

Abstract– Tens of thousands of people are applying for job to large companies every year. The goal of the recruitment process is to get new employees who can fit to the company's working culture. Due to the high number of applicants, the recruitment process takes a lot of time and requires high cost. This paper implements a popular combination of term frequency-inverse document frequency (TF-IDF) as the extraction method and support vector machine (SVM) as the classifier to filter the applicants' interview verbatim. SVM generally produces better accuracy in text classification compared to random forest or k-nearest neighbors (KNN) algorithm. However, TF-IDF has several developments to improve its flaws, one of them is term frequency-relevance frequency (TF-RF). As a comparison, in this study we use three extraction methods: TF only (without IDF), TF-IDF, and TF-RF. We use the verbatim from interview simulation at PT. Telkom as the data source. The results of SVM combined with TF-IDF achieves 86.31% of accuracy, whereas TF only achieves 85.06%, and with TF-RF gives 83.61% of accuracy. These results show TF-IDF still outperforms TF-RF in term of accuracy for this case.

Keywords: recruitment process, feature extraction, support vector machine, term frequency-inverse document frequency, term frequency-relevance frequency