Literature Review

There is a large amount of research related to Customer Relationship Management (CRM) that overcomes the limitations of their data sets and proposes various algorithms and techniques to identify customer segmentation of a particular data set. The following subsections discuss the related work of the most important challenges researchers face with their respective proposed solutions. In addition, we will focus our literature review to discuss related work that uses classification techniques namely Logistic Regression, Decision Tree, Random Forest Tree.

2.1 Customer Segmentation in Customer Relationship Management (CRM)

Segmentation can be seen as a simplification of the messy complexity of dealing with multiple individual customers, each with different needs and potential value. Traditional customer segmentation methods are generally based on experience classification methods or simple statistical methods. Traditional statistical methods group customers according to simple behavioral characters or attribute characters such as the category of the purchased product or the region in which it lives.[6] This method of segmentation cannot perform a more complex analysis that what kind of customers have high potential value and what kind of customers have credit. high. With the extensive implementation of EC and CRM, companies have accumulated a growing number of customer data. Traditional techniques such as multiple regressions cannot overcome this level of complexity. As a result, the reliability and validity of statistical functions used to generate segmentation or to build predictive models become possible contributing factors to dissatisfaction. CRM user [7].

Data mining can be thought of as a methodology and technology developed recently, becoming famous in the year l994. Sas Institute defines data mining as the process of selecting, exploring and modeling large amounts of data to uncover previously unknown data patterns. Thus, data mining can be thought of as a process and technology for detecting previously unknown things to gain a competitive advantage. Data mining uses neural networks, decision trees, link analysis, and association analysis to find useful trends and patterns from extracted data[6]. Data mining can generate important insights including predictive models and associations that can help companies understand their customers better. Many large companies today have terabytes of data, where they may be able to find more information about customers,markets, and competition than they need to. Data mining allows marketers to better extract valuable business information from the 'mountain data' in a company's systems. This is a potential solution to the big problems many companies face: abundant data and relatively dearths of staff, technology,and time to change. Numbers and records become meaningful. information about existing customers and potential customers. Data mining allows companies to measure consumer behavior based on 100 or more attributes, instead of the three or four associated with traditional statistical modeling. The more attributes a company uses, the greater the complexity of data and the greater the need for data mining tools. As practitioners enthusiastically seek out groups that benefit customers whose loyalties are stable, some academics are beginning to question whether the segment is actually an entity. A stable and many more fundamentals whether they really exist[7]. The segmentation method based on data mining created by this paper can solve the above problems because the model can learn new information that is entered later. and get new rules. It provides full support for dynamic management processes in acquiring customers, retaining customers and improving customer value, customer satisfaction and Promote customer loyalty. Establishing a mapping relationship between conception attributes and customers is a key step of the segmentation method based on data mining[18]. Customer data contains dispersive and continued attributes. Assigning each customer attribute as a dimension and designating each customer as a particle, the entire customer in the company can form a multidimensional space, which has been defined as the customer attribute space. The mapping relationship between customer attributes and conception categories can be constructed by analytical methods, or by sample learning methods. Analytical methods analyze the attribute character of each category of conception that must be possessed, further establishing a mapping of the relationship between attribute space and conceptionspace[12]. However many mapping relationships between attribute space and conception space are unclear, it is necessary to use sample learning methods to establish mapping relationships[11]. Sample learning methods automatically generalize the mapping relationship between attribute space and conception space by applying data mining technology to the same conception category. known in the company's database. The process of data mining is called sample learning. As for the rules in making customer classification [10].

1) Segmentation rule making

Sort customers by the Customer Segment Model and Segment Functions. After training the segment model, we get segment rules or network segment rules. We can effectively group new customers based on trained models.

2) Function Analysis

Functional analysis includes customer value analysis, credit analysis and promotion analysis, etc., based on basic mapping of relationships between customers and concepts. Furthermore, the need for new functions will be brought to CRM with the development of management practices. The new demand function will be added to the conceptual dimension and reconstruct the mapping relationship with customer characteristics.

Advantages of making a customer shake[10]

(1) Increase the promotional effect

Customer segmentation based on data mining can help companies to create appropriate promotional strategies, at the right time, with appropriate products and services, aimed at the appropriate customer.

(2) Analyze customer value and customer loyalty Customer value and customer loyalty are important to the company's management strategies and tactics. The company can confirm customer ratings according to expected value and loyalty analyzed with a segmentation model based on mining data.

(3) Analyze credit risk

Risk assessment is an effective way to evaluate a particular type of customer risk, usually the risk of default.

(4) Instruct the R&D of the new product

Companies can find out their customers' preferences by analyzing customers based on data mining, and ensuring that various requests will be realized in the new design.

(5) Confirming the target market

Customer segmentation based on data mining can make the target group of customers clear and find the market explicitly.

The key role of marketing is to identify customers or segments with the greatest potential for value creation and target them successfully with appropriate marketing strategies to reduce risk. These high lifetime value customers defect to competitors [10]. In this mode of construction, customer segmentation is the basic job of data mining according to known historical segmentation information.12 The training data used to build segment forecast modes can be historical data or exogenous data obtained from experience or surveys. Because customer behavior is uncertain and inconsistent, researchers and managers must build dynamic customer segmentation models to objectively reflect characteristics. In the customer-centric era, customer segmentation results are related to the determination of company strategies and tactics. Best practices require marketers to develop their understanding of customer segmentation based on data mining techniques and use output to develop marketing strategies on a regular basis. creative to maximize shareholder value[14].

2.2 Machine Learning in Customer Relationship Management (CRM) Framework

Machine learning and data mining help companies create tools that can create and take action based on customer knowledge and information. Customer information is the basis for maintaining a long-term relationship with customers and is also known as relationship and customer management(CRM). Classification and segmentation of customer data sets is used to maintain efficient relationships with customers and further increase profitability and productivity. In this paper, the authors propose segmentation of customers based on demographic properties such as gender, age and spending scores and analyze data sets for interesting facts. Derivative attribute data sets are investigated for classification. Classification is used to categorize each customer into several classes, namely, 'gold', 'silver', 'elite' and 'occasional'. Comparison of different classification algorithms is simulated with the WEKA tool. Multi-layer perceptron (MLP) was found to be the best classification algorithm with 98.33% accuracy compared to Naïve Bayes, regression and J48.4.

The digital revolution and the increasing amount of data generated by companies/organizations in recent decades have led to great interest in the field of machine learning and deeplearning. Technology organizations and companies use Machine L-based predictive analytics to gain an edge over their competitors. The purpose of ML techniques is to find 'hidden' information in data, which is almost impossible to do in traditional ways based on human analytical skills.5 According to (Cioca et al., 2013; Rahman and Khan, 2017). Machine Learning techniquesare used to mine data for business intelligence and sellable strategies for customers such as classifying them in different categories, creating promotional scheme strategies, and to improve customer relationship management (CRM) In the current scenario, business processes are becoming increasingly customer-oriented and placed as top management priorities.[6] Due to technological advances in e-commerce, M-commerce, virtual marketing and digital marketing, each product is just a click away from customers. Based on the statement from (Singh and Agrawal, 2019a; Adebiyi et al., 2016) This leads to very intense competition, which is necessary to ensure that consumers receive the highest possible quality standards to maintain them and reduce stirring rates[7]. Crm's focus is to expand customer service and support in customer retention. Customers are very important to every company and organization. For the identification and retention of their target customers, it is very important to have data analysis, which is used to explore valuable insights and trends to know the metrics and nature of the customer. According to (Singh et al., 2018; Yadav et al., 2018it is important to observe the most important key factors influencing a customer's purchasing decision to purchase any product and service[8]. Machine Learnings is one of the popular data analyses that governs the structure of analytical models, which is valuable for growth in purchasing behavior. ML techniques are widely used in customer segmentation prediction, customer life value (CLTV), churning, sales, etc. Based on a statement from (Sgaier et al., 2017) Customer segmentation is useful in understanding what demographic and psychographic subpopulation is inside your customers in business cases and leveraging this information to improve profits, image, value, and inventory management.[9] Companies in any business already realize that gaining new customers is not enough for lasting success and efforts need to be made to identify customer segmentation towards retention. In this paper, segmentation techniques and customer classification for business intelligence analysis purposesare proposed[10]. The demographic property of the customer is taken as a parameter of customer segmentation to know the analytics about the customer. That will help in CRM, efficiency and productivity of shopping malls. Furthermore, machine learning techniques are used to predict customer judgment[9].

2.2.1 Customer Relationship Management using Logistic Regression

Recurrence examination is a type of prescient display method that explores the relationship between the needy (target) and autonomous variables[11]. This procedure is used to anticipate, show timing and find causal relationships between these factors[12]. For example, the link between rush driving and the number of road accidents by drivers is best concentrated through relapse. Strategic regression is a machine learning grouping calculation used to forecast the likelihood of clear variables. In strategic relapsing, bound variables are paired variables that contain coded information 1 (really, achievement, and so on.) or 0 (no, disappointed, and so on.). The relapsing model strategically predicts P(Y=1) as component X. Calculated regression is one of the most well-known approaches to adjusting models for clear information, especially for double reaction information in data modeling.[12]

2.2.2 Customer Relationship Management using Decision Tree

The decision tree is said to be one of the common methodologies, used in predicting as well as estimating customer churn problems. Based on the method of dividing and conquering, decisions were developed.[13] But the decision tree does have some limitations such as not being usable for non-linear and complex relationships between attributes. However, it has been observed that the decision tree method does try to improve classification accuracy[13]. In the JST paper along with the decision tree it is used to perform customer churn predictions and it is known that decision trees outperform neural networks in terms of accuracy[14]. Presented a DT application classification methodology to analyze churn rates in the telecommunications industry. Here, the ID3 decision tree is used and observed that the customer area is one of the main classification features, the other gives two results to the customer for churn using several methods such as K-means clustering, QUEST, CART, Logistic Regression, neural network, exhaustive CHAId[13]. Here, it is observed that CHAID performs much better than the other methodologies mentioned. Note that its accuracy is about 60%, which is much better than other methodologies. In addition, other decision trees do not stand in front for exhaustive CHAID.[14]

2.2.3 Customer Relationship Management using Random Forest Tree

Random forests are supervised learning algorithms used for both classification and regression.[15] But in any case, this is mainly used for classification issues. As we know that forests are made up of trees and more trees means stronger forests. Similarly, random forest algorithms make tree decisions on data samples and then get predictions from each and ultimately choose thebest solution throughvoting. This is a better ensemble method than a single decision tree because it reduces over-fitting by flattening the results. It addresses the problem of overfitting by flattening or combining the results of different decision trees. Random forests work well for a large number of data items rather than a single decision tree. Random forests have fewer variances than single decision trees. Random forests are very flexible and have very high accuracy[10]. Scaling data is not required in random forest algorithms. It maintains good accuracy even after providing data without scaling. The Random Forest algorithm maintains good accuracy even as most ofthe data islost.[15] Complexity is the main drawback of random forest algorithms. Random forest construction is much more difficult and time consuming than decision trees. More computing resources are needed to implement the Random Forest algorithm. It's less intuitive if we have a large collection of decision trees[17]. The prediction process using random forests is very time consuming compared to otheralgorithms. Some conventional algorithms such as Decision Tree, Genetics algorithm, neural network and tree classification have been proposed. The above algorithm is able to estimate the churn rate. However, they have some problems such as a decision tree that is lacking with the same class probability problem that can drastically reduce performance[18]. Similarly, in the case of genetic algorithms, it is highly unlikely to recognize the possibilities associated with estimates that result in low performance and in the case of techniques such as state-of-art resulting in some errors. Therefore, the above discussion can conclude that conventional RF techniques do not produce efficient results in terms of large data sets and also perform very poorly when they are unbalanced. Our proposed MRF (Modified Random Forest) method performs better across a variety of parameters such as accuracy and other durability[19]. Our scheme mainly has an additional layer of RV (Random Variables) which helps to perform the model much better along with our methods of assisting in minimizing Gaussian noise and also helps in reducing regression and classification problems. Our methods help in building a variety of different trees from specific training data sets. The proposed method helps in focusing the estimation of consumer churn rates in telecommunication services. In research conducted by Irfan et al the Random Forest Algorithm (RF) worked well with 88.63% of correctly classified instances[20]. Creating an effective retention policy is an important task of CRM to prevent churners. After classification,the proposed model segments the data of moving customers by categorizing customers who quit in groups using cosine similarities to provide retention offers. group-based. The paper also identifies churn factors that are important in determining the root cause of churn[21]. By knowing the significant churn factors of customer data, CRM can increase productivity, recommend relevant promotions to groups of customers who are likely to churn. based on similar patterns of behavior, and excessively enhance the company's marketing campaigns. The proposed churn prediction model is evaluated using metrics, such as accuracy, precision,recall, f-measure, and the receiver's operating characteristic area (ROC).22 The results revealed that our proposed churn prediction model resulted in better churn classification using RF algorithms and customer profiles using k-means clustering. In addition, italso provides the factors behind customer churn churn through rules generated using the selected classification algorithm attributes.[17]