UNIVERSITI TEKNOLOGI MARA

CLASSIFICATION AND PREDICTION ON SCHOOL CHILDREN FOR FOOD INTAKE ATTITUDE TOWARD FOOD AND BEVERAGE ADVERTISING ON TELEVISION: KFC AS A CASE STUDY

AHMAD FIKRI BIN ANUAR

BACHELOR OF COMPUTER SCIENCE (HONS)

JANUARY 2017

SUPERVISOR APPROVAL

Classification and Prediction on School Children for Food Intake Attitude toward Food and Beverage Advertising on Television: KFC as a Case Study

By

Ahmad Fikri bin Anuar
2014290494

This thesis was prepared under the supervision of the project supervisor, Mr Khairul Nizam Bin Abd Halim. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfillment of the requirements for the degree of Bachelor of Computer Science (Hons).

Approved by

…….
Khairul Nizam Bin Abd Halim
Project Supervisor

JULY 6, 2017

STUDENT DECLARATION

I certify that this thesis and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

……

Ahmad Fikri bin Anuar

2014290494

JULY 6, 2017

ACKNOWLEDGEMENT

Alhamdulillah praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks goes to my supervisor, Mr Khairul Nizam Bin Abdul Halim for giving me supports and supervised me throughout completing this research to make sure that I delivered an excellent output.

I would also like to give a special thanks to my lecturer, Madam Rosniza Binti Roslan for guiding me through two semesters in completing this research and also to my examiner, Sir Ahmad Firdaus Bin Ahmad Fadzil for giving me amounts of useful comments in order for me to correct and strengthen my research.

Special appreciation and a thousand thanks also goes to my beloved parents and siblings for constantly praying for me, giving me support and encourage me to endure all the stress so that I can finish this research successfully and happily. A lot of thank also addressed to UITM Kampus Jasin, Melaka, for giving a chance on doing this project to gain knowledge and acknowledgement Last but not least, I would like to thank all my friends for being understanding and supportive throughout the semester.

ABSTRACT

Serious health problem in adulthood stage such as diabetes, hypertension, cardiovascular diseases are related to obesity in early childhood. Obesity has become a problem in Malaysia in context of healthy lifestyle and in estimation, Malaysia has highest rates of obesity in South-East Asia involving children. One of the most dominant mediums that promote unhealthy foods is through Television Food Advertising (TVFA) that aimed for children. A new approach were applied by using Artificial Intelligence (AI) strategy, from that the Naïve Bayes (NB) technique is used to predict the eating behaviour of children toward TVFA. Agile methodology is used as the project framework of the project study. Phase in agile is proceed one by one for each 5 phase of Agile. First phase is the Planning Phase where problem are identified, then the Analysis Phase to gather information about project, then Development Phase to design the system and produce prototype, followed with Testing Phase where all testing is done and lastly is to compile project finding in final year report as in the Documentation Phase. Five independent variables used in the model, are advertisement recognition, favourite advertisement, purchase request, product prefers and time watched TV. About 105 of school children of SK Merlimau of age 12 years old have been chosen as the target subject to realize the objectives of the prediction model. 80% of data collected were used as training data, and 20% were for the new data to be tested. 31 prediction models were produced by using this technique, and the results indicate that 78% accuracy from the data learnt. Although the accuracy result is not as expected (80% and above), Naïve Bayes could be implemented and may be continued by using other methods such as Support Vector Machine and Artificial Neural Network. The result finding for the system functionality is at best and functioning well. System can predict the expected outcome as data is learned before with appropriate variable. In the near future, hopefully there will be an extended work in terms of different technique and independent variables used to increase the accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Advertisement Recognition
DV	Dependent Variable
FA	Favorite Advertisement
IV	Independent Variable
ML	Machine Learning
MLR	Multiple Linear Regression
NB	Naïve Bayes
PP	Product Preference
PR	Purchase Request
SVM	Support Vector Machine
TV	Television
TVFA	Television Food Advertisement
TW	TV View Duration

CHAPTER 1

INTRODUCTION

This chapter describe the background and rationale of the study. Issues and problems are discussed in this chapter that led to this research. Details discussed include the project summary such as background of study, problem statement, research objectives, research scope, and research significance.

1.1 Background of the study

Obesity is the result of accumulation and increasing of excess body fat that become a multi-factor disease (Must A. & Strauss R.S., 1999). Obesity has become a problem in Malaysia in context of healthy lifestyle. In estimation, Malaysia is stated for the highest rates of obesity in South East Asia (Ng et al., 2015). Serious health problem in adulthood stage such as diabetes, hypertension, cardiovascular diseases are related to obesity in early childhood. The victim of these problems not only involves the adults but also children. In this context, obesogenic environment of childhood need serious attention as a modified risk factor (Ng et al., 2015).

Obesity came from various causes. Obesity influent the Body Mass Index (BMI) of an individual. Food advertising is an example factor toward children obesity. Food advertisement can be advertised in many ways. Food marketing focused on advertising as the most important marketing tools to achieve their market sales (Folkvord, Anschütz, Boyland, Kelly, & Buijzen, 2016) . Used of the advertiser is to induce craving for advertised food by persuasive message and also can give positive impact on their labels or brands (Folkvord et al., 2016)One of the determinants of the advertisement is through television broadcast. Medium that is the most dominant to encourage unhealthy foods are through Television Food Advertising (TVFA) that aim for children (Cairns, Angus, & Hastings, 2009;

Boyland, Harrold, Kirkham, & Halford, 2012). As well-known that TV is one of the popular medium of entertainment, TVFA often are involved. Animation used in TV commercial in food marketing purpose aims to attract children attention by combining the fictional world of advertisements with the reality (Ng et al., 2015). TVFA increased children's food consumption to a higher rate of childhood obesity when watched food advertisement compared to normal weight children (Halford, Boyland, Hughes, Oliveira, & Dovey, 2007),. Children's obesity on these researches is determined by the children behaviour on food intake by the time consumes on TV viewing duration.

Research done by Ng et al., (2015), stated that study of a given knowledge's gap involving obesegenic environment in Malaysia, aimed to evaluate four induction factors. The factors are advertisement recognition by children, what are their favourite advertisement, purchase request by them and lastly their product preference, with TV viewing duration. By using these inductions, children behaviour on food intake such as involvement in consuming the product can be determine.

1.2 Problem Statement

In the context of food advertising through television, there are several factors that involved the attitude of school children toward food intake that can lead to obesity (Increased BMI). From previous research done by Li et al. (2016) and Ng et al.(2015) there are several involvement of children attitude's for food intake toward food and beverage advertising.

Result of the research done by Ng et al.(2015) shows that children influenced by fun TV commercials, boosting purchase petition and gradually but firmly establish early childhood liking for unhealthy food by the food industries in Malaysia. Researches conducted are based on these induction factors; advertisement recognition, product preference, purchase request and lastly their favourite advertisement, with TV's watching time to determine the children's attitude. Li et al. (2016) research finding stated that unhealthy food

advertisement influenced children by the time consume to watch the advertisement and the persuasive marketing tactics. The variable to determine the behaviour is focusing on the time children consumes on TV viewing.

Those research have been using quantitative approach such as correlation and survey research. There is no experimental research done that focusing on system based on Artificial Inteligence in the research and there is no specific label or product of advertisement is stated in the research.

1.3 Objectives

The main objectives of this study are:

1. To identify independent variables that can be used to predict the behaviour of school children on food intake toward food advertisement (KFC) on television.
2. To design and develop a system to predict the behaviour of school children on food intake toward food advertisement (KFC) on television using Naïve Bayes.
3. To test the functionality of the system.

1.4 Project Scope

The scope of the project is to develop a system that can predict school children food intake attitude toward food advertisement on television (KFC advertisement). School children of SK Merlimau have been chosen as the target subject to satisfy the objectives of the system program. The subject to be question is mainly targeted for elementary school student age of 12 years old. The medium of question to be given are a set of questionnaire for the children. This method use is to gain information regarding the food intake by the children.

1.5 Significance of Project

This study will contribute in term of knowledge for predicting school children's attitude toward food intake based on TVFA specifically KFC as the case study. Experimental research approach is applied by using Artificial Intelligence strategy. This study will be the continues study from the previous research. Hopefully, this project may be a tool for communities in predicting the behaviour of children in context on food intake based on the commercialised food advertisement.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this chapter, literature review will discuss about the finding of the research based on the proposed project title that is the prediction on the behaviour of children food intakes based on advertisement in television focusing on food advertise. Figure 2.1 shows overall main idea of this project.

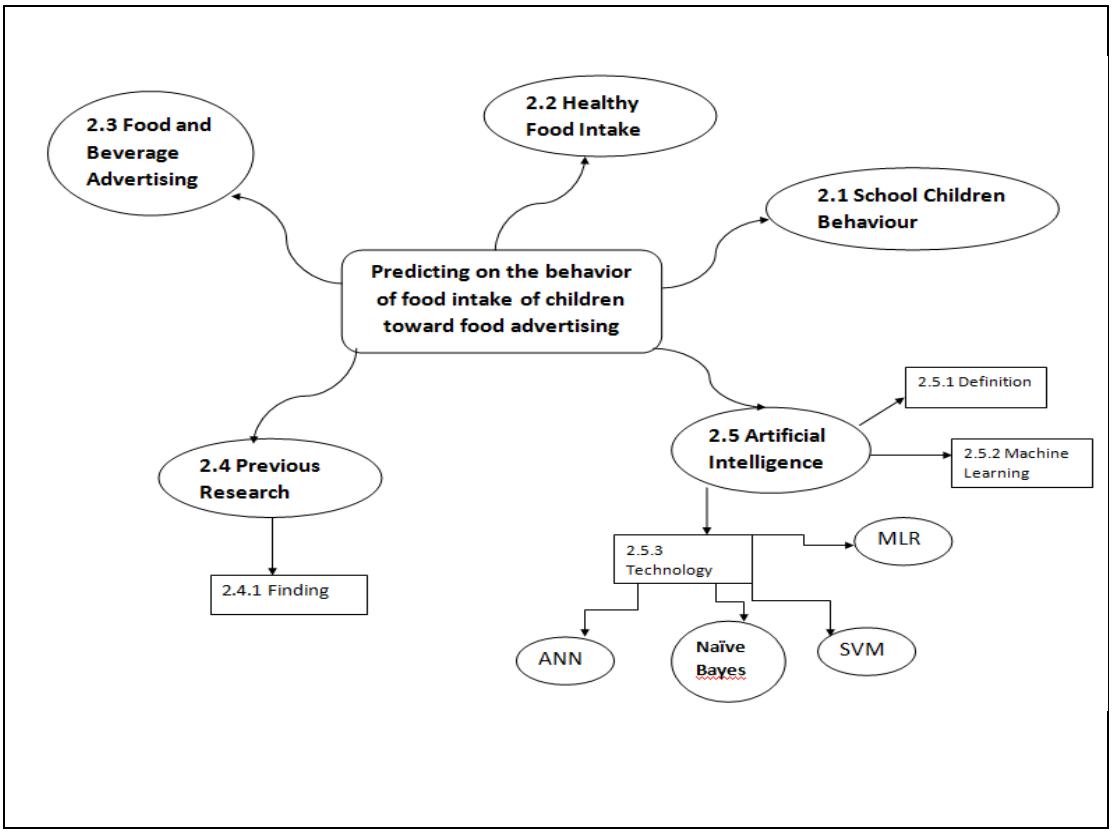


Figure 2.1 –Mind map of the project

2.2 School Children Behaviour

School children are children who are attending classes of education in attempt to get knowledge. Starting age of majority countries in Asia that receive formal education is starts at the age of 6 (Policy, n.d.) The range of age for primary school children can be in between 7 to 12 years old. For secondary school, the range of age can be in between 13 to 17 years old.

Children are exposed to different kind of food. They are freely to choose what they want to eat. In the context of behaviour, children likely to choose foods which contain a high in energy density and prefer food that has sweet taste more than bitter taste (Birch, 1999; Caton et al., 2014; Johnson, McPhee, & Birch, 1991; Kern, McPhee, Fisher, Johnson, & Birch, 1993). They are more reluctant preferring to choose based on the food taste. Mostly foods that are tastier in taste are unhealthy and bad for children health development.

During the stage of when children ages increases, they become more reluctant to eats novel foods and by 2–3 years of age many develop neophobia, even previously liked foods might be refused (Byrne, Magarey, & Daniels, 2014; Caton et al., 2014) This show children behaviour can be develop and affected by what can their eats.

2.3 Healthy Food Intake

To achieve healthy lifestyle, people need to eat and choose healthy food for body intake. There are different category of food mainly categorize as healthy and unhealthy food. Health problem will occur with unhealthy lifestyle. Researches by Ares et al. (2015) and Goetzke et al(2014) confirm, adequate nutrition is, along with physical activities, a critical aspect in influencing a person's health status and, ultimately, the physical and psychological wellbeing of the individual (Ares et al. 2015) nowadays we are facing the escalation of illnesses connected with unbalanced nutrition, including hiking obesity rates and increasing risks of chronic non-communicable diseases (NCD) (Ruini et al., 2016).

This trends is even more worrying if we consider that obesity rate is increasing rapidly among the youngest generations: according to the World Health Organization, more than 40 millions of children above 5 years were overweight or obese in 2011 (WHO, 2014). Obesity is the result of accumulation and increasing of excess body fat that become a multi-factor disease (Must A. & Strauss R.S., 1999). Obesity has become a problem in Malaysia in context of healthy lifestyle. In estimation, Malaysia is stated for the highest rates of obesity in South East Asia (Ng et al., 2015). Serious health problem in adulthood stage such as diabetes, hypertension, cardiovascular disease are associated with childhood's obesity. The victim of these problems not only involves the adults but also children. In this context, obesogenic environment of childhood need serious attention as a modified risk factor (Ng et al., 2015)

The result from a survey indicates that around 45 % of 19 to30 year-old men and 30% of same-aged-women take five or more daily servings of fruits and vegetables (Powell, Chriqui, Khan, Wada, & Chaloupka, 2013). For young adults, the average consume of beverages that is sugar-sweetened product is approximately two cups per day(Zizza, Sebastian, Enns, & Zeynep, 2015).

Furthermore, findings from other nationally representative data indicate that young adults, age 18–27 years, are frequent consumers of fast foods, with reported consumption frequency averaging 2·5 times per week(Quick, Wall, Larson, & Haines, 2013) Regularly eating dinner with others was significantly associated with numerous indicators of healthier dietary intake (i.e. greater intakes of fruits and vegetables), whereas eating on the run was associated with poorer dietary intake (i.e. greater intakes of soft drinks, fast foods and saturated fat(Laska, Hearst, Lust, Lytle, & Story, 2015). Figure 2.2 show the ideal food intake for healthy lifestyle.

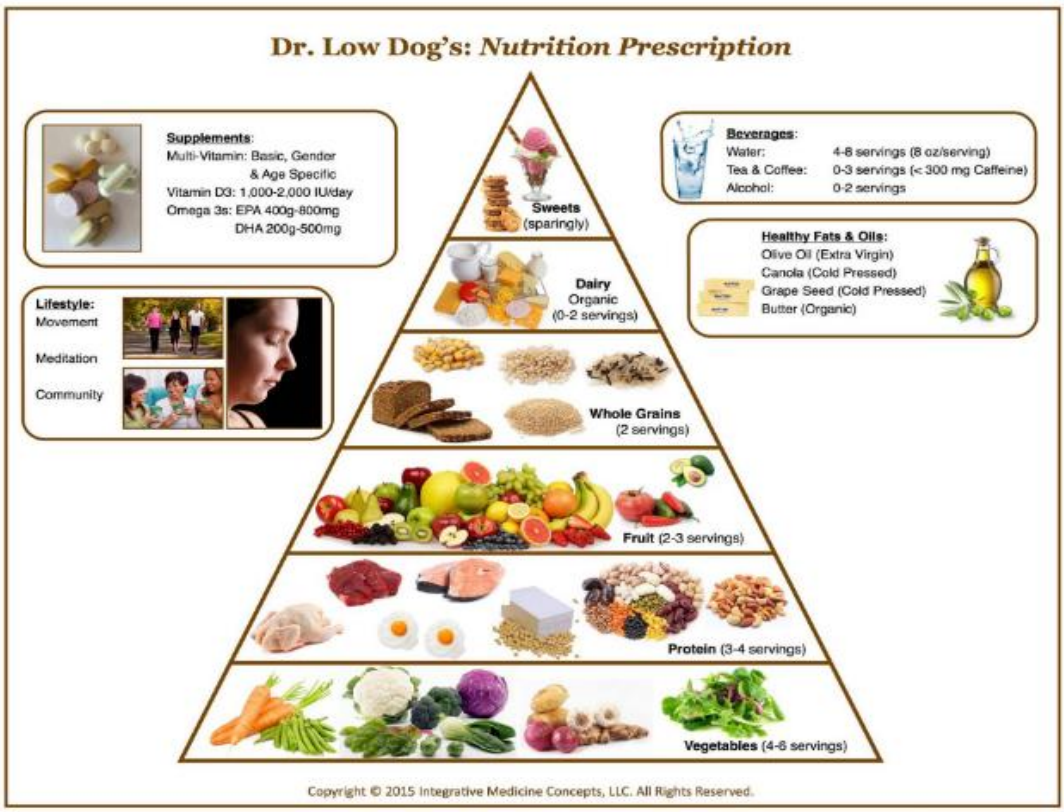


Figure 2.2 *General Pyramid of Food Intake*

(Source: Reality based Nutrition, 2015)

2.4 Food and Beverage Advertising

Advertisement is defined as an announcement or notice to public in order to promote a product, services or publicizing a job vacancy or an event. To commercialise a product and to attract the target audience is the objective of using an advertisement. . Also, the used of the advertisement is to induce the craving for the product by persuasive message and also can give positive impact on their labels or brands (Folkvord et al., 2016).

Food advertisement is advertised in different method. In a large country such as United States, advertising provides support for the media. Several media have been an advertising medium. Example, TV broadcast and much of cable TV, newspapers, magazines and the most is through internet usage. The media provide the significance of knowledge and entertainment for customers. The reason to use media as a medium is that an advertisers much pay media owners for displaying their product as in turn they do not have to charge the consumers the full cost of the media. Figure 2.3 shows the components in media advertising.

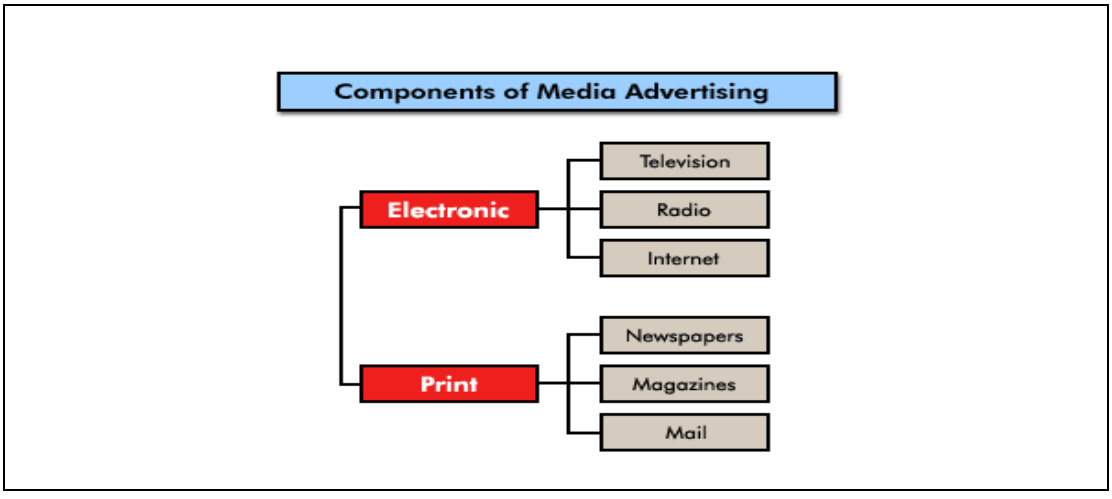


Figure 2.3 *Components of Media Advertising*

(Source: Resource Library ,2015)

Food marketing focused on advertising element as the most important marketing tools to achieve their market sales (Folkvord et al., 2016) .One of the determinants of the advertisement is through television broadcast. The most dominant medium to promote unhealthy foods and food choices are TV food advertising (TVFA) for children (Cairns, Angus, & Hastings, 2009; Boyland, Harrold, Kirkham, & Halford, 2012). As well-known that TV is one of the popular medium of entertainment, TVFA often are involved. Animation used in TV commercial in food marketing purpose aims to attract children attention by combining the fictional world of advertisements with the reality (Ng et al., 2015). TVFA increased children's food consumption to a higher rate of childhood obesity when watched food advertisement compared to normal weight children (Halford et al., 2007). Figure 2.4 shows the most the influential form of advertising that effected to the consumers.

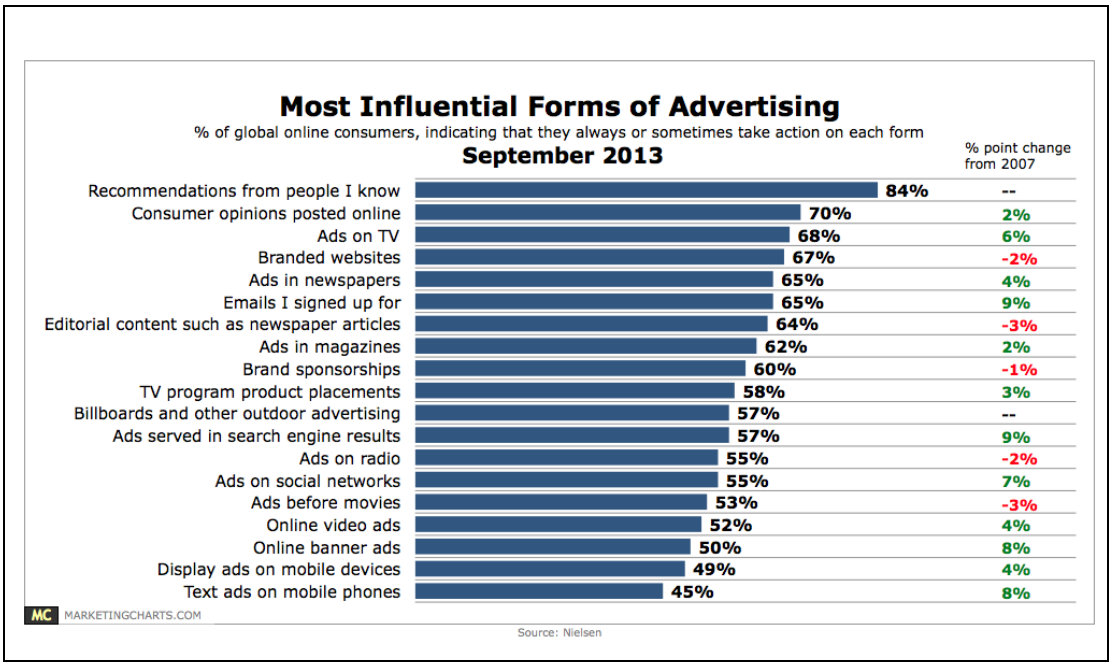


Figure 2.4*Most Influential Form of Advertising*

(Source: MarketingChart , 2013)

Table 2.1 show the different marketing techniques discovered by refer to previous researches.

Table 2.1 Marketing techniques

Author	**Marketing Techniques**	**Medium of marketing**	**Effect of advertising**
(Ng et al., 2015)	• food advertising represent food advertisements as entertain and exciting, and costumers of these goods sell as equally interesting • animation used to target the attention of children by combining the fictive world of advertisements and the reality	Television (TV)	• children remember an advertised product's slogans or taglines and used frequently • Increased children's food consumption when they watched food advertisements and with a greater likelihood for overweight.
(Li et al., 2016)	• Propaganda of energy -dense nutrient- poor foods together with celebrity endorsement and brand.	television (TV), convenience stores, and the internet compose the bulk of sites	• increase of inactive indoor lifestyles including watching TV, playing video games, and using the internet, sup-ports the possibility of more energy-dense nutrient-poor dietary intakes underlying the dramatic rise in childhood obesity

Author	Marketing Techniques	Medium of marketing	Effect of advertising
(Boyland et al., 2012)	• promotional characters, celebrity endorsers, premium offers, persuasive appeals , • cartoon or cartoon-related characters (e.g. Tony the Tiger or Ronald McDonald) are used in advertising or on packaging.	Webs, television, 'advergaming' – a game in which the advertised product is part of the game	• Children are influences by cartoon advertising.
(Karupaiah, Chinna, Mee, Mei, & Noor, 2008)	• 'High in Fat, Salt and Sugar' (HFSS) on children's television (TV) to	Children's television (TV)	• exposed children will choose advertised food products at significantly higher rates than children who were not exposed

2.5Previous Research

There are several researches have been conducted to discuss on the behaviour of children attitude toward food intake based on TVFA Most researchers found that advertisement on TVFA are act as the main criteria for obesity in childhood. However, there are still some others factor that can also be the cause of childhood obesegenic environment to happened that is stated in those researches. In this section, previous researchers' study findings will be extracted to show the factors which affect decision making in school choice and the problem related to their researches.

2.5.1 Previous Research Findings

Table 2.2 shows different research with different method, finding and their respective independent variables (IV) according to the same problem statement that involve TVFA and children food intake behaviour.

Table 2.2: Previous Research Findings

Author	Method	Finding	Independent Variable
(Ng et al., 2015)	Study design, ethics statement, and subject recruitment, Questionnaire development was facilitated using a food album of 24 advertised food products	Findings in this study revealed that Malaysian school children were more attracted to unhealthy TV advertisements than healthier core food	induction factors namely; a)advertisement recognition, b)favourite advertisement, c)purchase request, and d)product preference
(Li et al., 2016)	tests and logistic regression were applied to determine differences in the proportion of healthy food, unhealthy food and miscellaneous food advertisements for different channels, programs, dates, viewing periods and the use of persuasive marketing tactics	Intervening in the entrance of unhealthy foods into the market and establishing regulations related to food advertising especially unhealthy food advertisements are important strategies to prevent children's exposure to unhealthy food and childhood obesity.	Watching TV for a long period of time will increase the tendency toward and possibility of childhood obesity due to some unhealthy dietary behaviours, including more consumption of high-fat foods, fast foods and sugar-sweetened beverages

Author	Method	Finding	Independent Variable
(Boyland & Halford, 2013)	.Comparison between previous researches.	Children are exposed to extensive marketing activity through a variety of media and non-broadcast sources at all stages of their development. Its impact is readily demonstrated by their brand recognition, and its influence on eating behaviour (total intake and food choice)	Television food advertising and branding Nature.
(Karupaiah et al., 2008)	Statistical analyses, Content analysis, Design and Instrument	Food advertising frequency during children's prime time TV varied between TV channels and increased during weekends and holidays. It was found that advertisements of snacks are screened 5 times more frequently than fast foods and are the major component of food advertising on children's TV in Malaysia. The nutri-tional content of these snacks are a major concern as they are high in calories, sodium and salt per 100g of consumed food.	The study purpose was to report on the nature, type and frequency of food advertising directed to children during prime time by the different TV stations in Malaysia. Through content analysis, they were able to describe the nutritional value of advertised foods, and utilizing a Food Pyramid model we compared these foods with recommendations

2.6 Artificial Intelligence

2.6.1 Definition of AI

In Computer Science, Artificial Intelligence (AI) has been one of elective subject and has been studied for decades and still has the gap of knowledge in that area. This is partly because to how big and hazy the subject is. AI defines as the ability of machines to think and search algorithms used to play board games(Smith, McGuire, Huang, & Yang, 2006). The applications applied in almost every way computer is used. They also mention that AI term was first introduces by John McCarthy in 1956 in the first academic conference about the subject. In *As We May Think* by Vannevar Bush's, he suggested a system which raises people's own knowledge and understanding. Then, years later, a paper on the conception of machines wrote by Alan Turing that able to imitate the people's ability to thinks, such as play Chess (Smith, 2006).

The word 'artificial intelligence' itself meaning of the learning on computations that make it able to recognize cognition as brought up by Winston (1992). The objective of artificial intelligence can be observed from engineering and science prospect. AI also can succeed in helping expert to solve problems in analysis, help designing brand new devices, learning from examples and others, AI becoming less perceptible, yet more crucial. Thus, by using AI approach, prediction can be conducted based on by giving machine knowledge as human.

2.6.2 Machine Learning

Research in AI and computer science has studied the case in which the machine started with zero information about what is the problem. Machine learning is a part of AI that apprehensive with programs that learn from

knowledge it gain(Russell & Norvig, 1995). A good deal of background knowledge is basic for human learning.

Machine learning has been concerned with figuring the procedure of generalization as a search through conceivable theories but numerous technique do not include any searching by any means (Witten, Frank, & Hall, 2011). Inductive machine learning is one of the most important machine-learning tasks in which a generalization is gotten from arrangement of tests, and is formalized using distinctive models and methods (Kantardzic, 2003).

2.6.3 AI Technology

There are several AI technologies that have been used to predict things. As stated by Mellit et al., (2009), example of AI technologies are, expert system, artificial neural network, genetic algorithm and fuzzy logic. However, in this section only discussing on multiple linear regression (MLR), support vector machine (SVM), artificial neural network (ANN) and Naïve Bayes (NB) technology

a) Multiple Linear Regression

Multiple linear regression (MLR) is a method applied to design the linear relationship between a dependent variable and one or more independent variables where the dependent variables are sometimes called the predicted and the independent variables are called the predictors (Meko, 2015). MLR is the extension of the simple linear regression model as MLR needs more than a variable in order to do the prediction.

b) Support Vector Machine

SVM is a data mining tool and it has been used for study in varies fields(Cortes & Vapnik, 1995). The basic SVM is convex and extracted by minimize the risk of structural rather than the risk of empirical problem that

become the aim of other methods are, which able it to have fully solutions and solve the problem of over installation of the basic model (Xu & Dong, 2016). Those stipulate it perform well in generalization. Yet the basic model's complexity time is equal to O(n2). The support vector machine paradigm (SVM) is study for linear predictors in high dimensional feature spaces. Those raise both sample complexity and computational complexity challenges. The SVM algorithmic paradigm solved the sample challenges by finding the "large margin" separators. All in all, a half space separate and differentiates the training set with a large margin if all the examples are on the correct side and far away from separating hyper plane. Confining the algorithm to production of the large margin separator can produce a small sample complexity even if the dimensionality of the space will be high.

c) Artificial Neural Network

An artificial neural network (ANN) is a model of computation inspired based on the structural of human brain's neural networks. In the brain structural model, it made up of a huge amount of neurons as the basic computing devices that are connected to each other in a complexity of the communication network, through which the brain enable to bring out complex computations. Based on these computation paradigms, artificial neural networks is build that are modelled after this. A neural network can be related as graphs that directed to each nodes equivalent to neurons and edges correspond between them are linked. Each neuron accepts input as a weighted total of the outputs of the neurons connected to its edges. As mentions by Mia and Dhar, (2016), based on the functional of a human brain, ANN is a system that do non-linear mapping. A summation of three layers, each consists of one or more neurons, is the usual structure of an ANN. Some numerical values are demonstrated to the network through the neurons of the input layer. Each of it consist of input layer that can take only single input value and then will be brought to interconnector's hidden layers by synaptic weights to output layer. This shows in a way such that every connection of the layer is connected to every single neuron. Then, the output layers give out responses in numerical values.

The training of ANN is done by adjusting the weights or strength of the input connections, intermediate connection and output connection of the neurons which are able to do the storing works in memory as information. By able to achieve the ability learn problem, ANN able to produces the required responses based on what the given decision variables is. The number of the neurons on a single selected hidden layer can be different but choosing the right number, which can evade from over-fitting because of too many neurons and also the under-fitting due to too less neurons, is said to be a hard task to solve and it is determined based on the vector size of the input and vector space classifications of the input–output relation (Cao &Zhang, 2006; Karakurt&Hamzacebi, 2015).

The production of ANN are based on the training data size, training algorithm, transfer function, training time, structure of the network, learning function, values of weights and biases, and also can be determined on data representations and testing data size. Gradually by training an ANN, ability of the model to find logical solutions to the similar problems that are being solved, it is becoming more popular and gives out more contribution to study and research lately (Karakurt&Hamzacebi, 2015). In research done by Bardak, Tiryaki, Nemli, & Aydin, (2016) they are using this technique since ANNs are among the best and fastest in gathering and building information processing techniques in the field of the artificial intelligence for the prediction on bonding strength of the wood joints under varied conditions. ANNs are able to solved complex problems and increasing used for study or research due to the model in success to solved problem (Cao &Zhang, 2006; Karakurt&Hamzacebi, 2015).

The model is capable to acknowledge the work without having a detailed about the problem of system. Instead of acquiring the information, they are capable in learning the input and output relation through the variables of the previously data's recorded (Kalogirou, 2001).To able to accomplish this, the trained network with the data based on to the problem that are being solved by using a training algorithm. Adjustment on the connection weights that able the ANN to

provide outputs that are the same or close to targets is the training of the process (Hamed, Khalafallah, &Hassanien, 2004).

d) Naïve Bayes

Naïve Bayes or its classifier is classified as a family of simple probabilistic classifier based on Bayes' Theorem with strong independence assumptions between features; the independent variables in particular problem. Text categorization is usually used in Naïve Bayes.

Abstractly, Naïve Bayes is a conditional probability model given a problem instance to be classified, represented by vector $x = (x1, \dots, xn)$ representing some n features (independent variables), it assign to this instance probabilities

$$p(Ck|x1, \dots, xn) \quad (1)$$

for each of K possible outcome or classes Ck[7]

Using Bayes' Theorem, the conditional probability can be decomposed as

$$p(Ck|x) = \frac{p(Ck)p(x|Ck)}{p(x)} \quad (2)$$

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on C and the values of the features Fi are given, so that the denominator is effectively constant, The numerator is equivalent to the joint probability model

$$p(Ck, x1, \dots, xn) \quad (3)$$

which can be rewritten as follows, using the chain rule for repeated applications of the definition of conditional probability:

$$\begin{aligned} & p(Ck, x1, \dots, xn) \\ & = p(x1, \dots, xn, Ck) \quad (4) \\ & = p(x1|x2, \dots, xn, Ck)p(x2, \dots, xn, Ck) \end{aligned}$$

$$= p(x1|x2, \dots, xn, Ck)p(x2, \dots, xn, Ck)p(x3, \dots xn, Ck)$$

$$= p(x1|x2, \dots, xn, Ck)p(x2, \dots, xn, Ck)p(x3, \dots xn, Ck) \dots$$

$$p(xn-1|xn, Ck)p(xn|Ck)p(Ck)$$

Now the "naïve" conditional independence assumptions come into play: assume that each feature $F_j$is conditionally independent of every other feature $F_j$for j not equl to I, given the category C. this means that

$$p(xi|xi+1, \dots, xn, Ck) = p(xi|Ck) \quad (5)$$

Thus the joint model can be expressed as

$$p(Ck|x1, \dots xn) \propto p(Ck, x1, \dots, xn)$$

$$\propto p(Ck)p(x1|Ck)p(x2|Ck)p(x3|Ck) \dots \quad (6)$$

$$\propto p(Ck)\prod_{i=1}^{n} p(xi|Ck)$$

This means that under the above independence assumptions, the conditional distribution over the class variable C is:

$$p(Ck|x1, \dots, xn) = \frac{1}{Z}p(Ck)\prod_{i=1}^{n} p(xi|Ck) \quad (7)$$

where the evidence Z = p(x) is a scaling factor dependent only on x1,…,xn that is, a constant if the values of the features variables are known.

Table 2.3 shows the advantage of Naïve Bayes technique used in previous researches.

Table 2.3 Advantage of Naïve Bayes Technique using Previous Research

Author, Year	**Application**	**Technique and Method**	**Results**
Bilal, Israr, Shahid, & Khan, (2015)	Sentiment classification of Roman-Urdu opinions	Naïve Bayes, Decision Tree, KNN Pre-processing ↓ Features extraction and selection ↓ Classification ↓ Testing of models ↓ Analysis of results	NB: It's performance is best in terms of accuracy, precision, recall and F-measure. Naıve Bayes is a simple classifier but it can perform much better than other sophisticated classification algorithms It has good speed during learning and predicting. Its learning time is linear to the number of examples and its prediction time is independent of the number of examples More efficient in learning and classification than Decision Tree The fact is that it shows a good probability estimate for correct class, which enables it to perform the correct classification Another reason for Naıve Bayesian's good performance is that in a data set two attributes may depend on each other

			but thisdependence may distribute equally in each class. DT : The larger the training set the larger will be the tree size and more accurate results willbe obtained than the tree built from subsets
Baker & McHale, (2016)	An empirical Bayes model for time-varying paired comparisons ratings	Naïve Bayes Node allocation ↓ Empirical Bayes model extension: shrinkage ↓ Optimising the model parameter ↓ Connectivity	The advantages of adopting the empirical Bayes approach are first, that account is taken of players who do not play in many matches or have a high volatility in results, and second, that the strengths of players winning or losing all of their matches can still be estimated.

Author, Year	Application	Technique and Method	Results
Marucci-Wellman, Lehto, & Corns, (2015)	Semi-automated coding of short injury narratives from large administrative databases	Using Muiltinomial NB Extract Dataset ↓ Divide dataset into training set. ↓ Model development (Naïve Bayes) ↓ Evaluation of the human–machine coded approach	1.Multinomial Naïve Bayes model can outperform state-of-art methods of text classification for short snippets of text and when there are few training cases. 2. Naïve Bayesian algorithms was able to comprehensively and accurately classify the events leading to injury. Accurately categorizing the cause, source and location of injuries from hospital, trauma center records, or WC claims is an essential part of the analytic process of epidemiology and surveillance and provides critical information for preventing future events such as amputations 3. A particularly striking benefit of the Naïve Bayes algorithms for short narratives is the ability to filter out cases for manual review where there is low confidence in the algorithm assigned code.

2.7 Conclusion

This project is to predict level of food intake attitude of students in SK Merlimau of aged 12 year old based on their behaviour toward TVFA. The prediction is based on the independent variable of advertisement recognition, favourite advertisement, purchases request, preference on product, and time watching television especially when TVFA is aired and as the result predicted, the dependent variable will be the key. As it is become more serious in health problem for children in Malaysia, the chances in thought of reducing the rate of becoming the top country in obesity, an approach of technology in computer science are taken during this project. Although there are many other solution, Naïve Bayes approach as one of the AI method in prediction is used to build a model of prediction to predict school children behaviour.

CHAPTER 3

METHODOLOGY

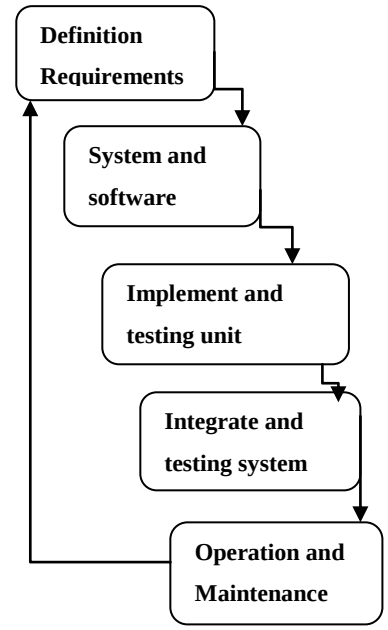
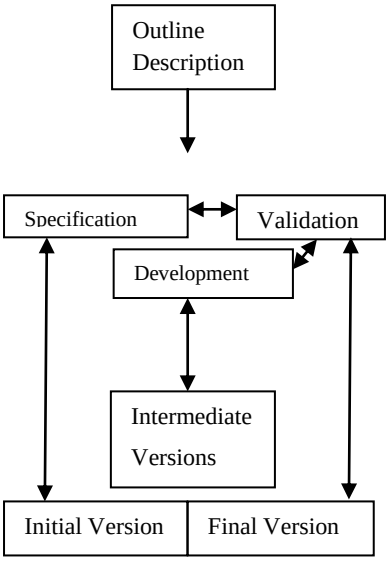
3.1 Introduction

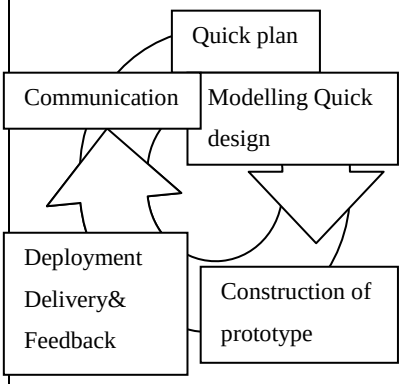
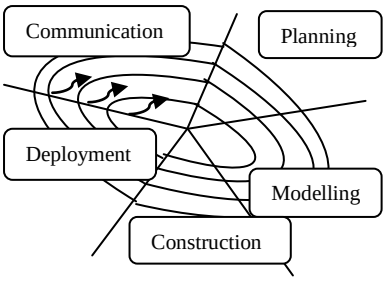
Contain of this chapter will be on the methodology used for application development as fulfilled all required objective purposed. As a main point for this chapter, agile model will be use as the project framework in order to make this project work properly. Agile SDLC model is a combination of iterative and incremental process models. It promotes nonstop iteration of development and testing throughout the software development life cycle of the project. Both development and testing activities in agile model are parallel and complimented to each other.

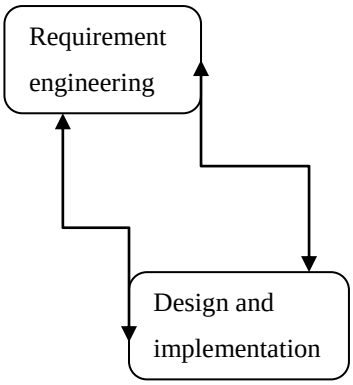
3.2 Project Framework

A generic process model of a framework for software engineering consist of five activities; communication, planning, modelling, construction, and deployment(Suryn, 2014). A project framework is needed for construction of a new system to be smooth. There are several process models as show in table 3.1 below(Bell & Parr, n.d.; Sommerville, 2010; Suryn, 2014).

Table 3.1 Process Model

Process Model	Step in Model	Principles of Model
Waterfall		•Contain well defined series of steps •each step creates a definite product with each product forms the basis for the next step •can check each of the step's correctness •Disadvantage: if in a particular stage, a problem occurs, there is no stepping back to redo and to solve the problem of the previous stages.
Incremental;		• Advantage; -cost is reduced -less redone work than waterfall -easier get customer feedback -More rapid delivery and deployment • Disadvantage; - process is not visible. - System structure tends to degrade as new increments are added .

Process Model	Step in Model	Principles of Model
Prototyping		• Advantage; - easier to determine the user requirements - development of systems are much faster and user friendly - effort in maintenance the system is reduced • Disadvantage - prototype and the actual system inconsistencies will occur. - Users could actually miss some of the shortcomings event in an artificial environment.
Spiral		• spiral model can be adapted to apply throughout the life of the computer software • The main difference between other models is its explicit recognition of risk

Process Model	Step in Model	Principles of Model
Agile	Agile Development 	• **1.** Top priority is to meet customer satisfaction through advance and constant delivery of valuable software, • **2.** Adapt with changing requirements, • **3.** Convey working software regularly, with preference to the shorter timescale. • **4.** Build projects around motivated individuals. • **5.** The most systematic and effectual method of relaying details to the development team. • **6.** Groupware is the priority measure of progress. • **7.** Agile processes encourage endurable development. • **8.** Continuous consideration toward smooth technical and sturdy design that further enhances agility. • **9.** Simple, maximize the amount of work not done, is crucial. • **10.** The best planning, essential, and designs emerge from self-organizing teams.

Therefore, agile is most suitable to be used in this project study and used as the methodology as agile give more accurate and best result and easy to analyze.

3.3 Detail of Phase in Agile Method

In this project, agile model will be use for the project framework. Since this project is limited in time and for one resource person project only, it is suitable to apply this model. Apart from that, agile model also can adapt with fast-moving technology, thus it easy to up-to-date and easy to find support components. Lastly, the best feature of this approach is the ability to adapt to changes in requirements.

3.4 Overview of Project Framework

3.4.1 Overview details

The frame contain of 5 phases that is planning, analysis, development, testing and documentation. Table 3.2.1-1 shows the overview of the project framework.

Table 3.2 Project Framework Overview

Phase	**Activity**	**Output**
Planning	• Able to identify the Project problem : Implementation of AI technology (Naïve Bayes Method) • And able to propose project title	• Able to list out the propose project problem statement, objective, project scope and project significant (Chapter 1)
Analysis	• Gather all the information in regard of the selected project. • Able to define the strategy and procedure by comparing with existing technology used by implement AI technology. • Able to define and identify the independent variable (IV) and dependent variable (DV) through the previous researches for the propose project title. • Able to identify the research design by using AI technology approach • Able to identify the research strategy by using Naïve Bayes method and technique	• Literature Review (Chapter 2)

Phase	Activity	Output
Development	• To collect data for training of prediction model • To design the classification algorithm for flowchart and pseudo code by using Naïve Bayes based on given IV. • To design and develop prediction model from the classification algorithm. • Choose the best prediction model through accuracy test case • Develop prediction prototype system	• Defined prediction model. • Prediction prototype system that well develop and functionality. (Chapter 4)
Testing	• To verify project requirement. • To verify system functionality	• Validation result (Chapter 5)
Documentation	• Compile the project findings.	• Final report

3.4.2 Planning

Table 3.3 below shows the framework of planning phase for this project.

Table 3.3 Planning phase framework

Phase	Activity	Output
Planning	• Able to identify the Project problem : Implementation of AI technology (Naïve Bayes Method) • And able to propose project title	• Able to list out the propose project problem statement, objective, project scope and project significant

The planning phase is where the project title and the domain problem are defined. Based on the literature review in chapter 2, the main problem found that there is no Artificial Intelligence strategy were used on the previous researches conducted. The domain problem defined as the propose project title is "*Classification and Predicting on School Children for Food Intake Attitude toward Food and Beverage Advertising on Television: KFC as a Case Study"*. This project will be focusing on the implementation of AI technique to do a prediction.

In this phase, we can get the problem statements, project objective and project significance. Problem statements define the research gap of the project in which is the implementation AI as prediction strategy to produce better result. The objectives are to find out the independent variables which will used to make prediction, to design and develop system. The scope of this project is to develop system that can predict on school children's food intake attitudes toward food and beverage advertising on television Lastly, the significance of the project is to contribute knowledge in which applying AI approach using experimental research approach and helpful to community for better and healthier environment for children.

3.4.3 Analysis

Table 3.4 below shows the framework of analysis phase for this project.

Table 3.4 Analysis phase framework

Phase	**Activity**	**Output**
Analysis	• Gather all the information that related to this project • Define strategy and procedure • Identify independent variables and dependent variable • Identify research design (approach)	• Literature Review (Chapter 2)

In analysis phase, activities involved are collections of informative data which are can be found in the previous research articles by defining strategy and procedure, identification of independent and dependent variable, research design and research strategy.

The data collected were on time consuming on watching television advertising of primary school children in Malaysia, and their behaviour on food and beverage intakes. These variables are being study on their obesegenic childhood environment based on previous research done. For this proposed project title, AI approach will be used to predict the behaviour for food intake of the school children. The strategy and procedure are then defined. Independent variables that will be used in this project are advertisement recognition by children, their favourite advertisement, purchase request and lastly product prefer by children, with time consume on watching TV to determine the children's attitude in choosing food (dependent variable). Research design or approach that will be used is AI which deploys classification algorithm and Naïve Bayes technique as research strategy. All these information were gathered from journal articles, book and websites.

The output of this phase is the literature review which is a critical analysis of published source, or literature, on a particular topic and provides a summary, classification, comparison and evaluation regarding the topic.

3.4.4 Development

Table 3.5 below shows the framework of development phase for this project.

Table 3.5 Development phase framework

Phase	Activity	Output
Development	• To collect data for training of prediction model • To design the classification algorithm for flowchart and pseudo code by using Naïve Bayes based on given IV. • To design and develop prediction model from the classification algorithm. • Choose the best prediction model through accuracy test case • Develop prediction prototype system	• Defined prediction model. • Prediction prototype system that well develop and functionality. (Chapter 4)

In development phase, flow chart and pseudo code are design using AI approach. Design and development phase involved the classification algorithm based on identified independent variables, designing prediction model from classification algorithm, and developing predictive prototype system. This phase will be later discussed on the next chapter.

The classification algorithms design will be represented in flow chart structure that using Naïve Bayes as the methods based on independent variables stated in this project. Then, the prediction method will be design based on the classification algorithm displayed by using use-case and sequence-diagram which will show the overall view of the system. Lastly, the development on the prototype system will be conducted.

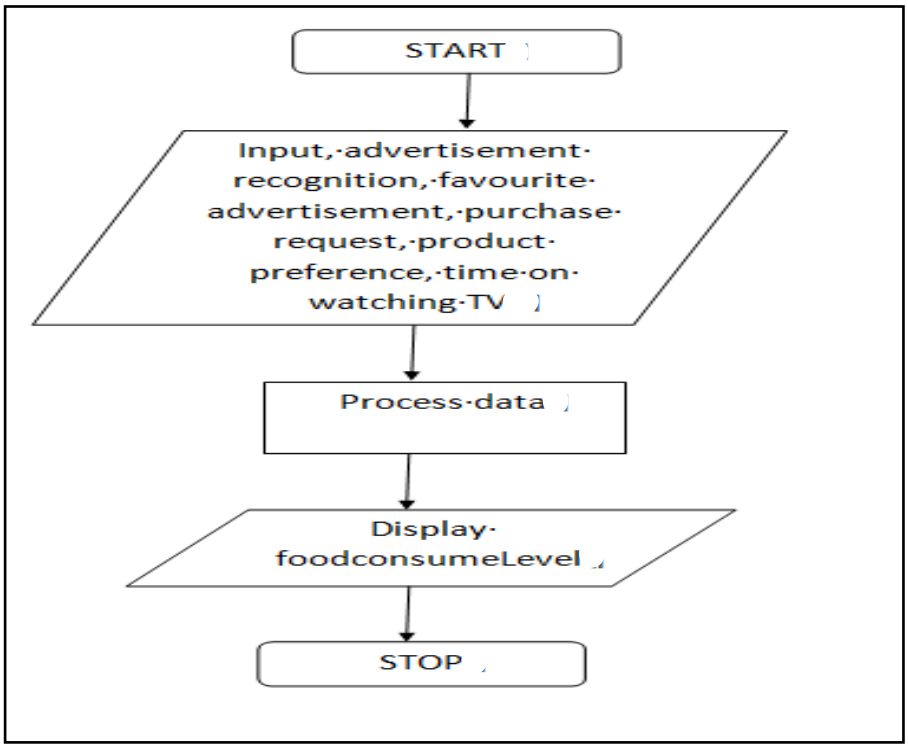


Figure 3.1 Simple flowchart of the system

The Figure 3.1 above will provide simple information about main activity in the system for the flowchart below.

3.4.5 Testing

Table 3.6 below shows the framework of testing phase for this project.

Table 3.6 Testing phase framework

Phase	**Activity**	**Output**
Testing	• To verify project requirement. • To verify prediction model. • To test the system functionality. • To test the accuracy of the system.	• Validation result (Chapter 5)

In testing phase, activities involved are verification of project requirement, prediction model, functionality and accuracy testing. Details on this phase will be further discussed in chapter 5.

Project requirement verification is verified against the system and user requirements. Prediction model is verified by analysing the system with the use case and sequence diagram, to determine whether the flow of the system match these model or not. To test the functionality of the system, it will make available for user to be tested and to get feedback on the system. Lastly, for accuracy testing, since 80% of the collected data from the survey will be used as data for machine learning, another 20% will be used as new data to determine the accuracy, if the accuracy is 80% and above, the system is said to be accurate and reliable. The combination of the independent variables (IV) will also be tested to identify which combination has the most accuracy percentage.

3.4.6 Documentation

Table 3.9 below shows the framework of documentation phase for this project.

Table 3.7 Documentation phase framework

Phase	**Activity**	**Output**
Documentation	• Compile the project findings.	• Final report

The documentation phase is the compilation of the entire project finding starting from the earliest stage (planning) until the finished product is produced. These findings will be combined in the final report.

3.5 Hardware and Software Requirement

The system required a suitable hardware and software to be function efficiently. Hardware is the physical parts or components of a computer and software the programs and other operating information used by a computer. For the purpose project the hardware and software used are listed as in the tables below. Below is the list of Hardware in Table 3.8

Table 4.2 List of Hardware

Component	Specification
Laptop Model : Acer Notebook	Processor: Intel(R) Core(TM) i5-3337U CPU capacity : 1.80 GHz RAM : 4.00 GB OS type : 64-bit, x64-based processor (14393.447) Edition : Windows 10 Pro
Printer Model : HP printer	Series number : Office Jet Pro 8720

Table 4.3 List of Software

Software	Description
Windows 10 Pro	Operating Systems
StarUML	Use case
Netbeans IDE 8.1	Programming language for Java and C++

This project used laptop which has the specifications as stated in Table 4.3 above. Software that will be used are google chrome to display the web pages, Netbeans is used for Java and C++ language (client-side scripting), Active Server Page (Server-side scripting) for database manipulation, and Hypertext Markup Language (HTML) for document representation on web pages. StarUML will be used to construct use-case and sequential diagram.

3.6 Conclusion

In this chapter, it is mainly discuss about the methodology that will be utilized as a rule to guarantee that the application work effectively. Agile method is chosen as the process model for this project because it is the best suited to this application development. The base on agile method is reasonable to build up this application. There are only five phases that involve in agile framework in this project which are planning, analysis, development, testing and documentation. Each of these phases has its own activities and deliverables to guarantee the outcome of this application can be accomplished.

CHAPTER 4

DESIGN AND DEVELOPMENT

4.1 Introduction

In this chapter will be reveals on the progression work of the prediction models for the calculation on accuracy based on 5 independent variable that is the purchase request, advertisement recognition, favourite advertisement, product preference and time watch TV.

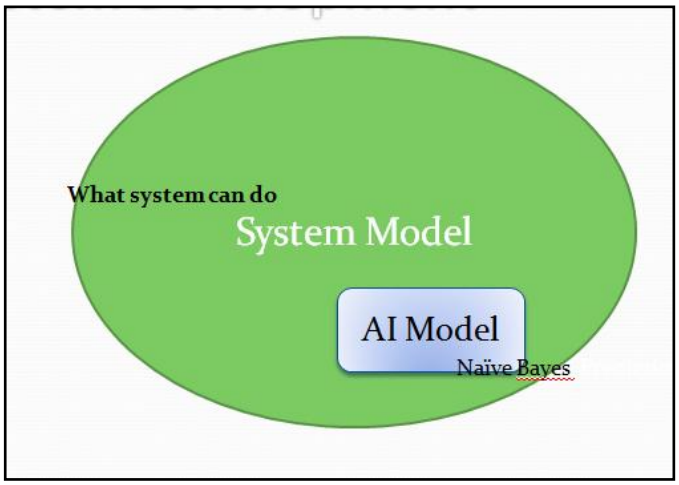


Figure 4.1 Whole System Model

In the Figure 4.1 above show the big picture of the design flow. The set of prediction model and the set of prediction system will be develop by using Naïve Bayes approaches and for interface is by using Netbeans stand alone system model.

4.2 Data Collection

In this study, a set of samplings applies for student of age 12 of SK Merlimau, Melaka have been taken. About 105 students is target as the respondent for the case study. A set of question is given to the respondent to get information regarding the topic case study. Questionnaire is used to collect information regarding the food intake by the target subjects. The data collected is used to test the accuracy by using Naïve Bayes method.

After collected all the needed data sets for the study, the data had been randomly picked as the learning data from 80% of data collection. Meanwhile, the remaining group has been used for testing data with 46 subjects from 20% of data collection. The process of prediction model had been learnt up until 63 models and six groups had been trained once for each model.

The Naïve Bayes methods use knowledge of probability and statistics based on applying Bayes' theorem, which can predict the class membership probabilities. Abstractly, Naïve Bayes is a conditional probability model given a problem instance to be classified. The conditional probability is described as:

$$p(a|b) = \frac{p(b|a)p(a)}{p(b)} \quad (1)$$

From equation 1, p(a|b) defines the probability of a given independent values b. In order to make the equation more simple, it can be written as

$$p(a,b) = p(a|b)p(b) = p(b|a)p(a) \quad (2)$$

Equation 2 proves from equation 1 can be written as we can reverse the conditioning between pairing of random variables. All variables were expected to be independent and simplifies classification. Thus, the result given by Naïve Bayes was more clear in categorizing and classification variables.

4.3 Development Process

4.3.1 Key Data Structure

Figure 4.2 below shows the key data structure of Naïve Bayes. It shows how the independent and dependent variables are constructed and joints together. In each category, there are associated independent variable with values attached to it which particularize the number of joints that category has with each categories of the dependent variable.

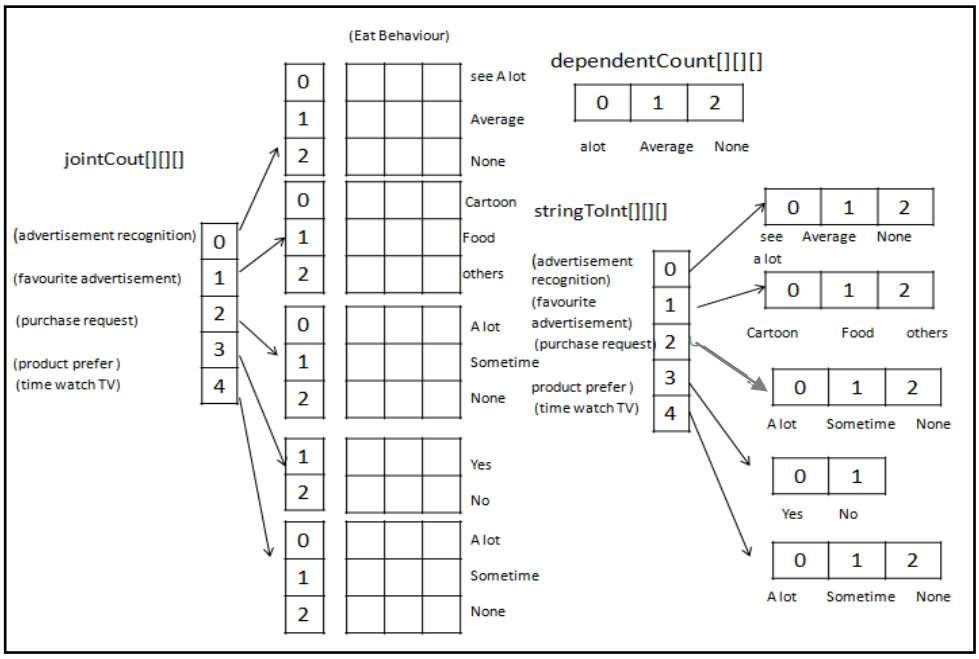


Figure 4.2 Key Data Structure of Naïve Bayer for 5 Independent variables and 1 Dependent Variable.

4.3.2 UML Class Diagram

Figure 4.3 below shows the class diagram for prediction models. The diagram consist of two classes which are the main class and the main application. Independent variables and operations involved in the main class for each model. For the data of respondets and outputs from the operations are included in the main application.

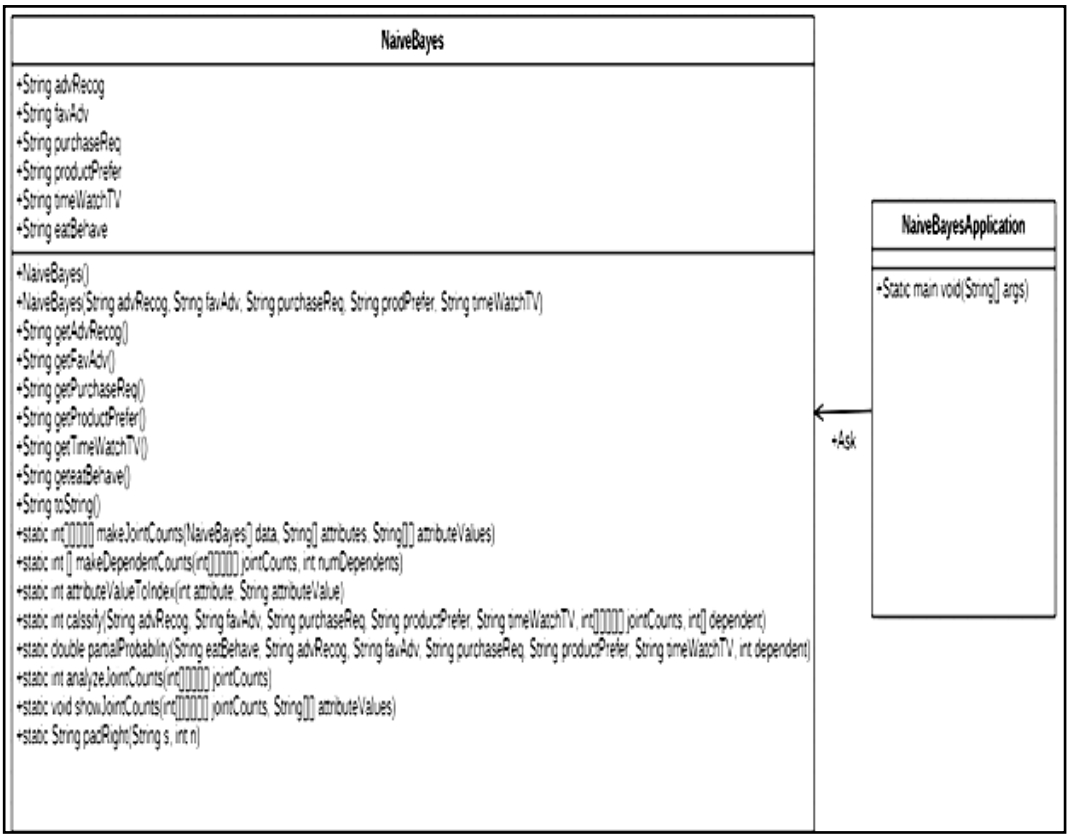


Figure 4.3 UML Class Diagram for 5 IV based on Prediction Model

Table 4.1 below will the guide for the combination of the independent variable used for the set of training data and new data.

Table 4.1 Combination of independent variable

Model	advertisement recognition	favourite advertisement	purchase request	product prefer	time watch TV
1	✓		✓	✓	
2		✓	✓	✓	
3	✓		✓	✓	✓
4		✓	✓	✓	✓
5	✓	✓	✓	✓	✓

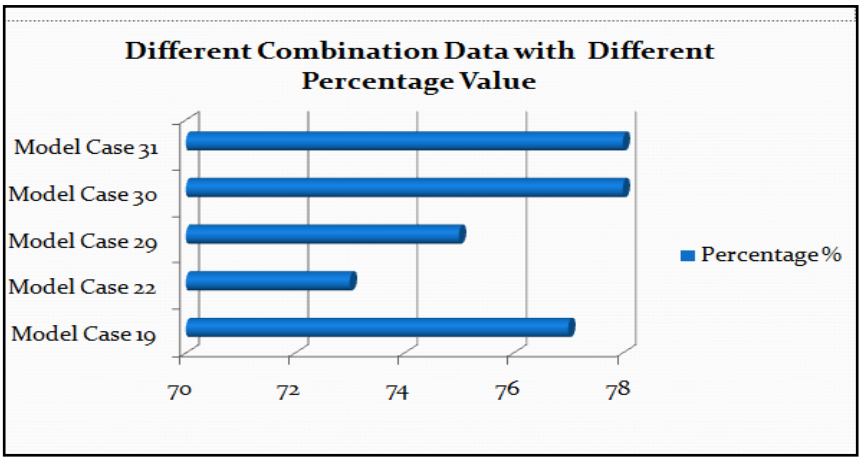


Figure 4.4 Different Combination of Independent Variable

Figure 4.4 shows the different combination of variable with different percentage value. The best 5 model out of 31 prediction model is choose as in the system development used.

Out of all 31 prediction models, only five had been chosen to be displayed for flexibility and usability for user. The process of choosing best prediction was from accuracy percentage of training datasets and new datasets. The least differ between training datasets and new datasets will be chosen as the closest to achieve accuracy of prediction model.This method was done in order to test whether the prediction model is greatly performed or not.

4.2.3 System Flow

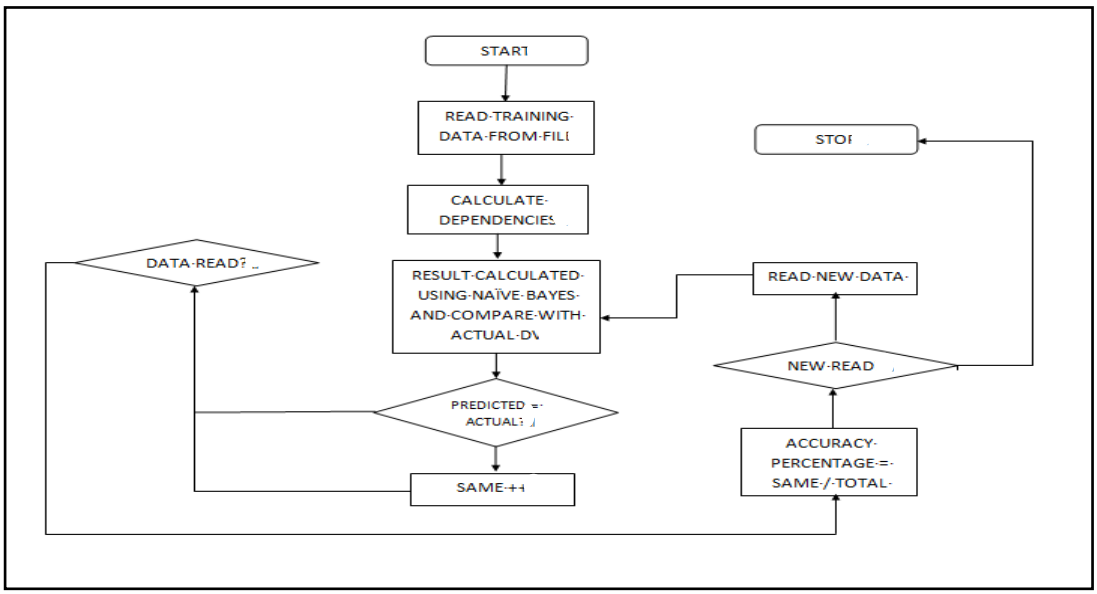


Figure 4.5 Flowchart for the Prediction Model

Figure 4.5 shows the flow of the system prediction model which accuracy testing for both training and new data is process. On the point of prediction, the system will start read all data from a file saved with data collected before. Then, the system will calculate the dependencies based on bayes theorem. The result will be shows by the system and compare with the actual dependent variable. If the predicted dependent variable is equal to the actual then the system will proceed as it save the result and continue with next step. After that, the accuracy testing for training data was determined by dividing value of the variable with the same total data in the file read at the beginning. Then, the data will be read from a new file which is the data that not used to define the dependencies. This data is used to test the dependencies of the training data to further determine the stability of the predictive model. The same happened to the training data, the results then be compared with the actual one and the accuracy determined by dividing the total of matched results with the total data in new file.

4.2.4 Use Case Diagram

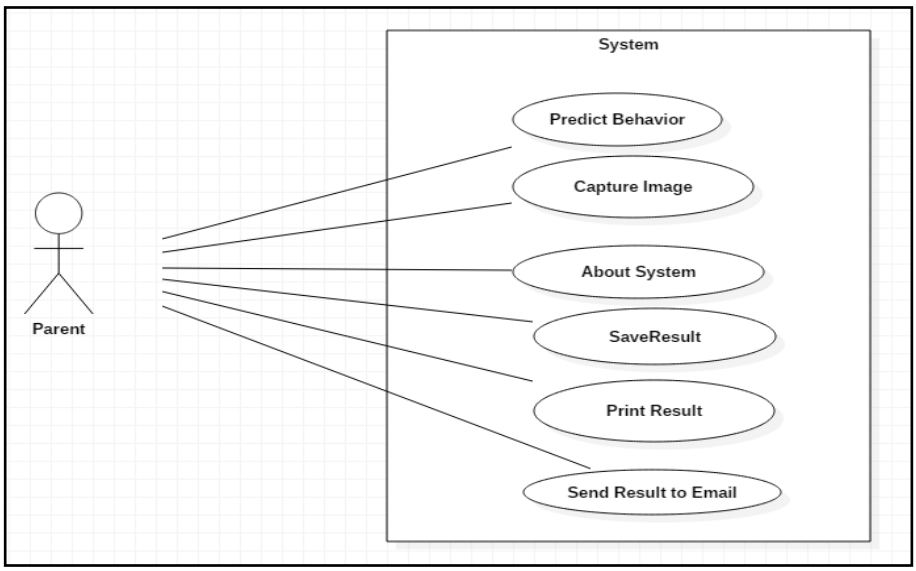


Figure 4.6 Use Case Diagram

Figure 4.6 show the use case diagram use in the project development. User can interact with the system as shown in the diagram.

4.4 System Interface

Before any development on the system start, interface for user have been sketch and roughly created as the actual representation of the real interface.

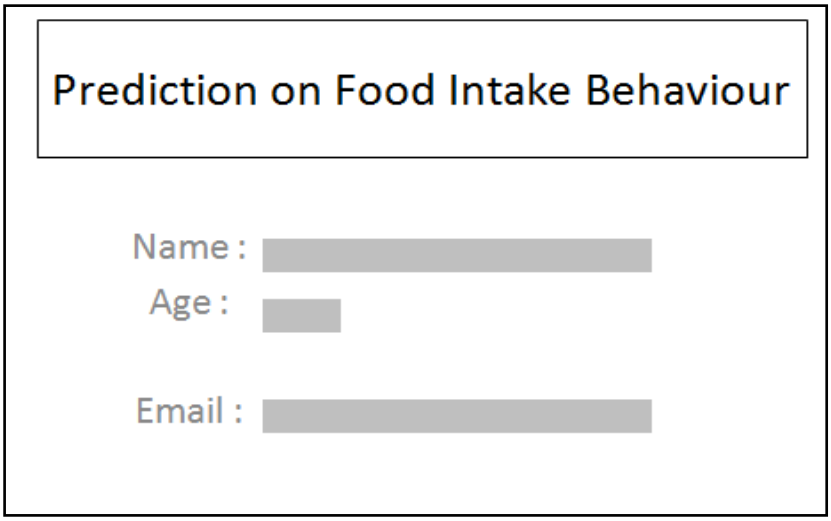


Figure 4.7 Login Interface

Figure 4.7 shows the login interface as a user need to insert required information ask by the system.

Figure 4.8 Main Menu Interface

Figure 4.8 shows the Main menu interface as user need to choose any one of the combination of IV.

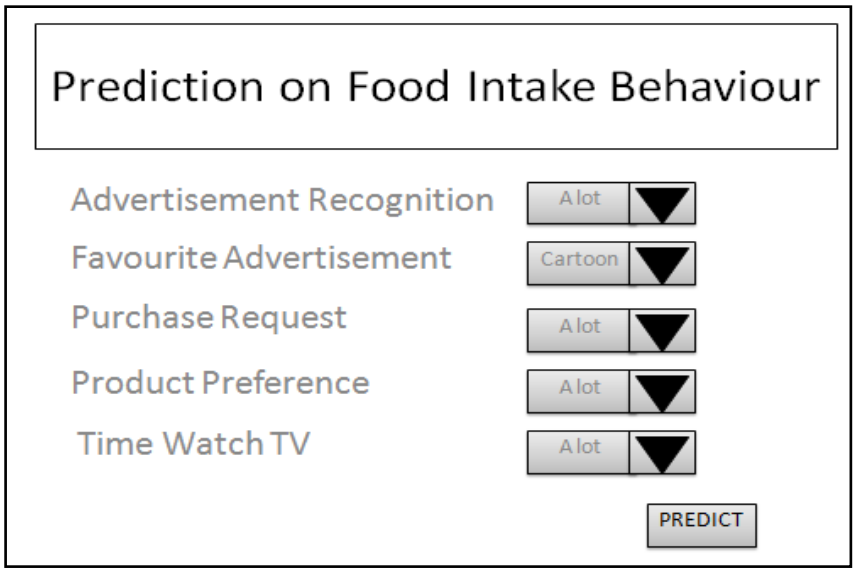


Figure 4.9 Prediction Interface

Figure 4.9 shows the prediction interface as user need to choose all requirement of the IV asked to be able to predict.

4.5 System Development Interface

After roughly plan and sketching for the user interface, after development, the following figure will show the actual interface for the system.

Figure 4.10 Main Page

Figure 4.10 shows the actual interface for the system. When starting the system, the first step user need to do is to insert all information needed by the system before proceeding to next step. User need to enter name and age of their child and their own email address. The information is then will be used later by the system.

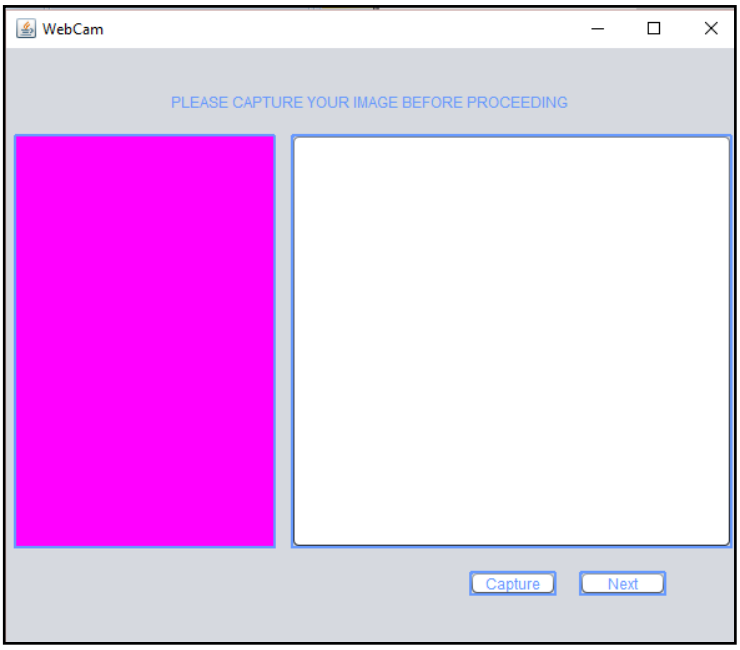


Figure 4.11 Web Camera

The next page provides user a chance to take a picture and use the webcam provided. The interface for the webcam is as in the Figure 4.11. User can either used the webcam or can just proceed for the next stage of the system. The system gives the chance for user to take the picture as prove that they used the system and also as motivation for them.

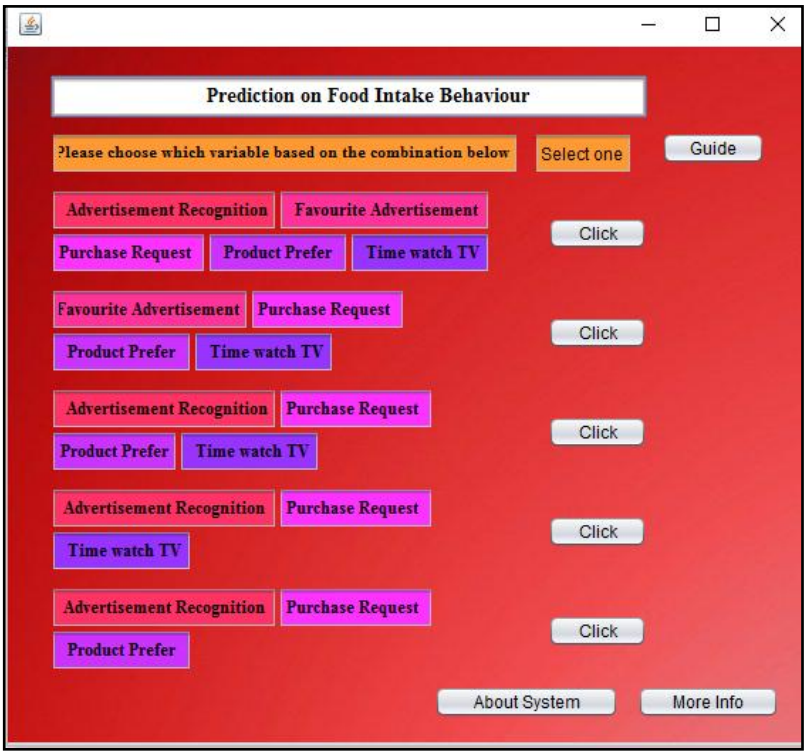


Figure 4.12 Main Menu

Figure 4.12 shows the system main menu. In this stage, the system provides option for the user as the user can choose different combination of different prediction model. The top best percentage result of prediction model has been used for the predictive system as shown on the figure. In this page, user can click the about system button for more information about the system then go to then page that is in Figure 4.13. Guide button is used to provide user more information about the system functionality and its combination of different prediction model as shown in Figure 4.14. Lastly the more info button provides is for user to get more information about their child health as shown in Figure 4.15.

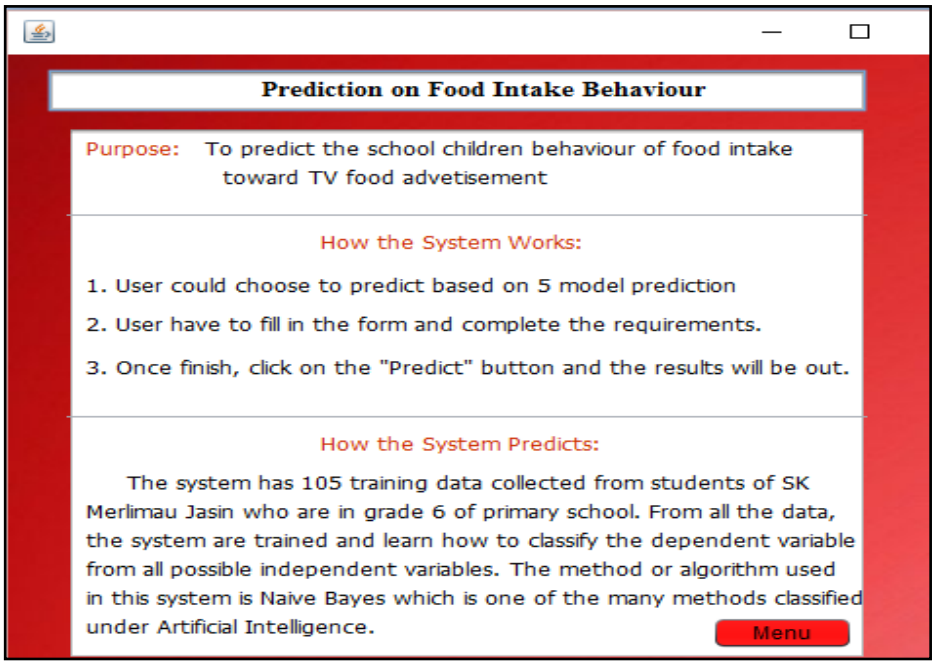


Figure 4.13 About System Page

Figure 4.13 shows the instruction and offer the knowledge of what the system can do.

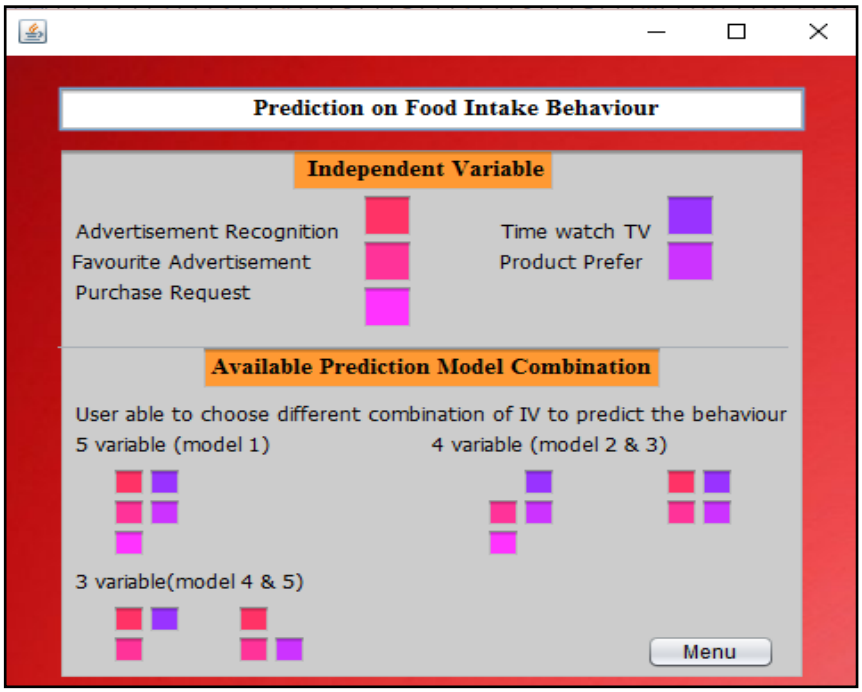


Figure 4.14 Guide page

Figure 4.14 shows the knowledge about the combination IV of study.

Figure 4.15 More Info Page

In Figure 4.15, this page shows that user can go directly to Kementerian Kesihatan page by only clicking the button. These give more valuable and trustable information provided by the Kementerian Kesihatan.

Figure 4.16 Prediction Model based on 5 Independent Variable

In Figure 4.16, this is where user can predict based on independent variable provided. User can choose and then predict by click the predict button and the result will shows at the bottom of the page. User can predict again by click the predict again button.

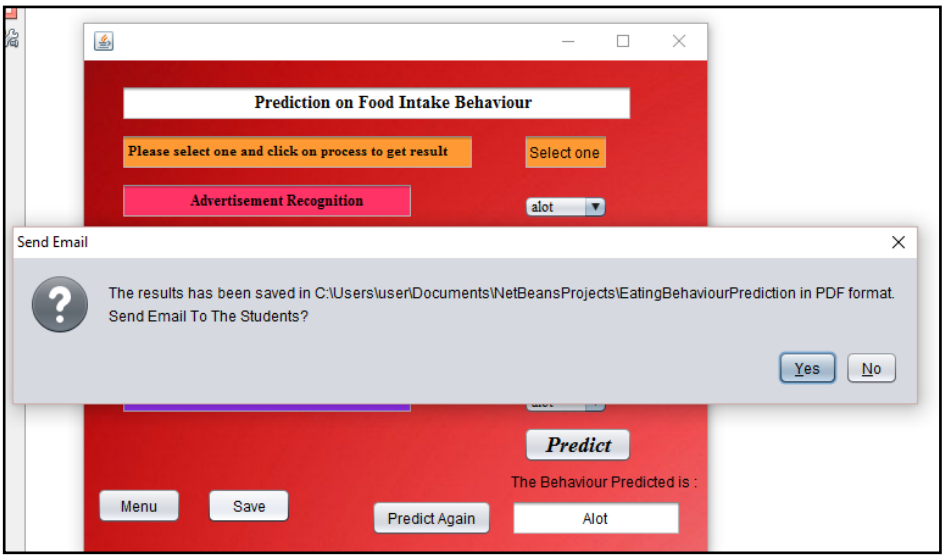


Figure 4.17 Save PDF and send email

After that, user can save the data predicted in PDF format while can either choose to send the result by email or not as shown in Figure 4.17 above.

4.6 Conclusion

In a conclusion, this chapter explained the design of prediction model and prediction system that were used in this project research. The development of system was challenging especially in finding the source code. However, it produces rapid prototyping system and saves time.

CHAPTER 5

RESULT AND FINDING

5.1 Introduction

In this chapter, data used for the prediction model and the result will be shown. All of the result achieved by testing using learning data that is 80% of the accuracy and other 20% is used for new data accuracy. There are 31 prediction model based on Naïve Bayes algorithm, and the testing and the accuracy should be 80% and above to show the best result. But in this case study, the highest accuracy recorded is shown as in Table 5.1. Therefore, the best highest 5 percentage of accuracy will be used in the system as the model of prediction.

5.2 Accuracy Testing

The performance of each model was evaluated by calculating the accuracy of the training data and the accuracy for new data testing. About a total of 105 data were collected. These data were collected from the same target subject. The data then will be used for the testing.

Accuracy percentage for training data was calculated from 85 data which constitute around 80% of the data collected from the sampling. The 85 data was picked randomly to be utilized as training data so that the training of the model can be various and not just centered on certain arrangement of the sampling. Meanwhile, accuracy % for new data was calculated from 20 data

which constitute around 20% from all data collected. These data are being tested to ensure the strength and accuracy of all models.

5.3 Result of Accuracy Table

Table 5.1 Accuracy Table

	Independent Variable					Percentage	
No.	advertisement recognition	favourite advertisement	purchase request	product prefer	time watch TV	80% (training Data)	20% (New Data)
1	✓					67.86	61.90
2		✓				66.67	61.90
3			✓			73.91	61.90
4				✓		67.86	61.90
5					✓	69.04	52.38
6	✓	✓				67.86	61.90
7	✓		✓			75.00	61.90
8	✓			✓		70.24	61.90
9	✓				✓	70.24	52.38
10		✓	✓			73.81	61.90
11		✓		✓		69.04	61.90
12		✓			✓	66.67	57.14
13			✓	✓		75.00	61.90
14			✓		✓	73.81	61.90
15				✓	✓	72.62	57.14
16	✓	✓	✓			75.00	61.90

17	✓	✓		✓		70.24	61.90
18	✓	✓			✓	67.86	57.14
19	✓		✓	✓		77.38	61.90
20	✓		✓		✓	72.62	57.14
21	✓			✓	✓	72.62	57.14
22		✓	✓	✓		76.19	61.90
23		✓	✓		✓	72.62	61.90
24		✓		✓	✓	72.62	57.14
25			✓	✓	✓	76.19	57.14
26	✓	✓	✓	✓		75.00	61.90
27	✓	✓	✓		✓	73.81	61.90
28	✓	✓		✓	✓	73.81	57.14
29	✓		✓	✓	✓	76.19	61.90
30		✓	✓	✓	✓	78.57	61.90
31	✓	✓	✓	✓	✓	78.57	61.90

Based on table 5.2, the 31 prediction models were established as sets of combination of five independent variables which include advertisement recognition (AR), favorite advertisement (FA), purchase request (PR), product preference (PP), and TV viewing duration (TW). All of these models are to predict eating behavior level that act as the dependent variable which consists of eat many of the product (A lot), eat in considerate amount(Average) , and not eat really often (Few).

The combination of all of the independent variables and one of the four combinations record the highest accuracy for both 80% training data and 20% new data with the accuracy percentage of 78.57% and 61.90%. This indicates

that the eating behavior level is more precise based on the accuracy testing result. However, there are several model recorded the result of the lowest for both 80% training data and 20% new data.

5.4 Functionality of System

When testing the functionality of system, it brings to the importance of each function included in system. Functionality testing had been done in order to check the system is working correctly based on the modules that have stated as in table 5.2.

Table 5.2 System functionality

Modules	**Expected results**	**Pass or Fail**
Run main page	Able to insert correct details of information needed.	Pass
Taking photo from webcam	The photo will be send into canvas in webcam page which can be save as photo in png format.	Pass
Run the menu page of the prediction	Run the system from index page from Naïve Bayes project in Netbeans.	Pass
Click each of combination of IV in main page	When choosing each the combination, user able to proceed to next session.	Pass
More Info Page	User able to go the each linked webpage.	Pass
Fill in the independent variables given	User must fill in the independent variables based on calcuator chose in order to predict body fat status.	Pass
Click 'Predict' button	User then will click the predict button to get the result after completely fill all requirement and result will be displayed.	Pass

5.5 Conclusion

As conclusion, Naïve Bayes is a method that can be predicted the behaviour of school children for food intake toward television advertisement. Prediction Model can be constructed by using Naïve Bayes approach. The best result shows percentage of accuracy of 78.57% indicate that the prediction model with all of the independent variable. Another prediction model only with four independent variables used except for the advertisement recognition, the percentage shares the same best result. Therefore, both model is conclude can be used for the prediction.

The aim of Naïve Bayes approach is to get the accuracy of above 80%. However, the highest recorded accuracy is only at 78.57%. Therefore, there a need to do a further research and study about the same topic study as there a need to get the accurate percentage on the result.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Introduction

This chapter will be the overall conclusion for the project. There are few limitation faced during development and study of the project. There are also need further study for future work about the same topic case for better understanding and a good result able to be produce.

6.2 Conclusion

Eating behaviour of school children predictive system can be good option to predict their behaviour based on TV food advertisement. This system is easy to understand since it is directly giving out the result of the behaviour and also know their status in managing healthy lifestyle after user completely enter all the requirement of independent variables. This system also helps user to differentiate the different result of behaviour based on different combination of independent variables. This brings the user some stability and flexibility in predicting the attitude in food intake. In addition, user can get the result of prediction in categorical value which is eat a lot, eat in average, and eat few and user are aware about their child's food intake attitude toward TV food advertisement. Moreover, this system is gave more understanding and also result percentage quiet accurate in performing the prediction model.

In a conclusion, this project was developed to help parent in predicting their children food intake behaviour. Nowadays, it is important to maintain a healthy life especially kids in growing stage, including taking care in food intake towards all health aspects. By using this system as alternative option, it can increase user interest in checking their children behavioural where it helps them to motivate themselves to be more aware about health. Basically, user can used the system to predict behavioural status without going to expert to get consultant or advised about their children food intake. Therefore, there is a need in predictive system to measure the value of behavioural children for food intake toward TV food advertisement as they can know their children attitude and can conduct a healthy childhood lifestyle.

6.2.1 Objective I

Identify and getting the independent variables to use in the prediction model is the first objective of the project. Data that is already gathered from participant in order to find the suitable variable for the topic study. The data information gathered is by searching and reading previous research, related articles and other source of knowledge. The objective is satisfied by taken previous research study topic and all related variables was induces for this topic and as the guided paper research of this project.

6.2.2 Objective II

Design and develop a system using AI approaches, Naïve Bayes prediction model was created and is the second objective of this project. Data collection was done in order to have the training for the prediction model. 31 sets of models were developed as to test the accuracy of each model and where only five best and highest accuracy percentages is used in the projects' system. Naïve Bayes is used as the classification algorithm for the prediction model. Therefore, to develop system smoothly, system flow is constructed and must be design properly.

6.2.3 Objective III

The third objective is to test the functionality of the system. Testing is done throughout the development and after the development all the function are properly functioning. This is objective is achieved when the system can be run and the result is shown in previous chapter 5 where there is the discussion on the system functionality.

6.3 Limitations

There are some limitations found during the development phases of this project. Firstly, lack of respondents in collecting data of independent variables. In order to get accurate result, a lot data needed to complete the prediction model. The reason that this study limited in target respondent is because of only one group of age 12 and only one primary school are is collected as the data. The next limitation is, the time taken throughout the study and development of this case study while doing in single-handled project. More time is needed to gather an accurate data to get more accurate result. Also needed is a helpful hand as a teammate that can be equally distributing the work load.

For prediction model, there were some combinations of independent variables with the dependent variables got lower result of accuracy percentage of training data and new data. The accuracy results got are still not understandable as the combinations are still vague.

6.4 Recommendations

The topic study has the potential to be expanded in future work. Therefore, the limitation state can be improvised and enhance to make it more interesting and more stable for the system. For example, the developer can increase the amount of data from getting more respondents, changing the independent variables with more reliably which will give higher result of accuracy percentage instead of making new research and also by using other AI techniques such as Support Vector Machine (SVM) to get accurate results.

This system also can be developed in mobile based in the future. The current version are being established is by using netbeans,so in future it can be settles in mobile application as it is more easy to carry and can be used anytime as today is which everyone owned a smart phones, tablets, mobile devices, and other technology devices. Thus, it can increase the ease of this system without carrying laptop every day.

REFERENCES

Baker, R. D., & McHale, I. G. (2016). An empirical Bayes model for time-varying paired comparisons ratings: Who is the greatest women's tennis player? *European Journal of Operational Research*, *0*, 1–6. http://doi.org/10.1016/j.ejor.2016.08.043

Bardak, S., Tiryaki, S., Nemli, G., & Aydin, A. (2016). Investigation and neural network prediction of wood bonding quality based on pressing conditions. *International Journal of Adhesion and Adhesives*, *68*, 115–123. http://doi.org/10.1016/j.ijadhadh.2016.02.010

Bell, D., & Parr, M. (n.d.). *Douglas bell & mike parr*.

Bilal, M., Israr, H., Shahid, M., & Khan, A. (2015). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and {KNN} classification techniques. *Journal of King Saud University - Computer and Information Sciences*, *28*(3). http://doi.org/http://dx.doi.org/10.1016/j.jksuci.2015.11.003

Birch, L. (1999). Development of food preferences. *Annual Review of Nutrition*. Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.nutr.19.1.41

Boyland, E. J., & Halford, J. C. G. (2013). Television advertising and branding. Effects on eating behaviour and food preferences in children. *Appetite*, *62*(October 2016), 236–241. http://doi.org/10.1016/j.appet.2012.01.032

Boyland, E. J., Harrold, J. A., Kirkham, T. C., & Halford, J. C. G. (2012). Persuasive techniques used in television advertisements to market foods to UK children. *Appetite*, *58*(2), 658–664. http://doi.org/10.1016/j.appet.2011.11.017

Byrne, R., Magarey, A., & Daniels, L. (2014). The "Good Eater": modelling maternal perception of food fussiness in Australian toddlers–the NOURISH and SAIDI cohorts. Retrieved from http://eprints.qut.edu.au/78033/

Cairns, G., Angus, K., & Hastings, G. (2009). The extent, nature and effects of food promotion to children: A review of the evidence to 2008, (December). Retrieved from http://www.who.int/dietphysicalactivity/Evidence_Update_2009.pdf

Catal, C., Sevim, U., & Diri, B. (2011). Practical development of an Eclipse-based software fault prediction tool using Naive Bayes algorithm. *Expert Systems with Applications*, *38*, 2347–2353. http://doi.org/10.1016/j.eswa.2010.08.022

Caton, S. J., Blundell, P., Ahern, S. M., Nekitsing, C., Olsen, A., Møller, P., … Hetherington, M. M. (2014). Learning to eat vegetables in early life: The role of timing, age and individual eating traits. *PLoS ONE*, *9*(5). http://doi.org/10.1371/journal.pone.0097609

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. article. http://doi.org/10.1007/BF00994018

Folkvord, F., Anschütz, D. J., Boyland, E., Kelly, B., & Buijzen, M. (2016). Food advertising and eating behavior in children. *Current Opinion in Behavioral Sciences*, *9*, 26–31. http://doi.org/10.1016/j.cobeha.2015.11.016

Halford, J. C. G., Boyland, E. J., Hughes, G., Oliveira, L. P., & Dovey, T. M. (2007). Beyond-brand effect of television (TV) food advertisements/commercials on caloric intake and food choice of 5-7-year-old children. *Appetite*, *49*(1), 263–267. http://doi.org/10.1016/j.appet.2006.12.003

Johnson, S., McPhee, L., & Birch, L. (1991). Conditioned preferences: young children prefer flavors associated with high dietary fat. *Physiology & Behavior*. Retrieved from http://www.sciencedirect.com/science/article/pii/003193849190590K

Karupaiah, T., Chinna, K., Mee, L. H., Mei, L. S., & Noor, M. I. (2008). What's on Malaysian television? - A survey on food advertising targeting children. *Asia Pacific Journal of Clinical Nutrition*, *17*(3), 483–491.

Kern, D., McPhee, L., Fisher, J., Johnson, S., & Birch, L. (1993). The postingestive

consequences of fat condition preferences for flavors associated with high dietary fat. *Physiology & Behavior*. Retrieved from http://www.sciencedirect.com/science/article/pii/003193849390045H

Kotsiantis, S. B. (2007). Supervised Machine Learning : A Review of Classification Techniques, *31*, 249–268.

Laska, M. N., Hearst, M. O., Lust, K., Lytle, L. A., & Story, M. (2015). How we eat what we eat: identifying meal routines and practices most strongly associated with healthy and unhealthy dietary factors among young adults. *Public Health Nutrition*, *18*(12), 2135–45. http://doi.org/10.1017/S1368980014002717

Li, D., Wang, T., Cheng, Y., Zhang, M., Yang, X., Zhu, Z., … Zeng, L. (2016). The extent and nature of television food advertising to children in Xi'an, China. *BMC Public Health*, 1–9. http://doi.org/10.1186/s12889-016-3468-0

Marucci-Wellman, H. R., Lehto, M. R., & Corns, H. L. (2015). A practical tool for public health surveillance: Semi-automated coding of short injury narratives from large administrative databases using Na??ve Bayes algorithms. *Accident Analysis and Prevention*, *84*, 165–176. http://doi.org/10.1016/j.aap.2015.06.014

Meko, D. (2015). Notes 11 - Multiple Linear Regression, (Weisberg 1985), 1–11.

Mellit, A., Kalogirou, S. A., Hontoria, L., & Shaari, S. (2009). Artificial intelligence techniques for sizing photovoltaic systems: A review. *Renewable and Sustainable Energy Reviews*, *13*(2), 406–419. http://doi.org/10.1016/j.rser.2008.01.006

Mia, M., & Dhar, N. R. (2016). Prediction of surface roughness in hard turning under high pressure coolant using Artificial Neural Network. *Measurement*, *92*, 464–474. http://doi.org/10.1016/j.measurement.2016.06.048

Ng, S. H., Kelly, B., Se, C. H., Sahathevan, S., Chinna, K., Ismail, M. N., & Karupaiah, T. (2015). Reading the mind of children in response to food

advertising: a cross-sectional study of Malaysian schoolchildren's attitudes towards food and beverages advertising on television. *BMC Public Health*, *15*(1), 1047. http://doi.org/10.1186/s12889-015-2392-z

Policy, E. (n.d.). Education Systems in ASEAN + 6 Countries : A Comparative Analysis Selected Educational Issues.

Powell, L. M., Chriqui, J. F., Khan, T., Wada, R., & Chaloupka, F. J. (2013). Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: A systematic review of prices, demand and body weight outcomes. *Obesity Reviews*, *14*(2), 110–128. http://doi.org/10.1111/obr.12002

Quick, V., Wall, M., Larson, N., & Haines, J. (2013). Personal, behavioral and socio-environmental predictors of overweight incidence in young adults: 10-yr longitudinal findings. *International*. Retrieved from http://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-10-37

Ruini, L., Ciati, R., Marchelli, L., Rapetti, V., Pratesi, C. A., Redavid, E., & Vannuzzi, E. (2016). Using an Infographic Tool to Promote Healthier and More Sustainable Food Consumption: The Double Pyramid Model by Barilla Center for Food and Nutrition. *Agriculture and Agricultural Science Procedia*, *8*, 482–488. http://doi.org/http://dx.doi.org/10.1016/j.aaspro.2016.02.049

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. *Neurocomputing* (Vol. 9). http://doi.org/10.1016/0925-2312(95)90020-9

Smith, C., McGuire, B., Huang, T., & Yang, G. (2006). The History of Artificial Intelligence. *University of Washington*, (December), 27.

Sommerville, I. (2010). *Software Engineering*. *Software Engineering*. http://doi.org/10.1111/j.1365-2362.2005.01463.x

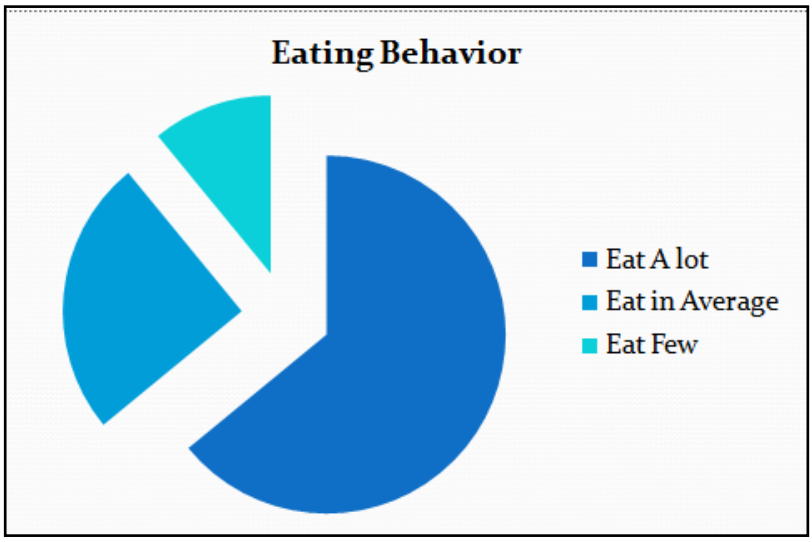
Suryn, W. (2014). *Software Quality Engineering: A Practitioner's Approach*. *Software Quality Engineering: A Practitioner's Approach* (Vol. 9781118592). http://doi.org/10.1002/9781118830208

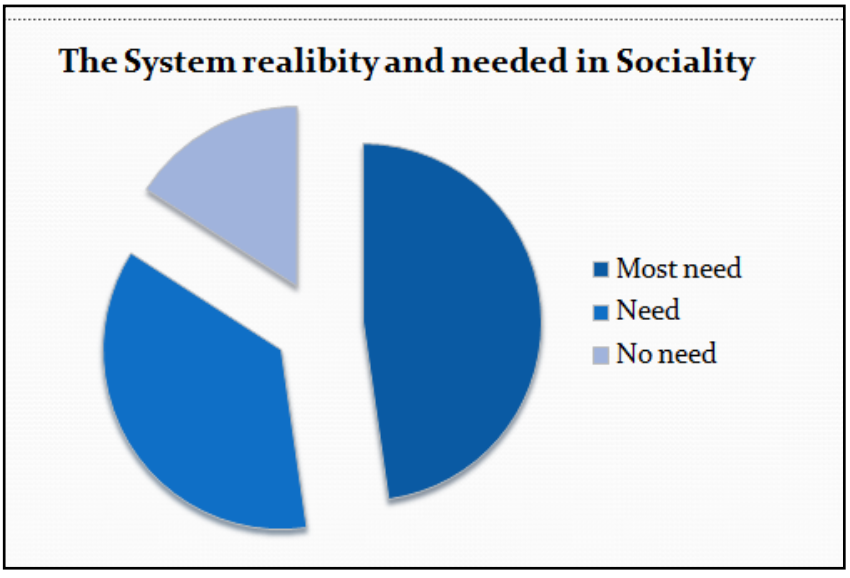
Xu, W., & Dong, L. (2016). A novel relative density based support vector machine. *Optik - International Journal for Light and Electron Optics*, *127*(22), 10348–10354. http://doi.org/10.1016/j.ijleo.2016.08.027

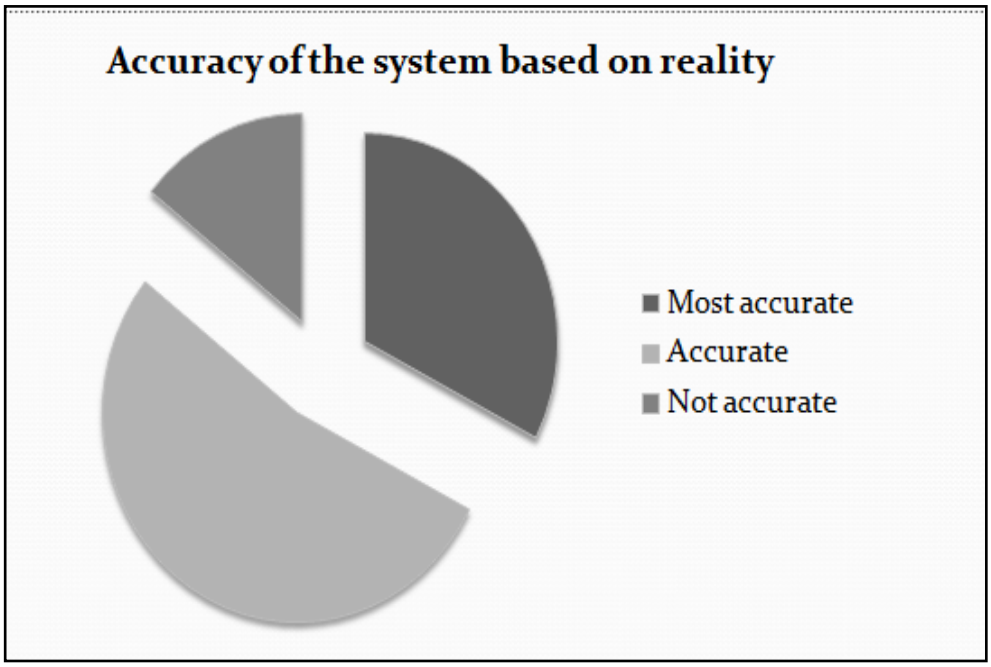
Zizza, C., Sebastian, R., Enns, C., & Zeynep, I. (2015). The Contribution of Beverages to Intakes of Energy and MyPlate Components by Current, Former, and Never Smokers in the United States. *Journal of the Academy*. Retrieved from http://www.sciencedirect.com/science/article/pii/S221226721501196X

APPENDICES

APPENDIX A: SURVEY ANALYSIS






Accuracy of the system based on reality
Most accurate
Accurate
Not accurate