Fanaticism Category Generation Using Tree-Based Machine Learning Method

A S Aribowo1,2, H Basiron1, N S Herman1, S Khomsah3

1Faculty of Information & Communication Technology Universiti Teknikal Malaysia Melaka, Malaysia.
2Universitas Pembangunan Nasional “Veteran” Yogyakarta, Indonesia
3Institut Teknologi Telkom Purwokerto, Jawa Tengah, Indonesia

Abstract. This research has produced categorization of political fanaticism on social media. The results of this study divide political fanaticism into 4 categories: positive, neutral, negative and very negative. The background of the research is that currently social media cannot recognize the elements of fanaticism and does not care about the spread of negative fanaticism. One of the best solutions to prevent the spread of negative fanaticism on social media is to develop a computer system which is able to recognize, categorize fanaticism and reject the entry of negative fanaticism. This research extracts features of sentiment, emotions, and expressions of hatred in opinions on social media. The results of the feature extraction are then processed into knowledge using tree-based machine learning methods, namely Random Forest and ID3. The validation result shows that the knowledge generated by the Random Forest process outperforms other techniques with an accuracy of 91.82%. Knowledge from the Random Forest can be used to categorize fanatical texts based on sentiment, emotion and hatred in further research.

1. Introduction

This article presents the results of the first stage of this study about political fanaticism categorization by using sentiment and emotion analysis on social media. This research aims to generate fanaticism knowledge from a fanatic text on social media. The fanaticism knowledge is useful for creating fanaticism level which will be used in the next stage of research, particularly about the political fanaticism detection model by implementing sentiment and emotion analysis. The data obtained from the first stage will be used as data training. If the second research is successful, the model can be used in the third stage of research about political fanaticism monitoring on social media.

1.1. Research Background

Social media are increasingly used in political communication. Indonesia entered the political year of 2019. Political leaders communicate with their supporters through social media. People use social media to give influence about political party or politicians to the other people. Political fanaticism arises because of excessive admiration, lack of sufficient use of reason and logic to idolize a politician or political party. This fanaticism can be positive if it is not destructive, does not attack other politicians, and does not contain hate speech. Fanaticism will be negative if it contains blasphemy, insults, hate speech towards certain characters or followers. Many negative cases arise due to this negative fanaticism, starting from friendships breakup [1], social unrest, and even murder [2].
1.2. Research Problem
One of the weaknesses of social media is that social media cannot recognize and categorize the existence of negative fanaticism in opinion texts. Social media does not care about the spread of political fanaticism. How to recognize an opinion text on social media, then classify the opinion text into positive or negative fanaticism. If negative fanaticism is identified, it can be rejected and cannot be disseminated. Social media is a media communication controlled by computers. Humans can recognize negative fanaticism in opinion, but humans cannot monitor every opinion that exists in social media because the number of opinions disseminated through social media is very large.

1.3. Problem Solution
One of the best solutions to prevent the spread of political fanaticism on social media is to develop a computer system that can recognize negative fanaticism. This system will classify a text containing negative fanaticism. The study about how a computer can recognize the meaning of the text is called text mining techniques. One part of text mining is sentiment analysis and emotion analysis. Sentiment analysis will calculate how positive and negative the polarity of the sentiment of a text. Emotion analysis will analyze what types of emotions are present in a text. In this research, detecting political fanaticism in opinion is by extracting sentiment and emotion features from an opinion on social media. These features are classified as positive or negative fanaticism groups. These features can be used to create fanatic text classifier that is useful to identify another text with fanatic tendencies.

2. Literature Review

2.1. Fanaticism
Webster’s dictionary defined a fanatic as a person with uncritical enthusiasm, extreme belief, whether in religion or politics. Several experts stated that the term fanaticism comes from the Latin adverb fānāticē (frenziedly, ragingly) and the adjective fānāticus (enthusiastic, ecstatic, furious)[1]. Some experts on fanaticism provide a philosophical definition of fanaticism and do not see fanaticism as a series of events that are not randomly connected. They regard it as a result of a disability of rationality. Philosophers treat fanaticism as a product of irrational commitment supported by religious dogma [2].

Fanaticism is extreme enthusiasm, emotion, and a sense of excessive love and interest due to people’s belief in a fanatical object (i.e. politician or public figure), and they tend to defend their beliefs [3]. There are several types of fanaticism such as the fanaticism of religion, group, ideology, ethnicity, race and political fanaticism. Fanaticism is not always negative but can also be positive [1]. Fanaticism in politics does exist and even a political expert says that fanaticism is emotion in politics [4]. Positive political fanaticism is showed by positive behaviors and certainly very constructive for both politician and supporters. The positive side is also good for the nation’s security and it is regarded as harmless. Negative political fanaticism will result in dissent, hatred, and hostility, which arise directly or through the communication media. Even fanaticism is the root of racism and radicalism [5]. The adverse effects of negative political fanaticism come to physical conflict and civil war [6].

Based on the research above, the study of fanaticism already existed and each expert has his perception. It can be concluded that fanaticism means excessive enthusiasm, extreme enthusiasm that is not critical, emotion, and excessive love and, interest in believing in fanatical objects and they tend to defend their beliefs without using the right reasons. There is little research that clearly explains what kind of emotion exists in fanatical texts and what emotion level exist in fanatical texts.

2.2. Fanaticism Detection on Text
Some studies categorize fanaticism in texts into 3 groups, namely Non-Fanatic, Code Attitude Fanaticism (CAF), and Code Red Fanaticism (CRF). The CAF group is fanaticism that indicates signs of hostility, but no violent tendencies. CRF is fanaticism that tends to use violent actions against adversaries. This research classifies the fanaticism using Case-Based Reasoning (CBR) and Naïve Bayes Classification [7].
This research was followed by a study on fanaticism detection in documents by using questioner. Fanaticism is also categorized into 3 groups, namely Non-Fanatic, Code Attitude Fanaticism, and Code Red Fanaticism. In this research, text classification uses machine learning (Random Forest) techniques [8].

This research was followed by a study about fanaticism detection by using a collected article from online Arabic newspapers and channel archives. Articles fall into two categories of classification only (fanatic and non-fanatic). This study used a technique that combines term-based analysis and background knowledge with boosting to improve the classification process. The machine learning applied three rule-based classification algorithms, namely, C4.5, RIPPER, and PART [9].

This research was continued in 2017 to analyze the detection and classification of fanaticism in texts from the Arabic-language Twitter media. This research developed a model to categorize whether a Twitter message is fanatic or non-fanatic. This work combined any classification algorithms, namely, TF-IDF, Support Vector Machine, and Naïve Bayesian with contextual knowledge and boosting to identify tweets of fanaticism. The maximum accuracy obtained is 82.1% [10].

The latest research about fanaticism detection and analysis needs to be improved on several sides:
1. In article [10] the detection of fanaticism is limited to the extraction of keyword features from the text. Fanaticism is emotion, it is necessary to involve emotion analysis to detect emotions in fanatic texts. In my article, emotion elements will be involved to improve accuracy.
2. Extension of categorization of fanaticism. Fanatical and non-fanatic categories in [10] cannot be used to categorize all forms of fanaticism. This category cannot accommodate positive fanaticism and is also not associated with emotion.

2.3. Sentiment Analysis
Sentiment analysis is the extraction of information aiming to extract information about the author’s feelings, both positively and negatively, by analyzing a large number of documents [11]. Sentiment analysis is a research about opinions, sentiments, and emotions that are expressed textually. If given a set of text documents containing opinions about an object, then sentiment analysis aims to extract the attributes and components of the document and to determine whether the comment is positive, neutral or negative [12].

2.4. Emotion Analysis
Emotion analysis is a field of study that analyzes emotions towards entities such as products, organizations, individuals, issues, and events. The emotion analysis is computational research of emotions expressed textually [12].

Many of English emotion analysis studies classify opinion based on Ekman's 6 (six) basic emotions: happiness, surprise, fear, anger, disgust, and sadness [13]. There is a research that categories emotion in tweet data into three-part [14]:
1. Positive emotions, such as like, happy, amazed or joy.
2. Negative emotions, such as sadness, anger or disappointment.
3. Neutral, do not express any sentiment.

Fanaticism is extreme enthusiasm, emotion, and a sense of excessive love and interest to a fanatic object. Based on this definition there are several emotions in fanaticism. Fanaticism analysis will apply emotions feature extraction. Emotions that represent fanatical sentences will be extracted as in emotions analysis. These emotions will be processed into knowledge to classify text fanatics based on emotions. The results will be used to complete and detail the classification process of keyword-based fanaticism analysis.

2.5. Hate Speech Analysis
The sentence containing elements of hostility, bad speech, tend to cause harm directed to a person or group of people is called hate speech [15]. Social media is a place to make, share, and exchange hates messages against groups that are perceived as enemies. Hate speech is directed at a group of people. There is a study mentioning that types of hates speech are related to ethnicity, nationality, religion, social groups, and social class [16].
Researches on the detection of hate speeches have been conducted. Hate speech is categorized into two different types of classification: hate and non-hate. There is also another statement that differentiates hate into three types: strong hate, weak hate, and no hate. [17].

3. Methodology
The essence of this first phase of research is how to make knowledge using the fanatic text feature. The steps of the research process are sorted according to Figure 1.

![Figure 1. Research Step](image)

3.1. Data Crawling from Social Media
This research used textual comments from facebook, twitter, and youtube. Indonesian presidential candidate’s name is used as the keyword during the searching process. Furthermore, the result of crawling is stored in the repository and is used as a dataset.

3.2. Analysis of the Level of Political Fanaticism
This research has prepared 5 (five) levels of fanaticism which will be tested to analyze whether those 5 (five) levels will be fulfilled by data and labelling from experts. The level of fanaticism prepared is as outlined in Table 1, following the philosophy of the experts about the types of fanaticism.

| Level Name             | Code | Information                        |
|------------------------|------|------------------------------------|
| Very Positive Fanaticism | VPF  | Very positive fanaticism          |
| Positive Fanaticism    | PF   | Positive fanaticism               |
| Neutral Fanaticism     | No F | No fanaticism tendency            |
| Negative Fanaticism    | NF   | Negative fanaticism tendencies    |
| Very Negative Fanaticism | VNF  | Negative fanaticism tendency with destruction potency |

3.3. Analysis of sentiments types
Sentiment polarity is divided into 3 levels, namely non-sentiment (neutral), positive and negative as shown in Table 2.
Table 2. Sentiment Polarity

| Polarity  | Information                 |
|-----------|----------------------------|
| Positive  | Positive sentiment          |
| Neutral   | There is no sentiment in the text |
| Negative  | Negative sentiment          |

3.4. Analysis of Emotion Type and Level

Emotion is divided into 6 types according to Ekman's 6 (six) basic emotions such as happiness, surprise, anger, disgust, fear, and sadness[13]. In this study, every type of emotion has 4 (four) levels of polarity as shown in Table 3.

Table 3. Level of Emotion

| Level of Emotion | Code | Information                                           |
|------------------|------|-------------------------------------------------------|
| High             | H    | The emotion value on the text approximately 100%-67%   |
| Average          | A    | The emotion value on the text approximately 66%-34%    |
| Low              | L    | The emotion value on the text approximately 33%-1%     |
| None             | N    | No emotion on the text                                 |

3.5. Analysis of hate speech types

The polarity of hate speeches was divided into two: no hate speeches (N) and hate speeches (Y).

3.6. Labelling

Labelling is the expert justification of the level of fanaticism, sentiment, emotion and hates speech. Labelling is a way to transfer knowledge from experts to machines. The expert will determine the label of each opinion sentences as shown in Figure 2.

Figure 2. Fanatic Text Labelling Form

3.7. Knowledge Generation

Knowledge instigation was done by processing the fanatic text feature in which the polarity of sentiment, the level of emotions, and the existence of hate speeches had been determined previously. These features are processed by using tree-based machine learning methods, namely ID3 and Random Forest. The results of each type of machine learning were tested by implementing Cross Validation and Confusion Matrix. The best machine learning method was the one that provided the highest accuracy. The results of the knowledge generation were the knowledge itself, the best machine learning method (highest accuracy), and fanaticism category.

4. Experiment

4.1. Data Crawling
Opinions were obtained from YouTube video comments, Twitter, and Facebook. The keywords for opinion data collection were the names of Indonesian presidential candidates in the 2019 election, namely Joko Widodo, Jokowi, Prabowo Subianto, Prabowo, Sandiaga Uno, Sandiaga and Maruf Amin. The data were obtained during the campaign period of presidential candidates in December 2018. Then, the opinions obtained were 3478 sentences.

4.2. Data Labelling
Experts observed every opinion text and gave label based on the level of text fanaticism in Table 1. They also justified the polarity of sentiments, determined the level of each type of emotion according to the rules in Table 2, and determined whether there is hate speech exist in it. The sample results from feature justifications are shown in Figure 3.

4.3. Knowledge Generation
Knowledge generation requires a machine learning process and a validity testing process. The machine learning used was ID3 and Random Forest. For comparison, this research also uses some standard machine learnings (Naive Bayes and K-Nearest Neighbor) to show that tree-based machine learning is better. Validity testing was done by applying K-Fold Cross-Validation and the accuracy was measured by using Confusion Matrix. The results of machine learning accuracy testing are depicted in Figure 4.

Based on observations using the 4 methods in Figure 4, it can be concluded that Random Forest provides the highest accuracy of 85.45%, followed by ID3 at 84.55%, K-NN at 82.73% and finally Naïve Bayes at 80.00%. From the test, it can be seen that the level of Very Positive fanaticism has never been properly classified because the class recall value is only 0.00%. Classification of Very Positive classes occupies the same class as Positive. This shows that Very Positive label can be pruned and changed to Positive label. If all Very Positive labels are changed to Positive, it is necessary to calculate the accuracy again. Figure 5 outlines the new accuracy results.
Based on observations conducted by using the 4 methods in Figure 5, in which the Very Positive label has been changed to Positive, it can be concluded that Random Forest provides the highest accuracy of 91.82%, followed by ID3 at 90.91%, Naïve Bayes at 88.18% and the last K-NN at 87.27%. Figure 6 shows the tree formation from one of the most complex trees in Random Forest method.

**Figure 5. Results of Improvement of Classification Tests Using 4 Methods and Accuracy**

![Table of Results](image)

**Figure 6. Tree Formation from Random Forest Method**

The tree formation in Figure 6 can be used as knowledge and categorize the fanatic texts. The root node of the tree is sentiment, which divides the root into 3 branches, namely positive, neutral and negative sentiments. Each branch is divided into the most dominant emotion, namely anger (for...
negative sentiment) and happiness (for positive sentiment), followed by other emotion types and hate-speech. There are several benefits of Random Forest: it can build several short trees in relatively small time, able to generate the rules of prediction created from the data training, and able to reduce the number of tests.

5. Discussion
Knowledge has been successfully generated. The best method to classify and to generate knowledge is Random Forest (accuracy 91.28%). The knowledge can be used as a category for classifying text fanaticism based on sentiment, emotion and hate speech. Some data are not properly classified to the label due to several reasons: Incorrect labelling; and incorrect weight of sentiment, emotion, and hate. Thus, labelling the classification process leads to another label. The research improvement needed by re-labelling or re-weighting all sentiment, emotion and hate speech features on data that are incorrectly classified.

6. Conclusion
This research was conducted to to generate fanaticism categorization in order to classify fanatic texts into four groups namely Positive Fanaticism, Non-Fanatic, Negative Fanaticism, and Very Negative Fanaticism. The knowledge can be used to classify fanatic texts based on sentiment, emotion and hate speech. There are several improvements needed for this research, one of which is how to get unique words for each type of classification class. Text extraction methods such as TF / IDF can be used. This unique word can be used by machine learning to determine the class of fanatic text that does not yet have a class label for the next study.

7. References
[1] K. Marimaa, “The Many Faces of Fanaticism,” in ENDC Proceedings, 2011, vol. 14, pp. 29–55.
[2] B.-K. H. Vo and N. Collier, “Twitter Emotion Analysis in Earthquake Situations,” Int. J. Comput. Linguist. Appl., vol. 4, no. 1, pp. 159–173, 2013.
[3] J. Eliani, M. S. Yuniardi, and A. N. Masturah, “Fanatisme dan Perilaku Agresif Verbal di Media Sosial pada Penggemar Idola K-Pop,” Psikohumaniora J. Penelit. Psikol., vol. 3, pp. 59–71, 2018.
[4] V. Heins, “Reasons of the Heart: Weber and Arendt on Emotion in Politics,” Eur. Leg., vol. 12, no. 6, pp. 715–728, 2007.
[5] S. I. Dewi and A. Aminulloh, “Social Media : Democracy in the Shadow of Fanaticism,” in The 3rd Conference on Communication, Culture and Media Studies, 2016, pp. 79–88.
[6] M. Hughes and G. Jhonson, Fanaticism and Conflict in The Modern Age, vol. 25. London : Frank Cass, 2005.
[7] A. Almonayyes, “Multiple Explanations Driven Naive Bayes Classifier.,” J. off Univers. Comput. Sci., vol. 12, no. 2, pp. 127–139, 2006.
[8] J. Kléma and A. Almonayyes, "Automatic Categorization of Fanatic Text Using Random Forests," Kuwait J. Sci. Eng., vol. 33, no. 2, pp. 1–18, 2006.
[9] A. Almonayyes, “Classifying Documents By Integrating Contextual Knowledge With Boosting,” no. November, pp. 28–29, 2016.
[10] A. Almonayyes, “Tweets Classification Using Contextual Knowledge And Boosting,” no. 4, pp. 87–92, 2017.
[11] S. Mukherjee and P. Bhattacharyya, “Sentiment Analysis : A Literature Survey,” 2013.
[12] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publisher, 2012.
[13] P. Ekman, “An Argument for Basic Emotions,” Cogn. Emot., vol. 6, no. 3–4, pp. 169–200, 1992.
[14] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” in Proceedings of the International Conference on Language Resources and Evaluation, 2010, pp. 1320–1326.
[15] R. Cohen-Almagor, “Fighting Hate and Bigotry on the Internet,” Policy & Internet, vol. 3, no. 3, pp. 89–114, 2011.
[16] N. D. Gitari, Z. Zuping, H. Damien, and J. Long, “A Lexicon-based Approach for Hate Speech
Detection,” Int. J. Multimed. Ubiquitous Eng., vol. 10, no. 4, pp. 215–230, 2015.

[17] F. Del Vigna, A. Cimino, F. Dell’Orletta, M. Petrocchi, and M. Tesconi, “Hate me, hate me not: Hate speech detection on Facebook,” in First Italian Conference on Cybersecurity (ITASEC17), 2017, vol. 1816, pp. 86–95.