

CHAPTER 1

INTRODUCTION

1.1 Background

In this globalization era, internet technology has grown rapidly and many things can be done using this technology. People can get many informations and facilities so fast is a prove that this technology has become a daily activity. With easy access, many people used internet technology as a business support. Hence this day, many bookstores offer online book sales to make it easier for people to get more information. But to be able to get real information that really needed, then users should be able to sort any websites which can be accessed and provide the right information. However many bookstore websites showed uncategorized books so the search was ineffective because it takes a long time.

To solve that problem, those books need to be classified into some categories so users can easily find the right book that needed. Book categories can be determined based on its description because some books at the same categories have similar description each of it.

This final project will use a machine learning algorithm which is Learning Vector Quantization (LVQ). LVQ algorithm will learn characteristic of categorized book descriptions on training datas and classify the test data as from LVQ training data before. And finally the book will be classified according to category that matches with the training datas using the learning logic from the algorithm, (case studies : <https://www.bukukita.com/>). Before the implementation of the LVQ algorithm in the book classification, it requires the help of other methods to process data, which is Text Preprocessing and Term Frequency-Inverse Document Frequency (TF-IDF)

1.2 Problem Formulation

Some questions that want to be proved in this project :

1. How to implement Learning Vector Quantization (LVQ) algorithm to classify the data into the right category?
2. How to use Text Preprocessing and Term Frequency-Inverse Document Frequency (TF-IDF) to help LVQ algorithm to preprocess data features?

1.3 Scope

In this project there are few limitations on certain things :

1. This project only has 3 variable categories and the total of the training data is 60 data.
2. This project only uses one algorithm which is Learning Vector Quantization (LVQ) and also the help from Text Preprocessing and TF-IDF to preprocess data features.

And also there are hypotheses that are made in this project :

1. A book can be classified in the wrong category if the description pattern of the book does not match the description patterns of other books.
2. The differences of learning rate / alpha could affect the classification result.

1.4 Objective

The purpose of this project is to prove that Learning Vector Quantization (LVQ) algorithm can help users classify books automatically based on book descriptions into suitable categories. The purpose of this project is to make users find the right book to buy a lot easier.