BASIC INFORMATION

Title of Project: Survey on Customer Churn Prediction in Telecom Industry

Student Name: Pratham Kumar

Branch : Artificial Intelligence & Data Science (B1-A)

Enrollment Number: 00119011921

Email ID: pratham.00119011921@ipu.ac.in

Contact Number: 8076123862

Google Drive Link:
https://drive.google.com/drive/folders/1Q4m1gUeq1Rp1hVTb0n6dleF6tZm7yXPw?usp=drive_link

Survey on Customer Churn Prediction in Telecom Industry

Abstract:

Churn, the phenomenon of customers switching from one service provider to another, is a critical concern for businesses operating in highly competitive industries. In the context of the Iranian telecommunications industry, understanding and predicting customer churn can significantly impact operational efficiency, customer retention strategies, and revenue generation. This abstract presents an overview of the Iranian Churn Dataset, a comprehensive collection of customer-related data specifically curated to analyze churn patterns within this industry.

The Iranian Churn Dataset comprises anonymized customer records, encompassing a diverse range of attributes such as demographic information, service usage patterns, billing details, customer complaints, and service cancellation records. The dataset spans a period of two years, allowing for longitudinal analyses and the identification of temporal churn trends.

To facilitate accurate churn prediction and analysis, the dataset includes features such as customer tenure, call duration, total bill amount, customer satisfaction ratings, and service subscription details. Additionally, it contains churn labels indicating whether a customer churned or remained loyal to their service provider during a specific period.

Keywords:

Customer churn prediction,

Predictive models,

Data mining,

Iranian telecommunications industry,

data analysis.

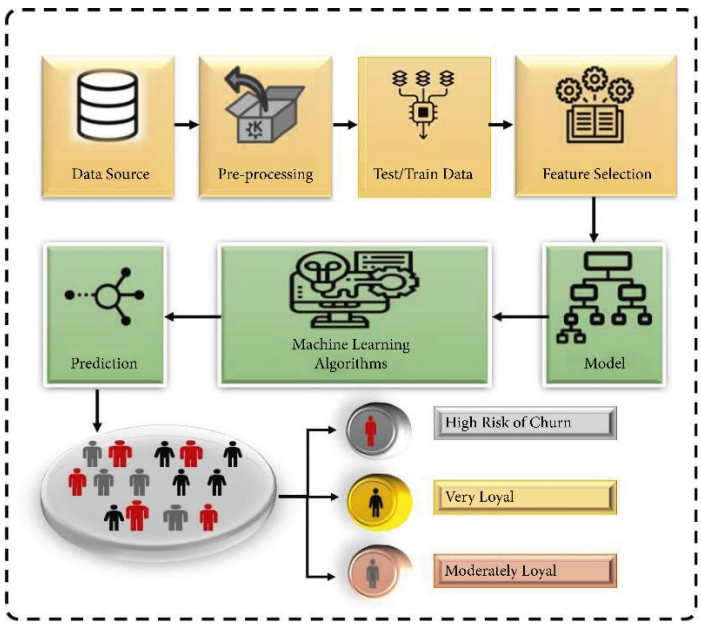
Introduction:

In today's highly competitive telecommunications industry, customer churn poses a significant challenge for service providers seeking to retain their customer base and sustain profitability. The Iranian telecommunications industry is no exception, with intense competition and evolving market dynamics driving the need for a deeper understanding of churn patterns. To address this, the Analyzing Churn Patterns in the Iranian Telecommunications Industry Dataset provides a comprehensive collection of data specifically curated to analyze churn behavior within this industry.

The dataset focuses on unraveling the complex factors contributing to customer churn in the Iranian telecommunications sector. By examining the characteristics, behaviors, and interactions of customers who churn, service providers can gain valuable insights into the underlying causes and devise effective strategies to reduce churn rates.

The Analyzing Churn Patterns in the Iranian Telecommunications Industry Dataset encompasses a diverse range of customer-related information, including demographic details, service usage patterns, billing information, customer complaints, and records of service cancellations. This wealth of data enables researchers and industry professionals to explore various dimensions that influence churn behavior and develop targeted approaches to retain customers.

Proposed Methodology



1.Datasets:

To facilitate accurate analysis and prediction, the dataset includes crucial features such as customer tenure, call duration, total bill amounts, customer satisfaction ratings, billing information, customer complaints, service subscription details and records of service cancellations. These variables provide insights into customer behavior, preferences, and satisfaction levels, allowing researchers and industry professionals to identify the key drivers of churn.

2.Preprocessing:

The preprocessing of the Iranian Churn Dataset involved several steps. Firstly, we retrive information about the dataset so that further operations can be performed. Next, numerical and categorical variables were normalized to a standardized scale. Missing data points were handled through appropriate techniques such as imputation or removal, maintaining the dataset's integrity. Further, Feature Scaling is done using Standard Scaler. It will standardize the data and make the mean of the data features nearly the same.

3.Classification:

a) **Logistic Regression :** It is a classification algorithm that works by estimating the probability of an instance belonging to a particular class. It models the relationship between the independent variables and the dependent binary outcome using a logistic function, making it suitable for binary classification tasks.

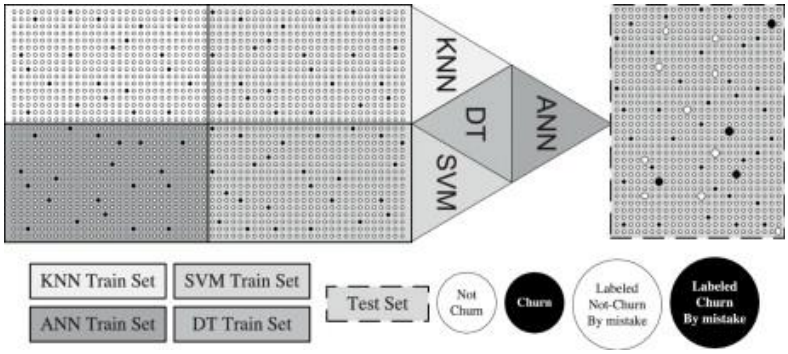
b) **Decision tree classifier:** The algorithm that uses a tree-like structure to classify data based on a series of decisions or rules. It recursively splits the data into subsets based on the features that provide the most information gain, leading to the creation of a predictive model. It is capable of handling both numerical and categorical data and is widely used for classification tasks in various domains**.**

c) **Perceptron:** It is a binary classification algorithm that iteratively learns a linear decision boundary between two classes. It receives input features,

assigns weights to them, and calculates a weighted sum. If the sum exceeds a threshold, the perceptron predicts one class; otherwise, it predicts the other. Through repeated iterations and weight adjustments, the perceptron aims to minimize classification errors and accurately separate the classes.

d) **Random Forest Classifier :** It utilizes an ensemble of decision trees to make predictions. Each tree in the ensemble is trained on a subset of features and data samples, and the final prediction is determined by a majority vote among the individual trees. This algorithm is effective for handling complex datasets, capturing feature interactions, and achieving high prediction accuracy in various classification tasks**.**

e) **K Nearest Neighbors Classifier:** The algorithm that works by assigning a class label to a new data point based on the majority class of its k nearest neighbors in the feature space. It calculates the distance between the data points and selects the k nearest neighbors to make predictions.

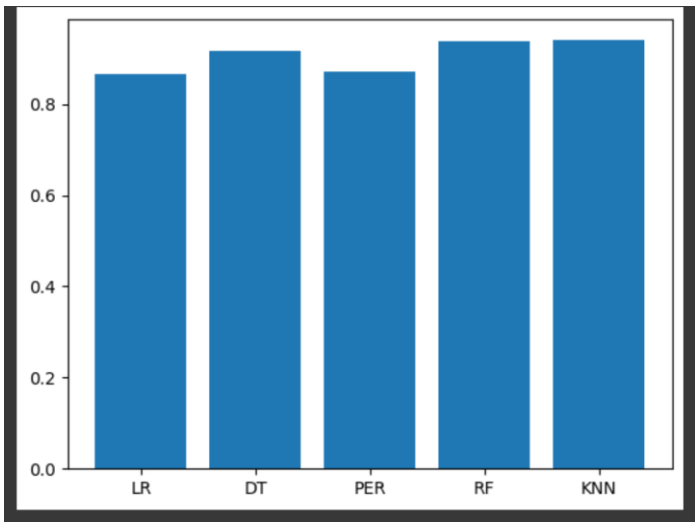


Performance measures must be done for each classifier to get the effectiveness of models or algorithms. It can be done using Confusion Matrix, Classification Report, F1 Score and Accuracy Score, etc.

Result:

After applying logistic regression, its k nearest neighbors, decision tree, perceptron, and random forest classifier to the Iranian Churn Dataset, each model demonstrated varying performance its k nearest neighbors achieved the highest accuracy(0.933), followed closely by random forest classifier. decision tree classifier exhibited moderate accuracy(0.917), while the perceptron(0.869) and logistic regression(0.866) had lower accuracies. These results suggest that k

nearest neighbors and random forest classifier models may be more suitable for predicting churn patterns in the Iranian telecommunications industry.



Future Work:

The Iranian Churn Dataset can be utilized in various ways in the future. Researchers and industry professionals can use the dataset to understand why customers switch from one telecom provider to another in Iran. By analyzing the data, they can predict churn behavior and identify key factors that influence customers' decisions. This information can help telecom companies develop effective strategies to retain customers and improve their services. Additionally, the dataset can be used to compare churn patterns over time and analyze seasonal trends.

Overall, the Iranian Churn Dataset has the potential to drive improvements in customer retention and satisfaction in the Iranian telecommunications industry.

References:

1. https://www.hindawi.com/journals/cin/2022/1703696/
2. https://www.sciencedirect.com/science/article/abs/pii/S1568494614004062
3. https://d1wqtxts1xzle7.cloudfront.net/54561005/IRJET-V3I4213-libre.pdf?1506592122=&response-content-disposition=inline%3B+filename%3DA_Survey_on_Customer_Churn_Prediction_in.pdf&Expires=1686662353&Signature=MDJQztY5-UiohQTuc-hwPoBcN5d-cHriWYVh8roW71Jrqn5dUavL~0OCYKc78CW8IgrGFp5ao3GxFl1Wm3t040caVFYs5o9kwjR6KAmoYWFOIkFhS2aG1OjMTNdwQEIWH3Y1qjZS0jP4RluSS-CEC-R3G2aYaIGsJiEg9~xEnB2s7sWqmixZIOOZzIz8AMaFtNqMkQEWxbMjMhKKyyWEKDvzwvMeu4YME3YkrPa4yY4mucNKSxiEGHWl-ZDfmAnL6OSZXhZGTjFcwzZjU1n6A2EPOOT9wMOvAtrrJ6ELz1hWU3yRkJO0W6lTlzcnRlvntmY2oHOPypmUMDavPb2XmQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
4. https://hrcak.srce.hr/clanak/220286
5. https://www.sciencedirect.com/science/article/pii/S1877050920306529
6. https://www.sciencedirect.com/science/article/abs/pii/S0360835220302102
7. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=33be72961a970b52516f494f9ef26712e0790e3b
8. https://ieeexplore.ieee.org/abstract/document/7707454
9. https://www.sciencedirect.com/science/article/abs/pii/S0148296318301231
10. http://section.iaesonline.com/index.php/IJEEI/article/view/2985/702
11. https://www.sciencedirect.com/science/article/abs/pii/S0957417411011353
12. https://archive.ics.uci.edu/dataset/563/iranian+churn+dataset
13. https://content.iospress.com/articles/informatica/infor484
14. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267935
15. https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=samples-telco-customer-churn