

# Evaluation of Malaysia's 15th General Election Prediction Models by using Cross Validation Technique

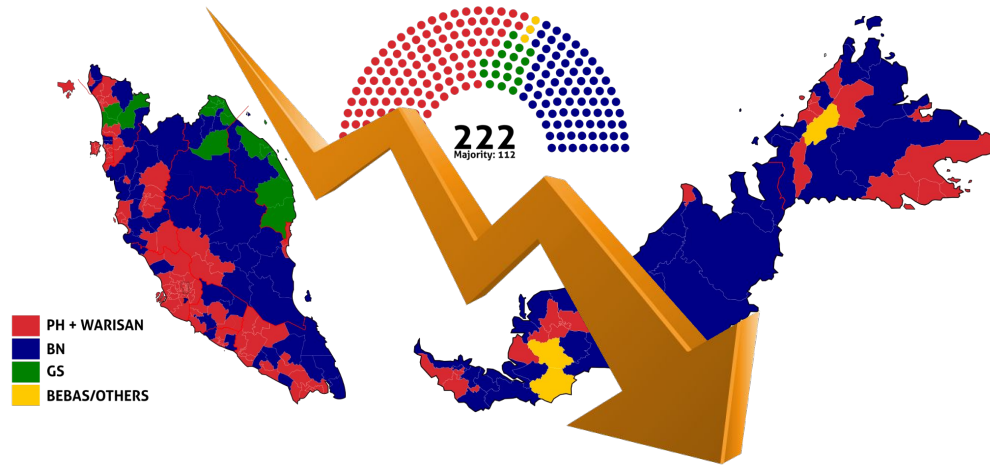
“Previous election result indicated the incumbency of BN as the ruling government ended when the party lost 54 numbers of parliament seats and from this figure 31 goes to PH with margin majority. Based on the ethnic composition, those are parliaments with slight majority of Malay/Bumi voters. Due to many issues related to the previous leadership, BN could possibly lost their Malay/Bumi votes share to the rival party.”

“The strategy to embark 3 corner fight which include another Malay influence party(whis is PAS) only give mileage to PH. However, based on the scenario from the past 3 years the marginal result only manifest political instability. We believe only politically stable government can revive the current economy.”

“Therefore by assuming the same voting pattern in the previous election, prediction model of incoming general election being developed. This model can be used to implement prescriptive analytics whereby simulation of result can be predicted based on engagement of political alliance”

## PROBLEM STATEMENT

“Political instability is jeopardizing nation building as so much time and money wasted on political driven effort. The problem occur when the new government on the previous general election was formed based on marginal parliamentary seats. This scenario triggered increasing pattern of parliament members from the losing alliance hopping to the opposing alliance which give him or her more benefits. Based on this fact, we believed strategizing political alliance guided by predicted best result scenario based on developed machine learning model could avoid similar issue from happening in the future, whereby improving political stability in the country”



# OBJECTIVE

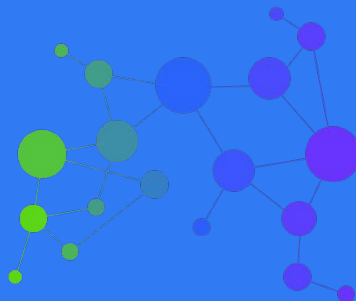
## DEVELOP

Development of **DATA MODEL** derived from multiple resources of open data for the implementation of machine learning.



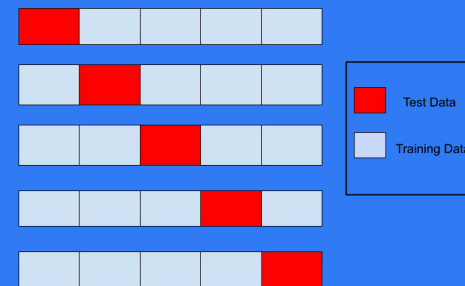
## TEST

Test developed **DATA MODEL** using 3 different **MACHINE LEARNING ALGORITHMS**.



## EVALUATE

Evaluate developed prediction models results using **CROSS VALIDATION** approach.



# RESEARCH BENEFICIARIES

## POLITICAL PARTIES

- UMNO
- PAS
- BERSATU
- AMANAH
- PKR
- DAP
- PBB
- WARISAN

## RESEARCH CENTER

- IDEAS
- INSAP
- THE CENTRE
- MERDEKA  
RESEARCH
- KRI
- EMIR RESEARCH

## GOVERNMENT

- MALAYSIA  
ELECTION  
COMMITTEE
- MALAYSIA ANTI  
CORRUPTION  
COMMISSION
- POLICE

# REFERENCES

## **SOCIAL SCIENCE PERSPECTIVE**

Reference on journals related to parameters on **VOTERS PROFILE** and **PREFERENCE** during the election.

## **ALGORITHM BASED PERSPECTIVE**

Reference on journal related to approach implementation of **BINARY TYPE CLASSIFICATION** and **MULTIPLE TYPE CLASSIFICATION** algorithms.

## **TOOLS BASED PERSPECTIVE**

Reference on journal related to tools utilisation to implement **DATA MODELLING** and **MACHINE LEARNING**.

## REFERENCES : SOCIAL SCIENCE PERSPECTIVE

“Ahmed, Skoric & Hilber, 2018, suggest that ML predictive model develop based on Twitter data from 3 country which include India, Pakistan and Malaysia produced satisfactory result for India and Pakistan only, however the model performance in Malaysia is considered unacceptable.”

“Chauhan & Sharma, 2020, suggest that election result can be forecasted by leveraging data mining approach from social media platforms”

“Kamakura, Mazzon & Bruyn, 2006, extracted political science theories from literature reviews to predicts electors preference using Nested Logit Factor Model ”

“Ryan Kennedy, Stefan Wojcikand & David Lazer, 2017, using election data which covering 86 countries suggest that political institutions, incumbent advantage, international linkage and aid did have strong influence to voters voting pattern.”

“M. Moniruzzaman & Kazi Fahmida Farzana, 2018, suggest that the dominance of UMNO in gaining support from Malay based voters had been weaken since 1999 when Anwar Ibrahim being ousted from the party and it become worse after Tun Mahathir in 2018 came into strategic partnership Pakatan Harapan”

## REFERENCES : ALGORITHMS BASED PERSPECTIVE

“Koli, Abdul M. & Ahmed, Muqem, 2018 in their study utilised features selection technique which include Random Forest, Decision Tree Classifier, Gradient boosting Classifier and Extra Gradient Boosting that produce 8 important parameters (Central Influence, Religion Followers, Party Wave, Party Abbreviations, Sensitive Areas, Vote Bank, Incumbent Party, and Caste Factor) to be implement in the ML models (Decision tree, Random forest, K-nearest neighbor and support vector machine) to predict election outcome for several states in India, especially Jammu and Kashmir.”

“Hayat Ullah, Bashir Ahmad, Iqra Sana, Anum Sattar, Aurangzeb Khan, Saima Akbar & Muhammad Zubair Asghar, 2021 in their study evaluated online reviews from Twitter to predict the user political affiliation . By implementing method proposed by opinion mining researchers, the data pre-processing annotated with different polarity classes and from here, different ML classifier (K-nearest neighbor, naïve Bayes, support vector machine, extreme gradient boosting) being implement to predict user political affiliation based on their online review”



## REFERENCES : TOOLS BASED PERSPECTIVE

“S. Asha Kiranmai & A. Jaya Laxmi, 2018, in their study to analyse power quality related data used WEKA as their analytics platform whereby data mining algorithms within that platform (J48, Random Tree and Random Forest decision trees) being implement and produced convincing result”

“Gnanambal, S; Thangaraj, M; Meenatchi, and V T; Gayathri, V., 2018, suggest that attributes selection is one of important process to be implement when doing data mining. They use WEKA to select attributes based on ranking produced by the platform. The verification process proceeded by comparing ML classification result performance before and after proposed attributes selection”

“Arif Senja Fitrani, Fajrillah Fajrillah and Wirda Novarika, 2019, used WEKA platform to predict participation of Governor And Vocational Governor Selection in Jemirahan Village, Jabon District, Indonesia. They implemented Naive Bayes classification using 6 pre-defined features for prediction.”



- | DATA SOURCING              | PROCESS                       | MODEL                                  | TRAIN   | EVALUATE  |
|----------------------------|-------------------------------|--|---|---|
| 1. Election Result         | 1. ETL Process                | 1. Feature Selection and Normalisation | 1. Train Developed Data Model using 3 Different ML Algorithms | 1. Evaluate Prediction Models Performance using Cross Validation Method |
| 2. Urbanisation Categories | 2. Data Cleaning              | 2. Data Model Development              |   |   |
| 3. Age Composition         | 3. Exploratory Data Analytics |  |   |   |
| 4. Ethnicity composition   | 4. Descriptive Analytics      |  |   |   |
|                            | 5. Diagnostic Analytics       |  |   |   |

# HIGHLIGHT ON FINDINGS

## MODEL RELIABILITY

**TRUE  
POSITIVE**

84.2% -  
Random Forest

**PRECISION**

84.2% -  
Random Forest

**ROC**

92.3% -  
Random Forest

## MODEL PROCESSING SPEED

**NAIVE BAYES**

Better model  
reliability with  
fastest processing

**LOGISTIC  
REGRESSION**

Good model  
reliability with fast  
processing

**RANDOM  
FOREST**

Best model  
reliability with  
faster processing

# PROPOSED PREDICTION MODEL OPERATIONAL FRAMEWORK

✓ PARLIAMENT	DUN	DM	LOKALITI
• 222	• 600	• 7500++	• 33000++
<div>✓</div> <ul style="list-style-type: none"><li>Static Data<ul style="list-style-type: none"><li>Result</li><li>Urbanisation</li><li>Age</li><li>Ethnic</li></ul></li><li>Dynamic Data<ul style="list-style-type: none"><li>News</li><li>Activities</li><li>Wishlist</li></ul></li></ul>	<ul style="list-style-type: none"><li>Static Data<ul style="list-style-type: none"><li>Result</li><li>Urbanisation</li><li>Age</li><li>Ethnic</li></ul></li><li>Dynamic Data<ul style="list-style-type: none"><li>News</li><li>Activities</li><li>Wishlist</li></ul></li></ul>	<ul style="list-style-type: none"><li>Static Data<ul style="list-style-type: none"><li>Result</li><li>Urbanisation</li><li>Age</li><li>Ethnic</li></ul></li><li>Dynamic Data<ul style="list-style-type: none"><li>News</li><li>Activities</li><li>Wishlist</li></ul></li></ul>	<ul style="list-style-type: none"><li>Static Data<ul style="list-style-type: none"><li>Result</li><li>Urbanisation</li><li>Age</li><li>Ethnic</li></ul></li><li>Dynamic Data<ul style="list-style-type: none"><li>News</li><li>Activities</li><li>Wishlist</li></ul></li></ul>
<div>✓</div> <ul style="list-style-type: none"><li>Model Train<ul style="list-style-type: none"><li>Static</li><li>Random Forest</li></ul></li><li>Enhanced Model Train<ul style="list-style-type: none"><li>Static+Dynamic</li><li>Random Forest</li></ul></li></ul>	<ul style="list-style-type: none"><li>Model Train<ul style="list-style-type: none"><li>Static</li><li>Random Forest</li></ul></li><li>Enhanced Model Train<ul style="list-style-type: none"><li>Static+Dynamic</li><li>Random Forest</li></ul></li></ul>	<ul style="list-style-type: none"><li>Model Train<ul style="list-style-type: none"><li>Static</li><li>Random Forest</li></ul></li><li>Enhanced Model Train<ul style="list-style-type: none"><li>Static+Dynamic</li><li>Naive Bayes</li></ul></li></ul>	<ul style="list-style-type: none"><li>Model Train<ul style="list-style-type: none"><li>Static</li><li>Random Forest</li></ul></li><li>Enhanced Model Train<ul style="list-style-type: none"><li>Static+Dynamic</li><li>Naive Bayes</li></ul></li></ul>

## CONCLUSION

“We believe the developed prediction models can be used to predict the upcoming election result with at least 80% accuracy. The accuracy however could be improvised if similar data from GE-12 and GE-13 integrated in the existing data model. ”

“In term of operational purpose, the stakeholder can adapt the proposed framework based on their data availability, scope and operation flow. The proposed model enable the stakeholder to integrate other sources of data such as detail profiling of the voters (B40, BR1M, TEKUN, etc) and locations (such as availability to public amenities) ”

“Similar approach could be implemented at DUN, DM and LOKALITI level, which dependent on user data availability”

“The insight obtained can be leverage to strategize political move for the upcoming election to ensure either any of the alliance could comfortably secured parliamentary seats and therefore promote political stability in Malaysia ”



## *Certificate of Participation*

This certificate is proudly presented to

**Izardy Amiruddin and Mohd Syahid Mohd Anuar**

for the paper entitled

**Evaluation Of Malaysia's 15th General Election Prediction  
Models By Using Cross Validation Approaches**

presented at the

**Razak Annual Technology, Informatics and Policy Seminar  
(RATIPS 2021)**

in conjunction with

**Razak Research Week 2021  
6<sup>th</sup> – 10<sup>th</sup> December 2021**

organized by

**Razak Faculty of Technology and Informatics,  
Universiti Teknologi Malaysia**

A handwritten signature in black ink, appearing to read 'Astuty', is placed above the name of the Dean.

**PROF. DR. ASTUTY AMRIN**

Dean

Razak Faculty of Technology and Informatics  
Universiti Teknologi Malaysia  
Kuala Lumpur