Autoencoder Based Churn Prediction In Telcom Using Machine Learning

Vasavi Mekala[1], Haneef Mohammad[2], Kamal Deepak Pedapalli[3], Varshitha Lingamaneni[4], V Rashmi[5]
[1, 2, 3, 4, 5] Dept of Information Technology
[1, 2, 3, 4, 5] Prasad V Potluri Siddhartha Institute Of Technology

Abstract- *Customer churn prediction is a critical challenge in the telecom industry, where retaining existing customers is more cost-effective than acquiring new ones. This paper explores the development of a churn prediction model using machine learning techniques to identify customers likely to switch service providers. The study focuses on data preprocessing, feature selection, and the application of advanced algorithms such as Hyper Parameterized Modified Artificial Neural Networks (HPMANN) to improve prediction accuracy. The proposed model aims to enhance customer retention strategies, reduce revenue loss, and improve long-term customer satisfaction. The results demonstrate the effectiveness of HPMANN in achieving higher accuracy compared to traditional methods like SVM and KNN.*

Keywords- Churn prediction, telecom industry, machine learning, artificial neural networks, feature selection, customer retention, HPMANN, SVM, KNN, data preprocessing, accuracy, precision, recall.

I. INTRODUCTION

In the telecom sector, "churn" keeps track of the costs of endorsers who go from one supplier to the next over a period of time. Because it is more beneficial to keep previous clients than it is to constantly enchant new ones. This is the foundation for creating a precise churn forecast form for identifying clients who are very likely to churn. The churn forecast model's main purpose is to identify clients who are likely to churn based on maintenance methods. Information collecting, preparation, planning, and estimation are all stages of a typical churn prediction model. It also denotes that recognizing the proper set of components has a significant impact on the size of true predictions.

The behaviour of churn customers has a detrimental impact on the company's performance, resulting in fewer sales and less service due to short-term consumers. Furthermore, the churner aids competitors in expanding unhappy clients through business promotions, but this result in revenue loss, has a gloomy impact on long-term customers, and intensifies the uncertainties, reducing the ratio of possible customers. In the telecommunications industry, churn is an important factor. The churning consumer decides to leave the service provider and plans to switch to a competitor in the market. The traditional methodologies are revealed, and client churns are classified into three groups: rotating churners, passive churners, and aggressive churners. Customer turnover is measured via churn management.

Churn prediction can be used in a variety of industries, including banking, life insurance, and others. This type of prediction allows the company to know if a customer is unhappy with their services in advance, allowing them to maintain their customer. In the existing research, customer turnover may be approached from two different perspectives. The researcher's focus is on improving customer churn prediction models, which are becoming more intricate and tend to improve predictive performance. Furthermore, researchers are more likely to understand churners and strive to provide client happiness. The forecast of customer turnover is seen as a supervisory problem, as it is determined by the customer's individual choice. Thus, customer churn prediction models are a hot topic among researchers, as they help managers better identify churners and make better judgments in the fight against customer churn.

The primary objective of this study is to develop a robust churn prediction model using machine learning techniques.

The model leverages data preprocessing, feature selection, and advanced algorithms like HPMANN to improve prediction accuracy. The proposed system aims to address the challenges of customer churn, which negatively impacts revenue, brand reputation, and long-term customer relationships.

II. LITERATURE SURVEY

For churn prediction, a variety of algorithms have been developed, the majority of which use machine learning techniques. Learning can take place in two ways: supervised and unsupervised. In supervised learning, a model is trained given input and output data that are known. The data is labelled, and these labels make use of the data to forecast future outputs based on fresh data. For unlabeled data, unsupervised learning is used. Based on statistical means for anticipating churns, this technique uncovers hidden patterns in the input data.

Recent advancements in machine learning have significantly improved churn prediction models in the telecom industry. Various studies have explored different algorithms and techniques to enhance prediction accuracy.

Hyper Parameterized ANN to classify Churn Customers:

Lalwani at el (2021) Customer churn prognosis is one among the telecom industry's most difficult concerns. The ability to predict client retention has significantly increased due to advances in machine learning and artificial intelligence. There are six phases to our proposed methodology.Information pre-processing and pattern evaluation are carried out in the first two phases. The gravitational search method is applied in the third step to assess feature extraction. Following that, the data was separated into two sections: a training phase and a test phase, with an 80/20 split. At the prediction stage, many standard predictive methods were opted on the training phase along with boosting and other ensemble methodologies were utilized to observe the variations in efficiency of the model. Furtherly, across validation method of k – fold is used for hyperparameter adjustment to overcome the model overfit on testing part.

Sebastian at el (2020) proposed three unique acquiesced churn prediction strategies in this research. Our ideas rely on the Minimax Probability, a resilient binary allocation optimization methodology that enhances efficacy in a probabilistic situation. Unlike other price - based techniques, which utilize cost metrics to distinguish amongst classifiers and/or to calculate the appropriate classification criterion given a probabilistic output, they modified this strategy and other versions to maximize the profit of a recall strategy in the optimal solution.The Tikhonov and LASSO establishers are added to a first approach that is built as a learning tool that does not involve a regularization term. Trials on standard churn prognosis datasets reveal that, when compared to alternative binary regulation algorithms, our proposal yields the highest return.

Sergue at el (2020) explained that they have focused on the properties of attributes that leads to churn rather than in prediction. The algorithms they have used revealed the parameters that lead to churn based on usage of a product. To discover this, they have used more sampling and least fitting and sequential time cross – verification along with other prediction models to predict and expand the churn possibility. This concluded that the usage of a non – sequential model executed greater than the prediction tools and stated that the combination of more sampling and least fitting resulted in greater yield for recall/ precision procedures. With the increasing popularity of the digital era, the researchers have conducted this thesis based on a well-known live example by collecting the data from a major SaaS cloud organization that deals with a market call system based on cloud which is namely, Aircall. The reason behind the focus on focusing to understand churn rate is due to the least retention of customers in that company.

Bharathi at el (2020) states that they have established this research aiming to enhance the significant cause of predicting retention of clients. As in practical example, a research that is conducted on telecom sector stated that identifying the retention of disappointed clients is larger than accessing to new clients. This scenario uses real-world data on customer turnover to anticipate and refine the use of classifiers in client retention prediction.So, by dividing and applying the developed process, the research is carried out in four stages, initiallyEM analysis for pre – processing, and for analyzing customer pattern KNN was opted. Furthermore, to seek an increased Cuckoo route optimization model with efficient SVM method is defined for choosing the attribute and to increase the forecasting of retention, respectively. Thus, the enhanced SVM is worked out as a standard prediction method using PSO methodology.

Deep Learning Approaches:

Kostic SM at el (2020) executed an experiment to predict churn rate in customers of a telecom industry depending on their network graph and derived a clustering approach on the nodes by digging deeper into the various parameters involved. This research primarily focuses on the possibility in identifying the vital nodes in the social graph that prone to churn retention. We illustrate that if a supervised telco operator loses a node, consumers who regularly engage with that vertex are more likely to quit the supervised telco operator circuit as well; hence, by reviewing current churn and prior call habits, we may forecast future customers who are likely to churn.The proposed strategy is broad enough to be used in any field where companionship ties are suspected of being a churn factor.

These studies highlight the importance of feature selection, data preprocessing, and advanced machine learning algorithms in improving churn prediction models.

III. METHODOLOGY

The proposed churn prediction model follows a structured approach, including data preprocessing, feature selection, and classification using HPMANN.

A. Data Preprocessing

Pre-Processing using Traditional Mechanisms:

The proposed system discovered that approximately 3512 entries in the Cell2Cell dataset have missing data during the data cleaning process, and it uses the traditional imputation technique to replace all missing values with their respective mean values. Since most of the attributes do not exhibit the skewness property, all missing values are replaced with their respective mean values. Assume that a property called "PrizmCode" contains four possible values, each of which is represented by a numerical value ranging from 0 to 3 based on the sorted data. Table 1 shows how this technique works.

Table 1: Illustration of Label Encoding Process

PrizmCode (Before Label Encoding)	PrizmCode (After Label Encoding)
Other	0
Rural	1
Sub Urban	2
Town	3

The dataset used in this study is preprocessed to handle missing values, label encoding, and feature scaling. Missing values are replaced with their respective mean values, and categorical variables are encoded using label encoding. Feature scaling is applied to standardize the data, ensuring that all features contribute equally to the model.

B. Feature Selection

Feature selection is performed using a modified BAT algorithm, which optimizes the selection of relevant features to improve model accuracy. The algorithm evaluates the fitness of each feature and selects the most important ones for the classification task.

The BAT method is a greedy approach similar to Particle Swarm Optimization (PSO) [20], and it has the following interactive parameters that change from iteration to iteration.

C. Classification using HPMANN

The proposed method used a three-step process to iteratively apply a classification algorithm and improved the model's accuracy as shown in figure 1.

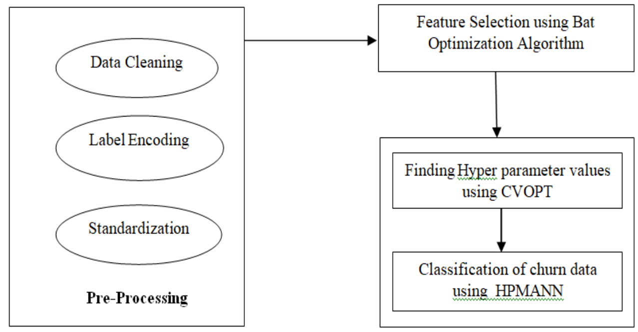

Figure 1:Block Diagram for Predicting Churn Customers

The proposed HPMANN model consists of a 5-layer neural network with one input layer, three hidden layers, and one output layer. The model is trained using the selected features, and hyperparameters are tuned using k-fold cross-validation to prevent overfitting.

IV. MODEL ARCHITECTURE

The architecture of the HPMANN model is designed to handle the complexity of churn prediction in the telecom industry. The model consists of the following components:

System has created a 5-layer network in the suggested system, with one input layer, three hidden levels, and one output layer. All of the layers in this model are dense layers with the following components (which get input from all other neurons in the previous layer) as discussed in table 2.

Table 2: Illustration of Components in Neural Networks

S.No	Name	Description
1	Units (or) Neurons	This parameter is important for controlling the model's capacity and complexity.
2	Activation Function	The neurons perform a transformation function on the combined weights before comparing them to the threshold value. Using activation functions, the neural network learns complicated patterns in a recursive manner.

3	Optimizer	It's an optimizer method that determines how much of a change in learning rate and weights is needed to reduce losses during the neural network model's training.
4	Weight Decay	It has a significant impact on the decrease of overfitting issues in neural networks. The L1L2 regularizer is a proposed model that calculates the "sum of absolute and squared weights."
5	Number of Hidden Layers	A single layer neural network can only solve simple problems, whereas a multi-layered neural network should be used to solve complicated problems.
6	Learning Rate	It is described as the "amount by which the weights are modified during neural network training." Internally, the optimizer controls this parameter.

- **Input Layer:** Receives the preprocessed data and selected features.
- **Hidden Layers:** Three dense layers with activation functions to learn complex patterns in the data.
- **Output Layer:**Produces the final churn prediction (churn or non-churn).

The model uses optimization techniques like Adam and regularization methods like L1L2 to reduce overfitting and improve generalization.

V. EVALUATION

The performance of the HPMANN model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The results are compared with traditional algorithms like SVM and KNN.

Table 3: Performance Metrics Comparison

Metrics/Methods	KNN	SVM	HPMANN
Accuracy	70	71	76
Precision	70	70	72
Recall	69	69	70
F1-Score	70	70	71

The HPMANN model outperforms traditional methods in terms of accuracy, precision, recall, and F1-score, demonstrating its effectiveness in churn prediction.

VI. RESULTS AND DISCUSSION

The accuracy measure is compared using the existing and suggested schemes on the datasets orange and cell2cell, respectively, as shown in Figures. The methods are plotted on the x-axis, while the accuracy is plotted on the y-axis. For the stated telecom datasets, conventional algorithms such as KNN, SVM, and ICHS-ESVM provide lower accuracy, however the suggested HPMANN technique provides higher accuracy. Using the optimal feature selection approach, it considerably lowers the misclassification error rate.

Comparison of KNN,SVM AND HPMANN:

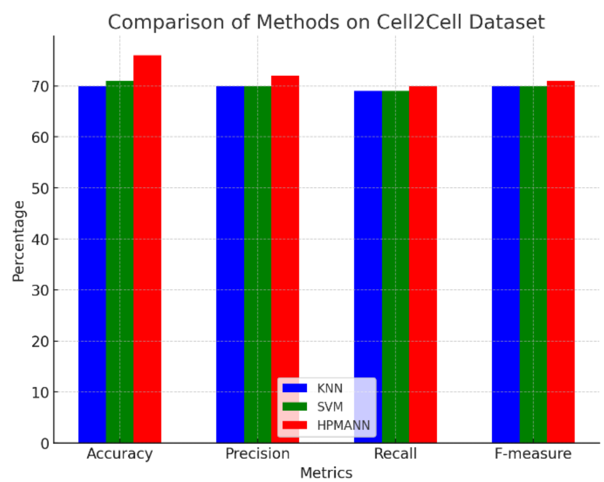

Figure 2: Comparison of Methods

Benchmarking Machine Learning Models Using Heatmap Visualization:

This heatmap visually represents the performance of different machine learning methods (KNN, SVM, and HPMANN) on the Cell2Cell Dataset across four evaluation metrics: Accuracy, Precision, Recall, and F-measure.

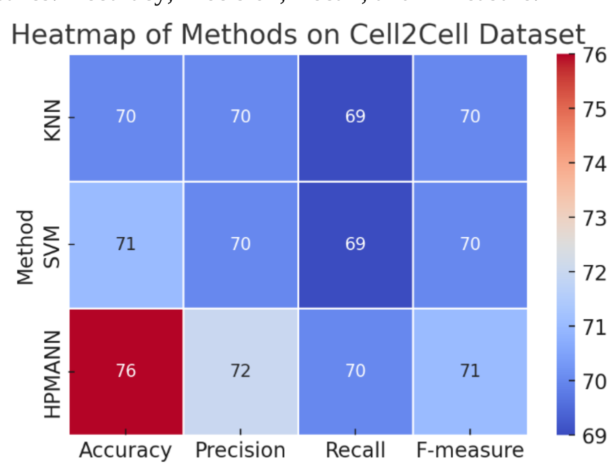

Figure 3: Heatmap comparing KNN, SVM, and HPMANN across all metrics.

The proposed HPMANN model achieves an accuracy of 76%, outperforming traditional methods like SVM and KNN. The model's ability to handle complex patterns in the data and its robust feature selection process contribute to its superior performance. The results indicate that HPMANN is a promising approach for churn prediction in the telecom industry.

VII. CONCLUSION

In the telecom industry, customer churn prediction is critical to churn control. Cell phone service providers must organise churn forecasting patterns that can consistently recognise clients who are going to quit in order to reduce the various costs associated with customer churn. Efficient feature sets are chosen to assist the identification of churners in classification models.

The HPMANN technique is offered as a way to significantly increase churn classification accuracy outcomes for a given telecom dataset. Filling in the missing values, label encoding, and standardisation procedures are used to pre-process the data in this study. It is used to improve the accuracy of churn prediction by removing missing values and redundancy features from a telecom dataset. These features are then considered throughout the feature selection process. The features are chosen using a Bat optimization approach that has been tweaked. Using the best fitness function, it is used to identify the most important and relevant attributes.

VIII. FUTURE SCOPE

A classification technique based on optimization could be developed in the future to handle the huge dataset.New optimized methods for efficient attribute selection on huge datasets can be developed, and improved classification algorithms can be used to efficiently anticipate customer turnover. Additional analysis with different datasets can be performed to expand the approach.

Future work could focus on developing more advanced optimization techniques for feature selection and exploring hybrid models that combine HPMANN with other machine learning algorithms. Additionally, the model could be tested on larger datasets and different industries to evaluate its generalizability.

REFERENCES

[1] Lalwani, P., et al. (2021). "Customer Churn Prognosis in Telecom Using Machine Learning."

[2] Sebastian, J., et al. (2020). "Minimax Probability-Based Churn Prediction in Telecom."

[3] Sergue, A., et al. (2020). "Attribute-Based Churn Prediction in SaaS Platforms."

[4] Bharathi, R., et al. (2020). "Hybrid Models for Customer Retention Prediction in Telecom."

[5] Kostic, S.M., et al. (2020). "Social Graph-Based Churn Prediction in Telecom."

[6] 6.Saias, José, Luís Rato, and Teresa Gonçalves. "An approach to churn prediction for cloud services recommendation and user retention." Information 13.5 (2022): 227.

[7] 7.Anudeep, Ch, et al. "Predicting Customer Churn in E-commerce Subscription Services using RNN with Attention Mechanisms." 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2024.

[8] 8.Belchior, Lúcia Madeira, Nuno António, and Elizabeth Fernandes. "Online newspaper subscriptions: using machine learning to reduce and understand customer churn." Journal of Media Business Studies (2024): 1-24.