Churners Prediction Based on Mining the Content of Social Network Taxonomy

Asia Mahdi Naser Alzubaidi, Eman Salih Al-Shamery

Abstract: Churner Customer is a main tricky and one of the most important issues for large companies, due to the straight impact on the incomes of the companies especially in the telecom domain, companies are searching for advance strategies to predict churn/non-churn customer. This research focuses on the construction of a predictive model to identify each customer as churner or not and gain additional insights about their service consumers. The main contribution is to overcome the limitation of independently based on data mining strategies by developing approaches and derived network metrics such as centrality and connectivity between customers to incorporate network mining with traditional data mining. Social network measurements e.g. Leverage, flow Bet, Page Rank, Cluster Coefficients and Eccentricity are joined with other attributes in the original network dataset to enhance the performance of the proposed methodology. The risk of churn can be predictive by preparing an extensive cleaning the raw data for churn modeling, It divides customers into clusters based on Gower distance and k-medoids algorithm to help understand and predict churner users, classification model using Extreme Gradient Boosting "XGBoost", assessment the model performance by computation the centralities metrics as new attributes appended to the original network dataset. Experiments conducted on Telecom shows that with an average value of all statistics accuracy not lower than 98.27%, while the average accuracy for the original dataset with it is clusters is not exceeded than 0.97%. The proposed method for churners detection which combines social impacts and network contents based on clustering significantly improved the prediction accuracy for telecom dataset as compared to prediction using the call log details, network information without implement of clustering , thus validate the hypothesis that combining social network attributes and Call/SMS information of the users for churn prediction could yields substantially improved of customer churn prediction.

General Terms: Theory and Applications of Data Mining, Dimensionality Reduction, Business Analytics, Machine Learning, Supervised Statistical Learning.

Keywords : Churn Prediction, Mobile Social Network Analysis, Churn in Telecom, Social Network Analysis, eXtreme Gradient Boosting algorithm (XGBoost), Centrality Metrics, Mobile Network.

I. INTRODUCTION

The communications sector has turned out to be one of the fundamental enterprises in developed countries. Potential churner customers in Telecom are customers who completely cease all their memberships and contracts and totally ending all activities with the telecom provider and never again acting as customers. Because of the high challenge between different telecom suppliers and the capacity of the customer to move from one service supplier onto the another, numerous methodologies have been proposed to gain more incomes for organizations, for example, get new customers, upsell the current customers, and increment the maintenance time of customers. Notwithstanding, contrasting these strategies taking the estimation of the rate of profitability of each into thought has demonstrated that the degree of consistency is the most beneficial methodology and demonstrates churn Customer causes genuine benefit misfortune when exceeds certain limits, on the other hand, companies realize that attracting new customers is much more expensive than preserving existing ones[1]. The issue of churn prediction has been studied extensively in the past in quite a lot of domains and social network analysis emerged as a key platform in churn prediction challenge to recognize the consequences of the strength of social relations and the structure of a social network in churn stems from the view that the attributes of the user in a network are much less significant than their associations with different subscribers in the Telecom network[2]. SNA is dependent on the supposition that the significance and effectiveness of analysis the ties amongst interacting individuals which are significant to discover the structure and interdependencies of customers. The evolution of the telecommunication sector has made a greater impact on the way in which individuals keep in touch with each other. When SNA utilized in telecommunications, it can help companies to appreciate the conduct of their customers and then predict the strength of ties between users and the possible influence of social events among them. Recent researchers in telecommunication companies use the SNA attributes equivalent to Calls/SMS/internet detail and call detail records (CDR) in conjunctions with networking features like centrality measures, similarity values, and client's network connectivity in the churn framework[3]. Some of the service providers and vendors are undoubtedly not able to extravagance their customers as interconnected peers who are influenced by their friend relations because the service providers lack the data needed to represent users as nodes, and form edges between them. Conversely, telecommunications carrier providers need to maintain the information of their client's mobile call/SMS documents for charging purposes. The identical information might, therefore, be used to form a social network of interconnected users and count on how they could inspiration of each different. Moreover, the diverse directions in relations to the connection in the network have distinct weighted values in the telecommunications scenario. Incoming calls are valued otherwise than outgoing calls because a customer can ship and receive calls within the

Asia Mahdi Naser Alzubaidi, Computer Science Department, College of Science, Karbala University, Karbala, Iraq. Email: asia.m@uokerbala.edu.iq

Eman Salih Al-Shamery, College of Information Technology, Software Department, Babylon University, Iraq. Email: emanalshamery@itnet.uobabylon.edu.iq

community, each incoming and outgoing calls are taken into consideration. Mobile networks take into accounts the best way in which customers are originating their calls, providing a good sense of how they operate throughout the social network. Furthermore, statistical studies have confirmed that more than 75% of customers will refer a friend before deciding on the purchase of a certain product or adoption of new service[4]. The contribution of this paper can be summarized as follows.

- Due to the heavy marketing and promotion schemes of other companies. The knowledge of potential churn customers plays as a weapon for competition and helps Telecom companies to reposition their products to make it more user-specific. Design and implement churn predictive framework that able to identify those customers who are at risk of churn and give an insight to telecom provider why their customers are churning so that a company's retention team can intervene and entice the customer to stay via a special offer such as upgrade handset at a discount price, cheaper monthly rate, etc.

- Identify potential churners based on relation information in social network analysis rather than traditional analytical that concerned to customers. The data may be stored according to the monthly or season of the year, these data will be analyzed, and useful information could be extracted from them.

- To demonstrate that Enriched data with social network attributes pattern yield more accurate predictive paradigm. By means of constructed the social network of all the customers in Telecom dataset and calculated centrality and connectivity features and contain them with initial network dataset to illustrate the influencing of social aspects in the enhancement of AUC and accuracy results because of the contribution of these social aspects in giving extra special information regards to the customers.

- Use a simple churn model as viable, i.e., an approach that required a minimal number of entering arguments and that had minimal runtime and storage complexity, so it might be implemented in real information and big data networks of telecommunications service providers. A churn detection procedure might be utilized in this sort of method that the provider could target the customers which could potentially depart with precise moves, such as promotions or reduction.

The structure of the article prepared in sections as illustrate in follows: Section 2, present previous studies about customer churn prediction in the Telco sector. Methodology, model building, Data preprocessing, executed methods are described in section 3. Section 4. Illustrated the experimental implementation and outcomes of churn system are discussed. Conclusions considered in section 5.

II. RELATED WORK

Churners and non-churners classification regard as predominant trouble for Telecom service providers and outlined as lacking customers because they go away to competitors. The capability of categorizing customer churn in advance, supplies the Telcom company with appropriate insight to retain and maintains their customer Vast variety of techniques have been applied to predict churn in Telecom corporations founded on SNA that proved its importance by using its high penetration cost in most of industries and the rapid expand in the number of empirical reviews employing network analysis, therefore looking at the precise structuring and contenting of the power ties between participants and at the impact of network content material/structure on model efficiency. Abdulrahim et al.[1] presented a data mining methodology for predicting churn of the customer in SyriaTel Telecom enterprise which has unbalanced and massive datasets. By utilizing four tree-algorithms were applied due to their diversity and applicability in the classification problem. These algorithms are decision Tree, Random Forest, XGBoost algorithm, and GBM tree algorithm. Extensive preparation of data and pick methods of the cellular social network attributes had the largest effect on uplifting the AUC value from eighty-four to ninety-three percentage and yield efficient results by applying XGBoost algorithm. Manďák[5]introduce thesis for suggesting and implement of churn prediction method focus on the CRISP-DM information mining method, cluster analysis centered on Gower distance metric in the k-medoids algorithm, decision tree, logistic regression, and random forests are utilized in classification models. The author creates a dashboard in the Qlik visualization tool to provide a person-pleasant instrument for understanding customer conduct. The target of the thesis is to make the evaluation by means of training the proposed churn model with entire customers in the Telecom dataset and within every customer in the cluster, it got here out that there's handiest minimal change between the two procedures. Gavril et al.[6]depicted a complicated process for detection churn customers in prepaid using small dataset and no missing values existed, with target churn variable of two values: churn/non-churn. Some samples include data that count the total number of income and outcome messages and voice mail for each user. the researcher used data reduction based on PCA algorithm to reduce data dimensionality. Three machine learning approaches had been used: support Vector machine, Neural Networks, and Bayes Networks to predict churn customers. The writer used AUC to measure the efficiency of three approaches. The AUC value for Bayes Networks was the larger. Prashanth et al. Gandhi et al. [7]This paper showed an adjusted version of the k-medoids algorithm for significant datasets evaluation and perform an assessment between the k-mean and the proposed data mining approach, experimental outcome has depicted that the modified k-medoid carried out higher as compare with k-mean in terms of cluster quality and complete run time metrics. moreover, there's additionally an area of improvement within the decision and plotting of the index's decision at the end of results analysis.[8] This paper supplies massive evidence that social ties play an influential function in affecting churn in networking and show an easy, amazing, diffusion-based process that exploits these influences to

determine a tremendous portion of churners in the Telcom community. This paper suggests that it will be very intriguing to be taught how the mapping function can also be assorted situated on customer/relation features and elevating the lift curve. Nankani[9]formulate a model that combining SNA characteristics with traditional Data Mining methodologies to enhance the accuracy of predictive models in deep data Mining space. This thesis tries to explore and analysis the model's behaviors with interactions between peoples and compared the models without interactions and demonstrate that using SNA methods gives deeper data understanding, as well as evaluate the strength of connections with the help of centrality measures by computing from SNA networks. These centrality measures are aggregated with other attributes in the original networking dataset to generate the predictive models, hence it enhances the classical DM with strong relationships and helps to understand the holistic view of big data. The investigation presented in this dissertation overcomes the hypothesis of data mining that all records and fields of datasets are independent by adding a relationship dependence.

model to identify each customer as potential churner or non-churner in Telecom datasets when they cut all ties with the current Telcom network. The main contribution of the article is to use the methodological process to investigate the impact of social networks attributes on customer churn behavior, only by pursuing each customer's social network via enriching the predictive model with different aspects of network centrality and connectivity features. Figure.1 detail the proposed churn framework, it contains all the fundamentals of data collection from various sources, data understanding, and data exploration with SNA techniques that provide hidden relationships which can compute and understand the visual representations of it. The framework of the proposed methodology was broken down into different steps as data collection, Data Enrichment with SNA features, Data Mining k-medoid clustering algorithm and XGBoost for a classification model, finally Evaluation of all the models developed during the research.

III. RESEARCH METHODOLOGY

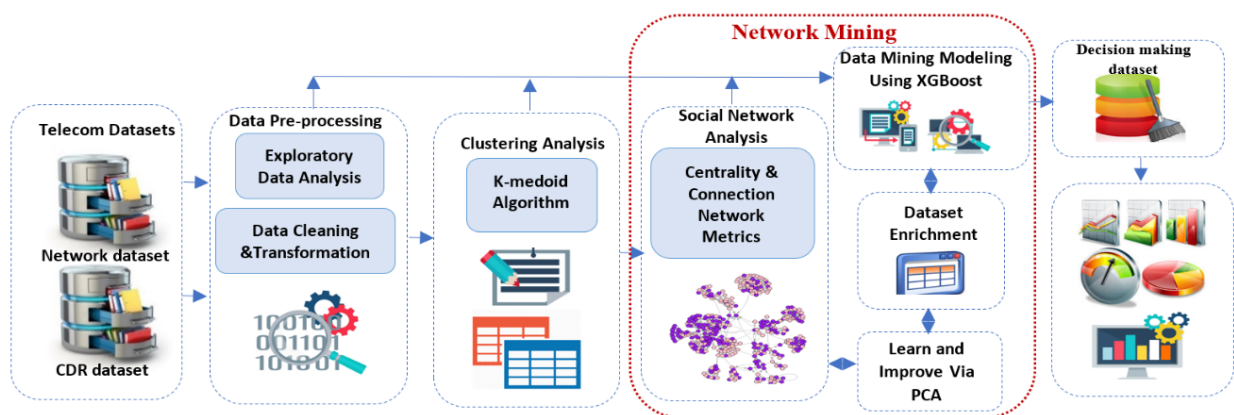This research focuses on the construction of a predictive

Figure 1. Proposed Churn Prediction Framework

Data cleaning approach by identifying a set of features that are as correlated as possible among themselves and as unrelated as possible with other feature sets in the same dataset. This was done by analytical the multivariate data and building classification models that able to identify redundant and inappropriate variables. On the other hand, Relevancy about churn feature which represents the potential predictor variable involves understanding the relationship between the target variable and all the other input variables in Telecom dataset. It can be calculated by classifying groups of attributes that are as correlated as possible to the target variable and the groups of attributes that are uncorrelated as possible with it. This research uses multiple correspondence analysis (MCA) as multivariate data analysis and data mining tool for finding and constructing a low-dimensional visual representation of variable that associations with churn variable as depicted in figure 2.
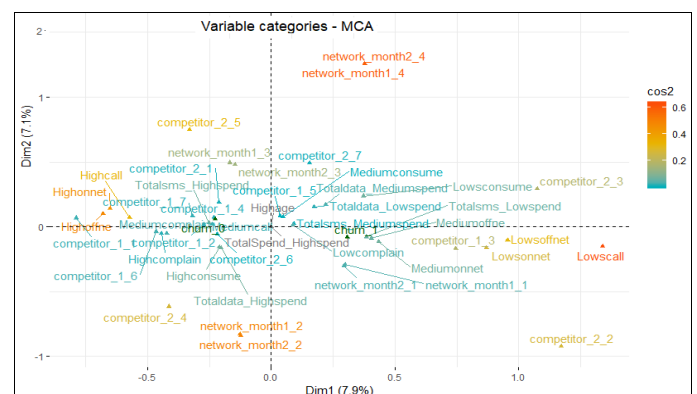

Figure 2. Variables Correlations with Churn Predictor using MCA

Network Mining
Studies show that social relationships play an important

role in influencing of customer churn in the telecoms network. Network mining in telco sector is a Data mining and SNA integration that gives potential understand and deep insights of the social influence and create models for correct identify the current set of churners and their underlying social networks from the Call Detail Records (CDR) data. Enrichment of Dataset represented by creation of the churning system follows conventional data mining methodology which consists of Telecom company considerate, dataset understanding, original and enriched datasets preparation, churn prediction modeling, evaluation and deployment. The original input dataset for a suggested churn predictive model is the customer historical activities that involved information from past Call, SMS, internet-traffic, vmail-messages, service-call for each mobile subscriber. However, the enrich dataset is the result of appended centrality and connectivity measures that aggregated with other attributes in the original networking dataset to generate the predictive models, hence it enhances the classical data mining with strong relationships and helps to understand the holistic view of big Telcom data, also to illustrate the influencing of social aspects in the enhancement of performance measures results due to the contribution of these social aspects in giving extra special information regards to the customers. Figure 3. depicted the enrichment step of data mining strategy.
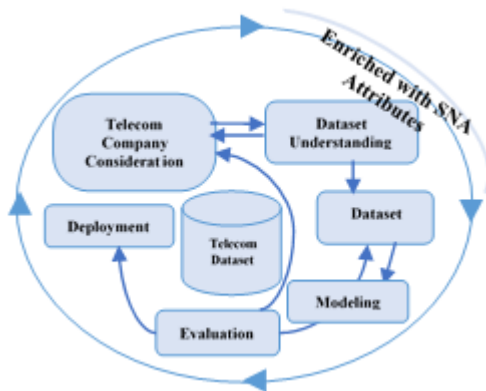

Figure 3: Enrichment of Data Mining Method

Analysis of social network determines the numerical value of a network node or edge that can be used to measure graph attributes such as the most important node, the influential node, the powerful link, the fill connected nodes. Telecom network dataset is represented by weighted and directed graph depending on the weighted edge, the weight of a path is the total number of common activities between every two customers, statistically computed via mean of normalized call/SMS. The Graph-based features centrality are depending on the concept of efficient diffusion of relations over the network. The interpretation of this characteristic in networks is that every edge in the network is evaluated for its content/structure information that subtly contributes to network processes[10]. The centrality of the network in social and biological sciences remains an open topic, and researchers are still seeking to improve existing measures as well as develop new ones[11]. We cover fundamental centrality definitions and computational algorithms in the field of Telco network, Numerous centrality measures that can be enrichment in Telecom dataset, the decision to choose which graph metric may the best suit to the analysis of given dataset attributes is crucial for inferring important functional properties of a network. Table(1) shown the centrality and connection measures that combining with original CDR dataset.

Table 1. shown the centrality and connection measures

No	Centrality and Connection Metrics
1	Degree Centrality
2	Betweenness Centrality
3	Eigen Centrality
4	Page Rank Centrality
5	Subgraph centrality
6	Eccentricity Centrality
7	Closeness Centrality
8	Leverage Centrality
9	Stress centrality
10	Flowbet Centrality
11	Gil-Schmidt Power Index
12	Geodesic K-path Centrality
13	Laplacian Centrality

This paper calculates about twenty-nine centrality and network attributes of each node in the entire network and then evaluate the contribution of this measurement in enhanced the performance of mining the telecom network based on principal component analysis (PCA). Only thirteen ones that denoted the best-evaluated centrality features above the red line are selected and added to the original dataset. Figure 4. shown the SNA attributes evaluation based on PCA.
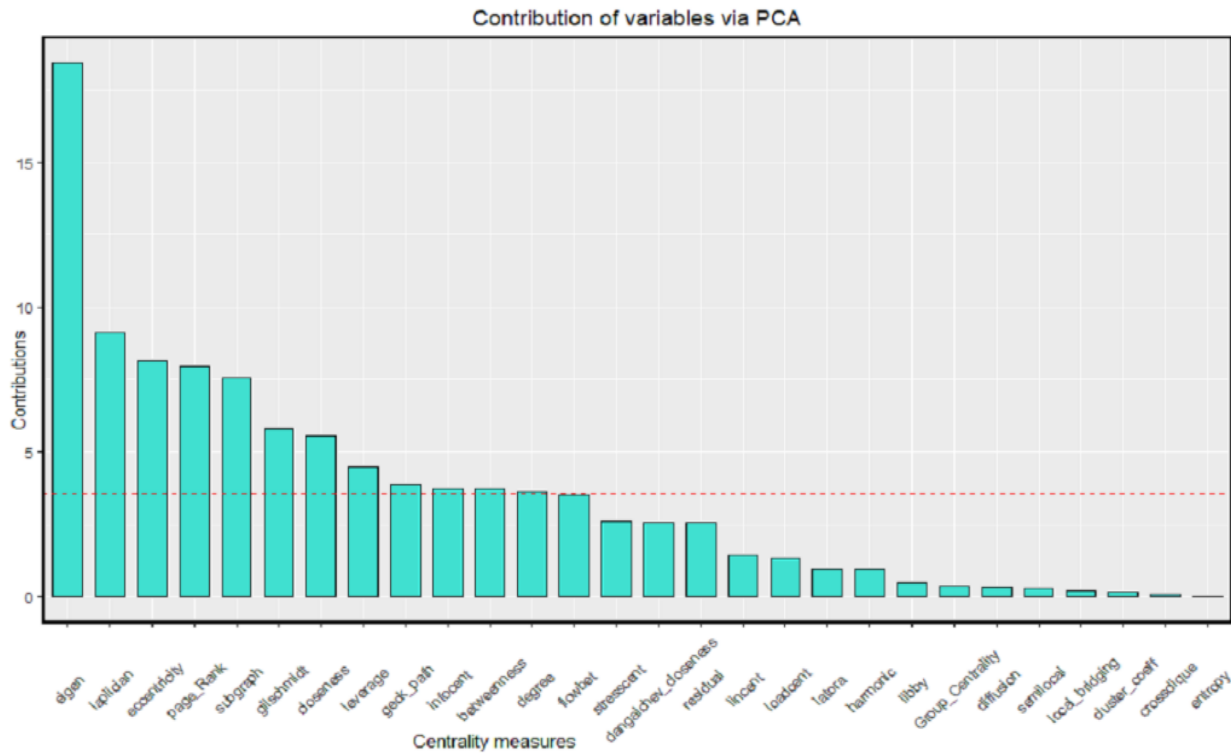
Figure 4. SNA centrality and connection attributes Evaluation via PCA

Cluster Analysis Using K-medoid Algorithm

Clustering plays a vital function in the study area of data analysis and machine learning; it can be used to verify the relationships between samples of data. The target of any clustering algorithm is to partition the observations of a dataset into meaningful subclasses known as clusters based on some similarity or dissimilarity measures, so the observations inside each cluster are close to every other, while observations in exclusive clusters are exceptional from each other[12]. The objective of any clustering algorithm is to ensure that the distance between data points in the cluster is very low as compared to the distance to other clusters. In other words, members of a cluster are very similar, and members of different clusters are extremely dissimilar[13]. Clustering is common purposes in many fields such as cluster evaluation Economic Science, biology, medicine, business and marketing, computer science or social science, documents classification, pattern recognition, digital image Processing, textual content mining[14]. To perform Clustering in the Telecom network, Customers should have clustered according to their profitability for the churn management system. knowing the profitable customer portion, then it has a chance to keep in hand the most important customers via suitable elevations and campaigns[15]. To perform cluster analysis, it is imperative to check some major assumptions of cluster analysis, such as the completely disappear of multicollinearity, outliers, and comparability the datatype of attributes. Multicollinearity represents the excessive correlations or associations between the independent variables. For this reason, a form of disturbance within the data, and if it presented in the knowledge then, the statistical inferences that concerning knowledge is probably not safe. The variable producing multicollinearity must be deleted or it must extend the dataset. Multicollinearity can be detected with the help of tolerance and its reciprocal technique, referred to as the variance inflation factor (VIF) as proven in equation1.

$$VIF_i = \frac{1}{1 - R_i^2} \qquad \dots (1)$$

If the value of VIF is 10 and above, then the multicollinearity is robust [5]. See table(2).

Attribute	VIF
network age	1.046782
Total SMS	1.083878
Total data	1.092086
Total consumption	1.050583
Unique_ call	1.590475
On_net	1.288429
Off_net	1.786582
Call_ complain	1.011974
network_month1	4.897691
network_month2	4.918685
competitor_1	1.033057
competitor_2	1.115942

For a confident classification of the clustering algorithm, it's better to check the appropriate number of clusters(k) earlier than divide the Telecom dataset into smaller, homogeneous partitions, from prior competencies it is not seeing how many clusters should be there. The correct number of clusters is commonly ambiguous and increasing k without penalty terms will continually diminish the heterogeneity in certain clusters[16]. Furthermore, is somehow particular and depends on the method used for measuring similarities and the parameters used for partitioning. There are several choices to learn how to decide an appropriate number of clusters, e.g. The elbow system, Calinsky criterion or Silhouette criterion, Gap statistic approaches[17]. Figure 5. View some procedure for an estimate of k worth.
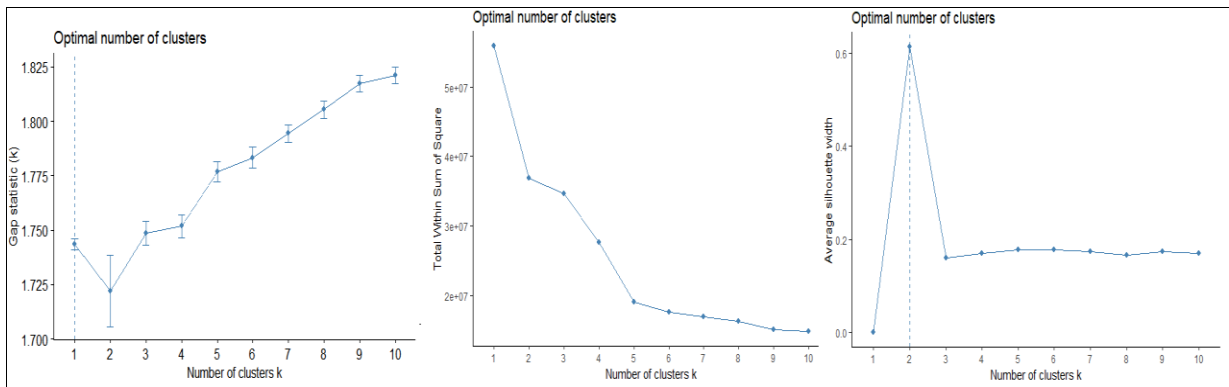
Figure 5. Gap Statistic, WCC, Silhouette for Estimate Number of Clustering

This research utilized probably the most k-medoids algorithms for clustering of Telcom dataset known as Partitioning around medoids (PAM). The large abilities of this procedure are that it less sensitive to outliers since an entity with an extremely large value may substantially distort the distribution of data[5].In Euclidean geometry the fundamental of k-Medoids clustering algorithms is that instead of taking the mean of objects as cluster center, this does not hold for arbitrary dissimilarities, it randomly takes reference factor of the exact object in the cluster referred to as medoid which is the most centrally located object in a cluster[18]. The partitioning around medoids algorithm is based on Repeated assign each closing object to the cluster with the minimization of the sum of dissimilarities between every object and its corresponding medoid and always try to find better medoid. In other words, the goal is to find k representative objects which reduce the sum of the dissimilarities of the objects to their nearest representative object. Consequently, the algorithm as overall referred to as a consultant object-established algorithm[19]. The distance between datapoint and medoid represented the cost of the current configuration is summed over all data points. See equation(2).

$$E = \sum_{j=1}^{k} \sum |p - o_j| \qquad \ldots (2)$$

Where E is cost function, p is the medoid point in space and is o_j the non- medoids datapoints. The k-medoids procedure is more potent to the presence of noise and, however, its processing time is more high-priced Also, it may obtain different results for different runs on the same dataset because the first k medoids are chosen randomly. The analyst wishes to verify the number of clusters before the run of the algorithm[12]. The appropriate estimation value of k is detected after try and error attempts, This research determines the efficient number of clusters in Telecom dataset equal to three clusters since it gives power performance see cluster three that directly made the classification decision without the need to implement any data mining techniques by making all customer with churn attribute. Figure 6. represents each cluster with it is a specific number of churn/non-churn customers s after strolling PAM. PAM procedure based on K-medoid detailed in the algorithm(1).
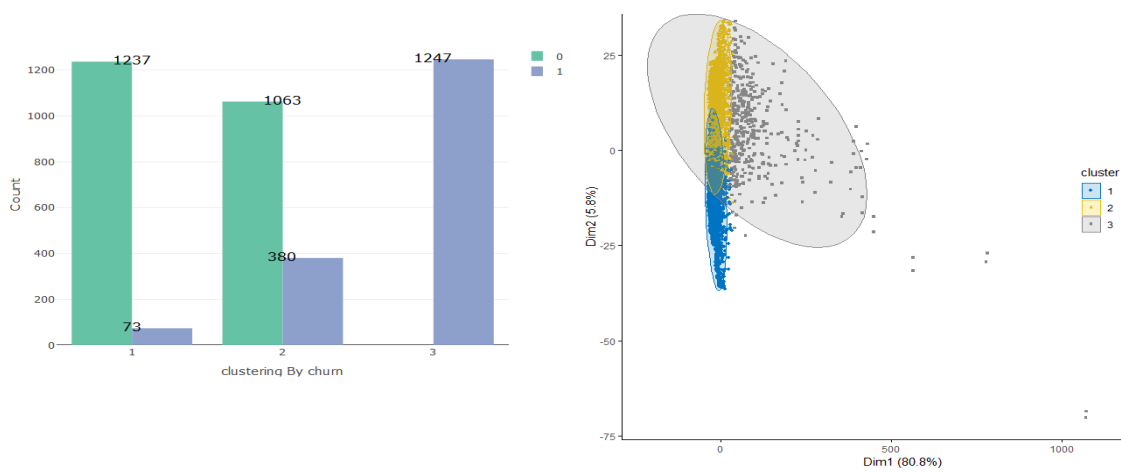

Figure 6. Number of churner/non-churner customers in each cluster

Algorithm (1)Clustering using K-medoid Procedure
Input: n of dataset attributes
Output: K clustering of the dataset based on customers attributes similarity.

1. **Initialize:** The appropriate number of clusters K using elbow system, Silhouette criterion and Gap statistic approaches
2. **Initialize:**
   Randomly choose

m of the n dataset as the medoids.

3. Scaling and centering numeric of the dataset attributes.
4. Select two random represented objects(o, m)as medoids datapoint.
5. Compute the distance between each selective objects and dataset attributes using Gower metric.
6. Associate each dataset attribute to the closest medoid.
7. Compute the total cost of the configuration, is the minimum sum of Euclidean distance between the dataset attributes and their medoid.
8. Select one of non-medoid point ó.
9. **While** the cost of configuration is decreased **Do**
10. **For** each medoid m and attribute point ó.
    - **8.1. Swap** ó with m
    - **8.2.** Associate each attribute point with the closest medoid
    - **8.3.** Recompute the total cost
    - **8.4.** Select the configuration with the lowest cost.
    - **8.5.** End for
11. **Until** the total cost for the current step is increased then undo the swap process.
12. Output the clustering vector which represents the degree of the customer to specific cluster.

the K-medoid algorithm is applied in the original CDR dataset and, in the Enrichment CDR dataset, each dataset yields three clusters dataset. At this step we have four CDR datasets could be represented by a different number of features with observation equal to 4000 customers. Table(3) display the CDR dataset with the attributes they have. the enrichment feature that is selected based on PCA are thirteen attributes, the four attributes that also added to enrichment dataset are call, callee, direction(incoming, outcoming) and duration.

Table 3. original and Enrichment CDR dataset with attributes after K-medoid algorithm

Dataset	# Attributes
Original CDR	13
Cluster1	13
Cluster2	13
Enrichment CDR	32
Cluster1_enrich	32
Cluster2_enrich	32

K-means clustering method is a popular clustering algorithm since it has less time complexity. However, it suffers from the sensitivity of outliers which may distort the distribution of data due to the extreme values. Due to the sensitivity of outlier in K-means, we are dealing with the K-medoids clustering method. Representation by K-medoids has advantages such as It presents no constraints on attributes types and preference of medoids are dictated by the location of a predominant fraction of points inside a cluster therefore, it is lesser sensitive to the presence of outliers and noise.

XGBoost Classifier Model
The proposed classification algorithm divided the Telecom dataset into two sets: the training set consists of 80% used to train the algorithm for churn predictive, and the testing group contains 20% of the dataset and used to evaluate the algorithm. In recent years, a new tree boosting method by the name eXtreme Gradient Boosting algorithm (XGBoost) emerged as the most useful, straightforward, efficient to deal with sparse data and robust solution by winning numerous data mining competitions and designed to be highly efficient, flexible and scalable in major distributed environments since it solve many data science problems in a fast and accurate way[20]. XGBoost is accurate execution of gradient boosting machines and it has demonstrated to support the restrictions for power computing of boosted trees algorithms Precisely, it was engineered to exploit every bit of storage resource for boosting the implementation of tree algorithm. XGBoost has both a linear model solution and tree learning algorithms. Therefore, it is at least 10 times faster than existing tree-decision algorithm duo to it is the ability to perform a parallel calculation[21]. XGBoost working by minimizes the regularized (L1-norm and L2-norm) objective function that combines a convex loss function represented by the differences between the predicted and target outputs and penalty terms for model complexity. The training procedures proceeds iteratively combine weak models, adding new tree-decisions that predict the residuals or errors then, bagging with previous trees to out the final prediction, weighting each of their predictions according to each week's learners performance and combine all learners predictions multiplied by their weights to obtain a single final weighted prediction that is much better than any of the base predictions themselves. It's called gradient boosting because it uses a gradient descent algorithm to reduced the loss when adding new models[22]. The XGBoost model is supervised learning typically indicates to the mathematical structure by which the prediction y_i is made from the input feature x_i, the popular model is linear combination of weight for each leaf node on the tree, where the predict value at the t-th step can be defined as in equation(3) uses K additive functions to predict the output.

$$\hat{y}_i^{(0)} = 0$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i)$$

$$\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i)$$

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad \dots(3)$$

Where $\hat{y}_i^{(t)}$ is the predicted class of the customer, x_i is the corresponding attribute vector, K is the total number of decision trees[23]. The objective function consisting of two parts: loss function and penalization terms to measure how well the model fits the training data. The learning process of training sets is based on minimization the objective function to make the bias-variance trade-off. See equation(4).

$$obj(t) = L(t) + \Omega(t) \quad \dots(4)$$

where L is the training loss function controls the predictive power, and Ω is the regularization terms control the simplicity. A common choice of the loss

function($L(t)$) is mean squared error (MSE), after taking the Taylor expansion of the loss function up to the second-order then, objective becomes as in equation(5).

$$obj(t) = \sum_{i=1}^{n}[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_t^2(x_i) + \Omega(f_t) + constant \quad \dots (5)$$

where the g_i and h_i the optimization goal for the new tree is defined as in equation (6)[24].

$$g_i = \partial_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)})$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}) \quad \dots (6)$$

After we remove all the constants, the specific objective at step t depicted in equation(7), which represents the optimization goal for the new tree.

$$obj(t) = g_i f_i(x_i) + \frac{1}{2} h_i f_t^2(x_i) + \Omega(f_t) \quad \dots (7)$$

One important advantage of this definition is that the value of the objective function only depends on g_i and h_i. The regularization term controls the complexity of the model, which helps to avoid overfitting. See equation(8).

$$\Omega(t) = \gamma T + \frac{1}{2}\lambda \sum_{t=1}^{T} w_t^2 \quad \dots (8)$$

Where T is the number of leaves, w_t^2 is the score at the t-th leaf. The loss function measures the differences between the observed and estimated factors. Ω denotes the runtime complexity of the model. The tree structure contains Internal node which split the flow of data points by one of the features using gain metric, and the Data points reach to a leaf will be assigned a weight that needs to predict it. mathematically the tree f_k can be defined as in equation(9)[23].

$$f_k(x_n) = \omega_{qk}(x_n) \quad \dots (9)$$

Where $qk(x_n)$ xn is the structure-function of the k decision tree that transforms the datapoint x_n to the equivalent leaf node, ω represent the quantized weight on leaves.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

The last stage of the predictive churn model is to assess the output of the churn prediction model. The selection of metric entirely depends on the kind and implementation plan of the prediction model. After completing the constructing of proposed churners classification system, performance evaluators on the tested Datasets will help in assessing the proposed customer churn prediction model. The comparison among the churn prediction based on original telecom dataset and the enrichment dataset after appended the centrality and network features to the original dataset using different assessment metrics e.g. accuracy, F-Score, and AUC. Table4. Shows the application of these evaluators in the suggested churn prediction system and the results of XGBoost algorithm after applying it in the original dataset, the richened dataset and the datasets resulted from clustering based on k-Medoids algorithms.

Confusion Matrix

A confusion matrix is a distinct method for brief the accuracy of a prediction model. In more Precise, the number of positive and negative predictions that summarize the number of actual values of classes. In general, it represented the actual and predicted levels for each class in square (NxN) matrix, where N is the count of classes that predicted. In case of churner classification problem in hand, N=two, and hence we get a 2x2 matrix. Some of the rates that often computed from matrices that can be verified based on a confusion matrix [14].

- **Accuracy:** count the correct predictions accomplished by the classification model over all kinds of predictions made. Overall, how often the classifier model is correct.
- **Precision:** the number of positive cases that correctly recognized.
- **Sensitivity:** the amount of actual positive cases that correctly recognized.
- **Specificity:** the amount of actual negative cases that correctly recognized. is the exact opposite of Recall or sensitivity
- **Prevalence:** How often does the yes condition occur in our sample

F-Score: Precision is invaluable for assessing the performance of data mining classifiers, but it surely leaves out some facts and for that reason will also be complicated. The recall is a portion of the true optimistic predictions to total positive observations in the dataset. Compute the percent of churn rate that appropriately categorized as churn/non-churn. The prediction models that have a low recall means it miss-classifies a great amount of the positive cases [5]. The F-Score can be outlined as in equation 10.

$$F - Score = 2.\frac{Precision \cdot Recall}{Precision + Recall} \quad \dots (10)$$

The area under Receiver Operating (AUC)

AUC Measures the subject underneath ROC curve, the diagonal line represents a random process, it has an AUC of zero:5, thus the AUC of a reputable churn classifier should be a lot higher, preferably virtually 1, as a worth of 1 represents ultimate classifier. The field-specific by means of AUC represents the chance that a random pair of churning and non-churning customers are properly identified, i.e. A positive instance receives a greater rating than a negative instance [13].

Lift Curve

The effectivity of the prediction model is expressed in the lift curve, which shows the fraction of all churners that may be caught when a designated fraction of subscribers used to be contacted. This is equal to the ratio between the sensitivity and the ratio of predicted churners after applying the churn model to the testing dataset. Formula 11 represents the lift value [14].

$$Lift = \frac{precision}{p/(p + N)} \quad \dots (11)$$

Table4. Evaluators Result for Churn Predictive Model

Evaluator	PAM Cluster based on Original Telecom Dataset			PAM Cluster based on Enrich Telecom Dataset		
	Original Dataset	Cluster1	Cluster2	Enrich Dataset	Cluster1_enrich	Cluster2_enrich
ACC	0.97000	0.98850	0.95830	0.96620	0.99240	0.98960
H	0.90005	0.78630	0.82761	0.89010	0.82567	0.96424
Gini	0.93347	0.81420	0.87320	0.92826	0.84615	0.97637
AUC	0.96673	0.90710	0.93660	0.96413	0.92308	0.98819
KS	0.93347	0.81420	0.87320	0.92826	0.84615	0.97637
MER	0.03000	0.01145	0.04167	0.03375	0.00763	0.01042
MWL	0.03222	0.01495	0.04842	0.03506	0.01451	0.00861
ER	0.03000	0.01145	0.04167	0.03375	0.00763	0.01042
F-Score	0.96296	0.85714	0.91667	0.95988	0.91667	0.97842
Youden	0.93347	0.81420	0.87320	0.92826	0.84615	0.97637
Lift	1.89700	0.63000	1.39500	2.29400	1.12000	4.17400

The above table figured the Performance statistics calculated from the confusion matrices of a predictive model based on XGBoost algorithm. Dataset enrichment with social network attributes has a strong predictive ability with an average value of all statistics accuracy not lower than 98.27%, while the average accuracy for the original dataset with it is clusters is not exceeded than 0.97%. Moreover, the results indicate that the use of PAM clustering method led to improve the performance of customer churn model as compared to the original dataset case when no clustering is used. The benchmarking that we need to perform is the one that compares original with enriching dataset and each cluster with it is corresponding cluster in rich dataset. ANOVA test based on fisher post hoc using to compare the means of proposed method see table(5) represent the calculated mean and see figure(7) represent the fisher coefficients output.

Table 5. Mean of methods using ANOVA Fisher post hoc test

Factor	Mean	StDev	95% CI
Original	0.781	0.559	(0.386, 1.176)
Cluster1	0.604	0.390	(0.210, 0.999)
Cluster2	0.708	0.453	(0.313, 1.102)
Enrichment	0.815	0.642	(0.420, 1.209)
Cluster1_enrich	0.668	0.431	(0.273, 1.062)
Cluster2_enrich	1.005	1.141	(0.610, 1.399)
Pooled StDev = 0.654369		**Analysis of Variance**	

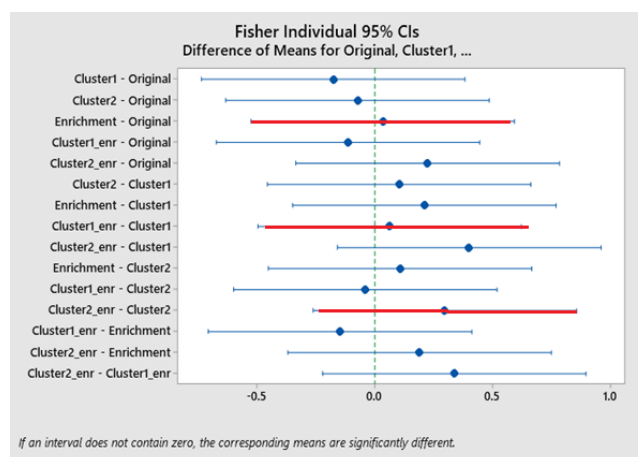| Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	5	1.086	0.2171	0.51	0.770
Error	60	25.692	0.4282		
Total	65	26.778			

Figure 7. Grouping Information Using the Fisher LSD Method

From the comparison above figure search about the comparison between the enrichment methods with original and not the opposite. The Enrichment-original, cenr-cluster_cluster1, enr-cluster2_ cluster2 are all in the positive side with red lines and this means enrichment approach outperforms the original in case of alone or in cluster form.

Big advantage after fitting XGBoost algorithm is a possibility to Measure importance of explanatory Features and therefore see which factor imports and how much it contributes in the prediction of the churn model. The attributes important can be calculated with different importance metrics, the Gain reflects the relative contribution of the feature to the churn model and calculated by taking each attribute's contribution for each tree in the XGBoost model. Coverage measure represents the relative number of customers related to this attribute. The Frequency metric measures the percentage of the relative number of times a specific attribute occurs in the trees of the XGBoost model. Figure8. Show the important attributes computed by XGBoost algorithm.
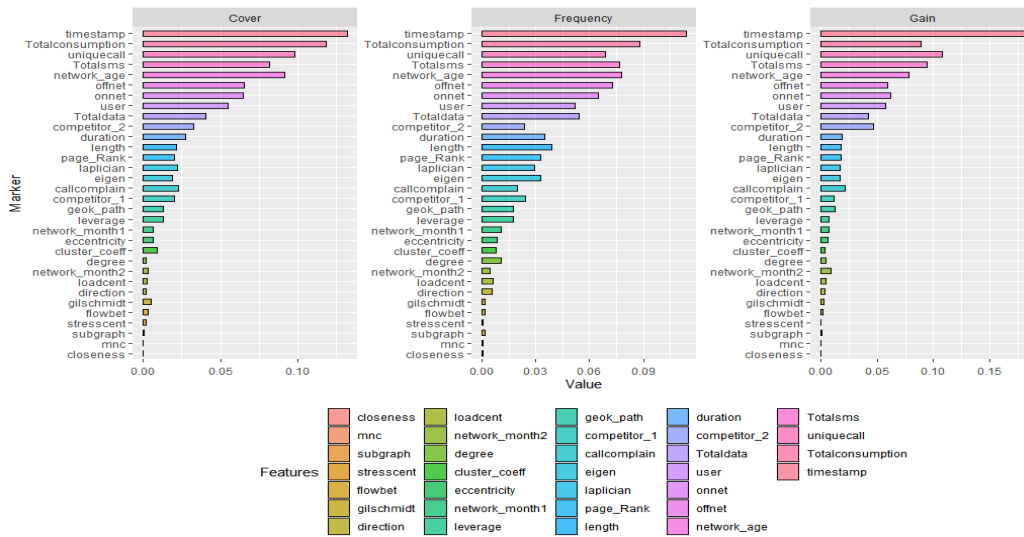
Figure8. Important attributes computed by XGBoost algorithm

The most important features as found from the above visualization using XGBoost learner that are most correlated with churn variable as compared with attributes selection based on MCA for content-based network features and important attributes extracted using PCA for content-structure based network it can be concluded that feature selection is ground for learner and potential uplift in classification model can be unlocked using important feature selection.

To sum up, Enrichment the original dataset with centrality and ties predictors was key in increasing the accuracy of the churn prediction model and suggests that the PCA method for evaluating the derived features for the model were valid and successful. Moreover, the basic classifier is modern XGBoost used to build a highly accurate classification model for the purpose of predicting customer churn. The method of exploring and selection of original attributes and appending with the social network features had a major influence on the success of the model, since the value of AUC in reached 99%. XGBoost tree model reached the best results in all metrics and it demonstrates that XGBoost is a promising approach for customer churn models. ANOVA comparison using post hoc Fisher test are used as statistical benchmarking to prove that enrichment approach outperforms any other methods in all performance measurements.

V. CONCLUSION

An enhancement strategy was developed in this research support a promote generating of new attributes using SNA techniques are related to the original Telecom dataset, focuses on the quantitative analysis of network attributes and implemented most of the metrics to quantify centrality and connectivity between network nodes. The SNA helps to identify the different types of nodes attributes in the network and on the well-organized network the most trusted nodes are central in the network. To partition the training dataset into clusters, presence of mixed data distance along with PAM algorithm was used, different criteria for choosing the optimal number of clusters in a dataset are depicted, these methods include the elbow, the silhouette and the gap statistic methods it finds the best value is three clusters. Enrichment

the original dataset with centrality and ties predictors was key in increasing the accuracy of the churn prediction model and suggests that the PCA method for evaluating the derived features for the model were valid and successful. Moreover, the basic classifier is modern XGBoost used to build a highly accurate classification model for the purpose of predicting customer churn. The method of exploring and selection of original attributes and appending with the social network features had a major influence on the success of the model, since the value of AUC in reached 99%. XGBoost tree model reached the best results in all metrics and it demonstrates that XGBoost is a promising approach for customer churn models.

REFERENCES

[1] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," J. Big Data, vol. 6, no. 1, p. 28, Dec. 2019.

[2] F. Aldahan and J. S. Grape, "Teknisk-natur vetenskaplig fakultet UTH-enheten," 2016.

[3] K. Dasgupta et al., "Social ties and their relevance to churn in mobile telecom networks," 2008, p. 668.

[4] R. Pagare and A. Khare, "Churn prediction by finding most influential nodes in the social network," in International Conference on Computing, Analytics, and Security Trends, CAST 2016, 2017, pp. 68–71.

[5] J. Manďák, "Proposal and Implementation of Churn Prediction system for Telecommunications Company," VŠB-TECHNICAL UNIVERSITY OF OSTRAVA FACULTY OF ECONOMICS DOCTORAL, Ostrava, 2018.

[6] I. Brandusoiu, G. Toderean, and H. Beleiu, "Methods for churn prediction in the pre-paid mobile telecommunications industry," in 2016 International Conference on Communications (COMM), 2016, vol. 2016-August, pp. 97–100.

[7] G. Gandhi and R. Srivastava, "ANALYSIS AND IMPLEMENTATION OF MODIFIED K-MEDOIDS ALGORITHM TO INCREASE SCALABILITY AND EFFICIENCY FOR LARGE DATASET," Int. J. Res. Eng. Technol., vol. 03, no. 06, pp. 150–153, Jun. 2014.

[8] N. Gamulin, M. Štular, and S. Tomažič, "Impact of Social Network to Churn in Mobile

Network," Automatika, vol. 56, no. 3, pp. 252–261, Jan. 2015.

[9] E. Nankani, "Deep Data Mining with Network Relationships," western Sydney, 2011.

[10] M. Dewing, "Social Media : An Introduction Social Media : An Introduction," Library of Parliament, no. 2010. pp. 1–2, 2012.

[11] M. Jalili et al., "CentiServer: A comprehensive resource, web-based application and R package for centrality analysis," PLoS One, vol. 10, no. 11, p. 8, 2015.

[12] A. K. Mallick and A. Mukhopadhyay, "Different Schemes for Improving Fuzzy Clustering Through Supervised Learning," in Communications in Computer and Information Science, vol. 1030, Springer Singapore, 2019, pp. 155–164.

[13] I. Panapakidis and G. Christoforidis, "Optimal Selection of Clustering Algorithm via Multi-Criteria Decision Analysis (MCDA) for Load Profiling Applications," Appl. Sci., vol. 8, no. 2, p. 237, Feb. 2018.

[14] J. Xiao, Y. Tian, L. Xie, X. Jiang, and J. Huang, "A Hybrid Classification Framework Based on Clustering," IEEE Trans. Ind. Informatics, no. August, pp. 1–1, 2019.

[15] T. Velmurugan, "A State of Art Analysis of Telecommunication Data by k-Means and k-Medoids Clustering Algorithms," J. Comput. Commun., vol. 06, no. 01, pp. 190–202, Dec. 2017.

[16] M. Yan, "Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion," Virginia Polytechnic Institute and State University, 2005.

[17] A. Kassambara, Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. 2017.

[18] E. Schubert and P. J. Rousseeuw, "Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms," arXiv:1810.05691v3[cs.LG], 2018.

[19] A. Batra, "Analysis and Approach: K-Means and K-Medoids Data Mining Algorithms," 5th IEEE Int. Conf. Adv., no. 274, pp. 274–279, 2011.

[20] T. Chen and C. Guestrin, "XGBoost," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 2016, pp. 785–794.

[21] L. Ruisen et al., "Bagging of Xgboost Classifiers with Random Under-sampling and Tomek Link for Noisy Label-imbalanced Data," IOP Conf. Ser. Mater. Sci. Eng., vol. 428, no. 1, p. 012004, Oct. 2018.

[22] R. Santhanam, N. Uzir, S. Raman, and S. Banerjee, "Experimenting XGBoost Algorithm for Prediction and Classification of Different Ramraj S, Nishant Uzir, Sunil R and Shatadeep Banerjee Experimenting XGBoost Algorithm for Prediction and Classi fi cation of Different Datasets," Int. J. Control Theory Appl., vol. 9, no. March, pp. 651–662, 2017.

[23] M. Hassan Elbedawi Omar, M. Borrotti, and A. Corti, "Customer Churn prediction based on eXtreme Gradient Boosting classifier," IMATI-CNR, Milano, 2018.

[24] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 2016, vol. 19, no. 6, pp. 785–794.