Use Machine Learning to Classify Weather Opinions from WhatsApp Group: A Model Design and Evaluation for Raining Event Prediction

Pou-Chang Chen,
Nogle Studio,
Nogle Taiwan Limited

Abstract

Machine learning technology can build a prediction model without much objective analysis of the weather situation from social media opinions. Social media have become one of reliable channels for people to receive weather information. In spite of there being much irrelevant information on social media, researchers found the weather signals can be extracted from these noises. These weather signals from social media are not all from educated meteorological forecasters but also from untrained masses as their weather opinions. Since the weather opinions from the untrained masses can be different from the reports of weather forecast offices, a different method should be proposed to find the pattern of the weather opinions. Machine learning has been used for weather predictions and climate models in weather forecast offices which forecasted the analyzed situations, but how can the unidentified weather situations from untrained masses be predicted? By labeling the targeted weather situations from social media, this research used machine learning algorithms to create a weather classification model.

In this research, the machine learning model for weather classification was to predict the weather opinions of the raining events from WhatsApp Group. Based on a convolutional neural network algorithm (CNN), the model utilized features from weather research and forecasting model (WRF) outputs and predicted raining events from social media opinions. The features included basic weather factors, cloud fractions, and probabilities of precipitation (POP), and an experiment to have different combinations of the features was discussed. According to the precision-recall curve analysis, this research found the cloud feature provided contribution to short term showers and POP with larger area weather information contributed to the longer term rain.

Key words: machine learning, social media, weather opinion, multi-class classification

I. Introduction

Machine learning (ML) has been applied to weather prediction and climate model in different topics. Based on the literature review of Bochenek and Ustrnul (2022) for ML applications in weather and climate researches, the most frequent research topics were wind prediction and model improvements (including weather ensemble prediction models and climate models). Researches have used ML algorithms to forecast the future weather patterns and used the patterns as feedback to improve numeric weather prediction models (NWP).

As a data-driven technology, ML can make the forecast without knowing the details of the weather situations. Supervised learning is a branch of ML models to make prediction with labels (prediction targets). For example, a cat-dog recognition model uses cat and dog images as the training input and labeled with “cat” or “dog”. Similarly in meteorological domain, without the analysis of weather situations, Roser and Moosmann (2008) used a ML algorithm to process images to classify different weather types, such as “clear”, “light rain”, and “heavy rain”. Especially considering deep learning model (DL), which is a branch of ML, Reichstein et al. (2019) suggested that DL may include both spatial and temporal information to detect rapid change extreme events. Although deep learning models (DL) are unlike NWP systems to include physical laws, Schultz et at. (2021) mentioned few studies which have involved physical constraints to let machine learning models “understand” physical laws. Furthermore, they thought DL models could provide more consistent forecast products than NWP models in which the workflow might include post-process data correction.

Social media has become one of primary sources for weather information, the abundant information from social media has been validated as an indicator for weather communication even if there are untrained masses’ opinions. The social media like Facebook and Twitter have become more and more reliable weather communication channels (Silva et al. 2013) even if the communication quality might be questionable (Eachus and Keim 2019). Silva and her colleagues interviewed professional weather forecasters to review the problems of weather reports from social media

(Silva et al. 2013), and they found the forecasters' major concerns were the accuracy and consistency. Because not all people had professional training for the weather reporting, the untrained couldn't measure the weather factors systematically as meteorological forecasters and identify the same weather situation in a consistent way. Thus, this research distinguished the weather communication terminologies into "weather opinions" which were from the untrained masses and "weather reports" which were created by weather forecast offices. The depiction of the idea is displayed in Figure 1. Despite of the weather opinions from untrained masses might be different from professional weather forecast offices, Silva and her colleagues found there were weather signals could be extracted from the noises of social media information. Ripberger et al. (2014) used statistical analysis for Tweets as a tornado indicator and found the amount of Tweets following the trend of tornados. Their work validated social media information as weather communication (in their Fig. 2) though not the weather situation itself.

When the weather communication can be predicted, it is possible to predict the unidentified weather situations before analysis. Doswell (2004) thought the untrained forecasters still relied on their experience and intuition instead of educational knowledge to make decisions, so that meant the weather opinions from the untrained masses could be different (might include error or bias) from educated weather forecasters. In this case, to predict the weather opinions from social media can't rely on traditional weather prediction approach. A traditional weather forecast product has many components to develop. Schultz et at. (2021) explained a NWP workflow involving data assimilation, model ensemble, physical parameterization etc. processes after the synoptic weather patterns were analyzed. In contrast, a DL model leveraged observation data to generate end user forecast product. Because ML model is a data-driven approach, it is possible to find the pattern from input data without knowing the weather situation details. Thus, this research utilized machine learning technology to use weather forecast model data for predicting weather situations from social media.

Different from the hybrid NWP-ML workflow in Schultz et at. (2021) to use DL in assimilation and prediction steps, this research introduced ML algorithm after NWP outputs have been created. This approach is trying to use ML to emulate human forecasters to interpret NWP forecast products. Using WRF outputs as features and weather information from social media, i.e. the messages from WhatsApp group, as labels, this research created a ML classification model to predict weather situations. The machine learning experiment configuration is introduced in section II. It explains how the opinions from social media were gathered as labels and describes the features from weather forecasting model. Section III introduces the result of experiment and discussed the evaluation requests to identify the different contributions from different weather features. Finally, the section IV says the conclusion to summarize the experiment and limitation of the research.

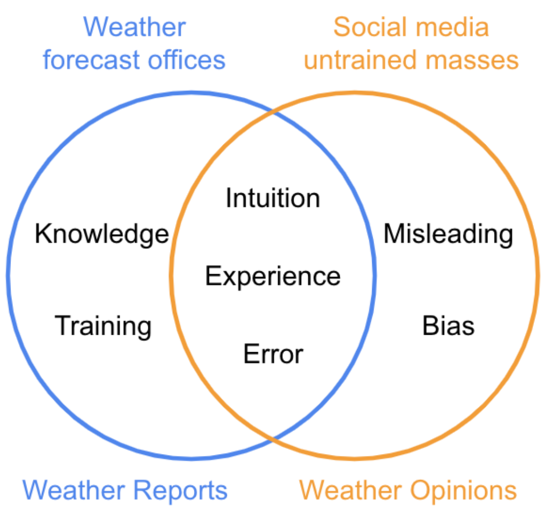

Figure 1. The graph of weather communication for Weather Reports and Weather Opinions. Even if Weather Reporters from Forecast offices include knowledge and training, intuition and experience are still used (Doswell 2004) and there might be forecast error. The weather opinions from untrained masses without knowledge and training, they can cause misleading and bias information (Silva et al. 2013).

II. Experiment Methodology

II.1 Model Design

In order to predict the weather opinions from social media, this research designed a weather classification model to predict raining events. The model includes 3 types of features for consisting of the raining environment: basic weather factors, cloud fractions, and probabilities of precipitation (POP). The first two groups were based on the city areas of social media groups, i.e. local small scale weather, and POP would include larger area weather information. By analyzing the different combinations of these features as model inputs, this research would discuss how the different features affecting the raining events. The structure of model framework is showed in Figure 2.

A supervised machine learning model was used to make a multi-class classification for the weather classification model. Supervised model is a model which has predict labels as prediction targets and uses features as model training input. For example, if we want to identify the future stock is going up or down, we use the historical market trend results as target labels and use the market price time series as input features to train a prediction model. When the target labels have more than two classes, the prediction model is a multi-class classification model.

To classify the raining events, a 1D-CNN model was applied to train the features from WRF model output and probability of precipitation (POP) data and provided with labels from social media group. This research used 2021 April to November, which is summer monsoon season (rainy season) in Southeast Asia (Takahashi and Yasunari 2006), data set for training and testing.

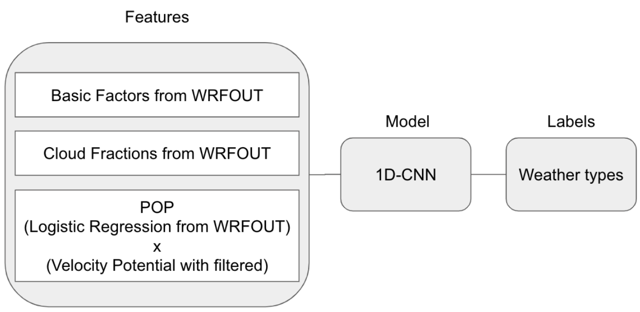

Figure 2. The feature and labels for 1D-CNN model are used in this research

II.2 Model Labels

The label data in this research was from one of clients WhatsApp group where the weather status was reported when the weather situations happened. There are several "flag members" would announce the rain starts and decide the watching session ends. Based on the group members reports, the weather situations were categorized into ["Clear", "Light Rain", "Medium Rain", "Heavy Rain"] 4 weather types. The following are the definition of each weather type:

- Clear: There was no "Rain" announcement in the session,
- Light Rain: There was one "Rain" announcement in the the session,
- Medium Rain: There were more than one "Rain" announcements in the session, and
- Heavy Rain: There was a "Super Rain" announcement in the session. When the "Super Rain" appeared, there would be no more announcement in the session.

The "Rain" and "Super Rain" announcements were preprocessed. The message texts from the messaging group were extracted, and the rain related keyword, such as "Rain", "Little Rain", "Medium Rain" were counted. The flag members would decide whether the rain stopped or started. Thus, the differences between "Light Rain" and "Medium Rain" were the times of rain not the amount of rain. In contrast, the "Heavy Rain" threshold was the amount of rain, i.e. it was announced "Rain" when the "flag members" determined enough.

This weather type labeling was different from classical quantitative precipitation estimation (QPE). Instead, it was similar to weather reporters announcement: "Clear", "Cloudy", "Light Rain" etc. Figure 3. presents the accumulated precipitation bar charts for different weather type cases which shows the variance presents in the same type of cases. Fig.1a and Fig.1d are Light Rain cases. Although they are both one Rain announcements (Little Rain), Fig.1d looks that there are two non-continuous rain events. In Medium Rain cases, Fig.1b has obvious two

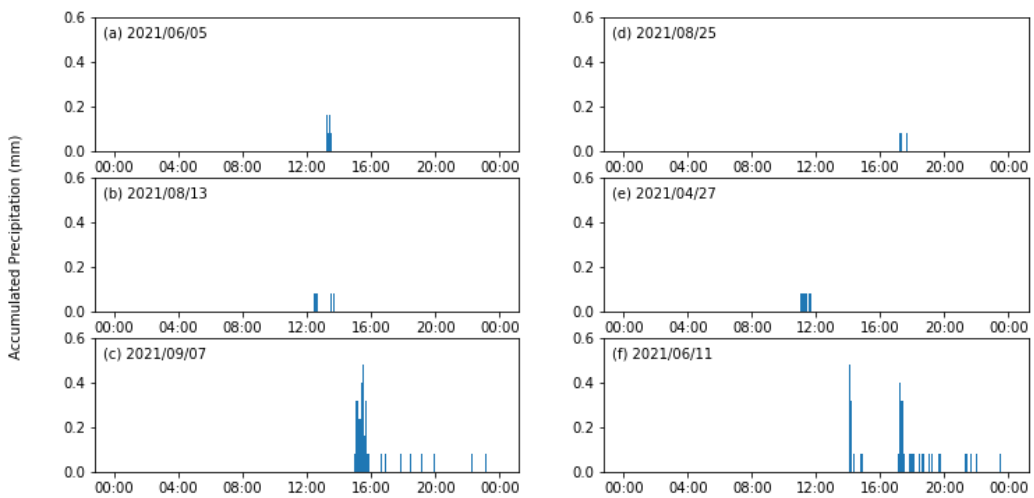

Figure 3. The accumulated precipitation bar charts for different weather types: Light Rain cases are (a) and (d), Medium Rain cases are (b) and (e), and Large Rain cases are (c) and (f).

events, but Fig.1e seems not clear. The gap length between rains can't be determined different weather types. The Heavy Rain case in Fig.1c shows a heavy rain at first and then multiple scatter rains later, but Fig.1f case has separated larger rain events. It seems there is no obvious rain pattern in the same weather types. Thus, this research tried to leverage ML technology to classify these different weather types based on the human announcements.

II.3 Model Features

Feature data were from WRF output and POP forecast data. The location of WRF output data was from the grid where the city of messaging group users was at. For the temporal resolution, the WRF outputs were generated every 3 hours, so there were 8 hourly time series data for each day. Every time series had 25 data points. The depiction of data set is in Figure 4. Because the client members usually used "cloud" as their forecast index, so this research included the cloud top temperate (ctt) and cloud fractions for the features besides the general weather factors. The output feature variables shows in Table 2. POP forecast data included short and long trend signals which would be discussed in Section II.2.B.

Feature Type	Variable Name	Description	Unit
Basic	tk	Temperature on 850hpa	K
Basic	ua	U-component of Wind on 850hpa	m s**-1
Basic	va	V-component of Wind on 850hpa	m s**-1
Basic	rh	Relative Humidity on 850hpa	%
Basic	pw	Precipitable Water	kg m**-2
Basic	slp	Sea Level Pressure	hPa
Cloud	ctt	Cloud Top Temperature	K
Cloud	low_cloudfrac	Low level cloud fraction	%
Cloud	mid_cloudfrac	Mid level cloud fraction	%
Cloud	high_cloudfrac	High level cloud fraction	%
POP	pop	Probability of Precipitation	%

Table 1. The variables were used for model features.

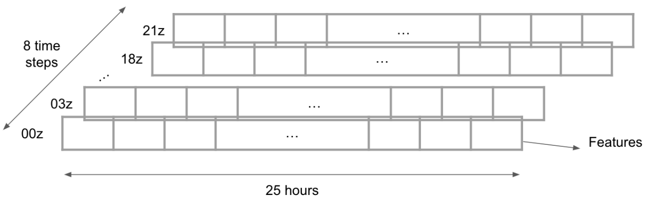

Figure 4. The graph of features data structure

II.3.A WRF Model

Weather and Research Forecasting model (WRF) is an open source weather forecasting model published in 1990's. Because WRF model output includes cloud forecasting, it can provide the effective features for the training model. The configuration information is in Table 2.

II.3.A.a Model Configurations

The WRF model was version 4.0 and run one domain (d01) with 74x62 horizontal grid points at 30 km. On the vertical coordinate, 33 vertical levels were used. Physics parameterization was NCAR Convection-Permitting Suite. The running horizontal domain was covering the Mainland Southeast Asia and indicated at Figure 5. centered at 12 'N and 105'E.

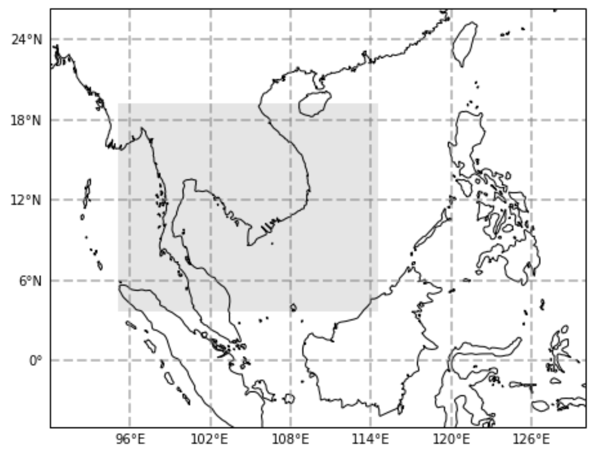

Figure 5. WRF model d01 domain area

II.3.A.b Initial and Boundary Conditions

The WRF model used NOAA GFS 0.5x0.5 degree data for initial and lateral boundary conditions. GFS is a global operational weather analysis data run by U.S. National Weather Service, and it provides for 120 hours forecasting every 6 hours. The horizontal resolution was about 28 km, vertical coordination was 64 levels (127 levels since 2021 February), and temporal resolution was 3 hour.

Configuration	Description
Input data	NOAA GFS (6-hour, 0.5′x0.5′)
Nesting	N/A
Domain	Domain 1: 30 km (74x62)
Vertical layer	33 levels
Physics parameterization	NCAR Convection-Permitting Suite

Table 2. The WRF configuration information

II.3.B POP (Probability of Precipitation) Model

The POP model created in this research was an unconventional approach to consider both smaller and medium weather system. For small scale, a Logistics Regression model was used for single grid tendency. For medium scale, potential function anomaly was considered to include broader information. Then the both scale information was combined into a POP value depending on the anomaly scale with the following equation:

$$POP = ShortTermPOP * LongTermPOP$$
$$= LogisticsRegression * PotentialFunction$$

When both small and medium scale weather system were dominant, POP will be large. If either one of them was not important, the POP would decrease.

II.3.B.a Logistics Regression

Based on WRF output data as features and GPM calibrated precipitation as label, an logistics regression model was used for training small scale POP. This research used GPM (Global Precipitation Measurement) satellite data as the rainfall label. GPM Mission is an international satellite data network for rainfall and snow every 3 hours. The GPM IMERG Late Precipitation L3 product was used and it had 0.1x0.1 degree spatial resolution. In order to integrate with WRF output features, bicubic interpolation was used for resizing GPM data into WRF output grid. For model features, the basic WRF outputs are listed in Table 1. and the output data were took at the grid of the messaging group city. The training dataset was in the period from 2009 April to 2020 November.

Logistics Regression use linear regression to classify the rainfall of the grid and logistic function is:

$$f(z) = \frac{1}{1 - e^{-x}}$$

The result of classification would be transformed into probability between 0 and 1.

II.3.B.b Potential Function

Since the Logistics Regression model can only present the daily pattern (rain in the night which is not showed) (Takahashi 2010), velocity potential anomaly was used for medium system to fit the rainfall events.

To simplify the feature extraction, a general indicator through whole rainfall season was selected. Based on Takahashi and Yasunari (2006) and Chen et al. (2012) researches, the rain system might cover about 10-20 Longitude degree. From the power spectrum of horizontal wind for 2020, waves of 16 to 30 for velocity potential anomaly were used as long term POP, showed as the stripe in Figure 6.

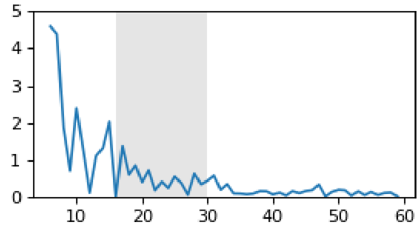

Figure 6. Power spectrum for potential velocity

II.4 Model Configuration

The classification model was used for predicting the weather type based on the weather features, so it's a multi-class classification problem. A 1D-CNN was used to train a wether type classifier which used 2 layers 1-d convolution. The model layer network is depicted in Figure 7. There were 2 convolutional layers and each of layers was followed by a ReLU activation layer. These layers helped to create various feature nodes. Dropout layer was to avoid overfitting. Maxpooling layer was for extracting the important features. The flatten layer was for vector data transforming to speed up computation. The last dense layers condensed the network information and used softmax function for the multi-class classification output (Clear, Light Rain, Medium Rain, and Large Rain).

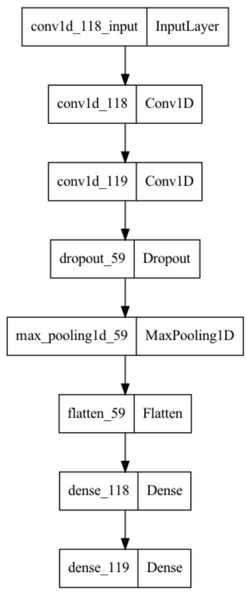

Figure 7. CNN model layer network

III. Evaluation and Discussion

III.1 Accuracy Analysis

To evaluate the classification model, accuracy analysis was used with different feature combinations to review the performance. Accuracy represents the correct prediction in total testing samples. The 25% of dataset was used as testing data and the different feature combinations are described in Table 3. Each of experiments was run 10 times and plotted in box plot as Figure 8.

Experiment Index	Feature Combination
Exp1	Basic
Exp2	Basic + Cloud
Exp3	Basic + Cloud + POP
Exp4	Basic + POP

Table 3. The different feature combination experiment settings

Comparing Exp1 to other models, Exp1 has lowest average Accuracy score. It's understandable that more features could provide more information to the model. Both cloud feature (Exp2) and POP feature (Exp4) can provide more information to the model and the effects are quite equivalent. However, when the two types of the features added together (Exp3), the contribution isn't accumulated so there might be some overlapped information brought from the features.

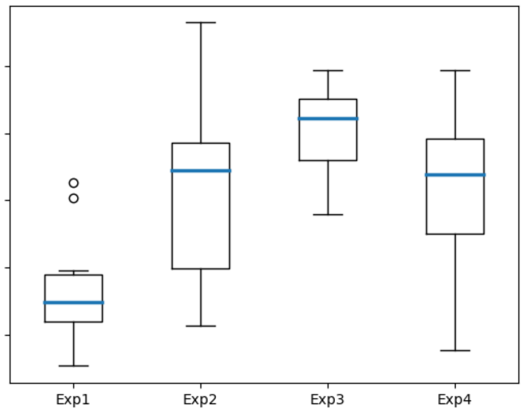

Figure 8. Box plot of model accuracy for different feature combinations.

III.2 Precision-Recall Curve Analysis

Since rain and no rain events were not quite equivalent in the social media reports, the evaluation of performance would use imbalance analysis. Because there were more no rain days, even if the model could predict no rain events, it couldn't guarantee the model being able to predict rain events well. Precision-Recall (PR) Curve is better for imbalanced dataset than Receiver Operating Characteristic (ROC) Curve. To compare the performance of different model, the area under curve (AUC) would be used. The larger AUC model has higher precision and recall values. The PR curve for different experiments are shown in Figure 9.

To compare AUC, all of experiments can handle no raining event prediction (AUC > 0.9). For cloud experiment (Exp2),

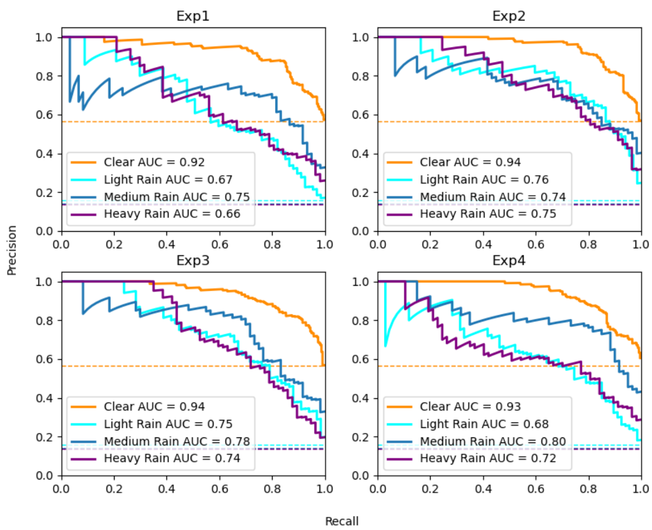

Figure.9 Precision-Recall Curves for different feature combination experiments

it has the best Light Rain AUC score, but the Medium Rain score doesn't help compared to Exp1. Because Light Rain represents the short term shower, so the larger area system (POP) has less effective and cloud fraction can represent the event. For POP experiment (Exp4), in contrast, it has the best Medium Rain AUC score, but the Light Rain score doesn't help compared to Exp1. Because Medium Rain is longer a raining event, and it would need more energy from larger area system (POP). For the combination experiment (Exp3, i.e. cloud plus POP), the Light Rain AUC score benefits from cloud feature and the Medium Rain AUC score benefits from POP feature. Although the Heavy Rain is not the best, the AUC score is higher than Exp1 and affected by both cloud and POP features. Heavy Rain represents the big amount of rain events, so it might give us a hint that the rainfall caused by both cloud fraction and larger area system effects.

To review the PR curves in charts, Exp4 has outstanding protrusion (in blue line) for Medium Rain which means the outperform comparing to other rain classes. On the other side, Exp2 has the outperform for Light Rain (in sky blue line). These both cloud and POP features may contribute the combination experiment (Exp3) to have a balanced performance in all weather types.

IV. Conclusion

Machine learning technology has been more and more reliable to apply in meteorological domain. This research fell into the group of using ML model with weather information to predict weather opinions from social media. A convolutional neural network with WRF output as features was developed to classify the weather types from social media. The first finding was that ML algorithm could have a promising forecast ability to predict weather community raining announcement. Even if there was no detail analysis processes about how the community determined the weather summary, ML could provide about 80% accuracy.

Secondly, by precision-recall curve analysis, both cloud feature and POP feature could help to the model performance and contribute in different weather types. Cloud features represented the short term shower, and POP feature which included larger system information would contribute longer term rainfall more. To have a balanced forecast model, it's better to include both types of features though the performance scores were not the best.

Finally, this research proposed a model design for ML weather classifier to predict social media weather opinions. In stead of common ML application to predict weather situations which is the goal of professional forecasters, this research predicted the weather opinions from social media users. Contrast to the rain announcements of this research were quite simple to identify the weather types, other communities could have unclear messages like "It seems to rain in this afternoon" or "It's about to rain". To deal with such unclear messages, using "opinion mining" (a branch of natural language process area) (Laan et al. 2017, Khairnar and Kinikar 2013) may preprocess meaningful texts to categorize different messages.

References:

Bochenek, B. and Ustrnul, Z., 2022. Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. *Atmosphere*, *13* (2), p.180.

Chen, T.C., Yen, M.C., Tsay, J.D., Alpert, J. and Tan Thanh, N.T., 2012. Forecast advisory for the late fall heavy rainfall/flood event in central Vietnam developed from diagnostic analysis. *Weather and forecasting*, *27* (5), pp.1155-1177.

Doswell III, C.A., 2004. Weather forecasting by humans—Heuristics and decision making. *Weather and Forecasting*, *19* (6), pp.1115-1126.

Eachus, J.D. and Keim, B.D., 2019. A survey for weather communicators: Twitter and information channel preferences. *Weather, climate, and society*, *11* (3), pp.595-607.

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J. and Tan, J., 2019. GPM IMERG final precipitation L3 half hourly 0.1 degree x 0.1 degree V06. *Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC)*, 10.5067/GPM/IMERG/3B-HH-L/06

Khairnar, J. and Kinikar, M., 2013. Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, *3* (6), pp.1-6.

Laan, A., Madirolas, G. and De Polavieja, G.G., 2017. Rescuing collective wisdom when the average group opinion is wrong. *Frontiers in Robotics and AI*, *4*, p.56.

NOAA Global Forecast System (GFS) was accessed on DATE from https://registry.opendata.aws/noaa-gfs-bdp-pds.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J. and Carvalhais, N., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, *566* (7743), pp.195-204.

Ripberger, J.T., Jenkins-Smith, H.C., Silva, C.L., Carlson, D.E. and Henderson, M., 2014. Social media and severe weather: do tweets provide a valid indicator of public attention to severe weather risk communication?. *Weather, Climate, and Society*, *6* (4), pp.520-530.

Roser, M. and Moosmann, F., 2008, June. Classification of weather situations on single color images. In *2008 IEEE Intelligent Vehicles Symposium* (pp. 798-803). IEEE.

Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A. and Stadtler, S., 2021. Can deep learning beat numerical weather prediction?. *Philosophical Transactions of the Royal Society A*, *379* (2194), p.20200097.

Silva, C.L., Ripberger, J., Jenkins-Smith, H.C., Friedman, J., Spicer, P. and Lamb, P.J., 2015. Utilization of real-time social media data in severe weather events.

Takahashi, H.G. and Yasunari, T., 2006. A climatological monsoon break in rainfall over Indochina—A singularity in the seasonal march of the Asian summer monsoon. *Journal of Climate*, *19* (8), pp.1545-1556.

Takahashi, H.G., Yoshikane, T., Hara, M., Takata, K. and Yasunari, T., 2010. High‐resolution modeling of the potential impact of land surface conditions on regional climate over Indochina associated with the diurnal precipitation cycle. *International journal of climatology*, *30* (13), pp.2004-2020.