Telecom Churn Prediction: A Reviews and Machine Learning Based Approach

Anirudh Bhattacharya
Department of Computer Engineering
K J Somaiya Institute of Technology
Mumbai, India
anirudh.b@somaiya.edu

Saumya Shah
Department of Computer Engineering
K J Somaiya Institute of Technology
Mumbai, India
saumya.ss@somaiya.edu

Mink Shethia
Department of Computer Engineering
K J Somaiya Institute of Technology
Mumbai, India
mink.shethia@somaiya.edu

Dr. Madhura Phadke
Department of Computer Engineering
K J Somaiya Institute of Technology
Mumbai, India
madhura.phadke@somaiya.edu

***Abstract*— Churn matters to both the firm and potential consumers in a competitive sector. Customer preference, geography, work, and other variables might affect turnover in the telecom sector. Thus, telecom customer turnover is intensively investigated. However, past churn rates alone cannot predict future churn. Other considerations must be addressed. A fix was suggested where past data may be used to estimate turnover for a quarter or year. For prediction value, utilize a Machine Learning algorithm. Customer input may change this value and affect turnover. Thus, forecasts and feedback will provide more accurate value. The system explained in this paper is the implementation of the proposed fix, and has more accuracy than previous systems.**

Keywords—Churn, Machine Learning, Django, React.JS, Web Application

I. Introduction

Customer churn is a significant problem for telecommunications companies, as customers frequently switch to competitors or cease using their services. To address this issue, there is a growing interest in utilizing machine learning techniques to predict customer churn and implement proactive retention strategies. Here, a novel method has been developed for predicting telecom customer churn based on customer feedback data. Customer feedback has been collected via surveys and this information has been incorporated into the system for churn prediction. This strategy aims to improve the precision of churn prediction and enable telecom companies to take targeted measures to retain customers who are at risk of churning. Using real-world data, we show that our approach outperforms existing methods in terms of prediction accuracy. Our findings indicate that incorporating customer feedback into machine learning models can significantly improve the accuracy of telecom churn prediction and assist businesses and their customers.

II. Related Work

In the past few years, telecom companies have been increasingly focused on addressing the challenge of customer churn, which refers to the situation where customers switch to a competitor or stop using their services altogether. Machine learning techniques were used to predict the churn rate and take proactive measures to retain them. A novel approach was proposed in [1] for telecom churn prediction that leverages customer feedback data [16]. Feedback was collected from customers through surveys [2] and this data was incorporated into the system, to be combined with the result from the machine learning model for the final prediction of customer churn. This approach aimed to improve the performance of the system and enable telecom companies to take targeted actions to retain customers. This approach was evaluated using real-world data and showed that it outperformed existing methods [3-12]. The results of this study suggest that incorporating customer feedback into machine learning models can significantly enhance the effectiveness of telecom churn prediction and help companies better manage customer churn.

III. Summary of Proposed System

The proposed system [1] uses Adaptive Boosting Classifier to fit a model on the Telco Customer Churn dataset by IBM, which contains 19 important features. The system suggests considering customer reviews as an additional factor to improve churn prediction accuracy. A web scraper can collect reviews from websites like TrustPilot, which can be analyzed using an NLP tool to determine if the review is positive or negative. The output of the NLP analysis can be used to adjust the predicted churn rate accordingly. The resultant prediction of churn for the can be obtained by combining the predicted rate with the adjusted value. A database schema is proposed to store the model output, reviews, and adjusted values.

IV. Technologies Used

The entire system has been developed in Python 3 and React.JS to build the server side and client side of the application respectively [13-15]. scikit-learn, an open-source machine learning library was used to pre-process the dataset and generate the machine learning model, which was then extracted using joblib. This model (extracted as a .sav file) was saved as a static file in the web application framework, namely Django.

Additionally, the web scraper in use is Scrapy, which has been integrated into Django in order to access the database-a

SQLite database, Django's default-using the ORM. The TextBlob library is used to find the emotion of a sentence and has been imported to a file within the system, and finally, the Django REST framework connects the React.JS frontend and Django backend.

V. Methodology

The system, developed using the Django framework, has three main applications within it. Firstly, it contains the core application, which contains the views and the models, which facilitates the ORM. Furthermore, this application also contains the function which runs the machine learning model (a static file), which saves its predictions to the "ModelPred" column in the table "churn." This table has been shown in Fig. 1 and the models have been shown in Fig. 3.

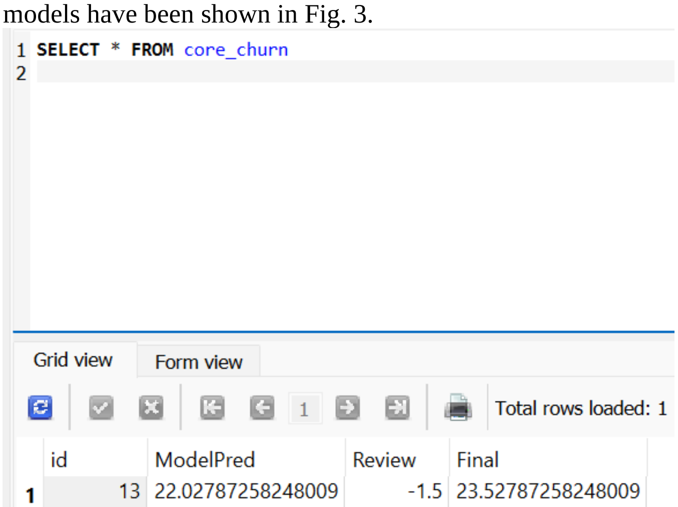


Fig. 1: Table "churn" of the database

The second application contains the web crawler, developed in Scrapy [17], which takes data from a trusted website, in this case, TrustPilot, and saves it to the table "reviews" in the database. The natural language processor, TextBlob [18], is also used in a function within this application. TextBlob then reads the reviews scraped and separates them as positive and negative reviews. The count of the positive and negative reviews is taken and their difference is then divided by ten. This value, saved in the column "Alt" in table "reviews" is mapped to the "Review" column in table "churn." This table has been shown in Fig. 2.

Fig. 2: Table "reviews" of the database

The third application is the Frontend, which contains the frontend, developed in React.JS and connected using the Django REST Framework, which creates APIs of the content to be displayed. The database is linked to the frontend and the values are visible on the screen.

```
core > models.py > Churn
from unittest.util import _MAX_LENGTH
from django.db import models

# Create your models here.
#Database coloumns defined
class Churn(models.Model):
    ModelPred = models.FloatField(max_length=100)
    Review = models.FloatField(max_length=100)
    Final = models.FloatField(max_length=100)

class Reviews(models.Model):
    PlaintextReviews = models.TextField(max_length=500)
    PosFeedback = models.FloatField(max_length=50)
    NegFeedback = models.FloatField(max_length=50)
    Alt = models.FloatField(max_length=100)
```

Fig. 3: The models in the system

In order to run the machine learning model, the scraper and the ORM, a custom command has been created, called "runmodel". Additionally, this command executes the mathematical formula to predict the final churn, which is given in equation (1). A larger number of negative reviews will result in a higher churn prediction and vice versa. We predict the accuracy of this system to beat the recorded accuracy in [1] by 6-7%.

$$ModelPred - Review = Final \tag{1}$$

The entire architecture of the system is shown in Fig. 4, with the tables of the database, the static files, the web scraper, and the input data.

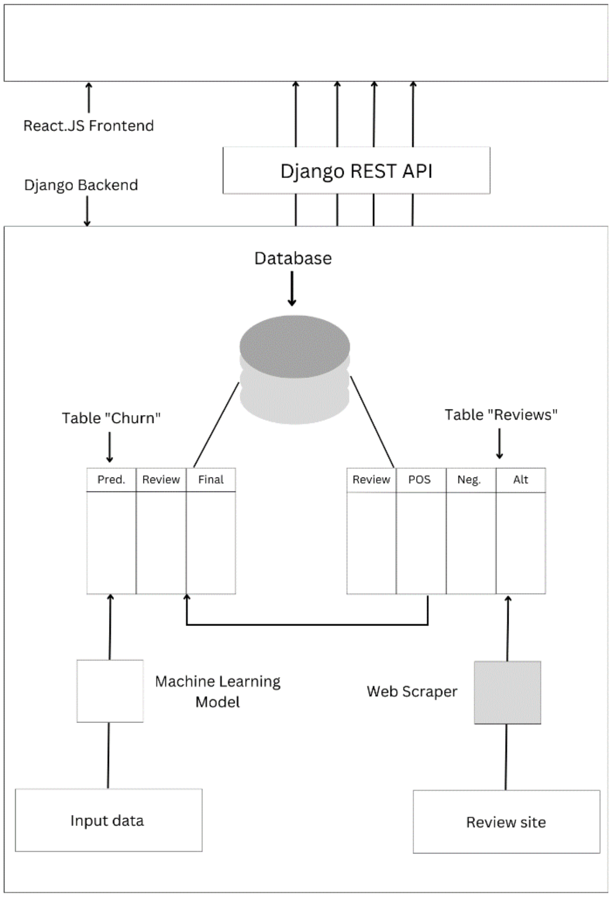
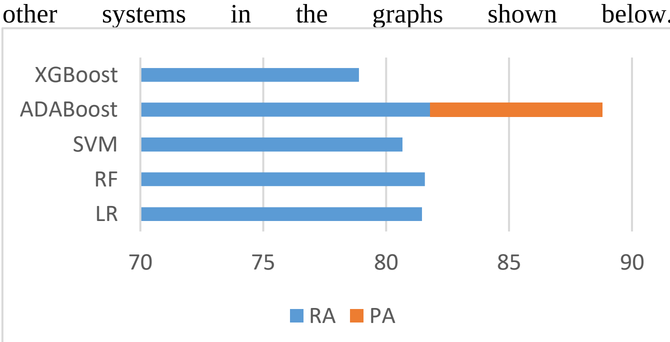


Fig. 4: The architecture of this system

VI. Result Analysis

This system's predicted accuracy has been compared to other systems in the graphs shown below.



LR: Logistic Regression
RF: Random Forest
SVM: Support Vector Machine
RA: Recorded Accuracy
PA: Predicted Accuracy

Fig. 5: Comparison with other algorithms on Telco Churn

In Fig. 4, this system has been compared to the other algorithms fitted on IBM's Telco Churn dataset and in Fig. 5, it is compared to other systems using machine learning techniques.

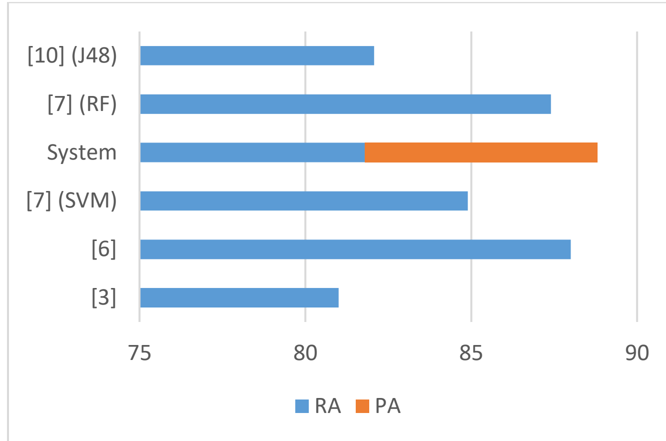


Fig. 6: Comparison with other papers

As seen in the above chart, this system performs very well when compared to other systems developed and surveyed in other papers.

VII. Conclusion

The prediction of customer churn, an important consideration in the telecommunication industry, is a field that has been studied by many researchers [3-12]. In [1], there was a proposal made to attempt to increase the accuracy of this prediction. In this paper, a working model of the proposed system has been developed, with the technologies of machine learning, web development, web scraping and natural language processing having been implemented. This system can yield a better prediction than previously researched methods and can be a significant asset to telecommunication companies and their customers.

In the future, we will try to generalize this prediction, make it usable for other fields, wherever legally possible.

References

[1] Dr. Madhura Phadke, Anirudh Bhattacharya, Mink Shethia, Saumya Shah (2022), Feedback Based Telecom Churn Prediction Using Machine Learning, 2022 5th International Conference on Advances in Science and Technology (ICAST)

[2] Jo Mackiewicz, Dave Yeats, Thomas Thornton (2016), The Impact of Review Environment on Review Credibility, IEEE Transactions on Professional Communication

[3] Pan Tang (2020), Telecom Customer Churn Prediction Model Combining K-means and XGBoost Algorithm, International Conference on Mechanical, Control and Computer Engineering (ICMCCE)

[4] Priya Gopal, Dr. Nazri Bin MohdNawi (2021), A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce, IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

[5] Peng Sun, Xin Guo, Yunpeng Zhang, Ziyan Wu (2013), Analytical Model of Customer Churn Based On Bayesian Network, 2013 Ninth International Conference on Computational Intelligence and Security

[6] Irfan Ullah, Basit Raza1, Ahmad Kamran Malik, Muhammad Imran, Saif ul Islam, Sung Won Kim (2019), A Churn Prediction Model using

Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector IEEEAccess

[7] Ahmad Hammoudeh, Malak Fraihat, Mahmoud Almomani (2019), Selective Ensemble Model for Telecom Churn Prediction, IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)

[8] Mr. Anurag Bhatnagar, Dr. Sumit Srivastava (2020), Performance Analysis of Hoeffding and Logisitic Algorithm for Churn Prediction in Telecom Sector, International Conference on Computation, Automation and Knowledge Management (ICCAKM) Amity University

[9] Xin Hu, Yanfei Yang, Lanhua Chen, Siru Zhu (2020), Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network, IEEE 5th International Conference on Cloud Computing and Big Data Analytics

[10] Ramakanta Mohanty, Jhansi Rani K (2015), Application of Computational Intelligence to predict churn and non-churn of customers in Indian Telecommunication, International Conference on Computational Intelligence and Communication Networks

[11] Pronaya Bhattacharya, Akhilesh Ladha, Ashwani Kumar, Ashwin Verma, Umesh Bodkhe, (2020) BaYcP: A novel Bayesian customer Churn prediction scheme for Telecom sector, 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)

[12] Yakub K. Saheed, Moshood A. Hambali (2021), Customer Churn Prediction in Telecom Sector with Machine Learning and Information Gain Filter Feature Selection Algorithms, 2021 International Conference on Data Analytics for Business and Industry (ICDABI)

[13] Ankit Verma, Chavi Kapoor, Abhishek Sharma, Biswajit Mishra (2021), Web Application Implementation with Machine Learning, 2nd International Conference on Intelligent Engineering and Management (ICIEM)

[14] Joel Vainikka (2018), Full-stack web development using Django REST framework and React (Thesis), Metropolia University of Applied Sciences

[15] Shen Zhao (2022), Design and Implementation of Big Data Crawling and Visualization System Based on COVID-19 Data, 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)

[16] Sara Motahari, Taeho Jung, Hui Zang, Krishna Janakiraman, XiangYang Li, Kevin Soo Hoo (2014), Predicting the Influencers on Wireless Subscriber Churn, IEEE WCNC'14 Track 4 (Services, Applications, and Business)

[17] David Mathew Thomas, Sandeep Mathur (2019), Data Analysis by Web Scraping using Python, Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]

[18] Orissa Octaria, Danny Manongga, Ade Iriani, Hindriyanto Dwi Purnomo, Iwan Setyawan (2022), Mining Opinion Based on Tweets about Student Exchange with Tweepy and TextBlob, International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)