iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

PREDICTIVE ANALYSIS FOR CUSTOMER RETENTION

D.R Nagamani [1]
[1]Assistant Professor, Bangalore Institute of Technology, Visvesvaraya Technological University, India
Mohan Kumar G[2]
Mohith HK Gowda[3]
Koushik S Murthy[4]
Karthik D[5]
[2,3,4]UG Student, Bangalore Institute of Technology, Visvesvaraya Technological University,India

ABSTRACT
Customer retention plays a pivotal role in sustaining the growth of service-oriented industries, especially within the competitive telecom sector. This study introduces a framework using predictive analytics to identify customers at risk of churn, leveraging data-driven insights and machine learning models. Utilizing the publicly available Telco Customer Churn dataset, the study involves data cleaning, feature selection, and classification model building. A frontend system enhances the usability of the predictions, enabling businesses to implement proactive retention strategies like personalized SMS communication. This approach offers a scalable solution for tackling churn effectively and boosting organizational performance.

Keywords:
Client Retention, Churn Forecasting, Telecommunications Sector, Predictive Modelling, Artificial Intelligence, Data Analysis, Categorization Techniques, Attribute Engineering, Retention Methods, Message-Based Outreach, Business Insights, Service-Oriented Enterprises, Telecom Churn Data, User-Focused Interface, Expandable Framework

INTRODUCTION
In a highly competitive market environment, retaining customers has emerged as a critical factor in sustaining profitability. The telecom industry faces unique challenges such as escalating customer acquisition costs and high churn rates. Customer churn, defined as customers discontinuing a service, directly impacts revenue. Identifying such customers in advance and implementing targeted retention strategies can preserve market share and foster customer loyalty.
This paper proposes the comprehensive churn prediction model integrating predictive analytics and machine learning. By utilizing the Telco Customer Churn dataset, the model explores data preprocessing, advanced feature engineering, and classification techniques to provide accurate predictions. Additionally, a user-centric application is introduced to facilitate actionable insights and enable direct retention efforts through personalized SMS communication. This synergy of analytics and actionable insights bridges the gap between theoretical modelling and practical business applications, positioning the framework as an adaptable solution for similar challenges in other service industries.

OBJECTIVES
The basic objective of this project is to develop a robust predictive framework for identifying customers at high risk of churn within the telecom sector by leveraging machine learning techniques on the Telco Customer Churn dataset. The project emphasizes thorough data preprocessing and analysis to uncover critical patterns thatenhance prediction accuracy. Additionally, it introduces a user-friendly frontend application for visualizing churn predictions and implementing actionable retention strategies. A key feature of this system is the integration of a personalized messaging platform, enabling proactive customer engagement through SMS notifications. The study also focuses on evaluating and optimizing various classification models to ensure high

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

performance and reliability. Ultimately, the proposed framework aims to offer a scalable and adaptable solution to mitigate churn challenges in the telecom industry and other service-oriented sectors.

METHODOLOGY

The research follows a systematic process, adhering to the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which includes the following phases:

1. **Business Understanding**: Define the goal of identifying potential customer churn as a classification problem, ensuring alignment with business objectives.
2. **Data Understanding**: Conduct exploratory analysis to recognize patterns in customer behavior, such as service usage, billing trends, and demographics, which inform feature selection.
3. **Data Preparation**: Handle the missing values, address outliers, encode categorical variables, and scale numerical features for seamless integration with ML models. To manage class imbalance,techniques such as SMOTE (Synthetic Minority Oversampling Technique) are applied.
4. **Modeling**: Exibit various ML algorithms, including logistic regression, decision trees, random forests, and ensemble methods like Naïve Bayes. Hyperparameter tuning and cross- validation are employed to enhance performance.
5. **Evaluation**: Evaluate the models using some metrics like accuracy, precision, recall, F1-score, and ROC-AUC to ensure reliable predictions. Tools such as confusion matrices and ROC curves provide a comprehensiveevaluation.
6. **Deployment**: Integrate the chosen model into an intuitive application that visualizes predictions and supports customer retention efforts through personalized SMS outreach. This practical implementation underscores the relevance of predictive modeling in real-world applications.

DATAANALYSIS AND INTERPRETATION

Data Analysis and Interpretation

The backend code employs Python and its robust libraries for data preprocessing, visualization, and model evaluation. The process is structured into the following stages:

1. **Data Loading**:
   The dataset is loaded using pandas, ensuring efficient handling of structured data. Key attributes influencing churn prediction are extracted for further analysis.
2. **Data Preprocessing**:
   - **Encoding**: Categorical variables are tranformed into numerical representations using LabelEncoder to ensure compatibility with machine learning algorithms.
   - **Scaling and Normalization**: Appropriate scaling techniques, like Min-Max scaling, are implemented to ensure uniform contributions from features.
   - **Handling Missing Values**: Missing data points are identified using imputation strategies such as filling with mean or mode values.
3. **Exploratory Data Analysis (EDA)**:
   Visualizations generated with seaborn and matplotlib identify patterns in customer behavior, such as correlations between service usage and churn likelihood. Heatmaps and distribution plots help pinpoint significant attributes.
4. **Model Development**:
   - **Algorithm Selection**: A RandomForestClassifier is employed due to its robustness and ability to handle complex feature interactions. Other ML models may also be incorporated as baselines.
   - **Data Splitting**: The datasets are split into training and testing subsets using train_test_split to evaluate model generalizability.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

5. **Model Evaluation**:
   Metrics like accuracy, precision, recall, F1-score, and ROC-AUC are calculated using the classification_report and metrics modules. The performance results are visualized through confusion matrices and ROC curves, providing a better view of the model's efficacy in differentiating between churned and retained customers.
6. **Feature Importance Analysis**:
   The RF algorithm outputs the relative importance of each feature, offering actionableinsights for customer retention strategies.
7. **Visualization of Results**:
   Matplotlib and Seaborn are utilized to create intuitive plots that display the distribution of churn across various factors, the correlation between features, and the performance metrics of the model.

SMS Operation

To integrate SMS notifications into your churn prediction system, you will initially need to create a Twilio account, which provides you with the necessary credentials, including an Account SID, Auth Token, and a dedicated phone number. These credentials allows you to use Twilio's Python library to send messages to customers.

To begin, install the Twilio Python package using pip install twilio. Once your churn prediction model generates the output (indicating whether a customer is likely to churn), you can use Twilio's API to send an SMS. For example, if the model predicts a high likelihood of churn for a particular customer, you can automatically send a message such as, "We've noticed you might be thinking about leaving. How can we assist you?" This can help in engaging with the customer proactively to address any potential concerns.

To send a message, create a function utilizing the Twilio Client class, passing in your Account SID and Auth Token. The client.messages.create() method allows you to send the message to the customer's phone number. By integrating this functionality, your churn prediction model will be able to trigger immediate responses, enhancing customer retention efforts. You can further tailor the content of these messages based on the likelihood of churn or specific customer attributes, ensuring a personalized approach.

Bellow code repesents the twilio code

```python
from twilio.rest import Client

account_sid = 'your_account_sid'
auth_token = 'your_auth_token'
client = Client(account_sid, auth_token)

def send_sms(to_phone_number, message):
    from_phone_number = 'your_twilio_phone_number'
    message = client.messages.create(
        body=message,
        from_=from_phone_number,
        to=to_phone_number
    )
    print(f"Message sent to {to_phone_number}: {message.sid}")

if predicted_churn:
    customer_phone = '+1234567890'
    send_sms(customer_phone, "Hello")
```

Figure 1: Twilio code

Vol-08 Issue 11, November -2024 ISSN: 2456-9348
Impact Factor: 7.936

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Out[30]: 0.820184790334044

Figure 2: Accuracy of Support Vector Machine

	AccountID	Churn	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count	account_segment	CC_Agent_Score	Marital_Status	rev_p
0	20000	1	4.0	3.0	6.0	Debit Card	0	3.0	3.0	Super	2.0	Single	
1	20001	1	0.0	1.0	8.0	UPI	1	3.0	4.0	Regular Plus	3.0	Single	
2	20002	1	0.0	1.0	30.0	Debit Card	1	2.0	4.0	Regular Plus	3.0	Single	
3	20003	1	0.0	3.0	15.0	Debit Card	1	2.0	4.0	Super	5.0	Single	
4	20004	1	0.0	1.0	12.0	Credit Card	1	2.0	3.0	Regular Plus	5.0	Single	
...	...	...	...	...	...	...	...	...	...	...	...	...	
11255	31255	0	10.0	1.0	34.0	Credit Card	1	3.0	2.0	Super	1.0	Married	
11256	31256	0	13.0	1.0	19.0	Credit Card	1	3.0	5.0	HNI	5.0	Married	
11257	31257	0	1.0	1.0	14.0	Debit Card	1	3.0	2.0	Super	4.0	Married	
11258	31258	0	23.0	3.0	11.0	Credit Card	1	4.0	5.0	Super	4.0	Married	
11259	31259	0	8.0	1.0	22.0	Credit Card	1	3.0	2.0	Super	3.0	Married	

Figure 3: Data set categories.

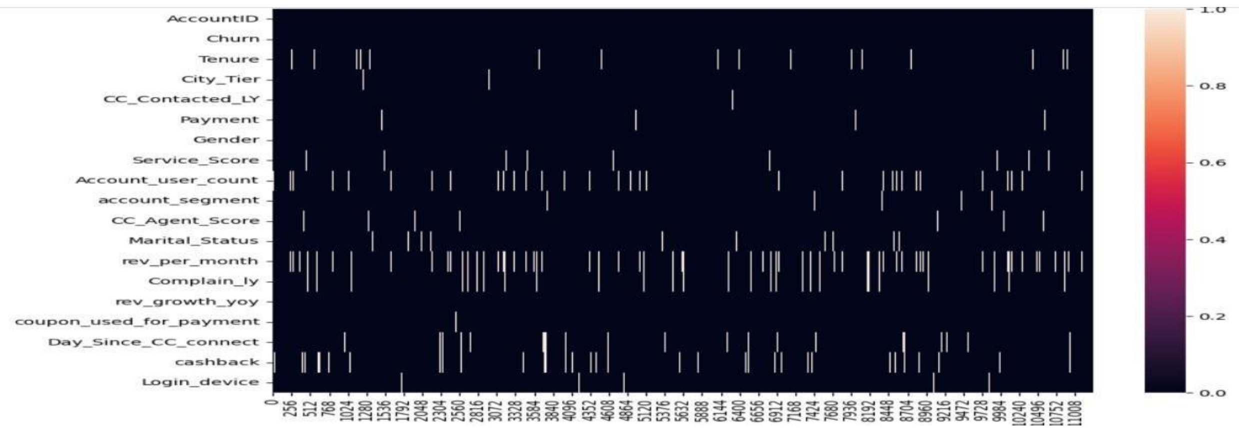

Figure 4: Impact of Different features

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

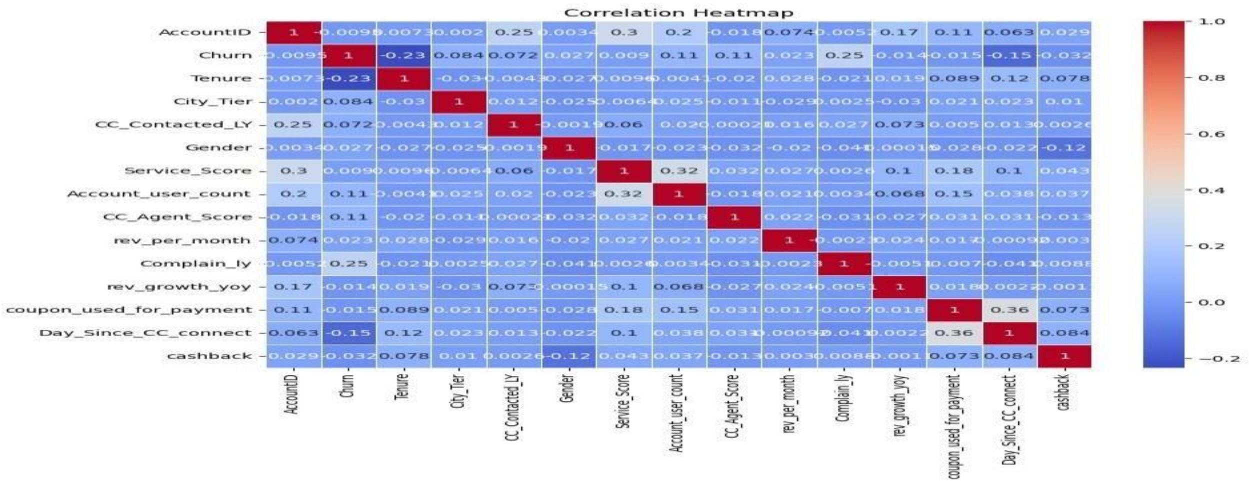

Figure 5: Correlation Heatmap

RESULTS AND DISCUSSION

```
RF's Accuracy is:  0.9617224880382775
              precision    recall  f1-score   support

           0       0.96      1.00      0.98      1381
           1       0.98      0.80      0.88       291

    accuracy                           0.96      1672
   macro avg       0.97      0.90      0.93      1672
weighted avg       0.96      0.96      0.96      1672
```

Figure 6: Random forest accuracy.

```
Naive Bayes's Accuracy is:  0.8397129186602871
              precision    recall  f1-score   support

           0       0.86      0.97      0.91      1381
           1       0.60      0.24      0.35       291

    accuracy                           0.84      1672
   macro avg       0.73      0.60      0.63      1672
weighted avg       0.81      0.84      0.81      1672
```

Figure 7: Naive Bayes's Accuracy

IJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

SVM's Accuracy is: 0.8881578947368421

	precision	recall	f1-score	support
0	0.89	0.98	0.94	1381
1	0.82	0.45	0.59	291
accuracy			0.89	1672
macro avg	0.86	0.72	0.76	1672
weighted avg	0.88	0.89	0.87	1672

Figure 8:Support Vector Machine accuracy.

DecisionTrees's Accuracy is: 88.45693779904306

	precision	recall	f1-score	support
0	0.91	0.96	0.93	1381
1	0.73	0.54	0.62	291
accuracy			0.88	1672
macro avg	0.82	0.75	0.78	1672
weighted avg	0.88	0.88	0.88	1672

Figure 9: Decision tree's accuracy.

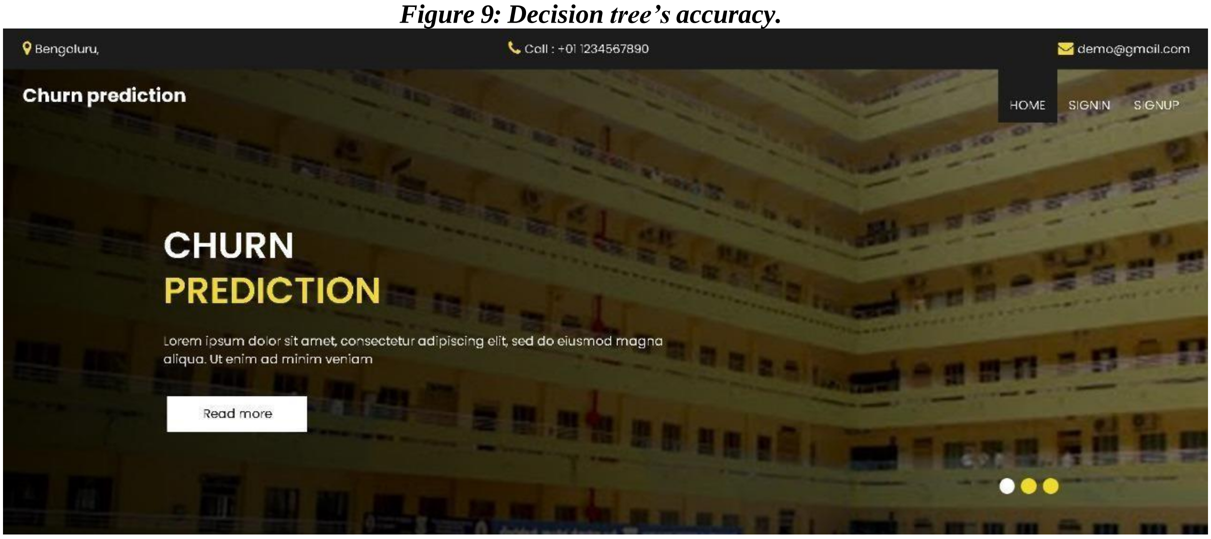

Figure 10: Home page of the front end

ISSN: 2456-9348
Impact Factor: 7.936

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Figure 11: Churn predection page

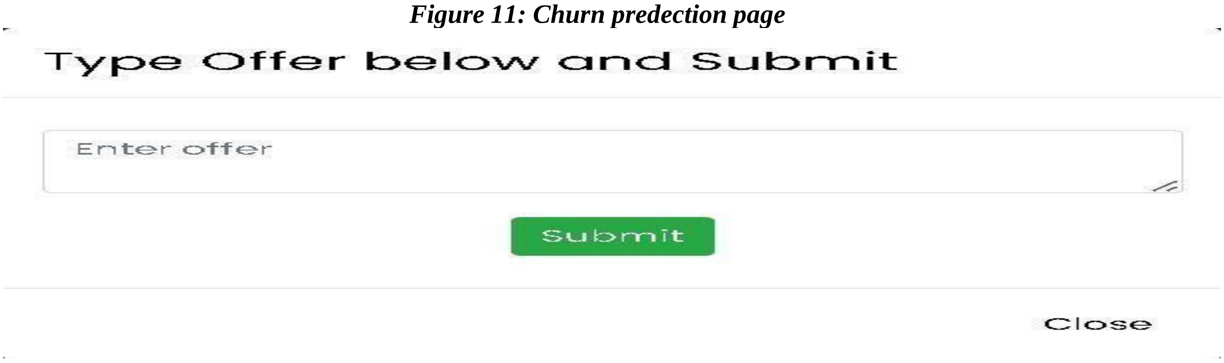

Figure 12: If churn, then offer is sent through this page

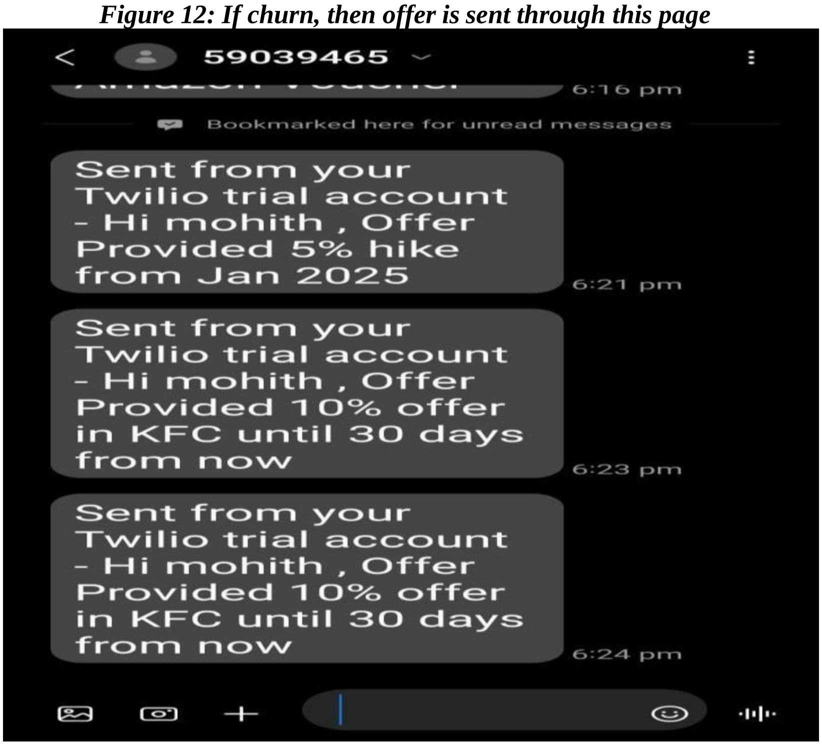

Figure 13: Message is sent to the churned customer as the company's wish.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

RESULTS AND DISCUSSIONS

The RF Classifier was utilized to predict churn, with the model trained under the training dataset and evaluated on test dataset. Using 20 estimators (decision trees) and a fixed random state for reproducibility, the model demonstrated an impressive predictive capability. The accuracy score for the RF model on the testdata was calculated to be 96%.

Additionally, the detailed classification report provided information into various performance metrics. These included precision, recall, and F1-score for each class in the dataset. The results highlighted the model's effectiveness in distinguishing between churned and non-churned customers, showcasing the Random Forest's ability to handle classification tasks in churn prediction with some balance between high accuracy and interpretability. These metrics confirm that, model is well-suited for the data and performs reliably in identifying churn behaviour.

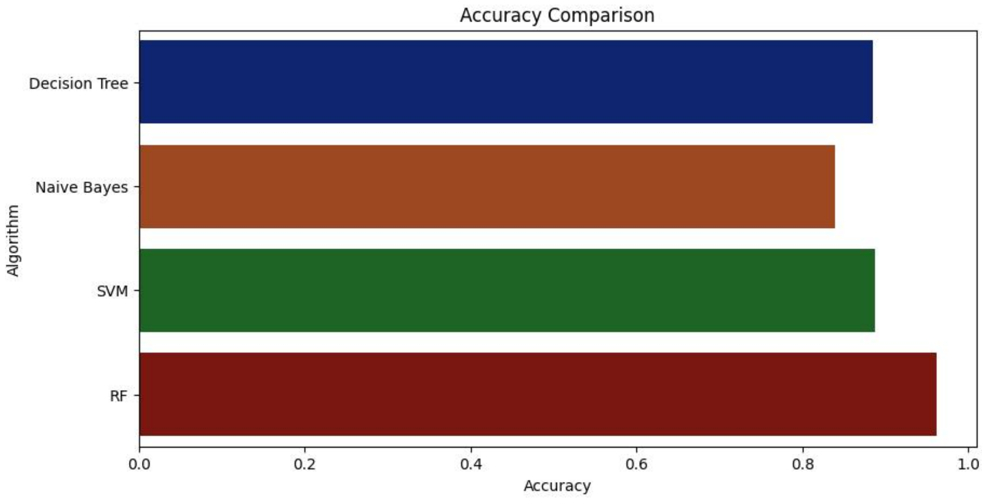

Figure 14: Accuracy Comparison of Different Algorithms used.

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to Dr. Nagamani for her invaluable guidance and support throughout this development of this project. Her expertise and encouragement wereinstrumental in shaping our understanding and approach to the problem.

We also sincerely thank the faculty and staff of Bangalore Institute of Technology for providing a conducive learning environment and the resources necessary to complete this project. Their encouragement and the skills we gained during the studies have been crucial in this endeavour.

CONCLUSION

Predicting customer churn is essential for businesses aiming to retain some of their most valuable customers by pinpointing thefactors that lead to attrition. This research highlights that variables such as customer tenure, contract types, monthly charges, and access to additional services like online security and technical support play an important role in influencing churn rates. By thoroughly checking these aspects, few businesses can proactively address customer needs with tailored strategies, including bundled services, incentives for long-term contracts, or loyalty rewards.

The application of ML models for churn prediction has yielded encouraging outcomes, particularly whensupported by techniques like data balancing and advanced feature engineering to improve prediction accuracy. This project emphasizes the critical role of strategic retention efforts in mitigating churn, as well as the transformative potential of data-driven insights in guiding business decisions.

As customer preferences and behaviors evolve, organizations must continuously induce their predictive models and retention strategies to remain competitive. By prioritizing customer satisfaction, offering personalized solutions, and

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

predictive analytics, some companies can significantly reduce churn, enhance customer loyalty, and achieve sustained growth.

REFERENCES

[1] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju "Churn prediction victimisation approachable support vector machine: associate analytical CRM application", Applied SoftComputing nineteen (2014) 31–40.

[2] Chih-Fong Tsai, Yu-Hsin metal "Customer churn prediction by hybrid neural networks", rofessional Systems with Applications thirty six (2009) 12547–12553.

[3] Abhishek and Ratnesh ,"Predicting client Churn Prediction in telecommunication Sector ictimisation numerous Machine Learning Techniques", within the proceedings of 2017 International Conference on Advanced Computation and Telecommunication, Bhopal, India, 2017.

[4] Abinash and Srinivasulu U ,"Machine Learning techniques applied to postpaid subscribers: case study on the telecommunication trade of Morocco", within the proceedings of 2017 International Conference on creative Computing and IP , Coimbatore, India, pp. 721-725, 2017

[5]Kavitha V, Hemanth G , Mohan S.V and Harish M , "Churn Prediction of client in telecommunication trade victimization Machine Learning Algorithms", International Journal of Engineering analysis & Technology(2278-0181), Vol. 9, Issue05, pp. 181-184, 2020.

[6]Krishna B.N, and Sasikala ,"Predictive Analysis and Modeling of client Churn in telecommunication victimization Machine Learning Technique ,"In the proceedings of International Conference on Trends in natural philosophy and IP , Tirunelveli, India, pp. 6-11, 2019.

[7] H. Jain, A. Khunteta, and S. Srivastava, "Churn prediction in telecommunication victimisation provision regression and logit boost," Procedia applied science, vol. 167, pp. 101–112, 2020.

[8] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system supported ensemble learning victimisation feature grouping," Applied Sciences, vol. 11, no. 11, p. 4742, 2021.

[9] R. Kamalraj, S. Neelakandan, M. Ranjith Kumar, V. Chandra Shekhar Rao, R. Anand, and H. Singh, "Interpretable filter primarily based convolutional neural network (IF- CNN) for aldohexose prediction and classification victimisation PD-SS rule," mensuration, vol. 183, Article ID 109804, 2021.

[10] A. De Caigny, K. Coussement, K. W. De Bock, and S. Lessmann, "Incorporating matter data in client churn prediction models supported a convolutional neural network," International Journal of statement, vol. 36, no. 4, pp. 1563–1578, 2020.

[11]Ahmed, A. A., & Maheswari, D. (2017). Churn predictionon large telecommunication knowledge victimisation hybrid firefly primarily based classification. Egyptian IP Journal, 18,215–220.

[12]Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K.,Hussain, A., et al. (2017). client churn prediction within the telecommunication sector employing a rough set approach. Neurocomputing, 237, 242–254.

[13]Sabbeh, S. F. (2018). Machine-learning techniques for client retention: A comparative study. International Journal of Advanced applied science and Applications, 9(2), 273–281.

[14]Andreea, D., Alexandra, A. M., Stancu, S., et al. (2020). Churn prediction in telecommunication industry: Model interpretability. Journal of japanese Europe analysis in Business and social science, 2020, 241442.

[15]Karanovic, M., Sladojevic, S., Arsenovic, M., & Stefanivic, D. (2018). Telecominication churn prediction—Deep learning approach. In twenty sixth Telecommunications forum TELFOR 2018, Serbia, Belgrade.