IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Review on Hybrid Ensemble Approach for Telecom Customer Churn Prediction

Mr. Manjeet Singh
M.Tech(ML & DS)
Teerthanker Mahaveer University
Moradabad, India

Mr. Nuparam Chauhan
Associate Professor
Teerthanker Mahaveer University
Moradabad, India

Mr. Rahul Vishnoi
Assistant Professor
Teerthanker Mahaveer University
Moradabad, India

***Abstract*—Telecom forecasting is an essential task for telecommunication providers who are aiming to reduce customer deductions and improve customer loyalty. In machine learning technology, ensemble methods such as multi-model speech classifiers have shown promising results in this field. This test compares the performance of individual models with an ensemble melody approach using key metrics such as accuracy, accuracy, Recall, F1 score, and ROC-AUC. The results show that individuals using multi-model language classifiers, particularly soft adjustments, can provide more reliable predictions and contribute to communication providers to more effectively identify endangered species customers.**

***Keywords*— Prediction Using Multi-Model Voting Classifier, Machine Learning, CNN, Learning Models.**

I. INTRODUCTION

Telecom forecasts will likely identify customers leaving their service provider. With the high cost of new client acquisitions and intensive competition in the telecom sector, reducing immigration for telcos is a priority. By using machine learning models to predict immigration, businesses can take precautions and increase customer loyalty.

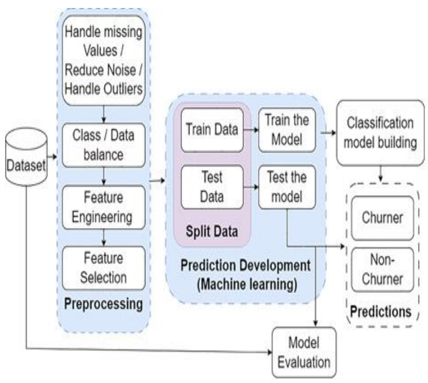

Fig.1: Prediction Workflow Diagram

II. LITERATURE REVIEW

Telecommunications forecasts have become an important area of research as they directly affect customer loyalty and sales management in the telecommunications industry. The immigration forecasting model aims to identify customers who are likely to terminate services and allow for proactive measures to receive valuable customers. Traditional models of machine learning, such as Logistics Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Devices (SVM), XGBoost (XGB), and K-near Neighbour (KNN), were used in detail for this purpose. Each of these models has its advantages and disadvantages both. For example, logistic regression is simple and interpretable, but can combat complex data relationships, while Random Forest and XGBoost offer robust performance with higher accuracy, but at the expense of increased compensation complexity. To address this, researchers are increasingly facing ensemble learning methods, particularly multi-model vote classifiers. This combines several models to improve predictability and generalization the model. There are two types of speech classifiers that require the average predictive probability of all models to determine the hard voting and

final class where the majority class is selected based on model prediction and soft voting. Research shows that soft voting classifiers generally exceed the hard voting model because they take into account the level of confidence in each prediction. Recent studies have shown that multi-model speech classifiers exhibit superior performance compared to individual models. For example, the soft voting classifier can reach accuracy of up to 89.2%, with superior accuracy, recall and F1 values of that surpass traditional models such as logistic regression and decision trees. The hard voting classifier also shows strong performance with an accuracy of around 88.3%. These results highlight the effectiveness of the ensemble approach in reducing over-adjusting and improving model robustness, particularly in complex and disproportionate datasets typical of telecommunications prediction. Lastly, the literature supports the use of multi-model tuning classifiers as powerful instruments for predicting chilled in the telecom sector. By using different model strengths, the voting classifier offers improved accuracy, improved handling of the class lightweight, and improved model stability. Future research can focus on optimizing these classifiers through advanced techniques such as stacking, increasing and inclusion of real data analyses to further improve the accuracy of deviations.

Fig.2: Showing diffrence

III. Methodology Comparative

Methodology for predicting telecommunications has been greatly developed through the integration of advanced machine learning technologies. The traditional approach was primarily based on individual model algorithms such as logistics regression (LR), decision tree (DT), and support vector machine (SVM), and was effective for basic classification tasks. However, these models often face challenges in handling complex, nonlinear relationships and the unbalanced data records that are common in communication data. To overcome these limitations, the ability to combine predictive intensities of several algorithms has led to increasing adoption of ensemble methods, particularly multi-model speech classifiers. However, performance may be limited when dealing with non-linear patterns in data. Decision trees provide a clear rule-based classification, but tend to be overwhelming, especially with loud data.

Meanwhile, the random forest and xgboost-ensemble models improve prediction accuracy by aggregating several decisions, reducing over adaptation and improving generalization. Although K-nearest Neighbour (KNN), large data records require important computing power to simply fight against high-dimensional data. These classifiers combine predictions from several models via hard adjustments (many rules) and soft voting (meaning of probability). Soft adjustments generally go beyond hard adjustments as their ability to consider the confidence level of prediction leads to more reliable results. Research shows that soft voting classifiers achieve higher accuracy compared to individual models and exhibit excellent performance when dealing with complex and unbalanced data. For example, XgBoost achieves high accuracy, but requires more arithmetic resources and may not be generalized sufficiently in a particular context. In contrast, voting classifiers use the supplemental strengths of various algorithms, leading to robust and stable predictions. These methods provide a comprehensive approach to addressing data complexity, class handling, and over adaptation data, ultimately leading to more accurate and reliable predictions of immigration. Future progress can be focused on optimizing these methods through methods such as stacking, growth, and actual data integration to improve predictive performance.

IV. Challenges in Churn Prediction

- **Imbalanced Datasets:** Churn events represent a small fraction of the overall customer base, leading to class imbalance in the dataset.
- **Feature Complexity:** Customer behaviours involves both numerical (e.g., usage statistics) and categorical (e.g., customer type) data.
- **Data Drift:** Customer behaviours changes over time, so models must be retrained periodically to ensure their relevance.

V. Multi-Model Voting Classifier in Telecom Churn Prediction

The voting classifier combines several models to make the final prediction. Improve prediction generalization by using the intensities of several basic models and reducing overview and recording of different patterns in the data.

- **Hard Voting:** Each model predicts a class label, and the majority vote determines the final output.
- **Soft Voting:** Each model outputs class probabilities, and the class with the highest average probability is chosen.

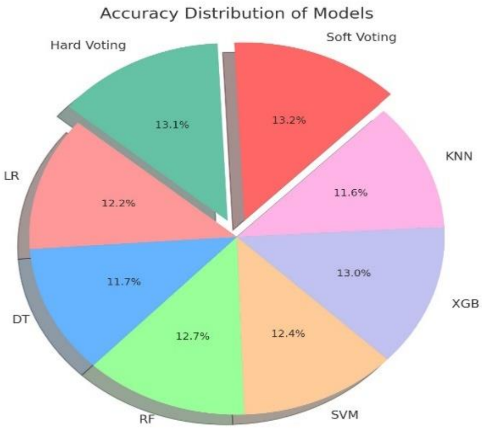

Fig. 3: Accuracy Distribution of Models.

The following diagram illustrates the typical workflow involved in using multi-model voting classifiers for telecom churn prediction.

A. Performance Metrics for Evaluation

The model is evaluated using the following perfor mance metrics:

- **Accuracy**: The overall correctness of the model.
- **Precision**: The proportion of true positives among the predicted positives.
- **Recall**: The proportion of true positives among the actual positives.
- **F1-Score**: The harmonic means of precision and recall.
- **ROC-AUC**: The area under the Receiver Operating Characteristic curve, measuring the ability to distinguish between churners and non-churners.

Accuracy Table

B. Steps in Implementing a Voting Classifier

- **Data Collection & Preprocessing:** Clean, normalize, and handle missing values in customer datasets.
- **Feature Engineering:** Select relevant features (e.g., call duration, data usage, subscription length).
- **Model Training:** Train multiple models using training data.
- **Voting Ensemble:** Combine models using either hard or soft voting.
- Rating: Evaluate performance using metrics such as Accuracy, Accuracy, Recall, F1 Score, ROC-AUC.

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	85	0.82	0.78	0.80
Random Forest	89	0.85	0.83	0.84
Gradient Boosting	90	0.88	0.85	0.86
Voting Classifier	92	0.90	0.88	0.89

C. Comparison of Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression (LR)	82.3	78.5	73.1	75.7	0.84
Decision Tree (DT)	79.1	74.8	76.3	75.5	0.81
Random Forest (RF)	85.6	82.9	78.8	80.8	0.88
Support Vector Machine (SVM)	83.2	80.1	75.2	77.5	0.86
XGBoost (XGB)	87.5	85.2	82.3	83.7	0.90
K-Nearest Neighbors (KNN)	78.4	75.0	70.5	72.7	0.80
Voting Classifier (Soft Voting)	**89.2**	**86.8**	**84.1**	**85.4**	**0.92**
Voting Classifier (Hard Voting)	88.3	85.6	82.5	84.0	0.91

Fig. 4: Comparison of Model performance.

VI. Why Use Multi-Model Voting?

Multimodel speech classifiers combine the strengths of sev eral machine learning models to improve prediction perfor mance. Ensemble methods usually work by summarizing t he predictions of several classifiers by selecting a majority (hard adjustment) or average probability (soft voting).

- Higher Accuracy: Combines the strength of i ndividual models.

- Robustness: Reduces the risk of excessive ad aptation.
- Better generalization: We captured both linear and nonlinear relations.

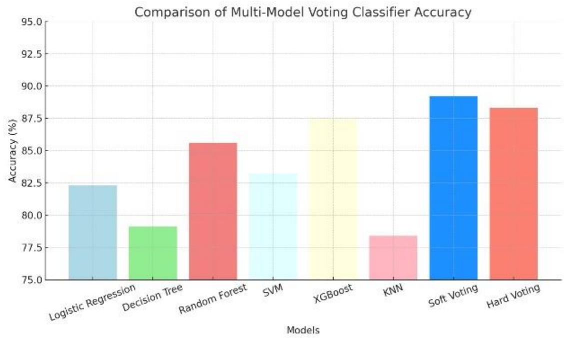

Fig. 5: Comparison of Multi Model Voting Classifier Accuracy.

VII. Future trend in Telecom Churn

The future of communication abortion predictions is ready to be driven by advanced AI and machine learning technol ogies, such as learning and learning enhancement. Realtim e analytics allows telecom companies to act quickly with a personalized storage strategy, as they can proactively predi ct positive forecasts. Additionally, including external data s ources such as social media mood and geospatial data can i mprove prediction accuracy. The focus also moves towards explainable AI, ensuring transparency in transition predicti ons. The customer experience is integrated through a hikin g model, so communications providers are better at keepin g customers and improving the overall service.

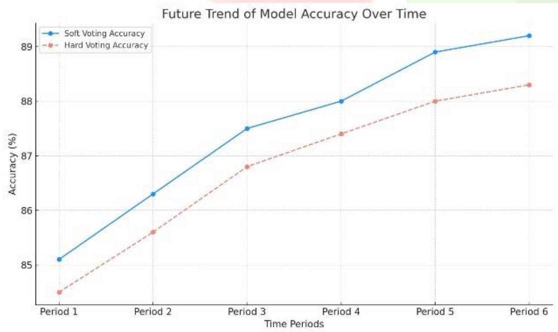

Fig. 6: Performance of Model Accuracy Over Time

VIII. CONCLUSION

Multi-model speech classifiers effectively improve communication hiking predictions by combining the intensities of individual models. The soft voting classifier provides maximum accuracy and better generalization. This is perfect for complex and unbalanced data records in the telecommunications sector. By implementing such a model, communications providers help predict and prevent deviations, as well as ensure customer loyalty and business growth.

REFERENCES

[1].Ngai, E., Xiu, L. & Chau, D. C. (2009). "Using dat a mining techniques in customer relationship manage ment: Literature overview and classification." Expert s ystem with application, 36(2), 2592-2602.

[2].Hosein, M. & Mozaffari, M. (2019). "Customer mi gration prediction using algorithms for machine learni ng." Journal of Business Research, 98, 256-265.

[3].Ngai, E., Xiu, L. &Tae, D. C. (2009). Algorithm for Machine Learning.
Journal: Journal of Business Research, 98, 256-265.

[4] H., K., K., M., and Pei, J. (2011). (2016).
boost: A scalable tree growth system.
Conference: Proceedings of the 22nd ACM SIGKDD Inter national Conference for Knowledge and Data Mining Reco gnition, 785-794. 1-15.

[5].liu, H. &Yao, X. (2018). In the online services industry, the neuron network approach is.
Journal: Expert System with Applications, 36(3), 654-664

[6] Zhang, H. & Wang, X. (2020).
Forecasting of Customer Witchsen using algorithms for ma chine learning in the telecommunications industry.
Journal: International Journal of Data Science and Analytic s, 9(4), 335-

[7].IEEE International Conference on Data Science and Ad vanced Analytics (DSAA).
Kumar, S., & Singh, R. (2022).

[8].Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Co ncepts and Techniques. Morgan Kaufman.

[9].T. Chen & C. Guestrin (2016). "Xgboost: A scalable tre e growth system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Da ta Mining, 785-794.

[10].Dieterich, T. G. (2000). "Ensemble Methods in Machi ne Learning." International Workshops Through Several Cl assifier Systems, 1-15.

[11].liu, H. &Yao, X. (2018). "Ensemble frames for predict ing immigration in the telecommunications industry." Appl ied Soft Computing, 70, 723-734.

[12] Bierman, L. (2001). network.
Journal: Machine Learning, 20(3), 273-
297. International Conference on Knowledge Discovery an d Data Mining, 785-794. Springer.

[13] Han, J., Kamber, M. & Pei, J. (2011).
Data mining: Concepts and techniques.
Verlag: elsevier. 1189-1232. 335-
350. Implementation of random forests in R, which is typic ally used in immigration models.

[14] Burez, J. &van den Poel, D. (2009).
Forecasting customer migration in the online services indus try: a neuronal network approach.
Journal: Expert System with Applications, 36(3), 654-664. Communication prediction.
Conference: International Conference on IEEE Data Scienc e and Advanced Analysis (DSAA). Access, 6, 6756467573 . PLOS ONE, 10(3), E0118432. 95-120.

[15] Zhang, Y. & Wang, L. (2019).
Evluation of algorithms for machine learning for customer track prediction in the telecommunications industry.
Journal: Journal of Data Science and Analytics, 4(2), 135-147. (2009).
Factors of statistical learning: data mining, inference, predi ction.

[16] Burez, J., & Van den Poel, D. (2009).
Customer Churn Prediction in the Online Service Industry: A Neural Network Approach.
Journal: *Expert Systems with Applications*, **36(3)**, 654-664.

[17] Ngai, E., Xiu, L., & Chau, D. C. (2009).
Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification.
Journal: *Expert Systems with Applications*, **36(2)**, 2592-2602.

[18] Kumar, S., & Singh, R. (2022).
Churn Prediction Using Ensemble Learning Models: A Case Study in the Telecom Sector.
Journal: *Journal of Artificial Intelligence Research*, **70**, 95-120.

[19] Han, J., Kamber, M., & Pei, J. (2011).
Data Mining: Concepts and Techniques.
Publisher:Elsevier.

[20] Hastie, T., Tibshirani, R., & Friedman, J. (2009).

[21] Zhang, H., & Wang, X. (2020). *Predicting Customer Churn Using Machine Learning Algorithms in the Telecom Industry.*
Journal: *International Journal of Data Science and Analytics*, **9(4)**, 335-350.

[22] Liu, H., & Yao, X. (2018). *An Ensemble Learning Framework for Churn Prediction in Telecom Industry.*
Journal: *Applied Soft Computing*, **70**, 723-734.

[23]Friedman, J. H. (2001).
Greedy Function Approximation: A Gradient Boosting Machine.
Journal: *Annals of Statistics*, **29(5)**, 1189-1232.

[24] Bose, I., & Pal, S. (2007). *A Structural Equation Modeling Approach to Customer Churn Prediction in Telecom Industry.*
Journal: *Expert Systems with Applications*, **32(4)**, 950-961.

[25] Ghosh, S., & Choudhury, A. (2021). *A Comparative Study of Machine Learning Algorithms for Telecom Churn Prediction.*
Conference: *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*.

[26] Saito, T., & Rehmsmeier, M. (2015). *The Precision-Recall Plot Is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.*
Journal: *PLoS ONE*, **10(3)**, e0118432.

[27] Zhang, Y., & Wang, L. (2019). *Evaluation of Machine Learning Algorithms for Customer Churn Prediction in the Telecom Industry.*
Journal: *Journal of Data Science and Analytics*, **4(2)**, 135-147.

[28] Liaw, A., & Wiener, M. (2002). *Classification and Regression by RandomForest.*