Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

A Model to Detect and Categorize Imbalance Darknet Traffic in Cyberspace

[1]Adepegba S.O., [1]Adeyemo A.B., [1]Adeniji D.O. & [2]Adepegba O.A[2],
[1]Department of Computer Science, University of Ibadan, Ibadan, Nigeria.
[2] Cybersecurity Programme Bowen University, PMB 284, Iwo Nigeria
Corresponding E-mail: adepegba.solomon@dlc.ui.edu.ng

ABSTRACT

Data imbalance issues hinder classification algorithms performance. When the real world data set is substantially uneven, the performance of predictive models is severely biased. Datasets are used by classification algorithms to generate the predictions needed for decision-making in a variety of applications, including health, information management, in the cyberspace, business planning, and others. Applications for cybersecurity that classify darknet traffic, detect fraud, and stop phishing attacks all naturally produce highly unbalanced data. The majority of these regions have an uneven spread of datasets commonly known as imbalanced data. When used for data classification, such datasets typically produce poor prediction accuracy. Hence, this research; A model to detect and categorize imbalance darknet traffic in cyberspace suggests using SMOTE-ENN (Synthetic Oversampling Technique Edited Nearest Neighbour) resample technique with an ensemble approach to tackle and address the problem of darknet traffic imbalance dataset. The developed model was evaluated by comparing performance of the base learners (classifier) with prediction accuracy of the ensemble learner (classifiers). The performance evaluation for the base classifiers (Support Vector Machine, Bagging and Adaboost) and the Ensemble classifier was based on performance metrics such as accuracy, precision, recall, F1-scorre, Optimized precision, and Balancer Accuracy Score. The results showed that the datasets generated by the developed model yielded better balanced accuracy score values. The implication of the better performances of the ensemble Learner model viz-a-viz those of the Adaboost across all the standard evaluation metrics is that the SMOTE-ENN ensemble Learner model has the potential of efficiently detecting darknet as-good-as or better than any single model that have been investigated in this study. Therefore, the developed model is justified for data resampling.

Keywords: Darknet, SMOTE-ENN, Ensemble Learner, Adaboost, Support Vector Machine (SVM), Bagging.

Adepegba S.O., Adeyemo A.B., Adeniji D.O. & Adepegba O.A. (2023): A Model to Detect and Categorize Imbalance Darknet Traffic in Cyberspace. Journal of Advances in Mathematical & Computational Science. Vol. 10, No. 4.
Pp 36-56. dx.doi.org/10.22624/AIMS/MATHS/V11N3P2 Available online at www.isteams.net/mathematics-computationaljournal.

1. INTRODUCTION

The Internet is one of the greatest inventions in human history. Its development still continues today. According to a study on digital life, as of January 2021, the number of people using the internet in the world has increased by 7.3 percent compared to the previous year and has been determined as 4.66 billion (Habibi et al., 2020). There are millions of web pages, databases and servers running constantly on the internet network. Websites that can be found on this network using known search engines are called the surface network. The network of hidden websites that could be accessed with a special web browser such as the Darknet, often known as Tor, is used to protect users' privacy and anonymity online (Sui et al., 2015).

Keyloggers, botnets, and ransomware are all examples of malware that may be found on the Darknet. Darknet traffic analysis aids in early malware detection before to an attack and the detection of harmful activities following an epidemic (Kaur & Randhawa, 2020). Classification of web traffic flows is used effectively in areas such as network monitoring, service quality, intrusion detection and network security. Network flows created by packets that have the same source IP, port, destination IP, port, TCP, and UDP information are determined by which application they belong (Cao et al., 2014).

The World Wide Web, also referred to as web, and the Internet are familiar to the majority of people. Due to the use of the search engine in Domain Name System (DNS) indexing through globally distributed Internet Protocol (IP) networks, we commonly access both via web browsers or other networked apps to exchange information in an open manner. This openly searchable and indexed address space is referred to as the "surface web" or "clearnet." The deep web, in contrast, is the area of the WWW address space that search engines do not index but that is nevertheless accessible to the general public. networks built up of personal deep web channels or The World Wide Web, also referred to as the web, and the Internet are familiar to the majority of people. Regular access to using both web browsers or other networked applications to disseminate publicly accessible information is aided by search engine indexing of the Domain Name System (DNS) through internationally spanned Internet Protocol (IP) networks. However, the surface web or clearnet is then openly searchable and indexable address space. In contrast, the deep web is the portion of the internet that search engines do not index but that is nevertheless publicly accessible. The term "darknets" refers to both private networks found in the deep web and networks made up of unallocated address space (Xing et al., 2020).

An overlay network that uses specialized software, user permission, or unconventional communication protocols can access the dark web. Numerous darknets give users the option of remaining anonymous while communicating, which facilitates a wide range of illicit activities, such as hacking, media piracy, terrorism, the trade of illegal products, human trafficking, and the production of child pornography (Aras et al., 2020). To more effectively identify and stop these criminal activities, researchers use machine learning and deep learning techniques to illuminate the darknet traffic. The CICDarknet dataset appears to be quite unbalanced, which may cause the results of their prediction performance to be inaccurate. This research strives to contribute by promoting accurate classification of traffic features from the well-studied CIC-Darknet2020 dataset (Muhammad et al., 2021). CICDarknet2020 is a collection of traffic features from two darknets, namely The Onion Router (Tor) and a Virtual Private Network (VPN), and also equivalent traffic generated over clearnet sessions using the same applications.

Hence this study aims to develop a SMOTE-ENN Enhanced Ensemble learning for the Darknet Traffic Detection. SMOTE- ENN is the proposed resampling technique with ensemble approach using majority voting method on the predicted values of the base models in order to achieve a more efficient predictive performance in order to give a better result.

2. RELATED WORKS

Li (2017), work on an Optimal Task Dispatching on Multiple Heterogeneous Multi-server Systems couple with Dynamic Speed and Power Management constructed task model, resource model, and analyse tasks' preference to reduce tasks' waiting time, the short coming of this study was that the optimization of cloud topology needed to decentralize load balancing for big data cloud was not considered.

According to Akinboro & Ayobami (2019), the study; Model for a Self-Adaptive Routing Optimization in Mobile Ad-hoc Network, the authors ssuggested a model that divides nodes into partitions and calculates the local best for each partition, which then communicates with one another to generate the global best based on computational time. The limitation with this study Optimization of throughput and network topology were not taken into account.

Fahrettin & Ahmed (2022) developed a Gradient Boosting model that detect darknet traffic and attacks in a dataset, Gradient Boosting Algorithm, the developed model was able to achieve the aim of the study, but the work was unable to put the issue of imbalance into consideration. Konstantinos Demertzis et al., (2021), developed a model to generalize new, unknown data sets that reduce the prediction error in data samples with unknown data set using Weight Agnostic Neural Networks Framework, the model performs as expected but the shortcoming of the study was the problem of Interpretability Optimization of Global to understands how the model makes decision which was not considered.

Zakartye & Jamaluddin (2020) worked on Application and Interpretation of Ensemble Methods for Darknet Traffic Classification, the developed model was designed for traffic tasks classification. The limitation with this study was the issue of balancing data and the problem of optimization of real life traffic whereby data can be mined from public computers which was not considered in this study.
Claude & Mourad (2018), worked on Darknet as the Source of Cyber Intelligence: Survey, Taxonomy and Characterization, the work was a survey on the previous work done on the same sample dataset but it was discovered that the Deployment of IPv6 and VoIP mobile- based darknet traffic analysis is not implemented in their study.

Muhammad and Muhammad (2021), developed a model called DarkDetect, the Darknet Traffic Detection and Categorization model was developed by adopting Modified Convolution-Long Short-Term Memory, Combination of classifiers in an ensemble approach for creating personalized generated dataset is not implemented and the challenges faced by the study was the issue of data imbalance.
Nhien & Mark (2022), developed a Darknet Traffic Classification and Adversarial Attacks Machine-Learning-Based Darknet Traffic Detection System for IoT Applications. The methodology was the use of a discriminator model of AC-GAN to classify darknet network traffic, the limitation was that the data Argumentation purposes was not implemented.

Qasem & Moez, (2022), developed model designed to composed contemporary actual IoT communication traffic, Optimization of Tor and VPN activities in communication networks is not implemented. Taking cognizance of reviewed literature, the following were observed: i. The issue of data imbalance was a major challenge ; ii. That the larger dataset would yield high evaluation values than a smaller dataset. at the same time, a data sparsity problem, in which specific points of data are missing and negatively impact the performance of the system, is also claimed to be possible when working with a big amount of data.

Thus, this study “SMOTE-ENN, which combines the capabilities of SMOTE and ENN to eliminate certain observations from both classes that are determined to have a different class from the observation's class along with its K-nearest neighbor the majority class, addresses these challenges. When the class of an observation and the majority class from the observation's K-nearest neighbor are different, the ENN removes both the observation and that of its K-nearest neighbor rather than only the observation and its 1-nearest neighbor who have different classes.

ENN should produce more thorough data cleansing as a result than just utilizing the conventional technique. The Majority Voting Another methodology for handling unbalanced data sets is Ensemble Learner, which combines the output or performance of several classifiers to enhance the performance of a single classifier. By combining different classifiers, this technique alters the generalization capacity of individual classifiers. It primarily integrates the numerous base learners, learns them in parallel, and combines them by training a meta-model to produce a prediction based on the predictions of the several base models.

3. METHODOLOGY

3.1 An Overview

The term "darknet" is frequently used to describe portions of the internet that are not indexed by search engines and demand particular software or authorization to access. As a result, no single "darknet dataset" can be said to include all darknet operations. There are, however, databases that concentrate on particular facets of the darknet, such as illegal marketplaces, online criminal activity, or anonymous messaging services. To better analyze criminal behavior, spot trends, and improve online security, these datasets are frequently gathered by research groups, law enforcement organizations, or cybersecurity companies.

3.2 A SMOTE-ENN Model Process

i. Darknet Dataset collection: The real dataset selected for this experiment is from Kaggle.com. This dataset was compiled to evaluate cutting-edge approaches to categorizing darknet traffic. Dataset compiled by the University of New Brunswick's Canadian Institute for Cybersecurity.

ii. Pre-processing: The dataset passed through the stage of pre-processing to meet the standard require by the pre-trained model such as reduction, normalization, standardization etc.

iii. Input: The dataset was inserted and fed in, then it goes into the algorithm.

iv. Detection: The darknet is detected and the dataset were then classified.

v. Feature Extraction: The pre-trained model extracts relevant features using the K nearest neighbor approach for recognition purpose.

vi. Recognition: The Machine Learning models will be sensitive to detection through feature selection and store the relevant feature output. The classifier checks the similarities considering close values between the data, the stored record and classifies.

i. Validation: This shows if the darknet dataset is correctly classified or not.

3.2.1 Research Approach

The approaches used in accomplishing the objectives of the research are as follows:

i. Data acquisition, description, extraction and classification were performed.

ii. Formulation of mathematical models for the algorithms

iii. A frame work for the SMOTE-ENN Algorithm for data resampling would be designed.

iv. An Ensemble Learner Model based on Majority Voting Technique was formulated.

v. Simulation of the SMOTE-ENN and the formulation of Majority Voting based Ensemble Learner Model would be performed using MATLAB R2020 software package.

vi. Evaluation of the Performance of SMOTE-ENN is performed by comparing the predictions of the classifiers: Adaboost, SVM, Bagging, and Enhanced Ensemble Learner with the balanced data using accuracy, precision, F1-score, Matthew Correlation Coefficient and balanced accuracy.

3.3.1 UML Diagram of the Proposed Modeling

The Ensemble Model approach is designed and represented using the Unified Modelling Language (UML) diagram paradigm. The UML diagram was used to depict the idea of the Ensemble Learner Model is the UML Activity diagram. The model design given in this thesis is depicted in Figure 3.3, which depicts the Unified Modelling Language Activity diagram. The figure's activity diagram serves as a model for the Ensemble Learner Model approach's activities. In essence, it is a flowchart that illustrates the movement of controls of one task to another. An operation on a few classes in a system that modifies the system's state is typically represented by an activity diagram. The activity diagram shows two (2) swim lanes;

Data Preparation swim-lane: This swim-lane depicts the activities that is related to the preprocessing of the CIC Darknet dataset including handling of class imbalance before the dataset will be presented to classification learning algorithms.

a. Model Training and Testing swim-lane: this swim-lane relates to the classification learning activities carried out to build the Ensemble Model.

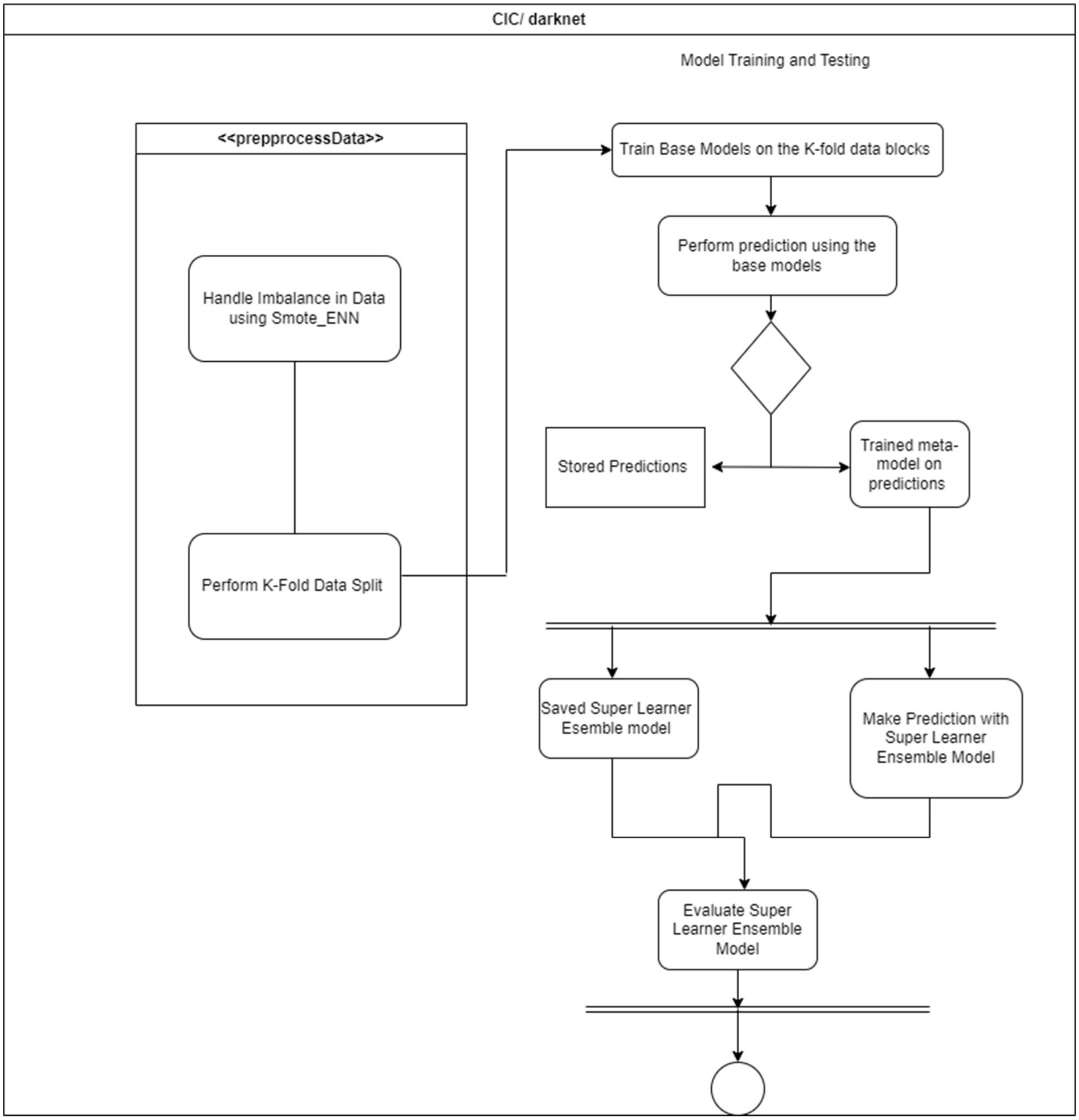

CIC/ darknet
Model Training and Testing
<<prepprocessData>>
Handle Imbalance in Data using Smote_ENN
Perform K-Fold Data Split
Train Base Models on the K-fold data blocks
Perform prediction using the base models
Stored Predictions
Trained meta-model on predictions
Saved Super Learner Esemble model
Make Prediction with Super Learner Ensemble Model
Evaluate Super Learner Ensemble Model

Figure 3.3: UML Activity Diagram of the Proposed Model

3.4. The Ensemble Learner Classifier

Pre-defining of the k-fold splits of the sample dataset is the initial step in the ensemble learner algorithm's working out. Following that, all of the base-learning algorithms are tested on the same splits of the dataset. Then, all of the out-of-fold prediction are saved and utilized to train a model that discovers the most effective way to integrate predictions from the base-classifiers. In essence, the Ensemble is essentially a particular stacking arrangement that is specifically designed for cross-validation with k-folds, with all of the base-classifiers taken into account for the predictive modeling.
The procedure for that of the ensemble learner classifier as used for predicting Darknet Dataset in this study is summarized in Algorithm 3.5 as follows:

3.4.1 Resampling Dataset with SMOTE-ENN Technique

The CIC Darknet dataset D is an imbalanced dataset which is highly skewed towards the legitimate class (0). In order to balance the dataset, the SMOTE-ENN technique will be applied to the original dataset. Developed by Batista et al. (2004), the SMOTE-ENN technique combines the capability of SMOTE to generate synthetic instances for the minority class (i.e., the darknet class in this study) and the capability of ENN to remove some observations from both classes that are identified as having different class between the observation's class and its K-nearest neighbor majority class (i.e., the non-darknet class, in this study).

4. RESULTS AND DISCUSSION

This section describes the System requirements and presents the results obtained from the approach adopted in this study, it shows the outcome from the experimental setup and the result of the experiments carried out (on sample datasets used) in this study based on the various selected standard performance metrics.

4.1. System Requirements

Programming Language of Choice Employed in the Implementation

In this work, the Matlab programming language was employed for implementation. Matlab was chosen because it offers far superior visualization of data capabilities than any other platform and because it is a far better language used for mathematical computing and mathematical-based algorithms.

4.1.1 Software and Hardware Requirements

For efficient execution, models were trained employing the MATLAB package on a Core i5-5200U processor with a 2.2GHz clock speed, 4GB of RAM, and 64-bit Windows OS, while students used the following MATLAB functions; The MATLAB functions fitrensemble, fitcsvm, and fitadbst, which catered to the basic learners and the ensemble learner, were used to develop Bagging, SVM, and Adaboost, respectively.

4.2 Data Preparation and Analysis

4.2.1 Exploratory Data Analysis

Exploratory data analysis (EDA) is a technique in data science used to understand the main characteristics of a dataset, identify patterns, and gain insights before applying more formal statistical methods or building predictive models. During EDA, data analysts explore the dataset through various visual and statistical techniques to answer questions, test assumptions, and generate hypotheses. The process involves examining descriptive statistics, visualizing data using histograms, scatter plots, and box plots, identifying missing values and outliers, and determining relationships between variables.

By conducting EDA, data analysts can gain a better understanding of the dataset, make informed decisions about data preprocessing steps, and identify possible pitfalls or limitations for subsequent modeling tasks. EDA is often considered a crucial step in the data analysis pipeline as it helps guide further analysis and ensures the quality and meaningfulness of subsequent findings. The sample datasets for this study is the CIC Darknet dataset, the CIC Darknet dataset available was an already transformed data. As previously stated in the data description section, there was no background information on the dataset's features, hence the interpretation of each feature cannot be identified for easy comprehension, the dataset contains some identified columns for example the target dataset was alphabetically labeled, some columns were duplicated from the original dataset and there was removal of unwanted columns (columns with '0s') and remapping the values of diagnosis column to (Darknet: 1 and no-Darknet: 0). After checking the various aspects of the dataset like null values count, missing values count, and info.

This dataset was perfect because of no Null and missing values. However, before proceeding to feed the dataset into machine learning algorithms, basic descriptive analysis (summaries and correlation plots) of the PCA (Principal Component Analysis) transformed dataset could be generated in order to observe the behavior of the dataset and determine if it is fit and prepared to be set and input into machine learning algorithms. Figure 4.1 shows the Exploratory Data Analysis (EDA) of the original CIC Darknet Data as retrieved from kaggle.com. The EDA shows the records of the dataset classes of both the majority and minority dataset consisting of 117,219 Non-darknet dataset and 24, 311 darknet dataset.

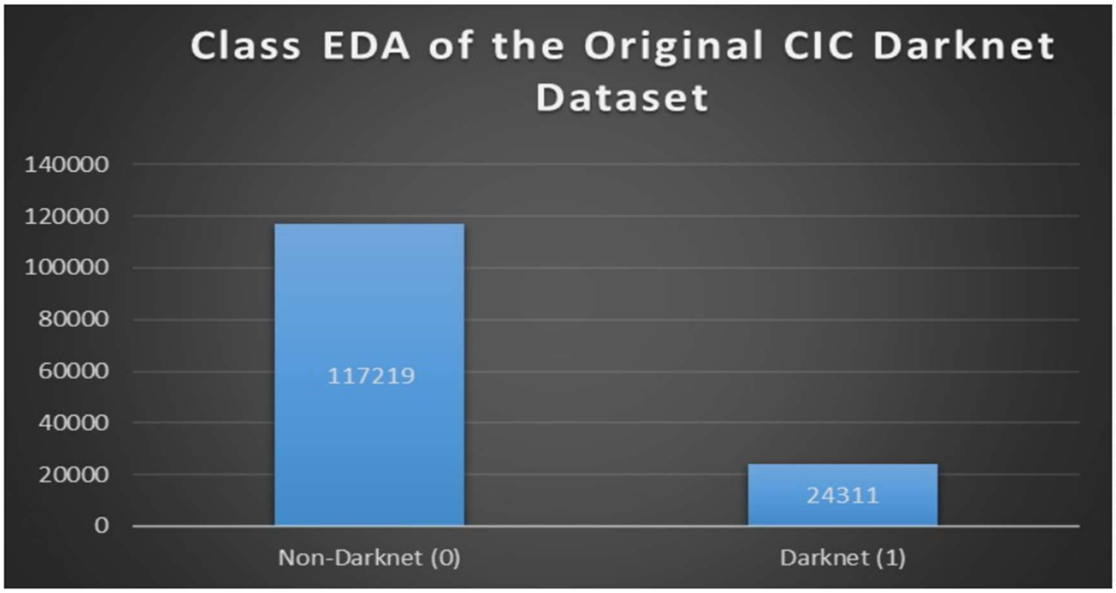


Figure 4.1: Class EDA of the SMOTE-ENN Resampled Dataset

4.3 Experimental Set up

The experimental set up carried out in the study on the sample datasets was used to deduce the results, evaluate and to compare the result for (with SMOTE-ENN) resampling technique. The evaluation of the models using k-fold cross-validation as explain and described in the system architecture. The training and testing takes the same process for all the classifiers, and then the experiment was conducted with the adoption of SMOTE-ENN and the Ensemble Learner classification with Majority Voting Approach. Because data imbalance is a major worry in this experiment, the problem that is faced with the majority and minority classes was not ignored. In order to better balance the dataset, we combine the ENN ability to remove certain observations from both classes that are recognized as having a different class from the observation's class and its K-nearest neighbor majority group with the strength of SMOTE that generate synthetic examples for the minority class. In this experiment, based (learners) classifiers were developed using training data, and predictions were assessed using 10-fold cross-validation. After that, an ensemble of predictions from the learned base classifiers was created using a layered generalization process based on a majority voting method.

4.3.1 Handling Imbalance in the Dataset through SMOTE-ENN Resampling Method

As observed from the raw dataset, the original dataset obtained for this study is highly imbalance. Therefore, to handle the imbalance in the dataset, the dataset was resampled for class-balancing using the adopted resampling method, Synthetic Minority over Sampling Technique with Edited Nearest Neighbour (SMOTE-ENN) which oversamples the minority class and under samples the majority class. The Synthetic Minority Oversampling Technique (SMOTE) aspect of the technique uses a KNN strategy to oversample the minority class by selecting K nearest neighbors, joining them, and creating synthetic samples in the space. The technique's Edited Nearest Neighbor feature eliminates any sample whose class label varies from at least two of its three closest neighbors.

The ENN approach discards cases of the majority class for whom the KNN prediction differs from the majority class's prediction. The EDA plot of the SMOTE-ENN resampled datasets is as shown in Figures 4.2. The resampled dataset now has 117,166 attack records for the Non-darknet class (0) while the darknet class (1) now has 101,386 record of attacks, as shown in Figure 4.1. This therefore means that after the resampling process of SMOTE-ENN method, the resampled dataset now has 218,552 records.

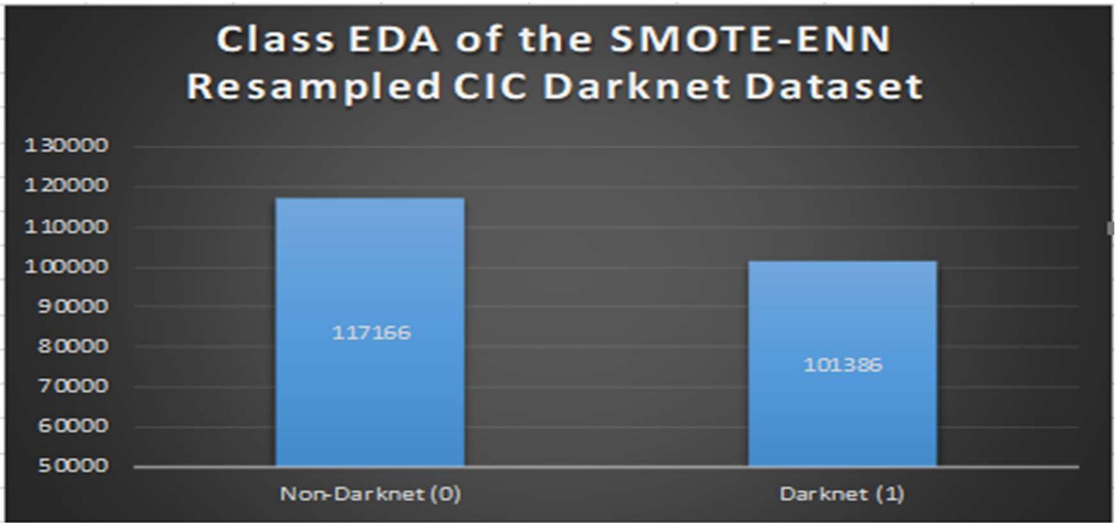


Figure 4.2: Class EDA of the SMOTE-ENN Resampled Dataset

4.4 Models Training and Evaluation

The resampled dataset was divided into two sets: a training set with 117,166 records and a test set with 101,386 records, comprising 70% and 30% of the dataset, respectively. The complete 70% of training set then given to the ensemble learner strategy (which uses three separate base learners) for the 10-fold cross validation (which 10% of the training dataset is kept out for an out-of-fold prediction and the remaining used for training). In order to generate their out-of-fold predictions, each base model is then trained on the training set and evaluated. Following that, the Ensemble learner meta-model is trained on all of the out-of-fold predictions and assessed on the held-out test set. For the training process, all features were taken into consideration.

Based on the made predictions, the performances displayed of each of the base models were evaluated, and the ensemble learner model's performance was also recorded. A number of performance metrics, including accuracy, precision, recall, Matthews Correlation Coefficient, balanced accuracy score, and F1 score, were used to assess the base-models' and super-learner ensemble's performances. Figures 4.3 shows the results of the displayed performance of each base classifiers and the ensemble learner based on accuracy performance. The result accuracy for all the learners shows 89.3068, 78.6304, 93.6658 and 90.7786 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively.

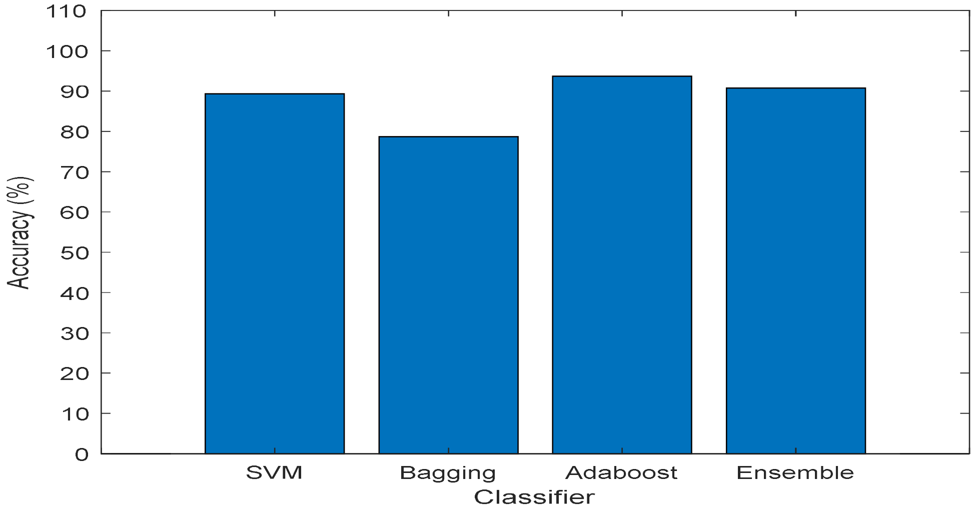


Figure 4.3: Accuracy Performance Evaluation of all the learners on CIC Darknet Dataset

Figures 4.4 shows the results of the displayed performance of each base classifiers and the ensemble learner based on precision as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Precision as the performance measure. The result of the Precision for all the learners shows 91.5039, 73.131, 93.3123 and 90.2053 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on precision indicates that the Adaboost classifier performs better than the other two base classifier, but the result of the ensemble classifier is better than any single classifier.

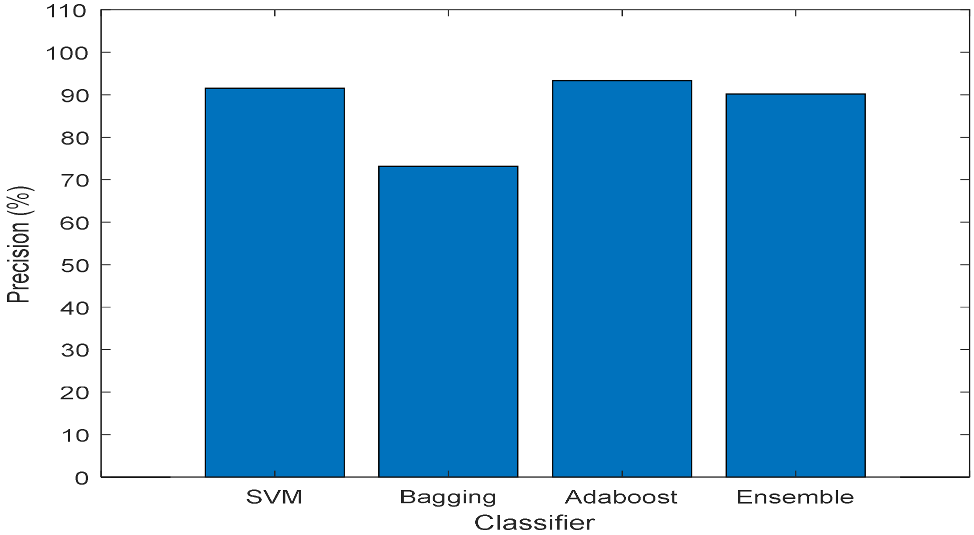


Figure 4.4: Precision Performance Evaluation of all the learners on CIC Darknet Dataset

Figures 4.4 shows the results of the displayed performance of each base classifiers and the ensemble learner based on precision as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Precision as the performance measure. The result of the Precision for all the learners shows 91.5039, 73.131, 93.3123 and 90.2053 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on precision indicates that the Adaboost classifier performs better than the other two base classifier, but the result of the ensemble classifier is better than any single classifier.

Figures 4.5 shows the results of the displayed performance of each base classifiers and the ensemble learner based on Recall as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Recall as the performance measure. The result of the Recall for all the learners shows 90.5313, 59.6364, 92.1324 and 88.345 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on Recall indicates that the Adaboost classifier performs better than the other two base classifiers too, but the result of the ensemble classifier is better than any single classifier performance.

Figures 4.5 shows the results of the displayed performance of each base classifiers and the ensemble learner based on recall as the performance metric.

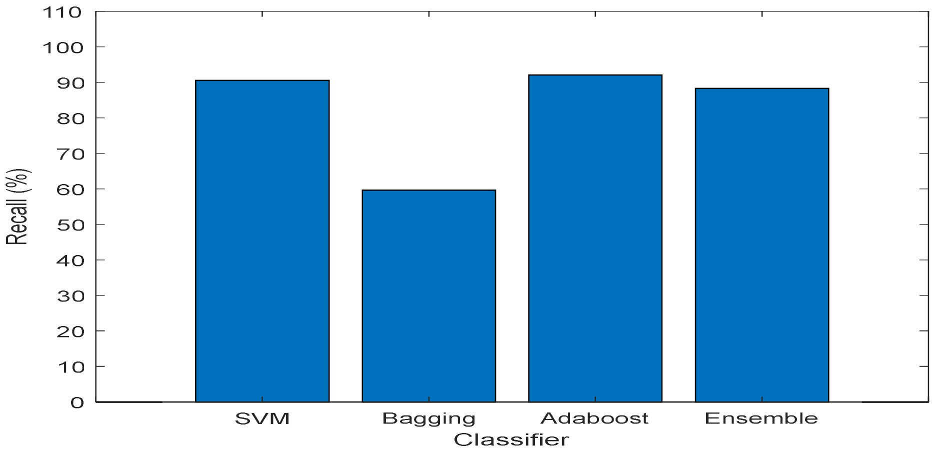


Figure 4.5: Recall Performance Evaluation of all the learners on CIC Darknet Dataset

Figures 4.6 shows the results of the displayed performance of each base classifiers and the ensemble learner based on F1-Score as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Precision as the performance measure. The result of the F1-Score for all the learners shows 91.015, 65.6979, 92.7186 and 89.2654 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on F1-Score indicates that the Adaboost classifier performs better than the other two base classifiers too, but the result of the ensemble classifier is better than any single classifier performance.

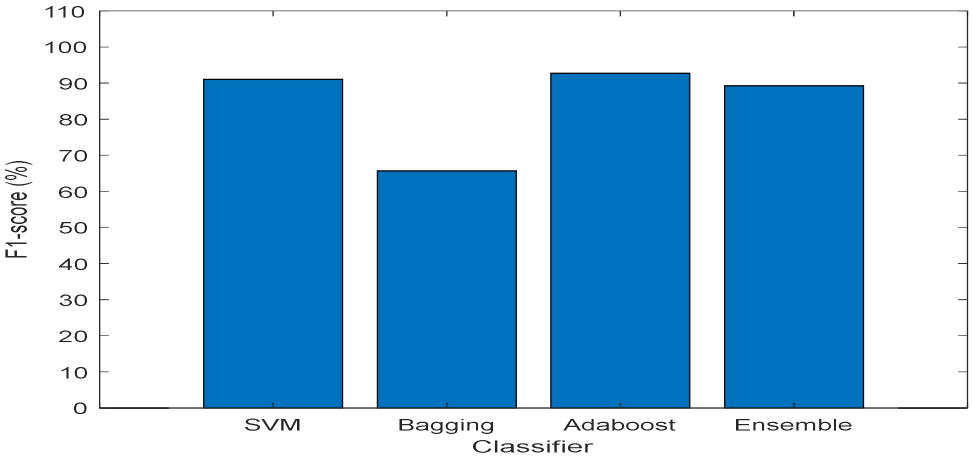


Figure 4.6: F1-score Performance Evaluation of all the learners on CIC Darknet Dataset.

Figures 4.7 shows the results of the displayed performance of each base classifiers and the ensemble learner based on Optimized Precision as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Precision as the performance measure. The result of the Optimized Precision for all the learners shows 89.3663, 78.9267, 93.7532, and 90.8757 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on Optimized Precision indicates that the Adaboost classifier performs better than the other two base classifiers too, Nevertheless, no matter how good the performance of a single classifier is, the result of the ensemble classifier is better than any single classifier performance.

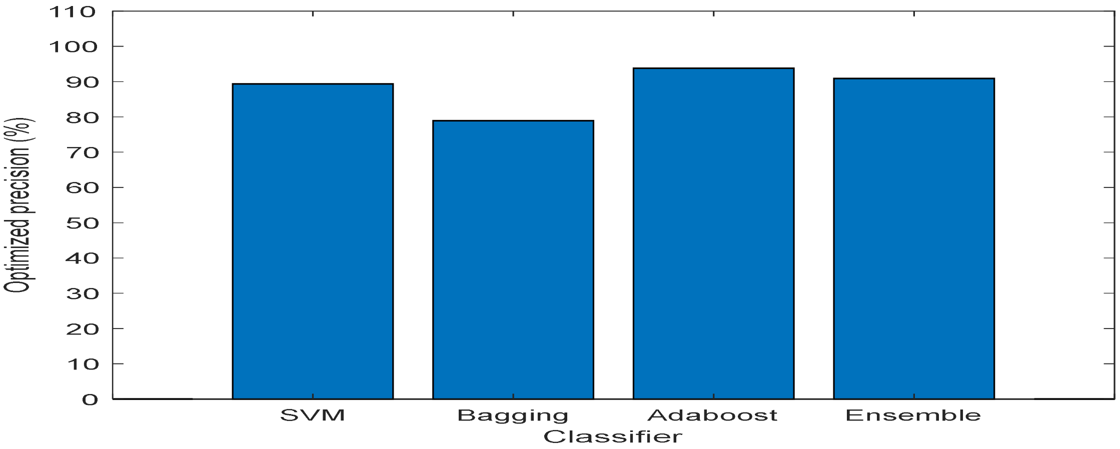


Figure 4.7: Optimized Precision Performance Evaluation of all the learners on CIC Darknet Dataset.

Figures 4.8 shows the results of the displayed performance of each base classifiers and the ensemble learner based on Optimized Precision as the performance metric. It shows the results of the displayed performance of each of the base classifiers and the ensemble learner based on Precision as the performance measure. The result of the Optimized Precision for all the learners shows 96.3893, 86.3109, 101.2968, and 98.2426 for the SVM Classifier, Bagging Classifier, Adaboost Classifier, and Ensemble Classifier respectively. The performance based on Optimized Precision indicates that the Adaboost classifier performs better than the other two base classifiers too, Nevertheless, no matter how good the performance of a single classifier is, the result of the ensemble classifier is better than any single classifier performance.

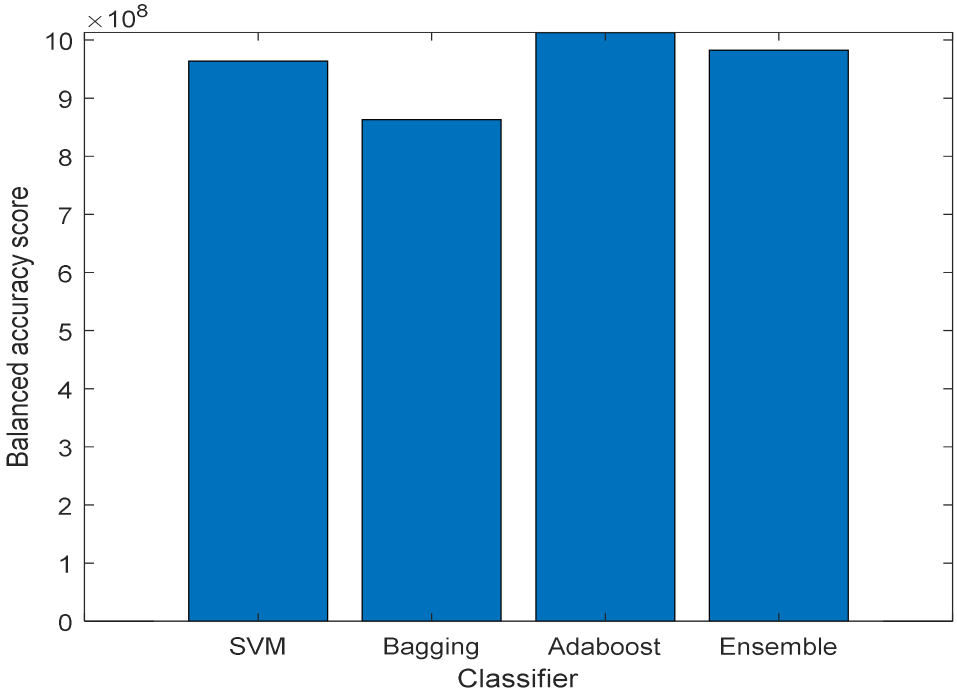
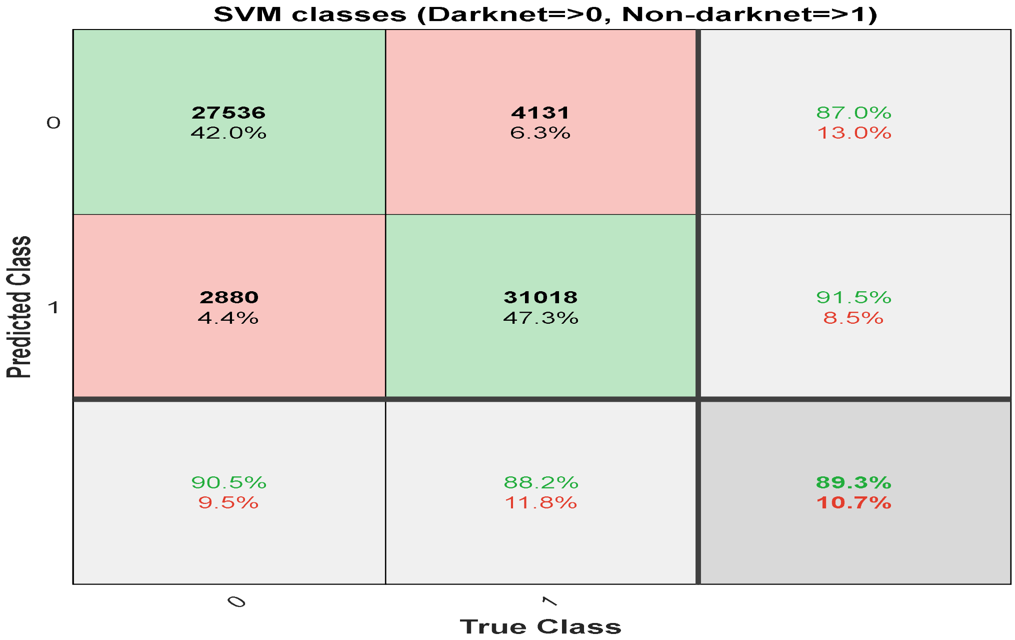


Figure 4.8: Balanced Accuracy Score Performance Evaluation of all the learners on CIC Darknet Dataset

Figures 4.9 shows confusion matrices result obtained from the Support vector machine base model. Accuracy shows how closely the model predict the true value. The experiment recorded from the SVM base learner 89.3068% as the best. Precision refers to how close the model was able to predict the same value to each other. The best result for the experiment one was 91.5039%. Recall is the measure of how many observations the model correctly predicted over the total amount of observation The best result for the experiment one was 90.5313%. Fl- Score: This is the harmonic mean of precision and recall. The best result for the experiment one way 91.0150%.



SVM classes (Darknet=>0, Non-darknet=>1)

Predicted Class \ True Class	0	1	
0	27536 42.0%	4131 6.3%	87.0% 13.0%
1	2880 4.4%	31018 47.3%	91.5% 8.5%
	90.5% 9.5%	88.2% 11.8%	89.3% 10.7%

Figure 4.9: Confusion Matrix of the Support Vector Machine Classifier

Figure 4.10 shows confusion matrices obtained from the Bagging base model.

Accuracy shows how closely the model predict the true value. The experiment recorded from the Bagging base learner 78.6304% as the best.

Precision refers to how close the model was able to predict the same value to each other. The best result for the experiment one was 73.131%.

Recall is the measure of how many observations the model correctly predicted over the total amount of observation The best result for the experiment one was 59.6364%.

Fl- Score: This is the harmonic mean of precision and recall. The best result for the experiment one way 65.6979%. Putting into consideration accuracy, precision, recall and F1-Score performance metrics.

The bagging classifier seems to give the lowest performance in all the performance metrics considered in this study.

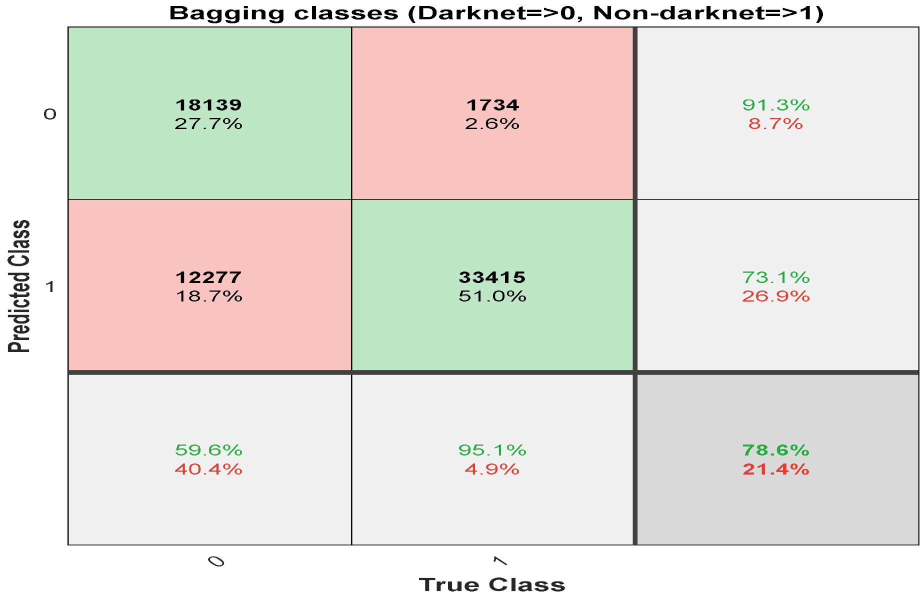


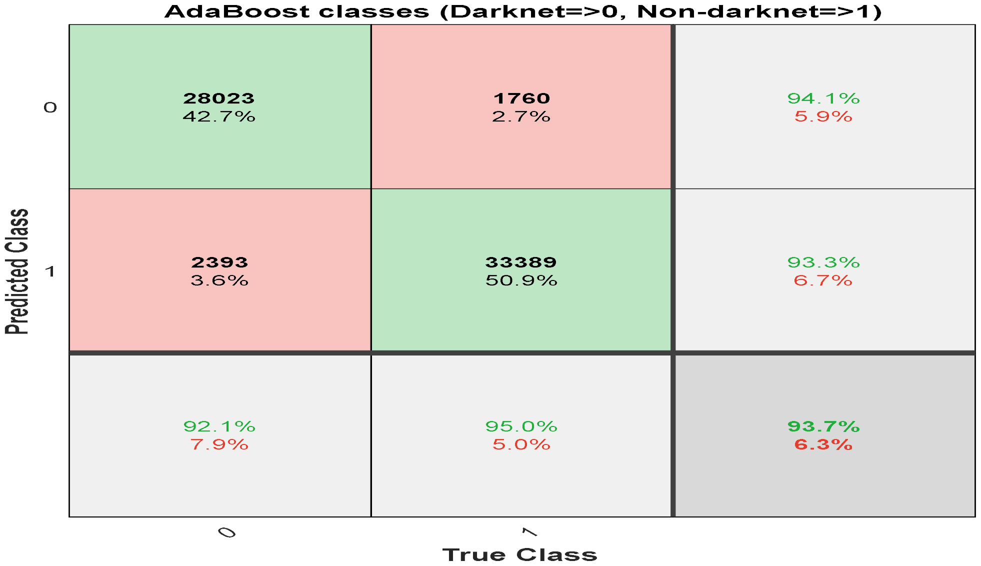
Figure 4.10: Confusion Matrix of the Bagging Classifier

Figures 4.11 shows confusion matrices obtained from the Adaboost base model.

Accuracy shows how closely the model predict the true value. The experiment recorded from the Bagging base learner 93.6658% as the best.

Precision refers to how close the model was able to predict the same value to each other. The best result for the experiment one was 93.3123%.

Recall is the measure of how many observations the model correctly predicted over the total amount of observation The best result for the experiment one was 92.1324%.
Fl- Score: This is the harmonic mean of precision and recall. The best result for the experiment one way 92.7186%. Consideration accuracy, precision, recall and F1-Score performance metrics adopted in this study, the Adaboost classifier give the best predictive performance. This is due to the its ability to increase the accuracy of the weak learners, its immunity from overfitting of data as it runs each model in a sequence and its weight associated with such learners.



	0	1	
0	28023 42.7%	1760 2.7%	94.1% 5.9%
1	2393 3.6%	33389 50.9%	93.3% 6.7%
	92.1% 7.9%	95.0% 5.0%	93.7% 6.3%

Figure 4.11: Confusion Matrix of the Adaboost Classifier

Figure 4.12: Shows the performance comparison of the base classifiers where the Ensemble Learner Consists of the performance of the Support Vector Machine Classifier, Bagging Classifier and the Adaboost classifier.

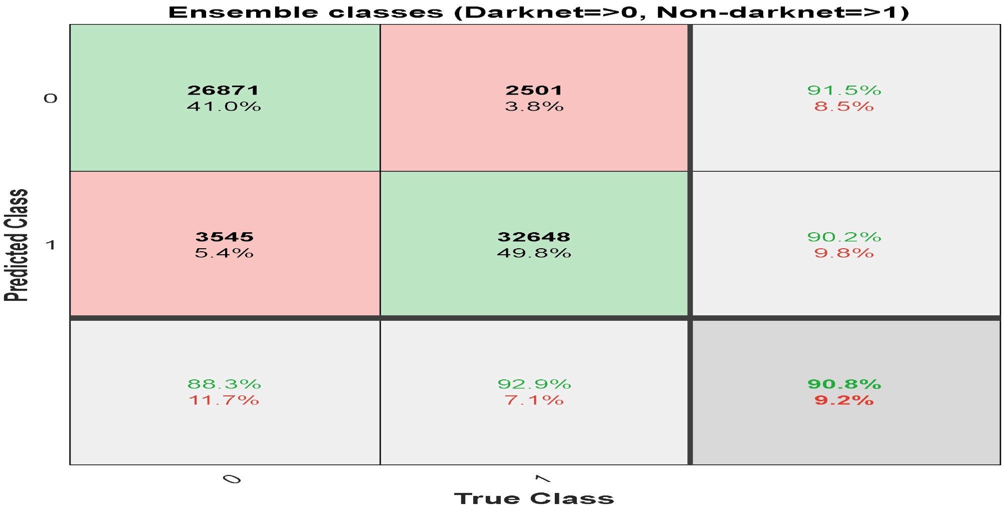
Accuracy shows how closely the model predict the true value. The experiment recorded from the Bagging base learner 90.7786% as the best.

Precision refers to how close the model was able to predict the same value to each other. The best result for the experiment one was 90.2053%.

Recall is the measure of how many observations the model correctly predicted over the total amount of observation The best result for the experiment one was 88.345%.

Fl- Score: This is the harmonic mean of precision and recall. The best result for the experiment one way 89.2654%. Consideration accuracy, precision, recall and F1-Score performance metrics adopted in this study, the Adaboost classifier give the best predictive performance. This is due to the its ability to increase the accuracy of the weak learners, its immunity from overfitting of data as it runs each model in a sequence and its weight associated with such learners.

In all the performance of the base learners class and the Ensemble learner, the Adaboost classifier seems to have a high performance compare to other classifiers notwithstanding, the results output of this model was not worse than the best performing model evaluated at the time of carrying out k-fold cross-validation and has better performance than any single model.



	0	1	
0	26871 41.0%	2501 3.8%	91.5% 8.5%
1	3545 5.4%	32648 49.8%	90.2% 9.8%
	88.3% 11.7%	92.9% 7.1%	90.8% 9.2%

Figures 4.12: Confusion matrix obtained from the ensemble learner predictions

Table 4.1: Performances of all the models in terms of accuracy, precision, recall, F1 score, and balanced accuracy score (BAS).

	SVM classifier	Bagging classifier	Adaboost classifier	Ensemble (SVM+ Bagging +Adaboost) Using Majority Voting
Accuracy (%)	89.3068	78.6304	93.6658	90.7786
Precision (%)	91.5039	73.131	93.3123	90.2053
Recall (%)	90.5313	59.6364	92.1324	88.345
F1-score (%)	91.015	65.6979	92.7186	89.2654
Optimized precision (%)	89.3663	78.9267	93.7532	90.8757
Matthew correlation coefficient	0.78618	0.59356	0.87264	0.8146
Balanced accuracy score	963893329	863109829.5	1012968764.5	982426444

Figure 4.13: Shows the performance comparison of the base classifiers where the Ensemble Learner. The Adaboost classifier seems to have a high performance compare to other classifiers notwithstanding, the results output of this model was not worse than the best performing model evaluated at the time of carrying out k-fold cross-validation and has better performance than any single model.

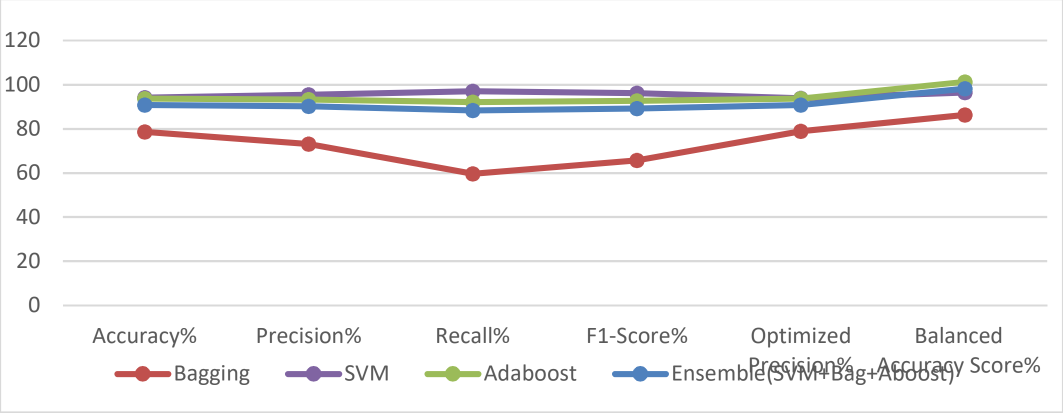


Figure 4.13: Performance comparison of the base classifiers with the Ensemble Learner

Figure 4.14 Shows the performance comparison developed model with the other existing works of the same CIC Darknet data samples. The performance metrics that are similar from the studies were picked for comparison. Observation shows the existing study use just one or two performance evaluation metrics (accuracy and F1-score) whereas the developed model based on its performance evaluation on six (6) metrics such as accuracy, precision, recall, f1-score, optimized precision and balanced accuracy score. This shows that the proposed model has better performance for most of the metric comparison. This gave a state of art performance.

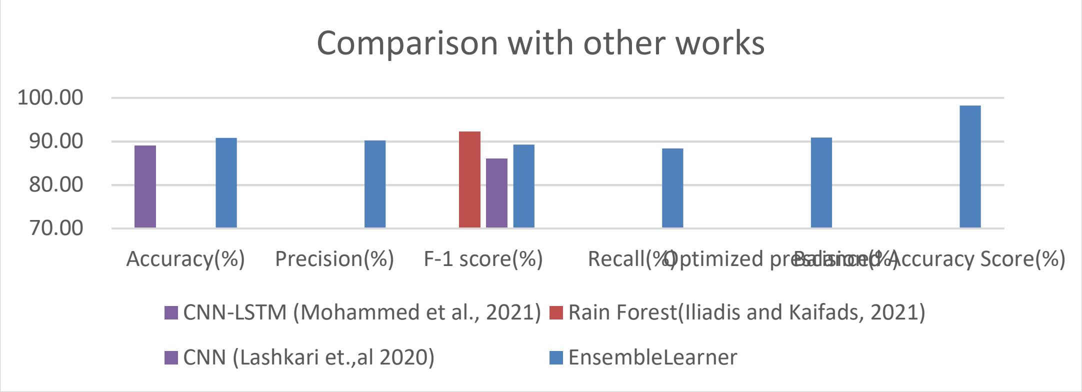


Figure 4.14: Performance comparison of the Developed Ensemble Model with other works

5. CONCLUSION

This research presents a resampling technique developed based on an adoption of SMOTE-ENN for solving classification problem of imbalance in datasets. This developed system was able to handle the issue with classification due to imbalanced dataset at the data level. According to Vishwa et al., (2019) it suggested that a promising avenue to the research of solving data imbalance problem is to investigate the effectiveness of a developed resampling technique by combining it with ensemble methods. Hence, the developed system was able to come up with combining the SMOTE-ENN resampling technique with an Ensemble method to provide a solution that addressed the ,problem of data imbalance experienced by many of the algorithms currently used. The results outcome of this model was not worse than the best performing model evaluated during k-fold cross-validation and has better performance than any single model. The built system was validated and tested the CUIC Darknet dataset. Comparative analysis shows the developed model can achieve better performance and adaptivity than conventional methods, in terms of classification accuracy, recall, F1-Score and Optimized precision, and BAS on sample datasets, validating effectiveness of this model, which is a large study with state-of-the-art performance

REFERENCES

1. Akinboro S.A., Olusesi A.T. 2019 A Model for Self-Adaptive Routing Optimization in MobileAd-Hoc Network International Journal of Swarm Intelligence Research 10(1):58-74 DOI:10.4018/IJSIR.2019010104.
2. Kaur, S., & Randhawa, S. (2020). Dark web: A web of crimes. Wireless Personal Communications, 112(4), 2131-2158.
3. Li, K., 2017. Optimal Task Dispatching on Multiple Heterogeneous Multiserver Systems with Dynamic Speed and Power Management. IEEE journal on Transactions on Sustainable Computing Volume: 2, Issue: 2, 01 April-June 2017 pg 167-182, DOI: 10.1109/TSUSC.2017.2706425
4. Muhammad B. S., Muhammad K. H., Ramzan T., Muhammad Y., and Muhammad U. S. 2018
5. Darkdetect: Darknet traffic detection and categorization using modified convolution-long short-term memory. IEEE Access, 9:113705–113713, 2021.
6. Olowookere T., & Adewale O. 2.020 A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach. Scientific African Journal Volume 8 (2020) e00464 https://doi.org/10.1016/j.sciaf.2020.e00464 2468-2276/
7. Qiang. Y., Xindong W. 2006. 10 Challenging Problems in Data Mining Research. InternationalJournal of Information Technology & Decision Making, Vol. 5, No. 4 597–604
8. Sui, D., Caverlee, J., &Rudesill, D. 2015. The deep web and the darknet. Washington DC: Publication Wilson Center.
9. Xing, L. Demertzis, K., and Yang, J. 2020. Identifying data streams anomalies by evolving spiking restricted Boltzmann machines. Neural Comput. Appl., vol. 32, no. 11, pp. 6699–6713, 06/2020, doi: 10.1007/s00521-019-04288-5.
10. Zakariye M. O., and Jamaluddin, I. 2020. An Overview of Darknet, Rise and Challenges and its Assumptions International Journal of Computer Science and Information Technology Research ISSN 2348-120X Vol. 8, Issue 3, pp: (110-116), July - September 2020.