

sarwas530 /
project-phase3

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

In



main ▾

project-phase3 / README.md



sarwas530 Update README.md

584d0f7 · 1 hour ago



133 lines (35 loc) · 3.46 KB

Preview

Code

Blame

Raw



Telecommunication analysis

Bussines understanding

The telecommunication market is growing rapidly, but with this expansion comes increased competition. Telecom companies are facing significant challenges due to revenue loss as customers leave for better offers and services from competing companies. This phenomenon, known as "churn," is becoming a critical issue for businesses, and they are now focusing on understanding customer behavior to address the root causes of churn.

Objective

- 1.How to reduce churn and increase retention by predicting high-risk customers and what makes them churn?
- 2.How can i optimize model performance with the right balance between precision and recall?
- 3.Building models that Minimize false positives.
- 4.How class imbalance can be handled to ensure accurate predictions for churners?
- 5.Develop targeted campaigns based on the model's predictions.

Data used

The data used is from SyriaTel Customer Churn.

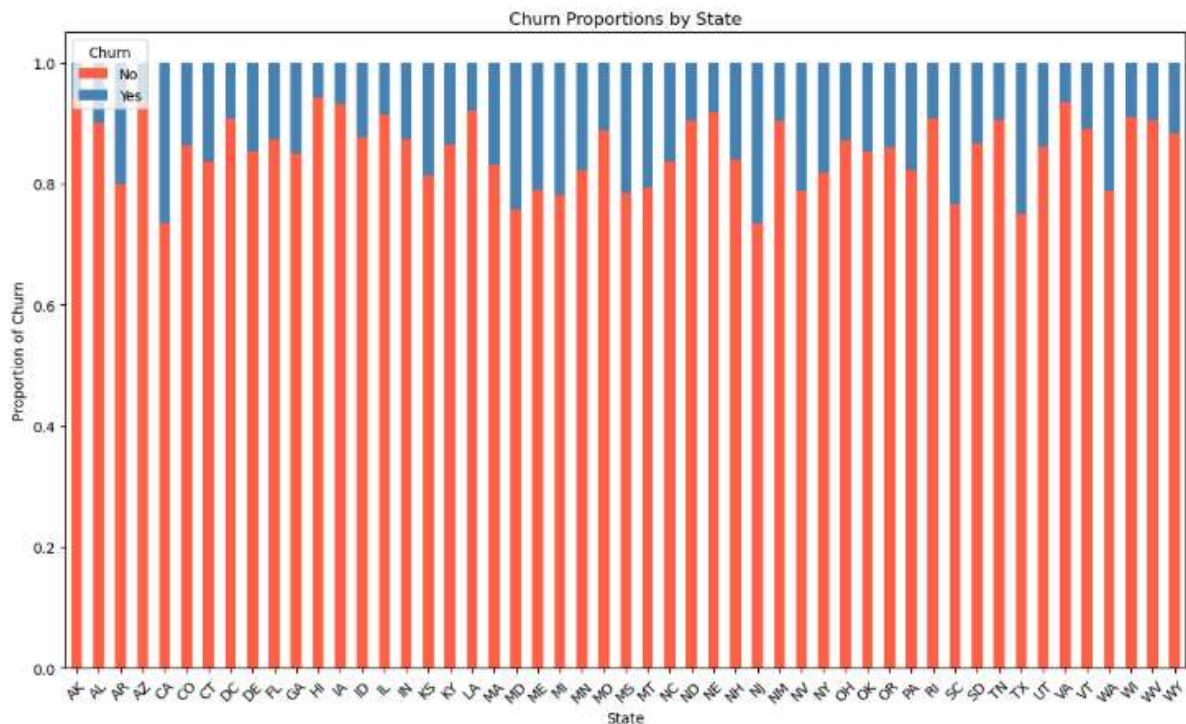
Key resource

pandas,
numpy,seaborn,matplotlib.pyplot,warnings,sklearn,sklearn.pipeline,sklearn.model_selection,
sklearn.linear_model,sklearn.metrics,sklearn.svm,sklearn.neighbors,sklearn.ensemble,catboost,

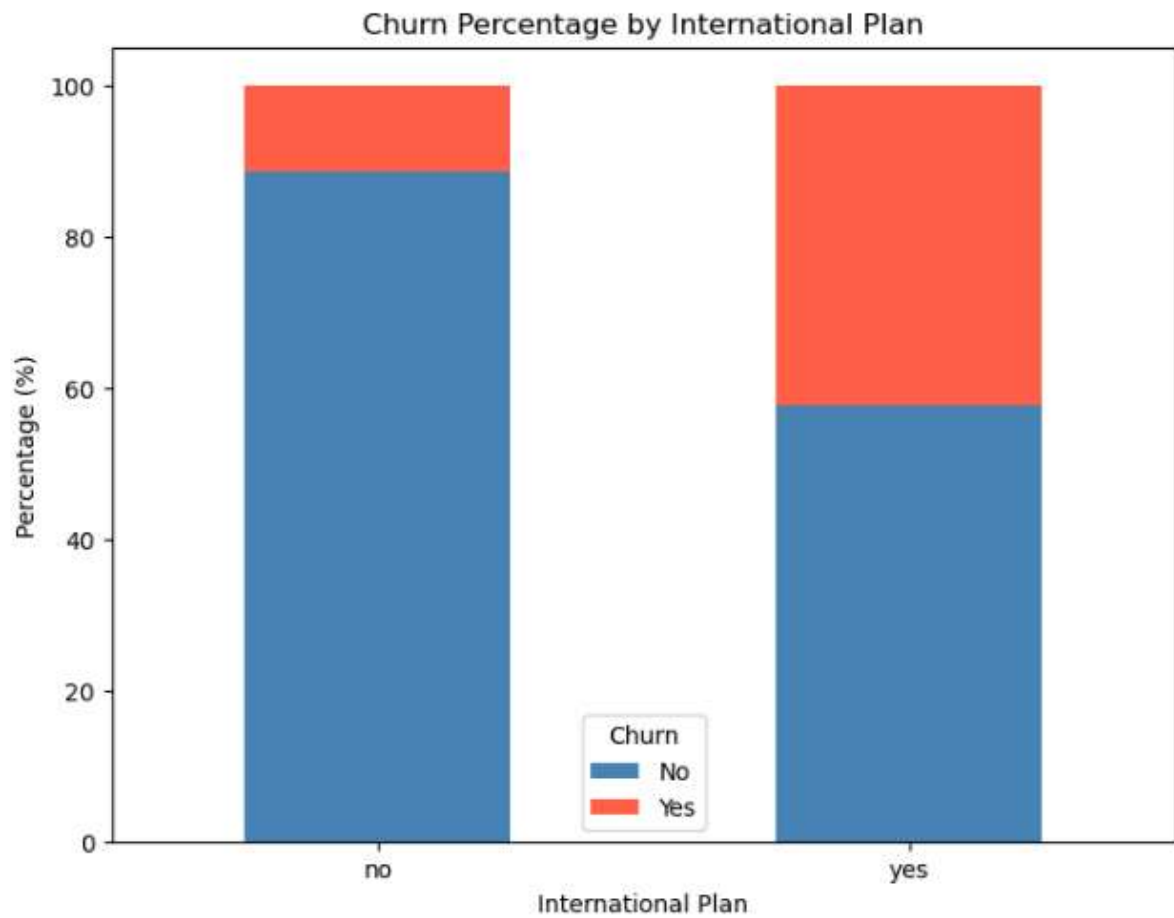
Models

The models used are: Logistic regression,SVM model,KNN model,Random forest and CatBoost.

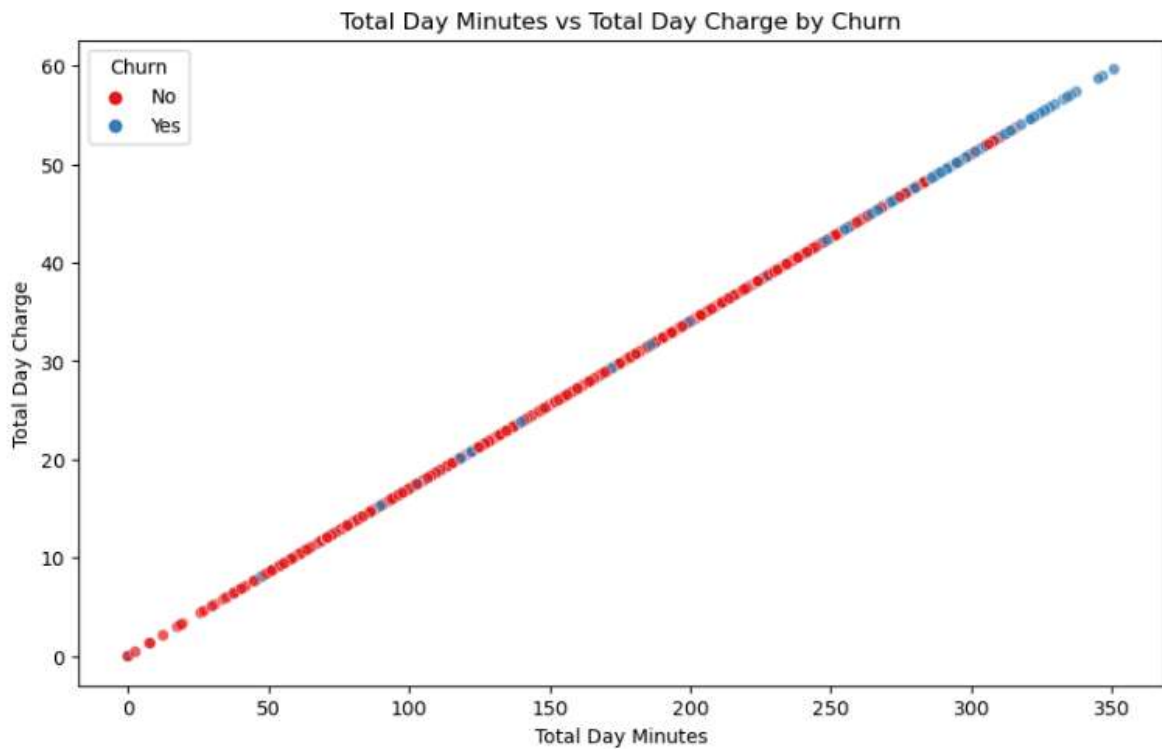
The graph shows the churn distribution by state in terms of percentage.



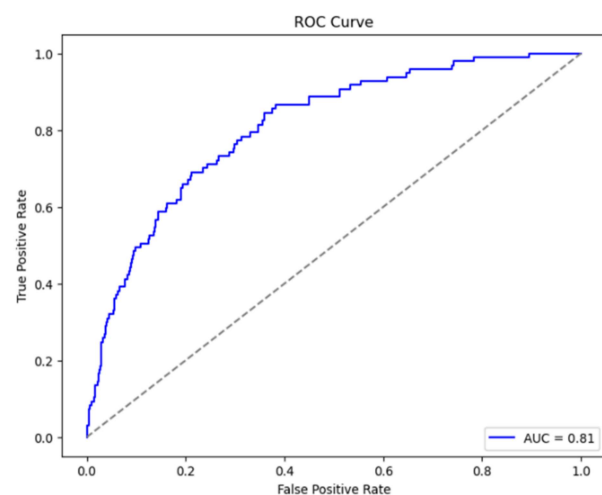
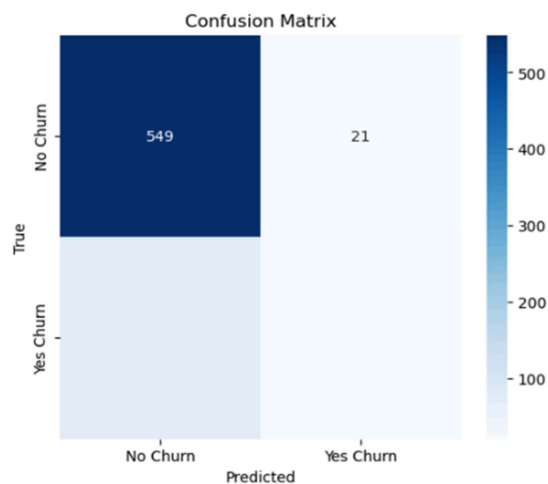
This plot explores the relationship between international calls and churn.



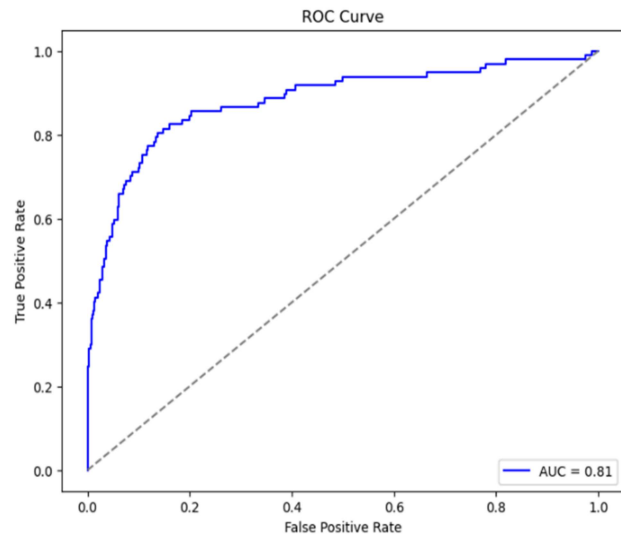
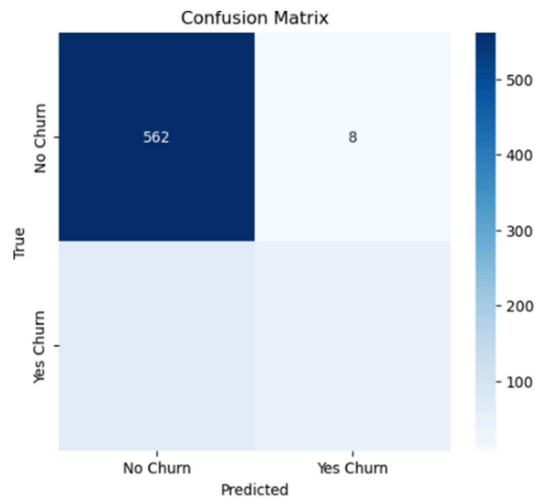
Total day minutes feature against day charge feature, colored by churn



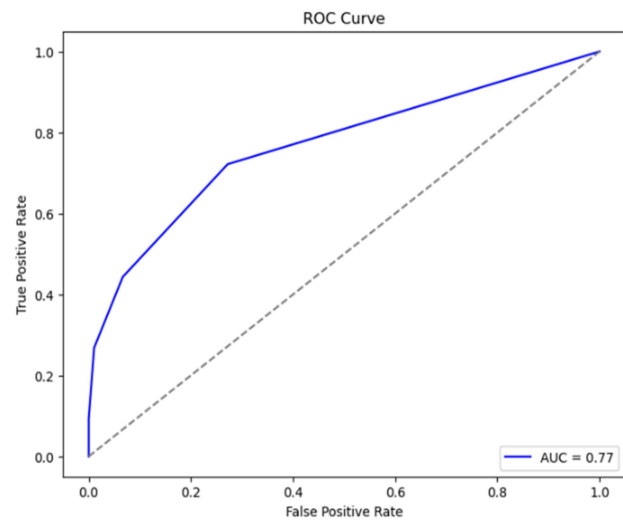
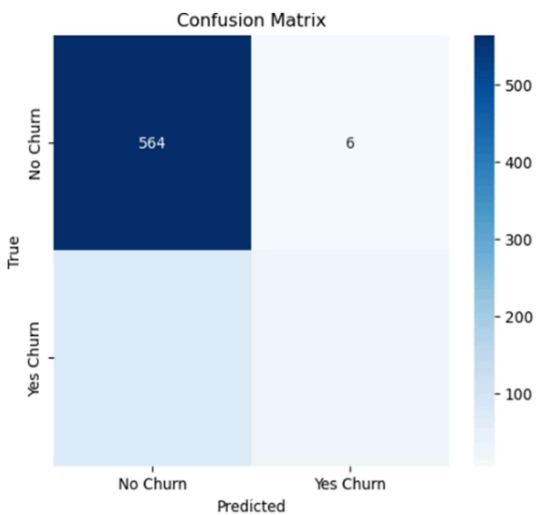
Logistic regression



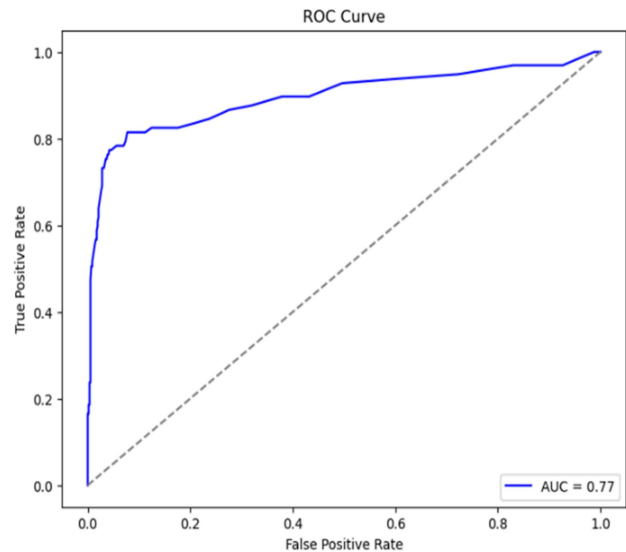
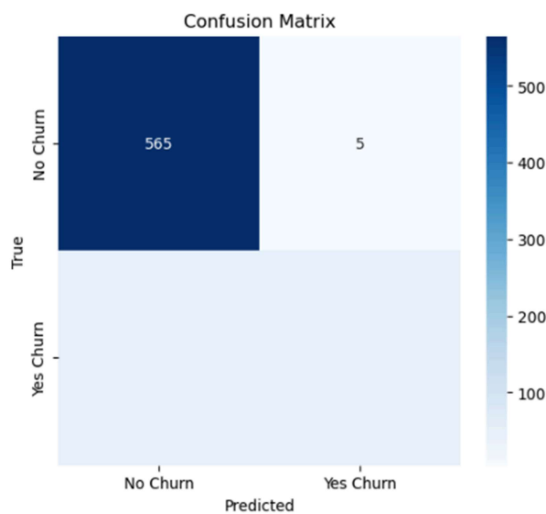
SVM Model



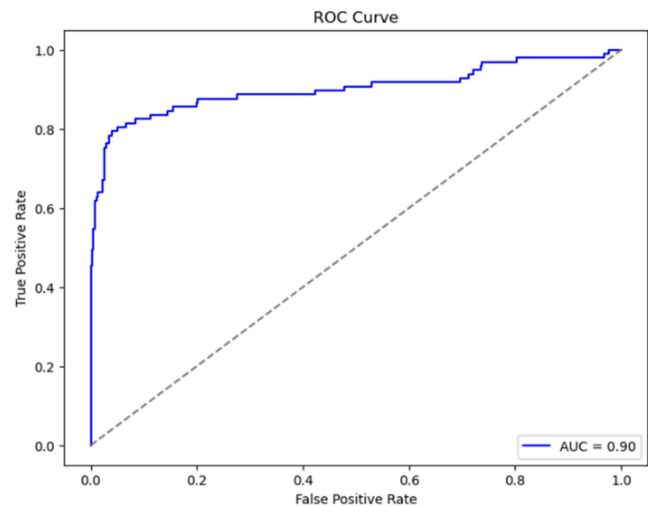
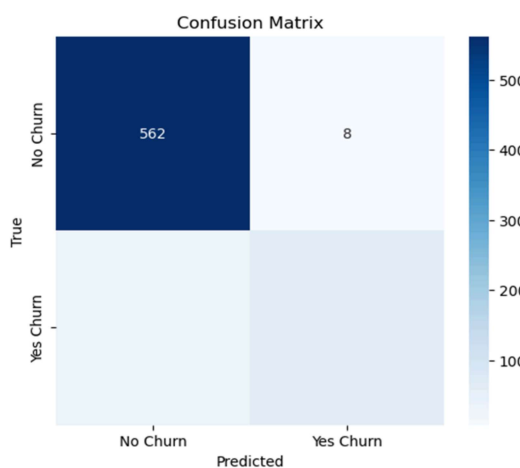
KNN Model



Random Forest



CatBoost



Conclusion

1. Logistic regression, SVM model, KNN model, Random forest and CatBoost have been tried to predict churn. Random forest and Catboost are considered to be the most suitable models based on the accuracy: Random forest (the accuracy is 92% and the recall of no churn is 99% while for yes churn is 59%). CatBoost (the accuracy is 94% and the recall for the 0 is 99% while for the 1 is 64%). They also handled class imbalance.

2.Feature importance was done for this 2 models(Random forest and CatBoost)and the features with the highest importance for Random Forest was customer service calls,total day charge and total day minutes. For the CatBoost was customer service calls,total day minutes,international plan and total day charge.

3.While interpreting results of random forest and catboost, as it has been mentioned that some features are more important for the churn modelling. Which means that such features should be considered while dealing with at risk customer, by offering discounts or personalizing plan. Focusing on such features can be useful on devising the plan to reduce churn.