Prediction of Online Shoppers Purchase Intention based on Web-server Log Data

Iswarya Hariharan
Computing and Software Systems
University of Washington Bothell
Bothell WA USA
iharihar@uw.edu

Dong Si
Computing and Software Systems
University of Washington Bothell
Bothell WA USA
dongsi@uw.edu

ABSTRACT

Online shopping in e-commerce has increased tremendously in the past few years and is very popular in recent times. Even though a potential has been created in the market, the conversion rates are relatively low. This paper proposes a system to predict the purchase intention of customers based on the navigation path in the e-commerce website and their activity during a session. We propose a robust model that deals with the class imbalance and achieves improvement with feature extraction in the pre-training phase. We have used bagging and boosting decision trees based ensemble models, Support vector machine, and Artificial Neural Network along with resampling techniques such as SMOTE, random oversampling, and random undersampling to achieve accurate results and predict the sessions that turn out to be revenue-generating which is the underrepresented class. The resampling techniques have found to increase the F1-score and Recall scores of the ensemble models thereby reducing the False Negatives which is of more importance so that e-commerce companies don't lose potential revenue-generating customer sessions. The classifier models are compared with the appropriate performance metrics for this imbalance situation and application. This can be used by e-commerce websites to predict the likelihood of a customer generating revenue for their website and accordingly provide them with custom alternatives to ensure that they end up buying what they want.

CCS CONCEPTS

• Machine Learning • Class imbalance • Sampling

KEYWORDS

Online purchasing intent, Machine Learning, Ensemble methods, Oversampling, Class imbalanced classification

1. INTRODUCTION

The spike in e-commerce usage in recent times has created potential in the online purchasing market. However, studies have shown that the corresponding conversion rates have not increased at the same rate which opens up the requirement of solutions to provide customized promotions to online shoppers similar to the physical retailing where salespeople suggest customized alternatives to shoppers by gaining experience over their requirements over time [1]. This emphasizes the importance of effective utilization of the customers' usage data within the website to propose solutions involving Machine learning models which could be used for ad-display and retargeting. Behavioral prediction systems imitating the behavior of a salesperson in virtual shopping environments are becoming popular nowadays to understand customer behavior with e-commerce companies investing in them [1]. The main mission of this project is to categorize the visits of a session established by a user based on the navigational pattern if its either revenue generating or not for the website. In this paper, we propose a solution that can be used to categorize the purchasing intent of the customers based on their navigation data within the website. We have used an online retailer's data and compared the performance of different machine learning models under different conditions. Popular models use Logistic regression and decision tree-based machine learning approaches to predict user activity. We have used artificial neural networks and support vector machines that capture the non-linear relationship between the input and the architecture [11]. We also use the Mutual Information based Feature selection procedure as a part of the data preprocessing phase which captures the non-linear relations between the features. Due to the class imbalance faced by this problem, we have used resampling techniques to improve the precision and recall of the models which is of more importance than accuracy to signify the underrepresented class. In the previous work, this dataset was used to predict the purchasing intention and thereby to model real-time behavioral machine learning models which can predict the abandonment measure of the user from the website indicating the likelihood of abandonment and recommend the right products based on this [1]. The authors used the C4.5 decision tree and Multi-Layer perceptron to predict the purchase intention of the dataset. In another paper, the authors have worked on applying association mining on this type of data by estimating the probability of purchase of the customers and extract useful knowledge from their profiles [11]. In another experiment, similar to our experiment, a supervised learning model was used to categorize the user sessions as buyer and browsing sessions with the historical data collected from an online book store [15]. Hidden Markov Model (HMM) was used to predict the actions of customers based on the frequently accessed paths of the visitors in a consecutive pattern and this was suggested for the potential customers [16].

2. DATASET

The dataset is obtained from an online bookstore and contains session information of 12,330 unique users to avoid any specific campaign or specific user profile. The binary classification problem predicts if a user would generate revenue for the company or not. The dataset contains 18 attributes that comprise of 10 numerical attributes and 8 categorical attributes. The first six numerical attributes correspond to the number of pages and duration spent in every type of page, 'Administrative', 'Informational', 'Product related'. They are obtained from the URL information from the web pages visited by the users. The other numerical attributes include 'Bounce rate', 'Exit Rate', and 'Page value' measured by the Google Analytics for the pages on the website which can be obtained when the website is registered by the company in Google analytics [13]. These metrics are stored in the database of the e-commerce and can be updated at regular intervals. The 'Special Day' feature is a numerical attribute that indicates the closeness of the site to a special day such as 'Mother's Day' during the time of the visit. The dynamics of the e-commerce website play a role to determine this value like the duration between the order and delivery date close to a special day. The categorical attributes include categories for each of the Operating system used by the user, region of origin of the request, Browser used by the user, traffic source of the origin of visit eg. SMS, Month of the visit, Binary value denoting if the day of the visit is a weekend and the type of user indicating if they are a new user, returning user, or others. The class label is the 'Revenue' feature that denotes a Boolean value of purchasing intent.

2.1 Dataset Class Imbalance

The dataset contains 10,422 negative class samples (84.5%) and 1908 positive class samples (15.5%) which denotes the presence of class imbalance problem with the skewed data distribution (Figure 1).

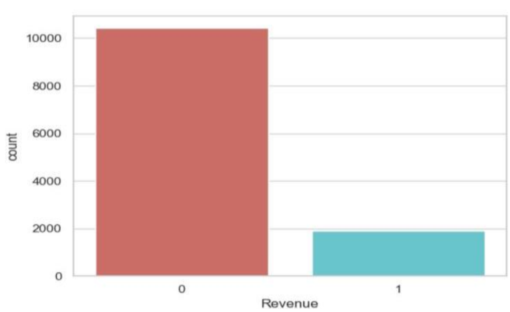

Figure 1: Dataset class imbalance problem depicted with the total number of samples based on the Revenue class label – if the user would generate revenue (value:1) or not (value: 0)

2.1.1 Handling class imbalance

Many methods can be used to handle this class imbalance problem. In this project, we will be using a hybrid approach with resampling techniques and cost-sensitive approaches on possible models to tackle this problem with this dataset. The resampling techniques that will be used include random oversampling, random undersampling, synthetic sampling with data-generation using SMOTE (Synthetic Minority Oversampling Technique) along with no resampling methods used to compare the performance of these methods. SMOTE uses the KNN algorithm to create random synthetic data points along the line segments to the k-nearest neighbors [14]. When using the SMOTE algorithm for synthetic data generation, the dataset is shuffled and split into training and validation sets, and then SMOTE is applied to the training dataset. This ensures that there is no information leakage from the validation set to the training set as opposed to when data is split after SMOTE is applied as the algorithm generates points based on K-NN theory [6]. The performance metric used here will not include accuracy as it is misleading due to the accuracy paradox with imbalanced classes [5]. Metrics such as Precision(exactness), Recall(completeness), F-score will be used. Each of the classifier models will be evaluated using ROC curves to assess the classifier's performance using the AUC (Area under the curve) along with AUC for Precision-Recall curves.

3. METHOD/APPROACH

The high-level representation of the workflow of the experiment to be performed is shown in Figure 2.

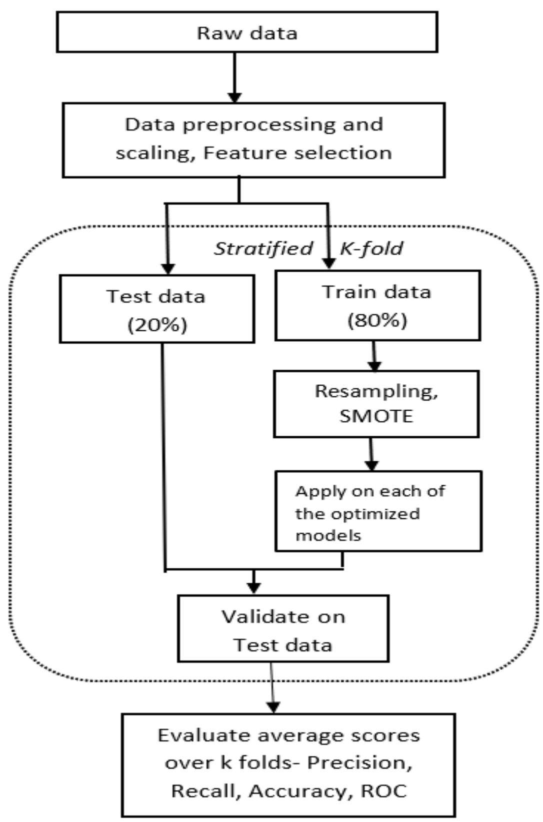

Figure 2: Flowchart depicting the proposed design of the model for solving the problem in this project

3.1 Data Preprocessing

The dataset is obtained from the public UC Irvine Machine Learning Repository [12]. The dataset is a .csv file that contains the 12,330 data objects. The data was checked for null values, missing values using the python package ‘pandas’. The categorical attributes such as ‘VisitorType’ and ‘Month’ are converted to discrete numerical values with ordinal encoding. The binary-valued attributes, ‘Weekend’ and ‘Revenue’ are encoded using label encoding.

3.2 Feature Selection

The attributes were analyzed initially individually to determine if feature selection techniques improve the classification performance of the system. As a first step, analyzing the categorical attributes manually brought out the understanding that ‘Month’, ‘VisitorType’, ‘TrafficType’, ‘browser’ influence the classification of the visit. This is is clearly in Figure 3 with bar charts plotted against the ‘Revenue’ class label categories.

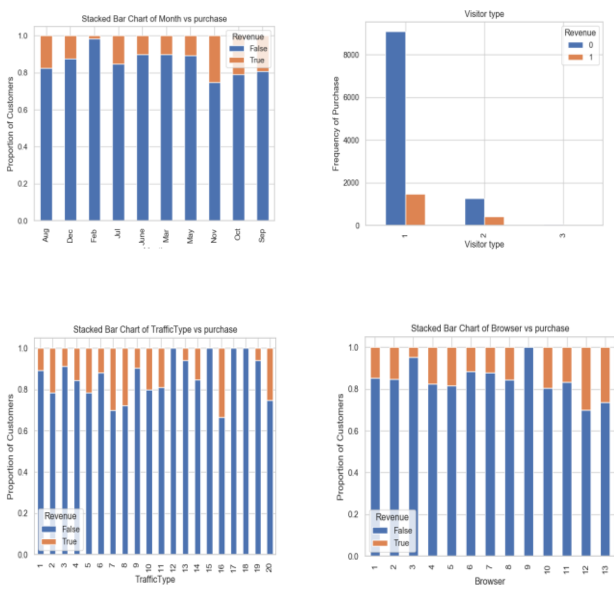

Figure 3: Bar charts of categorical features ‘Month’, ‘VisitorType’, ‘TrafficType’, Browser’ against revenue type

This derives to the conclusion that only some of the categorical attributes contribute to the classification accuracy more than the others as they each category fluctuates with their proportion for the revenue class label. Using a feature selection procedure that considers the mutual dependence or correlation of these input attributes individually to the class label is required [3]. The correlation matrix yields the linear dependence of features with the target class variable shown in Figure 4. However, there is no clear strong correlation of any feature with the target class variable.

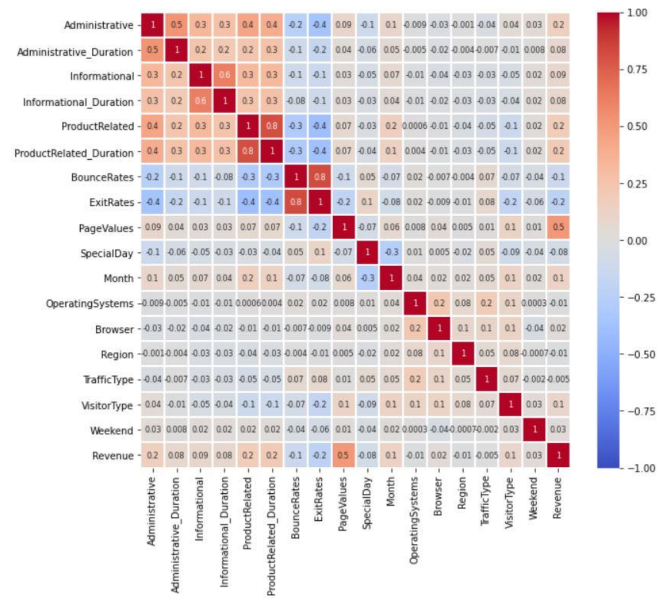

Figure 4: Correlation matrix showing the level of correlation of every feature with the other.

Mutual Information is the measure of mutual dependence between two variables that is based on entropy [1]. The entropy of a variable x with probability distribution *p*(X=*x)* or *p(x)* is given by,

$$H(X) = \sum_{x} [p(x)\log (p(x))]$$

Mutual Information is the mutual dependence between two variables X and Y given by,

$$I(X,Y) = H(X) - H(X|Y)$$

The popular feature reduction method, Principal Component Analysis is another alternative here, however, it would be a linear combination of all the attributes in the dataset. Hence, we are considering the feature selection method that uses Mutual Information (MI) to reduce the features based on its influence on the classification as it also considers the nonlinear relationship between the feature and the target variable [3,9]. Each of these methods is applied to the dataset to determine the score of the features. By sorting these features in descending order based on their scores serve as the selected feature subset for applying on the models. As Mutual Information works on categorical attributes, the continuous numerical attributes are converted to nine discrete levels using the adaptive binning method based on normalization [2]. This method is adopted after an analysis of the continuous attributes that they are skewed after applying the log transformation. The mean μ and standard deviation σ of each of these attributes are obtained and four intervals of size σ to the left of μ - $\sigma/2$ and four intervals to the right of $\mu + \sigma/2$ are discretized between values 1 to 9 with the middle interval being value 4. Figure 5 shows the values of the scores of Mutual Information. The top ten features with the highest MI scores are selected as the feature subset for each of the models.

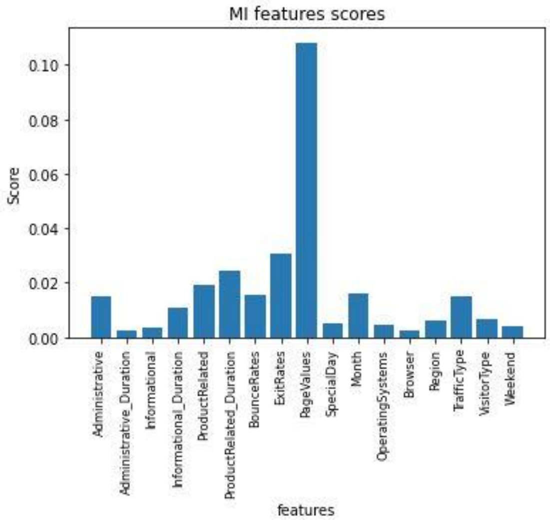

Figure 5: Results of feature scoring of Mutual Information (MI) of all features with respect to the target class feature.

3.3 Experiment Approach

The preprocessed data is trained on the machine learning models using five-fold validation where the training data and test data are 80% and 20% respectively. Five-fold cross-validation is used to avoid overfitting. Resampling methods such as random oversampling, random undersampling, SMOTE are applied followed by training the models with them. This is also done using imbalanced data not preprocessed with feature selection using Mutual information and resampling for comparing the results with the usage of resampling techniques. Each of the models used for k-fold cross-validation is optimized by tuning the hyper-parameters using Grid search. The models used are ensemble models based on boosting and bagging decision trees, voting ensemble classifier using these ensemble algorithms, Support vector machine(SVM) with a polynomial kernel of degree two and an Artificial Neural Network(ANN) with two hidden layers comprised of 30 and 10 neurons in the first and second hidden layers respectively. Support vector machines (SVM) finds the boundary to discriminate classes and will be used in this project. As SVM with non-linear kernels and Neural Networks capture non-linear relationships between the feature and target variables, we have used them in this experiment [11]. SVM has shown the ability to classify datasets in many literature studies and will be included in the analysis [10]. Ensemble models with decision trees based on bagging are Random forest, Bagging decision trees, and based on Boosting are Adaboost and Gradient boosting classifier. As decision trees are prone to overfitting, the number of trees used in these ensemble models has been hyper tuned and the best number avoiding overfitting is obtained using Grid search for each of the ensemble models. We have used the Random forest classifier as it has proven to exhibit a performance improvement over single tree classifier CART, C4.5 thereby yielding a generalized error rate comparable to Adaboost and is more robust to noise [8]. The test data is the validation dataset used in each of the k-fold validation and the classification metrics are noted for all the folds for each of the resampling techniques applied to the training data. Every resampling technique used in a model is evaluated based on the average of the scores from the five folds of the cross-validation on the dataset.

4. EXPERIMENT RESULTS

The models are evaluated based on the details of the confusion matrix as shown in Figure 6. The experiment aims to increase the True Positives (TP) as the 'True' revenue class label is the under-represented one. Resampling methods have been used to increase this number. This also has to ensure that resampling does not reduce the number of True Negatives(TN) as e-commerce companies have to be accurate with identifying the ones that would generate revenue and those who don't. Hence, accuracy has to be an evaluation metric in this problem. Area under Curve(AUC) of Receiver Operating Characteristic(ROC) curves plotted with True Positive rate against False Positives rate shows the increase in True positive rate through the validation of the model. The presence of an imbalance in the dataset leads to a necessity for the model to be designed with lesser False Negatives which requires higher Recall value without affecting the Precision [7]. In the real-case scenario, e-commerce companies don't want models to classify actual revenue-generating sessions to be classified as non-revenue generating as they would lose an opportunity to target those customers and earn revenue with advertising. Hence, the cost of a False Negative(FN) is more than the cost of a False Positive(FP). Thus, Area Under the Curve (AUC) value for the Precision-Recall curve gives a better idea of False negatives for a particular model than AUC for ROC curves.

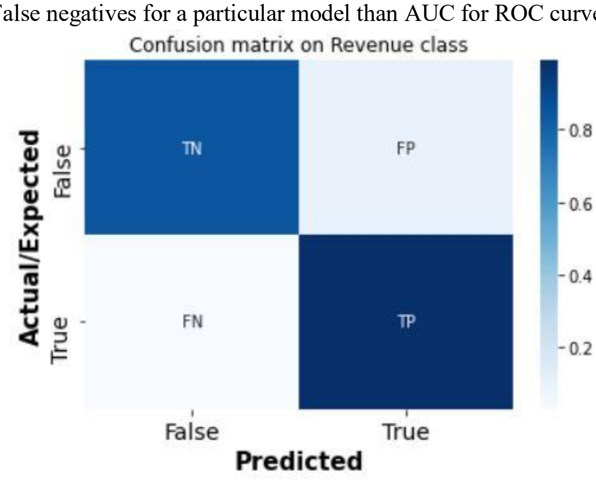

Figure 6: Confusion matrix representation. TP-True Positive, TN- True Negative, FP-False Positive, FN-False Negative

4.1 Area Under Curves (AUC)

Each of the six models with every resampling technique SMOTE, random Undersampling, Random oversampling, and no resampling applied on non-feature selected data is trained on the data in five-folds and evaluated on the validation set. The final evaluated score is the average of the five folds and this is used for comparison of the different models. Figure 7 shows the ROC AUC values of each of the models using different resampling techniques. This graph shows that the AUC values are more than 0.8 for all the models and SMOTE resampling has performed quite well in all the models. It has performed the best in the Random forest ensemble classifier. The false-positive and true positive AUC of the resampling methods have been very close to each other.

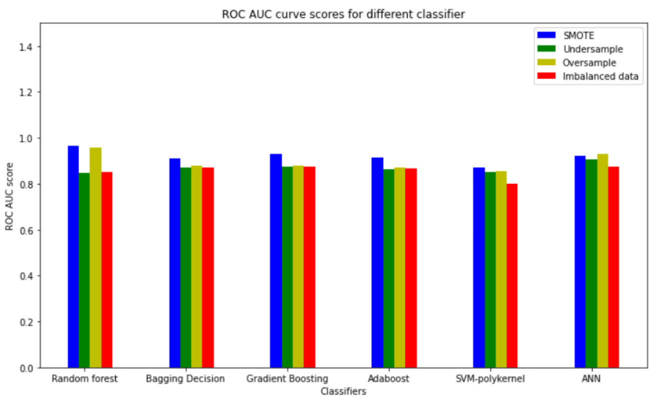

Figure 7: ROC AUC (Area Under Curve) values for all the models used for different resampling techniques.

The ideal scenario to have a high value of True positive and low False Negatives is observed with the Precision-Recall curve AUC values. This value for every model using the different resampling techniques is shown in Figure 8. Using this we can observe that the resampling techniques have shown to improve the precision and recall over the imbalanced dataset. Even though random oversampling and undersampling in all the models have pretty close values, SMOTE resampling has shown to have an edge over all of them except in ANN where all the resampling techniques are very close to each other. The random forest has once again shown the highest AUC values for SMOTE and oversampling.

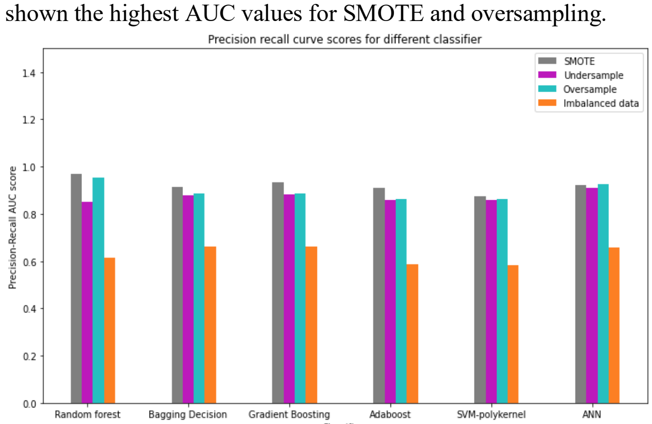

Figure 8: Figure 7: Precision-Recall curve AUC values for all the models used with different every resampling techniques.

4.2 Models and Evaluation

All the different classification metrics scores such as Accuracy, Precision, Recall, and F1-score are listed down for every model with different resampling techniques used in this experiment. The ensemble models involving bagging and boosting decision trees use 40 decision trees as the number of tree estimators which is obtained by hyper tuning. The voting ensemble classifier uses the majority votes of the four ensemble models for classification. The Support Vector Machine(SVM) with a polynomial kernel of degree two is hyper tuned with regularization parameter as 1000 and kernel coefficient to handle non-linear classification, 'gamma' set to 1.0. The Artificial Neural Network was built using Keras sequential classifier model with two hidden layers having 30 neurons and 10 neurons in the first hidden layer and second hidden layer respectively. Sigmoid and Relu activation functions were used along with Stochastic Gradient Descent as the optimizer. The hyper tuning of the parameters of all the models was done using Grid search on these models. The results of the different models on the resampled and preprocessed dataset with five-fold cross-validation are listed in Table 1.

Table 1: Results of models used for different resampling techniques. Average scores of Accuracy, F1-score, Precision, Recall are listed over five folds of the dataset.

Models	Resampling	Avg Accuracy	Avg. F1 score	Avg. Precision	Avg. Recall
Random Forest	SMOTE	0.92	0.92	0.93	0.91
	Undersample	0.76	0.76	0.77	0.75
	Oversampling	0.92	0.91	0.89	0.93
	None	0.88	0.56	0.68	0.47
Bagged Decision Tree	SMOTE	0.83	0.81	0.89	0.74
	Undersample	0.77	0.75	0.85	0.67
	Oversampling	0.79	0.76	0.87	0.68
	None	0.89**	0.58	0.71	0.49
Gradient Boosting	SMOTE	0.87	0.87	0.91	0.82
	Undersample	0.79	0.77	0.82	0.73
	Oversampling	0.79	0.77	0.85	0.70
	None	0.89**	0.58	0.72	0.48
Adaboost	SMOTE	0.86	0.85	0.87	0.82
	Undersample	0.78	0.78	0.80	0.75
	Oversampling	0.77	0.77	0.80	0.73
	None	0.88**	0.54	0.69	0.45
Voting ensemble	SMOTE	0.90	0.90	0.94	0.85
	Undersample	0.78	0.76	0.83	0.70
	Oversampling	0.83	0.82	0.90	0.74
	None	0.89**	0.56	0.75	0.44

Models	Resampling	Avg Accuracy	Avg. F1 score	Avg. Precision	Avg. Recall
SVM – poly kernel	SMOTE	0.78	0.75	0.89	0.65
	Undersample	0.76	0.73	0.86	0.64
	Oversampling	0.77	0.73	0.88	0.62
	None	0.88**	0.54	0.76	0.41
ANN	SMOTE	0.83	0.83	0.84	0.83
	Undersample	0.80	0.80	0.80	0.80
	Oversampling	0.83	0.84	0.84	0.83
	None	0.89**	0.61	0.71	0.53

**Accuracy paradox noticed

We can see that the resampling techniques applied to the dataset have shown to improve the precision and recall in all the models used. SMOTE has the highest F1-score of 0.92 in the Random forest classifier closely followed by random oversampling in the same dataset. We see the boosting based ensemble models have shown consistent F1-scores for all the resampling techniques. Even though random Undersampling and oversampling are pretty close in all these models, SMOTE has shown an edge over all of them. In terms of accuracy, we see that the accuracy of the imbalanced dataset without any resampling is high in all the models which prove the accuracy paradox as their F1-scores are very low in all the models. This proves that precision-recall curves and f1-scores would give a better idea of the model evaluation in this problem dataset. The voting ensemble classifier used here takes the majority votes of the ensemble models used to predict the class of the data object. SVM and ANN in this experiment have relatively good F1-scores for the resampling methods, however lower than the ensemble models used.

4.3 Previous work

Previously this dataset was used to predict if the user sessions would generate revenue or not and, those which are capable of generating revenue were used to predict if they would abandon the website without any purchase being made by taking into consideration their pageview details [1]. The former part of their work was handled by taking the top twelve correlated features with the target variable along with Maximum Relevance and Minimum redundancy feature selection method (mRmR) and followed by feature elimination using wrapper methods to obtain the feature subset. They handled the class imbalance using oversampling. They tested on C4.5 decision trees and Multi-Layer Perceptron (MLP) having ten, twenty, and forty neurons with one hidden layer. The MLP achieved a better F1-score of 0.86 and an accuracy of 0.87 than I could achieve with a sequential ANN in my experiment. However, their results from the C4.5 decision tree with F1-score of 0.82 and Accuracy of 0.8234 was lesser compared to the ensemble models used in our experiment with different resampling techniques. The previous work aimed to implement this in real-time to improve the conversion rates by offering alternate options to users similar t the ones that they intend to purchase. Hence, the abandonment prediction was estimated in real-time using Long short term Recurrent Neural Networks(LSTM RNN). In our experiment, we have used a normalization based adaptive binning approach for converting continuous numerical features to discrete ones and Mutual Information for data preprocessing followed by ensemble models, SVM and ANN with different resampling techniques to prove the effectiveness of the models. This is the main idea of this experiment which focuses on the resampling techniques using ensemble models along with Mutual information and adaptive binning based preprocessing.

4. FUTURE WORK

The dataset imbalance will be handled by using a hybrid approach which includes sampling methods along with a cost-sensitive based approach [4]. This imbalance leads to a necessity for the model to be designed with lesser False Negatives which requires higher Recall value without affecting the Precision [7]. F-measure that is a weighted average of precision and Recall will be used for performance measure. ROC curves will also be used to estimate the area under the curve for the classifiers and evaluate their performance. We plan to use a model with Random forest is proven to exhibit a performance improvement over single tree classifier CART, C4.5 thereby yielding a generalized error rate comparable to Adaboost and is more robust to noise [8]. Support vector machines (SVM) finds the boundary to discriminate classes and will be used in this project. Cost parameters will be updated in this model to achieve better performance. SVM has shown the ability to classify datasets in many literature studies and will be included in the analysis [10]. An ensemble model with Adaboost will be used to improve the performance of the decision of classification. We plan to use Artificial neural networks and evaluate the performance as it captures the non-linear relationship between the input and the architecture [11]. These models will be evaluated using the cross-validation approach so that the models do not overfit. We plan to use a k-fold validation approach to tackle this problem.

5. CONCLUSION AND DISCUSSION

We see that the resampling techniques applied on this dataset along with data preprocessing and feature subset selection using Mutual Information have shown to give better results for precision, recall, and F1-scores than the imbalanced dataset directly being used. Unlike the previous work, which used the top twelve correlated features from the correlation matrix followed by the wrapper method for feature subset selection, Mutual information-based feature selection captures the non-linear relationship between the feature and the target variable. The usage of ensemble models in our experiment has shown to have the highest scores of 0.92 Accuracy and F1-score with SMOTE resampling in the Random forest classifier model which is higher than the scores achieved in the previous work. We can say that ensemble classifiers have done pretty well with the resampling techniques in this dataset. Even though there is no clear winner among the resampling techniques in all the models, we can say that SMOTE has performed pretty well having a close edge with random oversampling technique in all the models. Even though SVM and ANN have been used, their scores weren't any better than the ensemble models used along with the previous work MLP with ten neurons doing pretty better than them. Future work could include the stacking of ensemble classifiers with giving weighted importance to the different models used. This will help in achieving better F1 scores as every model would be better with one class label than the other. Hence, understanding their distribution in the confusion matrix and giving them weighted importance improves the stacking ensemble classifier performance. A challenge faced here is that when e-commerce companies invest in the user-behavioral analysis they aim for real-time prediction. As most of the online searches for shopping end up being non-revenue generating, it is quite difficult to handle this dataset in realtime without resampling. Other types of resampling using Condensed nearest neighbor based Undersampling, ADASYN synthetic resampling which adds random small values to the points after SMOTE to make it more realistic can be researched upon. Feature selection methods using the wrapper and other filter methods can be researched. Moreover, the training time of SVM and ANN is too long which may not be the best option for real-time prediction. Hence, ensemble models would serve better in this situation if used in real-time. This experiment can be extended to research on the analysis of using the page view related data of every person classified as 'revenue-generating' and identify if they can be made provided with alternate options to ensure that they purchase based on what they are looking for. The models used in this experiment can be integrated with cost-sensitive approaches in the classifier models using weights to improve their performance [4]. This system could be used in real-time by e-commerce websites in the future to predict the likelihood of the customer abandoning their website and accordingly provide them with custom alternatives by adapting to their system.

REFERENCES

[1] Sakar, C., Polat, S., Katircioglu, M. and Kastro, Y. 2018. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications*. 31, 10 (2018), 6893-6908.

[2] Hanchuan Peng, Fuhui Long and Ding, C. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 27, 8 (2005), 1226-1238.

[3] Sakar, C., Kursun, O. and Gurgen, F. 2012. A feature selection method based on kernel canonical correlation analysis and the minimum Redundancy–Maximum Relevance filter method. *Expert Systems with Applications*. 39, 3 (2012), 3432-3437.

[4] Haibo He and Garcia, E. 2009. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*. 21, 9 (2009), 1263-1284.

[5] Akosa, J. 2017. Predictive Accuracy: A Misleading Performance Measure for Highly Imbalanced Data. *Proceedings of the SAS Global Forum (2017).* (2017), 942.

[6] The Right Way to Oversample in Predictive Modeling: 2020. *https://beckernick.github.io/oversampling-modeling/.* Accessed: 2020-05- 03.

[7] Jeni, L., Cohn, J. and De La Torre, F. 2013. Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. 2013 *Humaine Association Conference on Affective Computing and Intelligent Interaction.* (2013).

[8] Chen, C., Liaw, A. and Breiman, L. 2004. Using Random Forest to Learn Imbalanced Data. *CA: University of California, Berkeley.* (2004), Technical Report 666.[9] Guyon, I. and Elisseeff, A. 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res*. 3, 1157—1182.

[10] Salcedo-Sanz, S., Rojo-Álvarez, J., Martínez-Ramón, M. and Camps-Valls, G. 2014. Support vector machines in engineering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 4, 3 (2014), 234-267.

[11] Vieira, A. 2015. Predicting online user behaviour using deep learning algorithms. *CoRR*. 1511, 06247 (2015), 1-21.

[12] UCI Machine Learning Repository: Online Shoppers Purchasing Intention Dataset Data Set: 2020. *http://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset*. Accessed: 2020- 05- 04.

[13] Clifton, B. 2012. *Advanced Web metrics with Google Analytics*. Wiley.

[14] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16 (January 2002), 321–357. DOI:http://dx.doi.org/10.1613/jair.953

[15] Suchacka G, Skolimowska-Kulig M, Potempa 2015. Classification Of E-Customer Sessions Based On Support Vector Machine. *ECMS 2015 Proceedings edited by: Valeri M. Mladenov, Petia Georgieva, Grisha Spasov, Galidiya Petrova* (2015). DOI:http://dx.doi.org/10.7148/2015-0594

[16] Yeung WL. 2016. A review of data mining techniques for research in online shopping behaviour through frequent navigation paths. *HKIBS working paper series 075-1516.* Retrieved June 8, 2020 from Lingnan University website: https://commons.ln.edu.hk/hkibswp/76/