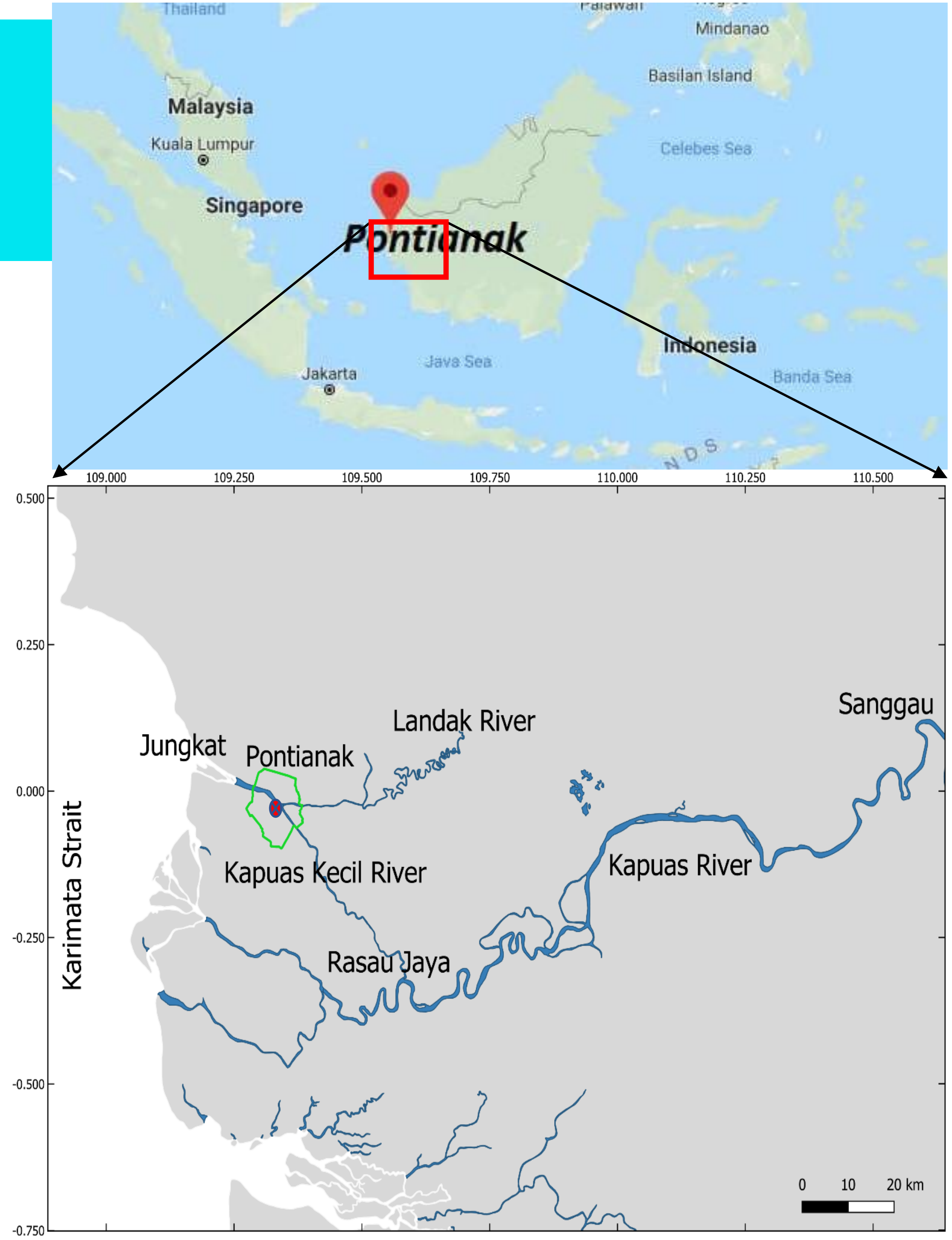
Machine learning models for predicting a daily maximum water level in a tidal river: A Kapuas Kecil River case study

Joko Sampurno*, Valentin Vallaeys, Randy Ardianto, Emmanuel Hanert
*joko.sampurno@uclouvain.be

International Conference on Radioscience, Equatorial Atmospheric Science and Environment (INCREASE)

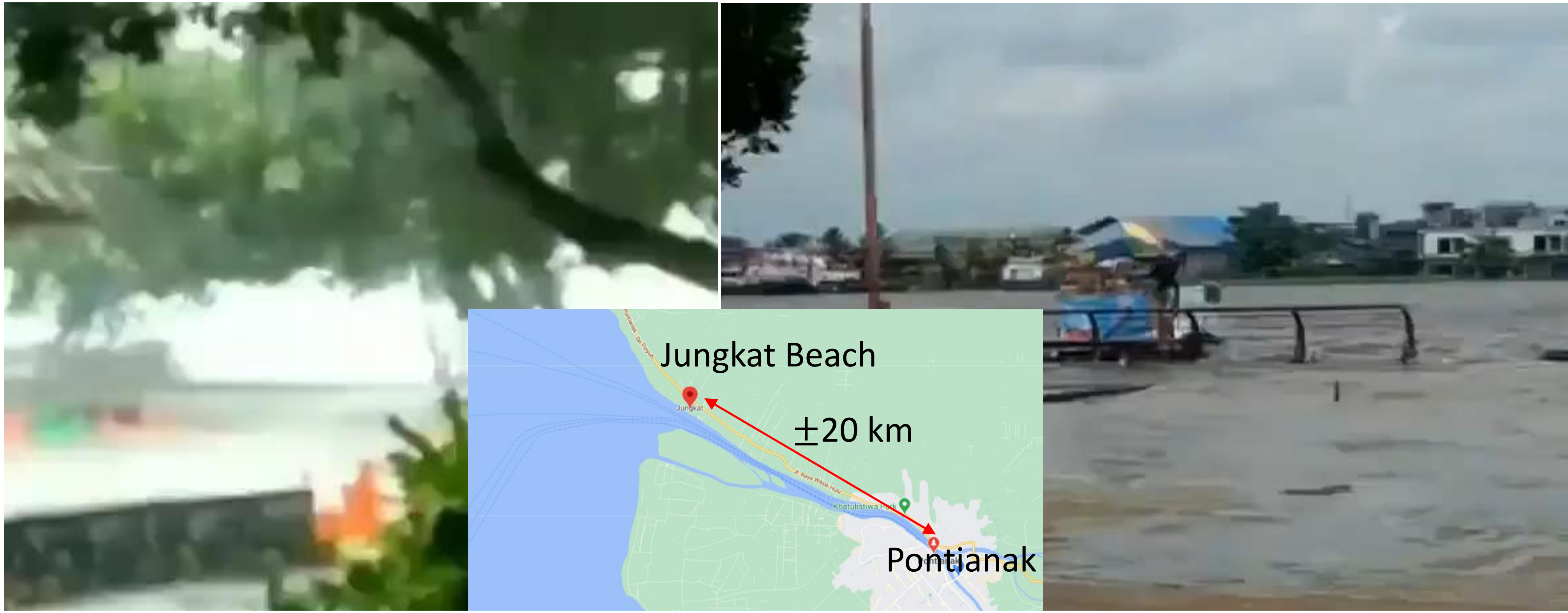
Background

- Pontianak is a low-lying city located near to estuary and crossed by a tidal river flow (the stream of Kapuas Kecil River).
- The city is vulnerable to compound floods, which endangering about 600.000 inhabitants who live there.



Jungkat Beach, Estuary of Kapuas River   Pontianak, located about 20km from river mouth

Compound flood?



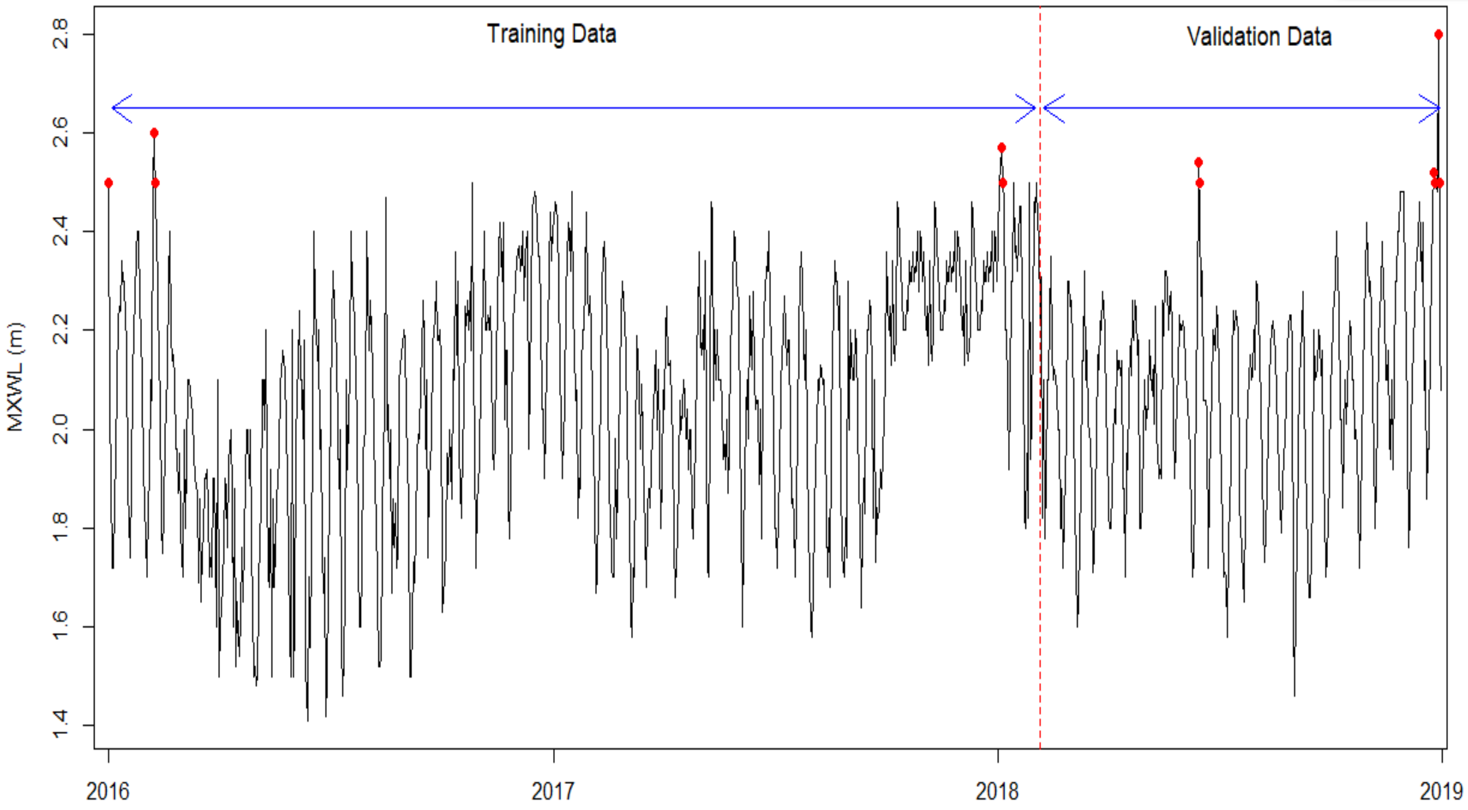
- Based on the IPCC standard (2012) about compound events: Compound flood is flood which is driven by a combination of two or more extreme events—such as storm surge, intense rainfall or high discharge—and this could amplify the hazardous effects (Seneviratne, Nicholls, et al., 2012).
- To mitigate the risk of the event, modeling the daily maximum water level (MXWL) is a tool that can be used to assess the future compound flood hazard.

*the Intergovernmental Panel on Climate Change (IPCC) 2012 report: https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf

Aim

This study aims to develop a daily maximum water level prediction model suitable for the Kapuas Kecil River, capable of predicting compound floods in Pontianak using machine learning algorithm.

Dependent vs. Independent Variables

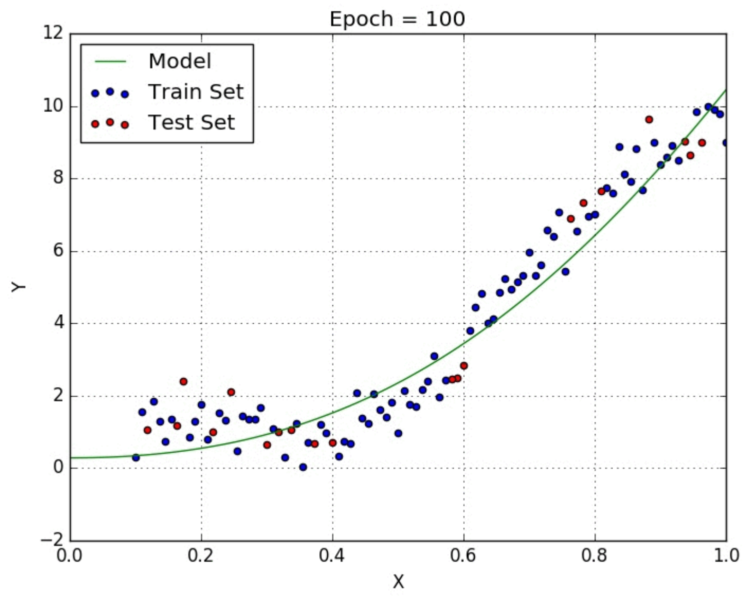


No	Variable	Description
1	Tide_max	Maximum Tide at River Mouth (m)
2	QSanggau	Daily Discharge of the Kapuas River at Sanggau in time (m3/s)
3	QSanggau1	Daily Discharge of the Kapuas River at Sanggau 1 day before (m3/s)
4	QSanggau2	Daily Discharge of the Kapuas River at Sanggau 2 day before (m3/s)
5	QLandak	Daily Discharge of the Landak River at Kuala Mandor in time (m3/s)
6	QLandak1	Daily Discharge of the Landak River at Kuala Mandor 1 day before (m3/s)
7	QLandak2	Daily Discharge of the Landak River at Kuala Mandor 2 day before (m3/s)
8	RR	Daily Precipitation (mm) at Pontianak
9	WS_AVG	Daily Average Wind Speed (m/s) at Pontianak
10	WS_MAX	Daily Maximum Instantaneous Wind Speed (m/s) at Pontianak
11	WD_AVG	Daily Average Wind Direction (degree, with range: 0 - 360) at Pontianak
12	TT_AIRMAX	Daily Maximum Air Temperature (degree Celsius) at Pontianak
13	TT_AIRAVG	Daily Average Air Temperature (degree Celsius) at Pontianak
14	TT_AIRMIN	Daily Minimum Air Temperature (degree Celsius) at Pontianak
15	RH	Daily Relative Humidity (%) at Pontianak
16	PP	Daily Atmospheric Pressure (millibars) at Pontianak
17	SR_AVG	Daily Average Solar Radiation (W/m2) at Pontianak
18	SR_MAX	Daily Maximum Solar Radiation (W/m2) at Pontianak
19	RR_J	Daily Precipitation (mm) at Jungkat
20	WS_maxJ	Daily Maximum Instantaneous Wind Speed (m/s) at Jungkat
21	WDJ	Daily Average Wind Direction (degree, with range: 0 - 360) at Jungkat
22	WS_avgJ	Daily Average Wind Speed (m/s) at Jungkat

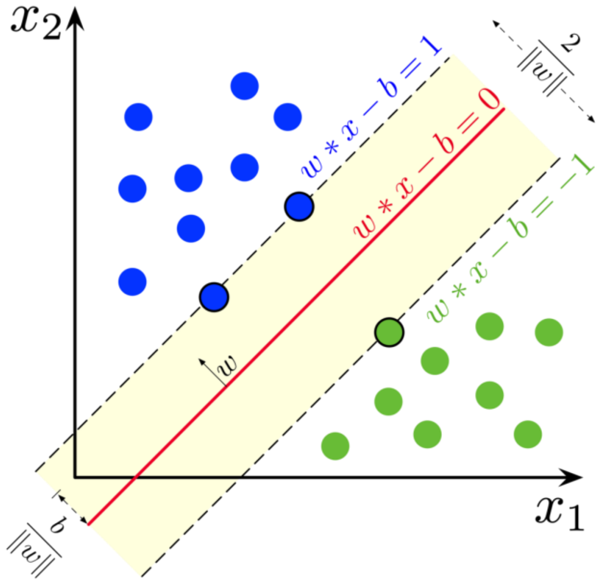
Algorithm

10 model scenarios with different algorithm were evaluated, where:

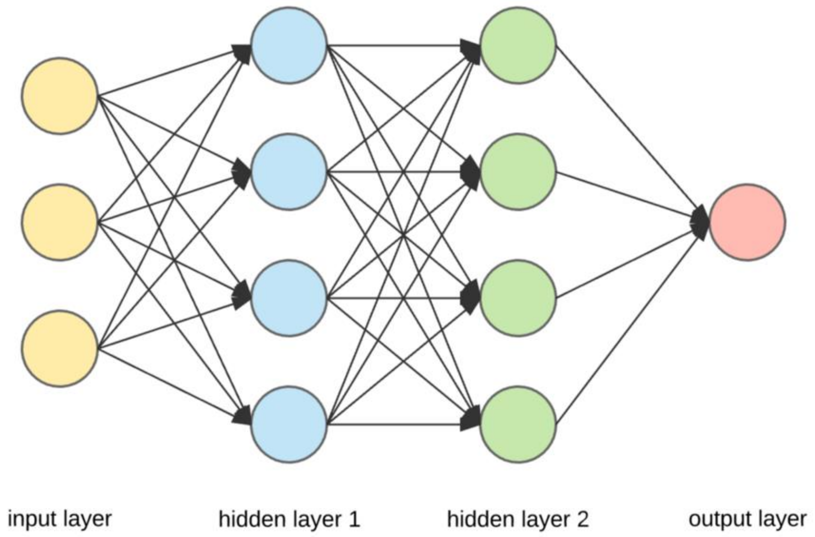
- Multiple Linear Regression (MLR): 1 scenarios
- Random Forest (RF): 3 scenarios
- Artificial Neural Network (ANN) : 3 scenarios
- Support vector machine (SVM) : 3 scenarios



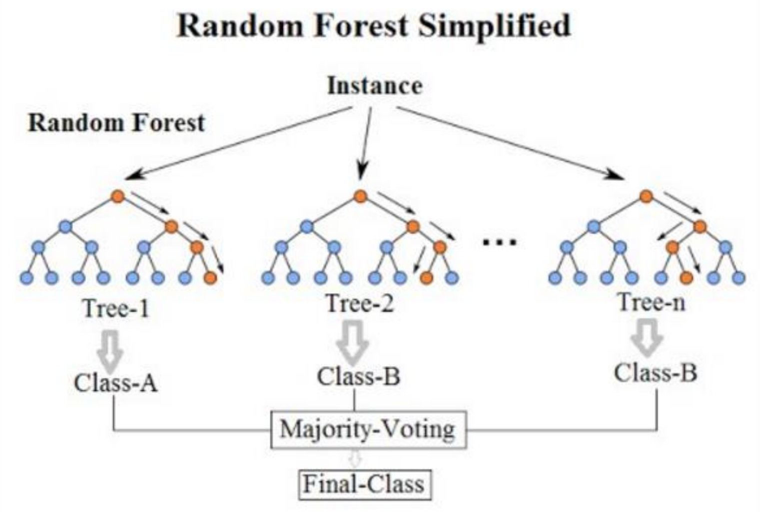
https://www.cs.toronto.edu/~frossard/post/multiple_linear_regression/



https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:SVM_margin.png

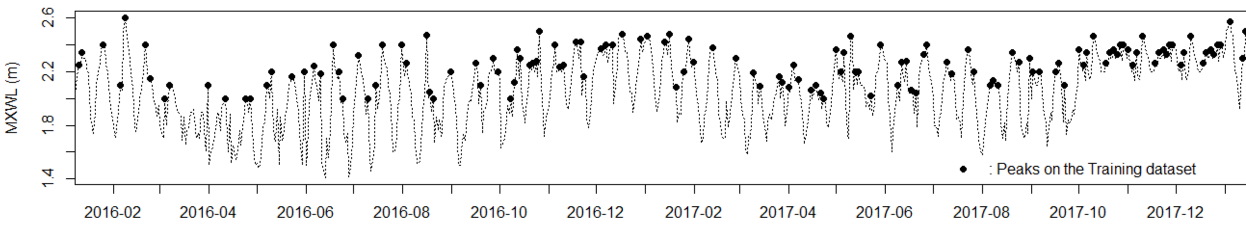


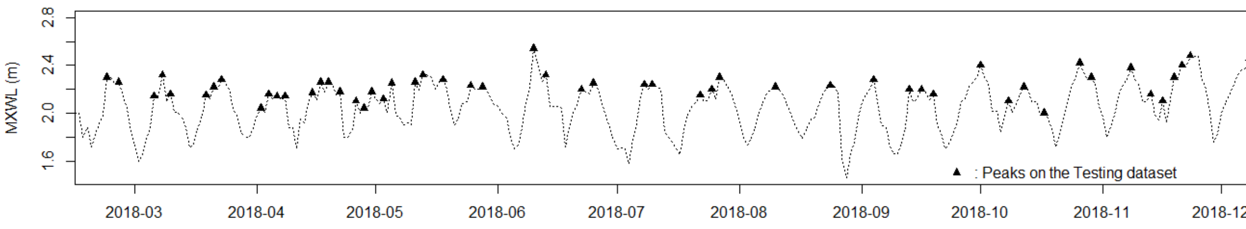
https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6



https://towardsdatascience.com/master-machine-learning-random-forest-from-scratch-with-python-3efdd51b6d7a

Metrics for Evaluation





Feature analysis

- Mutual Information (MI)

$$MI_{(X,Y)} = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)-p(y)}\right)$$

Model Performance Evaluation:

- Mean Absolute Error (MAE) → peaks
- Root Mean Square Errors (RMSE)→ peaks
- Pearson Correlation Coefficient (PCC)
- Nash–Sutcliffe Efficiency (NSE).

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| Y_{m(i)} - Y_{obs(i)} \right|}{n}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{m(i)} - Y_{obs(i)} \right)^2}{n}}$$

$$PCC = \frac{\sum_{i=1}^{n} \left(Y_{m(i)} - \overline{Y_m} \right) \left(Y_{obs(i)} - \overline{Y_{obs}} \right)}{\sqrt{\sum_{i=1}^{n} \left(Y_{m(i)} - \overline{Y_m} \right)^2 \sum_{i=1}^{n} \left(Y_{obs(i)} - \overline{Y_{obs}} \right)^2}}$$

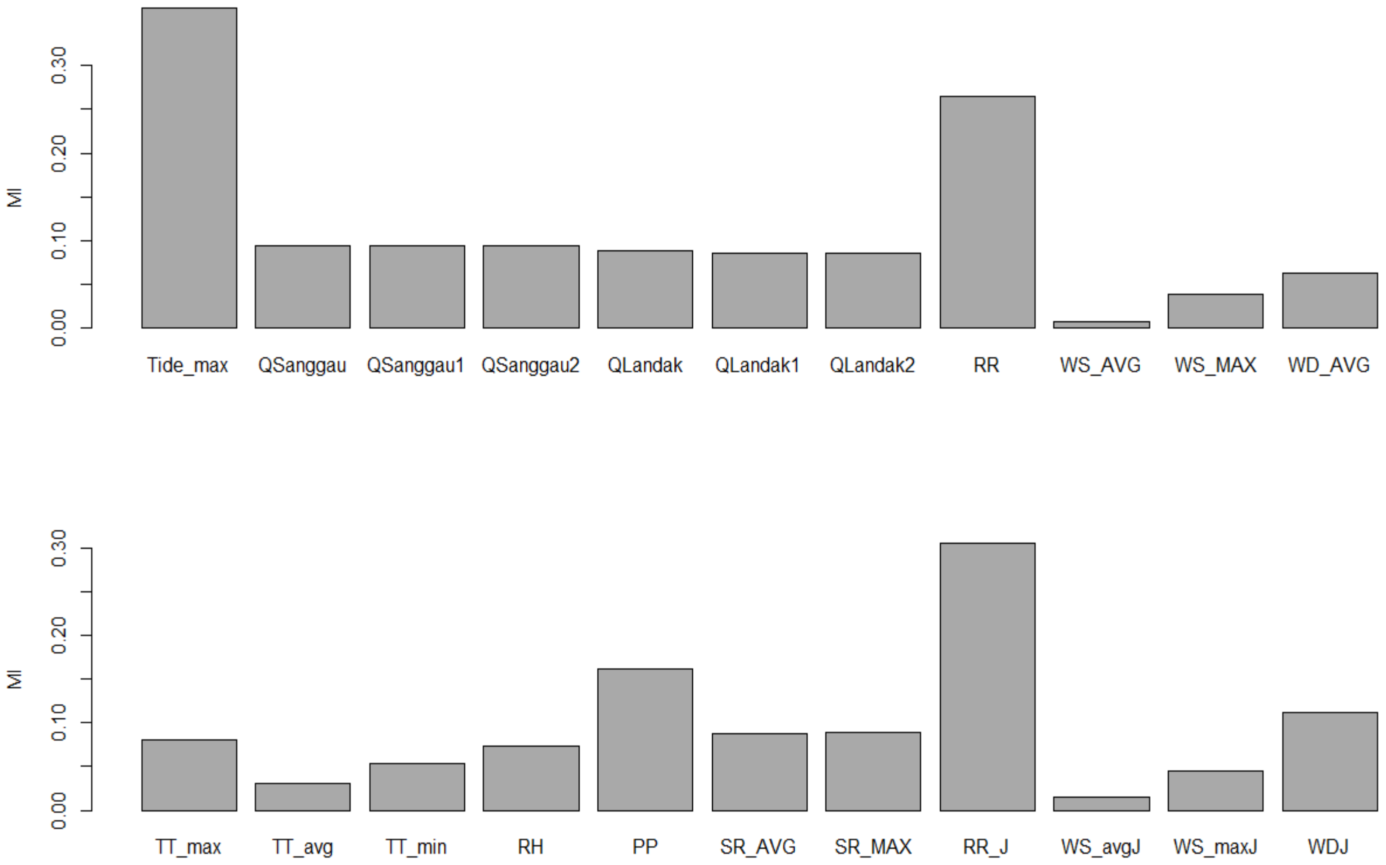
$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Y_{m(i)} - Y_{obs(i)} \right)^2}{\sum_{i=1}^{n} \left(Y_{obs(i)} - \overline{Y_{obs}} \right)^2}$$

The best model criterion:

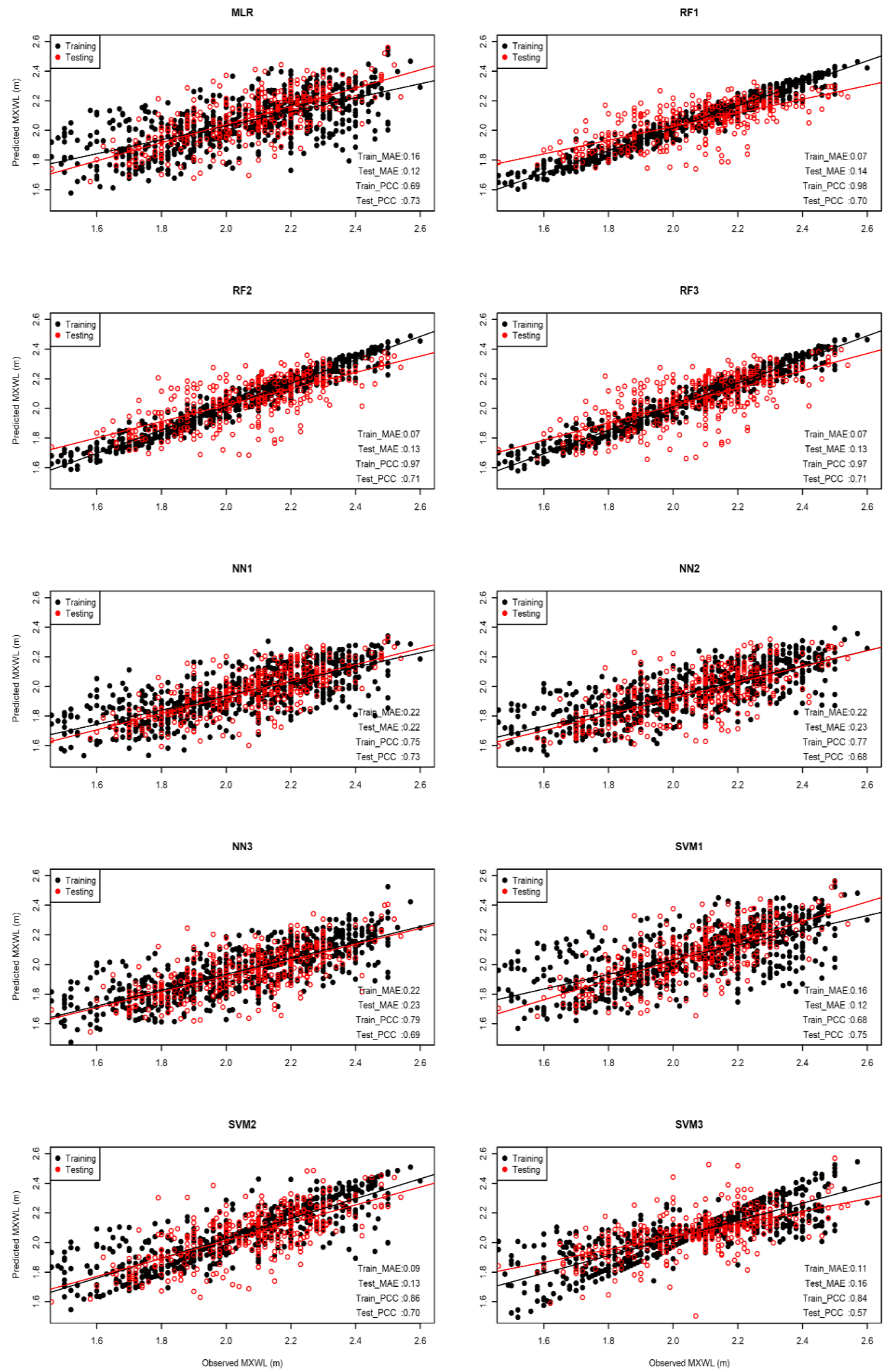
- The best model was selected based on its capability to predict flood events during the testing phase and its goodness-of-fit coefficients.

Results and Discussion

Feature analysis

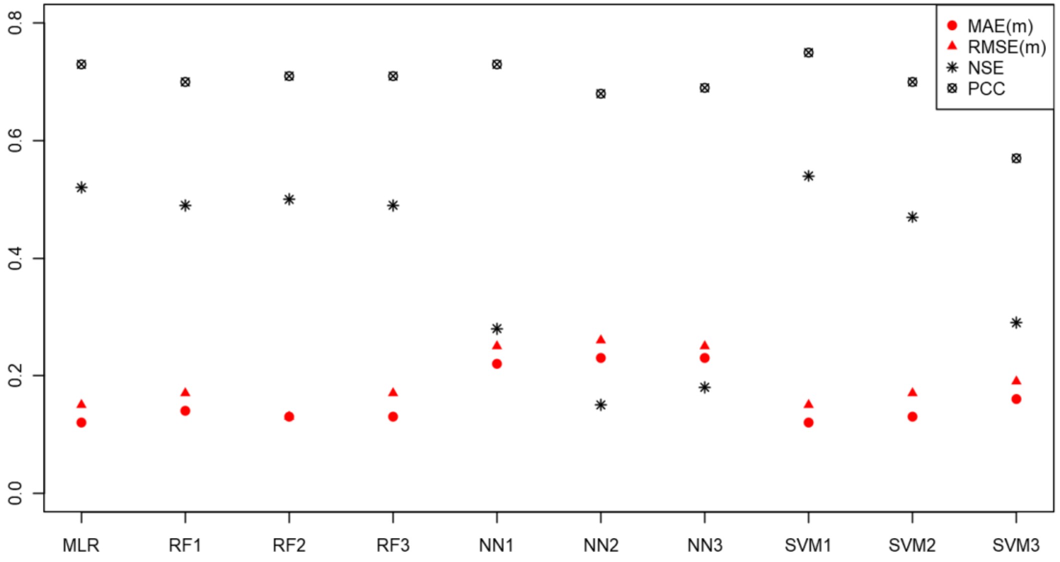


Model Performance:

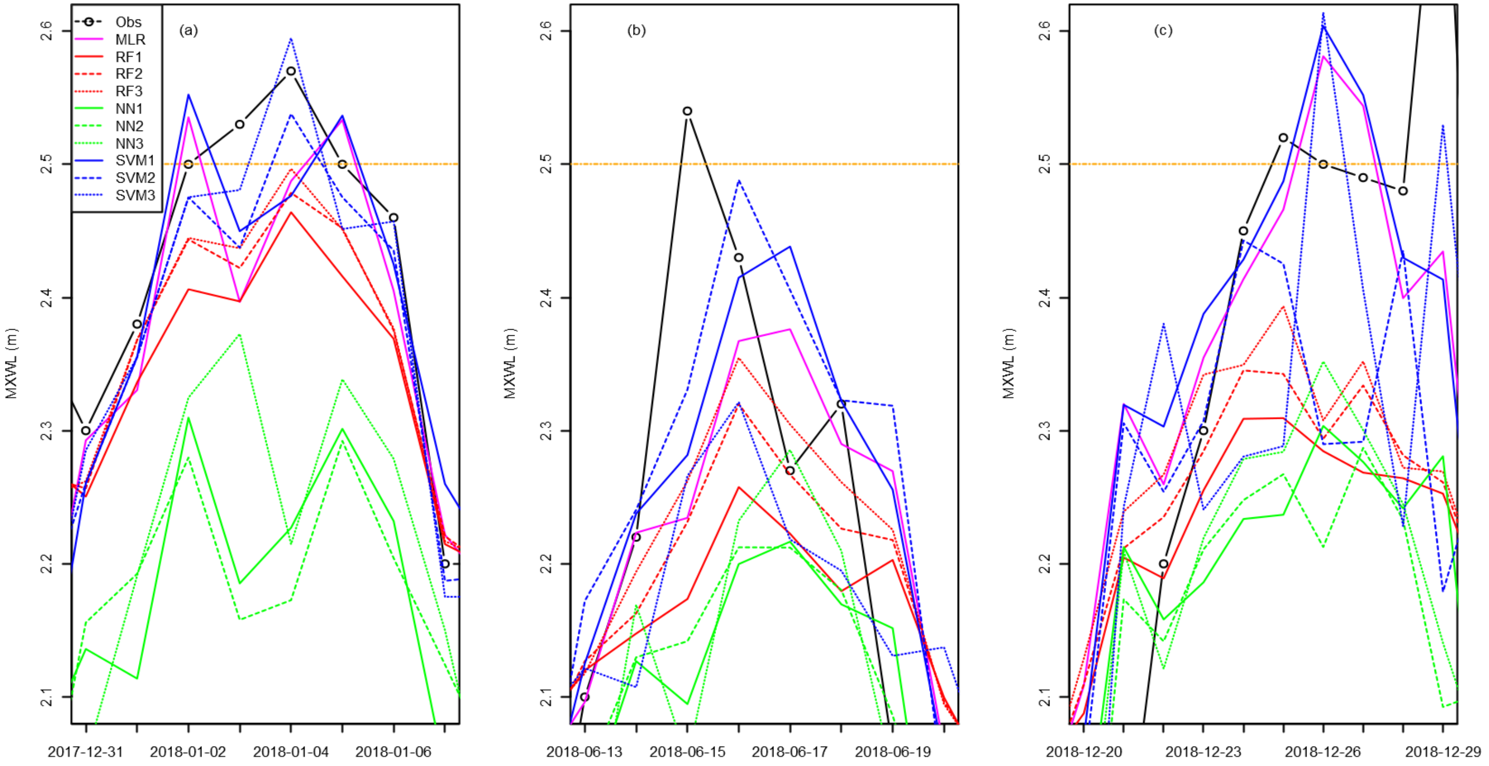


Testing Phase

Error metric



Prediction of Flood



Take-Home Messages

- There are five variables that strongly control the MXWL in Pontianak, i.e., maximum tide elevation, rainfall in both areas (river mouth and in the city), air pressure, and wind direction over the river mouth.
- Even though all models tested perform well both in training and testing sessions, we propose the support vector machine with a polynomial kernel function (SVM3 scenario) as the most suitable model to predict the future compound flood in the city of Pontianak.