Dietary Monitoring System Using Decision Tree to Control Human Obesity

Mohamad Faiz Mat Baseri1, Ahmad Fadli Saad1

1 Computing Science Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA. Perak Branch, Tapah Campus. Perak, Malaysia.

Abstract: Nowadays, obesity is one of the dangerous diseases in the world. Lack of dietary monitoring system will make it difficult for people with obesity to reduce their weight problems. The main objective of this research is to develop a dietary monitoring system that can be used by everybody especially for obesity’s people. The method used in this study is to identify the strength and weaknesses of the existing system which involves reviewing some articles, journals, magazines, and books. The survey was conducted which involves 10 people answering the questionnaire. Respondent answered was used to improve the quality of the system. Next method is utilizing a waterfall model as a method to develop a dietary monitoring system. The system applied the decision tree technique to a classified food calorie. Therefore, the decision tree technique was used in developing this system. The last method used in this study involves the participation of three respondents to evaluate the usability of the system. Respondents need to answer whether they satisfied with the system or they can give suggestions for future improvement. The results of this study show that obesity is a public health issue that is rapidly increasing and must be addressed seriously by developing the system. Significant by developed this system such as helping obesity’s people to diet by giving them the guideline. In conclusion, this system will help people to diet using the decision tree technique for classifying food calories.

Keywords: Decision Tree, Dietary, Monitoring System, Obesity.
1. Introduction

Overweight and obesity are classified as abnormal or unwanted fat that will cause risk to the body. A standard height and weight index are the Body Mass Index (BMI), which is commonly used to classify underweight, overweight, and obesity. Statistic to World Health Organization (WHO), overweight and obesity are global concerns, and more than 2 billion adults worldwide are predicted to be overweight and more than seven hundred million were obese. Obesity is defined as the condition of excessive fat accumulation to such an extent that affects the individual’s health. Obesity is a very dangerous disease because it can cause more easily for obesity’s people to get other diseases. Statistics have shown that between 1975 and 2016, the global incidence of obesity nearly tripled, and it is currently estimated that more than 2 billion people are suffering from excess body weight, and current predictions reveal that obesity will continue to grow until at least 2030. People who are obese find it difficult to control their weight because there is no suitable diet guide for them to follow. Because of that, some of the obese people cannot control their weight and become stressed when thinking about this problem. Most of the existing guides still use manual methods and obese people are given less exposure to technological systems. Mobile technology has a good potential in the process of facilitating weight loss management. The potential of mobile health systems to facilitate their ability to lose weight through strategies such as self-monitoring. Many benefits are gained by developing this system such as helping obese people lose weight. This research aims to develop a dietary system to help obesity’s people. Therefore, developing a mobile system is a way to control obesity. Since the major factor of obesity from food intake, the system may use the decision tree technique to classify food calories. In conclusion, this system will help obesity’s people to diet using the decision tree technique.

Nowadays, the rate of obesity increase rapidly around the world. Therefore, it is supposed to have a mobile system that can help people to diet to control their obesity. Mobile phone application systems, which are easy to get and available have been used to fix this problem [1]. However, there have also been questions regarding data quality and privacy in these systems, considering that mHealth systems are largely uncharted territories. Many of these systems, when created, have little to no input from specialists in dietary or physical activity.

People also tend to get obese since some of them need to stay at home during this pandemic COVID-19 [2]. People with overweight bodies are easy to get other chronic diseases. Obesity and being overweight have a significant impact on morbidity and mortality from various diseases such as type 2 diabetes, cardiovascular diseases, and metabolic syndrome [1]. More attention is needed since excess body weight is a serious global problem since it is likely to be related to the severity of the current COVID-19 pandemic.

However, the lack of dietary monitoring system will make it difficult for people with obesity to reduce their weight problems. For example, some apps can provide users with information on their percentage of body mass index (BMI) or body fat and help them monitor changes. Since the obesity disease increasing quickly cause of Pandemic Covid-19, it is more motivated to develop the system. Therefore, developing dietary monitoring system using the decision tree technique will guide obesity’s people to reduce their weight.

The scope of this research focusing on obesity’s people since this research was developed to help them getting the ideal body. Next, this research will implemented on the mobile application platform and also can open in the desktop. After that, this system will develop using the machine learning method which is a decision tree technique that can classify the data. Lastly, this research focusing on food intake and its calorie.

2. Literature Review

2.1. Obesity

Human obesity is complex and typically results from the interplay of genetic factors. For example, In Japanese SUMO Wrestlers, cases of obesity are caused predominantly by a high-fat diet. Genetic obesity is mostly due to a shift in the role of vital hypothalamus neuronal populations, the central brain region controlling energy and weight homeostasis [3]. From the research, Statistics indicate that obese and overweight diseases raise the risk of dying by nearly 50% compared with average weight subjects [4]. In obese cancer patients, a poor prognosis may also be associated with a reduced response to treatment either due to adipocyte-promoted resistance to chemotherapy or radio resistant
phenotypes [5]. Obesity is characterized as diet-derived excess fatty acids and an excessive increase in adipose tissue. Therefore, obesity is one of the dangerous diseases that should people overcome.

2.2. Machine Learning
Machine learning is a method that teaches the computer to learn with data given and aims to model deep data input relationships and recreate a knowledge system. Estimation, prediction, and classification are some examples of products of learning that can be used to make a decision. It enables prediction and classification from the dataset and can achieve high decision accuracy [6]. There are two types of machine learner supervised and unsupervised. Cox et al [7] in their research describe that “machine learning is supervised if the values it produces during training are checked against some target values, and the learner is provided with feedback while the unsupervised learner is not guided by such feedback and organize the data without explicit guidance.” Supervised learning may divide into problems with classification and regression. The classification method is used to predict discrete information while the regression method is used to predict ongoing data. Therefore, in the medical field, machine learning like deep learning has been increasingly used [8].

Two types of techniques are used in machine learning. First, supervised learning trains a model on known input and output data that can predict future outputs. Second, unsupervised learning seeks hidden patterns in input data or intrinsic structures.

2.2.1. Neural Network
A machine learning system that utilizes a network of functions to recognize and transform a single type of data input into the desired output is an artificial neural network learning algorithm or neural network. Human biology and the way human brain neurons function together to understand the inputs of human senses became the inspiration behind the creation of the artificial neural network. The neural network can be used in different machine learning algorithms as a component to process complex data inputs into a space that computers can understand. Many real-life issues such as image recognition and spam email filtering are being applied to neural networks. All the neurons influence each other in a Neural Network and they are all related. Every aspect of the dataset at hand and how the various sections of data may or may not relate to each other can be recognized and observed by the network. This is how Neural Networks in large data volumes can find highly complex patterns.

![General Structure Neural Network](image)

Figure 1. General Structure Neural Network

There is some benefit of the neural network. The first is the ability to operate with inadequate information. The performance generated by the data may be incomplete or insufficient after ANN training. The meaning of the missing information determines the lack of results. Next, strong tolerance for lack. The generation of output is not impaired by the corruption of one or more artificial
neural network cells. This makes it easier for the networks to tolerate defects. Finally, incremental corruption. A network slows over time and undergoes relative degradation. The network problem does not corrode instantaneously. The disadvantage of neural networks is hardware reliance. Artificial neural networks require processors with parallel processing power according to their structure. Then, the perception of the equipment is dependent. Therefore, the duration of the network is unsure. The network is limited to a certain sample error value, meaning that the training is completed. This value does not provide us with optimum efficiency.

Kumar et al [9] used machine learning techniques to perform Software Effort Estimation (SEE) and observe that “In terms of precision, machine learning approaches can surpass conventional methods. Generally, SEE is focused on strategies such as expert judgment, algorithmic models, and different methods of machine learning.” In his research, for machine estimating, he uses the ANN method. The complex relationship between the input and output of neural networks is recognized to produce accurate estimation as they map in a nonlinear fashion. Therefore, many researchers have used neural networks to make a prediction [9]. Several types of neural network models have been used to improve the estimation accuracy, such as feed-forward neural network (FFNN) with a back-propagation learning algorithm, neural network radial basis function, and general neural network regression (GRNN).

A Feed Forward Neural Network is a simple ANN that connects the neurons in such a way that, from the starting layer that is the input layer through the hidden layers, the information flow is only one way forward and eventually reaches the output layer.

Based on Figure 2, a network of Radial Basis Functions consists of 3 layers, which are layers of input, secret, and output. The network output is the total number of weighted outputs from the hidden layers and the hidden layer radial basis function. Each neuron corresponds to a predictor variable in the starting layer.

![Figure 2. Radial Basis Function Network](image)

Next, General Regression Neural Network (GRNN) consists of four-layer. The first layer consists of the neuron as a predictor variable. The second layer consists of a pattern neuron representing the training row. The third layer consists of two numerator and denominator node nodes. The number is the amount of weight multiplied by the actual output, while the denominator is the sum of weight from previous nodes. Figure 3 shows the architecture of GRNN.

Besides that, based on Nayak et al [10] study, it uses a neural network in food processing. ANN plays a vital role in processing food products such as fruits, vegetables, fish, and various food grains, as well as soft drinks, for food-based systems. The Neural Network plays a major role in fruit...
production, such as bananas, in fruits. Quality estimation of fruit is needed due to the difficulty of understanding the perceptions of current consumers. Nayak et al [10] have introduced an artificial neural network modeling framework for various drying of bananas' antioxidant activity and phenolic compounds. The purpose of this study is to investigate the impact of aeration conditions on the overall activity in bananas of antioxidants and phenolic compounds. The lower content of phenolic compounds and antioxidant activity is the result of this experiment.

Next, the neural network was used in vegetables to estimate carrot dynamic drying efficiency. This study showed that the neural network functions can be used for drying process control and online state estimation. Then, based on Nayak et al [10] study, the neural network approach is also used for predicting fish species in African lakes. This research uses backpropagation as a learning method for their proposed ANN model. Researchers also compare ANN and Discriminant Function Analysis (DFA) during this research and find that neural network accuracy is higher than DFA. Nowadays, for classification and prediction, the neural network is applied in many industries.

2.2.2. Support Vector Machine
The Support Vector Machine (SVM) is an algorithm for machine learning and is a type of algorithm supervised for learning and is used for classification purposes in many systems. The SVM tool utilizes the training dataset to create the algorithm for optimum accuracy. First, the SVM learns the behavior of the dataset and then generates a prediction hypothesis. Two-step classification, include studying the information and then testing the information [11].
There are a few advantages of SVM which are SVM works relatively well when there is a clear margin of distinction between groups. Next, SVM is more effective in high-dimensional spaces. Then, SVM is effective in cases where the number of dimensions is greater than the number of samples. The disadvantage of SVM is the algorithm is not optimal for the massive data set. Then, when the data set has more noise, SVM does not work very well, such as overlapping goal groups. Lastly, the SVM will not perform very well if the number of characteristics for each data point exceeds the number of training data samples.

In his research, Wang et al [12] intend to comprehensively examine the classification methods of detecting SVM machine learning. In this analysis, SVM was implemented in a QAM-DMT optical transmission connection based on the Mach- Zehnder modulator (M-ZM) and 10-km standard single-mode fiber (SSMF), so that different QAM formats can be loaded onto different carriers. Quadrature amplitude modulation (QAM) formats are very good for increasing transmission capacity due to highly improved spectral efficiency. The experimental results indicate that the SVM methods can well detect the non-linear damage of the signal in the case of a back-to- back and 10-km transmission. The high-order QAM signal makes full use of the amplitude and stage to carry the data, resulting in a substantial improvement in the capability of transmission. Furthermore, for all two training sets with different categories, SVM multi-classification based on one versus one (OvO SVM) provides an optimal decision function. The advantage of OvO SVM is that the number of samples per session is relatively small so that the training speed of a single decision surface is faster and the accuracy is greater [12].
SVM multi-classification may be based on the binary tree [12]. In his research, it is describe binary tree-based SVM multi-classification. The category contained in the node is split into two subclasses beginning from the root node, and the subclass is split into two subclasses until each has one category. The SVM classifier was trained to identify decision nodes in each binary tree. The complete tree requires a few SVM numbers, so the classifier speed of a complete binary tree with fewer support vectors is also faster.

2.2.3. Decision Tree
A decision tree is a diagram or graph that people use to decide a course of action or to show a statistical probability. It's tree-shaped. Any branch of the decision tree reflects a possible decision, outcome, or reaction. The outcomes are reflected by the tree's farthest branches. The decision framework allows users to take a problem with multiple potential solutions and view those solutions in a way that is clear or easy to understand, and often shows the relationship between different events or choices. Each result in the decision tree has an allocated risk and reward weight or number. If a person uses a decision tree to make a decision, they look at each outcome and compare the benefits and disadvantages. The tree itself can span as long or as short as needed to come to a proper conclusion.

![Decision Tree Architecture](image)

There are some advantages and drawbacks of the decision tree. Compared to other algorithms, a decision tree takes less time during pre-processing for data preparation. Next, a decision tree does not need normalization of data and does not need scaling of data as well. Then, the process of making a decision tree is therefore not dramatically impacted by missing values in the data. A weakness of the decision tree is that a minor change in the data can cause a large change in the layout of the decision tree that creates confusion. Then, to apply regression and forecast continuous values, the Decision Tree algorithm is not enough.

Based on the study, the Decision Tree provides an efficient model for data classification and regression. In his research, Chen et al [13] claimed that "the classification of the decision tree can be considered as a technique of "if-then" and involves three components that are feature selection, decision tree construction and decision tree pruning. Node and guided edge are also included in the decision tree.

Figure 8 shows the schematic diagram of the decision tree. The internal node represents the attribute or function and the leaf node is the class. Based on the learning algorithm, which selects the features from the root node, the decision tree classification method can be specified. Then, all root node data is split into two sub-nodes, with one group identified by each sub-node. Finally, when all root nodes are divided into sub-nodes, this process stops. Based on Chen et al [13] in their research found that a “decision tree” is based on the number of nodes or alternatives per decision. When the
number of choices is limited, a large number of alternatives can always be understandable. For instance, even if 30 different income intervals are specified, a decision tree containing a single decision that assigns subjects an income tax rate according to their income interval is simple to understand.”

![Decision Tree Schematic](image)

**Figure 8. Schematic of Decision Tree**

Besides that, Alvarez et al [14] in their research used a hybrid model of decision tree to predict the quality of water. There are many benefits to the decision-tree-based computer model, such as the ability to manage both data and standard features, insensitivity to missing values, and high performance. This study also found that, compared to ANN and SVM, decision tree-based models might have faster calculation speed and are more conducive to short-term prediction.

### 2.2.4. Naive Bayes

Naive Bayes is a classification technique based on Bayes' Theorem and the assumption of predictor independence. In other words, a Naive Bayes classifier assumes that the presence of one feature in a class has no bearing on the presence of any other feature. The Naive Bayes model is simple to construct and is especially useful for very large data sets. In addition to its simplicity, Naive Bayes has been shown to outperform even the most sophisticated classification methods [15].

![Bayes Theorem](image)

**Figure 9. Bayes Theorem**

### 2.3. Comparison Technique

Table 1 shows a good choice to apply decision tree technique in this research which are easier to understand compared to other technique. Decision tree also good in classifying and prediction technique. The data trained in decision tree also faster compare to other technique. It means the result will get more faster when using decision tree. Lastly, accuracy of decision tree also good for getting best result.
Table 1. Comparison Technique

<table>
<thead>
<tr>
<th>Decision Tree</th>
<th>Neural Network</th>
<th>Support Vector Machine</th>
<th>Naïve Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision trees have a natural flow that's easy to observe. They are also simple to program with if, then, and else statements for computer systems.</td>
<td>It is not easy to understand the neural network from visual representation. Creating computer systems from them is very difficult, and it is almost impossible to produce an interpretation from the model.</td>
<td>Decision trees have a natural flow that's easy to observe. They are also simple to program with if, then, and else statements for computer systems.</td>
<td>It is simple and quick to predict the class of test data set. It also excels at multi-class prediction.</td>
</tr>
<tr>
<td>More faster when train the data</td>
<td>Slower and less interpretable</td>
<td>Speed based on dataset. Slower when train a large of data.</td>
<td>Perform well in categorical input compare to numerical input</td>
</tr>
</tbody>
</table>

3. Methodology
Figure 10 shows the research design and methodology of the research. It consists two-phase which are preliminary study on the existing application and a study on developing dietary monitoring mobile application system.

Waterfall model used in developing dietary monitoring system using decision tree technique as stated in Phase-2. The waterfall model has seven-phase i.e., planning, defining requirements, design, software development, testing, deployment, and maintenance.
Phase 1: Planning
In the Planning phase, evaluate the terms and needs of the research. This includes calculating all material costs, creating a timetable with target goals, and creating the research’s teams and leadership structure. Planning also includes feedback from user. User are anyone who stands to benefit from the application. Try to get feedback from expert user, developers, and subject matter experts.

Phase 2: Define Requirements
Defining requirements is considered part of planning to determine what the application is supposed to do and its requirements. This research requires all information about calories needed and the total calorie in each food. This information gets from the internet. When the system was developed, the user's personal information such as name, height, weight, and gender was required for registration into the system.

Phase 3: Design
The design phase models the way a research will work. The design include system architecture, decision tree algorithm and flowchart of the system.

Phase 4: Software Development
This is the actual writing of the program. This research will be developed by single since this is small research. This phase will use an Access Control or Source Code Management application. These systems help developers track changes to the code. The coding process includes many other tasks. The coding will writing using java in android studio. Many things need to brush up on skill. Finding and fixing errors and glitches is critical. Tasks often hold up the development process, such as waiting for test results or compiling code so an application can run. Documentation can be a formal process, including wiring a user guide for the application. It can also be informal, like comments in the source code that explain why a developer used a certain procedure. Even companies that strive to create software that’s easy and intuitive benefit from the documentation. Documentation can be a quick guided tour of the application’s basic features that display on the first launch. In this research, the documentation was write in report. It tell how the software was develop.

Phase 5: Testing
It’s critical to test an application before making it available to users. Much of the testing can be automated, like security testing. Even though, the testing will done for usability testing. Give try for three user and ask them either they satisfy or not. This phase helps reduce the number of bugs and glitches that users encounter. This leads to a higher user satisfaction and a better usage rate.

Phase 6: Deployment of system
This research will focus on developing an image classifier model. The process of converting the whole requirement into a product, the code will be writing and compile. The development of the model started with pre-processing all the images. The training and testing of the images will be done locally using the terminal which results in producing a classifier on the local system. TensorFlow will be used in this research as an open source for deep learning tasks. As for the coding part of the model, the language that will be used is python3. Figure 3.8 shows the development phase of this research.

Phase 7: Operation and Maintenance
This research focus on development of the mobile application system. For future, this system will doing the maintenance if have some error new improvement in advancing this system.

4. Finding and Discussion
Figure 12 shows interface the dietary monitoring system. Figure 12.A show register interface design for dietary monitoring application system. User need to login first before going to the next page. When log in to the system, the register interface will showed. User need to register their personal information such as name, age, height, weight, and gender. Lastly, press the submit button.

The calories counter interface will showed calories needed by the user, as shown in Figure 12.B. It also show date, day and has feature to snap the picture. In Figure 12.C, food info interface show the picture was snap using the camera. Then the system classify food details and show to the user. User can know the food calories, carbs, proteins, and fats. While Figure 12.D shows some of the food calories that has been added as a record how much users take calories in a day.
A B C D

Figure 12. Interface Design

The usability evaluation for this system using three person. Liaw [18] (2008) emphasised that usability evaluation could produce the best result from the usability testing if not more than five experts participated at one particular time. The usability evaluation has been used to solve problems in software selection. The respondent were contacted face to face or through telephones. The usability evaluation of the e-LSO was conducted on a one-to-one basis. The three person who agreed to take part in the dietary monitoring system usability evaluation were given through face to face and date for conducting the evaluation was given.

The result of the usability of the dietary system has been reported. Three person were invited to test the tool in the usability evaluation of the DMS. These persons were newly selected and were not involved in developing the system. The three person were required to provide information about their existing evaluation practices and they need to answer the questionnaire about satisfying of the system. The questionnaire given contain five question and they need to answer all the question as the feedback for the system. The questionnaire question can see in appendix B. The feedback from the three person showed that this system in good condition. They also agree by using the decision tree technique was help in determines food calories and suggested food. From the findings of this study, it can be concluded that the usability is good and can use safely by the every user.

Table 2. Interface Functional Testing

<table>
<thead>
<tr>
<th>No</th>
<th>Interface</th>
<th>Expected Function</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Register</td>
<td>User should login to the system</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Calorie Counter</td>
<td>All the button can function well</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Food info</td>
<td>Show the right result</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Camera</td>
<td>Can function Well</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Recommended food</td>
<td>Show recommended food to user</td>
<td>Yes</td>
</tr>
</tbody>
</table>

5. Conclusion
Based on the testing from the user, it shows this system is good for an obese person to diet. It was discovered that the food calories classification using a decision tree improved the system quality. The usability system was tested by asking three-person to use the system. After that, the questionnaire was given to them to ask whether the quality of this system was good or not. Their answer has been recorded for future improvement.
Testing the system is used to evaluate the output of a software program by giving suitable input and comparing it to the functional requirements. The purpose of testing the system was to appraise whether the system can work well or not. This testing covers and examines the Application under Test’s user interface, database and other features. Testing were be carried by the own self using the mobile phone.

The register interface function well. User can register to the system by submitting their personal information such as name, age, height, weight and gender. In the calorie counter interface, it will show calories needed and calories consume based on their food intake. Next, the food info interface will show food calories, carbs and proteins. Then, the camera interface function well to snap the picture. Lastly, recommended food for dietary function well to recommend food for diet to the user. The limitation of this research is that the search process might have missed some relevant papers since there is a lack of research has been conducted in the developing obesity dietary control systems. Besides that, not all food can be snap by using this system. Only food that was been training for the system can check its calorie.

References


