CHAPTER I

INTRODUCTION

1.1 Background

Customer churn happens when customers are not using a company's product or service anymore, but instead they use competitor's product or service [1]. Of course, companies do not wish for their customers to churn, but rather retaining the loyal ones who have positive attitude towards the company, are committed to repurchase, and willing to make recommendations. Customer retention is more beneficial than acquiring new customers. In fact, if a company can retain just 5% of its customer, its profit will increase and it can reduce marketing and operational costs while at the same time increasing its sales [2]. A small improvement in customer retention can help a company save a lot of money [3].

Customer retention is important for most companies from various sectors, including the education sector. According to some studies, parents are the key decision-makers in deciding which school for their children [4]–[6]. The factors that influence parents' decision-making come from the school itself, like academic quality, safety and discipline, school's strategic location, the distance from home to school, curriculum, school's environment, students' performance, and of course teacher's quality. If a school provides excellent service in each factor, it can cause satisfaction for the parents and make them want to enroll their children to that school. This also applies to parents who want to re-enroll or retain their children at the same school for the next academic year after completing a level. For example, after completing kindergarten, parents will enroll their children to primary school.

In the education sector, it is called student retention, which is equivalent to customer retention in other business sectors.

Student retention prediction is necessary for schools because it can identify students who are most likely to retain or churn [7]–[10] and help school management to evaluate and improve its service, learning system, facility, and its business strategy and general [7], [9], [11]. This prediction also can show the causes or factors behind retain behavior [8]. Other studies use factors like academic performance, attendance, family, financial condition, behavior, parent involvement, and demographics (age, gender) as variables to predict student retention.

This research raises a case study of a private school that has a target of student retention. Usually at the beginning of an academic year, the school's management makes a list of student retention. In this list, the management will mark the ones who will most likely continue their education in the next academic year and who will quit. The list is made solely from daily observation on parents done by the customer service team only, regardless of other factors that may influence student retention. Therefore, it is challenging for the school to come up with the right customer relationship management strategy to overcome student retention issues.

The research is intended to use machine learning models as a new approach to predict student retention, and to determine which feature is the best predictor to student retention. In the future, this school can get a more precise result of prediction through machine learning, anticipate the parents and students who are showing signs of not continuing their education in the next academic year, and make a better student retention strategy. The experiment will use machine learning models such as Random Forest (RF), Logistic Regression (LR), Support Vector

Machine (SVM), and Neural Network (NN), then show the results of the four models, and to identify which feature is the best predictor to student retention.

1.2 Problem Identification

In this research, there are several identified problems to be answered:

1) The school has never achieved its maximum student retention target in kindergarten, primary, middle, and high school level.
2) This private school has never used any analytical approach such as using machine learning model to predict student retention.
3) It is not known what feature is best to predict student retention.

1.3 Scope of Work

The scope of work for this research are:

1) Collecting data from this school, including demographic data, enrollment data, parents' satisfaction survey data, and observation data of parents' behavior.
2) Perform feature engineering to the dataset.
3) Fitting the data into machine learning models Random Forest, Logistic Regression, Support Vector Machine, and Neural Network to get best prediction results from each model and determine which feature is the best predictor to student retention.

1.4 Research Questions

This research will try to answer these questions:

1) Among RF, LR, SVM, and NN, which machine learning model gives the highest score in predicting student retention?
2) Which feature is the best predictor based on its ability to predict student retention?

1.5 Research Objectives

The objectives of this research are:

1) To determine which machine learning model gives the best result in predicting student retention.
2) To determine which feature is the best predictor to student retention.

1.6 Structure Overview

This report is divided into five chapters, and each chapter has its own aim and content. The structure of this report is as follows:

Chapter I Introduction. This chapter will give an overview of the background of this research, the problems that are identified, and the objectives of why this research is conducted.

Chapter II Literature Review. This chapter will discuss about the theories and results from other research that will be used to answer the research questions.

Chapter III Research Methodology. This chapter will show the methods that will be used to answer the research questions.

Chapter IV Results. This chapter will exhibit the results of the research and provide argumentation to support the results.

Chapter V Conclusion. This chapter will provide conclusion for the research topic and point out recommendations for future research.

At the end of this thesis, there will be sections of references, appendixes, and writer's biography.