Telecommunication Service Subscriber Churn Likelihood Prediction Analysis Using Diverse Machine Learning Model

Ngurah Putu Oka H, Ajib Setyo Arifin
Department of Electrical Engineering
Universitas Indonesia
Jakarta, Indonesia
ngurah.putu@ui.ac.id, ajib@eng.ui.ac.id

***Abstract*— The biggest problem that occurs in the telecommunication industry is increased level of customer churn. This is a very important problem that must be resolved by the company because customers who stop will have an impact on company retention. The usage of the machine learning model will certainly be able to help to predict customer trends and making precise decisions in the future. To get good results, this study is analyzed with one algorithm that had never been analyzed in previous studies to make predictions, namely Deep Neural Network (DNN). DNN compared to models that have been tested before, Random Forest and Extreme Gradient Boosting (XGBoost). This research analyzed the importance of the features, the handling toward the selection of appropriate features, and simplified the process of gathering data. The proposed model was trained and tested over Google Colaboratory using TensorFlow backend. The testing that has been done produces very good results for the Deep Neural Network (DNN) model, with a process of 68 seconds and an accuracy of 80.62%. Extreme Gradient Boosting (XGBoost) produces 76.45% accuracy with a processing time of 175 seconds, and random forest produces 77.87% with a sufficiently long processing time of up to 529 seconds.**

Keywords*— *Churn Customer, Machine Learning, Tensorflow, Feature Importance, D-NN

I. Introduction

The penetration of cellular subscribers is very high. One of the challenges that cellular telecommunications operator companies will face in the future is an effort to reduce the number of customers who stop using company services and move to competitor companies [1]. The main concern of telecommunications companies is the right way to retain customers. With the development of technology, many companies in the telecommunications sector are using data printing techniques to predict customer churn rates [2]. Churn customers are customers who unsubscribe and move to another company, due to various factors. It is necessary to evaluate whether the big problems of churn customers and the company's management will make appropriate strategies to minimize the churn and retaining the customer [3]. Customer churn is a very important problem that must be faced by the company because the cessation of customers will have an impact on company retention. Customer churn can be caused by many things, ranging from competitive rates between operators, competitive features and facilities, to how providers serve, interact, and manage relationships with their customers [2]. In the telecommunications industry, the mobile cellular market is the segment that sees the fastest and nearly saturated growth. In retaining customers, each company has a different strategy in marketing. One way, companies must allow customers who want churn before asking this company to send proactive retention promotions. [2]. The predictive model will speed up the retention process and cellular telecommunications companies will achieve positive results in this competitive market. Positive results in question are able to compete well with competitors in the market by being able to predict customer churn and can create the best business strategy for customers.

Besides current technological advances in the big data field, there are many data mining and machine learning solutions that can be used to conduct data analysis [4]. Data mining is an analytical process designed to explore data to find patterns or systematic relationships that are consistent between variables and then validated [5]. Data mining techniques are very important for the prediction process that is important to improve the results obtained by machine learning. To extract knowledge from data, the process of mining data uses Machine Learning algorithms, statistics, pattern recognition, and visualization techniques [2]. By using the type of Machine learning algorithm technique, it is expected to be able to apply predictive models to find future trends and behaviors that enable telecommunications companies to make smarter decisions based on the knowledge extracted from data. Companies can use this technique to maximize profits. Also, this can be used to design appropriate strategies to retain customers who have potential churn. Data mining identifies the churn customer behavior of the extracted pattern based on the data used [4].

In this study, the churn level prediction process is carried out using machine learning. The dataset used in this study consisted of detailed call records and was obtained from one of the largest information technology companies in the world located in the United States, IBM. The amount of customer data used is 7040 cellular subscribers. Based on these data 75% is used for training data and 25% is used for testing data. Customers with each Churn dependent variable with two, Churn / Nonchurn classes, and 20 supporting variables for the processing in the machine learning model [2]. Before entering the stages of machine learning modeling, the preprocessing process is carried out in advance so that the data can be classified properly by the machine learning algorithm. This research conducted a process of feature importance analysis to overcome the lack of data and make it easier to do data gathering, besides that by knowing the feature importance, it can be seen what causes the most churn. Previous studies rarely analyzed this process. The machine learning model used in this study is the Deep Neural Network (DNN). To our knowledge, there are no previous studies that have conducted research on this machine learning model. To get maximum results this model will be compared with Random Forest and Extreme Gradient Boosting (XGBoost) that have been tested before.

The first part of this study discusses the background of why this study was carried out and how the strategies to overcome it. In the second part, it is explained about several related articles that have researched several machine learning models. The third section discusses the process flow used to complete this research. The fourth part compares and analyzes the machine learning model to get the best machine learning model, and in the end there are conclusions and some suggestions for further research.

II. Related Work

Previous studies about churn prediction in the telecommunication industry mainly applied machine learning techniques such as support vector machine, decision trees, and logistic regression to predict churn levels[6]. However, to our knowledge, no studies are comparing the Machine Learning Model of the Neural Networks and XGBOOST collected from a customer of one of IBM's technology companies to predict churn levels.

In 2017 [7] , Studied the churn of telecommunications company customers in India. According to their research, Churn Prediction will be more realistic and accurate. this is very important for the early detection of high-risk customers who leave the service. according to this study, Ensemble-based Classifiers namely Bagging, Boosting, and Random Forest are used for Churn Prediction in the telecommunications industry. Ensemble-based classifiers have been compared with well-known classifiers namely Decision Tree, Naïve Bayes Classifier, and Support Vector Machine (SVM). The results of their experiments show that Random Forests have fewer error rates, lower specificity, high sensitivity, and greater accuracy compared to other methods.

Another research about the churn rate in India telecommunication industry is studied by [8]. CRM is very important because it carries out a long-term relationship between the customer and the company. Churn can hamper the growth of the customer table and it is the biggest challenge to maintain telecommunications companies. The research they propose is 2 models that predict customer churn with a high degree of accuracy, namely the logistic regression model which is a non-linear classification with a sigmoid activation function. The second model is a full-edged Neural Multilayer Perceptron Network (MLP) with normalized input feature vectors stacked with three hidden layers and employs binary cross-entropy as a loss function. Using this predictive model, organizations can conduct marketing research and study the specific needs of customers in detail.

As can be seen on [4], according to their analysis Decision-makers and businesses, emphasizing that getting new customers is more difficult than retaining existing customers. Churn customers provide a factor behind customers who rotate in the telecommunications sector. The model proposed by them displays first-class customer data using a classification algorithm, where the Random Forest (RF) algorithm performs better. After classification, the proposed model segmented mixing customer data by categorizing customers who are mixed in groups using cosine similarity to provide group-based retention offers.

In [9] they propose telecommunications operators, where traditional sources of revenue, voice, and SMS, are shrinking because customers are using over-the-top (OTT) applications such as WhatsApp. In this challenging technological development, telecommunications companies need to maintain or grow to support the market of the OOT, by providing a user-friendly experience on their network. However, they collect and predict the quality and experience of users in telecommunications networks directly, that is the problem they discuss in this research. they provide advice to discuss, in the real-time experience of cellular customers in assessing which conditions cause users to communicate with telecommunications customer service centers. For this purpose, they follow supervision for learning and prediction in the Random Limited forest model we use, as a proxy for bad experiences, which addresses customer transactions in telco data feeds before users make calls to customer service centers. We assess that it is recommended that they use the rich datasets provided by major African telecommunications companies and the new big data architecture for training and assessment of predictive models. Our empirical studies prove effective solutions in predicting user experience by considering whether customers will have a conversation based on the current context. This promising result was released for service users, which will help telecommunications companies to reduce churn rates and increase the number of their customers.

In another study [10], Churn levels, which are used to predict growth, are currently considered as important as metrics as well as finance. By winning a competition in the market, the company is eager to maintain the lowest possible churn rate. Thus, churn predictions are very important, not only for existing customers but also to predict future customer trends. In their research, they showed the prediction of stirring in the Telco dataset using the Far Learning assessment. Multilayered Neural Networks are designed to build a non-linear classification model. The published model then issues the final weights for these features and predicts churn expenditure for this customer. Because this model also provides churn factors, it can be used by companies to analyze the reasons for those factors and take steps to retention.

According to [11] analysis and prediction of churn customers in the telecommunications sector is now a problem because the telecommunications industry needs to analyze various customer behavior to predict customers who will leave a subscription from a telecommunications company. So data mining techniques and algorithms play an important role for companies in today's commercial conditions. In this paper we compare the classification of models such as Logistic Regression, SVM, Random Forest and Gradient Tree Improvement and also compare the performance of this model. The best model will help reduce the level of churn by doing targeted analysis.

According to [12], They explained the experiment was carried out on data provided by the telecommunications company - Orange, to predict churn. The preprocessing stage of the experiment includes the removal of missing values and redundant data, Lasso and manual engineering features. Convolutional Neural Network is applied as a classifier on preprocessed one-dimensional datasets with fairly good accuracy. It can be concluded that the proposed model can be applied in telecommunications systems for detection.

Another Study [13], focussed their research on New telecommunications companies that provide various options and services to attract various customers to switch from support services to their services because customers may not like old licenses and services. Therefore, the goals of success are companies that maintain their existing goals. Therefore it is necessary or companies to learn in advance which customers can switch from their services to services. They do an analysis predicting which customers might churn or not use customer data with the help of Apache PIG. This will help telecommunications companies to know in advance which customers can switch from their services to their purchasing services. Make them able to do something for customers and keep them first. The company will avoid big losses.

III. The Proposed Method

In this study, the prediction process will use machine learning. By comparing 3 algorithms from machine learning. The algorithms used are Random Forest, XGBoost, and Deep Neural Network. Before entering the TV, data crowding, Standardization, and Normalization data are carried out to facilitate data processing. The dataset used is a dataset from IBM with 7040 customers. To ensure the efficiency of the model formed, the data will be collected into learning data (train dataset) and test data (dataset test). The division of data used varies depending on the algorithm used. Learning (train dataset) learning to train to learn from machine learning and tests (test dataset) to test the performance of the machine learning model used. Following the process block diagram in Fig 1.

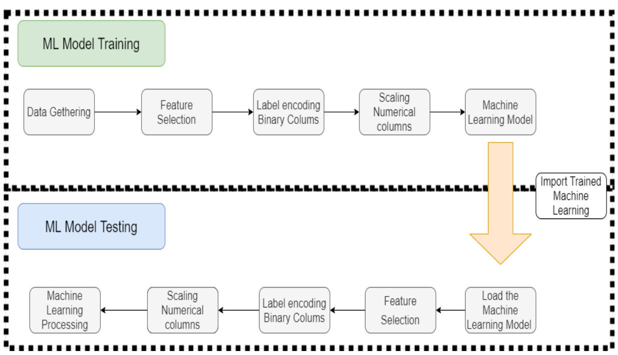

Fig. 1. Block Diagram Process Machine Learning Model

A. Data Gathering

The data collection process is used from a customer dataset of one of the largest information technology companies in the world located in the United States namely the International Business Machines Corporation (IBM), data that contains customer data that is telecommunications network from IBM. The data used is a dataset that was updated in 2018. The collected data totaled 7043 customers with 21 features, 73.4% of non-churn data and 26.6% of churn data. This data is used because it has a variety of features and no missing pieces of data, it is very suitable to be used as a model for machine learning.

B. Feature Selection

Feature selection is a very important part of machine learning, this process is crucial for the performance of the model. The data features that are used to train the machine learning model, have a big influence on the development model that will be achieved. Irrelevant features can be used negatively in the modeling process. The choice of features should be the first step in designing the model [14]. Process Selection Feature before modeling data [14]:

- Reduces Overfitting: Data that is not used can be filtered so that not too much data is processed, thus fewer opportunities to make decisions with data as outliers
- Improves Accuracy: Selection of the right data, it certainly will improve modeling accuracy.
- Reduces Training Time: Selection of the right data, it certainly will improve modeling accuracy

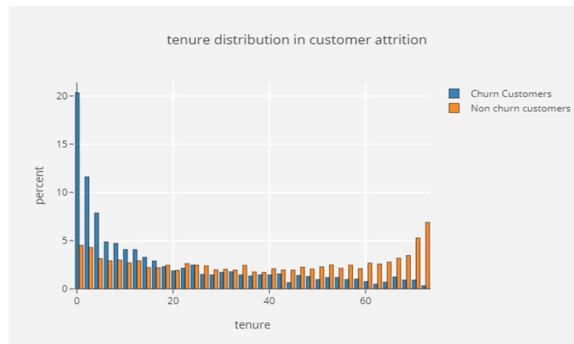

Fig. 2 Tenure Distribution in Customer Attrition

Fig 2 shows some sample data processing in the Feature Selection process. Feature selection is the process of selecting features from data that has been collected. Feature Selection is an activity that generally can be done by preprocessing and aims to select features that are influential and rule out features that have no effect in a data modeling or analysis activity.

C. Label Encoding Binary Columns

In machine learning techniques all variables used for input and output must be converted to numeric. Then the categorical data must be coded into numeric before doing an evaluation of a model. The most popular techniques are integer encoding and hot encoding. In the development of newer techniques namely embedding learning can provide a middle ground that can process these two methods [15]. Label Encoding Binary Columns is a common process of data preprocessing, this process is done by standardizing data. Data in the form of strings will be converted into data in binary form (0 1) to facilitate data processing and speed up time processing in the machine learning model.

D. Scaling Numerical Columns

Scaling Numerical Columns is part of preprocessing data. In this process, the data normalization is done in which the process of equalizing the range of data in each column is uniform. The root-mean-square for a (possibly centered) column is defined as [16],

$$\sqrt{\Sigma(x^2)/(n-1)} \tag{1}$$

where x is a vector of the non-missing values and n is the number of non-missing values [16].

E. Machine Learning Model

The algorithm method will be used to process machine learning. In this study, three machine learning model algorithms were compared, namely Random Forest, Xgboost, and Deep Neural Network.

1. Deep Neural Network Model

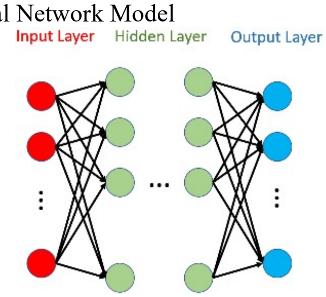

Fig. 3 Deep Neural Network Model

Artificial Neural network (ANN) is a computation algorithm which mimics the way how human brain works. ANN is structured by three main interconnected layers: input, hidden, and output layer. Each layer consisted of at least one neuron that contains bias value as a processing unit to process the value which is passed into it. A weight value is also available in a link between two interconnected neurons. ANN is used to learn the pattern from given input and its related output and then use it on other input with unknown output. ANN learn the pattern by feeding the input into the hidden layer and then provide the final result at the output layer. Then, loss value is calculated based on the distance between the real output and the expected output. This loss value is then used to update the internal structure of the neural network (bias and weight)

There is only one input layer and one output layer in the ANN system. However, it is possible to build an ANN with more than hidden layer as depicted in fig 3. Adding more layer to the hidden layer can improve the performance of the neural network by extracting more abstract feature from given input-output. This structure is usually called as deep neural network (DNN). D-NN has been proven to generate more accurate prediction compared to the shallower architecture. However, adding more elements to the hidden layer adds up to the complexity of the system and thus resulted in higher processing time.

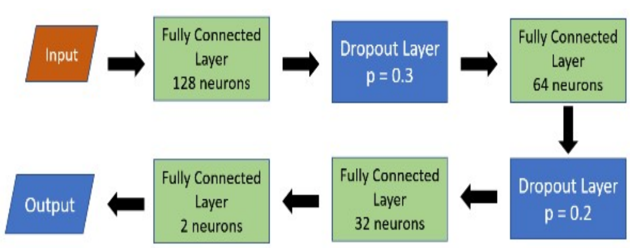

Fig. 4 Flow Diagram Deep Neural Network

In this paper, we propose a DNN algorithm which is consisted of four fully connected (dense) layer and two dropout layer (fig 4). We use 128 neurons at the first fully connected layer and only two neurons at the last layer (which is directly translated to the output: churn and not churn). We use dropout layer to reduce the overfitting effect that might occur as we use a complex architecture. This helps neural network to adapt to the change of input data in which the accuracy in the testing stage will not be totally different with the training stage. The dropout layer itself works by dropping some neurons from its previous fully connected layer. This is why dropout layer only placed after two hidden layers that have the most number of neurons in the network.

2. Extreme Gradient Boosting (XGBoost) Model

The gradient boosting model was developed into Extreme Gradient Boosting (XGBoost), consisting of several weak prediction models which are ensemble models. XGBOOST will increase the use of good settings to avoid overfitting thereby increasing the performance of this model. XGBOOST generates predictions by transferring the function as agreed in (2), which consists of the training data function and the regularization function. [16].

$$\mathrm{obj} = \sum_{i=0}^{n} \mathrm{l}(y_i, \gamma_i^{(t)}) + \sum_{i=1}^{t} \varphi(f_i) \qquad (2)$$

This model makes it impossible to make all tree ensemble models at the same time because this model is practical. XGBOOST uses additional training data strategies, it aims to approve the knowledge, and add more trees to the furniture one by one, each tree will optimize the objective function [16].

3. Random Forest Model

Random Forest can be used to handle imbalanced data, works like a decision tree that uses modeling in the form of a decision tree, and then assigns value to each tree. Two ways to handle imbalance data in Random Forest, namely Balanced Random Forest (BRF) based on sampling techniques and Weight Random Forest (WRF) based on cost-sensitive learning [15]. In general, the pseudocode for the Random Forest (RF) algorithm is as follows:

- Let N be the sum for training data and the number of variables for the classifier is M.
- Determine the value of the variable m that will be used to determine the decision at the tree node; m <M.
- Choose a training set to form a decision tree (Boostrap) by selecting n times randomly (with replacement) from all N training cases. Use the rest of the unselected cases (Out of Bag) to estimate the error of the tree, by predicting its class.
- For each tree node, randomly select (without replacement) the m variable that will be used to base the decision tree on that node. Calculate the best split based on this variable m in the training set.
- Each tree is formed without pruning.
- Prediction results are obtained from the calculation mode of each decision tree.

IV. Experiment and Result

Experiments were carried out on the proposed model using machine learning techniques on a data set. The next section presents the results obtained using different machine learning techniques. Further, it provides the factor and compares the models behind the customer churn. The proposed model was trained and tested over google collaboratory using TensorFlow backend [1].

A. Dataset Descripcion

In this study, the dataset used is telecommunications customer data from information technology companies in the United States namely IBM [2]. The data used is a dataset that was updated in 2018. This data is used because it has a variety of features and no missing pieces of data, it is very suitable to be used as a model for machine learning. This dataset provides 7043 customers with 21 features, 73.4% of non-churn data and 26.6% of churn data. From this data 75% is used as training data and 25% is used as testing data.

TABLE I. FEATURE IMPORTANCES

Fitur	Nilai ROC
Contract_Month-to-month	0.1984
tenure	0.1498
InternetService_Fiber optic	0.1384
Contract_Two year	0.1108
TotalCharges	0.0749
PaymentMethod_Electronic check	0.0737
MonthlyCharges	0.0628
InternetService_No	0.0486
Contract_One year	0.0365
PaperlessBilling	0.0244

Table I shows the results of the analysis for feature importance. This important feature is very important to be analyzed because thus it can be determined what features will greatly affect the level of customer churn besides this process will facilitate the data gathering process. Based on the analysis results in Fig 10, the most influential features in this study contract month to month, Internet Servie, and Tenure.

B. Simulatiom Result

In this study, the proposed churn prediction model is evaluated using AUC , Accuracy , and Spesifity .

1. Peformance Result

TABLE II. RANDOM FOREST MODEL

Measurement	Value	Standart Deviation
Accuracy	77.87%	±1.5%
AUC	82.78%	±1.4%
specificity	86.54%	±0.8%

TABLE III. XGBOOST MODEL

Measurement	Value	Standart Deviation
Accuracy	76.45%	±1.1%
AUC	84.47%	±2.7%
specificity	95.26%	±1.2%

TABLE IV. RANDOM FOREST MODEL

Measurement	Value	Standart Deviation
Accuracy	80.62%	±1.0%
AUC	84.36%	±2.6%
specificity	95.35%	±0.8%

[1] https://colab.research.google.com/notebooks/intro.ipynb#recent=true
[2] https://www.kaggle.com/blastchar/telco-customer-churn

From the table above, it is explained that D-NN has a fairly low error rate and has a higher specificity than the comparison model.

2. Confusion Matrix

TABLE V. RANDOM FOREST MODEL

	Nonchurn	Churn	precision
predict. Nonchurn	1266	266	82.64%
predict. Churn	197	284	59.04%
class recall	86.50%	51.64%	

TABLE VI. RANDOM FOREST MODEL

	NonChurn	Churn	precision
predict. Nonchurn	1413	335	80.84%
predict. Churn	69	195	73.86%
class recall	95.34%	36.79%	

TABLE VII. RANDOM FOREST MODEL

	NonChurn	Churn	precision
predict. Nonchurn	1413	338	80.70%
predict. Churn	69	192	73.56%
class recall	95.34%	36.23%	

Based on the confusion matrix in tables IV, V, and VI the results of the analysis show the level of precision against non-churn shows good results in Random Forest that is 82.64%, for a good level of precision churn is xgboost which is 73.86%. For Recall Nonchurn D-NN showed a very good result that is 95.34% and for churn, Random Forest gets better results 51.64%. Based on the analysis that has been made the prediction result of non-churn is better than the churn level because the amount of non-churn data is more than the churn data.

C. Performance Analysist

In this study, what will become a reference for analysis is the accuracy and processing time of data processing obtained from each machine learning model.

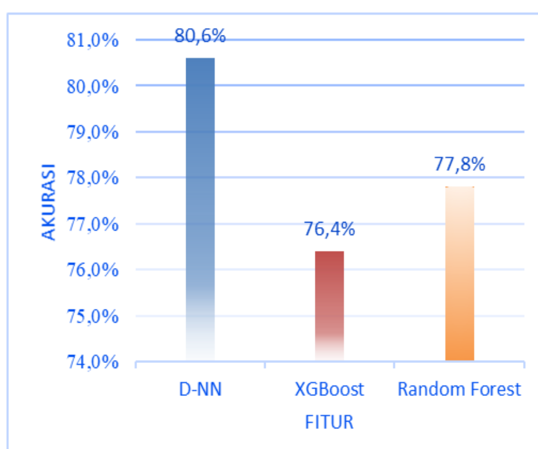

Fig. 6 Accuracy Comparison

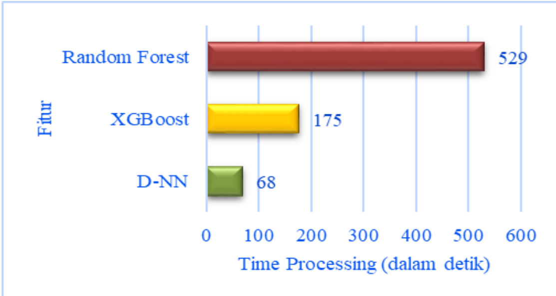

Fig. 7 Time Processing Comparison

Can be seen in Fig 6 and 7, Deep Neural Network (DNN) Is a Good Machine Learning Algorithm Model used for the Churn Prediction Model in this study when Seen Deep Neural Network Getting 80.62% accuracy and the fastest processing time is 68 seconds. On the other hand, the accuracy of Random Forest is not much different from XGBoost, but it requires a long processing time of up to 529 seconds.

V. Conclusion

In this paper, the prediction of customer churn rates in telecommunications companies, using machine learning algorithm modeling. To get accurate prediction results, this study tested the Deep Neural Network (D-NN) machine learning model. Because this model has never been tested in previous studies, to measure the accuracy of this model, the writer compares D-NN to several models that have been previously studied, namely Random Forest and Extreme Gradient Boosting (XGBoost). This research also analyzed the importance of this feature aims to simplify the process of gathering data and can manage the features that most influence the churn level of the customer. Based on the analysis, the most influential features in this study are contracts month to month, Internet Service, and Tenure. From the results obtained the Deep Neural Network has very good results with an accuracy of 80.62% with the fastest processing time of 64 seconds. For Random Forest and XGBoost have an accuracy that is not much different, for Random Forest 77.87% and XGBoost 76.45%. However, the random forest process takes up to 529 seconds while XGBoost only takes 175 seconds. This proves DNN is feasible to use to predict churn levels.

But the factors that need to be considered in future research are, First, the limitations of the dataset which are still using common datasets and with a small amount of data, in the future can take a more recent dataset with a greater amount of data to get a better analysis of the results than now.

Acknowledgment

This research is funded by PUTI Prosiding Universitas Indonesia Grant 2020 with contract number: NKB-1152/UN2.RST/HKP.05.00/2020.

References

[1] P. E. Putra and R. Gadang, "The Design Of A Churn Prediction Model Using Data Mining For Customer Retention (A CASE OF POSTPAID CUSTOMERS OF PT . XYZ CELLULAR)," vol. 7, no. X, pp. 1–16, 2018.

[2] I. Brându·oiu, G. Toderean, and H. Beleiu, "Methods for churn prediction in the 3re-3aid Mobile telecommunications Industry," pp. 97–100, 2016.

[3] A. L. Saonard, "Modified Ensemble Undersampling-Boost to Handling Imbalanced Data in Churn Prediction," 2020.

[4] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, vol. 7, pp. 60134–60149, 2019.

[5] E. N. Anggraini et al., "Big Data : Clustering Using Expectation-Maximization Algorithm with Gausians Mixture Models For Analysis of tren Product from ecommerce in Indonesia," J. Sci. Technol. JST-1903-2020, 2020.

[6] F. Napitu, M. A. Bijaksana, A. Trisetyarso, and Y. Heryadi, "Twitter opinion mining predicts broadband internet's customer churn rate," 2017 IEEE Int. Conf. Cybern. Comput. Intell. Cybern. 2017 - Proc., vol. 2017-Novem, pp. 141–145, 2018.

[7] A. Mishra, "A Comparative Study of Customer Churn Prediction Classifiers," no. Icici, pp. 721–725, 2017.

[8] S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra, and S. Kumar, "Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron(MLP)," Proc. 2nd Int. Conf. Green Comput. Internet Things, ICGCIoT 2018, pp. 436–438, 2018.

[9] E. Diaz-Aviles et al., "Towards real-time customer experience prediction for telecommunication operators," Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015, pp. 1063–1072, 2015.

[10] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning," 2018 Int. Conf. Smart Comput. Electron. Enterp. ICSCEE 2018, pp. 1–6, 2018.

[11] A. Gaur and R. Dubey, "Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques," 2018 Int. Conf. Adv. Comput. Telecommun. ICACAT 2018, pp. 1–5, 2018.

[12] M. Karanovic, M. Popovac, S. Sladojevic, M. Arsenovic, and D. Stefanovic, "Telecommunication Services Churn Prediction - Deep Learning Approach," 2018 26th Telecommun. Forum, TELFOR 2018 - Proc., pp. 420–425, 2018.

[13] A. Tiwari, R. Sam, and S. Shaikh, "Analysis and prediction of churn customers for telecommunication industry," Proc. Int. Conf. IoT Soc. Mobile, Anal. Cloud, I-SMAC 2017, pp. 218–222, 2017.

[14] R. Shaikh, "Feature Selection Techniques in Machine Learning with Python," 2018. [Online]. Available: https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e. [Accessed: 27-Apr-2020].

[15] J. Brownlee, "3 Ways to Encode Categorical Variables for Deep Learning," Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python/. [Accessed: 26-Apr-2020].

[16] A. R. Becker, R. A., Chambers, J. M. and Wilks, "Scaling and Centering of Matrix-like Objects," Web, 1988. [Online]. Available:https://stat.ethz.ch/R-manual/R patched/library/base/html/scale.html.