JOURNAL OF DATA SCIENCE | Vol.2023:02
eISSN:2805-5160

Features Selection in the Proposed Draft Sheet C1 for General Elections in Indonesia

Erin Efriansyah[1], Tri Basuki Kurniawan Kurniawan[1*], Edi Surya Negara[2] and Muhamad Akbar[3]

[1] Magister of Information Technology, University of Bina Darma, Palembang, Indonesia
[2] Computer Science, University of Bina Darma, Palembang, Indonesia
[3]Faculty of Engineering Science, University of Bina Insan, Lubuklinggau, Indonesia

Email: tribasukikurniawan@binadarma.ac.id

Abstract

Elections in Indonesia are held every five (5) years. In 2019, the elections were held simultaneously so that general elections could be carried out efficiently because it reduces time wastage. In the implementation of the General Election, using sheet C1 as a sheet to fill in the calculation results, the obstacles that often occur in the implementation of the election are often technical problems such as filling out the C1 sheet, which is still manual, and error input values. This study aims to create a new approach using the features in the C1 draft proposal for general elections to reduce mistakes and manual voting considerations. This study uses an artificial neural network method as a feature of the numerical value prediction process from the proposed new C1 sheet. The neural network method used is the Backpropagation technique, where the machine will recognize each number so that the calculation process will be faster. With the election feature, the C1 design for this gets the results that this feature can detect the writing of numbers accurately with an accuracy rate of 98%.

Keywords

Election, New design C1 sheet, Artificial Neural Network, Backpropagation

Introduction

Sheet C1 is required in the implementation of general elections in Indonesia. Form C1 is a record that contains the results of TPS vote counting (Polling Places). After the C1 form is recorded at the TPS, it is recapitulated to the next level in stages. Errors are frequently discovered in the input process for acquiring votes.

This error is caused by several factors, including human error, such as an unclear reading of the writing on sheet C1, resulting in a multi-interpreted and invalid data reading. Technological innovation is required to reduce this error and improve the accuracy and validity of voting results. Furthermore, technological innovation is necessary to accelerate the election implementation

Submission: 20 February 2023; **Acceptance:** 21 March 2023

Copyright: © 2023. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

process and use time better. Using the pattern recognition feature to identify the data in the C1 form is one of the technological innovations that can be implemented in inputting data to calculate the voting results.

Recognizing objects pattern recognition is a branch of science that classifies and quantifies the quantitative value of an object's features (characteristics) or main properties (Pangaribuan & Sagala, 2017). Pattern recognition aims to determine groups (Syafitri, 2011) or pattern categories based on the features and characteristics of these patterns or to distinguish an object from other objects (Romario, Ihsanto, & Kadarina, 2020). Octariadi (2020) used backpropagation as a research method to solve problems discussed in his paper. It researches particularly in fields involving grouping and pattern recognition. Furthermore, Herman, Syafie, & Indra (2018) research on Handwritten Number Recognition Using Artificial Neural Networks yielded an 11.67% accuracy.

Human handwriting patterns take many forms, as explained in Rosnelly's (2018) research entitled Recognizing handwritten numeral patterns on checks using Neocognitron, so it will be a problem if human handwriting cannot be recognized. We need a pattern recognition feature for this. Putra (2018) explained pattern recognition in her research Detection and Identification of Motor Vehicle Number Plates Using Backpropagation Neural Networks, which groups numeric and symbolic data (including images) automatically by machines (computers). The goal of grouping is to recognize objects based on the characteristics that these objects/patterns possess.

A pattern is a defined entity identified by its characteristics (features). These characteristics differentiate a pattern from others (Putra, 2018). The image (image) is captured with a camera and then subjected to a feature extraction process (PRASASTI et al., 2020) to measure specific features and properties. Mellyssa's (2019) study, Recognizing Banknote Denominations, employs a Backpropagation Artificial Neural Network to detect a value from a number on a nominal banknote with an accuracy rate of 78.8%.

As a result, the researcher conducted this research on sheet C1, which initially used handwriting, and the proposed C1 sheet uses a pattern-shaped number provision. The researcher obtained samples of C1 sheets from the KPU vote count website and created a new proposed C1 sheet of 100 sheets. Furthermore, the data from the KPU website's vote counting sheet is transferred to the new C1 sheet template by writing in a predetermined pattern and shading with various writing instruments. Colored markers, pencils, and pens to test whether stationery with different numbers will still be recognized when processed are used.

Methodology

According to Sugiyono (2014), the research method is a scientific method of gathering data for specific purposes and applications. Based on this, four keywords require attention: scientific method, data, purpose, and usability. The scientific method implies that research activities are based on scientific characteristics such as rationality, empirical evidence, and systematicity. The author's methods include library research, analysis, experimentation, and design.

The Artificial Neural Network (ANN) method with the Backpropagation technique is used in this feature selection research on the proposed C1 sheet design for general elections in Indonesia (Shanmuganathan, 2016). A programming language is required to implement this technique. The programming language used in this study is Python, and the data processing application is Anaconda Navigator with Jupyter Network tools (Santoso & Umam, 2018).

A literature review was conducted by searching for literature related to this research, such as searching literature on the Artificial Neural Network Method to learn about backpropagation techniques (Amalia, 2017). The following method is analysis. It is accomplished by analyzing and developing data processing algorithms and pattern recognition and developing and implementing the Backpropagation technique (Ghojogh et al., 2021).

The third approach is experimentation. The authors used this method to test how the Backpropagation technique (Lesnussa et al., 2018) is implemented in processing election result data by transferring the image of numbers from sheet C1 on the KPU SITUNG to the proposed sheet C1 design. The design method is the final method. Processes such as model design, coding, and testing are carried out using this method.

Preparation

At this stage, the researcher collected samples of C1 sheets obtained from the KPU vote count website, and then the researcher made a new proposed C1 sheet of 100 sheets, as shown in Figure 1.

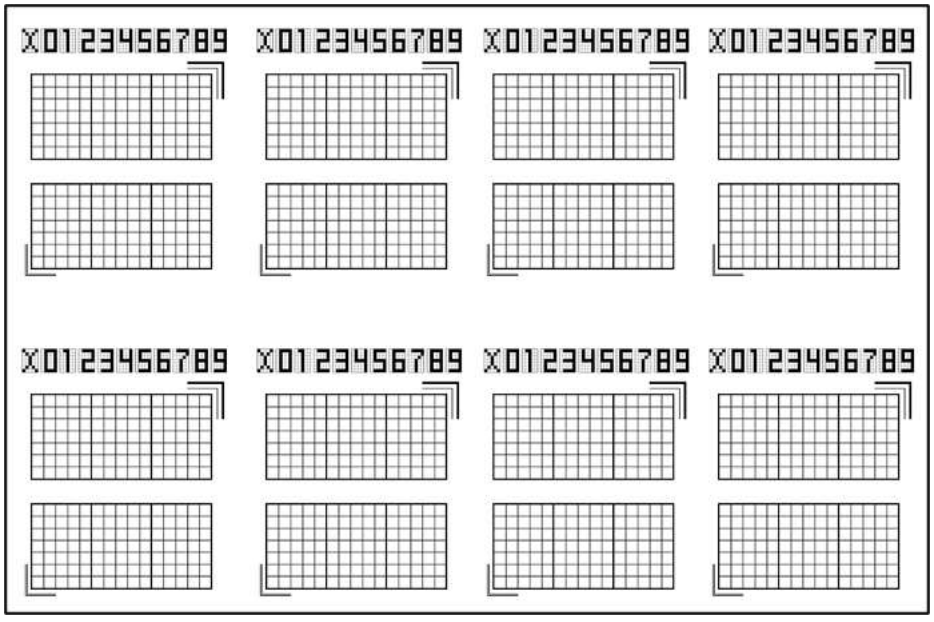

Figure 1. New C1 Sheet Design

The shaded number pattern on the template is photographed as shown in Figure 2, cropped in each numbered box, and saved in.jpg format with a resolution of 15x21 pixels, as depicted in Figure 3.

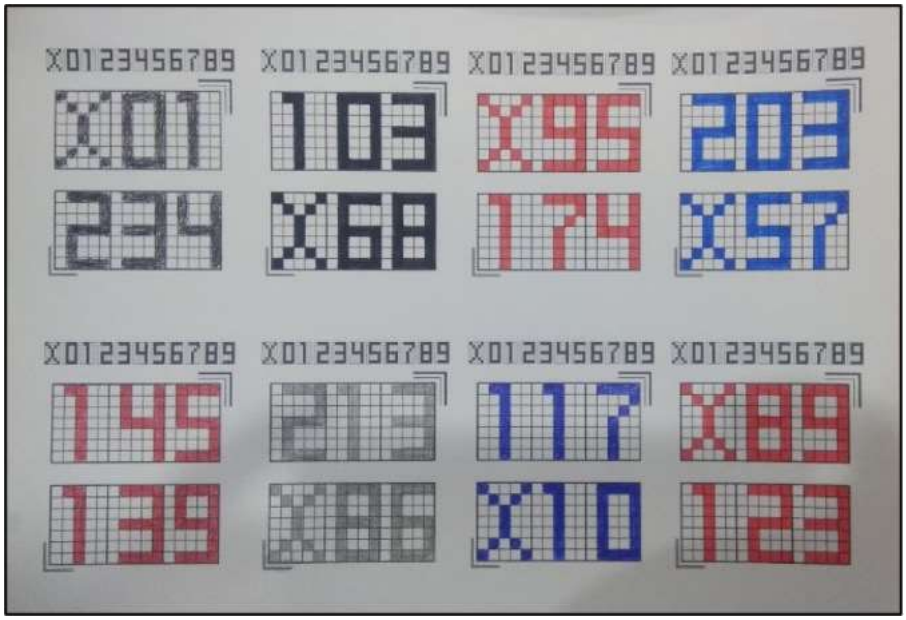

Figure 2. Shaded Number Pattern

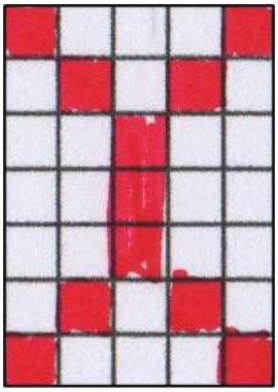

Figure 3. Number pattern that has been cropped

Data Processing

Preprocessing is the initial stage of processing shaded data by changing the image's color to grayscale. Because the image's quality and brightness differ from the photo taken, the next step is to do character segmentation to change the image to a black-and-white image or a binary image to reduce noise and make the image clearer.

Feature Extraction

Extraction of Characteristics To characterize objects from the results of this process, the image will have one value with the numbers 1 and 0 for each pixel. The number 1 represents the value of a contained pattern, and the number 0 represents an empty pattern, as shown in Figure 4.

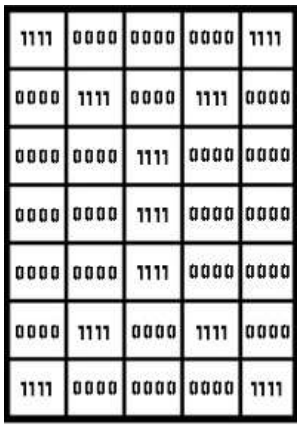

Figure 4. Number values in the image

Furthermore, the image is processed using an information processing algorithm designed to mimic the workings of the human brain by learning bit values (0 and 1) in solving a problem through changes in synaptic weights. Based on the past, artificial neural networks can recognize activities. Artificial neural networks will study past data to decide on previously studied data.

Results and Discussion

Data Selection

The first stage of this research's implementation was to prepare data selection using data obtained from the KPU election calculation website, which was then transferred to the proposed C1 sheet design template. The research data used sheet C1 in the Penukal Abab Lematang Ilir (PALI) District of the Province of South Sumatra, totaling 589 TPS. The numeric characters were transferred to the new C1 sheet template with various writing tools, including pencils, pens, and colored markers.

Dataset

Training/Dataset/Data Set Data is a data set derived from past information and managed to become information for performing data mining science techniques. Data transferred to the new C1 sheet template is then edited and cropped before being grouped by the same number as the test material, as depicted in Figure 5.

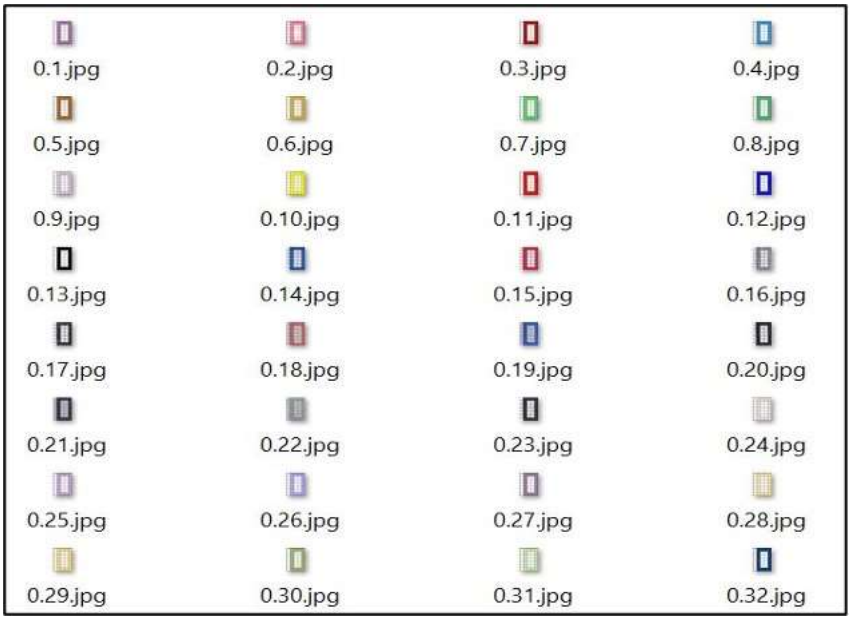

Figure 5. Image Data Sets

Preprocessing

At this stage, each existing dataset is preprocessed, with all datasets resized to 15x21 pixels so that each cropped image has the same pixel size before processing and testing the data.

Classification

The classification stage aims to divide the dataset into two parts, namely training, and testing; later, the results of testing the accuracy of the dataset in this study can be obtained from this division. The accuracy value obtained from this training was 89.13%, as shown in Figure 6.

```
Splitting data training dan testing ...
Data Training: [ 230 ] data
Data Testing: [ 154 ] data
Splitting data Done!


-- Training data --
Accuracy: 89.13
Classification Report:
              precision    recall  f1-score   support

           0       0.56      1.00      0.72        22
           1       1.00      0.92      0.96        13
           2       1.00      0.85      0.92        20
           3       1.00      0.88      0.93        24
           4       0.89      1.00      0.94        24
           5       0.93      0.96      0.95        27
           6       1.00      0.85      0.92        20
           7       1.00      1.00      1.00        24
           8       0.84      0.84      0.84        19
           9       1.00      0.47      0.64        19
          10       1.00      0.94      0.97        18

    accuracy                           0.89       230
   macro avg       0.93      0.88      0.89       230
weighted avg       0.93      0.89      0.89       230
```

Figure 6. Classification Output Results

Test Result

At this point, dataset testing is performed using the Backpropagation method to determine the accuracy of characters collected on the Jupyter notebook. The accuracy rate obtained from the results of this test was 98.5%, as shown in Figure 7.

```
---- Test data ----
Accuracy: 98.05
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        13
           1       1.00      1.00      1.00        22
           2       0.94      1.00      0.97        15
           3       1.00      0.91      0.95        11
           4       1.00      1.00      1.00        11
           5       1.00      0.75      0.86         8
           6       1.00      1.00      1.00        14
           7       1.00      1.00      1.00        11
           8       1.00      1.00      1.00        16
           9       0.89      1.00      0.94        16
          10       1.00      1.00      1.00        17

    accuracy                           0.98       154
   macro avg       0.98      0.97      0.97       154
weighted avg       0.98      0.98      0.98       154
```

Figure 7. Test Output Results

When using the backpropagation method, a confusion matrix is displayed in addition to the level of accuracy to measure the performance of the classification model used. The confusion matrix works by comparing actual values with predicted values. The goal is to see the model's accuracy from the other side (Xu, Zhang, & Miao, 2020), as shown in Figure 8.

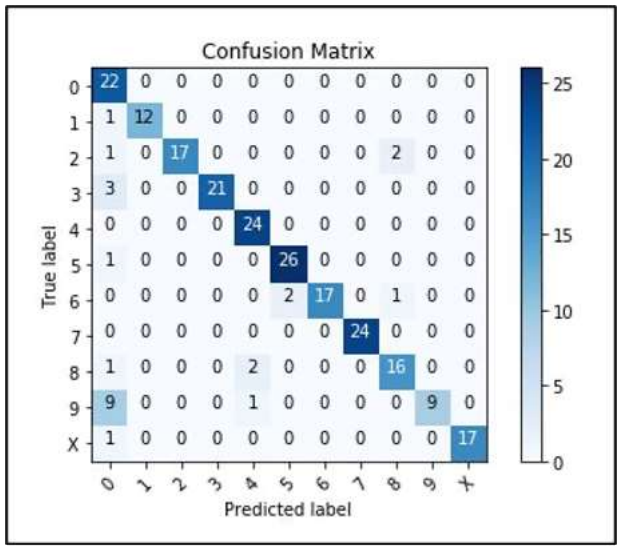

Figure 8. Confusion Matrix Backpropagation Method

From the confusion matrix image above, data with value five (5) can be predicted correctly. There are 26 data correctly predicted numbers. There is only 1 case where the number 5 is predicted as 0. Then there is the number 9, which is correctly predicted by as many as 9 numbers and is predicted as the number 0 with as many as 9 digits.

Testing Using Cross-Validation

Cross-validation is an additional data mining technique that aims to achieve maximum accuracy (Pratiwi, 2017). This method is known as k-fold cross-validation because the experiment is repeated k times for the same model with the same parameters. An accuracy rate of 95% is obtained in cross-validation testing, as shown in Figure 9.

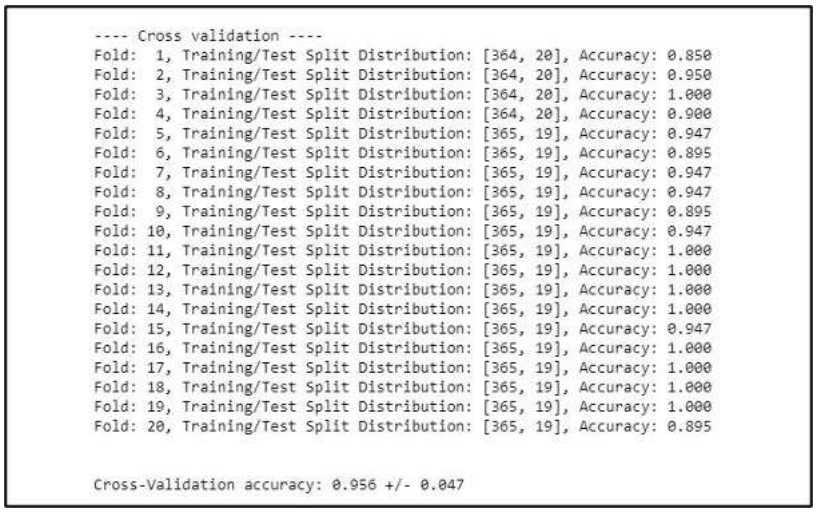

Figure 9. Testing Using Cross-Validation

Conclusions

According to the research findings, an Artificial Neural Network (ANN) with the Backpropagation method can detect the number pattern on the proposed C1 sheet design. Furthermore, pattern recognition using the Backpropagation method with tools in the Anaconda application, namely Jupyter Notebook, yields outstanding accuracy results with a 98.5% accuracy rate.

References

Amalia, S. (2017). Pengenalan Digit 0 Sampai 9 Menggunakan Ekstraksi Ciri MFCC dan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Teknik Elektro ITP*, *6*(1), 1–8. https://doi.org/10.21063/jte.2017.3133601

Amrutha, J., & Ajai, A. (2018). *Performance analysis of Backpropagation Algorithm of Artificial Neural Networks in Verilog*. https://doi.org/10.1109/RTEICT42901.2018.9012614

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2021). *Restricted Boltzmann Machine and Deep Belief Network: Tutorial and Survey*. Retrieved from http://arxiv.org/abs/2107.12521

Herman, H., Syafie, L., & Indra, D. (2018). Pengenalan Angka Tulisan Tangan Menggunakan Jaringan Syaraf Tiruan. *ILKOM Jurnal Ilmiah*, *10*, 201. https://doi.org/10.33096/ilkom.v10i2.317.201-206

Lesnussa, Y. A., Mustamu, C. G., Kondo Lembang, F., & Talakua, M. W. (2018). Application of Backpropagation Neural Networks in Predicting Rainfall Data in Ambon City. *International Journal of Artificial Intelligence Research*, *2*(2). https://doi.org/10.29099/ijair.v2i2.59

Mellyssa, W. (2019). Pengenalan Nominal Uang Kertas menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Litek : Jurnal Listrik Telekomunikasi Elektronika*, *16*, 1. https://doi.org/10.30811/litek.v16i1.1463

Octariadi, B. C. (2020). Pengenalan Pola Tanda Tangan Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. *Jurnal Teknoinfo*, *14*(1), 15. https://doi.org/10.33365/jti.v14i1.462

Pangaribuan, Y., & Sagala, M. (2017). Menerapkan Jaringan Saraf Tiruan untuk Mengenali Pola Huruf Menggunakan Metode Perceptron. *Jurnal Teknik Informatika UNIKA Santo Thomas*, *2*(479), 53–59. Retrieved from http://103.76.21.184/index.php/JTIUST/article/view/191

PRASASTI, A. L., IRAWAN, B., FAJRI, S. E., RENDIKA, A., & HADIYOSO, S. (2020). Perbandingan Ekstraksi Fitur dan Proses Matching pada Autentikasi Sidik Jari Manusia. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *8*(1), 95. https://doi.org/10.26760/elkomika.v8i1.95

Pratiwi, L. (2017). PERBANDINGAN METODE CROSS VALIDATION DAN GENERALIZED CROSS VALIDATION DALAM REGRESI NONPARAMETRIK BIRESPON SPLINE. *Jurnal Varian*, *1*, 43. https://doi.org/10.30812/varian.v1i1.49

Putra, K. N. A. (2018). *Deteksi Dan Identifikasi Plat Nomor Kendaraan Bermotor Menggunakan Jaringan Syaraf Tiruan Backpropagation*. Retrieved from https://repository.unsri.ac.id/8208/

Romario, M., Ihsanto, E., & Kadarina, T. (2020). Sistem Hitung dan Klasifikasi Objek dengan Metode Convolutional Neural Network. *Jurnal Teknologi Elektro*, *11*, 108. https://doi.org/10.22441/jte.2020.v11i2.007

Rosnelly, R. (2018). Pengenalan Pola Angka Tulisan Tangan Pada Cek Menggunakan Neocognitron. *CSRID (Computer Science Research and Its Development Journal)*, *10*(1), 23. https://doi.org/10.22303/csrid.10.1.2018.58-67

Santoso, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics Teori dan Implementasi menggunakan Python & Apache Spark*. Jogjakarta: Penebar Media Pustaka.

Shanmuganathan, S. (2016). *Artificial Neural Network Modelling: An Introduction*. https://doi.org/10.1007/978-3-319-28495-8_1

Sugiyono. (2014). *Metode Penelitian Kuantitatif Kualitatif dan E&D*. Bandung: Alfabeta.

Syafitri, N. (2011). Pengenalan Pola Untuk Deteksi Uang Koin. *Semin. Nas. Teknol. Inf. Komun. Dan Ind. III*. Semin. Nas. Teknol. Inf. Komun. dan Ind. III.

Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, *507*, 772–794. https://doi.org/https://doi.org/10.1016/j.ins.2019.06.064