WEATHER PREDICTION USING NEURAL NETWORK

Leon Franciskus Anggoro[1], Edward Effenlie[2], Harura Cendekia Hasanasultoni[3]
[1,2,3]Bina Nusantara University, Kemanggisan, School of Computer Science, Indonesia
David David S.Kom, M.T.I.[4], PANJI ARISAPUTRA, S.Kom, M.Kom.[5]
[4,5]Binus Undergraduate Program - Doctor of Computer Science

ABSTRACT

Good weather is mandatory to carry out the daily needs of humans to work efficiently and safely. Predicting the weather is one of the most studied and researched topics in order to study how weather works on our planet. To accurately predict the weather, scientists use a math based model to predict the weather. We propose multiple machine learning algorithms that can predict the weather faster and more reliable than the current math based models. In this search we use Recurrent Neural Network, Decision Tree, and K-nearest neighbour. The data we used was collected from an open source website called Kaggle, The data collected was processed so the models can use the data. The model that can predict the weather accurately is the KNN model with an accuracy 68.51%.

1 INTRODUCTION

In carrying out daily activities, humans constantly need information about weather conditions, both short-term weather information and long-term weather. In its application, a dynamic weather pattern is a challenge that we constantly faced in providing weather information that is fast and accurate.

Weather prediction or weather forecasting is the application of science and technology to predict the conditions of the atmosphere for a certain place and time. Weather prediction is a very complex process that usually involves collecting data from various sources such as satellite, radar, and weather balloon, followed by the analysis of this data using mathematical models.

Accurate forecasting and prediction of weather is still a major challenge for the scientific community[1], especially for tropical regions like Indonesia. There has been a surge in climate change-induced extreme weather events such as floods, droughts, and storms in the past decade (2010-2020)[2].

Accurate weather prediction is an important part of modern society. Accurate prediction can save lives and minimise losses from extreme weather events[3]. With improvements made to the warning system, we can reduce the extent of damage by improving the resilience of the at-risk population and reinforcing the preparedness to cope with these extreme weather events. Weather predictions are also incredibly crucial for other fields like agriculture, allowing farmers to track when is the best time for planting and help them to reduce their vulnerability to climate change risk[2].

The need for more accurate weather prediction has been further emphasised in recent years, as climate changes have led to more frequent severe weather events. Climate change is altering the global climate system, leading to changes in temperature, precipitation, and other weather-related variables. These challenges make it all that more challenging to accurately predict the weather , further highlighting the need for continued research and innovation in this field.

This paper aims to highlight the usage of machine learning models for weather prediction tasks. The methods that our models will be using are K-nearest neighbors, Decision tree, and Neural Network. The purpose of using multiple methods to give a base accuracy, time complexity, and efficiency for each method and compare them to each other.

2 LITERATURE REVIEW

The paper “Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere” by Jonathan A. Weyn, Dale R. Durran, Rich Caruana [4] was conducted to improve efficiency and cutting the cost on the expensive numerical weather prediction model by using machine learning. The data from European Centre for Medium-Range Weather Forecasts (ECMWF) was used and pre-process into inputs the model can use. The algorithm used was Convolutional Neural Network (CNN) [4]. The attribute used by the model geopotential height (sea-level height), geopotential thickness, and temperature [4]. The authors conclude the usage of CNN is much more efficient and less expensive than the numerical model in terms of forecasting and disaster management.

In the paper “A Novel Approach for Weather Forecasting using Machine Learning Techniques" by B.S.Panda, D.Maneesha, P.V.Lasyasri , P.Goutham, K.Suresh, Ch.Pranavsankar [5], the authors used machine learning models to improve weather forecasting because Numerical Weather Prediction (NWP) is a human based process that needs experts to observed the data and interpret it. The algorithms used by the authors were linear regression and deep neural network [5]. The data used in the model was data mined to reduce the cost of manually collecting the data and pre-process into inputs [5]. The conclusion of this paper is that the proposed models exceed the current NWP in accuracy and efficiency with less human intervention.

In the paper “Predictive Analytics In Weather Forecasting Using Machine Learning Algorithms” by Mohd Wazih Ahmad, T.Gopi Krishna [6], the author used machine learning to predict the weather to increase agriculture efficiency output in India. The algorithms used for the models were linear regression and decision tree [6]. The data was collected by using data mining and pre-process into inputs the models can use. The results from the two models were fairly similar with decision tree leading if more features were added.

In the paper ”Weather Forecasting using Neural Network” by Priyanka Mahajan, Chhaya Nawale, Siddheshwar Kini, Prof. Krishnanjali Shinde [7], the author proposed a machine learning model to improve efficiency of weather forecasting while comparing it to statistical weather prediction models. The algorithm used was neural network (backward propagation). The data was mined from official weather gathering websites to decrease cost. The author concludes the usage of machine learning based on neural networks is worth developing with higher accuracy than the statistical model.

With many different methods of approach to predict the weather of our planet, these related work papers helped us identify each of the model's strengths and weaknesses. Various algorithms have their own complexity and challenges in making them. The most common algorithm that we found in these papers for weather prediction was neural network and linear regression-based models with them being reliable and efficient in weather forecasting.

3 METHODOLOGY

3.1 Dataset

This paper uses an open-source dataset for weather prediction provided from Kaggle by Greeg Titan and Elmajo Adriel. The dataset contains about 590k data in row and separated by 12 columns which are date, min temperature, max temperature, avg temperature, avg humidity, rainfall, duration of sunshine, max wind speed, wind direction at maximum speed, avg wind speed, most wind direction, station id which record the data. In this research, min temp, max temp, max wind speed, avg humidity are used as input to train the weather prediction model. According to our findings, 80% of the data will be used for training and 20% of the data will be used for testing purposes because it can provide better results without the model. Below is the correlation between each parameters

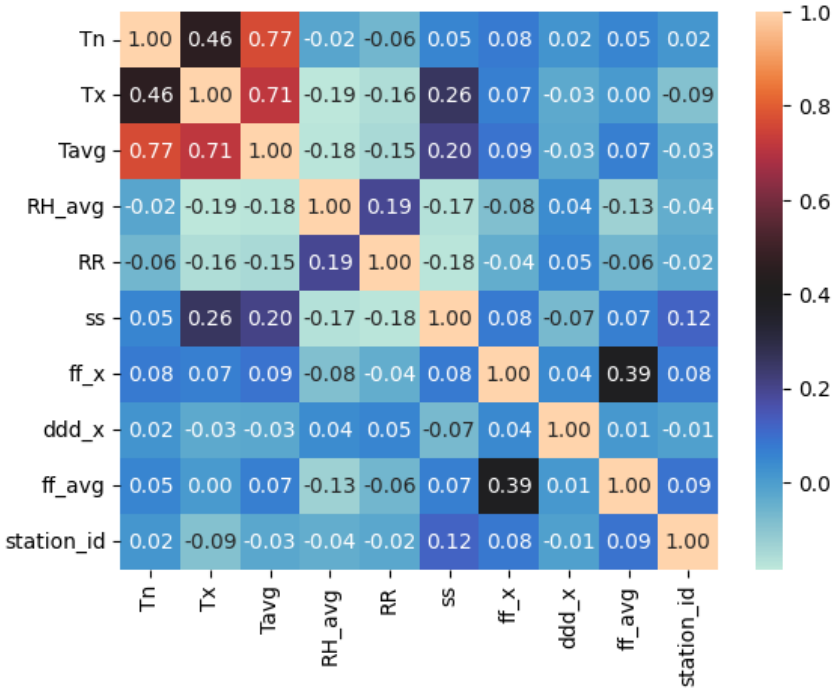

Fig. 1 Correlation Table

3.2 Pre-processing

The data used in this paper are noisy. It is important to pre-process the data that we need, converting raw data into clean data to help improve the accuracy of our model. Such as minimising the outliers based on our prior knowledge about weather, scaling the minimum value and the maximum value into 0 until 1 using Minmax Scaler. We noticed that our data has about 12% of NULL data. We have 590k worth of data so we can remove those NULL values without causing a decrease of accuracy.

3.3 Machine Learning Algorithms

3.3.1 K-Nearest Neighbour

K-Nearest Neighbour (KNN) is a supervised machine learning algorithm that is used for classification and regression problems, but usually used for classification problems, perfect for our weather prediction model. KNN predicts the testing data by calculating the distance between both the testing data and the training data to identify the nearest neighbour of a given point[8]. We develop KNN using the sci-kit learn package in Python with the total of 5 neighbours (n), with the other settings set to default.

3.3.2 Decision Tree

Decision Tree is a supervised machine learning algorithm that is used for classification problems [6]. Decision tree accepts an input as yes or no, meaning it only can produce a 0 or 1 output. It is a very powerful machine learning algorithm that can be improved with increasing the number of nodes of each leaf to increase the accuracy of the model. By using a python library called sci-kit learn, we can easily implement this algorithm into a classification model.

3.3.3 Recurrent Neural Network

Recurrent Neural Network(RNN) is part of the deep learning models that can be used for classification and regression. RNN uses many nodes to communicate with each other based on weighted values. These nodes are organized in layers such as input layers, hidden layers, and output layers based on the weighted values that have been assigned to each layer[9]. We used a python library called keras to implement the input layers, hidden layers, and the output layer with various values with an optimizer called Adam.

4 DISCUSSION AND RESULTS

Table 1. Accuracy of the models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
RNN	28.71	26.34	13.18	17.57
Decision Tree	63.37	69	66.19	67.57
KNN	68.51	71.5	75.45	73.42

The experiment was done using the Python language version 3.11.3 kernel, we use the untrained model from sci-kit learn and keras. The model evaluation result was taken using metrics library from sci-kit learn. In this experiment, the performance metrics that we used are accuracy, precision, recall, and F1 score. These criteria were used to test whether the models performance is dependable enough for predicting weather in the real world.

From table 1, we can see that RNN has the lowest accuracy when compared to the two other models while KNN achieves the highest accuracy with 68.51% accuracy. The second highest is Decision Tree at only 5% behind the performance of KNN at 63.37%. Accuracy is the most straightforward of the 4 metrics used in the evaluation where it measures the overall correctness of predictions made by the classification model

Precision is the model's value when it comes to predicting positive instances out of all actual positive instances. The Higher precision means the lower false positive in the model's to predict the actual outcome. The model with the highest precision score is KNN with a score of 71.5% while the second highest is the decision tree model with a precision score of 69%.

Recall score is the ability of the model's to trace back and recall the same pattern in the training set. The higher the recall score, the less likely of the model making a wrong prediction. Based on table 1 data, out of the 3 models the one score the lowest is RNN with an recall score of 13.18% while KNN reached a score of 75.45% scoring the highest to predict the positive instances.

F1 score is the combination of the recall score and precision of the model and is the most used and useful evaluation metric, unlike the accuracy metric, F1 score can also be used for an imbalance dataset. In our case we had an imbalance in our dataset where the negative label for weather prediction outweighs the number of positive instances by a huge margin. From table 1 we can see that overall KNN has the highest performance in predicting weather instances. The model that underperforms is RNN ranking last out of the three models.

Here are some Confusion Matrices for each model that predicts the outcome.

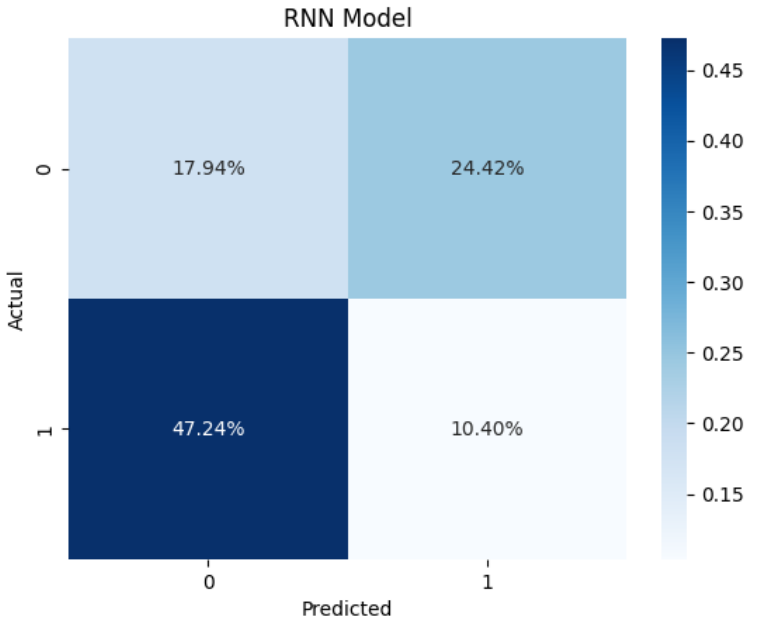

Fig 2. RNN Model

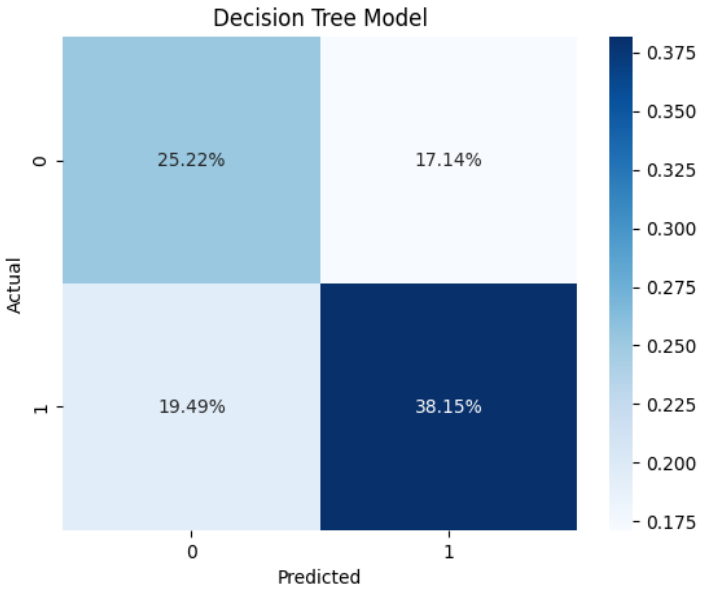

Fig 3. Decision Tree Model

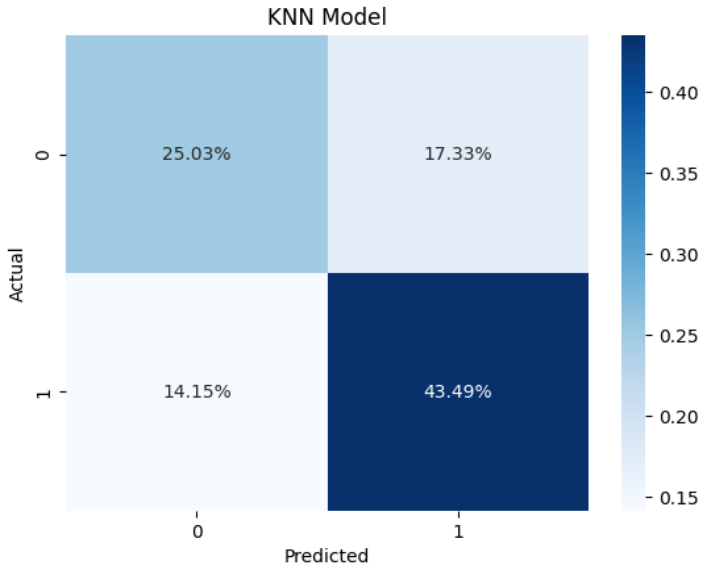

Fig 4. KNN Model

5 CONCLUSION

We believe that there might be several reasons for the low performance of our RNN model. The first is that we believe that our dataset might not be compatible with our model. Based on Fig. 1, the data has low correlation between each other, meaning it could lead to an imbalance when training the models..The second factor that might contribute to this is the lack of hypertuning/features of the RNN model. When we compare our model with others, they have higher accuracy than ours. It means that adding more layers and features to the current RNN model could lead to a higher accuracy. This also applies to the decision tree model, the current decision tree only has basic features from sci-kit learn library, meaning it also can score higher than KNN.

In conclusion, picking the right dataset with high correlation between each data and improving the basic models with hypertuning each model can increase its reliability when predicting the weather of our world.

6 REFERENCE

1. https://www.researchgate.net/profile/Jugal-Kishore-Mani/publication/311225409_Accuracy_of_Weather_Forecast_for_Hill_Zone_of_West_Bengal_for_Better_Agricultural_Management_Practices/links/583fa31208ae61f75dc79837/Accuracy-of-Weather-Forecast-for-Hill-Zone-of-West-Bengal-for-Better-Agricultural-Management-Practices.pdf
2. https://www.frontiersin.org/articles/10.3389/fenvs.2022.935696/full
3. https://journals.ametsoc.org/configurable/content/journals$002feint$002f19$002f17$002fei-d-14-0032.1.xml?t%3Aac=journals%24002feint%24002f19%24002f17%24002fei-d-14-0032.1.xml#d115794919e194
4. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020MS002109
5. http://csjournals.com/IJCSC/PDF11-2/5.%20Panda.pdf
6. https://eudl.eu/pdf/10.4108/eai.7-12-2018.159405
7. https://www.ijert.org/weather-forecasting-using-neural-network
8. https://www.researchgate.net/publication/360365118_Smart_Weather_Prediction_Using_Machine_Learning
9. https://www.researchgate.net/publication/367072321_Frost_prediction_using_machine_learning_and_deep_neural_network_models