Customer Persistence Modelling for Insurance Firm
Prof. Dr. Mrs M Vijayalakshmi¹ Aakash Mishra² Rohit Singh³ Nitin Lulla⁴ Suraj Bhatia⁵
¹,2,3,4,5Department of Information Technology
Vivekanand Education Society’s Institute of Technology, Mumbai, India

Abstract— Customer relationship management (CRM) is a system for managing a company’s interactions with current and future policy holders. It involves using technology to organize, automate and synchronize sales, marketing, customer service, and technical support. CRM systems track and measure marketing campaigns over multiple networks. These systems can track customer analysis by customer clicks and sales. Places where CRM is used include call centres, heavily utilized in social media, direct mail, data storage files, banks, and customer data queries. Problem statement -Within a company's customer relationship management strategy, finding the policy holders most likely to leave is a central aspect. There arises a need for a dynamic modelling approach for predicting individual policy holders’ risk of leaving an insurance company. Churn prediction is a very powerful weapon to know soon which policy holders will abandon the company, in order to increase the degree of retention, to build a solid CRM strategy and to save acquisition costs between five and fifteen times more expensive to gain new policy holders than to retain the current ones.

Key words: Churn, Churn Rate, Training Data, Testing Data, Confusion Matrix

I. INTRODUCTION

Data mining [1][2], the analysis step of the "Knowledge Discovery in Databases" process, is an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records unusual records and dependencies.

We can summarize the economic value of customer retention as:

1) Lowering the need to seek new and potentially risky policy holders, which allows focusing on the demands of existing policy holders.
2) Long term policy holders tend to buy more.
3) Positive word of mouth from satisfied policy holders is a good way for new policy holders' acquisition.
4) Long term policy holders are less costly to serve, because of a larger database of their demands.
5) Long term policy holders are less sensitive to competitors' marketing activities.
6) Losing policy holders results in less sales and an increased need to attract new policy holders, which is five to six times more expensive than the money spent for retention of existing policy holders.
7) People tend to share more often negative than positive service experience with friends, resulting in negative image of the company among possible future policy holders.

II. LITERATURE SURVEY

A. Origin of Concept:
The planning that helps management to cope with the uncertainty of the future, depending mainly on data from the past and present and analysis of trends. The typical approach to the problem of churn prediction is using a sufficiently large data set that contains churning and non-churning customers. This set is being analysed to construct a classifier. The work of a classifier is to decide, given a customer data set, if churn is more or less likely. Such classifiers are constructed using, for instance, neural networks, Bayesian statistics or decision trees constructed with the heuristics like CART or C4.5.

Policy holders who churn can be broadly classified as: Mandatory and Voluntary.

Mandatory churn is a case where the customer misuses the service or does not pay his bills. Whereas, in Voluntary churn it is the decision of the customer to switch to another company or service provider. There are three types of churn:

1) Active / deliberate - the customer decides to quit his contract and to switch to another provider. Reasons for this may include: dissatisfaction with the quality of service (e.g. not fulfilling service level agreements), too high costs, not competitive price plans, no rewards for customer loyalty, no understanding of the service scheme, no information about reasons and predicted resolution time for service problems, no continuity or fault resolution, privacy concerns, etc.

2) Rotational / incidental - the customer quits contract without the aim of switching to a competitor. Reasons for this are changes in the circumstances that prevent the customer from further requiring the service, e.g. financial problems, leading to impossibility of payment; or change of the geographical location of the customer to a place where the company is not present or the service is unavailable.

3) Passive / non-voluntary - the company discontinues the contract itself.

B. Approaches That Can Be Used For Predicting Churn:

1) For Clustering [8]:

1) K-Means [6][7] - In k-means method the algorithm usually works on an iterative mode. The premise for the iteration is the assignment of the data point to each cluster, based on the minimum Euclidean distance from the K-cluster centroids to data points. The cluster centroids become more defined as the data points in the cluster change based on the minimum distance calculation.
2) X-Means [3]- X-means is K-Means extended by an Improve Structure part. Extending K-means with Efficient Estimation of the Number of Clusters. Despite its popularity for general clustering, K-means suffers three major shortcomings; It scales poorly computationally, the number of clusters K has to be supplied by the user, and the search is prone to local minima. X-MEANS propose solutions for the first two problems, and a partial remedy for the third.

3) K-Medoids – It is a classical partitioning technique of clustering that clusters the data set of n objects into k clusters known a priori. A useful tool for determining k is the silhouette. It is more robust to noise and outliers as compared to k-means because it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances.

2) For Classification:
1) Decision trees - A decision tree is a flowchart-like structure in which each internal node represents a “test” on an attribute, each branch represents the outcome of the test and each leaf node represents a class label. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities.

2) Naïve Bayes [4] - Naïve Bayes is a type of supervised-learning module that contains examples of the input-target mapping the model tries to learn. Such models make predictions about new data based on the examination of previous data. The Naïve Bayes algorithm uses the mathematics of Bayes’ Theorem to make its predictions.

3) K-Nearest Neighbour [5] - In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

3) Other Methods:
1) Use Ensemble Classification – The Approach of using 2 or more classifiers together to increase the accuracy of the prediction is called ensemble classification.

III. PROPOSED SYSTEM
In an insurance firm, policy holders play a vital role in overall progress and development of the organisation. In such a scenario accurately predicting the policy holders who are about to churn the organisation, can provide a great competitive advantage and thus necessary preventive measures can be taken to retain those policy holders.

A. Challenges
1) Predicting the churn rate in order to anticipate.
2) Detecting which policy holders are about to abandon.
3) Knowing the real value of the potential loss of those policy holders.
4) Taking the necessary preventive measures to reduce or avoid their abandonment.

B. Proposed Solution
1) The project will apply advanced analysis techniques into practice with the aim of identifying which variables define the churners and which don't.
2) Applying the Clustering algorithm to divide the policy holders into different clusters.
3) Finding the most optimal and accurate solution to the churn prediction results using combination of classifiers.
4) Classifying the policy holders further based on probabilities into High, Medium, and Low Churners.
5) Displaying the results in the form of Bar Graph, Pie Chart, Line Chart and Scatter Diagram.

C. Flow of the System:

![Flow of the System](image)

1) Block 1: Collection of Data:
   - Attributes to be considered:
     1) Customer behaviour identifies which parts of the service a customer is using and how often is he using them. Interesting are product-specific ownership (which product/service is owned/on loan), total product ownership (number of products owned/on loan), inter purchase time (time between the purchase of two different articles).
     2) Customer perceptions are defined as the way a customer apprehends the service. They can be measured with customer surveys and include data like overall satisfaction, quality of service, problem experience, satisfaction with problem handling, pricing, location convenience, image/reputation of the company, customer perception of dependency to the vendor, etc.
3) Customer demographics are some of the most used variables for churn prediction. They include age, gender, level of education, social status, geographical data, etc.

4) Macro environment variables identify changes in the world, different experiences of policy holders, which can affect the way they use a service. For example, in the telecommunication industry people who have survived a natural disaster and could rely on their mobile phones during it are more likely to continue using the service.

2) Block 2: Pre-Processing:
After successful collection of data, Pre-processing the data to remove outliers, inaccurate, missing values present, as raw data is high susceptible to noise, missing values and inconsistency. The quality of data affect the mining result. So to improve the quality of the data collected, pre-processing is required.
- Tasks Involved:
  1) Data Cleaning – Involves handling of missing values, noisy data, inconsistent data.
  2) Data Integration – Data analysis will involve data integration, which combines data from multiple sources into coherent store.
  3) Data Transformation – Data are transformed or consolidated into forms appropriate for mining. It involves tasks like Normalization, Smoothing, Aggregation, and Generalization.
  4) Data Reduction – Mining on huge amounts of data may take a very long time. Data reduction techniques reduces the representation of data without compromising the quality of data.

3) Block 3: Clustering the Dataset:
After pre-processing, the data is ready to be used by the system for further analysis. Clustering is proposed to identify set of similar policy holders so that it becomes easy to identify the reasons of churning at the later stage. X-Means clustering algorithm is proposed to be used to accurately cluster the dataset in appropriate number of cluster, as X-Means has built in cluster evaluation (determination) algorithm based on Bayesian Information Criteria (BIC). Euclidean distance is used as measure for distance calculation between different attributes. Based on closest centroid to the distance calculated, the attributes are clustered in different groups.

4) Block 4: Classify The Clustered Data:
Ensemble classification is used to accurately classify the records as churners and non-churners. Algorithms used are K-Nearest Neighbour and Naïve Bayes Classification. Each cluster is given to the K-Nearest Neighbour algorithm as input to classify it as churners or non-churner. The output of this is given as input to Naïve Bayes algorithm to eliminate the non-churners which are incorrectly classified as churners by K-Nearest Neighbour, increase a level of classification and thus improve accuracy.
- Data used for testing accuracy: Churners -25 Non-Churners - 18
- Confusion Matrix after K-Nearest Neighbour –
  - 22 (True positive) 3 (True negative)
  - 8 (False Positive) 10 (False Negative)
- Confusion Matrix after Naïve Bayes –
  - 20 (True positive) 5 (True negative)

5) Block 5: Classifying The Churners As HIGH, MEDIUM And LOW Churners:
Once the final records related to churners are obtained after Naïve Bayes Algorithm along with their probabilities of churn, The Churners are to be classified as High probability churners or Medium probability churners or Low probability churners.
Threshold used for Classification:
- X > 0.500 and X<= 0.660 – LOW
- X > 0.660 and X<= 0.810 – MEDIUM
- X > 0.810 – HIGH
  Where, X is probability of an x churner

6) Block 6: Representing The Results In The Form Of Graphs And Charts:
The results obtained of overall Churners and Non-Churners, Cluster wise churners and non-churners and Churners as HIGH, MEDIUM and LOW churners are to be represented in an attractive manner to the user. The graphs generated should be such that it should give a clear idea of the user about the policy holders he/she has to focus more on retaining.

IV. IMPLEMENTATION
Java is used as front-end language for complex analysis of the data and determining the results, because of its robustness, architectural neutrality, quick processing of data and built in support for data mining. MS-SQL Server is used in backend to collect the Training and Testing data as well as store the results of churn process back in to the database.
For Clustering, choice is provided between K-Means and X-Means for the manager. If he/she wants to specify the exact number of clusters to be formed then K-means is to be used otherwise if he/she wants the algorithm to decide upon the number of clusters X-Means is to be used. Combination of K-Nearest Neighbour and Naïve Bayes is used to improve the accuracy of churning as proved by the Confusion Matrix generated specified in Block 4 of FLOW OF THE SYSTEM. JFreeChart Application Programming Interface (API) is used for generating bar graphs and pie charts of the results obtained.

Fig. 2: Loading the Training and Testing Data from MS SQL Server
V. CONCLUSION

System developed will help the Insurance firm to quickly predict the policy holders about to churn the organisation, represented in a meaningful and easy to understand way. Also focus their attention first on the policy holders who have very high probability of churning classified in the category of HIGH Probability churners, viewing at the Cluster wise report for Churners, organisation can determine the reasons for churning looking at the attributes of the specific policy holder and determining the common characteristics among group of churners.

REFERENCES