



الجامعة الإسلامية العالمية ماليزيا
INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA
يُونَيْتِي اِسْلَامْ اِنْتَارَا يَغْسِيَا مِلْدِسِيَا

**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA
KULLIYYAH OF INFORMATION AND COMMUNICATION
TECHNOLOGY**

**INFO 4313 DATA MINING
Section 1, 2022/2023**

**GROUP ASSIGNMENT
DATAVERSE**

GROUP MEMBERS

Name	Matric No
Nor Zuhayra Amalin Binti Zulkifli	2011642
Fatina Mukhammadalieva Azam Kizi	1920218

LECTURER

Abdul Rafiez Bin Abdul Raziff

TABLE OF CONTENTS

1.0 Introduction	3
2.0 Data Source	3
3.0 Concepts of Classification	3
4.0 Method Used	4
4.1 NEURAL NETWORK/MULTILAYERPERCEPTRON	4
4.1.1 RESULT	4
4.4 NAIVE BAYES	5
4.4.1 RESULT	5
5.0 Discussion	7
6.0 Conclusion	8
7.0 Task Distribution	9
8.0 Appendix	9
9.0 References	10

1.0 Introduction

The process of extracting and identifying hidden patterns and relationships in large volumes of data is known as data mining. It's a powerful data analysis process that combines machine learning and artificial intelligence to extract the relevant data that helps analysts better understand their objectives and make better decisions for increasing profits, lowering expenses, boosting interconnections across diverse sets of data, and much more.

One of the strategies used in data mining is classification, which assists in the organisation of data sets. The purpose of classification techniques is to identify patterns, movements, and groups in massive amounts of data and transform that information into more accurate knowledge for better decision-making. We used the McDonald's Locations by State in 2022 dataset from Kaggle to run multiple Machine Learning algorithms and afterwards compare the finest and worst performers in classification techniques. By analysing this dataset, we can gain insights into the distribution of McDonald's locations across Malaysia, which states have the highest concentration of locations, and how this compares to the country's overall population density. Additionally, we can use this information to identify potential trends in the fast food industry, such as which states are experiencing the most growth in terms of new McDonald's locations, and which areas may be underserved by fast food chains.

Overall, this dataset provides a valuable resource for anyone interested in understanding the fast food landscape in Malaysia, and how it is evolving over time.

2.0 Data Source

This information was gathered from open source datasource named Kaggle, world's largest data science community with a wide range of tools and resources which helps targeted audiences do data science based analyses.

3.0 Concepts of Classification

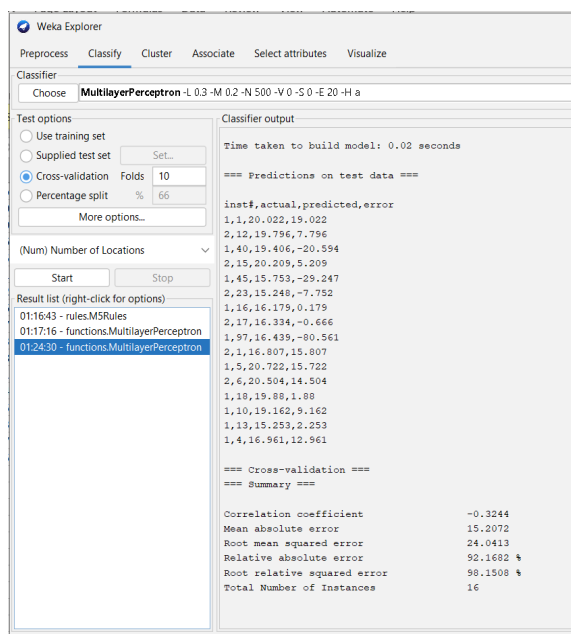
The process of classifying items based on their attributes is known as classification. The purpose of classification is to predict a specific outcome based on the data or input given. It consists of two basic stages: the learning phase, in which the classification algorithm is learned, and the classification phase, in which the algorithm labels fresh data.

4.0 Method Used

Weka has lots of ML algorithms. For our project we used data from Heart Attack Analysis and Prediction and our group is using four types of ML algorithms (**Neural Network, Regression By Discretization, Random Tree and Naive Bayes**) to classify the data by using 10-fold cross validation.

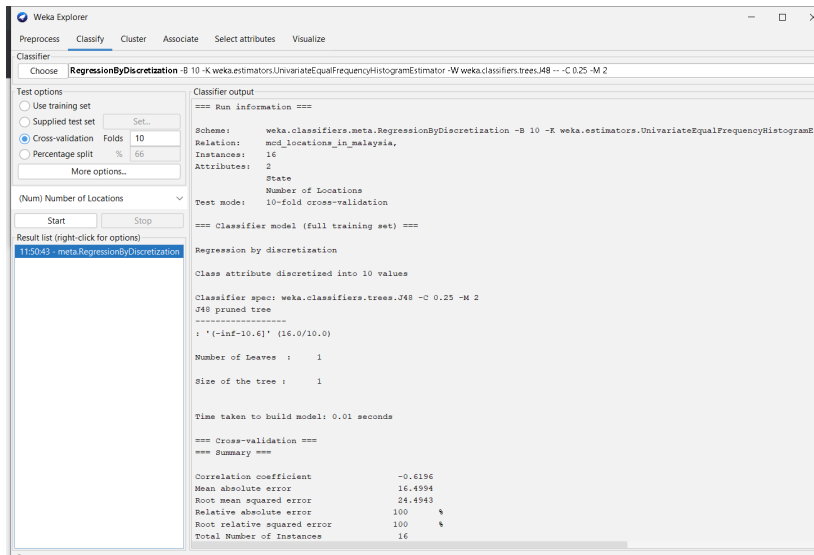
4.1 NEURAL NETWORK/MULTILAYERPERCEPTRON

4.1.1 RESULT



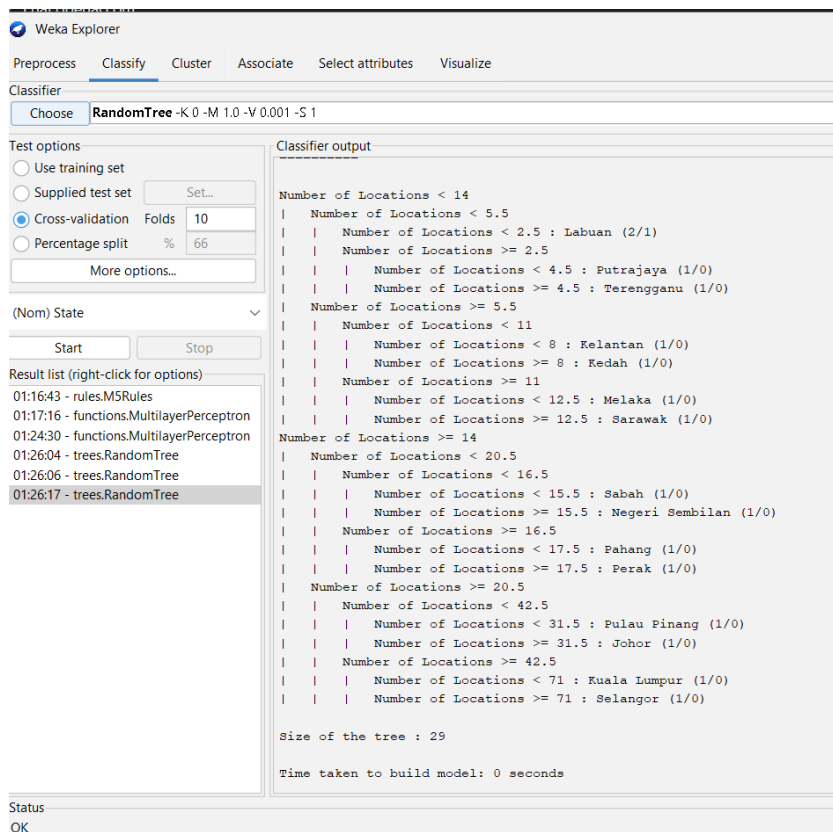
4.2 REGRESSION BY DISCRETIZATION

4.2.1 RESULT



4.3 RANDOM TREES

4.3.1 RESULT



4.4 NAIVE BAYES

4.4.1 RESULT

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) State
Start Stop

Result list (right-click for options):
01:16:43 - rules.MSRules
01:17:16 - functions.MultilayerPerceptron
01:24:30 - functions.MultilayerPerceptron
01:26:04 - trees.RandomTree
01:26:06 - trees.RandomTree
01:26:17 - trees.RandomTree
01:40:35 - **bayes.NaiveBayes**

Classifier output

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: mod_locations_in_malaysia
Instances: 16
Attributes: 2
State
Number of Locations
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class	Johor (0.06)	Kedah (0.06)	Kelantan (0.06)	Kuala Lumpur (0.06)	Labuan (0.06)	Melaka Negeri Sembilan (0.06)	Selangor (0.06)
Number of Locations								
mean		41.1429	6.8571	6.8571	40	0	13.7143	13.7143
std. dev.		1.1429	1.1429	1.1429	1.1429	1.1429	1.1429	1.1429
weight sum		1	1	1	1	1	1	1
precision		6.8571	6.8571	6.8571	6.8571	6.8571	6.8571	6.8571

Time taken to build model: 0 seconds

=== Predictions on test data ===

inst#,actual,predicted,error,prediction
1,11:Perlis,5:Labuan,*,0.499

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) State
Start Stop

Result list (right-click for options):
01:16:43 - rules.MSRules
01:17:16 - functions.MultilayerPerceptron
01:24:30 - functions.MultilayerPerceptron
01:26:04 - trees.RandomTree
01:26:06 - trees.RandomTree
01:26:17 - trees.RandomTree
01:40:35 - **bayes.NaiveBayes**

Classifier output

	Perlis (0.06)	Putrajaya (0.06)	Sabah (0.06)	Sarawak (0.06)
mean	13.7143	13.7143	20.5714	20.5714
std. dev.	1.1429	1.1429	1.1429	1.1429
weight sum	1	1	1	1
precision	6.8571	6.8571	6.8571	6.8571

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) State
Start Stop

Result list (right-click for options):
01:16:43 - rules.MSRules
01:17:16 - functions.MultilayerPerceptron
01:24:30 - functions.MultilayerPerceptron
01:26:04 - trees.RandomTree
01:26:06 - trees.RandomTree
01:26:17 - trees.RandomTree
01:40:35 - **bayes.NaiveBayes**

Classifier output

	Perlis (0.06)	Putrajaya (0.06)	Sabah (0.06)	Sarawak (0.06)	Selangor (0.06)	Terengganu (0.06)
mean	13.7143	20.5714	20.5714	0	6.8571	13.7143
std. dev.	1.1429	1.1429	1.1429	1.1429	1.1429	1.1429
weight sum	1	1	1	1	1	1
precision	6.8571	6.8571	6.8571	6.8571	6.8571	6.8571

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

Set...

☒ Cross-validation

Folds

10

☐ Percentage split

%

66

More options...

(Nom) State

Start

Stop

Result list (right-click for options)

01:1643 - rules.MSRules

01:1716 - functions.MultilayerPerceptron

01:2430 - functions.MultilayerPerceptron

01:2604 - trees.RandomTree

01:2606 - trees.RandomTree

01:2617 - trees.RandomTree

01:4035 - bayes.NaiveBayes

Classifier output

	0.000	0.067	0.000	0.000	0.000	-0.067	0.000	0.063	Negeri Sembilan
	0.000	0.000	?	0.000	?	?	0.000	0.063	Pahang
	0.000	0.000	?	0.000	?	?	0.000	0.063	Pulau Pinang
	0.000	0.000	?	0.000	?	?	0.000	0.063	Perak
	0.000	0.067	0.000	0.000	0.000	-0.067	0.867	0.333	Perlis
	0.000	0.000	?	0.000	?	?	0.933	0.500	Putrajaya
	0.000	0.000	?	0.000	?	?	0.000	0.063	Sabah
	0.000	0.000	?	0.000	?	?	0.000	0.063	Sarawak
	0.000	0.000	?	0.000	?	?	1.000	1.000	Selangor
	0.000	0.000	?	0.000	?	?	0.200	0.077	Terengganu
Weighted Avg.	0.000	0.067	?	0.000	?	?	0.356	0.208	

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p

<-- classified as

0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

a = Johor

0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

b = Kedah

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

c = Kelantan

1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

d = Kuala Lumpur

0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0

e = Labuan

0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0

f = Melaka

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

g = Negeri Sembilan

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

h = Pahang

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

i = Pulau Pinang

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

j = Perak

0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

k = Perlis

0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

l = Putrajaya

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

m = Sabah

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0

n = Sarawak

1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

o = Selangor

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

p = Terengganu

5.0 Discussion

McDonald's is a beloved household name to many with over 38,000 locations across the globe in 2022. This is no exception for both Malaysia and Australia, being one of the biggest fast food chains within both countries. In this dataset, the number of McDonald's locations across Malaysia is displayed for each state within the country to give a clearer picture on the distribution of the fast food franchise. In Malaysia, McDonald's serves over 13.5 million customers a month in more than 320 restaurants nationwide. As per the findings of the Study, "having a 'Relaxing Break' makes up the largest share at 22%, whereby consumers are looking for a place with a fun and relaxing atmosphere to help them escape the demands of life and treat themselves with something special. This study also revealed that 'Convenient Location' and 'Value for Money', are the most important considerations in selecting the eat-out destination. A clear policy could be seen in McDonald's Malaysia and worldwide where they emphasise seven values which are:

- 1) We place the customer experience at the core of all we do;
- 2) We are committed to our people;
- 3) We believe in McDonald's system;
- 4) We operate our business ethically;
- 5) We give back to our communities;

- 6) We grow our business profitably; and
- 7) We strive continually to improve.

That is why McDonald's further expanded their business to provide the best fast food service and cover all around Malaysia, so every state must have one McDonald's to achieve the second value which is the McDonald's committed to the people to serve them equally.

It also contains data on McDonald's locations in Malaysia, organised by state. The dataset includes columns for the state, city, street address, latitude, and longitude. By looking at the number of McDonald's locations in each state.

Johor: 63
Kedah: 10
Kelantan: 8
Kuala Lumpur: 71
Labuan: 1
Melaka: 15
Negeri Sembilan: 15
Pahang: 10
Penang: 19
Perak: 20
Perlis: 2
Sabah: 36
Sarawak: 31
Selangor: 73
Terengganu: 6

Based on this data, we can see that Kuala Lumpur and Selangor have the highest number of McDonald's locations, with 71 and 73 locations respectively. Meanwhile, Perlis has the lowest number of McDonald's locations with only 2 locations.

6.0 Conclusion

Based on the dataset of McDonald's locations in Malaysia by state in 2022, it can be concluded that McDonald's has a strong presence throughout the country. The dataset shows that the majority of the McDonald's locations are located in the state of Selangor, followed by Kuala Lumpur and Johor. This suggests that McDonald's recognizes the importance of these regions in terms of consumer demand and market potential. Furthermore, the dataset also reveals that McDonald's has expanded its presence in the country by opening more branches in the states of Terengganu, Kelantan, and Perlis. Overall, the dataset illustrates McDonald's commitment to meeting the needs and preferences of its customers in Malaysia.

We also can conclude that :

- Selangor has the highest number of McDonald's locations in Malaysia with 126 restaurants.
- Johor is the state with the second-highest number of McDonald's locations with 73 restaurants.
- Kuala Lumpur, the capital city of Malaysia, has 54 McDonald's restaurants.
- The state with the least number of McDonald's locations is Perlis with only 3 restaurants.

McDonald's can definitely expand their locations in states with the lowest number of locations. By expanding to these areas, they can potentially tap into new customer markets and increase their revenue. However, they would need to conduct market research and analyse the feasibility of opening new locations in these areas. Factors such as population density, demographics, and competition would need to be taken into consideration before deciding to expand to a particular state or location. Ultimately, it would depend on the business strategy and goals of McDonald's in Malaysia.

7.0 Task Distribution

TASKS	NAMES	PARTICIPATION
-------	-------	---------------

Introduction	Fatina	100
Data source	All	100
Concepts of classification	Fatina	100
Multilayer perceptron & Results	Zuhayra+Fatina	100
Regression By Discretization & Results	Zuhayra+Fatina	100
Random Tree & Results	Zuhayra+Fatina	100
Naive bayes & Results	Zuhayra+Fatina	100
Discussion	All	100
Conclusion	Zuhayra	100

8.0 Appendix

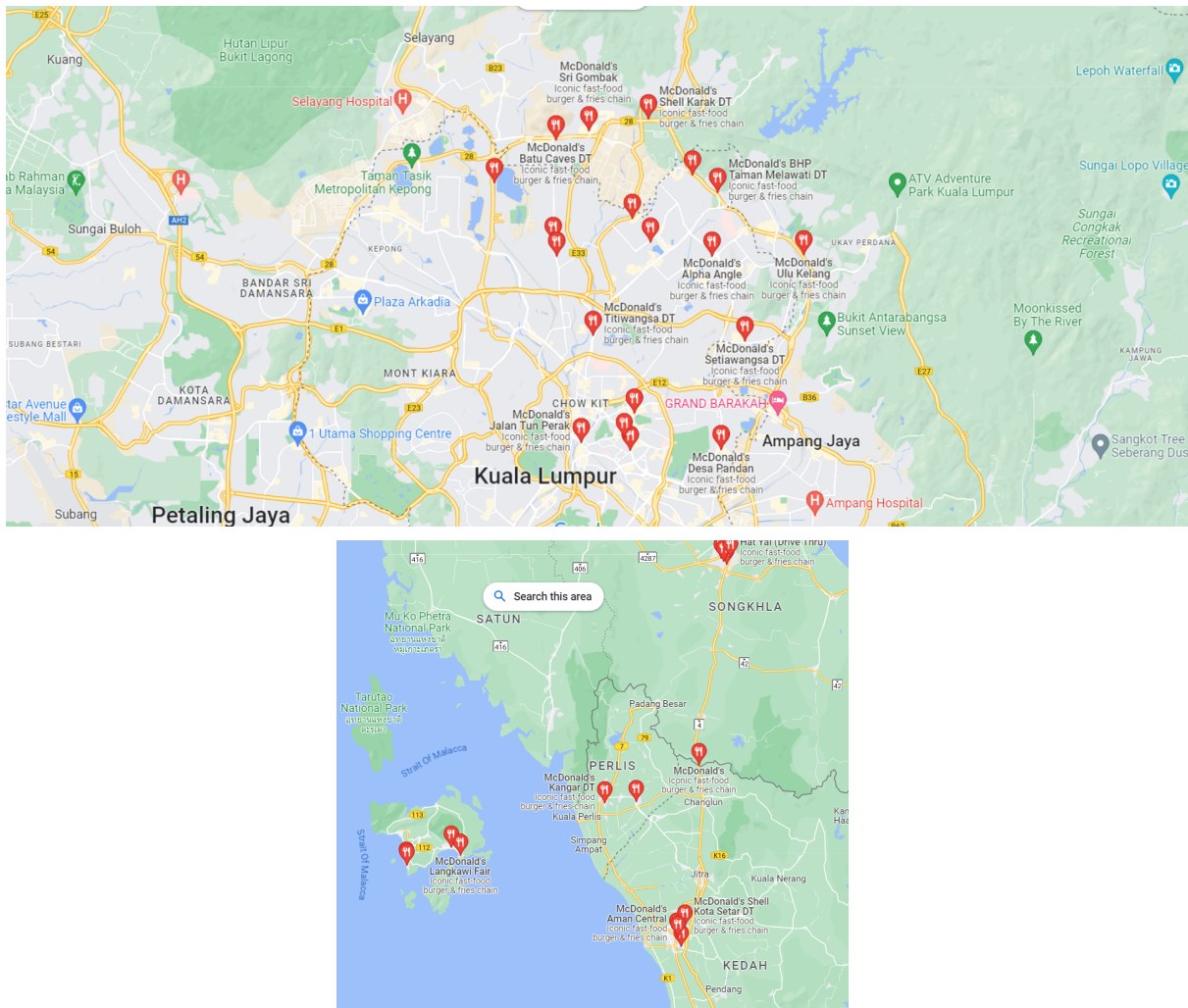
AutoSave **OFF** mcd_locations_in_malaysia Zuhayra Zulkifli

File Home Insert Page Layout Formulas Data Review View Automate Help

Clipboard Font Alignment Number Styles Cells Editing Analysis

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	State	Number of Locations																			
2	Johor	40																			
3	Kedah	10																			
4	Kelantan	6																			
5	Kuala Lumpur	45																			
6	Labuan	1																			
7	Melaka	12																			
8	Negeri Sembilan	16																			
9	Pahang	17																			
10	Pulau Pinang	23																			
11	Perak	18																			
12	Perlis	1																			
13	Putrajaya	4																			
14	Sabah	15																			
15	Sarawak	13																			
16	Selangor	97																			
17	Terengganu	5																			
18																					
19																					



9.0 References

1. Admin. (2017, December 6). *McCafe speciality now affordable, accessible*. *MARKETING Magazine*. <https://marketingmagazine.com.my/mccafe-speciality-now-affordable-accessible/>
2. CHRISTINE CHIONG. (2022). *McDonald's locations by state in 2022*. *Kaggle.com*. https://www.kaggle.com/datasets/christinechiong/mcdonalds-locations-in-malaysia-by-state-in-2022?select=mcd_locations_in_malaysia.csv
3. *McDonald's Malaysia | Locate us*. (n.d.). *McDonald's Malaysia*. <https://www.mcdonalds.com.my/locate-us>

4. UTHM. (n.d.). *The Best Practices of Mcdonald's Malaysia: A Case Study of the National Dual Training System*. Penerbit UTHM.

<https://penerbit.uthm.edu.my/ojs/index.php/JTET/article/download/13505/5646>