

FAITHKAMANDE / Machine-Learning-project

Q

Type / to search

+

<>

CodeIssuesPull requestsActionsProjectsWikiSecurityInsightsSettings

Machine-Learning-project

Public

1 Branch

0 Tags

Q

Go to file

t

Go to file

+

Add file **About**Code

...

FAITHKAMANDE

Added Notebook PDF

5a7039c · now

🕒 28 Commits

	.gitignore	Clean repo: removed ch...	yesterday
	Phase 3 project.ip...	Updated notebook	2 hours ago
	Phase 3 project.pdf	Added Notebook PDF	now
	Presentation.pdf	Added Presentation slid...	1 hour ago
	README.md	Update README.md	1 hour ago
	bigml.csv	Added csv and notebook	3 days ago

No description, website, or topics provided.

- README
- Activity
- 0 stars
- 0 watching
- 0 forks

Releases

No releases published  
[Create a new release](#)

Packages

No packages published  
[Publish your first package](#)

Languages

README

Machine-Learning-project

## OVERVIEW

---

This project focuses on predicting which customers are likely to stop using a telecom company's services (customer churn). The dataset consists of 3,333 customer records, containing details such as call usage, subscription plans, and customer service interactions. The main goal is to build a machine learning classification model that can proactively identify customers at risk of churn. Early identification enables the company to take targeted actions — like offering personalized services or special deals — to retain valuable customers. By analyzing usage behavior, plan details, and customer service interactions, we aim to discover key patterns and insights that contribute to customer churn, thereby improving customer satisfaction and reducing churn rates.

## BUSINESS QUESTION

---

"Which customer behaviors and service factors most influence churn in a telecommunications company, and how accurately can we predict churn to drive proactive customer retention strategies?"

## OBJECTIVE

---

- Develop a binary classification model to predict customer churn.
- Identify and rank the most significant features influencing churn.
- Evaluate model performance using appropriate classification metrics (e.g., Recall, F1-score, ROC-AUC).
- Provide actionable insights into customer behaviors that lead to churn.
- Recommend targeted retention strategies to reduce churn and increase customer lifetime value.

- Enable the telecom company to proactively intervene with at-risk customers, improving overall customer satisfaction.

## DATA

---

- [Churn in Telecom's dataset](#)
- 3,333 telecom customers
- 21 columns including: Call usage (day, evening, night, international)  
Subscription plans (international plan, voicemail plan) Customer service interactions Account length and call counts

## CLEANING DATA

---

- Removed irrelevant columns (phone number).
- Checked for missing values — none detected.
- Handled categorical variables via binary encoding (yes/no to 1/0).
- Log-transformed skewed variables (customer service calls, voicemail messages).

## NORMALIZATION AND STANDARDIZATION

---

- Analyzed feature distributions to decide on scaling or transformations.
- Standardization considered where necessary for model optimization.

## ANALYSIS AND FEATURE SELECTION

---

- Visualized churn distribution.

- Analyzed relationships using boxplots and chi-square tests.
- Identified key churn indicators (e.g., international plan, customer service calls).
- Feature importance analyzed using Random Forest and Logistic Regression.

## TRAIN - TEST SPLIT

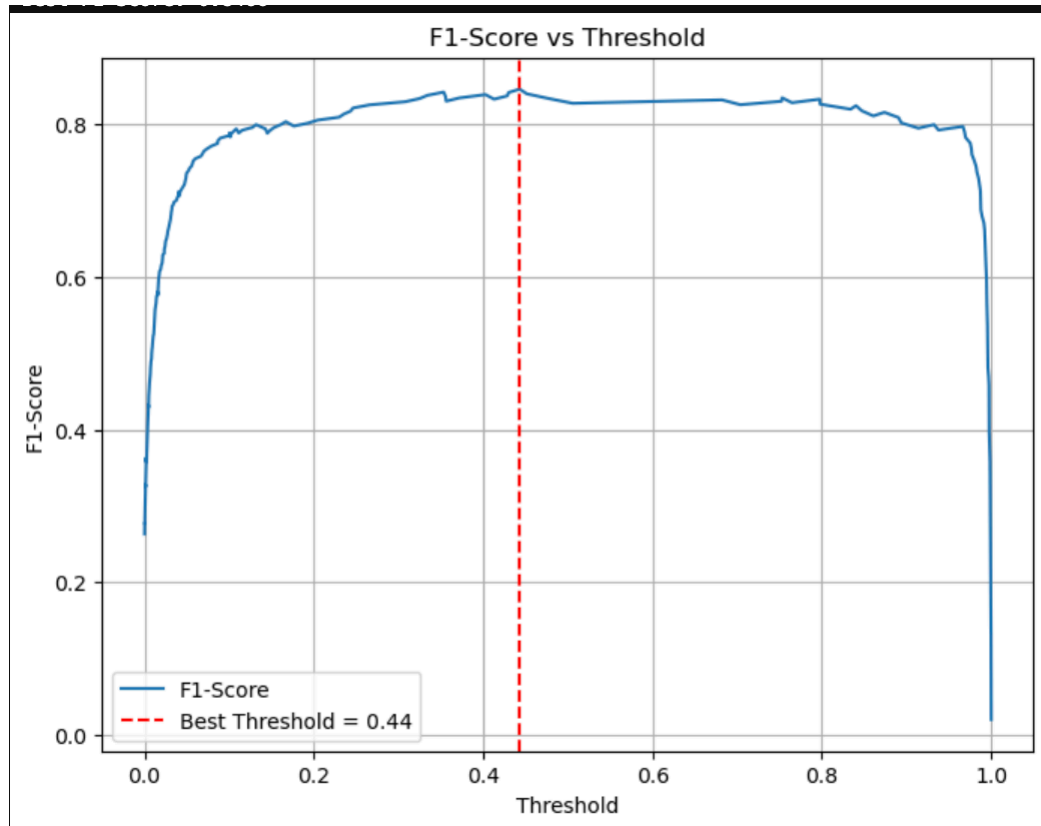
---

- Split data into 80% training and 20% testing sets.

## MODELLING

---

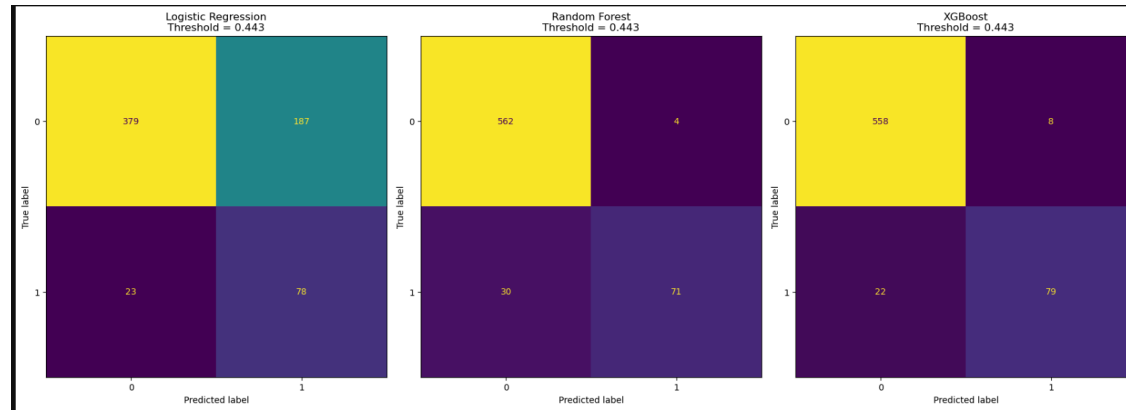
- Trained multiple models: Logistic Regression (baseline, interpretable) Random Forest Classifier (robust and non-linear) XGBoost Classifier (advanced boosting method)
- Handled class imbalance using: `class_weight='balanced'` in models. Threshold tuning (set at 0.4430) to improve recall.



## EVALUATION

- Evaluated models using: Accuracy (overall correctness) Recall (churner detection rate) Precision (correct churn predictions) F1-Score (balance between precision and recall) ROC-AUC (discrimination ability)

- Achieved: 78% Recall on churn class 90% Precision 95% Accuracy 0.81 AUC [



## OBSERVATIONS AND RECOMMENDATIONS

### OBSERVATIONS.

#### International Plan:

- Customers with an international plan are significantly more likely to churn.
- This suggests potential dissatisfaction with the pricing or quality of international services.

#### Voice Mail Plan:

- Having a voice mail plan reduces churn risk.
- Voice mail users are more engaged and possibly more satisfied customers.

#### Customer Service Calls:

- Frequent customer service calls are associated with a higher churn risk.

- This likely reflects dissatisfaction with service or unresolved issues.

### **Total Day Minutes:**

- Heavy daytime usage correlates with increased churn risk.
- High usage customers may have greater service expectations and are more sensitive to quality or pricing issues.

### **Total Evening Minutes:**

- Higher evening call minutes are also associated with a moderate increase in churn risk.

### **International Calls**

- customers making more international calls are less likely to churn, suggesting deeper engagement with the service.

### **Total International Minutes:**

- However, higher total international call duration slightly increases churn risk, possibly due to higher costs or dissatisfaction with long call sessions.

### **Three machine learning models were trained and evaluated to predict customer churn:**

- GBoost outperformed other models, achieving the highest accuracy (95.2%) and AUC (0.9246), indicating excellent classification ability.
- Threshold tuning to 0.443 improved Recall across models, ensuring more churners were correctly identified without major losses in Precision.

- Logistic Regression struggled with lower precision and F1-Score, while Random Forest and XGBoost maintained a strong balance between Precision and Recall.
- XGBoost is selected as the final model for its superior performance, offering the best trade-off to support proactive customer retention strategies.

## RECOMMENDATIONS

---

### Investigate and Optimize International Plans:

- Review pricing and service quality for international plans to improve customer satisfaction and reduce churn.

### Enhance Customer Service:

- Focus on improving customer service quality and responsiveness.
- Prioritize customers with frequent service interactions for proactive retention efforts.

### Promote Voice Mail Plan Adoption:

- Encourage customers to adopt voicemail plans through promotions or bundled services, as this reduces churn risk.

### Loyalty Programs for Heavy Users:

- Introduce loyalty rewards, special discounts, or premium service tiers targeting high day-time and evening call users.
- Address their high service expectations to prevent churn.



**Monitor International Usage Patterns:**

- Identify customers with high international call durations and offer special packages or discounts to mitigate potential churn risks.

**Deploy the XGBoost model for Churn Prediction:**

- Deploy the XGBoost model in production to proactively flag high-risk churn customers, given its strong performance in Precision, Recall, and AUC.