Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTING RAINFALL FROM WEATHER OBSERVATIONS USING SVM APPROACH FOR IDENTIFY THE PARAMETER OF FUEL MOISTURE CODE AS FIRE WEATHER INDEX

DARWIS ROBINSON MANALU[1,2] , MUHAMMAD ZARLIS[1], HERMAN MAWENGKANG[1], OPIM SALIM SITOMPUL[1]

[1]Program Studi Doktor (S3) Ilmu Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, North Sumatera-20222, Indonesia
[2]Universitas Methodist Indonesia, Medan, Sumatera Utara, Indonesia

ABSTRACT

The Fine Fuel Moisture Code (FFMC) is a numeric rating of the dampness substance of litter and other restored fine fills. This code is a pointer of the general simplicity of start and the combustibility of fine fue. In this study we observed the rainfall time series as a parameter to get the index of FFMC. The main goal of this study to predict the amount of rainfall in a particular division or state well in advance. We predict the amount of rainfall using past data to generate the parameter of FFMC using SVM model in North Sumatera. Based on the result, the various visualizations of data are observed in Aek Godang, North Sumatera which helps in implementing the approaches for rainfall prediction to evaluate the parameter of fuel moisture code as fire weather index. The analysed individual year rainfall patterns for 2017, 2018, 2019, the approximately close means, noticed less standard deviations.

Keywords: *Rainfall, Predicting, FFMC, SVM Model, North Sumatera*

1. INTRODUCTION

Monsoon prediction is clearly of great importance for Northern Sumatera, Indonesia for identify the parameter of fire weather index. As we know that the fire weather index is a one factor to identify the forest fire in Indonesia. Forest fire in Indonesia was affected a large portion of the Indonesian population causing economic hardship and disruption to commerce and short, and long-term health problems [1]–[3].

To estimate the risk of wild fire, a general system is used to rate the fire danger, called Fire Weather Index (FWI). This system consists of six components, Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought code (DC), Initial Spread Index (ISI), Buildup Index (BUI), and Fire Weather Index (FWI), that account for the effects of fuel moisture and wind on fire behavior [4], [5]. In this study we focused to compute the DC as a part of FFMC parameter based on the rainfall parameter. Rainfall can be generated the information of DC [6], [7]. The Drought Code (DC) is a numeric rating of the common moisture content material of deep, compact organic layers. This code is a useful indicator of seasonal drought outcomes on forest fuels and the amount of smoldering in deep duff layers and giant logs [8].

The main issue of this study, the precipitation can increase best lifeless fuel moisture greater swiftly than any other factor. Fine dead fuels react very unexpectedly to precipitation and reach their saturation factors quickly. In North Sumatra, rainfall is one indicator which very important to identify the DC, FFMC and DMC, so we need to predict the rainfall using some model in machine learning. The more important of rainfall can be describe of indicator of FFMC to evaluate the potential of Fire Index Weather. Most attention to rainfall data not only used to justify the indicator of climate change, but also can generated to evaluation of fire weather index for justify of forest fire in some area.

Two types of rainfall predictions can be done, they are long term predictions as Predict rainfall over few weeks/months in advance and Short-term predictions by predict rainfall a few days in advance in specific locations. Indonesia meteorological department provides forecasting data required for project[9]. In this study, we are planning to work on long term predictions of rainfall. The main goal of this study is to predict the amount of rainfall in a particular division or state well in advance. We predict the amount of rainfall using past data to generate the parameter of FFMC.

In this study, we have a limited distribution rainfall data as one indicator component to evaluate the fire weather index by compute of FFMC and in general issue forest fire identify based on the hot spot in some region. The main issue of how to evaluate the forest fire not only come from the hot spot distribution but also come from the potential rainfall indicator to making justify fire weather index.

2. DATA AND METHODS

The dataset of this study, has average rainfall from 2017-2020 for each station every month in North Sumatra as show in Table 1.

No	Station	Latitude	Longitude
1	Aek Godang	1,55	99,45
2	Balai Besar	3,53	98,64
3	Belawan	3,78	98,71
4	Binaka	1,16	97,70
5	Fl Tobing	1,55	98,88
6	Geophysics Deli Serdang	3,50	98,56
7	Gunung Sitoli	1,16	97,7
8	Klimatologi Deli Serdang	3,62	98,70
9	Kualanamu	3,64	98,89

In this study, we converting data in to the correct format to conduct experiments by make a good analysis of data and observe variation in the patterns of rainfall. Finally, we try to predict the average rainfall by separating data into training and testing. We apply various statistical and machine learning approaches (SVM) in prediction and make analysis over various approaches. By using various approaches, we try to minimize the error [10].

This model is one of the most effective machine learning algorithms, both practical and theoretical. In the early 90's to early 2000's, SVM arguably took over the role of Neural Networks as the most favorable algorithm due to its speed, accuracy[11], and guarantee to always produce optimum global solutions - until around 2006-2007, neural networks (which wrapped in the name Deep Learning) takes over that role, especially for large scale problems, ie, amount of training data> ~ 10,000. SVM with a focus on (linear) binary classification problems, whether it is gradual, intuitive, mathematical, and at the same time simple implementation using the libsvm / liblinear library to solve handwritten digit recognition problems. The real strength of SVM is to use a kernel trick to solve non-linear SVM problems without a kernel trick. [12].

Variable / Constant: scalars are written in lowercase letters / symbols, vectors are written in bold letters / symbols, and matrices are written in bold letters /symbols. Example: a (scalar), a (vector), A (matrix). Vector norm: ‖ a ‖. Example in Euclidean space: $||a|| = \sqrt{a_1^2 + \cdots + a_d^2}$, which is known as the length of the vector. Scalar Space (real): R

Vector Space (real): $R^d, d > 1$, Dot Product: $(a, b) = a\ b = \sum_{j=1}^{d} a_j b_j$, Orthogonality: Two vectors $a, b \in R^d$ are said to be orthogonal (perpendicular to each other) when $a\ b = 0$

Input Chamber: X, Output / Label Space: Y, Loss Function: $L : YxY \rightarrow R$, The function L accepts 2 inputs from space Y and returns a real number output.

Indicator function: $1_{a \neq b}$ (+1 if $a \neq b$, -1 if vice versa),

Expected Risk: $R(f) = E_{x \sim p}[L(f(x), c(x))]$,

Empirical Risk: $R(f) = \frac{1}{n} \sum_{i=1}^{n} L(f(x), c(x_i))$, where $f \in H$ is the hypothesis and $c \in C$ is the concept. Classification problem is a problem to categorize an entity into certain groups. For this purpose, it focuses on the problem of binary classification: classifying an entity into True (+1) or False (-1) groups. Machine learning aims to find optimal solutions for functions[13]

$f : x \ \{+1, -1\}$ is given n samples $\{(x_i, y_i)\}\frac{n}{i=1}$, where $x_i \in X$ and $y_i \in \{-1,1\}$. In general, x is a sub-vector space.

More specifically, the machine learning algorithm looks for a function f: x → {-1,1} that minimizes empirical risk:

$$f^* \coloneqq min \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i)$$

where L (a, b) = 1a ≠ b is a loss function.

The SVM algorithm is an algorithm that solves optimization problems based on equations (4) and (9) above. More specifically, given n samples $\{(X_i, Y_i)\}_{i=1}^{n}$, SVM tries to solve the problem as follows:

$$(w, c) \coloneqq \arg\min \frac{1}{2} \|w\|_2^2$$
$$subject\ to\ y_i(\langle w, x_i \rangle + c) \geq 1, \forall_i = 1, \ldots . n \ldots \quad (1)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, we can summarize the above expression with only 1 equation in Lagrangian form: where $\alpha_i, \forall_i = 1, \dots, n$ are Lagrange multipliers. Because later we will use a special optimization tool to solve the problem (10), by requiring the calculation of the slope against w and c from equation (1): $\frac{\partial L}{\partial w} = 0$ dan $\frac{\partial L}{\partial c} = 0$.

$$\frac{\partial L}{\partial w} = 0 \dots(2)$$

$$\rightarrow w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \dots.(3)$$

$$\rightarrow w = \sum_{i=1}^{n} \alpha_i \, y_i x_i \quad \dots(4)$$

and

$$\frac{\partial L}{\partial c} = 0$$

$$\Rightarrow \; - \sum_{i=1}^{n} \alpha_i \, y_i = 0 \dots..(5)$$

Furthermore, have you found the optimal w and c? Then one must know the value of the Lagrange SVM characteristics as described in the previous section, summarized as following[14]. In principle, SVM is a linear classifier, pattern recognition is done by transforms the data in the input space to a higher dimensional space, and optimization is done on vector space recently. This differentiates SVM of the pattern recognition solution on generally, who do the optimization parameters in the transformed space which is a lower dimension than the input space dimension; and implementing a Structural Risk strategy Minimization (SRM); d). Basically, the working principle of SVM is only able to handle two class classification.[15].

The Fine Fuel Moisture Code (FFMC) speaks to the dampness substance of litter and cured fine fills, 1-2 cm profound. It communicates the ease of start and fuel combustibility. FFMC is touchy to day-by-day changes in temperature, precipitation, relative mugginess, and wind speed. Time slack is 2/3 days, which suggests that it takes two-thirds of a day for the fine powers to react to a alter within the climate [6], [7], [16].

3. RESULT AND DISCUSSION

Based on the Data has individual months, annual, combinations of 3 consecutive months. For some of the data is from 2017 to 2020.

All the attributes have the sum of amount of rainfall in mm. The bar graphs showing distribution of amount of rainfall. In this study, distribution of amount of rainfall yearly, monthly, groups of months. Distribution of rainfall in subdivisions, districts form each month, groups of months. Heat maps showing correlation between amount of rainfall between months.

Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

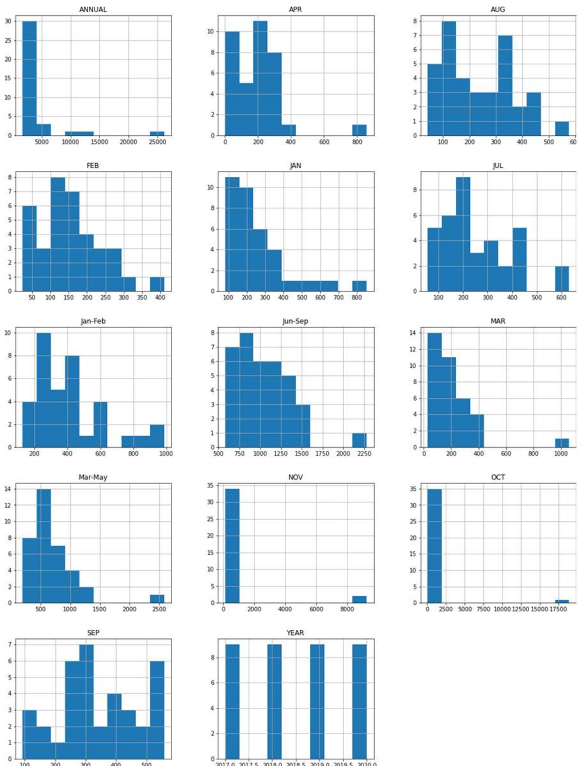


Figure 1. Visual Dataset Of Rainfall 2017-2020

The observation of the dataset shown in figure 2.

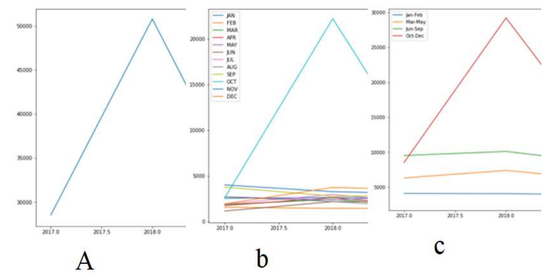


Figure 2. Distribution Of Rainfall Over Years.

In Figure 2a. the histograms show the distribution of rainfall over months. Observed increase in amount of rainfall over months July, August, September. Shows distribution of rainfall over years. The Observed high amount of rainfall in 2018 in Figure 2b.
The graphs clearly show that amount of rainfall in high in the months July, Aug, Sep which is monsoon season in North Sumatera.

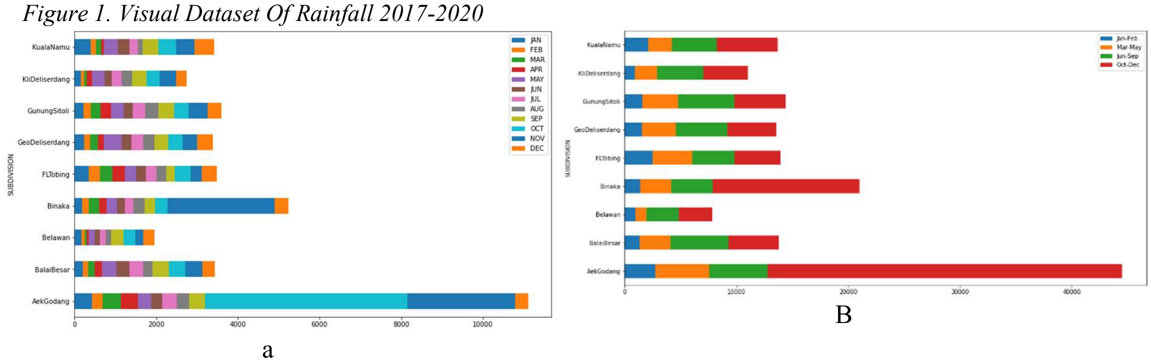


Figure 3. Graphs Show The Distribution Of Rainfall Over Months.

Above two graphs show that the amount of rainfall is reasonably good in the months of March, April, may in eastern North Sumatera

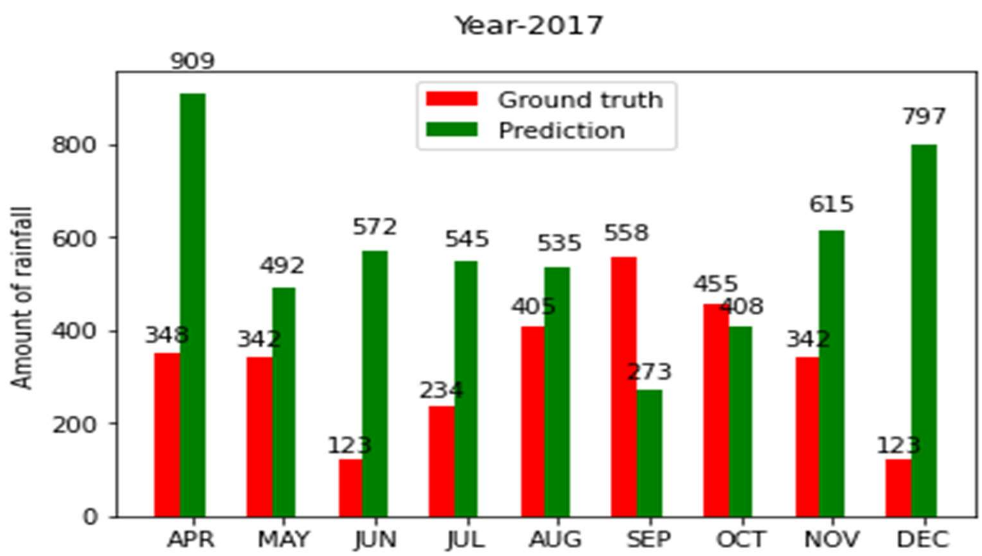
Figure 4. Heat Map Of Rainfall Over Month

Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

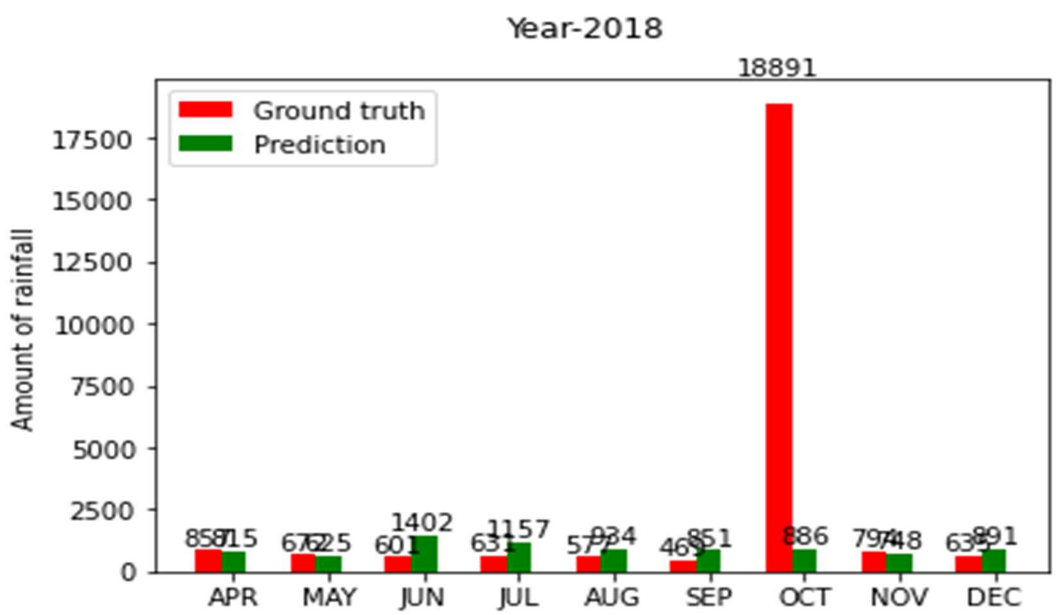
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Heat Map shows the co-relation (dependency) between the amounts of rainfall over months. From above it is clear that if amount of rainfall is high in the months of July, August, September then the amount of rainfall will be high annually. It is also observed that if amount of rainfall in good in the months of October, November, December, then the rainfall is going to be good in the overall year.

For prediction we formatted data in the way, given the rainfall in the last three months we try to predict the rainfall in the next consecutive month. For all the experiments we used 80:20 training and test ratio by used SVM. Testing metrics, we used Mean absolute error to train the models.



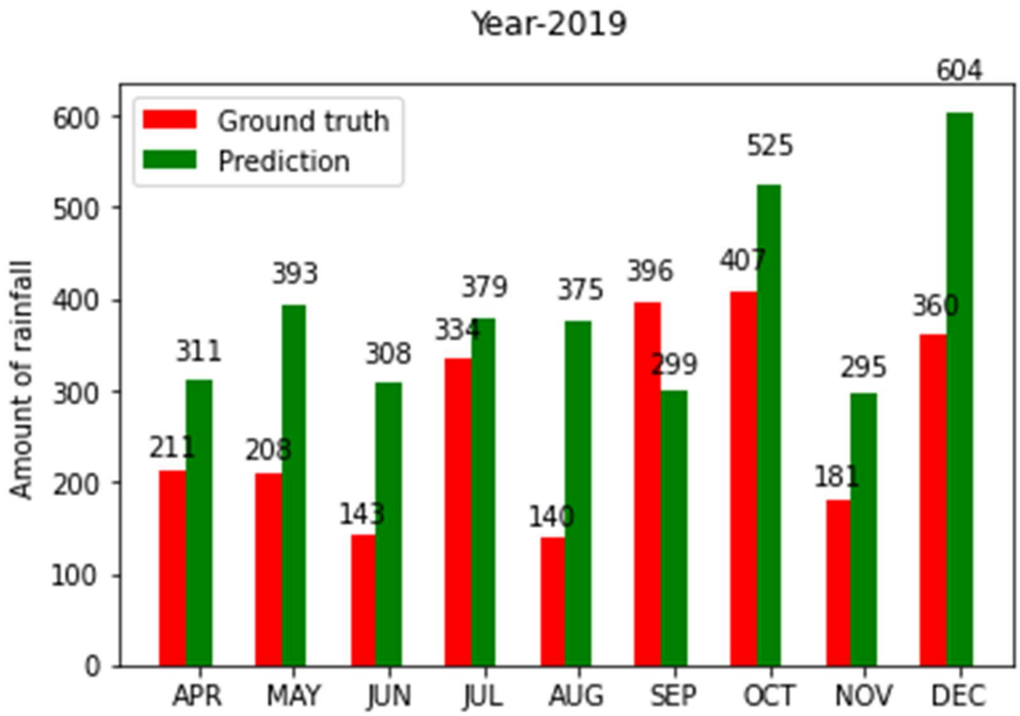
a. 2017



b. 2018

Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

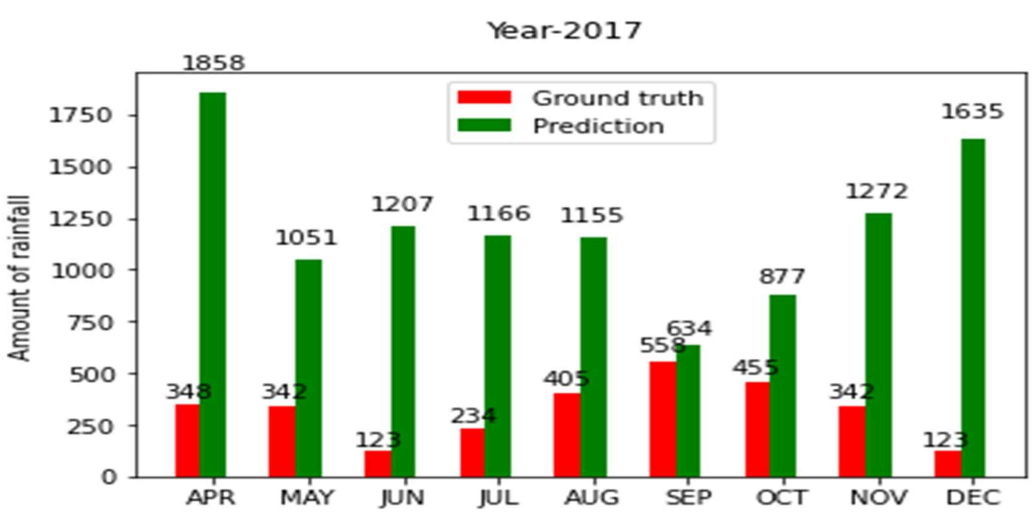


c. 2019

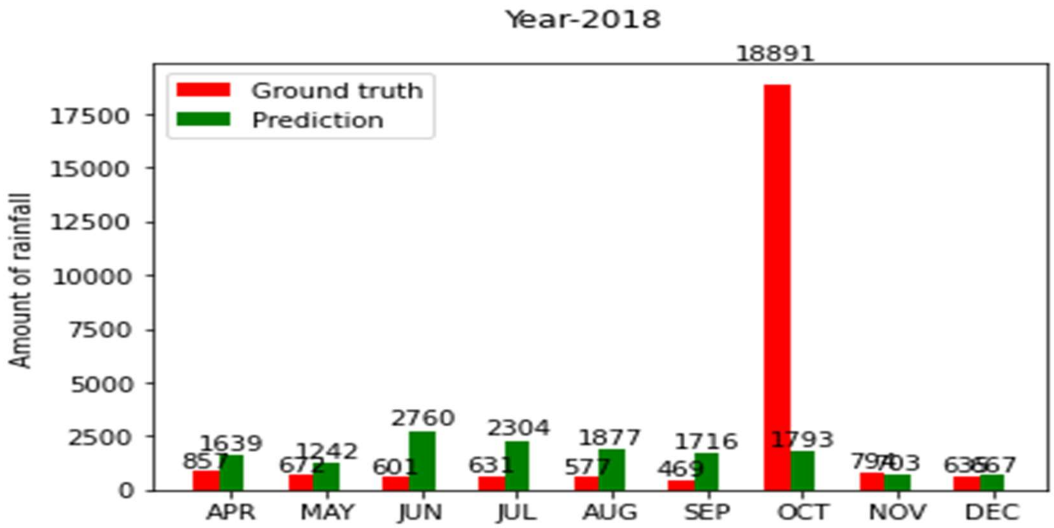
Figure 5. Predicted based on SVM Model

We also shown the amount of rainfall actually and predicted with the histogram plots. We did two types of trainings once training on complete dataset and other with training with only Aek Godang data. All means are standard deviation observations are written, first one represents ground truth, second one represents predictions.

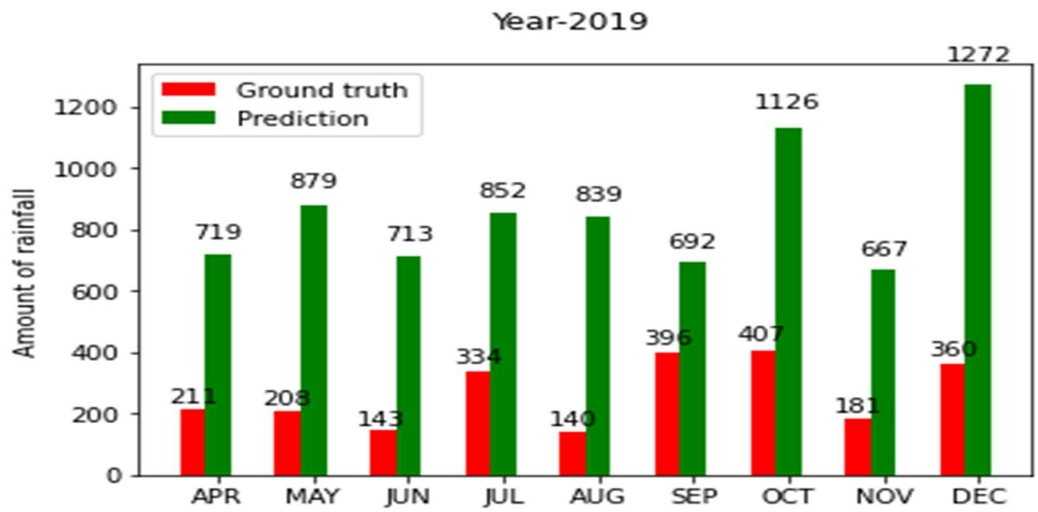
The Mean Absolute Error (MAE) based on the linear regression is 189.3450. Based on the SVM model we got the mean of 2017 is 326.0.333, Standard deviation 136.87. in 2018 the mean 2681.155 and standard deviation 5732.2387. The result of mean 2019 is 264.866 and Standard Deviation 102.56



A



B



C

Figure 6. Aek Godang prediction using SVM model

In this study, we compute the rainfall from all station in figure 5 to validate the fire weather index in situ. The red graph indicated the ground thruth and green graph is prediction. Based on the SVM model, we classify the rainfall prediction for each year. In year of 2017, the maximum amount of rainfall in April with 909 mm, and the minimum amount of rainfall is 272 in September. In year 2018, the significant amount of rainfall October with 18891 mm. and in year 2019 the maximum prediction of amount of rainfall in December with 604 mm.

Observed MAE is very high which indicates machine learning models won't work well for prediction of rainfall. Aek Godang data has a single pattern that can be learned by models, rather than learning different patterns of all states so has high accuracy. Analysed individual year rainfall patterns for 2017, 2018, 2019. Approximately close means, noticed less standard deviations.

4. CONCLUSION

Based on the result of this study, various visualizations of data are observed which helps in implementing the approaches for prediction of indicator impact of fire weather index. Prediction of amount of rainfall for both the types of dataset can be describe the parameter issue of FFMC to compute the DC as one indicator of fire weather index. Observations indicates machine learning models won’t work well for prediction of rainfall due to fluctutaions in rainfall. The result of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

predicted of Aek Godang rainfall show that the potential rainfall in future indicate more rainfall, it means that the precipitation can increase best lifeless fuel moisture greater swiftly than any other factor and Fine dead fuels react very unexpectedly to precipitation and reach their saturation factors quickly. The performance of SVM show that better model. The Aek Godang rainfall data can be learned by model, rather than learning different patterns of all states, so has high accuracy and can be used to identify the fuel moisture code to getting the good performance of fire weather index to evaluate the forest fire in North Sumatera.

REFERENCES

[1] CIFOR, "Fires in Indonesia: causes, costs and policy implications," *Fires Indones. causes, costs policy Implic.*, no. 38, 2002.

[2] P. Rosul, "Analysis of Policy Effectiveness on Forest Fires in Riau, Indonesia," no. May, 2015.

[3] I. Prasasti, R. Boer, and L. Syaufina, "Application of Cmorph Data for Forest/Land Fire Risk Prediction Model in Central Kalimantan," *Int. J. Remote Sens. Earth Sci.*, vol. 11, no. 1, p. 41, 2017.

[4] M. Elhag and S. Boteva, "Assessment of Forest Fire Rating Systems in Typical Mediterranean Forest, Crete, Greece," *Nat. Hazards Earth Syst. Sci.*, no. October, pp. 1–27, 2017.

[5] R. H. Ziel, P. A. Bieniek, U. S. Bhatt, H. Strader, T. S. Rupp, and A. York, "A comparison of fire weather indices with MODIS fire days for the natural regions of alaska," *Forests*, vol. 11, no. 5, 2020.

[6] A. P. Dimitrakopoulos, "Evaluation of the Canadian fire weather index system in an eastern Mediterranean environment," *https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/met.214*, 2010. .

[7] Nwcg, "Fire Weather Index (FWI) System | NWCG," *https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system*, 2021. .

[8] W. J. D. Groot, R. D. Field, M. A. Brady, O. Roswintiarti, and M. Mohamad, "Development of the Indonesian and Malaysian fire danger rating systems," *Mitig. Adapt. Strateg. Glob. Chang.*, vol. 12, no. 1, pp. 165–180, 2007.

[9] Sunardi, Dwiyanto, M. Sinambela, Jamaluddin, and D. Robinson Manalu, "Prediction of Domestic Passengers at Kualanamu International Airport Using Long Short Term Memory Network," vol. 4, no. 2, pp. 2–5, 2019.

[10] K. Tarigan, M. Sinambela, M. Panjaitan, P. Simangunsong, and H. K. Siburian, "Machine Learning for Waveform Spectral Analysis on Nuclear Explosion Signal and Performance of Broadband Vertical Component," *J. Phys. Conf. Ser.*, vol. 1120, no. 1, pp. 748–752, 2018.

[11] D. Ding, X. Wu, J. Ghosh, and D. Z. Pan, "extraction and classification Machine Learning based Lithographic Hotspot Detection with Critical-Feature Extraction and Classification," no. May 2014, 2009.

[12] Hartono, O. S. Sitompul, Tulus, and E. B. Nababan, "Biased support vector machine and weighted-SMOTE in handling class imbalance problem," *Int. J. Adv. Intell. Informatics*, vol. 4, no. 1, pp. 21–27, 2018.

[13] A. M. Elshewey, "Forest Fires Detection Using Machine Learning Techniques," vol. XII, no. Ix, pp. 510–517, 2020.

[14] D. R. Manalu, M. Zarlis, H. Mawengkang, and O. S. Sitompul, "Forest Fire Prediction in Northern Sumatera using Support Vector Machine Based on the Fire Weather Index," *AIRCC Publ. Corp.*, vol. 10, no. 19, pp. 187–196, 2020.

[15] N. Kerdprasop, P. Poomka, P. Chuaybamroong, and K. Kerdprasop, "Forest fire area estimation using support vector machine as an approximator," *IJCCI 2018 - Proc. 10th Int. Jt. Conf. Comput. Intell.*, no. September, pp. 269–273, 2018.

[16] AICC, "AICC - Predictive Services - Fuels / Fire Danger - FWI Defined." .