AN ENSEMBLE-BASED REGRESSION MODEL FOR PERCEIVED STRESS PREDICTION USING RELEVANT PERSONALITY TRAITS

CHANG HON FEY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

AN ENSEMBLE-BASED REGRESSION MODEL FOR PERCEIVED STRESS PREDICTION USING RELEVANT PERSONALITY TRAITS

CHANG HON FEY

DESSERTATION SUBMITTED IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SOFTWARE ENGINEERING

FACULTY OF COMPUTER SCIENCE AND INFORMATION
TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: CHANG HON FEY

Registration/Matric No: WGC140003

Name of Degree: Master of Software Engineering

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

An Ensemble-Based Regression Model for Perceived Stress Prediction using Relevant Personality Traits

Field of Study: Human Computer Interaction

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every right in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature	Witness's Signature
Date:	Date:
Name:	Name:
Designation:	Designation:

ABSTRACT

This study compared various machine learning methods to develop an accurate predictive system to predict perceived stress in regression problem with relevant personality traits. The machine learning methods that were identified and being compared including the single regression models (Multiple Linear Regression, Support Vector Machine for regression, Elastic Net, Random Forest, Gaussian Process Regression, and Multilayer Perceptron), homogeneous ensemble models (Bagging, Random Subspace, and Additive Regression), and heterogeneous ensemble models (Voting and Stacking). The dataset for the training and testing the predictive methods was taken from a study which the survey was distributed to the public in Melbourne, Australia and its surrounding districts. The selected predictors for perceived stress include gender and six personality traits, namely; mastery, positive affect, negative affect, life satisfaction, self-esteem, and perceived control of internal states. The predictive performances of all the predictive methods were compared, and the benchmark single model was identified. The ensemble instances with certain combinations of single models as base learners and with certain meta learners were proven to perform better than the benchmark single model. The implications and recommendations were discussed in this study.

ABSTRAK

Kajian ini membandingkan pelbagai kaedah pembelajaran mesin untuk membangunkan sistem ramalan yang tepat bagi ramalan persepsi tekanan bagi masalah regresi dengan meggunakan sifat keperibadian yang berkaitan. Kaedah pembelajaran mesin yang dikenal pasti dan dibandingkan termasuk model regresi tunggal (Multiple Linear Regression, Support Vector Machine for Regression, Elastic Net, Random Forest, Gaussian Process Regression, and Multilayer Perceptron), kaedah ensemble homogen (Bagging, Random Subspace, and Additive Regression), dan kaedah ensemble heterogen (Voting and Stacking). Dataset yang digunakan untuk melatih dan menguji kaedah ramalan telah diambil dari suatu kajian yang soal selidiknya telah diedarkan kepada orang awam di Melbourne, Australia dan daerah sekitarnya. Peramal yang dipilih bagi persepsi tekanan termasuk jantina dan enam sifat keperibadian, seperti penguasaan, perasaan positif, perasaan negatif, kepuasan hidup, harga diri, dan persepsi kawalan dalaman. Keputusan ramalan bagi semua kaedah ramalan telah dibandingkan, dan model tunggal penanda aras telah dikenalpasti. Kaedah ensemble dengan kombinasi model tunggal tertentu seperti 'base learners' dan dengan 'meta learners' tertentu terbukti dapat meramal lebih baik daripada model tunggal penanda aras. Implikasi dan cadangan telah dibincangkan dalam kajian ini.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my main supervisor Dr. Mumtaz Begum Binti Peer Mustafa for her continuous guidance, patience, motivation, full support, prompt response, and professionality throughout the whole process of my research. I would like to thank my co-supervisor Dr. Asmiza Binti Abdul Sani for her guidance and advice in helping me to beautify and restructure my research chapters. I also thank Dr. Mehdi for teaching me about the WEKA software. Besides, I would like to express my gratitude to my spiritual parents, Rev. Ezekiel Chong and Pr. Charlotte Tsen for the spiritual support and prayers during the times of hardship throughout my research. Nevertheless, I would like to express my sincere appreciation to my beloved and supportive wife, who took care of our kids and the housework so that I could focus on the research. Finally, I give all the glory to the Lord God Jesus Christ for giving me chance to further study and leading me through every step in my life.

DEDICATION

For my beloved wife who is the faithful supporter behind my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

ML	:	Machine Learning
MLR	:	Multiple Linear Regression
MLP	:	Multilayer Perceptron
SVM	:	Support Vector Regression
ELN	:	Elastic Net
RF	:	Random Forest
GPR	:	Gaussian Process Regression
RSS	:	Random Subspace
BG	:	Bagging
AR	:	Additive Regression
MAE	:	Mean Absolute Error
RMSE	:	Root Mean Squared Error

CHAPTER 1: INTRODUCTION

Lifestyles today are full of intensifying stress. The negative consequences of stress are quite worrying with, many mental health problems such as depression, hopelessness, and suicidal ideation caused by stress (Pearlin, Menaghan, Lieberman & Mullan, 1981; Marshall, Davis, Sherbourne & Morland, 2000; Ciarrochi, Deane & Anderson, 2002; Schönfeld, Brailovskaia, Bieda, Zhang & Margraf, 2016). Stress identification is important, because it helps to determine whether an individual need special treatment due to stress. However, stress identification alone is not enough, before one was diagnosed to have high degree of stress, one may have dwelled in serious stress-related problems, such as mental health problems.

Due to the growing awareness of stress-related health problems, prediction of perceived stress is urgently needed so that early intervention can be conducted before the mental health problems manifested. When developing the predictive model of stress, selecting the right predictors (attributes) is very important because it will help to eliminate redundant predictors and improve the prediction accuracies. Many criteria could be the predictors of stress, for example, family background, demographic characteristics, social-economic factors, and so on. Some studies found that personality traits are important predictors of stress (Pallant & Lae, 2002; Lazarus, 2006; Schaefer et al., 2017), however, most of the studies only focused on several personality traits, but not the comprehensive list of personality traits which are relevant to stress predictions.

Most importantly, the predictive performance (or accuracy) of the predictive model is the main concern of the predictive study. In the past, most of the social science research uses Multiple Linear Regression (MLR) to build the predictive models for stress in regression problems. Nowadays, with the advancements in computer science, many have started using other Machine Learning (ML) models to improve the performances of the stress-related predictive models, but majority focused on classification problems (Subhani, Mumtaz, Saad, Kamel & Malik, 2017; Smets et al., 2015; Bogomolov, Lepri, Ferron, Pianesi & Pentland, 2014a, 2014b). Further research is needed to find out the suitable ML regression models for the prediction of stress-related regression problems. Besides, the performance of the ensemble regression models (which normally being used to improve the performance of the single learning model) in predicting regression problems of stress are also left unknown.

1.1 Background of the Study

In the past, plenty of social and behavioral science research was done to explore the important topic of stress, such as definitions and process of stress (Selye, 1936; Lazarus & Folkman, 1984; Butler, 1993), measurements of stress (Holmes & Masuda, 1974; Cohen, Kamarck, & Mermelstein, 1983; Brown & Harris, 1989; Karasek et al., 1998; Muscatell & Eisenberger, 2012; Gianaros & Wager, 2015), and predictors of stress (Greer, 2008; Shah, Hasan, Malik & Sreeramareddy, 2010; Heinze, Stoddard, Aiyer, Eisman & Zimmerman, 2017). Those research founds that perceived stress is more applicable in explaining stress (Lazarus & Folkman, 1984; Butler, 1993; Monroe & Kelley, 1997) and the reactions under stress cannot be predicted without references to personality traits (Lazarus, 2006).

The traditional way of prediction in social and behavioral science focuses on identifying the predictors of perceived stress and understanding the relationships between the predictors and perceived stress. However, this was not adequate to improve the predictive performance of a perceived stress prediction system. Indeed, the Machine Learning (ML) models from Computer Science are focusing on improving the predictive performances of the predictive models. Most of the stress-related predictive research uses

ML classification models to predict the categorical outcomes (Scherer et Al., 2008; Plarre et al., 2011; Sharma & Gedeon, 2012; Smets et al, 2015; Subhani et al., 2017), but very little research focuses on using ML regression models to predict the stress-related numerical outcome like perceived stress. For other domains, the commonly used ML regression models (refer to Chapter 2) that performed better than others in the comparison studies were the Multiple Linear Regression (MLR), Support Vector Machine for regression (SMOreg or SVM), Elastic Net, Random Forest, Gaussian Process Regression (or Kriging), and Multilayer Perceptron (MLP). Predictive studies for perceived stress were commonly done using MLR (Moon, Seo & Park, 2016; Heinze, Stoddard, Aiyer, Eisman & Zimmerman, 2017), but the predictive performances of SVM, Elastic Net, Random Forest, Gaussian Process Regression, and Multilayer Perceptron for the prediction of perceived stress are unknown.

Besides, some stress-related predictive studies that focused on classification problems found that ensemble models out-performed the single models (Chowdary, Devi, Mounika, Venkatramaphanikumarm & Kishore, 2016; Rosellini, Dussaillant, Zubizarreta, Kessler & Rose, 2018). Ensembles have also been studied for regression task, such as rainfall forecasting (Wu & Chen, 2009), wind and solar power forecasting (Ren, Suganthan & Srikanth, 2015), financial domains (Jiang, Lan, & Wu, 2017), and imbalanced regression tasks (Moniz et al., 2017). However, regression ensemble studies that related to stress are very limited and need more exploration, especially for perceived stress.

1.2 Research Motivation

Current sensory, physical and physiological measures could only detect stress level but is unlikely to make prediction on possible stress. Stress prediction is important because many mental health problems require early prevention. When the stress level can be detected by physical measures, which mean that the consequences of the negative stress have partially manifested physically, and the condition of the mental health problems can be more difficult to be treated. Choosing the appropriate ML models to predict perceived stress with relevant personality traits could help to develop an accurate perceived stress prediction system, which focuses on the individual's inner thought because, only from within, the perceived world of an individual that the true meaning of the event that the individual experiences can be understood. Besides, early prevention actions can be taken to deal the stress before the crisis occurs. The perceived stress prediction system can be embedded in different devices and being used by different parties to help individuals maintain and restore good mental health.

1.3 Problem Statement

Stress plays important role to motivate people to better achievement, however, excessive stress can seriously destroy ones psychologically and physically. Researchers from social and behavioral science have been developing instruments to measure the perceived stress and identifying the predictors of perceived stress. In other words, social and behavioral science focuses on identifying the relationship between the predictors and perceived stress. However, an accurate perceived stress predictive system is needed to predict the stress an individual has perceived in advance and alert the concerned parties if one's perceived stress was predicted to exceed a potential degree, so that early interventions can be planned and implemented to avoid the negative consequences of over-stress, such as depression, hopelessness, and suicidal ideation.

For the development of a perceived stress prediction system, an accurate predictive model is required, and the suitable dataset with the relevant predictors is needed to train the predictive model. Researchers from Computer Science have developed many Machine Learning (ML) models which may produce accurate predictive results for certain datasets. As perceived stress is measured in numerical scale, predictive of perceived stress by the ML is performed using the linear regression. However, predicting perceived stress in regression lacks in predictive study with ML regression models, i.e. it is unclear that which ML regression model would be most suitable for the perceived stress prediction.

Selecting a single model for perceived stress prediction may not be enough for accurate predictive results. According to Chowdary et al. (2016) and Rosellini et al. (2018), ensemble models may perform better than single model as they are well-known for providing advantage over single models in reducing the variance and bias in learning tasks. However, not all ensemble models can improve the performance of the single models in perceived stress prediction. On top of that, no prior works looks at the use of ensemble model with the appropriate base learners regression for improving the prediction for perceived stress scale.

1.4 Research Objectives

This study primarily aimed to identify the most accurate regression model as the benchmark model and to develop a suitable ensemble model to improve the predictive performance over the benchmark model in predicting perceived stress with relevant personality traits. To achieve the main goal, this study has identified the following sub-objectives:

1. To identify the personality traits that are relevant for predicting the perceived stress scale.

2. To determine the most suitable single regression model for predicting perceived stress scale to be used as the benchmark.
3. To identify and develop a suitable ensemble regression model for improving the prediction of perceived stress using relevant personality traits.
4. To compare the prediction performances of the proposed ensemble regression models with the benchmark single model.

1.5 Research Questions

1. What are the personality traits that uniquely predict perceived stress?
2. What are the most commonly used ML regression models?
3. What are the most commonly used ensemble regression models?
4. What is the most suitable ML regression model in predicting perceived stress?
5. What is the most suitable ensemble model in predicting perceived stress?
6. Does ensemble model perform better than the benchmark ML regression model in predicting perceived stress?

1.6 Research Scope

This study focused on the perceived stress measured by Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983). The dataset was taken from Pallant's (2013), which consist of males and females, with ages ranging from 18 to 82. In this study, the single regression models and ensemble models were chosen based on their performances in other domains from the literature due to limited studies were done to use ML models in predicting regression problem of stress-related domains. Through the literature, only five commonly used regression models and six ensemble regression models were focused.

1.7 The Structure of the Study

This section describes the purposes of each chapter of this study in brief. Totally there are six chapters being arranged for this study.

Chapter 1 (the current chapter) has included the introduction, study background, research motivation, problem statement, and the objectives, questions and scope of this study.

Chapter 2 focuses on the important work related to improving the predictive performances of perceived stress predictions. Specifically, it reviews the literature about the definitions of stress, measurements of perceived stress, personality traits, ML single regression models, ensemble models, ensemble framework, and evaluation method.

Chapter 3 describes the details of the four important steps for the proposed research methodology in this study, which are literature review, data collection, model development, and evaluation.

Chapter 4 describes the experimental design of this study by explaining the implementation procedure that directly reflected from the research methodology discussed in Chapter 3.

Chapter 5 discusses the results and findings from the implementation of the experimental design of this study and the comparison of the predictive performances of all developed single models and ensemble models in predicting the perceived stress with relevant personality traits.

Chapter 6 summarizes and concludes the research objectives achieved, discusses the implications, and provides recommendations for future researchers to overcome the limitations of this study.

CHAPTER 2: LITERATURE REVIEW

The main purpose of this study is to develop an accurate system to predict stress. However, there are many definitions to stress and many methods can be used to make prediction, therefore, this chapter is going to review the existing literature thoroughly and find out the most suitable solutions to solve the problems of this study. Definitions of stress would be reviewed first to adopt a suitable stress definition for this study, because selecting the wrong definition of stress would cause the study to select the wrong dataset, wrong predictors and wrong predictive methods, in other words, it would lead the whole research heading to the wrong direction. After identifying the most suitable stress definition, next will be the review on the measurements and predictors of the defined stress. Following that, this chapter also reviews the predictive methods including the single predictive models that could be used to predict the defined stress, and the ensemble models to improve the prediction accuracies of the predictive models, as well as the evaluation methods.

2.1 Definitions of Stress

Depression is considered as one of the most widespread illness and increasing globally (World Health Organization, 2012; Ciarrochi, Deane & Anderson, 2002). Many mental health problems such as depression, hopelessness, and suicidal ideation are caused by stress (Pearlin, Menaghan, Lieberman & Mullan, 1981; Marshall et al., 2000; Ciarrochi, Deane & Anderson, 2002; Schönfeld, Brailovskaia, Bieda, Zhang & Margraf, 2016). Lazarus, Speisman, Mordkoff and Davison (1962) stated that stress is commonly known as a central problem in our life. Stress is never easy to be defined, indeed it is very complicated. Different people from different field of studies under different conditions gives it a different meaning. Classically, stress is defined at least in three different ways as in Table 2.1, which are response-based, stimulus-based, and relation-based definitions of stress. Each definitions of stress will be reviewed further in following sub-sections.

Table 2.1: Classical definitions of stress

Definition (Butler, 1993; Brüggemann & Santos, 2016)	**Tradition (Cohen, Gianaros & Manuck, 2016)**	**Description**
Stimulus-based definition	Epidemiologic tradition	The stress posed by external stimulus or individual life events (Butler, 1993; Cohen et al., 2016).
Response-based definition	Biological tradition	Stress is the nonspecific response of the body to any demand (Selye, 1936)
Relation-based definition	Psychological tradition	Stress was defined as "a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her wellbeing" (Lazarus & Folkman, 1984, p. 19)

2.1.1 Stimulus-based Definition of Stress

Epidemiologic model focuses on external sources of stress, which is the stress posed by individual life events, and suggests that stress is cumulative, where-by each additional event added to one, the amount of stress will be added to one's overall burden of adaptation (Butler, 1993; Cohen et al., 2016). According to Cohen, Kessler and Gordon (1995), Adolf Meyer began his work with the interest in the stress posed by life events in 1930s. In the late of 1940s, a substantial of research body which was highly influenced by Meyer's ideas had documented the stressful life events that associated with variety of physical illness and later filled out a life chart as part of their medical examination of the patients. In 1957, Hawkins, Davies and Holmes (1957) developed the Schedule of Recent Experiences (SRE) to systematize Meyer's life chart. The scale was used by many and found relationship between diseases and stressful life events, like heart disease and skin disease (Holmes & Masuda, 1974).

Later, Social Readjustment Rating Scale (SRRS; Holmes & Masuda, 1974), a subsequent modification of the SRE was developed and gathered 43 stressful life events, such as, divorce, marriage, pregnancy, death of spouse, being fired at work, trouble with boss, retirement, and so on. Each event was given a standardized score based on judges' normative evaluation of the rate of difficulty required to adapt to the event. Another example of method for assessing stressful life events is the Life Events and Difficulties Schedule (LEDS; Brown & Harris, 1989), which is a structured survey used to investigate the details of the related events and the ambient conditions. Any event meets or exceeds the LEDS-defined threat-severity threshold marks the presence of sufficient stress to put one at risk of disease. The empirical evidence showed single severe event is enough to predict depressive episodes or increasing the risk for a range of psychiatric and physical disorders (Brown & Harris, 1989). According to Cohen et al. (2016), single consensually determined threatening events are sufficient to generate substantial levels of threat, which last for months or even years.

2.1.2 Response-based Definition of Stress

The response-based definition of stress is commonly used in biological or physiological tradition and it is promoting stressful life events which promotes biological response that are conducive to disease, for example, immune, altered metabolic, respiratory, and cardiovascular functioning (Cohen et al., 2016). Selye (1936) proposed the response-based definition of stress as "the nonspecific response of the body to any demand". He has developed the General Adaptation Syndrome (GAS) model to describe the physiological response to stress in three stages (alarm, resistance, and exhaustion). Firstly, when the body is alerted, it will respond with alarm reactions. Next, when the body is preparing to deal with the stress, autonomic activities will be triggered. Lastly, if the stress exceeds certain level that the body can handle, the system may be destroyed or affected.

The GAS concept is similar with the flight-or-flight response that was underscored earlier by Cannon (1929), which is a physiological response of animals in the reactions towards perceived danger or harmful event. Psychological responses may follow the similar course, such as a person may cope with or adapt to the stress, but if the stress is beyond the capacity that a person can cope with, the consequences may not be known or seen externally, and one may not even realize that one is in a dangerous condition. Besides, ability to cope with stress may vary with the person's characteristics and depends on many factors which involve a complicated process.

In the past, biological research on stress in humans emphasized on laboratory studies, in which participants are exposed to experimental challenges or stressors, and then types of autonomic and neuroendocrine responses, systemic biological and cellular changes (such as altered metabolic, immune, respiratory, and cardiovascular functioning) that are conducive to disease typically assessed in such studies (Cohen et al., 2016). Later, McEwen (1998) broadened the biological view of stress in terms of dysregulated systems by equating stress with overactivation of hypothalamic–pituitary–adrenal (HPA) and sympathoadrenal medullary (SAM). Recent biological human stress research has characterized the brain systems that appraise psychological and social stressors, such as using functional magnetic resonance imaging (fMRI) to assess the activities of the brain while one completes process threatening stimuli that are modeled from laboratory-based studies of physiological stress (Muscatell & Eisenberger, 2012; Gianaros & Wager, 2015).

2.1.3 Relation-based Definition of Stress: Perceived Stress

Lazarus' (1976) cognitive theory of stress states that it is not the event that causes one stress; rather it is one's perception of the event, which is an essential factor that influences the impact the event has on one's life. In other words, it is one's appraisal of the event determines whether the event is considered stressful to oneself. In congruent to Lazarus' (1976) cognitive theory of stress, Stuber et al. (1997) found that the predictors of posttraumatic stress symptoms are mainly subjective factors (e.g. subjective appraisal and anxiety) instead of the objective stressors of medical sequelae. Besides, Salvador (2005) found that the neuroendocrine response depends more on subjective factors related to the perception of the situation rather than on the end results. Those researches seem reflecting that one's perception of stress plays an important role in psychological as well as physiological stress response.

Lazarus (2006) stated that psychological noxiousness is not easy to be specified as physiological noxiousness does; "the degree and kind of stress response, even to singularly powerful stress conditions, are apt to vary from person to person, and these variations need to be understood" (p. 54). He also mentioned that "the existence of substantial individual differences means that a stimulus alone is insufficient to define stress" (p. 54). In psychological perspective, a stressful experience cannot be inferred by uniform reference to any event, and the same event may be stressful for some people but not everyone (Cohen et al., 2016).

Lazarus and Folkman (1984) defined stress as "a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her wellbeing" (p. 19). Butler (1993) agreed with Lazarus and Folkman's (1984) definition of stress and stated that response-based or stimulus-based stress definition alone has limitations, as she emphasized that stress is a dynamic process that reflecting both external and internal factors, such as one's characteristics and circumstances, as well as the interactions between them. She proposed to understand stress from cognitive factors in psychological well-being, such as beliefs, attitudes, and thoughts. She concluded that the cognitive factors influence both the response and stimulus sides of the equation (Butler, 1993).

During the confrontation with stress, if one feels no control over the situation, one may develop sense of helplessness, which can negatively affect one's motivation to cope the stress (Lazarus and Folkman, 1984). Before the actual confrontation with stress, there will be a period of stress anticipation; research found that anticipation of a threat produces more harmful effects than the actual confrontation with the stressors, and long anticipation is more stressful than short anticipation (Lazarus, 1966; Nomikos, Opton Jr & Averill, 1968; Feldman, Cohen, Hamrick & Lepore, 2004). Feldman et al. (2004) suggested that the stress process may be best studied during a period of stress anticipation. The stress anticipation period is a crucial time which determines whether one will continue to the stress confrontation. Nomikos et al. (1968) found that most of the stress reaction occurred during the periods of stress anticipation, rather than during the actual stress confrontation. All these findings were reflecting that one's perception of stress plays an important role to determine how stressful one is, rather than the stimulus or response of stress determines how stressful one is.

Lazarus and Folkman (1984) introduced a new term "perceived stress" and defined it as "the thoughts or feelings that one has about how much stress one has perceived within a period or at a specific point of time". It incorporates feelings about one's confidence to handle the unpredictability and uncontrollability of one's life and how often one must struggle with the problems. In other words, it is assessing how one perceived of one's stressfulness and one's capacity to manage it (Michalos, 2014).

Herbert and Cohen (1996) mentioned that "individuals are the best source for information on appraisal, since only they have the necessary awareness of their motives, commitments, and concerns that give meaning to the situation" (p. 318). Monroe and Kelley (1997) also stated that it is only from within the perceived world of an individual that the true meaning of the event that the individual experiences can be understood, therefore, this is where the subjective measures of appraisal should base on. Therefore, it is a need to have a good measure of the perceived stress in term of interview or self-administrated questionnaire which allows individual to provide information about his/her current perception of stress. Following section will review the measures of perceived stress.

2.2 Perceived Stress and Perceived Stress Scale (PSS)

In the context of this research, based on the findings from previous section, perceived stress was being selected as the predictive output and domain of this study. This section will review the measurements developed to measure perceived stress and to identify the most suitable measurement for perceived stress, because the dataset that would be taken in this study must consist the construct of perceived stress that was measured by the same measurement.

Researchers measured perceived stress in specific domain using related perceived stress measure, for example, perceived job stress was measured with Karasek Job Control Questionnaire (JCQ; Karasek et al., 1998) and individuals' appraisals of the negative impact associated to specific social roles like work, marriage, or parenthood (Lepore, 1995). Besides event-dependent measures, global (event-independent) perceived stress measures were developed to measure perceived stress in a wide range of domains. An adaptation of the JCQ has been used to measure non-job-related stress (Kamarck, Muldoon, Shiffman & Sutton-Tyrrell, 2007). The most widely used global perceived stress measure is Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983), which is a self-administrated questionnaire to measure “the degree to which situations in one’s life are appraised as stressful” (p. 385).

“In all comparisons, the PSS was a better predictor of the outcome in question than were life-event scores” (Cohen et al., 1983; p. 385). Compare with using objective tests in measuring the number of significant life events occurred within a specific timeframe, PSS was found to be a better predictor of health outcomes (Cohen et al., 1983). Cohen et al. (1983) found correlations between perceived stress and physical symptomology, as well as behavioral and psychological outcomes. Hence, Cohen et al. (1983) stated that PSS “can be used as an outcome variable, measuring people’s experienced levels of stress as a function objective stressful events, coping resources, personality factors, etc.” (p. 393) and “provides a potential tool for examining issues about the role of appraised stress levels in etiology of disease and behavioral disorders” (p. 394).

Besides, Cohen and colleagues (Cohen et al., 1983; Cohen & Williamson, 1988; Cohen, Tyrrell & Smith, 1993) had successfully used prospective designs and controlling for other possible predictors of psychological outcomes to address the confounding appraisal issue as measured by the PSS with antecedents and psychological outcomes. They have demonstrated that the scores on PSS could predict various outcomes without depending on the measures of psychological and physical symptoms assessed at baseline (Herbert & Cohen, 1996). More than thirty years after the PSS is developed, it is still globally used and top cited (currently cited by 13,642 resources from Google Scholar) assessment of one’s perception of stress and the stress related health outcomes (Morgan, 2014; Garber, 2017; Dobkin, Zhao & Monshat, 2017).

2.3 Personality Traits and Perceived Stress

Lazarus (2006) mentioned that it is a need to understand human variation if investigators want to understand or deal effectively with ones, because the stimulus and response of stress may be different for different persons. “To have rule-based definition, we must identify the characteristic that make some people vulnerable to the stimulus as a stressor, and others not vulnerable, or less so” (Lazarus, 2006; p. 54). Solid evidence of individual differences in response found in research results showing that, with the same threat of failure, some experimental subjects did much better while others did much worse; it was as if the stress condition pushing some of them upward while pushing other downward (Lazarus & Eriksen, 1952; Lazarus, Deese & Osler, 1952). As a result of the research, Lazarus (2006) stated that “it became increasingly clear that reactions under stress cannot be predicted without references to personality traits and processes that account for the individual differences in the ways people respond to a so-called stressful stimulus” (p. 55).

In the early 80’s, Pearlin et al. (1981) found personality traits (mastery and self-esteem) and social support act as mediators and moderators of the relationship between exposure to stressors and depression. Later, Lazarus (2006) also listed some of the personality traits which were found to help people resist deleterious effects of stress, such as optimism, ability to think constructively, hope, hardiness, learned resourcefulness, self-efficacy, and sense of coherence. In a longitudinal research, Schaefer et al. (2017) found that most people will experience a diagnosable mental disorder and only minority who have an advantageous personality traits during childhood, and negligible family history of mental disorder experience enduring mental health. If experiencing diagnosable mental disorder is the norm, and only people with advantageous personality traits during children could avoid such conditions and endure extraordinary mental health, then truly those personality traits are very important and worth to be further studied (Schaefer et al., 2017). The advantageous personality traits that were found in Schaefer et al.'s (2017) study are: little evidence of strong negative emotions in childhood, significantly less socially isolated in childhood, significantly higher levels of childhood self-control, and having fewer relatives with mental health issues. Again, it is getting clearer that personal traits could be important predictors of stress which play important role in determining whether one could resist stress.

In previous sections, different stress definitions were discussed, whereby response-based or stimulus-based stress definition alone has limitations while perceived stress (relation-based definition of stress) is found better to explain the concept of stress. In concurring research, a number of stress predictors or personality traits were found significantly correlated with the perceived stress that was measured with PSS, the most suitable and widely used measure of global perceived stress (refer to previous section), such as mastery (Pearlin et al., 1981; Pallant & Lae, 2002), perceived control of internal states (PCOIS; Bretherton & McLean, 2015), self-esteem (Pearlin et al., 1981; Robins, Hendin & Trzesniewski, 2001; Pallant & Lae, 2002), life satisfaction (Chang, 1998; Rey & Extremera, 2015; Tang & Chan, 2017), optimism (Scheier & Carver, 1985; Chang, 1998; Pallant & Lae, 2002), negative affect (Ezzati et al., 2014; Robles et al., 2016; Schaefer et al., 2017), and positive affect (Curtis, Groarke, Coughlan & Gsel, 2004; Ezzati et al., 2014).

Some researchers have concluded that gender is a significant predictor of perceived stress (Ezzati et al., 2014; Robles et al., 2016; Nwoke, Onuigbo & Odo, 2017). The results seem to suggest that male and female respondents will respond differently to stressors and stressful situations. Nwoke et al. (2017) mentioned that the possible explanation for females reporting more stress than males is that females are easily emotional and can be more emotionally upset than males in stressful situations. Other than that, smoking behavior is one of the predictors of perceived stress as well. Smokers are found to have higher perceived stress than ex-smokers and nonsmokers (Cohen & Lichtenstein, 1990; Ng & Jeffery, 2003). Table 2.2 shows the predictors of perceived stress with their findings in brief. Identifying the potential predictors from different research will allow more relevant predictors to be added to the predictive model so that the prediction accuracy can be enhanced. The following sections will review the research gaps or limitations of the predictive research that have been done to predict perceived stress.

Table 2.2: Predictors of perceived stress

Authors	**Predictors of Perceived Stress**	**Findings**
Pearlin et al. (1981); Pallant and Lae (2002)	Mastery	Higher scores on Mastery Scale were associated with lower scores on the Perceived Stress Scale.
Bretherton and McLean (2015)	Perceived control of internal states (PCOIS)	Perceived control of internal states was significantly negatively related to perceived stress.
Pearlin et al. (1981); Robins, Hendin and Trzesniewski (2001); Pallant and Lae (2002)	Self-esteem	Self-esteem negatively associated with perceived stress.
Chang (1998); Rey and Extremera (2015); Tang and Chan (2017)	Life satisfaction	Life satisfaction negatively associated with perceived stress.
Scheier and Carver (1985); Chang (1998); Pallant and Lae (2002)	Optimism	Optimism negatively associated with perceived stress.
Ezzati et al. (2014); Robles et al. (2016); Schaefer et al. (2017)	Negative affect	Negative affect positively associated with perceived stress.
Curtis, Groarke, Coughlan and Gsel (2004); Ezzati et al. (2014)	Positive affect	Positive affect negatively associated with perceived stress.
Ezzati et al. (2014); Robles et al. (2016); Nwoke, Onuigbo and Odo (2017)	Gender	Females reporting more perceived stress than males.
Cohen and Lichtenstein (1990); Ng and Jeffery (2003)	Smoking behaviour	Smokers have higher perceived stress than ex-smokers and non-smokers.

2.4 Predicting Perceived Stress with Relevant Personality Traits

Establishing the predictors of perceived stress helps to reflect how the stress perception originates and motivates the interventions to resist stress (Lebois, Hertzog, Slavich, Barrett & Barsalou, 2016). Since a wide range of personality traits were found associated to perceived stress, they could be the potential predictors of perceived stress and form a good model with high predictive performance. Several studies were conducted to predict perceived stress; the predictors, measure of perceived stress, predictive model, and research gaps (or limitations) of those studies are shown in Table 2.3.

Majority of the studies focused on special population like people with epilepsy (Moon, Seo & Park, 2016), people in emerging adulthood who had exposed to violence during adolescence (Heinze, Stoddard, Aiyer, Eisman & Zimmerman, 2017), caregivers of children with learning disabilities (Isa et al., 2017) and medical undergraduates (Shah, Hasan, Malik & Sreeramareddy, 2010) which may perceive higher stress than undergraduates from other courses. Moon et al. (2016) mentioned that one of the limitations of their study is their samples were taken from a tertiary care hospital, and the predictors of perceived stress may differ from the people with epilepsy. Pearlin (1999) mentioned that "social stress is not about unusual people doing unusual things and having unusual experiences" (p. 396). Rather, stress theories focus on how ordinary people deal with difficulties in the society (Aneshensel & Avison, 2015). In addition, people in disadvantage situations will not only suffer from a proliferation of stressors but also from a relative lack of multiple protective factors (Pearlin, 1999). Therefore, if the study is to find out the general predictive personality traits of the global perceived stress, then it must recruit the samples from the general population to generate more applicable results, because population in special situation may have different predictors which are not applicable for the general or other populations.

According to the findings from the research in Table 2.3, the predictors of the research varies, such as mental health disorders (Moon et al., 2016), school related factors (Heinze et al., 2017), groups of stressors (Shah et al., 2010), anger regulation strategies (Yamaguchi, Kim, Oshio & Akutsu, 2017), minorities status stress (Greer, 2008), and coping styles (Isa et al., 2017). Lebois et al. (2016) claimed that their exploratory study was "the first to provide a comprehensive assessment of the features that predict perceived stress, we assessed a non-clinical sample in the laboratory", however, their sample size is just as small as 12 participants and their measure of perceived stress was not the globally used PSS, but a single question, "If you were actually in this scenario, how much stress would you experience? (1-7 scale: 1 = low, 4 = medium, 7 = high)" and the participants were to answer the same question after reading each stressful event scenario provided. Due to the small sample size and the newly developed measure of perceived stress, the study needs a bigger sample size and to verify the reliability of the newly used measure of perceived stress.

In conclusion, there is a need to conduct the research to predict the perceived stress of the general population using a comprehensive list of personality traits as predictors. The next section discusses the focus of the current study, which is the prediction of perceived stress using ML regression models.

2.5 Predicting Perceived Stress using Machine Learning Regression Models

Machine Learning (ML) is well known in predictive analytics to automatically mine and detect patterns, make intelligent decisions based on data, and build predictive models without being explicitly programmed (Kamber, 2011; Kitchin, 2014). ML offers a large body of models which generally categorized under several techniques, such as classification, clustering, regression, simulation, content analysis and recommenders (Fontama et al., 2015). Among these techniques, classification and regression are commonly used for predictive modeling (Fontama et al., 2015). Classification models are used to predict categorical or ordinal value; while regression models are used to predict continuous (numerical) output (or response variable) but the input variables can be numeric or categorical.

Apparently, all the predictive research of perceived stress found in Table 2.3 was using the commonly used Multiple Linear Regression (MLR) model, which is one of the regression models. Generally, PSS is designed to access “the degree to which situations in one’s life are appraised as stressful” (Cohen et al., 1983; p. 385), which its outcome in nature is numerical instead of ordinal value. As Nuñez-Gonzalez and Graña (2015) stated in their proposed experiment to predict the ratings given by the users in social networks, “because of the range of the ratings we cannot assume that all failures are the same, in other words, if we have to predict a rating of ‘2’ marks, making a prediction of ‘3’ marks is a smaller error than making a prediction of ‘5’ marks” (p. 66). Therefore, predicting perceived stress is a regression problem rather than a classification problem and that is the reason all the predictive research of perceived stress found in Table 2.3 were using MLR (regression model) as their predictive model.

ML has provided many regression models, and each has its advantages and disadvantages. As no single model works best for every problem (especially for predictive modeling) and the size and structure of the dataset may vary the selection of the suitable models, (Elite Data Science, 2017, September 16), therefore different models must be tried for the same problem to evaluate the performance of the models to select the model that could outperform others in the same problem. All the predictive research of perceived stress found in Table 2.3 adopted Multiple Linear Regression (MLR) directly and did not compare the predictive performance of MLR with other regression models have left the performance of other regression models unknown and restricted the predictive model to perform better. Most of the stress-related predictive research were using ML classification models to predict categorical outcomes (Scherer et Al., 2008; Plarre et al., 2011; Sharma & Gedeon, 2012; Bogomolov et al., 2014a, 2014b; Smets et al, 2015; Chowdary et al., 2016; Subhani et al., 2017; Rosellini et al., 2018), but very little research exists on using ML regression models to predict stress-related numerical outcome like perceived stress.

Since very limited literature that compares between different regression models in predicting stress-related outcomes, therefore the potential regression models must be identified through the literature that conducted comparison study between different regression models in predicting other regression problems. Table 2.4 shows the ML regression models which were found highly predictive and out-performed other regression models in predicting none-stress-related regression domains, such as Support Vector Regression (SVM), Multilayer Perceptron (MLP), Multiple Linear Regression (MLR), Elastic Net (ELN), Gaussian Process Regression (GPR) and Random Forest (RF). Table 2.5 shows the brief description of those regression models. Those single models would be compared to predict the perceived stress in current study based on their predictive performances. However, RF would be explained in next section because it is a type of ensemble model, though it can be also taken as a single model as it is not required to be built with base learner.

Table 2.3: Predictive analysis for perceived stress

Authors	Predictors	Perceived Stress Measure	Population	Predictive Model	Remarks
Moon, Seo and Park (2016)	Neurological disorders depression, sleep-related impairment, generalized anxiety disorder, seizure control	PSS	People with epilepsy	MLR	Clinical sample.
Heinze, Stoddard, Aiyer, Eisman and Zimmerman (2017)	Gender, age, highest parent occupational prestige score, race, school, depression, violent behaviour, school relations, school attitudes	PSS	People in emerging adulthood who had exposed to violence during adolescence	MLR	Sample who exposed to violence during adolescence.
Shah, Hasan, Malik and Sreeramareddy (2010)	Demographic variables and groups of stressors (i.e. academic, psychosocial, and health-related)	PSS	Medical undergraduates in a Pakistani Medical School	MLR	Only gender was significant ($p < 0.05$) with PSS score. Predictors: a group of stressors instead of personality traits.
Yamaguchi, Kim, Oshio and Akutsu (2017)	Anger-in, anger-out, anger-control	PSS	American and Japanese adults	MLR	Predictors: anger regulation strategies instead of personality traits.
Lebois, Hertzog, Slavich, Barrett and Barsalou (2016)	Expectation violation, self-threat, coping efficacy, bodily experience, arousal, negative valence, positive valence, perseveration	1-item of perceived stress question	12 university students	MLR	Exploratory study designed to provide a comprehensive assessment of the features that predict perceived stress in a non-clinical sample in the laboratory, participants answered the same single question (If you were actually in this scenario, how much stress would you experience? [1-7 scale: 1 = low, 4 = medium, 7 = high]) instead of using PSS after reading each scenarios of stressful events.
Greer (2008)	Gender, age, SAT scores, minorities status stress	PSS	African American students at a historically Black college and university	MLR	Predictors: racial and ethnic-related stressors instead of personality traits.
Isa et al. (2017)	Coping styles (use of instrumental and emotional support, behavioral disengagement, religion), number of children under a caregiver	PSS	Malay caregivers of children with learning disabilities in Kelantan	MLR	Predictors: coping styles instead of personality traits.

Table 2.4: Best regression models used in predicting none-stress-related domains

Authors	**None-stress-related Regression Domain(s)**	**Regression Models that were being Compared**	**Best Regression Model(s)**
Graczyk et al. (2009)	Market value of properties	MLP, RBF, M5P, M5R, MLR, SVM	SVM
Liu et al. (2009)	Cost of product life cycle	IBK, LWR, M5P, MLP, SVM	SVM and MLP
Graczyk et al. (2010)	29 benchmark regression datasets	LRM, SVM, M5P, MLP, RBF, RBI, RBD, IRP, SON	SVM
Trawiński et al. (2012)	29 benchmark regression datasets	MLP, RBF, RBI, RBD, iRP, SON	MLP
Mendes-Moreira et al. (2012)	Long term travel time	PPR, SVM, RF	RF
Forkuor et al. (2017)	Soil properties	MLR, RF, SVM, SGB	RF and MLR
Khair et al. (2017)	Stream flow	MLR, MLP, SVM	SVM
Domingo et al. (2017)	Biomass losses and CO2 emissions	MLR, SVM, RF, LWR, MDL, RLM, KNN, WKNN	MLR
Estelles-Lopez et al. (2017)	Meat spoilage	OLS, SLR, PLS, PCR, SVM, KNN	RF
Lichtenberg and Şimşek (2017)	60 publicly available datasets from varying domains	ELN, RF, OLS	ELN
Keshtegar (2018)	Solar radiation	Kriging or GPR, RSM, MAR, M5P	GPR

ELN = Elastic Net; IBK = Instance-Based K-Nearest Neighbours; IRP = Multilayer Perceptrons trained with iRProp+; KNN = K-Nearest Neighbours; GPR = Gaussian Process Regression; LTA = N-Latest Transactions in an Area; LWR = Locally Weighted Regression; M5P = M5 Model Tree; M5R = M5 Rules; MAR = Multivariate Adaptive Regression; MDL = Linear Model with a minimum length principle; MLP = Multilayer Perceptron; MLR = Multiple Linear Regression; NSP = N-Nearest Similar Properties; OLS = Ordinary Least Squares Regression; PCR = Principal Component Regression; PLS = Partial Least Square Regression; PPR = Projection Pursuit Regression; RBD = Decremental Radial Basis Function Neural Network; RBF = Radial Basis Function Neural Network; RBI = Incremental Radial Basis Function Neural Network; RF = Random Forest; RSM = Response Surface Method; SGB = Stochastic Gradient Boosting; SLR = Stepwise Linear Regression; SON = Self Organizing Modular Neural Network; SVM = Support Vector Regression; WKNN = Weighted K-Nearest Neighbours;

Table 2.5: Brief description of the identified single regression models

Regression Models	Abbreviation	Description
Multiple Linear Regression	MLR	Standard statistical model used to build linear model predicting a value of the outcome while knowing the values of the other variables. It uses the least mean square method to adjust the parameters of the linear model (Bańczyk, Kempa, Lasota & Trawiński, 2011).
Multilayer Perceptron	MLP	Artificial neural networks that consist of multiple layers and usually interconnected in a feed-forward way, where each neuron on the layer has directed the connections to the neurons of the subsequent layer (Bańczyk et al., 2011).
Support Vector Machine for Regression	SVM (or SMOreg)	The SVM which performs linear regression in the high-dimension feature space using insensitive loss, and, at the same time, tries to reduce model complexity.
Elastic Net	ELN	Elastic Net is a regularized regression method that linearly combines the limitations of the LASSO (least absolute shrinkage and selection operator) and ridge methods.
Gaussian Process Regression (a.k.a Kriging)	GPR	It is nonparametric kernel-based probabilistic model which implements Gaussian processes for regression purposes.

2.6 Ensemble Models

Ensemble models are well-known for providing advantage over single models in reducing the variance and bias in learning tasks. Besides, Moniz, Branco and Torgo (2017) also found that smaller datasets were prone to larger improvements in predictions using ensemble models. According to Al-Abri (2016), as illustrated in Figure 2.1, the ensemble models are categorized into two main categories, which are homogeneous (using the same base learner on different distributions) and heterogeneous (using multiple base learners) ensembles. There are three types of homogeneous ensembles (Bagging, Randomization, and Boosting) and two types of heterogeneous ensembles (Voting and Stacking). The commonly used randomization models are Random Forest (Mendes-Moreira et al., 2012; Forkuor et al., 2017; Estelles-Lopez et al., 2017) and Random Subspace (Dapeng, 2017; Pham, Prakash & Bui, 2017; Suganya, & Ebenezer, 2017), and the commonly used

Boosting for regression task is Additive Regression (Pérez et al. 2017; Burke, 2017; Liu, Shang & Cheng, 2017). Sub sections below briefly describe the commonly used ensemble models.

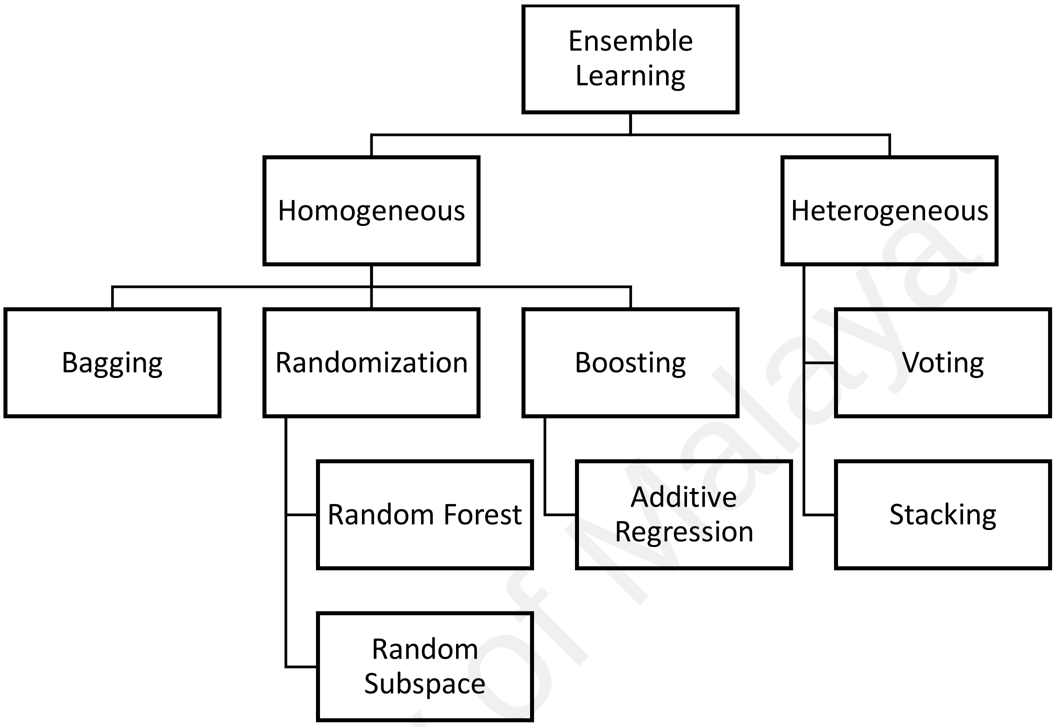


Figure 2.1: Ensemble Learning Hierarchy (Al-Abri, 2016)

2.6.1 Bagging (BG)

Bagging also named as Bootstrap Aggregation because it is the application of bootstrapping and aggregating concepts to reduce the variance for the models that have high variance. It operates by taking a base learning model and invoking it multiple times with different training sets, and then integrates the outputs of different models into a single prediction model using either weighted or average vote (Breiman, 1996).

2.6.2 Random Forest (RF)

Random Forest is an extension of Bagging that specifically designed for decision tree classifiers. It constructs bunch of decision trees and outputs the mean prediction of the individual trees. It is different to Bagging in the way that it splits the node of a tree and randomly picks the sub-features that it searches for instead of looking for the best point to split the node (Breiman, 2001). Besides, it does not need to base on another single learning model to build the ensembles.

2.6.3 Random Subspace (RSS)

Random Subspace is like Bagging as it is also called Feature Bagging. It is different to Bagging in the way that the features or predictors are randomly sampled with replacement for each learner. It tries to ensure that individual learners not to over-focus on features that are highly predictive only at certain training sets (Ho, 1998).

2.6.4 Additive Regression (AR)

Additive Regression is designed to enhance the performance of a regression base learning model by iterating the models. In each iteration, a model will be created to fit the residuals left by the model from the previous iteration, and the final prediction is done by adding up the predictions of each model (Friedman, 2002).

2.6.5 Voting

Voting is like Bagging except that it builds the final model by averaging the outputs from the models produced by different base learning models (Major & Ragsdale, 2001).

2.6.6 Stacking

Stacking combines multiple models via meta learner. In first level training, the base level models are trained using original dataset. In second level training, the meta learner is trained using the outputs from the base level model as features. After that, the predictions from the second level training would be used as the inputs to train a higher-level learner (Wolpert, 1992).

2.7 State of Art of Ensemble Models

King, Abrahams and Ragsdale (2014) have proposed their ensemble framework which using three single models (MLR, Artificial Neural Networks or ANN, Regression Trees or CART) with four ensemble models (BG, RSS, Voting, and Stacking) which were mentioned in the current study, producing ten ensemble implementations as shown in Figure 2.2 for the advanced skier days prediction.

Specifically, there were three BG instances, three RSS instances, three Stacking instances, and one Voting instance. The BG and RSS ensemble models both required an instance to be created from one of the single models. However, Stacking and Voting ensembles were created differently. Three stacking instances were created with all three single models simultaneously as base learners and one of these single models as the meta learner, in turn. The one Voting instance was created with all three single models simultaneously as base learners, but without meta learner. Each ensemble instance was cycled ten times and their predictive performance was calculated for the model comparison.

Among the three single models, MLR was the best performer. However, among the ten ensemble implementations, nine instances achieved improvements over the prediction performance of the MLR alone, except the BG-MLR instance. The best ensemble implementation was Stacking with all models as base learners with ANN as meta learner.

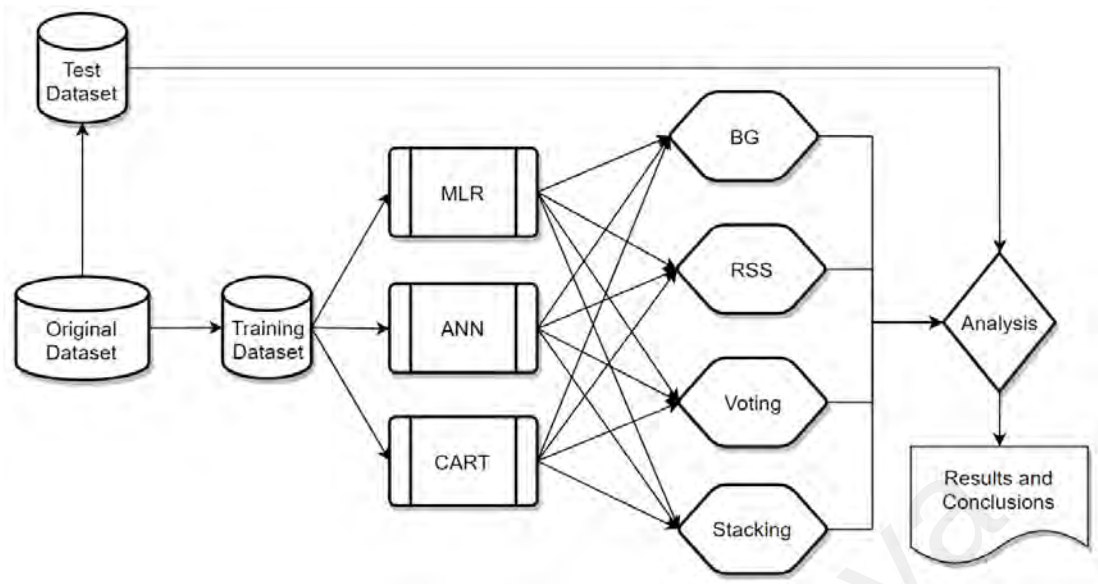


Figure 2.2: Ensemble Framework (King et al., 2014)

2.8 Predicting Perceived Stress using Ensemble Regression Models

Ensembles have been studied for stress-related classification tasks, for example: Plarre et al., (2011) found that prediction of psychological stress using J48 Decision Tree with Adaboost (ensemble model) gaining higher predictive accuracy than using single J48 classifier; Chowdary et al. (2016) proved that Pegasos (a modified model of stochastic gradient) combined with Adaboost ensemble achieved good results in detecting the stress suffered by IT professionals; Rosellini et al. (2018) showed that the super learner model (an ensemble model suited to develop risk scores) achieved a better cross-validated performance than 39 individual models in predicting posttraumatic stress disorder (PTSD). Besides, ensembles have been studied for regression tasks too, such as rainfall forecasting (Wu & Chen, 2009), wind and solar power forecasting (Ren, Suganthan & Srikanth, 2015), financial domains (Jiang, Lan, & Wu, 2017), and imbalanced regression tasks (Moniz et al., 2017). However, regression ensemble studies that related to stress are very limited and need more exploration.

2.9 Chapter Summary

After reviewing the related works, it was found that perceived stress is the most suitable definition for explaining psychological stress. Besides, PSS was found as the most suitable measurement developed to measure perceived stress. The predictors that were found related to perceived stress are personality traits (mastery, PCOIS, self-esteem, life satisfaction, optimism, negative affect, positive affect), gender, and smoking behavior. The commonly used single models for predicting regression problem are MLR, SVM, ELN, RF, GPR, and MLP. Furthermore, the commonly used ensemble regression models to improve the predictive accuracies of the single models are BG, RSS, AR, Voting, and Stacking. Through the literature review, basically, the targeted outcome, its predictors, and the predictive models were already identified. However, there is a need to identify the benchmark single model as well as the ensemble model for perceived stress prediction, so that an enhanced model can be developed to improve the predictive performance. Next Chapter will present the research methodology about how to predict the perceived stress with the personality traits using the single models and ensemble models that were identified from the literature and how to evaluate the predictive models.

CHAPTER 3: RESEARCH METHODOLOGY

There are four important phases adopted for the research methodology in this study as depicted in Figure 3.1. Firstly, literature review was carried out in Chapter 2 to identify the definitions of stress, measure of perceived stress, predictors of perceived stress (personality traits), suitable single regression models, and ensemble models for perceived stress prediction. Second step is the data collection stage, where the suitable dataset that consists of the scores of relevant personality traits and perceived stress would be selected and preprocessed. Besides, analysis would be done to test the reliability of the measures for all the constructs. Next, during model development, the relevant personality traits would be selected through the attribute selection process to identify the predictors of the perceived stress. The selected predictors would be used for the development of single models and ensemble models. Finally, the predictive methods would be evaluated using 10-fold cross-validation and the predictive performances of all predictive methods would be compared. Chapter 4 will discuss the implementation procedures of the current chapter.

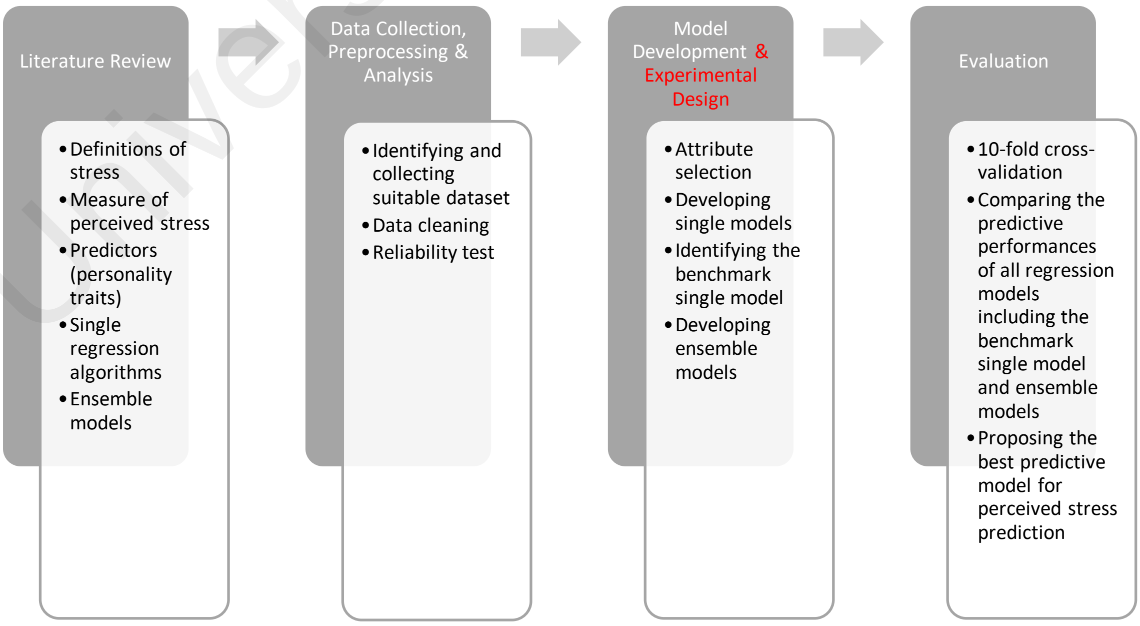


Figure 3.1: The proposed research methodology

3.1 Dataset Collection, Preprocessing and Analysis

This study is to understand the perceived stress of the general public instead of some special populations like clinical patients, people who were having some bad experiences or people who are under very stressful working or family environments. Hence, the special populations may not be applicable to the global perceived stress because they may have different characteristics which are more vulnerable or different from the general population. Therefore, the targeted population had to be the general public.

The collected dataset would have to be preprocessed first before it can be used (refer to Chapter 4 for the details). Besides, the measurements (scales) that were used to measure the related constructs in the dataset would have to be validated through reliability analysis, and the measurements of the constructs that do not achieve Cronbach's Alpha values above or equal to 0.70, would have to be removed from the dataset. Next chapter discusses the details of the implementation part of data preparation, measurements and reliability analysis.

3.2 Model Development

Before developing the predictive models, the suitable attributes would be selected first. There are two phases for model development in this study, first phase was to develop single models and to identify the benchmark single model. Second phase was to develop the ensemble models.

3.2.1 Attribute Selection

Attribute selection is the process to identify the relevant predictors (or attributes) and to remove the redundant and irrelevant predictors from the training dataset. Selecting lesser and relevant predictors could help predictive models to perform faster and more effectively (Karagiannopoulos et al., 2007). This study would use M5 attribute selection method that proposed by Yang, Tang and Yao (2012) to select the relevant predictors of perceived stress.

M5 attribute selection method uses Akaike information criterion (AIC; Hall et al., 2009) to select attributes for linear regression. AIC is calculated using the residual sums of squares from the regression below (Akaike, 1974):

$$\mathrm{AIC} = \mathrm{n}{*}\ln(\mathrm{RSS}/\mathrm{n}) + 2{*}\mathrm{K} \qquad (1)$$

where n is the number of samples, RSS is the residual sums of squares, and K is the number of parameters in the model. AIC compares the models with different permutation of attributes and selects the model which gives the lowest loss of information (Deshpande, 2012, December 11).

M5 attribute selection method iterates through the attributes and removes the attribute with the smallest standardized coefficient until no improvement is observed in the estimate of the error given by the AIC (Hall et al., 2009). In other words, M5 generates decision trees with linear regression function in their notes, and in each iteration, AIC uses the best first search strategy and separate-and-conquer method to build a M5 model tree and creating the best leaf into a rule (Remeseiro et al., 2018). After this, the selected predictors would be used to build the predictive models in the following sections.

3.2.2 Single Model Development and Experiments

In this phase, the six identified single regression models, MLR, MLP, SVM, ELN, GPR, and RF (RF is an ensemble model, but it can be categorized as a single model because it does not require any base learner or meta learner, and it could be used as base learner or meta learner of other ensemble models) would be built with the selected predictors using 10-fold cross-validation. The predictive performances of the single models would be compared and ranked, and the benchmark single model would be identified.

3.2.3 Ensemble Model Development and Experiments

As the ensemble framework proposed by King et al. (2014) and discussed in section 2.7, the BG, RSS, and AR ensemble models require an instance to be built from one of the single models, in turn. Next, each Stacking and Voting instances required to be trained with one or more than one single models simultaneously as base learners (details of the arrangements for the combinations of the base learners will be explained in Chapter 4). However, each Stacking instance also requires one of the single models as its meta learner, in turn.

3.3 Evaluation

Cross-validation was commonly used in studies to perform model evaluation (Lichtenberg & Şimşek, 2017; Forkuor et al., 2017; Graczyk et al., 2010), especially 10-fold cross-validation, is a well-known strategy to avoid over-fitting (Fontama et al., 2015). The 10-fold cross-validation splits the original dataset into ten samples for model training and model testing, and to evaluate how accurate a predictive model will perform. Besides, if there is not enough data to be separated to training and testing sets, cross-validation can still split the data accordingly without losing the testing capability. 10-fold cross-validation would be used in the development of each single model and ensemble model of this study to perform model evaluation and to avoid model over-fitting.

In ML predictive analysis, performance measures are used to measure the prediction accuracies of the ML models. The commonly used performance measures in regression problems are Mean Absolute Error (MAE; Keshtegar, Mert & Kisi, 2018; Khair et al., 2017; Liu et al., 2009) and Root Mean Squared Error (RMSE; Keshtegar et al., 2018; Khair et al., 2017; Liu et al., 2009). Besides MAE and RMSE, there are many other performance measures that were used in different studies, and each has its own advantages and disadvantages. Furthermore, certain ML models may only perform well in term of some performance measures but not all. In this study, many models would be trained and tested, if multiple performance measures were used, it may create confusion, for example, Model A performed better than Model B in term of MAE, but Model B performed better than Model A in term of RMSE. If more than hundreds of models must be compared with and different models performed well in different performance measures, it may not be able to make a conclusion that which is the best ML method.

Therefore, to make the comparisons easier, it is important to choose the most suitable performance measure. According to Brassington (2017), whose research was to find out among the commonly used MAE and RMSE, which one is the better measure for assessing model performance, and the researcher found that MAE was preferable because it was shown to be an unbiased estimator and had a lower sample variance compared to RMSE. Besides, Willmott and Matsuura (2005) indicated that MAE is more natural in measuring average error, while RMSE is unambiguous because RMSE is a function of three characteristics of a set of errors, rather than one average error, which misinterpreted measure of average error. As a result, MAE is preferable to be used as the performance measure in this study. Below is the formula of MAE (Graczyk et al., 2009), where y_i is actual value, $\dot{y}_i$ is predicted value, and *N* is number of cases in the testing set.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \dot{y}_i| \qquad (2)$$

The predictive performances of the single models would be compared in term of MAE to identify the benchmark single model and to rank the single models according to their predictive performances. After that, the predictive performances (in MAE) of the ensemble models would be compared with each other based on the improvements over the MAE of the benchmark single model to identify the best ensemble model.

CHAPTER 4: MODEL DEVELOPMENT AND EXPERIMENTAL DESIGN

This chapter explains the implementation procedures and experimental design that directly reflect from the research methodology discussed in Chapter 3. Therefore, the structure of this chapter is similar to Chapter 3, but the focus is on implementation part.

4.1 Dataset Collection, Preprocessing and Analysis

Currently there is no suitable dataset in Malaysia that consists the related attributes and it is expensive to develop own one as it takes time and resources to collect the data as well as to validate the data. Therefore, the dataset of this study was taken from a real data file (http://spss5.allenandunwin.com.s3-website-ap-southeast-2.amazonaws.com /data_ files.html) collected by Pallant's (2013) students which their study was designed to explore the factors that related to psychological adjustment and wellbeing. Their dataset was being adopted because their survey contains a variety of validated scales that measure the personality traits as well as perceived stress. The targeted population of the dataset was the members of the general public in Melbourne, Australia and its surrounding districts. The content of the dataset will be explained in the following sub-sections.

4.1.1 Demographic Attributes

Table 4.1 shows the demographic attributes of the samples in the dataset, which could be important predictors for perceived stress, for example, gender, age, highest education level, whether the respondent staying with children, does the respondent smoke and number of cigarrettes smoked per week.

Table 4.1: Demographic variables and coding instructions

Demographic Variable	Name in Dataset	Coding Instructions
Gender	sex	1 = male; 2 = female
Age	age	in years
Do you have any children currently living at home with you?	child	1 = yes; 2 = no
Highest level of education	educ	1 = primary; 2 = some secondary; 3 = completed high school; 4 = some additional training; 5 = completed undergraduate; 6 = completed postgraduate
Do you smoke?	smoke	1 = yes; 2 = no
Cigarettes smoked per week	smokenum	Number of cigarettes smoked per week

4.1.2 Measurements

Table 4.2 shows the measurements that were used to measure the related constructs (perceived stress and the identified personality traits) in this study. The number of items per measurement ranges from 5 to 18 items. Majority of the measurements are using 5-point scale from 1 (strongly disagree) to 5 (strongly agree), only 2 measurements are using 4-point scale and 1 measurement is using 7-point scale.

Table 4.2: Measurements and coding instructions

Construct	Name in Dataset	Measurement	No. of Items	Coding Instructions
Perceived stress	Tpstress	Perceived Stress Scale (Cohen et al., 1983)	10	1 = never, 5 = very often
Optimism	Toptim	Life Orientation Test (Scheier, Carver & Bridges, 1994)	6	1 = strongly disagree, 5 = strongly agree
Mastery	Tmast	Mastery Scale (Pearlin & Schooler, 1978)	7	1 = strongly disagree, 4 = strongly agree

Positive affect	Tposaff	Positive and Negative Affect Scale (Watson, Clark & Tellegen, 1988)	10	1 = very slightly, 5 = extremely
Negative affect	Tnegaff	Positive and Negative Affect Scale (Watson et al., 1988)	10	1 = very slightly, 5 = extremely
Life satisfaction	Tlifesat	Satisfaction with Life Scale (Diener, Emmons, Larson & Griffin, 1985)	5	1 = strongly disagree, 7 = strongly agree
Self-esteem	Tslfest	Self-esteem Scale (Rosenberg, 1965)	10	1 = strongly disagree, 4 = strongly agree
Perceived control of internal states	Tpcoiss	Perceived Control of Internal States Scale (PCOISS; Pallant, 2000)	18	1 = strongly disagree, 5 = strongly agree

4.1.3 Data Cleaning

Originally, the dataset was in SAV format, and it was being converted to CSV format using SPSS software. The names of the attributes (in short form) are shown in Table 4.1 and Table 4.2. Some records from the attributes of 'child', 'smoke' and 'smokenum' contained missing data, and those missing data were replaced manually with the related values, for example, 'child' (replaced with 2, assumed that the respondent did not stay with any child), 'smoke' (replaced with 2, assumed that the respondent did not smoke), and 'smokenum' (replaced with 0 number of cigarette smoked per week). The collected dataset contained 439 records, however, there were 10 records that contained many missing data and those records were being removed. The remaining sample size was 429.

However, there were still some records having missing data for different attributes, therefore, Microsoft Azure Mahcine Learning Studio was being used to replace the missing data using Probabilistic PCA which was widely used for data preprocessing (Karhunen, 2011; Severson, Molaro & Braatz, 2017) as depicted in Figure 4.1. After replacing the missing data uisng Probabilistic PCA, the dataset was converted to ARFF (Attribute Relation File Format) format as it is the main file format for the Weka software which would be used in the following sections. The final dataset consist of 42 per cent males and 58 per cent females, with ages ranging from 18 to 82 (mean = 37.4).

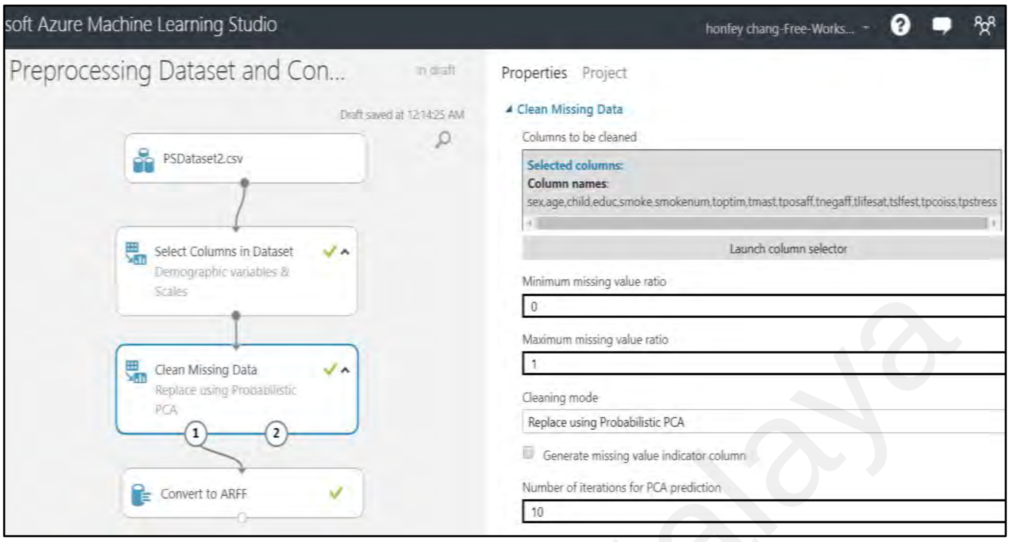


Figure 4.1: Replacing missing data with Probabilistic PCA and converting the dataset to ARFF format using Microsoft Azure Machine Learning Studio

Figure 4.2 shows part of the dataset in ARFF format. ARFF format was specifically established for the use of Weka software. There are three sections in the ARFF file, such as @RELATION, @ATTRIBUTE and @DATA sections. @RELATION section simply defines the name of the dataset and does not contribute any other function. @ATTRIBUTE section is the important declarations of the names and data types of the attributes whereby each line of data from the @DATA section represents a record of the attribute values following the order of the attribute declarations set in the @ATTRIBUTE section.

Weka (Waikato Environment for Knowledge Analysis) software was used to implement the experiments in this study as recommended by Murphy (2015) that Weka is an easy to use powerful tool for ML and data mining, Besides, Weka consisted all the single models and ensemble models that were identified from the literature. Workbench application was chosen as the environment for model development in this study as it provided the Preprocess, Classify and Experiment modules for the ease to develop the models. Figure 4.3 shows the graphical user interface (GUI) of the Weka Workbench application.

```
@RELATION Unnamed

@ATTRIBUTE sex NUMERIC
@ATTRIBUTE age NUMERIC
@ATTRIBUTE child NUMERIC
@ATTRIBUTE educ NUMERIC
@ATTRIBUTE smoke NUMERIC
@ATTRIBUTE smokenum NUMERIC
@ATTRIBUTE toptim NUMERIC
@ATTRIBUTE tmast NUMERIC
@ATTRIBUTE tposaff NUMERIC
@ATTRIBUTE tnegaff NUMERIC
@ATTRIBUTE tlifesat NUMERIC
@ATTRIBUTE tslfest NUMERIC
@ATTRIBUTE tpcoiss NUMERIC
@ATTRIBUTE tpstress NUMERIC

@DATA
2,24,1,5,2,0,22,22,49,39,23,35,51,29
2,48,1,2,2,0,19,19,15,14,33,31,47,19
1,41,1,2,2,0,26,26,49,36,33,40,63,31
```

Figure 4.2: Dataset in ARFF format

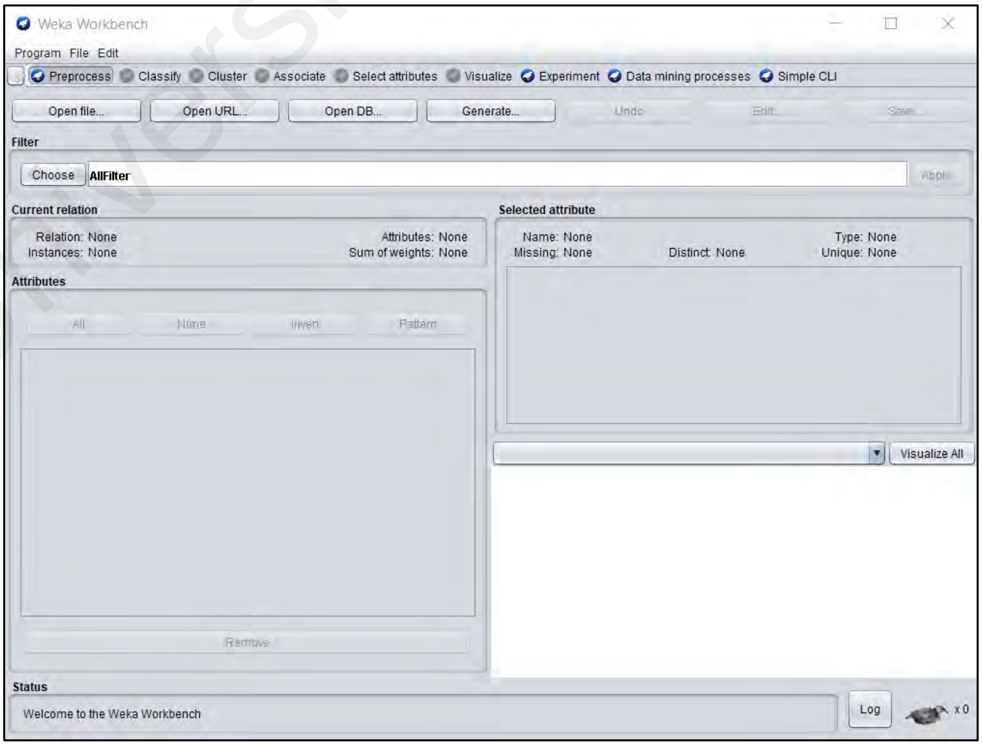


Figure 4.3: GUI of Weka Workbench application

First of all, the cleaned dataset in ARFF file was loaded to the Preprocess module of Weka Workbench, as shown in Figure 4.4. The 14 attributes were shown in the Attributes panel on the bottom left and some basic statistic information of the selected attribute was shown on the panel on the right, such as minimum, maximum, mean and standard deviation of the selected attribute.

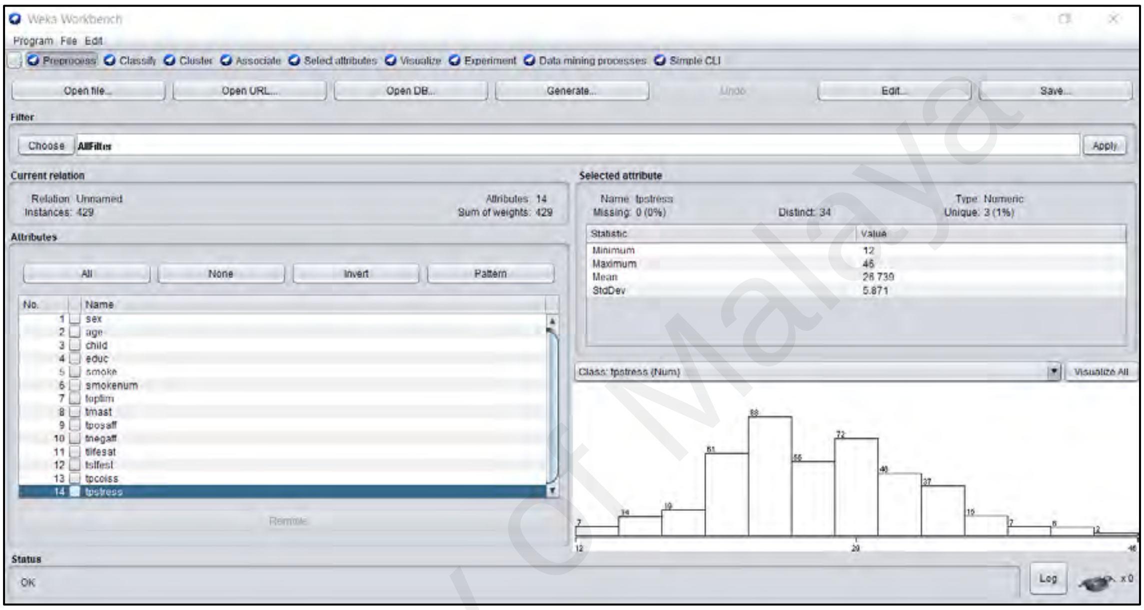


Figure 4.4: The attributes of the dataset loaded in the Preprocess module of Weka Workbench

4.1.4 Reliability Analysis

After data cleaning, reliability analysis was done (using IBM SPSS Statistic Software) to analyze the reliability of the measurements used to confirm whether the scales adopted in the dataset are reliable to be used in this study. If any of the scales used in the dataset is not reliable, the related variable that was measured by the scale would not be included in this study. Table 4.3 shows the reliability of all measurements that were used in the dataset collected in this study. All measurements showed adequate reliability, which their Cronbach's Alpha values are above 0.70.

Table 4.3: Reliability of the measurements

Measurement	Cronbach's Alpha
Perceived stress	0.855
Optimism	0.799
Mastery	0.764
Positive affect	0.872
Negative affect	0.876
Life satisfaction	0.892
Self-esteem	0.852
Perceived control of internal states (PCOIS)	0.901

4.2 Model Development

This section will discuss the implementation procedures for attribute selection as well as the single and ensemble model development. Ensemble model development will be separated to two sections, which are homogeneous and heterogeneous ensemble model development.

4.2.1 Attribute Selection

Next process was to select the predictors of perceived stress from the 14 attributes using M5 attribute selection method that is embedded in Linear Regression model. The Classify module of Weka Workbench was selected as shown in Figure 4.5. After that, the Linear Regression classifier was selected (the default classifier was ZeroR), and the attribute selection method of the classifier was changed to M5 method as shown in Figure 4.6. Following that, the Linear Regression classifier was being built using 10-fold cross-validation and Figure 4.7 shows the result for the selected predictors of perceived stress in the Linear Regression model.

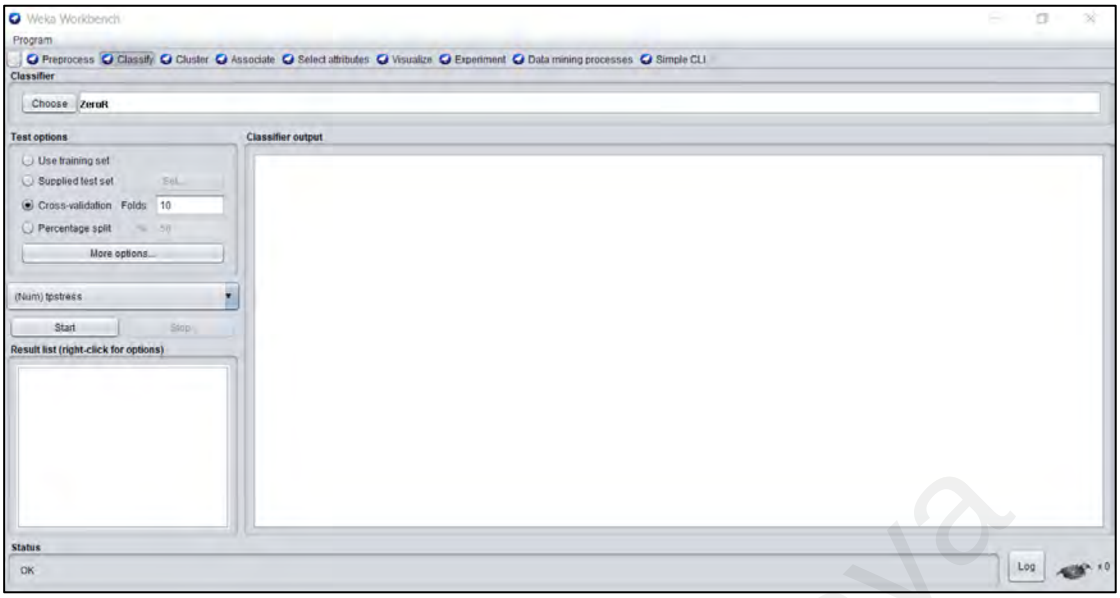


Figure 4.5: GUI of Classify module in Weka Workbench

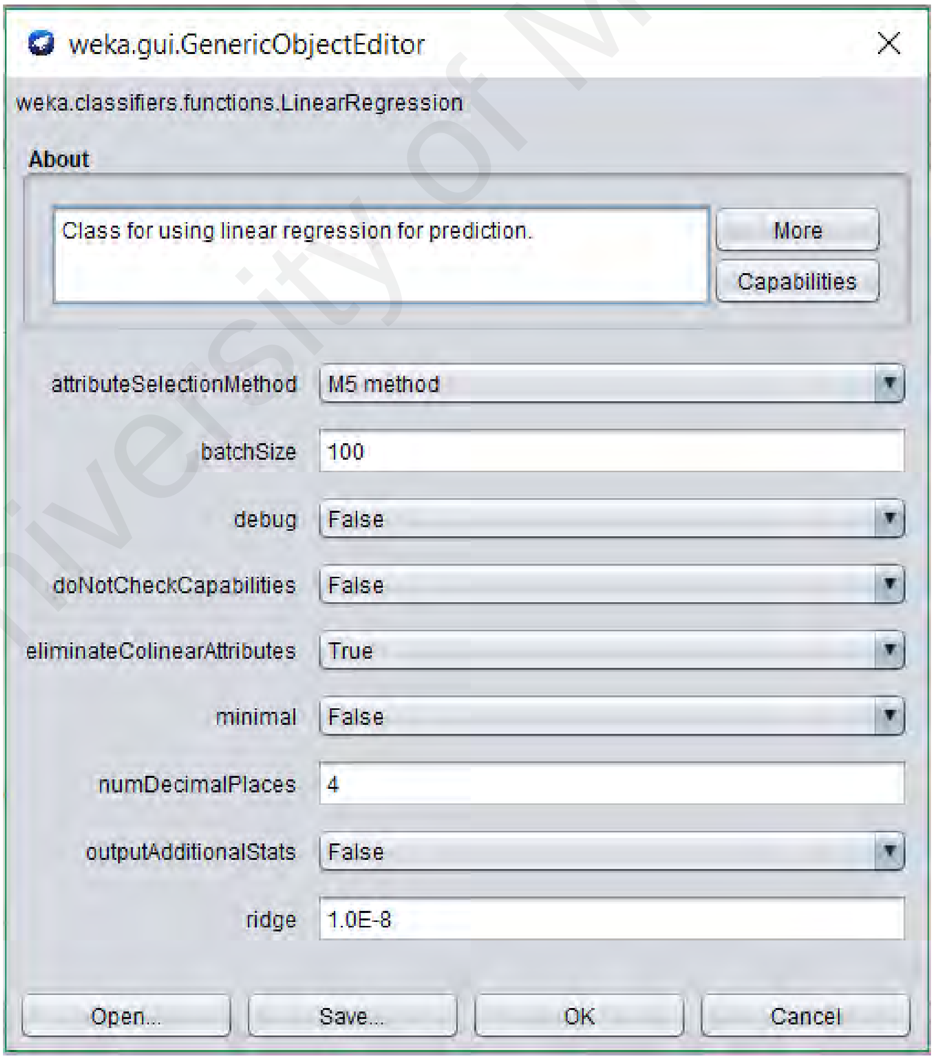


Figure 4.6: Linear Regression classifier with M5 attribute selection method

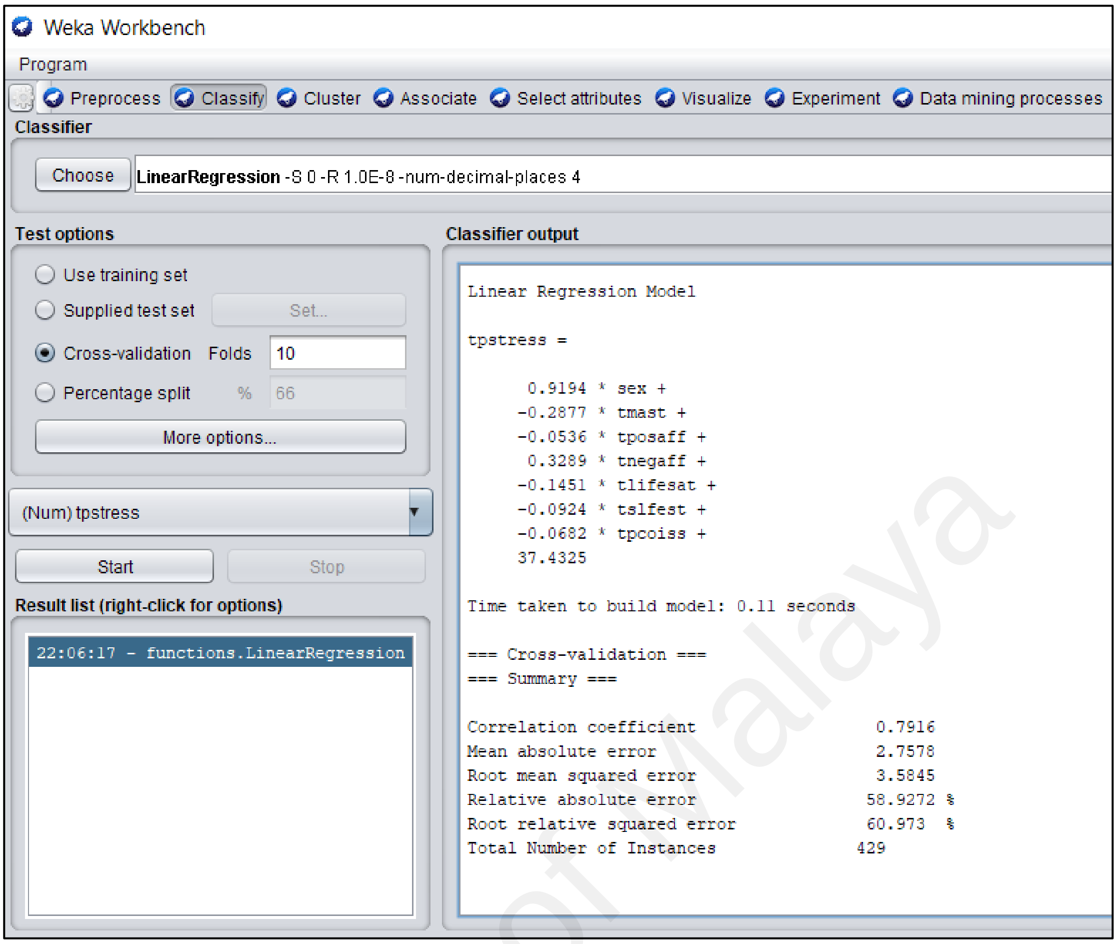


Figure 4.7: Result for the selected predictors of perceived stress

Among six demographic attributes and seven personality trait attributes, there were only one demographic attribute (gender) and six personality trait attributes (mastery, positive affect, negative affect, life satisfaction, self-esteem, and PCOIS) were being selected through the M5 method. Below is the built Linear Regression model of the perceived stress with the predictors selected using M5 attribute selection method as shown in Figure 4.7:

$$
\begin{aligned}
\text{Perceived Stress} \quad = \quad & 0.9194 * \text{Gender} + \\
& -0.2877 * \text{Mastery} + \\
& -0.0536 * \text{Positive Affect} + \\
& 0.3289 * \text{Negative Affect} + \\
& -0.1451 * \text{Life Satisfaction} + \\
& -0.0924 * \text{Self Esteem} + \\
& -0.0682 * \text{PCOIS} + \\
& 37.4325
\end{aligned}
$$

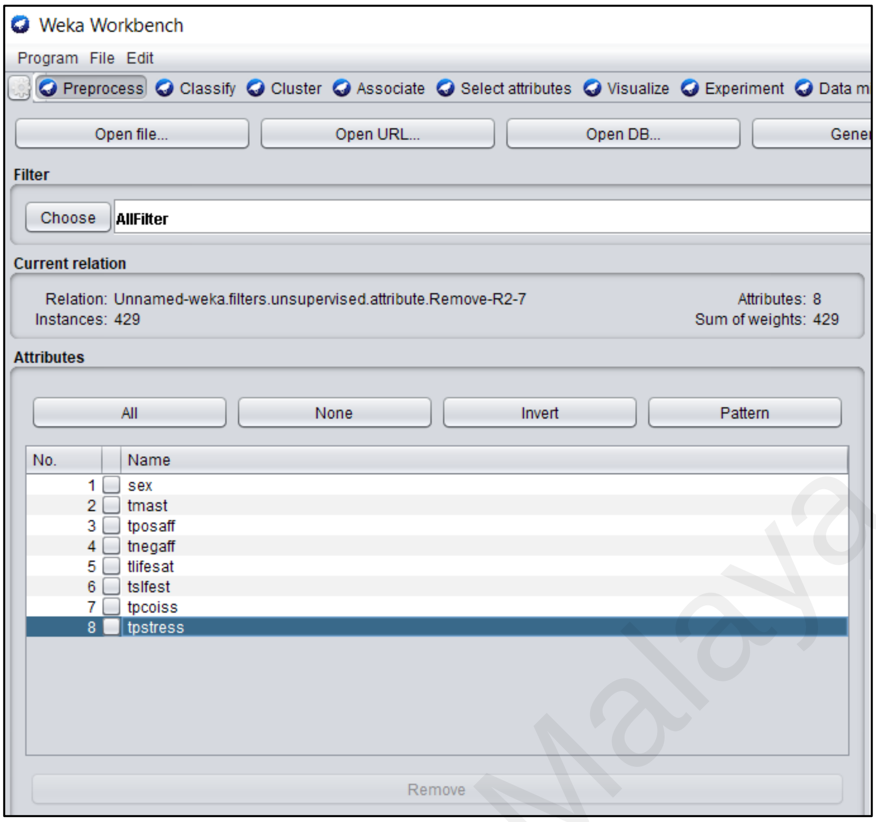


Figure 4.8: The seven selected predictors and the output perceived stress

As a result, the rest of the attributes were being removed from the Preprocess module and only eight attributes were left (seven predictors and one perceived stress) as shown in Figure 4.8. The latest dataset with eight attributes was saved in a new ARFF file and ready for the use of model development purpose.

4.2.2 Single Model Development

The Experiment module of Weka Workbench (refer to Figure 4.9) was chosen to run the experiments in this study. Weka Experiment module allows the user to run, analyze, and compare multiple models against one or several datasets in a convenient way. Before starting the experiment, the new ARFF file was loaded through the Setup tab of the Experiment module, the latest dataset with the seven predictors and one perceived stress output would be used to train and test the models. For identifying the benchmark single model, the six identified single regression models (MLR, MLP, SVM, ELN, GPR, and RF) with default parameters would be added to the Weka Experiment module as shown in Figure 4.10. Besides, 10-fold cross-validation was selected to evaluate the models.

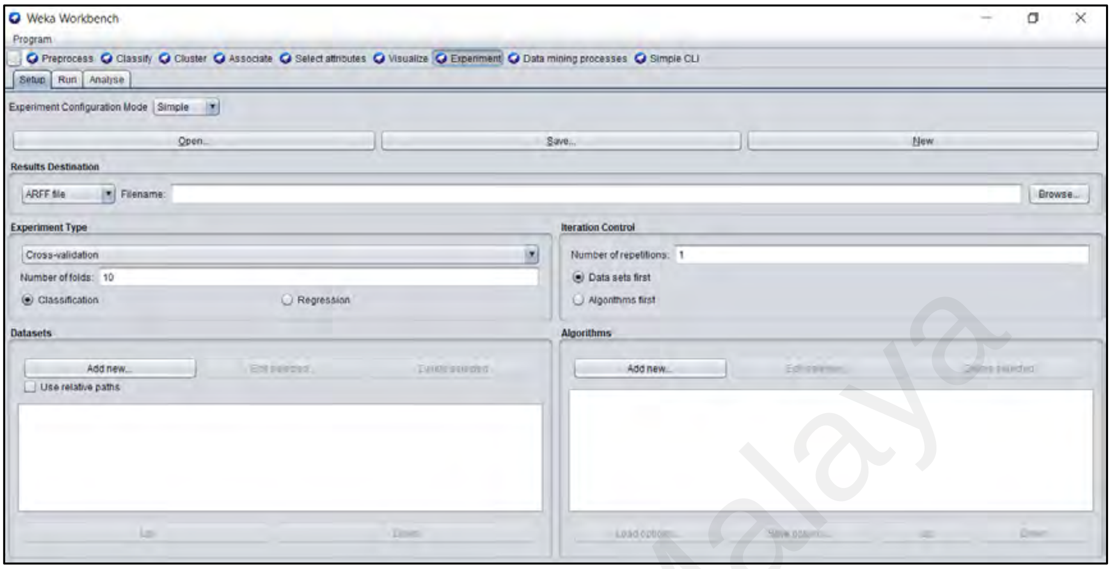


Figure 4.9: Experiment module of Weka Workbench

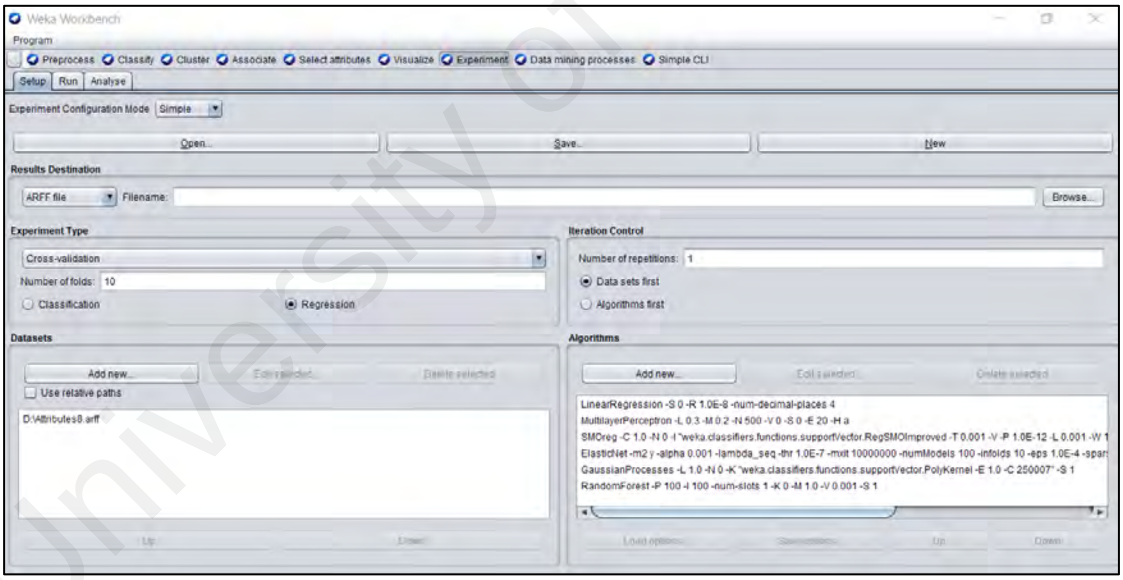


Figure 4.10: Six single regression models were added to the Weka Experiment module

Table 4.4: Names of the single models in Weka software

Identified Single Regression Model	Name of the Model in Weka Software
MLR	functions.LinearRegression
MLP	functions.MultilayerPerceptron
SVM	functions.SMOreg
ELN	functions.ElasticNet
GPR	functions.GaussianProcesses
RF	trees.RandomForest

Some models in Weka software have different names compared to the identified models in this study, but they are the same models. Table 4.4 shows the names of the identified models in Weka software. After running the experiment to build the six single models, the results were analysed using the experiment analyser as shown in Figure 4.11. The predictive performances of those models in MAE were being compared with each other's to identify the benchmark single model as well as their rank positions according to their predictive performances.

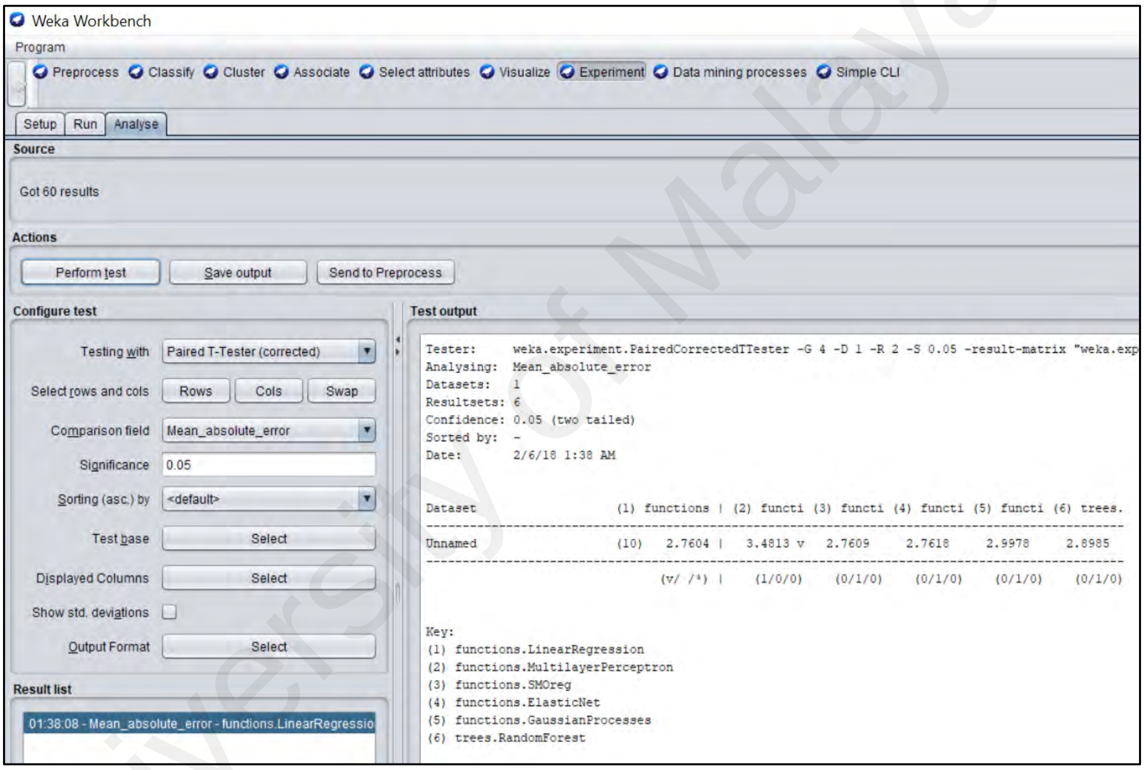


Figure 4.11: MAE of the single models

Table 4.5 shows the predictive performances of the single models in MAE and their rank positions (for the use to develop the ensemble models in next section); whereby lower MAE means that the prediction made smaller errors or more accurate. The result shows that MLR was the benchmark single model in this study. Following best single model was SVM, then ELN, RF, GPR, and the last one was MLP. Figure 4.12 summarizes the experimental design for the single model development of this study.

Table 4.5: Rank positions of the single models in predicting perceived stress according to MAE

Single Base Model	MAE	Ranking
MLR	2.7604	1
SVM	2.7609	2
ELN	2.7618	3
RF	2.8985	4
GPR	2.9978	5
MLP	3.4813	6

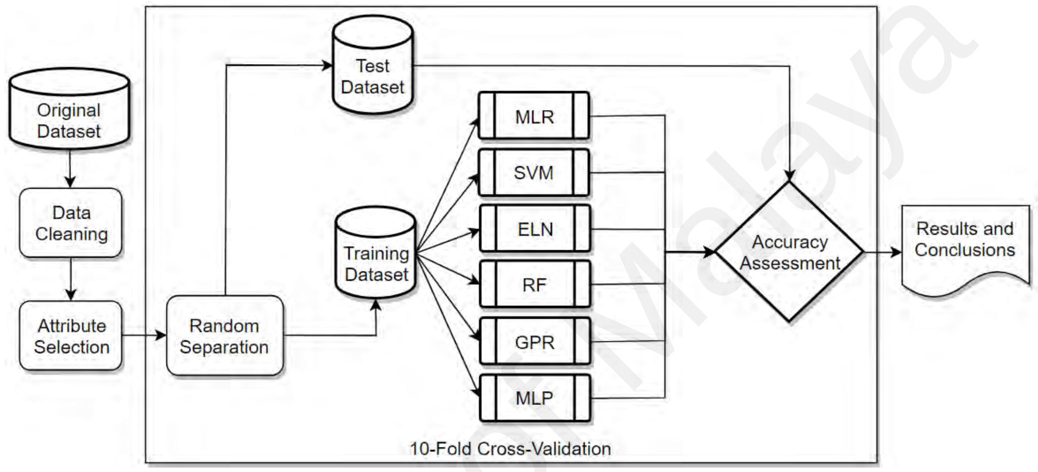


Figure 4.12: Experimental design for the single model development

4.2.3 Homogeneous Ensemble Model Development and Experiments

The commonly used homogeneous ensemble models that were identified from the literature are BG, RSS, and AR. Table 4.6 shows the names of the identified homogeneous ensemble models in Weka software. As the ensemble framework (refer to Figure 2.2) proposed by King et al. (2014), each homogeneous ensemble models required an instance to be built from each of the single models. Since there were six commonly used single models being adopted in this study (MLR, SVM, ELN, RF, GPR, and MLP), therefore, there would be six BG instances, six RSS instances, and six AR instances being developed, as depicted in Figure 4.13. Each instance would be created with the selected predictors using 10-fold cross-validation to predict the perceived stress, and their predictive performances in MAE would be compared.

Table 4.6: Names of the homogeneous ensembles in Weka software

Name of Homogeneous Ensembles	Name of the Ensembles in Weka Software
BG	meta.Bagging
RSS	meta.RandomSubSpace
AR	meta.AdditiveRegression

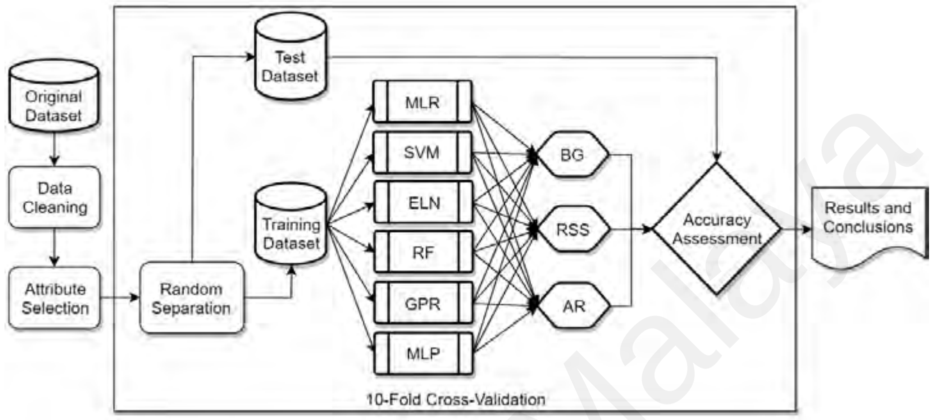


Figure 4.13: Experimental design of homogeneous ensembles (BG, RSS, and AR)

The six instances of BG ensemble were built first, followed by the instances of RSS and AR ensembles. Figure 4.14 shows the example for setting up a BG instance where the BG (Bagging) ensemble was selected with MLR (Linear Regression) model as the base learner, other parameters remained with default values. The setup of the other five BG instances was done in the same manner, except that their base learners were different.

Figure 4.15 shows the six BG instances that were selected in the Experiment module of Weka Workbench and Figure 4.16 shows the predictive performances of the BG instances in MAE. Figure 4.17 shows the six RSS instances that were selected in the Experiment module of Weka Workbench and Figure 4.18 shows the predictive performances of the RSS instances in MAE. Figure 4.19 shows the six AR instances that were selected in the Experiment module of Weka Workbench and Figure 4.20 shows the predictive performances of the AR instances in MAE.

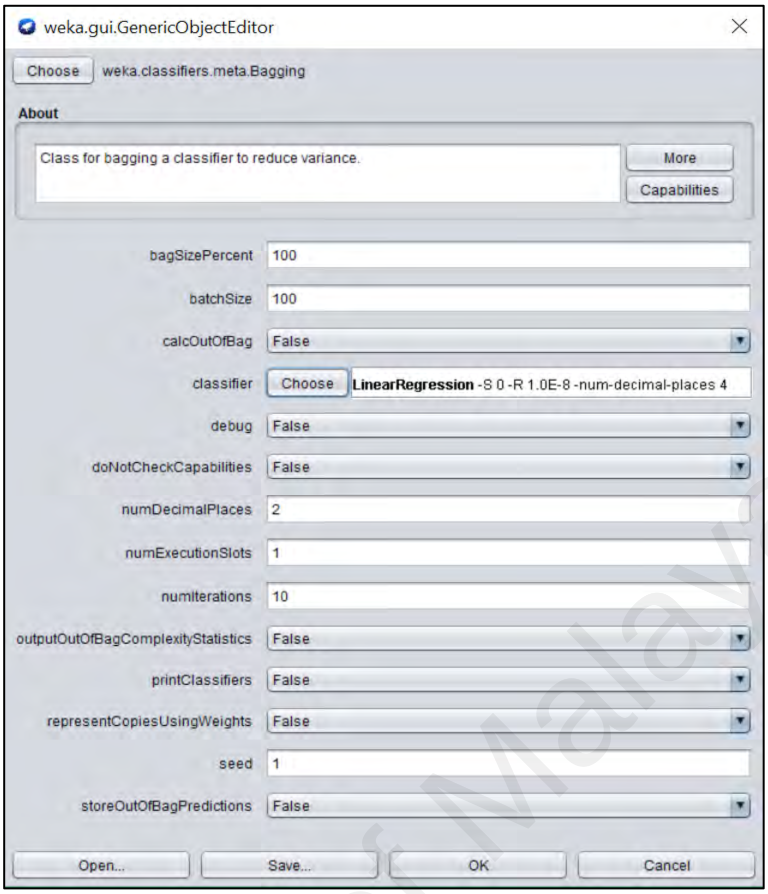


Figure 4.14: Setting up the BG instance with MLR as the base learner

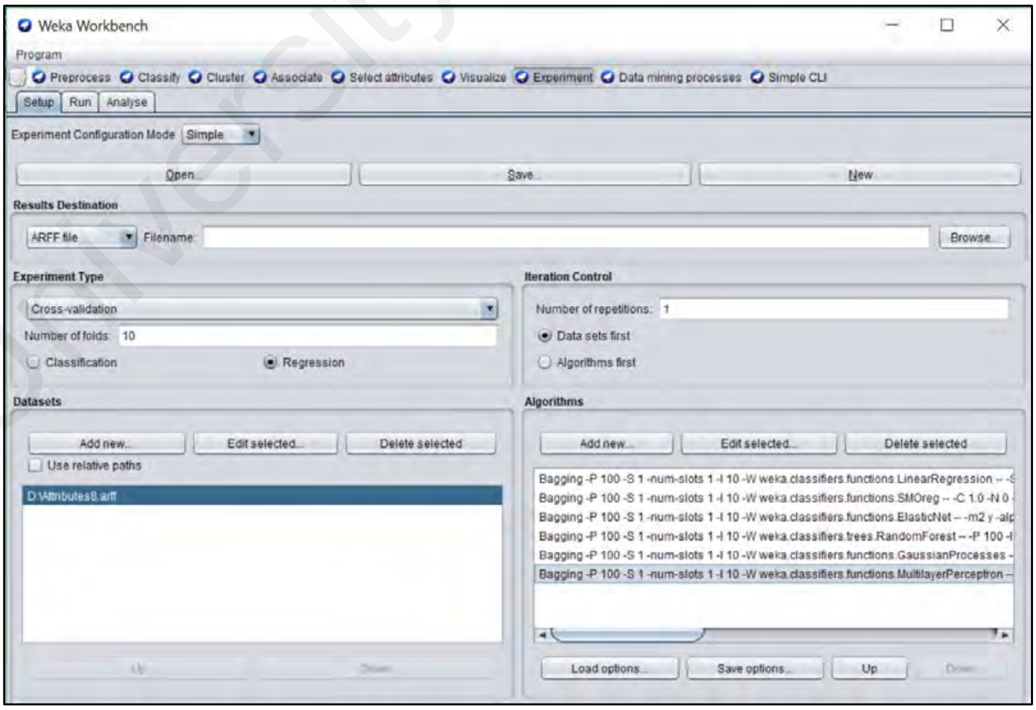


Figure 4.15: Six BG instances with different single model as base learner were added to the Weka Experiment module

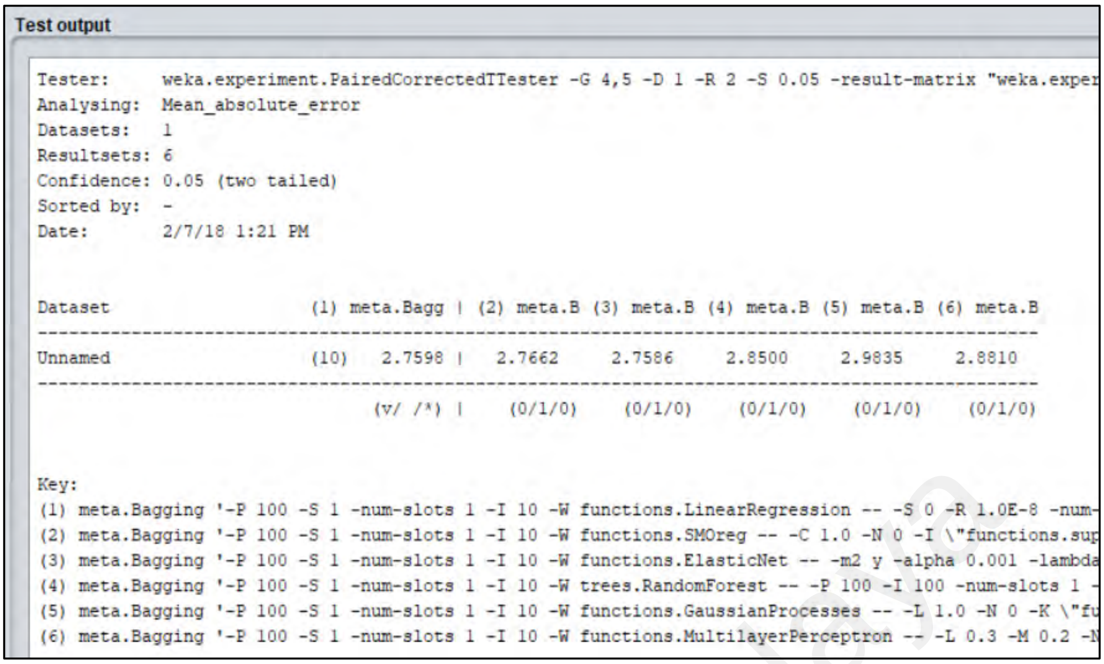


Figure 4.16: MAE of the BG instances

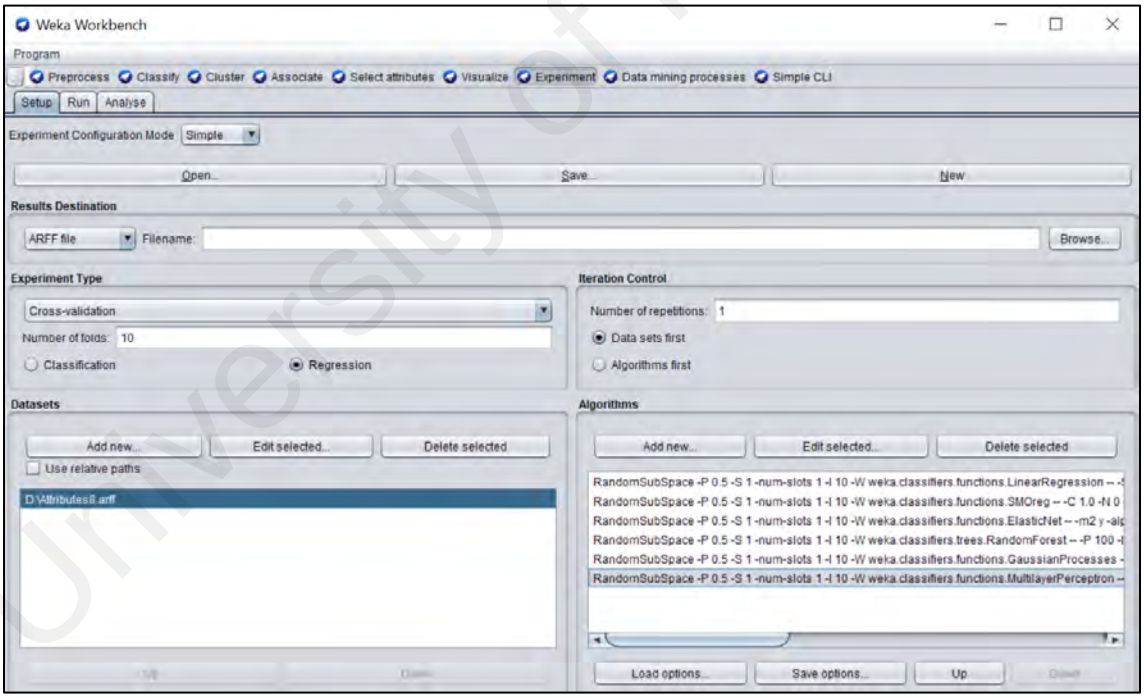


Figure 4.17: Six RSS instances with different single model as base learner were added to the Weka Experiment module

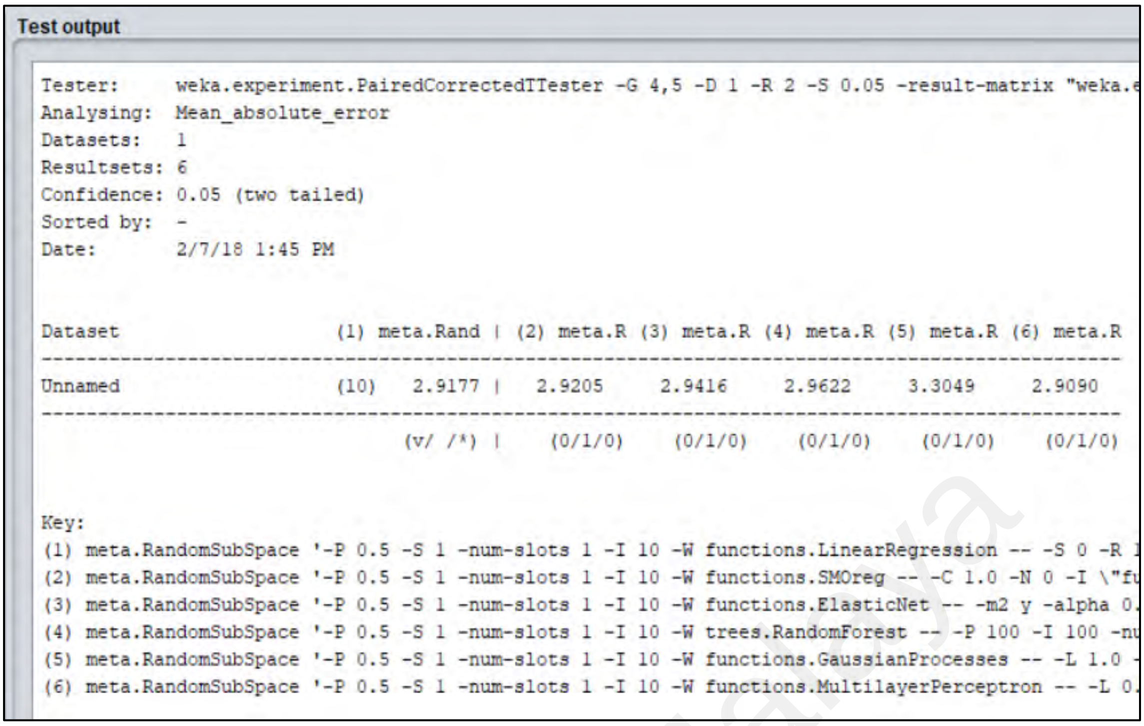


Figure 4.18: MAE of the RSS instances

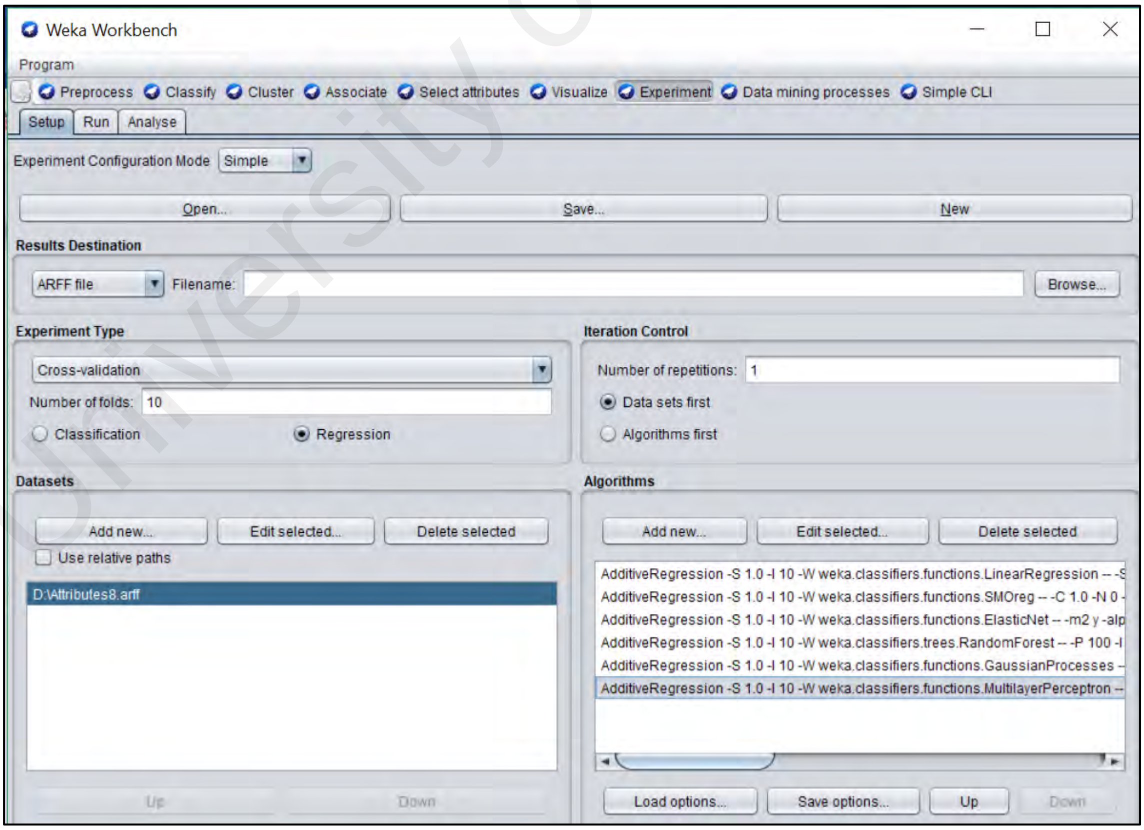


Figure 4.19: Six AR instances with different single model as base learner were added to the Weka Experiment module

```
Test output

Tester:     weka.experiment.PairedCorrectedTTester -G 4,5 -D 1 -R 2 -S 0.05 -result-matrix "weka.e
Analysing:  Mean_absolute_error
Datasets:   1
Resultsets: 6
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       2/7/18 1:53 PM


Dataset                  (1) meta.Addi | (2) meta.A (3) meta.A (4) meta.A (5) meta.A (6) meta.A
-----------------------------------------------------------------------------------------------
Unnamed                  (10)   2.7604 |   2.7586     2.7618     2.9124     2.9226     3.5785 v
-----------------------------------------------------------------------------------------------
                              (v/ /*) |    (0/1/0)    (0/1/0)    (0/1/0)    (0/1/0)    (1/0/0)


Key:
(1) meta.AdditiveRegression '-S 1.0 -I 10 -W functions.LinearRegression -- -S 0 -R 1.0E-8 -num-dec
(2) meta.AdditiveRegression '-S 1.0 -I 10 -W functions.SMOreg -- -C 1.0 -N 0 -I \"functions.suppor
(3) meta.AdditiveRegression '-S 1.0 -I 10 -W functions.ElasticNet -- -m2 y -alpha 0.001 -lambda_se
(4) meta.AdditiveRegression '-S 1.0 -I 10 -W trees.RandomForest -- -P 100 -I 100 -num-slots 1 -K 0
(5) meta.AdditiveRegression '-S 1.0 -I 10 -W functions.GaussianProcesses -- -L 1.0 -N 0 -K \"funct
(6) meta.AdditiveRegression '-S 1.0 -I 10 -W functions.MultilayerPerceptron -- -L 0.3 -M 0.2 -N 50
```

Figure 4.20: MAE of the AR instances

4.2.4 Heterogeneous Ensemble Model Development and Experiments

The commonly used heterogeneous ensemble models that were identified from the literature are Voting and Stacking. Table 4.7 shows the names of the identified heterogeneous ensemble models in Weka software.

Table 4.7: Names of the heterogeneous ensembles in Weka software

Name of Heterogeneous Ensembles	Name of the Ensembles in Weka Software
Voting	meta.Vote
Stacking	meta.Stacking

However, when developing the instances for heterogeneous ensemble models (Voting and Stacking), it was slightly different from the ensemble framework proposed by King et al. (2014). King et al. (2014) directly built the instances of Voting and Stacking from all the single models simultaneously as the base learners. The problem was, building the instances of Voting and Stacking from all six identified single models as base learners simultaneously would leave the predictive performances of the instances that built with only one, or two to six single models as base learners simultaneously unknown. Since the predictive performances of each single model were known, the instances of heterogeneous ensembles would be built with the best single model alone first as base learner, and the number of the single models would be gradually increased following the order of their predictive performances from best to least as in shown in Table 4.5, so that the performances the heterogeneous ensembles with different combinations of single models as base learners could be evaluated. There are six combinations of single models as the base learners of heterogeneous ensembles, as shown below:

1) MLR
2) MLR, SVM
3) MLR, SVM, ELN
4) MLR, SVM, ELN, RF
5) MLR, SVM, ELN, RF, GPR
6) MLR, SVM, ELN, RF, GPR, MLP

The Voting and Stacking instances would be built with the six combinations of base learners above, in turn.

However, a meta learner is required in the second level training of the Stacking instances. All the single models could be the meta learner of Stacking instances, hence, there would be six categories of Stacking instances, which each category of stacking instance required one of the single models as meta learner, such as Stacking-MLR, Stacking-SVM, Stacking-ELN, Stacking-RF, Stacking-GPR and Stacking-MLP. The prefix and suffix of the Stacking instance category name is separated by a hyphen, for example, Stacking-MLR, which the prefix (Stacking) represents the name of the ensemble model and the suffix (MLR) represents the meta learner. Each heterogeneous ensemble instance would be built with the selected predictors using 10-fold cross-validation to predict the perceived stress, and their predictive performances in MAE would be compared. Figure 4.21 shows the experimental design of the development of six Voting instances and 36 Stacking instances.

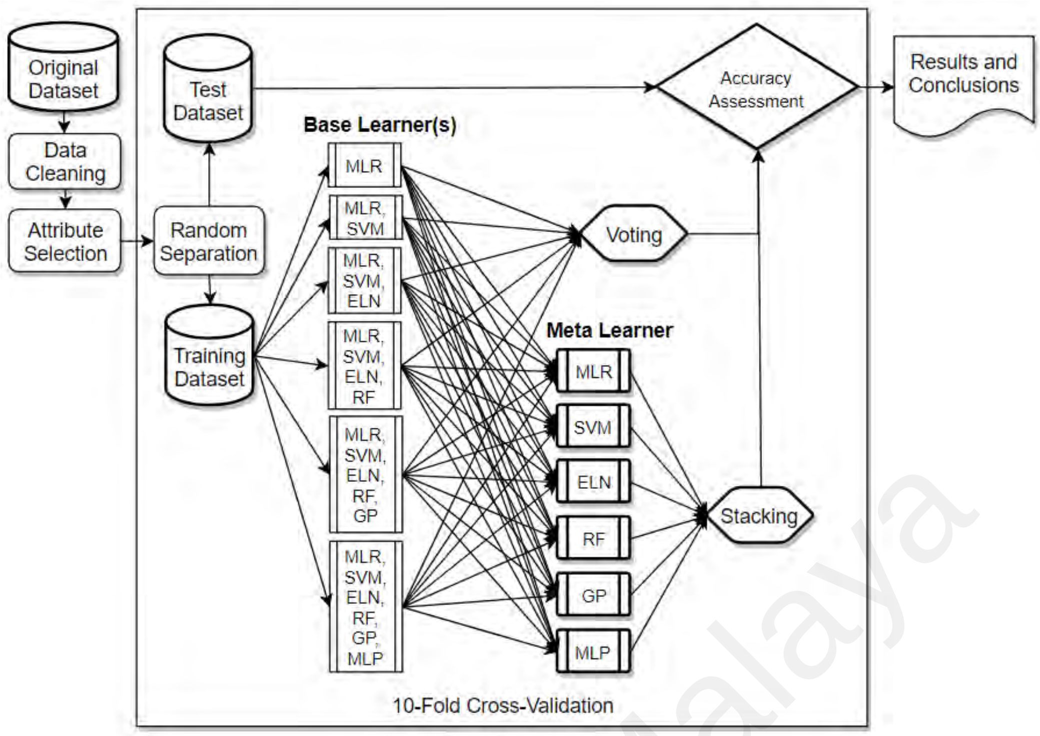


Figure 4.21: Experimental design of heterogeneous ensembles (Voting and Stacking)

Figure 4.22 shows the example for setting up a Voting instance where the six single models were selected as the base learners, other parameters remained with default values. The other five Voting instances were setup in the same way, except that their base learners were different. Figure 4.23 shows the six Voting instances that were selected in the Experiment module of Weka Workbench and Figure 4.24 shows the predictive performances of the Voting instances in MAE.

Figure 4.25 shows the example for setting up a Stacking instance with MLR as meta learner and with six single models as base learners, other parameters remained with default values. The six sets of Stacking instances (with MLR, SVM, ELN, RF, GPR, and MLP, respectively) were setup in the same way, with six combinations of single models as base learners, in turn. Figure 4.26 shows that six Stacking-MLR (Stacking with MLR as meta learner) instances with different combinations of single models as base learner(s) were added to the Experiment module of Weka Workbench and Figure 4.27 shows the predictive performances of the Stacking-MLR instances in MAE. Figure 4.28 to Figure

4.36 show the six Stacking-SVM instances, six Stacking-ELN instances, six Stacking-RF instances, six Stacking-GPR instances, six Stacking-MLP instances, with different combinations of single models as base learner(s) were respectively added to the Experiment module of Weka Workbench. Figure 4.29 to Figure 4.37 show the predictive performances of the Stacking-SVM, Stacking-ELN, Stacking- RF, Stacking-GPR, and Stacking-MLP instances in MAE respectively.

4.3 Evaluation

All the single models and ensemble models were evaluated using 10-fold cross-validation and their predictive performances in MAE were collected. The details of the results were discussed in Chapter 5.

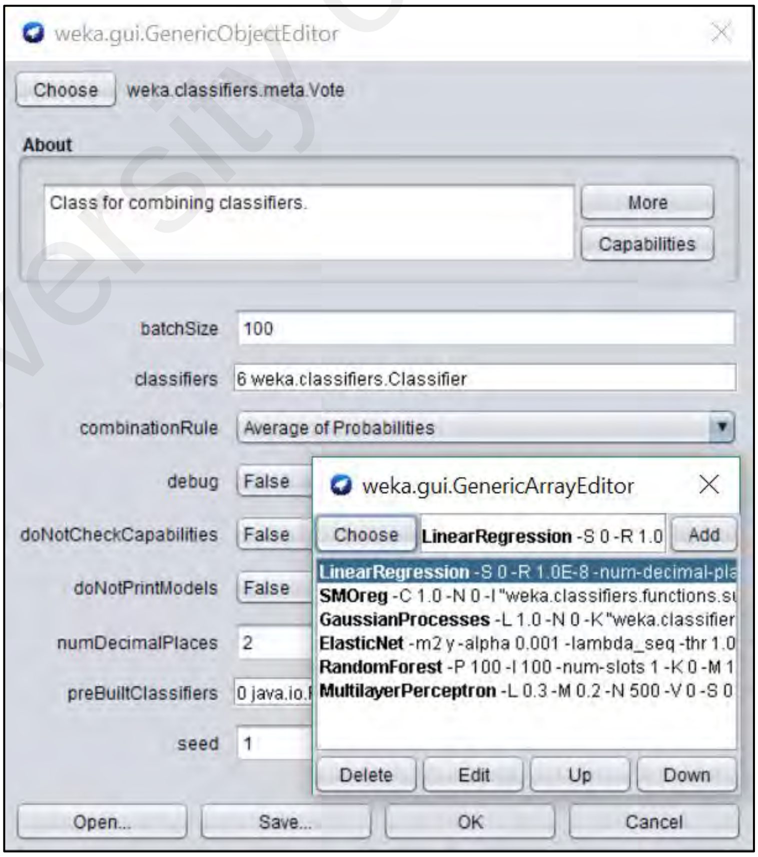


Figure 4.22: Setting up the Voting instance with six single models as the base learners

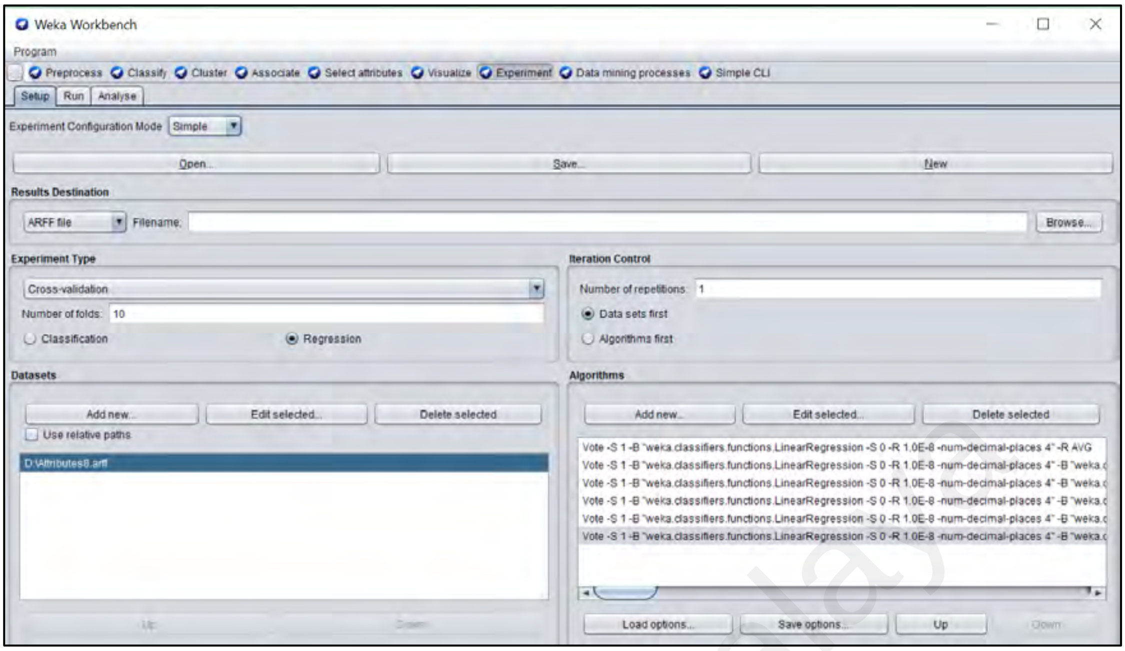
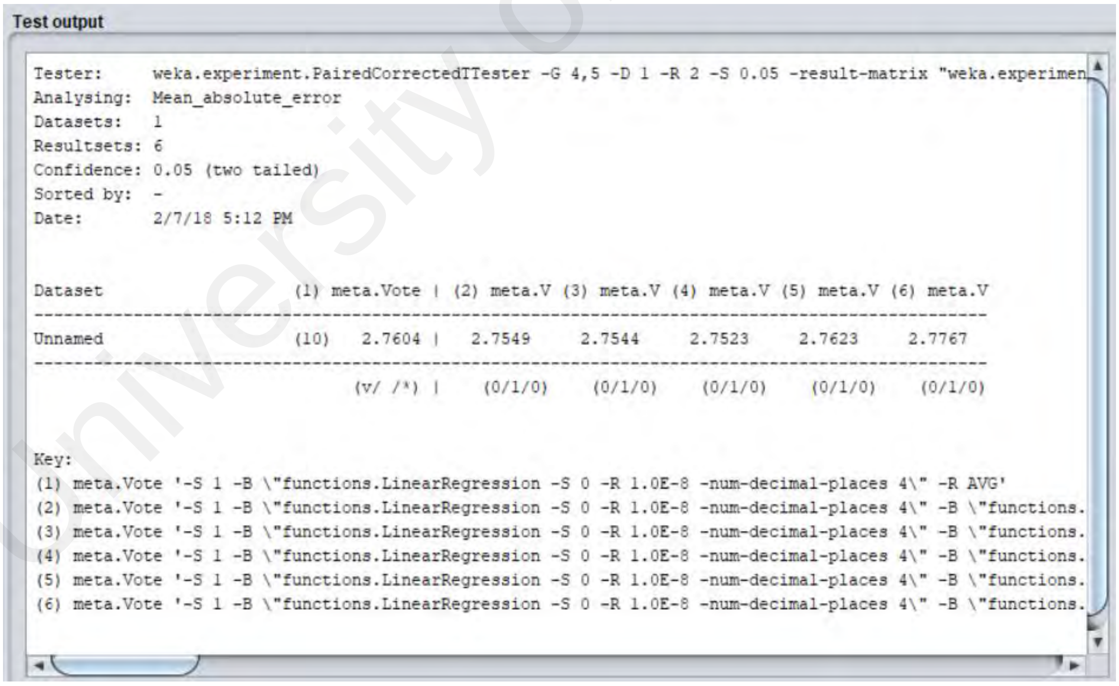


Figure 4.23: Six Voting instances with different combinations of single models as base learners were added to Weka Experiment module



Test output

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5 -D 1 -R 2 -S 0.05 -result-matrix "weka.experimen
Analysing:  Mean_absolute_error
Datasets:   1
Resultsets: 6
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       2/7/18 5:12 PM


Dataset                   (1) meta.Vote | (2) meta.V (3) meta.V (4) meta.V (5) meta.V (6) meta.V
------------------------------------------------------------------------------------------------
Unnamed                   (10)   2.7604 |   2.7549     2.7544     2.7523     2.7623     2.7767
------------------------------------------------------------------------------------------------
                                (v/ /*) |    (0/1/0)    (0/1/0)    (0/1/0)    (0/1/0)    (0/1/0)


Key:
(1) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -R AVG'
(2) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -B \"functions.
(3) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -B \"functions.
(4) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -B \"functions.
(5) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -B \"functions.
(6) meta.Vote '-S 1 -B \"functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4\" -B \"functions.
```

Figure 4.24: MAE of the Voting instances

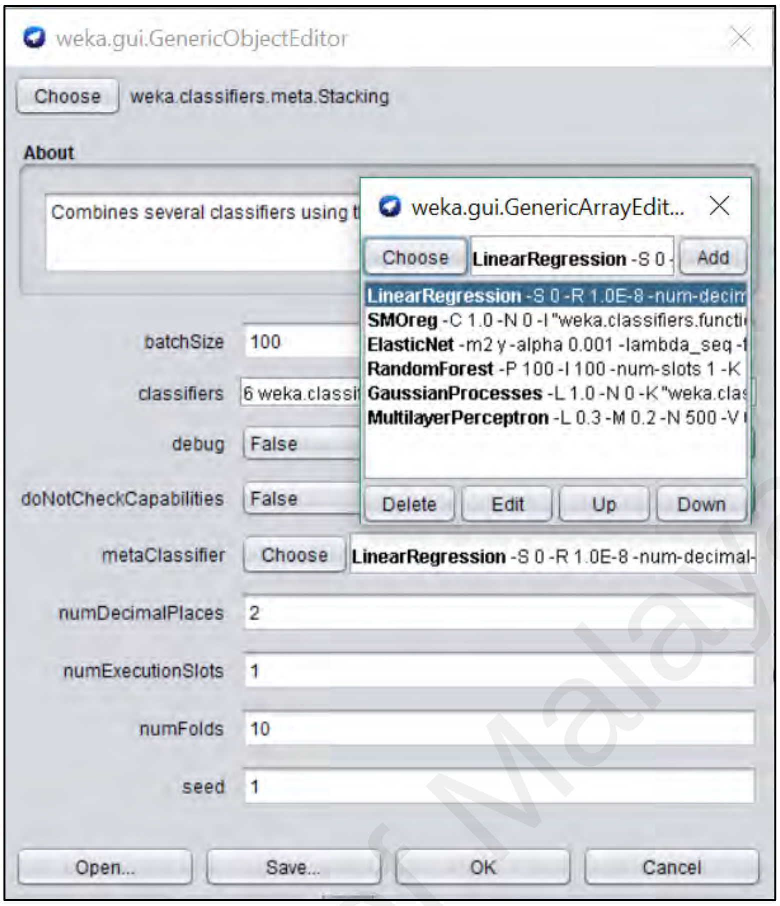


Figure 4.25: Setting up the Stacking instance with six single models as the base learners and MLR as meta learner

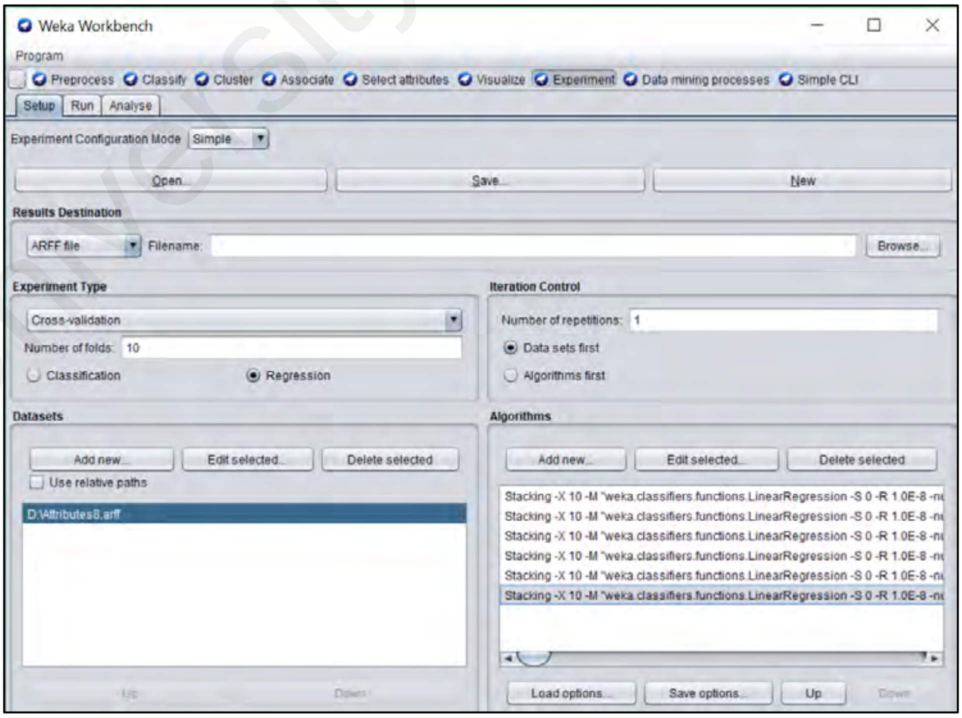


Figure 4.26: Six Stacking instances with MLR as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

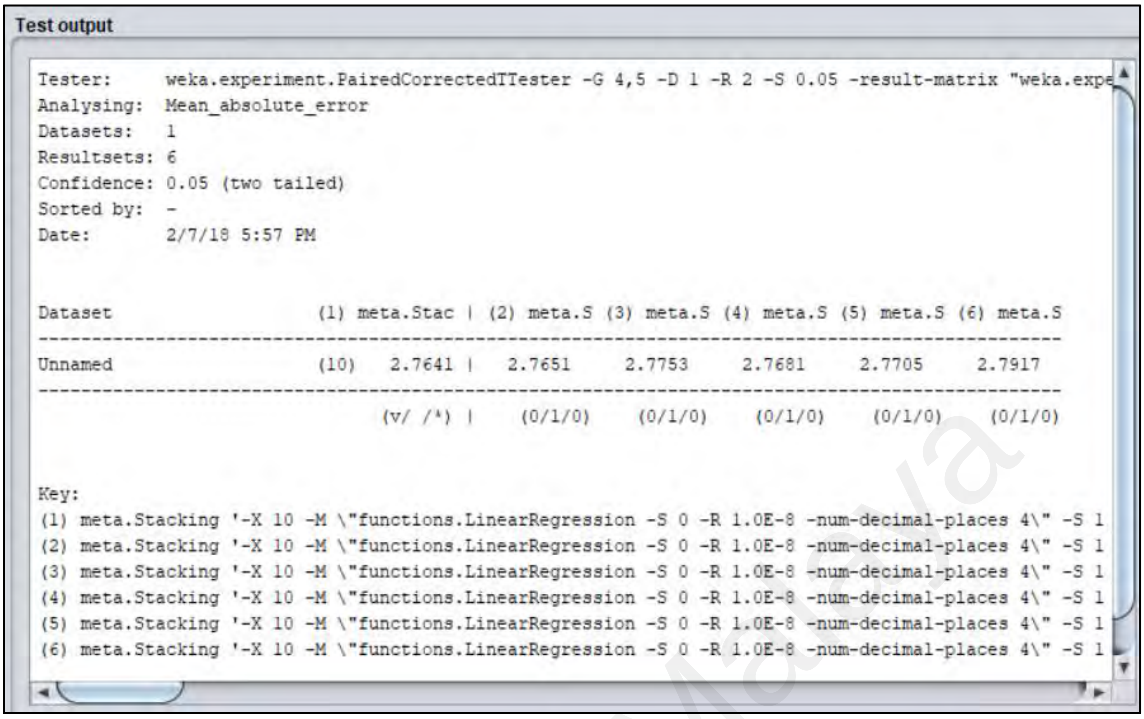


Figure 4.27: MAE of the Stacking instances with MLR as meta learner but with different combinations of single models as base learners

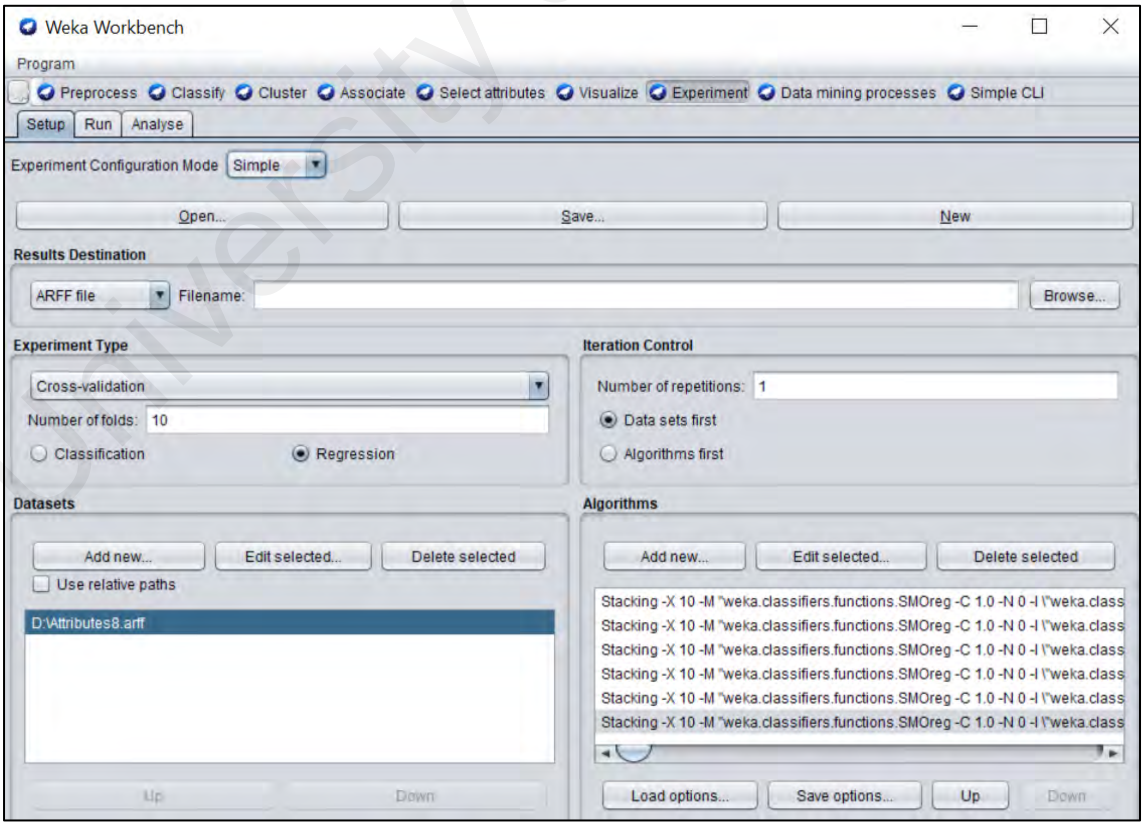


Figure 4.28: Six Stacking instances with SVM as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

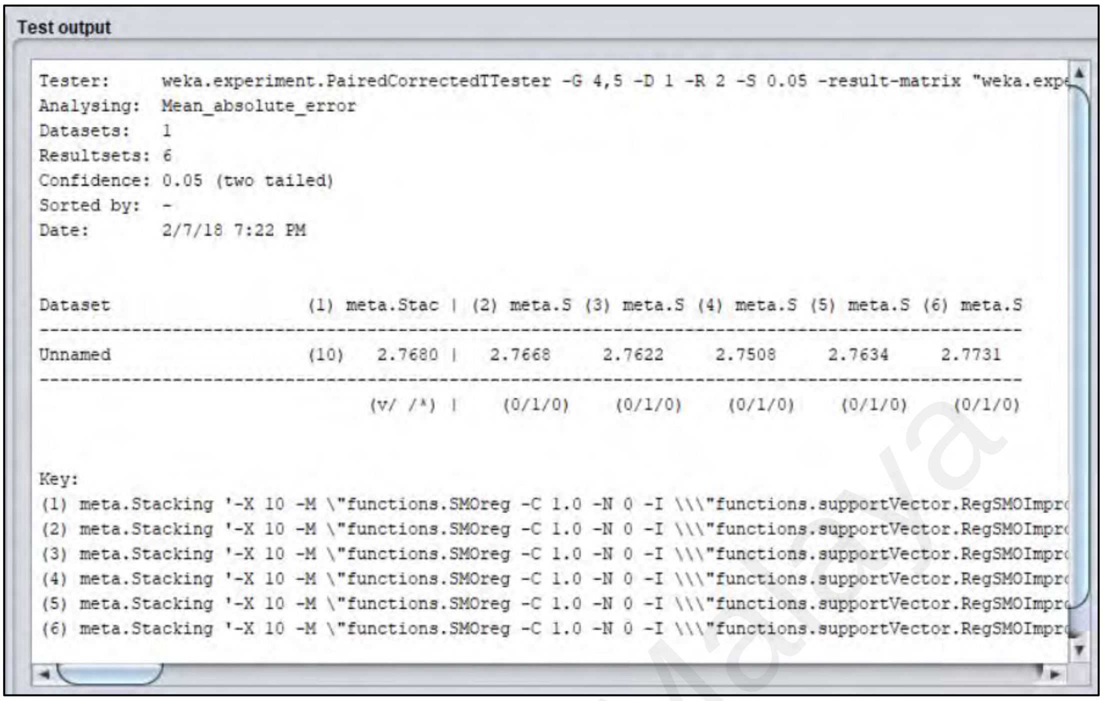


Figure 4.29: MAE of the Stacking instances with SVM as meta learner but with different combinations of single models as base learners

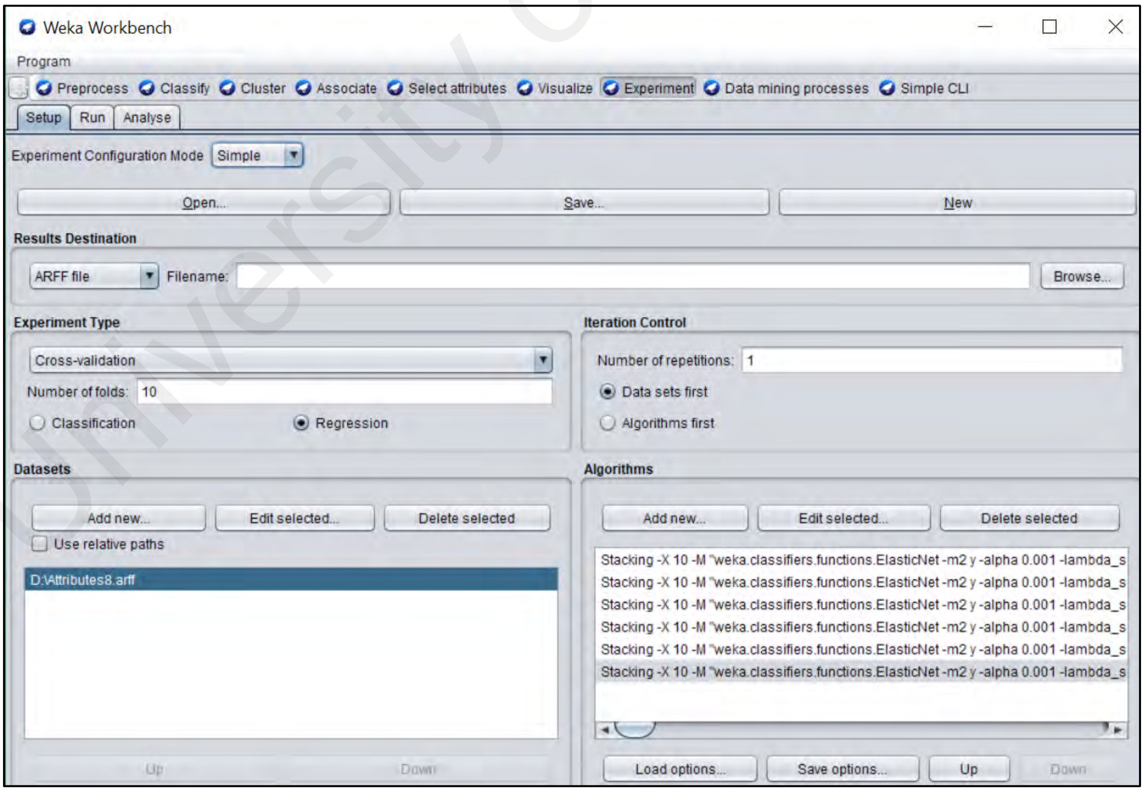


Figure 4.30: Six Stacking instances with ELN as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

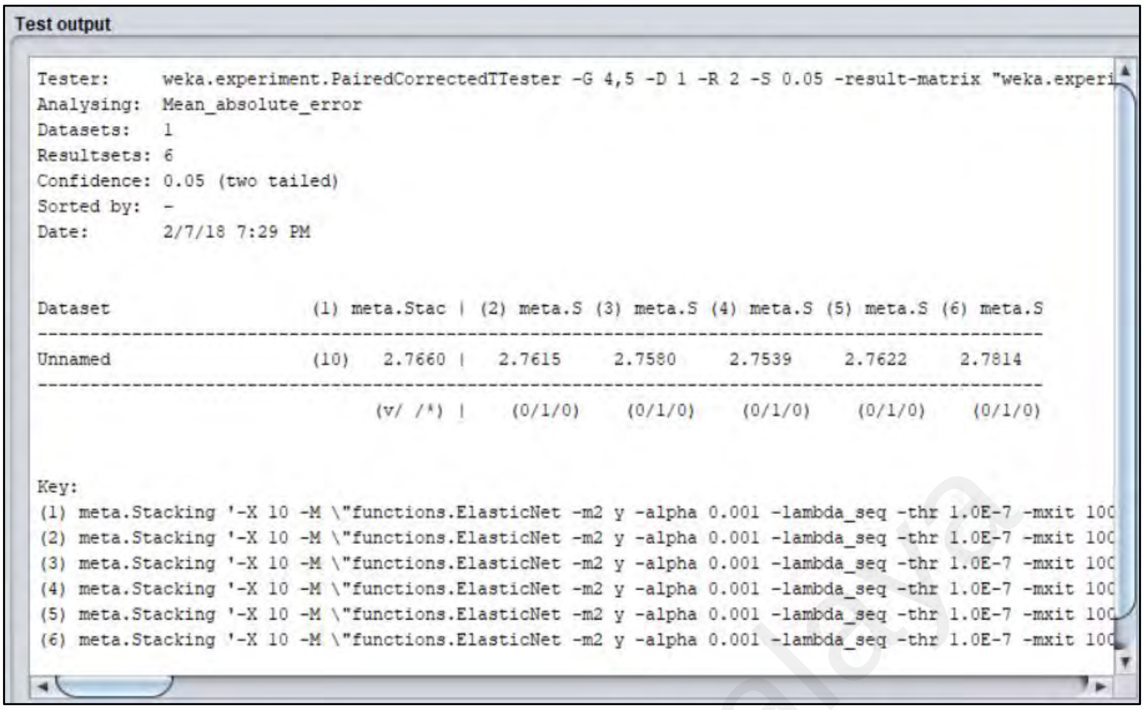


Figure 4.31: MAE of the Stacking instances with ELN as meta learner but with different combinations of single models as base learners

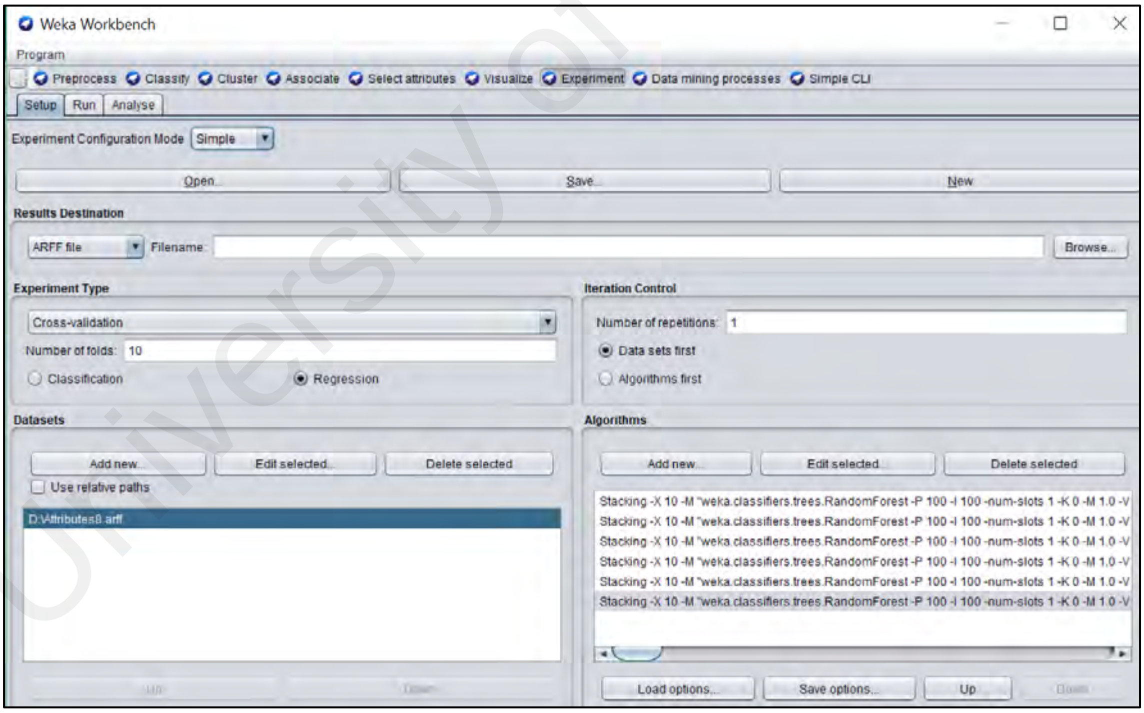


Figure 4.32: Six Stacking instances with RF as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

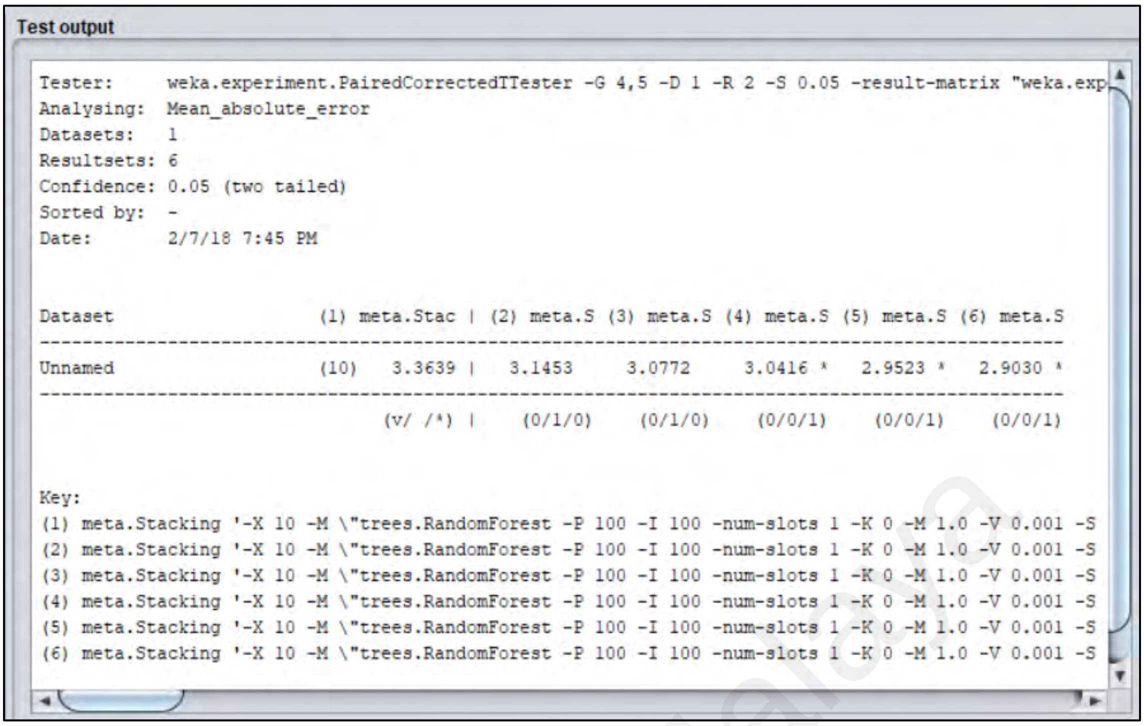


Figure 4.33: MAE of the Stacking instances with RF as meta learner but with different combinations of single models as base learners

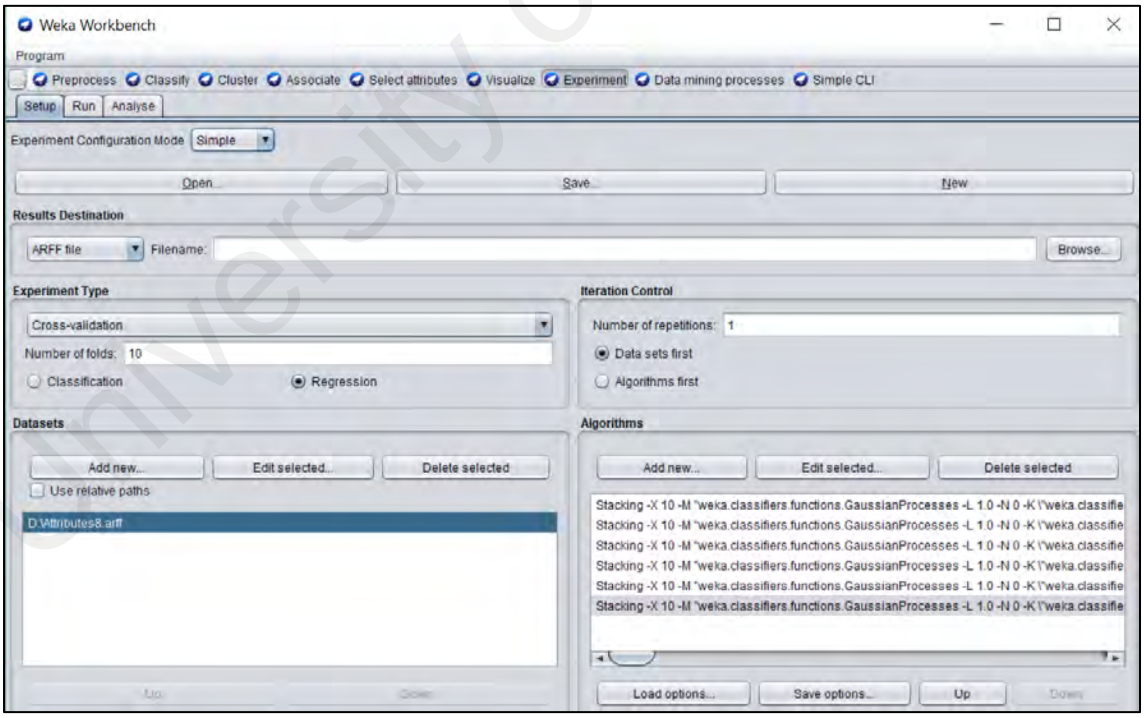


Figure 4.34: Six Stacking instances with GPR as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

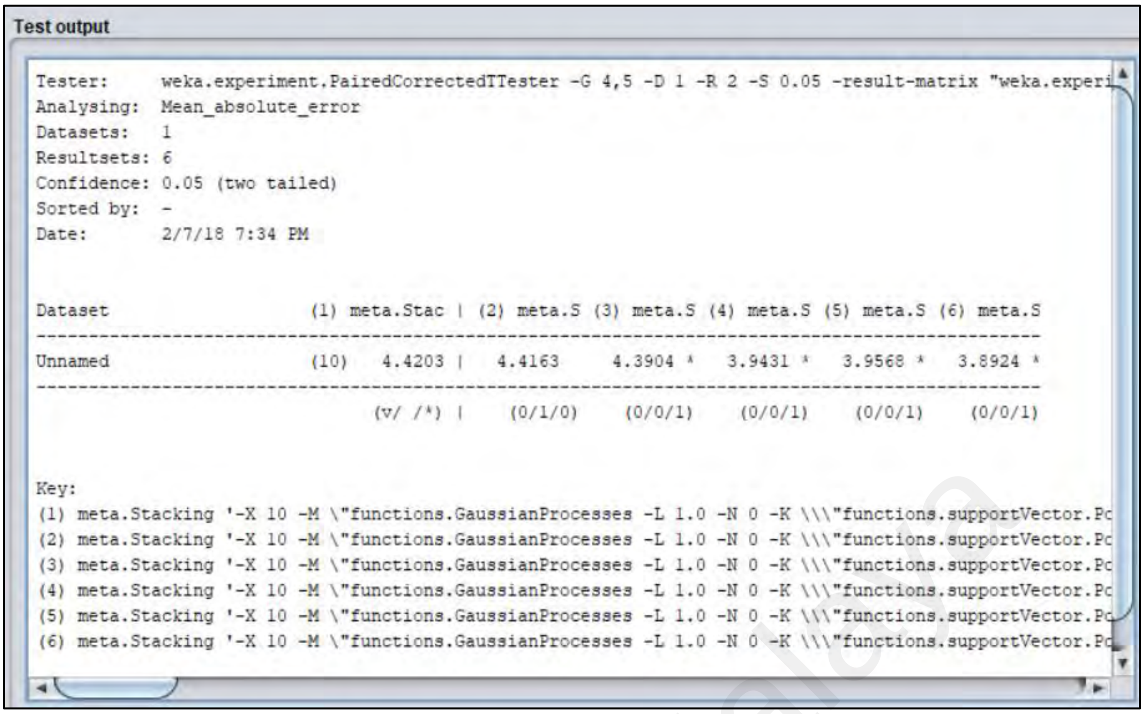


Figure 4.35: MAE of the Stacking instances with GPR as meta learner but with different combinations of single models as base learners

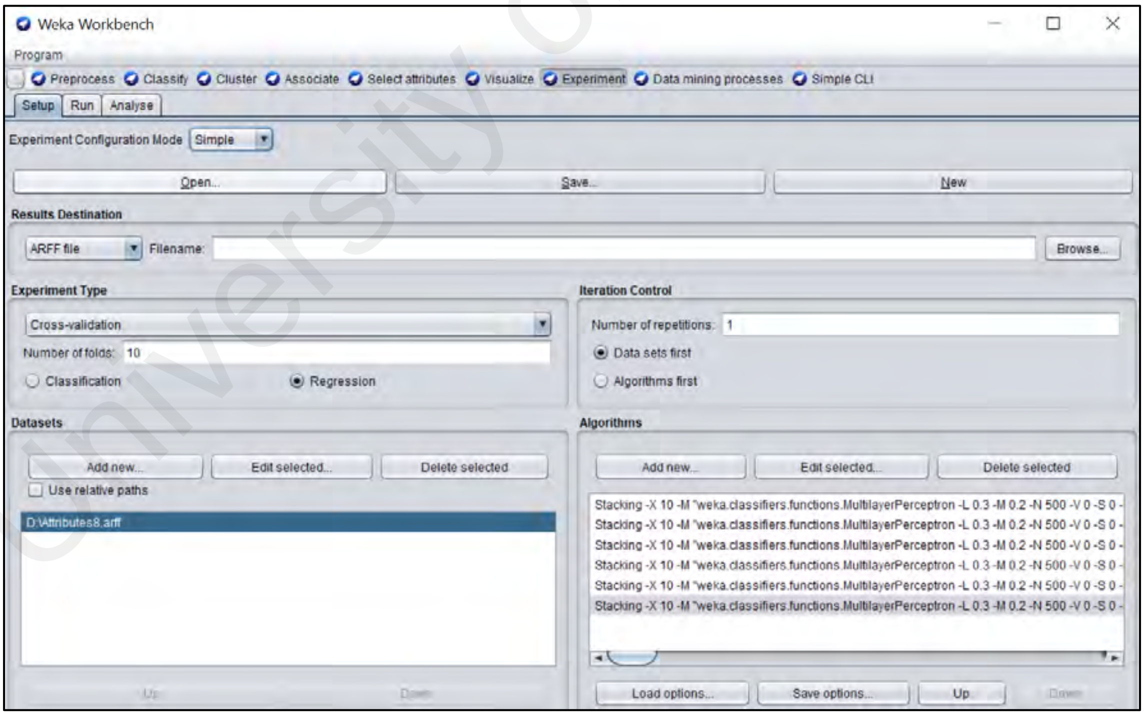


Figure 4.36: Six Stacking instances with MLP as meta learner but with different combinations of single models as base learners were added to Weka Experiment module

```
Test output

Tester:     weka.experiment.PairedCorrectedTTester -G 4,5 -D 1 -R 2 -S 0.05 -result-matrix "weka.ex
Analysing:  Mean_absolute_error
Datasets:   1
Resultsets: 6
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       2/7/18 7:40 PM


Dataset                   (1) meta.Stac | (2) meta.S (3) meta.S (4) meta.S (5) meta.S (6) meta.S
-------------------------------------------------------------------------------------------------
Unnamed                   (10)   3.0895 |   3.1299     3.0272     3.0279     3.0291     3.1406
-------------------------------------------------------------------------------------------------
                                (v/ /*) |    (0/1/0)    (0/1/0)    (0/1/0)    (0/1/0)    (0/1/0)


Key:
(1) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
(2) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
(3) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
(4) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
(5) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
(6) meta.Stacking '-X 10 -M \"functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -
```

Figure 4.37: MAE of the Stacking instances with MLP as meta learner but with different combinations of single models as base learners

CHAPTER 5: RESEACH FINDINGS AND DISCUSSION

The purpose of this chapter is to report and discuss the results collected from the experiments designed in Chapter 4, which apply various single regression models and ensemble regression models on perceived stress prediction using relevant personality traits. Firstly, the findings for the predictors of perceived stress will be discussed. Next, the predictive performances of the single models will be compared to identify the benchmark single model. After that, the predictive performances of the homogeneous and heterogeneous ensemble models according to their base learners will be compared. Finally, comparison of the predictive performances between all the regression methods, and the improvement of the ensemble models over the benchmark single model will be discussed.

5.1 Predictors of Perceived Stress

Among six demographic attributes and seven personality trait attributes, there were only one demographic attribute (gender) and six personality trait attributes (mastery, positive affect, negative affect, life satisfaction, self-esteem, and PCOIS) selected from the attribute selection process. The predictors (attributes) of perceived stress that were selected were in accordance to the findings from the literature as shown in Table 2.2, except optimism and smoking behaviour which were not selected. All the selected predictors negatively associated with perceived stress, except negative affect and gender (which females reporting more perceived stress than males) positively associated with perceived stress. The seven predictors were used to train and test all the predictive methods (will be discussed in following sections) in this study.

5.2 Single Regression Models

Figure 5.1 shows the predictive performances of the six single models in MAE; whereby lower MAE means that the prediction made smaller errors and more accurate. The results show that MLR was the best single model, followed by SVM, and then ELN, RF, GPR, and the last one was MLP. The MAE of SVM just differed 0.0005 from MLR, which may indicate that both MLR and SVM were the good single models in predicting perceived stress. However, the benchmark single model in this study is still MLR.

MLR is an effective method for prediction using data taken from survey and useful for modelling responses to survey questions as function of (external) sample data and/or other survey data (Benton et al., 2016, February). One possible reason why MLR outperformed other single models in this study is because the data used is a survey scales such as the Likert scale (low, moderate or high), in which the response of each questions is bounded by the perception of the respondents that is less complex and tends to be linear (Der & Deary, 2006).

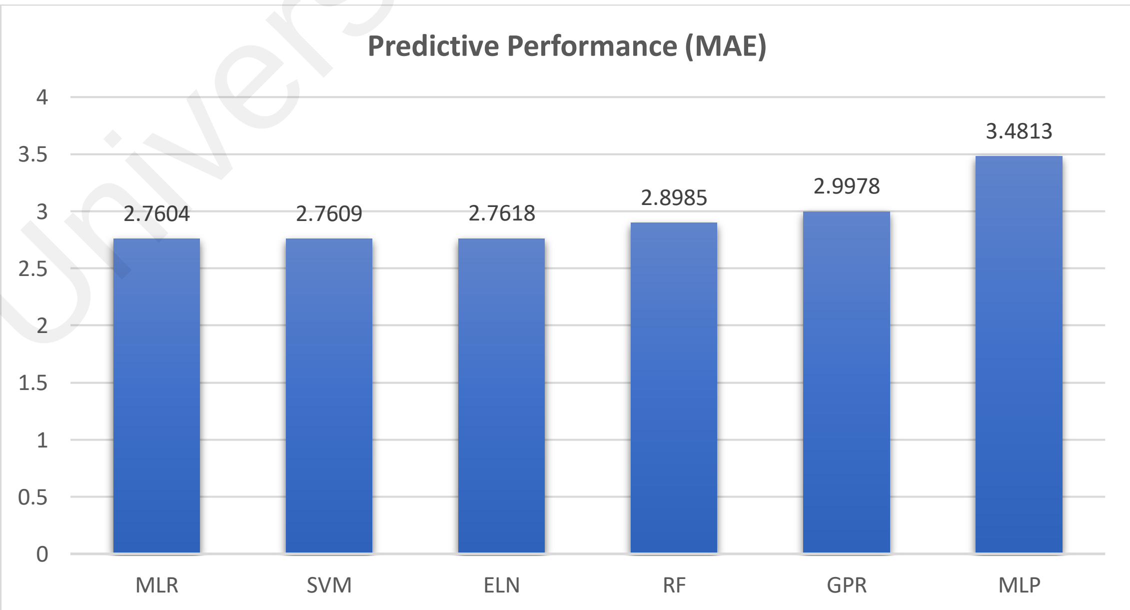


Figure 5.1: Predictive performances of the single models

5.3 Ensemble Regression Models

Following sub-sections discuss the predictive performances of the instances of each ensemble model used in this study.

5.3.1 Bagging (BG)

Figure 5.2 shows the predictive performances of the six BG ensemble instances built to predict the perceived stress. The results show that the best BG instance was BG-ELN (BG ensemble with ELN as base learner), however, the worst BG instance was BG-RF. The best BG-ELN instance achieved about 10.71% improvement in MAE over the worst BG-RF instance.

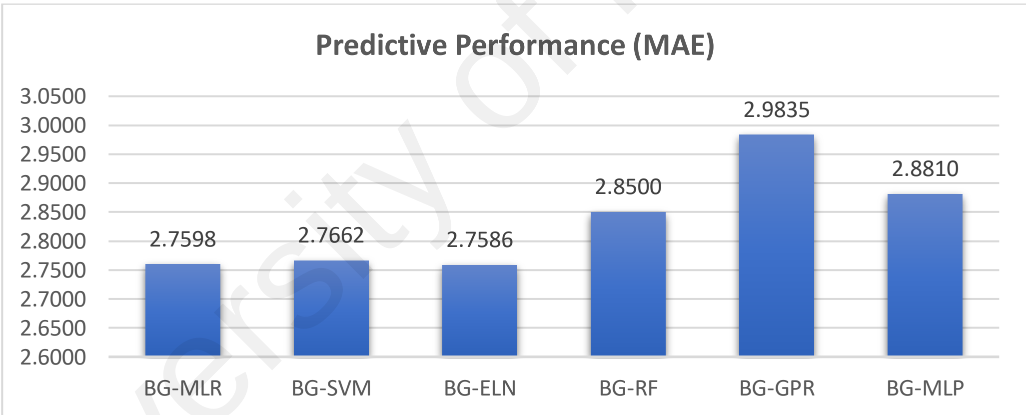


Figure 5.2: Predictive performances of the BG ensemble instances

5.3.2 Random Subspace (RSS)

Figure 5.3 shows the predictive performances of the six RSS ensemble instances built to predict the perceived stress. The results show that the best RSS instance was RSS-MLP (RSS ensemble with MLP as base learner), however, the worst RSS instance was RSS-GPR. The best RSS-MLP instance achieved about 12% improvement in MAE over the worst RSS-GPR instance.

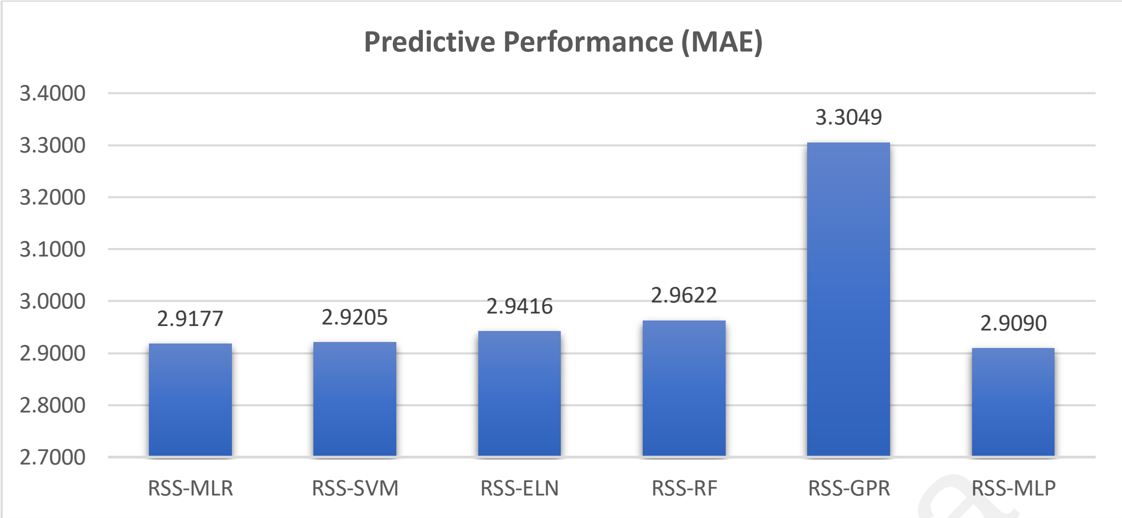


Figure 5.3: Predictive performances of the RSS ensemble instances

5.3.3 Additive Regression (AR)

Figure 5.4 shows the predictive performances of the six AR ensemble instances built to predict the perceived stress. The results show that the best AR instance was AR-SVM (AR ensemble with SVM as base learner), however, the worst AR instance was AR-MLP. The best AR-SVM instance achieved about 22.91% improvement in MAE over the worst AR-MLP instance.

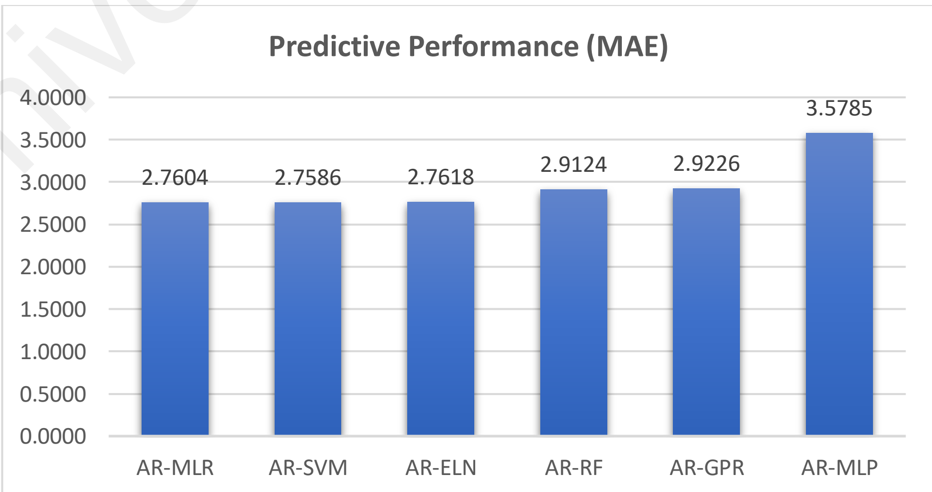


Figure 5.4: Predictive performances of the AR ensemble instances

5.3.4 Voting

Figure 5.5 shows the predictive performances of the six Voting ensemble instances (with different combinations of single models as base learners) built to predict the perceived stress. The results show that the best Voting instance was built with four base learners simultaneously (MLR, SVM, ELN, and RF), however, the worst Voting instance was built with all six base learners simultaneously. The best Voting instance achieved about 0.88% improvement in MAE over the worst Voting instance.

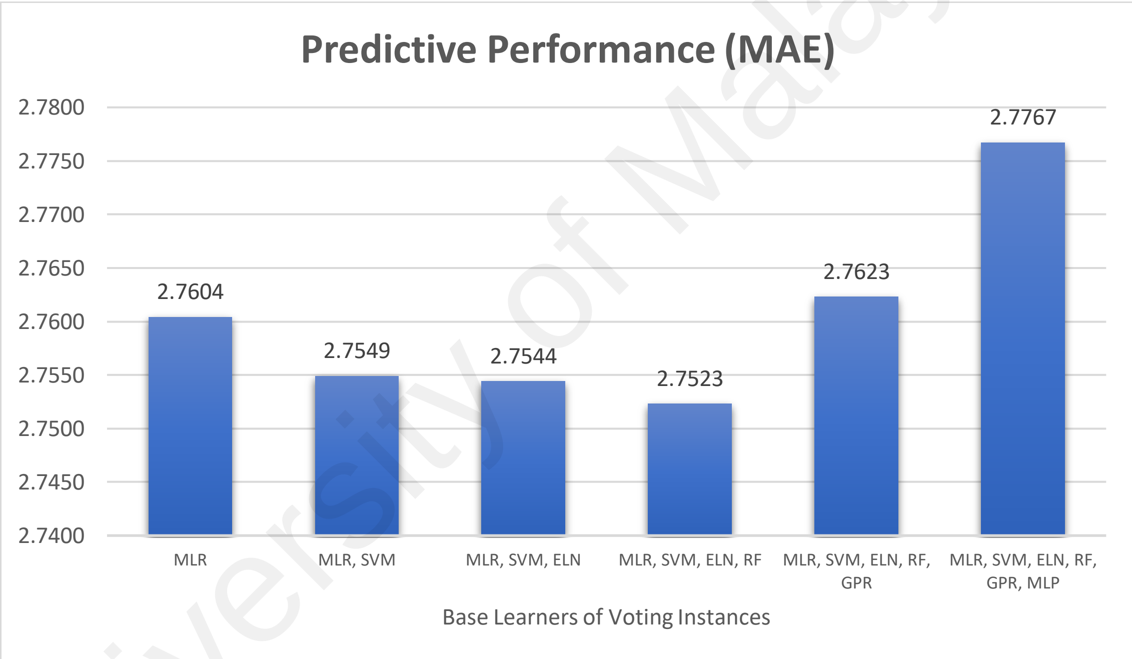
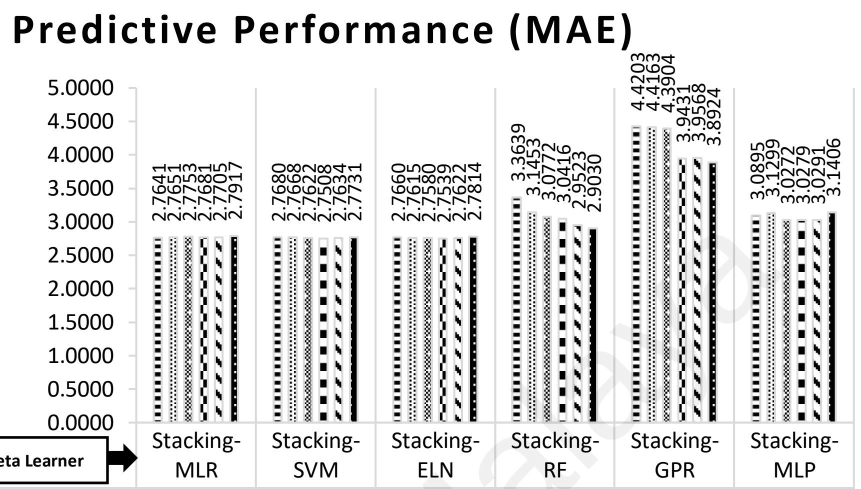


Figure 5.5: Predictive performances of the Voting ensemble instances

5.3.5 Stacking

Figure 5.6 shows the predictive performances of the 36 Stacking ensemble instances built to predict the perceived stress. The results show that the instances Stacking-MLR, Stacking-SVM, and Stacking-ELN categories were performing well while instances of Stacking-GPR category performed poorly. The best Stacking instance was Stacking-SVM with four single models (MLR, SVM, ELN, and RF) as base learners, while the worst Stacking instance was Stacking-GPR with only MLR as base learner. The best Stacking instance (MAE = 2.7508) achieved about 37.77% improvement in MAE over the worst Stacking instance (MAE = 4.4203).



Base Learners	Stacking-MLR	Stacking-SVM	Stacking-ELN	Stacking-RF	Stacking-GPR	Stacking-MLP
MLR	2.7641	2.7680	2.7660	3.3639	4.4203	3.0895
MLR, SVM	2.7651	2.7668	2.7615	3.1453	4.4163	3.1299
MLR, SVM, ELN	2.7753	2.7622	2.7580	3.0772	4.3904	3.0272
MLR, SVM, ELN, RF	2.7681	2.7508	2.7539	3.0416	3.9431	3.0279
MLR, SVM, ELN, RF, GPR	2.7705	2.7634	2.7622	2.9523	3.9568	3.0291
MLR, SVM, ELN, RF, GPR, MLP	2.7917	2.7731	2.7814	2.9030	3.8924	3.1406

Figure 5.6: Predictive performances of the Stacking ensemble instances

5.4 Comparison of the Predictive Performances between the Homogeneous Ensemble Models According to their Base Learners

Figure 5.7 shows the predictive performances of three types of homogeneous ensemble instances (BG, RSS, and AR) according to six base learners (MLR, SVM, ELN, RF, GPR, and MLP), as each homogeneous ensemble instance required to be built with a single model as base learner to predict the perceived stress. For the homogeneous ensemble instances that are built with the same base learner, BG instances performed best when built with ELN as base learner, while AR instances performed best when built with SVM as base learner. RSS instances did not perform outstandingly when built with any base learner. Eventually, the best homogeneous ensemble instances were AR-SVM and BG-ELN, of which both their MAEs are 2.7586. The worst homogeneous ensemble instance was AR-MLP, of which its MAE is 3.5785. The best homogeneous ensemble instances (AR-SVM and BG-ELN) achieved about 22.91% improvement in MAE over the worst homogeneous ensemble instance (AR-MLP).

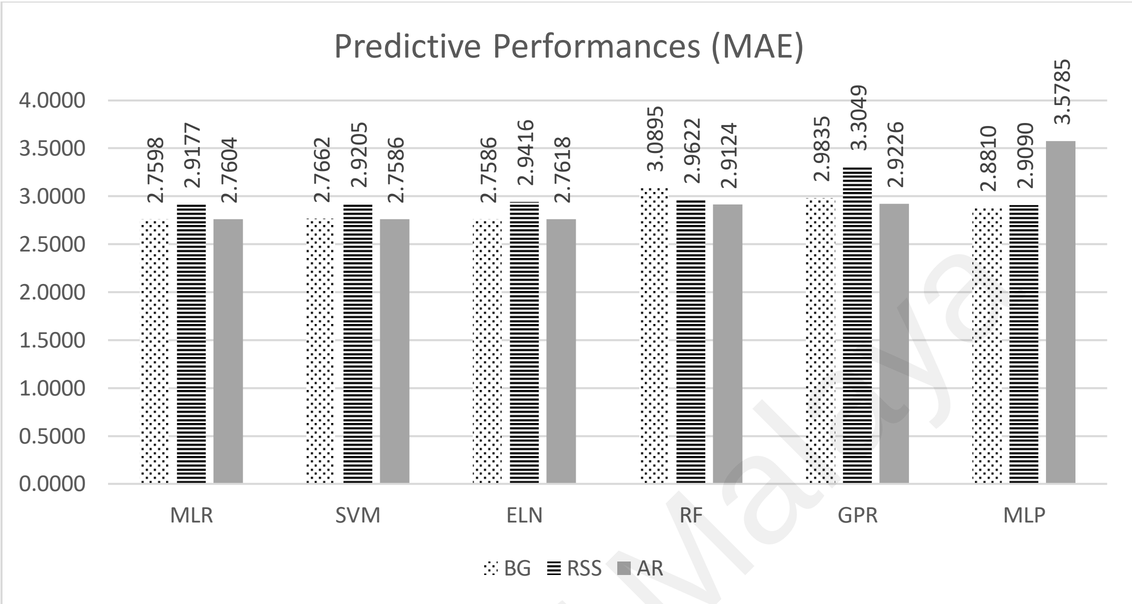


Figure 5.7: Predictive performances of the homogeneous ensemble instances according to six base learners

5.5 Comparison of the Predictive Performances between the Heterogeneous Ensemble Models According to their Base Learner Combinations

Figure 5.8 shows the predictive performances of two types of heterogeneous ensemble instances (Voting and Stacking) according to each of the six base learner combinations. There are six types of Stacking instances which each type of Stacking instances was built with one of the six meta learners (MLR, SVM, ELN, RF, GPR, and MLP). Therefore, Figure 5.8 shows seven types of heterogeneous ensemble instances (Voting, Stacking-MLR, Stacking-SVM, Stacking-ELN, Stacking-RF, Stacking-GPR, and Stacking-MLP) which required to be built with one of the six combinations of base learners to predict the perceived stress.

For the heterogeneous ensemble instances that built with the same combination of base learners, Voting instances performed best when built with first (MLR), second (MLR and SVM) or third (MLR, SVM, and ELN) combination of base learners, while Stacking-SVM instances performed best when built with forth (MLR, SVM, ELN, and RF) or sixth (MLR, SVM, ELN, RF, GPR, and MLP) combination of base learners, and Stacking-ELN instances performed best when built with fifth (MLR, SVM, ELN, RF, and GPR) combination of base learners. Stacking-GPR instances performed worst when built with any combination of base learners.

Eventually, the best heterogeneous ensemble instance was Stacking-SVM instance with four base learners (MLR, SVM, ELN, and RF), which its MAE is 2.7508. The worst heterogeneous ensemble instance was Stacking-GPR instance with only one base learner (MLR), which its MAE is 4.4203. The best heterogeneous ensemble instance achieved about 37.77% improvement in MAE over the worst heterogeneous ensemble instance.

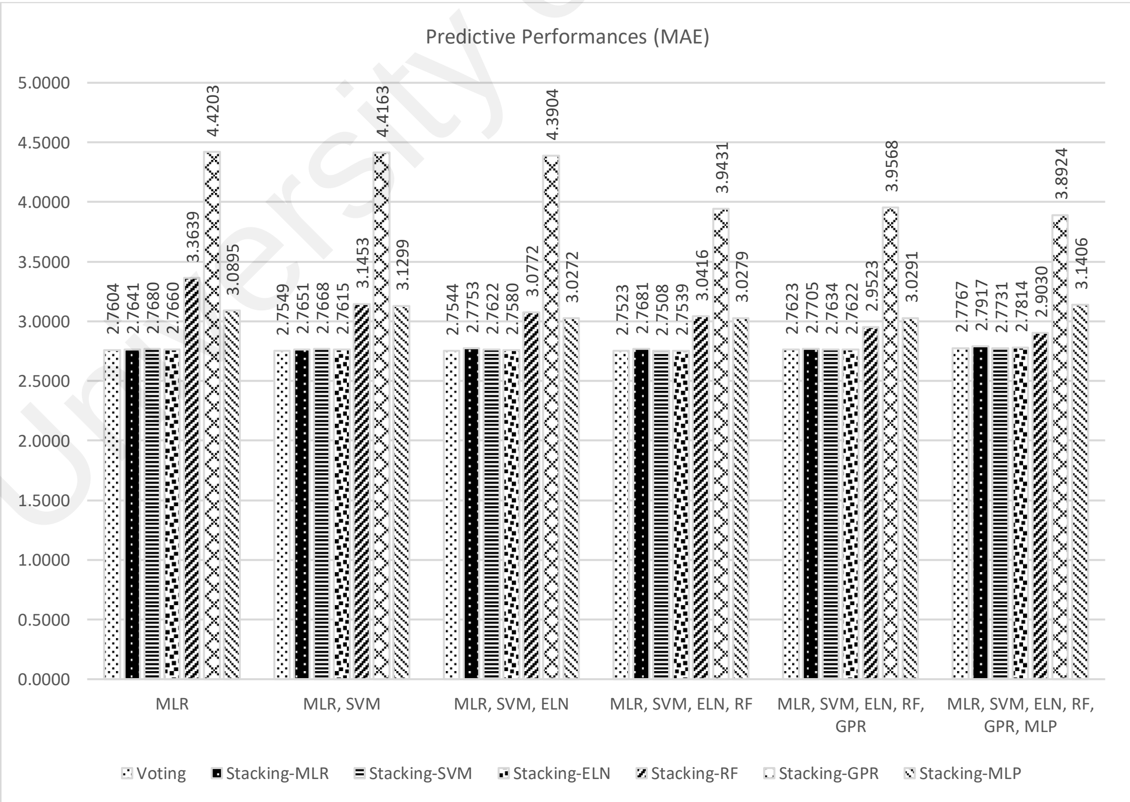


Figure 5.8: Predictive performances of the heterogeneous ensemble instances according to their base learner combinations

5.6 Comparison of the Predictive Performances between All the Regression Methods

Table 5.1 shows the rank positions of the predictive methods (single models and ensemble models with different base learners and meta learners) according to their predictive performances in MAE. Totally there were 66 methods. First six methods were only using the six single models respectively to predict perceived stress. The following predictive methods were using ensemble models, which started with homogeneous ensembles (BG, RSS, and AR) that required a single model as base learner, in turn. After that, the heterogeneous ensembles (Voting and Stacking) were built with each combination of base learners in turn, and with a single model as meta learner for the second level of training.

Table 5.1: Rank positions of the single and ensemble models in predicting perceived stress according to MAE

Ensemble Method	Meta Learner	Base Learners						MAE	Ranking
		MLR	SVM	ELN	RF	GPR	MLP		
-	-	✓						2.7604	9
-	-		✓					2.7609	10
-	-			✓				2.7618	12
-	-				✓			2.8985	31
-	-					✓		2.9978	42
-	-						✓	3.4813	54
BG	-	✓						2.7598	8
BG	-		✓					2.7662	19
BG	-			✓				2.7586	7
BG	-				✓			2.8500	29
BG	-					✓		2.9835	41
BG	-						✓	2.8810	30
RSS	-	✓						2.9177	35
RSS	-		✓					2.9205	36
RSS	-			✓				2.9416	38
RSS	-				✓			2.9622	40
RSS	-					✓		3.3049	52
RSS	-						✓	2.9090	33
AR	-	✓						2.7604	9
AR	-		✓					2.7586	7
AR	-			✓				2.7618	12
AR	-				✓			2.9124	34
AR	-					✓		2.9226	37
AR	-						✓	3.5785	55
Voting	-	✓						2.7604	9

Voting	-	✓	✓					2.7549	5
Voting	-	✓	✓	✓				2.7544	4
Voting	-	✓	✓	✓	✓			2.7523	2
Voting	-	✓	✓	✓	✓	✓		2.7623	14
Voting	-	✓	✓	✓	✓	✓	✓	2.7767	26
Stacking	MLR	✓						2.7641	16
Stacking	MLR	✓	✓					2.7651	17
Stacking	MLR	✓	✓	✓				2.7753	25
Stacking	MLR	✓	✓	✓	✓			2.7681	22
Stacking	MLR	✓	✓	✓	✓	✓		2.7705	23
Stacking	MLR	✓	✓	✓	✓	✓	✓	2.7917	28
Stacking	SVM	✓						2.7680	21
Stacking	SVM	✓	✓					2.7668	20
Stacking	SVM	✓	✓	✓				2.7622	13
Stacking	SVM	✓	✓	✓	✓			2.7508	1
Stacking	SVM	✓	✓	✓	✓	✓		2.7634	15
Stacking	SVM	✓	✓	✓	✓	✓	✓	2.7731	24
Stacking	ELN	✓						2.7660	18
Stacking	ELN	✓	✓					2.7615	11
Stacking	ELN	✓	✓	✓				2.7580	6
Stacking	ELN	✓	✓	✓	✓			2.7539	3
Stacking	ELN	✓	✓	✓	✓	✓		2.7622	13
Stacking	ELN	✓	✓	✓	✓	✓	✓	2.7814	27
Stacking	RF	✓						3.3639	53
Stacking	RF	✓	✓					3.1453	51
Stacking	RF	✓	✓	✓				3.0772	47
Stacking	RF	✓	✓	✓	✓			3.0416	46
Stacking	RF	✓	✓	✓	✓	✓		2.9523	39
Stacking	RF	✓	✓	✓	✓	✓	✓	2.9030	32
Stacking	GPR	✓						4.4203	61
Stacking	GPR	✓	✓					4.4163	60
Stacking	GPR	✓	✓	✓				4.3904	59
Stacking	GPR	✓	✓	✓	✓			3.9431	57
Stacking	GPR	✓	✓	✓	✓	✓		3.9568	58
Stacking	GPR	✓	✓	✓	✓	✓	✓	3.8924	56
Stacking	MLP	✓						3.0895	48
Stacking	MLP	✓	✓					3.1299	49
Stacking	MLP	✓	✓	✓				3.0272	43
Stacking	MLP	✓	✓	✓	✓			3.0279	44
Stacking	MLP	✓	✓	✓	✓	✓		3.0291	45
Stacking	MLP	✓	✓	✓	✓	✓	✓	3.1406	50

Overall, the results show that some ensemble instances with certain single model as meta learner and with certain combination of single models as base learners performed better than using the single model alone, for example, BG-ELN (MAE = 2.7586) performed better than ELN (MAE = 2.7618); AR-SVM (MAE = 2.7586) performed better than SVM (MAE = 2.7609); and BG-MLR (MAE = 2.7598) performed better than MLR (MAE = 2.7604). The best predictive method in predicting perceived stress was Stacking instance with four single models (MLR, SVM, ELN, and RF) as base learners and with SVM as meta learner (MAE = 2.7508), however, the worst predictive method was Stacking instance with MLR as base learner and with GPR as meta learner (MAE = 4.4203). The best predictive method achieved about 37.77% improvement in MAE over the worst predictive method. The instances with the same Stacking ensemble model but with different meta learner and base learners caused them to be the best and the worst predictive methods, therefore, choosing a set of the base and meta learners is very important.

Table 5.2: The top ten predictive methods in predicting perceived stress according to their predictive performances in MAE

Ensemble Method	Meta Learner	Base Learners						MAE	Ranking
		MLR	SVM	ELN	RF	GPR	MLP		
Stacking	SVM	✓	✓	✓	✓			2.7508	1
Voting	-	✓	✓	✓	✓			2.7523	2
Stacking	ELN	✓	✓	✓	✓			2.7539	3
Voting	-	✓	✓	✓				2.7544	4
Voting	-	✓	✓					2.7549	5
Stacking	ELN	✓	✓	✓				2.7580	6
BG	-			✓				2.7586	7
AR	-		✓					2.7586	7
BG	-	✓						2.7598	8
-	-	✓						2.7604	9
AR	-	✓						2.7604	9
Voting	-	✓						2.7604	9
-	-		✓					2.7609	10

Table 5.2 summarizes the results from Table 5.1 by only listing the top ten predictive methods according to the predictive performances in MAE in ascending order. Top six predictive methods were Stacking and Voting instances, this shows that Stacking and Voting ensemble models (both are heterogeneous ensembles) were the most suitable methods to predict perceived stress in this study. Following that were BG and AR

instances (both were homogeneous ensembles). The single models, MLR and SVM were ranked at ninth and tenth positions respectively. Coincidently, top three instances were all with MLR, SVM, ELN, and RF simultaneously as their base learners. This seems to show that MLR, SVM, ELN, and RF were the best combination of base learners. In other words, it is the concern about the suitability of the base learners for the ensemble models that improve the predictive performance, instead of the quantity of the base learners used to build the ensemble models. Besides, the most suitable meta learners for Stacking ensemble models were SVM and ELN.

5.7 Ensemble MAE Improvements over the Benchmark Single Model

Figure 5.9 shows the MAE improvement percentages of the ensemble instances over the benchmark single model, MLR. Due to many ensemble instances were built in this study and majority experienced MAE deterioration instead of improvement, therefore, only the predictive methods that gained MAE improvements over MLR are shown in Figure 5.9. The results show that Stacking-SVM with four base learners (MLR, SVM, ELN, and RF) simultaneously achieved the highest MAE improvement (0.35%) over the benchmark single model (MLR), following by other Voting and Stacking instances. Besides the heterogeneous ensemble instances, the AR and BG instances from homogeneous ensembles also achieved a little MAE improvement over the benchmark single model. In this study, heterogeneous ensemble models were proved to perform better than homogeneous ensemble models and single models.

Likewise, Aldave and Dussault (2014) also found the similar results from their experiments, which their best stacking model has outperformed other ensemble models, however it could only perform as good as the best of the single models. This shows that more future work is needed to improve the accuracy of ensemble models in regression problems, as Mendes-Moreira et al. (2012) has mentioned that the benefits of ensembles with respect to single models has been reported not only in terms of increased accuracy but also robustness.

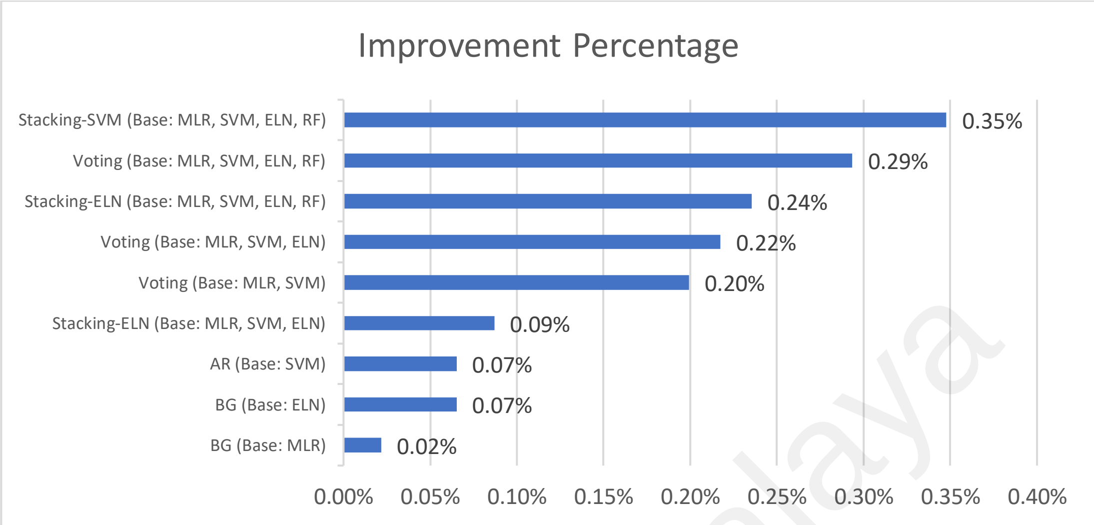


Figure 5.9: Ensemble MAE improvement over the MLR model

5.8 Chapter Summary

This chapter has compared and discussed the predictive performances (MAE) of various regression methods (including single models and ensemble models) used to predict the perceived stress using relevant personality traits. MLR outperformed the other five single models (SVM, ELN, RF, GPR, and MLP) and be the benchmark single model in the current study. Among all the ensemble models that were being tested in this study, Stacking is the most suitable ensemble model for the prediction of perceived stress. The Stacking instance with four single models (MLR, SVM, ELN, and RF) simultaneously as base learners and with SVM as meta learner managed to achieve the highest MAE improvement over the benchmark single model (MLR). Overall, the results show that using the suitable ensemble model with the suitable base and meta learners could perform better than just using a single model in predicting perceived stress with the relevant personality traits.

CHAPTER 6: CONCLUSION AND RECOMMENDATION

This chapter presents the summary (by revisiting the research objectives), conclusion (contribution and significance), implications from the current study, and recommendations for future work.

6.1 Summary

The present study aimed to identify the most accurate regression methods (including the single and ensemble models) that can be used to predict perceived stress with relevant personality traits. This study was conducted using the real dataset (429 sample size after data cleaning) collected from the survey that was distributed to the general public in Melbourne, Australia and its surrounding districts. The dataset contains a variety of validated scales that measuring personality traits as well as perceived stress. The summary of the findings according to the study objectives is shown below:

<u>Objective 1: To identify the personality traits that are relevant in predicting the perceived stress scale.</u>

As discussed in Section 5.1, the personality traits that were relevant to the perceived stress of the respondents in this study were mastery, positive affect, negative affect, life satisfaction, self-esteem, and PCOIS. Only negative affect was positively associated with perceived stress, the rest of the personality traits were all negatively associated with perceived stress. Besides, gender was the only background characteristic that was relevant to the perceived stress whereby females reporting more perceived stress than males.

Objective 2: To determine the most suitable single regression model for predicting perceived stress scale to be used as the benchmark.

There were six commonly used single regression models (MLR, SVM, ELN, RF, GPR, and MLP) identified from the literature. In the current study, MLR performed better than other five single models in term of predictive performance (MAE) in predicting perceived stress, therefore, it was chosen as the benchmark single model for predicting perceived stress with the relevant personality traits (refer to Section 5.2 for the details).

Objective 3: To identify and develop a suitable ensemble regression model for improving the prediction of perceived stress using relevant personality traits.

There were six commonly used ensemble models (BG, RSS, AR, Voting, and Stacking) identified from the literature. In the current study, Stacking and Voting were the most suitable ensemble models for predicting perceived stress because the top six predictive performances were all achieved by the Stacking (with different meta learner) and Voting instances with different combinations of single models as base learners. The best predictive performance was achieved by Stacking instance with four single models (MLR, SVM, ELN, and RF) simultaneously as base learners and with SVM as meta learner (refer to Section 5.6 for the details).

Objective 4: To compare the prediction performances of the proposed ensemble regression models with the benchmark single model.

Overall, there were 66 predictive methods being built from the single and ensemble models in this study. When comparing the predictive performances of all the methods in this study, the Stacking instance with four single models (MLR, SVM, ELN, and RF) simultaneously as base learners and with SVM as meta learner managed to achieve the highest MAE improvement over the benchmark single model (MLR). Besides, some Voting, Stacking, AR, and BG instances with different combination of base learners also out-performed benchmark single model. The results show that using the suitable ensemble model with a set of base and meta learners could perform better than just using a single model in predicting perceived stress with the relevant personality traits (refer to Section 5.7 for the details).

6.2 Conclusion

The past studies that related to the prediction of stress were mostly focused on the classification problems, where classification models were used to solve the problems. However, for the prediction of perceived stress (regression problem), the questions regarding what single regression models could perform better and whether the ensemble regression models can outperform the single models were left unknown. This study was designed to answer those questions and the goals were achieved by conducting a comprehensive comparison study between the predictive performances of all the commonly used single regression models and ensemble models identified from the literature. The results of this study fill the gaps of the stress related research so that the solution will not be limited only to classification models, but also using single regression models and more advancing ensemble regression models to improve the prediction performances. Moreover, this study also provides important insights into the roles of relevant personality traits as the predictors of perceived stress.

6.3 Implications and Contributions

It is important to understand the predictors of perceived stress, once the predictors are established, they can reflect how the perception of stress originates, and can be used to motivate interventions to resist stress. The identified best ensemble model in this study can be used to build a more accurate predictive method for the development of perceived stress prediction system using the relevant personality traits as predictors. The system can be embedded in different devices and being used by many parties to help their employees, organization members or their loved ones to gain better mental health by executing the early prevention actions once the targeted persons were being predicted with a high degree of perceived stress. It could prevent the consequences of the negative stress from silently developing to severe stages which could affect the physical and mental health silently. Hence, many mental health problems such as depression, hopelessness, and suicidal ideation can be prevented.

Besides, the proposed experimental design of this study (especially the development of the predictive methods and the comparison study between the predictive performances of all the methods) can be extended to the regression problems of other research fields which the implementations of the ensemble models may help to improve their predictive performances. The researchers from the social or behavioral science research could relate stress-related studies to the identified personality traits and develop good predictive models to achieve accurate predictions.

6.4 Limitations and Recommendations

There are several recommendations that could be given to the future researchers based on the limitations in the present study. In this study, only several single models and ensemble models were being compared, and only gender and six personality traits were being selected as the predictors of perceived stress. Future studies should include other potential single models and ensemble models as well as more relevant personality traits and other attributes to build better predictive methods. Future research should also use more datasets from the general populations of different countries to verify the findings, because the datasets from different countries may give different results. There was very little improvement of predictive accuracy for the best ensemble instance over the benchmark single model in this study, therefore, future research should improve or propose a better ensemble model to improve the predictive accuracy over the benchmark single model.

For the researchers who want to develop the perceived stress prediction system, such as the application in the mobile device, it may cause the participants to feel burdensome and obtrusive easily when the participants are being interrupted to complete a self-report which consists of many items (demographic questionnaire and 8 scales, which each scale consists of 5 to 20 items). Besides, recall bias and social desirability bias might occur because the measures were conducted in the form of self-report. Therefore, future researchers are encouraged to develop the measurements with lesser items but without decreasing its capability in measuring the constructs. If the sensory devices can replace the measurements to measure personality traits, that would be a great help for the prediction of perceived stress.

REFERENCES

Abolghasemi, A., & Varaniyab, S. T. (2010). Resilience and perceived stress: predictors of life satisfaction in the students of success and failure. *Procedia-Social and Behavioral Sciences*, 5, 748-752.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Aldave, R., & Dussault, J. P. (2014). Systematic ensemble learning for regression. ArXiv Preprint ArXiv:1403.7267.

Al-Abri, E. S. (2016). *Modelling atmospheric ozone concentration using machine learning algorithms* (Doctoral dissertation, Loughborough University).

Aneshensel, C. S., & Avison, W. R. (2015). The stress process: An appreciation of Leonard I. Pearlin. *Society and Mental Health*, 5(2), 67-85.

Aneshensel, C. S., & Mitchell, U. A. (2014). The stress process: its origins, evolution, and future. In *Sociology of Mental Health* (pp. 53-74). Springer International Publishing.

Bańczyk, K., Kempa, O., Lasota, T., & Trawiński, B. (2011, April). Empirical comparison of bagging ensembles created using weak learners for a regression problem. In *Asian Conference on Intelligent Information and Database Systems* (pp. 312-322). Springer, Berlin, Heidelberg.

Benton, A., Paul, M. J., Hancock, B., & Dredze, M. (2016, February). Collective Supervision of Topic Models for Predicting Surveys with Social Media. In *Association for the Advancement of Artificial Intelligence* (pp. 2892-2898).

Blevins, C. L., Sagui, S. J., & Bennett, J. M. (2017). Inflammation and positive affect: Examining the stress-buffering hypothesis with data from the National Longitudinal Study of Adolescent to Adult Health. *Brain, Behavior, and Immunity*, 61, 21-26.

Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. S. (2014a). Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM International Conference on Multimedia* (pp. 477-486). Association for Computing Machinery.

Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. S. (2014b). Pervasive stress recognition for sustainable living. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on* (pp. 345-350). IEEE.

Brassington, G. (2017, April). Mean absolute error and root mean square error: which is the better metric for assessing model performance?. In *EGU General Assembly Conference Abstracts* (Vol. 19, p. 3574).

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Bretherton, S. J., & McLean, L. A. (2015). Interrelations of stress, optimism and control in older people's psychological adjustment. *Australasian Journal on Ageing*, 34(2), 103-108.

Brown, G. W., & Harris, T. O. (Eds.). (1989). *Life events and illness*. Guilford Press.

Brüggemann, C., & Santos, F. (2016). When Stress at Work Becomes a Problem: Coping Approaches Told by Middle Managers.

Butler, G. (1993). Definitions of stress. *Occasional paper (Royal College of General Practitioners)*, (61), 1.

Burke, S. (2017). *Investigation of Pigmentation-Associated Polymorphisms in a Novel Indigenous American Population* (Doctoral dissertation).

Cannon, W. B. (1929). Bodily changes in pain, fear, hunger, and rage. *New York, Appleton*.

Carey, M. P., Kalra, D. L., Carey, K. B., Halperin, S., & Richards, C. S. (1993). Stress and unaided smoking cessation: A prospective investigation. *Journal of Consulting and Clinical Psychology*, 61(5), 831-838.

Chan, P. K., & Stolfo, S. J. (1997). On the accuracy of meta-learning for scalable data mining. *Journal of Intelligent Information Systems*, 8(1), 5-28.

Chang, E. C. (1998). Dispositional optimism and primary and secondary appraisal of a stressor: Controlling for confounding influences and relations to coping and psychological and physical adjustment. *Journal of Personality and Social Psychology*, 74(4), 1109.

Chowdary, E. D., Devi, K. A., Mounika, D., Venkatramaphanikumar, S., & Kishore, K. K. (2016, August). Ensemble classification technique to detect stress in IT-professionals. In *Inventive Computation Technologies (ICICT), International Conference on* (Vol. 3, pp. 1-6). IEEE.

Ciarrochi, J., Deane, F. P., & Anderson, S. (2002). Emotional intelligence moderates the relationship between stress and mental health. *Personality and Individual Differences*, 32(2), 197-209.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.

Cohen, S., Gianaros, P. J., & Manuck, S. B. (2016). A stage model of stress and disease. *Perspectives on Psychological Science*, 11(4), 456-463.

Cohen S, Kessler RC, Gordon LU. Strategies for measuring stress in studies of psychiatric and physical disorders. In: Cohen S, Kessler RC, Lynn UG, editors. Measuring stress: A guide for health and social scientists. New York: Oxford University Press; 1995. pp. 3–26.

Cohen, S., Tyrrell, D. A., & Smith, A. P. (1993). Negative life events, perceived stress, negative affect, and susceptibility to the common cold. *Journal of Personality and Social Psychology*, 64(1), 131.

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the US In S. Spacapam & S. Oskamp (Eds.), The social psychology of health: Claremont symposium on applied social psychology (pp. 31-67).

Cox, T., & Macay, C. J. (1976). Psychological model of Occupational stress, A Paper presented to medical research council meeting. *Mental health in Industry, London. November. Nakajima, H (1994) Stress and Health, World Health Magazine*, 47(2), 2.

Curtis, R., Groarke, A., Coughlan, R., & Gsel, A. (2004). The influence of disease severity, perceived stress, social support and coping in patients with chronic illness: a 1 year follow up. *Psychology, Health & Medicine*, 9(4), 456-475.

Dapeng, Z. (2017). 79. Analysis of Micro-blog Comments Tendentious Vector Based on Machine Learning. *Boletín Técnico, ISSN: 0376-723X, 55*(17).

Decision tree learning. (2017, September 29). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:39, September 30, 2017, from https://en.wikipedia.org/w/index.php?title=Decision_tree_learning&oldid=802954821

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, 21(1), 62.

Deshpande, B. (2012, December 11). How to apply feature selection in linear regression with RapidMiner. Retrieved February 04, 2018, from http://www.simafore.com/blog/bid/111764/How-to-apply-feature-selection-in-linear-regression-with-RapidMiner

Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4), 407-422.

Diener, E., Emmons, R.A., Larson, R.J. & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49, 71-6.

Dobkin, P. L., Zhao, Q., & Monshat, K. (2017). Who experiences depressive symptoms following mindfulness-based stress reduction and why? *International Journal of Whole Person Care*, 4(1), 13-31.

Elite Data Science. (2017, September 16). Modern Machine Learning Algorithms: Strengths and Weaknesses. Retrieved September 27, 2017, from https://elitedatascience.com/machine-learning-algorithms

Ezzati, A., Jiang, J., Katz, M. J., Sliwinski, M. J., Zimmerman, M. E., & Lipton, R. B. (2014). Validation of the Perceived Stress Scale in a community sample of older adults. *International journal of geriatric psychiatry*, 29(6), 645-652.

Feldman, P. J., Cohen, S., Hamrick, N., & Lepore, S. J. (2004). Psychological stress, appraisal, emotion and cardiovascular response in a public speaking task. *Psychology & Health*, 19(3), 353-368.

Fernandez, R., & Picard, R. W. (2003). Modeling drivers' speech under stress. *Speech Communication*, 40(1), 145-159.

Fontama, V., Barga, R., & Tok, W. H. (2015). *Predictive Analytics with Microsoft Azure Machine Learning 2nd Edition*. Apress.

Forkuor, G., Hounkpatin, O. K., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in South-Western Burkina Faso: A comparison of machine learning and multiple linear regression models. *PloS One*, 12(1), e0170478.

Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA workbench. *Data mining: Practical machine learning tools and techniques. Burlington: Morgan Kaufmann*.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.

Garber, M. C. (2017). Exercise as a stress coping mechanism in a pharmacy student population. *American Journal of Pharmaceutical Education*, 81(3), 50.

Gebhart, G. F., & Schmidt, R. F. (2013). *Encyclopedia of Pain*. Springer Reference.

Gellman, M. D., & Turner, J. R. (2013). *Encyclopedia of Behavioral Medicine*. Springer Reference.

Gianaros, P. J., & Wager, T. D. (2015). Brain-body pathways linking psychological stress and physical health. *Current Directions in Psychological Science*, 24(4), 313-321.

Graczyk, M., Lasota, T., & Trawiński, B. (2009). Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA. *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, 800-812.

Graczyk, M., Lasota, T., Telec, Z., & Trawiński, B. (2010). Nonparametric statistical analysis of machine learning algorithms for regression problems. *Knowledge-Based and Intelligent Information and Engineering Systems*, 111-120.

Gravetter, F. J., & Forzano, L. A. B. (2018). *Research methods for the behavioral sciences*. Cengage Learning.

Greer, T. M. (2008). Racial and ethnic-related stressors as predictors of perceived stress and academic performance for African American students at a historically Black college and university. *The Journal of Negro Education*, 60-71.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10-18.

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.

Hawkins, N. G., Davies, R., & Holmes, T. H. (1957). Evidence of psychosocial factors in the development of pulmonary tuberculosis. *American review of tuberculosis*, 75(5), 768.

Heinze, J. E., Stoddard, S. A., Aiyer, S. M., Eisman, A. B., & Zimmerman, M. A. (2017). Exposure to violence during adolescence as a predictor of perceived stress trajectories in emerging adulthood. *Journal of applied developmental psychology*, 49, 31-38.

Herbert, T. B., & Cohen, S. (1996). Measurement issues in research on psychosocial stress. *Psychosocial Stress: Perspectives on Structure, Theory, Life-course, and Methods*, 295-332.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844.

Isa, S. N. I., Ishak, I., Ab Rahman, A., Saat, N. Z. M., Din, N. C., Lubis, S. H., & Ismail, M. F. M. (2017). Perceived Stress and Coping Styles among Malay Caregivers of Children with Learning Disabilities in Kelantan. *The Malaysian journal of medical sciences: MJMS*, 24(1), 81.

Karagiannopoulos, M., Anyfantis, D., Kotsiantis, S. B., & Pintelas, P. E. (2007). Feature selection for regression problems. *Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications, Athens, Greece*, 2022.

King, M. A., Abrahams, A. S., & Ragsdale, C. T. (2014). Ensemble methods for advanced skier days prediction. *Expert Systems with Applications*, 41(4), 1176-1188.

Nuñez-Gonzalez, J. D., & Graña, M. (2015, June). Graph-based learning on sparse data for recommendation systems in social networks. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 61-68). Springer, Cham.

Holmes, T. H., & Masuda, M. (1974). Life changes and illness susceptibility. In *B. S. Dohrenwend & B. P. Dohrenwend (Eds.), Stressful life events: Their nature and effects* (pp. 45–72). New York, NY: Wiley.

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, 11(2), 213-218.

Jiang, M., Lan, M., & Wu, Y. (2017). ECNU at SemEval-2017 Task 5: An Ensemble of Regression Algorithms with Effective Features for Fine-Grained Sentiment Analysis in Financial Domain. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 888-893).

Kamarck, T. W., Muldoon, M. F., Shiffman, S. S., & Sutton-Tyrrell, K. (2007). Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men. *Health Psychology*, 26(3), 324.

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., & Amick, B. (1998). The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology*, 3, 322-355.

Karhunen, J. (2011). Robust PCA methods for complete and missing data. *Neural Network World*, 21(5), 357-392.

Keshtegar, B., Mert, C., & Kisi, O. (2018). Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renewable and Sustainable Energy Reviews*, 81, 330-341.

Khair, A. F., Awang, M. K., Zakaraia, Z. A., & Mazlan, M. (2017). Daily streamflow prediction on time series forecasting. *Journal of Theoretical and Applied Information Technology*, 95(4), 804-809.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Lazarus, R. S. (1966). Psychological stress and the coping process. *New York, NY: McGraw-Hill*.

Lazarus, R. S. (1976). Patterns of adjustment (3rd ed.). *New York, NY: McGraw-Hill*.

Lazarus, R. S. (2006). *Stress and emotion: A new synthesis*. Springer Publishing Company.

Lazarus, R. S., Deese, J., & Osler, S. F. (1952). The effects of psychological stress upon performance. *Psychological bulletin*, 49(4), 293-317.

Lazarus, R. S., & Eriksen, C. W. (1952). Effects of failure stress upon skilled performance. *Journal of Experimental Psychology*, 43(2), 100-105.

Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. *New York: Springer*.

Lazarus, R. S., Speisman, J. C., Mordkoff, A. M., & Davison, L. A. (1962). A laboratory study of psychological stress produced by a motion picture film. *Psychological Monographs: General and Applied*, 76(34), 1-35.

Lebois, L. A., Hertzog, C., Slavich, G. M., Barrett, L. F., & Barsalou, L. W. (2016). Establishing the situated features associated with perceived stress. *Acta psychologica*, 169, 119-132.

Lepore, S. J. (1997). Measurement of chronic stressors. *Measuring stress: A guide for health and social scientists*, 102-120.

Levi, L. (1971). Society, stress and disease. Vol. 1. The psyehosocial environment and psychosomatic diseases. Proceedings of an International Interdisciplinary Symposium held in Stockholm, April 1970. In *Society, stress and disease. Vol. 1.*

The psyehosocial environment and psychosomatic diseases. Proceedings of an International Interdisciplinary Symposium held in Stockholm, April 1970.. New York, Toronto, London: Oxford University Press, 37 Dover Street, W. 1..

Lichtenberg, J. M., & Şimşek, Ö. (2017, August). Simple Regression Models. In *Imperfect Decision Makers: Admitting Real-World Rationality* (pp. 13-25).

Liu, H., Gopalkrishnan, V., Quynh, K. T. N., & Ng, W. K. (2009). Regression models for estimating product life cycle cost. *Journal of Intelligent Manufacturing*, 20(4), 401-408.

Liu, M., Shang, Z., & Cheng, G. (2017). How Many Machines Can We Use in Parallel Computing for Kernel Ridge Regression? *Indiana University–Purdue University Indianapolis*, 1-25.

Major, R. L., & Ragsdale, C. T. (2001). Aggregating expert predictions in a networked environment. *Computers and Operations Research*, 28(12), 1231–1244.

Marshall, G. N., Davis, L. M., Sherbourne, C. D., & Morland, L. (2000). *A Review of the Scientific Literature as it Pertains to Gulf War Illnesses: Stress*. RAND.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, 338(3), 171-179.

Mendes-Moreira, J., Soares, C., Jorge, A. M., & Sousa, J. F. D. (2012). Ensemble approaches for regression: A survey. *ACM Computing Surveys (CSUR)*, 45(1), 10-40.

Michalos, A. C. (2014). *Encyclopedia of quality of life and well-being research*. Springer Reference.

Moh, J. (2016, September 19). Suicide trend worry in urban areas. New Straits Times. Retrieved from https://www.nst.com.my/news/2016/09/174207/suicide-trend-worry-urban-areas

Moniz, N., Branco, P., & Torgo, L. (2017, October). Evaluation of Ensemble Methods in Imbalanced Regression Tasks. *In First International Workshop on Learning with Imbalanced Domains: Theory and Applications* (pp. 129-140).

Monroe, S. M., & Kelley, J. M. (1997). Measurement of stress appraisal. In *S. Cohen, R. C. Kessler, & L. U. Gordon (Eds.), Measuring stress: A guide for health and social scientists* (pp. 122-147). New York: Oxford University Press.

Moon, H. J., Seo, J. G., & Park, S. P. (2016). Perceived stress and its predictors in people with epilepsy. *Epilepsy & Behavior*, 62, 47-52.

Morgan, A. L. (2014). *A comparison of coping and perceived stress among students by first generation and continuing generation status, race/ethnicity, and gender*. East Carolina University.

Murphy, A. (2015). *The use of ensemble techniques in multiclass speech emotion recognition to improve both accuracy and confidence in classifications* (Doctoral dissertation).

Muscatell, K. A., & Eisenberger, N. I. (2012). A social neuroscience perspective on stress and health. *Social and Personality Psychology Compass*, 6(12), 890-904.

Ng, D. M., & Jeffery, R. W. (2003). Relationships Between Perceived Stress and Health Behaviors in a Sample of Working Adults. *Health Psychology*, 22(6), 638-642.

Nomikos, M. S., Opton Jr, E., & Averill, J. R. (1968). Surprise versus suspense in the production of stress reaction. *Journal of Personality and Social Psychology*, 8(2p1), 204-208.

Nwoke, M. B., Onuigbo, E. N., & Odo, V. O. (2017). Social support, self-efficacy and gender as predictors of reported stress among inpatient caregivers. *The Social Science Journal*, 54(2), 115-119.

Okely, J. A., Weiss, A., & Gale, C. R. (2017). The interaction between stress and positive affect in predicting mortality. *Journal of Psychosomatic Research*, 100, 53-60.

Pallant, J. F. (2000). Development and validation of a scale to measure perceived control of internal states. *Journal of Personality Assessment*, 75(2), 308-337.

Pallant, J. F. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).

Pallant, J. F., & Lae, L. (2002). Sense of coherence, well-being, coping and personality factors: further evaluation of the sense of coherence scale. *Personality and Individual Differences*, 33(1), 39-48.

Pearlin, L. I. (1999). *The stress process revisited. In Handbook of the sociology of mental health* (pp. 395-415). Springer US.

Pearlin, L. I., & Bierman, A. (2013). Current issues and future directions in research into the stress process. In *Handbook of the sociology of mental health* (pp. 325-340). Springer Netherlands.

Pearlin, L. I., Menaghan, E. G., Lieberman, M. A., & Mullan, J. T. (1981). The stress process. *Journal of Health and Social behavior*, 337-356.

Pearlin, L. I., & Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior*, 19, 2–21.

Pham, B. T., Prakash, I., & Bui, D. T. (2017). *Spatial prediction of landslides using hybrid machine learning approach based on Random Subspace and Classification and Regression Trees*. Geomorphology.

Plarre, K., Raij, A., Hossain, S. M., Ali, A. A., Nakajima, M., Al'absi, M., ... & Siewiorek, D. (2011). Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on* (pp. 97-108). IEEE.

Poole, D. L., & Mackworth, A. K. (2010). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press.

Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health? *Psychological Bulletin*, 131(6), 925.

Rahe, R. H., & Lind, E. (1971). Psychosocial factors and sudden cardiac death: A pilot study. *Journal of Psychosomatic Research*, 15(1), 19-24.

Rebbeck, T. R., Weber, A. L., Spangler, E., & Zeigler-Johnson, C. M. (2013). What stresses men? Predictors of perceived stress in a population-based multi-ethnic cross sectional cohort. *BMC public health*, 13(1), 113-121.

Remeseiro, B., Barreira, N., Sánchez-Brea, L., Ramos, L., & Mosquera, A. (2018). Machine Learning Applied to Optometry Data. In *Advances in Biomedical Informatics* (pp. 123-160). Springer, Cham.

Ren, Y., Suganthan, P. N., & Srikanth, N. (2015). Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 50, 82-91.

Rey, L., & Extremera, N. (2015). Core self-evaluations, perceived stress and life satisfaction in Spanish young and middle-aged adults: An examination of mediation and moderation effects. *Social Indicators Research*, 120(2), 515-524.

Pérez, A., Chien, L. C., Harrell, M. B., Pasch, K. E., Obinwa, U. C., & Perry, C. L. (2017). Geospatial associations between tobacco retail outlets and current use of cigarettes and e-cigarettes among youths in Texas. *Journal of Biometrics & Biostatistics*, 8(5), 375-380.

Plarre, K., Raij, A., Hossain, S. M., Ali, A. A., Nakajima, M., Al'absi, M., ... & Siewiorek, D. (2011, April). Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on* (pp. 97-108). IEEE.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27(2), 151-161.

Robles, Z., Garey, L., Hogan, J., Bakhshaie, J., Schmidt, N. B., & Zvolensky, M. J. (2016). Examining an underlying mechanism between perceived stress and smoking cessation-related outcomes. *Addictive Behaviors*, 58, 149–154.

Rogers, A. H., Seager, I., Haines, N., Hahn, H., Aldao, A., & Ahn, W. Y. (2017). *Minority stress, emotion regulation, and substance misuse in sexual minorities* (Doctoral dissertation, Department of Psychology, The Ohio State University).

Rosenberg, M. (1965). *Society and the adolescent self image*. Princeton, NJ: Princeton University Press.

Rosellini, A. J., Dussaillant, F., Zubizarreta, J. R., Kessler, R. C., & Rose, S. (2018). Predicting posttraumatic stress disorder following a natural disaster. *Journal of Psychiatric Research*, 96, 15-22.

Rule-based machine learning. (2017, March 18). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:26, September 30, 2017, from https://en.wikipedia.org/w/index.php?title=Rule-based_machine_learning&oldid=770852385

Salvador, A. (2005). Coping with competitive situations in humans. *Neuroscience & Biobehavioral Reviews*, 29(1), 195-205.

Schaefer, J. D., Caspi, A., Belsky, D. W., Harrington, H., Houts, R., Horwood, L. J., ... & Moffitt, T. E. (2017). Enduring mental health: Prevalence and prediction. *Journal of Abnormal Psychology*, 126(2), 212-224.

Scheier, M.F. & Carver, C.S. (1985). Optimism, coping and health: An assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219–47.

Scheier, M.F., Carver, C.S. & Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 6, 1063–1078.

Scherer, S., Hofmann, H., Lampmann, M., Pfeil, M., Rhinow, S., Schwenker, F., & Palm, G. (2008, May). Emotion Recognition from Speech: Stress Experiment. In *LREC*.

Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X. C., & Margraf, J. (2016). The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. *International Journal of Clinical and Health Psychology*, 16(1), 1-10.

Schulz, H., Vögele, C., & Meyer, B. (2009). Optimism, self-efficacy, and perceived stress as predictors of self-reported health symptoms in college students. *Zeitschrift für Gesundheitspsychologie*, 17(4), 185-194.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature*, 138(3479), 32.

Selye, H. (1974). Stress without distress. *New York*, 26-39.

Selye, H. (1976). *Stress in Health and Disease*. Stoneham, MA: Butterworth.

Selye, H. (1978). *The Stress of Life: Rev Ed*. McGraw-Hill.

Severson, K. A., Molaro, M. C., & Braatz, R. D. (2017). Principal Component Analysis of Process Datasets with Missing Values. *Processes*, 5(3), 38.

Shah, M., Hasan, S., Malik, S., & Sreeramareddy, C. T. (2010). Perceived stress, sources and severity of stress among medical undergraduates in a Pakistani medical school. *BMC Medical Education*, 10(1), 2-9.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, 108(3), 1287-1301.

Simpson, C., & Carter, P. (2013). Short-term changes in sleep, mastery & stress: Impacts on depression and health in dementia caregivers. *Geriatric Nursing*, 34, 509-516.

Smets, E., Casale, P., Großekathöfer, U., Lamichhane, B., De Raedt, W., Bogaerts, K., ... & Van Hoof, C. (2015, September). Comparison of machine learning techniques for psychophysiological stress detection. In *International Symposium on Pervasive Computing Paradigms for Mental Health* (pp. 13-22). Springer, Cham.

Strahan, R. & Gerbasi, K. (1972). Short, homogeneous version of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191–193.

Stuber, M. L., Kazak, A. E., Meeske, K., Barakat, L., Guthrie, D., Garnier, H., ... & Meadows, A. (1997). Predictors of posttraumatic stress symptoms in childhood cancer survivors. *Pediatrics*, 100(6), 958-964.

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., & Malik, A. S. (2017). Machine Learning Framework for the Detection of Mental Stress at Multiple Levels. *IEEE Access*, 5, 13545-13556.

Suganya, M., & Ebenezer, D. D. (2017). High dimensional data partitioning with an adaptive ensemble construction and analysis scheme. *Advances in Natural and Applied Sciences*, 11(3), 117-124.

Tang, K. N., & Chan, C. S. (2017). Life satisfaction and perceived stress among young offenders in a residential therapeutic community: Latent change score analysis. *Journal of Adolescence*, 57, 42-53.

Tavana, M. (Ed.). (2014). *Handbook of Research on Organizational Transformations through Big Data Analytics.* IGI Global, Hershey PA USA.

Trawiński, B., Telec, Z., Krasnoborski, J., Piwowarczyk, M., Talaga, M., Lasota, T., & Sawiłow, E. (2017, July). Comparison of expert algorithms with machine learning models for real estate appraisal. In *INnovations in Intelligent SysTems and Applications (INISTA), 2017 IEEE International Conference on* (pp. 51-54). IEEE.

Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

Wolf, L., Stidham, A. W., & Ross, R. (2015). Predictors of stress and coping strategies of US accelerated vs generic baccalaureate nursing students: an embedded mixed methods study. *Nurse Education Today*, 35, 201–205.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.

World Health Organization. (2012). Depression: A global crisis. *World Mental Health Day, 2012*.

World Health Organization. (2017). Suicide (Fact sheet No. 398). Retrieved from http://www.who.int/mediacentre/factsheets/fs398/en/

Wu, J., & Chen, E. (2009). A novel nonparametric regression ensemble for rainfall forecasting using particle swarm optimization technique coupled with artificial neural network. *Advances in Neural Networks–ISNN 2009*, 49-58.

Yamaguchi, A., Kim, M. S., Oshio, A., & Akutsu, S. (2017). The role of anger regulation on perceived stress status and physical health. *Personality and Individual Differences*, 116, 240-245.

Yang, X., Tang, K., & Yao, X. (2012). A learning-to-rank algorithm for constructing defect prediction models. *Intelligent Data Engineering and Automated Learning-IDEAL 2012*, 167-175.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit‐formation. *Journal of Comparative Neurology*, 18(5), 459-482.