Prediction of Customer Churn by Machine Learning

Mandar Vaidya

x17153409

MSc. in Data Analytics

17th August 2018

Abstract

One of the most vital responsibilities of many organizations is customer retention. Customer attrition or about to attrite is known as churn, churn rate of almost 40%, telecommunication sector takes the first place. In this research, advanced data mining and deep learning predictive models are used to predict customer churn to accurately valuate who are likely to churn and approximate time of the likely churn. Many companies are shifting towards customer retention than customer acquisition like before because the cost of retention is far greater than acquisition. Prediction techniques like Logistic Regression, K-means Clustering, Random Forest, Singular Value Decomposition based on new features will help in prediction churn with the help of different combination of variables. The dataset used in this research is obtained from IBM Watson Analytics website. It contains data of 7,044 customers and their usage behavior of 2 months.

keywords**:** Customer Churn, Data Mining, Machine Learning, Churn Rate

Contents:

1. <u>Introduction:</u>

1.1 Overview

Customer Churn is extensively familiar concept in today's global world especially in industries like gaming, banking, telecommunications. Churn is defined as a customer leaving a service. Churn prediction process has become a highly debated research area these days. Researchers from various fields have tried to analyze this area by integrating various techniques and recommended solutions to decision makers about churners. Due to global competition, especially telecom companies have shifted their focus on customer retention than customer acquisition. As, customer retention is more expensive than customer acquisition. I last decade or so, lot of changes in telecom industry like, uplifting of market restrictions of opening competition in telecom market, new technologies, new lenient laws because of such massive competition, churn of customers causes a substantial loss in this sector which comes a serious concern for the companies.

Telecommunications companies are thus focusing more on defensive marketing strategies.

One of the important task is to identify customers who are on verge of churn before they are going to churn. To analysis about churners and their behavior various data mining, machine learning, deep learning techniques are implemented for the prediction. Conventional churn prediction techniques are simple and robust but their limitations to the interpretation of the reasons for churning is very low.

Churners can be defined into two types: (i) Voluntary and (ii) Involuntary.

Voluntary churners are most difficult to regulate, they are prompt to change the service with the provider with ease where as Involuntary churners are the easiest to identify because the telecom companies decide to remove them from the subscribers lists.

Voluntary churners can be sub-divided into two more categories, (i) Deliberate and (ii) Incidental

Deliberate churn occurs for the reason of the technology, for example; bad service, price sensitivity, customers opting for newer technology, customers opting for better technology, convenience parts, social factors etc. Deliberate churn is the main area of focus and that's why most it is the most researched category churn management solutions try to analyze.

Incidental churn occurs unexcepted, it is not planned by the customers it happens because something happened in their lives, for example, location change, financial issues etc.

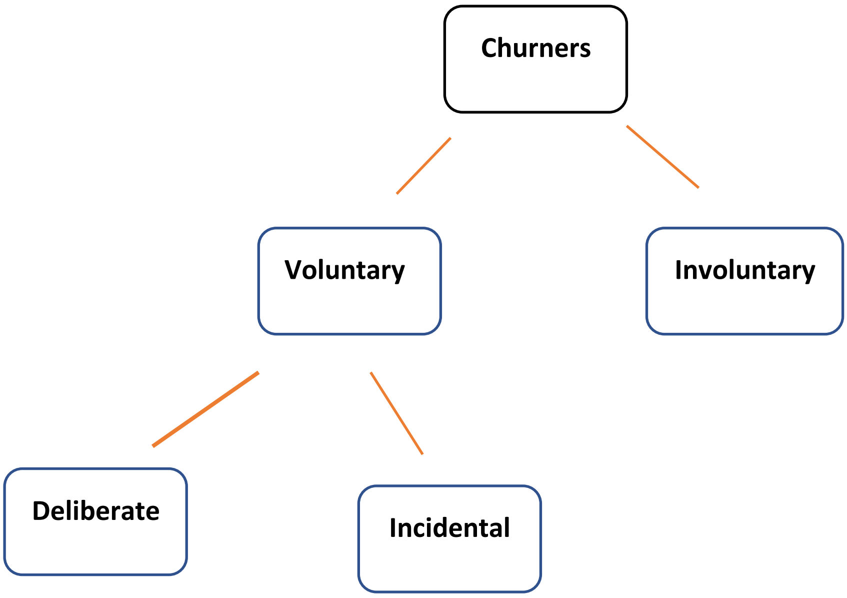

Fig.1 Churn Classification

Retaining customers who are about too or mostly likely to churn, decision makers, marketing team must be able to predict in advance which customers are going to churn and which strategies will have a positive impact on every customer. With this knowledge, customer churn can be analyzed and can be eliminated to an extent. This prediction of churn analysis can be analyzed through Data Mining, Machine Learning Algorithms with the help of classifier for better accuracy.

Churn prediction analysis understand specific customer behaviors and various attributes which are dependent on customer churn. The more accurate the technique the more chance of retention of customers. If marketing team is unaware of a customer who is about/likely to churn, no prevention action will be taken against that customer. More importantly, incentives, offers, retention-focused schemes need to be executed to make the active customers happy, thus by resulting in less churning and reduced revenues. As it is mentioned in various sources, churn rate of almost 30-40%, the telecommunications take the first place on the list of churners and it is showing no signs of decline also.

1.2 Research Question

How will organizations benefit to cut down churners and churn rate from predictive model with the help of Machine Learning algorithm

2. <u>Literature Review:</u>

(Khan and Kozat; 2017) investigated churn detection as well as prediction by continuous cellular network data. The cellular customer data can be divided into two parts: (i) Static and (ii) Sequential. Age, gender, start date etc. which remain constant are known as Static data where as bill payments, type of services, bill payments etc. which are not constant are known as Sequential customers data respectively. A comparison was by Machine Learning algorithms like Random Forest, Support Vector Machine, Gradient Boosting Trees etc. Classifier was used to get accurate results. Cross-validation is used to verify the model persistence. Libraries like Keras and Scikit-Learn are used for the experiments. Since, sequential algorithms do not store and process all the previous data, it is more efficient for training and testing data, although real-life scenario might be different. By using ROC curves and AUC scores, analysis was compared between different algorithms which not only predict the churn time but also time of churn in advance. To extend the research, other type of features need to be consider like categorical, numerical etc. for real-time analysis and prediction for further adaptation.

(Yihui and Chiyu; 2016) suggests, dimensionality plays a decisive role in reduction method especially in field of telecom. In this paper, indicator system of customer churn is studied.
Orientation Ordering Pruning Method (OOPM) is a feature selection pruning technique was proposed by the author. Pruning classifier is the replacement of feature selection method an indicator system was formulated for customer churn. And, with the help of machine learning technique like Random Forest and Transduction (FE_RF&T) was used for the extraction of various other features of customer churn. Analysis showed that, OOPM has better results than FE_RF&T method. To improve the model efficiency, feature selection and feature extraction is widely used. With the help filter model and wrapper model, original features are maintained by feature selection. Whereas, transformation of feature space to a database with the help of projection matrix is done by feature extraction. As, OOPM selects a basic classifier, feature selection method was applied to it. With the help of proximity matrix and dimension scaling of Random Forest and Transduction, feature extraction was done. Compared to PCA, FE_RF&T, Random Forest, OOPM showed better results. Further analysis can be done on real-time big datasets.

(Brândușoiu et al.; 2016) suggests for dimensionality reduction, and elimination of multicollinearity, principal component analysis was performed. Machine Learning techniques like Neural Networks, Support Vector Machine, Bayesian Networks was performed in discrete variables. For accurate churn detection, predictive model is must. It will accelerate the retention process and will help the companies in achieving better results. After training and testing of the dataset into various parts, SVM has a better accuracy overall for prediction of churners and non-churners respectively. With the help of different marketing approaches, decision makers can retain churners. Analysis can be performed on real-time data and customer relationship management companies for customer acquisition.

(Meher et al.; 2017) suggests large data volumes with high dimensionality of the customer data is challenging. In this paper, by reducing the dimensionality and decision tree algorithm, churn management system was prepared. Aggregated key performance indicators (KPIs) were consider as key features and feature selection techniques was used for in-depth analysis. Apache Oozie workflow engine was used for yielding faster results. Practical issues like class imbalance, high dimensionality, campaign actionability were covered during the process.
KPIs were extracted using PIG script using Hadoop cluster. Then, KPIs were divided into two categories: (i) Snapshot and (ii) Trending. Snapshot contain single value while trending contains multiple KPIs respectively. Wrapper method was used for forward selection and new techniques like recency-based filtering and computation-intensive filtering is used for feature extraction for space. The above techniques were used because, when some KPIs have unique feature of patterns. For example; when selected together, they show better meaningful behavioral patterns then considered distinctly. Machine Learning techniques for analysis like Logistic Regression, Random Forest, Support Vector Machine, Random Sup-space. Decile Wise Cumulative Coverage (DWCC) was to be used for accessing the overall model quality. For future work, Principal Component Analysis (PCA) can be used for further dimensionality reduction as well as Incremental learning can also be used with the help new instances when model is trained and tested every time for robust patterns.

(Vafeiadis et al.; 2015) different a comparative study of various machine learning methods.
Machine Learning techniques like Naive Bayes, Logistic Regression, Bayes Classifiers, Support Vector Machine, Decision Trees and Artificial Neural Networks classifiers were applied and then cross-validation was done on the domain dataset. After that, boosting was compared of the above classifiers for performance improvement and then main part parameter

combinations were performed on Monte Carlo Simulations for wider range of parameters. To improve classification performance through various machine learning models, boosting was performed. Through F-measure effectiveness of the classifiers was measured.
Weak classifiers were combined for building a more accurate classifier for training the dataset. AdaBoost algorithms is the most and powerful boosting algorithm and better than many classifiers. In this project, AdaBoost.M1 with Decision Tree and Back-Propagation algorithm (BPN) both as weak classifiers. R language was used for simulation for testing of classifiers. For cross validation, 100 Monte Carlo realizations were generated from various parameter for each specific classifier. For tuning, free parameters is must. This work, provided some idea about churning prediction problem and supported the idea of boosting techniques. For future work, different boosting algorithms other than AdaBoost and weak learners for Ada.BoostM1 algorithm needs to be examined. And, for more accuracy, large dataset will be used for maximizing the statistical importance of all the classifiers.

(Quershi et al.; 2013) says, churn prediction has come out as one of the most important Business Intelligence application who are about to churn. Machine Learning algorithms such as Artificial Neural Networks, Decision Trees, Regression Analysis are done of the dataset. To solve the problem of class imbalance, re-sampling method was applied. As many companies are moving or have already moved on towards customer strategy model from business strategy model. A churn prediction model was proposed which helps in customer who are on the risk of churning and must be retained and dealing with the class imbalance by re-sampling methods were two important things discussed in this paper. To check the performance of various prediction models, F-measure, Precision, recall were the evaluation methods used. For determining which variable has most predictive significance to the target variable Pearson's Correlation test was done. After applying the machine learning algorithms, SPSS was used to check the various variations of decision trees; Chi-squared Automatic Interaction Detector (CHAID), Exhaustive CHAID, Classification and Regression Trees (CART), Quick, Unbiased and Efficient Statistical Tree (QUEST) was performed. After detailed analysis of everything, Exhaustive CHAID was the most accurate variation of decision tree. After adding five new variables, recall rose highest above another measure.
In future work, same approach would be done on bigger datasets and analysis will be done for longer time.

(Brandusoiu et al.; 2013) suggests, predictive models can play a decisive role in churn prediction. In this paper, Support Vector Machine (SVM) machine learning algorithm was implemented with four kernel functions for predictive analysis. Gain measure was done at last to compare all the models. Four SVM kernels used in this paper are RBF, Linear, Polynomial and Sigmoid. After implementing SVM, RBF, linear and Polynomial kernels performed almost the same while sigmoid performed the lowest. To cross check more, gain measure was performed to compare various SVM models. By using, cumulative gain chart, which is compared with random expectation line shows models that use RBF and polynomial kernel functions perform better than linear and sigmoid. In future work, decision makers can build different approaches with different predictors to retain churners and thus by improving model performance.

(Brandusoiu et al.; 2016) predicted an advanced methodology to predict churners using machine learning algorithm by applying Neural Networks. ROC curve and gain measure was used for the evaluation and performance analysis. The multilayer perceptron (MLP) have been for classification for linearly inseparable data. For training data, MLP neuron of neural networks with the help of Back Propagation (BP) algorithm have been used. By using back propagation algorithm, multilayer perceptron can be extended to multiple layers. For deeper analysis, the gradient descent, the weight initialization and convergence criteria algorithms were implemented. And lastly, pruning strategy was checked for sensitivity. In future work, this model can be implemented for acquisition, cross and up selling the products.

(Ismail et al.; 2015) proposed a Multilayer Perceptron (MLP) neural network for prediction of churners. Performance of the models were measured on 3 measures; sensitivity, specificity and accuracy. Accuracy is measured by the number of correct predictions divided by number of all predictions. Sensitivity is measured by the number of true positive predictions divided by number of all positive predictions. And, specificity is measured by the number of true negative predictions divided by number of all negative predictions. MINTAB was used for the analysis of logistic regression. For training algorithm, various MLP neural network were implemented. After thorough analysis, the results are compared against Multiple Regression and Logistic Regression Analysis. With the use of Levenberg Marquardt (LM) machine learning algorithm, neural network had a good overall accuracy. In future work, this model can be used for real-time data and big dataset for robust performance.

(Chouiekh; 2017) focused on various types of customers like unreliable group of customers, less committed to a provider, prepaid customers, who end service without prior notice. Various Machine Learning techniques was used for implementation like Naive Bayes, Decision Tree, Gradient Boosting Classifier, Support Vector Machine, K-nearest neighbors, Logistic Regression. For better evaluation of all the classifiers, four independent measures were analyzed; precision, recall, accuracy and F1 measure. To evaluate each measure, True Positive, True Negative, False Positive, False Negative scores were calculated. Calculation of each measure was done separately and then performance rate was compared between different algorithms. Analysis showed all the algorithms have roughly the same accuracy except slightly more in the range of 7-8% for Decision Tree. Although for F1 score, Logistic Regression and Gradient Boosting Classifier have the highest value. However, when the recall measure was added, Gradient Boosting Classifier showed the best accuracy compared to others. In the future work, deep learning techniques like image recognition and natural language processing can be used for more detailed performance as different regions, different locations have different churn rates, thus decision makers need to check out this situation and stop consumers from churning.

(Umayaparvathi and Iyakutti; 2016) recommends for telecom industries to take all the necessary steps regarding customer churn. In this paper, a churn prediction methodology with the help of predictive models were proposed. As well as, detailed analysis of various attributes was studied for accuracy. Two datasets were used for cross validation. Machine learning classifiers like; Decision Tree, Support Vector Machine, Gradient Boosting, Random Forest, K-nearest Neighbor, Ridge Regression Classifier and Logistic Regression were implemented. Through analysis were done with the help of metrics like; Recall, Precision, F1-score, Confusion Matrix and Accuracy score. For coding, Python was used with scikit-learn, pandas, NumPy and various other libraries. The entire workflow was implemented in I Python Notebook for easy browser interaction. For validation of above mentioned classifiers, 10-fold cross validation and stratified cross validation was implemented to avoid unbalancing of the datasets. Based on the analysis, Gradient Boost classifier outperforms other classifiers. After analyzing each attribute, it was observed only 6 attributes have maximum importance of churn prediction. In future work, call records, customer demographic can be observed instead of using all the predicted variables.

(Dalvi et al.; 2016) mentioned that the limelight has shifted to customer retention from customer acquisition. In this paper, machine learning algorithm like, Logistic Regression and Decision Trees are applied, and a comparative study has been made based on the accuracy of the available data. R programming was used for building the churn prediction model. Shiny package was used for web interface. The system was divided into 3 parts for detailed analysis: View performance analysis, testing and training. View performance analysis was obtained by applying machine learning algorithms. Testing is that part where attributes are examined for prediction of churn and training is the building of model where all the analysis is done on the dataset. In this paper, just a statistical analysis was compared between Logistic Regression and Decision Tree. Decision makers can focus on retention policies and reason behind churn for the future work by analyzing various attributes.

(Yabas and Cankaya; 2013) implemented various machine learning techniques on public dataset for the KDD Competition. Focus was on meta-classifiers were analyzed based on their performances. Although competition is over, it is still open for submissions and possible further improvements. Results are compared with the help of ROC curve as it is independent from class ratios. Meta-classifiers proposed from the preprocessed dataset. After testing, a combination of meta-classifiers for good performance. WEKA tool was used for voting classifiers. Variety of classifiers with plenty of algorithms was used for in depth analysis. Ada-Boosting, Support Vector Machine, Neural Network were slow in model building so was eliminated. Decision Trees, Logistic Regression, Bagging with Decision Tree and Random Forest was implemented. Self-made meta classifier was used for churn analysis and prediction, voting classifier was performed on that classifier. Proposed model equals the ROC area score and results showed proposed models, classifier performance increases along with the voting meta-classifier. However, threshold value used for classifiers is not yet determined, thus new threshold value gives new results. In future work, multiple classifiers need to apply for various algorithms.

(Prashanth et al.; 2017) implemented data mining techniques like Logistic Regression, Random Forest and Deep Neural Network was used for churn prediction. In this paper, a comparative study of linear and non-linear techniques was carried out. Parameters were selected through 5-fold cross-validation. Performance metrics like accuracy, specificity, sensitivity, ROC and area under curve compared. After analysis, non-linear techniques performed better than linear techniques. Random Forest have the best accuracy, specificity and area under the curve, while

Recurrent Neural Network have the best sensitivity. In future work, above techniques can be used on different datasets with different parameters.

(Rodan et al.; 2015) ensembled a Multilayer Perceptron (MLP) then a negative correlation learning for numerous parameters for churn prediction. In NCL, simultaneously all the networks are trained and calculated in correlation penalty measures in the error functions.
Penalty factor and number of hidden nodes were compared by 5-fold cross validation with size of networks equals to 10. The ensembles measures were then compared with data mining techniques like; Random Forest, C4.5 algorithm, K-nearest Neighbor, Support Vector Machine. As churn data is unbalanced, NCL was compared with other techniques for class imbalance; AdaBoost, MLP for cross-sensitive classification, Bagging algorithm.
Analysis showed, NCL have the highest accuracy but accuracy is not the best evaluation metric due to class imbalance. Churn rate for Naive Bayes showed the highest significance but it was worst in terms of accuracy. Thus, giving a massive advantage of Negative Correlation Learning over Naive Bayes. Overall, Negative Correlation achieved better performance with the ensemble of MLP. In future work, other machine and deep learning models can be examined for better performance.

(Huang et al.; 2012) did a detailed comparative study between various machine learning techniques with the evolutionary data mining algorithm. Techniques like Linear Classification, Support Vector Machine, Neural Networks, Multilayer Perceptron were implemented with the help of new variables. As, Churn Prediction is common, it has affected land-line telecom industry a lot. Henley Segmentation, call and bill details, complaint info etc. are the new variables used in this paper. Data mining techniques were used as predictors and comparative study showed new proposed variables were more effective with the Decision Tree and Support Vector Machine for analyzing churn rate and for churn probability Logistic Regression was preferred. In this future work would be to add new variables like; fault reports, contract info etc. with the help of dimension reduction and more methods on selection techniques.

(Zhang et al.; 2010) suggests many factors affect customer churn. In this paper, network attributes are taken for analysis. These network attributes are linked with prediction models that can be useful in building machine learning algorithms. This effect the accuracy of churn prediction and thus by comparing and structuring them for model effectivity. Prediction models like Logistic Regression, Decision Tree, Neural Network are compared and training with SAS

Enterprise Miner. Variety of different combinations was chosen and with the help of Gini Reduction, Decision tree was split to run multi-layer perceptron. From the analysis, it was concluded that network attributes improve accuracy of prediction model. In future work, topologies, calling behaviors that are important variables of network attributes can be analyzed with other network and traditional based attributes.

3. **Methodology:**

The dataset that would be used in this research was acquired from IBM Watson Analytics community. The dataset consists of 7044 customers (active and disconnected).
The dataset has the following information:

- Churn column is the key column as it shows which customers have left within the last month.
- Customer that have signed up for which services like (Phone Service, Multiple Lines, Internet, Online Security, Online Backup, Tech Support, Streaming TV, Streaming Movies, Device Protection).
- Information about customer account like (Contract, Payment Method, Paperless Billing, Monthly Charges, Total Charges)
- Customers demographic information like (Gender, Senior Citizen, partner, Dependents).

Churn is the key attribute. The customers in the dataset are classified by a dichotomous variable called Churn (Yes or No). A customer will be categorized as Active (Yes) if he/she continues to use the same network (non-churner). Whereas, a customer will be categorized as Non-Active (No) of if he/she discontinues the network (churner).

Following algorithms will be implemented in the above dataset:

3.1 **Determining the Important Variables:**

For the basic overview of the dataset, before applying the machine learning algorithms, one of the vital steps is to select right group of variables. Those variables are known as predictors. To determine which variable, whether a variable has any importance in doing prediction analysis, we need to calculate its p-value with the selected target variable. p-value in simple terms is defined as the probability that the in statistical terms the sample data which is observed has the null hypothesis value true. If the value of p is less than 0.05, null hypothesis is rejected. Thus, to get the decide the best set of variables for the prediction, variables whose p-value is above 0.05 were rejected.

3.2 Determining the Correlation between the Variables:

Next basic overview of the dataset is to determine the correlation. As, it is one of the most important measure to test the dependence between the two variables. Although there are many statistical tests to find out about correlation. In this research, Kendall Tau correlation test is preferred because the standard error is known, and corresponding population parameter is at a better estimation rate. As, it is used in large samples and when the values of the selected variables are generally in the same range. Tau correlation is analytical and elucidatory as the percentage of the data points of the selected pair of the variables are mostly positive. So, the tau-value of 0.6 or greater than 0.6 means the correlation is positive and tau-value of -1 means the correlation is negative between the selected variables. Variables with the highest Kendall Tau coefficient status will be selected for further analysis.

3.3 Checking about Class Imbalance:

Different dataset has different features. Some of the features, might pose problem while implementing the machine learning algorithms, to extract meaningful data from the dataset. Next vital step for the overview of the dataset is to check whether, class is imbalance or not. Because if it is imbalanced, then ratio of output categories will be one-sided and then the machine learning techniques will only predict the majority class thus by giving not so accurate analysis. To check about class imbalance, re-sampling is done. Re-sampling can be done in two ways: (i) over-sample and (ii) under-sample. Random over-sampling is generally preferred above under-sample because it is used to increase the strength of minority class. Thus, technique will provide a detailed view of the class imbalance thus by helping in accurate analysis of churners.

3.4 Logistic Regression:

In our research, regression techniques like Linear or Logistic can be used. But, as the target variable in our dataset (churn) is categorical, logistic regression is preferred. If the variable was continuous dependent, then linear regression would have been a better choice. In the research data, (churn) variable is dichotomous while most of the other variables are independent but continuous in nature, that's why logistic regression was used for analysis. While performing the data mining techniques, customer churning or likely to churn was observed and conditional probability model was formed for detailed analysis. As in the research dataset with the use of continuous variables, by using logistic function a straight line linear regression equation is passed to get the conditional probability model.

3.5 K-Means Clustering:

K-means clustering is one of the most widely machine learning algorithm used for clustering. It will classify the data into k different clusters. Before executing, the value of k needs to be specified. First, we need to identify the k-number of seeds, this can be done by k different observations and, assign them as seeds. Based on their proximity, remaining observations were assign the rest of the seeds. The proximity is calculated using distance or similarity. In this research, Manhattan Distance and Cosin Similarity will be used. The cluster optimization, the k means algorithm goes through n number of iterations. In each iteration, each cluster with its centroid is recalculated as iteration, which are known as iteration points. And, those iteration points are designated as seeds. From this recall and F-measure values will be calculated on those bases of selected variables.

3.6 Random Forest:

Random Forest is an ensemble of decision trees. They are created individually by means of randomly drawing the same number of samples from the training dataset. Decision Trees include 3 types of nodes: (i) root node, (ii) Internal node and (iii) End nodes. It has only one root node which is the set of whole data. Internal node is a main split which splits all the arrival samples by a selected feature. The root-to-end-node is known judge rule. Decision tree technique is used because it applies to top-to-bottom greedy algorithm. After this, the best split feature is chosen by each internal node to divide as well as arrival samples into 2 parts or maybe more. The split feature is chosen by Iterative Dichotomizer 3 (ID3). Random Forest replicates the above steps to build multiple decision tree which can used for 5-fold cross validation between the variables. After this, Quick, Unbiased and Efficient Statistical Tree (QUEST) will be performed for prediction accuracy with the help of recall and specificity rate for the selected variables.

3.7 Singular Value Decomposition:

SVD is a dimension reduction technique that help us understand which principal components can contribute a robust computing framework which can provide us a better analysis with improved accuracy. This, left singular vectors can be imagined as the weights which can be projected on frequencies terms, to get the SVD projections. These components are serial numbers which can be use in implementation of predictive models. When normalized, SVD components can be help in calculating important measures like similarity scores and Euclidean distance between the variables. Thus, by finding the variables which are same and later which can be used in clustering.

4. Project Plan:

As per Gantt Chart the future 14 weeks plan is shown below. The Gantt Chart provided below shows the tasks and timelines for the research project within the required time.

Task Name	Start Date	Days to Finish
Research Project	23-Sep	5
Data Collection	23-Sep	12
Data Cleaning and Extraction	19-Oct	1
Data Transformation	20-Oct	4
Data Modeling	25-Oct	13
Data Evaluation	9-Nov	2
Data Mining Techniques	12-Nov	10
Deep Learning Techniques	19-Nov	9
Project Milestone	25-Nov	9
Research Project	30-Nov	4
Literature Review	6-Dec	2
Proof Reading	8-Dec	1
Final Report Preparation	12-Dec	3

Fig.2 Gantt Chart Tasks

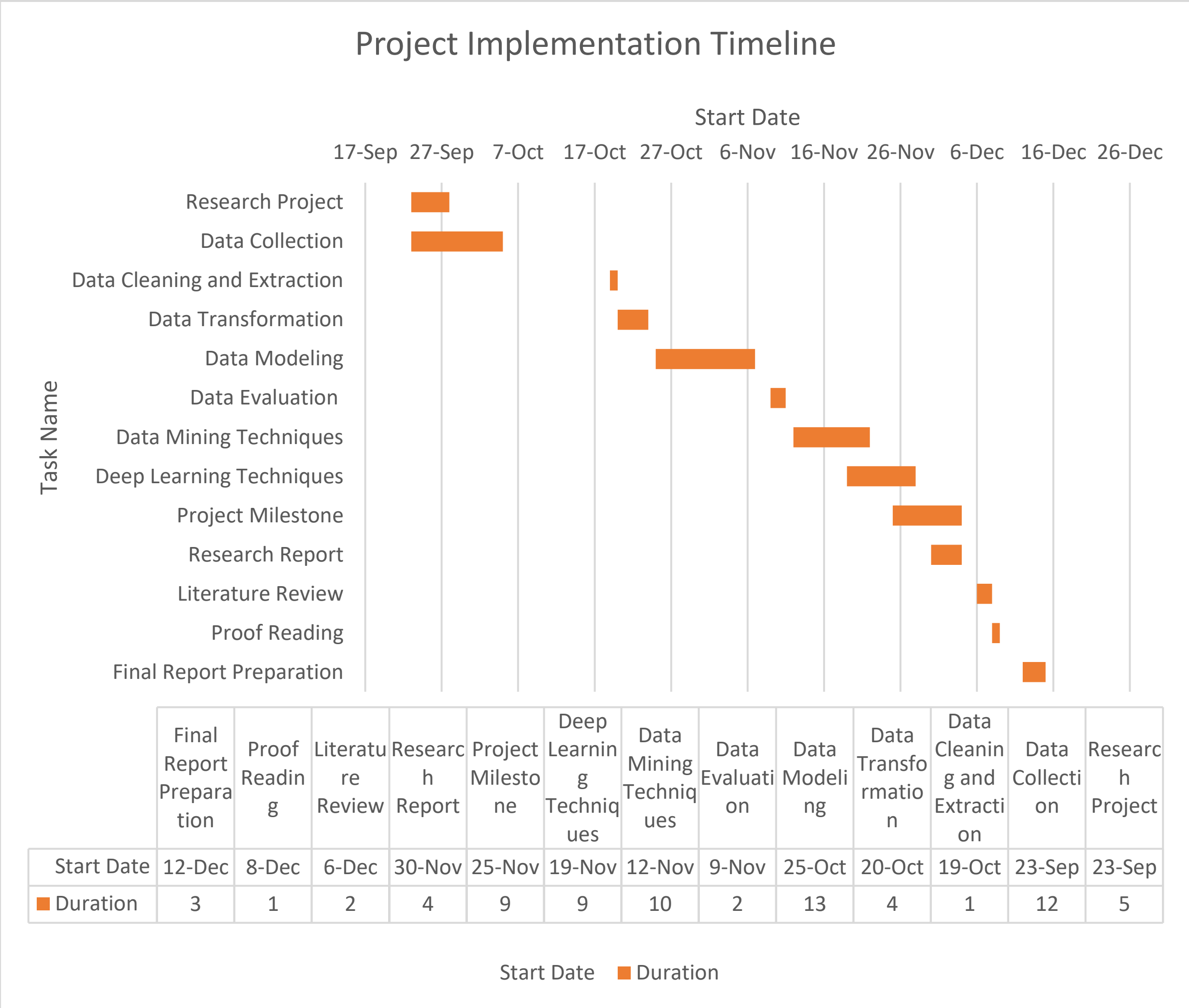

Fig.3 Gantt Chart Timeline

References:

Khan, F. and Kozat, S.S., 2017, May. Sequential Churn Prediction and Analysis of Cellular Network Users–A multi-class, multi-label perspective. In 2017 25th Signal Processing and Communications Applications Conference, SIU 2017. Institute of Electrical and Electronics Engineers Inc.

Yihui, Q. and Chiyu, Z., 2016, August. Research of indicator system in customer churn prediction for telecom industry. In Computer Science & Education (ICCSE), 2016 11th International Conference on (pp. 123-130). IEEE.

Brândușoiu, I., Toderean, G. and Beleiu, H., 2016. Methods for churn prediction in the pre-paid mobile telecommunication industry.

Meher, A.K., Wilson, J. and Prashanth, R., 2017, July. Towards a Large Scale Practical Churn Model for Prepaid Mobile Markets. In Industrial Conference on Data Mining (pp. 93-106). Springer, Cham.

Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C., 2015. A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, pp.1-9.

Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A. and Rehman, A., 2013, September. Telecommunication subscribers' churn prediction model using machine learning. In Digital Information Management (ICDIM), 2013 Eighth International Conference on (pp. 131-136). IEEE.

Brandusoiu, I. and Toderean, G., 2013. Churn prediction in the telecommunications sector using support vector machines. Margin, 1, p.x1.

Brandusoiu, I.B. and Toderean, G., 2016. Churn prediction in the telecommunications sector using neural networks. Acta Technica Napocensis, 57(1), p.27.

Ismail, M.R., Awang, M.K., Rahman, M.N.A. and Makhtar, M., 2015. A multi-layer perceptron approach for customer churn prediction. International Journal of Multimedia and Ubiquitous Engineering, 10(7), pp.213-222.

Chouiekh, A., 2017, April. Machine Learning techniques applied to prepaid subscribers: case study on the telecom industry of Morocco. In 2017 Intelligent Systems and Computer Vision (ISCV) (pp. 1-8). IEEE.

Umayaparvathi, V. and Iyakutti, K., 2016, March. Attribute selection and Customer Churn Prediction in telecom industry. In Data Mining and Advanced Computing (SAPIENCE), International Conference on (pp. 84-90). IEEE.

Dalvi, P.K., Khandge, S.K., Deomore, A., Bankar, A. and Kanade, V.A., 2016, March. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In Colossal Data Analysis and Networking (CDAN), Symposium on (pp. 1-4). IEEE.

Yabas, U. and Cankaya, H.C., 2013, December. Churn prediction in subscriber management for mobile and wireless communications services. In Globecom Workshops (GC Wkshps), 2013 IEEE (pp. 991-995). IEEE.

Prashanth, R., Deepak, K. and Meher, A.K., 2017, July. High accuracy predictive modelling for customer churn prediction in telecom industry. In International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 391-402). Springer, Cham.

Rodan, A., Fayyoumi, A., Faris, H., Alsakran, J. and Al-Kadi, O., 2015. Negative correlation learning for customer churn prediction: a comparison study. The Scientific World Journal, 2015.

Huang, B., Kechadi, M.T. and Buckley, B., 2012. Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), pp.1414-1425.

Zhang, X., Liu, Z., Yang, X., Shi, W. and Wang, Q., 2010, July. Predicting customer churn by integrating the effect of the customer contact network. In Service Operations and Logistics and Informatics (SOLI), 2010 IEEE International Conference on (pp. 392-397). IEEE.