CHAPTER I

INTRODUCTION

1.1 Background

Education is one of the most important factors in the quality of a nation. The more educated the citizen, the more prosperous the nation would be. Indonesian National Education aims to educate the life of the nation, to form a religious society, to uphold diversity, democratic and dignified, advanced-civilization, and to improve the welfare of humanity physically and mentally [1]. Moreover, education in this 21st century has developed to be more than just delivering information or giving the students access to information. Education should be able to prepare students to face current complex and changing world challenges.

All educational institutions are responsible for the quality of educational services they provide. Academic achievement serves as an important instrument for evaluating educational quality and its assurance [2]. It reflects the outcomes that show how well an individual has achieved set of goals within educational settings like schools, colleges, and universities [3]. Gradually, the study about academic success has become a central research topic for scholars and experts [4]. Therefore, students' academic achievement can be seen as an essential indicator of students' success that can determine the quality of education.

Mathematical knowledge is a key area for acquiring skills that is relevant to 21st century education. It forms the cornerstone of scientific and technological understanding, which is very crucial for the nation's socio-economic development [5]. Many studies indicate that Mathematics influences all facets of human life at

multiple levels, underscoring it importance in understanding the world around us [5], [6], [7], [8]. Among all subjects taught in formal education, Mathematics' achievement has a special concern in Indonesian education. Mathematics is considered as the most difficult subject among all the various subjects taught in schools by most of Indonesian students [9]. The latest Indonesian National Assessment 2021 was conducted to more than 259 thousand schools with 6.5 million students. The result recorded that in numeracy aspect, 2 out of 3 students had not yet reached the minimum competency [10]. Referring to the results of the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS), Mathematics' scores for both assessments have not shown significant improvement since 2000. Both PISA and TIMSS are international assessments that have been widely used across the world to measure 15-years old students' and grade four ability in Mathematics. Overview of Indonesian Mathematics' PISA and TIMSS ranking[11], and the latest Mathematics PISA trends [12] can be seen on Table 1.1 and Figure 1.1 respectively.

Table 1. 1. Mathematics ranking of Indonesian students by PISA and TIMSS

Year	PISA	PISA Score	TIMSS	TIMSS Score
1999/2000	39 of 41	367	34 of 38	403
2003	38 of 40	360	35 of 46	411
2006/2007	50 of 57	391	36 of 49	397
2009	61 of 65	371	-	-
2011/2012	64 of 64	375	38 of 42	386
2015	63 of 70	386	45 of 50	397

Figure 1. 1 Indonesian 2003-2018 PISA score trends
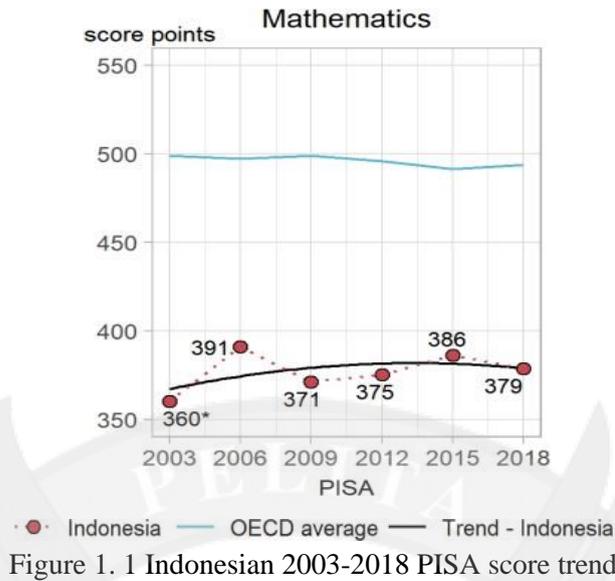
PISA and TIMMS evaluation pointed out that Indonesian students still lacked logic and reasoning in completing Mathematics competency test [13]. Following up the results, the Director of the National Educational Evaluation Centre (NEEC), Nizam argued that majority of Indonesia students are good at answering questions that require memorization only. Whereas the 21st century requires the competence of good literacy and numeracy that includes reasoning, critical, creative, communicative, and collaborative thinking [14]. Based on the parameter of PISA and TIMSS, Mathematical literacy and numeracy have now become the main focuses of the government in the renewal of national curriculum into Merdeka Curriculum. It aims to improve students' thinking skills in using mathematical concepts, procedures, facts, and tools to solve contextual problems [15]. Therefore, various efforts are continually made by schools in Indonesia as to improve the students' ability in learning Mathematics since primary until higher education.

In a private school at Tangerang where this research is initialized, the researcher, who is a Mathematics teacher, found out that Mathematics is considered

as the most challenging subject compared to other subjects. Figure 1.2 shows the distribution of Mathematics Report scores of 280 grade 6 students on year 2019/2020. It is clearly seen in Figure 1.2 that Mathematics score is the lowest compared to other subjects and it also has the most variance between the low-achievers and high-achievers. This causes Mathematics to be the most challenging subject for students, including teachers and parents in assisting the students' learning. Considering the importance of critical thinking skills as one of the school's visions and adjusting to the Singaporean and National 2013 Revised Curriculum that are used in the school, the teachers are continually trying their best to facilitate the students' Mathematics learning from elementary until secondary level.
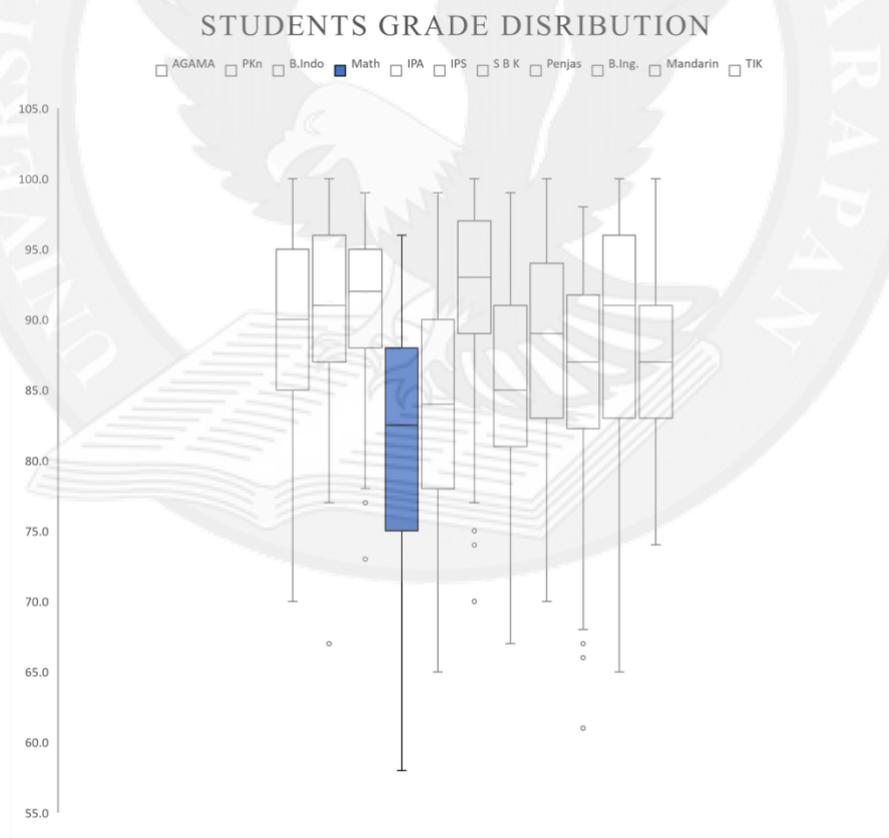
Figure 1.2 Students' grade distribution on all subjects

Many factors must have influenced the students' academic achievement in Mathematics. Gender, economic, social, and cultural status, time spent learning Mathematics, affective, and socioeconomics features have significant effect on students' mathematical skills [16]. A recent study by Alyahyan and Düştegör [17] identified five key factors that often explored by researchers in predicting students' performance: academic achievement, demographic factors, learning environment, psychological traits, and learning activities. By knowing the factors that affect students' achievement, school can plan for the best intervention to improve students' academic achievement in Mathematics. Thus, it is important to analyze the factors that influence students' Mathematics achievement before providing intervention to the students learning.

Many times, intervention on students learning would be given after the teacher evaluates several tests near the end of the first term. This is not effective to help the students because the content itself is getting more advanced. It is also not efficient because the time for further intervention in the second term is getting shorter at the end of academic year. In fact, early and accurate detection on students who perform low, medium, or high achievement will otherwise help the school to provide earlier, relevant, and focus intervention. It can facilitate them to enhance their learning and achievement, especially for the students who are at risk of failure. For example, the school might provide personalized learning system to mitigate failure for the low-performed students at early stage, or scholarship recommendation/financial supports for high-performed students which will benefits the students, parents, and the school itself. This is supported by recommendation from [5] who conducted systematic review about factors affecting

students' achievement in Mathematics in developing countries. Based on ERIC data base and Google Scholar/Search, it concluded that all the factors that influence students' achievement should be addressed early in the students' career so there will be more success on the enrolment to higher education.

In the process of learning over long periods of time, a large amount of data about students is inevitably collected by the schools. This data includes academic and non-academic records, or other information the contribute to it. Regularly, after reporting the students' academic achievement to their parents every semester, those data will be archived in the curriculum database system every academic year and rarely be used. Since the amount of data recorded will increase every year, there comes a great opportunity to utilize the increasing data for the benefit of schools' practices rather than just storing it as archives. Therefore, students' large amount of data recorded about the academic achievement of students and other many related components could definitely be analyzed optimally by using data mining.

Data mining is widely known as the process of mining knowledge from a large amount of data. It is powerful in uncovering hidden information within large data volumes and it has been widely used in various fields, such as marketing analysis, medicine, manufacturing, engineering, education, etc. In other words, data mining can be said to be the process of applying various algorithms to discover patterns and relationships within dataset. It offers more comprehensive understanding of the data compared to traditional statistical methods, uncovering unseen patterns, creating predictive models, facilitating improved data driven decision-making, and guiding future improvements[18]. Therefore, the large

amount of data recorded about the students' academic achievement and other many related factors could be utilized optimally by using data mining.

The use of data mining tools and technologies in educational settings is known as Educational Data Mining (EDM). It has become prominent research field that focuses on extracting valuable insights from educational databases for a variety of purposes [19]. Along with the increasing utilization of technology in educational settings, EDM can recommend suggestions that can help schools to improve the education service. EDM explores diverse data features from educational environments to improve understanding about students and their learning contexts [20]. It offers a more thorough and precise understanding of the data than traditional statistical approaches. It also uncovers hidden pattern and develop predictive models that empowers educators to make informed decisions for future improvements [18]. Numerous studies on EDM have proven to improve decision making for educational institutions.

Previous studies in EDM mostly focused on three main categories, which are detecting the causal structure and relation of features, predicting student performance/achievement/learning outcome/dropout rates, and observing key features that have a high influence on student performance and learning behaviors [21]. Predictive models built by EDM are frequently used in to forecast expected future trends, events, and behaviors, including students' performance based on historical data [22]. Various classification algorithms such as Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbor (K-NN), and Artificial Neural Network (ANN) have been widely implemented in EDM to predict the student achievement [23]. Moreover, to enhance the classifiers

performance, the ensemble methods such as Bagging, Boosting and Stacking that combine more algorithms were applied [24]. A recent systematic literature review about predicting students' performance from 2015-2021 by [23] highlighted that when a tremendous amount of data is collected to make prediction on a binary or multi-class classification, imbalanced distribution might cause bias and erroneous, but not many methods or algorithms found to deal with imbalanced issue comprehensively. Thus, the use of EDM in predicting students' achievement on imbalanced class is a great potential be investigated as to improve the prediction accuracy of the models.

Research in the fields of EDM is conducted in various contexts, such as different complexities of the datasets, different significations of the contributions, different performances of the algorithms used [21]. Different algorithms handle a problem differently, so it is a promising idea to observe the best algorithms in predicting the students' achievement. Following its emerging significant contributions on different areas including education, Abbor in [18] also initiated the discussion on the use of EDM on more specialized and oriented work of educational domain, such as Mathematics education. There are not many studies focusing on implementing EDM in Mathematics area. Refers to some systematic literature reviews on EDM in predicting students' achievement [17], [23], [24], [25], [26], it can be seen that EDM is mostly utilized only in on population of university students, some in higher school, but very rare in secondary, even elementary students. Most studies have focused on predicting the final GPA, success/failure on certain courses or exam in higher education, students drop-out rates, but not specific in Mathematical achievement for Elementary level that also significance as the earliest

level of education. In fact, there are quite a number of graduate students from elementary level who have not assisted optimally in their Mathematics achievement before continuing to secondary level. For instance, lack of Mathematics ability that can lead to the high are risk of failure in secondary, minimum personalized learning system to mitigate failure for the low-performed students, or lack of earlier intervention of recommendation/financial supports to high-performed students that can bring benefits the students and the stakeholders. Thus, this research will focus on the implementation of EDM techniques to predict the students' academic achievement in Mathematics at Elementary level.

Considering the background description above, this research will compare several techniques in this research to find the best algorithm to predict the students' academic achievement in Mathematics. The predictive model from EDM can bring useful knowledge for the school's decision making. By then, schools can categorize the abilities of students who are performing low, medium, high and provide early intervention in the learning process. Therefore, this research will implement EDM techniques to predict the students' academic achievement in Mathematics in Elementary school.

1.2 Problem Identification

The identification of problems in this research are:

a. The importance of Mathematics' competence has not followed by the improvement in students' academic achievement. Lower achievement in Mathematics represents the low level of thinking and reasoning skills among the students in the school and as a nation. Moreover, Mathematics

becomes the most challenging subject that also has a wide gap among the high-, middle-, and low-level students.

b.   Late and inefficient intervention by teachers/schools lead to the need for an early and accurate mapping of students who have low, middle, high achievement in Mathematics.

c.   Increasing number of data recorded by school every semester that is not followed by optimal utilization of it. There is a need to optimize the utilization of technology in educational settings using EDM techniques to provide data-driven decision making in the school.

d.   The fact that EDM is emerging, and it is utilized mostly for higher education only. There are still inadequate experiments or studies about implementation of EDM in Mathematics for elementary level, a learning period and specific subject matter that are also crucial for the success of education in the next level. Moreover, there are still open opportunities of utilizing EDM in Mathematics subject area with more combination of learners that can increase the accuracy of the prediction on imbalanced classes.

1.3   Problem Limitation

Limitation of problems in this research are:

a.   Considering the number of students features of both academic records and demographic records, the size of the dataset is relatively small because it comes from one academic year only (before pandemic). Not many features can be recorded well during online learning, such as

extracurricular activities, student' attendance list, the reliability of the students test results, etc.

b. The number of subjects and attributes that are limited and specific to one private school in Tangerang, that will not generalize the results to other educational settings.

c. The process of data collection from different departments in the school took more time and so required careful handling in the data selection process to avoid erroneous data.

1.4 Research Problem

The research problems that will be answered in this research are:

a. How to implement EDM in predicting students' academic achievement in Mathematics at Elementary School?

b. What is the best algorithm in predicting students' academic achievement in Mathematics at Elementary School?

1.5 Research Objectives

The objectives of this research are:

a. Implementing EDM to predict students' academic achievement in Mathematics at Elementary School.

b. Finding the best algorithm to predict students' academic achievement in Mathematics at Elementary School.

1.6 Research Outline

The outline of this research consists of:

Chapter 1 – Introduction. This chapter talks about how the background of problems in this research is formulated that cause this research to be conducted. It discussed the current context of students' achievement in Mathematics that provides the challenges that this research aims to address. It continues with the clear articulations of problem identification, limitation of the problems, central issue of research problems, purposes that becomes objectives , and research outlines that tells the roadmap of the whole structure of research.

Chapter 2 – Literature Review. This chapter discusses the existing knowledges, theories, definition, methods that are related to the topic of this research. It starts form broader perspectives of academic achievement then narrows down to the highlight of its application in the mathematics teaching in the Elementary and factors that influence it. Then, the chapter transitions to the discussion about data mining and how it impacts on the field of education, especially in the prediction of Mathematics achievement of students in the school. It ends by addressing the gap and its conceptual framework before continuing to the methodology in the next chapter.

Chapter 3 – Research Methodology. This chapter describes how the methodology if this research is conducted to achieve the goals of the research. It gives outlines and procedures used in this research, starting from the flowchart diagram, environment within which the whole Educational Data Mining methods are conducted. Then, it is followed by the details about from data selection stage until applying machine learning algorithms and its evaluation and interpretation.

Chapter 4 – Result and Discussion. This chapter presents the core findings of the thesis. It tells about the detail of experiments being conducted using classification

algorithms without resampling techniques, and within resampling techniques (oversampling, under-sampling, and hybrid-sampling). The best algorithm for predicting student's achievement in Mathematics is then identified using the evaluation metric chosen in the previous chapter.

Chapter 5 – Conclusion and Recommendations. This chapter synthesis the main findings and draws conclusion based on the evidence presented in previous chapters. It reflects the implications of the results and how they contribute to the field of educational data mining and prediction of students' achievement. Recommendation on the practical applications based on findings are suggested as to propose areas for future research. The research ends by presenting a comprehensive list of references, acknowledging the sources that cited in the research.