[Churn Prediction using Neural Network based Individual and Ensemble Models](F://Academic/4th%20Year/7th%20Semester/Sessional/CSE%204120%20Technical%20Writing%20&%20Seminar/Paper/OK%20Churn_Prediction_using_Neural_Network_based_Individual_and_Ensemble_Models.pdf)

# Proposal

This paper evaluates existing individual and ensemble Neural Network based classifiers and proposes an ensemble classifier which utilizes Bagging with Neural Network in order to improve performance measures resulting in better accuracy for churn prediction.

An average accuracy of 81% is achieved by the proposed model.

## Keywords

### Churn Prediction

Churn prediction involves identifying at-risk customers who are likely to cancel their subscriptions or close/abandon their accounts. A churn model works by passing previous customer data through a machine learning model to identify the connections between features and targets and make predictions about new customers.

### Retention:

The continued use, existence, or possession of something or someone.

### Data mining

Data mining is the process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal of extracting information from a data set and transforming the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases". Data mining is a process applied on data, having previously unknown information which may be useful for taking some decision. More precisely we can say that it works on the large datasets from databases to extract the relevant patterns which may help to design different strategies for a profitable business.

Association rules, Support Vector Machine (SVM), Artificial Neural Network (ANN) and Decision Trees (DT) are commonly used data mining techniques.

### Machine Learning

Machine Learning is emerged from the study and construction of algorithms which emerged from the studies of computational learning and pattern recognition in artificial intelligence.

Algo’s are divided into 2 parts

* Supervised
  + Classification: Classification models are trained on data from past customers to generate a model which is then applied to classify unseen patterns. (clustering algorithms: clustering algorithms focus on similar features to group data in a cluster and then categorize unseen data into any one of the related cluster.)
  + Regression: In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The most common form of regression analysis is linear regression, in which one finds the line that most closely fits the data according to a specific mathematical criterion. For example, the method of ordinary least squares computes the unique line that minimizes the sum of squared
* Unsupervised

### Overfitting

In machine learning, overfitting occurs when an algorithm fits too closely or even exactly to its training data, resulting in a model that can’t make accurate predictions or conclusions from any data other than the training data.

### UnderFitting

Underfitting is a scenario in data science where a data model is unable to capture the relationship between the input and output variables accurately, generating a high error rate on both the training set and unseen data.

### Pros and cons

The pros and cons of something are the arguments for and against it. When you’re considering the pros and cons of something, you’re thinking about its good and bad points. The phrase pros and cons came from the Latin words “prō,” which means “for” and indicates favor towards something, and “contrā,” which means against, indicating opposition.

### Zero\_one loss

In multilabel classification, the zero\_one\_loss function corresponds to the subset zero-one loss: for each sample, the entire set of labels must be correctly predicted, otherwise the loss for that sample is equal to one

### Negative log-likelihood

Negative log-likelihood is a measure of how likely a given model is to produce the observed data. It is obtained by taking the negative of the natural logarithm of the probability (or likelihood) of the data. The log of a probability is always negative, so the negative log-likelihood is always positive. Negative log-likelihood is used for optimization because most optimizers minimize a function, and minimizing the negative log-likelihood is equivalent to maximizing the likelihood.

## Factors of churn prediction

* **churn rate** (The churn rate, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity. It is most commonly expressed as the percentage of service subscribers who discontinue their subscriptions within a given time period)
* **prediction performance** (Performance prediction in computer science refers to estimating the execution time or other performance factors of a program on a given computer1.Performance evaluation measures include2:)
* **customer retention capabilities** (Customer retention is the ability of a company to keep its customers loyal and satisfied over time1234. It is measured by the rate at which customers stay with or return to a business in a given period, also known as churn rate35. Customer retention is important for increasing revenue and reducing customer defections24. Customer retention programs include activities and actions that aim to stimulate customer loyalty and brand loyalty4.)

These are the main drivers to predict customer behaviour.

## Solution to retain customers

Many researchers have proposed various solutions for retaining customers by applying different techniques.

* Inductive algorithms,
* hybrid oversampling and under sampling,
* SMOTE oversampling technique,
* echo state network (ESN) with SVM training algorithm,
* clustering techniques,
* artificial neural networks,
* decision trees,
* random forest algorithms,
* PSO,
* mRMR,
* GA and
* ensemble classifiers

have been used for achieving performance in churn prediction. Early prediction of customer churn increases the chances to retain the customer.

## Research Contributions

1. This research evaluates different versions of Neural Network based individual classifiers for churn prediction. These classifiers include Neural Network (NN), Multi Layer Perceptron (AutoMLP) and Deep Learning (DL).
2. These classifiers are then evaluated in an ensemble setting where well-known ensemble methods are utilized such as Bagging, AdaBoost and Majoirty Voting.
3. Results are computed and evaluated using various performance measures such as accuracy, precision, recall, f-measure and statistical measures such as kappa, absolute error, relative error and classification error.

## Platform

1. Big ML
2. Azure ML

platforms were used for churn prediction by focusing on user services based on demographic and behavioural data.

## Classifier

### Deep Learning

Deep learning is the subset of machine learning methods based on neural networks with representation learning. The adjective "deep" refers to the use of multiple layers in the network. Methods used can be either supervised, semi-supervised or unsupervised.

DL is basically a machine learning algorithm subjective to the area of AI. It is basically an inspired version of ANN. This model consists of larger architecture in which multiple nodes are connected to one another and they work in a manner like neuron of the human brain.

### Neural Nets

The next step is building model on NN which is also one of the machine learning techniques used commonly. Neural nets work in the same concept as the neuron cells of the human brain work. They make a network in which different adjacent layers are connected together. These layers connectively make a deep network. In NN weight of the signals move across the neurons is collected at the output then results are derived from it. NN can work on both supervised and unsupervised ML.

### AutoMLP

AutoMLP is our third machine learning classifier. In this classifier NN can adjust its size and learning rate during the training of classifier. This algorithm use some of the ideas from genetic algorithms and some from stochastic optimization.

## Ensemble Classifier

Classifiers are the individual models use training dataset to train them. Among all classifiers some are classified as ensemble classifier because they work in combination with some other classifier and use the working strategy to combine each of them to get the higher precision, recall and classification performance.

Three different types of ensemble classifier are going to be used (Bagging, AdaBoost & Majority Voting)

### Bagging

Bootstrap aggregation classifier work for the classification and prediction problem solving. It reduces classification error and improves the accuracy in comparison to individual classifiers.

Majority Voting

Majority voting, also known as plurality voting, works on multiple classifier feed, gets the high frequency vote from each classifier and calculates the best voting results for the classifiers.

AdaBoost

AdaBoost is basically an iterative algorithm that works with some weak classifiers to solve the classification problems by building a strong classifier rather than regression. AdaBoost improves the performance of machine learning classifiers up to many folds3. AdaBoost works on the basis of weights assigned to each instance in the training dataset.