# Sentiment Analysis for Decision Support Systems of Employee

**Saeful Bahri[1], Miftah Farid Adiwisastra[2], Tuti Alawiyah[2*], Dini Silvi Purnia[1], Bambang Kelana Simpony[2]**

[1]Department of Information System, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta, Indonesia
[2]Department of Information System Kampus Kota Tasikmalaya, Universitas Bina Sarana Informatika, Tasikmalaya, Jawa Barat, Indonesia

*tuti.tah@bsi.ac.id

**Abstract**. This paper presents sentiment analysis that will be used as Decision Support in employee recruitment. Sentiment analysis used Term Frekuensi.Index Document Frekuensi (TF.IDF) weight calculations. Weighting results were classified using the Support Vector Machine (SVM) method into several categories, namely negative sentiment, positive sentiment and neutral. the results of this study showed an accuracy value of 0.65 which was the best accuracy for text classification

## 1. Introduction

Employee recruitment activities become important in a company to support its business activities and productivity [1]. But these companies often get employees who do not fit the specified criteria because the company does not have fixed standards. Also, the objective assessment does not cause injustice in the process of employee recruitment which causes the quality of the employee received is not sometimes as expected.

This study intended to build a system that measured the sentiment of prospective employees that was useful as one component in supporting the decision to recruit employees objectively to get quality employees by company needs.

Sentiment analysis in this study used TF.IDF (Term Frekuensi. Index Document Frekuensi) weighting to find the weight value of each word. TF.IDF was one of the most popular weighting methods used in giving weight to a text [2]. After the weighting process, then the text or words would be classified using the SVM (Support Vector Machine) algorithm. SVM was the best classification algorithm in reducing errors in the classification results[3].

The objectivity in job interviews was very important. Therefore we needed a system to analyze sentiment in job interviews. Based on the model that had been formed then it would be implemented into an application.

## 2. Methods

This study used data taken from the writing or essay of prospective new employees for further analysis using sentiment analysis. Sentiment Analysis was a way to extract one's feelings, opinions and habits from an article [4]. Sentiment analysis could be used as an alternative psychological assessment of someone from writing activities [5].

Sentiment analysis had been used in various fields including sports which were used to evaluate the match based on the sentiment of the fans [5], the political field was used to analyze the level of candidate selection in an election based on writing or tweet on social media tweeter [6] [7] [8] [9], and in the social field it was used to see the response of the community to the surrounding social conditions [4] [10] [11]. Sentiment analysis could also be applied as a method to support a decision [12] [13].

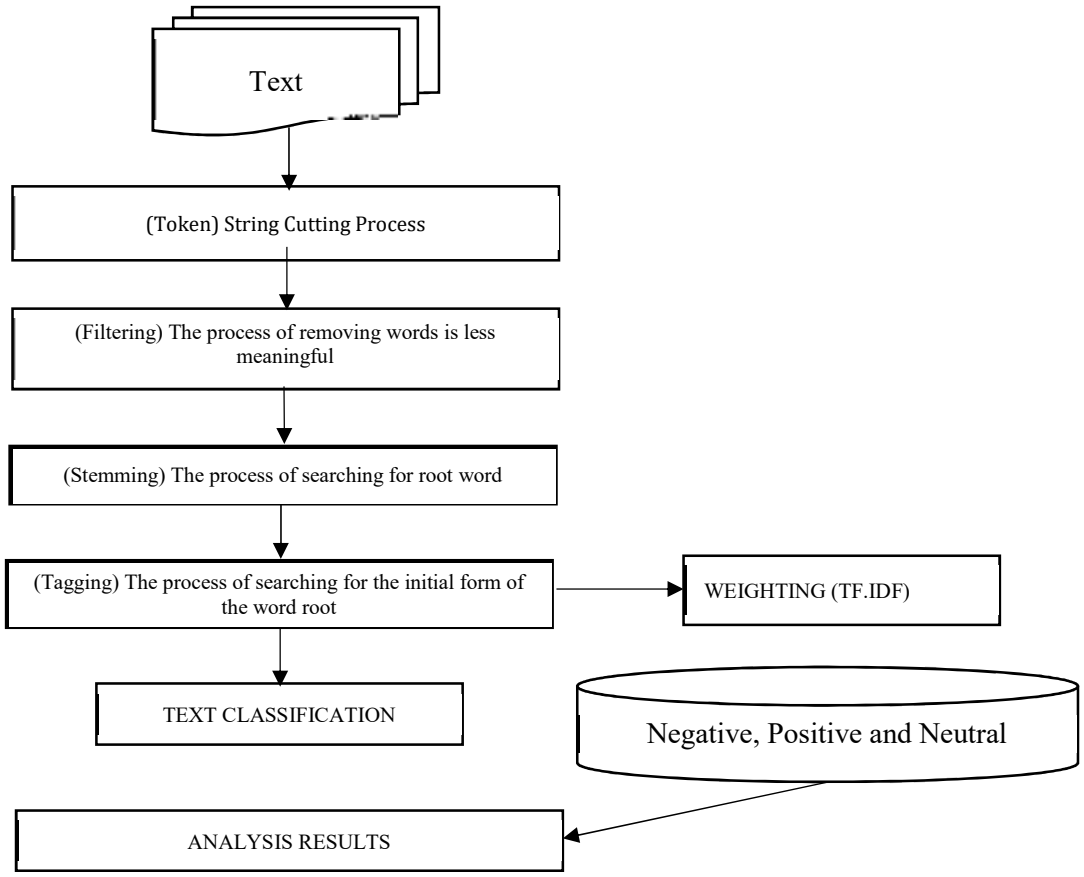


**Figure 1.** Research Stages

## 3. Result and Discussion

In the preprocessing stage, the document or corpus text that has been collected will go through a token process that is cleaning by cutting the string into word fragments. Furthermore, the filtering process is to get rid of less meaningful words then the process of stemming is done which is the process of finding the basic root of each word.

After the preprocessing stage is completed, in the next stage the weight search uses TF.IDF then a text classification is performed. In weighing the results of the essay, there are several documents (d). The document in question is the content of several prospective employees. As an example, we present the problem as follows:

(d1) I am a great person.
(d2) I'm also a caring person
(d3) someone who wants to study
(d4) someone who works hard

**Table 1.** The results of the stemming process

| term (t) | d1 | d2 | d3 | d4 |
|---|---|---|---|---|
| I | 1 | 1 | 0 | 0 |
| Am | 1 | 1 | 0 | 0 |
| Person | 1 | 1 | 0 | 0 |
| Someone | 0 | 0 | 1 | 1 |
| Study | 0 | 0 | 1 | 0 |
| Great | 1 | 0 | 0 | 0 |
| Works | 0 | 0 | 0 | 1 |
| Hard | 0 | 0 | 0 | 1 |

**Table 2.** The results of document frequency (*df*)

| term (*t*) | *df* |
|---|---|
| I | 2 |
| am | 2 |
| person | 2 |
| someone | 2 |
| study | 1 |
| great | 1 |
| works | 1 |
| hard | 1 |

Then calculate inverse document frequency (*idf*) using equations

$$idf = \frac{1}{df} \quad or \quad idf = log\left(\frac{N}{df}\right),\ N = count\ of\ d$$

**Table 3.** The results of inverse document frequency (*idf*)

| term (*t*) | *Df* | *idf* | term (*t*) | *df* | *idf* |
|---|---|---|---|---|---|
| I | 2 | 1/2=1 | I | 2 | Log(4/2)=0.301 |
| am | 2 | 1/2=1 | am | 2 | Log(4/2)=0.301 |
| person | 2 | 1/2=1 | person | 2 | Log(4/2)=0.301 |
| someone | 2 | 1/2=1 | someone | 2 | Log(4/2)=0.301 |
| study | 1 | 1/1=1 | study | 1 | Log(4/1)=0.602 |
| great | 1 | 1/1=1 | great | 1 | Log(4/1)=0.602 |
| works | 1 | 1/1=1 | works | 1 | Log(4/1)=0.602 |
| hard | 1 | 1/1=1 | hard | 1 | Log(4/1)=0.602 |

**Table 4.** The results of Term Frekuensi.Index Document Frekuensi (*TF.IDF*)

| term (*t*) | *TF.IDF* | | | |
|---|---|---|---|---|
| | *d1* | *d2* | *d3* | *d4* |
| I | 0,602 | 0,602 | 0 | 0 |
| am | 0,602 | 0 | 0 | 0 |
| person | 0,602 | 0,602 | 0 | 0 |
| someone | 0 | 0 | 0.602 | 0.602 |
| study | 0 | 0 | 0,602 | 0 |
| great | 0,602 | 0 | 0 | 0 |
| works | 0 | 0 | 0 | 0,602 |
| hard | 0 | 0 | 0 | 0,602 |

### *3.1. Model Testing*

Model testing is performed to find the accuracy value of the SVM classification algorithm. The value obtained will be a reference for researchers in developing the next model, the following is the result of testing the model.

The accuracy value obtained from 4 times of testing using cross-validation method where the amount of data distribution is 2000 divided into 4 fold. From the 4 fold all act as training and testing. Validation results could be seen in Figure 2. Based on the graph in Figure 2, the average accuracy value is 0.65%.

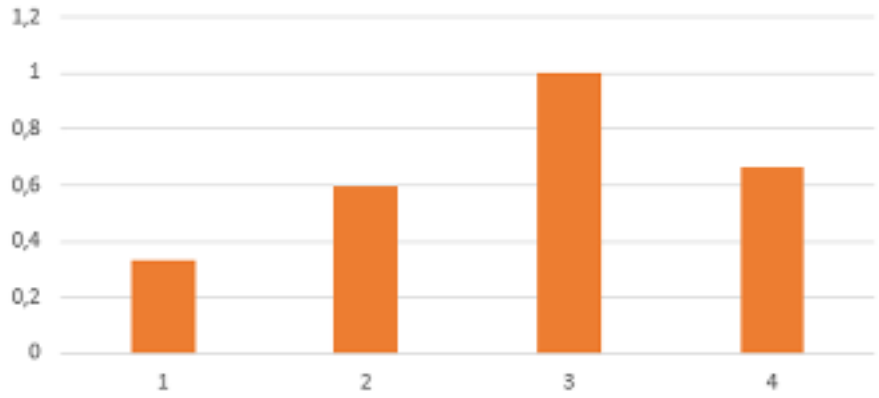


**Figure 2.** Accuracy Graph

### *3.2. Implementation*

After testing the built model, the model was then applied to an application program. Pageviews of job interview sentiment analysis could be seen in Figure 3.

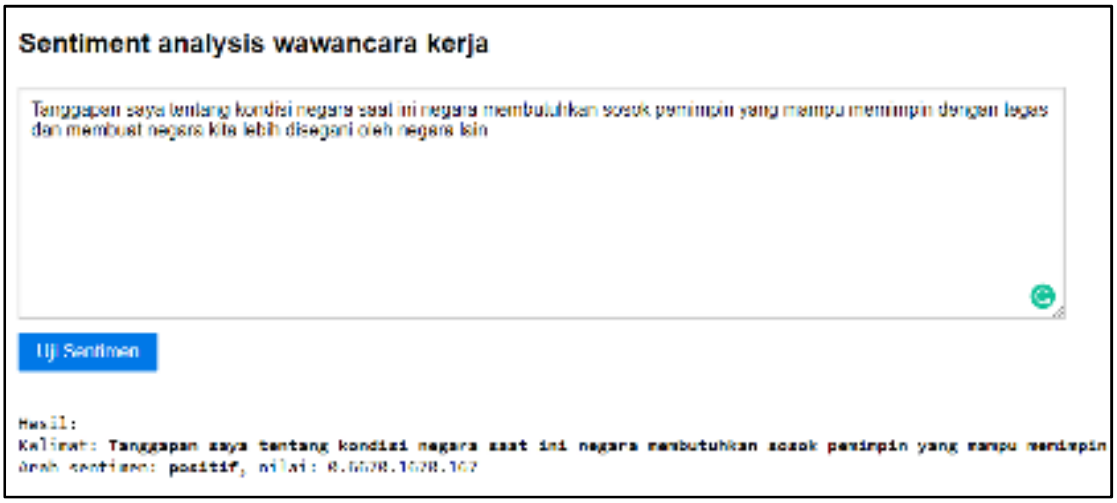


**Figure 3.** Display of Job Interview Sentiment Analysis Page

## 4. Conclusions

The results of this study indicate that the classification of sentiment analysis with SVM, from the results of 4 times the test has quite significant results with the accuracy value obtained is an average of 0.65

From the results obtained, the effectiveness and objectivity in the acceptance of employees can be achieved so that a company will easily determine employees by the expected criteria.

## References

[1] O. Fajarianto, M. Iqbal, and J. T. Cahya, "Sistem Penunjang Keputusan Seleksi Penerimaan Karyawan Dengan Metode Weighted Product," J. Sisfotek Glob., vol. 7, no. 1, pp. 49–55, 2017.

[2] A. Fathan Hidayatullah, A. Sn, J. Teknik, I. Fakultas, and T. Industri, "ISSN: 1979-2328 UPN "Veteran," Semin. Nas. Inform., vol. 2014, no. semnasIF, pp. 115–122, 2014.

[3] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," World Wide Web, vol. 20, no. 2, pp. 135–154, 2017.

[4] N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis," Telemat. Informatics, vol. 35, no. 1, pp. 136–147, 2018.

[5] Y. Yu and X. Wang, "World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets," Comput. Human Behav., vol. 48, pp. 392–400, 2015.

[6] D. F. Budiono, A. S. Nugroho, and A. Doewes, "Twitter Sentiment Analysis of DKI Jakarta ' s Gubernatorial Election 2017 with Predictive and Descriptive Approaches," in 2017 International Conference on Computer, Control, Informatics and its Applications, 2017, pp. 89–94.

[7] M. Ramzan, "Are Tweets The Real Estimators Of Election Results ?," 2017, no. August, pp. 10–12.

[8] B. Joyce and J. Deng, "Sentiment analysis of tweets for the 2016 US presidential election," in 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), 2017, pp. 1–4.

[9] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation," in Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015, 2016, pp. 1348–1353.

[10] R. Archana and S. Chitrakala, "Explicit sarcasm handling in emotion level computation of tweets - A big data approach," in Proceedings of the 2017 2nd International Conference on Computing and Communications Technologies, ICCCT 2017, 2017, pp. 106–110.

[11] S. Roh, "Big Data Analysis of Public Acceptance of Nuclear Power in Korea," Nucl. Eng. Technol., vol. 49, no. 4, pp. 850–854, 2017.

[12] C. Lee and I. Paik, "Stock Market Analysis from Twitter and News Based on Streaming Big Data Infrastructure," no. iCAST, pp. 312–317, 2017.

[13] C. Study, B. Bri, and M. Indonesia, "Association Perception Customer Feedback with Text Network Analysis in Social Media," vol. 0, no. c, 2017.

[14] J. Philbin, "Near Duplicate Image Detection : min-Hash and tf-idf Weighting," 2008.