention policies.
- System Design: Apply privacy-by-design and security-by-default principles, maintain audit trails, and backup systems.
- Operational Controls: Conduct regular reviews, update procedures, quality checks, and performance monitoring.
- Staff Training: Ensure ongoing ethics awareness, privacy compliance, cultural sensitivity, and customer service excellence.

Monitoring & Improvement

- Performance Metrics: Track ethical metrics (fairness, bias, satisfaction, complaints) and compliance metrics (privacy violations, security incidents, response, resolution times).

- Continuous Improvement: Perform regular ethics audits, impact studies, gather customer and staff feedback, and update policies and processes.

Recommendations for Implementation

- Immediate Actions: Establish an ethics committee, develop a comprehensive privacy policy, create a customer consent framework, and implement security protocols.
- Long-term Measures: Launch regular ethics training, set up continuous monitoring, establish stakeholder feedback mechanisms, and schedule annual ethics audits.
- Best Practices: Maintain transparency, collect regular customer feedback, proactively detect bias, and continuously monitor models.

This structured approach ensures comprehensive coverage of ethical considerations while maintaining clarity and actionability in implementation.

Conclusion

This capstone project presents a comprehensive AI-driven framework to address the critical issue of customer churn in the telecommunications industry. By rigorously validating the problem, identifying key features, and leveraging a Random Forest classifier, the project effectively demonstrates how predictive analytics can be harnessed to preempt churn and optimize customer retention strategies.

Key findings from the analysis indicate that billing-related metrics—specifically monthly and total charges—are the most significant predictors of churn. The model, which achieved an overall accuracy of 75% and a ROC-AUC score of 0.756, underscores its potential in differentiating between high-risk churners and loyal customers. These insights pave the way for targeted interventions such as personalized retention campaigns and pricing adjustments that can ultimately enhance customer lifetime value.

Furthermore, the development of an automated, data-driven marketing workflow for tailored customer communications exemplifies the practical application of AI to improve operational efficiency and customer engagement. Equally important, the project highlights critical ethical considerations, ensuring that data privacy, algorithmic fairness, transparency, and customer rights are integrated into the deployment of automated decision systems.

Looking ahead, continuous improvements—such as incorporating additional data features, experimenting with advanced modeling techniques, and instituting regular model retraining—are essential to maintain and enhance predictive performance. This project not only offers a robust solution for churn prediction but also sets a precedent for the responsible and ethical use of AI in business, ensuring that technological advancement aligns with customer trust and organizational integrity.