# CLASSIFICATION ANALYSIS OF SLEEP QUALITY ON THE ACTIVE PERFORMANCE

By:

**TAN HANN WOEI**

(Matrix No.: 128972)

Supervisor:

**DR. LOH WEI PING**

May 2019

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfillment of the requirement to graduate with honors degree in
**BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)**

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

**Declaration**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed……………………………………………… (TAN HANN WOEI)

Date…………………………………………………

**Statement 1**

This journal is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/ references are appended.

Signed……………………………………………… (TAN HANN WOEI)

Date…………………………………………………

**Statement 2**

I hereby give consent for my journal, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed……………………………………………… (TAN HANN WOEI)

Date…………………………………………………

## Acknowledgments

First of all, I wish to express my appreciation to School of Mechanical Engineering, Universiti Sains Malaysia for providing me with a good opportunity to conduct Final Year Project. Final Year Project is executed in the period of two semesters to preparation, set-up and perform the experiment which is quite relevant to real life.

I also feel very grateful and like to express my greatest respect to my supervisor, Dr. Loh Wei Ping who always provides me with a lot of useful guidance and suggestions throughout the project with her patience and knowledge. She leads me to the solution when I am wondering about the problems faced throughout my Final Year Project.

Lastly, I would also want to thank my seniors and course mates who provide me with a lot of support throughout this project.

**Table of Contents**

**List of Figures**

## List of Tables

**Nomenclature/Symbols**

| | |
|---|---|
| REM | - Rapid Eye Movement |
| NREM | - Non-Rapid Eye Movement |
| WEKA | - Waikato Environment for Knowledge Analysis |
| USM | - Unoverisit Sains Malaysia |
| OSA | - Obstructive Sleep Apnea |
| RBD | - REM Sleep Behaviour Disorder |
| VD | - Vascular Depression |
| HAM-D | - Hamilton Depression Rating Scale |
| HAM-A | - Hamilton Anxiety Scale |
| PSQI | - Pittsburgh Sleep Quality Index |
| ESS | - Epworth Sleepiness Scale |
| PSD | - Polysomnography |
| NVD | - Non-Vascular Depression |
| SRBD | - Sleep-related Breathing Disorder |
| TEA | - SRA Test of Educational Ability |
| OS | - Operating System |
| SMO | - Sequential Minimal Optimatization |
| IBk | - Instance-based learner |
| LWL | - Locally Weighted Learning |
| LMT | - Logistic Model Trees |
| TSP | - Total Sleep Period |
| AWP | - Awake Period |
| RSP | - REM Sleep Period |
| LSP | - Light Sleep Period |
| DSP | - Deep Sleep Period |
| SC | - Sleep Cycle |

## Abstrak

Tidur yang secukupnya adalah satu aspek yang penting untuk sentiasa aktif dalam aktiviti harian. Kualiti tidur yang bagus memang membantu tumpuan otak kita, meyimpan memori dan kekal aktif dalam aktiviti fizikal. Kebanyakan penyelidikan sebelum ini hanya mengkaji faktor mempengaruhi corak tidur dan mengabaikan kesan corak tidur terhadap prestasi harian. Maka, penyelidikan ini dilaksanakan untuk (i) menentukan faktor utama yang mepengaruhi qualiti tidur, (ii) membahagikan corak qualiti tidur, dan (iii) mengkaji hubungan antara corak tidur dan prestasi aktif. Projek kajian ini mengandungi satu ujikaji dan penyelidikan soal selidik terhadap qualiti bagi prestasi aktif pada waktu pagi. Analisis data dijalankan berdasarkan pada pendekatan terhadap perlombongan data menggunakan perisian Waikato Environment for Knowledge Analysis (WEKA). Terdapat 20 orang peserta sukarela merupakan pelajar Kampus Kejuruteraan Universiti Sains Malaysia. Dua aplikasi jejakan tidur Android, iaitu PrimeNap dan Runtastic digunakan untuk mengukur corak tidur. Data tersebut digabungkan dalam Google Form bersama dengan soal selidik tentang prestasi aktif, pembahagian data kepada tiga kelas menunjukkan ketepatan dari 63.2-92.8% untuk kedua-dua PrimeNap dan Runtastic. Data tersebut dianalisis lebih mendalam melalui proses penemuan ilmu, dan tempoh tidur keseluruhan, kitaran tidur dan tempoh tidur pengerakan mata cepat didapati sebagai tiga atribut yang paling bererti untuk menentukan kualiti tidur. Walaupun demikian, qualiti tidur tidak dapat menggambarkan keadaan prestasi aktif harian.

**Abstract**

Sufficient sleep is an important aspect to maintain active for daily activities. Good sleep quality essentially helps our brain to concentrate, store memory and stay active for physical activities. Most of the previous studies on sleeping analysis focused on the factors affecting sleeping patterns instead of the sleep quality effecting on daily performances. Therefore, this study is conducted to (i) determine the significant factors affecting the sleep qualities, (ii) classify sleeping quality patterns and (iii) investigate the relationship between sleeping patterns and active performances. This research project involved an experimental and a survey study on the sleeping qualities for active performances during the daytime. Data analysis was performed based on data mining approach using the Waikato Environment for Knowledge Analysis (WEKA) software. There were 20 voluntary participants which are the students of University Sains Malaysia Engineering Campus. Two Android sleep tracking apps: PrimeNap and Runtastic were applied to measure the sleeping patterns. The data was then compiled on Google Form along with the active performance survey questionnaires, the classification of the data into three classes show accuracies ranged from 63.2-92.8% on both PrimeNap and Runtastic respectively. The data is further analysed in the knowledge discovery process and the total sleep period, sleep cycle and REM sleep period were identified as the three most significant attributes corresponding to determine the sleep quality. However, the sleeping quality could not describe the conditions of daily active performances.

# CHAPTER 1

# INTRODUCTION

## 1.0 Introduction

Sleep is a vital natural body mechanism synchronized with other body functions like breathing, heart beating and body temperature changes. Everyone needs an adequate sleep period and good sleep quality to have better effect on physical activities. A complete sleep cycle consists of rapid eye movement (REM) and non-REM (NREM) cycles that are repeated four to six times throughout the sleep per night. The light sleep and deep sleep are parts of the NREM sleep [1]. A good sleep is evaluated by the sleep quality, sleep cycle, sleep period for total sleep period and classified sleep period (for awake, REM sleep, light sleep and deep sleep). Another way to judge optimal sleeping pattern is from the efficiencies of active performances during daytime. There are many factors that could affect the sleep quality and sleep efficiency of an individual, especially the sleeping pattern. Good sleep quality and healthy lifestyle can be achieved through habitual sleeping pattern like always sleep for more than 8 hours a day. Therefore, the optimal sleeping pattern determined can lead to high active performance during daytime.

Previous researchers had reported that every adult (18–64 years old) requires seven to nine hours of sleep each night [2]. Most of the related studies were focused on sleep performance corresponds to the total sleep period and external factors, like environment and pre-sleep activities. Existing studies favoured the application of data analysis like t-test, multivariate analysis of variance, chi-squared analysis and multiple stepwise regression analyses to evaluate the relationship of the factors affecting the sleep quality and the sleeping patterns [3], [4]. Literature works were successful in determining the factors affecting sleep quality that causing sleep problems like sleep apnea, insomnia, parasomnias, circadian disorders and narcolepsy. However, the existing studies did not measure the effects of the sleeping patterns on the physical active performance during the daytime. Apparently, the sleep quality and sleep efficiency were not properly judged from the comparison between the sleeping patterns and the following day’s active performance.

Therefore, this project attempts to fill the gap by applying the data mining approach to study the sleeping quality and active performance to identify the relationship between two subjects. The purpose of the project is to (i) determine the factors affecting sleeping patterns, (ii) predict optimal sleeping patterns and (iii) investigate the relationship between sleeping patterns and active performances. An experimental and survey study on 20 participants from University Sains Malaysia Engineering Campus were conducted for a duration of 2 weeks. Two Android applications: PrimeNap and Runtastic will be employed to measure and record the sleeping patterns of the participants. The participants were required to fill up questionnaire surveys designed to investigate the details of their sleeping patterns and their following day's active performances. The data collected were subsequently processed using data mining concepts (data pre-processing and classification) in WEKA software. The sleeping patterns data are later classified into the good, moderate and bad sleep quality using the 5 classifiers of WEKA software: Bayes, function, lazy, rules and tree. The significant attribute analysis and correlation analysis are performed to discover the knowledge regarding significant factors correspond to the sleep quality. The relationship between sleep quality and active performance was determined through the classification of active performance based on the three classes of sleep quality: good, moderate and bad.

## 1.1 Project Background

The project is an experimental and analytical project that predicts the optimal sleeping quality for active performance during the daytime. There will be 20 targeted student participants whom invited to participate in the experiment from University Sains Malaysia Engineering Campus. These participants were selected for having odd sleeping patterns throughout their undergraduate studies in USM, burdened with assignments and projects tasks throughout the study. The experiment will be carried out for two weeks to record the sleeping patterns using two pre-installed android applications: PrimeNap and Runtastic Sleep Better to Android OS smartphone. PrimeNap was used to record the sleep period, sleep quality and sleep cycle. The application also categorizes the sleep periods into awake, REM sleep, light sleep and deep sleep by period and percentage. Runtastic, however, acts as a compatible app to the PrimeNap for comparison and verification of the results obtained. On the following day of the sleep experiment, the participants were requested to fill in a Google Form survey questionnaire with their sleeping patterns recorded on two applications and the active performance at three different time sessions (morning, afternoon and evening). measured on Likert scale of 1-5. Based on the data, the sleeping patterns will be classified by sleep quality categorized as good, moderate and bad. The data collected will be used for data mining analysis aided by WEKA software to interpret and predict the optimal sleeping patterns.

The data mining processes include pre-processing, classification and knowledge discovery analyses. During the pre-processing, the data will undergo processes such as data cleaning, data transformation, data reduction and discretization. The pre-processing stage solves the problems of the incomplete, noisy and inconsistent data. Then, data classification process is carried to divide the pre-processed data into sleep quality classes: good, moderate and bad. The evaluated data will be checked on the its reliability based on the ZeroR algorithm. After the data mining process, the knowledge discovery will be carried out on the processed data. The The classified data will be analysed based on the active performance evaluated from survey studies.

## 1.2 Problem Statement

Based on most of the previous studies, sleep quality is mostly judged depending on the total sleep period of the participants. There is some lacking in terms of factors that should be considered like the sleeping pattern, sleep cycle, sleep period for different sleep categories (REM sleep, deep sleep, light sleep and awake). Besides, majority studies rarely applied data mining approach to classify the data obtained based on the three classes of sleep quality: good, moderate and bad.

## 1.3 Objectives

The purposes of the project are to:

(i) determine the significant factors affecting the sleep qualities,

(ii) classify sleeping quality patterns and

(iii) investigate the relationship between sleeping patterns and active performances.

## 1.4 Scope of Work

The study involves an experiment and questionnaire survey to track the sleeping patterns of 20 invited Universiti Sains Malaysia Engineering Campus students. An experimental study involves sleep pattern tracking using two Android applications: PrimeNap and Runtastic Sleep Better. Questionnaire surveys will be conducted using Google Form designed to query the active performance of the participants in terms of the type of activity, duration of active tasks and level of the active period on the Likert scale (1 to 5) at three different time phases: morning (7am-12pm), afternoon (12pm-5pm) and evening (5pm-10pm). The data collected will then be analysed with the data mining procedure in three stages. Stage 1 involves data pre-processing that includes data filtration and removal of the outliers and extreme values. Stage 2 involves the classification of the pre-processed data into the predefined class attribute, the sleep quality: good, moderate and bad. In Stage 3, classification reliability analysis is performed by comparing other classification algorithms with ZeroR algorithm, the base classifier algorithm. The significance attribute analysis and correlation analysis are performed to determine the significant data attributes. Lastly, the relationship between the sleep quality and the active performance is verified through the classification analysis of active performance corresponds to the sleep quality.

## 1.5 Thesis Outline

This thesis is organized into five chapters:

- Chapter 1 begins with a general background relating to sleeping patterns and active performance.
- Chapter 2 reviews the previous studies and converting details into relevant information about the factors and effects of the sleeping patterns.
- Chapter 3 describes the PrimeNap and Runtastic used in this study for data collection and the data mining approach for data classification analysis.
- Chapter 4 presents all the results of the analyses in values and statements to verify the objectives.
- Chapter 5 concludes the achievement of the research project and its limitations and potential improvements to be done in the future.

# CHAPTER 2

# LITERATURE REVIEW

## 2.0 Overview

There are lots of works that reported on sleep analysis. Some researchers explored the sleep problems from the patients who have diseases like Alexithymia, obstructive sleep apnea (OSA), vascular depression (VD) sleep behaviour disorder and epileptic seizure. field. There are researchers who are more interested to investigate the factors affecting the sleep quality instead of the effects of sleep quality.

## 2.1 Sleep Quality

Elhefny et al. [3] studied the prevalence of sleep-related breathing problems and sleep disturbances among health-related employees and workers at Fayoum University hospitals. Their study showed the risk of sleep disorder was observed to be more in rural living, educated and married health care workers. Urbanization and large scale of industrialization in rural areas are the main incidences of sleep problems among the rural living. The environmental factors are causing quite some effects on the sleeping quality based on this study.

Murphy et al. [5] examined the relationship between alexithymia and sleep quality. Alexithymia is an inability to identify and express or describe one's feelings. Regression analyses were used to predict sleep quality from age, gender and alexithymia total scores. Their study confirmed that alexithymia is associated with poor sleep quality ($p < 0.001$), but the impact did not vary across genders (standardized $\beta = 0.154$, $t = 1.302$, $p > 0.05$).

In view of the motion perspectives, Trujillo et al. [1] analyzed breathing and heart rate during sleep in a common bed overnight. The measurement is performed with the help of sensors placed between the mattress and the frame. A two-stage pattern classification algorithm has been implemented using statistical analysis to recognize the position of patients. The authors proved that some specific sleeping positions can be effective in reducing sleep disorder symptoms.

Murawski et al. [6] performed meta-analysis aimed to quantify the efficacy of behavioural and cognitive sleep interventions in adults with poor sleep health, who do not have a sleep disorder. Subjective sleep quality and sleep duration were determined as the only two parameters of sleep health that improved significantly following cognitive and/or behavioural intervention. The study found that a large number of adults who report poor quality sleep, but do not have a sleep disorder.

Cooray et al. [7] proposed a fully-automated framework for REM sleep behaviour disorder (RBD) detection consisting of automated sleep staging followed by RBD identification. There is evidence suggests that the Rapid-Eye-Movement (REM) Sleep Behaviour Disorder (RBD) is an early predictor of Parkinson's disease. The study carried out an automated multi-stage sleep classification to validate a fully-automated pipeline for identifying individuals with RBD. The study found the algorithms which outperform individual metrics and demonstrates that sleep architecture and the atonia levels between sleep stages can help distinguish RBD individuals from the healthy controls (HCs).

Kazemi et al. [8] measured the cognitive performance, melatonin rhythms, and sleep after different consecutive night shifts (7 vs. 4) night shift work among control room operators within the context of the Iranian petrochemical industry. There are different methods applied for different subjects like the cognitive performance was assessed using the n-back task and continuous performance test. The saliva was collected and tested by enzyme-linked immunosorbent assay to evaluate melatonin. The Pittsburgh Sleep Quality Index and Karolinska Sleepiness Scale were used to assess sleep and sleepiness respectively. The results show that sleep quality was higher among participants in the 7N shift pattern and their melatonin rhythm and sleepiness were more organized, they were more likely to adapt themselves to the night shift as a result of longer periods of the night shift. Therefore, the odd sleeping habit can be adapted after it was performed for a certain period.

Chen et al. [9] investigated the sleep status of patients with vascular depression (VD) and analyze the characteristics of sleep by polysomnogram. All subjects were evaluated by the Hamilton Depression Rating Scale (HAM-D), Hamilton Anxiety Scale (HAM-A), Pittsburgh Sleep Quality Index (PSQI) and Epworth sleepiness scale (ESS). All these patients were monitored by polysomnography (PSG). The characteristic PSG findings of the non-vascular depression (NVD) patients were shortening REM sleep latency and REM disinhibition. Characteristics of VD with PSG patients were sleep-related breathing disorders (SRBD) and daytime sleepiness, and disorders of 24-h sleep structure. From the case study, a hidden individual factor like the disease may be lead to different sleep quality compared to a normal person.

Faulkner et al. [10] examined the effect of light interventions on sleep quality, duration and timing, and effect moderators. They involved the controlled studies in intrinsic circadian rhythm disorders (such as advanced or delayed sleep) and in neuropsychiatric disorders with assumed high prevalence of circadian dysregulation. The case study found the meaningful improvements in some sleep parameters can be achieved through altering light exposure patterns.

Ong & Gillespie [11] reviewed and assessed the current selection of sleep analysis smartphone applications (apps) available. The iOS and Google Play mobile app store were searched for sleep analysis apps targeted for consumer use. A total of 51 unique sleep apps in both iOS and Google Play stores were included. Based on the case study, more than 65% of sleep apps report on sleep structure, including duration, time awake, and time in light/deep sleep, while reporting of REM was limited. The availability of extra features was variable, ranging from 4% to 73% of apps. There is a lot of apps reviewed in the study, but the functionalities and accuracies of the apps were not tested and verified.

Sigl-Glöckner & Seibt [12] studied the relationship between sleep and cognition. In this case study, there were applications of various techniques related to ‘in vivo brain imaging’, from single synapse to large scale network activity to verify the relationship. Their study performed an optical investigation of the sleeping brain and obtained important new information on the underlying physiology and the role of sleep in plasticity. The study is also found that sleep helps clearance of toxic/ metabolic homeostasis besides the improvement in learning and memory.

Gupta et al. [4] performed an assessment of sleep schedule, pre-sleep behaviour, co-sleeping and parent’s perception of sleep of school going children. Sleep patterns were assessed using the validated Hindi version of Childhood-Sleep-Habit-Questionnaire. A comparison was made between the urban and rural group and between boys and girls. Interaction of gender, domicile and school-type was examined on the sleep patterns. From the results, television watching before bedtime was more common among urban school children and they had shorter total sleep time which leads to the signs of sleep deprivation. The study found that the interaction between gender, domicile and school-type did not have any significant effect on sleep patterns.

Lund et al. [13] characterized sleep patterns and predictors of poor sleep quality in a large population of college students. One thousand one hundred twenty-five students aged 17 to 24 years from an urban Midwestern university completed a cross-sectional online survey about sleep habits. In the study, students reported disturbed sleep and over 60% were categorized as poor-quality sleepers by the Pittsburgh Sleep Quality Index (PSQI), bedtimes and rise times were delayed during weekends, and students reported frequently taking prescription. The results demonstrate that insufficient sleep and irregular sleep–wake patterns present at alarming levels in the college student population and close relationships between sleep quality and physical and mental health, intervention programs for sleep disturbance.

El Halal & Nunes [14] investigated the association between sleep duration and weight-height development in children and adolescents. A non-systematic search in the MEDLINE database was performed using the terms anthropometry, body composition, overweight, obesity, body mass index, growth, length, short stature, sleep, children, and infants and adolescents, limited to the last 5 years. The study verified the association between shorter sleep duration and risk of overweight and obesity is well established for all pediatric age groups.

Nguyen-Michel et al. [15] compared the motor semiology of sleep behaviour disorder (RBD) during rapid eye movement (REM) with epileptic seizures in non-REM and REM sleep. This study used video recordings to compare nocturnal motor events from 15 patients with epilepsy and 15 patients with RBD. One interesting finding of the study is the association of sniffing, coughing, and changes in respiration with seizures but not with RBD events. Changes in respiratory rate and amplitude are common in REM sleep with higher variability of rate and amplitude than NREM sleep, but the changes in epileptic seizures are qualitatively more marked and longer, infrequent in normal sleep. From the results, the illness usually leads to negative influence towards sleep quality.

Based on the collective findings from the literature studies, the efficiency of the sleeping pattern (sleep quality) was commonly measured based on the sleep factors that include but the effect of the sleeping patterns on daytime active performance was not measured. Most of the studies also did not classify the sleep period into different categories, like deep sleep, light sleep and REM sleep.

### 2.2 Active Performance

Adelantado-Renau et al. [16] analyzed the association of sleep patterns with academic and cognitive performance in adolescents. A sample of 269 adolescents (140 boys) aged 14 years from the baseline data of the Deporte, ADOlescencia y Salud study completed questionnaires about sleep quality, cognitive performance, and leisure time sedentary behaviours. The results of the study showed that sleep quality (but not sleep duration) was associated with all the academic performance indicators (all $p < 0.05$) and showed no differences regarding cognitive performance. The Spanish version of the ‘‘SRA Test of Educational Ability’’ (TEA) which provides general measures of three areas of intelligence and skills of learning: verbal, numeric and reasoning abilities are applied to determine the cognitive performance in this study. However, only normal statistical analysis is performed to evaluate the results in this study.

Jike et al. [17] examined the dose-response relationship between long sleep duration and health outcomes including mortality and the incidence of diabetes mellitus, hypertension, cardiovascular diseases, stroke, coronary heart diseases, obesity, depression and dyslipidemia. The relationship between short sleep duration and health outcomes has been examined. Meta-regression analyses found statistically significant linear associations between longer sleep duration and increased mortality and incident cardiovascular disease.

Whibley et al. [18] determined whether sleep impacts the diurnal pattern of next-day OA-related pain and fatigue. Multilevel linear regression models examined interactions between sleep variables and time of next-day symptom reports. The study showed that the pain is heightened after a night of poor sleep but that the effect dissipates across the course of a day. The pain intensity varied across a given day, on average, the variation did not meet criteria for clinically important differences. From the results, the sleep is insignificant correspond to the pain and fatigue.

The previous studies did not prove the effects of the sleeping patterns on the active performance. More details and methods should be included in the analysis to improve the accuracy of the results.

# CHAPTER 3

# RESEARCH METHODOLOGY

## 3.0 Overview

The entire research project was presented at three levels: data collection, data mining and knowledge discovery. Initially, the data collection process that involves experimental and survey studies were performed. The data collected was divided into two categories: sleeping patterns and active performance. The Waikato Environment for Knowledge Analysis (WEKA) software is used to run the data mining processes in three stages: data pre-processing, classification and classification reliability analyses. The knowledge discovery analysis considering the significance attribute, correlation and classification analyses for active performance were conducted. The flow of the research project implementation was presented in figure 3.1.

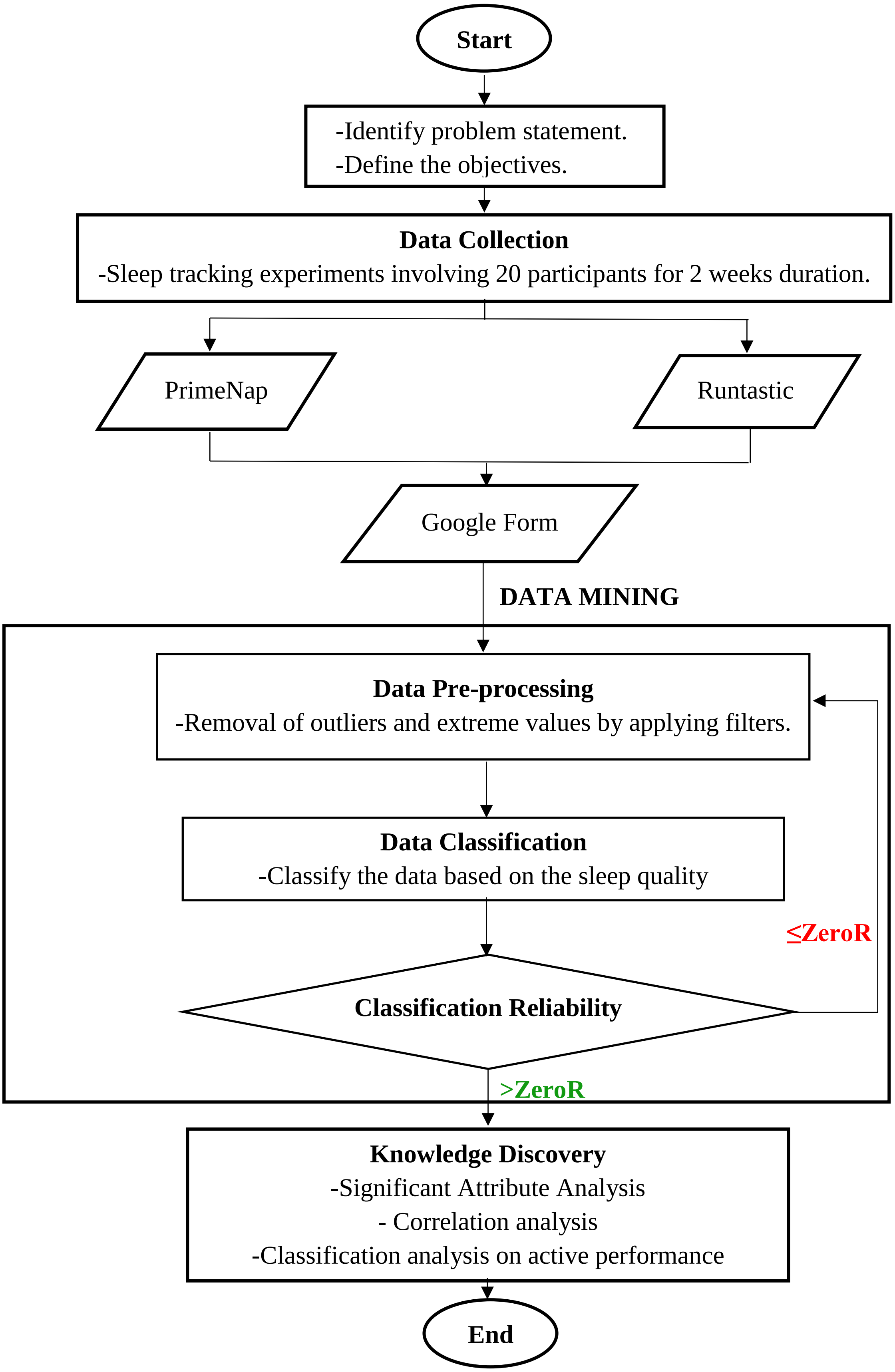


*Figure 3.1: Flowchart of the research project implementation.*

## 3.1 Data Collection

### 3.1.1 Sleep Tracking

The data collection began with sleep tracking experiment. The experiment was participated by 20 Universiti Sains Malaysia (USM) Engineering Campus undergraduate students for a duration of 2 weeks on a voluntary basis. All participants were required to install two Android sleep tracking apps: PrimeNap and Runtastic Sleep Better in their individual Android OS smartphones. Both the Android apps were initiated before the participants sleep.

The smartphone is placed at the position beside the sleeping pillow as shown in figure 3.2. The participants were advised to keep their phones in flight mode to minimize the battery consumption and the radiation emitted by the phones. The sleep tracking apps were operated to participants' movements based on the built-in sensors. The running apps will be terminated upon waking up from their sleep the next day.

Figure 3.2: Position of the smartphone during the experiment.

Data recorded in Android apps that include PrimeNap and Runtastic were filled into the questionnaire forms prepared using the Google Form as shown in Appendix 1. A sample data obtained from PrimeNap is shown in figure 3.3 which included the date of experiment, sleep time, wake up time, total sleep period, awake period, REM (Rapid Eye Movement) sleep period, light sleep period, deep sleep period, sleep cycle and sleep quality. For the Runtastic, the data obtained are almost similar to PrimeNap except that it has additional data attributes like the sleep efficiency instead of REM sleep period and sleep cycle as shown in figure 3.4. The data description of the collected data to track sleeping patterns is shown in table 3.1.

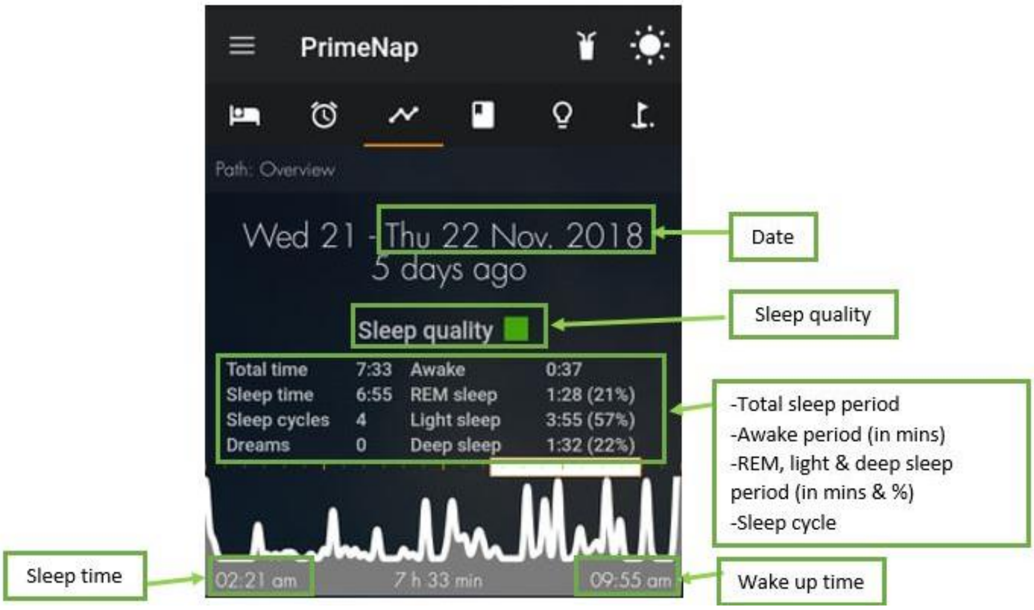


Figure 3.3: Data collected shown on the PrimeNap interface.

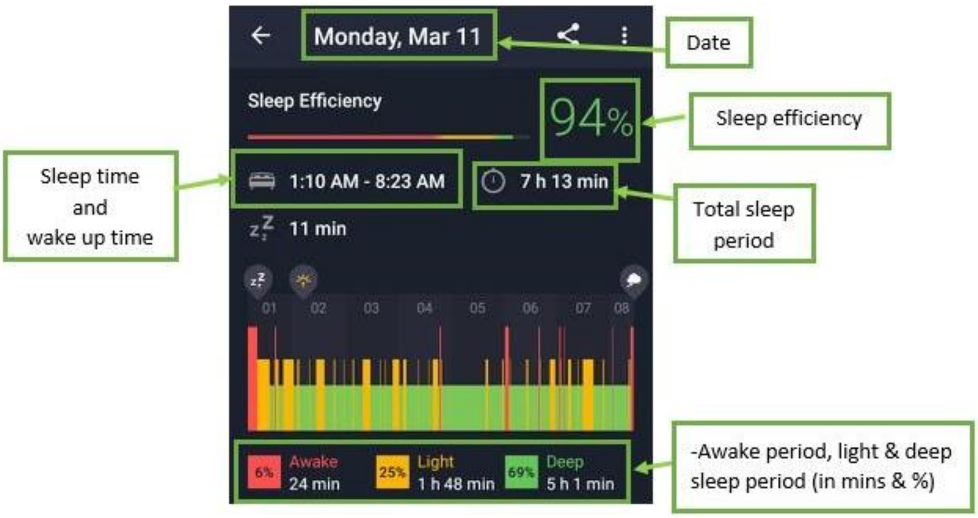


Figure 3.4: Data collected shown on the Runtastic interface.

Table 3.1: Study data description regarding the sleeping patterns.

| No. | Attribute | Description | Scale Type | Data Range |
|---|---|---|---|---|
| 1 | T | Timestamp | Numeric | [13/3/2019 11:25:03 PM- 7/5/2019 4:07:03 PM] |
| 2 | N | Name | Nominal | {CHIANG HAN XI, CHONG JIA JUN, …, YEO YING HENG} |
| 3 | G | Gender | Nominal | {Male, Female} |
| 4 | A | Age | Numeric | [24] |
| 5 | D | Date | Numeric | [12/3/2019-25/3/2019] |
| 6 | DIW | Day in week | Nominal | {Monday, Tuesday, Wednesday, …, Sunday} |
| 7 | NED | No. of experiment day | Numeric | [1st day- 14th day] |
| 8 | ST [P] | Sleep time [PrimeNap] | Numeric | [12:01:00 AM-11:47:00 PM] |
| 9 | WT [P] | Wake up time [PrimeNap] | Numeric | [5:33:00 AM-12:48:00 PM] |
| 10 | TSP (min) [P] | Total sleep period (minutes) [PrimeNap] | Numeric | [150-613] |
| 11 | AWP (min) [P] | Awake period (minutes) [PrimeNap] | Numeric | [7-70] |
| 12 | RSP (min) [P] | REM sleep period (minutes) [PrimeNap] | Numeric | [30-205] |
| 13 | RSP (%) [P] | REM sleep period (%) [PrimeNap] | Numeric | [13-38] |
| 14 | LSP (min) [P] | Light sleep period (minutes) [PrimeNap] | Numeric | [54-364] |
| 15 | LSP (%) [P] | Light sleep period (%) [PrimeNap] | Numeric | [27-84] |
| 16 | DSP (min) [P] | Deep sleep period (minutes) [PrimeNap] | Numeric | [8-188] |

| | | | | |
|---|---|---|---|---|
| 17 | DSP (%) [P] | Deep sleep period (%) [PrimeNap] | Numeric | [2-49] |
| 18 | SC [P] | Sleep Cycle [PrimeNap] | Numeric | [2-7] |
| 19 | SQ [P] | Sleep Quality [PrimeNap] | Nominal | {Bad, Moderate, Good} |
| 20 | ST [R] | Sleep time [Runtastic] | Numeric | [12:01:00 AM-11:47:00 PM] |
| 21 | WT[R] | Wake up time [Runtastic] | Numeric | [5:33:00 AM-12:48:00 PM] |
| 22 | TSP (min) [R] | Total sleep period (minutes) [Runtastic] | Numeric | [150-614] |
| 23 | AWP (min) [R] | Awake period (minutes) [Runtastic] | Numeric | [2-477] |
| 24 | LSP (min) [R] | Light sleep period (minutes) [Runtastic] | Numeric | [14-367] |
| 25 | DSP (min) [R] | Deep sleep period (minutes) [Runtastic] | Numeric | [0-538] |
| 26 | AP (%) [R] | Awake period (minutes) [Runtastic] | Numeric | [1-94] |
| 27 | LSP (%) [R] | Light sleep period (minutes) [Runtastic] | Numeric | [3-99] |
| 28 | DSP (%) [R] | Deep sleep period (minutes) [Runtastic] | Numeric | [0-96] |
| 29 | SE (%) [R] | Sleep efficiency (%) [Runtastic] | Numeric | [54-99] |

### 3.1.2 Active Performance

The data of the active performance after the participants awake the next day was also investigated through the Google Form survey questionnaire. The active performance is divided into 3 sessions: morning (7am-12pm), afternoon (12pm-5pm) and evening (5pm- 10pm). In the questionnaire, the Likert scale of 5 levels was used to determine the level of active performance, as shown in figure 3.5. The Likert-type or frequency scale was chosen by fixed choice response formats and are designed to measure attitudes or opinions (Bowling, 1997; Burns, & Grove, 1997). The descriptions of the data collected for active performance is shown in table 3.2.

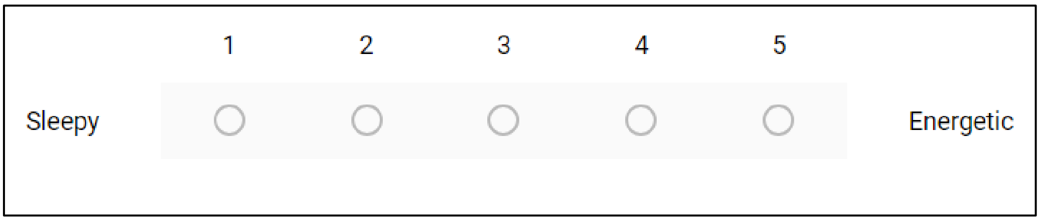


Figure 3.5: Likert scale applied in the questionnaire.

Table 3.2: Study data description regarding the active performance.

| No. | Attribute | Description | Scale Type | Data Range |
|---|---|---|---|---|
| 1 | T | Timestamp | Numeric | [13/3/2019 11:25:03 PM- 7/5/2019 4:07:03 PM] |
| 2 | N | Name | Nominal | {CHIANG HAN XI, CHONG JIA JUN, …, YEO YING HENG} |
| 3 | D | Date | Numeric | [12/3/2019-25/3/2019] |
| 4 | DIW | Day in week | Nominal | {Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday} |
| 5 | NED | No. of experiment day | Numeric | [1st day- 14th day] |
| 6 | AP (M) | Active Performance (Morning: 7am-12pm) | Numeric | [1-5] |
| 7 | AP (A) | Active Performance (Afternoon: 12pm-5pm) | Numeric | [1-5] |
| 8 | AP (E) | Active Performance (Evening: 5pm-10pm) | Numeric | [1-5] |

### 3.2 Data Mining

The data mining approach was conducted at three stages that include data pre-processing, classification and classification reliability analyses.

#### 3.2.1 Data Pre-processing

The data mining process was aided by the Waikato Environment for Knowledge Analysis (WEKA) software. The first step is data pre-processing that includes data filtration and removal of the outliers and extreme values. This process can help to remove the unwanted data to simplify the process of the data classification. During the data pre-processing stage, an unsupervised attribute filter interquartile range shown in equation (3.1) was applied.

$$Interquartile\ Range\ (IQR) = Q3 - Q1 \tag{3.1}$$

whereby Q3 is the upper quartile and Q1 is the lower quartile.

The outliers of the data were determined based on the interquartile range shown in equations (3.2) and (3.3).

$$Outlier < Q1 - 1.5(IQR) \tag{3.2}$$

$$Outlier > Q3 + 1.5(IQR) \tag{3.3}$$

The extreme values of the data, whereas, were determined based on the interquartile range shown in equations (3.4) and (3.5).

$$Outlier < Q1 - 3(IQR) \tag{3.4}$$

$$Outlier > Q3 + 3(IQR) \tag{3.5}$$

The outliers and extreme values are values beyond the equations (3.2) and (3.3). They were removed on unsupervised data instance filter ‘RemoveWithValues’. The results obtained after the filter is named as the pre-processed data.

### 3.2.2 Data Classification

During the data classification process, the pre-processed data will be classified into the predefined class attribute. In this study, the predefined class is sleep quality. There are three classes in the sleep quality attribute: good, moderate and bad. A total of 29 algorithms from five built-in classifiers of WEKA was applied to perform the classification. The reason for applying all algorithms was to compare the algorithms and obtain the best algorithms which can provide the best classification analysis. The entire list of algorithms used is as shown in table 3.3. The data classification analyses are carried out on the raw data, pre-processed data and significant data of PrimeNap and Runtastic at 10-fold cross validation mode to compare the classification accuracies.

In order to select only significant data attributes collected from the PrimeNap and Runtastic, insignificant data attributes that include personal information like the name, age and gender were removed. There are two sets of similar data attributes analysed by two apps: the PrimeNap and the Runtastic data. The classification accuracies of two sets of data are compared to determine the reliability of the apps.

Most of the data attributes are present in both apps like the sleep time, wake up time, total sleep period, awake period, light sleep period, deep sleep period. There are only some differences between the data attributes collected for PrimeNap and Runtastic. PrimeNap also records the REM sleep period and sleep cycle whereas, Runtastic records the sleep efficiency. In order to standardize the two apps' comparisons, only similar data attributes like total sleep period, awake period, light sleep period, deep sleep period were included in the analysis.

Table 3.3: List of 29 algorithms applied in the data classification analysis.

| Classifier | Algorithm |
|---|---|
| Bayes | BayesNet |
| | NaiveBayes |
| | NaiveBayesMultinomial |
| | NaiveBayesMultinomialText |
| | NaiveBayesMultinomialUpdateable |
| | NaiveBayesUpdateable |
| Function | Logistic |
| | MultilayerPerceptron |
| | SimpleLogistic |
| | SMO |
| Lazy | IBK |
| | KStar |
| | LWL |
| Rules | DecisionTable |
| | JRip |
| | OneR |
| | PART |
| | ZeroR |
| Trees | DeicisionStump |
| | HoeffdingTree |
| | J48 |
| | LMT |
| | RandomForest |
| | RandomTree |
| | REPTree |

### 3.2.3 Classification Reliability

The data mining analysis is proceeded with the classification reliability analysis to verify the results obtained from the classification analysis. The analysis is carried out in the WEKA experimenter platform using the experimental data collected from PrimeNap because PrimeNap has higher overall classification accuracy than the Runtastic. The algorithm with the highest accuracy from each classifier was selected and compared with the ZeroR algorithm (benchmark). The classification reliability is verified if an algorithm shows higher classification accuracy than the base classifier algorithm accuracy.

## 3.3 Knowledge Discovery

There were three analyses performed in the knowledge discovery process. First, the significant attribute analysis was conducted to determine the most significant attribute. The algorithms of “Tree” classifier including the J48, REPTree and RandomTree which could construct tree diagrams were employed for the analysis. The position or frequency of a data attribute in the tree diagram determines its significance in the classification analysis. The higher the data in the tree diagram, the more significant it is.

The analysis proceeds with a correlation analysis between attributes identified from the tree diagram. The correlation analysis between data attributes is performed using the PrimeNap experimental data with Microsoft Excel spreadsheet. The correlation analysis corresponds to the most significant attribute was performed. Correlation value ranges from -1 to 1. The closer the value to 1 or -1, the higher the correlation. In other words, the correlation is lower when the value is closer to 0. The high value of the correlation between the range 0.8-1 is desired in this study to verify the correlation between the attribute with sleep quality. The higher the value of the correlation of an attribute, the more significant the particular attribute with the sleep quality

The classification analysis of active performance into sleep quality categorization is carried out. The purpose was to verify the relationship between active performance and sleeping patterns. The same 29 classifier algorithms were applied for evaluations.

# CHAPTER 4

# RESULTS AND DISCUSSION

## 4.0 Overview

Data mining analysis were executed at three levels: data pre-processing, classification and classification reliability. At data pre-processing, 13 outliers and extreme values were identified. Data classification on the pre-processed data shows an accuracy range of 63.2%-92.8% (Table 4.1). During the classification reliability analysis, the best algorithms from each classifier which are BayesNet (91.58%), MultilayerPerceptron (91.45%), LWL (88.99%), OneR (93.01%) and REPTree (93.00%) has much higher classification accuracy than the benchmark algorithms, ZeroR (63.29%). Therefore, the classification analyses are verified to be reliable.

The knowledge discovery is performed through significance attribute analysis, correlation analysis and classification analysis of active performance. Through the significance attribute analysis, the total sleep period is verified to be the most significant attribute corresponds to the sleep quality. Next, the sleep cycle and REM sleep period are verified as the second and third most significance attribute. The classification analysis of the active performance proves that the sleeping patterns could not be described by the active performance.