COMMUNICATIONS IN SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

Weather Forecasting using Artificial Neural Network to Improve Aviation Safety in Indonesia

Muhammad Gilang Pratama[a], Hary Fadly Yusuf[b], Heru Setiawan[c]

[a]*Department of Electrical Engineering and Information Technology, Gadjah Mada University, Yogyakarta 55281, Indonesia, muhammadgilangpratama@mail.ugm.ac.id*
[b]*Department of Electrical Engineering and Information Technology, Gadjah Mada University, Yogyakarta 55281, Indonesia, haryfadlyyusuf@mail.ugm.ac.id*
[c]*Department of Electrical Engineering and Information Technology, Gadjah Mada University, Yogyakarta 55281, Indonesia, herusetiawan@mail.ugm.ac.id*

Article history:
Received: xx xxxx xxxx / Received in revised form: xx xxxx xxxx / Accepted: xx xxxx xxxx *(to be inserted by publisher)*

Abstract

Indonesian aviation has rapidly increased since 2010 and growth almost 94 million in 2018. Air transportation has become the favorite choice of Indonesian people travel. On other hand, flight accidents still happen in Indonesia because of bad weather. So to overcome the problem above it's necessary to have a system that can make weather predictions to reduce flight accidents. Air Navigation operators can use the prediction result to recommend for pilot to improve aviation safety. The system builds using Artificial Neural Network with the dataset from BMKG. Dataset range is from 1 March 2021 to 31 Maret 2022. Artificial Neural Network implementation using python library called Scikit Learn. MLP Regressor is used as a model to optimize the value of the squared error using stochastic gradient descent as a weight optimization. Research result especially for maximum wind speed and average wind speed can be applied to make a decision for aviation operator. Based on maximum wind speed and average wind speed prediction result, it will be safe for aviation since maximum wind speed and average wind speed no higher then 4.12 m/s.

Keywords: weather forecast, artificial neural network, aviation

1. Introduction

Indonesian aviation has rapidly increased since 2010. Based on data from Badan Pusat Statistik (Central Statistics Agency) of Indonesia, the number of domestic passengers was only 48 million in 2010 and increased to 94 million in 2018 [1]. Air transportation has become the favorite choice of Indonesian people to travel across areas and abroad with nearly double growth within ten years. With its growing number of passengers, flight safety and security become necessary concerns of this modern transportation. Airplane safety features, pilot and crew capabilities, also AIRNAV, and ground support are the main factor of aviation safety. Besides that, there is the weather, another factor that has a huge impact on flight safety, and in many ways has become the highest challenge to pilots, crew, and AIRNAV because cannot be totally controlled until today.

According to the data from the Federal Aviation Administration (FAA), unfavorable weather is responsible for about 70% of aircraft delays. This will result in flight time inaccuracies, which will affect flight service users' pleasure. Flight safety is also influenced by the weather. According to the National Transportation Safety Committee (NTSC), there were 41 incidents in 2016, with 26 of them being serious, and the weather was a factor in 12.33 percent of all aviation accidents. Accurate weather data, which can be used as a reference for the flight plan, is required to solve the issues. The objective is that by referring to weather data, aviation stakeholders can ensure and improve the safety of air transportation.

Weather forecasting is an effort that has been made so far to provide an overview of the best flight routes and times for consideration by pilots and AIRNAV in ensuring flight safety. Many studies have been done on weather forecasting, some of them have enough prediction, and some others still have not correct enough. In this research, we conducted Artifical Neural Network (ANN) approach which is one of the artificial representations of the human brain which always tries to simulate the learning process of the human brain for purpose of learning weather conditions and giving a weather forecast that can be used to improve aviation safety in Indonesia. The data source that used in this research is taken from open public data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) that consists of rainfall (mm), maximum wind speed (m/s), wind direction at maximum speed (°), and the average of wind speed (m/s) closely related to the aviation safety factors.

2. Related Works

In the past few years, there are much research about improving aviation safety. Michael Schultz, et al conduct

research to predict and understand weather impact on airport performance through machine learning [2]. They use recurrent and convolutional neural networks considering weather data to know the impact on London Gatwick Airport. The weather data put from the local meteorological report and airport performance is derived from both flight plan data and report delays. The result shows that machine learning method above can quantify the correlation between decreased airport performance and the severity of local weather events.

Another research conducted by Mickael Rey, et.al. discusses safety risk identification from flight data analysis [3]. The research aims to reduce the number of flight incidents using a data-driven approach with the use of data analysis methods and machine learning tools. A boosted tree classifier was trained to classify flights as safe or risky.

Filip Skultety, et.al. conduct a research to know the effect of the dangerous weather phenomena on flight delays in Europe [4]. The main idea is to determine the impact of severe weather on air transport in Europe during the summer months. They analyze data from the Performance Review Unit (PRU) Dashboard from 2013 to 2019. The paper uses Exponential Triple Smoothing (ETS) Holt-Winters model to forecast the weather-related delays for the next 3 years. There are several recommendations from the paper such as rerouting to avoid the thunderstorm areas using a change of rerouting, different airways, and flight levels overflying or flying around known or predicted areas of dangerous meteorological phenomena.

3. Literature Review

3.1. Artificial Neural Network

Artificial Neural Network (ANN) is a technology developed from the concepts of biological neural systems. This technology simulates the brain and nervous system electrical activity that consists of perceptron as connected processing elements system. Perceptrons are arranged in a layer that its output serves as the input of the next other layers. A perceptron may be connected to all or a subset of perceptron in the subsequent layer. Data signals that are received by perceptron simulate the excitation of nerve cells then weighted and transferred the information to the network or brain. Adjustment input value from processing element i_n is multiplied by connection weight $w_{n,m}$ simulates ANN as the strengthening neural pathways to the brain. Simulation of asynchronous activity of the human nervous system in ANN theory should activate the processing elements with the input signal in an asynchronous manner[5].

ANN model simulates the human brain process with mathematics and computations. This model enriches artificial intelligence-related research such as image, recognition, voice recognition, robotics, etc. The process inside ANN model such as data collection, analysis, network structure, hidden layers, network simulation, and. weights trade-offs are computed through learning and training. ANN model is categorized as: (1) static ANN, (2) dynamic ANN, and (3) statistical ANN. Static ANN is known as multilayer perceptron neural network, dynamic ANN is known as recurrent neural network and tapped delay lines, and dynamic ANN is knowns as generalized regression neural network and radial basis function model[6].

3.2. Google Collaboratory

Google Collaboratory is a platform for disseminating research on machine learning provided by Google. Known as Google Colab, this open-source project is a browser-based tool that integrates interpreted languages, libraries, and tools for visualization like Jupyter Notebook[7]. Google Colab provides python 2 and 3 pre-configured runtimes with Machine Learning (ML) and Artificial Intelligence (AI) essential libraries such as TensorFlow, matplotlib, and Keras. This Google service provides Graphics Processing Unit (GPU) accelerated runtime on Virtual Machine (VM) hosted by Google Cloud Platform. As presented above, this cloud service is configured with the leading Artificial Intelligence (AI) libraries and a robust GPU processor.

3.3. Scikit Learn

Scikit Learn is an open-source Python library to build any machine learning algorithm [8]. It was developed by David Cornapeau in 2013 when he did Google Summer of Code. Scikit-learn can run some machine learning algorithms such as classification, clustering, prediction, regression, model selection, and preprocessing. There are several advantages of Scikit Learn: easy and free to use, a handy tool that can do multiple things, and providing complete documentation. Scikit learns work based on mathematical, statistical, and general purpose algorithms. Implementing machine learning using scikit learn still needs another library like pandas, numpy, and matplotlib.

3.4. MLP Regressor

One system that can be used to implement Neural Network is using MLP Regressor. MLP Regressor is a multilayer fast forward neural network training system that implements a multi-layer perceptron regression algorithm to solve a Multilayer Perceptron Regression Task. There is six step to implementing MLP Regressor: import the library, set up the data for the classifier, use MLP classifier and calculate the score, set up the data for the regressor, use MLP Regressor and calculate the score, and plot the model. MLP Regressor result is a number between 0 and 1. The result that is closed to 0 means that the result is better.

4. Materials and Methods

4.1. Dataset

The data source to be used is weather data taken from the official BMKG website. The data available on the BMKG website includes the value of : Wind direction at maximum speed, most wind direction, rainfall value, maximum wind speed, average wind speed average, average of humidity, maximum and minimum temperature, average temperature. In this research, we only use 4 data such as rainfall (mm), maximum wind speed (m/s), wind direction at maximum

speed (°), and wind speed average (m/s).

Table 1. BMKG Dataset

Date	Rainfall	Max Wind Speed	Wind Direction	Avg Wind Speed
01-03-2021	0	3	S	2
02-03-2021	0	3	C	1
…	…	…	…	…
30-03-2022	0	3	C	3
31-03-2022	24	2	S	2

BMKG dataset still a raw dataset so we need to do some data preprocessing activities such as remove missing value and remove outlier. How we conduct data preprocessing will be explain in implementation section.

4.2. Method

We conducted Artificial Neural Network approach which is one of the artificial representations of the human brain which always tries to simulate the learning process of the human brain. ANN was developed as a generalization of mathematical models of human comprehension (human cognition) that are based on the idea that information processing takes place in simple units known as neurons. Signals go between nerve cells via connecting junctions, each of which has a corresponding weight, and each nerve cell is an activation function of the weighted summation signal that enters it to determine its output signal.

Any patterns of input and output information given to the ANN is processed in the neuron. These neurons gathered in the neuron layers. The layers that make up the ANN can be divided into 3, namely:

- The input layer, the units in the input layer are called input units. These input units receive data patterns from the outside that describe a problem.
- Hidden layer, the units in the hidden layer are called hidden units. Where the output cannot be directly observed.

Output layer, the units in the output layer are called output units. Output from this layer is an ANN solution to a problem.

5. Implementation

The implementation of the case discussed is using the Multi Layer Perceptron (MLP) Regressor. The model we use optimizes the value of the squared error using stochastic gradient descent as a weight optimization. We also use "sklearn neural network with MLP Regressor" python library. The data set that we use has been preprocessed first by deleting or changing the value to 0 for the missing values. We perform tests on each of the parameters, to get the best results.

5.1. Defines the Function and Dataset

At figure below we use pandas, matplotlib.pyplot, and numpy as library function for running scenario. We import iklim_data.csv from BMKG dataset and this data still has missing values stored as NaN. Dataset is preprocessed from missing value at the next step.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import t
np.random.seed(11)

df = pd.read_csv("iklim_data.csv",header=0, sep=";", encoding='cp1252')
df.head()
```

	Tanggal	Curah Hujan (mm)	Kecepatan angin maksimum (m/s)	Arah angin saat kecepatan maksimum (°)	Kecepatan angin rata-rata (m/s)
0	01/03/2021	NaN	3	270	2
1	02/03/2021	0.0	3	200	1
2	03/03/2021	0.0	2	200	1
3	04/03/2021	2.2	3	260	1
4	05/03/2021	NaN	3	200	2

Fig. 1. Defines the Function and Dataset

5.2. Preprocessing

```
[4] iklim.isnull().sum()

Tanggal             0
Curah Hujan (mm)    76
dtype: int64

[5] iklim.fillna(0,inplace=True)
    iklim.isnull().sum()

/usr/local/lib/python3.7/dist-packages/pandas/core/frame.py:5182: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  downcast=downcast,
Tanggal             0
Curah Hujan (mm)    0
dtype: int64
```

Fig. 2. Preprocessing

Dataset with missing value can destroy the whole process of data analysis. Null values pose the biggest risk when they are too massive and changing the whole data statictically. The data set that we use has some missing value as NaN that come from real field dataset record. This dataset the preprocessed by deleting or changing the value to 0 for the missing values.

5.3. Cross Validation

```
df_train_ch = iklim[["Curah Hujan (mm)"]]
df_train_ch = df_train_ch.rename({"Curah Hujan (mm)" : "x1"}, axis=1)
df_train_ch["x2"] = df_train_ch["x1"].shift(-1)
df_train_ch["x3"] = df_train_ch["x2"].shift(-1)
df_train_ch["x4"] = df_train_ch["x3"].shift(-1)
df_train_ch["x5"] = df_train_ch["x4"].shift(-1)
df_train_ch["y"] = df_train_ch["x5"].shift(-1)
df_train_ch.head()
```

	x1	x2	x3	x4	x5	y
0	0.0	0.0	0.0	2.2	0.0	0.0
1	0.0	0.0	2.2	0.0	0.0	17.1
2	0.0	2.2	0.0	0.0	17.1	14.3
3	2.2	0.0	0.0	17.1	14.3	12.8
4	0.0	0.0	17.1	14.3	12.8	14.6

Fig. 3. Cross Validation

Cross-validation process resampling different portion of the data to test and traing model on different iteration. This method is mainly used in prediction or estimation how accurate a predictive model will perform.

5.4. Import sklearn Library and Defines Number of Data Train and Data Testing

```
[11] from sklearn.neural_network import MLPRegressor
     from sklearn.datasets import make_regression
     from sklearn.model_selection import train_test_split
     X, y = df_train_ch[["x1","x2","x3","x4","x5"]],df_train_ch["y"]

[12] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=100, random_state=11)
```

Fig. 4. Import sklearn Library and Defines Number of Data Train and Data Testing

At the next step we implement Multi Layer Perceptron (MLP) Regressor by importing MLPRegressor from sklearn.neural_network. We use this method to optimizes the value of the squared error using stochastic gradient descent as a weight optimization.

5.5. Make Weather Prediction

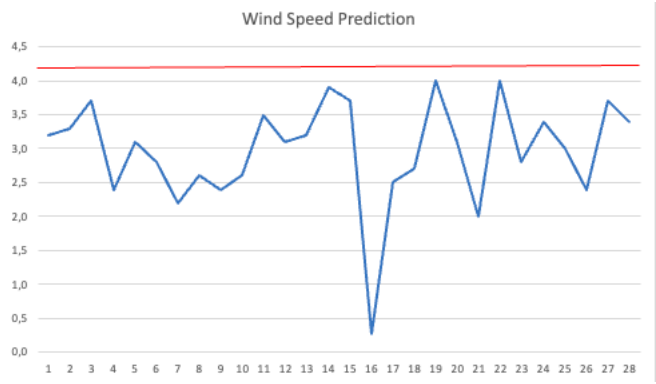

Fig. 5. Maximum Wind Speed Prediction

After Artificial Neural Nework model has been created, then we use it to make a prediction for maximum wind speed and average wind speed. Prediction result will show in dashboard, so Aviation Operator can understand easily.

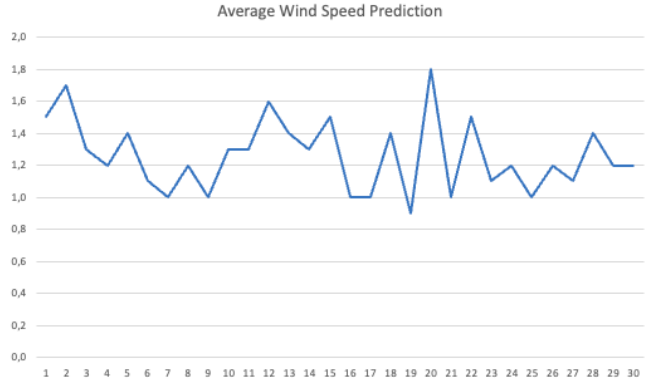

Fig. 6. Average Wind Speed Prediction

Maximum wind speed is one of parameters to decide plane can depart or not. Based on data, flight can depart if maximum wind speed is lower than 4.12 m/s. Wind speed prediction result above is never more than 4.12 m/s so it will be safe for plan to depart.

6. Results and Discussion

The results of the experiments that we carried out on the four parameters that we used had different accuracy. In the first test, the rainfall parameter has an RMSE value of 10.0248 where this value is very high so that the resulting accuracy is also very low. The regression value obtained is also very low until it touches the minus number at -0.98249. The second test, on the maximum wind speed parameter, the accuracy results are much better than the rainfall parameter, which has an RMSE of 0.8866 and a regression score of 0.109. The next test, the wind direction parameter at maximum speed, has a very poor accuracy value with RMSE touching 65.528 and a regression value of -0.2488. And in the last test, the average wind speed parameter has the best accuracy value compared to the previous 3 parameters. The RMSE score is 0.57688, and the regression score is 0.024.

Based on the dataset used and the results of the test, the rainfall parameters and wind direction parameters at maximum speed cannot be relied upon to be used as benchmarks for weather predictions, which was proven by the lack of accuracy in both parameters. Meanwhile, the parameters of maximum wind speed and average wind speed have the possibility to be relied on. With more varied adjustments, there is still room to improve the accuracy of both. In its influence in the world of aviation, wind conditions are sometimes one of the crucial things.

VII. CONCLUSION

From the subject matter we researched regarding the impact of weather on aviation safety, the solution we provide has the possibility to be applied in the real world of aviation with some important notes.

The results of our experiments, the parameters we use are not all reliable, and the next two parameters have little room to develop and have the possibility to be applied. More and richer data is also needed to further determine the reliability of the methods we apply (e.g. real time weather data). Real time weather data is important and very necessary when it comes to flight safety, because it can help pilots and air traffic controllers (ATC) in making flight planning decisions to determining the coordinates of aircraft maneuvers.

As a suggestion for future development, testing the configuration of the multi layer perceptron method is more numerous and varied so that it has a greater possibility of getting higher accuracy and much lower error.

References

1. Badan Pusat Statistik, “Jumlah Penumpang Pada Keberangkatan di Indonesia.” 2018. [Online]. Available: https://www.bps.go.id/indicator/17/655/1/jumlah-penumpang-pada-keberangkatan-di-bandara-indonesia.html
2. M. Schultz, S. Reitmann, and S. Alam, “Predictive classification and understanding of weather impact on airport performance through machine learning,” Transp. Res. Part C Emerg. Technol., vol. 131, p. 103119, Oct. 2021, doi: 10.1016/j.trc.2021.103119.
3. M. Rey, D. Aloise, F. Soumis, and R. Pieugueu, “A data-driven model for safety risk identification from flight data analysis,” Transp. Eng., vol. 5, p. 100087, Sep. 2021, doi: 10.1016/j.treng.2021.100087.
4. F. Škultéty, M. Jarošová, and J. Rostáš, “Dangerous weather phenomena and their effect on en-route flight delays in Europe,” Transp. Res. Procedia, vol. 59, pp. 174–182, 2021, doi: 10.1016/j.trpro.2021.11.109.
5. S. Walczak and N. Cerpa, “Artificial Neural Networks,” in Encyclopedia of Physical Science and Technology, Elsevier, 2003, pp. 631–645. doi: 10.1016/B0-12-227410-5/00837-1.
6. A. Malekian and N. Chitsaz, “Concepts, procedures, and applications of artificial neural network models in streamflow forecasting,” in Advances in Streamflow Forecasting, Elsevier, 2021, pp. 115–147. doi:

10.1016/B978-0-12-820673-7.00003-2.

7. T. Carneiro, R. V. Medeiros Da Nobrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho, “Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications,” IEEE Access, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.
8. R. Garreta and G. Moncecchi, Learning scikit-learn : Machine Learning in Python. Bimbingham: Packt Publisher, 2013.