UNIVERSITI TEKNOLOGI MARA
Cawangan Kedah
Kampus Sungai Petani

2021 ICMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

Date:
4 - 5 August 2021

"DRIVING RESEARCH TOWARDS EXCELLENCE"

Virtual Conference

The 5th International Conference on Computing, Mathematics and Statistics 2021
(iCMS2021)

Research and Innovation: Opportunities and Challenges for the Academia

Professor Dr Yap Bee Wah
Faculty of Computer and Mathematical Sciences, UiTM Shah Alam
IBDAAI, Universiti Teknologi MARA

Cawangan Kedah
Kampus Sungai Petani

Fakulti
Sains Komputer
Dan Matematik

IBDAAI
Institute for Big Data Analytics and Artificial Intelligence

CONTENTS

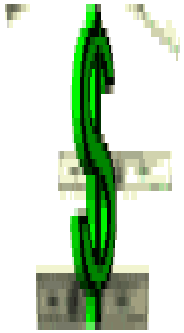
RESEARCH INNOVATION

data driven research

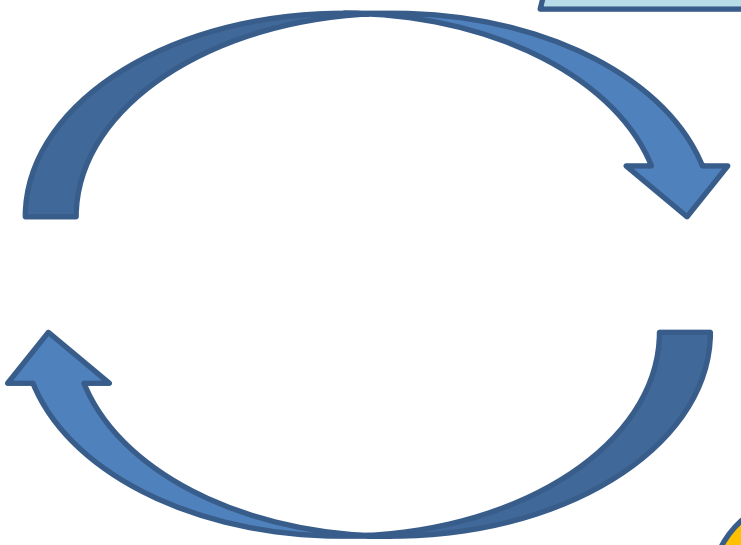
Research Grants

1. Research and Innovation

Research involves investigation of a phenomenon that leads to knowledge generation.



Research



Innovation is the creation or improvement of a product, process or service.

Innovation

- **Research** -transformation of money into knowledge.
- **Innovation** - transformation of knowledge into money

Dr Geoffrey Nicholson, "Father of Post-it Notes", on 3M & Innovation

2. Research Process

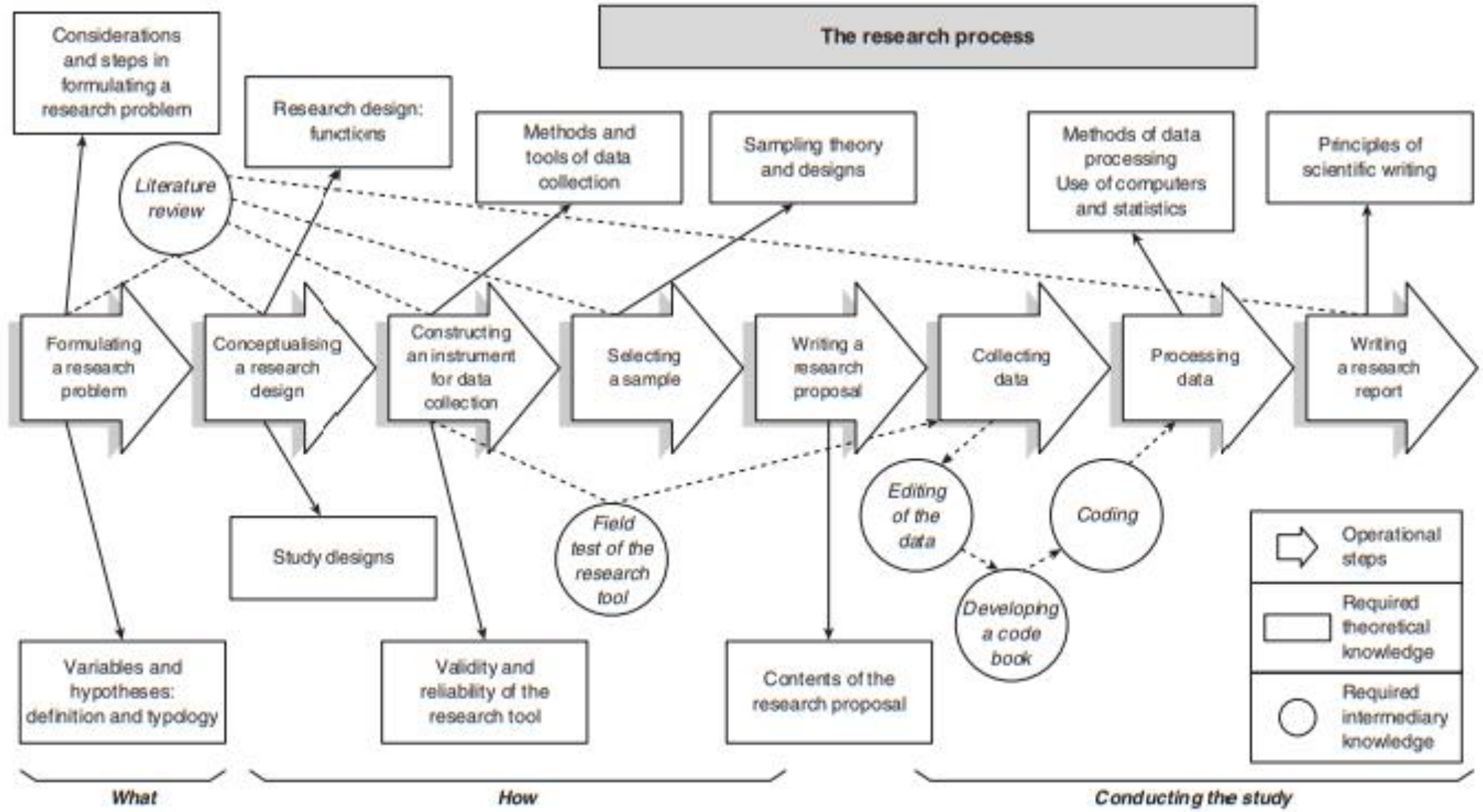


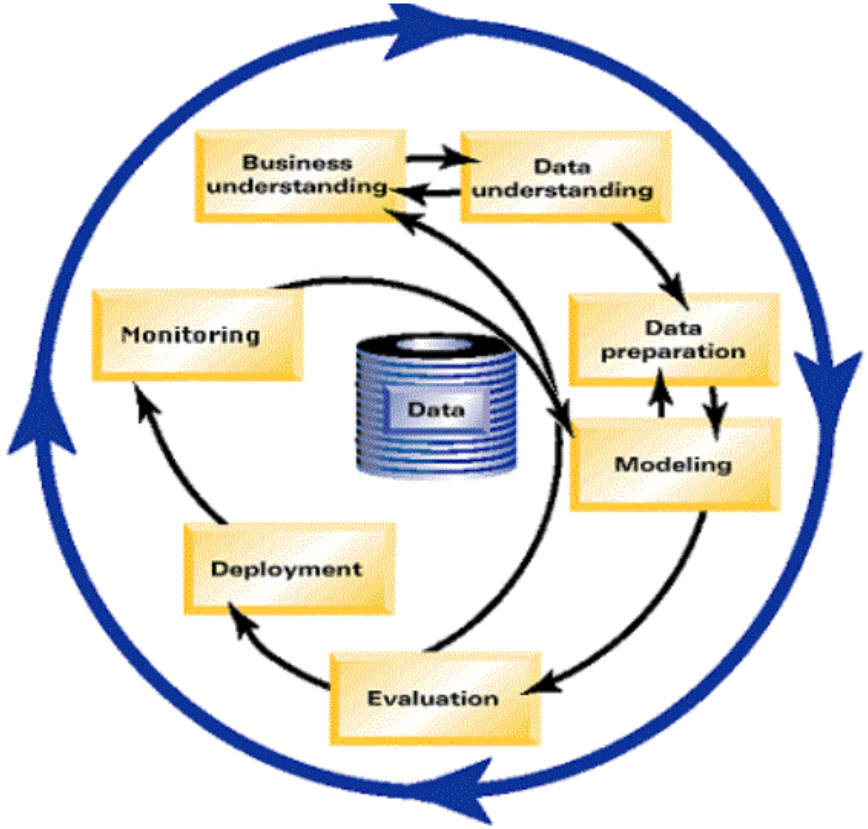
FIGURE 2.2 *The research process*

Source: Ranjit Kumar (2011), Research Methodology: a step-by-step guide for beginners. Chapter 3.

This type of research is usually theory driven.

Data Driven Research Process

- Follows CRoss-Industry Standard Process for Data Mining
- It is a standard framework develop to help to carry out DM projects.
- It was developed by a consortium of companies mainly in Europe.
- It involves 6 phases/steps.
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
  - Deployment



Data Driven Research

Reports/Publications

sqoop

hadoop APACHE HBASE

cloudera IMPALA

APACHE kafka A distributed streaming platfo

FLUME

HIVE

HDFS

APACHE Spark

Extraction

Store

Process

python

R

rapidminer

IBM SPSS Modeler

123 abc

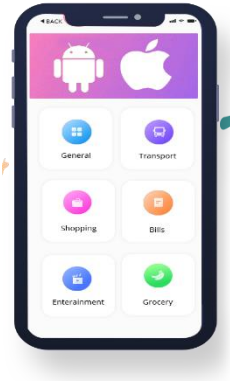
APACHE Spark

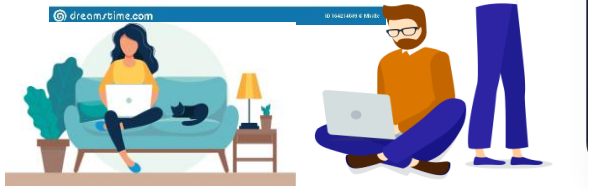
Data Visualization

Data Analytics

Innovation

Mobile/web application





Data Professor

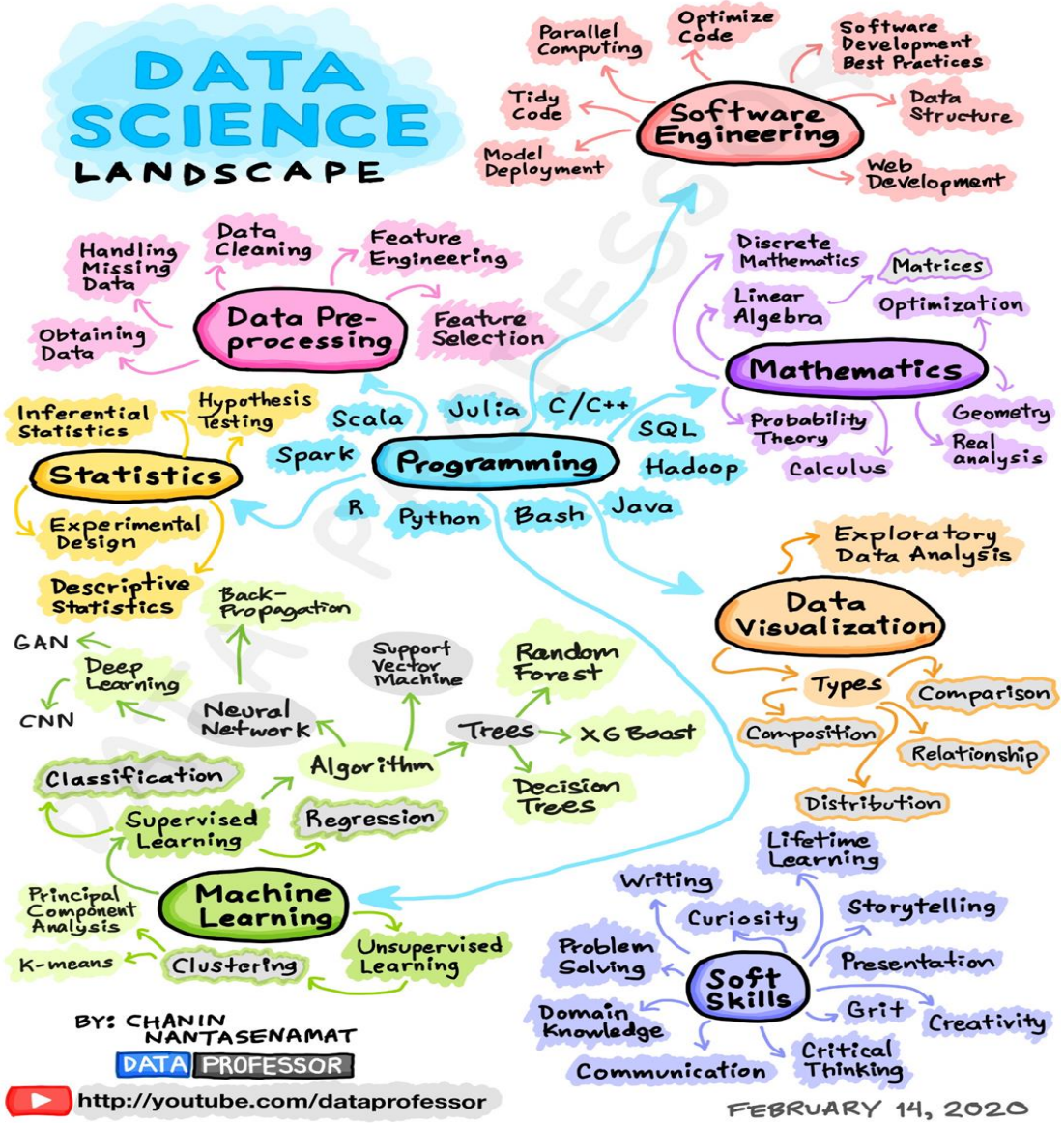
Data Science, Machine Learning, Bioinformatics, Research and Teaching are my passion. The Data Professor YouTube...

www.youtube.com

Connect with Me on Social Network

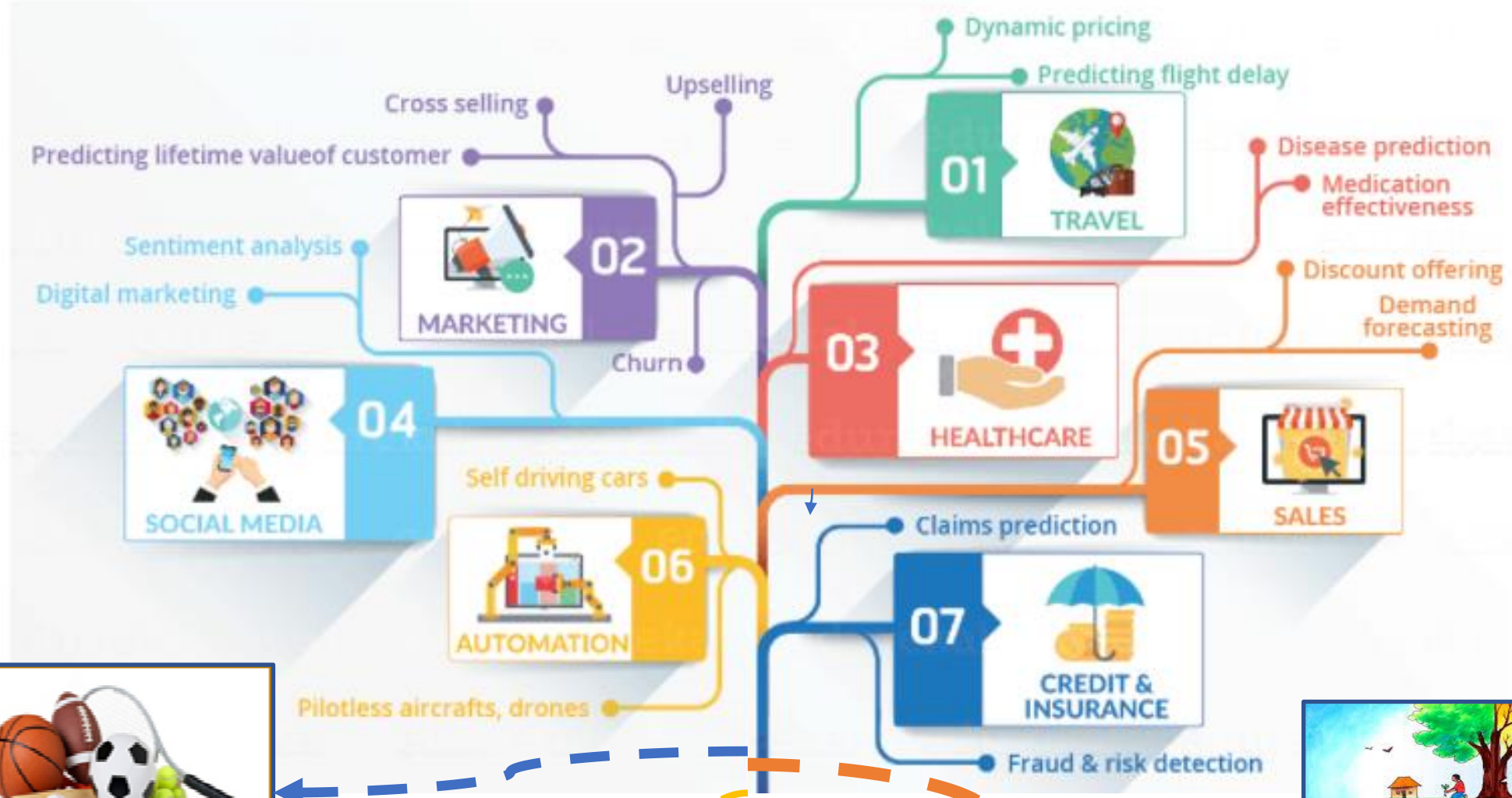
✅ YouTube: http://youtube.com/dataprofessor/
✅ Website: http://dataprofessor.org/ (Under construction)
✅ LinkedIn: https://www.linkedin.com/company/dataprofessor/
✅ Twitter: https://twitter.com/thedataprof/
✅ FaceBook: http://facebook.com/dataprofessor/
✅ GitHub: https://github.com/dataprofessor/
✅ Instagram: https://www.instagram.com/data.professor/

Associate Professor
Dr Chanin Nantasenamat
Mahidol University



https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b

Data Science Research Applications

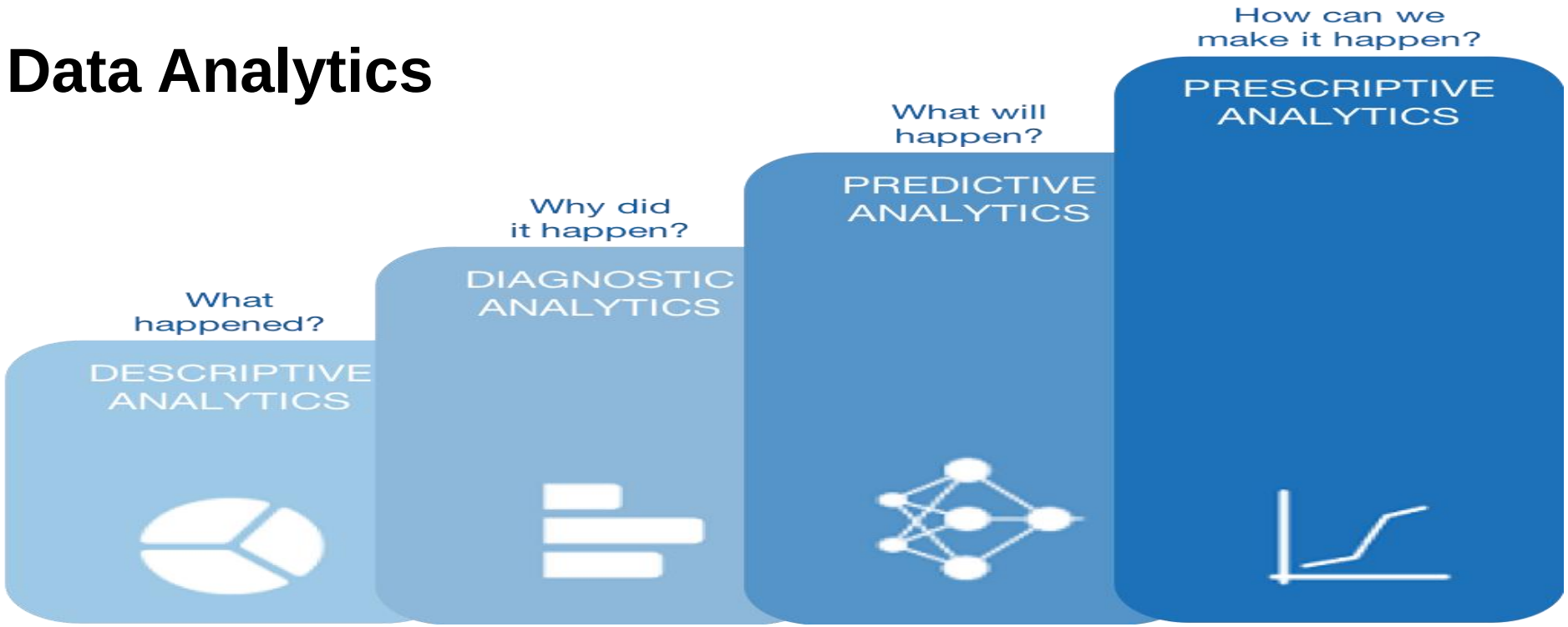


Sports

Environment

Source: google.com/search?source=univ&tbm=isch&q=data+science+images&sa=X&ved=2ahUKEwiE-MzYl57wAhWYzTgGHUVnBp4Q7Al6BAgCEFA&biw=1024&bih=465#imgrc=G1c5jaX6Kvis2M

Data Analytics



Reporting **summary statistics** based on your data.

Data Visualization to get insights from your data. Statistical analysis

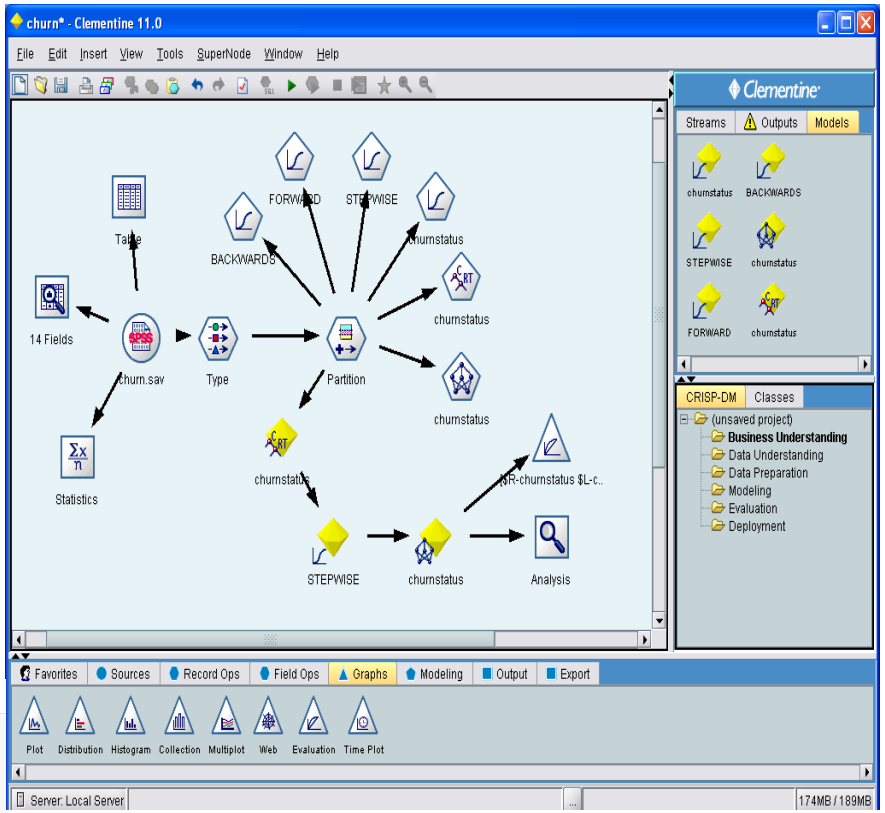
Evaluating several prediction/ machine learning models.

Recommend the best prediction model.

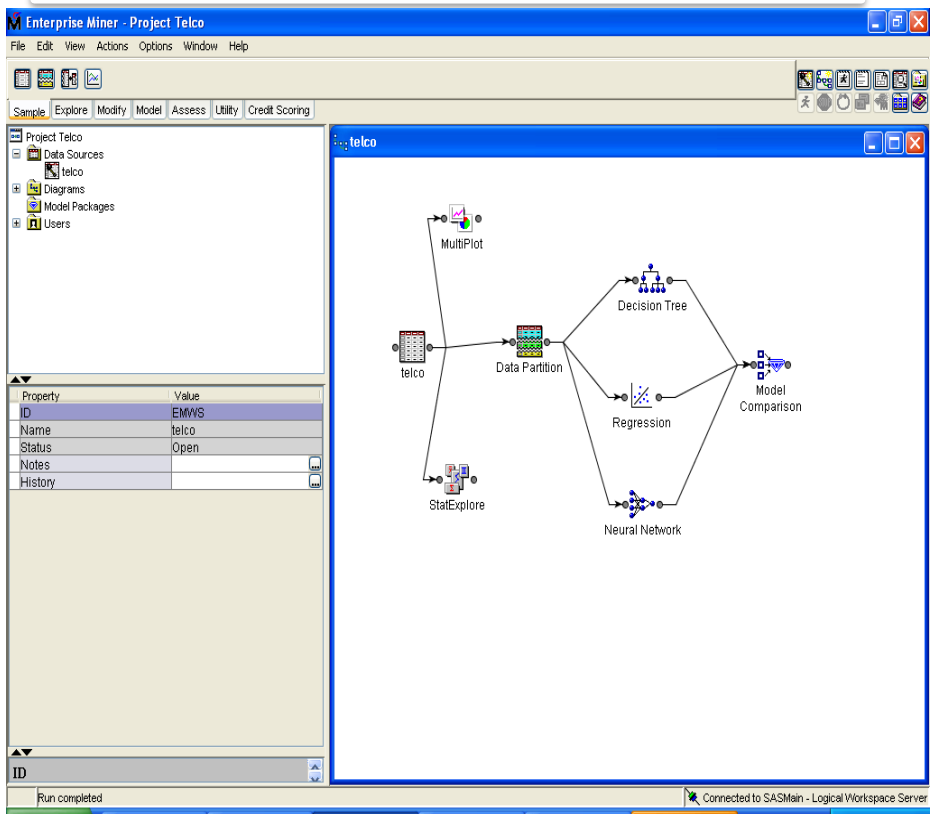
Opportunities

DATA ANALYTICS TOOLS- data analysis is faster and easier

IBM SPSS MODELER 18

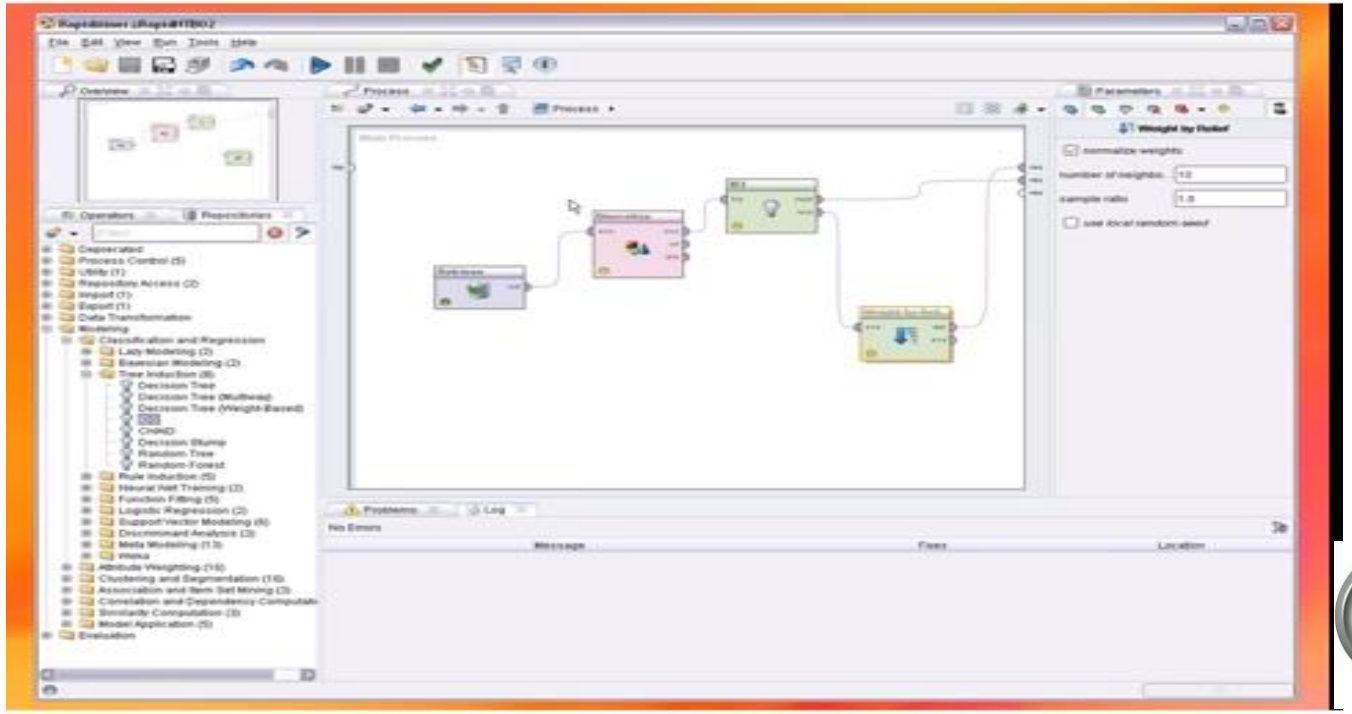


SAS ENTERPRISE MINER



Licensed software

MATLAB
Programming language



rapidminer

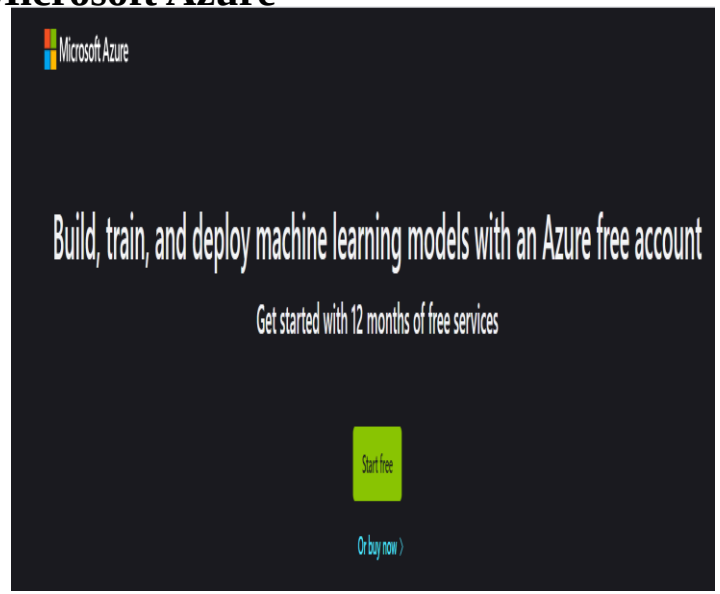
https://rapidminer.com/

WEKA
The University of Waikato

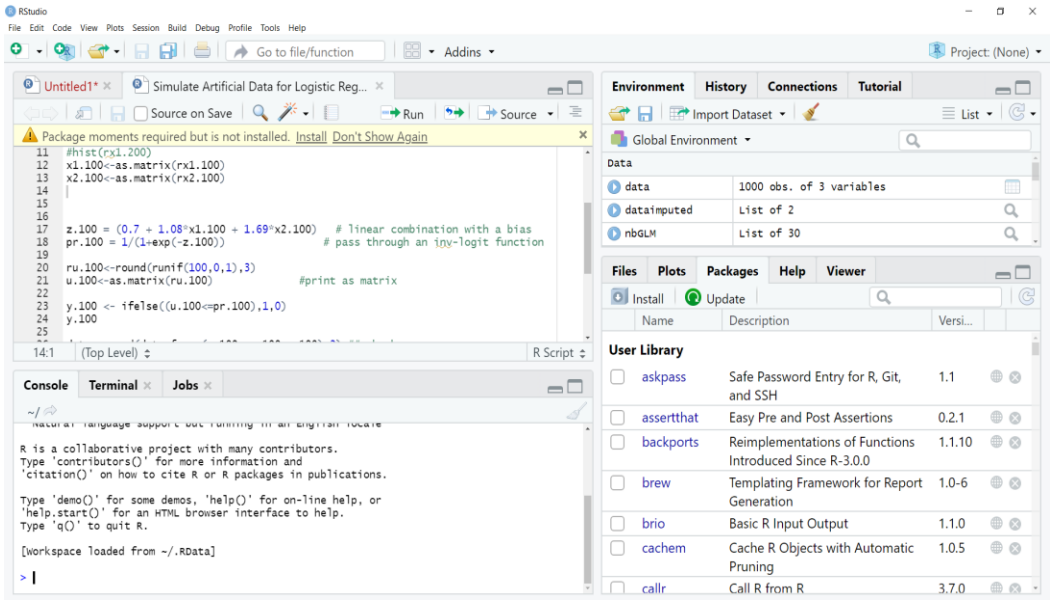
R

python

Microsoft Azure



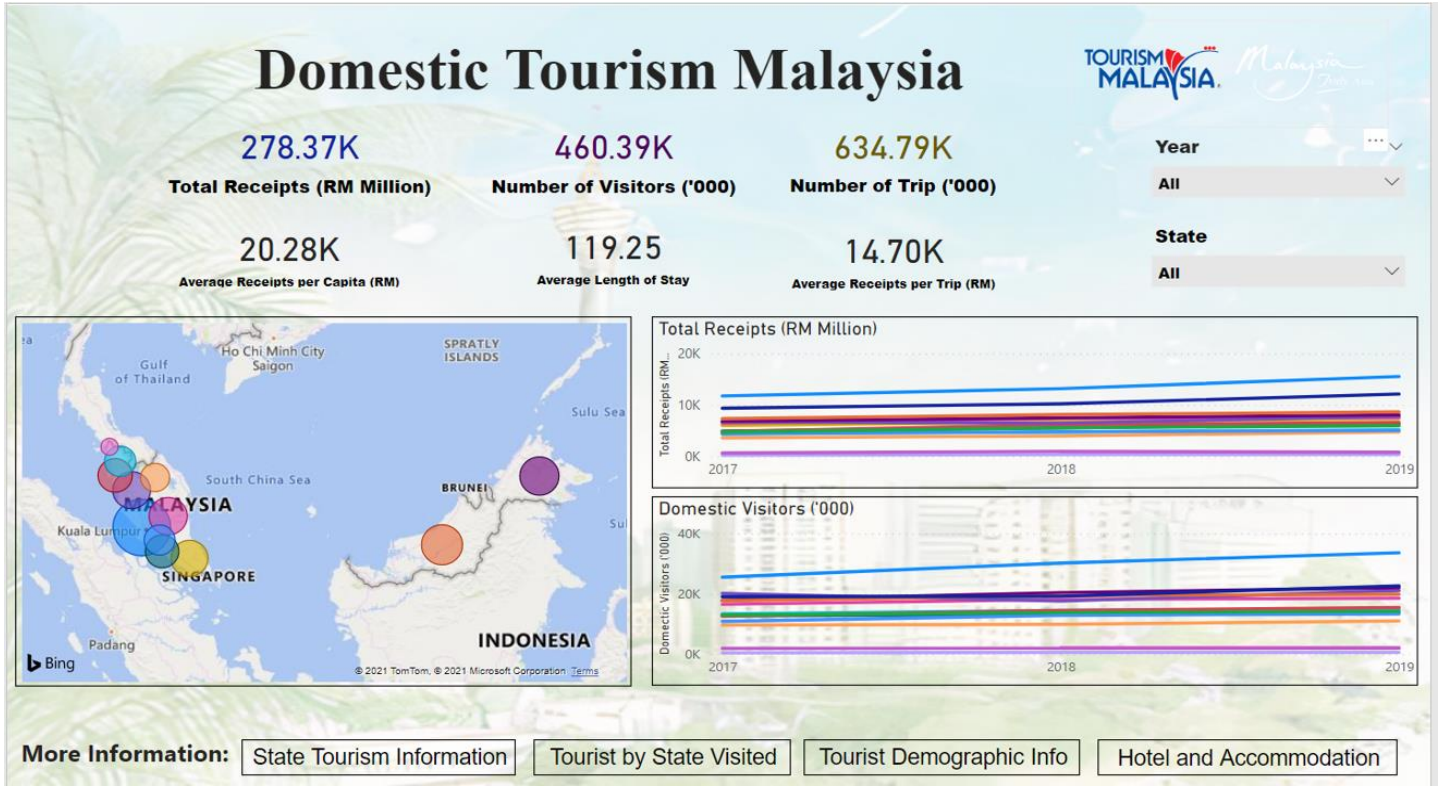
R studio



DECISION SUPPORT SYSTEMS

Power BI

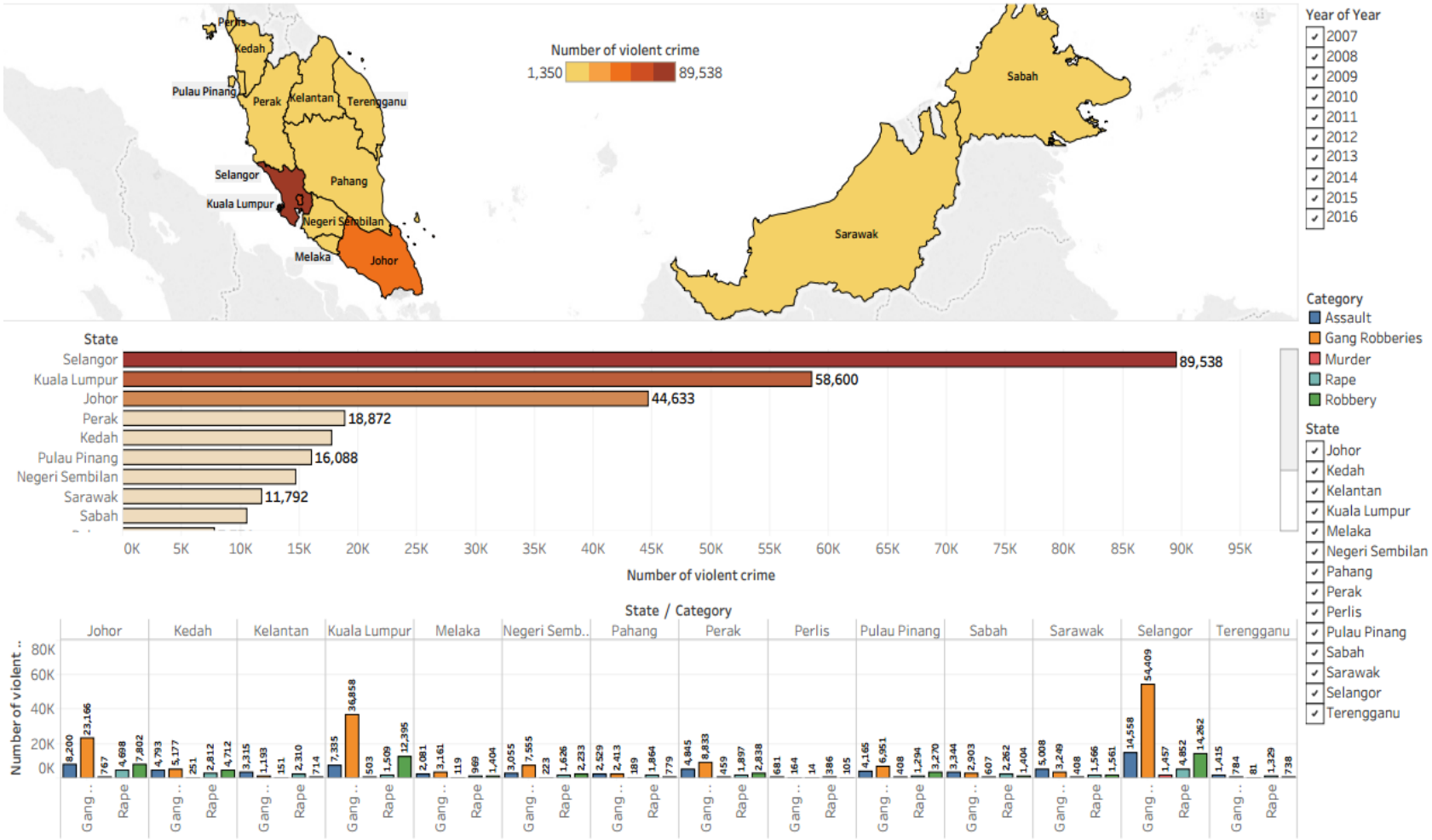
+ableau



Dashboard using Power BI

Shareh Zulhelmi
Master of Data Science, UiTM

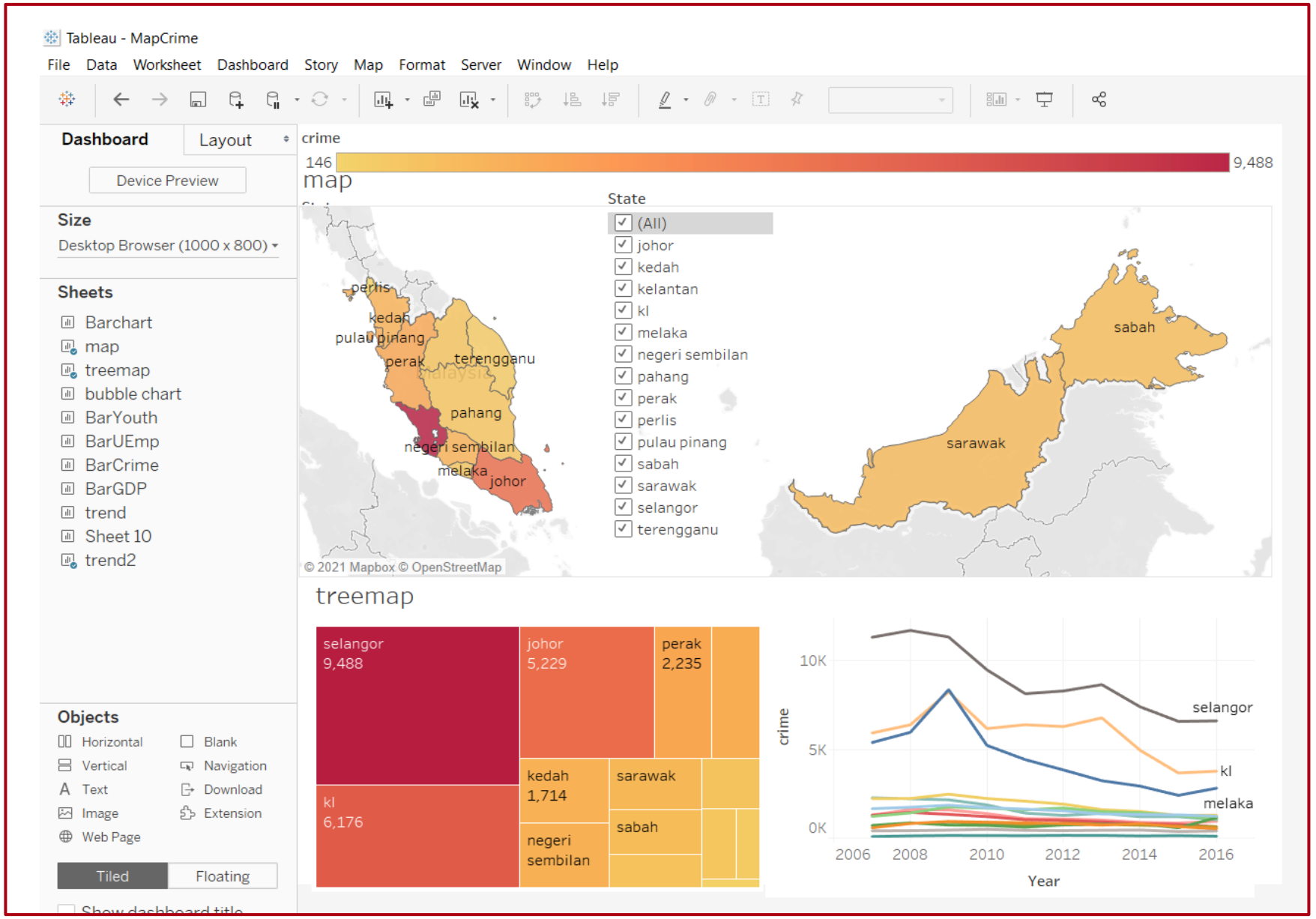
Data Visualization of Violent Crime in Malaysia -Namelya Anuar



https://www.tableau.com/academic/teaching

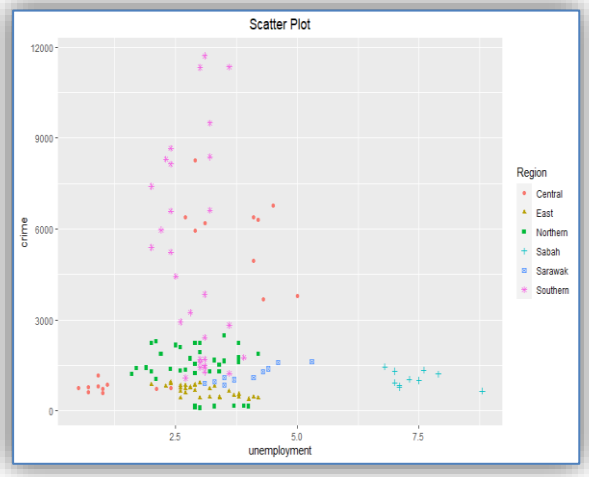
Tableau dashboard

Data Visualization

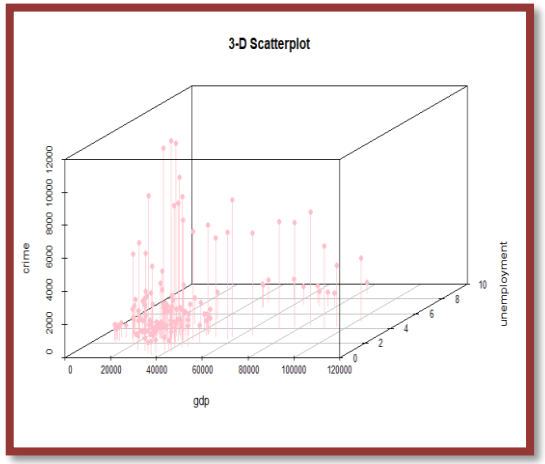


Data Visualization

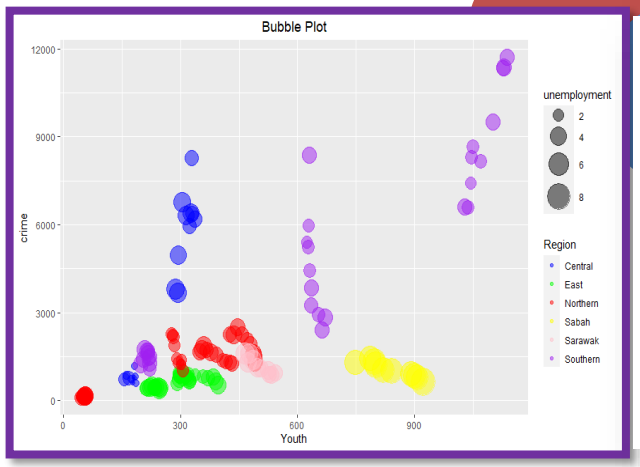
SCATTER PLOT



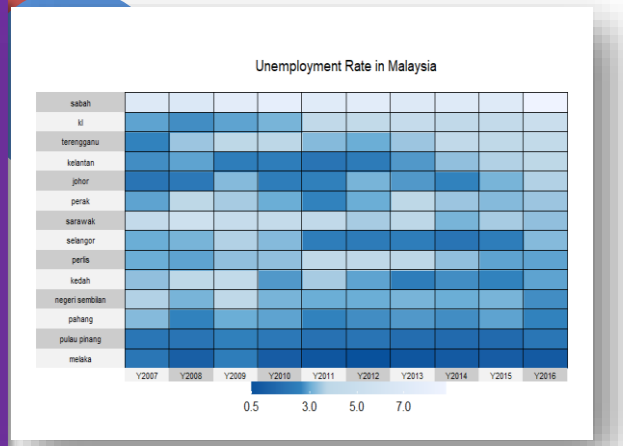
3-D PLOT



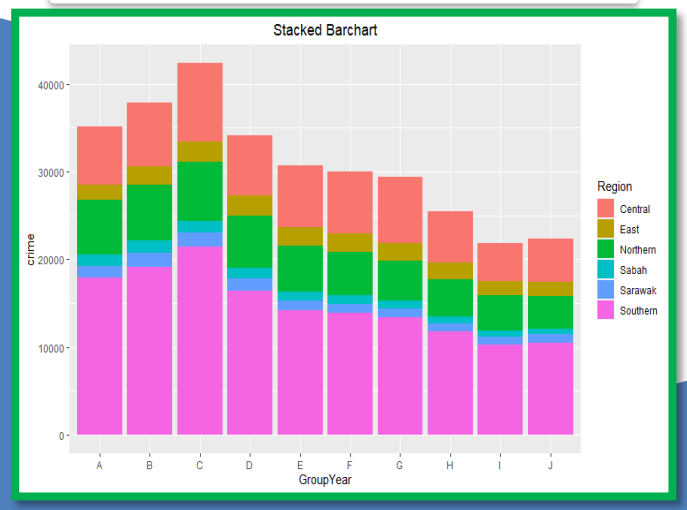
BUBBLE PLOT



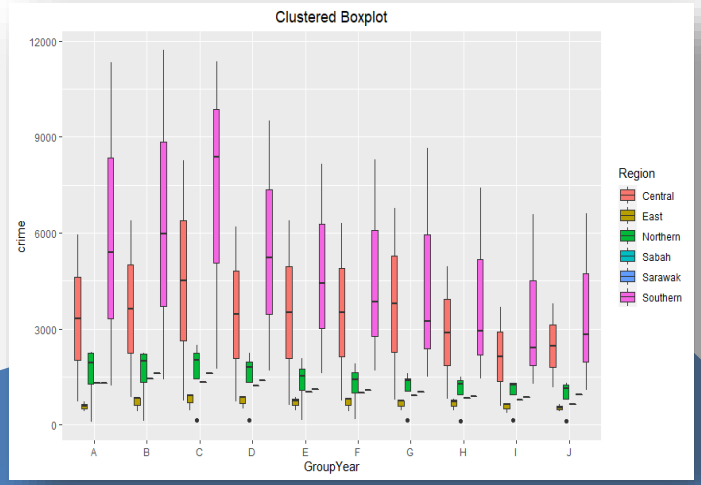
HEAT MAP



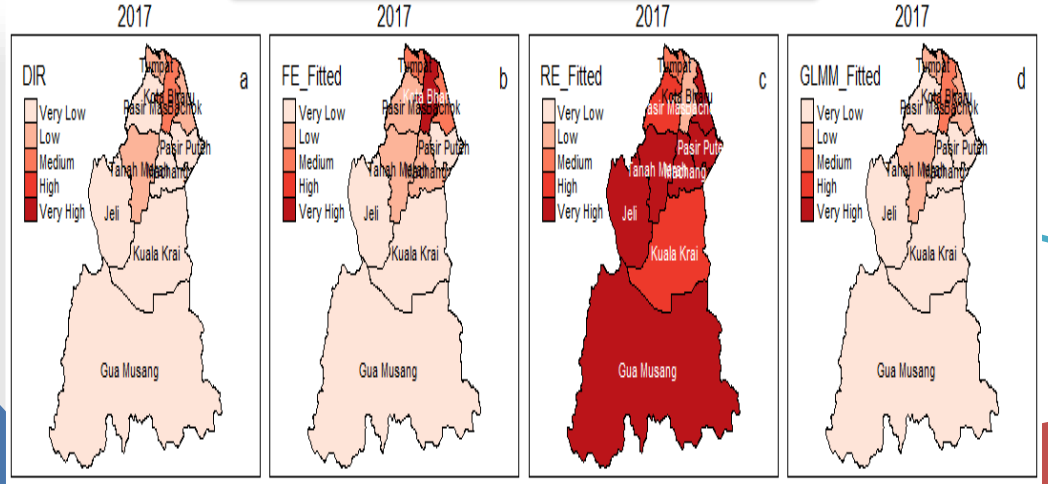
STACKED BAR CHART



CUSTERED BOX PLOT



SPATIAL MAPS



Opportunities for Research Publications and Innovation:

Focus on:
Predictive Analytics
(or **Supervised Learning**)

DATA MINING PROCESS

Step 1:
Identify the Problem & goal of Data Mining

Step 2:
Data Understanding (Selection of Variables)

Step 3:
Data Preparation (Data Integration, data cleaning, data imputation)

Step 4:
Data Mining

- Supervised Learning
  - ✓ Logistic Regression
  - ✓ Decision Tree
  - ✓ ANN
  - ✓ SVM
- Unsupervised Learning
  - ✓ K-means
  - ✓ Fuzzy C-means
- Association Analysis
  - Apriori algorithm

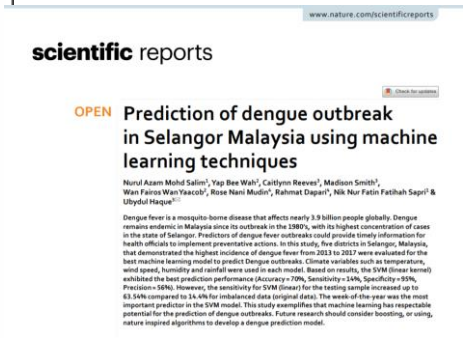
Step 5:
Evaluation of results

Step 6:
Deployment

Innovation

Predictive Model Apps

Publications



www.nature.com/scientificreports

scientific reports

OPEN **Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques**

Nurul Azam Mohd Salim[1], Yap Bee Wah[2], Caitlynn Reeves[3], Madison Smith[3], Wan Fairos Wan Yaacob[1], Rose Nani Mudin[4], Rahmat Dapari[5], Nik Nur Fatin Fatihah Sapri[5] & Ubydul Haque[3,6]

Dengue fever is a mosquito-borne disease that affects nearly 3.9 billion people globally. Dengue remains endemic in Malaysia since its outbreak in the 1980's, with its highest concentration of cases in the state of Selangor. Predictors of dengue fever outbreaks could provide timely information for health officials to implement preventative actions. In this study, five districts in Selangor, Malaysia, that demonstrated the highest incidence of dengue fever from 2013 to 2017 were evaluated for the best machine learning model to predict Dengue outbreaks. Climate variables such as temperature, wind speed, humidity and rainfall were used in each model. Based on results, the SVM (linear kernel) exhibited the best prediction performance (Accuracy = 70%, Sensitivity = 14%, Specificity = 95%, Precision = 56%). However, the sensitivity for SVM (linear) for the testing sample increased up to 63.54% compared to 14.4% for imbalanced data (original data). The week-of-the-year was the most important predictor in the SVM model. This study exemplifies that machine learning has respectable potential for the prediction of dengue outbreaks. Future research should consider boosting, or using, nature inspired algorithms to develop a dengue prediction model.

Challenges in Research & Innovation for Academia

Research Skills

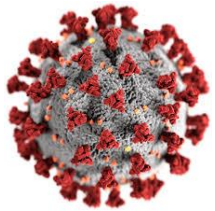
Researchers

- Work with **experts** in the area (mentor-mentee)
- **Upskill** – acquire *programming*, *data analytics*, *writing skills*
- Establish **collaborations**
- **Committed** team members

University

- **Good leadership** (visionary, exemplary, etc)
- **Strategic** research planning & implementation
- **Recognition**
- **Rewards**
- Provide **funding (especially young lecturers)**

Data Source



Covid-19 data

Environmental data

Credit card charges

Students database

Garuda Indonesia

malaysia AIRLINES

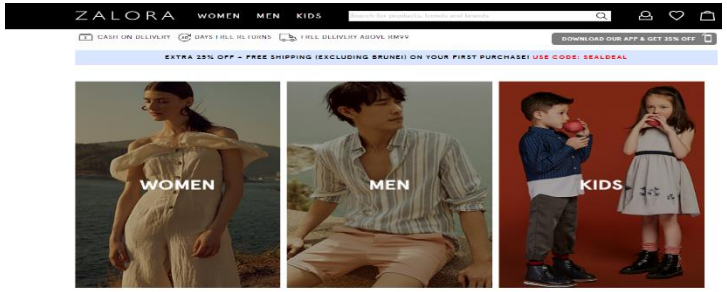
Airline reservations

traffic data

facebook

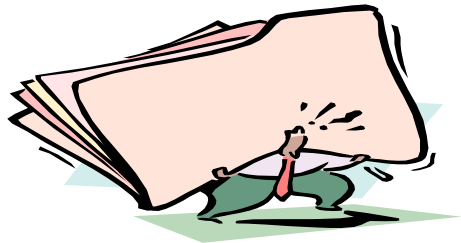
Telco customer data

amazon Try Prime



Lazada

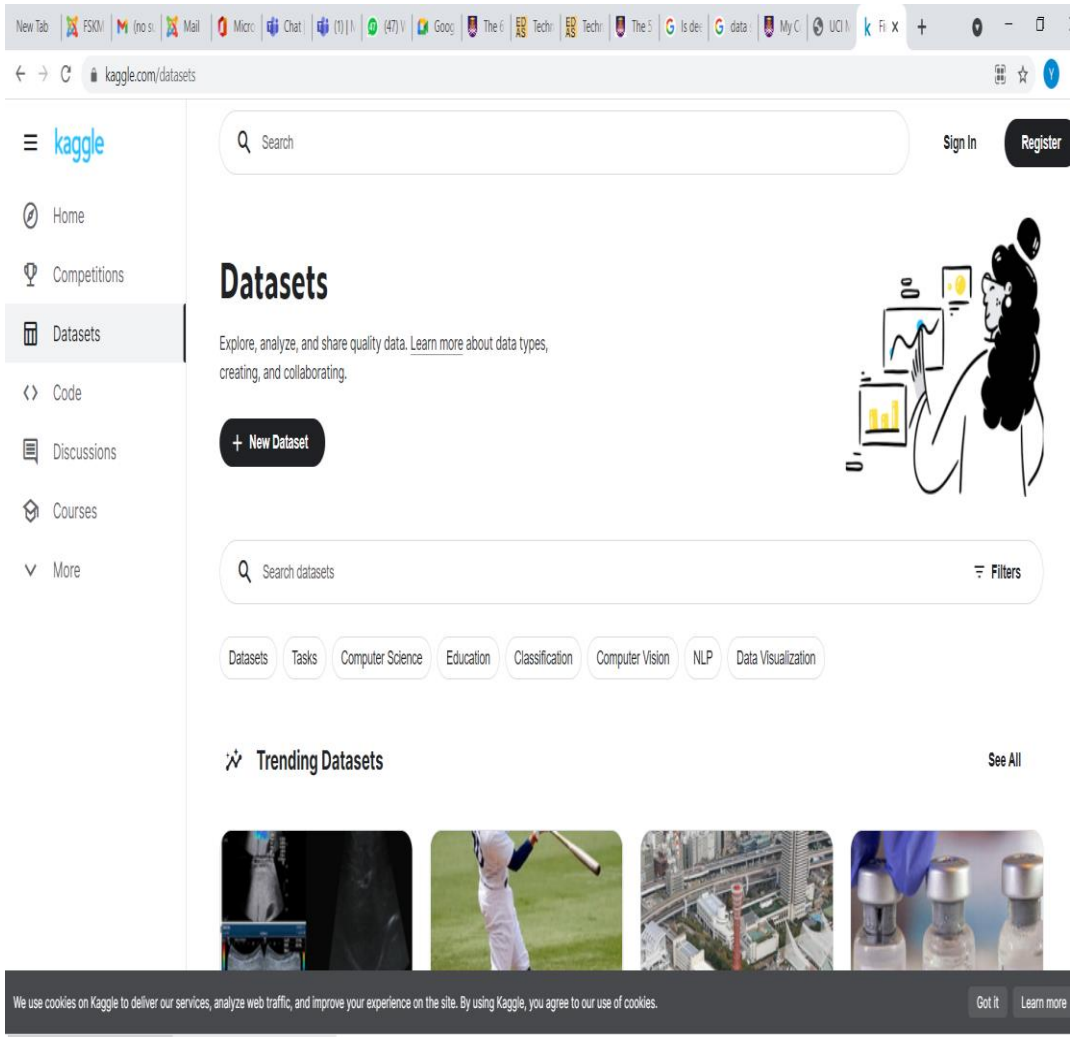
Online purchase data



Tax returns

Online public data for Research/Projects

https://archive.ics.uci.edu/ml/index.php



https://www.kaggle.com/datasets

Simulate your own data using R to test your method

Generate Simulated Dataset for Linear Model in R

When the real dataset is hard to find, simulate it.

Raden Aurelius Andhika Viadinugroho Jun 18 · 6 min read

Motivation

In these recent years, research about Machine Learning (ML) has increased along with the increased computation capability. As a result, there is much development in some of the ML models — if not inventing a new model — that performs better than the traditional model.

One of the main problems that the researchers usually encountered when trying to implement the proposed model is the lack of the proper real-world dataset that follows the model's assumptions. Or in the other case, the real-world dataset exists, but the dataset itself is very expensive and hard to collect.

```r
library(car)
library(MASS)
library(lmtest)
library(tseries)
library(ggfortify)

set.seed(1234)

##Linear Regression
#Generate the independent variable and the error
x1=rnorm(100,50,9)
x2=rnorm(100,200,64)
error=rnorm(100,0,16)
#Generate the dependent variable (b0=150, b1=-4, b2=2.5)
y1=150-(4*x1)+(2.5*x2)+error
#create the model
m1=lm(y1~x1+x2)
summary(m1)
autoplot(m1)
```

https://towardsdatascience.com/generate-simulated-dataset-for-linear-model-in-r-469a5e2f4c2e

Stage 1 Business Understanding

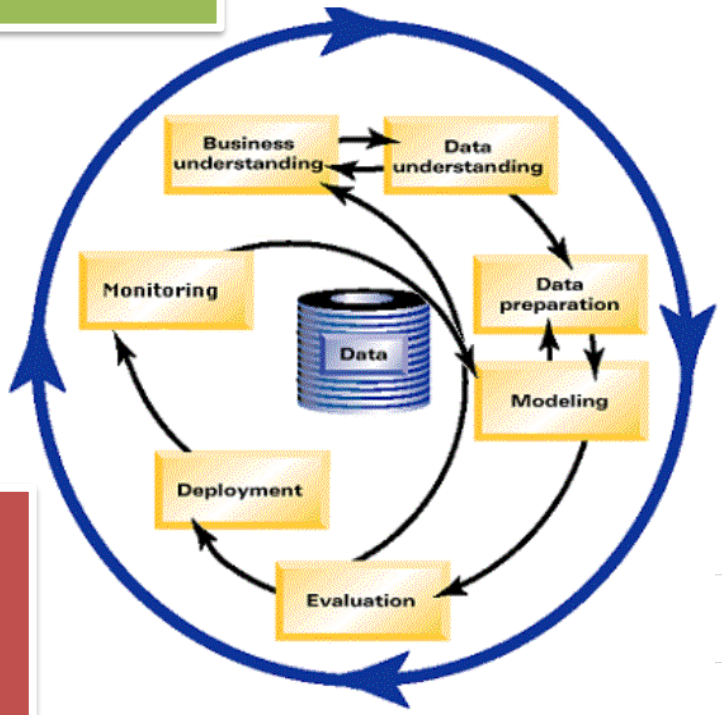
- Perform a situation assessment.
- Identify business (research problem) objective(s).
- Determine data mining goals.
- Produce a project plan.

Stage 2 Data Understanding

- Identify types of data.
- Identify target variable(for predictive analytics).
- Data Audit (data errors, missing values, outliers, imbalance data)
- Data Exploration (charts)

Stage 6 Deployment

Classify **new cases** using model selected in Stage 5.



CRISP-DM Process

Stage 3 : Data Preparation

- Variable selection (feature selection)
- Data Imputation for missing values
- Data transformation
- Creating new variable(s)

Stage 5 Evaluation

Performance measures:

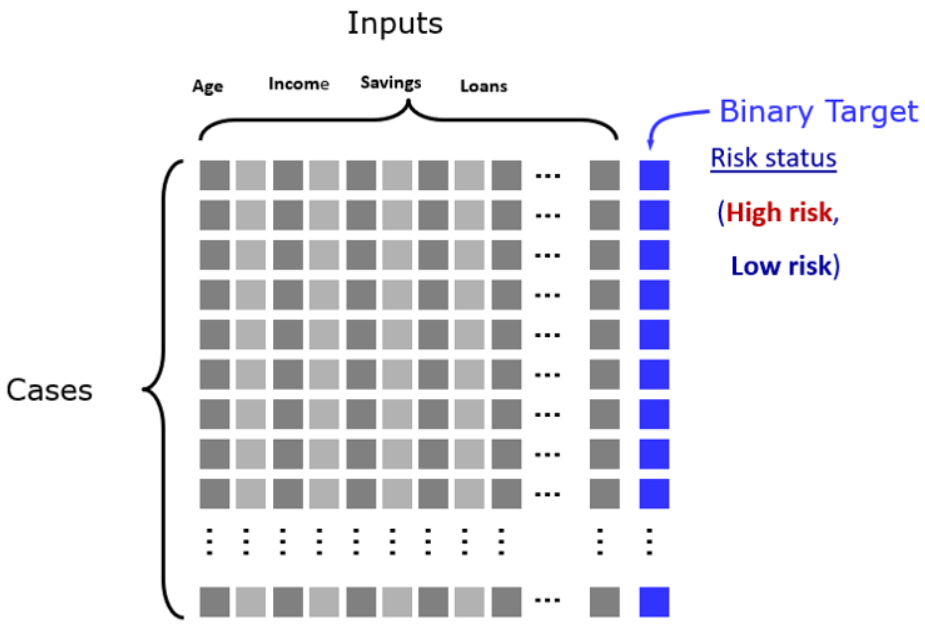
(for binary target)

- Accuracy
- Sensitivity
- Specificity
- Precision

Stage 4 : Data Modeling

Data	Outcome (or Target) Variable (Y)	Predictive models	Predictor Variables(X)
Cross-sectional data	Categorical	**Logistic and Multinomial Regression**, Decision Tree, ANN, SVM, Naïve Bayes, Random Forest, k-NN	Continuous or categorical
Cross-sectional data	Continuous	**Linear Regression**, Regression Tree, k-NN, ANN, SVR	Continuous or categorical
Time series data	Continuous	**Multivariate Time Series Models**	Continuous or categorical data

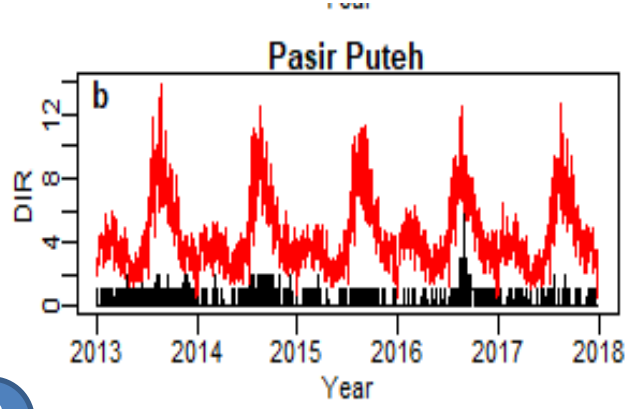
Types of data



1 **Cross-sectional data**

Year	State	Youth	gdp	unemployment	crime
2007	johor	625.2	18753	2	5390
2008	johor	629.1	20162	2.2	5965
2009	johor	629.8	18878	3.2	8374
2010	johor	628.4	22035	2.4	5229
2011	johor	632.1	24350	2.5	4428
2012	johor	635.7	25442	3.1	3842
2013	johor	634.9	26308	2.8	3239
2014	johor	653.4	28089	2.6	2933
2015	johor	662	29558	3.1	2413
2016	johor	671.7	31952	3.6	2820
2007	kedah	349.7	12160	3.3	1661
2008	kedah	354.8	13023	3.8	1742
2009	kedah	360.8	12481	4.2	1859
2010	kedah	367.7	14034	2.8	1714
2011	kedah	378.6	15563	3.5	1635

3 **Panel data**



2 **Time series data**

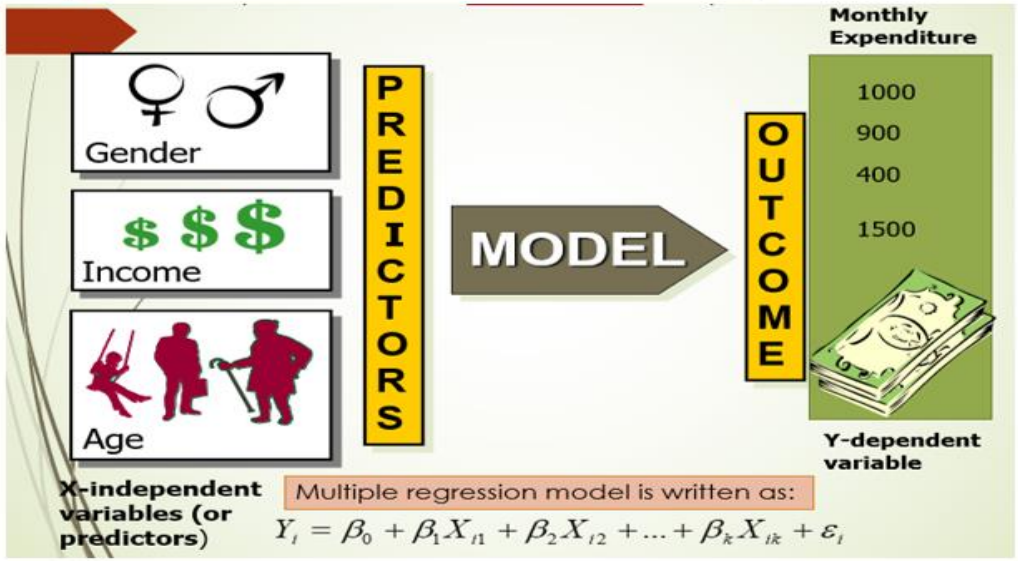
4 **Experimental data**
Treatment and Placebo group

	center_id	center_size	attphys_id	patient_id	gender	dob	treatment	week	convulsions
1	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Pre-treatm...	2
2	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Week 1	6
3	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Week 2	4
4	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Week 3	4
5	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Week 4	6
6	07057	Small	COMX	1FSL	Male	05/26/1990	Anticonvulsant	Week 5	3
7	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Pre-treatm...	4
8	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Week 1	7
9	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Week 2	5
10	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Week 3	7
11	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Week 4	6
12	07057	Small	COMX	30QU	Female	06/07/1977	Placebo	Week 5	6
13	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Pre-treatm...	5
14	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Week 1	8
15	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Week 2	5
16	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Week 3	4
17	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Week 4	5
18	07057	Small	COMX	3974	Male	03/02/1972	Anticonvulsant	Week 5	3
19	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Pre-treatm...	2
20	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Week 1	1
21	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Week 2	1
22	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Week 3	2
23	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Week 4	2
24	07057	Small	COMX	4EVE	Female	01/18/1964	Anticonvulsant	Week 5	1
25	07057	Small	COMX	98PC	Female	10/15/1986	Placebo	Pre-treatm...	1
26	07057	Small	COMX	98PC	Female	10/15/1986	Placebo	Week 1	3

Cross-sectional data

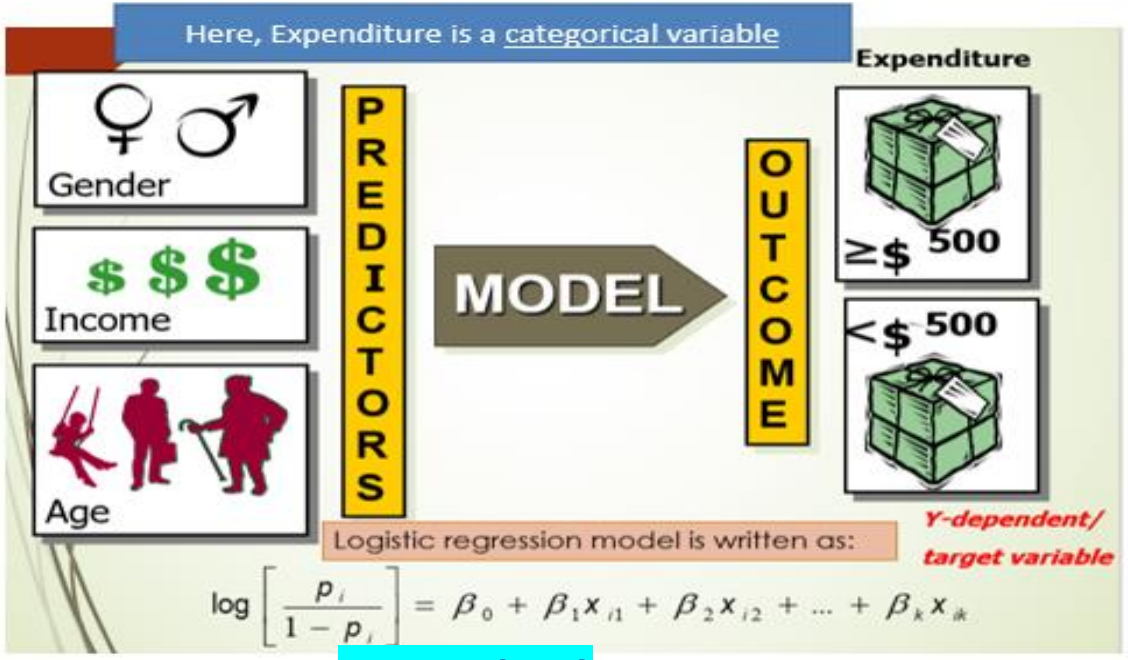
Selection of model depends on your **target (dependent variable)**

Here, Expenditure is a continuous variable



Y is continuous

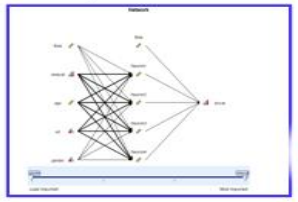
Here, Expenditure is a categorical variable

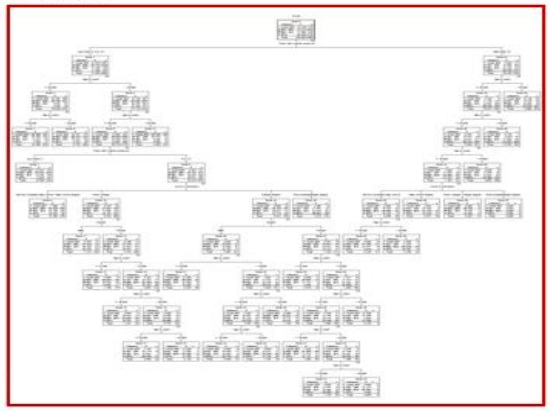


Y is binary (0,1)

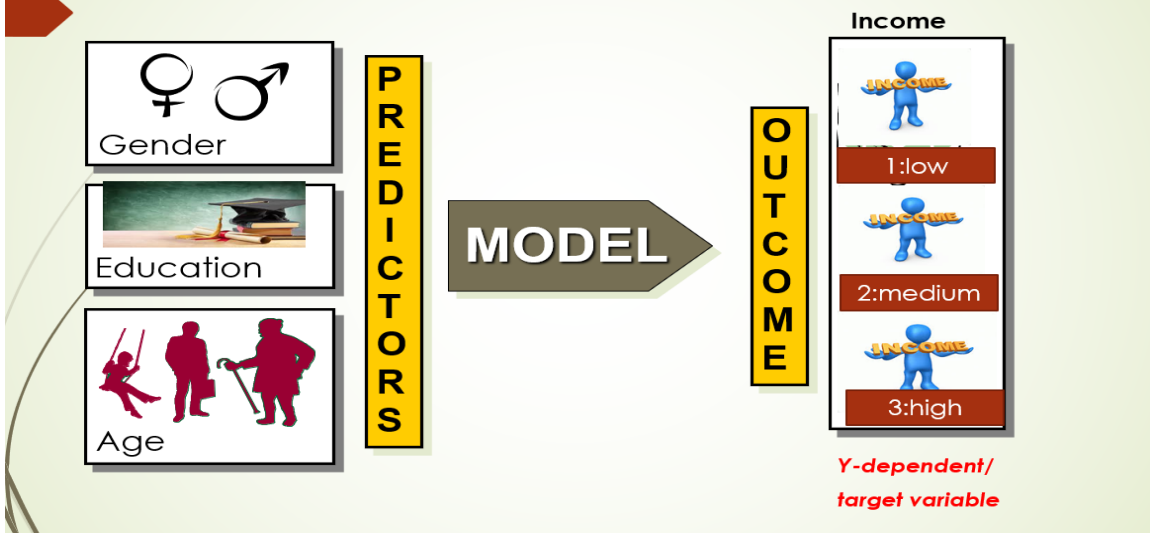
Multi-class classification model

- Multinomial Logistic Regression
- Ordinal Logistic Regression
- Decision Trees
- ANN
- SVM
- Bayes' Network





Multi-class classification problem



Y is multi-class (1,2,3)

Predictive Analytics for Cross-sectional data (Supervised Learning)

Machine Learning classifier
(for categorical target, Y)

- ✓ Decision tree
- ✓ Random forest
- ✓ Logistic Regression
- ✓ Artificial Neural Network
- ✓ Support Vector Machine
- ✓ k-NN
- ✓ Bayes Network

(b) Machine Learning Classifier
(for continuous target, Y)

- ✓ Regression tree
- ✓ Random forest
- ✓ Linear Regression
- ✓ Artificial Neural Network
- ✓ Support Vector Regression
- ✓ k-NN

Cross-sectional data

Y is continuous

Common method

Multiple Regression

Alternative model

Lasso*

Adaptive Lasso*

Elastic net *

Ridge regression*

Regression Tree

Random Forest

K-NN

Support Vector Regression (SVR)

Table 1 Popular Penalized Regression Methods

Method	Penalty
LASSO	$\sum_{j=1}^{p} \mid \beta_j \mid < t$
Adaptive LASSO	$\sum_{j=1}^{p} \left(\mid \beta_j \mid / \mid \hat{\beta}_j \mid \right) < t$
Elastic net	$\sum_{j=1}^{p} \mid \beta_j \mid < t_1$ and $\sum_{j=1}^{p} \beta_j^2 < t_2$

*Introduced penalty to prevent overfitting and improve model performance

Y is binary (0,1)

Common method

Logistic Regression

Alternative model

Lasso*

Adaptive Lasso*

Elastic net*

Ridge*

Decision Tree

Random Forest

K-NN

Support Vector Machine (SVM)

1

LOGISTIC REGRESSION MODEL

Example: Prediction of Churn: Y=1 (Churn) or 0(Active)

$$P(event) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_kX_k}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_kX_k}} \text{ where } \mathrm{P(event)} = \mathrm{Prob(Y=1)}$$

The binary logistic model is:

$$\ln\left(\frac{\text{Prob(event)}}{1-\text{Prob(event)}}\right) = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_kX_k$$

	Actual				Predicted		Predicted
	Y	Age	IncomeCat	Usage	P(Y=1)	P(Y=0)	Class
Ali	1	35	1	300	0.85	0.15	1
James	1	45	2	200	0.25	0.75	0
Siti	0	30	1	150	0.65	0.4	1
Mary	0	25	3	350	0.9	0.1	1
Salmah	0	28	1	100	0.14	0.86	0

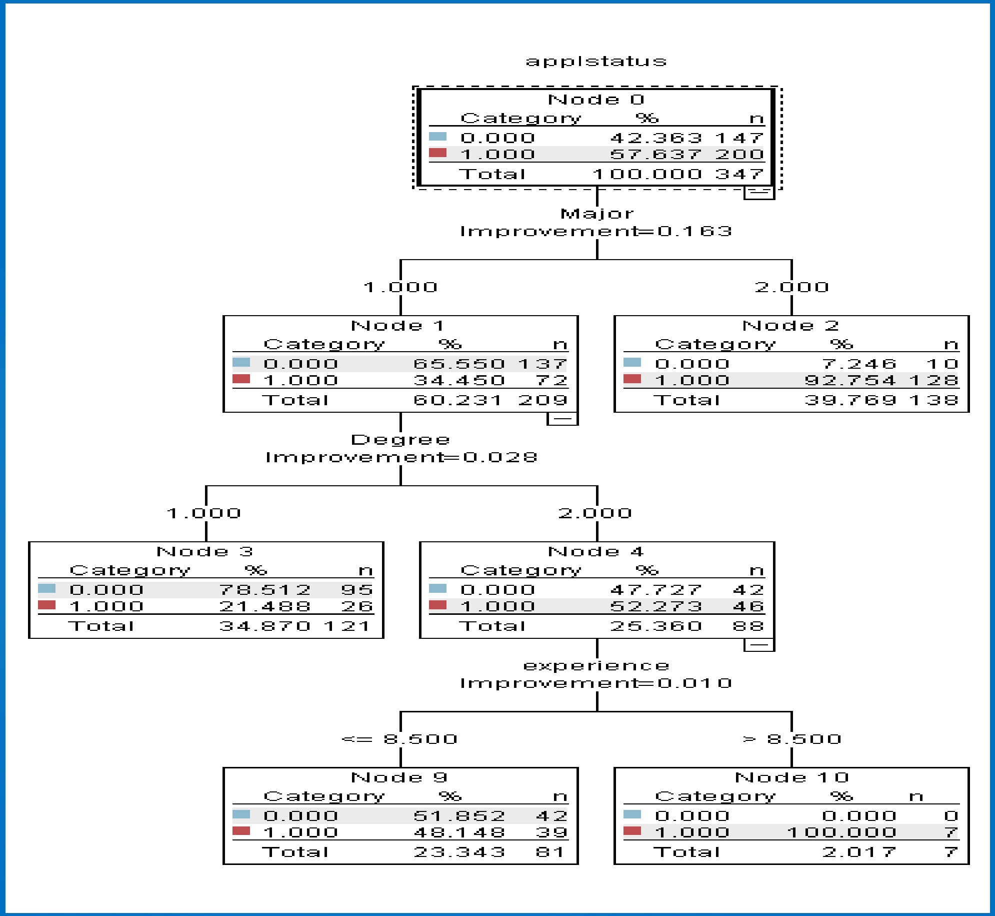
Confusion matrix

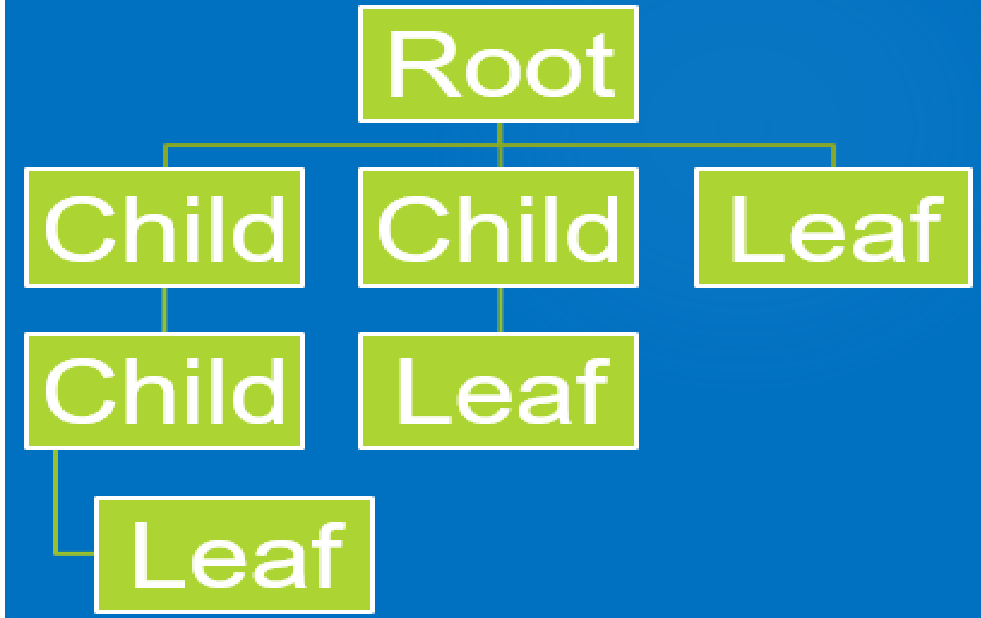
	Predicted	Y	
Actual	1	0	Total
1	1	1	2
0	2	1	3
Total		2	5

Acc	40
Sen	50
Spec	33.33
Precision	50

2

Decision Tree Structure





Decision trees algorithms

CART

- construct decision tree for categorical & continuous target variable
- Uses Gini measure for classification tree.
- Performs only binary split

C5

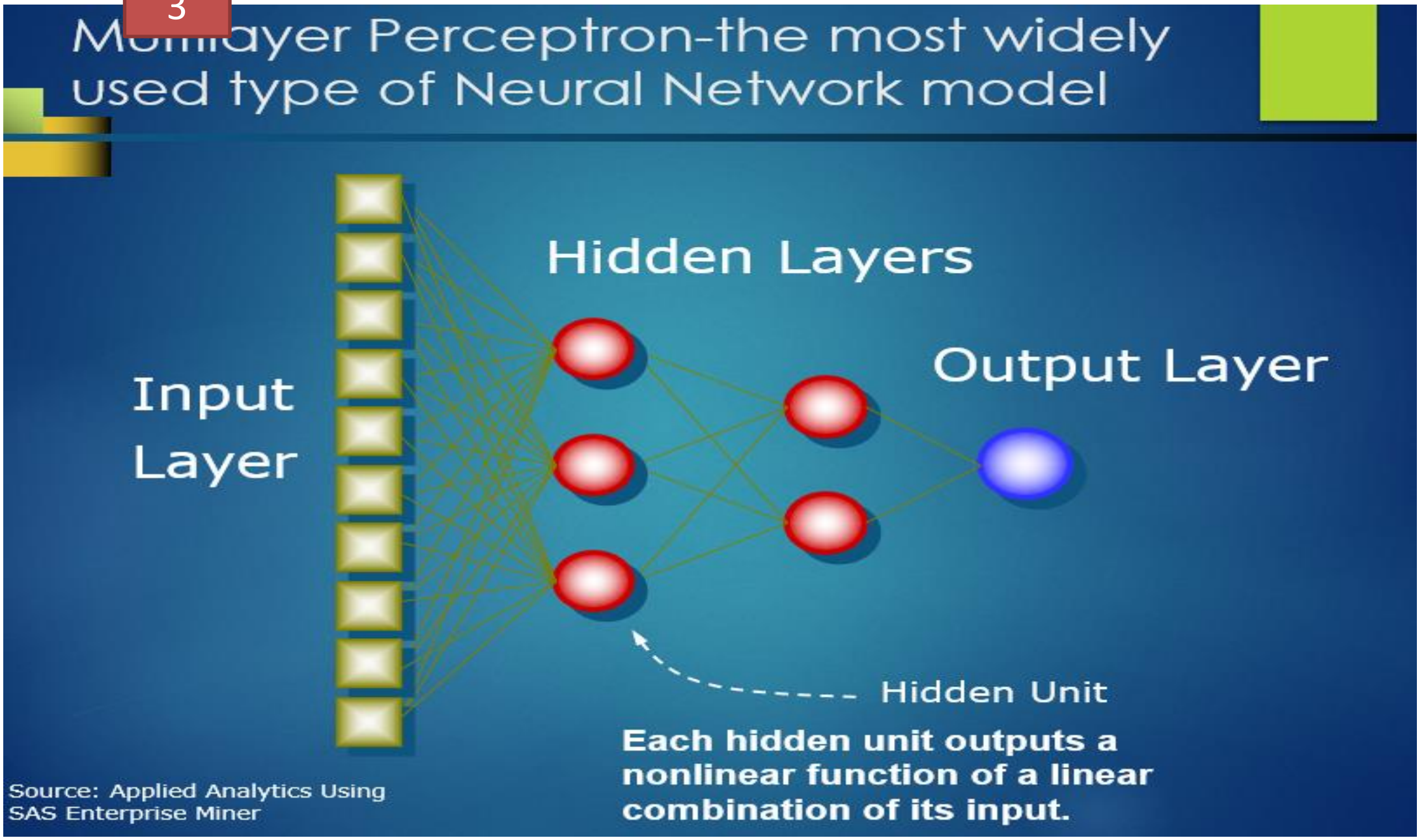
- construct decision tree for categorical target variable.
- Uses entropy measure for splitting nodes .
- Performs multi-way split

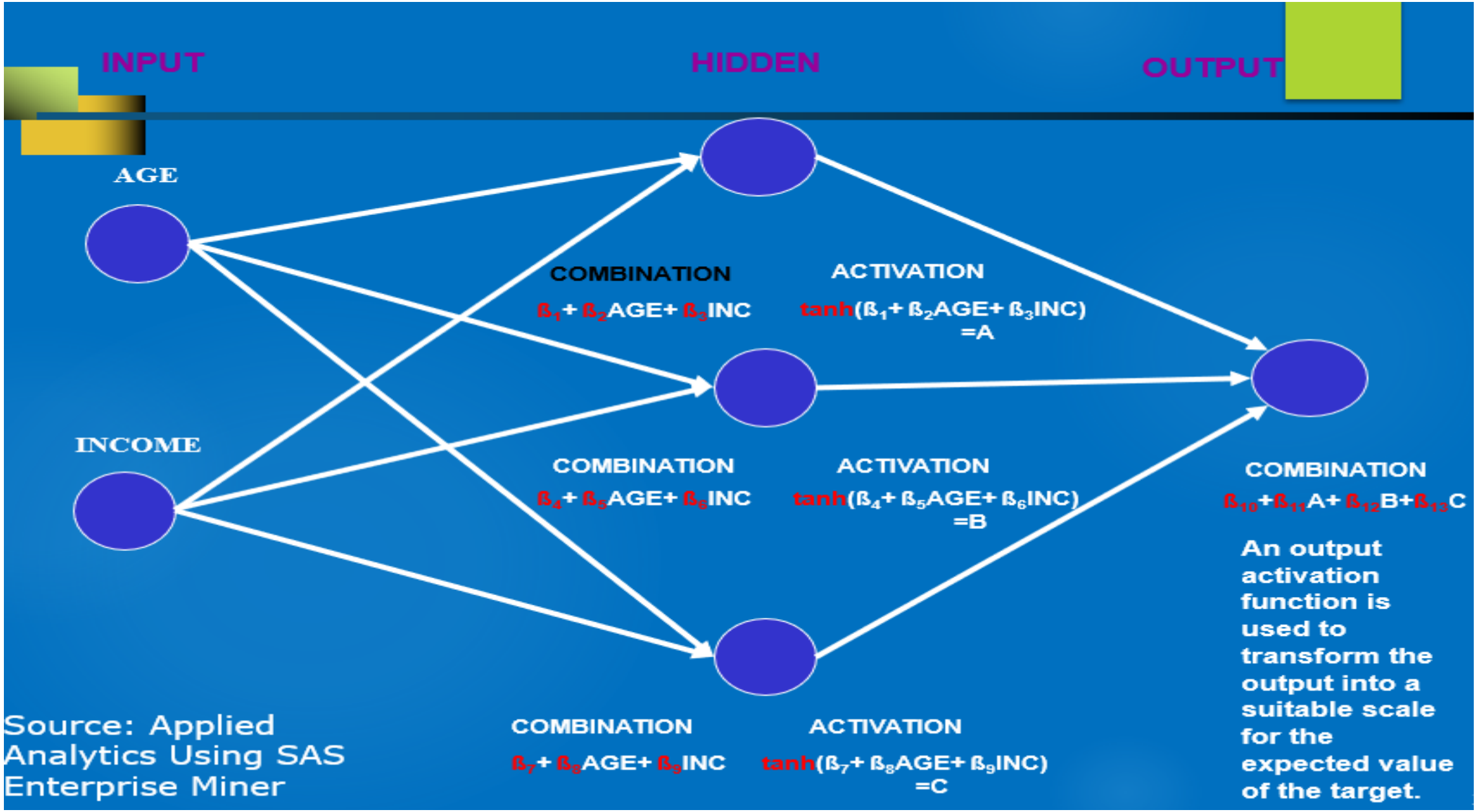
CHAID

- construct decision tree for categorical and continuous target variable.
- Split algorithm (Chi-Square test) designed for categorical inputs so continuous inputs must be discretized.
- Performs multi-way split

3

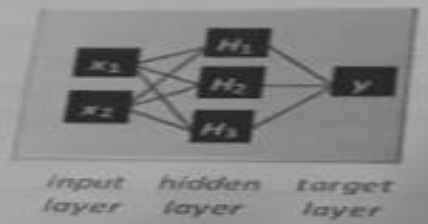
Multilayer Perceptron-the most widely used type of Neural Network model



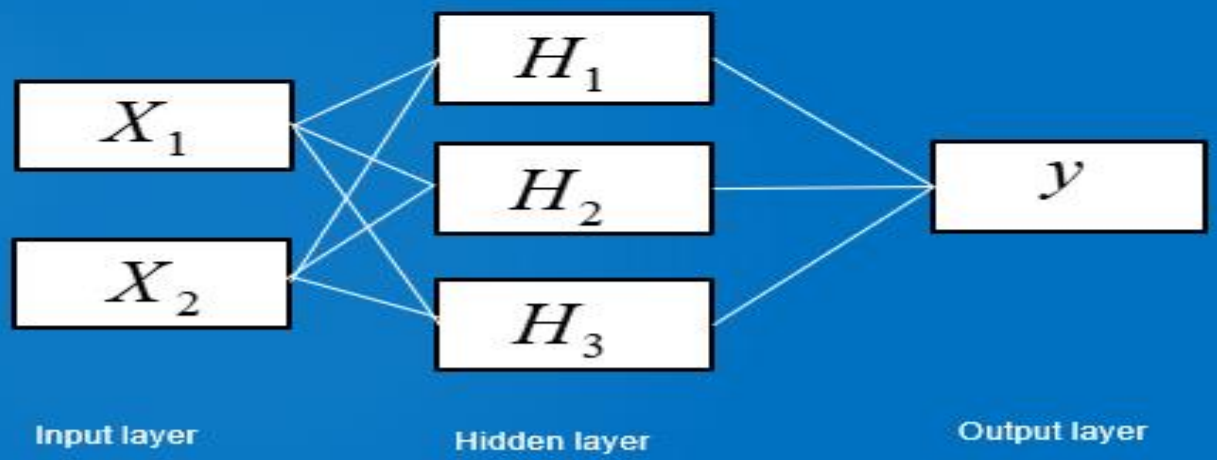

INPUT
HIDDEN
OUTPUT
AGE
INCOME
COMBINATION
ß1+ ß2AGE+ ß3INC
ACTIVATION
tanh(ß1+ ß2AGE+ ß3INC)
=A
COMBINATION
ß4+ ß5AGE+ ß6INC
ACTIVATION
tanh(ß4+ ß5AGE+ ß6INC)
=B
COMBINATION
ß7+ ß8AGE+ ß9INC
ACTIVATION
tanh(ß7+ ß8AGE+ ß9INC)
=C
COMBINATION
ß10+ß11A+ ß12B+ß13C
An output activation function is used to transform the output into a suitable scale for the expected value of the target.
Source: Applied Analytics Using SAS Enterprise Miner

Neural Network Diagram

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$
$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

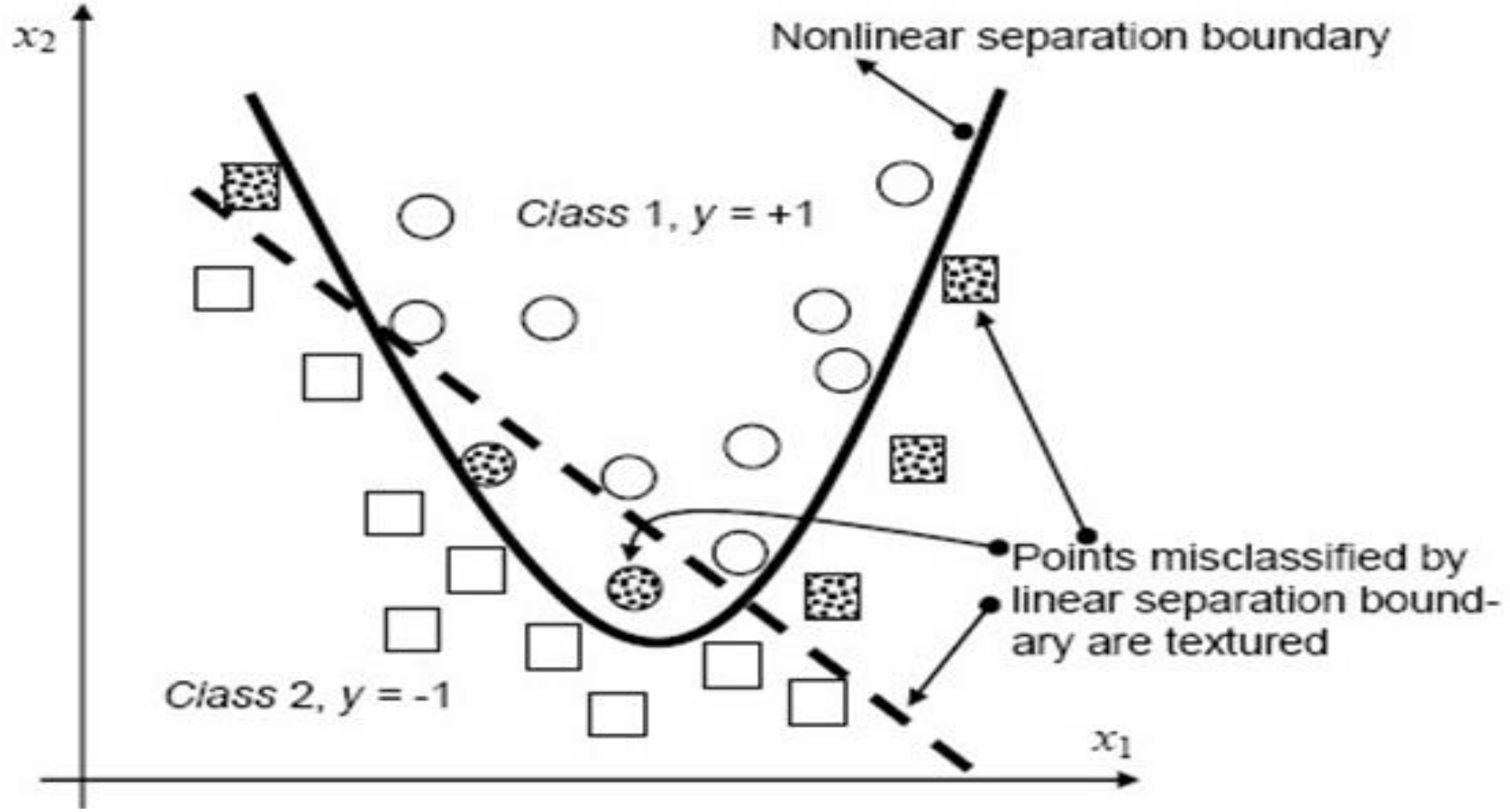


Source: Applied Analytics Using SAS Enterprise Miner

4

Support Vector Machines

Data mining algorithms that can perform linear or non-linear classification



SVM(Support Vector Machine)

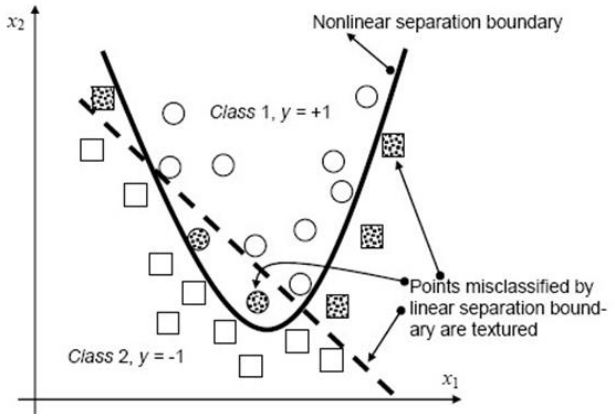
The mathematical function used for the transformation is known as the **kernel** function. SVM in IBM® SPSS® Modeler supports the following kernel types:

- Linear
- Polynomial
- Radial basis functi
- Sigmoid

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^p$$

$$K(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|^2/(2\sigma^2)}$$

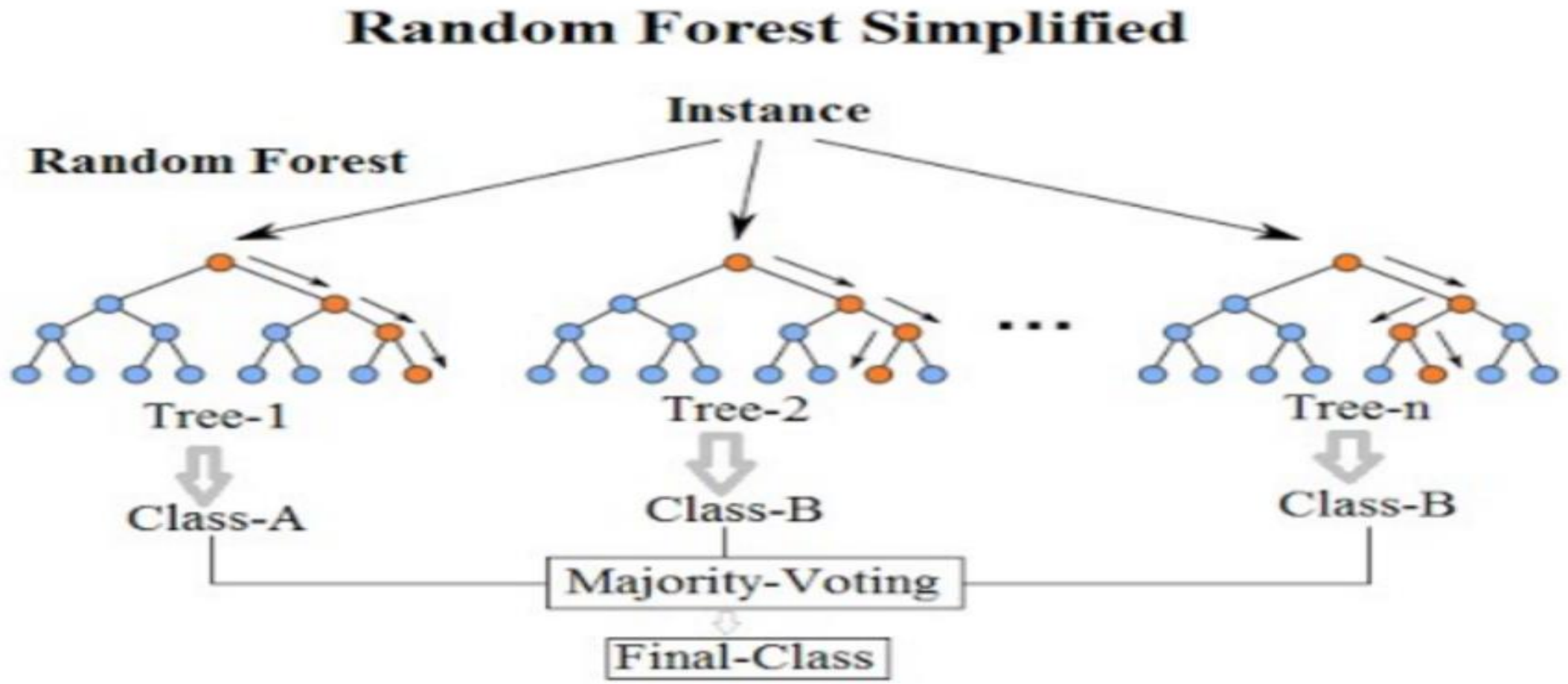
$$K(\mathbf{x}, \mathbf{y}) = \tanh(k\mathbf{x} \cdot \mathbf{y} - \delta)$$



A linear kernel function is recommended when linear separation of the data is straightforward. In other cases, one of the other functions should be used. You will need to experiment with the different functions to obtain the best model in each case, as they each use different algorithms and parameters.

5

Random Forest



Random Forest node (ibm.com)

1.11. Ensemble methods — scikit-learn 0.24.1 documentation (scikit-learn.org)

Naïve Bayes Classifier

- A probabilistic framework for solving classification problems
- Conditional Probability:

$$P(Y \mid X) = \frac{P(X,Y)}{P(X)}$$

condition

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$$

- Bayes theorem:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

Proof:

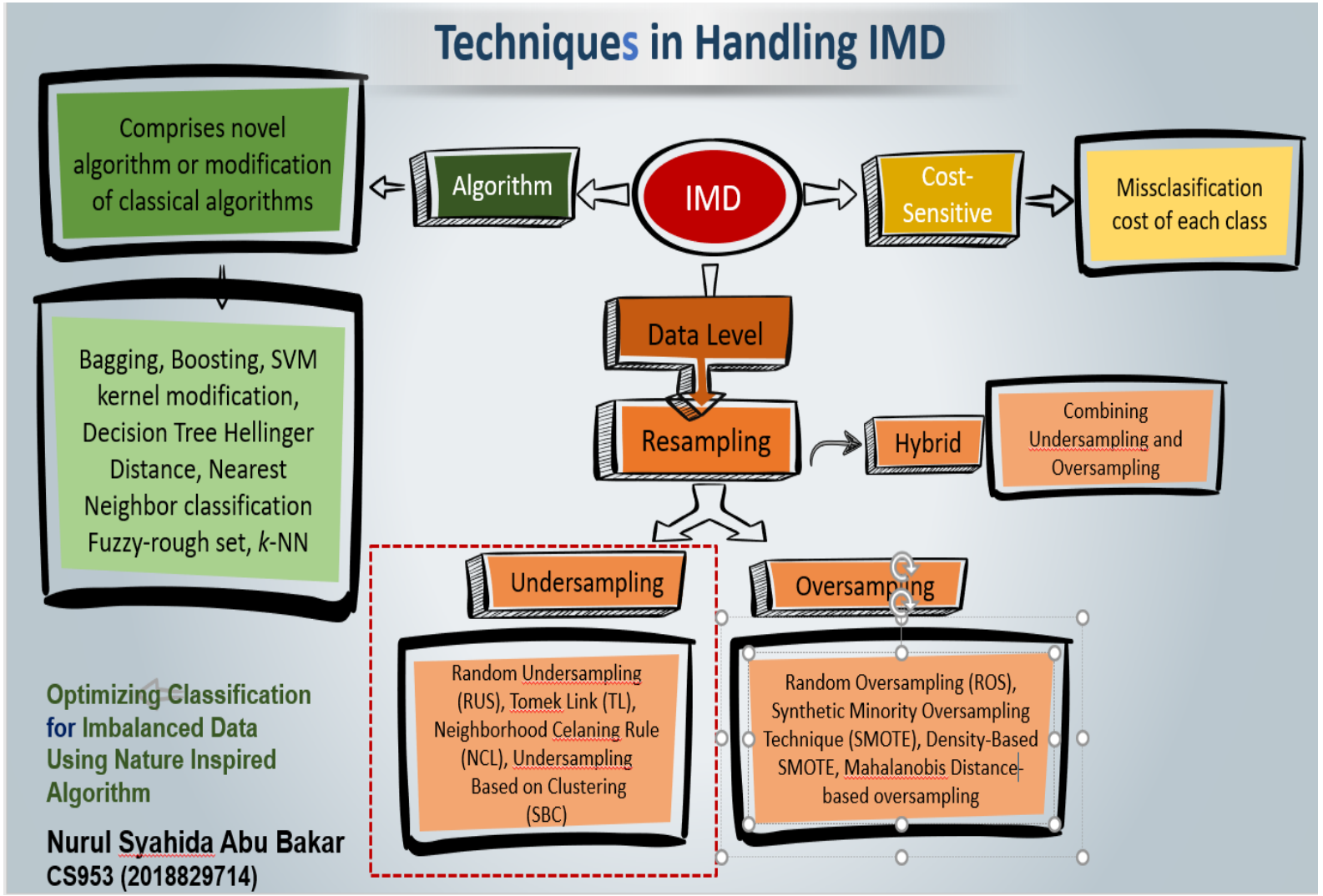
$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

$$= \frac{\frac{P(X \cap Y)}{P(Y)} P(Y)}{P(X)}$$

$$= \frac{P(X \cap Y)}{P(X)}$$

Research Ideas

Imbalanced data



Journal of Hydrology 598 (2021) 126382

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

XGBoost-based method for flash flood risk assessment

Meihong Ma[a], Gang Zhao[b,*], Bingshun He[c], Qing Li[c], Haoyue Dong[a], Shenggang Wang[d], Zhongliang Wang[a,e,*]

[a] School of Geographic and Environmental Sciences, Tianjin Normal University, Tianjin 300387, China
[b] School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK
[c] China Institute of Water Resources and Hydropower Research, Beijing 100038, China
[d] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[e] Tianjin Key Laboratory of Aquatic Science and Technology, Tianjin Chengjian University, China

ARTICLE INFO

This manuscript was handled by, Editor-in-Chief

Keywords:
Flash flood
Risk assessment
XGBoost
Yunnan

ABSTRACT

Flash flood risk assessment, a widely applied technology in preventing catastrophic flash flood disasters, has become the current research hotspot. However, most existing machine learning methods for assessing flash flood risk rely on a single classifier, which is suitable for processing small sets of sample data, but the resulting prediction accuracy and generalization ability are insufficient. Meanwhile, machine learning methods that integrate multiple classifiers are thus far unknown. Extreme Gradient Boosting (XGBoost) is an excellent algorithm for ensemble learning methods which has achieved remarkable results in many fields. It not only optimizes the algorithm but also automatically applies the CPU's multi-threading to perform parallel calculations, thus greatly improving the model training speed and prediction accuracy. Therefore, this article introduces the XGBoost model for the assessment of flash flood risk, and then combines the two input strategies and the Least squares support vector machine (LSSVM) model to verify its optimal effect, thus proposing the XGBoost-based method for flash flood risk assessment. Subsequently, an attribution analysis was implemented to assess the possible errors of this approach; and finally, a county-level flash flood risk map for Yunnan Province, China, was generated based on the proposed method. The results demonstrate that: (1) XGBoost performs well, with an accuracy of 0.84 in the testing period, and its five indices (precision, recall, accuracy, kappa, and F-score) are all higher than those of LSSVM. (2) The XGBoost-based approach provided the reliable flash flood risk maps, which were validated by another flash flood inventory, although some errors may be attributed to critical environmental factors and statistical disaster location accuracy. (3) The high-risk counties (including high-risk and highest-risk) accounted

EFFICIENT **MACHINE LEARNING DATA IMPUTATION** METHOD FOR HEALTHCARE PREDICTIVE ANALYTICS

NURUL AZIFAH BINTI MOHD PAUZI
2019656214
(PhD Statistics)

Hybrid Machine Learning Method

Aydilek & Arslan (2012)

A novel hybrid approach to estimating missing values in databases using **K-nearest neighbors** and **neural networks**

Methods

- Neural Networks with K-Nearest Neighbors (NN-KNN)
- Neural Networks with Genetic Algorithm (NN-GA)

Findings

Hybrid NN-KNN provides better imputation accuracy compared to hybrid NN-GA

Al-Milli & Almobaideen (2019)

Hybrid Neural Network to Impute Missing Data for IoT Applications

Methods

Neural Network with **Genetic (NN-GA)** Algorithm

Findings

NN-GA is able to impute missing data with high classification accuracy compared to the results without data imputation

Sanjar et al. (2020)

Missing data imputation for geolocation-based price prediction using **KNN-MCF** method

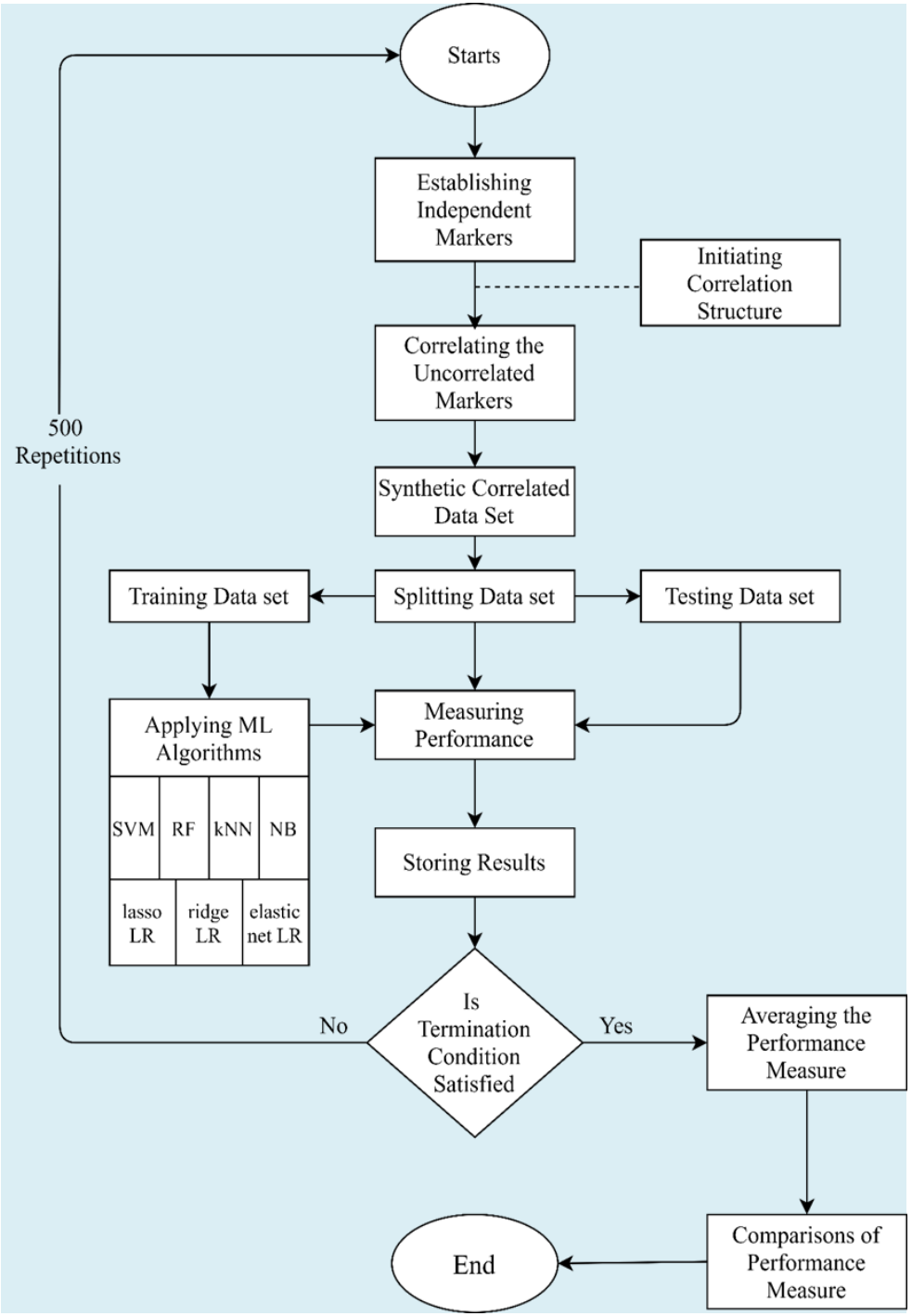
Methods

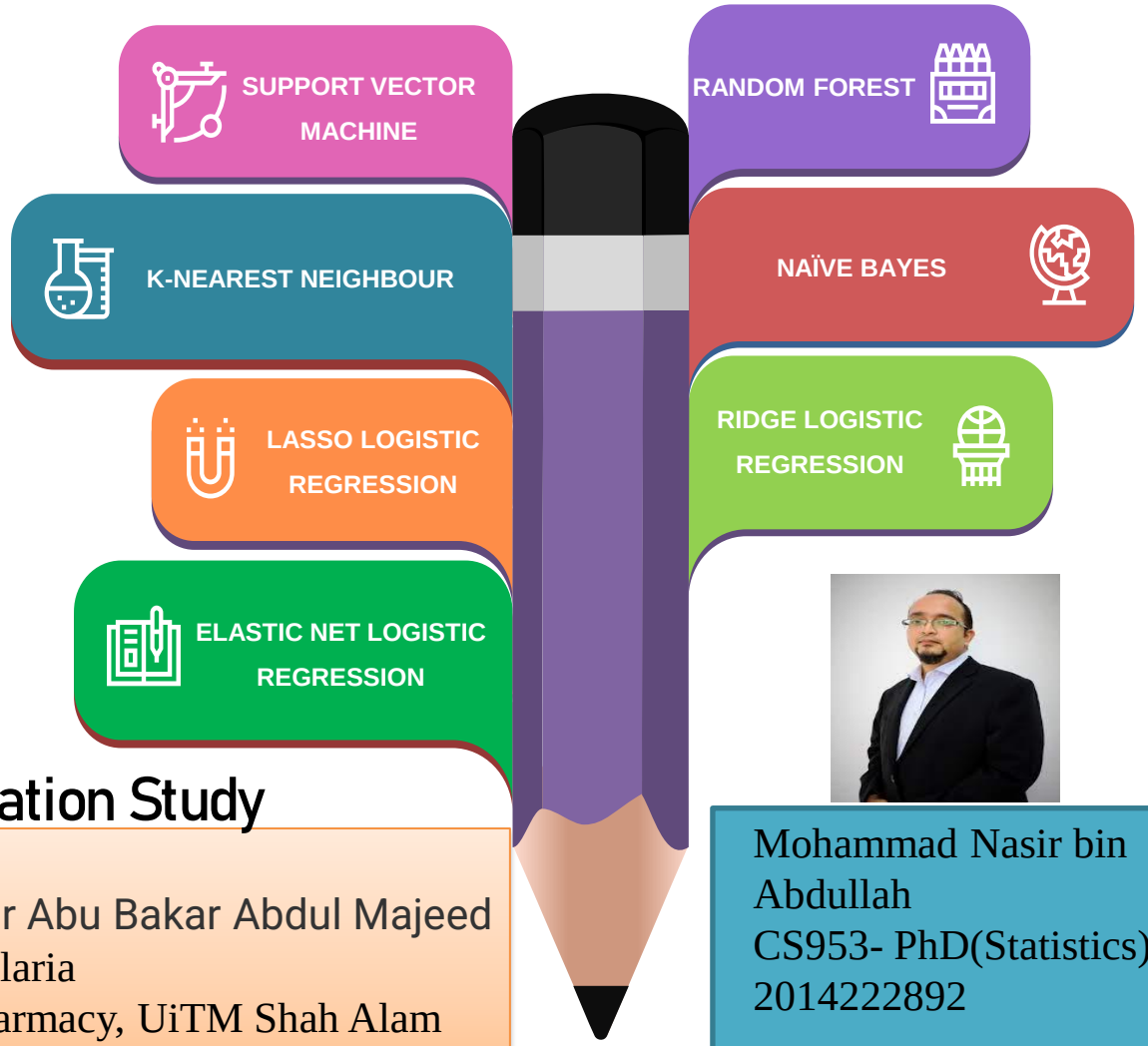
- **KNN-MCF**
- **Mean Imputation**
- **KNN Imputation**

Findings

KNN-MCF performed better compared to the mean and KNN Imputation

MACHINE LEARNING AND PENALIZED REGRESSION MODELS FOR HIGH DIMENSIONAL DATA ANALYSIS ON MULTI OMICS BLOOD-BASED BIOMARKERS FOR ALZHEIMER'S DISEASE





Simulation Study

Co-SV:
Prof. Dato' Dr Abu Bakar Abdul Majeed
Dr Yuslina Zalaria
Faculty of Pharmacy, UiTM Shah Alam

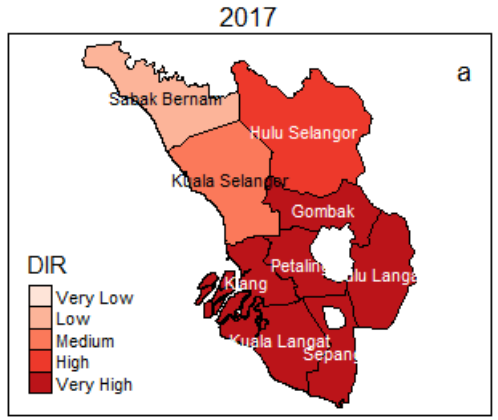
Mohammad Nasir bin Abdullah
CS953- PhD(Statistics)
2014222892

Spatial Modelling

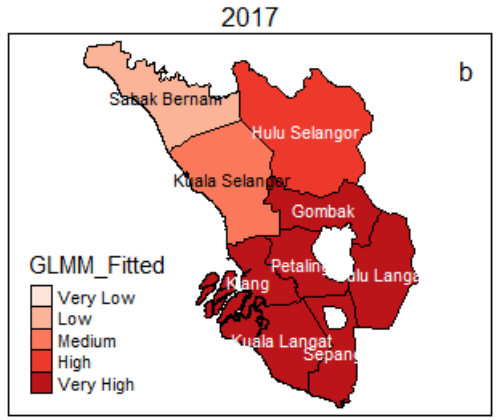
- ***Bayesian Hierarchical Modeling*** with Marcov Chain Monte Carlo (MCMC)
- Application in dengue disease - to predict dengue cases and examine influential covariates associated with the risk of dengue outbreak.
- How can ***Spatio-temporal Bayesian model of Generalized Linear Mixed Model*** improve prediction? - Over space and time the introduction of ***spatial random effects with Conditional Autoregressive Structure (CAR) of Bayesian framework*** into the linear predictor allows the variability of the heterogeneity factors to be captured in the previous trend of the dengue counts to be derived in the posterior predictions. – i.e: In dengue disease, develop a disease spatial map, accurate prediction and identify if any association between dengue prevalence, temperature, rainfall and humidity at finer scale.
- Statistical packages for implementing such Bayesian models using MCMC include ***WindBugs, CrimeStat*** and many packages available via ***R programming language.***

Nik Nur Fatin
CS953- PhD(Statistics)
2016334351

Dr Wan Fairos Wan Yaacob
UiITM Cawangan Kelantan



Actual



Spatial-temporal Bayesian GLMM

Struktur Organisasi

Pejabat Timbalan Naib Canselor (Penyelidikan & Inovasi)

MENJALANKAN FUNGSI NAIB CANSELOR

TIMBALAN NAIB CANSELOR (PENYELIDIKAN & INOVASI)

- **PENGARAH RMC**
  - OPERASI PENYELIDIKAN
  - PEROLEHAN GERAN
  - PEMANTAUAN PENYELIDIKAN
  - ETIKA & PENERBITAN
  - DATA PENYELIDIKAN
  - INOVASI
- **PENGARAH ReNeU**
  - *INDUSTRIAL TECHNOLOGY*
  - *CYBER TECHNOLOGY*
  - *LOGISTICS & TRANSPORTATION*
  - *HEALTH & WELLNESS*
  - *ENERGY & ENVIRONMENT*
  - *SOCIAL CREATIVITY & INNOVATION*
- **PENGARAH BITCOM**
  - PENGKOMER-SIALAN DAN PEMBANGUNAN PERNIAGAAN
  - PERUNDANGAN KOMERSIAL DAN PENGURUSAN HARTA INTELEK
  - PERUNDINGAN DAN PERKHIDMATAN INOVATIF
- **PENGARAH ARI**
  - PENTADBIRAN / LATIHAN & PERUNDINGAN
  - PENYELIDIKAN & JARINGAN
  - PASCASISWAZAH & INOVASI
  - PUSAT PENYELIDIKAN
- **PENGARAH CoE (13)**
  - IIESM
  - IoS
  - AuRIns
  - I-PPerForM
  - iPROMISE
  - IBE
  - MRI
  - MITRANS
  - SMRI
  - SRI
  - IBSD
  - IBDAAI
  - CaVaLRI
- **PENGARAH PENERBIT UiTM**
  - PENTADBIRAN
  - EDITORIAL
  - PEMASARAN
  - JURNAL
- **KETUA BAHAGIAN PENTADBIRAN & SUMBER MANUSIA**
  - PENGURUSAN SUMBER
  - PEMBANGUNAN SUMBER MANUSIA
  - TADBIR URUS
  - PENGURUSAN KEWANGAN
  - PENTADBIRAN AM, SEKRETARIAT, PENGURUSAN KENDERAAN
  - DISIPLIN & TATATERTIB
  - PENGURUSAN KUALITI & STRATEGIK
- **KETUA UNIT MAKLUMAT STRATEGIK**
  - PENGURUSAN DATA STRATEGIK
  - PEMBANGUNAN DATA STRATEGIK
- **KETUA UNIT KOMUNIKASI PENYELIDIKAN & VISIBILITI**
  - PERANCANGAN VISIBILITI
  - PENERBITAN MEDIA
- **KETUA UNIT ICT**
  - SISTEM OPERASI
  - SISTEM PELAPORAN
  - APLIKASI VISIBILITI

UiTM di hatiku

University Research &Innovation Ecosystem

Research Management & Monitoring

Research Excellence

IP & Commercialization

Consultancy

ReNeU

FUNCTION

- Serves as a platform for researchers and lead in bridging university research towards excellence.
- Increase visibility & impact of research through strategic and sustainable networking and collaboration (both local and global).
- Escalate the competitiveness in grant application (International, National & Private Grants).
- Empowering research output specifically indexed journal publications as well as output that can be translated into government policy, industrial applications and for community well-being.

Vision

To be the leading Technology Transfer Centre in Malaysia that creates impact to society through successful commercialization of UiTM innovations

Mission

To accelerate the translation of important discoveries arising from UiTM research and innovation activities into business opportunities for the benefit of the university, the country and the global community

Funding Opportunities

Pejabat
Timbalan Naib Canselor
(Penyelidikan dan Inovasi)

GERAN NASIONAL

(ReNeU & RMC)

Professor Dato' Dr Abu Bakar Abdul Majeed
Pengarah RMC, UiTM

Professor Dr Nooritawati Md Tahir
Pengarah ReNeu, UiTM

GERAN KEMENTERIAN PENGAJIAN TINGGI (KPT)

- Geran Penyelidikan Fundamental (FRGS –RACER)
- Geran Penyelidikan Fundamental (FRGS)
- Dana Pembudayaan Penyelidikan (RAGS)
- Geran Penyelidikan Pembangunan Prototaip (PRGS)
- Geran Penyelidikan Transdisiplinari (TRGS)
- Geran Penyelidikan Jangka Panjang (LRGS)

GERAN KEMENTERIAN PENGAJIAN TINGGI (KPT)

- Geran Penyelidikan Research Acculturation Collaborative Effort (RACE)
- Malaysia Laboratories For Academia-business Collaboration (MyLaB)
- Geran Penyelidikan Sukan Kementerian Pengajian Tinggi
- Geran Konsortium Kecemerlangan Penyelidikan (KKP)

R&D&C&I FUNDING FRAMEWORK RMKE-12

	AIF	TeD1	TeD2	BGF
	Individu/ Persatuan/ Koperasi/ NGO yang berdaftar/ Pemilikan Tunggal/Perkongsian Liabiliti Terhad	Start-up/ IHL/ Politeknik /Kolej Komuniti/GRI/ Agensi STI	Start-up/ IHL/Politeknik /Kolej Komuniti/ GRI/ Agensi STI (Termasuk Bukan Warganegara)	Start-up/ IHL/Politeknik/ Kolej Komuniti/ GRI/ Agensi STI
	RM500,000.00	RM1,000,000.00	RM3,000,000.00	RM4,000,000.00
	12 - 18 bulan	24 bulan	36 bulan	36 bulan

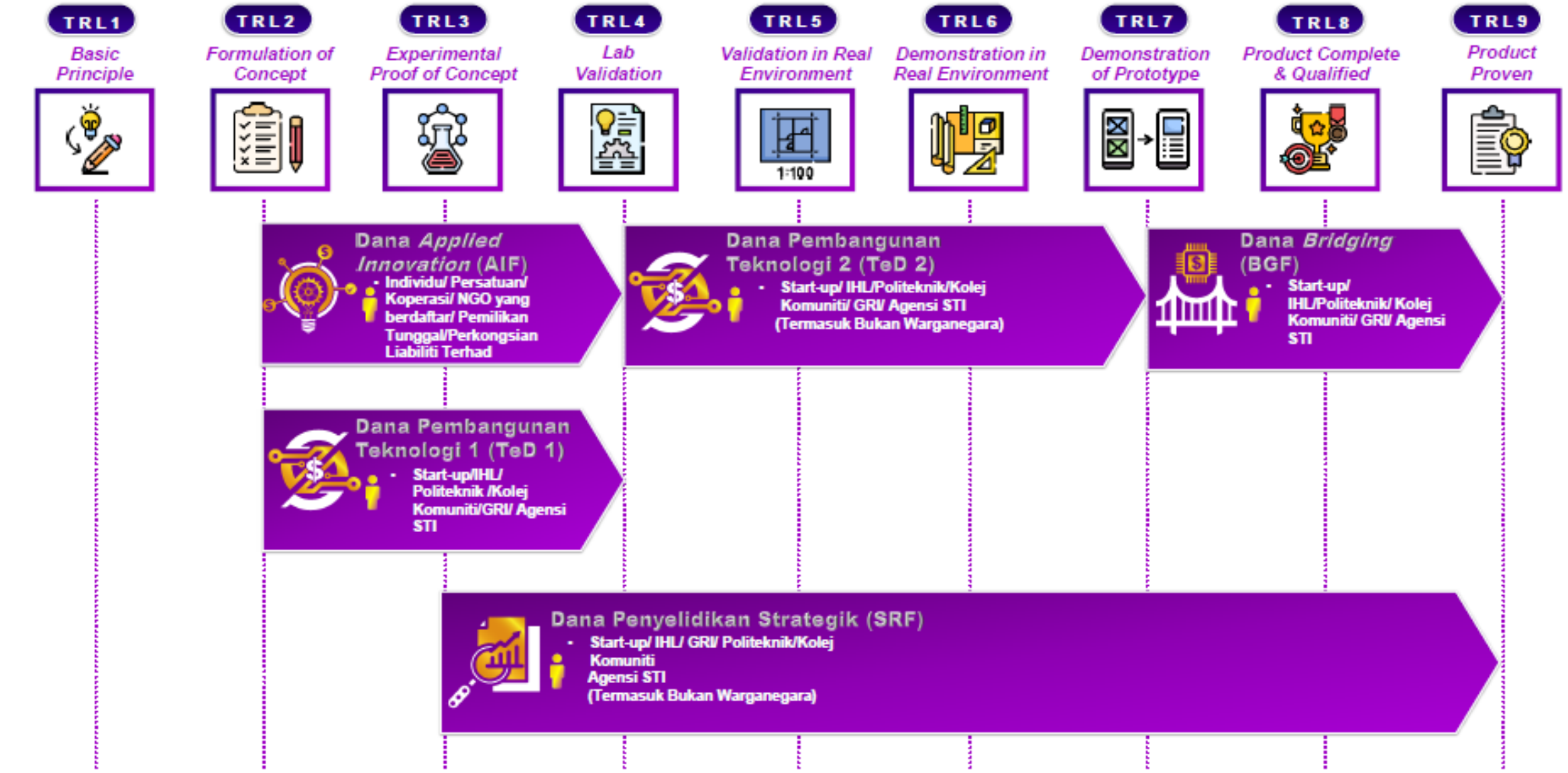
SRF

Start-up/ IHL/ GRI/ Politeknik/Kolej Komuniti
Agensi STI
(Termasuk Bukan Warganegara)

RM15,000,000.00

36 bulan

MOSTI R&D&C&I STAGES AND FUNDING FACILITIES FRAMEWORK



UNIVERSITI TEKNOLOGI MARA
Pejabat
Timbalan Naib Canselor
(Penyelidikan dan Inovasi)

GERAN UNIVERSITI

GERAN UNIVERSITI

- Geran Penyelidikan LESTARI
- Geran Penyelidikan LETARI SDGTriangle@UiTM
- Geran Inisiatif Penyeliaan (GIP)
- Geran Penyelidikan Global Research Reputation (GRR)
- Geran Penyelidikan Strategic Research Partnership (SRP)

GERAN UNIVERSITI

- Geran Penyelidikan Penyelidik Muda Berbakat (YTRG)
- GERAN PENYELIDIKAN MyRA

  i) Geran Penyelidikan MyRA

  ii) Geran Penyelidikan MyRA Sains Sosial

  iii) Geran Penyelidikan MyRA Lepasan PhD

  iv) Geran Penyelidikan MyRA Road to HICoE

SENARAI GERAN INDUSTRI/AGENSI

- MAKNA CANCER RESEARCH AWARD

- MALAYSIA TORAY SCIENCE FOUNDATION (MTSF)

- YAYASAN PENYELIDIKAN ANTARTIKA SULTAN MIZAN (YPASM)

- L'OREAL-UNESCO FOR WOMEN IN SCIENCE FELLOWSHIP

- NEWTON-UNGKU OMAR FUND

- CREST R&D GRANT

SENARAI GERAN ANTARABANGSA

- TURTLE CONSERVATION FUND

- CRP : INTERNATIONAL CENTRE FOR GENETIC ENGINEERING AND BIOTECHNOLOGY (ICGEB)

- HUMAN FRONTIER SCIENCE PROGRAM

- INTERNATIONAL SOCIETY FOR INFECTIOUS DISEASES (ISID)

- ORGANIZATION FOR THE PROHIBITION OF CHEMICAL WEAPONS (OPCW)

- NAGAO NATURAL ENVIRONMENT FOUNDATION
- INTERNATIONAL EDUCATION RESEARCH FOUNDATION (IERF)
- FULBRIGHT-MCMC U.S. SENIOR SPECIALIST GRANT

- TWAS-COMSTECH JOINT RESEARCH GRANTS

- JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE (JSPS)

- TERRA VIVA GRANTS

- National Institutes of Health (NIH)

- National Science Foundation

- The Nippon Foundation

- WELLCOME TRUST

- NATIONAL GEOGRAPHIC SOCIETY

- QATAR NATIONAL RESEARCH FUND

- HUMAN FRONTIER SCIENCE PROGRAM
- ISESCO-COMSTECH RESEARCH GRANTS
- KURITA WATER AND ENVIRONMENT FOUNDATION

Commercialize Product (BITCOM UiTM)

Paina Sauce

PAINA- PINEAPPLE CHILLI SAUCE, A PRODUCT OF FSG
The development of PAINA was carried out by Dr. Azizah Othman, Dr. Fadhilah Jailani, and Dr. Siti Roha Ab. Mutalib from the Department of Food Science and Technology, Faculty of Applied Sciences since 2017.

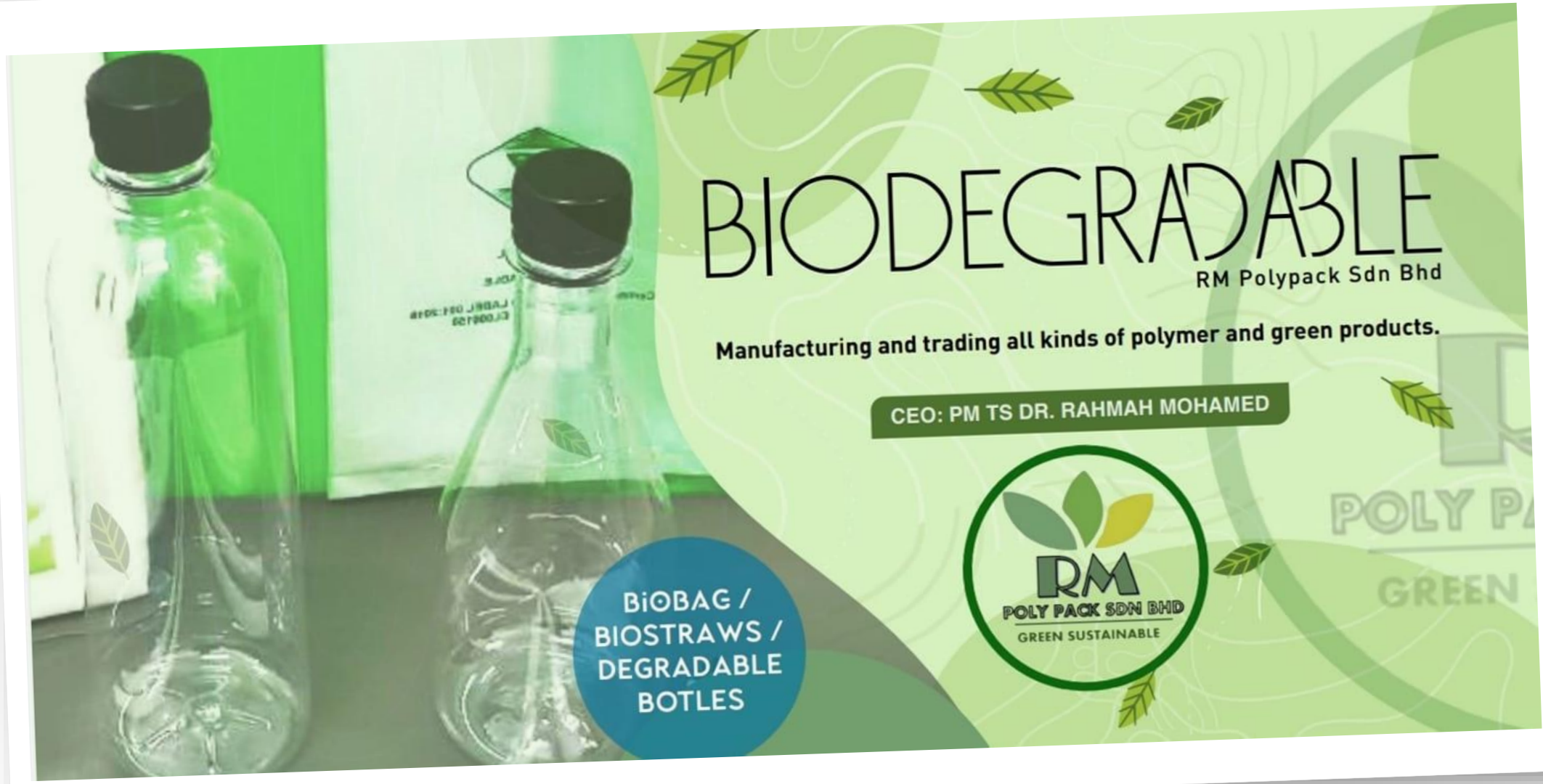
PAINA was developed to offer healthy and multi-purpose condiments to the consumer. The recipe using PAINA as the main ingredient was successfully created by Assoc. Prof Dr. Mohd Hafiz Mohd Hanafiah, Noradzhar Baba, and Hamizad Abdul Hadi from Faculty of Hotel and Tourism Management. The incorporation of pineapple puree in the PAINA sauce creates a more unique taste and aroma in many cuisines.

Unizzol Hand sanitizer

Dr Azri's Perfume

Una Coffee

UnaCoffee adalah produk hasil penyelidikan pensyarah Sains dan Teknologi Makanan, Fakulti Sains Gunaan, UiTM Shah Alam di bawah kelolaan syarikat start-up Fav Food Industries.



Hasli penyelidikan PM Dr Rahmah

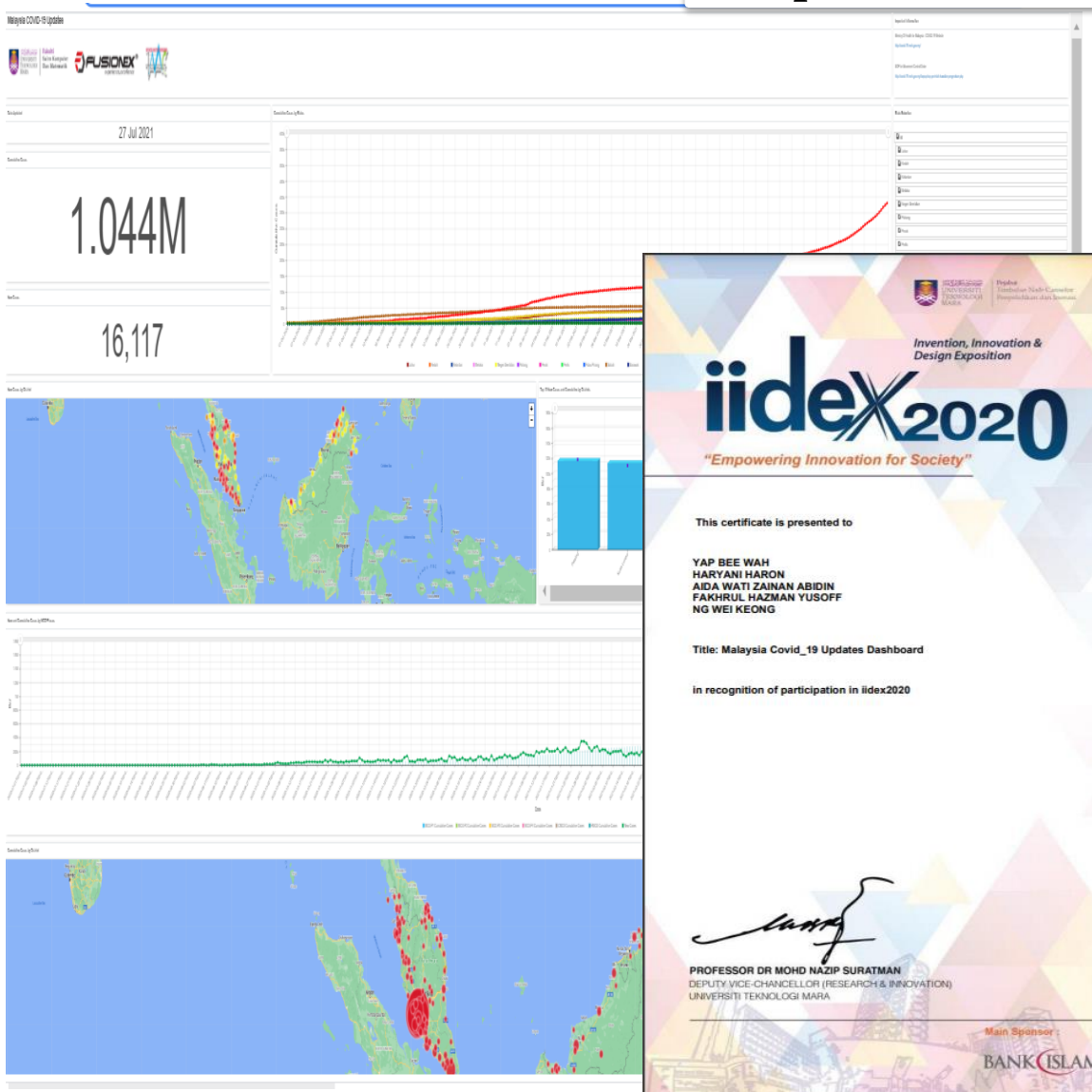
Genotyping kit

Hasil penyelidikan iPROMISE, di ketuai oleh Prof Dato Mohd Zaki Salleh

Education kits

Sample current research

Malaysia Covid-19 Updates dashboard



https://mycovid.uitm.edu.my/

MAKLUMAT PROJEK

Tajuk Projek	Enhancing Future Food Security Through Sustainable Aquaponics System in Rural Environment
Organisasi Pemohon	Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia
Ketua Projek	Ts Dr Saiful Farik Bin Mat Yatin
Kuantum (RM)	MYR 300,000
Kolaborator *(Jika ada)*	ANGKATAN KOPERASI KEBANGSAAN MALAYSIA BERHAD (ANGKASA) PAYER MAJU ENTERPRISE
Lokasi Projek *(lokasi projek akan dilaksanakan)*	Kg Lompat, Mukim Songsang 28000 Temerloh Pahang Daruk Makmur
Tempoh Pelaksanaan Projek *(tidak melebihi 12 bulan)*	12 bulan
Bidang Projek	☐ Tenaga ☐ Perkhidmatan Perniagaan & Kewangan ☐ Kebudayaan, Keseniaan & Pelancongan ☐ Perubatan & Penjagaan Kesihatan ☑ Teknologi Sistem Pintar ☐ Bandar Pintar & Pengangkutan ☐ Air & Makanan ☐ Pertanian & Perhutanan ☐ Pendidikan ☐ Alam Sekitar & Biodiversiti

2

GAMBAR PRODUK/ TEKNOLOGI

Aquaponics Rafting System

Ts Dr Saiful Farik Bin Mat Yatin
Felo IBDAAI, UiTM
&
Fakuti Pengurusan Maklumat

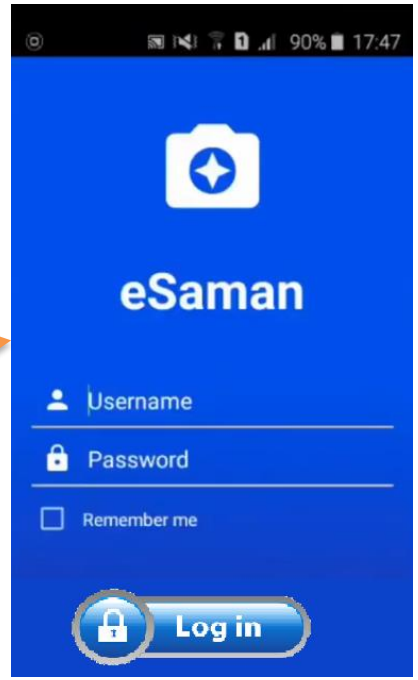
A PROTOTYPE OF INTELLIGENT DATA DRIVEN PREDICTIVE MAINTENANCE FOR OIL & GAS

Industry Project with CeRDAS, UTP

PM. Ts. Dr. Chin Kim On
Faculty of Computing and Informatics
Universiti Malaysia Sabah
kimonchin@ums.edu.my

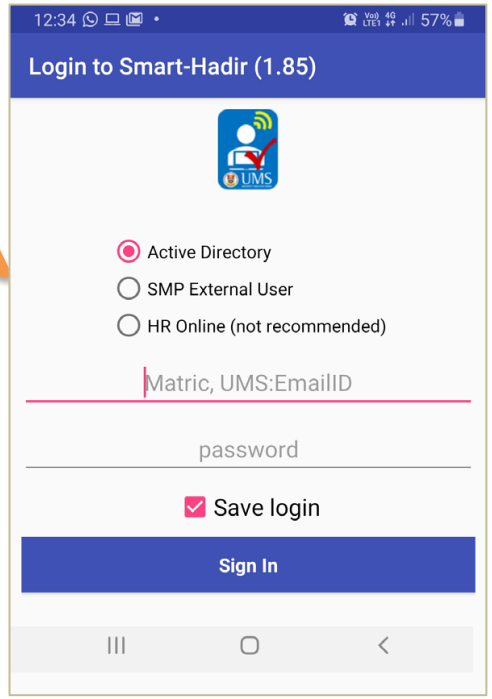
Extended project
PRGS0013-ICT-1/2020
(RM 58,450)

Applied Science
e-Science 01-01-10-sf0235 (2016)
(RM 246,800)



e-Saman UMS - Android based Summon System using Image Processing and Neural Networks

Output – 1 paten filing in progress



Smart-Hadir - Mobile based Attendance Management System

Output – 1 Scopus Q2 paper, 1 bronze award, 1 copyright filling in progress

environment | climate | health | Food security | Agriculture

Smart Sensing & Monitoring

Smart Analytics & Cloud Computing

IOT, Big Data & Cloud Computing

Computing. 2021 Jan 8 : 1–39. PMCID: PMC7791158
doi: 10.1007/s00607-020-00877-8 [Epub ahead of print]

An intelligent healthcare system for predicting and preventing dengue virus infection

Sandeep Kumar Sood,[1] Vaishali Sood,[2] Isha Mahajan,[2] and Sahil[3]

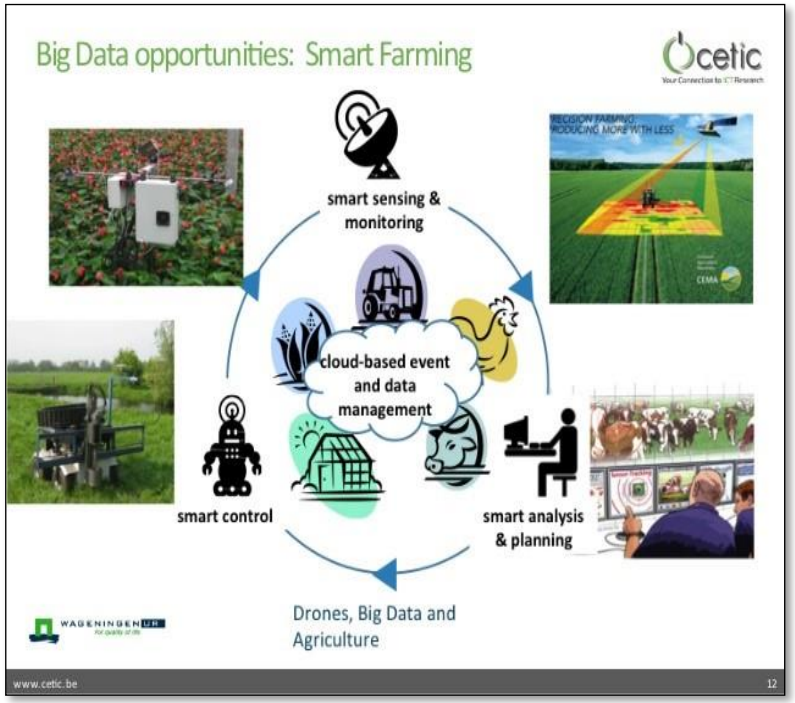
▸ Author information ▸ Article notes ▸ Copyright and License information Disclaimer

Abstract

Go to:

Dengue is a mosquito-borne pandemic viral infection, which transmits to humans from Female Aedes albopictis or Aedes agypti mosquitoes. It progressively deteriorates the health of infected individuals and poses a high threat of human morbidity and mortality. This paper proposes an intelligent healthcare system which identifies, monitors, and alerts dengue virus (DeV) infected individuals and other stakeholders in real-time and control the DeV infection outbreak using cloud computing, internet of things and fog computing paradigms. The proposed system uses Naive Bayesian Network (NBN) for diagnosing the possibly DeV infected individuals and generating real-time alerts for suggesting and alerting the concerned stakeholders for taking on-time necessary actions at the fog subsystem. The proposed system also uses Social Network Analysis at the cloud subsystem, to provide Global Positioning Systems (GPS)-based global risk assessment of the DeV infection on Google Maps (Google-based web map service) and control DeV infection outbreak. The analysis of the experimental results acknowledges the efficiency of the NBN-based DeV infection diagnosis, alert generation, and GPS-based risk assessment functionality, of the proposed system, via various statistical measures and experimental approaches.

Keywords: Dengue virus, Cloud computing, Fog computing, Internet of things (IoT), Naive Bayesian network (NBN), Global positioning system (GPS), Social network analysis (SNA)



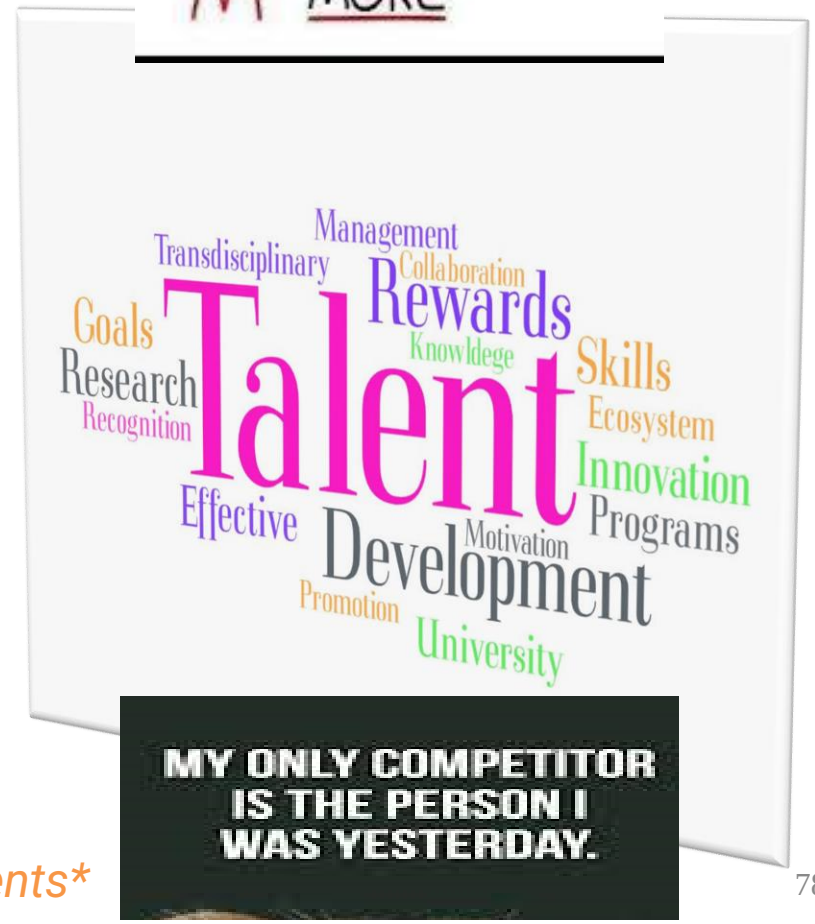
https://www.whizlabs.com/blog/relationship-between-iot-big-data-cloud-computing/

Push-Pull and Key Success Factors

Push factors	Pull factors	Key Success Factors
✓ Leadership: Top-Down ✓ Organizational environment ✓ University research ecosystem	✓ Funding ✓ Promotion ✓ Recognition ✓ Commercialization support	✓ Good leadership ✓ Good project management ✓ Good reward system/practices ✓ Talent Development ✓ Smart partnership

Conclusion

- New generation of researchers with **inter- and trans-disciplinary skills.**
- **Data-intensive research-** support upskilling of lecturers.
- New **incentives** and **measures** for evaluating and rewarding both **individual** and **collective contributions** to research
- Good **ecosystem** for well co-ordinated research strategies, incentives, and monitoring to ensure successful outputs.



Passion is the catalyst for perseverance and achievements

Ackowledgement

My sincere thanks to:

- YBhg Prof Dr Mohamad Abdullah Hemdi (Rektor UiTM, Cawangan Kedah)
- ICMS2021 Conference Committee (Dr Ida Normaya Mohd Nasir, Puan Norashikin Nasaruddin, En Kamarul Ariffin Mansor)
- Prof. Dr Mohd Nazip Suratman, TNCPI, UiTM
- Prof. Dato' Dr Abu Bakar Abdul Majeed-Pengarah RMC,UiTM
- Prof. Dr Nooritawati Md Tahir, Pengarah ReNeU, UiTM
- Prof. Madya Ir. Ts. Dr Syed Shatir Asghrar Syed Hasan, Pengarah BITCOM
- Prof. Dr. Haryani Haron-Dean, FSKM, UiTM
- Prof. Dr Jasni Mohamad Zain, Pengarah IBDAAI, UiTM
- Dr Wan Fairos Wan Yaacob, UiTM Cawangan Kelantan
- Nurul Azifah Mohd Pauzi
- Mohamad Nasir Abdullah
- Ezzatul Akmal Kamaru Zaman
- Collaborators & Colleagues