Data Mining Approach on Assessing the Human Motion Balancing Mechanisms

By,

KOGULAN A/L LETCHUMANAN

(Matrix no : 120378)

Supervisor:

Dr.Loh Wei Ping

May 2017

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

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

DECLARATION

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

Signed _____ (Candidate)

Date _____.

Statement 1

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

Signed _____ (Candidate)

Date _____.

Statement 1

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

Signed _____ (Candidate)

Date _____.

Acknowledgement

First of all, I would like to thank my Final Year Project (FYP) supervisor Dr Loh Wei Ping and who always be the guidance for me to complete this project. The time and effort that she inputted had been the greatest motivation for me to complete this project. I would like thank Dr. Mohamad Ikhwan Zaini Ridzwan who the coordinater for FYP to allow me to conduct this project.

I would also take this opportunity to thank all the technicians that have assisted me especially Mr. Baharom, who is the technician in charge of the Proton Lab. Without his contribution, I couldn't have conducted the experiment for FYP.

Finally, my sincere thanks to my friends and most importantly my parents for their continuous support and motivation. It is by far with their combined presence that I have made it this far. Thank you to everyone who helped me in this part of my life.

Title of Project: Data Mining Approach on Assessing the Human Motion Balancing Mechanisms

Date of Submission (Academic Year): 22th June 2017 (2016/2017)

Candidate (Matric No.): Kogulan A/L Letchumanan (120378)

Name of Supervisor: Dr. Loh Wei Ping

TABLE OF CONTENTS

DECLARATION II

ACKNOWLEDGEMENT III

TABLE OF CONTENTS IV

LIST OF FIGURES V

ABSTRAK VI

ABSTRACT VII

1. Introduction

2. Methodology

2.1 Introduction

2.2 Data collection

2.3 Data preprocessing

2.4 Data processing

2.5 Knowledge discovery

3. Result & Discussion

4. Conclusion

5. References

List of Figures

Figure 1 Field of view for an in-line came to the object.....14

Figure 2 Example of images snapped for 1st inclination angle and 2nd Inclination angle.....15

Figure 3 Example of image snapped from 3rd inclination angle.....15

Figure 4 Example of decision tree obtained from Weka software.....16

Figure 5 Examples of result obtained from Weka software using decision tree 17

Figure 6 Examples of graph obtained from Microsoft Excel 18

Figure 7 Stability clasification result.....18

Figure 8 Decision tree for stability classification 19

Figure 9 Classification for distance between both legs 20

Figure 10 Decision tree for distance between both legs 20

Abstrak

Pendekatan perlombongan data boleh digambarkan sebagai proses menukar saiz besar dan tidak terkawal data ke dalam bentuk data yang lebih mudah dan berguna. Projek ini menggunakan pendekatan penambangan data untuk mencapai kestabilan manusia apabila gerakan terlibat. Dalam projek ini, kita melibatkan aktiviti-aktiviti yang dilakukan oleh manusia yang normal untuk mengetahui postur sempurna untuk mereka melakukan aktiviti itu. Perlombongan data terdiri daripada beberapa peringkat iaitu pengumpulan data, data pra pengolahan, pemprosesan data dan penemuan pengetahuan. Data dianalisis dengan menggunakan sekumpulan orang yang terdiri daripada 10 orang dari umur yang berbeza dan penampilan fizikal. Dengan menerapkan pendekatan perlombongan data kepada usul manusia, kita boleh mengelakkan kecederaan dan mengambil langkah keselamatan yang sewajarnya.

Abstract

Data mining approach can be described as the process of converting a big and uncontrolled size of data into a more simple and useful form data. This project uses data mining approach to access human stability when a motion involved. In this project, we involve the activities perform a normal human being to find out the perfect posture for them to do that activity. Data mining comprises of few stages which are data collection, data preprocessing, data processing and knowledge discovery. Data are analysed by using a group of people consist of 10 people from different age and physical appearance. By applying data mining approach to human motion, we can avoid injuries and take proper safety precautions.

1.0 Introduction

Data mining can be defined as the process of analyzing the data collected using different method of study to obtain useful information from it for further knowledge discovery. The method used in this study are data collection, data preprocessing, data processing and knowledge discovery. [1] Data mining technique had been used in various sector and applications for example bike-sharing travel behavior, elite athlete injuries and others. [2] [3]

Data mining also had been very useful in the field of combining complex networks, it is easier to use data mining approach to solve complex problem, even though there are advancement in the computer technology but there is also got increases in the size data that need to be extracted. Hence, this is where data mining plays a vital role by provide a enough space for the complex data to be extracted. [4] Data mining also has a big influenced on a nowadays social media. It helps to reduce the noise, large, unstructured and dynamic data in the social media by using tie strength recognition to solve the uncertainties arise where handling with complex data. [5]

Data mining approach used to find accident patterns in the extractive industry used two method of study which are selection of variables and methodology design. For selection of variables is basically data collection with respect to the variables fixed that causes accident in the industry which are age group, day which accident occur frequently, working environment, specific physical activity, deviation of accident occurs, contact with victims, days on leave, type of injury, part of body. Whereas, the methodology design is the process where the data collected being processed and useful information extracted to conduct the study. Another example is the promises and challenges using data mining approach. In this study, the method proposed to be used for the data mining are predictive modeling, clustering, data

summarization, dependency modeling, change and deviation detection. In this study, predictive modeling is the data collection depending on the variable, data summarization and clustering are data processing, dependency modeling and deviation detection are knowledge discovery.

Data collection are the process of collecting data which are necessary and useful information from the targeted variables by conducting experiments. In our study, we are required to video captured a group of people which consist of 10 people from different physical appearance such as different height, different weight and different abilities running on treadmill at same angle of inclines. This 10 people consist of different age. Our study is mainly focus changes on the body postures of the person at one and half minutes' duration. Hence, the person movement is recorded and the image is snap out for every interval of 5 seconds to be processed.

The improvement of our study compares to other study, our study had data preprocessing. The reason for us to conduct data preprocessing are the data obtain in the real are basically incomplete, noisy, inconsistent. Thus, the outcome will have a lot irrelevant data, outliers and not accurate. If the data obtain is less quality, then the data mining results also will be less quality. Hence, we conduct data cleaning, data reduction, data integration and data transformation to improve the quality of the data. Data cleaning is where we identify the outliers and noisy data and remove them and correct not consistent data. Data reduction is where we reduce the data from image form to a simpler form by extracting information to construct graphs. Data integration is where we join data for all the video taken to produce an analysis. Data transformation is where the information from the analysis are normalized and smoothen.

Human motion is inconsistent since each of us have different speed of movement. Everyone differ from each other in physical appearance. The classification of human motion is very hard especially the experiment conducted by capturing subject which is human in 3-

dimension while data processed is only in 2-dimension. Hence, the will a great number of loss in depth of information from the video extracted to image. Thus, this may reduce the number of readings taken. The problem faced in this study is there may be missing of information when the transformation of data takes place.

The general objective of this study is to use data mining approach to obtain best result for a person body balance. The specific objectives of this project are to estimate the balanced position for human body by using data mining approach and to identify the best posture for a person when he is running on an inclined treadmill which is external disturbances for the person. A lot of previous work has been conducted on studying the human motion for example human motion estimation model based on 2D movement data to study the suitable human posture for different daily activities [6], 3D motion tracking by using particle filter algorithms [7], effect of distributed vibration power absorption on human arm at different postures [8], human action recognition using a sequence of key poses and atomic motions [9] and effect of PCL properties and balancing techniques at human leg and ligaments [10]. In this project, we analyses the effect of external perturbation on human body by using data mining techniques that is what different from previous work different.

Data mining approach had been used for finding accident pattern in extractive industry. The main objective of this work is to figure out patterns of non-fatal accidents in the Portuguese mineral extractive industry, hereafter called simply extractive industry. The main method used in this study is selection of data variables and methodology design. Selection of data variables is where the data is collected and tabulated. Methodology design is where all the collected data is being processed and required information was extracted to do further knowledge discovery. The limitations of this study are the data collected didn't undergo the process of data preprocessing. Other than that, there was a study had been done to explain the promises and challenges using data mining approach. This aim of this work is to provide an overview of this

growth of this research area, briefing of the basic techniques, and provide an explanation of how they are used in some applications. The limitation of this study is this study can't develop mining algorithms for classification, clustering, dependency analysis, and changes detection that measure to large databases.

The inspiration to conduct this study are usefulness of data mining approach to study human motion and human balance. Besides that, these data are obtaining from people from different physical appearance which enables us to find out how different people react when their stability affect by external disturbances. Thus, we will be able to figure out the impact of data mining in human motion analysis. The study outcome can be used for future application, to reduce injuries faced by athletes and to design robotic arm to be used in industry.

2.0 Methodology

2.1 Introduction

In this project, the major motive is tabulate and conclude the effect of any perturbation factor to human body using data mining approach. Data mining can be defined as a process of converting data into database and further study the outcome to obtain knowledge about data being studied. Data mining is the analysis stage of the “knowledge discovery in databases” process. In this project, the iPhone camera used to capture video of the 10 people from different age and physical apperance running on a treadmill which has three inclination angle that is 8°, 18° and 28° to obtain a raw data and the raw data is further processed to more simpler form by considering different parameter such as position of stronger leg, distance between the persons both leg, body mass index (BMI), time required per reading of the recording and further process using Waikato Environment for Knowledge Analysis (WEKA) software to obtain result.

The study of this project is divided into four level of process which are data collection, data preprocessing, data processing followed by knowledge discovery.

2.2 Data collection

In this study, data collection is the process of collecting raw data which contain useful information that can be extracted to became a simpler form of data that can be processed. IPhone 6s camera which has 12-megapixel, $f/2.2$ aperture and auto image stabilization camera that able to do 720p HD video recording at 30 fps used to record 10 people that are gathered who from different age and physical appearance to collect data for this experiment who are from different age and physical appearance. Each person required to run on the treadmill with the three-different inclined angle which are 8°, 18° and 28° angle about time interval 15 seconds per reading. There will three reading taken for each of the inclination angle. The iPhone camera is placed at the distance of 1.5 meter from the person who is running on the treadmill at the side of the person when the recording taken. This is because at that distance full body clear view can be obtained. The distance between camera and the treadmill is fixed throughout the experiment conducted and the position of treadmill is fixed throughout the experiment. This is because to avoid disruption in the raw data collected. The camera is placed exactly in-line by fixing it on tripod stand with the body of the person before running to avoid mistake in data taken and for better view of the person. The height and weight of the person are recorded and body mass index of the person are calculated.

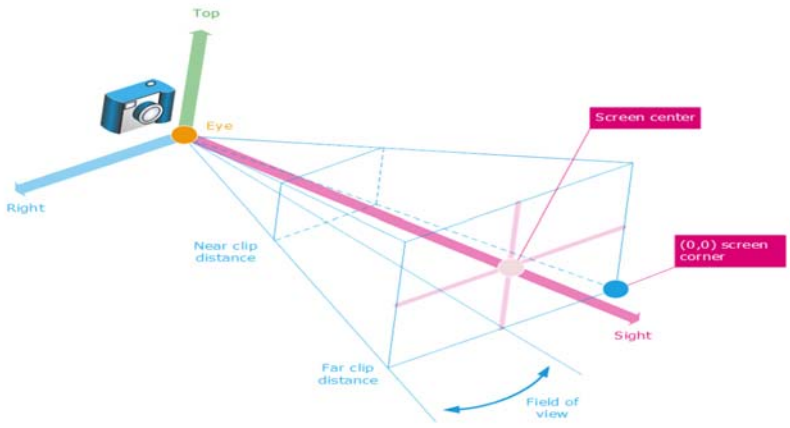

Figure 1 Field of view for an in-line came to the object

2.3 Data Preprocessing

Data preprocessing basically can be defined as the process of removing or filtering the irrelevant data and only keep the useful data for further analyzed. In the study, five images is then snapped out for each reading taken at 3,6,9,12,15 seconds of the recording taken. This step is repeated to all three-inclination angle. Then, the images snapped out are input into Adobe Photoshop Cc software the x and y coordinate of hip, knee and ankle of each person for each image are then measure using ruler tool in the Adobe Photoshop Cc software. After that, all the data collected are transfer into table in Microsoft Excel software which includes person as subject, position of stronger leg for hip, knee and ankle for all three inclinations for all reading s taken, distance between both feet for all three-inclination angle, body mass index (BMI), prediction of stability according to the result obtained into table. This data will useful for further analysis. The Microsoft Excel readings are separated for each to obtain average to maintain the accuracy of the result obtained.

Figure 2 Example of images snapped for 1st inclination angle and 2nd Inclination angle

Figure 3 Example of image snapped from 3rd inclination angle

2.4 Data Processing

The process of grouping the data preprocessed by the information of data analyzed which matches and meets our requirement of study called data processing. The grouping can be done by image classification using Weka software. The data file from the Microsoft Excel (.xlsx) are save as CSV (.csv) file to be opened in Weka software. Waikato Environment for Knowledge Analysis (Weka) is a one of machine learning software that are developed using Java programing language that have visualization tool and programmed for data analysis and predictive modelling. After the data file opened in the Weka software, filter is applied and two different type of classification which classification for stability and classification for distance between leg. Decision tree are obtained and the percentage of accuracy are obtained from the result from Weka software.

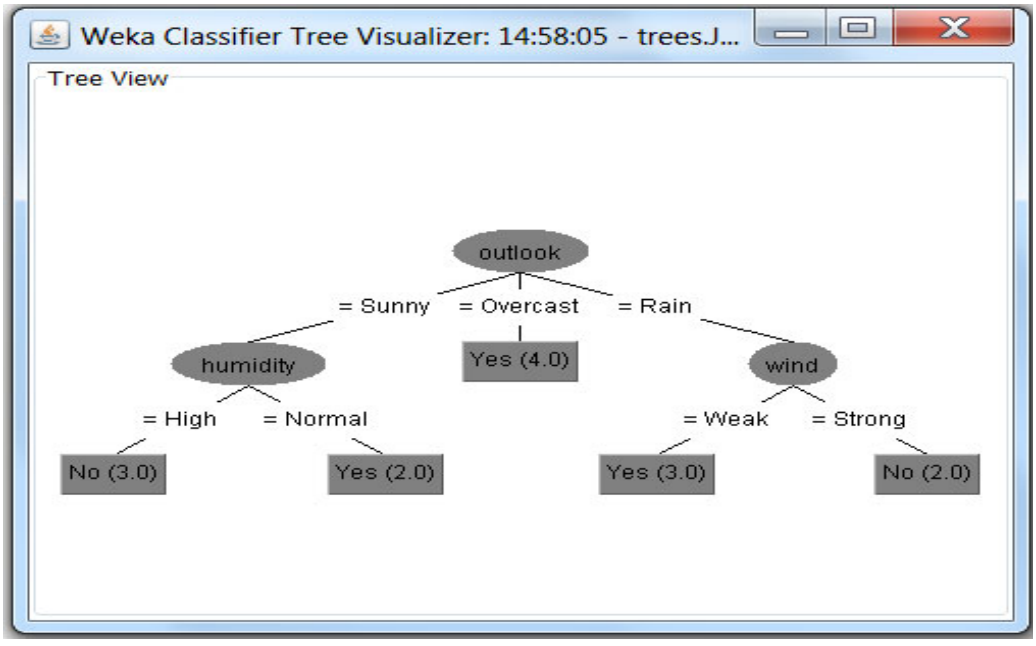

Figure 4 Example of decision tree obtained from Weka software

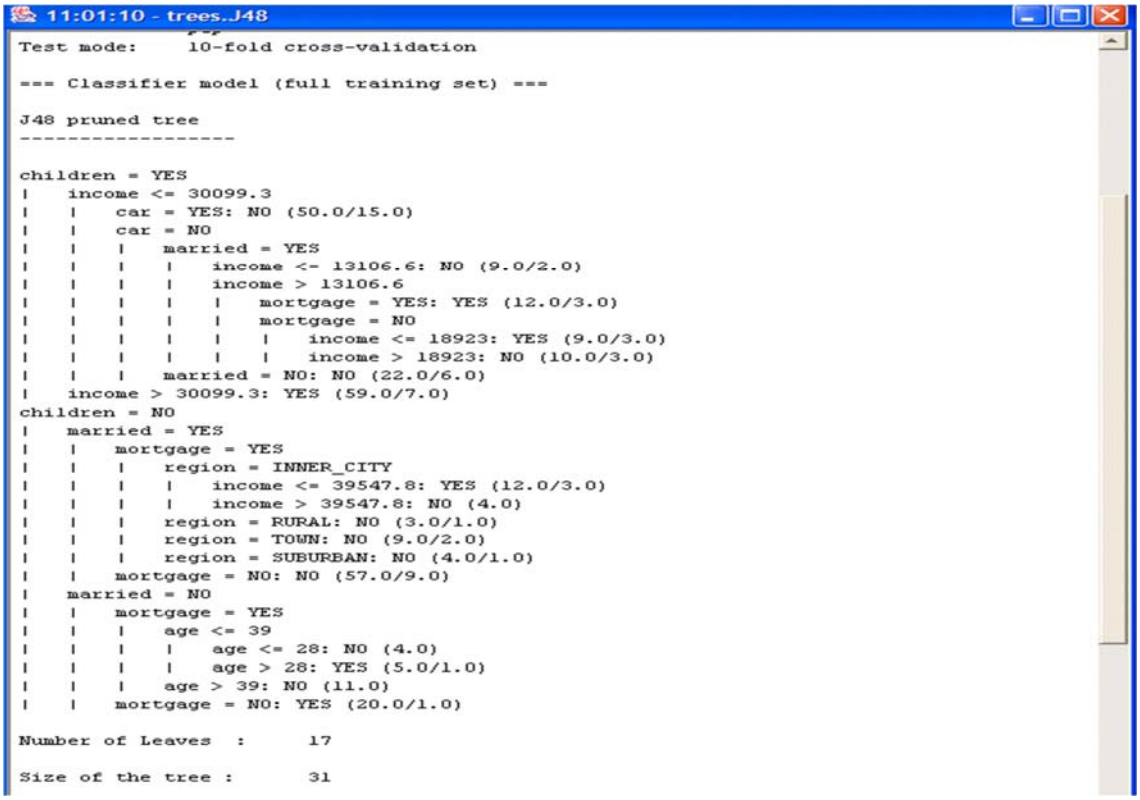

Figure 5 Examples of result obtained from Weka software using decision tree

2.5 Knowledge discovery

Knowledge discovery basically is the process of understanding the data, analyzing and interpretation of data and conclude based of information obtained. For this study, graphs are plotted based on data grouped of stability, distance between both leg and graph showing the relation between body mass index (BMI) and stability. A conclusion is made based on the results obtained.

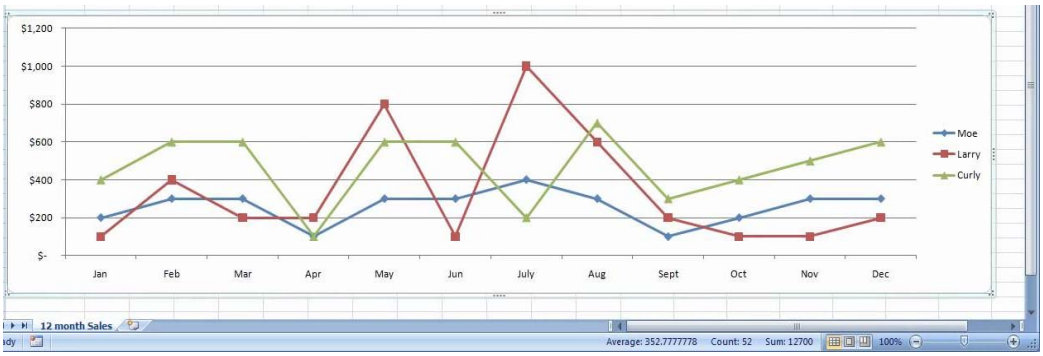

Figure 6 Examples of graph obtained from Microsoft Excel

3.0 Result and Discussion

3.1 Classification for stability

```
=== Summary ===

Correctly Classified Instances          48               96      %
Incorrectly Classified Instances         2                4      %
Kappa statistic                          0.918
Mean absolute error                      0.0267
Root mean squared error                  0.1333
Relative absolute error                  5.5457 %
Root relative squared error             27.2155 %
Total Number of Instances               50

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.933    0.000    1.000      0.933    0.966      0.921    1.000     1.000     Not Stable
                 1.000    0.067    0.909      1.000    0.952      0.921    1.000     1.000     Stable
Weighted Avg.    0.960    0.027    0.964      0.960    0.960      0.921    1.000     1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 28  2 |  a = Not Stable
  0 20 |  b = Stable
```

Figure 7 Stability clasification result

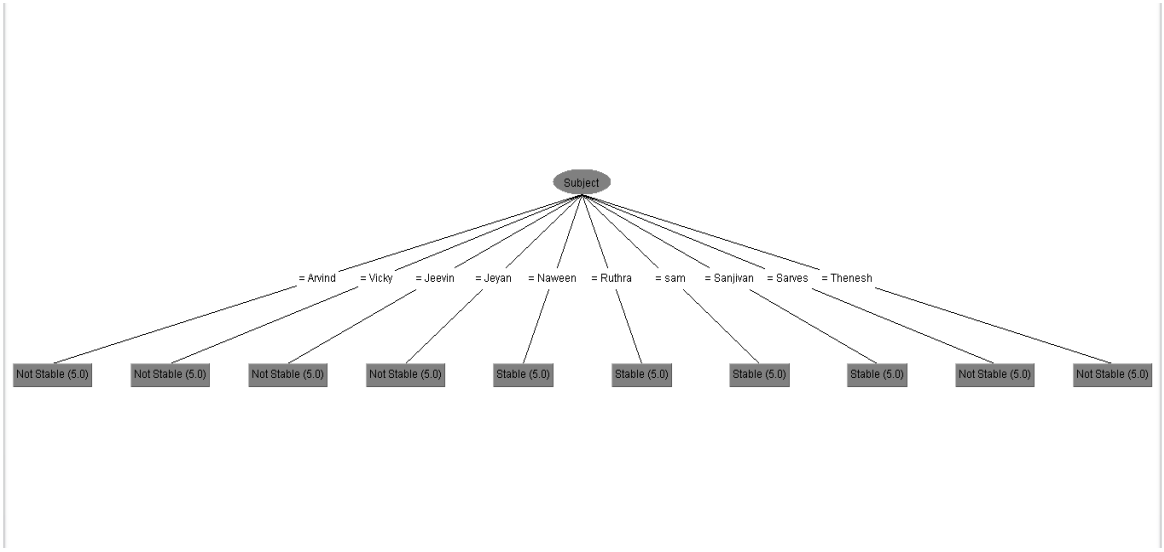

Figure 8 Decision tree for stability classification

Thus, the result obtained from Weka software states that 96 % accuracy for prediction made based on the data obtained from the excel file are correct. The prediction is made based on the position of hip, knee and ankle position. When a person is static, the position of body won't change move. Hence, the person is at stable condition. As the angle of inclination increases, the 10-people running on treadmill have huge different in their hip, knee and ankle from their initial positions. This proves that a person can't stay still as the external factor which in this case angle of inclination affects the stability of the human.

3.2 Classification for distance between both leg

```
Stability = Not Stable
|   1st distance <= 4.12
|   |   1st distance <= 2.506667
|   |   |   1st distance <= 1.54: Thenesh (3.0)
|   |   |   1st distance > 1.54: Jeevin (8.0/3.0)
|   |   1st distance > 2.506667
|   |   |   2nd distance <= 2.616667: Jeyan (5.0/2.0)
|   |   |   2nd distance > 2.616667
|   |   |   |   2nd distance <= 3.306667: Vicky (3.0)
|   |   |   |   2nd distance > 3.306667: Sarves (5.0/2.0)
|   1st distance > 4.12: Arvind (6.0/1.0)
Stability = Stable
|   2nd distance <= 1.673333: Sanjivan (5.0)
|   2nd distance > 1.673333
|   |   1st distance <= 2.806667
|   |   |   2nd distance <= 2.496667: Ruthra (5.0)
|   |   |   2nd distance > 2.496667: Naween (5.0)
|   |   1st distance > 2.806667: sam (5.0)

Number of Leaves  :     10

Size of the tree :      19


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          28               56      %
Incorrectly Classified Instances        22               44      %
Kappa statistic                          0.5111
Mean absolute error                      0.0986
Root mean squared error                  0.2787
Relative absolute error                 54.2248 %
Root relative squared error             91.9214 %
Total Number of Instances               50
```

Figure 9 Classification for distance between both legs

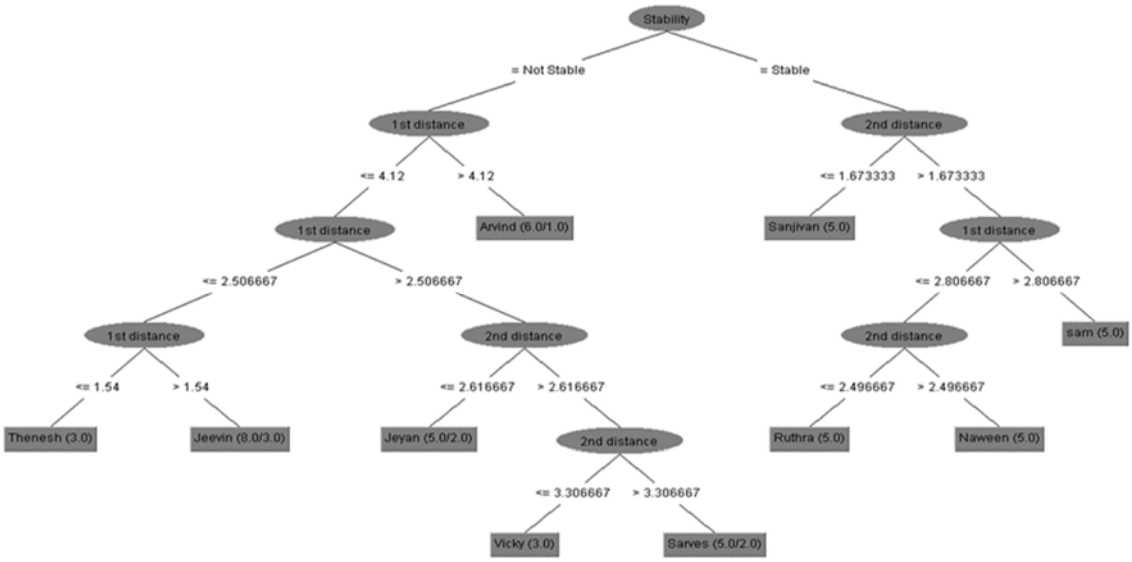

Figure 10 Decision tree for distance between both legs

This indicate that distance between both legs play a vital role for stability of a person when an external factor effects. This different is caused due to as the angle of inclination increases, a person need to place a bigger foot step when he is running to stable himself. This

eventually proves that a person can’t stay stable as the running is a continuous action, the person need avoid injuries. By remain still, the person will fall and will have injuries.

4.0 Conclusion

Data mining approach is a very useful way to extract data from the a set of information. Data mining also helps to transform a large set of uncontrolled data to be simplified and converted into a more smaller and simpler form. Human motion are related to stability of a person, a person stay stable when he is static but he needs to always stay stable in order to avoid injuries data mining approach can give a person the ideal position for each of a persons daily activities. Based on our result , as the inclination angle increases, the stability of the person decreases.

5.0 Reference

[1] P. S. Usama Fayyad, "Data mining and KDD: Promise and challenges," *Future Generation Computer Systems,* p. 115, 1997.

[2] L. d. F. I. Maria Bordagaray, "Capturing the conditions that introduce systematic variation in," *Transportation Research Part C,* pp. 232-248, 2016.

[3] Y.-M. C. Y.-C. L. Y.-J. H.-J. H. P.-H. C. Shen-Kai Chen, "INVESTIGATION OF MANAGEMENT MODELS," pp. 220-226, 2005.

[4] D. P. S. M. N. K. B. M. Zanin, "Combining complex networks and data mining:," *Physics Reports,* p. 38, 2016.

[5] M. K. Sohrabi, "A comprehensive study on the effects of using data mining techniqto predict tie strength," *Computers in Human Behavior,* pp. 534-541, 2016.

[6] W. C.K. Chan, "Human motion classification using 2D stick-model matching," *Applied Mathematics and Computation,* pp. 70-89, 2016.

[7] D. L. D. H. S. S. Jigang Liu, "3D Human motion tracking by exemplar-based conditional," *Signal Processing,* pp. 164-177, 2015.

[8] S. R. M. T. S. Adewusi, "Distributed vibration power absorption of the human hand-arm," *International Journal of Industrial Ergonomics,* pp. 363-374, 2013.

[9] L. Z. P. S. S. Guangming Zhu, "Human action recognition using multi-layer codebooks," *Signal Processing: Image Communication,* pp. 19-30, 2016.

[10] P. H. D. W. M. V. J. Zelle, "Numerical analysis of variations in posterior cruciate ligament properties and," *Medical Engineering & Physics,* pp. 700-707, 2010.