1. INTRODUCTION

Statistics Indonesia or locally known as Badan Pusat Statistik (BPS) stated that in 2019 there were 5 out of every 100 workers in Indonesia who were still unemployed [1]. This figure is relatively high. The impact can be seen in tens of thousands of people applying for employment in PT. Telkom Indonesia every year as an example. The large number of applicants makes the recruiting process more complicated. The higher time that increases costs should be reduced by using the text classification method. However, not much recent research has focused on the classification of job interview texts.

This study focuses on the classification of the PT interview texts. The Telkom recruitment process, which consists of nine parameters of categories of work culture: Solid, Speed, Intelligent, Imagine, Focus, Action, Integrity, Enthusiasm, Totality. Each label has two classes that represent a greater number of classes, a greater probability of being chosen. The goal is to rank each given interview text entry, whether it corresponds to a lower or higher class.

Most text classification systems are integrated into four sentences: feature extraction, dimension reduction, classifier selection, and evaluation [2]. In this study, the dimension reduction method is not used because it is not really necessary due to the limited data sources available. There are several approaches to text classification, which are vector space models, probabilistic models, and inference network models [3]. The most common techniques for extraction characteristics are the Term Frequency-Inverse Document Frequency (TF-IDF) and the standard Term Frequency (TF) [2]. In some cases, the TF-IDF has several drawbacks that reduce its ability to determine the value of different conditions normally. Therefore, the research [4] introduced a developed frequency-based model called the Term Frequency-Relevance Frequency.

As there are not many studies that focus on classification interview texts, we tried to propose a basic method for further comparison in the classification of interview texts specifically. The survey studied in the study [2] states that the Support Vector Machine (SVM) generally offers better results in text classification compared to other traditional classification algorithms. Therefore, this study uses SVM as the classification method. For best results, we compared the three characteristics of the feature extraction process: TF, TF-IDF, and TF-RF to determine which combination is best suited to classify text interviews with limited data.

The purpose of this study is to build a system to implement job classifier text classifications into two classes using TF-IDF and SVM and to analyze the classification performance results obtained from applying TF-IDF compared to TF methods. RF and TF. In the end result investigation process, various process scenarios are applied to obtain the best end results.