DOI: 10.14421/ijid.2020.09108 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 9, No. 1, 2020, Pp. 52-58

Veil and Hijab: Twitter Sentiment Analysis Perspective

Lusiana Lestari[1], M. Didik R. Wahyudi[2], Usfita Kiftiyani[3]
Informatics Department, Faculty of Science and Technology
State Islamic University of Sunan Kalijaga Yogyakarta, Indonesia
[1]10lestarilusiana@gmail.com, [2]m.didik@uin-suka.ac.id, [3]usfita.kiftiyani@gmail.com

Article History
Received July 2nd, 2020
Revised Sept 22nd, 2020
Accepted Sept 22nd, 2020
Published Sept, 2020

***Abstract*— Controversies about veil and hijab are often occur in society. Especially in today's digital era, public opinion expressed through social media can greatly influence the others opinions, regardless of whether it is positive or negative. Therefore, this research was aiming to conduct an approach through analysis sentiment of public opinion about the veil and hijab to know how much accurate the sentiment analysis predict the positive, negative, or other sentiments with using Twitter data as the research object. The algorithm used in this study is Support Vector Machine (SVM) because of its fairly good classification model though it trained using small set of data. The SVM on this research was combined with Radial Base Function (RBF) kernel because of its numerical difficulties that are fewer than linear and polynomial kernel and also because this research doesn't have a large feature. The amount of data used is 3556 tweets data. Tweets data, which is numbered 1056, is classified manually for the learning process. The remaining 2500 data will be classified automatically with the classifier model that has been created. A total of 1056 tweets data that have been classified manually is separated into training and testing data with a ratio of 8: 2. The result of the sentiment analysis process using Support Vector Machine algorithm RBF kernel with C=1 and γ=1 has an accuracy score of 73.6% with precision to negative opinions are 62%, positive opinions are 83%, neutral opinions reach 53% and irrelevant opinions that talk about hijab and veil reach 98%. It shows that sentiment analysis can be used for predicting the negative, positive or other sentiments of a sentence based on a certain topic, in this case veil and hijab.**

Keywords- Sentiment Analysis; Twitter; Support Vector Machine; Veil; Hijab

1 INTRODUCTION

Social media lately has become something that is in great demand, both as a source of information and as a medium for sharing opinions and daily life. Social media nowadays seems to be one of the main needs that cannot be separated from everyday life. Data released by Facebook and Twitter companies, Indonesia is among the top five users of their products. For Facebook, in 2017 it reached 115 million, up 40 percent from 2016, which reached 82 million. The number of Facebook users is expected to continue to grow each year. The same increase also occurred on Twitter in Indonesia. Even though the number is not mentioned, Twitter users in Indonesia are relatively large and 77% are classified as active users[1].

Previously the number of characters on Twitter was limited to only 140 characters, but now the maximum number of characters has been increased to 280 characters. Through Twitter, users can share every day's activity, share photos, or express opinions about something. Usually, a topic that is currently a hot topic on Twitter will be a trending topic. Twitter has become one of the media used by the public to participate in expressing their opinions on topics that are currently being discussed.

The rise of the delivery of opinions through social media Twitter can be one of the media to analyze the tendency of information on a topic whether it tends to be positive, negative, or neutral[2]. Those who need information about public opinion on certain keywords can use Twitter social media as the data analysis.

Veil (*Cadar*) user (*niqabis*) communities often used social media to introduce their community[3]. Along with the increasing popularity of its community, public opinion about veil is also increased. And many of these opinions are expressed through social media including twitter. Based on the meaning of hijab-that are either covers or obstructions- veil can be grouped as one of them [4]. Therefore, the discussion about veil is usually associated with the discussion about hijab.

Even further, the controversy over hijab and veil is growing[4], [5]. Thus, make many people interest to research hijab or veil. Some studies about hijab or veil were discussed through religion and culture perspective[6], norm and historical perspective[4], its function[3], and so on. Therefore, the author conducting research about hijab and veil by analyzing the sentiment impact of twitter users' discussion.

The research from [6] mostly stated about the law of origin from hadith and Qur'an verses, and the cultural origin for a woman wearing veil or hijab. As told in [4], after September 11th, 2001 tragedy that demolishes WTC Building in America, the existence of hijab and veil was greatly rejected almost in all regions of Europe and America, they treat veil as the symbol of terrorists and radicals. And in [3], they said that *niqabis* use social media to show that women that wear veil can still be attractive and make their appearance more powerful and forming a strong character. These three studies show that there is a negative of positive public opinion regarding veil and hijab, which is why the author tries to analyze the sentiment impact of these two words from social media Twitter to know a new perspective.

Sentiment analysis also called opinion mining, is one branch of data mining that aims to analyze, understand, process, and extract textual data in the form of opinions on entities such as products, services, organizations, individuals, and certain topics [7].

Classification is a job evaluating data objects for put it in a certain class from several available classes. In the classification there are two main tasks carried out, namely: first, the construction of the model as a prototype to be stored as memory and second, the use of the model to perform the introduction/classification/prediction of another data object so that it is known which class of data objects are in an easy model saved[8]

Natural Language Processing (NLP) is an attempt to extract a more complete picture of the text. NLP usually using linguistic

concepts such as part-of-speech (nouns, verbs, adjectives, etc.) and grammatical structural (both represented in the form of noun phrases, prepositional phrases, or the existence of dependency relationships such as the subject of or object from) [9].

Research conducted by Buntoro (2017) discusses the sentiment analysis of DKI governor candidates for 2017 on twitter. The data used in this study were tweets in Indonesian with the keywords AHY, Ahok, and Anies totaling 300 tweets. This study uses the Lexicon Based method to determine the sentiment class and the Naïve Bayes Classifier and Support Vector Machine method to classify tweets into positive, negative, and neutral classes. The results of this study show the highest accuracy using the Naïve Bayes Classifier method with 95% accuracy [10].

Novantirani conducted subsequent research. This study analyzes public opinion on Twitter regarding the use of public land transportation within cities. The results obtained that sentiment analysis of Twitter data regarding the use of public land transportation in the city can be done with the Support Vector Machine method, with an accuracy of 78.12% in the dataset Transjakarta. The results of the sentiment analysis can help in the assessment and evaluation of the use of public land transportation in the city [11]

This research will conduct an analysis of the search results for the tweets with the keywords "veil" and "hijab". Therefore, for knowing the tendency of tweet's opinions regarding these two keywords on Twitter, the data will then be classified into four sentiment classes, namely positive, negative, neutral, and irrelevant[12]. The method that will be used in this research is Support Vector Machine. This method produces a good classification model even though it is trained by using a small set of data. It is hoped, with this research, it can help those who need information analysis on certain topics on Twitter social media.

2 METHOD

This research consists of data collection, data selection, data preprocessing, and sentiment analysis process. For more details can be seen in Figure 1.

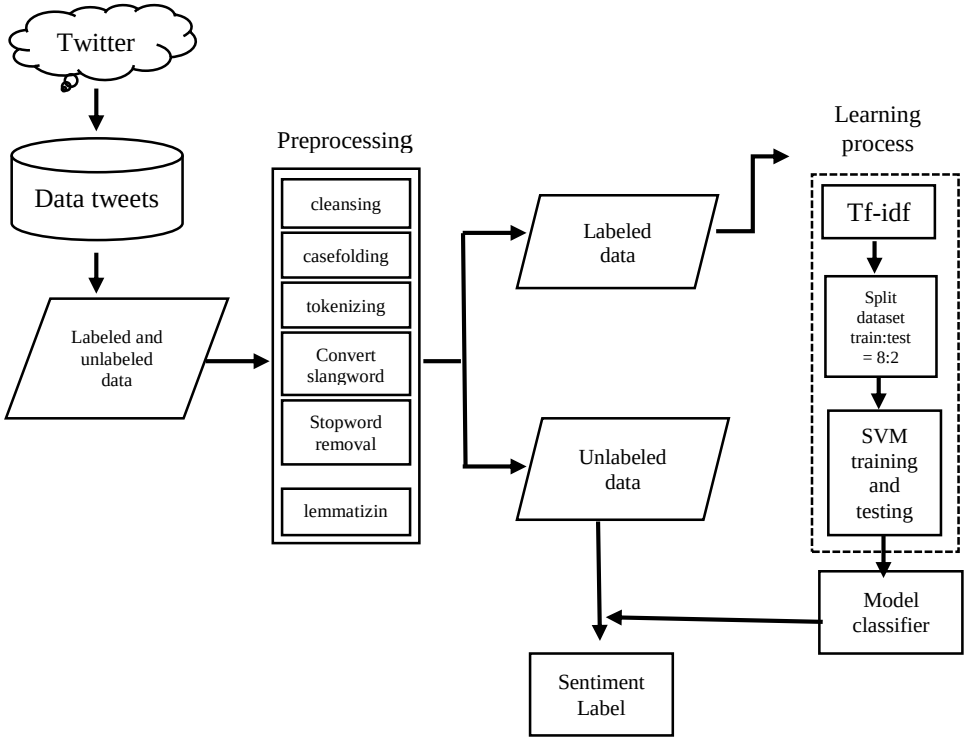

Figure 1. Research Flow

2.1 Data Collection

The data used in this study is Indonesian language twitter data search results with the keywords "*cadar*" and "*hijab"*. *Cadar* is the Indonesian language's term for veil. Data is taken by the crawling method using Twitter API. Twitter data is then stored in a MySQL database. The data used in this study were 3556 data, which is a combination of tweets with the keywords "veil" and "*hijab*".

2.2 Data Selection and Labeling

2.2.1 Data Selection

Not all of the tweets' data that have been collected can be used for further research processes. This is because some tweets consist of only one or two letters or words, do not speak Indonesian, and some tweets are repeated. Therefore, a selection of data tweets is performed to determine the data to be used in the next step.

This selection process is only carried out on data that will be used in the learning process. Data to be labeled automatically, chosen randomly. After going through the selection process, 1056 data will be taken to be used in the supervised learning process and 2500 data will be taken randomly without going through the selection process.

2.2.2 Data Labeling

Data labeling is not done on the whole data, but only on data that will be used at the training stage. The data to be labeled manually is as much as 1056 data tweets. Each tweets will be grouped into four classes, namely positive, negative, neutral, and irrelevant.

2.3 Preprocessing

At this stage, the tweet data will be cleaned up. The steps in cleaning the data include cleansing, case folding, tokenizing, convert slang word, stop word removal and stemming/lemmatizing. All these processes are carried out with the aim that the quality of the classification carried out for the better [13]. The preprocessing steps in this study are as follows.

2.3.1 Cleansing

Data tweets usually contain html tags, URLs, usernames, punctuation marks, or other non-letter characters. At this stage, the components are cleaned.

2.3.2 Case folding

At this stage, all letters in the tweets data are converted to lowercase letters. The existence of capital letters will affect the calculation of *tf-idf*. Therefore, it is done to generalize letters into lowercase letters.

2.3.3 Tokenizing

Tokenization is the process of dividing text, sentences or paragraphs into certain tokens or parts. Usually, this token separation refers to the presence of spaces or punctuation.

2.3.4 Convert Slang word

Sentences on Twitter usually contain non-standard words that are not following the standard Indonesian spelling. Therefore, at this stage, the conversion of non-standard words into the EYD standard.

2.3.5 Remove Stop word

Stop words are common words that often appear which are usually ignored because they are considered to have no meaning. Examples of words that are categorized as stop words are people pronouns and conjunctions, like me, you, and from, and others. This process is done by utilizing the Sastrawi stop word library.

2.3.6 Lemmatizing

This process is the process of changing a word into its basic word form.

2.4 Sentiment Analysis Process

2.4.1 Weighting words with Tf-Idf

Tf-Idf is the process of finding the value of data based on the frequency of occurrence in the document [14]. TF-IDF can be calculated using Formula 1.

$$\text{tf-idf}(d, t) = \text{tf}(t) * \text{idf}(d, t) \quad (1)$$

where the formula of inverse document frequency is as shown in Formula 2:

$$\text{Idf}(d,t) = log(\frac{N}{df(d,t)}) + 1 \quad (2)$$

2.4.2 Support Vector Machine Algorithm

SVM tries to find hyperplane by maximizing the distance between classes (margins). SVM will find a hyperplane that separates between classes. SVM only finds one hyperplane whose position is right in the middle between two classes[8]. That is, the distance between hyperplane and data objects of different classes adjacent (outermost), is identical as given in Figure 2.

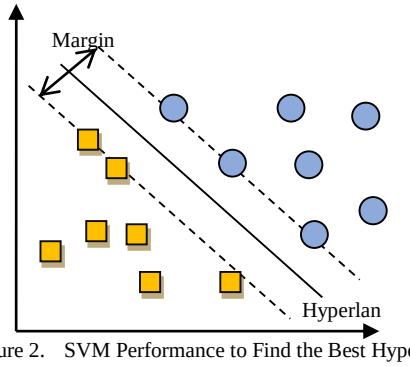

Figure 2. SVM Performance to Find the Best Hyperplane

This research will experiment using SVM with Radial Base Function (RBF) kernel with C=1; 10 and γ=0.2; 0.5; 1. The RBF kernel was used because of its numerical difficulties that are fewer than linear and polynomial kernel and also because this research doesn't have a large feature that is suitable for using RBF kernel[15].

3 Result and Discussion

The data used in this research are Twitter posts with the keywords "veil" and "hijab" in Indonesian Bahasa. Data is taken by crawling method using Twitter API and Tweepy library. From the resulting 3556 Twitter posts. Examples of tweets data are presented in Table 1.

Table 1 Example of Tweets Data

No	Tweets
1	Adem yak kalo liat perempuan pake cadar?
2	Ihh... mengerikan... Cadar kok dipakai untuk sprt itu... Kirain tambah kemayu.. eh jd teroris https://t.co/Khu749x3uk
3	@ZheeOmega Kalian boleh-boleh saja dan sah-sah saja ngomong pakai Bahasa Arab, memakai jubah/abaya atau bahkan cadar. Itu hak masing masing individu. https://t.co/Bd1LDPnjuh
4	Liat orang pake cadar pengen gua bawa ke kantor polisi rasanya. Jadi bener2 benci sebencinya itu liat org yg bunuh org gaberdosa.
5	BERITA BAIK WARGA PASIR GUDANG serta kawasan2 sekitar Cadar Queen serendah RM 19 Cadar Patchwork 2 beg RM 90 Pesta Cadar Patchwork akan berlangsung di TS HOTEL... https://t.co/VRLBuK9Bek

The data will be divided into two, namely training data and test data. Training data are data to which sentiment labels are done manually that will be used to do machine learning. From those tweets data, 1056 data are for training. Test data is data that will be given a sentiment label by machines that have been trained with test data. The test data were taken as much as 2500. Both the training data and test data were taken randomly. Training data as shown in Table 2 below:

Table 2 Training Data

No.	Tweets	Label
1	Adem yak kalo liat perempuan pake cadar?	1
2	Ihh... mengerikan... Cadar kok dipakai untuk sprt itu... Kirain tambah kemayu.. eh jd teroris https://t.co/Khu749x3uk	-1
3	@ZheeOmega Kalian boleh-boleh saja dan sah-sah saja ngomong pakai Bahasa Arab, memakai jubah/abaya atau bahkan cadar. Itu hak masing-masing individu. https://t.co/Bd1LDPnjuh	0

No.	Tweets	Label
4	Liat orang pake veil pengen gua bawa ke kantor polisi rasanya. Jadi bener2 benci sebencinya itu liat org yg bunuh org gaberdosa.	-1
5	BERITA BAIK WARGA PASIR GUDANG serta kawasan2 sekitar Cadar Queen serendah RM 19 Cadar Patchwork 2 beg RM 90	2
6	bukan hanya yang pake hijab yang merasa tenang, yang ngeliat juga tenang | kalo mau tenang dan menenangkan, ya berhijab syar'i aja :D	1
7	Gambar Tutorial Hijab Modern Paris Segi Empat Simple Terbaru 2017 https://t.co/WEoigVzdOC https://t.co/jBXJVah95m	2
8	sejak memilih untuk berhijab, hati lebih tenang,semoga pilihan ini diridhoi allah.. Aminn :) Terimakasih Hijab ^^ #HIJ48	1

In the preprocessing stage, the tweets data is cleared from html tags, URLs, usernames, punctuation marks or other non-letter characters. Table 3 shows the comparison of data before and after going through the preprocessing stage.

Table 3 Preprocessing Results

	Tweets
Before	Ihh... mengerikan... Cadar kok dipakai untuk sprt itu... Kirain tambah kemayu.. eh jd teroris https://t.co/Khu749x3uk
	Saya khawatir dengan adanya kasus-kasus ini membikin stigma buruk para pemakaian cadar kian melebar.
	sejak memilih untuk berhijab, hati lebih tenang,semoga pilihan ini diridhoi allah.. Aminn :) Terimakasih Hijab ^^ #HIJ48
After	Ihh mengerikan Cadar kok dipakai untuk sprt itu Kirain tambah kemayu eh jd teroris
	Saya khawatir dengan adanya kasus kasus ini membikin stigma buruk para pemakaian cadar kian melebar
	sejak memilih untuk berhijab hati lebih tenang semoga pilihan ini diridhoi allah Aminn Terimakasih Hijab HIJ

The next process is Case folding. At this stage, all letters in the tweet's data are converted to lowercase letters. The existence of capital letters will affect the calculation of TF-IDF.

Data tweets that have been converted into lowercase letters then proceed with the tokenization process. Tokenization is the process of dividing text, sentences, or paragraphs into certain tokens or parts. Tokenization refers to the presence of spaces or punctuation.

Sentences on Twitter usually contain non-standard words that are not in accordance with the standard Indonesian spelling. Therefore, at this stage, the conversion of non-standard words is carried out in the form of EYD (Ejaan Yang Disempurnakan)/ Indonesian Dictionary standard. The following Table 4 shows the change of words after going through the process of converting slang words.

Table 4 Comparison of the Results of the Convert Slang Word

	Tweets	Changed Words
Before	ihh mengerikan cadar kok dipakai untuk sprt itu kirain tambah kemayu eh jd teroris	kok, sprt, jd
	saya khawatir dengan adanya kasus kasus ini membikin stigma buruk para pemakaian cadar kian melebar	membikin
	sejak memilih untuk berhijab hati lebih tenang semoga pilihan ini diridhoi allah aminn terimakasih hijab hij	-
After	ihh mengerikan cadar mengapa dipakai untuk seperti itu kirain tambah kemayu eh jadi teroris	mengapa, seperti, jadi
	saya khawatir dengan adanya kasus kasus ini membuat stigma buruk para pemakaian cadar kian melebar	membuat
	sejak memilih untuk berhijab hati lebih tenang semoga pilihan ini diridhoi allah aminn terimakasih hijab hij	

It is followed by stop word removal. Stop words are common words that often appear which are usually ignored because they are considered to have no meaning. Example words categorized as stop words are pronouns people and conjunctions, like me, you, and from, and others.

The next process is lemmatizing. This process is the process of converting a word into its basic form using the Indonesian Spacy lemmatizing library

After the preprocessing is completed, the next process is the sentiment analysis process. In this process, the initial step is to weight the TF-IDF. TF-IDF is the process of finding the value of data based on the frequency of occurrence in the document. Each word in the table above then counts the frequency of occurrence. After calculating the occurrence frequency and IDF, the weight of each word TF-IDF is calculated in each document. TF-IDF value is obtained from the product of TF and IDF multiplication of each term. The results of the TF-IDF calculation are carried out the normalization process with Euclidian Normalization. The TF-IDF value for each term after undergoing the normalization process is presented in Table 5.

Table 5 TF-IDF Values After Normalization Process

No	Term	TF-IDF			
		D1	D2	D3	D4
0	ada	0	0,28833	0	0
1	adem	0	0	0	0,40726
2	baru	0	0	0,33333	0
3	buat	0	0,28833	0	0
4	buruk	0	0,28833	0	0
5	veil	0,10998	0,18404	0	0,25995

The next process is SVM process. SVM tries to find a hyperplane by maximizing the distance between classes (margins). SVM will find a hyperplane that separates between classes. Table 6 below shows the stages of how SVM found hyperplanes in sentences that had gone through a weighting process with TFIDF.

Table 6 Sentences in Vector Form

	Sentences Vector	label
D1	[0, 0, 0, 0, 0, 0.10998, 0.30998, 0.30998, 0, 0, 0, 0.30998, 0.30998, 0, 0, 0.30998, 0, 0, 0.30998, 0, 0, 0, 0.30998, 0, 0, 0, 0.30998, 0, 0, 0.30998, 0.30998, 0, 0, 0]	1
D2	[0.28833, 0, 0, 0.28833, 0.28833, 0.18404, 0, 0, 0, 0, 0, 0, 0, 0.57665, 0, 0.28833, 0.28833, 0, 0.28833, 0, 0, 0, 0.22732, 0, 0, 0, 0, 0.28833, 0, 0, 0, 0, 0]	1
D3	[0, 0, 0.33333, 0, 0, 0, 0, 0, 0.33333, 0.33333, 0.33333, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.33333, 0, 0, 0.33333, 0.33333, 0, 0.33333, 0, 0, 0, 0.33333, 0, 0]	1
D4	[0, 0.40726, 0, 0, 0, 0.25995, 0, 0, 0, 0, 0, 0, 0, 0.40726, 0, 0, 0, 0, 0, 0, 0.40726, 0, 0, 0.32109, 0, 0, 0, 0, 0, 0, 0, 0, 0.40726, 0.40726]	1

The tweet data used for the training set process in this study amounted to 1056. This data is a combination of search results data from Twitter with the keywords "*cadar*" and "*hijab*" that have been labeled sentiment. Each sentiment data has the same amount of data as presented in Table 7.

Table 7 Number of Tweets for Each Class

No	Sentiment	Number of Tweets
1	Positive	264
2	Negative	264
3	Neutral	264
4	Irrelevant	264
Total		1056

In the learning process, the data is divided into two, namely training data and testing for accuracy, with the division of classification proportions 8:2. This process will produce a classifier model that will be used to classify the tweets' data without labels at a later stage. The algorithm used is SVM with the RBF kernel function. The parameters needed in the RBF are the cost parameter (*C*) and the gamma parameter (γ). Table 8 shows the comparison of accuracy values with different combinations of *C* and **γ** values.

Table 8 Comparison of Accuracy Values

C	Γ	Accuracy
1	0.2	66%
1	0.5	71%
1	1	73.6%
10	0.2	68%
10	0.5	70%
10	1	72%

The results of the classification process using the Support Vector Machine method with a value of C = 1 and γ = 1 of 73.6%. The details of the results of this classification can be seen in Table 9.

Table 9 Classification Report

Sentiment	Precision	Recall	F1-score	Support
Negative	0.62	0.76	0.68	54
Neutral	0.58	0.60	0.59	53
Positive	0.83	0.64	0.72	53
Irrelevant	0.98	0.94	0.96	52
Avg / total	0.75	0.74	0.74	212

These results indicate that there are still various opinions (positive, negative, or neutral opinion) regarding hijab and veil as described in previous research. The previous research from[4] described that from a historical perspective there are some negative opinions about hijab and veil as well as positive opinions. And from the other research [3], [6] some positive and neutral opinions just treat hijab and veil as the law from Qur'an and Hadith.

The training process that has been carried out on labeled data produces a classifier model that can be used to predict sentiment classes in tweets data that have not yet had sentiment labels. There are 2500 data tweets with unknown types of sentiment. These data have gone through the same preprocessing stage, as did the training data. At this stage, classification of sentiment types will be carried out on 2500 data. The results of sentiment class predictions on 2500 data tweets are presented in Figure 3.

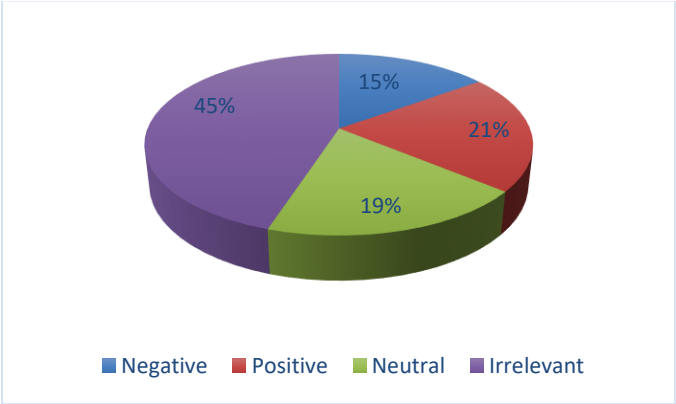

Figure 3. Percentage of the Number of Tweets in each Sentiment Class

From Figure 3, it can be concluded that irrelevant classes dominate the tendency of tweets sentiment with the keywords "veil" and "hijab". The tweets that are categorized as irrelevant class in this study include tweets about product sales, news titles that do not contain opinions, and tweets that only consist of one word. Tweets with irrelevant labels dominate by 45%, then tweets with positive opinions by 21%, neutral 19%, and tweets with negative opinions have the fewest number, namely only 15%.

4 CONCLUSION

Based on the research that has been done, our finding is Support Vector Machine method combined with RBF kernel, a value of C = 1 and 1 = 1 in 1056 data tweets with a composition of 80% training data and 20% testing data, resulting in an accuracy of 73.6%, 75% precision, 74% recall and 74% f1. And result with precision to negative opinions are 62%, positive opinions are 83%, neutral opinions reach 53% and irrelevant opinions that talk about hijab and veil reach 98%. It shows that sentiment analysis can be used for predicting the negative, positive or other sentiments of a sentence based on a certain topic, in this case veil and hijab.

The results of labeling automatically on 2500 data tweets with the classifier model created in the training process resulted in percentage value of tweets. They are irrelevant category with 45%, then positive tweets of 21%, neutral 19%, and tweets negative has the fewest amount which is only 15%.

REFERENCES

[1] H. Maulana, "Media Sosial sebagai Sumber keberagamaan Alternatif Remaja dalam Fenomena Cyberreligion," *Dakwah J. Kaji. Dakwah dan Kemasyarakatan*, vol. 22, no. 1, pp. 1–34, 2018, doi: http://doi.org/10.15408/dakwah.v22i1.12044.

[2] B. A. Sevsa and M. D. R Wahyudi, "Analisis Sentimen pada Indeks Kinerja Dosen Fakultas Saintek UIN Sunan Kalijaga Menggunakan Naive Bayes Classifier," *J. Buana Inform.*, vol. 10, no. 2, p. 112, 2019, doi: 10.24002/jbi.v10i2.2250.

[3] P. A. R. Dewi, "Niqab sebagai Fashion: Dialektik Konservatisme dan Budaya Populer," *Scriptura*, vol. 9, no. 1, pp. 9–15, Aug. 2019, doi: 10.9744/scriptura.9.1.9-15.

[4] L. Rasyid and R. Bukido, "Problemtika Hukum Cadar Dalam Islam: sebuah Tinjauan Normatif-Historis," *J. Ilm. Al-Syir'ah*, vol. 16, no. 1, pp. 74–92, 2018, Accessed: Aug. 12, 2020. [Online]. Available: https://www.researchgate.net/publication/326805010_problemtika_hukum_cadar_dalamislam_sebuah_tinjauan_normatif-historis.

[5] W. Nuroniyah, "Dekonstruksi Hijab," *Al-Manahij*, vol. 11, no. 2, pp. 263–280, Dec. 2017, doi: 10.24090/mnh.v11i2.2017.pp263-280.

[6] M. Mujahidin, "Cadar: Antara Ajaran Agama dan Budaya," *JUSPI (Jurnal Sej. Perad. Islam.*, vol. 3, no. 1, p. 11, Jul. 2019, doi: 10.30829/juspi.v3i1.3142.

[7] B. Liu, *Sentiment Analysis(Introduction and Survey) and Opinion Mining*. 2012.

[8] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. ANDI Publisher, 2012.

[9] S. R. Kao, Anne, Poteet, *Natural Language Processing and Text Mining*. New York: Springer-Verlag, Inc., 2007.

[10] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J. Maret*, vol. 1, no. 1, pp. 32–41, 2017.

[11] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.

[12] M. D. R. Wahyudi, "Analisis sentimen ujaran kebencian pemilihan presiden 2019 menggunakan algoritma Naïve Bayes," *Jnanaloka (Jurnal Open Access Yayasan Lentera Dua Indones.*, no. 2020, pp. 5–10, 2020.

[13] A. F. Hidayatullah, "Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016.

[14] M. D. R. Wahyudi, "Evaluation of TF-IDF Algorithm Weighting Scheme in The Qur ' an Translation Clustering with K-Means Algorithm," 2020.

[15] V. Apostolidis-afentoulis, "SVM Classification with Linear and RBF kernels Konstantina-Ina Lioufi," *ResearchGate*, no. July, pp. 0–7, 2015, doi: 10.13140/RG.2.1.3351.4083.