

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

The number of final project documents at Soegijapranata Catholic University's Faculty of Computer Science increasing every year. This final project consists of various fields of science that have been studied previously. One of the main problems that happens is the task of organizing text data. Categorization of this text is done based on certain similarities. The main approach is the categorization of the text, the task of trying to automatically assign the document into each category. Final project are often grouped by technical domain and sub-domain.

The problem that often occurs is the difficulty of classifying final report based on existing scientific categories. Actually the determination of this document can be classified manually, but with the amount of data that very much will take a long time, so this becomes very inefficient, and can also cause errors (human error). For that classification is an appropriate method to identify the final project report. The best Text Categorization system uses a machine learning approach. It starts from the classifier, learns the rules from the example, and evaluates them on a set of test documents.

The classification used in this project uses the Support Vector Machine (SVM) algorithm. SVM algorithm is one of machine learning algorithm. SVM will study the characters of each category in the training data and test the data from what SVM has learned before. With this machine learning, the final report can be classified into 4 predefined categories as default (case studies: Faculty of Computer Science Soegijapranata Catholic University).

## 1.2 Scope

Some questions that want to be proved in this project

1. How to processing data using TF-IDF as term weighting?
2. How to implement Support Vector Machine (SVM) to classify the data into category?

## 1.3 Objective

The objective of this project is to make a program that can automatically classifying title of final report into a predefined category using the Support Vector Machine (SVM) method. The benefit of this project is make classifying job faster and easier.

