

I. INTRODUCTION

The recruitment process is important for a company because the quality of employees influences the overall company's performance. Therefore, companies are very selective in finding promising applicants [1]. However, with a large number of applicants on every recruitment process, companies have difficulty in getting employees who meet their criteria. Moreover, the recruitment process also requires enormous resources from the process, costs, as well as human resources [2]. The applicant's personality can be an important factor for companies to determine whether they can work well or not. According to N. R. Ngatirin, Z. Zainol, and T. L. C. Yoong [3], personality represents a combination of features and qualities that build the characteristics of an individual. Personality traits can be used to understand human behavior regarding many things, including how they work with their environment. Psychological testing is one of the efforts that can be done to determine the applicants' personality, but requires a lot of time and cost in its implementation. Therefore, we need a system that can classify the personality of the applicants as well as reduces the time and cost required. Personality classification is done by classifying applicant text using one of the text mining methods. The data used for personality classification is interview verbatim which has been converted from interview recordings into text.

Multinomial Naive Bayes and Decision Tree are the classification methods used in this study, Multinomial Naive Bayes is used because each feature stands individually so that its implementation can produce a good performance, multinomial naive Bayes also can predict with less time so it will reduce time spent to classify the personality and acceptance of applicants to the company because the text mining will have more than a hundred features to naive calculate with the label. Therefore, multinomial Bayes will be suitable to be implemented in this research. Decision Tree is used because it is easy to implement and the classification process can be easily understood and the learning speed is quite fast. Decision tree also suitable for this research that gives output categorical, more than that we can calculate or see what the most relevant word that can associate with personality traits by visualizing the tree itself.

Research conducted by Rintaspon Bhannarai and Chartchai Doungsa-ard states that personality tests can be performed to predict agile people who are suitable for an agile methodology based on the theory of big five personality traits. With k-nearest neighbor as the classification technique used, the study gave an accuracy of 65.71% with its best k as 1 [4].

Research conducted by Harshal Chaudhari, Nalini Yadhav, and Yash Shukla predicts what types of jobs are suitable for each individual based on their resumes, the classification is done using the Naive Bayes Classifier by calculating the likelihood of candidates getting the jobs based on the tokens calculated from the resumes. Tokens are calculated based on academic grades, hobbies, professional experiences, projects, publications, awards, etc. The expected result from this study is that the created model can help companies to select candidates closely and employ competent applicants in the right position [5].

To be able to classify personalities using text data, personality theory that can be used as a basis for the personality classification is needed. For example, the personality theory of the Big Five Personality Traits, is a personality theory that categorizes people into five different characteristics [6]. The five characteristics are Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. But in this study, the personality traits that support the applicant's personality classification process are Conscientiousness and Agreeableness. The dataset used is the text data that needs to be done through the cleansing process first, so that the data can be used for the applicant's personality classification process. The final result of the classification of applicants is whether the applicant is suitable or not for the company based on the personality theory. Several limitations arise in this study, that is the difficulty of collecting the required dataset because the dataset needed is the data result of interviews between applicants and companies, the difficulty lies in the code of ethics of psychology that forbids anyone other than the parties involved to find out the contents of the interview so there are many procedures needed to be followed to collect the interview data from the company, whereas making the interview data manually requires a significant amount of money to pay volunteers to be interviewed and psychologists to identify the personality of the volunteer, the result is that the data collected is limited.

The objective of this research is to be able to classify applicants' personalities based on existing text data. The steps taken are processing the text data that will be used to classify the personality, the features used in the classification process are a collection of words that have been filtered beforehand, and calculating the accuracy of the classification results from the two methods used to find out which method is better to use for the classification and whether the system built is good enough to be implemented or not.