International Journal of Computer Techniques -– Volume 8 Issue 4, August 2021

RESEARCH ARTICLE OPEN ACCESS

Identification of Classification Anomalies in Students' Areas of Specialization using Ensemble Classifiers

Anthony Irungu Njina*, Mary Asunta Gaceri**
* *Department of Computer Science, Meru University of Science and Technology*
Email: freenjina@gmail.com
** *Department of Information Technology, Meru University of Science and Technology*
Email: gacerimariah@gmail.com

--************************----------------------------------

Abstract:

In this paper, we propose an ensemble of different classifiers and examine the distribution of anomalies in the classification reports of individual model results. The ensemble is constructed using three base classifiers: Multinomial Naïve Bayes (MNB), Support Vector Machines (SVM), and Random Forest (RF). We expect improved accuracy as a result of the combined prediction power of different algorithms, and hypothesize that a concurrence of high error rates within a class is an indication of classification anomaly. Results showed an improved accuracy for classes where individual F1 scores were within the range of the average F1 scores. However, we could not make similar observations for classes with low support, and in the classes identified with possible instances of misclassification. The results obtained from this experiment suggest that an ensemble model with data preprocessing is a more accurate model for predicting students' subject area combinations.

Keywords — **Misclassification, Anomaly Detection, Ensemble Learning, Students Tracking.**

--************************----------------------------------

I. INTRODUCTION

An anomaly is a data instance whose characteristics or features deviate significantly, such that it creates doubt as to whether it was generated using a similar mechanism as the other instances in the same data set. Anomaly detection refers to the task of identifying or the problem of finding these instances in a given dataset. Terminologies like outliers, noise, exceptions, rare events are also used interchangeably in different literature to refer to anomalies within datasets [1],[2]. Anomalies manifest as point, contextual, or collective anomalies. Point anomalies occur when a data instance is anomalous compared to the rest of the data. They represent the simplest form of anomaly and are common in a majority of studies on anomaly detection. Contextual or conditional anomalies occur where a data instance is normal in one context but anomalous in another. Collective anomalies occur where a group of related data instances is anomalous in comparison to the rest of the dataset. There are two main reasons for identifying anomalies in a dataset. In classification tasks, anomalies affect how classifiers learn especially in supervised learning. Therefore, once anomalies instances are identified, removing them is an important step for producing an accurate model. Secondly in tasks such as intrusion detection, fraud detection, or medical diagnosis the anomalous instance is usually the observation of interest as it carries valuable information [2].

many challenges that make the task of identifying anomalies problematic. In real-life situations, data is usually extremely huge, noisy, and unlabeled. There is usually no normal region that is defined in any given dataset. The boundary between normal and outlier behavior is usually not clear. In other applications, the normal behavior keeps evolving with malicious advisories (attacks, new disease symptoms, etc.) mimicking normal activities. Lastly, the exact concept of an outlier is different for different application domains. The other challenge is that labeled data for training or validating results is commonly not available.

Noise and outliers affect how different machine learning algorithms model the data. However, a universal method of detecting anomalies or noise within a large dataset does not exist. This is because every dataset emanating from real-life scenarios exhibits unique characteristics requiring a unique machine learning approach.

A. Ensemble learning

Ensemble learning entails the use of different learning models, based on different algorithms to solve a problem. Instead of making one hypothesis from training data, ensemble methods construct a set of hypotheses and combine them to develop an enhanced generalization ability. Classification algorithms such as SVM, Random Forest, and Naïve Bayes differ in their hypothesis space, the model quality criteria, and search

strategies [3]. The base classifiers chosen for ensemble learning should be as accurate as possible and also as diverse as possible. [2] proposed an ensemble model which automatically and systematically selects the results from its constituent detectors. They advise that a selective combination of the base learners is vital in building effective anomaly ensembles. In ensemble learning, combining many base learners does not necessarily result in improved performance. [2] advises that trusting results from all constituent detectors may deteriorate the overall performance since some detectors could return erroneous results based on the quality of the training dataset and also the underlying assumptions of individual detectors.
In this study, several classification algorithms will be used to identify instances that are likely to be misclassified in a dataset containing labels composed of two teaching subjects. The next sections describe the classification problem, followed by a literature review of related work, a methodology and tool used to solve the problem, a summary of results and discussion, and finally a conclusion and recommendations for further studies.

B. Description of the Problem

The main task entails tracking students taking Bachelor of Education Arts and Bachelor of Education Science. The basic data requirements in this task involve identifying the two teaching subjects for the graduated and current students while referring to a set of course units studied. The curriculum implementation practices in the school had not anticipated such requirements during registration due to some factors. First, students are admitted into either the Science or the Arts program depending on whether they meet a set cut-off points plus a specific minimum grade in two teaching subjects in science or in Arts respectively. In many cases, the students meet this requirement in more than two subjects, and the choice of the specific teaching subjects is deferred to a later date (sometimes as late as in the second semester of study).
Secondly, the curriculum is implemented using a unit-based approach whereby a student chooses individual course units for a given semester, based on individual financial constraints, and more often, personal preferences. The consistency of students' progression relies on the consistency of advisory given by program course advisors, and on the consistency of students in seeking such advisory when selecting course units at the beginning of every semester.
Thirdly, the institution has not put in place a mechanism of updating enrolment data to capture details of teaching subjects, once students choose specific teaching subjects. Notably, the students are categorized as either Science or Arts upon completion of the program depending on the subjects covered. This categorization is vague and not sufficient for the current task. For example, students with Mathematics and Business units were categorized as Arts, while students with Mathematics and Computer Studies units were categorized as science, and in other cases, students with Business and Computer studies were categorized as Arts.
Finally, due to constant revisions and evolution of academic programs, different course units were not coded consistently, whereby courses had codes with inconsistent prefixes. Thus, a simple glance at course units covered by the students would not be not sufficient in identifying their teaching subjects. Furthermore, students could change their choice of teaching subjects before completing the course or could choose to take extra or optional units. Moreover, a significant number of course units are common to all students and thus do not have any relationship with any given combination of teaching subjects.
In summary, the task involved categorizing 3799 students into 41 different subject area combinations. Each student is registered into an average of 38 course units. A total of 777 different course units were covered by students in all subject area combinations. Using domain knowledge on different course units, a team manually categorized 2613 students into 41 different subject areas. In this paper, we shall explore the use of various classifiers to build an improved classification model and compare the results of the different classifiers to identify instances that were possibly misclassified.

II. RELATED WORK

The use of ensemble learning in anomaly detection has been widely investigated in various disciplines such as engineering, health, and agriculture. [1] explored the use of ensemble approaches for anomaly detection in the building energy domain. Their approach sought to enhance the performance of a pattern-based anomaly classifier using the ensemble anomaly detection (EAD) framework. Using majority voting, the ensemble improved the sensitivity of the pattern-based anomaly classifier by 3.6% and reduced the false positive rate by 2.7%. However, the study did not achieve an optimal combined threshold of the EAD using the optimal threshold values of the base classifiers.
[4] proposed an approach for detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. The ensemble classifier consisted of two-class and one-class classifiers. The model was able to function without requiring adjusting of parameters by a domain expert and was also adaptive to different driving scenarios. [5] used an ensemble of classification models based on the artificial immune system to identify mammography anomalies for breast cancer detection. They found out that the ensemble models achieved very high classification rates even when training and testing were done on two completely independent and heterogeneous datasets.
[6] proposed a methodology of detecting anomalies in the quality of water using seven classification algorithms on the same classification task. They found out that not all seven algorithms produced good results given the large data set. They also advised that feature selection is a very important stage in deciding an appropriate algorithm for anomaly detection.
[2] proposed ensemble approach for anomaly detection called SELECT, which automatically and systematically selects the results from base detectors. The model then combined the results in a fully unsupervised fashion, and yielded superior performance compared to the individual detectors alone, the

full ensemble, an existing diversity-based ensemble, and an existing weighted ensemble approach.

[7] used ensemble model of Decision Trees, Naïve Bayes, and Random Forest as the base classifiers to improve the quality of student data by eliminating noisy instances. By empirically comparing the technique with single model-based techniques, they found out that using ensemble models gives better predictive accuracies. After eliminating the noisy instances, the authors were able to generate association rules for understanding the factors influencing student outcomes

In this study, an ensemble model of Multinomial Naïve Bayes (MNB), Support Vector Machines (SVM), and Random Forest (RF) shall be used to identify classification anomalies in assigned labels of teaching subjects in the students' data. The original data set of 777 features shall be used to test the validity of the base learners. Thereafter, pre-processing of data shall be done to reduce the number of features using domain knowledge on the various course units taken by a student. First, a considerable number of units are common to all students, and thus, their presence in the data set represents noise that affects the performance of the classification algorithms. Secondly, up to a significant extent, course units can be grouped using their prefixes. These pre-processing tasks shall be described in detail in the next section.

III. METHODOLOGY

This section describes the machine learning algorithms and tools used in this study. The initial problem of classifying 2613 instances with 777 attributes or features into 41 classes, using shall be described using a probabilistic method. We shall then introduce two other classifiers (support vector machine, and random forest) that are based on different machine learning techniques and compare their results. Finally, three classifiers shall be used to build a Voting classifier where predictions shall be based on majority voting. The outliers identified in the results of the ensemble model shall be used as confirmation of classification anomalies in the original data set. Finally, the performance metrics used in the anomaly detection classifiers shall be described.

A. *Naïve Bayes Classifiers*

Naive Bayes classifiers are modeled using simple, but very effective probabilistic algorithms based on the Bayes Theorem. The dataset comprises 2613 instances described using 777 attributes (course units) that were used to categorize them into 41 possible classes (subject areas). Using features $X_0, \dots, X_n$ and classes $C_0, \dots, C_k$, the model shall determine the probability of specific features appearing in each class. Therefore, for each possible subject area combination **c**, we shall calculate

$P(\,C_i\,|X_0, \dots, X_n)$ Where n=777 and k=41

From the basic representation of the Bayes rule,

$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

We state that $P(A|B)\alpha\, P(B|A) \times P(A)$ by considering $P(B)$ as the normalization term.

Thus, $P(\,C_i\,|X_0, \dots, X_n) \propto P(\,X_0, \dots, X_n|C_i\,) \times P(C_i)$

The term $P(C_i)$ represents the proportion of the entire dataset that is of class C_i. The terms $P(\,X_0, \dots, X_n|C_i\,)$ were computed by assuming that the features $X_0, \dots, X_n$ are conditionally independent given each class. In this case, course units are not clustered together for a particular subject combination. This is because students could select a minimum of one up to a specified maximum of units in each semester such that the presence of a course unit in a selection does not influence the presence or absence of another course unit.

Thus, $P(\,X_0, \dots, X_n|C_i\,) = P(\,X_0|C_i\,) \times P(\,X_1|C_i\,) \times \dots \times P(\,X_n|C_i\,)$

Therefore

$$\boldsymbol{P(\,C_i\,|X_0, \dots, X_n) \propto P(C_i)\prod_{j=1}^{n} P(\,X_j|C_i\,)}$$

The data set in this case is distributed according to multivariate Bernoulli distribution. There exist multiple features, where each of the features assumes a binary variable. For each of the 777 possible features (course units) that describe the dataset, a binary value of 1 is assigned where a student is registered for a certain course unit, and a binary value of 0 is assigned where a student is not registered for a certain course unit.

The computation of the individual $P(\,X_j|C_i\,)$ terms are done by observing the probability of a course unit X_j given a subject area combination, C_i. On average, a student is registered into 38 out of 777 possible course units. Therefore, in a particular class comprising of students from a specific subject area, even given a few outliers (elective units, wrong advisory and so on), the majority of the $P(\,X_j|C_i\,)$ terms equal to 0. That means that no student ever registered for a course X_j belonging to the subject area C_i. Therefore, using such terms will end up nullifying the entire calculation of the Cartesian product in equation 3. This problem is solved by applying sample correction in the computation (smoothing) such that none of the probability estimates in $P(\,X_j|C_i\,)$ terms equal to zero. Maximum A Posteriori (MAP) decision rule is used to determine the most likely class by selecting the class that returns the highest probability.

$$y = \underset{c_i}{argmax}\, P(C_i)\prod_{j=1}^{n} P(\,X_j|C_i\,)$$

The Multinomial Naïve Bayes (MNB) and Bernoulli Naïve Bayes (BNB) algorithms are applicable probabilistic models for the dataset used in this classification problem. However, empirical comparisons of the two models show that MNB is usually superior to BNB. A high dimensional dataset commonly presents a big challenge for learning algorithms in a classification task. Tang, Kay, and He (2017) used an automated feature selection in their study on text categorization to reduce the feature size and to speed up the learning process of classifiers. This is because there will be fewer parameters to be estimated and also fewer probabilities to be computed. The approach entailed selecting only a subset of 1original features for input to the learning algorithms. Provided that the size if the feature subset is correctly chosen, the accuracy of a classifier will also be significantly improved [8]. Therefore, the MNB will be selected as one of the base classifiers of the ensemble model.

B. Random Forest Classifier

A random forest is considered an ensemble model in itself since it is a collection of decision trees combined to make a more accurate model. Random forest works by constructing multiple decision trees during training and returning the mode of the classes yielded by individual trees. A random forest is composed of an ensemble of binary trees $\{T_1(F), \ldots, T_n(F)\}$, where F = $\{f_1, \ldots, fn\}$ is an n-dimensional feature vector. The ensemble produces n outputs $\{\hat{Y}_1=T_1(F),\ldots, \hat{Y}_n=T_n(F)\}$. The final prediction $\hat{Y}$ is made by averaging the values predicted by each tree [1].

Random Forests are commonly prone to overfitting, and they tend to be poor at predicting underrepresented classes in unbalanced datasets. These weaknesses can be controlled by defining the maximum depth of the trees, and the maximum number of features to be considered at each split. By considering all features, the performance of the random forest reduced drastically when applied to the initial dataset of 777 different course units. Reducing dimensions is thus a crucial step in achieving more accurate results from a random forest classifier.

C. Support Vector Machine (SVM) Classifier

Support Vector Machine (SVM) is a supervised learning method for classification. In SVM, hyperplanes are constructed to separate various classes of instances in high-dimension space. Finding the hyperplane is an optimization problem, which maximizes the distance between a hyperplane and the nearest data point in different classes.

SVMs have been observed to achieve reliable predictions in classification problems, especially in anomaly detection, and with results comparable with those from other classifiers. In anomaly detection via SVM, if a new instance is located above the hyperplane, it would be reported as an anomaly, while marked as normal otherwise While using high dimensional datasets, SVM based approaches are used in reducing dimensionality, and redundant information [10],[11],[9],[13]. The two common variants of SVM are linear SVM and non-linear SVM. Based on initial experiments using the original data set, linear SVM outperformed non-linear SVM was therefore chosen for the ensemble model.

D. Voting Classifier

Classification algorithms such as SVM, Random Forest, and Naïve Bayes differ in their hypothesis space, the model quality criteria, and search strategies [3]. The base classifiers chosen for ensemble learning should be as accurate as possible and also as diverse as possible. Thereafter several combination techniques can be used. The three classifiers were combined in such a way to produce an ensemble majority voting classifier that is superior to any of the individual classifiers. In anomaly detection majority voting has been widely chosen as the combination technique [1]. The majority voting classifiers combine the three base algorithms and assign the class that receives the largest number of votes. [12] shown that the ensemble majority voting classifier always achieves better accuracy as it is analyzed in the relative confusion matrices.

IV. RESULTS AND DISCUSSION

The main objective of this study was to identify classification anomalies in a dataset of students' teaching subjects. The dataset did not have labels of whether instances were misclassified or not. Therefore, we used classification models to reclassify the instances using three base classifiers with stratified sets of training and test data at 75% and 25 % respectively. Using the original data set of 2613 rows and 777 features, initial tests using returned an average score of 84% (MNB), 86% (SVM), and 89% (RF).

Pre-processing of the data reduced the number of dimensions (attributes) to 52. This was done by first dropping the course units that were considered core to all students regardless of their teaching subjects, and then combining the ones that were considered to be related into a single feature. The classifiers performed better using the new data with average scores of 92%, 96%, and 97% respectively. For the ensemble model, using the three base classifiers, 5 classes were dropped due to low support, resulting in a dataset of 2615 rows and 52 feature attributes. The voting classifier yielded marginal improvement with average scores of 98 %.

A. Metrics used

The confusion matrix shows a representation of actual labels against labels predicted by the classification model. The diagonal elements of the matrix represent true-positives (TP) for each class. The elements along the row, excluding the diagonal element represent the false-positives (FP) for a class corresponding to that row. The elements along a column excluding the diagonal element represent the number of false-negative (FN) for a class corresponding to that column.

We evaluated the results of each class using precision, recall, and F_1 scores, summarized in the classification reports. Recall (R) is the ability of a classification model to identify all relevant instances. Precision (P) is the ability of a classification model to return only relevant instances, and finally, F1 score is a single metric that combines recall and precision using the harmonic mean and is usually recommended as an appropriate metric for evaluating performance [14]. The three metrics are computed as follows;

$$R = \frac{TP}{TP + FN} \qquad P = \frac{TP}{TP + FP} \qquad F_1 = 2\frac{P \times R}{P + R}$$

From the results, classes were categorized into three groups. The first category included classes with high precision and high recall. In these classes the majority of the **many** instances identified for a particular class were **correct**. The second category was for classes with high precision and low recall, in which case, a majority of the **few** instances identified for a particular class are **correct**. The third category included classes with high recall and low precision, where a majority of the **many** instances predicted for particular class were **wrong**. Table 1 below shows the scores of the 25 classes with high precision and high recall. The average classification accuracy of the three base models suggested that the likelihood on classification anomalies within these classes were low.

Table 1: Classes with high precision and high recall

Class	MNB			SVM			RF			Support
	P	R	F1	P	R	F1	P	R	F1	
Agriculture_Biology	1	1	1	1	1	1	.99	1	.99	66
Agriculture_Chemistry	1	1	1	1	1	1	1	1	1	6
Agriculture_Geography	1	1	1	1	1	1	1	1	1	5
Agriculture_PE	1	1	1	1	1	1	.67	1	.8	2
Biology_Chemistry	.95	1	.98	.97	.97	.97	.97	.97	.97	40
Biology_Mathematics	1	1	1	1	.95	.97	.9	.95	.93	20
Business_Geography	1	1	1	.91	1	.95	1	1	1	10
Business_History	1	1	1	1	1	1	1	1	1	4
Business_PE	1	1	1	1	1	1	1	1	1	2
Chemistry_Geography	1	1	1	1	1	1	1	1	1	1
Chemistry_Mathematics	1	.92	.96	.96	1	.98	.96	1	.98	24
Chemistry_PE	1	1	1	1	1	1	1	1	1	1
Chemistry_Physics	1	1	1	1	1	1	1	1	1	12
Computer_Geography	1	1	1	1	1	1	1	1	1	1
Computer_PE	1	1	1	1	1	1	1	1	1	1
Computer_Physics	.5	1	.67	1	1	1	1	1	1	1
English_Literature	.99	.97	.98	1	.92	.96	.99	1	.99	75
Geography_History	.86	.72	.79	.92	1	.96	.97	.95	.96	61
Geography_Kiswahili	1	1	1	1	1	1	1	1	1	16
History_Kiswahili	1	1	1	1	1	1	1	.97	.98	31
History_Religion	.73	.94	.82	.99	.94	.96	.95	.97	.96	100
Kiswahili_PE	1	1	1	1	1	1	1	1	1	1
Kiswahili_Religion	1	.98	.99	.98	.98	.98	.98	.98	.98	49
Mathematics_Physics	1	1	1	1	1	1	.97	1	.98	31
PE_Physics	1	1	1	1	1	1	0	0	0	1

Table 2: Classes with high precision and low recall or vice-versa

Class	MNB			SVM			RF			Support
	P	R	F1	P	R	F1	P	R	F1	
Agriculture_Mathematics	0	0	0	0	0	0	0	0	0	1
Biology_Geography	1	1	1	1	0.67	0.8	1	0.33	0.5	3
Business_Computer	0.85	1	0.92	0.9	0.82	0.86	1	0.82	0.9	11
Business_Kiswahili	1	1	1	0.75	0.75	0.75	1	0.75	0.86	4
Business_Mathematics	1	0.64	0.78	1	1	1	0.96	1	0.98	22
Business_Religion	0.86	1	0.92	0.5	1	0.67	0.86	1	0.92	6
Computer_Mathematics	1	0.75	0.86	0.8	1	0.89	0.89	1	0.94	8
Geography_Mathematics	0.83	0.83	0.83	0.71	0.83	0.77	1	1	1	6
Geography_Religion	1	0.62	0.76	1	0.92	0.96	0.96	0.92	0.94	26
Mathematics_PE	.67	1	.8	.67	1	.8	1	1	1	2

Table 2 shows the results of the 10 classes that were identified as having a high recall and low precision and vice versa from the initial results of the base classifiers. The classes with low support (with<5) were ignored since the poor results were attributed to low training. Using the average F_1 score of the ensemble model at 0.98, 7 classes were identified with scores less than 0.97. Comparing these with the results in table 2 above, 5 out of the original 6 classes were retained and two new classes being identified.

Table 3: Extract of the ensemble model's classification report

Teaching Subjects	P	R	F1	support
Agriculture_Chemistry	0.86	1	0.92	6
Business_Computer	1	0.73	0.84	11
Business_Geography	0.83	1	0.91	10
Business_Mathematics	0.92	1	0.96	22
Business_Religion	1	0.83	0.91	6
Computer_Mathematics	0.89	1	0.94	8
Geography_Religion	1	0.92	0.96	26

Repeated trials using the ensemble model yielded different set of classes that were likely to have been misclassified. However, only 3 out of the 7 classes in table 3 were consistently identified as misclassified as shown in table 4 below.

Table 4: Extract of repeated trials using the ensemble model.

	Trial 1			Trial 2			Trial 3			
Teaching Subjects	P	R	F1	P	R	F1	P	R	F1	support
Buss_ Comp	1	.73	.84	1	.82	.9	.89	.73	.8	11
Buss_ Math	.92	1	.96	.88	1	.94	.95	.91	.93	22
Comp_ Math	.89	1	.94	.86	.75	.8	.7	.88	.78	8

V. CONCLUSION

In this paper, we explored the use of different algorithms to identify classification anomalies in student data. The results show that the pre-processing of data using domain knowledge improved the performance of classification algorithms by an average of 10%. The use of an ensemble model improved the performance by 1% and also aided in identifying 7 classes with classification anomalies. Further, repeated tests and comparisons of the confusion matrix could not yield the same classes as being misclassified, since only 3 of the 7 classes were consistently identified as misclassified in the repeated trials. Therefore, we concluded that due to the high imbalances in the number of features within classes, the classification anomalies are likely to have been masked in classes with higher support. Similar limitations of machine learning in detecting anomalies while using unbalanced data were observed by [14]. Future research will explore the identification of classification anomalies using clustering techniques.

REFERENCES

[1] Araya, D. B., Grolingera, K., ElYamany, H. F., Capretz, M. A., & Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings, 144*, 191-206.

[2] Rayana, S., & Akoglu, L. (2016). Less is more: Building selective anomaly ensembles. *ACM Transactions on Knowledge Discovery from Data (TKDD), 10*(4), 42.

[3] Li, Y., Gao, J., Li, Q., & Fan, W. (2014). Ensemble Learning. In C. C. Aggarwal, *Data Classification, Algorithms and Applications* (pp. 502 - 528). New York : Chapman and Hall/CRC.

[4] Theissler, A. (2017). Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. *Knowledge-Based Systems, 123*, 163-173.

[5] Magna, G., Casti, P., Jayaraman, S. V., Salmeri, M., Mencattini, A., Martinelli, E., et al. (2016). Identification of mammography anomalies for breast cancer detection by an ensemble of classification models based on artificial immune system. *Knowledge-Based Systems, 101*, 60-70.

[6] Muharemi, F., Logofătu, D., & Leon, F. (2019). Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecommunication*.

[7] Satyanarayana, A., & Nuckowski, M. (2016). Data mining using ensemble classifiers for improved prediction of student academic performance.

[8] Tang, B., Kay, S., & He, H. (2017). Toward Optimal Feature Selection in Naive Bayes for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*.

[9] Shen, X., Li, Z., Jiang, Z., & Zhan, Y. (2013). Distributed SVM Classification with Redundant Data Removing. In Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom). *IEEE International Conference on and IEEE Cyber, Physical and Social Computing* (pp. 866-870). IEEE.

[10] Shen, X., Mu, L., Li, Z., Wu, H., Gou, J., & Chen, X. (2016). Large-scale support vector machine classification with redundant data reduction. *Neurocomputing, 172*, 189-197.

[11] Dong, R., Meng, H., Long, Z., & Zhao, H. (2017). Dimensionality reduction by soft-margin support vector machine. *In Agents (ICA), 2017 IEEE International Conference on Agents (ICA)* (pp. 154-156). IEEE.

[12] Anagnostopoulos, T. T., & Skourlas, C. (2014). Ensemble Majority Voting Classifier for Speech Emotion Recognition and Prediction. *Journal of Systems and Information Technology*.

[13] Wang, Y. W., & Feng, L. Z. (2018). A new feature selection method for handling redundant information in text classification. *Frontiers of Information Technology & Electronic Engineering, 19*(2), 221-234.