

The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.



**PREDICTING OCCUPATIONAL ACCIDENT AT  
AUTOMOTIVE MANUFACTURING INDUSTRY IN MALAYSIA  
USING DECISION TREE TECHNIQUE**

**SITI NOR FARAH JAWAHIR BT FADZIL**



**MASTER OF SCIENCE (DECISION SCIENCE)  
UNIVERSITI UTARA MALAYSIA  
2022**



PUSAT PENGAJIAN SISWAZAH  
(CENTER FOR GRADUATE STUDIES)  
UNIVERSITI UTARA MALAYSIA

PERAKUAN KERJA/DISERTASI  
(Certification of Dissertation Work)

Kami, yang bertandatangan, memperakukan bahawa  
(We, the undersigned, certify that)

**SITI NOR FARAH JAWAHIR BINTI FADZIL**

calon untuk Ijazah

(candidate for the degree of)

**MASTER of SCIENCE (DECISION SCIENCE)**

telah mengemukakan tesis/disertasinya yang bertajuk

(has presented his/her project paper of the following title)

**PREDICTING OCCUPATIONAL ACCIDENT AT AUTOMOTIVE MANUFACTURING  
INDUSTRY IN MALAYSIA USING DECISION TREE TECHNIQUE**

seperti yang tercatat di muka surat tajuk dan kulit tesis/disertasi

(as it appears on the title page and front cover of project paper)

bahasa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan, dan liputan bidang ilmu yang memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **24 FEBRUARI 2022**

(That the thesis/dissertation is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on **24 February 2022**)

Pengerusi Viva  
(Chairman for Viva)

: Dr. Zurina Hanafi

Tandatangan :  
(Signature)

Penilai Dalam 1  
(Internal Assessor 1)

: Dr. Ch'ng Chee Keong

Tandatangan :  
(Signature)

Penilai Dalam 2  
(Internal Assessor 2)

: Dr. Mazlan Mohd Sappri

Tandatangan :  
(Signature)

Penyelia Utama  
(Principal Supervisor)

: Ts.Dr.Jastini Mohd Jamil  
(Signature)

Tandatangan :  
(Signature)

Penyelia Kedua  
(Co-Supervisor)

: Prof. Madya Ts. Dr.Izwan Nizal  
Mohd Shahraneer

Tandatangan :  
(Signature)

Dekan  
Pusat Pengajian Sains  
Kuantitatif  
(Deputy Dean, School of  
Quantitative Sciences)

: Prof. Madya Dr. Mohd Kamal  
Mohd Nawawi

Tandatangan :  
(Signature)

Tarikh  
(Date)

: **03 Mac 2022**

## **Permission to Use**

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the university library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or in their absence, by the Dean of Awang Had Salleh Graduate School or Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

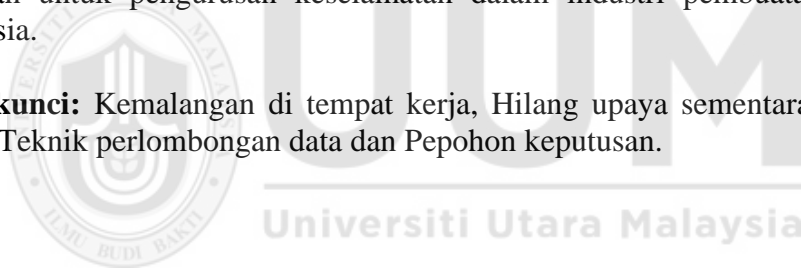
Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences  
UUM College of Arts and Sciences  
University Utara Malaysia  
06010 UUM Sintok  
Universiti Utara Malaysia

## Abstrak

Menurut Jabatan Statistik Malaysia (DOSM) pada tahun 2018, industri pembuatan menyumbang 91.2% kes hilang upaya sementara dan 6.9% kes hilang upaya kekal. Walaupun terdapat peningkatan jumlah penyelidikan berkenaan kemalangan tempat kerja dalam industri pembuatan automotif di Malaysia, setiap kajian mempunyai objektif dan kaedah kajian yang berbeza. Kajian ini meramalkan kecenderungan hilang upaya sementara dan hilang upaya kekal dengan mengenal pasti secara tepat ciri-ciri kemalangan di tempat kerja yang berlaku dalam industri pembuatan automotif di Malaysia. Teknik pepohon keputusan telah digunakan untuk membina pemodelan ramalan kemalangan pekerjaan di industri pembuatan automotif. Model pepohon keputusan dibina dengan pelbagai algoritma (Khi-kuasa dua, Indeks Gini dan Entropi), bilangan cabang (dua dan tiga) dan pembahagian data (80/20, 70/30 dan 60/40). Model yang berbeza dibandingkan untuk menentukan model terbaik bagi meramal dan mengenal pasti kesan daripada kemalangan pekerjaan. Model terbaik tersebut ialah model pepohon keputusan dengan tiga cabang menggunakan Khi-kuasa dua sebagai kriteria sasaran nominal dan pembahagian data 60/40. Nilai ketepatan bagi data ujian ialah 75.52%. Pembolehubah terpenting dalam model ini ialah *jenis kemalangan*, *punca kemalangan* dan *jenis pekerjaan*. Kajian ini menghasilkan satu set faktor yang signifikan dalam menerangkan sistem keselamatan tempat kerja dan membina model ramalan untuk meramal kesan kemalangan pekerjaan. Kajian ini boleh dijadikan garis panduan untuk pengurusan keselamatan dalam industri pembuatan automotif di Malaysia.

**Kata kunci:** Kemalangan di tempat kerja, Hilang upaya sementara, Hilang upaya kekal, Teknik perlombongan data dan Pepohon keputusan.





## Abstract

According to the Department of Statistics Malaysia (DOSM) in 2018, manufacturing industry accounted for 91.2% of temporary disability cases and 6.9% of permanent disability cases. Even though there is an increasing number of research on analyzing occupational accidents at automotive manufacturing industry in Malaysia, each research aimed for different purposes and methods. This study predicts the tendency of temporary and permanent disability by accurately identifying the characteristics of workplace accidents that occurred within automotive manufacturing in Malaysia. Decision Tree was used to build the predictive modelling of occupational accidents at automotive manufacturing industry. Decision Tree models were constructed with various algorithms (Chi-square, Gini Index and Entropy), numbers of tree branches (two and three) and data partitions (80/20, 70/30 and 60/40). The different models were compared to determine the best model for predicting and identifying the effects of occupational accidents. The best model was a three-branch decision tree model using Chi-Square as the nominal target criterion and 60/40 data partition. The testing accuracy value is 75.52%. The most important variables in the model were *types of accident*, *cause of accidents* and *job types*. This study produced a set of significant factors in explaining safety workplace system and built a predictive model for predicting effect of occupational accidents. It can be served as a guideline to safety management in automotive manufacturing industry in Malaysia.

**Keywords:** Workplace accident, Temporary disability, Permanent disability, Data mining technique and Decision tree.

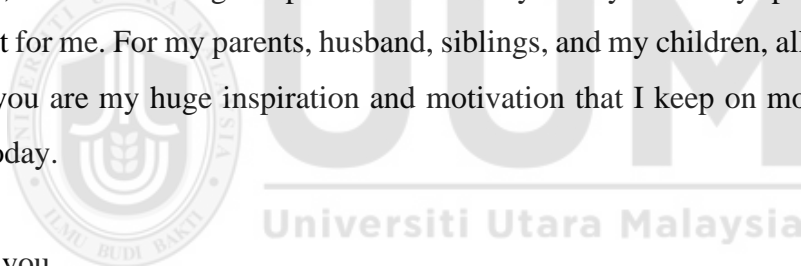
## Acknowledgement

All praise is just to Allah. I would like to express my appreciation and gratitude to Allah for bestowed me upon His blessings, grace, mercy, physical and mental well-being by providing idea and good health to me throughout completing this project paper successfully without any serious obstacles.

Upon completing this project paper, I would like to thank my supervisor, Ts. Dr. Jastini Binti Mohd Jamil and Prof. Madya Ts. Dr. Izwan Nizal Bin Mohd Shahraneer for being such a supportive and dedicated supervisor by always guiding and helping me a lot in completing this project paper and hence, with their support I manage to complete this project paper as well as to my examiners, Dr. Ch'ng Chee Keong and Dr. Mazlan Bin Mohd Sappri for correcting and guiding my dissertation paper.

Finally, I would like to give special thanks to my family for always provide an endless support for me. For my parents, husband, siblings, and my children, all I can say is that all of you are my huge inspiration and motivation that I keep on moving and trying until today.

Thank you.



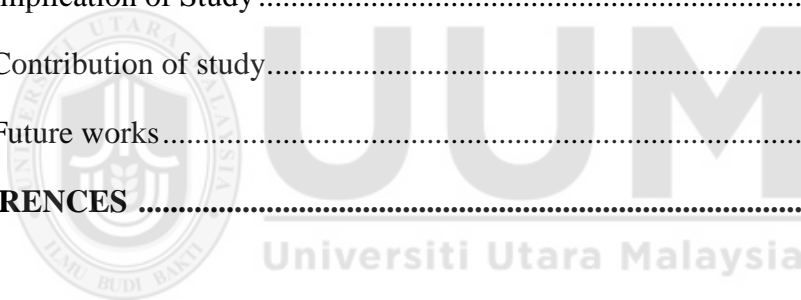
# Table of Contents

Permission to Use .....	i
Abstrak .....	ii
Abstract .....	iii
Acknowledgement.....	iv
Table of Contents .....	v-vi
List of Tables.....	ix
List of Figures .....	x
List of Appendices.....	xi
<b>CHAPTER ONE INTRODUCTION .....</b>	<b>1</b>
1.1 Background .....	1
1.2 Data Mining .....	6
1.3 Problem Statement .....	7
1.4 Research Question.....	9
1.5 Research Objectives .....	10
1.6 Significance of study .....	10
1.7 Scope of the study .....	10
<b>CHAPTER TWO LITERATURE REVIEW .....</b>	<b>12</b>
2.1 Introduction .....	12
2.2 Manufacturing Industry .....	12
2.3 Occupational Accident at Manufacturing Industry .....	12
2.3.1 Effect of occupational accident in manufacturing .....	13
2.4 Data Mining.....	14
2.5 Prediction Method .....	16
2.5.1 Neural Network.....	16
2.5.2 Multiple Linear Regression .....	16
2.5.3 Decision Tree .....	17



2.5.3.1 Splitting rule.....	18
2.5.3.2 Pruning .....	20
2.5.3.3 Accuracy Measure.....	21
2.6 Application of Data Mining in Occupational Accident Cases .....	22
2.7 Previous studies related to occupational accident in Malaysia .....	24
2.8 Summary .....	28
2.8.1 Summary of data mining application in workplace accident analysis.....	28
2.8.2 Summary of previous studies about occupational accident in Malaysia ...	30
2.9 Conclusion.....	32
<b>CHAPTER THREE METHODOLOGY .....</b>	<b>33</b>
3.1 Introduction .....	33
3.2 Developed Framework .....	34
3.3 Description of Method.....	35
3.3.1 Project Understanding.....	36
3.3.2 Data Understanding .....	36
3.3.2.1 Dataset Characteristics .....	36
3.3.3 Data Pre-processing .....	39
3.3.3.1 Missing Values Handler.....	39
3.3.3.2 Data Partitioning .....	41
3.3.4 Modelling.....	43
3.3.4.1 Decision Tree Models .....	43
3.3.4.2 Splitting Criteria.....	43
3.3.5 Evaluation / Model Comparison .....	44
3.3.5.1 Interpretation Process.....	46
3.3.6 Deployment .....	46
3.4 Conclusion.....	47
<b>CHAPTER FOUR RESULT AND FINDINGS.....</b>	<b>48</b>
4.1 Introduction .....	48
4.2 Findings .....	48

4.2.1 Descriptive Analysis .....	48
4.2.2 Decision tree models comparison .....	56
4.2.3 Best model analysis: Decision tree with three branches using Chi-Square splitting method with 60/40 data partition .....	58
4.2.3.1 Variable importance for best model .....	59
4.2.3.2 Tree Diagram .....	60
4.2.3.3 Node rules .....	61
4.2.3.4 Scoring Ranking Overlay .....	67
4.3 Summary .....	68
<b>CHAPTER FIVE CONCLUSION .....</b>	<b>69</b>
5.1 Introduction .....	69
5.2 Discussions .....	69
5.3 Implication of Study .....	71
5.4 Contribution of study.....	72
5.5 Future works.....	73
<b>REFERENCES .....</b>	<b>74</b>



## List of Tables

Table 1.1 Accidents reported to Department of Occupational Safety and Health Malaysia by sectors from 2013 until 2018.....	2
Table 2.1 A confusion matrix for model evaluation .....	21
Table 2.2 Summary of data mining application in workplace accident analysis .....	28
Table 2.3 Summary of previous study about occupational accident in Malaysia.....	30
Table 3.1 Cross Industry Standard Process Methodology .....	35
Table 3.2 Independent and dependent variable description .....	37
Table 3.3 Statistic of missing data .....	41
Table 3.4 Dataset Summary .....	42
Table 3.5 Confusion Matrix .....	46
Table 4.1 Model comparison.....	57
Table 4.2 Description of miss-classification and accuracy rate for Decision Tree model with three branches using Chi-Square splitting rule and 60/40 data partition .....	58
Table 4.3 Variable importance of Decision tree model with three branches using Chi-Square splitting rule .....	59
Table 4.4 Node rules of the highest prediction rate of temporary disability (low prediction of permanent disability) .....	62
Table 4.5 Node rules of the highest prediction rate of permanent disability (low prediction of temporary disability) .....	65

## List of Figures

Figure 1.1. Occupational accidents by sectors for the category of temporary disability in 2018.....	3
Figure 1.2. Occupational accidents by sectors for the category of permanent disability in 2018.....	3
Figure 3.1 Phase of CRISP-DM reference model.....	33
Figure 3.2 The developed framework.....	34
Figure 3.3 Data partition setting for model's performance.....	42
Figure 3.4 Data partition setting for model's performance.....	44
Figure 3.5 Overview of diagram that consist whole models.....	45
Figure 4.1 Statistical distribution of effect of accident.....	49
Figure 4.2 Statistical distribution of effect of accident by gender.....	50
Figure 4.3 Statistical distribution of effect of accident by age group.....	51
Figure 4.4 Distribution of effect of accident by causes of accident.....	52
Figure 4.5 Distribution of effect of accident by job types.....	53
Figure 4.6 Distribution of effect of accident by type of manufacturing.....	54
Figure 4.7 Distribution of working environment.....	55
Figure 4.8 Distribution of type of injury.....	56
Figure 4.9 Tree diagrams of the best decision tree model.....	60
Figure 4.10 Score Rankings Overlay window of Cumulative % Response (Depth = 20) for the best model.....	67
Figure 4.11 Score Rankings Overlay window of Cumulative % Response (Depth = 60) for the best model.....	68

## List of Appendices

Appendix 1. Tree Diagram of Decision Tree Best Model; Chi-Square splitting rule with 60 % train data 40% test data.....	80
Appendix 2. Node rules of Decision Tree Best Model .....	81



# CHAPTER ONE

## INTRODUCTION

### 1.1 Background

With the aim of becoming an industrialized country, industrialization has become an important part in the Malaysian development strategies. Manufacturing industry has shown to be one of the significant supports and a major contributor to Malaysian economy (Said et al., 2012). Apparently, this industry has offered a vast amount of job opportunities in the recent years. According to Department of Statistic Malaysia (DOSM), total employees engaged in the manufacturing sector in December 2018 was 1,075,635 persons, an increase of 1.7 per cent or 18,044 persons as compared to 1,057,591 persons in December 2017 (Department of Occupational Safety and Health, 2018). Instead of improving income distribution and improving quality of life, economic growth through industrialization unfortunately also resulted in growing in number of occupational accidents. In accordance with estimation made by International Labor Organization (ILO), every year about 270 million occupational accidents occur (Amirah et al., 2013).

Normally, factors such as large employment of new workers, machinery and equipment are associated with rapid expansion of manufacturing industries. Regrettably, the application of new technologies would expose new hazards to the unskillfulness workers (Said et al., 2012). Moreover, hiring new worker might as well expose higher risk of accident as they are not familiar to the hazard of workplace environment.

## REFERENCES

- Abdullah, M. S., Othman, Y. H., Osman, A., & Salahudin, S. N. (2016). Safety Culture Behaviour in Electronics Manufacturing Sector (EMS) in Malaysia: The Case of Flextronics. *Procedia Economics and Finance*, 35(16), 454–461. [https://doi.org/10.1016/s2212-5671\(16\)00056-3](https://doi.org/10.1016/s2212-5671(16)00056-3)
- Abidin, A., Awang Lukman, K., Sajali, H., Syed Abdul Rahim, S. S., Robinson, F., Hassan, M. R., Hayati, F., Ibrahim, M. Y., & Jeffree, M. S. (2021). Prevalence of occupational injury and determination of safety climate in small scale manufacturing industry: A cross-sectional study. *Annals of Medicine and Surgery*, 69(June), 102699. <https://doi.org/10.1016/j.amsu.2021.102699>
- Ali, D., Yusof, Y., & Adam, A. (2017). Safety Culture and Issue in the Malaysian Manufacturing Sector. *MATEC Web of Conferences*, 135. <https://doi.org/10.1051/mateconf/201713500031>
- Amirah, N. A., Asma, W. I., Muda, M. S., & Mohd Amin, W. A. A. W. (2013). Safety culture in combating occupational safety and health problems in the Malaysian manufacturing sectors. *Asian Social Science*, 9(3), 182–191. <https://doi.org/10.5539/ass.v9n3p182>
- Arockia Christopher, A. B., & Appavu Alias Balamurugan, S. (2014). Prediction of warning level in aircraft accidents using classification techniques: An empirical study. *Advances in Intelligent Systems and Computing*, 243, 1217–1223. [https://doi.org/10.1007/978-81-322-1665-0\\_126](https://doi.org/10.1007/978-81-322-1665-0_126)
- Ayob, A., Shaari, A. A., Zaki, M. F. M., & Munaaim, M. A. C. (2018). Fatal occupational injuries in the Malaysian construction sector-causes and accidental agents. *IOP Conference Series: Earth and Environmental Science*, 140(1). <https://doi.org/10.1088/1755-1315/140/1/012095>



- Barros, A. G., Said, J., Gomes, G. R. R., & Almeida, F. M. de. (2018). Data Mining Technique for Preventional Analysis of Work Accidents. *International Journal of Advanced Engineering Research and Science*, 5(9), 22–26. <https://doi.org/10.22161/ijaers.5.9.4>
- Bascompta, M., M<sup>a</sup> Rossell, J., & Francisco Anticoi, H. (2018). *Analysis of occupational accidents in underground 1 and surface mining in Spain using data mining techniques Lluís Sanmiquel*. <https://doi.org/10.20944/preprints201801.0231.v1>
- Boden, L. I., Reville, R. T., & Biddle, J. (2005). The Adequacy of Workers' Compensation Cash Benefits. In *Workplace Injuries and Diseases: Prevention and Compensation* (Vol. 3). <https://doi.org/10.17848/9781429454919.ch3>
- Department of Occupational Safety and Health. (2018). *Occupational Accidents Statistics by Sector Until October 2018 Note: Statistics show cases investigated by DOSH only Occupational Accident Statistics by Sector updated every 2 months*. BM (</index.php/ms/statistik?kemalangan?pekerjaan/mengikut?sektor>) ENG (</index.php/en/occupational?accident?statistics/by?sector>)
- Düntsch, I., & Gediga, G. (2019). Confusion Matrices and Rough Set Data Analysis. *Journal of Physics: Conference Series*, 1229(1). <https://doi.org/10.1088/1742-6596/1229/1/012055>
- Ganesh, C. S., & Krishnan, R. (2016). A review of occupational injury research in Malaysia. In *Medical Journal of Malaysia* (Vol. 71, pp. 100–104).
- Ghousi, R. (2015). Applying a decision support system for accident analysis by using data mining approach: A case study on one of the Iranian manufactures. *Journal of Industrial and Systems Engineering*, 8(3), 59–76.
- Guardiola Mouhaffel, A., Martínez Domínguez, C., Arcones, B., Morán Redonda, F.,

- & Díaz Martín, R. (2017). Using multiple regression analysis lineal to predict occupation market work in occupational hazard prevention services. *International Journal of Applied Engineering Research*, 12(3), 283–288.
- Guneri, A. F. (2016). *Analysis of the relation between the characteristics of workers and occupational accidents using data mining A Multi Criteria Decision Making Model For Risk Assessment Of Technology Investment Project View project Muhammet Gül Tunceli Üniversitesi 18 PUBL*. <http://www.turjoem.com>
- Hee, O. C. (2014). Factors Contribute to Safety Culture in the Manufacturing Industry in Malaysia. *International Journal of Academic Research in Business and Social Sciences*, 4(4). <https://doi.org/10.6007/ijarbss/v4-i4/753>
- Hui-Nee A. (2014). Safety Culture in Malaysian Workplace: An Analysis of Occupational Accidents. In *Health and the Environment Journal* (Vol. 5, Issue 3).
- Jamil, J. M., & Shaharane, I. N. M. (2014). Comparative analysis of data mining techniques for business data. *AIP Conference Proceedings*, 1635(August 2015), 587–593. <https://doi.org/10.1063/1.4903641>
- Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. In *Knowledge Engineering Review* (Vol. 21, Issue 1, pp. 1–24). <https://doi.org/10.1017/S0269888906000737>
- López-Arquillos, A., & Rubio-Romero, J. C. (2016). Analysis of Workplace Accidents in Automotive Repair Workshops in Spain. *Safety and Health at Work*, 7(3), 231–236. <https://doi.org/10.1016/j.shaw.2016.01.004>
- Obi, A. N., Azuhairi, A. ., & Huda, B. (2017). Factors associated with work related injuries among workers of an industry in malaysia. *International Journal of Public Health and Clinical Sciences*, 4(2), 97–108.

- Osman, R., Awang, N., Abdul, S., Syed Hassan, H., & Yusof, N. M. (2015). Level Of Awareness On Behaviour-Based Safety (Bbs) In Manufacturing Industry Towards Reducing Workplace Incidents. In *International Journal of Education and Research* (Vol. 3, Issue 1). www.ijern.com
- Pandis, N. (2016). Multiple linear regression analysis. In *American Journal of Orthodontics and Dentofacial Orthopedics* (Vol. 149, Issue 4, p. 581). American Association of Orthodontists. <https://doi.org/10.1016/j.ajodo.2016.01.012>
- Panigrahi, R., & Borah, S. (2018). Rank Allocation to J48 Group of Decision Tree Classifiers using Binary and Multiclass Intrusion Detection Datasets. *Procedia Computer Science*, 132, 323–332. <https://doi.org/10.1016/j.procs.2018.05.186>
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: A systematic literature review. *PeerJ Computer Science*, 6, 1–43. <https://doi.org/10.7717/PEERJ-CS.267>
- Rivas, T., Paz, M., Martín, J. E., Matías, J. M., García, J. F., & Taboada, J. (2011). Explaining and predicting workplace accidents using data-mining techniques. *Reliability Engineering and System Safety*, 96(7), 739–747. <https://doi.org/10.1016/j.ress.2011.03.006>
- Said, S. M., Said, F., & Zairihan, A. H. (2012). The determinants of industrial accidents in the Malaysian manufacturing sector. *African Journal of Business Management*, 6(5), 1999–2006. <https://doi.org/10.5897/AJBM11.2439>
- Sanmiquel, L., Bascompta, M., Rossell, J. M., Anticoi, H. F., & Guash, E. (2018). Analysis of occupational accidents in underground and surface mining in Spain using data-mining techniques. *International Journal of Environmental Research and Public Health*, 15(3), 1–11. <https://doi.org/10.3390/ijerph15030462>
- Sarkar, S., Patel, A., Madaan, S., & Maiti, J. (2017a). Prediction of occupational

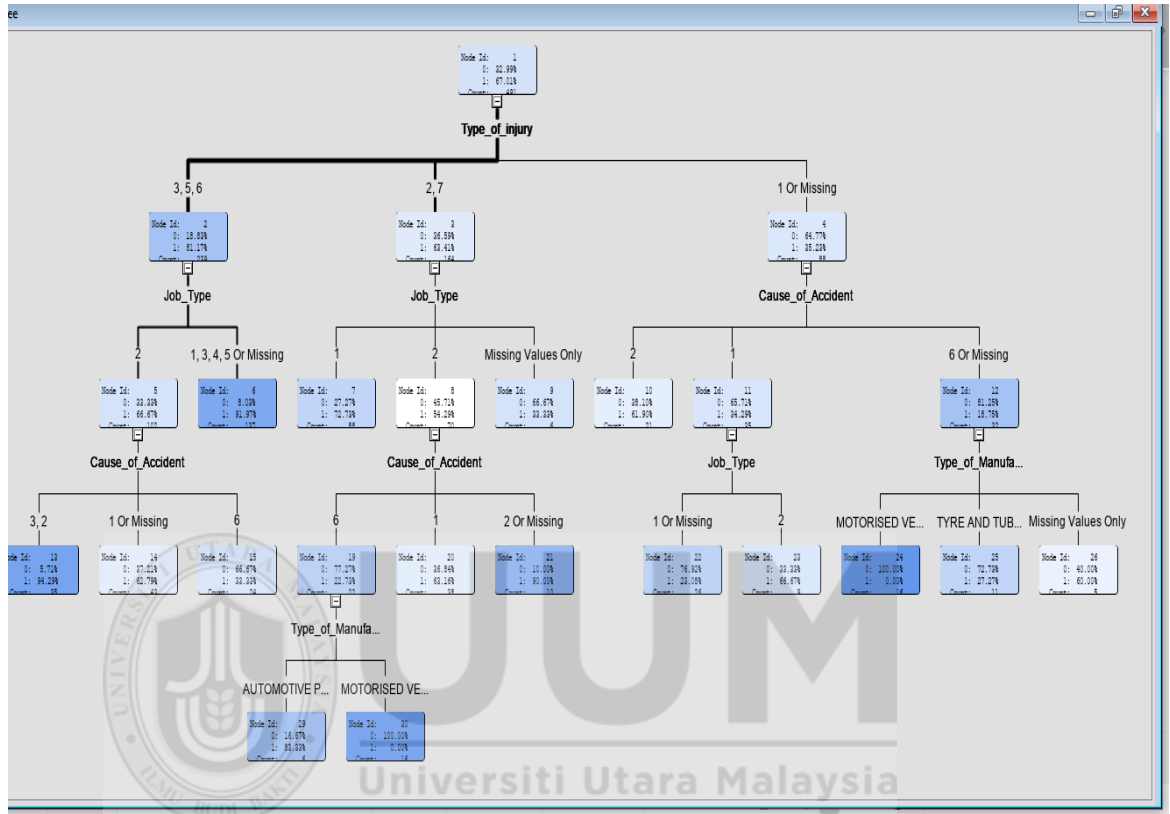
- accidents using decision tree approach. *2016 IEEE Annual India Conference, INDICON 2016, December*. <https://doi.org/10.1109/INDICON.2016.7838969>
- Sarkar, S., Patel, A., Madaan, S., & Maiti, J. (2017b, January 31). Prediction of occupational accidents using decision tree approach. *2016 IEEE Annual India Conference, INDICON 2016*. <https://doi.org/10.1109/INDICON.2016.7838969>
- Shaharane, I., Dillon, T., & Hadzic, F. (2011). Assertaining Data Mining Rules Using Statistical Approaches. *2011 International Symposium on ...*, 1(Icccc 2009), 181–189. <http://www.ipcsit.com/vol1/33-A1672.pdf>
- Shirali, G. A., Noroozi, M. V., & Malehi, A. S. (2018). Predicting the outcome of occupational accidents by CART and CHAID methods at a steel factory in Iran. *Journal of Public Health Research*, 7(2). <https://doi.org/10.4081/jphr.2018.1361>
- T. Akomolafe, D., & Olutayo, A. (2013). Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. *American Journal of Database Theory and Application*, 1(3), 26–38. <https://doi.org/10.5923/j.database.20120103.01>
- Tamizharasi, K. (2014). *Employee Turnover Analysis with Application of Data Mining Methods*. 5(1), 562–566. [www.ijcsit.com](http://www.ijcsit.com)
- Tetik, Y. O., Kale, O. A., Bayram, I., & Baradan, S. (2021). Applying decision tree algorithm to explore occupational injuries in the Turkish construction industry. *Journal of Engineering Research*, November. <https://doi.org/10.36909/jer.12209>
- Weedmark, D. (2018). The advantages & disadvantages of a multiple regression model. In *Sciencing* (p. 7). <https://sciencing.com/advantages-disadvantages-multiple-regression-model-12070171.html>
- Yedla, AD. (2019). *Predicting Safety Outcomes in Mining Industry-A Machine Learning Approach*.

<https://search.proquest.com/openview/02a5bf3e209f9fc2d7c9ca10ae7f5a8f/1?q-origsite=gscholar&cbl=18750&diss=y>

- Yedla, Anurag, Kakhki, F. D., & Jannesari, A. (2020). Predictive modeling for occupational safety outcomes and days away from work analysis in mining operations. *International Journal of Environmental Research and Public Health*, 17(19), 1–17. <https://doi.org/10.3390/ijerph17197054>
- Yeow, J. A., Ng, P. K., Tai, H. T., & Chow, M. M. (2020). A Review on Human Error in Malaysia Manufacturing Industries. *Journal of Information System and Technology Management*, 5(19), 01–13. <https://doi.org/10.35631/jistm.519001>
- Yilmaz AssistProfDr, F. (2014). Analysis of Occupational Accidents in Construction Sector in Turkey. In *Journal of Multidisciplinary Engineering Science and Technology (JMEST)* (Vol. 1). [www.jmest.org](http://www.jmest.org)
- Zakaria, N. H., Mansor, N., & Abdullah, Z. (2012). Workplace Accident in Malaysia: Most Common Causes and Solutions. In *Business and Management Review* (Vol. 2, Issue 5). <http://www.businessjournalz.org/bmr>

# Appendices

Appendix 1. Tree Diagram of Decision Tree Best Model; Chi-Square splitting rule with 60 % train data 40% test data



Appendix 2. Node rules of Decision Tree Best Model

\*-----\*

Node = 6

\*-----\*

if Type of injury IS ONE OF: 3, 5, 6

AND Job Type IS ONE OF: 1, 3, 4, 5 or MISSING

then

Tree Node Identifier = 6

Number of Observations = 137

Predicted: Effect of accident=1 = 0.92

Predicted: Effect of accident=0 = 0.08

\*-----\*

Node = 7

\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Job Type IS ONE OF: 1

then

Tree Node Identifier = 7

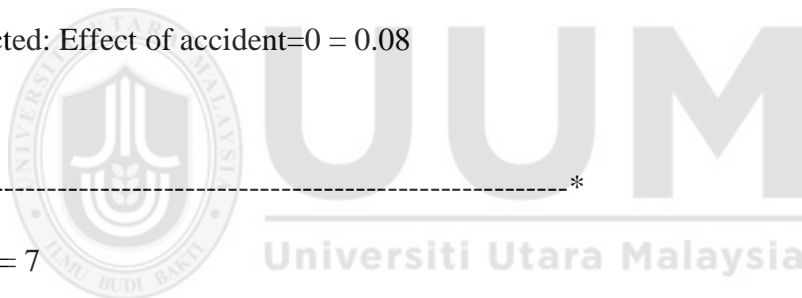
Number of Observations = 88

Predicted: Effect of accident=1 = 0.73

Predicted: Effect of accident=0 = 0.27

\*-----\*

Node = 9





\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Job Type equals Missing

then

Tree Node Identifier = 9

Number of Observations = 6

Predicted: Effect of accident=1 = 0.33

Predicted: Effect of accident=0 = 0.67

\*-----\*

Node = 10

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Cause of Accident IS ONE OF: 2

then

Tree Node Identifier = 10

Number of Observations = 21

Predicted: Effect of accident=1 = 0.62

Predicted: Effect of accident=0 = 0.38

\*-----\*

Node = 13

\*-----\*

if Type of injury IS ONE OF: 3, 5, 6

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 3, 2

then

Tree Node Identifier = 13

Number of Observations = 35

Predicted: Effect of accident=1 = 0.94

Predicted: Effect of accident=0 = 0.06

\*-----\*

Node = 14

\*-----\*

if Type of injury IS ONE OF: 3, 5, 6

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 1 or MISSING

then

Tree Node Identifier = 14

Number of Observations = 43

Predicted: Effect of accident=1 = 0.63

Predicted: Effect of accident=0 = 0.37

\*-----\*

Node = 15

\*-----\*

if Type of injury IS ONE OF: 3, 5, 6

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 6

then

Tree Node Identifier = 15

Number of Observations = 24

Predicted: Effect of accident=1 = 0.33

Predicted: Effect of accident=0 = 0.67

\*-----\*

Node = 20

\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 1

then

Tree Node Identifier = 20

Number of Observations = 38

Predicted: Effect of accident=1 = 0.63

Predicted: Effect of accident=0 = 0.37

\*-----\*

Node = 21

\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 2 or MISSING

then

Tree Node Identifier = 21

Number of Observations = 10

Predicted: Effect of accident=1 = 0.90

Predicted: Effect of accident=0 = 0.10

\*-----\*

Node = 22

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Job Type IS ONE OF: 1 or MISSING

AND Cause of Accident IS ONE OF: 1

then

Tree Node Identifier = 22

Number of Observations = 26

Predicted: Effect of accident=1 = 0.23

Predicted: Effect of accident=0 = 0.77

\*-----\*

Node = 23

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 1

then

Tree Node Identifier = 23

Number of Observations = 9

Predicted: Effect of accident=1 = 0.67

Predicted: Effect of accident=0 = 0.33

\*-----\*

Node = 24

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Type of Manufacturing IS ONE OF: MOTORISED VEHICLES

AND Cause of Accident IS ONE OF: 6 or MISSING

then

Tree Node Identifier = 24

Number of Observations = 16

Predicted: Effect of accident=1 = 0.00

Predicted: Effect of accident=0 = 1.00

\*-----\*

Node = 25

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Type of Manufacturing IS ONE OF: TYRE AND TUBE INDUSTRIES

AND Cause of Accident IS ONE OF: 6 or MISSING

then

Tree Node Identifier = 25

Number of Observations = 11

Predicted: Effect of accident=1 = 0.27

Predicted: Effect of accident=0 = 0.73

\*-----\*

Node = 26

\*-----\*

if Type of injury IS ONE OF: 1 or MISSING

AND Type of Manufacturing equals Missing

AND Cause of Accident IS ONE OF: 6 or MISSING

then

Tree Node Identifier = 26

Number of Observations = 5

Predicted: Effect of accident=1 = 0.60

Predicted: Effect of accident=0 = 0.40

\*-----\*

Node = 29

\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Type of Manufacturing IS ONE OF: AUTOMOTIVE PART

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 6

then

Tree Node Identifier = 29

Number of Observations = 6

Predicted: Effect of accident=1 = 0.83

Predicted: Effect of accident=0 = 0.17

\*-----\*

Node = 30

\*-----\*

if Type of injury IS ONE OF: 2, 7

AND Type of Manufacturing IS ONE OF: MOTORISED VEHICLES or MISSING

AND Job Type IS ONE OF: 2

AND Cause of Accident IS ONE OF: 6

then

Tree Node Identifier = 30

Number of Observations = 16

Predicted: Effect of accident=1 = 0.00

Predicted: Effect of accident=0 = 1.00

