American Journal of Electronics & Communication

Subhalaxmi Chakraborty et. al., American Journal of Electronics & Communication, Vol. I (3), 22-29

Fake Note Detection using Machine Learning Techniques

Subhalaxmi Chakraborty, Soumyadip Banerjee, Biman Kumar Singha, SayaniGhati

Department of Computer Science and Engineering, UEM Kolkata

Abstract- In recent scenario detection of fakenote has become a genuine problem in the area of the financial sector as per the of various countries. In this paper, we have proposed a machine learning model that is capable of eradicating the fake note problem. In this paper, we have used a dataset of fake note images having a size of 1500. Hence exhaustive experiments have been conducted using various machine learning algorithms for proper authentication of the banknote. Here we considered K-Nearest Neighbour, Naive Bayes and random forest classifier technique yielding various result in terms of accuracy, precision and recall and f-score. It is observed that the K- nearest neighbour technique shows better performance compared to the other applied algorithm having an accuracy of 99%. Moreover, it is observed that it gives a result on determining whether a note is fake or real by output 0 when the note is fake and it gives output 1 when the note is real. Hence K-nearest neighbour gives there result more accurately than other classifiers. The rules are given by machine learning classifier techniques also tested and found that they are accurate enough to be used for prediction and compare their performance to see which classifier performs best on determining the fakenote and showing their performance by bar-graphrepresentation.

***Keywords* -** KNN , Naive Bayes ,Random Forests, Accuracy, Precision, Fake note

I. Introduction

In recent days machine learning techniques are tremendously used in fake note detection. Kim et al. applied both ensemble method boosting for improving the performance of traditional neural network in bankruptcy prediction and concluded that ensemble boosted neural networks performance is better than traditional neural networks [1]. The experimental method has set up using a method that Hold on method. The method is one in which the dataset is separated into subsets (70:30ratio) called the training set and testing set respectively. The training set is used to train the classifier while testing set is used to estimate the error of trained classifier [2]. A naive Bayes classifier is used when features are independent to each other in each class. It is based on estimating P(X|Y), probability of X given that Y is occurred. It classifies the process into training and prediction step [3]. Sun et al. applied support vector machine for financial distress prediction. SVM is a supervised machine learning technique that uses classification algorithm for two group classification problems. SVM helps to finding out more accurate result towards class prediction[4]. Danenas et al. applied linear SVM for credit risk evaluation to show that it performs better than logistic regression in terms of accuracy [5]. Boyacioglu et al. applied ML, K means cluster, learning vector quantization etc. Those techniques helps the model far better for banknote detection purpose [6]. Preserve genuinity of printed banknotes is one of the critical thing. It has a major role in financial activities of a country. Aoba et al. used three layered perception and euro banknote recognition with various RBF Ker tools [7]. Zhou et al. has done the performance evaluation of corporate fake note prediction by imbalanced dataset by applying sampling method. Sampling method gives more accuracy to find out the prediction about fake note [8]. Yu et al. applied Leave-One-Out incremental extreme machine learning for bankruptcy prediction and has given specific financial indicator [9]. Guangli et al. hasapplied decision tree and logistic regression to performs better. Decision tree helps to find out the prediction of fake note more accurately and gives desired output. Logistic Regression also performs well that estimates the probability of class ownership [10]. The dataset used for carrying out the experiments is taken from UCI machine learning dataset that gives huge number of entries of dataset for bankruptcy prediction with total five attributes [11]. Zhang et al. applied logistic regression and artificial neural network for bankruptcy prediction and it shows ANN performs better than LR and Logistic Regression is not that much accurate as like as Artificial Neural Network [12].Mayadevi A. Gaikwad et al. has done automatic fake currency detection technique using image processing. They have used coins, banknote and electronic data as currency and also the idea of image segmentation and characteristic extraction has been applied for achieving better performance[13].Renuka Nagpureet al. has applied image processing and java-based application that helps to recognize a bank note based on its denomination on an application window. They have also worked for various image recognition method that efficiently helps in currency recognition and fake note detection[14]. M.Deborahet al. has applied some image enhancement techniques that contains image segmentation, cropping, smoothing, contrast stretching, de-blurring and adjusting for finding fake currency. They have also used image acquisition, edge detection and Peak Signal to Noise Ratio (PSNR)technique to identify fraudulent currency [15]. Faiz M. Hasanuzzaman et al. has done banknote recognition based on robust and effective component. They have used a camera-based computer vision technology to automatically recognize banknotes for assisting visually impaired people. SURF technique is applied there to repeatability, distinctiveness and robustness of local image feature extraction in banknote recognition [16]. Mohammad H Alshayeji et. al. has done counterfeit currency detection using bit- plane slicing technique. This technique consists of decomposing original images of 256 grey levels into their equivalent 8 binary images. They have used MATLAB function for edge detection, image acquisition etc. [17].Komal Vora et. al. has done currency recognition system based on frequency domain feature extraction method and implementation of OCR. An optimal and efficient implementation of two-dimensional discrete wavelet transform to develop a currency recognition system and non-textural features are used for checking authenticity. The concept of histogram equalization is also used for better result [18].Ankush Singh et al. has done fake currency detection using image processing and cloud storage and taken images of currency for detection. They have applied a proposed solution in form of an mobile app coupled with cloud storage. Image processing algorithms are applied to extract the features such as security threat. SVM or Support Machine Vector algorithm is used for better result [19]. Devid Kumar et al. has applied computer vision technology and feature extraction in fake currency detection. They mainly have used ORB (Oriented Fast and Rotated BRIEF) and Brute-Force matcher approach to extract the feature of paper currency. Some sort of image processing techniques and feature extraction algorithms are applied to detect fake currency [20]. M. Laavanya et al. has applied deep learning and image processing technique on real time fake currency detection to avoid less efficiency and time consuming. They have used the concept of deep convolutional neural network and feature extraction. A transfer learning using Alex net that is popular in deep learning is adopted for detecting fake currency [21]. Navya Krishna et al. has applied convolutional neural network for designing Automatic Fake Currency Recognition system(AFCRS) that performs better than previously used image processing techniques. They have trained an artificial neural network and make a neural network to predict a class in that an image belongs to. VGGNet in CNN is chosen for the model to perform better [22]. Dr. P. Mangayarkarasi et al. applied image acquisition, feature extraction and comparison method for recognizing fake Indian currency note. They have produced the result in the form of text and voice by using image segmentation and edge detection technique. Brute Force Classification algorithm is used to calculate hamming distance using the descriptor that returns the point with minimum hamming distance applied on the notes [23]. Snehlataet al. used Principle Component Analysis (PCA) technique and image pre-processing for fake currency detection. PCA is used to detect the feature of currency through modelling. PCA is a method of identification of data patterns in that data are expressed in order to highlight similarities and difference. MATLAB is used for numerical computation and image processing[24].P. Gayathri et al. used texture classification and some image processing methods and convolutional neural network for fake Indian currency detection. They have applied the concept of ANN and CNN to make result and performance more accurate and efficient. The usage of image re-scaling and image shearing makes it more efficient and helps to gaining desired result [25]. The goal is to detect a note is fake or not. This thing is occurred in banking purpose and other economical aspects. In this Paper we have to take three machine learning classifiers to train a dataset and then we test a note by this classifier to check it is fake or not. The entire dataset has been divided into 70:30 ratio means 70% data is used for training purpose and remaining 30% data is used for testing purpose. The same thing is happening with those three classifiers and finally we observe which classifier predicts at its best by measuring classifier performance. We comment on best classifier and represent the bar- graph according to classifiers performance. So we used some packages like Sklearn, matplotlib and functions like Gaussian NB, RFC, KNN etc. in python programming language. Here data pre- processing is used before applying classification techniques on dataset that cleans out the entire data from missing values and null values. For pre-processing some python module has been used and some particular mean strategy is applied for pre-processing that replace all missing or null values with the value of mean of that particular column of the dataset.

II. Description of Dataset

The dataset used for fake note detection is taken from UCI dataset. This dataset has total 1372 instances. The dataset has 5 attributes. Among five first four attributes are used as input data and the last column that is used for target purpose means the output column. Those first four columns are Variance, Skewness, Kurtosis, Entropy. The last column “Class" describes the value ‘0’ and ‘1’ where ‘0’ describes the note is real against the input data and if it is ‘1’ that describes the note is fake.

Variance: - It is a measure of ‘spread’ of a distribution about its average value.

Skewness: -Skewness tells about the direction of variation of the lack of symmetry.

Kurtosis: - Kurtosis is a parameter that describes the peakof distribution.

Entropy: -Image entropy is the amount of information which must be coded for bya compression algorithm.

Class: - Class contains two values that is 0 and 1 where 0 represents real banknote and 1 represent fake note.

Excluding the ‘class’ remaining four attributes represents the input of a note that is given to check it is fake or real and the target column represents the output.

III. Method of problemsolving

This experiment is performed using a method in machine learning called classification. During classification three different types of algorithm is used to find the output. The total dataset that is used for the measurement of fake note detection is divided into two subsets in 70:30 ratio. First subset is training set and the other one is testing set. Training set is used to train the dataset to gain experience and the testing set is used to check the performance and estimate the error rate. Before applying some classification algorithm some data pre-processing is taken into consideration.

i) Datapre-processing

Some packages are used to pre-process the dataset before applying some classification technique. ‘simpleimputer' package is used for pre-processing. It is used to replace the missing or null values with the average of the values present in that column. It is used to avoid errors duringmeasurement.

ii) Performing classificationtechniques

Three classification techniques are used like Random forest, KNN and naive Bayes classifier. For each classifier dataset is divided into two subsets and then applying different techniques. Then testing the note against some input data to measure the performance of classifier and find the outcomes. After that some metric values are find out.

iii) Performancemeasure

To measure the performance of every classifier some metric values are taken into consideration like accuracy , precision , recall and F1- score.

Accuracy = (TP + TN) / (TP + TN + FP +FN)

Accuracy is defined as the percentage of correct predictions for the test data. It is calculated by dividing the number of correct predictions by the number of totalpredictions.

Precision = TP / (TP + FP)

Precision is the ratio of true positives to the total of the true positives and false positives.

Recall = TP / (TP + FN)

Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

F1-score = 2*((Precision * Recall) / (Precision + Recall))

F1-score is the harmonic mean of precision and recall and gives a better measure of the incorrectly classified cases than the accuracy metric.

TP = True Positive TN = True Negative FP = False Positive FN = False Negative

iv) PerformanceRepresentation

The all metrics from each classifier are taken and represented in bar graph which clearly determines the performance of every classification technique.

Machine Learning Classifiers Techniques

There are three different classifier techniques are used as mentioned below:

1) Random Forest Classifiertechnique

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. Like a forest have many trees, a random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the overfitting by averaging the result. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. It takes less training time as compared to otheralgorithms. For classification purpose, random forest algorithm required Gini index formula that is used to decide how nodes on a decision tree branch.

$$\text{Gini} = 1 - \sum_{i=1}^{c}(p_i)^2 \text{..............................} \quad (i)$$

This formula uses the class and probability to determine the Gini of each branch on a node, determining which of the branches is more likely to occur. Here, pi represents the relative frequency of the class that is observing in the dataset and c represents the number of classes. Entropy is also used to determine how nodes branch in a decision tree.

$$\text{Entropy} = \sum_{i=1}^{c} -p_i * \log_2(\, p_i\,) \text{ ......} \quad (ii)$$

2) *Naive Bayes Classifiertechnique*

Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' Theorem**. It is not a single algorithm but a family of algorithms where all of them share a common principle, means every pair of features being classified is independent of each other. A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem. It is mainly used in *text classification* that includes a high- dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.** The formula of Baye's theorem is given by :

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \text{.......................} \quad (iii)$$

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing theevidence.

P(B) is Marginal Probability: Probability of Evidence.

The formula of Baye's theorem in following way:

$$P(y \mid X) = \frac{P(X \mid y)P(y)}{P(X)} \text{ .... (iv)}$$

Here, y is class variable and X is a dependent feature vector where: X = (x1, x2 , x3…….. , xn)

$$P(y \mid x_1,.., x_n) = \frac{P(x1|y)P(x2|y)......P(xn|y)P(y)}{P(x1)P(x2)....P(xn)} \text{ ....(v)}$$

That can be expressed as:

$$P(y \mid x_1,……, xn) = \frac{P(y)\prod_{i=1}^{n} P(xi|y)}{P(x1)P(x2)....P(xn)} \text{ .... (vi)}$$

3) KNN classifiertechnique

KNearestNeighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. KNN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using KNN algorithm.It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. For performing KNN algorithm Euclidean distance is taken to find out distance between testing data point to each and every training points.

Distance is,

$$D = \sqrt{\sum_{i=1}^{k}(xi - yi)^2} \text{..............(vii)}$$

Here, k is the distance between k data points.

4) Data Pre-processing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and important step while creating a machine learning model. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. At first the dataset is taken at the format of .csv file. That dataset may contain some missing values and noises which may affect the overall calculationsandotheraspects. Somepackageisusedfordata pre-processing like “simple imputer”. Pre-procesing refers to the transformations applied to our data before feeding it to the algorithm.

In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.For achieving better results from the applied modelin Machine Learning Papers the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null valueshave tobemanagedfromtheoriginalrawdataset.Aotheras pectisthat data setshould be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen. Mean strategy is taken to fill out those missing values to make the dataset without null values that prevents various kinds of data hazards and data duplicity.

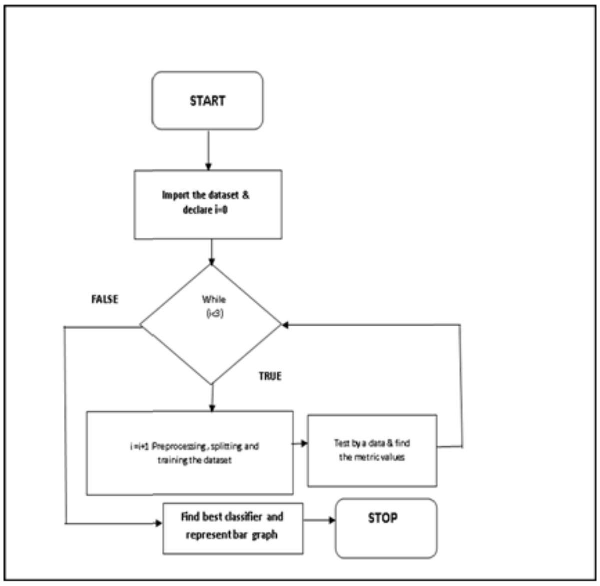

Fig. 1. Flow Chart for Fake note Detection

algorithm

IV. Result andDiscussion

After performing three types of classification algorithm on the dataset the performance of those classifiers are different from each other. The result of three classifiers is shown below. From the below mentioned table, it is seen that KNN classifier gives highest accuracy 99.76% that is higher than other two classification technique. The Naïve Bayes Classifier have the ccuracy of 85.92% and Random Forest classifier has 99.27% accuracy that is nearest to the KNN classifier.F1-score value is found out from accuracy, precision and recall that determines which classifier performs best. From the result of those different classifier techniques a bar graph is obtained that tells which classifier performs best.

Table: I: Comparative analysis of various classification techniques

Technique name	Accuracy	Precision	Recall
RandomForest classifier	99.27%	98.85%	99.42%
Naïve Bayesclassifier	85.92%	84.22%	82.29%
KNN classifier	99.76%	99.50%	99.8%

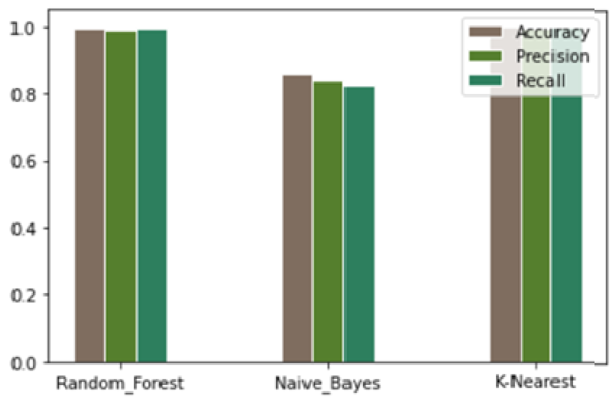

Fig . 2. Performance Measurement of various classification techniques in graphical representation

The bar graph represents the performance percentage ofRandom Forest, Naïve Bayes and KNN classifier in terms of accuracy, precision and recall. It has a scale between 0 to 1 with having an interval of 0.2 that helps to measure the performance metrics of each classifier. Three different colours are associated for accuracy, precision and recall. This graph helps to visualize the performance of each classifier and also helps to determine the classifier that performs best. Those three classifiers give the same output against particular input data of a given note and according to accuracy and other metrics they perform differently. The four input values thoseare separated by comma represents the values of four attributes of a note. Those are Variance, Skewness, Kurtosis and Entropy respectively. Output represents the prediction of the class ofnotedepending on four input values. The class is 1 that determines the note isfake. According to the bar-graph, the performance of KNN classifier is best than other two.

V. Conclusion and Future Scope

In this proposed work of fake note detection using various machine learning classification technique we analysed on the basis of Random Forest, Naïve Bayes, KNN. So, it works fine against all input of banknote and gives the expected result that whether the note is fake or real. Moreover, we determine which classifier performs better by finding metric values like accuracy, precision and recall represent the bar-graph by those values with respect to the performance of classifiers. In this regard, we obtained that KNN classification shows better result terms of accuracy compared to other models used, based on same fake note dataset. This work can be extended in the field of latest technologies like deep neural network, texture classification, convolutional neural network, information retrieval etc.

References

[1] Myoung-Jong-Kim, Dae-Ki Kang, Ensemble with neural networks for bankruptcy prediction, Expert Systems with Applications 37 (2010) 3373-3379.
[2]https://raw.githubusercontent.com/lucko515/classification-python/master/Banknotes%20Authentication%20Classification/data_ anknote_authentication.csvaccessed on (5/5/2020).
[3] Anamika Yadav, Aleena Swetapadma, Combined DWT and Naive Bayes Based Fault Classifier for protection of Double Circuit Transmission Line, IEEE International Conference on Recent Advances and Innovations in Engineering(ICRAIE-2014), May 09- 11, 2014, Jaipur India.
[4] Jie Sun, Hui Li, Financial distress prediction using support vector machines: Ensemble vs. Individual, Applied Soft Computing 12(2012) 2254-2265.
[5]PauliusDanenas, Gintautas Garsva, Selection of Support Vector Machines based classifiers for credit risk domain, Expert Systems with Applications 42(2015) 3194-3204.
[6] MelekAcarBoyacioglu, Yakup Kara, Omer KaanBaykan, Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund(SDIF) transferred banks in Turkey, Expert Systems with Applications 36(2009) 3355-3366.
[7] M. Aoba, T. Kikuchi, Y. Takefuzi, Euro banknote recognition system using a three layered perceptron and rbf networks, IPSJ Transactions on mathematical modelling and its applications,vol.44(may 2003) No. SIG 7(TOM 8). Machine learning and cybernetics,2260-2264.
[8] Ligang Zhou, Performance of corporate bankruptcy prediction models on imbalanced dataset. The effect of sampling methods, Knowledge-Based Systems 41(2013) 16-25.
[9] Qi Yu, YoanMiche, Eric Severin, Amaury Lendasse, Bankruptcy prediction using extreme learning machine and financial expertise, Neurocomputing 128 (2014)296-302.
[10] GuangliNie, Wei Rowe, Lingling Zhang, Yingjie Tian, Yong Shi, Credit card churn forecasting by logistic regression and decision tree, Expert System with Application 38(2011) 15273-15285.
[11] https://www.kaggle.com/ritesaluja/bank-note-authentication-uci- dataaccessed on (5/5/2020).
[12] Guoqiang Zhang, Michael Y. Hu, B. Eddy Patuwo, Daniel C. Indro, Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, European Journal of Operation Research 116 (1999) 16-32.
[13] Gaikwad, Mayadevi A., Vaijinath V. Bhosle, and Vaibhav D. Patil., "Automatic Indian New Fake Currency Detection.",International Journal of Engineering Research & Technology (IJERT), Vol. 6, Issue 11, November, 2017.
[14] Renuka Nagpure, Shreya Sheety, Trupti Ghotkar, "CurrencyRecognition and Fake Note Detection", IJIRCCE, vol. 4, 2016.
[15] M. Deborah and SoniyaPrathap ―Detection of Fake currency usingImage Processing. IJISET- International Journal of InnovativeScience, Engineering & Technology, Vol. 1, Issue 10, 2014.
[16] Faiz M. Hasanuzzaman, Xiaodong Yang, and YingLi Tian, SeniorMember, IEEE Robust and Effective Component-based BanknoteRecognition for the Blind IEEE Trans Syst Man Cybern C ApplRev. 2012 Nov; 42(6): 1021–1030.
[17] Mohammad H Alshayeji, Mohammad Al-RousanandDunyaT.Hassoun, Detection Method for Counterfeit Currency Based on Bit Plane Slicing Technique, International Journal of Multimedia andUbiquitous Engineering Vol.10, No.11 (2015).
[18] Komal Vora, Ami Shah, Jay Mehta, A Review Paper on CurrencyRecognition System, International Journal of ComputerApplications (0975 – 8887) Volume 115 – No. 20, April 2015.
[19] Ankush Singh, Prof. Ketaki Bhoyar, Ankur Pandey, Prashant Mankani, Aman Tekriwal – Detection of Fake Currency using Image Processing, IJERT, vol. 8, Issue 12, 2019.
[20] Devid Kumar, Surendra Chauhan – Indian Fake Currency Detection using Computer Vision, IRJET, vol. 7. Issue 5, 2020.

[21] M. Laavanya, V. Vijayaraghavan- Real Time Fake Currency Note Detection using Deep Learning, IJEAT, vol. 9, Issue 1S5, 2019.
[22] Navya Krishna G, Sai Pooja G, Naga Sri Ram B, Yamini Radha V, Rajarajeshwari P- Recognition of Fake Currency Note using Convolutional Neural Networks, IJITEE, vol. 8, Issue 5, 2019, 2278-3075.
[23] Dr. P. Mangayarkarasi, Akhilendu, Anakha A S, Meghashree K, Faris A B – Fake Indian Currency Note Recognition, IRJET, vol. 7, Issue 5, 2020.
[24] Snehlata, Vipin Saxena – An Efficient Technique for Detection of Fake Currency, IJRTE, vol. 8, Issue 3, 2019, 2277-3878.
[25] P. Gayathri, P. Soniya, V. YaminiPriya, G. Srinivas, V. Ravindra – Texture Classification for Fake Indian Currency Detection, IJERT, vol. 9, Issue 6, 2020, 2278-0181.