



A Decision Guidance for Solving Success Rate Political Campaign Using Distance Weighted kNN in Nassi-Shneiderman Framework

Aryo Nugroho^{1,2*} Surya Sumpeno^{1,3} Kimiya Fujisawa⁴
 Mauridhi Hery Purnomo^{1,3,5}

¹*Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

²*Department of Information System, Universitas Narotama, Surabaya, Indonesia*

³*Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

⁴*Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology, Japan*

⁵*University Center of Excellence on Artificial Intelligence for Healthcare and Society (UCE AIHeS)*

* Corresponding author's Email: aryo.nugroho14@mhs.ee.its.ac.id

Abstract: Campaigns on online social media are becoming intriguing because all parties can compete with each other. Various content can be created to support or defeat one another. Interaction is important to control the battlefield in the face of disinformation. The campaign team must be able to increase interaction and reporting it as a form of achievement. The reality is that the campaign team has not been able to provide a measurable report on the success rate of political campaign interaction on social media. This is because there is no standardization in measuring these interactions. This fact inspires to build a model for measuring campaign interaction on social media. This model will be able to help the campaign team. This study uses a dataset from a governor election campaign in Indonesia which is then tested on a presidential election campaign dataset in America. The model was built using machine learning with the k classification method Nearest Neighbor (kNN) and Distance Weight kNN (DWkNN). The research stages were arranged using the Nassi-Shneiderman Framework. This framework describes the stages and comparison of the use of the KNN and the DWkNN from the training to testing stages in an easy to understand manner. The kNN and DWkNN training stages showed excellent accuracy results of nearly 100%. Furthermore, at the testing stage using the cross-validation method with a variety of fold 5 to 20 variations showed excellent results, namely 99.89% on the governor election dataset and a range of 98% on the American presidential election dataset. This classification model has been tested using several datasets from several candidate social media accounts. We also created a dataset campaign interaction. This study shows that the proposed model can outperform previous studies. So, that it provides novelty in the form of a campaign interaction guide model for the campaign team.

Keywords: Election campaign, Nassi-shneiderman, K-nearest neighbour, Distance weight.

1. Introduction

One of the most popular applications is social media. Supported by the use of smartphones [1] that make humans closer to this application [2]. Social media developed from a friendship application platform to be used for business purposes [3]. Promotion of products or services finds new media that can be easily accessed by its users. The use of this promotion is also used in political campaigns.

Election campaigns are one of the promotion models to increase the popularity and electability of a candidate or a party. The campaign activities that were initially in the form of the face-to-face switch made use of social media applications on the internet. Campaign patterns have evolved using new ways of communicating with voters [4].

The internet and smartphone technology [5] provides more opportunities for candidates to socialize their vision and mission. The internet can reach a wider audience with a lower budget. Candidates can reduce campaign costs using the

Table 1. The Previous works in campaign in social media

Authors	Features	Evaluation Method	Description
J. Ohme [3]	First Time Voter, Social Media Content-Type	Statistical Analysis	Predicting Campaign Participation in Social Media
D. Ktoridou [5]	Frequency, Type of content	Statistical Analysis	Social Media for Presidential Elections in Cyprus
T. Paterson [6]	Economic Growth, Cyber-Amplified Religious Intolerance	Systematic Literature Review	Indonesia Political Campaign during Cyberspace Growth
C. Fox [8]	Candidate Ethnic Politicization	Statistical Analysis	Evidence from Indonesia
W. Budiharto [9]	Sentiment Analysis: Positive and Negative	Prediction and Analysis	Indonesia Presidential Election
A. Khan [10]	Time, Number of Degree	Growth-Based Popularity Prediction	Popularity Predicting and Modelling Bipartite Networks
I.F. Roji [12]	Political issues	Reusable Component-Based Software Engineering Framework for accuracy level and classification	Indonesian people tweet about political campaign Classification using SVM
R. Hidayatillah [13]	Like, Retweet, Reply	Classification: Accuracy 76.74%	Level of Political Participation on two candidates Classification Naïve Bayes
F.I. Puspitasari [14]	Egoism, Altruism, Perception of ease	Statistical Analysis	The moderating effect of social media for recommending product and services
S. Z. M. Zain [15]	Influence, Agitate, Interactivity	Systematic Literature Review	Social Media Marketing using Hashtag (#) approach to attract Twitter
F. M. F. Wong [18]	Parody, Political, Media, Others	Principal Component Analysis Sentiment Analysis Classification	Quantifying Political Leaning from Tweets, Retweets, and Retweeters Classification using SVM
K. C [19]	Information Produced	Statistical Analysis	Peer-Produced Political Information During the 2016 U.S Presidential Campaign
M. Bay [20]	Hyper targeting and Psychometrics	Rawlsian Approach	Political and Commercial Campaign
J. McSwiney [21]	Mobilization and Social Movement	Statistical Analysis	2019 Australian federal election
D. Sameh [22]	Extraversion, Neuroticism, Psychoticism	Classification, Deep Learning	Behaviour Analysis in Voting Classification Naïve Bayes, SVM and Deep Learning
A. Nugroho [27]	Like, Retweet, Reply	Classification: Accuracy 80.14%	Campaign Interaction in Twitter Classification Naïve Bayes Kernel Estimator
D. Grimaldi[30]	Positive, Negative, and Neutral	Classification Accuracy 86%	2019 Spanish Presidential Election Classification using kNN

internet [6]. On the other hand, it can also reach several regions at the same time. However, this advantage also has its challenges.

Campaigns on social media are like a battlefield. The content disinformation is increasingly rising, which must also be resisted by the campaign teams [7]. Campaigns that are deemed successful are those

that can attract the presence or active participation of voters. Meanwhile, a campaign can be considered a failure if it is not able to attract the attention of voters to participate in the activity. Therefore, a successful campaign usually has a campaign strategy guide on social media. The need for this strategy guide can be assisted by information technology. On an advanced

scale, an automatic sign is carried out as in the enterprise resources planning scale [8].

In general, campaign teams do not have clear guidelines for measuring their success. This could be due partly to the lack of understanding of the campaign team with the social media campaign battlefields. Then the preparation of the campaign team was not mature due to the very short nomination on time. So, that to overcome these shortcomings, it is necessary to build a guide using artificial intelligence.

This study was conducted in Indonesia, which geographically has an archipelago-shaped area with a large population. The diversity of regions, ethnicities, and religions has made campaign topics in Indonesia very diverse[9]. In some areas, the internet has also reached remote areas. So, that its affordability can be relied upon. Also supported is smartphone ownership as a tool to access the internet. There are several types of general elections in Indonesia. Election activities are carried out for the election of members of the representative council, president, regional election for governor or mayor or regent [10]. In 2015 there were 269 regional head elections, namely governors, regents, and mayors. In 2017 there were 201 regional head elections in Indonesia. Whereas in 2020 there will be 270 regions that will hold simultaneous regional head elections, with details of 9 provinces, 224 districts, and 37 cities. This diversity also contributes to interaction patterns that are useful for data collection in research. Campaigns on social media can use several popular applications such as Facebook, Instagram, and Twitter [11]. Facebook users must create an account, and interactions can be enjoyed with a more flexible display. Users can invite each other to join in mutual friendship. Instagram has different characteristics by displaying images as a point of view. A picture can represent an event that intrigues the audience.

Another well-known application is Twitter which has a slightly different character. Twitter is designed so that humans can voice their thoughts in a short format. This format is often known as a tweet and has quite extensive fans. This model allows users no need to think about writing long sentences. It can even be developed as a chatbot or virtual assistant [12].

In general, Twitter is a social network application that is always used in election campaigns in any country. Some previous research shows various campaign activities that utilize Twitter. Starting from raising prospective voters, efforts to increase electability to behaviour analysis in directing choices [11]. However, not much research has focused on measuring the success rate of political campaigns in

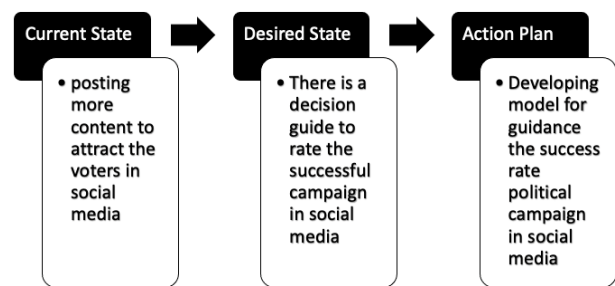


Figure. 1 Gap analysis

social media.

Some research related to Campaigns in Social media is generally done by statistical analysis methods, scenario, systematic literature review, and classification (see Table 1). In this paper, we are interested in solving the problem by providing a guide to measuring the success of campaign interactions with classification. Fig. 1 illustrates the gap analysis chart. The current state shows the current political campaign conditions. The desired state is the desired condition to be achieved. So, the action plan is a solution to the gap that will be resolved.

This study will use the main dataset from the Ganjar Pranowo campaign in Indonesia. This dataset has been used in previous studies using the Naïve Bayes algorithm [14].

Finally, previous work was developed by improving research procedures and improving classification performance. The research procedure was developed using the Nassi-Shneiderman framework. The model performance is built using the K-Nearest Neighbour (kNN) classification algorithm. Also, this algorithm attempts for better performance by using distance weights. Finally, we tested these two algorithms using cross-validation. Apart from using data from the election campaign in Indonesia, this research also uses data from the presidential campaign in America in 2020. The data taken from the campaign accounts of Obama, Joe Biden, and Kamala Harris are also tested to ensure this research can provide a good contribution and novelty.

The discussion in this paper is divided into several sections. Section 2 discusses the state-of-the-art. Section 3 explains the proposed methods. Nassi-Shneiderman Framework will be used in four experiments consisting of two datasets. Section 4 discusses the results and discussion of this research. Then close with conclusions in the last section.

2. Previous research

Several Previous research about campaigns and classification and frameworks have been described in this section.

2.1 A campaign in social media

Facebook is a social networking application that has the largest users. At the beginning of 2020, there were no less than 2.5 billion active monthly users from a personal friendship web application, and it develops into a group to have a marketplace place. This feature has opened opportunities as a marketing tool by promoting any product or service through social media platforms [15]. This pattern is known as social media promotion [16], which is then used in conducting political campaigns on social media through the Facebook application.

An Instagram application with the characteristics of the image is also used as a promotional and marketing media. A study shows the use of endorsements on Instagram. The endorsement is made to raise the number of users to bring public attention. But many Instagram accounts have fake accounts. This study can detect and classify genuine followers and fake followers using web scraping [17].

Twitter is a social media application that is unique in communication patterns [18]. Individuals who want to interact on Twitter must have an account. Furthermore, the account must be known by other account holders, to follow the account voluntarily. A Twitter account should ideally have as many followers as possible to get broad interaction. So that the Twitter account owner also needs to promote his account to be known in the community through various activities.

Campaigns on Twitter have a unique role in all countries [19]. Twitter was originally designed for users to chat briefly, known as tweets. This design was liked and used widely in campaigns. Various posts can be done to attract interest in interacting. Interactions that occur on the Twitter application include online responses to comments or sharing with other parties with shares. The frequency of this interaction can be collected and studied as a campaign pattern [11].

Twitter has been introduced as a media campaign in various countries. The increasing number of young voters who are digital natives also encourages the use of this application [4]. The study of the presidential election in Cyprus shows that many political parties use it for a one-way communication pattern. Several candidates used it to disseminate information on their programs and campaigns. The form of communication has become more dynamic and active and has increased interaction with voters [6]. This dynamic interaction was also demonstrated in a study on the 2016 presidential campaign in America. This research shows the dynamic contribution of voters in providing information related to candidates through

updating on Wikipedia [20]. This campaign has become a dynamic political communication between candidates. Various communication strategies, such as psychometric, are used to draw individual attention to a candidate. This strategy includes the dissemination of information and disinformation on social media as a means of increasing interaction [21]. However, these efforts and studies have not shown how to measure campaign success.

A study has been conducted to analyze social media networks in Australia. The study was conducted in the Australian general election with four parties. This study successfully identified the lines of communication and the mobilization of voters via Facebook and Twitter [22]. Mobilization through party organizations can be proven through the results of this study. Social media succeeded in bringing the direction of mobilization to social movements rather than political movements. However, this research has not revealed the intensity of interaction that can be used in scaling the gain of campaigns through social media.

Research [23] shows the use of several classification methods to recognize the personality of Twitter users on voting techniques. However, this previous study has not shown the level of interaction of the responses of existing posts. So, this study is expected to be able to make additional contributions to interactions on Twitter.

A study shows the quality of communication through a community's Twitter data. The quality of communication is measured by the number of words for each tweet which are then given a low and high score. This study has not shown the quality of interaction through responses in the form of likes or shares [24]. However, this study inspires to measure interactions more fully by involving responses from users.

2.2 Classification

Classification is a way in Data Mining to classify the data we will analyze. Data that is classified with a supervised pattern always has a label in each class [25]. The purpose of this classification is to find patterns from data then analyze and help solve problems. Membership of each data in a class is mutually exhaustive and mutually exclusive. So that each data can only be a member of a class. A data member must not be a member of more than one class or not be a member of any class.

Various classification methods use a similar approach to learning. Each classification method uses a learning algorithm to form a model that explains the relationship between variables (attributes) and

classes (targets) of data that the class already knows. The resulting model is expected to be able to recognize new data that is not yet known to its class. So, the primary purpose of the classification model is to form a model that has useful generalizations.

In general, the classification method can form a model. This model is a function of mapping the input-output of training data and uses a model that has been formed to estimate the output of new input. For example, in the Decision Tree method, the model is in the shape of a tree while in the Naïve Bayes method, the model used is a probability function.

Several previous studies related to data classification on Twitter have been conducted using various algorithms. A study with data from Twitter has been conducted related to spam detection. This spam can be a severe disruption of information dissemination. The results of this study resulted in two contributions, which were done by decision tree and random forest algorithms. The first contribution is sentiment analysis by classifying spam/non-spam information. The second is to produce a classification set feature that is useful for detecting tweets in a spam /non-spam group by comparing it with the dataset that has been generated[26]. However, this study only classified two labels, spam/non-spam, also how much interaction has not occurred.

Other studies have shown the use of Naive Bayes in Sentiment Analysis. The study was conducted on travel applications with destinations in several cities in India. Cities that were the object of this study were Bangalore, Chennai, Delhi, Gurugram, J&K, Mumbai, and Noida. This analysis was carried out with the help of the R. Naive Bayes statistical program used in classifying opinions and sentiments on Twitter. There are four types of labels used, namely positive, negative, neutral, joy, surprise, sadness, anger, disgust, and fear. The results of this study indicate the satisfaction of visitors to the destination city above. The accuracy resulting from this classification is 83%. This study shows that data from Twitter can be classified to help measure customer satisfaction [27]. However, this research has not been able to show how the campaign plan and the measure of the success of the interaction.

The Ganjar Pranowo dataset for this research has been used before using Naïve Bayes [28]. The results obtained show an accuracy of 76.74%. The next step, Naïve Bayes Kernel Estimator is used to improve performance, which gets a yield of 80.14%. From the results of previous studies [28], and referring to a study related to the classification of spam detection on Twitter [29], the k-Nearest Neighbour (kNN) shows better results. A study of the 2019 presidential election in Spain showed the use of the kNN was

```
#Distance Weight k-nearest neighbour

Define DWkNN(p,t,k)
#Parameters
#p: training points
#t: a table of data
#k: number of nearest neighbour
|
distances ← array_of_data
while group in p
  while feature in p[groups]
    #calculate the distance with euclidean
    euclid ← squareroot((feat[0]-p[0])**2
      +(feat[1]-p[1])**2)
    #add a table of distance
    distances.add(euclid, groups)

#sort the distance and choose the first k distance
distances ← sorted(distances)[:k]

frequency_a ← 0 # weighted sum of groups 0
frequency_b ← 0 # weighted sum of groups 1

while j in distances
  if j[1] equal to 0:
    increment frequency_a with (1 /j[0])
  else if j[1] equal to 1:
    increment frequency_b with (1 /j[0])

return 0 if frequency_a>frequency_b else 1
```

Figure. 2 Pseudo-code of distance weight kNN

able to achieve an accuracy of up to 95% [30]. So, the kNN was chosen as the algorithm used in the development of this model. The kNN also has the ability to good artificial intelligence. The k-Nearest Neighbour (kNN) method works based on the assumption that the data will be in the same class as the surrounding data. So, kNN works based on the principle of nearest neighbor by the name it bears. The focus of the neighborhood in kNN is based on the principle of distance.

Distance measurement can use various distance formulations. Some distance formulas that can be used are Euclidean, Manhattan, Mahalanobis, or Minkowski. The most commonly used distance formula is to use the Euclidean distance formula as in Eq. (1):

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

This mathematical formula (1) shows that d is the distance, while p and q represent the two points measured by the distance. Distance measurement with the Euclidean equation [31] is widely used because of the concept of a straight distance between two points to produce the smallest value.

2.3 Distance weight

The Distance Weight kNN (DWkNN) modification to the distance by giving more weight to closer points. Then give a lighter weight to the point

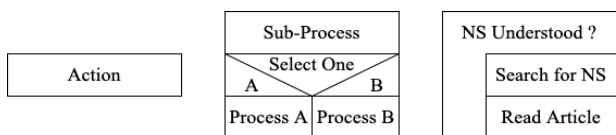


Figure. 3 Nassi-shneiderman block: standard process block, simple branching block, and test-first loop block

that it is in a further position. This concept is known as distance weight [32]. Various functions can be used in weighting. The simplest function is to use the inverse from a distance, which is used in this study. The detail of the pseudo-code of Distance Weight kNN can be seen in Fig. 2.

2.4 Nassi-shneiderman diagram

The flow of problem-solving in a research process can be solved using various frameworks [33]. This framework can be a step guide in conducting studies. One framework that can be used is the Nassi-Shneiderman diagram.

This diagram was developed by Isaac Nassi and Ben Shneiderman in 1972. Nassi-Shneiderman is one technique to improve user understanding of logical processes by using diagram components [34]. This framework will be proposed for use in this study.

Nassi-Shneiderman diagrams can be arranged from several process blocks, such as:

1. Standard process block
The process blocks are steps that must be done and do not require any conditions. If a process block is found, the command in the block is done until it is finished. Then, we can go to the next block. This block diagram can be seen on the left side of Fig. 3
2. Simple branching block
The simple branching block is a block that gives the program flexibility to choose two paths of choice. The path chosen depends on the conditions that are met in the statement. This block diagram can be seen in the center side of Fig. 3
3. Test-first loop block
This block allows the action to be repeated repeatedly until the requested conditions are met. In the test-first loop block, the action in it will be carried out, then check with the requested conditions. If the conditions requested have not yet been fulfilled, the action will continue until the conditions are met. The diagram used is like on the right side of Fig. 3.

2.5 Validation

The development of a model will be implemented if the study can ensure its accuracy through a validation process. In machine learning models performance can be very high but does not guarantee its performance for new data[35]. Validation must be done correctly to minimize bias and variance. Bias is the difference between the actual value and the model output. Variance is the difference between the model output in the training data and test data.

The validation process is testing a model with all or part of the existing data in training. Another way that can be taken is to use new data in training. Some validation techniques that have been known include substitution and n-fold cross-validation, the simplest technique is a substitution, using all data to train the model. So, the output level is evaluated based on the output and the actual value of all the data. Whereas the n-fold cross-validation technique divides the D dataset randomly into n subsets (usually called fold) that are mutually independent; f_1, f_2, \dots, f_n so that each fold contains 1/n parts of data.

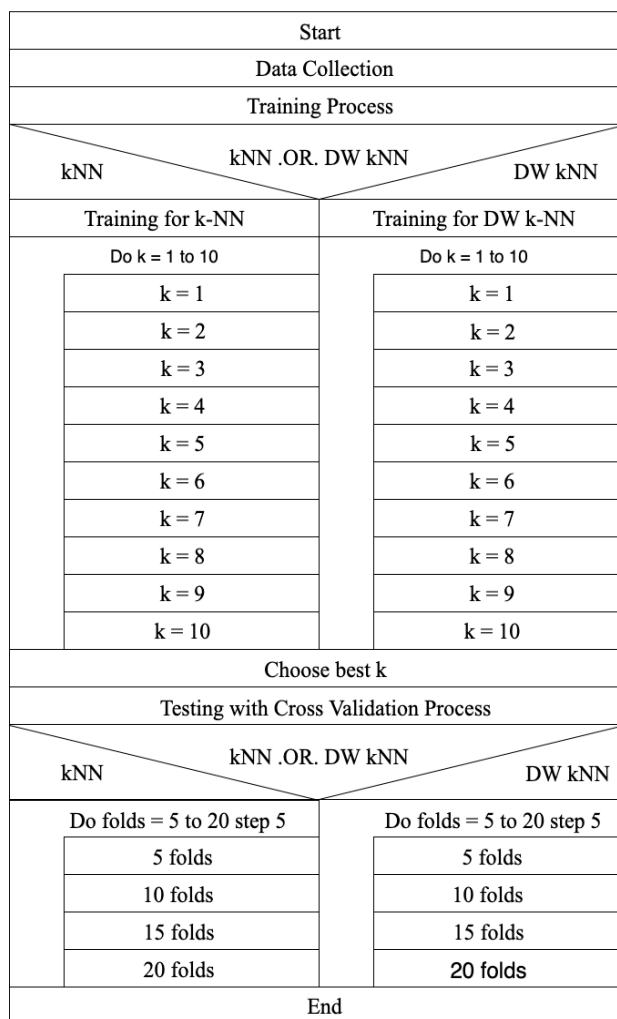


Figure. 4 Nassi-shneiderman framework

Furthermore, a set of datasets is built: D_1, D_2, \dots, D_n which each contains $(n-1)$ fold for training data and one fold for test data[36].

3. Methodology

Following the discussion above, this study comparing an improvement of the kNN method. The method comparing kNN and distance weight k-nearest neighbor (DWkNN).

Meanwhile, the research step will use the Nassi-Shneiderman framework. The framework used is illustrated by the Nassi-Shneiderman Framework shown in Fig. 4. In Fig. 4 stages start from the Start Process Block and end at the End Process Block. The stages of which will be explained in the following sections.

3.1 Data collection

The first phase of the study was to collect data on campaign activities that took place in the province of Central Java, Indonesia. Data was taken from a @ganjarpranowo Twitter account. This is a personal account owned by Ganjar Pranowo, a public figure who had been the governor. Ganjar Pranowo is running again to become governor of Central Java province in the next leadership period. This figure is very active in telling his daily activities. This post can evoke the attention of his followers. Data collection from this account was carried out during the campaign period. Retrieval of data using existing regulatory procedures on Twitter. Twitter allows data retrieval using the application program interface (API) interface that has been provided, by registering in advance.

Table 2. Attribute and label description

Type	Name	Description	Data Type
Attribute	Like	Number of likes in a post	Numeric
Attribute	Comment	Number of comments in a post	Numeric
Attribute	Share	Number of shares in a post	Numeric
Label	Interaction	Low/ Medium/ High	String

The next stage is data cleaning. This stage was carried out to remove several attributes that are not related to this study. This study aims to classify interactions.

So, the result of cleaning up this data results in a dataset containing three attributes. Then from the three attributes that have been obtained are given a label. The dataset attributes and their labels can be described in Table 2. Description of Attributes and Labels. This dataset has three attributes with names like, comment, and share with the label in the form of interaction. Each attribute uses a numeric data type that indicates the number of activities.

A similar procedure was performed on the collection of Donald Trump, Joe Biden, and Kamala Harris datasets. Some of these datasets will be trained and tested to demonstrate the reliability of the proposed model.

3.2 Training process

In this stage, we will conduct training from the dataset using two-choice methods, namely conventional kNN and distance weight kNN (DWkNN). Both options will be made by conducting training at $k = 1$ to $k = 10$ and recording the training results.

3.3 Choose the best k

In this Process Block, the best results from a series of training sessions will be selected for both methods. The best percentage value will be selected as a reference in determining the value of k .

3.4 Testing with cross-validation

Validation is the process of testing the performance of an algorithm. The validation method used in this research is cross-validation. The working principle of cross-validation is to divide the dataset into two parts, namely training data and testing data. At 5 folds cross-validation, the data is divided into five parts. Then 80% of the data is used for training and the remaining 20% as testing. This process will be carried out up to 5 times until all data records get turned into data testing. This study will use folds between 5 and 20 with intervals of 5 in cross-validation with similar logic.

4. Results and discussion

The stages of the study were carried out following those described in the Nassi-Shneiderman Framework.

Table 3. Training results

k (1)	kNN Precision (2)	DWkNN Precision (3)	Best Result (4)
1	100.00	100.00	(2),(3)
2	99.97	99.98	(3)
3	99.89	99.93	(3)
4	99.82	99.87	(3)
5	99.82	99.87	(3)
6	99.71	99.75	(3)
7	99.75	99.75	(2),(3)
8	99.60	99.60	(2),(3)
9	99.60	99.60	(2),(3)
10	99.51	99.66	(3)

This dataset is then used in training using two methods, namely conventional kNN and DWkNN. The results obtained at the training stage are tabulated in Table 3.

In evaluating the performance of supervised learning algorithms such as kNN, the Confusion Matrix reference is used. Confusion Matrix represents the predictions and actual conditions of the data generated by the algorithm. The best accuracy is 100% while the worst is 0%.

Based on Table 3, it can be concluded that changes in the k value of kNN result in a decrease in algorithm performance. At k = 1 it reaches the highest value then decreases. At k = 4 it decreases quite drastically and then steadily decreases to k = 10. While the DWkNN tends to decrease at k between 1 to 9 then slightly increases at k = 10.

So, from the results of the training above, the one with the best performance is chosen. At k=1 and k=2 have nearly the same good value. However, because the value of k = 1 was feared of overfitting, it was decided to take k = 2. After k is determined, the next step is testing with n-folds cross-validation. The test is carried out using n-fold values between 5-20 with intervals of 5 with the results tabulated in Table 4.

Following the results shown in Table 4, kNN has the highest precision value of 99.88% in n-folds = 15. Whereas the DWkNN precision shows rising results from n folds 5 up to 15. But then it drops at n folds 20. The highest precision DWkNN results occur at n-folds 15 which is 99.89%.

In general, for n-folds 5 to 15, the best value is

Table 4. n-fold cross-validation

n-folds	kNN Precision	DWkNN Precision	Best Result
5	99.86	99.87	DWkNN
10	99.84	99.84	DWkNN
15	99.88	99.89	DWkNN
20	99.84	99.84	k-NN DWkNN

Table 5. Training with k=2

Data set	kNN	DWkNN
Donald Trump	99.29	100
Joe Biden	98.97	100
Kamala Harris	98.93	100

Table 6. Testing with 20 folds cross-validation

Data set	kNN	DWkNN
Donald Trump	98.37	98.70
Joe Biden	97.70	97.95
Kamala Harris	97.56	97.78

obtained by the DWkNN algorithm method. In comparison, the highest 20 n-folds were obtained by the kNN and DWkNN algorithm method. The results of this study can guide the success of campaigns related to social networks.

Furthermore, experiments were also carried out using a dataset taken from Twitter for Donald Trump, Joe Biden, and Kamala Harris on the American Presidential Campaign in 2020. Experiments were carried out in the same stages, in training with k = 1 to 10 and testing with cross-validation with n = 5 to 20 with intervals of 5.

The accuracy results obtained in the training experiment are summarized by showing the best results, namely, k = 2 which can be seen in Table 5. While the testing experiments are summarized by only displaying the best accuracy results at n = 20 as shown in Table 6.

From Table 5 it can be concluded that the DWkNN can give better results than KNN with 100% accuracy at k = 2. Whereas in Table 6, testing using 20 folds cross-validation shows that DWkNN is superior in accuracy for all datasets.

The results of the experiment using the dataset of Ganjar, Donald Trump, Joe Biden, and Kamala Harris show that DWkNN provides the best results in both the training model and its testing.

This study shows that the model built with the DWkNN can provide new solutions and contributions compared to previous studies, as shown in Table 7.

Table 7. Comparing with previous research

Previous Research	Classification Method	Accuracy %
[13]	Naïve Bayes	76.74
[27]	Naïve Bayes Kernel Estimator	80.14
[30]	kNN	86

Several previous studies have shown results ranging from 76.74% with the Naïve Bayes method to 86% using kNN. So, that the results of this study have been able to provide better results.

5. Conclusion

The results of this study indicate that the DWkNN model development has succeeded in providing better accuracy. These results can be achieved in the training and testing stages. The choice of $k = 2$ is carried out with the precautionary principle to avoid the possibility of overfitting when using $k = 1$. However, $k = 2$ shows good results for the entire dataset. The testing stages using cross-validation showed good accuracy results in various n-fold scenarios between 5 and 20, and got the best results at 20-folds.

The contributions made to this research were in the form of datasets from various campaign activities, and modeling using the DWkNN. An additional contribution is the use of the Nassi-Shneiderman Framework. This framework can help guide researchers in simple and easy to follow form.

Overall, this research can answer existing problems and provide results in the form of the availability of guidelines that can measure the success of political campaigns on social media. Experiments have also been conducted using other campaign datasets to prove that this model is universally applicable to various campaign activities. That it provides novelty in the form of a campaign interaction guide model for the campaign team.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Aryo Nugroho and Mauridhi Hery Purnomo; methodology, Aryo Nugroho, Surya Sumpeno and Mauridhi Hery Purnomo; software, Aryo Nugroho; validation, Aryo Nugroho, Surya Sumpeno and Mauridhi Hery Purnomo; formal analysis, Aryo Nugroho and Surya Sumpeno; investigation, Aryo Nugroho; resources, Aryo Nugroho, Kimiya Fujisawa, Surya Sumpeno and Mauridhi Hery Purnomo; data curation, Aryo

Nugroho; writing—original draft preparation, Aryo Nugroho; writing—review and editing, Aryo Nugroho, Kimiya Fujisawa, Surya Sumpeno and Mauridhi Hery Purnomo; visualization, Aryo Nugroho; supervision, Kimiya Fujisawa, Surya Sumpeno, Mauridhi Hery Purnomo; project administration, Aryo Nugroho and Mauridhi Hery Purnomo; funding acquisition, Aryo Nugroho, Surya Sumpeno and Mauridhi Hery Purnomo.

Acknowledgments

The first author would like to express my sincere gratitude for BPPDN Scholarship. Further, the first author would like to thank AONO Cloud Experience for the infrastructure of this research. To conclude, we cannot forget to thank our family and friends for all the unconditional support in this very intense research.

References

- [1] A. G. Sooi, A. Nugroho, M. N. A. Azam, S. Sumpeno, and M. H. Purnomo, "Virtual artifact: Enhancing museum exhibit using 3D virtual reality", In: *Proc. of 2017 TRON Symposium (TRONSHOW)*, pp. 1–5, 2017.
- [2] K. Michael, "Are You Addicted to Your Smartphone, Social Media, and More? The New AntiSocial App Could Help", *IEEE Consumer Electronics Magazine*, Vol. 6, No. 4, pp. 116–121, 2017.
- [3] A. A. Alalwan, N. P. Rana, Y. K. Dwivedi, and R. Algharabat, "Social media in marketing: A review and analysis of the existing literature", *Telematics and Informatics*, Vol. 34, No. 7, pp. 1177–1190, 2017.
- [4] J. Ohme, "When digital natives enter the electorate: Political social media use among first-time voters and its effects on campaign participation", *Journal of Information Technology & Politics*, Vol. 16, No. 2, pp. 119–136, 2019.
- [5] W. A. Kristiana, M. M. Achlaq, B. Anindito, A. Nugroho, C. Darujati, and M. N. A. Azam, "UID Beacon Advertisements for Lecture Schedule Information", In: *Proc. of 2018 5th International Conf. on Electrical Engineering, Computer Science and Informatics (EECSI)*, pp.270–276, 2018.
- [6] D. Ktoridou, E. Epaminonda, and A. Charalambous, "Social Media in Election Campaigns: The Case of the 2013 Presidential Elections in Cyprus", *IEEE Technology and Society Magazine*, Vol.37, No. 2, pp.32–39, 2018.

- [7] T. Paterson, “Indonesian cyberspace expansion: a double-edged sword”, *Journal of Cyber Policy*, Vol. 4, No. 2, pp. 216–234, 2019.
- [8] A. Nugroho, D. Rizaludin, S. Soebandhi, L. Junaedi, S. Winardi, and M. N. Al-Azam, “Automatic Sign of Commencement of Work from Enterprise Resource Planning”, In: *Proc. of 2020 International Conf. on Smart Technology and Applications (ICoSTA)*, pp. 1–6, 2020.
- [9] C. Fox, “Candidate-centric systems and the politicization of ethnicity: evidence from Indonesia”, *Democratization*, Vol. 25, No. 7, pp. 1190–1209, 2018.
- [10] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis”, *Journal of Big Data*, Vol. 5, No. 1, p. 51, 2018.
- [11] A. Khan, J. P. Li, N. Ahmad, S. Sethi, A. U. Haq, S. H. Patel, and S. Rahim, “Predicting Emerging Trends on Social Media by Modeling it as Temporal Bipartite Networks”, *IEEE Access*, Vol. 8, pp. 39635–39646, 2020.
- [12] Mirwan, A. Nugroho, F. Hendarta, R. Hidayatillah, F. Hassan, and K. P. Nana, “Virtual Assistant Using Lstm Networks in Indonesian”, In: *Proc. of 2018 International Seminar on Research of Information Technology and Intelligent Systems*, pp. 652–655, 2018.
- [13] I. F. Roji, S. W. Sihwi, and A. Doewes, “Development of Regional Public Election Analytics Application in Indonesia Using Data from Twitter”, In: *Proc. of 2018 International Seminar on Research of Information Technology and Intelligent Systems*, pp. 394–399, 2018.
- [14] R. Hidayatillah, M. Mirwan, M. Hakam, and A. Nugroho, “Levels of Political Participation Based on Naive Bayes Classifier”, *Indonesian Journal of Computing and Cybernetics Systems*, Vol. 13, No. 1, pp. 73–82, Jan. 2019.
- [15] F. I. Puspitasari, S. Soebandhi, D. H. Subagyo, and A. Nugroho, “Altruism and Egoism in e-WOM: The moderating effect of Facebook perceived ease of use”, In: *Proc. of 2019 International Seminar on Application for Technology of Information and Communication*, pp. 536–541, 2019.
- [16] S. Z. M. Zain, A. R. C. Hussin, and M. H. Selamat, “Derivation of Hashtag (#) Factors for Hashtag Marketing Model (HASHMAM) in Social Media Platform”, In: *Proc. of 2019 6th International Conf. on Research and Innovation in Information Systems*, pp.1–6, 2019.
- [17] M. I. Akrianto, A. D. Hartanto, and A. Priadana, “The Best Parameters to Select Instagram Account for Endorsement using Web Scraping”, In: *Proc. of 2019 4th International Conf. on Information Technology, Information Systems and Electrical Engineering*, pp. 40–45, 2019.
- [18] B. Alkouz, Z. A. Aghbari, and J. H. Abawajy, “Tweetluenza: Predicting flu trends from twitter data”, *Big Data Min. Anal.*, Vol. 2, No. 4, pp. 273–287, 2019.
- [19] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang, “Quantifying Political Leaning from Tweets, Retweets, and Retweeters”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 8, pp. 2158–2172, Aug. 2016.
- [20] K. C., “The Dynamics of Peer-Produced Political Information During the 2016 U.S. Presidential Campaign”, In: *Proc. of the ACM on Human-Computer Interaction*, 2019.
- [21] M. Bay, “Social Media Ethics: A Rawlsian Approach to Hypertargeting and Psychometrics in Political and Commercial Campaigns”, *ACM Transactions on Social Computing*, Vol. 1, No. 4, pp. 1–14, 2018.
- [22] J. McSwiney, “Social networks and digital organisation: far right parties at the 2019 Australian federal election”, *Information, Communication & Society*, Vol. 0, No. 0, pp. 1–18, 2020.
- [23] Modern Academy for Computer Science and Management Technology, D. Sameh, G. Khoriba, Helwan University, M. Haggag, and Helwan University, “Behaviour Analysis Voting Model Using Social Media Data”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 2, pp. 211–221, 2019.
- [24] B. S. Abdur Rahman Crescent Institute of Science & Technology, M. Khatoon, W. Banu, and B. S. Abdur Rahman Crescent Institute of Science & Technology, “An Effective Way of Detecting Communities in Social Network”, *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 4, pp. 199–211, 2018.
- [25] Beni-Suef University, N. Hassan, W. Gomaa, Beni-Suef University, G. Khoriba, Helwan University, M. Haggag, and Helwan University, “Credibility Detection in Twitter Using Word N-gram Analysis and Supervised Machine Learning Techniques”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 1, pp. 291–300, 2020.
- [26] M. Ahsan and T. P. Sharma, “Spams classification and their diffusibility prediction on Twitter through sentiment and topic models”, *International Journal of Computers and Applications*, Vol. 0, No. 0, pp. 1–11, 2020.

- [27] P. Bhardwaj, S. Gautam, and P. Pahwa, “A novel approach to analyze the sentiments of tweets related to TripAdvisor”, *Journal of Information and Optimization Sciences*, Vol. 39, No. 2, pp. 591–605, 2018.
- [28] A. Nugroho, R. Hidayatillah, S. Sumpeno, and M. H. Purnomo, “Klasifikasi Interaksi Kampanye di Media Sosial Menggunakan Naïve Bayes Kernel Estimator”, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, Vol. 8, No. 2, pp. 107–114, 2019.
- [29] A. Pinandito, R. S. Perdana, M. C. Saputra, and H. M. Az-zahra, “Spam detection framework for Android Twitter application using Naive Bayes and K-Nearest Neighbor classifiers”, In: *Proc. of the 6th International Conf. on Software and Computer Applications*, Bangkok, Thailand, pp. 77–82, 2017.
- [30] D. Grimaldi, J. D. Cely, and H. Arboleda, “Inferring the votes in a new political landscape: the case of the 2019 Spanish Presidential elections”, *Journal of Big Data*, Vol. 7, No. 1, p. 58, 2020.
- [31] A. E. Geyer, R. Nikjah, S. A. Vorobyov, and N. C. Beaulieu, “Euclidean and Space-Time Block Codes: Relationship, Optimality, Performance Analysis Revisited”, *IEEE Transactions on Communications*, Vol. 63, No. 8, pp. 2912–2923, 2015.
- [32] B. Wang, X. Gan, X. Liu, B. Yu, R. Jia, L. Huang, and H. Jia, “A Novel Weighted KNN Algorithm Based on RSS Similarity and Position Distance for Wi-Fi Fingerprint Positioning”, *IEEE Access*, Vol. 8, pp. 30591–30602, 2020.
- [33] R. L. Purvis and W. Leigh, “Adapting Structured Analysis & Design Tools for Event-Driven Systems”, *Journal of Computer Information Systems*, Vol. 37, No. 3, pp. 21–28, 2016.
- [34] J. Jurinová, “Increasing IT skills in algorithmization for students of informatics”, *R&E-SOURCE*, 2017.
- [35] P. Tamilarasi and R. U. Rani, “Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning”, *2020 Fourth International Conf. on Computing Methodologies and Communication*, pp. 1034–1038, 2020.
- [36] T.-T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation”, *Pattern Recognition*, Vol.48, No. 9, pp. 2839–2846, 2015.