Effective Soil Type Classification Using Convolutional Neural Network

International Journal of Informatics and Computation (IJICOM)
Vol. 3, No.1, Agustus 2021
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI : 10.35842/ijicom

Effective Soil Type Classification Using Convolutional Neural Network

Hamzah[1], Mohammad Diqi[2] Antomy David Ronaldo[3]

Abstract

Soil classification is a growing research area in the current era. Various studies have proposed different techniques to deal with the issues, including rule-based, statistical, and traditional learning methods. However, the plans remain drawbacks to producing an accurate classification result. Therefore, we propose a novel technique to address soil classification by implementing a deep learning algorithm to construct an effective model. Based on the experiment result, the proposed model can obtain classification results with an accuracy rate of 97% and a loss of 0.1606. Furthermore, we also received an F1-score of 98%.

Keywords:
Soil, Classification, Deep Learning, CNN

1. Introduction

Soil has a significant role in the life of living things and plants. Soil has different properties and characteristics, such as red, black, yellow ground. However, for the chemical and physical elements of the earth, we use the LIBS method to analyze the soil and present the application of LIBS in soil characteristics. The LIBS system has a unique model for different soil types to determine the yield of nutrients and toxic elements in soil samples. Within a few years, the development of LIBS techniques for soil analysis. The process of the LIBS method is straightforward, so the ability to detect soil type is less efficient [1].

The features of the soil are crucial in analyzing and delivering helpful information. The texture of the earth and the color of the ground are traditionally determined in a variety of ways. Multiple ways to choose soil type have employed soil texture in other attributes. As a result, the process of soil categorization and qualitative distinction the based on soil color constancy. The paper the USDA triangle is less effective for classifying because the process takes energy and time. Hence, researchers use many methods based on computer vision and image processing for soil classification [2].

Traditionally, the development and classification of soils with the current need to measure and interpret changes in soil function. Some of the main properties of soil, such as carbon, nitrogen, and pH. Soil surveys to classify the origin of the ground take time and effort. The overall accuracy according to the number of soil type samples the predicted correctly. To ensure the reliability of this model per soil type by analyzing the matrix and accuracy of the predictions. Soil variable analysis uses the survey method, and soil classification can use cluster analysis. The result accuracy level is so low that both approaches are less efficient in classifying [3].

Corresponding Author:
1 Antomy David Ronaldo[1], Universitas Respati Yogyakarta,
2 Hamzah[2], Universitas Respati Yogyakarta (hamzah@respati.ac.id)
3 Mohammad Diqi[3], Universitas Respati Yogyakarta,

Soil classification is essential for determining potential use, land restrictions, and adequate management for each condition. Such information helps to determine the most intense use of soil and know the level of soil quality. However, portable X-ray fluorescence spectrometers and infrared spectroscopy have increasingly been applied to soil type and attribute predictions. The prediction model has been good, but the accuracy level is less than perfect [4].

Previous research on soil classification does conduct online using the UAS sensor placement system. These sensors can be deployed in difficult-to-reach locations, allowing for better data collecting. A paper developed a sensor system for soil classification. However, the UAS sensor system is less effective because it needs to the redesigned, so it takes a lot of time to fix it [5]. Another paper proposed Cone Penetrometer Test (CPT), but the trial using this method is still difficult to recognize some types of soil. The category of CPT method cannot provide an accurate prediction of soil type classification, so it is a poor result.[6]. A study also explored an automated system to classify the soil, namely the classification system using mining methods. But the technique is not suitable because it only knows the type of soil only [7].

To deal with soil classification, several researchers explored machine learning algorithms to deal with soil classification. An article proposed a method to deal with soil classification using the SVM by extracting five different soil types. However, the study remains shortcomings when training a tiny dataset. In traditional learning, the statistical category requires a much larger dataset to affect accuracy [8]. Learning models need to measure the accuracy and loss scores by tuning different hyperparameters. Moreover, some conventional learning method remains drawback to obtaining a better accuracy [9].

The current paper proposes various techniques to address soil classification. A paper introduces method algorithm CNN, compression spectral imaging system to classify soil. The experiment's findings suggest that this algorithm can speed up feature discriminability and soil categorization [10]. A paper method algorithm DNN model uses to classify soil type show good performance [11]. A paper method algorithm RNN, this model looks pretty effective in complex modeling to know the results of soil classification. In this study, despite showing significant variations, the RNN model displays a strong ability in soil classification [12].

Therefore, we introduce a classification model using deep learning to create soil type classification issues to solve the problem. In the soil type classification problem, this study has some contributions as follows:

1. We design a classification model to determine the soil type. Instead of using conventional models, we created models using deep learning methods to analyze soil types. Based on the classification result, our proposed model can achieve the best and most efficient results.
2. We evaluate the model to determine better and more efficient results on the classification of soil types. Then, we present an evaluation of the metrics to prove the quality of the model created. This study the conducted using al large datasets to produce classification models learning algorithms.
3. We test the proposed model to achieve high accuracy results to quickly and accurately classify soil types. To achieve the best training model classifier, we also set many parameters to obtain the best accuracy value.

Organization: The remainder of this paper the written as follows: Part II delves deeper into related research. Part III explains how this study problem the defined. Part IV describes the experimental setup, including feature learning methodologies, datasets, and data pre-processing, while Part V gives the study's findings and extensive analysis. Part VI summarizes the findings and identifies several unsolved issues in soil classification research.

2. Related Works

In soil classification problems, various articles have proposed models to classify better to know the accuracy level. A study introduced an analysis of the main components in grouping soil types to help create a soil type classification model [13]. A soil classification system facilitates soil identification and predicts accuracy, strong predictive capacity, simplicity, precision, and adaptability to classification systems [14]. Characteristics to provide an information system for soil classification are very useful as a medium of learning. This study dataset used 150 soil samples to be analyzed using USDA Soil Taxonomy to obtain results that show the most common soil orders are Aridisols, Inceptisols, Entisols, Vertisol, Mollisols, and Alfisols [15].

The soil classification system a used to group soil types into categories based on soil properties [16]. This paper presents the results of using near-infrared and visible light (VIS-NIR) spectroscopy to communicate soil and soil profile examination [17]. To classify the soil used 291 soil samples. Based on the validation findings obtained by 67 % using the best vis-nir categorization method. The results obtained are still not so maximal because the final product is still low [18].

A paper proposed a learning algorithm using SVM to classify soil type [19]. The proposed system has added a soil prediction feature based on color and texture on the ground [20]. Soil classification using the SVM method using dataset consists of 175 soil samples containing sandy clay, peat, humus clay, clay sand, clay peat, and clay. Based on testing using the SVM algorithm, the highest level of accuracy achieved 95% [21].

Classification of soil according to soil nutrition is helpful to predict soil type. A study proposed a technique to determine the accuracy level using the KNN [22]. Machine learning models the trained to classify soil based on soil texture. The KNN algorithm method is still used to this day, namely to train soil classification data [23]. A paper used data from 383 samples to test into the KNN algorithm, from the classifying result obtained the best accuracy of 94% [24].

The current paper explored deep learning to establish a classification model to improve accuracy in image classification [25]. A report presented an RNN algorithm method to improve classification performance to obtain accurate results [26]. Another study introduced the GRU-RNN way to classify soils with pH and moisture values. Test results using LSTM and GRU-RNN methods received classification results of 0.920 for the LSTM method and 0.957 for the GRU-RNN process [27].

Soil is the main component for classification, and soil color describes soil attributes. A paper developed a more efficient algorithm for detecting soil color using the KNN algorithm. In this algorithm, the results obtained are good enough to classify the soil type based on soil color [28]. In the following paper, the classification of soil uses CNN algorithms. This algorithm the used to study the category of soil texture based on hyperspectral data. The result of this classification model reaches 70%, so it is less effective to determine the effect because the model made is still weak to classify the soil texture [29].

Based on research from several models of soil classification, there are still deficiencies in the level of curation. Therefore, we propose a model that can classify soil types using CNN because the CNN method has a high degree of accuracy in organizing soil types.

3. Proposed Method

Based on our review, CNN is a growing algorithm to deal with various problems in computer science [35][36][37]. The study focused on soil type classification using CNN algorithms based on features on datasets that include black soil and red soil. In this study, the dataset consisted of training data and testing data. A digital image can then be represented by a matrix consisting of M= columns and N= rows. Coordinates and intensity or color are the two characteristics that define pixels. The coordinates (x,y) hold the value f. (x,y) [30]. Therefore, can write images into a matrix:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & f(0,M-1) \\ f(1,0) & f(1,1) & f(1,M-1) \\ f(N-1,0) & f(N-1,1) & f(N-1,M-1 \end{bmatrix}$$

Based on the above formula, an image of $f(x,y)$ can write into mathematical functions such as the following:

Table 1: Mathematic notation

Notation	Description
X	training data
Y	testing data
M	number of row pixels in the image array
N	number of column pixels in the image array
G	grayscale value

In this study, we adopt the CNN algorithm to classify the soil type. The CNN algorithm is a Multiplayer Perceptron (MLP) development method designed to manage two-dimensional data images. The CNN algorithm was employed to classify soil types in this investigation. Several parts of CNN consist of three layers in the CNN architecture: the input layer, hidden layer, and output layer. The Convolution Layer is part of the stage in CNN architecture. This stage aims to convolute the picture data so that features from the input image may extract. Convolution produces a linear transformation of the input data based on the data's comprehensive information. The revolution kernel is specified by the layer's weight, allowing the convolution kernel to learn using CNN inputs [31].

Convolutional Layer consists of neurons arranged to form a filter with length and height (pixels). I will shift this filter to all parts of the image. Each shift will perform a "dot" operation between the input and the value of the filter resulting in an output.

$$Y_k = f(W_k * x) \tag{1}$$

Table 2: Mathematic notation

Notation	Description
Y_k	Output feature map
x	Input image
W_k	Convolutional filter
$f(.)$	Nonlinear activation function

A Pooling Layer is a layer that uses a function with a feature map as input and processes it with a variety of statistical operations based on the closest pixel value. Pooling layers does use to retrieve the maximum or average value of the pixel part of the image. We implement max pooling and average pooling methods with sizes 2x2 and stride 2. In the max-pooling method, the retrieved value is the largest in the 2x2 area, and the average pooling will take the average value of each feature map.

$$(n_h - f + 1)/\, s \, * \, (n_w - f + 1)/s \, * \, n_c \qquad (2)$$

Table 3: Mathematic notation

Notation	Description
n_h	Height of feature map
n_w	Width of the feature map
n_c	Number of channels in the feature map
f	Size of filter
s	Stride length

A Fully Connected Layer map of the map generates in the previous stage in a multidimensional array. The feature map will be "leveled" or reshaped before entering the fully connected layer stage, resulting in a vector utilized as input from the completely corresponding layer [32].

$$FC(W,x)_{(ij)} = \sum_k^K W_k \ .(r.x_{(ij)}) + b \qquad (3)$$

Table 4: Mathematic notation

Notation	Description
FC	Fully connected
W	Matrix of convolutional kernels
x	Input feature map
i,j	Embodies a receptive field around the position
k	Dimensions convolutional kernel
W_k	Convolutional filter
r	Variable
b	Trainable

4. Experimental Setup

1. Main Idea

   The main goal of this paper is to create a color-based classification model to classify soil types using CNN's algorithm. We model the data and analyze the data generated by CNN's algorithm. The purpose of this method is to organize the soil type. CNN's algorithms are also often used to identify objects. The high level of accuracy achieved by CNN's algorithm is suitable for handling classification difficulties [33].
2. Dataset

   In this study, we gather a dataset from a benchmark. To train our model, we separate the dataset into two, namely red soil and black soil. There are two elements to a dataset: training data and testing data. Data training is data with a class, while data testing is the data that will search for the type. The training data is used to build classification models while testing data is used to measure the performance of classification models.

Table 5. Soil dataset in this experiment

Dataset Label	**Soil features**	
	Training	Testing
Red Soil	212	47
Black Soil	184	48

3. Data Pre-Processing

   At this stage, we prepare a dataset of soil types taken from benchmarks. The dataset is raw data that does not produce good accuracy when applied to any classification system because it validation from the results. Our goal is to demonstrate some pre-processing techniques, namely mean normalization, standardization, and zero component analysis. We adopt three pre-processing methods to turn every piece of information in a data set into a vector for the model to understand [34].

4. Classification Method

   We trained the model that extracted the feature after the pre-processing stage to build the soil classification model. We created a model using deep learning to train a dataset consisting of black soil and red soil data. In this study, we designed a model with CNN to study the results of soil type classification. To optimize the training model and value, we tuned the hyperparameter on CNN to produce an optimal classification model.

   In the training process, we do pre-process. First, we prepare a soil type dataset as a material for classifying. Then, the raw dataset used the pre-processing method to convert the dataset into a vector. We use three pre-processing techniques, namely mean normalization, standardization, and zero component analysis.

   For the testing phase, we use vectors as inputs for feature extraction. There are two components of the dataset, namely dataset training and dataset testing. Then, we do model testing using training datasets from our classification to get maximum results.

5. Result & Analysis

In the classification process, our approach can achieve a classification model with the CNN method. In training, we check the data set using three optimizers, namely Adam, RMSProp, and SGD. We also use the loss function to estimate losses, compare, and measure the predicted result's good or bad. Therefore, when modeling the trained, the interconnection weight of neurons will gradually match until a good prediction the obtained.

In this study, we train our model by setting several parameters to obtain the highest level of accuracy. We set epoch = 80, batch size = 32, and learning rate= 0.6 during the training and testing phase.

Table 6. Training loss and training accuracy result with various optimizer setting

Hyperparameter	Optimizer	Training Loss	Training Accuracy
Epoch = 80 Batch size = 32 Learning Rate = 0.6	Adam	0.1877	0.9800
	RMSprop	0.3040	0.9276
	SGD	0.6784	0.6479

Table 7. Testing loss and testing accuracy result with various optimizer setting

Hyperparameter	Optimizer	Testing Loss	Testing Accuracy
Epoch = 80 Batch size = 32 Learning Rate = 0.6	Adam	0.0815	0.9784
	RMSprop	0.2699	0.9139
	SGD	0.6905	0.5268

In table 6, we have conducted training on the CNN model that has the created. We use three optimizers, namely Adam, RMSprop, and SGD. All three optimizers the used to find out how accurately the model the made. We conducted tests on training data using optimizer Adam obtained results on training loss = 0.1877 and training accuracy = 0.9800. Next, we conducted a test using RMSProp optimizer got the results of training loss = 0.3040 and training accuracy = 0.9276. then we also do tests using SGD optimizer obtained training loss results = 0.6784 and training accuracy = 0.6479.

In table 7, we test the model that we have created using three optimizers. Our first optimizer conducted testing using adam optimizer results obtained testing loss = 0.0815 and testing accuracy = 0.9784. Our second optimizer uses RMSprop results obtained testing loss = 0.2699 and testing accuracy = 0.9139. Our third optimizer conducted a test using the SGD optimizer obtained test loss results = 0.6905 and testing accuracy = 0.5268.

We calculate Accuracy, Precision, Recall, F1 Score, Support, and Confusion Matrix to assess the model's performance in evaluation measures. This study estimates accuracy predicts the accuracy of the overall data on soil type. Precision describes the accuracy between the requested data and the predictive results provided by the model. Recall defines the success of a model in finding information. F1 Score for comparison of average precision and recall weighted, we can use as a reference to determine accuracy.

Table 8. Classification terms for the statistical measures for the Soil dataset.

Classification Report	Precision	Recall	F1-Score	Support
Black_Soil	0.96	0.98	0.97	47
Red_Soil	0.98	0.96	0.97	46
Accuracy	-	-	0.97	93
macro avg	0.97	0.97	0.97	93
weighted avg	0.97	0.97	0.97	93

Our proposed model can achieve a TP score and the highest TN score based on confusion matrix calculations. The picture below shows Confusion Matrix using CNN algorithm obtained TP = 46 and TN = 44. Based on Confusion Matrix, the results received better and effective in soil classification.

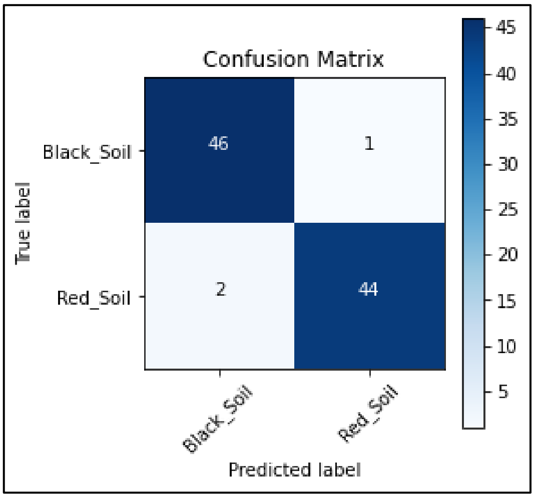

Fig 1: the Confusion Matrix of the Soil Classification model.

6. Conclusion

Soil classification is a growing research area in the current era. Various studies have proposed different techniques to deal with the issues, including rule-based, statistical, and traditional learning methods. However, it takes a lot of time and effort to classify the soil type. The study used CNN to build a soil type classification model to produce a higher accuracy with tiny loss. Using the proposed model, we used CNN's algorithm because classifying the soil type does not take time and effort to determine classification results.

Based on the experiment results, we achieve trade-offs between accuracy and performance time by adjusting the hyperparameter to optimize model performance. We set epoch = 80, batch size=32, and learning rate=0.6. In the training process, the model can produce an accuracy = 98% and loss = 0.1877. The classification model can be a promising solution to address soil type classification. In addition, our propose model can get TP = 46 and TN = 44.

As a direction for future research, more research into semantic networks to train nodes could be beneficial. Exploration of picture categorization is essential to generate improved predictions. The next can utilize GAN or GCN algorithm to improve the classification result.

Acknowledgment

This study is conducted under the Department of Informatics. Respati University of Yogyakarta, Indonesia.

References

[1] Paulino R.V.B, et al., “Applications of laser-induced breakdown spectroscopy for soil characterization, part II: Review of elemental analysis and soil classification,” European Journal of Soil Science, Vol. 71, issue 5, pp. 805-818, 2019.

[2] Srivastava, Pallavi, Aasheesh Shukla, and Atul Bansal., "A comprehensive review on soil classification using deep learning and computer vision techniques," Multimedia Tools and Applications, pp. 1-28, 2021.

[3] Fiona M. Seaton, et al., “Soil health cluster analysis based on national monitoring of soil indicators,” European Journal Soil Science, 2020.

[4] Benedet et al, “Soil subgroup prediction via portable X-ray fluorescence and visible near-infrared spectroscopy.” Geoderma, 365, 114212, 2020.

[5] Plowcha A, et al. “Online Soil Classification Using a UAS Sensor Emplacement System.” Springer, pp 174-184, 2020.

[6] Yin K. S, et al., “Marine soil behavior classification using piezocone penetration test (CPT) and borehole records,” Canadian Geotechnical Journal, Vol. 58, No. 2, pp. 190-199, 2021.

[7] Priyanka C. J, et al., “An Analytical Approach for Soil and Land Classification System Using Machine Learning,” International Journal of Scientific Research in Science, 2020.

[8] Amol D Vibhute, et al., “Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data,” International Conference on Man and Mchine Interfacing (MAMI), 2015.

[9] Zhai, Y, et al,. “Soil texture classification with artificial neural networks operating on remote sensing data,” Computers and Electronics in Agriculture, Vol. 54, No. 2, pp. 53-68, 2006.

[10] Yu, Yue, et al. "Compressive spectral imaging system for soil classification with three-dimensional convolutional neural network." Optics Express, Vol. 27, No.16, pp. 23029-23048, 2019.

[11] Ou, Depin, et al. "Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction." Geoderma, Vol. 385, No. 114875, 2021.

[12] Zhu, Jian-Hua, Musharraf M. Zaman, and Scott A. Anderson. "Modelling of shearing behavior of a residual soil with recurrent neural network," International journal for numerical and analytical methods in geomechanics, Vol. 22, No. 8, pp. 671-687, 1998.

[13] Bellinaso, H., Demattê, J. A. M., and Romeiro, S. A., “Soil spectral library and its use in soil classification,” Revista Brasileira de Ciência do Solo, Vol. 34, No. 3, pp. 861-870, 2010.

[14] Moreno-Maroto, J. M., Alonso-Azcárate, J., and O'Kelly, B. C., “Review and critical examination of fine-grained soil classification systems based on plasticity,” Applied Clay Science, 200, 105955, 2021.

[15] Mohammed S, et al., “Characterization of major soil orders in Syria,” Eurasian Soil Science, Vol. 53, pp. 420-429, 2020.

[16] Park, J., and Santamarina, J. C., “Revised soil classification system for coarse-fine mixtures,” Journal of Geotechnical and Geoenvironmental Engineering, Vol. 143, No. 8, 2017.

[17] Debaene, G., et al., “Visible and near-infrared spectroscopy as a tool for soil classification and soil profile description,” Polish Journal of Soil Science, Vol. 50, No. 1, 2017.

[18] Vasques, G. M., et al., “Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths,” Geoderma, Vol. 223, pp. 73-78, 2014.

[19] Harlianto, P. A., Adji, T. B., and Setiawan, N. A., “Comparison of machine learning algorithms for soil type classification,” International Conference on Science and Technology-Computer (ICST) IEEE, pp. 7-10, 2017.

[20] Abimala, T., S. Flora Sashya, and K. Sripriya., "Soil Classification & Crop Suggestion based on HSV, GLCM, Gabor Wavelet Techniques and Decision Tree Classifier in Image Processing," EasyChair, 3544, 2020.
[21] Srunitha, K., and S. Padmavathi., "Performance of SVM classifier for image based soil classification." International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), IEEE pp. 411-415, 2016.
[22] Abraham, et al., "Classification of soils into hydrologic groups using machine learning," Data, Vol. 5, No.1, 2020.
[23] Saranya, N., and A. Mythili., "Classification of Soil and Crop Suggestion Using Machine Learning Techniques," Engineering Research & Technology, pp. 671-673, 2020.
[24] Rahman, et al., "Soil classification using machine learning methods and crop suggestion based on soil series," International Conference of Computer and Information Technology (ICCIT) IEEE, pp. 1-4, 2018.
[25] Li, Xiufang, et al., "Pixel Dag-recurrent neural network for spectral-spatial hyperspectral image classification," IGARSS IEEE International Geoscience and Remote Sensing Symposium, pp. 2718-2721. IEEE, 2019.
[26] Zhang, X., et al., “Spatial sequential recurrent neural network for hyperspectral image classification,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 11, No. 11, pp. 4141-4155, 2018.
[27] Yang, Chenghao, Tiantian Wang, and Jing Liang., "Soil pH and Humidity Classification Based on GRU-RNN Via UWB Radar Echoes," International Conference in Communications, Signal Processing, and Systems. Springer, Singapore, pp. 2124-2131, 2019.
[28] Maniyath, S. R., et al., “Soil color detection using Knn classifier,” International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C) IEEE, pp. 52-55, 2018.
[29] Riese, F. M., and Keller, S., “Soil texture classification with 1D convolutional neural networks based on hyperspectral data,” arXiv preprint arXiv, Vol. 1901,04846. 2019.
[30] Kang, Chenyao, and Chu He., "SAR image classification based on the multi-layer network and transfer learning of mid-level representations," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1146-1149, 2016.
[31] Rawat, Waseem, and Zenghui Wang., "Deep convolutional neural networks for image classification: A comprehensive review," Neural computation, Vol. 29, No. 9, pp. 2352-2449. 2017.
[32] Shang, Ronghua, et al. "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," Knowledge-Based Systems, Vol. 194, 105542, 2020.
[33] Zhou, Xiaokang, Yue Li, and Wei Liang., "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2020.
[34] Pal, Kuntal Kumar, and K. S. Sudeep., "Preprocessing for image classification by convolutional neural networks," IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1778-1781, 2016.
[35] Wanda, P. and H. Jie. “DeepSentiment: Finding Malicious Sentiment in Online Social Network based on Dynamic Deep Learning.”, 2019.
[36] Wijaya, Nurhadi., “Evaluation of Naïve Bayes and Chi-Square performance for Classification of Occupancy House,” International Journal of Informatics and Computation, [Sl], v. 1, n. 2, p. 46-54, Feb. 2020. ISSN 2714-5263
[37] Diqi, Mohammad, Muhdalifah, Mujastia., “Design and Building Javanese Script Classification in The State Museum of Sonobudoyo Yogyakarta,” International Journal of Informatics and Computation, [Sl], v. 1, n. 2, p. 35-45, Feb. 2020. ISSN 2714-5263.