



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 11, November 2020

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Customer Behaviour Prediction Using Machine Learning

Poonam W. Jadhav, Rahul S. Gaikwad

PG Student, Dept. of C.S.E.,G.C.O.E.,DBATU University, Lonere, India

Assistant Professor, Dept. of C.S.E.,G.C.O.E, DBATU University, Lonere, India

ABSTRACT: Nowadays, customers have become more interested in the quality of service (QoS) that organizations can provide them. Services provided by different vendors are not highly distinguished which increases competition between organizations to maintain and increase their QoS. Customer Relationship Management systems are used to enable organizations to acquire new customers, establish a continuous relationship with them and increase customer retention for more profitability. CRM systems use machine-learning models to analyze customers' personal and behavioral data to give organization a competitive advantage by increasing customer retention rate. Those models can predict customers who are expected to churn and reasons of churn. Predictions are used to design targeted marketing plans and service offers. This paper tries to compare and analyze the performance of different machine-learning techniques that are used for churn prediction problem. Ten analytical techniques that belong to different categories of learning are chosen for this study. The chosen techniques include Discriminant Analysis, Decision Trees (CART), instance-based learning (k-nearest neighbors), Support Vector Machines, Logistic Regression, ensemble-based learning techniques (Random Forest, Ada Boosting trees and Stochastic Gradient Boosting), Naïve Bayesian, and Multi-layer perceptron. Models were applied on a dataset of online shoppers from UCI machine learning repository contains 12330 sessions. Results show that both random forest outperform all other techniques with almost the same accuracy 96%. Both Multi-layer perceptron and Support vector machine can be recommended as well with 94% accuracy. Decision tree achieved 90%, naïve Bayesian 88% and finally logistic regression and Linear Discriminant Analysis (LDA) with accuracy 86.7%.

KEYWORDS: Customer relationship management (CRM); customer retention; analytical CRM; business intelligence; machine-learning; predictive analytics; data mining; customer churn

I. INTRODUCTION

For any business, customers are the basis for its success and revenue and that is why companies become more aware of the importance of gaining customers' satisfaction. Customer relationship management (CRM) supports marketing by selecting target consumers and creating cost-effective relationships with them. CRM is the process of understanding customer behavior in order to support organization to improve customer acquisition, retention, and profitability. Thus, CRM systems utilize business intelligence and analytical models to identify the most profitable group of consumers and target them achieve higher customer retention rates. Those models can predict customers with high probability to churn based on analyzing customers' personal, demographic and behavioral data to provide personalized and customer-oriented marketing campaigns to gain customer satisfaction. The lifecycle of business – customer relationship includes main stages:

4) 1) *identification*; 2) *attraction*; 3) *retention*; and 4) *development*.

1) *Customer identification/acquisition*: This aims to identify profitable customers and the ones that are highly probable to join organization. Segmentation and clustering techniques can explore customers' personal and historical data to create segments/sub-groups of similar customers .

2) *Customer attraction*: The identified customer segments

/ sub-groups are analyzed to identify the common features that distinguish customers within a segment. Different marketing techniques can be used to target different customer segments such targeted advertising and/or direct marketing.

3) *Customer retention*: This is the main objective of CRM as retaining existing customers is at least 5 to 20 times more cost effective than acquiring new ones depending on business domains. Customer retention includes all actions taken by organization to guarantee customer loyalty and reduce customer churn. Customer churn refers to customers moving to a competitive organization or service provider. Churn can be for better quality of service, offers and/or benefits. Churn rate is an important indicator that all organizations aim to minimize. For this sake, churn prediction is an integral part of proactive customer retention plan. Churn prediction includes using data mining and predictive

analytical models in predicting the customers with high likelihood to churn/defect. These models analyze personal and behavioral customer data for tailored and customer-centric retention marketing campaigns.

4) *Customer development*: The main objective of this phase is to increase the amount of customer transactions for more profitability. For this sake, market basket analysis, customer lifetime value, up, and cross selling techniques are used. Market basket analysis tries to analyze customers' behavior patterns to maximize the intensity of transactions. Analyzing customer lifetime value (CLTV) can help identifying the total net income expected from customer. Up and/or Cross selling include activities that increase the transactions of the associated services/products.

Customer retention and churn prediction have been increasingly investigated in many business domains, including, but not limited to, telecommunication, banking, retail and cloud services subscriptions. Different statistical and machine-learning techniques are used to address this problem. Many attempts have been made to compare and benchmark the used techniques for churn prediction. In a comparison between (Decision trees, Logistic regression and Neural Network) models was performed. The study found that neural network perform slightly higher than the other two techniques. Another comparison between a set of models against their boosted versions is discussed in. This study included two-layer Back-Propagation neural network (BPN), Decision Trees, SVM and Logistic Regression.

Most of comparisons in the literature did not consider a study that covers the various categories of learning techniques. The bulk of the models applied for churn prediction fall into one of the following categories:

1) Regression analysis, 2) Decision tree-based, 3) Support Vector Machine, 4) Bayesian algorithm, 5) Instance – based learning, 6) Ensemble learning, 7) Artificial neural network, and 8) Linear Discriminant Analysis.

This study presents a comparative study of the most used algorithms for predicting customer churn. The comparison is held between algorithms from different categories. The main goal is to analyze and benchmark the performance of the models in the literature. The selected models are:

- 1) Regression analysis: logistic regression.
- 2) Decision tree–CART.
- 3) Bayes algorithm: Naïve Bayesian.
- 4) Support Vector Machine
- 5) Instance – based learning: k-nearest Neighbor.
- 6) Ensemble learning: Ada Boost, Stochastic Gradient Boost and Random Forest.
- 7) Artificial neural network: Multi-layer Perceptron.
- 8) Linear Discriminant Analysis.

A. Contribution

The key contribution of this paper is the analysis of most common learning techniques in the state of the arts and the evaluation of their accuracy.

II. MACHINE LEARNING FOR CHURN PREDICTION

Machine-learning techniques have been widely used for evaluating the probability of customer to churn. Based on a survey of the literature in churn prediction, the techniques used in the bulk of literatures fall into one of the following categories 1) Regression analysis; 2) Tree – based; 3) Support Vector Machine; 4) Bayesian algorithm; 5) Ensemble learning; 6) Sample – based learning; 7) Artificial neural network; and 8) Linear Discriminant Analysis. A brief introduction of the chosen algorithms is presented in this section

1) *Regression analysis*: Regression analysis techniques aim mainly to investigate and estimate the relationships among a set of features. Regression includes many models for analyzing the relation between one target/response variable and a set of independent variables. Logistic Regression (LR) is the appropriate *regression analysis* model to use when the dependent variable is binary. LR is a predictive analysis used to explain the relationship between a dependent binary variable and a set of independent variables. For customer churn, LR has been widely used to evaluate the churn probability as a function of a set of variables or customers' features.

2) *Decision Tree*: Decision Tree (DT) is a model that generates a tree-like structure that represents set of decisions. DT returns the probability scores of class membership. DT is composed of: a) **internal Nodes**: each node refers to a single variable/feature and represents a test point at feature level;

b) **branches**, which represent the outcome of the test and are represented by lines that finally lead to c) **leaf Nodes** which represent the class labels. That is how decision rules are established and used to classify new instances. DT is a flexible model that supports both categorical and continuous data. Due to their flexibility they gained popularity and became one of the most commonly used models for churn prediction.

3) *Support Vector Machine*: Support Vector Machine (SVM) is a supervised learning technique that performs data analysis in order to identify patterns. Given a set of labeled training data, SVM represents observations as points in a high-dimensional space and tries to identify the best separating hyperplanes between instances of different classes. New instances are represented in the same space and are classified to a specific class based on their proximity to the

separating gap. For churn prediction, SVM techniques have been widely investigated and evaluated to be of high predictive performance.

Bayes Algorithm: Bayes algorithm estimates the probability that an event will happen based on previous knowledge of variables associated with it. Naïve Bayesian (NB) is a classification technique that is based on Bayes' theorem. It adopts the idea of complete variables independence, as the presence/absence of one feature is unrelated to the presence/absence of any other feature. It considers that all variables independently contribute to the probability that the instance belongs to a certain class. NB is a supervised learning technique that bases its predictions for new instances based on the analysis of their ancestors. NB model usually outputs a probability score and class membership. For churn problem, NB predicts the probability that a customer will stay with his service provider or switch to another one.

- 1) **Instance – based learning:** Also known as **memory- based learning**, new instances are labeled based on previous instances stored in memory. The most widely used instance based learning techniques for classification is K-nearest neighbor (KNN). KNN does not try to construct an internal model and computations are not performed until the classification time. KNN only stores instances of the training data in the features space and the class of an instance is determined based on the majority votes from its neighbors. Instance is labeled with the class most common among its neighbors. KNN determine neighbors based on distance using Euclidian, Manhattan or Murkowski distance measures for continuous variables and hamming for categorical variables. Calculated distances are used to identify a set of training instances (k) that are the closest to the new point, and assign label from these. Despite its simplicity, KNN have been applied to various types of applications. For churn, KNN is used to analyze if a customer churns or not based on the proximity of his features to the customers in each classes .
- 2) **Ensemble – based Learning:** Ensemble based learning techniques produce their predictions based on a combination of the outputs of multiple classifiers. Ensemble learners include bagging methods (i.e. Random Forest) and boosting methods (i.e. Ada Boost, stochastic gradientboosting).

a) Random Forest

Random forests (RF) are an ensemble learning technique that can support classification and regression. It extends the basic idea of single classification tree by growing many classification trees in the training phase. To classify an instance, each tree in the forest generates its response (vote for a class), the model choses the class that has receive the most votes over all the trees in the forest. One major advantage of RF over traditional decision trees is the protection against overfitting which makes the model able to deliver a high performance.

b) Boosting – based techniques (Ada Boost and Stochastic GradientBoosting)

Both AdaBoost (Adaptive Boost) and Stochastic Gradient Boosting algorithms are ensemble based algorithms that are based on the idea of boosting. They try to convert a set of weak learners into a stronger learner. The idea is that having a weak algorithm will perform better than random guessing. Thus, Weak learner is any algorithm that can perform at least a little better than random solutions. The two algorithms differ in the iterative process during which weak learners are created. Adaboost filters observations, by giving more *weight* to problematic ones or the ones that the weak learner couldn't handle and decrease the correctly predicted ones. The main focus is to develop new weak learns to handle those misclassified observations. After training, weak learners are added to the stronger learner based on their alpha weight (accuracy), the higher alpha weight, the more it contributes to the final learner. The weak learners in AdaBoost are decision trees with a single split and the label assigned to an instance is based on the combination of the output of all weak learners weighted by their accuracy.

On the other hand, gradient bosting gives importance to misclassified/difficult instances using the remaining errors (pseudo-residuals) of the strong learner. At each iteration, errors are computed and a weak learner is adjusted to them. Then, the contribution of the weak learner to the strong one is the minimization of the overall error of the strong learner . For churn prediction Adaboost and Sochastic gradient have been used for churn prediction.

4) **Artificial neural network:** Artificial Neural Networks (ANNs) are machine-learning techniques that are inspired by the biological neural network in human brain. ANNs are adaptive, can learn by example, and are fault tolerant. An ANN is composed of a set of connected nodes (neurons) organized in layers. The input layer communicates with one or more hidden layers, which in turn communicates with the output layer. Layers are connected by weighted links. Those links carry signals between neurons usually in the form of a real number. The output of each neuron is a function of the weighted sum of all its inputs. The weights on connection are adjusted during the learning phase to represent the strengths of connections between nodes. ANN can address complex problems, such as the churn prediction problem. Multilayer perceptron (MLP) is an ANN that consists of at least three layers. Neurons in each layer use supervised learning techniques . In the case of customer churn problem, MLP has proven better performance over LR.

5) **Linear Discriminant Analysis:** Linear Discriminant Analysis (LDA) is a mathematical classification technique that searches for a combination of predictors that can differentiate two targets. LDA is related to regression analysis. They both attempt to express the relationship between one dependent variable and a set of independent variables. However, unlike regression analysis, LDA use continuous independent variables and a categorical dependent variable (target). The output label for an instance is estimated by the probability that inputs belong to each class and the instance

is assigned the class with the highest probability. Probability in this model is calculated based on Bayes Theorem. LDA can be used for dimensionality reduction by determining the set of features that are the most informative. LDA has been used in for different classification tasks including customer churn .

III. METHODOLOGY

The first step before applying the selected analytical models on the dataset, explanatory data analysis for more insights into dataset was performed. Based on the observations data was preprocessed to be more suitable for analysis.

1. Data Collection and its description

Data collection is important step for research in area. It is a process of gathering and measuring information related to our area. If we collect faulty data, it results in difficulty in prediction of customer. So, accurate data collection is important for maintaining integrity of prediction. It enables us to predict customer behavior accurately. We have semi structured data in the form of comma separated file. It does not fit into formal structure of data models associated with relational databases, but It has self-describing structure. Here we have collected Online Shoppers Purchasing Intention dataset from UCI Machine learning repository .

Online Shoppers Purchasing Intention Dataset contains 12330 sessions in the dataset. out of which, 84.5% (10,422) were negative class samples that indicates customer did not Purchase the product and the rest (1908) were positive class samples that end with Purchase of the product. Attribute values of this dataset are integer, real. This dataset is donated on 2018-08-31 and have 18 attributes. Revenue is our target variable. Details of these features is described below:

- a) Administrative: It indicates the number of different types of administrative pages visited by the visitor in that session.
- b) Administrative Duration: Includes total time spent in administrative pages.
- c) Informational: It indicates the number of different types of informational pages visited by the visitor in that session.
- d) Informational Duration: Includes total time spent in informational pages.
- e) Product Related: It indicates the number of different types of product related pages visited by the visitor in that session.
- f) Product Related Duration: Includes total time spent in product related pages.
- g) Bounce Rate: The value of " Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave (" bounce") search without purchase of product.
- h) Exit Rate: The value of " Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.
- i) Page Value: It is the average value for a web page that a user visited before completing an e-commerce transaction.
- j) Special Day: It contains details of the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction
- k) Operating system: Includes details of type of operating system used.
- l) Browser: Includes details of type of browser used.
- m) Region: Includes details of region of customer.
- n) Traffic type: Includes traffic type details.
- o) Visitor type: Indicates type of visitor like returning visitor, new visitor or other.
- p) Month: Month of the year
- q) Weekend: Boolean value indicating whether the date of the visit is weekend.
- r) Revenue: Boolean value indicating whether customer purchase the product.

2. Data Preprocessing

Second most important step of proposed system is data pre-processing. Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and it may contain many errors. So, data pre-processing is a method to resolve such issues. Data pre-processing prepares raw data for further processing. Data goes through a series of pre-processing steps like data cleaning, data integration, data transformation, data reduction, data discretization etc. In proposed model, data cleaning and data transformation is done to prepare data for further processing.

1) Data Cleaning

It is a process of finding and removing corrupt or incorrect data from dataset. Data of proposed system is cleansed by finding missing values and then filling in missing values.

2) Data Transformation

It is a process of converting data from one format to another format. Dataset contains two or more labels in one or more than one column. These labels are in the form of words or numbers. Labels in the form of words or numbers are human readable but not machine readable. So, to make data machine readable label encoding is done. Label encoding is process of converting human readable labels into machine readable form Online shoppers purchasing intention dataset contains some features like Weekend, Revenue which have binary values like True-False. Some features like Month have many labels like Jan, Feb, Mar, Apr etc. So, we converted it to machine readable form with label encoding. For example, True/False converted to 1/0 (Here, label encoding is done according to alphabetical order or numerical order).

3. Feature Selection

Feature selection is alternative name of variable selection or attribute selection. It is the process of selection of subset of attributes in data (such as columns in dataset) that are relevant to the proposed model. Feature selection and dimensionality reduction are different. Both methods reduce the number of attributes in the dataset, but a dimensionality reduction method reduce number of features by creating new combinations of attributes, whereas attribute selection method includes necessary feature and excludes unnecessary features present in the data without changing them.

Dataset contains many features that may be relevant or irrelevant, important or unimportant. Feature selection method helps to remove features that are irrelevant or unimportant for model. It helps to create an accurate predictive model by removing unnecessary overhead of processing of unimportant features. Simpler model is easy to understand and explain. So, Feature selection makes model simpler by selecting fewer attributes which reduce complexity of the model.

i. Correlation

Correlation is process of finding mutual relationship or association among each other. It can be positive or negative. It shows that ProductRelated Duration and ProductRelated are positively correlated to each other. Similarly, BounceRates and Exit Rates are positively correlated to each other.

After Correlation Matrix, feature ranking is done according to feature importance as shown in figure 1. According to feature ranking, we removed less important features and noted results. PageValues (with feature importance score 0.377216) is most important feature among all features. SpecialDay is least important feature with

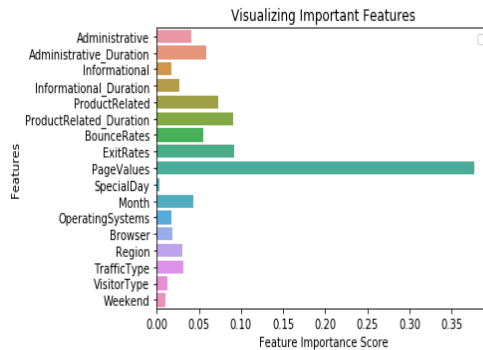


Figure 1: Feature Importance graph for features of dataset

feature importance score 0.003404. Some other less important features that can be removed are Weekend (0.010041), VisitorType (0.012600).

4. Modelling

The last module in proposed system is the prediction algorithm. In this module, we predicted behaviour of customer that whether he will purchase or not purchase. We applied 10 classification algorithms like adaboost classifier, Decision Tree classifier,GBM (Gradient Boosting Machine), KNN (K-Nearest Neighbour), Logistic Regression, Multi-Layered Perceptron Neural Network (MLPNN), Nave Bayes (NB), Random Forest Algorithm (RFA), Support Vector Classification algorithm (SVC), XGB (XGBoost) .Random forest algorithm performed very well in all cases of data splits i.e. 80-20, 70-30, 60-40 splitting. It gave best prediction accuracy even after removal of less important features. So, Random Forest algorithm is used in proposed system.

VisitorType Other (i.e. 1 in figure) didnt end with purchase and Purchase done by New visitor (i.e.Type0 in figure). To improve Purchase, companyneeds to focus on these categories of visitor type.

5. Training and Testing

Finally training of model is done by splitting data in the form of 80% training 20% testing 70% training 30% testing

60% training 40% testing

Accuracy calculated for all 10 folds and mean accuracy is considered as final accuracy

IV. RESULT AND DISCUSSION

The experimentation of proposed model has been done in python programming language with Online Shoppers Purchasing Intention dataset of UCI Machine Learning Repository. Label encoding is done on dataset to make it machine readable. It is the most important preprocessing step in machine learning. Once data becomes ready, 10 different classification algorithms are applied and their accuracy is tested to get better prediction results. These classification algorithms are adaboost classifier, Decision Tree classifier, GBM (Gradient Boosting Machine), KNN (K-Nearest Neighbour), Logistic Regression, Multi-Layered Perceptron Neural Network (MLPNN), Nave Bayes (NB), Random Forest Algorithm (RFA), Support Vector Classification algorithm (SVC), XGB (XGBoost). Accuracy of models during 10 folds with 80% training data and 20% testing data Similarly, Accuracy of models during 10 folds with 70% training data and 30% testing data and accuracy of models during 10 folds with 60% training data and 40% testing data Finally, to compare all classification algorithms and to find out one with better prediction accuracy. We calculated minimum accuracy, maximum accuracy and mean accuracy for all 10 classification algorithms. It is summarized that Random Forest Algorithm performs better in all cases of data split. Feature importance graph shown in Figure 1 states that Page Value is most important feature as well as Special Day and Weekend are less important feature. So, we removed Special Day and Weekend. As a result, unnecessary overhead of system gets reduced and accuracy of system increased to 90.42986861119473.

Confusion Matrix

```
[[3998  157]
 [ 330  447]]
```

Figure2: Confusion Matrix for RFA

Figure 2 shows confusion matrix for random forest algorithm. Confusion matrix is used to describe performance of classification algorithm. In our case, it describes performance of Random Forest Algorithm. First row indicates negative class (Not Purchase) and Second row indicates positive class (Purchase). The value 447 indicates that customer end with Purchase and model predicted it as Purchase. The value 3998 indicates that customer did Not Purchase, and model predicted it as Not Purchase. The value 157 indicates that customer did not purchase but model predicted it as Purchase (FP). The value 330 indicates that customer purchased product, but model predicted it as Not Purchase (FN). Figure 3 shows classification report for Random Forest Algorithm. Classification report for Random Forest Algorithm shows the scores of precision, recall, f1 and support for Random Forest Algorithm. Here 0 indicates value of Not Purchase class and 1 indicates value of

Classification Report

	precision	recall	f1-score	support
0	0.92	0.96	0.94	4155
1	0.74	0.58	0.65	777

Figure3: Classification Report for RFA

Purchase class.

Precision: It is the ability of classifier to label Positive as Positive and Negative as Negative (i.e Negative sample should not be labelled as Positive).

$$\text{Precision} = \frac{TP}{TP+FP}$$

- Recall: It is the ability of finding all positive samples.

$$\text{Recall} = \frac{TP}{TP+FN}$$

- F1-score: It is a mean of precision and recall.

Support: It indicates the number of samples of true responses that is present in class.

V. CONCLUSION

To conclude with the analysis, we have understood that customers purchase chances are more if Bounce Rate is below 0.050 and exit Rate below 0.075. The Chances of product purchase is high if product Related Duration is between 0-30000 seconds and Product Related pages are between 0-300. During data analysis, we observed that customers have preferred Operating System 1,2,3,4 is most frequently used in all region. Browser 2 is used by many customers. Finally, we understood that online purchasing must be emphasized and improved more among New customers (Type 0) and other (Type 1) customers whereas the use of promo codes must be emphasized with both Visitor Types. The future work of this paper can be suggesting promotional tools for improving the sales profit, predicting which products the customer buy most and providing marketing strategies for improving the sales.

REFERENCES

- [1] Dr.K. Maheswari and P.Packia Amutha Priya, ‘ ‘Predicting Customer Behavior in Online Shopping Using SVM Classifier’’, 2017 IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNIQUES IN CONTROL, OPTIMIZATION AND SIGNAL PROCESSING.
- [2] Mahendra Pratap Yadav, Mhd Feeroz and Vinod Kumar Yadav, ‘ ‘Mining the customer behavior using web usage mining in e-commerce’’, ICCNT’12 26th 2Sdl July 2012, Coimbatore, India.
- [3] Farshid Abdi and Shaghayegh Abolmakarem, ‘ ‘Customer Behavior Mining Framework (CBMF) using clustering and classification techniques’’, Journal of Industrial Engineering International. Received: 4 June 2017 / Accepted: 2 August 2018.
- [4] Femina Bahari Tand Sudheep Elayidom M., ‘ ‘An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour’’, International Conference on Information and Communication Technologies (ICICT2014).
- [5] Pooja Sharma, Vidyalakshmi Nair, Amalendu Jyotishi, ‘ ‘Patterns of Online Grocery Shopping in India: An Empirical Study’’, CONIAAC ’14, October 10 - 11 2014, Amritapuri, India.
- [6] Ge Yunshengi, Zhang Qianqian and Kong Jie, ‘ ‘Research on the prediction of user behavior based on neural network’’, ICIIP ’18, May 1920, 2018, Guilin, China.
- [7] Zoltan Balogh, ‘ ‘Analysis of Public Data on Social Networks with IBM Watson’’, Acta Informatica Malaysia, 2(1):10-11(2018).
- [8] Zhang R, ‘ ‘Research on big data feature analysis based on kernel discriminant analysis and neural network [J]’’, Multimedia Tools and Applications, 2018(10).
- [9] <https://stats.stackexchange.com/questions/117654/what-does-the-numbers-in-the-classification-report-of-sklearn-mean>
- [10] <https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>
- [11] Rana Alaa El-Deen Ahmida, M. Elemam Shehaba, Shereen Morsya and Nermeen-Mekawiea, ‘ ‘Performance Study of Classification Algorithms for Consumer Online Shopping Attitudes and Behavior Using Data Mining’’, (CSNT), 2015 Fifth International IEEE Conference on 4-6 April 2015, Electronic ISBN: 978-1-4799-1797-6, Print on Demand (PoD) ISBN: 978-1-4799-1798-3.
- [12] https://scikitlearn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.
- [13] M.N.Saroja, S.Kannan, K.R. Baskaran, ‘ ‘Analysing the Purchase Behavior of a Customer for Improving the Sales of a Product’’, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018.
- [14] Mr. Shrey Harsh Baderiya, Prof. Pramila M. Chawan, ‘ ‘Customer buying Prediction Using Machine-Learning Techniques: A Survey’’, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 10—Oct 2018.
- [15] <https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/>
- [16] <https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>
- [17] Harsh Valecha, Aparna Varma and Ishita Khare, ‘ ‘Prediction of Consumer Behaviour using Random Forest Algorithm’’, 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON).
- [18] Sahar F. Sabbeh, ‘ ‘Machine-Learning Techniques for Customer Retention: A Comparative Study’’, International Journal of Advanced Computer Science and Applications, Vol. 9, No. 2, 2018.



INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details