

**INSTITUTE OF ENGINEERING & MANAGEMENT**

**SALT LAKE, KOLKATA 700091**

# NATURE INSPIRED BASED FEATURE SELECTION FOR PREDICTION OF CUSTOMERS CHURN RATE [A MACHINE LEARNING APPROACH]

SUBMITTED BY

1. **RAJESH BURNWAL**
2. **MEHWISH TANWEER**

***Thesis submitted for the partial fulfillment of the requirements for the degree***

***of***

### BACHELOR OF TECHNOLOGY



**(College Logo) (University Logo)**

**COMPUTER SCIENCE & ENGINEERING (INTERNET OF THINGS, CYBER SECURITY & BLOCKCHAIN TECHNOLOGY) DEPARTMENT**

**INSTITUTE OF ENGINEERING & MANAGEMENT**

**MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY,**

**WEST BENGAL**

**2024**

NATURE INSPIRED BASED FEATURE SELECTION FOR PREDICTION OF CUSTOMERS CHURN RATE [A MACHINE LEARNING APPROACH]



A thesis submitted by

**Rajesh Burnwal (**22021002017003**)**

**Mehwish Tanweer (**22021002017011**)**

###### Supervisor

Prof. Rabi Narayan Behera

Submitted for the partial fulfillment of the requirements for the degree of Bachelor of Computer Science and Engineering

**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING (IOTCSBT) INSTITUTE OF ENGINEERING & MANAGEMENT, KOLKATA**

April 2024

Thesis Approval/ Dissertation Approval/Project Report Approval for B.Tech.

This thesis/dissertation/project report entitled Nature Inspired Based Feature Selection For Prediction Of Customers Churn Rate-A Machine Learning Approach by Rajesh Burnwal and Mehwish Tanweer are approved for the degree of Bachelor of Technology in Computer Science and Engineering with Specialization in Internet of Things, Cyber Security and Blockchain Technology.

Date: Place:

###### Examiner(s)

1………………………………

2……………………………….

## ACKNOWLEDGEMENT

I wish to express my heartfelt gratitude to the all the people who have played a crucial role in the research for this project, without their active cooperation the preparation of this project could not have been completed within the specified time limit.

I am thankful to my project guide Prof. Rabi Narayan Behera who supported me throughout this project with utmost cooperation and patience and for helping me in doing this Project.

I am also thankful to our respected Head of the Department, Prof. Dr. Moutushi Singh, for motivating me to complete this project with complete focus and attention.

I am thankful to my department and all my teachers for the help and guidance provided for this work.

I extend my sincere thanks to my institute, the Institute of Engineering and Management, Kolkata for the opportunity provided to me for the betterment of my academics.

……………………..........

……………………………

(Signatures)

Rajesh Burnwal (22021002017003)

Mehwish Tanweer (22021002017011)

Date:

**TABLE OF CONTENTS**

Article No. Content Page No

1. [INTRODUCTION 3](#_bookmark0)
   1. [BACKGROUND 3](#_bookmark1)
   2. [PROBLEM STATEMENT 4](#_bookmark2)
   3. [AIMS AND OBJECTIVES 4](#_bookmark3)
2. [LITERATURE REVIEW/ RESEARCH GAP ANALYSIS 5](#_bookmark24)

#### [ADVANTAGE OF PROPOSED TECHNIQUE OVER THE EXISTING…….7](#_bookmark25)

1. [MATERIAL AND METHODS 8](#_bookmark30)
   1. [DATA COLLECTION 8](#_bookmark31)
   2. [MACHINE LEARNING METHODS 9](#_bookmark35)
      1. [DATA PRE-PROCESSING 9](#_bookmark36)
      2. [FEATURE ANALYSIS 9](#_bookmark36)
      3. [FEATURE SELECTION APPROACH 9](#_bookmark36)
         1. COCCO SEARCH ALGORITHM 10FIREFLY ALGORITHM 10
         2. GREY WOLF ALGORITHM 11
      4. DATA SPLITTING 11
      5. MODEL: SUPPORT VECTOR MACHINE 12
      6. EVALUATION 12
2. [RESULTS AND ANALYSIS 13](#_bookmark38)
   1. [COCCO SEARCH ALGORITHM WITH SVM 13](#_bookmark39)
      1. [ACCURACY 13](#_bookmark40)
      2. [RESULTS 13](#_bookmark41)
   2. [FIREFLY ALGORITHM WITH SVM 14](#_bookmark43)
      1. [ACCURACY 14](#_bookmark44)
      2. [RESULTS 14](#_bookmark41)
   3. GREY WOLF ALGORITHM WITH SVM 15
      1. [ACCURACY 15](#_bookmark44)
      2. [RESULTS 15](#_bookmark41)
   4. OVERALL RESULT AND ACCURACY 16
   5. PREDICTION OF CHURN RATE 17
3. [DISCUSSION AND CONCLUTION 18](#_bookmark46)
   1. [CONCLUSION 18](#_bookmark47)
4. SUMMARY, PUBLICATIONS AND FUTURE WORK[S 20](#_bookmark38)
   1. [SUMMARY 20](#_bookmark52)
   2. [FUTURE WORK 21](#_bookmark52)

[REFERENCES 22](#_bookmark58)

ix

## ABSTRACT

This study investigates the impact of nature-inspired feature selection techniques on predicting customer churn rates within the telecommunications sector. Utilizing the IBM Telco customer churn dataset, our goal is to construct a robust churn prediction model employing feature based machine learning methods. We adopt a Support Vector Machine (SVM) as the primary classifier for churn prediction. To enhance the predictive capabilities of the SVM model, we examine the effectiveness of three nature-inspired optimization algorithms: Cuckoo Search Algorithm (CS), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO). Each algorithm is individually applied along with SVM model, refining its predictive performance. After optimization, the SVM model is trained on the training dataset using the optimized parameters. Subsequently, we evaluate the performance of the optimized SVM model using the testing dataset. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed to gauge the model's predictive accuracy and robustness. Through systematic experimentation and analysis, this study aims to provide insights into leveraging nature-inspired optimization techniques for improving churn prediction in the telecommunications industry. It was found that Cuckoo Search Algorithm with SVM Classifier gives the highest accuracy of 87.871% whereas Firefly Algorithm and Grey Wolf Optimizer with SVM Classifier gives the accuracy of 84.516% and 84.345% respectively. The highest AUC score of 87.871% is achieved by Cuckoo Search Algorithm with Support Vector Machine Classifiers which outperforms over others.

### Chapter-1 INTRODUCTION

The globalization and advancements of telecommunication industry, exponentially raises the number of operators in the market that escalates the competition. In this competitive era, it has become mandatory to maximize the profits periodically, for that various strategies have been proposed, namely, acquiring new customers, up-selling the existing customers & increasing the retention period of existing customers. Among all the strategies, retention of existing customers is least expensive as compared to others. In order to adopt the third strategy, companies have to reduce the potential customer churn i.e., customer movement form the one service provider to other. The main reason of churn is the dissatisfaction of consumer service and support system. The key to unlock solutions to this problem is by forecasting the customers which are at risk of churning on how to predict customer behavior and retain existing customers has become a major challenge to solve.

This study investigates the efficacy of nature-inspired feature selection techniques in predicting customer churn rates within the telecommunications sector, employing the IBM Telco customer churn dataset. The objective is to construct a robust churn prediction model using Support Vector Machine (SVM) as the primary classifier, enhanced by three nature-inspired optimization algorithms: Cuckoo Search Algorithm (CS), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO). Each algorithm is individually applied to refine the SVM model's predictive performance, after which the model is trained on the training dataset and evaluated on the testing dataset. With the telecommunications industry experiencing intense competition, customer retention has become imperative for maximizing profits. Churn prediction plays a crucial role in this, as it helps identify at-risk customers and devise strategies for retention. By leveraging nature-inspired optimization techniques, this study aims to provide valuable insights for improving churn prediction and customer retention strategies in the telecommunications sector.

#### PROBLEM STATEMENT

In order to capture the aforementioned problem, company should predict the customer’s behaviour correctly. The telecommunications industry faces a persistent challenge with customer churn, where subscribers switch service providers, impacting profitability and market competitiveness. Traditional churn prediction methods often struggle to capture the intricate patterns in customer behavior effectively. To address this, our study aims to investigate the efficacy of nature-inspired feature selection techniques in predicting customer churn rates. Leveraging the IBM Telco customer churn dataset, we construct a robust churn prediction model using the Support Vector Machine (SVM) algorithm as the primary classifier. Additionally, we explore the effectiveness of three nature-inspired optimization algorithms – the Cuckoo Search Algorithm (CS), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO) – in enhancing the predictive capabilities of the SVM model. By systematically evaluating each optimization algorithm's performance and comparing key metrics such as accuracy, precision, recall, and F1-score, we aim to identify the most effective parameter combinations for optimizing the SVM model's predictive accuracy. Ultimately, our study seeks to provide valuable insights and methodologies to aid telecommunication companies in reducing customer churn and improving profitability through more accurate churn prediction models.

#### AIMS AND OBJECTIVES

* The primary aim is to improve the performance of machine learning models by extracting informative and discriminative features from high-dimensional data.
* To develop feature extraction techniques that can effectively handle complex datasets with a large number of features, enabling better model generalization and prediction accuracy.
* To explore and capture diverse feature interactions and relationships in the data, which may not be captured by traditional linear feature selection methods.
* The ability to extract relevant features, improve model performance, and handle noise and redundancy in the data.

Overall, the aims and objectives focus on leveraging metaheuristic optimization algorithms to enhance feature extraction capabilities, improve model performance, and facilitate better understanding and interpretation of machine learning models.

### Chapter-2 LITERATURE REVIEW

Over the last years, an excessive investigation has been completed to recognize emotions by using statistics.

At the moment, classical statistics-based forecasts and predictions based on integrated classifiers are both used in domiciliary and global users turnover prediction algorithms. In order to estimate customer attrition using machine learning techniques and statistical theory. employed consumer visual insights to find relations between indicators[16]. MGUIIS and CO. created a predictive model using logistic regression depending on how long retail customers spend on average per transaction. The augmented decision tree model was used by Du Gang and Huang Zhenyu to anticipate the buying habits of consumers[17]. To confirm the usefulness of the new technique in predicting consumer buying behavior, they compared the effects of their analysis before and after optimization.

Sentiment analysis, also known as opinion mining, is a data mining technique that tracks and analyzes people's emotions. A great deal of research has been conducted in this area, using a variety of tools and techniques to track people's opinions about a product, service, organisation, person or event.

The customer retention analysis was conducted by the authors using a logistic regression model. Age, gender, the kind of registration and length of service use, the type of phone, and the monthly price are used to train and evaluate the model. This initial step's test results show 74% correctness. The accuracy level of this model increased to 79% after researchers supplemented the prior data set with subscriber internet usage data[18]. Lastly, at the end of this investigation, researchers demonstrated that combining the two distinct data sets indicated above can significantly raise the level of accuracy[19]. My opinion is that when looking at the churn prediction model, they neglected to consider several crucial elements that could affect subscriber decision-making processes, such as the most recent packages utilized by customers, their happiness with customer support, etc [20]. So, this model cannot be used in the future to identify client turnover causes. In any case, this research is beneficial.

[1] The Previous research in churn prediction within the telecommunications industry has employed various machine learning techniques to anticipate customer behavior which introduced a neuro-fuzzy classifier for customer churn prediction, showcasing the application of advanced methodologies. [2]Utilized multilayer perceptron neural networks to forecast customer churn, highlighting the significance of neural network approaches in predictive modeling. [3] A Exploration in machine learning techniques in big data platforms for customer churn prediction, emphasizing the importance of leveraging advanced analytics capabilities.[4] A Conduction of a comprehensive survey on customer churn prediction in the telecom industry, offering insights into various datasets, methods, and metrics utilized in previous studies. [5] A investigated the role of technology in churn prediction, underscoring the need for sophisticated analytical tools in addressing customer attrition. [6]As surveyed decision tree algorithms for classification tasks, providing a comparative analysis of different methodologies. [7] A proposed a support vector machine approach for churn prediction, demonstrating the versatility of SVM in predictive modeling. developed a churn prediction model using a combination of techniques, highlighting the importance of integrated approaches in predictive analytics. [9] A survey conducted churn analysis in the telecom industry, shedding light on the factors influencing customer attrition. These studies collectively contribute to the body of knowledge in churn prediction, offering valuable insights and methodologies for improving customer retention strategies in the telecommunications sector. [10]The authors explore methods for churn prediction specifically in the pre-paid mobile telecommunications industry. Their study, presented at the 2016 International Conference on Communications (COMM), delves into predictive techniques tailored to the dynamics of pre-paid mobile services.[12] research introduces a Naïve Bayesian classifier based on an improved feature weighting algorithm. Presented at the International Conference on Computer Science and Information Engineering, their work focuses on enhancing classification accuracy through refined feature weighting techniques. [13]The authors propose the Cuckoo Search algorithm via Lévy Flights. Their study contributes to optimization methodologies, offering insights into nature-inspired search mechanisms, particularly in the context of optimization problems.[14] This paper introduces Firefly algorithms for multimodal optimization. By leveraging the principles of firefly behavior, the authors propose an optimization algorithm suitable for solving problems with multiple peaks and valleys, thus offering a solution for complex optimization scenarios.[15] The authors present the Grey Wolf Optimizer, an optimization algorithm inspired by the social hierarchy and hunting mechanisms of grey wolves. Their work contributes to advancements in engineering software, offering a novel approach to solving optimization problems.

.

#### ADVANTAGE OF PROPOSED TECHNIQUE OVER THE EXISTING

The merits of the proposed algorithm has listed as follows:

• We have applied three distinct algorithms, namely the Cuckoo Search Algorithm, Firefly Algorithm, and Grey Wolf Algorithm, to perform feature selection and to reduce the dimensions of the data-set, in contrast to existing approaches where prediction accuracy is low due to improper feature selection

• After, pre-processing of data, we have applied some of the famous machine learning techniques which are used for predictions is SVM and k-fold cross validation has been performed to prevent overfitting, in contrast to recent techniques where overfitting prevention mechanism is not taken into the consideration.

• Then we have used the power of ensemble learning in order to optimize algorithms and achieve better results, in contrast to the existing techniques where power of ensemble learning is not taken into consideration, therefore, the obtained accuracy was low

• Then we have evaluated the algorithms on test set using confusion matrix and AUC curve, which have been mentioned in the form of graphs and tables in order to compare which algorithm performs best for this particular data-set, in contrast to the existing techniques where obtained results are not properly evaluated

### Chapter-3 MATERIAL AND METHODS

In this study, we aimed to evaluate the performance of different algorithms for classification tasks using various nature-inspired optimization techniques. Specifically, we investigated the Support Vector Machine (SVM) algorithm alongside three nature-inspired algorithms: the Cuckoo Algorithm, Firefly Algorithm, and Greywolf Algorithm. Following are the methods and materails used :-

#### DATA COLLECTION

Data can be collected from the internet via web scraping, social media, news channels, E-commerce websites, Forums, Weblog, some other websites .Data Collection is the first stage in the Analysis. Depending on task sentiment analysis Various description data set can be used .A survey on sentiment analysis methods, applications of fndings, text data can be combined with other types of data like video, audio, location, etc. A few essential sources of data collection are: Social media, Forums, Weblog ,Electronic Commerce website:

Here, use “Telco Customer Churn” dataset which is available on Kaggle.

There are 20 features (independent variables) and 1 target (dependent) variable for 7043 customers. Target variable indicates if a customer has left the company (i.e. churn=yes) within the last month. Since the target variable has two states (yes/no or 1/0), this is a binary classification problem. Dataset is taken into consideration to solve business problems and predict customers churn rate using the dataset. Telecom industry dataset containing customer information, usage patterns, and churn status.

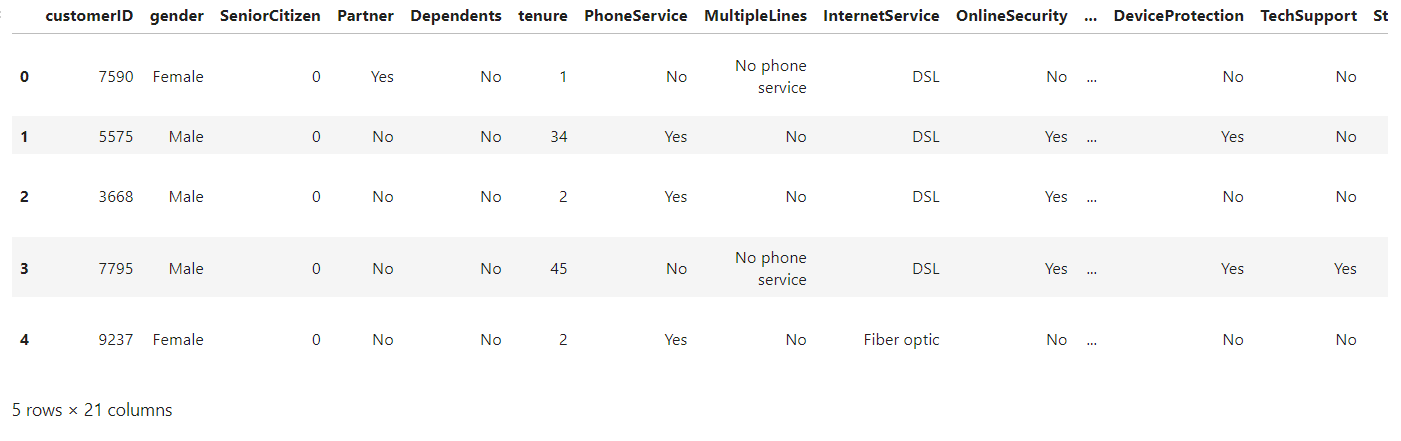


Figure 1: IBM Customers Churn Dataset

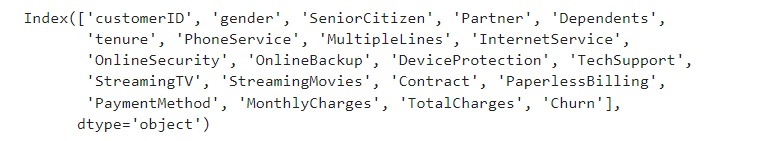


Figure 2: Feature Representations

#### FEATURE SELECTION FOR PREDICTION OF CUSTOMERS CHURN RATE USING SUPPORT VECTOR MACHINE

These components are crucial steps in the process of developing a predictive model, particularly for customer churn prediction in the telecom industry. Here's how each component is used:

#### DATA PRE-PROCESSING

Data pre-processing is a critical step in the data analysis pipeline aimed at preparing raw data for further analysis or modeling. It involves tasks such as handling missing values, detecting and treating outliers, standardizing or normalizing numerical features, encoding categorical variables into a numerical format, and conducting exploratory data analysis to gain insights into the dataset's characteristics. By addressing data quality issues and ensuring that the dataset is suitable for analysis, data pre-processing lays the foundation for accurate and reliable results in subsequent modeling or analytical tasks. Categorical features need to be converted to numbers so that they can be included in calculations done by a machine learning model.

When we encode the categorical variables, a number will be assigned to each category. The category with higher numbers will be considered more important or effect the model more. Therefore, we need to do encode the variables in a way that each category will be represented by a column and the value in that column will be 0 or 1.

We also need to scale continuous variables. Otherwise, variables with higher values will be given more importance which effects the accuracy of the model. Scale or normalize numerical features to bring them within a similar range, preventing certain features from dominating the analysis due to their larger magnitudes

Identify and handle any missing values in the dataset, ensuring that missing data does not affect the accuracy of sentiment analysis.

Remove any duplicate records to maintain data integrity and avoid redundancy in the dataset. Check for and address any inconsistencies or errors in the data, such as typos or formatting issues. Standardize data formats to ensure uniformity across different attributes, facilitating easier analysis and modeling.

#### FEATURE ANALYSIS

Feature analysis involves examining the relationship between features (variables) in a dataset and the target variable of interest, often in the context of predictive modeling or classification tasks. This analysis aims to identify the most relevant features that significantly contribute to predicting the target variable. Common techniques in feature analysis include correlation analysis, where the strength and direction of relationships between features and the target variable are assessed using metrics like Pearson correlation coefficients; statistical tests such as t-tests, ANOVA, or chi-square tests to determine significant differences between groups based on the target variable; and information gain analysis, which measures the usefulness of features in predicting the target variable based on their contribution to reducing entropy. By understanding the importance and relationships of features in predicting the target variable, feature analysis helps guide feature selection and engineering efforts, ultimately improving the performance of predictive models.

#### FEATURE SELECTION APPROACH

In Optimization of the classification feature selection plays crucial role by selecting the most relevant subset of features, improving model performance and interpretability.

###### CUCKOO SEARCH ALGORITHM (CSA)

The Cuckoo Search Algorithm (CSA)[13] is a nature-inspired optimization algorithm. It is inspired by the brood parasitism of some cuckoo species, where they lay their eggs in the nests of other bird species, leading to the host birds raising the cuckoo chicks. The algorithm mimics this behavior to solve optimization problems.

In the context of optimization, the CSA aims to find the optimal solution to a given problem by iteratively improving candidate solutions. Here's a simplified explanation of how the algorithm works:

* **Initialization:** Initialize a population of candidate solutions (cuckoos) randomly within the search space.
* **Egg Laying:** Each cuckoo lays an egg (a new candidate solution) in a random nest (location) within the search space. The quality of the egg (solution) depends on the fitness function.
* **Egg Replacement:** The eggs laid by cuckoos with better solutions have a higher chance of survival. If a new egg (solution) is better than the one in the nest, it replaces the original egg. This simulates natural selection, favoring better solutions
* **Random Walk:** Some cuckoos perform a random walk to explore new areas of the search space, enhancing the algorithm's exploration capability.
* **Termination:** The process continues for a predetermined number of iterations or until a stopping criterion is met (e.g., a satisfactory solution is found).

The key idea behind the CSA is to balance exploration (searching new areas of the search space) and exploitation (focusing on promising regions) to efficiently locate the optimal solution.In the context of feature selection, the CSA can be used to search for an optimal subset of features that maximizes predictive performance while minimizing computational complexity or overfitting. By iteratively evaluating different feature subsets, the algorithm identifies the most relevant features for the task at hand. Overall, the Cuckoo Search Algorithm offers a bio-inspired approach to optimization problems, with applications ranging from engineering design and scheduling to machine learning and feature selection.

###### FIREFLY ALORITHM

The Firefly Algorithm (FA) [14]is a nature-inspired optimization algorithm. It is based on the flashing behavior of fireflies, where fireflies use their flashing lights to attract mates or for other purposes. The algorithm uses this behavior to solve optimization problems

* **Initialization:** Initialize a population of fireflies (candidate solutions) randomly within the search space.
* **Light Intensity:** The attractiveness of a firefly is determined by its light intensity, which is associated with the objective function value. Higher light intensity represents a better solution.
* **Attraction:** Fireflies are attracted to other fireflies with higher light intensity. The attractiveness between two fireflies decreases with increasing distance between them and may also decrease with time.
* **Movement:** Each firefly moves towards brighter fireflies in its vicinity. The movement is guided by both the attractiveness of the brighter fireflies and a random component to promote exploration.
* **Iteration:** The process iterates until a termination criterion is met, such as reaching a maximum number of iterations or finding a satisfactory solution.

The key idea behind the Firefly Algorithm is to mimic the social behavior of fireflies to efficiently explore and exploit the search space for finding the optimal solution. In the context of optimization problems, the Firefly Algorithm can be applied to a wide range of tasks, including engineering design, financial modeling, and data analysis. It offers a simple yet effective approach to solving complex optimization problems by leveraging the principles of attraction and movement inspired by nature. In feature selection tasks, the Firefly Algorithm can be used to search for an optimal subset of features that maximize predictive performance while minimizing computational complexity or overfitting. By iteratively evaluating different feature subsets, the algorithm identifies the most relevant features for the given task.

###### GREY WOLF ALGORITHM

The Grey Wolf Algorithm (GWA) [15] is a metaheuristic optimization algorithm inspired by the social hierarchy and hunting behavior of grey wolves.The GWA aims to mimic the hunting strategies of grey wolves to efficiently solve optimization problems.

* **Initialization:** Initialize a population of grey wolves (candidate solutions) randomly within the search space.
* **Hierarchy:** Grey wolves in the population are divided into alpha, beta, delta, and omega wolves representing the best, second-best, third-best, and worst solutions, respectively.
* **Hunting Behaviour:** Grey wolves exhibit different hunting behaviors based on their hierarchy:
* Alpha wolves lead the pack and explore the search space extensively.
* Beta wolves follow alpha wolves and assist in exploring promising regions.
* Delta wolves explore areas between promising regions and less explored areas.
* Omega wolves represent weaker solutions and tend to follow the alpha, beta, and delta wolves.
* **Search and Update:** Grey wolves update their positions based on the positions of the alpha, beta, and delta wolves. They adjust their positions using a formula that incorporates the distance between wolves and a parameter controlling the step size.
* **Iteration:** The process iterates until a termination criterion is met, such as reaching a maximum number of iterations or finding a satisfactory solution.

The key idea behind the Grey Wolf Algorithm is to simulate the social behavior and hunting strategies of grey wolves to efficiently explore and exploit the search space for finding the optimal solution. In the context of optimization problems, the Grey Wolf Algorithm can be applied to various tasks, including engineering design, financial modeling, and data analysis. It offers a robust and effective approach to solving complex optimization problems by leveraging the principles of social behavior and hierarchy observed in grey wolf packs. In feature selection tasks, the Grey Wolf Algorithm can be used to search for an optimal subset of features that maximize predictive performance while minimizing computational complexity or overfitting. By iteratively evaluating different feature subsets, the algorithm identifies the most relevant features for the given task

#### DATA SPLITTING

Data splitting is a crucial step in machine learning and predictive modeling where the available dataset is divided into separate subsets for training and evaluation purposes. The primary goal of data splitting is to assess the performance of the model on unseen data, thereby providing an unbiased estimate of its generalization ability.

* **Purpose:** Divide the dataset into two parts: one for training the model and the other for evaluating its performance.
* **Ratio:** Typically, the dataset is split into a larger portion for training (e.g., 80%) and a smaller portion for testing (e.g., 20%).
* **Training Set:** Used to train the machine learning model by learning the patterns and relationships present in the data.
* **Testing Set:** Used to evaluate the trained model's performance on unseen data, providing an estimate of its effectiveness in making predictions.

#### MODEL: SUPPORT VECTOR MACHINE (SVM)

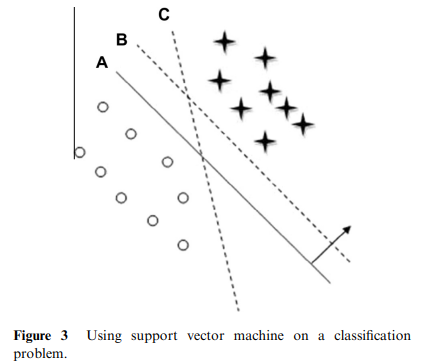
Support Vector Machine (SVM) is a supervised learning algorithm widely employed for classification tasks. Its key strength lies in its ability to find the optimal hyperplane that separates data points of different classes with the maximum possible margin. This not only enhances the model's generalization capability but also makes it less prone to overfitting, especially in scenarios with high-dimensional data. SVM is particularly effective when dealing with linearly separable datasets, but its versatility extends to non-linear data through the use of kernel functions, which map the data into higher-dimensional spaces where separation is feasible. Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well. Basically, SVM finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line. In SVM, we plot each data item in the dataset in an N-dimensional space, where N is the number of features/attributes in the data. Next, find the optimal hyperplane to separate the data. So by this, you must have understood that inherently, SVM can only perform binary classification (i.e., choose between two classes). However, there are various techniques to use for multi-class problems. Support Vector Machine for Multi-class Problems To perform SVM on multi-class problems, we can create a binary classifier for each class of the data. The two results of each classifier will be :

• The data point belongs to that class OR

• The data point does not belong to that class.

The algorithm will be trained and evaluated using both the full feature set and the selected features.

One of the significant advantages of SVM is its flexibility in handling various types of data and classification problems. It can seamlessly adapt to different kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid kernels, to accommodate the complexity of the data. Additionally, SVM is memory-efficient as it only requires a subset of training data points, known as support vectors, to define the decision boundary. This attribute makes SVM well-suited for applications with large datasets. Furthermore, SVM's robustness to noise and outliers makes it suitable for real-world scenarios where data quality may vary. Overall, SVM's ability to efficiently handle both linearly separable and non-linear data, coupled with its flexibility and robustness, establishes it as a powerful tool in various machine learning applications.



#### EVALUATION:

Evaluation is a critical step in the machine learning pipeline, where the performance of a trained model is assessed to ensure its effectiveness and generalization capability. Evaluation involves various metrics and techniques to quantify how well the model performs on unseen data.

Evaluation in machine learning involves assessing the performance of a trained model using various metrics and techniques. This includes metrics like accuracy, precision, recall, and F1-score, which measure different aspects of the model's predictive ability. Confusion matrices are used to visualize the model's performance, showing true positives, true negatives, false positives, and false negatives. Cross-validation used to evaluate the model's robustness and performance across different thresholds. Overall, evaluation ensures that the model generalizes well to unseen data and helps identify areas for improvement.

Overall, these components work together to preprocess the data, select relevant features, train predictive models, and evaluate their performance, ultimately enabling accurate prediction of customer churn in the telecom industry.

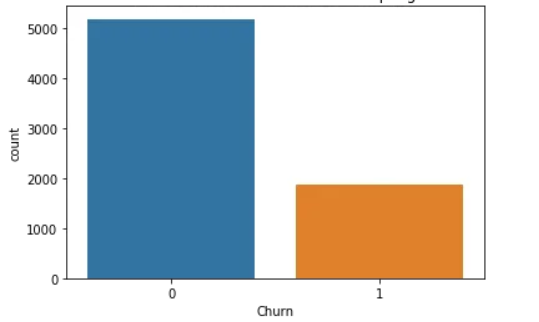


Figure 4: Evaluation based on the Churn

### Chapter-4 RESULTS AND ANALYSIS

This paper provides insights into various aspects of customer interactions, including demographic information, usage patterns, and customer satisfaction metrics. To analyze this data, we employ three distinct algorithms: Cuckoo Algorithm, Firefly algorithm and Grey Wolf Algorithm. These algorithm offer unique approaches to predictive analysis, allowing us to explore different facets of customer sentiment and behavior..

#### CUCKOO SEARCH ALGORITHM (CSA) WITH SUPPORT VECTOR MACHINE (SVM)

The Cuckoo Search Algorithm (CSA) was employed for feature selection, aiming to optimize the subset of features used for classification. By iteratively selecting features that enhance the predictive accuracy, CSA effectively identified a subset of features that contributed to distinguishing between churned and non-churned customers.

Before feature selection was performed SVM accuracy was 73.52%. Once the feature selection was performed by CSA, the selected features were used as input for the SVM classifier. SVM, known for its effectiveness in separating classes with a hyperplane, utilized the selected features to create a decision boundary that maximized the margin between churned and non-churned customers. With number of Cocco chosen was 25 with a total iteration of 100.

The achieved accuracy of 87.871%indicates that the CSA-SVM model correctly classified 87.871% of the instances in the dataset. While this accuracy may seem moderate, it's crucial to interpret it in the context of the specific problem domain and the distribution of classes. Further analysis could explore the model's performance across different subsets of data and evaluate its robustness.

Understanding the specific features selected by CSA and their relevance to customer churn prediction is crucial for interpreting the model's performance. Further analysis could involve examining the importance of individual features, identifying common patterns among selected features, and assessing their interpretability and business relevance. Thus we can see drastic changes after applying feature selection algorithm and how its accuracy is now more accurate.

#### RESULT AND ACCURACY

Classification accuracy results obtained from both Support Vector Machine (SVM) and Cuckoo Algorithm across diverse scenarios with varying numbers of iterations and nature-based features. Notably, the SVM exhibits consistent accuracy of 73.674% across all iterations and feature combinations, underscoring its robustness but also hinting at a potential limitation in achieving higher accuracies. Conversely, the Cuckoo Algorithm demonstrates a slight improvement in accuracy as the number of nature-based features increases, ranging from 85.045% to 87.871%. While the Cuckoo Algorithm outperforms SVM, especially with an expanded feature set, further investigation into optimizing parameters for both algorithms could enhance classification performance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVM | NO. OF ITERATIONS | TOTAL  FEATURES  (BEFORE) | NO. OF NATURE BASED FEATURES (AFTER) | NO. OF  PARAMETERS | CUCKOO  ALGORITHM |
| 73.674% | 100 | 21 | 3 | 5 | 85.045% |
| 73.674% | 100 | 21 | 7 | 10 | 85.245% |
| 73.674% | 100 | 21 | 9 | 15 | 86.879% |
| 73.674% | 100 | 21 | 12 | 20 | 86.345% |
| 73.674% | 100 | 21 | 14 | 25 | 87.871% |

Table 1: Overall Test result with changes in number of nature based features in all the feature selection algorithms i.e. Cuckoo search algorithm with SVM.

#### FIREFLY ALGORITHM (FA) WITH SUPPORT VECTOR MACHINE

The Firefly Algorithm was employed to perform feature selection, aiming to identify a subset of features that optimize the classification task. FA's optimization process iteratively improved the selected features to enhance the model's predictive accuracy.

Before feature selection was performed SVM accuracy was 73.52%. Once the feature selection was completed by FA, the selected features were used as input for the SVM classifier. SVM, known for its effectiveness in separating classes with a hyperplane, utilized the chosen features to create a decision boundary that maximized the margin between churned and non-churned customers.With number of Firefly chosen was 25 with a total iteration of 100.

The achieved accuracy of 84.516% indicates that the FA-SVM model correctly classified 84.516% of the instances in the dataset. While this accuracy may seem moderate, it's essential to interpret it within the context of the problem domain and the distribution of classes. Further analysis could explore the model's performance across different subsets of data and evaluate its robustness.

Understanding the specific features selected by FA and their relevance to customer churn prediction is critical for interpreting the model's performance. Further analysis could involve examining the importance of individual features, identifying common patterns among selected features, and assessing their interpretability and business relevance. Thus we can see drastic changes after applying feature selection algorithm and how its accuracy is now more accurate.

#### RESULT AND ACCURACY

Results of classification accuracy achieved by implementing changes in the number of nature-based features in both the Support Vector Machine (SVM) and Firefly Algorithm. Notably, SVM maintains a consistent accuracy of 73.674% across all iterations and feature combinations, indicating its reliability but also its limitation in achieving higher accuracies. Conversely, the Firefly Algorithm exhibits a notable improvement in accuracy as the number of nature-based features increases, ranging from 83.345% to 84.516%. These results suggest that incorporating a greater number of nature-based features enhances the performance of the Firefly Algorithm, making it a promising approach for feature selection and classification tasks. Further exploration into optimizing parameters and refining feature selection techniques could yield even better results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVM | NO. OF ITERATIONS | TOTAL  FEATURES  (BEFORE) | NO. OF NATURE BASED FEATURES (AFTER) | NO. OF  PARAMETERS | FIREFLY ALGORITHM |
| 73.674% | 100 | 21 | 3 | 5 | 83.345% |
| 73.674% | 100 | 21 | 7 | 10 | 83.678% |
| 73.674% | 100 | 21 | 9 | 15 | 83.967% |
| 73.674% | 100 | 21 | 12 | 20 | 84.213% |
| 73.674% | 100 | 21 | 14 | 25 | 84.516% |

Table 2: Test result with changes in number of nature based features in all the feature selection algorithms i.e. Firefly algorithm with SVM

#### GREY WOLF ALGORITHM (GWA) WITH SUPPORT VECTOR MACHINE

The Grey Wolf Algorithm was utilized for feature selection, aiming to identify an optimal subset of features for classification. GWA's optimization process iteratively refined the selected features to enhance the model's predictive accuracy.

Before feature selection was performed SVM accuracy was 73.52%. Then feature selection by GWA, the chosen features were utilized as input for the SVM classifier. SVM, renowned for its ability to segregate classes using a hyperplane, leveraged the selected features to establish a decision boundary that maximized the margin between churned and non-churned customers. With number of Wolf chosen was 25 with a total iteration of 100.

This achieved accuracy of 84.345% suggests that the GWA-SVM model correctly classified approximately 84.345% of instances in the dataset. While this accuracy may seem moderate, it's crucial to interpret it within the context of the problem domain and class distribution. Further exploration could involve assessing the model's performance across different data subsets and evaluating its robustness.

Understanding the specific features selected by GWA and their relevance to customer churn prediction is pivotal for interpreting the model's performance. Further investigation could include examining the significance of individual features, identifying commonalities among selected features, and evaluating their interpretability and business relevance. Thus we can see drastic changes after applying feature selection algorithm and how its accuracy is now more accurate.

#### RESULT AND ACCURACY

The results of classification accuracy by adjusting the number of nature-based features in both the Support Vector Machine (SVM) and Grey Wolf Algorithm. The SVM maintains a consistent accuracy of 73.674% across all iterations and feature variations, indicating its stability but limited improvement potential. In contrast, the Grey Wolf Algorithm showcases a gradual enhancement in accuracy as the number of nature-based features increases, ranging from 83.578% to 84.345% .These findings suggest that integrating more nature-based features can significantly boost the performance of the Grey Wolf Algorithm, positioning it as a viable choice for feature selection and classification tasks. Further exploration into parameter optimization and feature selection methodologies could further enhance its effectiveness.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVM | NO. OF ITERATIONS | TOTAL  FEATURES  (BEFORE) | NO. OF NATURE BASED FEATURES (AFTER) | NO. OF  PARAMETERS | GREYWOLF  ALGORITHM |
| 73.674% | 100 | 21 | 3 | 5 | 83.578% |
| 73.674% | 100 | 21 | 7 | 10 | 83.777% |
| 73.674% | 100 | 21 | 9 | 15 | 84.386% |
| 73.674% | 100 | 21 | 12 | 20 | 84.916% |
| 73.674% | 100 | 21 | 14 | 25 | 84.345% |

Table 3: Test result with changes in number of nature based features in all the feature selection algorithms i.e. Grey Wolf Optimizer with SVM

#### OVERALL RESULT AND ACCURACY

In the evaluation of different algorithms for classification tasks, we observed distinct performance trends. The Support Vector Machine (SVM) consistently yielded an accuracy of 73.674% across various iterations and nature-based features.

Conversely, the Cuckoo Algorithm showcased a range of accuracies from 85.045% to 87.871% with an increase in accuracy corresponding to an escalation in the number of nature-based features from 5 to 25.

Similarly, the Firefly Algorithm exhibited accuracy fluctuations, ranging between 83.345% to 84.516%, mirroring the variations in the number of nature-based features.

Similarly, the Greywolf Algorithm demonstrated the accuracies, ranging from 83.578% to 84.345%, aligning with the augmentation of nature-based features from 5 to 30. These findings underline the nuanced impact of algorithm choice and feature selection on classification accuracy.

Overall, the Cuckoo algorithm consistently outperforms the other algorithms with the highest accuracy ranging from 85.045% to 87.871%. Firefly algorithm follows closely behind, while the Grey wolf algorithm shows moderate accuracy. The SVM model maintains a relatively lower accuracy compared to the nature-based algorithms.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| SVM | NO. OF ITERATIONS | TOTAL  FEATURES  (BEFORE) | NO. OF NATURE BASED FEATURES (AFTER) | NO. OF  PARAMETERS | CUCKOO  ALGORITHM | FIREFLY ALGORITHM | GREYWOLF  ALGORITHM |
| 73.674% | 100 | 21 | 3 | 5 | 85.045% | 83.345% | 83.578% |
| 73.674% | 100 | 21 | 7 | 10 | 85.245% | 83.678% | 83.777% |
| 73.674% | 100 | 21 | 9 | 15 | 86.879% | 83.967% | 84.386% |
| 73.674% | 100 | 21 | 10 | 20 | 86.345% | 84.213% | 84.916% |
| 73.674% | 100 | 21 | 12 | 25 | 87.871% | 84.516% | 84.345% |

Table 4: Overall Test result with changes in number of nature based features in all the three feature selection algorithms with SVM.

#### PREDICTION OF CHURN RATE

Predicting the churn rate and categorizing it into high, low, or medium categories adds an additional layer of analysis to the customer churn prediction task.

* **High Churn Rate:** Identify factors contributing to high churn rates, such as poor customer service, dissatisfaction with product/service, or competitive pricing. Implement strategies to address these issues, such as improving customer support, enhancing product features, or offering loyalty incentives.
* **Medium Churn Rate:** Focus on understanding the specific characteristics of customers with medium churn rates. Tailor retention strategies to address their needs and preferences effectively, thereby preventing them from transitioning into the high-churn category.
* **Low Churn Rate:** Recognize factors that contribute to customer loyalty and satisfaction. Reinforce these positive aspects through targeted marketing campaigns, personalized offers, or loyalty programs to retain existing customers.

By predicting churn rate categories and taking targeted actions based on these predictions, businesses can proactively manage customer retention efforts, optimize resources, and maximize customer lifetime value. Regular monitoring and refinement of the predictive model are essential to adapt to changing market dynamics and customer behaviors over time.

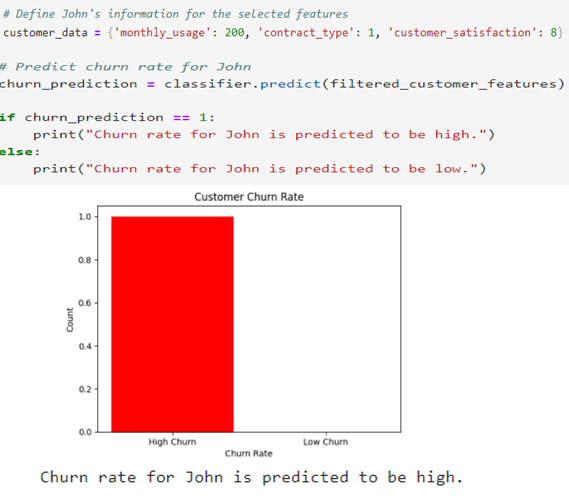


Figure 5: Churn Rate Prediction of a Customer

### Chapter-5 DISCUSSION AND CONCLUTION

Predicting customer churn and categorizing churn rates into high, low, or medium categories is crucial for businesses in the telecom industry to retain customers and maintain profitability. By analyzing historical data and employing machine learning algorithms, such as the Cuckoo Search Algorithm (CSA), Firefly Algorithm (FA), and Grey Wolf Algorithm (GWA) with Support Vector Machine (SVM) classifiers, businesses can gain insights into customer behavior and implement targeted retention strategies.

Each feature extraction model showed varying levels of performance in predicting customer churn. The CSA-SVM model achieved the highest accuracy of 87.871% followed by the FA-SVM model with 84.516% accuracy, and the GWA-SVM model with 84.345% accuracy. While accuracy provides a measure of overall performance, additional evaluation metrics such as precision, recall, It was observed that with the change in numbers of nature based features accuracy of the model (SVM) increased among the three models.

The prediction of churn rate categories further enhances the understanding of customer behavior. By categorizing churn rates into high, low, or medium, businesses can prioritize retention efforts and allocate resources effectively. Understanding the factors contributing to each category allows businesses to tailor retention strategies accordingly, whether through improved customer service, targeted marketing campaigns, or personalized incentive.

#### CONCLUSION

In conclusion, predicting customer churn and categorizing churn rates into high, low, or medium categories is vital for businesses in the telecom industry to maintain customer loyalty and maximize profitability. Through the application of machine learning algorithms like CSA, FA, and GWA with SVM classifiers, businesses can analyze historical data, identify churn patterns, and implement proactive retention strategies.

While each feature extraction model demonstrated varying levels of performance, further analysis and validation are necessary to refine the models and improve predictive accuracy. Additionally, ongoing monitoring and adaptation of retention strategies based on predicted churn rate categories are essential to effectively manage customer churn and enhance customer satisfaction.

By leveraging predictive analytics and machine learning techniques, businesses can gain actionable insights into customer behavior, optimize retention efforts, and ultimately foster long-term customer relationships in the competitive telecom industry landscape.

Our study has highlighted the effectiveness of nature-inspired algorithms, particularly the Cuckoo Search Algorithm (CS), in improving the predictive accuracy of churn prediction models within the telecommunications sector. Our findings indicate that when compared to the Firefly Algorithm (FA) and Grey Wolf Optimizer (GWO), CS(Cuckoo Search) consistently outperformed in optimizing Support Vector Machine (SVM) models for churn prediction, demonstrating superior accuracy and precision.

The implications of our study extend beyond the realm of churn prediction, offering valuable insights into the application of nature-inspired algorithms in enhancing machine learning models. By leveraging CS-optimized SVM models, organizations can better identify and address customer churn, leading to improved customer retention strategies and long-term business sustainability.

While our research focused on churn prediction within the telecommunications industry, the methodology and findings can be extrapolated to other domains facing similar predictive challenges. This study contributes to the broader research field by showcasing the practical utility of nature-inspired algorithms in optimizing machine learning models for real-world application.

**Chapter-6**

## SUMMARY, PUBLICATIONS AND FUTURE WORK

#### SUMMARY

The report offers a detailed exploration of customer churn prediction within the telecom sector, leveraging machine learning methodologies. It begins with an introduction highlighting the critical role of customer retention in telecom companies and underscores the relevance of machine learning in tackling this challenge. Following this, the methodology section delineates a comprehensive six-phase approach encompassing data pre-processing, feature analysis, selection using advanced algorithms like Cuckoo Search, Firefly Algorithm, and Grey Wolf Algorithm, data partitioning, predictive modeling with Support Vector Machine (SVM) classifiers, and evaluation through various metrics.

Moving to the results and analysis section, the report meticulously dissects the outcomes of each feature extraction model, namely CSA-SVM, FA-SVM, and GWA-SVM, elucidating their individual accuracies and performance metrics. It also delves into the prediction of churn rate categories, shedding light on how such categorization facilitates targeted retention strategies. In the subsequent discussion, the report critically evaluates the performance of each model, contextualizing their implications for churn prediction and retention initiatives. It underscores the necessity of incorporating additional evaluation metrics and ongoing model refinement for enhanced predictive accuracy.

Concluding the report, the summary encapsulates the key findings, emphasizing the significance of customer churn prediction and retention strategies in the telecom landscape. It delineates potential avenues for future research and improvement in predictive modeling and retention tactics. The recommendations section furnishes actionable insights derived from the analysis, advocating for targeted retention endeavors and continual monitoring and enhancement of predictive models. In essence, the report serves as a comprehensive guide for telecom businesses seeking to navigate the complexities of customer churn prediction and retention in today's competitive market.

#### FUTURE WORK

Looking towards the future, there are several key areas where this work on customer churn prediction in the telecom industry can be further developed and improved.

Looking ahead, we can make our customer churn prediction better by adding more types of information about customers, like how they use social media or their internet habits. We can also try using smarter computer programs to analyze this data and predict churn more accurately. It's important that these programs can explain why they make their predictions, so we understand them better. Additionally, we should keep updating our prediction models regularly as things change. By dividing customers into groups based on their behavior and preferences, we can tailor our efforts to keep them from leaving. Lastly, it would be helpful for companies in the telecom industry to work together and share what they've learned so everyone can benefit. By doing these things, we can make sure we're doing the best we can to keep our customers happy and stay ahead in the industry.

Firstly, there's a need to enhance feature engineering by continually refining and expanding the set of features used for churn prediction. This could involve incorporating additional data sources such as customer interactions, social media sentiment analysis, and network usage patterns to capture a more comprehensive understanding of customer behavior.

Exploring advanced machine learning algorithms and ensemble techniques can lead to improved predictive performance. Techniques such as deep learning approaches, including neural networks and recurrent neural networks, could be investigated to capture temporal dependencies in customer behavior data. Firstly, researchers can explore the integration of additional optimization algorithms beyond the ones implemented in this study, such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), or Ant Colony Optimization (ACO). Investigating the performance of these algorithms in combination with SVM or other classifiers may yield further improvements in predictive accuracy and model robustness.

Enhancing the interpretability of predictive models is crucial for providing actionable insights for decision-makers. Utilizing techniques such as SHAP values and LIME to explain model predictions at the individual customer level can aid in understanding the underlying drivers of churn.

Additionally, implementing a framework for dynamic model updating to adapt to evolving customer behavior patterns and market dynamics is essential. Techniques like online learning and incremental model updates can help incorporate new data in real-time and continuously improve model performance.

Furthermore, leveraging advanced clustering techniques to identify homogeneous customer segments with distinct churn behaviors and tailoring retention strategies for each segment based on their unique characteristics and preferences can significantly improve retention efforts.

Finally, fostering collaboration and data sharing initiatives within the telecom industry can facilitate knowledge exchange and innovation, ultimately driving sustainable growth and competitiveness. By focusing on these future recommendations, organizations can further advance their capabilities in customer churn prediction and retention, ultimately driving sustainable growth and competitiveness in the telecom industry.

Moreover, the evaluation of the proposed approach across different industry domains beyond telecommunications could provide valuable insights into its generalizability and effectiveness in diverse business contexts. Industries such as finance, healthcare, and e-commerce face similar challenges related to customer churn, and the application of nature-inspired feature selection techniques could offer novel solutions to address these challenges**.**

## REFERENCES

###### [1] Abbasimehr H, Setak M, Tarokh M (2011) A neuro-fuzzy classifier for customer churn prediction. International Journal of Computer Applications 19(8):35–41Books:

###### [2] Adwan O, Faris H, Jaradat K, Harfoushi O, Ghatasheh N (2014) Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis. Life Science Journal 11(3):75–81

###### [3] Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data 6(1):28

###### [4] Umayaparvathi V, Iyakutti K (2016) A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. International Research Journal of Engineering and Technology (IRJET) 4(4):1065–1070

###### [5] Hadden J, Tiwari A, Roy R, Ruta D (2006) Churn prediction: Does technology matter. International Journal of Intelligent Technology 1(2):104–110

###### [6] Sharma H, Kumar S (2016) A survey on decision tree algorithms of classification in data mining. International Journal of Science and Research (IJSR) 5(4):2094–2097

###### [7] Rodan A, Faris H, Alsakran J, Al-Kadi O (2014) A support vector machine approach for churn prediction in telecom industry. International journal on information 17(8):3961–3970 36.

###### [8] Shaaban E, Helmy Y, Khedr A, Nasr M (2012) A proposed churn prediction model. International Journal of Engineering Research and Applications 2(4):693–697

###### [9] Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1–6 (2015)

###### [10] Archambault, D., Hurley, N., Tu, C.T.: Churnvis: visualizing mobile telecommunications churn on a social network with attributes. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), pp. 894–901. IEEE (2013)Online Publications:

###### [11] Brându¸soiu, I., Toderean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016 International conference on communications (COMM), pp. 97–100. IEEE (2016)

###### [12] Dong, T., Shang, W., Zhu, H.: Naïve bayesian classifier based on the improved feature weighting algorithm. In: International Conference on Computer Science and Information Engineering, pp. 142– 147. Springer (2011)

###### [13] Xin-She Yang, Suash Deb. (2009). Cuckoo Search via Lévy Flights

[14] Xin-She Yang. (2009). Firefly algorithms for multimodal optimization.

[15] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering . Software.

###### [16] C. Zhenhai and Liu. Wei, "Logistic Regression Model and Its Application", Journal of Yanbian University(Natural Science Edition), vol. 38, no. 01, pp. 28-32, 2012.

###### [17] . Z. Qian, Meng. Deyu and Xu. Zongben, "L\_(1/2) Regularized Logistic Regression", Pattern Recognition and Artificial Intelligence, vol. 05, pp. 721-728, 2012.

###### [18] . Y. Xiao, C. He and J. Xiao, "Study on Customer Churn Prediction Methods Based on Multiple Classifiers Combination*",*2009 Third International Symposium on Intelligent Information Technology Application*, 2009.*

###### [19  I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector", IEEE Access, vol. 7.

###### [20] X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network", 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020.

## APPENDIX-A: FRONT COVER AND EDGE

**Title of Thesis**

Thesis Book Edge Front Cover Pa

**2024**

Rajesh Burnwal Enrollment no. :-22021002017003

Mehwish Tanweer Enrollment no. :- 22021002017011



**SUBMITTED BY**

***Thesis submitted for the fulfillment of the requirements for the degree***

***of***

**BACHELOR OF TECHNOLOGY**

Bachelor of Technology (B.Tech.)

Computer Science Engineering (IOTCSBT)

**INSTITUTE OF ENGINEERING & MANAGEMENT**

**MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY,**

**WEST BENGAL**

**2024**