CHAPTER 1
INTRODUCTION

1.1. Background

Caffeine is popular worldwide and has many health benefits due to the presence and number of compounds. The compound is a xanthine alkaloid in the form of crystals and tastes bitter which works as a psychoactive stimulant and mild diuretic. Caffeine was discovered by a German chemist, Friedrich Ferdinand Runge in 1819. Caffeine is found naturally in foods such as coffee beans, tea leaves, colas, guarana, and maté.In plants, it acts as a natural pesticide that paralyzes and kills certain insects that eat these plants. It is commonly consumed by humans by extracting it from coffee beans and tea leaves. Caffeine is a central nervous system stimulant drug in humans and can temporarily ward off drowsiness.Drinks that contain caffeine, such as coffee, tea, and soft drinks, are very popular. Caffeine is the most widely consumed psychoactive substance in the world. Unlike other psychoactive substances, caffeine is legal and unregulated in most jurisdictions around the world. In North America, 90% of adults consume caffeine daily.

The use of data mining techniques has made it possible in the process of collecting, storing, and processing large and complex amounts of data. In classifying caffeine in beverages, the use of data mining techniques was carried out by (Hemanenda 2013) using the LVQ learning vector quantization algorithm. The data used is caffeine data, totaling 619 types of various types of drinks that contain caffeine. At the trial, an analysis of the results of the accuracy of the classification of caffeine content is carried out based on the correct amount of data compared to the test data. From the analysis results obtained the minimum support value with the best accuracy results.

The research that has been carried out by the two researchers uses data of the same type, namely data on caffeine levels. This research focuses on measuring the performance of the KNN and SVM method algorithms. The test parameters used 4 attributes in the drink type data. The trials in this research used several scenarios, such as the number of test documents and training documents, and other determinants to obtain varying results between the two methods used in this research. The hope of this research is that it can be useful as a reference for similar research in order to find out the best method to use in classification.

1.2. Problem Formulation

From the background above, the formulation of the problem in this study is, what is the accuracy of the results of caffeine classification for types of drinks using the K-Nearest Neighbor method compared to the Support Vector Machine?

1.3. Scope

The limitation of the problem in classifying this project only discusses the caffeine content, not the other ingredients. Then the final results that have been classified only show the true or false results of the contents in the caffeine content. And the dataset used is also limited to only being downloaded from the kaggle site.

1.4. Objective

to compare the classification algorithm techniques in order to assess Orange's effectiveness in categorizing the caffeine level of various types of beverages. Understanding the accuracy, precision, recall, and f-score of this algorithm.