Developing promotional model using Customer Lifetime Value score to avoid Customer Churns

MSc Research Project
MSc. In Data Analytics

Shrey Sanjay Shah
Student ID: X18192271

School of Computing
National College of Ireland

Supervisor: Mr. Hicham Rifai

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name:	Shrey Sanjay Shah		
Student ID:	X18192271		
Programme:	MSc. In Data Analytics	**Year:**	2019-2020
Module:	MSc. Research Project		
Supervisor:	Mr. Hicham Rifai		
Submission Due Date:	28th September 2020		
Project Title:	Developing promotional model using customer lifetime value score to avoid Customer churn		
Word Count:	6120	**Page Count:**	16

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shrey Sanjay Shah

Date: 28/09/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Developing promotional model using Customer Lifetime Value score to avoid Customer Churns

Shrey Sanjay Shah
x18192271

Abstract

Customer churn is one of the biggest problems that every telecom company is facing and measures are taken to avoid it. An approach to avoid customer churns has been implemented in this paper using the Customer Lifetime Value (CLV) data to give maximum waivers to eligible customers. CLV is basically defined as the total revenue that a customer has earned from start of his association with the company. Customer Lifetime Value (CLV) data has been calculated individually for each customer. These CLV scores are grouped in appropriate clusters with the help of k-means clustering algorithm. Then, all the probable churners are predicted with the help of decision tree resulting in an accuracy of 98%. All the customers that are predicted to be probable churners are given appropriate discounts or waivers based on the cluster assigned to them. Also, transfer learning has been implemented to check the consistency of the model over multiple datasets and it was observed that the performance of the model did not degrade significantly. These waivers and offers are provided to the customers well in advance before the customer churns, with an intention to retain maximum customers and avoid potential future churners.

Keywords – Telecom, Customer Churn, k-means clustering, decision tree, Customer Lifetime Value, Transfer Learning.

1 Introduction

Customer churn is one of the biggest financial concern among the rapidly growing telecom companies. Customer churn is defined as a situation where an existing customer switches to some other service provider that is offering better services. According to claims, telecom companies have to face financial losses of up to 1.6 trillion USD per year due to customer churn.[1] This figure is approximately 5 times more than the cost required by the companies to retain their existing customers. Also, to acquire a new customer and bring them to the same level as any existing customer, it could cost up to 16 times more. And if the company can succeed in retaining even a small number of customers of approximately 5% it could directly result in increase in overall profits ranging anywhere between 25% - 95%.[2] Thus, it becomes very important to figure out potential churners. In one of the approaches by (Gaur and Dubey, 2018), an attempt was made to predict the potential churners with the help of multiple

[1] https://www.superoffice.com/blog/reduce-customer-churn/#:~:text=Companies%20lose%20%241.6%20trillion%20per,level%20as%20an%20existing%20customer.
[2] https://hbswk.hbs.edu/archive/the-economics-of-e-loyalty

strategies and compare the performance of all the models. It was also observed that with proper feature selection and pre-processing, the models were able to predict the probable churners with some decent accuracy. Generally, there are various reasons leading to customer churning out of the company like bad customer support, competitive pricing, network coverage issues, etc. but the primary reason is that the customers are not able to cope with exorbitant charges. If these customers are given some waivers or promotional offers well in advance before they churn, it could most probably result in a smaller number of overall churners. But these offers cannot be generalised because every customer is different and has different customer value based on the lifespan and revenue earned by that customer and thus the offers should be personalised for every customer. This could be achieved by calculating the Customer Lifetime Value (CLV) score. CLV score is defined as the overall revenue earned to the company by a particular customer. There are various factors affecting the CLV scores like the lifespan of the customer, service plan, value-added services used by customer, etc. All this combined together contribute in calculation of CLV score for each customer. An approach by (Yonghui and Wanli, 2010) where various strategies are proposed to calculate the CLV and overcome the limitations of the existing conceptual CLV models. Various attributes such as profits earned, retention rate, discount rate, lifespan, etc. were used to build up the CLV calculation model. Thus, after successful calculation of CLV score, it becomes easier to segregate customers into higher valued customers and lower valued customers. Once the CLV scores are calculated, the customers with similar CLV scores could be clustered with the help of some clustering algorithm like k-means clustering. An approach by (Tolon and Aslan, 2018) where an attempt is made to categorise customers considering the associated product from around 30 products sold by the company. K-means clustering approach was adopted for implementation of the same and it was observed to have very good performance in accurately clustering the customers into higher and lower valued customers. On successful grouping of clustering algorithm, the customers who have earned higher profits for the company and been associated with the company from longer time period are all in higher clusters and likewise there are multiple clusters with the lowest cluster comprising of the customers who have earned least profits for the company. Now, after clustering it is possible to give waivers to each cluster separately based on the average CLV score in that cluster. Higher the average CLV score in a particular cluster, higher the overall waiver grant.

1.1 Research Question

"To what extent the CLV score could be used to personalise waivers and promotional offers for avoiding customers churning out of the telecom company?"

1.2 Research Objectives

Research Objectives for addressing the above-mentioned research question are derived as

- Analysing existing and already implemented approaches to tackle customer churn with the help of CLV and deriving the limitations and drawbacks of the same.
- Develop a comprehensive model for telecom industry using CLV calculation model, Decision tree for predicting churners and k-means clustering for grouping customers with similar CLV and finally providing customers with appropriate promotional offers.
- Implement transfer learning in order to assure the consistency of the model.
- Evaluate the performance of the models.

The next section of the paper that is Section 2 will widely discuss all the previous attempts and researches in the domain of customer churns and CLV. Section 3 will discuss about the research methodology. Section 4 will briefly discuss the system design and architecture. Section 5 will mainly discuss the implementation of the hybrid model and its integration in detail. Section 6 will discuss the evaluation of implemented model and comparison with existing models. Section 7 focuses on conclusions, limitations and future scope of the model.

2 Related Work

2.1 Predicting probable churners

Relevant feature selection plays a major role in accurately predicting probable churners. An attempt by (Varun *et al.*, 2019) to effectively choose features that are responsible to explain most part of the target variable for making the churn prediction. The data for this research was made available by Columbia with almost 3400 observations and 20 features. The pre-processing was a two-step process, where in the first step, data was cleaned. For data cleaning, multiple approaches like deleting identical or duplicate entries, handling missing or null values and removing all the redundant data were used. In the second step, feature reduction was performed, where correlation matrix was used. All the highly correlated attributes were either replaced or eliminated. After feature reduction, the remaining data was capable to explain 92% of the target variable. Thus, after appropriate pre-processing and feature reduction final data was ready for churn prediction.

Another similar attempt by (Gaur and Dubey, 2018) where the authors went a step further and implemented the churn prediction model. The data used for this research was a telecom dataset comprising of around 7000 data observations and just over 20 features. Similar to previous research feature reduction was done with the help of correlation matrix. After pre-processing, different machine learning models were used for churn prediction namely Random forest, logistic regression, gradient boosting as well as SVM. On comparison, it was observed that gradient boosting outperformed all other models with an accuracy of approximately 85%.

An attempt by (Lee *et al.*, 2018), where the authors implemented different churn prediction models in the online gaming industry and compare them. The three models used were Random forest, XG Boost and generalized boosting regression. Random forest was observed to have the best accuracy of 86%.

Another approach by (Hung, Yen and Wang, 2006) where the authors advanced further and used neural network for churn prediction to further improve the model performance. For the purpose of this research, data from a Taiwanese telecom company was used comprising data of customers over 150 thousand. The customers were first clustered with k-means clustering based on various parameters such as lifespan of customer, billing information, etc. Bach propagation neural network was compared with decision tree if it could provide any enhanced performance. It was observed that the neural network had slight edge over decision tree when evaluated with HIT and LIFT ratio.

Another attempt by (Swetha, Usha and Vijayanand, 2018), where the random forest algorithm was modified with an intention to improve overall performance of the model. Dataset was provided by a French telecom company Orange. The model was modified to fit one extra layer of random variables which could potentially reduce the problem of overfitting. The modified model had an accuracy of 92%.

2.2 Customer Lifetime Value

Customer lifetime Value (CLV) is one of the key attributes for the proposed research. There are various methods for CLV calculation. An attempt by (Yonghui and Wanli, 2010), where the customers are grouped into high and poor quality customers. The high quality being the customers who have already earned more than average profits to the company and poor quality being the customers who have earned less than the average expected profits to the company.

Another approach by (Chuang and Shen, 2008), where the authors try to recognize high value customers with an aim to further targeted marketing in order to improve the lifetime of each high valued customer. For calculation of CLV, RFM value was calculated for every customer after which it was possible to calculate individual CLV. Based on calculated CLV, customers are segmented, and then Artificial neural networks are used to identify the higher CLV group. This higher CLV group is the targeted group for promotional marketing.

One of the attempts by (Wang, Sanguansintukul and Lursinsap, 2008) using neural networks to predict CLV. For pre-processing, feature reduction was implemented using PCA and the features were reduced to 27 from 166. Artificial Neural network and Multi level perceptron neural networks are used to estimate the CLV scores. The implemented model was able to give an accuracy of 89%

An attempt by (Desirena *et al.*, 2019), tried to implement 2 stage stacked neural network to increase overall CLV prediction performance. Collaborative Metric Learning is implemented at first stage to identify 5 items responsible for maximizing the CLV out of 18 items available. At the second stage, using neural network-based survival analysis one product responsible for maximizing the profits is selected among the 5 items selected in stage 1.

Another approach by (Mauricio *et al.*, 2016) to forecast the CLV scores in a direct selling company consisting data of almost fifty sellers chosen on random basis and 4500 transactions. For recognizing average service life time of the company, binomial logistic regression was used. Also, the features crucial for churn prediction were identified. Churning customer was predicted with an accuracy of 94%. Following churn prediction, multiple linear regression was implemented to recognize all the attributes responsible for customer profits. Finally, customer profit prediction model was implemented using Multi-Layer Perceptron Neural network.

2.3 CLV Based Customer Segmentation

Grouping customers considering their calculated CLV scores is another key step because it is only after grouping that personalised offers can be given to each group. One approach for customer segmentation by (Ben Mzoughia and Limam, 2014) groups customer considering various other attributes than CLV like lifetime of the customer and count of transactions. Multicriterion segmentation model was used for grouping. This approach aimed to overcome limitations of traditional approaches of grouping customers solely on the CLV attribute. Also, to overcome some limitations of Multicriterion segmentation, genetic algorithm was used.

Another approach by(Tolon and Aslan, 2018) tried to cluster products from a B2B company using k-means clustering. Firstly, products were clustered based on target customers into 5 groups. Then the customers were grouped with k-means clustering algorithm. Kruskal Wallis Test was used for evaluation in which the null hypothesis was rejected meaning that the grouping of the customers was accurate enough. Also, in k-means clustering the difference between the class was 99% and within the class was 0.85% representing good clusters.

A unique approach of classifying the data based on missing data by (Nadella, Gupta and Pudkaje, 2020) done with the help of deep learning. In this research, an attempt is made to predict the gender of the subscriber in order to boost targeted marketing. Since gender is confidential information and not disclosed easily, it is essential to predict the gender for improved performance of any system. Various features were considered from a telecom company's dataset and the accuracy achieved in predicting the gender was 81.2%.

Another contribution by (Ali *et al.*, 2011) where the authors try to group the customers in real time with live data being fetched from a telecom company. The pre-processing like handling outliers and missing values is also done in real time using density decaying factor. D-stream clustering approach was implemented for clustering customers.

Another approach by (Abedzadeh and Nematbakhsh, 2012) to prevent customer churn in a Refah departmental store dataset comprising of approximately 15000 observations with 8 features. Firstly, the probable churners were identified with the help of decision tree. CLV score was then calculated for these customers with the help of C5.0 decision tree and then the customers were segmented based on these CLV scores. Finally, strategies were selected individually in order to avoid churns with the help of genetic algorithm.

2.4 Summary of previous work and proposed model overview

In general, most of the approaches were targeted on implementing one specific model in a particular area like Churn prediction or CLV score calculation or Clustering customers. Most of the previous approaches have managed to obtain a decent accuracy but with limited goals. Also, decision tree (Swetha, Usha and Vijayanand, 2018) and logistic regression (Mauricio *et al.*, 2016) were observed to have the best performance in predicting the probable churners. K-Means clustering was observed to be the best approach for clustering the customers based on calculated CLV scores (Tolon and Aslan, 2018). Thus, from all the previous researches, best models are used in this research and an attempt is made to implement a comprehensive model for churn avoidance in telecom industry.

3 Research Methodology

Knowledge Discovery Databases or KDD approach was followed for the purpose of this research. All the steps of the KDD methodology were implemented so as to acquire meaningful insights and analyse the data. The following section describes all the steps taken in this research.

3.1 Dataset Description and Selection

The dataset has been uploaded on KDD championship data[3] for developers consisting of small and large data. The dataset used in this paper was derived from KDD Orange data, cleaned and redundant data was eliminated and uploaded on BigML [4] [5]. This is a dataset of

[3] https://www.kdd.org/kdd-cup/view/kdd-cup-2009/Data
[4] https://bml-data.s3.amazonaws.com/churn-bigml-80.csv

French telecom company Orange which was made available to public. The dataset comprises of 20 features and around 3300 observations. This data was further modified in this research for making the data fit for models implemented.

3.2 Data Pre-processing, Transformation and Feature Engineering

- In data pre-processing, firstly, all the datatypes of the attributes was changed according to the requirement.
- Also, some string values were converted to Boolean values for example, feature Churn had values True and False and feature International Roaming had values Yes and No, all these values were converted to 0s and 1s.
- Also, the dataset was checked if it had any NULL or missing values.
- Some new features were calculated like total_mins, total_charge and total_calls since these features were missing and were very crucial further in the CLV calculation model.
- Some redundant features like State, Area code were eliminated to reduce the complexity.
- After calculation of some new features, the datatypes were changed where required.
- Number of calculations were performed to calculate CLV and as a result to store all these calculations, new features were added like avg_purchase_value, avg_purchase_freq_rate, customer_value, etc. After successful calculations, these features were no longer needed and hence eliminated.
- Also, same process of addition and elimination of features was implemented for clustering model.
- Finally, an attribute named offer was added that specifies how much discount the customer was eligible for in the upcoming bill cycle.

3.3 Data Mining

- After all the required data pre-processing and transformations, the data mining models could be applied on this dataset.
- A hybrid machine learning model was developed for the same where CLV calculation model, decision tree and k-means clustering were merged together.
- These models were observed to outperform any other models in the previous works.

3.4 Data Evaluation

- The dataset was split in 66% - 33% train and test data which performed best, after trial and error of multiple combinations.
- The evaluation matrix that was used to evaluate the performance of decision tree are accuracy, precision, recall, F1 score, etc.
- Also, for evaluation of K-Means Clustering, scatter plot was viewed and elbow method was used to find the ideal number of clusters.

---

[5] https://bml-data.s3.amazonaws.com/churn-bigml-20.csv

4 Design Specification

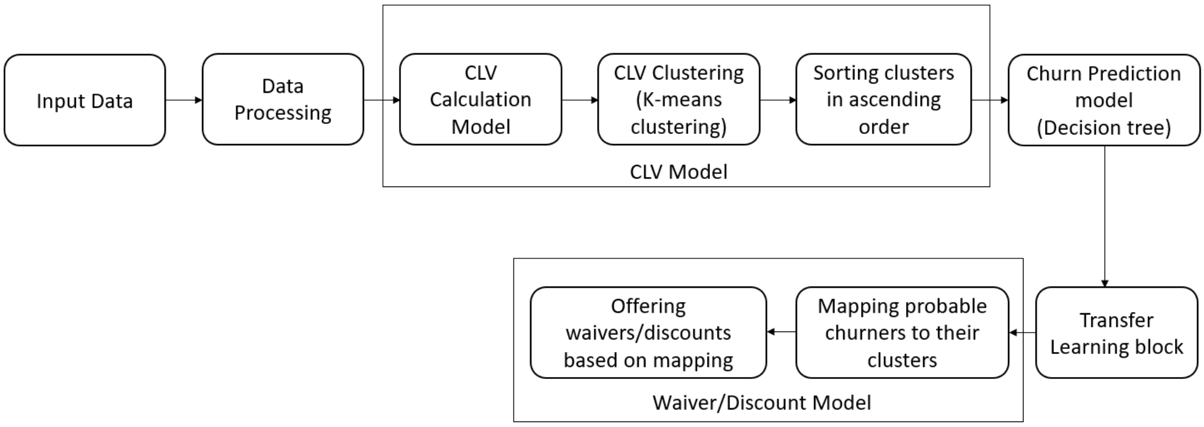

Figure 1: Model Architecture

The hybrid model was subdivided into three main components:
4.1 CLV model
4.2 Churn Prediction model
4.3 Waiver/Discount model

4.1 CLV Model:

The CLV model further comprises of 2 sections

4.1.1 CLV Calculation model:

In CLV Calculation model, 4 main values were calculated as follows:

- Average Purchase Value: Average Purchase value (APV) was calculated by averaging the total money spent by a particular customer during each day of the month. This gave a rough estimate of how much an average customer spends per day.

- Average purchase frequency rate: Average Purchase Frequency rate (APFR) was calculated by dividing number of customers and number of orders. This gave an estimate of how many orders an average customer makes each month.

- Customer Value: Customer Value (CV) was calculated by averaging APV and APFR. This gave an estimate of how much revenue an average customer earns to the company in a day.

- Average Customer Lifespan: Average Customer Lifespan (ACL) was derived by adding up all the customer lifespans and averaging it with the total number of customers. This gave an estimate of how long an average customer's lifespan is.

- Customer Lifetime Value: Customer Lifetime Value (CLV) was defined as a product of CV i.e. what was an average customer's value in a particular week and ACL i.e. what was an average customer’s lifespan. This resulted in the actual profit that a particular customer had given to the company from the start of his association with the company.

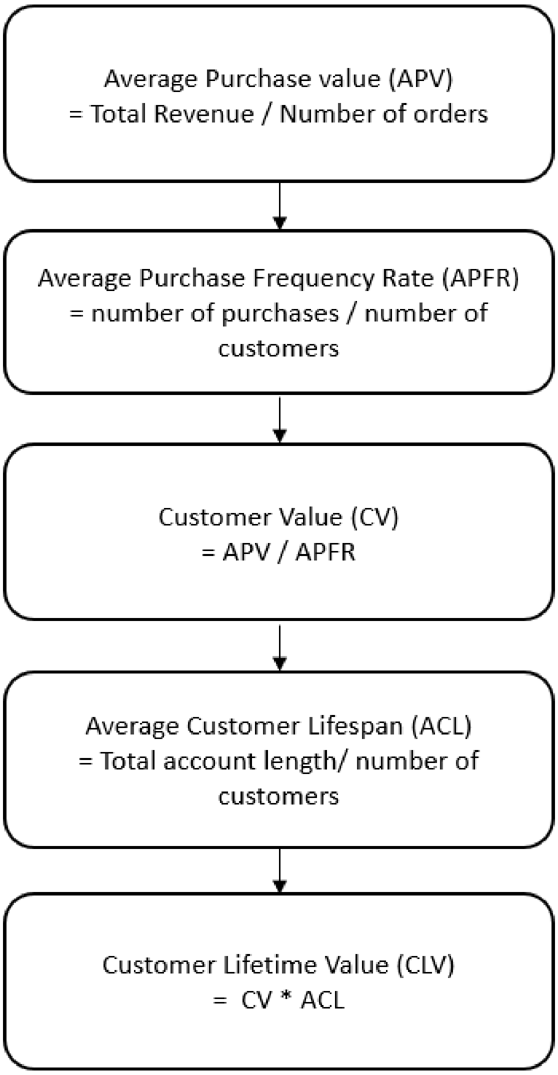

Figure 2: CLV Calculation Model

4.1.2 CLV Clustering model:

Once the CLV scores were calculated, the customers with similar CLV could be grouped together with the help of k-means clustering. This could bring all the customers with similar clv into same group making it easier to assign similar promotional offers as most of these customers had earned almost same revenue to the company.

K-Means Clustering:

K-means clustering algorithm clustering algorithm is used to group *n* number of observations in *k* number of clusters. These clusters were mostly based on one column and observations were grouped in each cluster in a way that they were closest to the mean of that particular cluster. Also, these clusters were created in a random order and need to be sorted in ascending order so as to perform further actions.

4.2 Churn Prediction Model:

Churn prediction was the heart of this hybrid model. It was used to predict which customer was most likely to churn out of the company. Various features were used to predict a churning customer. Decision tree classifier was one of the approaches that observed higher performance than other machine learning algorithms. For identifying the columns that help the maximum in explaining the target variable, correlation matrix was used so as to eliminate redundant columns and highly correlated columns that explain the same thing.

Decision Tree Classifier:

Decision tree classifier is one of the supervised machine learning algorithms which constantly keeps splitting the data based on a particular parameter passed. There are two types of decision trees namely classification trees and regression trees. When the target variable or the output variable is a categorical variable, mostly classification trees are used, whereas for continuous data types, regression trees are used. In this research, the output variable was Customer churn and it was a categorical variable i.e. it had 2 values either the customer will churn or not churn. Thus, Decision tree classifier was the perfect fit for the proposed research. In decision tree classifier, the tree was developed with the help of an iterative process of splitting the data into groups or partitions known as binary recursive partitioning. This process continuously executes splitting the data further and can be stopped after a certain interval to stop any kind of overfitting of data.

4.3 Waiver/Discount Model:

Finally, once the probable churners and their respective clusters were identified, these churners could be given some waivers and discounts on their following bill cycle based on the cluster they belong to. This was done by mapping the probable churners to their respective CLV-based clusters and then identifying the mean of that cluster in order to give some appropriate waivers such that the company doesn’t have to face any sort or losses and at the same time the chances of retaining that customer increases. Finally, after considering all these parameters, appropriate discount was updated for every churner and offered to the customer before the customer was likely to churn so as to avoid customer churns significantly.

5 Implementation

This section will discuss implementation process from the start to end that was followed for the purpose of this research. The implementation of models was executed on Jupyter notebook itself using python because of its ease of use as well as large number of machine learning options available with a number of libraries. The datasets used for this research had multiple columns and the output or the target column is a categorical variable. The research was implemented in multiple parts:

5.1 CLV Model:

5.1.1 CLV Calculation Model

In CLV Calculation model, 4 main values were calculated as follows:

- Average Purchase Value (APV): For calculating APV, one extra column that was to be computed was 'total_charge' which was sum of day, evening, night and international charges for every customer. Also, a customer pays for entire month regardless he used these services every day or no. Hence the number of orders was 30, considering 30 days of a month. Thus, total charge was divided by number of orders i.e. 30 to achieve APV.

- Average purchase frequency rate (APFR): APFR was assumed to be 1.25, assuming that the customer made some purchases or used services on his telecom plans at least 80% of the month that is 24 days. Although since the customer pays for the entire month, regardless he used the telecom services everyday or no, the number of customers was considered to be 30, i.e. he was a customer for all the 30 days of the month. So, APFR was calculated as number of purchases (i.e. 24 in this case) divided by number of customers (i.e. 30 in this case), which was 1.25.

- Customer Value (CV): CV was calculated by APV/APFR which gave us an estimate of revenue earned by a particular customer to the company.

- Average Customer Lifespan: Average Customer Lifespan (ACL) can be derived by adding the 'account_length' of each customer and averaging it with the total number of customers.

- Customer Lifetime Value: Customer Lifetime Value (CLV) was calculated as a product of CV and ACL.

5.1.2 CLV Clustering model

In CLV clustering model, all the customers were grouped based on their CLV scores. The customers were grouped into 5 main clusters with each cluster having different customers with similar scores as shown in Figure 3. Also, one biggest challenge was that these clusters were assigned in random order i.e. it was not possible to track which cluster had customers with highest CLV scores and which cluster had customers with lowest CLV scores. This resulted in increased complexity while offering waivers because it was not possible to identify if the customer was eligible for higher discounts, only using cluster number.

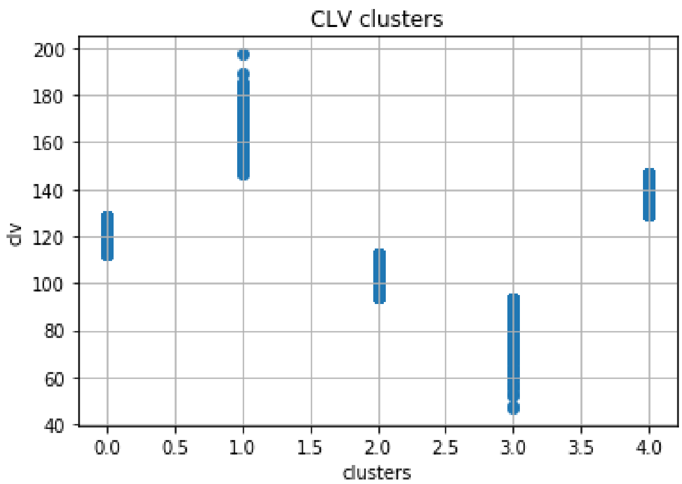

Figure 3: Clusters after K-Means Clustering

For overcoming this limitation of k-means clustering, sorting was performed based on the average of CLV scores in a particular cluster. These clusters were then sorted in ascending order based on the average of CLV scores as shown in Figure 4. Thus, after sorting, it was easy to understand that customer in cluster 0 had one of the lowest CLV scores and was not so valuable customer as the customer in cluster 4, who had one of the highest clv scores and likewise as the cluster number increases, the waivers or discounts increases simultaneously.

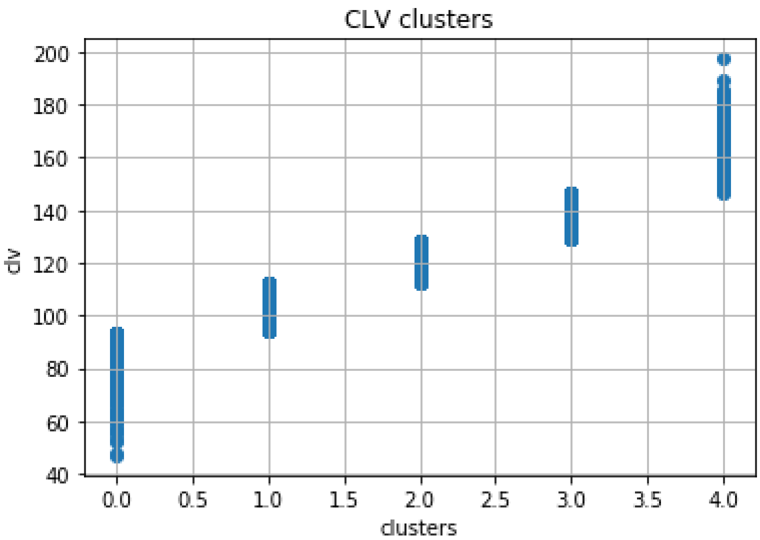

Figure 4: Clusters after sorting

5.2 Churn Prediction model:

After successful calculation of CLV scores and clustering of customers based on it, next step was to identify all the probable churners. Decision tree classifier was used for identifying the churners with the help of multiple independent variables. All the variables were checked with the correlation matrix so as to eliminate all the highly correlated and redundant variables. For example, ‘total day minutes’ and ‘total day charge’ were found to be highly correlated and thus total day charge was eliminated in favour of ‘total day minutes’ and likewise night, evening and international charges were eliminated. After elimination of variables, ‘Churn’ was selected as the output variable and all the remaining variables were considered as input variables to the decision tree model.

5.3 Waiver/Discount Model:

In this model, all the probable churners that were predicted in the previous model were mapped to their CLV clusters. Then based on the CLV clusters, discounts were allocated to each churner before churning out of the company with a motivation to retain that customer. These discounts were given on assumption of 10% discount for churners from cluster 0, 20% discount for churner from cluster 1, and likewise 50% discount for churners from cluster 4. These assumptions were made considering the value of churner to the company i.e. the churners in cluster 4 are most valued customers and given the maximum discount in order to maximise the chances of retaining those customers.

Also, the churn prediction model was applied on another similar dataset after making all the required changes to ensure the consistency and portability of model over multiple datasets.

6 Evaluation

6.1 CLV Clustering

Clustering is an unsupervised learning algorithm, and unlike supervised it was not possible to evaluate its performance by comparing predicted and original values. Evaluating the performance of clustering algorithm can be done by seeing how well the clusters are separated from each other. It could be easily observed that the customers were perfectly fit in clusters with the help of scatter plot shown in Figure 3. For the purpose of this experiment, the elbow method was used to determine ideal value of k i.e. ideal number of clusters that the CLV score will be grouped in. The elbow method helped in selecting best possible k value by fitting the model with multiple k values. Inertia score was used in this process which was nothing but the measure of sum of distances of all the points from the centroid in a particular cluster. Figure 5 was the derived elbow method for this research and k values 3, 4 and 5 were considered to be optimal options. K value of 5 was considered for this research because more the number of clusters more will be the groups for which the offers could be customized.

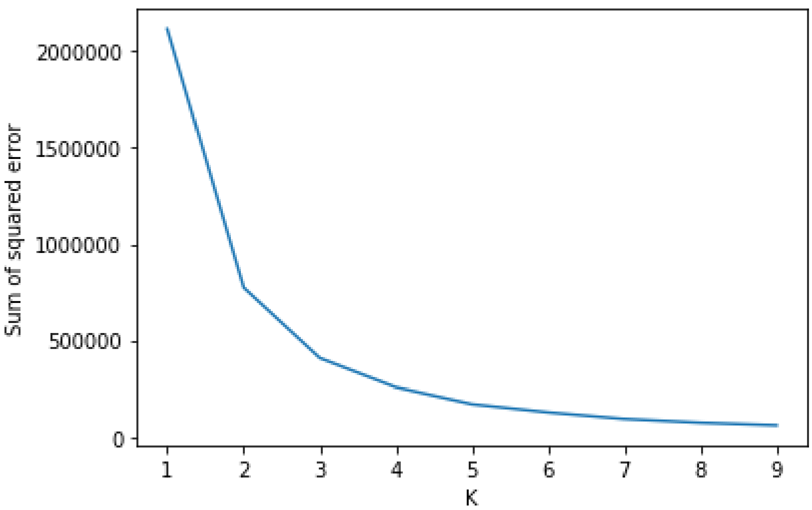

Figure 5: The Elbow Method

6.2 Churn Prediction

For Churn prediction, Decision tree classifier was used. Decision tree classifier was implemented after CLV calculation model unlike any previous researches in an attempt to improve the overall performance of the model further. The performance of the model was evaluated using the following metrics:

- Accuracy score: It was observed that the model is able to predict the probable churners with 98% accuracy.

- Precision: Precision was calculated as the ratio of accurately predicted positive values to all the predicted positive values. It was used to identify out of all the predicted churners, how many customers actually churned out of the company. Of all the predicted churners, 98% actually churned out of the company.

- Recall: Recall was calculated as the ratio of accurately predicted positive values to all the predicted values. It helped to identify out of all the actual churners, how many were successfully predicted by the model. Of all the actual churners, 98% were successfully predicted.

- F1 Score: The F1 score was calculated as a weighted average of the precision and the recall. F1 Scores closer to 1 were considered to be good. In this research, the F1 score of 98% was achieved.

6.3 Transfer Learning

Trained Model was implemented on different dataset of the same company in order to test the robustness and consistency of the model over multiple datasets. On performing transfer learning, the performance of the model didn't degrade much, which proves the capability of the model to perform well across diverse scenarios. Below is the comparison table.

Evaluation Metrics	Primary data	Transfer Learning
Accuracy	98.06%	94.00%
Precision	98.07%	94.55%
Recall	98.06%	94.00%
F1 Score	98.02%	94.19%

6.4 Discussion

This research had implemented a hybrid machine learning approach in order to achieve a telecom churn avoidance model. The novelty of this research lied in the comprehensiveness that this model was able to provide with a significant increase in performance of individual models. Previous researches in the same domain were mostly focused on individual models like churn prediction, CLV calculation or CLV clustering. Unlike previous conservative researches, this research made an attempt to come up with an all in one solution for telecom industry which integrated all the required models together to ultimately aid in clustering customers based on their CLV and mapping churners in each cluster to deal in separate way with each customer. Though, there were some data restrictions, the implemented model is able to provide a decent performance with some of the individual models of this research outperforming the previous researches significantly. While some models had never been implemented before, it was hard to compare with any previous researches. The main purpose of this model was to avoid the customer churn before the churning process even starts. This was done by offering the waivers to the probable churners well in advance before they think of churning out of the company. Since increased telecom bills is one of the main reasons affecting customer churns, an assumption is made that offering waivers to probable churners will decrease the churn rate significantly. Also, due to some limitations of data, system was unable to provide more specific promotional offers like free calling minutes or free data pack.

7 Conclusion and Future Work

The main aim of this research was to develop a comprehensive telecom churn avoidance model integrating all the essential models. In this research, some of the existing models were improved and few new models were implemented to achieve an all in one churn avoidance hybrid model for telecom industry. CLV model was implemented considering many features that were available in the dataset. K-Means Clustering was used for clustering the customers based on their CLV scores. An additional cluster sorting model was used in order to ease the process of mapping customers to their respective CLV based clusters. Also, churn prediction model was implemented after all these modifications and calculations, which made the data stronger to be able to predict probable churners with even better accuracy. The churn prediction model was implemented using decision tree classifier which saw a decent improvement in performance over any of the previously implemented models. Also, waiver model was implemented on many assumptions such as CLV score of the customer, average CLV score of the customer's cluster and reasons for churn. Based on this, waivers were given corresponding to CLV cluster of customers. Higher the CLV of the customer, higher the overall waiver for the following bill cycle. Also, this model was capable enough to predict the churners well in advance due to which it was possible to offer discounts to probable future churners. This could potentially minimize the overall churn rate.

With some minor tweaks and modifications in the code, this model can be implemented on various other applications like avoiding churners in banking and inventory industry or even used to avoid probable credit and loan defaulters. Also, if this system was to be implemented in any environment with live data coming continuously, Apache Spark could be used on top of the same model to perform all the operations, clustering and predictions in real-time. Also due to some limitations of data, some features couldn't be implemented. If the telecom services usage data was made available, the offers could be personalised even further like if the customer is using more data services, the customer could be offered with some free data packs instead of giving discounts on the bill. Also, the gender data is a key information and could play a crucial role in improving the overall performance of the model further. Although, these improvements could potentially be implemented in future, but for now this comprehensive solution could significantly affect the overall customer churn rate in telecom industry.

Acknowledgements

I would like to express my gratitude to Prof. Hicham Rifai for his consistent guidance, support, advice, direction and motivation in this research without which it wouldn't have been feasible.

References

Abedzadeh, N. and Nematbakhsh, M. A. (2012) 'Using CLV for modelling churn and customer retention', *International Journal of Electronic Marketing and Retailing*, 5(2), pp. 128–146. doi: 10.1504/IJEMR.2012.051017.

Ali, M. H. *et al.* (2011) 'Applicative implementation of D-stream clustering algorithm for the real-time data of telecom sector', *Proceedings - International Conference on Computer Networks and Information Technology*. IEEE, pp. 293–297. doi: 10.1109/ICCNIT.2011.6020884.

Chuang, H. M. and Shen, C. C. (2008) 'A study on the applications of data mining techniques to enhance customer lifetime value-based on the department store industry', *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, ICMLC*, 1(July), pp. 168–173. doi: 10.1109/ICMLC.2008.4620398.

Desirena, G. *et al.* (2019) 'Maximizing customer lifetime value using stacked neural networks: An insurance industry application', *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019*, pp. 541–544. doi: 10.1109/ICMLA.2019.00101.

Gaur, A. and Dubey, R. (2018) 'Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques', *2018 International Conference on Advanced Computation and Telecommunication, ICACAT 2018*. IEEE, pp. 1–5. doi: 10.1109/ICACAT.2018.8933783.

Hung, S. Y., Yen, D. C. and Wang, H. Y. (2006) 'Applying data mining to telecom churn management', *Expert Systems with Applications*, 31(3), pp. 515–524. doi: 10.1016/j.eswa.2005.09.080.

Lee, E. *et al.* (2018) 'Profit Optimizing Churn Prediction for Long-term Loyal Customer in Online games', *IEEE Transactions on Games*. IEEE, 12(1), pp. 1–1. doi: 10.1109/tg.2018.2871215.

Mauricio, A. P. *et al.* (2016) 'Predicting customer lifetime value through data mining technique in a direct selling company', *ICIMSA 2016 - 2016 3rd International Conference on Industrial Engineering, Management Science and Applications*. IEEE, pp. 1–5. doi: 10.1109/ICIMSA.2016.7504027.

Ben Mzoughia, M. and Limam, M. (2014) 'A multicriterion segmentation approach based on CLV components', *CINTI 2014 - 15th IEEE International Symposium on Computational Intelligence and Informatics, Proceedings*, pp. 191–195. doi: 10.1109/CINTI.2014.7028674.

Nadella, M., Gupta, N. and Pudkaje, T. (2020) 'Subscriber Gender Prediction in Telecom Using Deep Learning', *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, pp. 1–4. doi: 10.1109/wiecon-ece48653.2019.9019953.

Swetha, P., Usha, S. and Vijayanand, S. (2018) 'Evaluation of churn rate using modified random forest technique in telecom industry', *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018 - Proceedings*. IEEE, pp. 2492–2497. doi: 10.1109/RTEICT42901.2018.9012251.

Tolon, M. and Aslan, D. (2018) 'Comparing Customer Segmentation With CLV Using Data Mining and Statistics: A Case Study', *Journal of Business Research - Turk*, 10(4), pp. 887–900. doi: 10.20491/isarder.2018.554.

Varun, E. *et al.* (2019) 'An Efficient Technique for Feature Selection to Predict Customer Churn in telecom industry', *1st IEEE International Conference on Advances in Information Technology, ICAIT 2019 - Proceedings*, pp. 174–179. doi: 10.1109/ICAIT47043.2019.8987317.

Wang, Y., Sanguansintukul, S. and Lursinsap, C. (2008) 'The customer lifetime value prediction in mobile telecommunications', *Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology, ICMIT*, pp. 565–569. doi: 10.1109/ICMIT.2008.4654427.

Yonghui, Z. and Wanli, H. (2010) 'Customer lifetime value and the evaluation model in telecom industry', *Proceedings - 2010 International Symposium on Computational Intelligence and Design, ISCID 2010*. IEEE, 2, pp. 109–112. doi: 10.1109/ISCID.2010.116.