Journal of Information Systems and Informatics

Vol. 5, No. 4, December 2023 e-ISSN: 2656-4882 p-ISSN: 2656-5935

DOI: 10.51519/journalisi.v5i4.592 Published By DRPM-UBD

Comparing the Prediction of Numeric Patterns on Form C1 Using the K-Nearest Neighbors (K-NN) Method and a Combination of K-Nearest Neighbors (K-NN) with Connected Component Labeling (CCL)

Uci Suriani[1], Tri Basuki Kurniawan[2]

[1] Master of Informatics Engineering Study Program, Postgraduate, Bina Darma University,
Palembang, Indonesia 30264
Email: [1]uci.suryani1@gmail.com, [2]tribasukikurniawan@binadarma.ac.id

Abstract

Indonesia's elections serve as a cornerstone of its democratic system, with the active participation of its citizens being of paramount importance. To bolster transparency and civic engagement during these elections, the SITUNG system (Election Result Information System) is employed for the tabulation of election results. However, the current tabulation process remains manual, potentially leading to data entry errors and a reduced accuracy of election outcomes. This research endeavor seeks to enhance the efficiency and accuracy of election result tabulation by employing the K-Nearest Neighbors (K-NN) method for recognizing numeric patterns on Form C1, both independently and in combination with Connected Component Labeling (CCL). The K-NN method demonstrates a commendable 60.0% accuracy in recognizing numeric patterns from the original Form C1 data. However, when combined with CCL, the accuracy drops to 51.2%. This research makes a significant contribution by simplifying the tabulation process and improving the accuracy of election results in Indonesia through the application of the K-NN method. The technology is anticipated to fortify democracy by promoting a more transparent and participatory electoral process for the citizens.

Keywords: Predicting numeric patterns, K-Nearest Neighbors Method (K-NN), Connected Component Labeling (CCL), General Election Commission Form C1

1. INTRODUCTION

Indonesia adheres to a democratic system, characterized by a presidential model that facilitates the transition of political positions through General Elections (Pemilihan Umum) [1]. Elections represent the method for selecting individuals to occupy specific political roles. In today's digital era, virtually all human activities are influenced by advancements in information and communication technology, exemplified by systems like SITUNG [2]. SITUNG is a system employed for aggregating election results. Although SITUNG has been in use since 2004, it initially operated offline. In 2014, it transitioned to an online platform with the aim of allowing the public to monitor the vote tallying process [3]. However, the vote tallying process still relies on manual input of numbers from Form C1 into the SITUNG system."

The rapid evolution of technology has ushered in new disciplines, among them artificial intelligence, which is a branch of artificial neural networks [4]. Artificial neural networks excel in problem-solving, particularly in the realms of classification and pattern recognition [15], [16]. As a result, many researchers have delved into studies that leverage these capabilities. For example, there's research focused on recognizing numeric patterns, evaluated using 100 test images. The top three classification results for identifying handwritten Arabic (Indian) numerals using the K-NN method were 86% with k=1, 84% with k=3, and 83% with k=5 [5]. Additionally, a study involving distance transform in kernel discriminant analysis for numeric pattern recognition achieved a direct accuracy rate of 95.5%, and a combined accuracy of 87.98% [6]. Subsequently, character segmentation was carried out employing the Connected Component Labeling (CCL) method [7]. This segmentation technique is used to isolate characters within an image, preventing them from merging together.

The utilization of CCL as a character segmentation process aims to isolate objects recognized as characters, such as those found in vehicle license plates [8]. During testing, the segmentation method using CCL successfully achieved an 80% accuracy rate. Once the segmented characters are obtained, the subsequent step involves character recognition [9].

Considering recent advancements in various academic fields, it has become evident that Form C1 should be subject to scrutiny, especially in response to the multitude of issues that arose during the 2019 elections, specifically the inaccuracies in data input during election result tabulation [1]. Hence, this research is focused on Form C1, with an emphasis on pattern recognition. Originally, Form C1 was manually transcribed, but it has since been digitalized into pattern-based representations. To facilitate the recognition of numerical patterns in election results, patterns for each digit have been generated using a variety of colors. This has been achieved by shading with pens, pencils, markers, or coloring tools to represent the numbers 0 to 9, as well as the letter X, which is employed to fill in vacant columns in the vote tally. The identification of these patterns during the pattern recognition process necessitates a data matching method [10]. Each digit boasts its own unique characteristics, making a data matching method crucial for accurate numeric recognition. As such, this research employs the K-Nearest Neighbors (K-NN) method, complemented by a combination of K-NN and Connected Component Labeling (CCL), in order to enable machine-based recognition of numeric patterns.

2. METHODS

Drawing upon a comprehensive review of relevant literature in various journals and previous research experiments related to the subject of numeric pattern recognition, this study focuses on the recognition of numeric patterns using data from Form C1 ballots provided by the General Election Commission (KPU). The research's conceptual framework is presented in Figure 1.

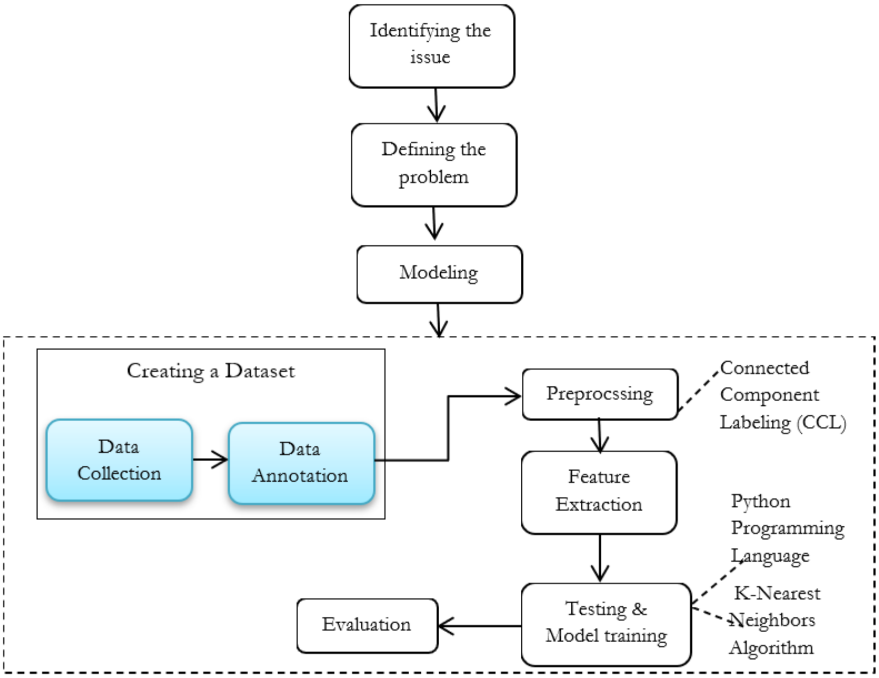


Figure 1. Illustrates the flowchart of the conceptual framework.

Within this framework, we undertake the process of selecting the appropriate machine learning algorithm to predict numeric patterns on the Election Form C1. The essential points of this framework can be elucidated as follows.

a. Data Collection: The dataset used comprises information extracted from General Election Commission Form C1.
b. Data Annotation: This stage is of paramount importance, as it involves data labeling. It serves to determine whether the images in the Form C1 format are legible or not. The process encompasses various adjustments to the Form C1 images, such as converting them to grayscale and reducing their resolution.

c. Preprocessing: This stage involves two methods: preprocessing the raw data before machine learning, either with or without the application of the CCL algorithm.

d. Feature Extraction: This step holds significant importance in the classification of Form C1 images. It entails the transformation of unstructured data into a structured format that can be processed by machine learning algorithms for classification into predefined categories.

e. Training and Testing: The implementation process is carried out using the Python programming language. In this phase, classification algorithms are tested to determine which one performs most effectively in recognizing image content, such as the K-Nearest Neighbors (K-NN) method and the combination of K-NN with CCL.

f. Evaluation: The model is evaluated in this phase to draw well-founded conclusions regarding its practical usability.

2.1. Creating the Dataset

To acquire the numeric patterns essential for this study, the researcher has devised a numeric pattern format that will be used as a template for generating numeric patterns. The procedural steps for establishing a dataset that can be effectively utilized in this research are outlined in Figure 2.

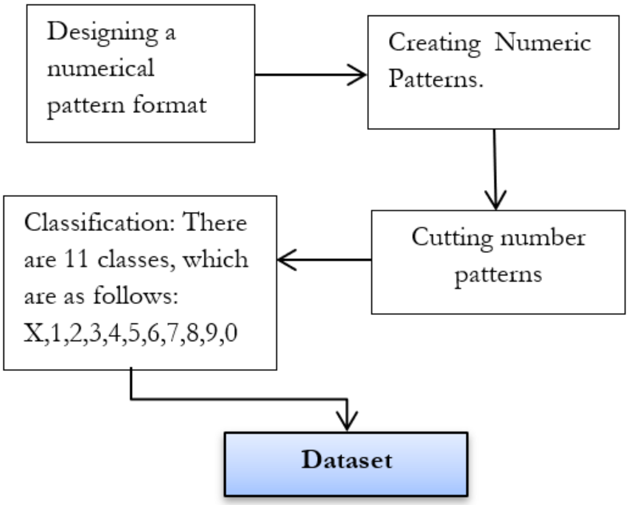
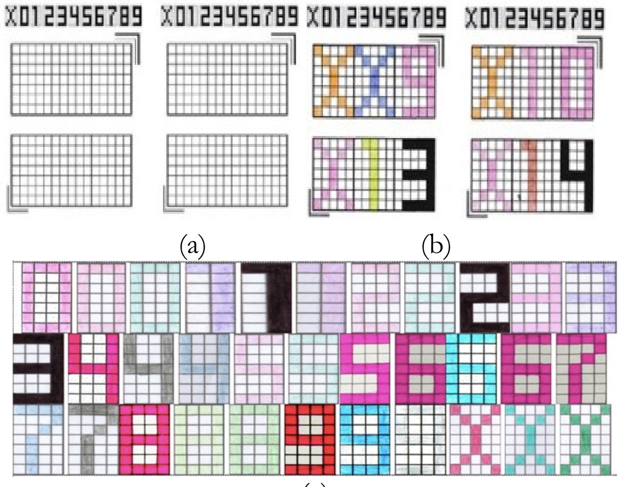

Designing a numerical pattern format
Creating Numeric Patterns.
Cutting number patterns
Classification: There are 11 classes, which are as follows: X,1,2,3,4,5,6,7,8,9,0
Dataset

Figure 2. The process of creating the dataset.

In order to enable numeric pattern recognition within the system, it is necessary to create numeric patterns beforehand. The format for generating numeric patterns in the new Form C1 template is illustrated in Figure 3.



(a) (b)

(c)

Figure 3. The Numeric Pattern Creation Process. (a) Initial image format. (b) Image after coloring. (c) Final numeric pattern result.

In the numeric pattern format, each digit is outlined to resemble the numbers 0-9 and the letter X, as illustrated in Figure 3 (point b). Each digit is displayed in various distinct colors, serving the dual purpose of determining which colors the system recognizes better and assessing how lighting affects the recognition of these numeric patterns. The numeric patterns are captured using either a camera or a scanning machine to acquire the digit patterns, which are then stored in digital file format. Once the numeric patterns are digitized into images, each digit is uniformly cropped to a size of 15x21 pixels. Several examples of these cropped numeric patterns can be seen in Figure 3 (point c).

The results of digit segmentation are categorized into their respective classes. This categorization comprises 11 classes: X, 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Subsequently, each class is used as both training and testing data. To facilitate pattern recognition, specific criteria are assigned to each digit within the new format of Form C1. In the process of pattern recognition, both the K-NN method and a combination of the K-NN method with CCL are employed, allowing for pattern recognition in a computer or future vote counting system.

2.2. Image Processing

In this study, we utilized the Connected Component Labeling (CCL) method for image processing. CCL is a technique that can be applied to classify regions or objects within digital images. This technique is based on the theory of pixel connectivity in images. Pixels within a region are considered 'connected' when they adhere to adjacency rules or rules of pixel proximity. These rules rely on the inherent pixel neighborhood relationships. After completing the CCL process, each numeric image is labeled to prepare it for the subsequent preprocessing stages.

In this study, we employed the Connected Component Labeling (CCL) method for image processing. CCL is a technique used to classify regions or objects within digital images, based on the theory of pixel connectivity. Pixels within a region are considered 'connected' when they adhere to adjacency rules, which are determined by pixel proximity. This means that connected pixels have inherent neighborhood relationships with one another. Once the CCL process is completed, each numeric image is labeled in preparation for subsequent preprocessing stages.

2.3. Classification

In this study, classification is performed using the K-NN algorithm, which is based on the concept of the nearest neighbors. This classification method allows for the consideration of more than one neighbor, a technique commonly referred to as K-NN classification [8]. Through a more modern approach, K-NN classification aims to identify a group of k objects within the training set that are closest to the object being tested, based on their prevailing class in the surrounding context [11]. This approach involves three key elements: a group of labeled objects, a distance calculation or mathematical equation for measuring the distances between objects, and the value of k (indicating the number of neighbors).

This algorithm is straightforward, operating based on the shortest distance from the query instance to the training samples to determine their proximity [12]. The steps for computing the K-Nearest Neighbors (K-NN) method include the following [13]:

a. Determining the parameter 'k';
b. Calculating the distance between the data to be evaluated and all training samples;
c. Sorting the resulting distances;
d. Selecting the closest distances up to the 'k' order;
e. Matching the corresponding classes;
f. Counting the number of classes from the nearest neighbors and assigning that class to the data being evaluated (Equation 1).

$$d_i = \sqrt{\sum_{i=1}^{p} (X_{2i} - X_{1i})^2} \tag{1}$$

Where:

$X1$ = Sample Data
$X2$ = Test Data
i = Data Variable
d = Distance
p = Data Dimension

2.4. Building the Model

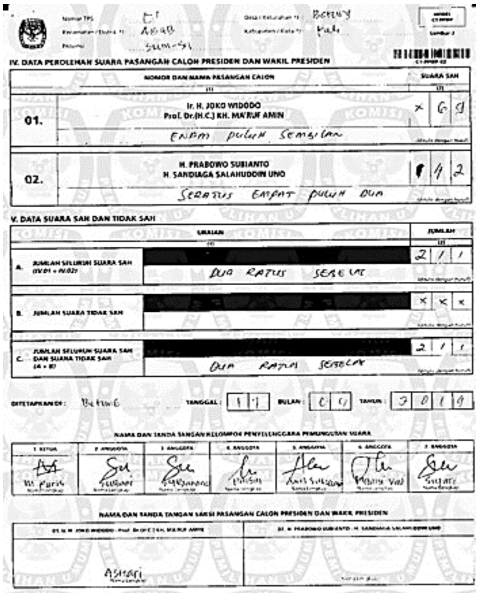
In the modeling phase, this research employs Google Colab as a tool to facilitate data processing. The first step in the modeling process involves establishing a connection between Google Colab and Google Drive to access a previously uploaded dataset. Executing this command provides a URL containing an authorization code. The 'AngkaDB' dataset is created from segments of newly formatted number images from C1, which are first outlined and then organized according to their respective categories. During the labeling stage, each number in the dataset is assigned a label for individual number recognition.

Data modeling is highly important for visualizing and obtaining initial insights into data patterns. In this research, the chosen method for modeling is K-Nearest Neighbors (KNN). This algorithm is simple, operating based on the shortest distance between a query instance and training samples to determine its neighbors. The KNN method is used to predict numeric patterns on Form C1 and plays a key role in the election result tallying process.

3. RESULTS AND DISCUSSION

This research focuses on identifying numeric patterns on Form C1 and involves creating a new format for Form C1 using a pattern recognition approach, with reference to the original Form C1 (see Figure 4 for details).

The objective of pattern recognition is to identify the distinctive features of the election result numbers on Form C1. This facilitates quicker vote tabulation and minimizes data entry errors. In this study, we focus on the numeric patterns present in the new format of Form C1, which incorporates various colors. To enable the computer to recognize these numbers, we need to define the characteristics of each number, taking into account factors like color and shape. The characteristics primarily pertain to numeric values.



Nomor TPS: 01
Kecamatan / Distrik: ASAB
Provinsi: SUM-SL
Desa / Kelurahan: Betung
Kabupaten / Kota: Pali

MODEL C1-PPWP
Lembar 2

IV. DATA PEROLEHAN SUARA PASANGAN CALON PRESIDEN DAN WAKIL PRESIDEN

	NOMOR DAN NAMA PASANGAN CALON	SUARA SAH
	(1)	(2)
01.	Ir. H. JOKO WIDODO Prof. Dr.(H.C.) KH. MA'RUF AMIN	× 6 9
	ENAM PULUH SEMBILAN	
02.	H. PRABOWO SUBIANTO H. SANDIAGA SALAHUDDIN UNO	1 4 2
	SERATUS EMPAT PULUH DUA	

V. DATA SUARA SAH DAN TIDAK SAH

	URAIAN		JUMLAH
		(1)	(2)
A.	JUMLAH SELURUH SUARA SAH (IV.01 + IV.02)	[redacted]	2 1 1
		DUA RATUS SEBELAS	
B.	JUMLAH SUARA TIDAK SAH	[redacted]	× × ×
C.	JUMLAH SELURUH SUARA SAH DAN SUARA TIDAK SAH (A + B)	[redacted]	2 1 1
		DUA RATUS SEBELAS	

DITETAPKAN DI: Betung TANGGAL: 1 1 BULAN: 0 4 TAHUN: 2 0 1 9

NAMA DAN TANDA TANGAN KELOMPOK PENYELENGGARA PEMUNGUTAN SUARA

1. KETUA	2. ANGGOTA	3. ANGGOTA	4. ANGGOTA	5. ANGGOTA	6. ANGGOTA	7. ANGGOTA
M. Faris	Sudari	Sugiharno	Ehsan	Amrh Sulaiman	Wahyu Vidi	Sugiarti

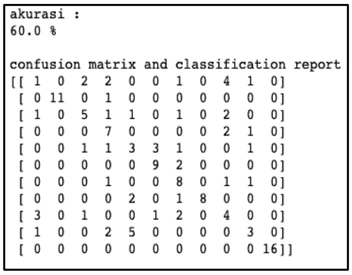
NAMA DAN TANDA TANGAN SAKSI PASANGAN CALON PRESIDEN DAN WAKIL PRESIDEN

01. H. IR. JOKO WIDODO - Prof. Dr.(H.C.) KH. MA'RUF AMIN	02. H. PRABOWO SUBIANTO - H. SANDIAGA SALAHUDDIN UNO
ASHARI	

Figure 4. An example of an authentic Form C1 document

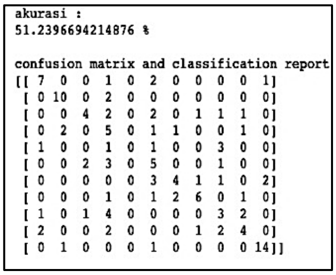
3.1 Comparing the Outcomes of the K-NN Method and the K-NN Combined with CCL

From the AngkaDb dataset, the data accuracy reaches 60.0%. This indicates that not all predictions of numeric patterns are entirely accurate. Additionally, we use a confusion matrix to assess precision and recall [14]. Precision measures the level of accuracy in matching user-requested information with the system's responses, while recall evaluates the system's success in retrieving information. The purpose of employing this function is to ensure that we don't blindly rely on the model's accuracy, but instead, we can justify its correctness by examining the model's accuracy results from different angles.



```
akurasi :
60.0 %

confusion matrix and classification report
[[ 1  0  2  2  0  0  1  0  4  1  0]
 [ 0 11  0  1  0  0  0  0  0  0  0]
 [ 1  0  5  1  1  0  1  0  2  0  0]
 [ 0  0  0  7  0  0  0  0  2  1  0]
 [ 0  0  1  1  3  3  1  0  0  1  0]
 [ 0  0  0  0  0  9  2  0  0  0  0]
 [ 0  0  0  1  0  0  8  0  1  1  0]
 [ 0  0  0  0  2  0  1  8  0  0  0]
 [ 3  0  1  0  0  1  2  0  4  0  0]
 [ 1  0  0  2  5  0  0  0  0  3  0]
 [ 0  0  0  0  0  0  0  0  0  0 16]]
```

(a)

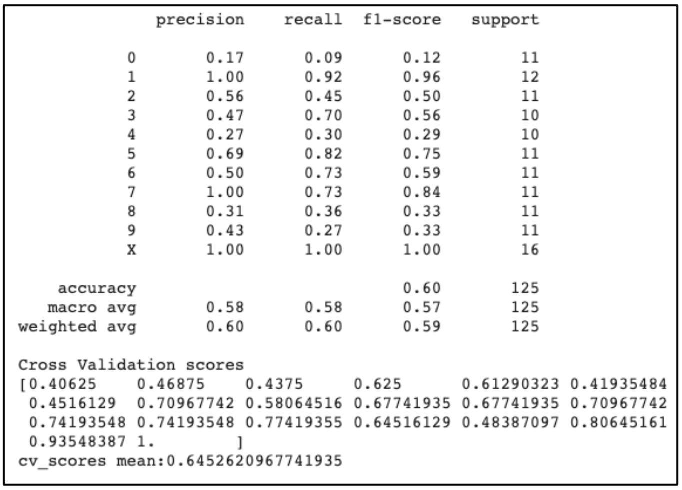


```
akurasi :
51.2396694214876 %

confusion matrix and classification report
[[ 7  0  0  1  0  2  0  0  0  0  1]
 [ 0 10  0  2  0  0  0  0  0  0  0]
 [ 0  0  4  2  0  2  0  1  1  1  0]
 [ 0  2  0  5  0  1  1  0  0  1  0]
 [ 1  0  0  1  0  1  0  0  3  0  0]
 [ 0  0  2  3  0  5  0  0  1  0  0]
 [ 0  0  0  0  0  3  4  1  1  0  2]
 [ 0  0  0  1  0  1  2  6  0  1  0]
 [ 1  0  1  4  0  0  0  0  3  2  0]
 [ 2  0  0  2  0  0  0  1  2  4  0]
 [ 0  1  0  0  0  1  0  0  0  0 14]]
```

(b)

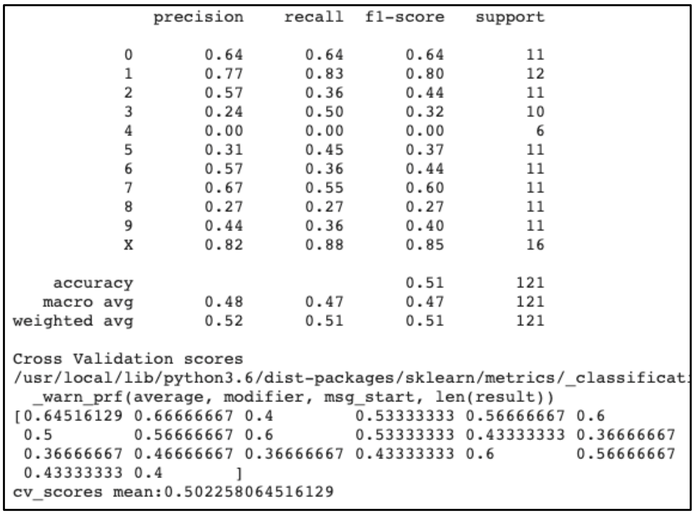
Figure 5. Confusion Matrix for (a) K-Nearest Neighbor (K-NN) and (b) K-Nearest Neighbor (K-NN) with Connected Component Labeling (CCL).

In Figure 5, we can observe variations in the accuracy of digit predictions. In confusion matrix (a), it is evident that digit 1 is correctly predicted only once, while digits 2 and 3 are predicted twice, digit 6 once, digit 8 four times, and digit 9 once. If '0' is present, it indicates that the training digit is not misclassified as a digit in another category. The same principles apply to the other categories as well, as seen in confusion matrix (b). Alongside the confusion matrix, there is the concept of cross-validation, which determines the number of folds to be used (Refer to Figures 6 and 7).



```
              precision    recall  f1-score   support

           0       0.17      0.09      0.12        11
           1       1.00      0.92      0.96        12
           2       0.56      0.45      0.50        11
           3       0.47      0.70      0.56        10
           4       0.27      0.30      0.29        10
           5       0.69      0.82      0.75        11
           6       0.50      0.73      0.59        11
           7       1.00      0.73      0.84        11
           8       0.31      0.36      0.33        11
           9       0.43      0.27      0.33        11
           X       1.00      1.00      1.00        16

    accuracy                           0.60       125
   macro avg       0.58      0.58      0.57       125
weighted avg       0.60      0.60      0.59       125

Cross Validation scores
[0.40625    0.46875    0.4375     0.625      0.61290323 0.41935484
 0.4516129  0.70967742 0.58064516 0.67741935 0.67741935 0.70967742
 0.74193548 0.74193548 0.77419355 0.64516129 0.48387097 0.80645161
 0.93548387 1.        ]
cv_scores mean:0.6452620967741935
```

Figure 6. Cross-Validation of the K-Nearest Neighbors Method

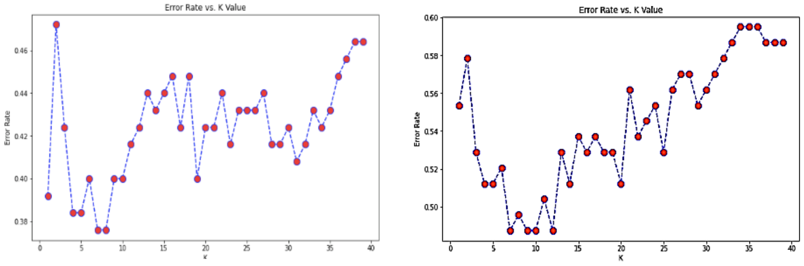


```
              precision    recall  f1-score   support

           0       0.64      0.64      0.64        11
           1       0.77      0.83      0.80        12
           2       0.57      0.36      0.44        11
           3       0.24      0.50      0.32        10
           4       0.00      0.00      0.00         6
           5       0.31      0.45      0.37        11
           6       0.57      0.36      0.44        11
           7       0.67      0.55      0.60        11
           8       0.27      0.27      0.27        11
           9       0.44      0.36      0.40        11
           X       0.82      0.88      0.85        16

    accuracy                           0.51       121
   macro avg       0.48      0.47      0.47       121
weighted avg       0.52      0.51      0.51       121

Cross Validation scores
/usr/local/lib/python3.6/dist-packages/sklearn/metrics/_classificat
  _warn_prf(average, modifier, msg_start, len(result))
[0.64516129 0.66666667 0.4        0.53333333 0.56666667 0.6
 0.5        0.56666667 0.6        0.53333333 0.43333333 0.36666667
 0.36666667 0.46666667 0.36666667 0.43333333 0.6        0.56666667
 0.43333333 0.4       ]
cv_scores mean:0.502258064516129
```

Figure 7. Cross-Validation of the K-Nearest Neighbors (K-NN) Method Combined with Connected Component Labeling (CCL).

3.2 Assessing the Error Rate

The error rate is a measure used to calculate the rate of errors in predictions. It is presented in the form of a curve that distinctly illustrates the differences between those who use the K-NN method and those who use the combination of the K-NN method with CCL (see Figure 8).



(a) (b)

Figure 8 depicts the error rates for (a) the K-NN method and (b) the combination of the K-NN method with CCL.

After conducting tests using both the K-Nearest Neighbors (K-NN) method and the combination of K-NN with CCL, a comparison reveals that the accuracy achieved with the K-NN method is higher than when using the combination of K-NN and CCL. The accuracy obtained with the K-NN method is 60.0%, whereas it is only 51.2% with the combination of K-NN and CCL. To assess the performance of the classifiers in both experiments, you can refer to the comparison presented in Table 1.

Table 1. Method Comparison Results

Class	Method K-NN				Combination of K-NN Method with CCL			
	Precision	Recall	Fi-score	Support	Precision	Recall	Fi-score	Support
0	0.17	0.09	0.12	11	0.64	0.64	0.64	11
1	1	0.92	0.96	12	0.77	0.83	0.8	12
2	0.56	0.45	0.5	11	0.57	0.36	0.44	11
3	0.47	0.7	0.56	10	0.24	0.5	0.32	10
4	0.27	0.3	0.29	10	0	0	0	6
5	0.69	0.82	0.75	11	0.31	0.45	0.37	11
6	0.5	0.73	0.59	11	0.57	0.36	0.44	11
7	1	0.73	0.84	11	0.67	0.55	0.6	11
8	0.31	0.36	0.33	11	0.27	0.27	0.27	11
9	0.43	0.27	0.33	11	0.44	0.36	0.4	11
X	1	1	1	16	0.82	0.88	0.85	16
Acc	0.6				0.512			

4. CONCLUSION

This research centers on the recognition of numeric patterns by introducing a new format for Form C1 sheets in elections. The process entails the creation of this new format, outlining each digit, converting them into digital formats such as jpg or pdf, and uniformly cropping them. Subsequently, the cropped results are categorized into 11 classes and employed as a dataset for modeling.

The K-Nearest Neighbors (K-NN) method and the combination of K-NN with Connected Component Labeling (CCL) are employed to predict numeric patterns. The research findings indicate an accuracy rate of 60.0% when utilizing the K-NN method, while the combination of K-NN with CCL achieves 51.2%. From the research outcomes, it can be concluded that the K-NN method is more effective at recognizing numeric patterns on Form C1 compared to the combination of K-NN with CCL. The higher accuracy rate of the K-NN method demonstrates its potential for future application in the election result tabulation process.

This study makes a substantial contribution to streamlining the election result tabulation process and improving accuracy through the utilization of the K-NN method. However, to achieve further accuracy improvements, additional research and the selection of an appropriate method tailored to the specific data conditions and characteristics are essential.

REFERENCES

[1] M. Haboddin, *Pemilu dan partai politik di Indonesia*. Universitas Brawijaya Press, 2016.

[2] A. S. Alam and M. I. Sultan, “Keterbukaan informasi publik melalui sistem penghitungan (situng) online hasil pilkada terhadap pengetahuan, sikap, dan perilaku masyarakat di Kota Palu,” *KAREBA J. Ilmu Komun.*, pp. 92–103, 2016.

[3] A. Priyono, “Eksistensi Putusan Badan Pengawas Pemilu Republik Indonesia Nomor 07/LP/PP/ADM/RI/00.00/V2019 Terhadap Pelanggaran Tata Cara dan Prosedur Dalam Input Data Sistem Informasi Penghitungan Suara (SITUNG) Pemilu 2019 yang Tidak Dilaksanakan Oleh KPU Berdasarkan Undang-Undang Nomor 7 Tahun 2017 Tentang Pemilu,” 2020.

[4] M. Fadhilla, M. R. A. Saf, and D. S. S. Sahid, “Pengenalan kepribadian seseorang berdasarkan pola tulisan tangan menggunakan jaringan saraf tiruan,” *J. Nas. Tek. Elektro Dan Teknol. Inf. JNTETI*, vol. 6, no. 3, pp. 365–373, 2017.

[5] R. S. Akbar Eko Adi, “Studi Analisis Pengenalan Pola Tulisan Tangan Angka Arabic (Indian) menggunakan Metode K- Nearest Neighbors dan Connected Component Labeling,” *Din. Rekayasa*, no. Vol 12, No 2 (2016): Dinamika Rekayasa-Agustus 2016, pp. 45–51, 2016.

[6] A. Husain, A. H. A. Prastian, and A. Ramadhan, “Perancangan Sistem Absensi Online Menggunakan Android Guna Mempercepat Proses

Kehadiran Karyawan Pada PT. Sintech Berkah Abadi," *Technomedia J.*, vol. 2, no. 1, pp. 105–116, 2017.

[7] R. Akbar and E. A. Sarwoko, "Aplikasi Pengenalan Pola Tulisan Tangan Angka Arabic (Indian) menggunakan Metode Connected Component Labeling dan Template Matching," 2016.

[8] A. Budianto, R. Ariyuana, and D. Maryono, "Perbandingan K-Nearest Neighbor (KNN) Dan Support Vector Machine (SVM) Dalam Pengenalan Karakter Plat Kendaraan Bermotor," *J. Ilm. Pendidik. Tek. Dan Kejuru.*, vol. 11, no. 1, pp. 27–35, 2018.

[9] A. Budianto, T. B. Adji, and R. Hartanto, "Deteksi nomor kendaraan dengan metode connected component dan SVM," *J. Teknol. Inf. Magister*, vol. 1, no. 01, pp. 106–117, 2016.

[10] H. Masrani and I. R. Ilhamsyah, "Aplikasi Pengenalan Pola pada Huruf Tulisan Tangan Menggunakan Jaringan Saraf Tiruan dengan Metode Ekstraksi Fitur Geometri," *Coding J. Komput. Dan Apl.*, vol. 6, no. 2, 2018.

[11] Vinita Chandani, Romi Satria Wahono, and Purwanto Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.

[12] M. Lestari, "Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung," *Fakt. Exacta*, vol. 7, no. 4, pp. 366–371, 2015.

[13] L. A. R. Hakim, A. A. Rizal, and D. Ratnasari, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)," *JTIM J. Teknol. Inf. Dan Multimed.*, vol. 1, no. 1, pp. 30–36, 2019.

[14] R. K. Dinata, F. Fajriana, Z. Zulfa, and N. Hasdyna, "Klasifikasi Sekolah Menengah Pertama/Sederajat Wilayah Bireuen Menggunakan Algoritma K-Nearest Neighbors Berbasis Web," *CESS J. Comput. Eng. Syst. Sci.*, vol. 5, no. 1, pp. 33–37, 2020.

[15] A. Muzakir, H. Syaputra, and F. Panjaitan, "A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia," Sci. J. Informatics; Vol 9, No 2 Novemb. 2022, doi: 10.15294/sji.v9i2.35149 , Oct. 2022.

[16] U.Ependi, A.F.Rochim, A.Wibowo, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods," *International Journal of Intelligent Engineering and Systems*, vol. 16, no.3, pp. 345-361, 2023.