Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhanced Customer Churn Prediction for CRM Using PBGL-WPLM and Optimized ANN Models

[1]Corresponding Author*: Dr. N. Senthil Madasamy, [2]Dr. A. Noble Mary Juliet, [3]Dr. J. Bhavithra, [4]J Ramprasath, [5]Dr. Ranjana R

Associate Professor, Department of Computer Science and Engineering,

Dr.Mahalingam College of Engineering and Technology.

[1]senkav1293@gmail.com

Associate Professor, Department of Computer Science and Engineering,

Dr.Mahalingam College of Engineering and Technology,

[2]cse.julie@gmail.com

Department of Computer Science and Engineering,

Dr.Mahalingam College of Engineering and Technology,

[3]bavi.rr@gmail.com

Associate Professor, Department of Information Technology,

Dr Mahalingam College of Engineering and Technology,

[4]jrprasath@gmail.com

Associate professor ,Department of Artificial intelligence and Data Science

Dr.Mahalingam college of engineering and technology

[5]rr.aids@drmcet.ac.in

ARTICLE INFO	ABSTRACT
Received: 23 Oct 2024 Revised: 01 Jan 2025 Accepted: 15 Jan 2025	In today.s digital landscape, the information sector is the dominant service provider, results intense competition among them. To maintain strong market presence, service providers need to concentrate on customer churn, retention and customer satisfaction. Customer relationship management (CRM) strategies were particularly developed to improve , maintain and establish long term relationship with customer. Churn occurs when a customer switches from one service provider to another one. Predicting customer churn is challenging due to the lack of identifying churn reasons for customer churn. As a result, determining the efficiency of a prediction model is also dependent on how well the data can be interpreted in order to identify potential reasons for churn. Here Populace

Based Gradual Learning — Wide and Profound Learning Models(PBGL-WPLM) method, Structure Optimized Simulated Annealing ANN, and Structure Optimized Hybrid SA-PBIL ANN methods are evaluated using the CRM dataset from the American media transmission associations and compared with novel technology called Populace Based Gradual Learning — Wide and Profound Learning Models Network Factorization(PBGL-WPLM-NF). Agitate is determined as the quantity of clients who leave the help somewhere in the range of 61 and 90 days subsequent to being tested.

Keywords: Customer Relationship Management Churn, Artificial Intelligence, Populace Based Gradual Learning-Wide and Profound Learning Models with Matrix Factorization, .

1. INTRODUCTION

Customer Relationship Management (CRM) focuses on increasing, maintaining, and developing long-term customer relationships. CRM is based on the collecting of information prior to making judgments. This never-ending cycle of churning has an impact on the overall profitability and image of the organization. As a result, it is preferable to prevent consumers from churning and instead focus on churn prediction [1].

The lack of a single cause for client turnover poses the most significant barrier to churn prediction. This is usually the result of a combination of several different factors. It is difficult to identify these causes because they are dependent on the organizational services used by consumers as well as the customer's individual perspectives. In this area of knowledge, firms must prioritize early churn prediction, identifying the primary causes of churn, and predicting remedies to prevent churn. All of these steps can be carried out using the data accessible within the organization. However, the nature of the data significantly complicates the prediction procedure. Churn prediction applications include both physical and intangible service-oriented domains, product-based organizations, and telecommunications services.

As a result, determining the efficiency of a prediction model is also dependent on how well the data can be interpreted in order to identify potential reasons for churn. Disturb models hope to recognize early beat signs and clients who are presumably going to uninhibitedly leave. In this study, a methodology for evaluating statistical models for classification using a composite indicator is presented. With substantial study in Artificial Intelligence, it is now possible to get to the root of the problem that causes client churn[2].

Holding existing customers is more savvy than procuring new ones, making it worthwhile for organizations. Clients with long haul experience are more averse to be affected by contenders' way of behaving. Furthermore, positive informal exchange can draw in new clients and they can support to increment organization income. Loss of clients could bring about expanded open door costs. Foreseeing client wear down likewise intends to further develop maintenance. Obviously, holding current clients is more financially savvy than enrolling new ones. The instruments utilized for creating and applying client maintenance models (agitate models) are imperative for BI applications. In a powerful market, low consumer loyalty to new items, regulations, and contest systems can prompt an expanded stir. Early agitate models attempt to recognize shoppers who are bound to willfully leave[3].

The company's success depends on increasing client awareness and retention with contemporary technologies. In today's competitive market, organizations must predict and manage churn to retain consumers and maximize profits. Build an accurate Customer Churn Prediction (CCP) model. The stir models aim to identify clients who exhibit early agitation and are likely to leave on their own. This paper proposes a system for examining measurable models for grouping using a composite marker. Advancements in artificial Intelligence have made it possible to identify the root cause of customer dissatisfaction. Numerous machine discovery calculations have been presented to address the issue of stirring expectations.

The ANN is very helpful and inventive when applied to AI critical thinking. A model of a data director carries out roles like those tracked down in organic diary frameworks. As of late, there has been a flood of interest in the cerebrum's usefulness all over the planet. These classifiers are normally used to characterize people as churners or non-churners. A CRM dataset is utilized to gather information from American telecom firms that center around stir forecasting[4].

TF-IDF is a keen methodology that concludes the repetition reports by picking the overall meaning of all repetitions. It uses phrase go over, and thus, report re-iterate fitting, beside novel understandings by virtue of text. The TF is as yet hanging out there by broadening the TF with the IDF. The TF-IDF has been changed by various arranged specialists; thusly, there are exceptional recipes. To find the immense words connected with client beat in the Voice of Client, this text for churners and non-churners was broke down openly.

The RF will create numerous arrangement trees. To order the item from an information vector, it very well might be put down with each tree viewed as in the woodland. Each tree will give further order as the tree will 'decide in favor of' a certain Class.

ANN is a nonlinear model that is easier to use and understand than traditional quantitative approaches. ANN is a nonparametric model, but other approved procedures demand a more defined, quantifiable foundation. ANN is a black box learning approach that cannot investigate input-yield relationships or address flaws. Several strategies, like highlight decision, have been combined with ANN to adapt to this test. The primary benefit of applying the ANN is that it is equipped to make the models be utilized effectively and precisely in the perplexing normal frameworks that have huge data sources. [21].

Metaheuristics are progressed heuristics that furnish ideal answers for streamlining issues with restricted calculation or information. Factual calculations can precisely expect close optimality in light of authentic perceptions. [5].

Hybrid neural network approaches to churn prediction were presented. The Simulated Annealing (SA) optimization procedure can be used to identify the fewest number of excitations. The term "annealing" derives from metallurgy, and it refers to the process of cooling to reduce material defects. SA is a popular heuristic method for tackling global optimization problems. Optimization is the process of maximizing or minimizing the algorithm's primary target function. The algorithm's convergence is determined by how the problem is programmed and evaluated[6].

2. LITERATURE REVIEW

Domingos et al., fostered the Nick of Time (NIT) procedure, which was a more sensible answer for the issue than other state of the art Client Stir Expectation (CSE) strategies. Be that as it may, the JIT, similar to different models of conventional agitate forecast, requires adequate authentic information. To conquer this hole in customary models, the work used cross-organization information, which is information from another association. This happened as a feature of the JIT's endeavors to fix issues in the broadcast communications industry. The model's presentation was assessed utilizing freely accessible datasets from two telecom administrators.

It was observed that on an empirical evaluation of a NIT-CSE:

(i) Using cross-organization datasets for training allowed for evaluation of prediction model performance

(ii) Manifest the mixed ensemble-based NIT-CSE model was well-suited to this compared to the other individual classifier or homogeneous techniques[7].

Farhad et.al., had proposed the new weka tool that was used for performing different operations on the dataset and for converting data into a nominal form. This is divided into different classes along with cross validation was carry out on the agitate class. The Hoeffding, first and foremost, tree calculation and the Strategic calculation were applied to the dataset. There was a classifier execution evaluator that was utilized for the characterization of information. For noticing the outcomes, a message watcher was utilized, and a graphical portrayal was pushed off to show examinations. The Disarray framework had proposed that the Strategic calculation was better in the forecast of client agitate contrasted with the Hoeffding tree calculation.

Hao et.al., had developed another rule-based model for client churn forecast using the Artificial Neural Network and uneven set theory for purchaser reviews dataset. Using this framework helped in developing intelligent decision support systems.

Höppner et al., fostered the TR-LSTM, a procedure for bearing-based significant design that is utilized to produce assumptions by mining instances of client conduct behind course data. This strategy kills three unmistakable kinds of components that are heading-based and utilize Long and Passing Memory Mind Association (LSTM) in driving successive showings. The discoveries of the preliminaries in view of genuine client-bearing datasets show that the proposed TR-LSTM beats check strategies. The procedure adds one more instrument for scholastics and specialists to outperform assumptions.

Jiang et al., introduced another model called Altered Irregular Woodland (MRF), which is useful in enhancing accuracy and avoiding the issue of relapse. Such an approach is used to arrange orange datasets. For evaluating the suggested computation, conduct a thorough assessment of the present and proposed tactics used in the two situations, such as upselling and stirring. The proposed model is then compared to the model of beat expectation. According to the findings of this study, the proposed model outperforms the current technique, which includes a conventional irregular woodland, in terms of characterisation precision and AUC.

Jung et al., had constructed models of churn prediction that are interpretable and profitable is proposed for selecting the churn model that is profitable. Another new classifier has been presented, integrating an EMPC metric into the construction of the model. The ProfTree makes use of the evolutionary algorithm to learn the decision trees that are profit-driven. For the benchmark work, using real-life datasets from different telecommunication service providers has proved that ProfTree has a significant improvement when compared to a classic method that is accuracy-driven.

In [9], the authors have done a similar examination of client expectation models in the telecom business. They have utilized three best-fit calculations, specifically KNN, RF, and SVM, alongside an improvement calculation for hyperparameter tuning. They have inferred that the essential renditions of these calculations perform less than the mixture (RF with matrix inquiry streamlining calculation) with a low-proportion undersampling technique.

Kumar et al., tended to the client stir issue in the broadcast communications business in light of the fact that the pace of beat is a lot higher than in different ventures (typically running somewhere in the range of 10 and 60 percent). Anticipating client weakening quite a bit early can assist you with keeping up with your current clients. The proposed approach, known as the XGBoost calculation, is a model that beats other state-of-the art calculations.

Nabahirwa et al., proposed the Multi-Objective-Cost-Delicate Subterranean Insect Settlement Streamlining strategy, merging an expense-based and non-overwhelmed hereditary computation highlight choice with MO- ACO cost-sensitive learning. This Excavator was tests revealed that the model worked well in acquiring collector working trademark bend values for churner expectation, which was advantageous in the intensely competitive media communications industry.

Özmen et al., represented hybrid machine learning model structures are often be basic, and they perform well in the specialized data form of a customer churn prediction task. At the same time, the vast majority of them are easily understandable. Researchers frequently use these single machine-learning models as a benchmark.

Praveen et al., delivered EML models are utilized to further develop the agitate forecast model's exactness and generalizability. EML has been found to consolidate the advantages of various AI models with unmistakable structures, lessening individual blunders and bringing about better anticipated exactness than an independent model. As of late, with the fast extension of profound learning, models addressed by profound brain networks have quickly obtained interest in the client agitate expectation field. Notwithstanding their low interpretability, profound brain networks extensively limit counterfeit element designing exertion because of their brilliant versatile component extraction ability.

Ramesh et al. proposed the need for customized and high-return products and services is increasing. Banks must not only seek new revenue streams but also adopt a user-centered business strategy. As a result, competition in the banking business has become increasingly strong, both from established banking institutions competing for consumer resources and from new entrants into the banking market.

Swetha, et al., has evaluated some research has revealed that just a small number of thoughts rush toward the final assumption, which was assessed in expulsion tests using the Thought Head Pruning technique (Hao et al., 2021). This study conducts four social affairs involving expulsion primers, each of which is replicated on separate occasions. The number of thinking heads employed in the four social affairs of expulsion tests is 2, 4, 8, and 16, whereas the other association hyperparameters are consistent.

When contrasted with other single-man-made intelligence models, the supporting method showed an impressive improvement in exactness. Seyed et al. (2021) presented a client-beat-hope model in light of XGBOOST, which achieved an AUC of 0.78.

Since about 2014, the contemplation component has been used to improve learning. The profound learning community has acknowledged it as a critical notion for accelerating profound brain networks because to its ability to forcibly handle data streams and channel low-significant improvements by doling out integrate loads (Tang, Q et Colombini, 2021). Transformer, a key achievement in consideration system research, totally replaces RNN with a self-consideration component for machine interpretation tasks, allowing it to better capture global conditions between information and results.

Cenggoro et al. (2021) made a reasonable client agitate forecast model utilizing the vector implanting strategy in profound learning, and the model had a F1 score of 81.16%. Wael acquainted support learning with the client agitate expectation local area interestingly. Their work zeroed in on the model's strength. DQN is utilized to prepare on the client agitate informational collection, and it is exhibited that DQN, as a functioning student, is more powerful than other AI models when the information design changes.

The poll revealed that machine learning and artificial intelligence play an increasingly important role in customer attrition analysis. Furthermore, machine learning algorithms outperform individual performance. Deep learning algorithms are well suited for image-based reviews and sentiment analysis. Additionally, feature engineering is suggested as a method of model optimization. To aid readers and researchers, this paper conducts a study of the best-fit algorithms for predicting customer turnover using machine learning. The study's findings provide valuable insights into the sector, allowing them to identify customer churn early on and retain clients.

3. METHODOLOGY

The ANN classifiers are utilized for characterizing the churners and the non-churners and have been portrayed here. The ANN functions similarly to the brain sensory system., with interconnected neurons and nodes that solve specific problems. It facilitates message transmission and reception, allowing actions to be taken based on processed information. As a result, Artifical Neural Networks are an effective tool for processing and analysing complex operations similar to those performed by human brains. For the assessment of the various classifiers, here CRM dataset has been utilized . The figure depicts the entire workflow of this study.

Dataset portrayal

The data source for CRM gives information of the American telecom affiliations zeroing in on client beat supposition. It contains 51,306 endorsers, including 34,761 churners and 16,545 non-churners, from July 2021 to January 2022. In the dataset, endorsers in tantamount relationship for a surprisingly long time are viewed as developed clients, and Blend is figured considering the client leaving the help inside 31 to 60 days after the client was examined.

Ascribes like Orientation, Senior Resident, Accomplice, Reliant, residency, Telephone Administration, Different Lines, Network access, Online Security, On the web Reinforcement, Gadget Assurance, Technical support, Streaming television, Streaming Motion pictures, Contract, Paperless Charging, Installment Technique, Month to month Charges, All out Charges are considered for assessing agitate.

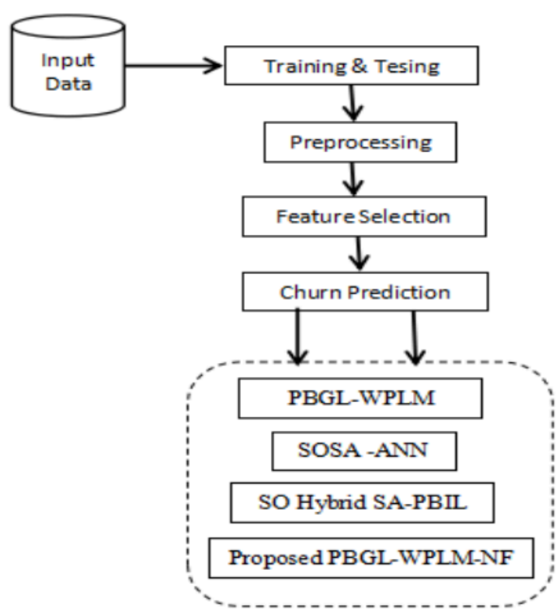


Figure 1. overall workflow

3.1 Structure Optimized Simulated Annealing Artificial Neural Networks

The process of metallurgical annealing along with controlled cooling, which takes effect based on the transitioning probability towards a solution that may be worse. This probability is directly get affected by temperature. If the temperature is high, possibility of it shifting towrads a worse solution increases. Due to this feature, the SA could traverse the entire search space. Compared to the Swarm Intelligence and Genetic algorithms, the SA does not face the premature convergence problem.

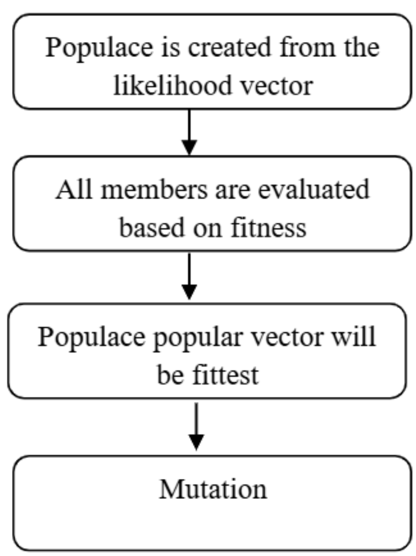
The SA will work on one single probable solution (called the state) and will attempt to get a better solution by means of improving it. Such improvement may be conducted by generating another new successor in its current neighborhood and further transitioning it in a probabilistic manner. If S indicates the current state and the successor or neighbor generated based on the current state, the move from S to S will take place based on the fitness function: in case the fitness for S is better than that of S, the transition will take place and if not it may occur. In case the fitness of the successor is lower than its current state, the transition probability will correspond to the temperature and fitness value.

3.2 Structure Optimized Hybrid SA-PBIL ANN

Structure Optimized Hybrid SA-PBIL ANN is combination of Simulated Annealing (SA) and Population-Based Incremental Learning (PBIL) to enhance and facilitate the structure and performance of artificial neural network. This method used to avoid the local optima and helps to getover the obstacles of traditional ANN. This strong structure supports to predict the complex patterns and classifications. The experimental setup of this model entails iteratively adjusting the ANN structure and hyper parameters using optimization techniques. This iterative process is repeated with appropriate network configuration until a satisfactory solution is obtained .

3.3 Populace Based Gradual Learning(PBGL)

The PBGL strategy utilizes transformative figuring out how to assess the dispersion calculation. This populace based approach for steady learning falls under the classification of metaheuristics. Recognize close ideal answers for complex, NP-difficult issues. In huge and discrete spaces, addressing issues can take quadratic time as the quantity of factors increments. This kind of GA includes expressly dealing with a populace's measurements, bringing about the development of the whole populace instead of individual individuals.



PBGL outperformed standard GAs and real-time benchmarks such as power system development and stabilizers using incremental learning. It also successfully trained models for responsive stabilizers with variations to patterns. The basic PBGL algorithm ideas are listed below.

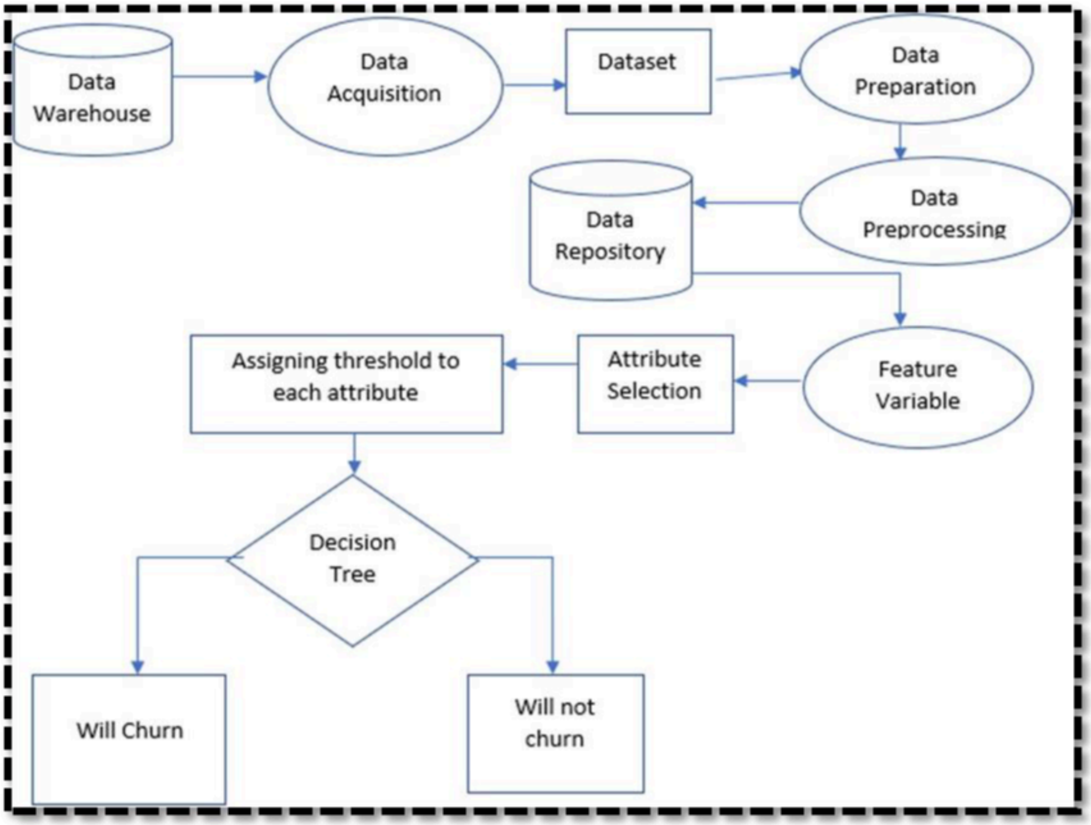


Figure 3. Framework of customer churn

Setting up an solution structure. A solution structure on which a PBGL calculation will work, similar to the Hereditary Calculation. The issue's choice factors have been recognized and encoded as neighboring parallel sub-strings that structure a total string known as the arrangement vector. The genuine idea of these decision factors can influence the length of a similar string. It could be sufficient on the off chance that the choice variable, which can take on values going from 0 to 100, has a substring of seven double digits. The course of action vector will be arbitrarily designated matched values (0 or 1). Decoding the sub-strings and assigning appropriate values to the variables will result in a single answer. Analysts must select a size for the finite solution vector set, which represents the population of feasible solutions.

Evaluation Function. A metaheuristic-suggested solution is evaluated using an evaluation function. The function is usually mathematical, however in the case of dynamic or complicated stochastic systems, simulation models are employed. The function should be decreased or maximized.

Likelihood vector. This design contains a few components, including arrangement vectors. Dissimilar to twofold whole numbers, every component will have a likelihood value. The component esteem addresses the certified likelihood of a particular digit in a similar arrangement vector, which could incorporate '1'. A low value in component I shows an unfortunate likelihood of tracking down the digit '1' in an answer vector. The likelihood vector will act as the basis for the PBIL calculation. Vector components are introduced with a likelihood of around 0.5 each. An answer vector can change the items in a component throughout a cycle.

Mutation: This is normally propagated at a likelihood of 0.02. “Likelihood (0 or 1)" can refer to either '0' or '1' with a chance of 50%.

Convergence. If the probability algorithm converges, all elements lies between 0 and 1. Limit settings are determined to forestall superfluous emphasess. In the event that the component arrives at a worth underneath 0.05, it shows combination to nothing. In the event that the component has a worth higher than 0.95, it has united to 1. Assuming any of these circumstances are met, the calculation will stop.

3.4 The Simulated Annealing (SA) algorithm

Simulated Annealing can be used to obtain the near-global optimal value for any given function. This name is derived from the annealing technique, which involves heating and controlled cooling of crystals to reduce defects and achieve a state of stability and strength. Simulated Annealing (SA) is an upgraded version of the traditional local search approach. A local search heuristic gradually improves the initial result by incorporating little changes, such as modifying a single variable or exchanging two values.

Starting with the initial solution, a trial solution from the neighborhood set will be chosen and replaced if it outperforms the current one. This process is known as the neighborhood move. However, this approach may lead to a local optimum rather than a global one.

An underlying reproduced strengthening temperature will be given.

• The populace size (N) is set, underlying population is produced aimlessly. To encode chromosomes, A decimal value will be used. The quantity of factors to advance is not set in stone by the quantity of qualities on every chromosome. The accuracy of the iterative number M is described.

• The populace's chromosomes are assessed, and their wellness capability is assessed by ANN, preparing for the goal capability's base streamlining issue.

3.5 PBGL-WPLM algorithm

This paper presents an algorithm for predicting churn using a wide and deep learning architecture. The Rectified Linear Unit is the nearly all often employed commencement meaning in deep learning models, and it also serves as the classification function. This approach is capable of compounding the benefits of memory and generalization with less feature engineering, suitable for evaluating telecom sector data. Data is automatically analyzed to identify features and capture non-linear or context-dependent correlations.

The features were processed by a broad and Deep Learning Model Network which comprises of a extensive module and a deep component, as illustrated in Figure .1. The wide component is a generalized linear model designed for large-scale regression and classification. Memorizing feature interactions adds non-linearity to a generalized linear model. The deep neural network transforms high-dimensional features into dense embedding vectors.

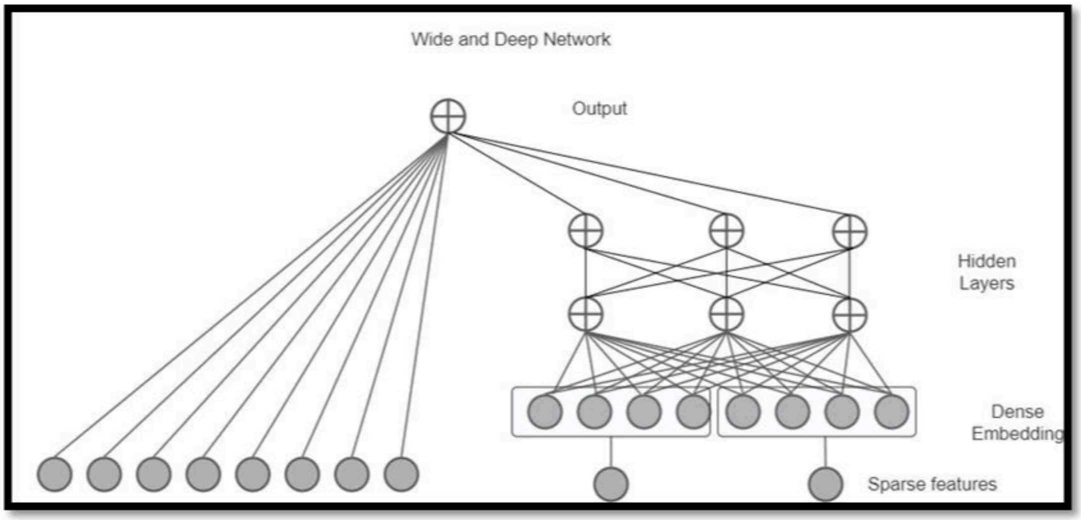


Figure 4. Deep and Wide Learning Network

This component of the WPLM Network consisted of two types of layers: embedding and concealed layers. Embedding layers contained two embeddings that corresponded to two types of features: churn and no-churn. All features were placed in the broad part, which contained the crossing features and was combined with the output of the deep part in the final layer to generate a vector. The ReLU served as the activation function in subsequent layers. The embedding vectors are created at random, and during training, the values that minimize the final loss function are discovered. The dense embedding vectors are fed through the network to generate the output. A weighted sum is used to combine wide and deep networks and generate predictions for joint training.

The moderate-layer design of the brain network has been addressed, utilizing every lattice section. The underlying and last layers of the Brain networks were set as per the information datasets and their sizes. Hence, to recognize the ideal design of the dataset, there were particular likelihood vectors were analyzed. The PBIL procedure proposed had been settled inside the internal and external circles. This internal circle was utilized to recognize the preparing mistakes in the applied design. This external circle was utilized for streamlining the design of the brain organization. Inside the external layer, the worldwide best will be taken to be the molecule that has a base mean square blunder esteem, and Every likelihood vector has a related parallel population. The inward circle has the designs prepared utilizing the PBIL. The organization has a moderate layer with a limit of 9 and at least 2 hubs.

The deep component is optimized with the following steps:

- Initialize probability vector
- Generate random solutions.
- When termination conditions are not met, evaluate solutions and select the best one.
- Update or mutate probability vector.

3.6 Proposed Populace Based Gradual Learning — Wide and Profound Learning Models – Network Factorization (PBGL-WPLM-NF)

The PBGL-WPLM produces only the mutate probability vector, where as, it doesn't provide consumer's network to monitor the dissatisfaction. So the Network factorization is applied , to find the an impetus, empowering the framework to measure the client's careful motivation behind the buy, check various pages, waitlist, and rank the right item or administration, and suggest numerous choices accessible. When the result matches the necessity, the lead converts into an exchange and the arrangement clicks.

This mathematical model helps the structure by disengaging a section into various extra veritable entries through a planned rectangular social occasion of numbers or works to track down the components or information central to the relationship among clients and things.

PBGL-WPLN-NF:

Expecting to be that the suggestion framework incorporates m clients and n wares, $Rc\times p = [R_{ij}]c\times p$ addresses the client service attain network. Rjk addresses the rating of client j to thing k, where $Rjk \in [1,4]$. For the most part, there many void components in $Rc\times p$, furthermore, it will cause a fights. The connection among clients and items can be addressed by a lattice C: C= [Cjk]cxp, the worth of C_{jk} = 0 or 1, and it is 0, it intends that there is no connection between clients. This can be represented in the below Table 1.

Table 1: Customer and commodity scoring and relationship matrix

Customer-commodity scoring matrix					Customer-commodity relationship matrix				
	C1	C2	C3	C4		C1	C2	C3	C4
P1		5		3	P1		1		1
P2	4		2		P2	1		1	
P3	5			1	P3	1			1
P4	4		4		P4	1		1	

Network $Rm\times n$ can be factored into client include grid $C_{m\times k}$ and thing highlight $P_{k\times n}$, separately. k addresses the dimension of vector; as a general rule, it is a lot more modest than m and n, and afterward dimensionality decrease can be understood. Ci and Pj addresses the potential component spaces of comparing clients Ci and things Pj, individually, invalid worth in scoring lattice can be checked through CP and some time later, the figure scoring association could be obtained.

$$p(U \mid \sigma_U^2) = \prod_{i=1}^{m} \mathcal{N}(U_i \mid 0, {\sigma_U}^2 I \quad) \qquad (1)$$

In standard proposition estimation, clients are liberated from all, it dismisses the clients' communal connections. If communal association between clients occurs, tendencies of the clients or their alternative of things will impact with everyone. Subsequently, it is vital to facilitate social associations into the idea estimation; thusly, proposition precision will be incredibly gotten to the next level. The contingent dissemination of noticed social relationship can be characterized:

$$p(C \mid \mathrm{P}, Z, \sigma_C^2) = \prod_{i=1}^{m} \prod_{k=1}^{m} \mathcal{N}\left[\left(c_{ik} \mid g\left(U_i^T Z_k\right), \sigma_C^2\right)\right]^{I_{ik}^C} . \# \qquad (2)$$

Assume that P and U follow the round Gaussian earlier conveyance with mean 0 :

$$\mathcal{L}(R,C_1,C_2,U,V,Z_1,Z_2) = \frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\left(r_{ij} - g\left(U_i^T V_j\right)\right)^2 + \frac{1-\lambda_C}{2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^{C}\left(C_{ik}^{*} - g\left(U_i^T Z_k\right)\right)^2 \# \quad (7)$$

$$+\frac{\lambda_U}{2}\parallel U \parallel_F^2 + \frac{\lambda_V}{2}\parallel V \parallel_F^2 + \frac{\lambda_Z}{2}\parallel Z \parallel_F^2, \#$$

$$p(U \mid \sigma_U^2) = \prod_{i=1}^{m} \mathcal{N}(U_i \mid 0, \sigma_U^2 I), \quad (3)$$

Then, through simple Bayesian inference, the following results can be obtained:

$$p(P, Z \mid C, \sigma_C^2, \sigma_U^2, \sigma_Z^2) \propto p(C \mid U, Z, \sigma_C^2) p(U \mid \sigma_U^2) p(Z \mid \sigma_Z^2) \quad (4)$$

Here, N(x|μ, σ2) exhibits that x follows a Gaussian course whose mean is μ and change a record capacity, and if client I has a score of thing j, its worth is 1; regardless, it is 0. Expect to be C and P submit to the round Gaussian scattering before mean 0.

$$\ln p(U,V,Z \mid C,R,\sigma_C^2,\sigma_R^2,\sigma_U^2,\sigma_V^2,\sigma_Z^2) = -\frac{1}{2\sigma_R^2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\left(r_{ij} - g\left(U_i^T V_j\right)\right)^2 - \frac{1}{2\sigma_C^2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^{C}\left(c_{ik}^{*} - g\left(U_i^T Z_k\right)\right)^2$$
$$-\frac{1}{2\sigma_Z^2}\sum_{k=1}^{m} Z_k^T Z_k - \frac{1}{2}\left(\left(\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\right)\ln \sigma_R^2 + \left(\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\right)\ln \sigma_C^2\right) \quad (5)$$
$$-\frac{1}{2}(m l\ln \sigma_U^2 + n l\ln \sigma_V^2 + m l\ln \sigma_Z^2) + C$$

where C is a consistent which relies upon no limit, and the most outrageous back scattering capacity should be identical to the base objective ability, which is according to the accompanying:

$$\ln p(U,V,Z \mid C,R,\sigma_C^2,\sigma_R^2,\sigma_U^2,\sigma_V^2,\sigma_Z^2) = -\frac{1}{2\sigma_R^2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\left(r_{ij} - g\left(U_i^T V_j\right)\right)^2 - \frac{1}{2\sigma_C^2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^{C}\left(c_{ik}^{*} - g\left(U_i^T Z_k\right)\right)^2$$
$$-\frac{1}{2\sigma_Z^2}\sum_{k=1}^{m} Z_k^T Z_k - \frac{1}{2}\left(\left(\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\right)\ln \sigma_R^2 + \left(\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}\right)\ln \sigma_C^2\right) \quad (6)$$
$$-\frac{1}{2}(m l\ln \sigma_U^2 + n l\ln \sigma_V^2 + m l\ln \sigma_Z^2) + C$$

where λC ∈ [0, 1] is used to change the effect of the client scoring structure and social relationship matrix on the proposition result. When λC ∈ 1, it implies that the social connection between clients isn't thought of , when λC ∈ 0, it implies that the client scoring lattice has an ace piece of 0, and the rest implies that a social relationship is coordinated. λC ∈ σ2 x/σ2 C, λU∈ σ2 X/σ2 U, λV ∈ σ2 R/σ2 V, λZ ∈ σ2 X/ σ2 Z, ‖·‖2 F represents regularization.

The gradient descent algorithm can be used to solve the objective function as follows:

$$\frac{\partial \mathcal{L}}{\partial U_i} = \lambda_C \sum_{j=1}^{n} I_{ij}^{R} g'(U_i^T V_j)\big(g(U_i^T V_j) - r_{ij}\big)V_j + \beta(1-\lambda_C)\sum_{j=1}^{m} I_{ik}^{C^1} g'(U_i^T Z_k^1)\big(g(U_i^T Z_k^1) - C_{ik}^{1*}\big)Z_k^1$$
$$+(1-\beta)(1-\lambda_C)\sum_{j=1}^{m} I_{ik}^{C^2} g'(U_i^T Z_k^2)\big(g(U_i^T Z_k^2) - C_{ik}^{2*}\big)Z_k^2 + \lambda_U U_i$$
$$\frac{\partial \mathcal{L}}{\partial Z_k^1} = \lambda_C \sum_{i=1}^{m} I_{ik}^{C^1} g'(U_i^T Z_k^1)\big(g(U_i^T Z_k^1) - C_{ik}^{1*}\big)U_i + \lambda_Z^1 Z_k^1 \quad (8)$$
$$\frac{\partial \mathcal{L}}{\partial Z_k^2} = \lambda_C \sum_{i=1}^{m} I_{ik}^{C^2} g'(U_i^T Z_k^2)\big(g(U_i^T Z_k^2) - C_{ik}^{2*}\big)U_i + \lambda_Z^2 Z_k^2$$

DJwhere $g'(x) = ex/(1 + ex)^2$ represents derivative of logistic function g(x). For reducing model complexity, the corresponding parameter setting is Λu= λV= λ1 Z = λ2 Z.

4. RESULTS AND ANALYSIS:

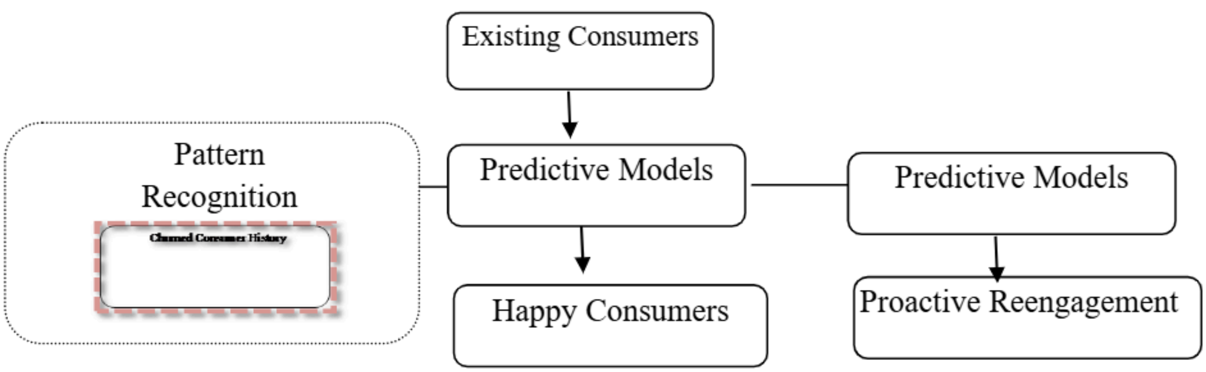


Figure 5. Churn rate Prediction with AI

Setup:

The CRM dataset was used to assess the performance of PBGL WPLM Network, Structure Optimized Simulated Annealing ANN, and Structure Optimized Hybrid SA-PBIL ANN and proposed method PBGL WPLM-NF. Churn is calculated based on customers abandoning the service between 31 and 60 days after sampling.

Churn Rate can be calculated as, total customer lost / Total number of customer at the starting time * 100;

Profit rate can be measures as, total number of customer at beginning-end of the month- new customers / total number of customers at beginning.

Results:

The trials were repeated 15 times, with the average findings shown. The average standard deviation was approximately 2.8%. Table 2 presents a summary of the results. Figures 6 to 10 display classification accuracy, recall, precision, and f measure for both churn and no-churn cases.

Table 2. summary of the results

	SO-SA ANN	SO-PBIL ANN	PBGL WPLM Network	PBGL- WPLM -NF
Classification Accuracy (%)	89.70	90.51	92.13	93.21
Recall for Churn (%)	91	92	94	96
Recall for No Churn (%)	82	86	92	97
Precision for Churn (%)	95	94	98	99
Precision for No Churn (%)	86	87	89	91
F-Measure for Churn (%)	93	94	96	97
F-Measure for No Churn (%)	84	86	92	93

The following graphical representation depicts the evaluation of models performance.

Where class.Acc - Classification Accuracy, R C - Recall for Churn, R N C - Recall for No Churn, P C - Precision for Churn, P N C - Precision for No Churn, F-M C - F-Measure for Churn,

F-M N C - Measure for No Churn. Figures 7, 8, 9, 10, individually compared based on Accuracy, Precision, Recall, F1 Score.

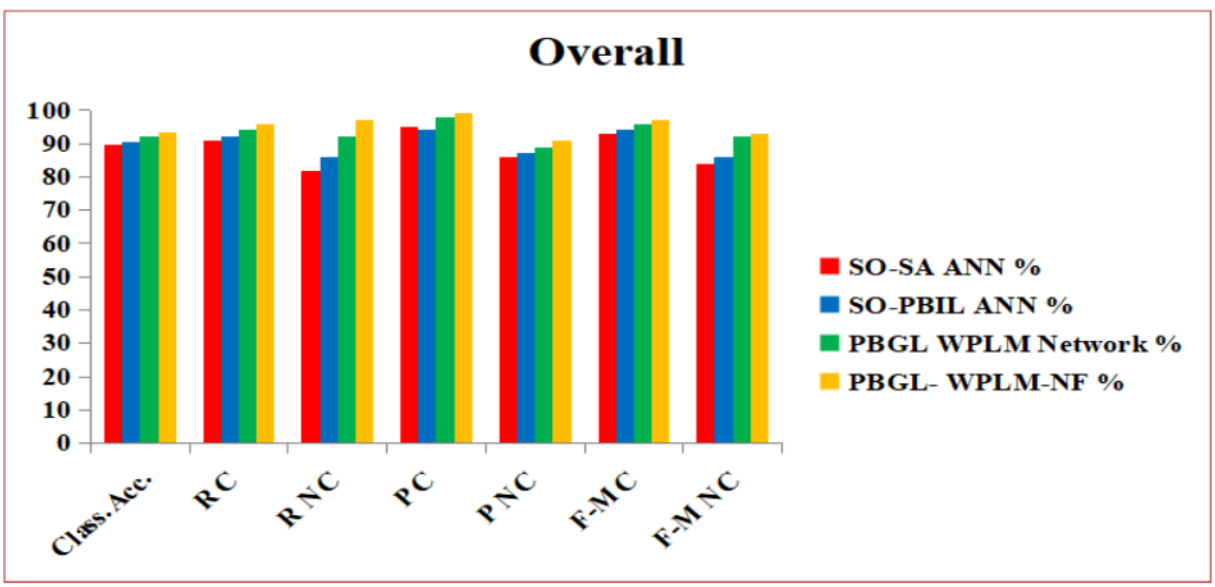


Figure 6. Overall model performance

Figure 7. Accuracy of experimented models

Accuracy (ACC): This implies the degree of results (TP, as well as TN) which was among the total broke down events. The precision that is the best will be 1, and the most clearly terrible will be 0. The proposed PBGL WPLM NF network achieves higher accuracy than others.

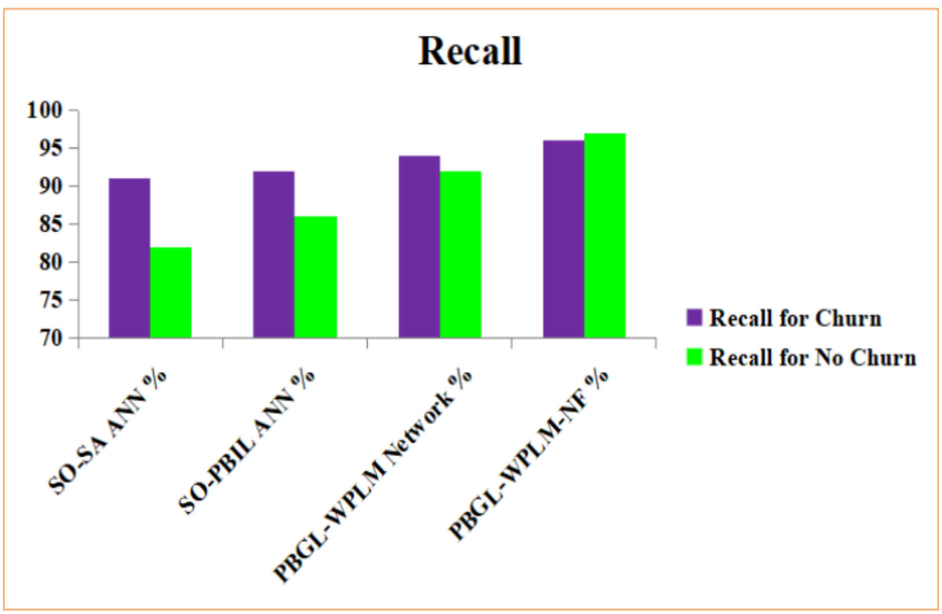


Figure 8. Recall for experimented models

Recall (Genuine Positive Rate/Awareness): This is processed as the positive expectations that were partitioned by the absolute up-sides. The best review for this is 1, and the most exceedingly awful is 0. Figure 8, shows that the PBIL WDLM NF network outperforms than other models.

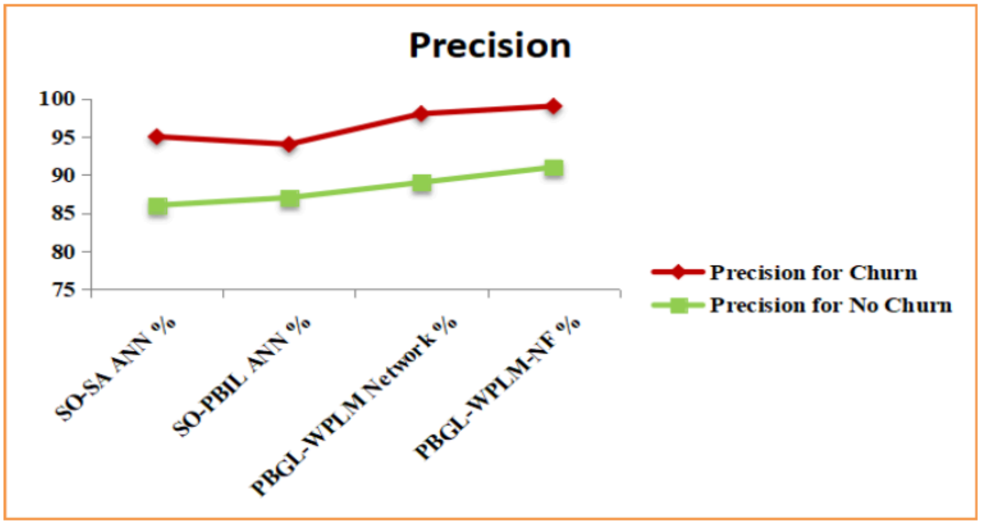


Figure 9. Precision for experimented models

Precision (the Positive Farsighted Worth): This is enlisted as the genuine number of right specific conjectures that are isolated by the complete positive gauges. The best is 1, and the most ridiculously dreadful is 0,

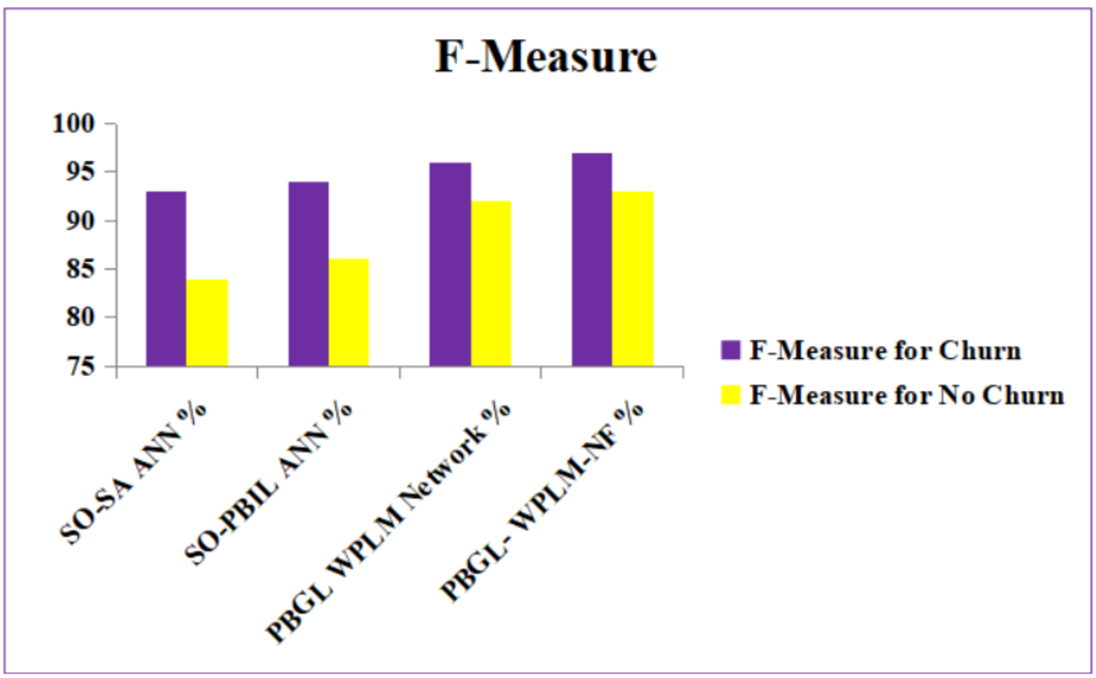


Figure 10. F-Measure for PBGL –WPLM Network

F-measure: This is characterized as the real weighted consonant mean for accuracy and review. Typically, this is utilized for consolidating both Review and Accuracy into one single measure to contrast with various ML calculations. From the figures 6 to 10, this study observed that the proposed system PBGL WPLM NF Network outperforms other existing systems SO-SA ANN, SO-PBIL ANN, PBGL WPLM Network.

CONCLUSION:

Today's technological developments are constantly changing the world and influencing peoples preferences. To remain competitive in the market, business must understand the customer preferences and behavior toward their products or services. The aforementioned study is resource expensive, and needs strong support in terms of experimental skills and infrastructure. The proposed method utilized various deep learning architectures and aims to analyse various AI driven models. As a result, this study explored CRM dataset to predict customer churn preferences. The resultts were dissected that PBGL WPLM NF network achieved accuracy of 93.21%, recall for churn 96%, precision for churn 99%, f-measure for churn 97% than other models SO-SA ANN, SO-PBIL ANN, PBGL WPLM Network. The accuracy of other models SO-SA ANN is 89.70%, SO-PBIL ANN 90.51%, PBGL WPLM Network is 92.13% is smaller than proposed model. Investigating customer churn rate is tie consuming and it needs proper guidance from the dataset. To determine its effectiveness, the proposed method should be tested again using real time data. The use of CRM in various domains such as banking and e-commerce would be beneficial.

REFERENCES

[1] Ahmed, M, Afzal, H, Siddiqi, I, Amjad, MF & Khurshid, K 2020, 'Exploring nested ensemble learners using overproduction and choose approach for churn prediction in telecom industry', Neural Computing and Applications, vol. 32, no. 8, pp. 3237-3251.

[2] Amin, A, Al-Obeidat, F, Shah, B, Al Tae, M, Khan, C, Durrani, HUR & Anwar, S 2020, 'Just-in-time customer churn prediction in the telecommunication sector', The Journal of Supercomputing, vol. 76, no. 6, pp. 3924-3948.

[3] Bhatnagar, A & Srivastava, S 2020, 'Performance Analysis of Hoeffding and Logisitic Algorithm for Churn Prediction in Telecom Sector', In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM) pp. 377-380.

[4] Cenggoro, T. W., Wirastari, R. A., Rudianto, E., Mohadi, M. I., Ratj, D., & Pardamean, B. (2021). Deep learning as a vector embedding model for customer churn. Procedia Computer Science, 179 , 624–631.

[5] Correia, A. d. S., & Colombini, E. L. (2021), " Attention, please! a survey of neural attention models in deep learning", arXiv preprint arXiv:2103.16775

[6] Dalli, A. (2022). Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. Mathematical Problems in Engineering, 2022

[7] Dhote, S, Vichoray, C, Pais, R, Baskar, S & Shakeel, PM 2020, 'Hybrid geometric sampling and AdaBoost based deep learning approach for data imbalance in E-commerce', Electronic Commerce Research, vol. 20, no. 2, pp. 259-274.

[8] Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. Computation, 9 , 34.

[9] Edvaldo Domingos, Blessing Ojeme, Olawande Daramola 2021, "Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector", Computation, Vol. 9 no. 3, pp. 34-41.

[10] Farhad Shaikh, Brinda Pardeshi, Ajay Jachak, Akash Bendale, Nandakishor Sonune, Mangal Katk ar 2021, "Customer churn prediction using nlp and machine learning: an overview" International Journal Of Advance Scientific Research And Engineering Trends, Vol. 6, no. 2, pp. 40-45

[11] Hao, Y., Dong, L., Wei, F., & Xu, K. (2021). Self-attention attribution: Interpreting information interactions inside transformer. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 12963–12971). volume 35.

[12] Höppner, S, Stripling, E, Baesens, B, vanden Broucke, S & Verdonck, T 2020, 'Profit driven decision trees for churn prediction', European journal of operational research, vol. 284, no. 3, pp. 920-933.

[13] Jiang, W., Luo, Y., Cao, Y., Sun, G., & Gong, C. (2020). On the build and application of bank customer churn warning model. International Journal of Computational Science and Engineering, 22 , 404–419.

[14] Jung, H., & Lee, B. G. (2020). The impact of transforming unstructured data into structured data on a churn prediction model for loan customers. KSII Transactions on Internet and Information Systems (TIIS), 14(12), 4706-4724.

[15] Kumar, H & Yadav, RK 2020, 'Rule-Based Customer Churn Prediction Model Using Artificial Neural Network Based and Rough Set Theory', In Soft Computing: Theories and Applications, pp. 97-108, Springer, Singapore

[16] Li, D, Deng, L, Su, Q & Song, Y 2020, 'Providing a guaranteed power for the BTS in telecom tower based on improved balanced owl search algorithm', Energy Reports, vol. 6, pp. 297-307.

[17] Nabahirwa Edwine, Wenjuan Wang, Wei Song, Denis Ssebuggwawo 2022, "Detecting the risk of customer churn in telecom sector: a comparative study", Mathematical Problem Engineering 2022 , ID: 8534739, 16 pages.

[18] Navid Khaledian, Farhad Mardukhi 2022, "CFMT: a collaborative filtering approach based on the nonnegative matrix factorization technique and trust relationships" J. Ambient Intell. Hum. Comput. 2022, pp. 1-17

[19] Özmen, M, Aydogan, EK, Delice, Y & Toksari, MD 2020, 'Churn prediction in Turkey's telecommunications sector: A proposed multiobjective–cost-sensitive ant colony optimization', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 1, p. e1338.

[20] Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, Pratyush Sethi 2022, "Customer churn prediction system: a machine learning approach" Computing (2022), pp. 1-24.

[21] Rai, S, Khandelwal, N & Boghey, R 2020, 'Analysis of Customer Churn Prediction in Telecom Sector Using CART Algorithm', In First International Conference on Sustainable Technologies for Computational Intelligence, pp. 457-466. Springer, Singapore.

[22] Ramesh, P, Jeba Emilyn, J & Vijayakumar, V 2021, 'Hybrid Artificial Neural Networks using Customer Churn Prediction', Wireless Personal Communications, vol. 124, issue no. 2, pp. 1695-1709

[23] Seyed Mohammad Sina Mirabdolbaghi, Babak Amiri 2022, " Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions",Discrete Dynam Nat. Soc., 2022, ID 5134356, 20 pages

[24] Sharma, T, Gupta, P, Nigam, V & Goel, M 2020, 'Customer Churn Prediction in Telecommunications Using Gradient Boosted Trees', In International Conference on Innovative Computing and Communications, pp. 235-246, Springer, Singapore.

[25] Sri Sai Surya P, Anitha K, 2022, "Comparative analysis of accuracy and prediction of customer loyalty in the telecom industry using novel diverse algorithm" ,2022 International Conference on Business Analytics for Technology and Security (ICBATS), IEEE- 2022, pp. 1-7

[26] Swetha, P & Dayananda, RB 2020, 'Customer Churn Prediction and Upselling using MRF (Modified Random Forest) technique', International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, vol. 9, no. 3.

[27] Tang, Q, Xia, G, Zhang, X & Long, F 2020, 'A Customer Churn Prediction Model Based on XGBoost and MLP', In 2020 International Conference on Computer Engineering and Application (ICCEA), pp. 608-612.

[28] Wael Fujo, S., Subramanian, S., & Ahmad Khder, M. (2022). Customer churn prediction in telecommunication industry using deep learning. Information Sciences Letters, 11 , 24.

[29] T. Korpinar, Z. Korpinar, M. Inc, and D. Baleanu, "Geometric phase for timelike spherical normal magnetic charged par- ticles optical ferromagnetic model," Journal of Taibah Uni- versity for Science, vol. 14, no. 1, pp. 742–749, 2020.

[30] Bahrami, S & Doulati, F 2016, 'Prediction of pyrite oxidation in a coal washing waste pile using a hybrid method', coupling artificial neural networks and simulated annealing (ANN/SA). J Clean Product 2016; vol. 137, pp. 1129–37.

[31] Rere, L.R.; Fanany, M.I.; Arymurthy, A.M. Simulated Annealing Algorithm for Deep Learning. *Procedia Comput. Sci.* **2015**, *72*, 137–144.

[32] Thamilarasi V., Roselin. R., "Application of Machine Learning in chest X-Ray images", International book & River publisher series in computing and information science and Technology, pp: 52-1,2023. ISBN : 9788770228114, 9788770228107.

[33] Sivasankar, E & Vijaya, J 2019, 'Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network', Neural Computing and Applications, vol. 31, no. 11, pp. 7181-7200.

[34] https://www.kaggle.com/datasets/shubh0799/churn-modelling