BAB I

Diabetes has been considered as one of the leading health problems globally these days. Diabetes is a disease where someone has high levels of blood glucose, and the pancreas gland could not produce enough insulin [1][3]. World Health Organization (WHO) stated, in 2017, there are 425 million people who were suffering from diabetes, 75% of them are ranged from 20 to 64 years old [2]. In 2017, Indonesia was ranked sixth out of ten countries with the highest number of diabetic patients, which is 10.3 million patients and is expected to increase to 16.7 million patients by 2045 [3]. Back in 2016, the WHO reported that diabetes was ranked seventh as the leading cause of death throughout the world [3]. In a few years, diabetes will be more dangerous than AIDS.

Nowadays, multiple machine learning techniques and appliances are applied in various field, including the medical field. In the medical field, machine learning is applied to predict various diseases, and one of them is diabetes. However, the medical dataset is typically high dimensional, incomplete, and noisy; therefore, data pre-processing such as the dimensionality reductions method and fill in the missing value are well suited to tackle this problem.

The accuracy is critical. That is why machine learning is used in this field. Several approaches have been used in this field. One of them is KNN. It makes an appropriate search algorithm since it's computationally tractable even for large datasets, such as medical datasets.

Since the dataset contains many predictive features, both PCA and AE are investigated to tackle such an issue. In [4], it has been proved that by combining KNN and KMC could provide higher accuracy. However, the KNN algorithm takes the majority of the vote to classify a point, making it a problem if there is an imbalance in a dataset [15]. Therefore, in this research, WKNN is applied as the classification method