Flying Safer Around Miami: Predictive Machine Learning Model for Real-Time Meteorological Aerodrome Report (METAR)

A Submission to Ocean Data Challenge hosted through Desight.AI
Dominikus Brian 钟鸿盛 | domi@dreambrook.tech
www.dreambrook.tech

Introduction

In the aviation world, Meteorological Aerodrome Report (METAR) is ubiquitous and plays crucial role in ensuring flight safety. The emergence of machine learning (ML) techniques, has now allowed historical weather reports to become more valuable and more meaningful.[1-4] In this study we delve deep into the various features of typical METAR formatted data to then reveals numerous useful insights that helps in weather forecasting for use in aviation. A collection of past METAR data were processed and curated into datasets used for training neural network-based machine learning models for local weather forecasting. The goal of this study is two-fold. First is to meticulously explore relationships across different meteorological data and uncover statistical patterns amongst different factor that influence weather, Second, is to demonstrate the utilization of past historical METAR report data in order to build a fast and predictive machine learning model and AI-driven weather forecast application. The dataset used for this particular study is based on METAR data published by Miami International Airport from January 2014 to December 2023. We believe the insights derived from this study could further empower ML application to make "Flying Safer Around Miami".

Key Findings

In this report we highlight five main key findings from the exploratory data analysis and machine learning model development regarding the Miami International Airport weather condition based on METAR formatted dataset. The report presents several key findings regarding weather prediction and trends, particularly around Miami International Airport, using a decade of historical METAR data and various analytical techniques:

(1) A specific type of Neural Network called the Long-Short Term Memory Neural Network was employed to create sets of models that capable of predicting next hour weather condition around Miami International Airport with good accuracy. We reveal that overall sets of temporal models employing shorter memory spans, perform better overall in predicting various weather factors compared to those with longer memory spans. Our result is comparable to others in literature that try predict similar weather feature using LSTM.[1,2]
(2) Among the nine primary features analyzed (Wind Direction, Wind Speed, Wind Variations, Visibility, Cloud Types, Cloud Heights, Temperature, Dew Point, and Air Pressure), Wind Direction and Cloud Types are the most challenging to predict accurately. Conversely, Temperature, Dew Point, Air Pressure, and Visibility are predicted with high accuracy, while the remaining features show moderate prediction accuracy.

(3) Yearly weather trends indicate an increase in mean temperatures, from an average of 25.01°C in 2014 to 26.24°C by 2023. In addition, the data analyzed clearly indicate seasonal variations that include stronger winds in the second half of the year, poorer visibility across fall and winter, with decrease of air pressure and more predominantly eastward wind direction nearing the year end.

(4) Time-Correlation study for each of the weather factors, shows decreasing temporal relationships for Wind Direction, Visibility, and Cloud Height and are negligible beyond 48 hours. Other weather factors like Temperature, Wind Speed, Air Pressure, and Cloud Coverage exhibiting cyclic correlations, indicating significant influence from past data on future predictions. Cross-correlation analysis between different weather factors reveals moderate to strong correlations between Temperature, Dew Point, and Cloud Coverage, which make it easier to train ML model, but no significant correlation for Cloud Types.

(5) We also identify that across the dataset weather phenomena like rain and mist, which require additional attention, constitute a small percentage of overall reports, indicating generally good operational conditions around the airport.

In the next section, we deliberately extend the explanation and discussion leading to the above several key findings and present the associated

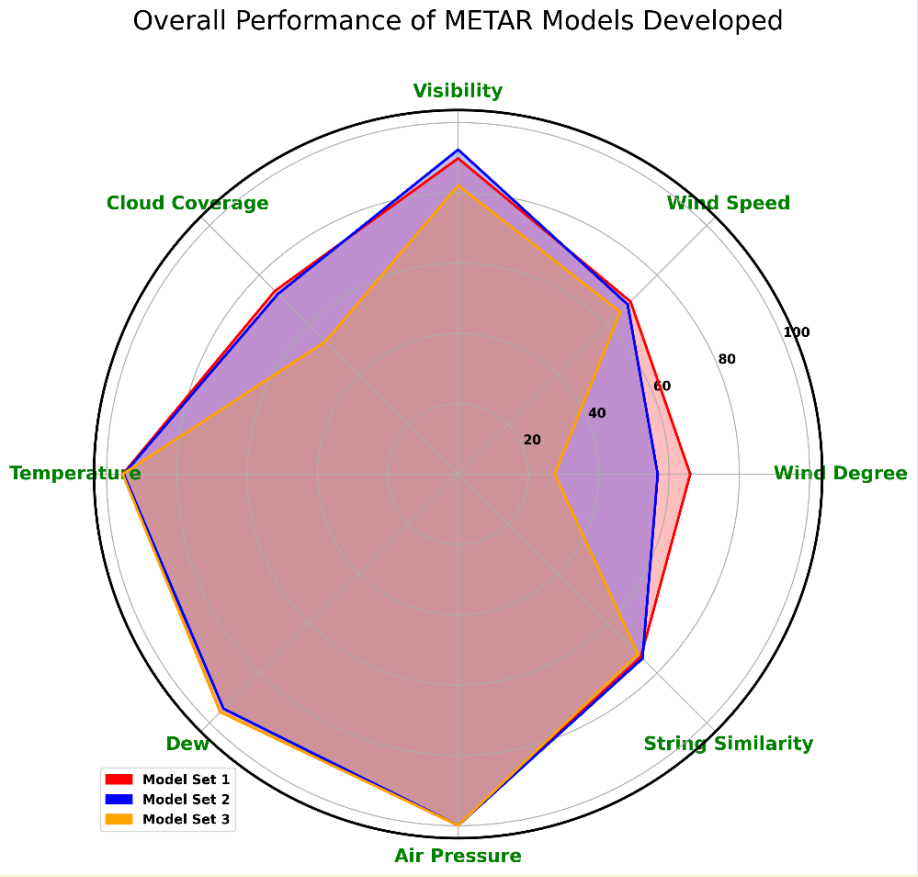


Figure 1. Comparison of overall performance of three Machine Learning Model Sets developed in this work. The presented Model Set correspond to a set of ML Models based on the LSTM structure with different length of memory spans. Model Set 1, 2, and 3 employs memory span of the length 6, 12, and 24 hours, respectively.

Discussions

We first begin with addressing the choice of primarily using the LSTM structure for developing the models. This decision is based on the outcome of our exploratory data analysis that reveals characteristics time-dependent (or temporal) relationships across different features in the dataset. LSTM as a specific type of recurrent neural network architecture in deep learning that have the capability to retain important information over long sequence of data. Making it an ideal general starting points for forecasting time-series data across different weather features. After performing preliminary scan of parameters set in the way LSTM model can be build, we unravel that using shorter memory perform better in overall prediction across different weather Factors. In this respect, we have trained and tested models for various weather factors having memory span of either 3, 6, 12, 24, 48, or 96 hours. In Fig 1 we visualize the results of three model sets developed here, each containing an independent LSTM model for each weather factor. Each Model Set contains LSTM models that all correspond to the same length memory span, albeit having different hyperparameters (parameters used to tune the model performance) optimized specifically for each weather factor such as temperature, wind speed, etc. Model Set 1, 2, and 3 each represent result for LSTM models employing past 6, 12, 24 hours memory spans, respectively. The reported data in Fig.1 is for 100 independents next hour prediction runs. For smaller number of runs, generally the averaged score is higher on all respects.

In order to predict a basic METAR formatted weather data (See Fig A1. of Appendix for illustration and explanation of the data), we here identify nine primary weather factors that serve as the prediction targets of our ML model. Out of 9 Primary Features mentioned above, Wind Direction and Cloud Types turns out to be the two most challenging to predict with high accuracy. Temperature, Dew Point, Air Pressure, and Visibility are predicted with superb accuracy, while the rest can be predicted with moderate accuracy. Fig.1 visualize overall performance of each model set upon the respective weather factor dimension. Fig. 2 provide snapshots of the string output of the METAR prediction made by ML models developed in comparison with the actual reported METAR data.

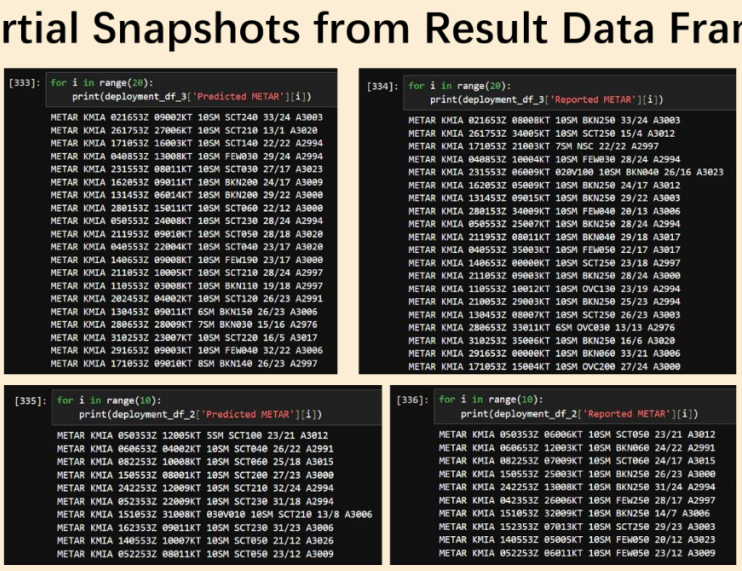


Figure 2. Snapshots displaying the string formatted output of the Predicted METAR data (left panels) and Reported METAR data (right panels) for a subset of independent next hour METAR predictions using Model Set 1 (upper panels), and Model Set 2 (lower panels).

We begin our exploratory data analysis by first performing data occurrence count and distribution across the dataset. The resulting histogram distribution for Wind Direction (in Degree), Wind Speed, Visibility, Temperature, Dew Point, and Air Pressure is shown in Fig.3 below. Specifically for cloud coverage we performed further analysis to identify pattern across cloud types (Fig. 4a), cloud types at different altitude layers (Fig. 4b) and its daily cloud cycle (Fig. 4c). Our data analysis on seasonal or cycles of weather profile across years, months, and throughout the day revealed that, in general weather in the surrounding of Miami International airport inherently demonstrated clear distribution trends shifts. The results of these variations are presented in Fig.5. First and foremost, in Fig. 5a. year-on-year mean temperature observed to increases steadily. We observed that the yearly average temperature is 25.01°C Degree throughout 2014 and passed through 26°C by the year of 2020 peaked at yearly average of 26.24°C by 2023. Notably the standard deviations of the data also are decreasing, meaning in absolute terms the higher temperature is observed to be more persistent over the years.

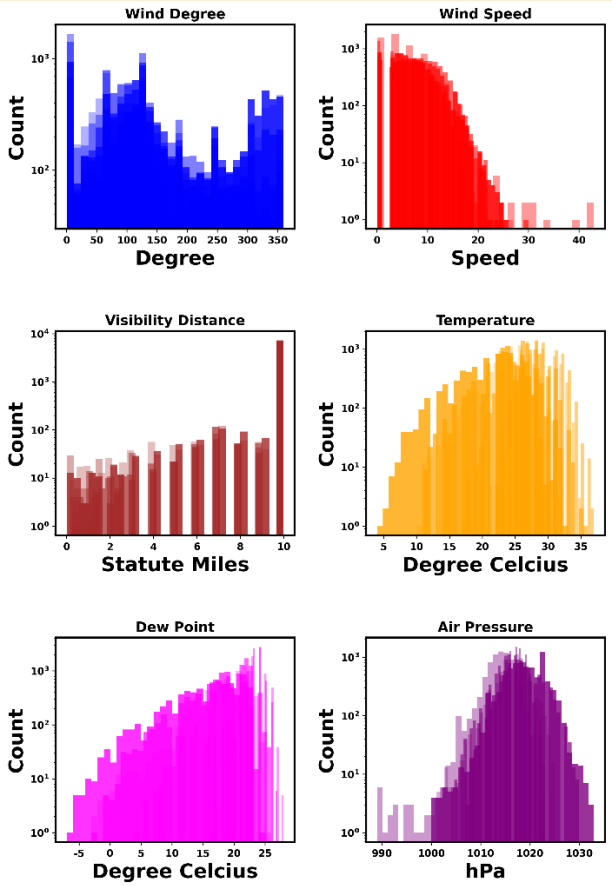


Figure 3. Data points distribution histogram for six weather factors in our METAR dataset. The data are plotted on a semi-logarithmic scale and with increasing transparency based on months, from January to December.

Secondly, there are also the increasing average temperature Strong winds more than double the typical wind speed tend to happens nearing the second semester of the year. Visibility is observably poorer more often in Fall and Winter season. Air Pressure decreases nearing the winter season. Nearing the end of the year, the wind blows toward east direction more often, indicating macroscopic weather pattern at play.

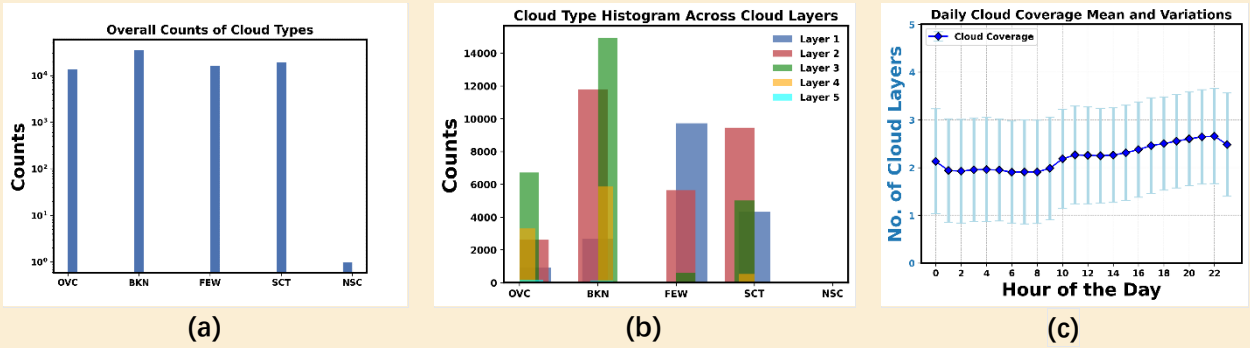


Figure 4. Overall cloud histogram counts by types (a), how it is distributed across altitude layers (b), and averaged Daily cycle of Cloud Coverage Layers.

In this study, we performed two sets of correlation studies. First one is a so-called autocorrelation function that checks on the correlation between a feature data point at particular time with the data points of the same feature from previous times. The second correlation study is cross-correlation check between different weather factor and data features. The results for our correlation studies are displayed in Fig. 6 for Wind, Temperature, Air Pressure, Dew Point, and Visibility and in Fig. 7 for cloud layers. Details on the way these correlations calculated is provided in the methodology section.

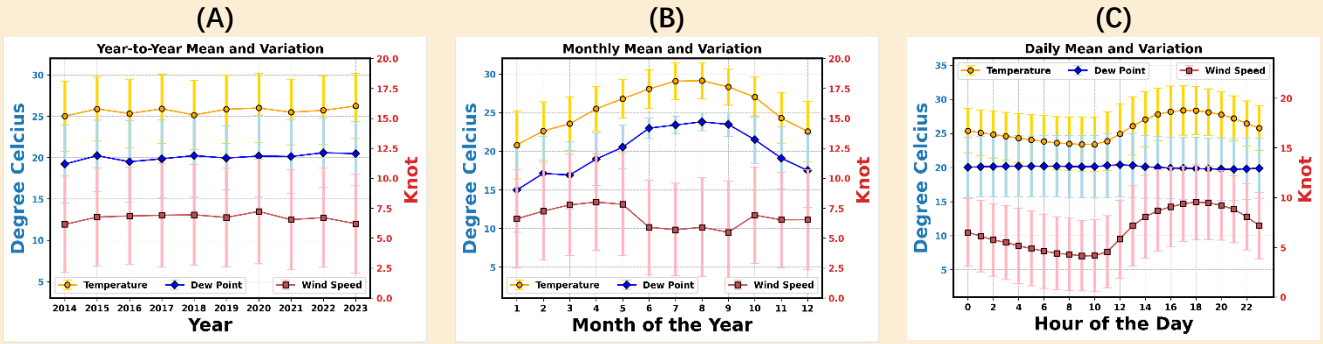


Figure 5. Mean and Variation analysis on seasonal or cycles of weather profile across (a) years, (b) month of the year, and throughout (c) hours of the day.

The first correlation study exhibit that, for Wind Direction, Visibility, and Cloud height, the temporal relationship (or memory if you may) is rather weak, diminishing to negligible value after about 48 hours. Dew points data, also become less memory dependent over time, but remain significant. Analysis on different layers of cloud, also suggest that cloud layer at lower altitude have smoothly decreasing correlation that makes it more predictable compare to the higher cloud layers. Which could also be an artifact to the measurement schemes used around the airport weather reporting station, or perhaps traced back to fluid dynamics of cloud formation that become more turbulent as it goes higher up. Temperature, Wind Speed, Air Pressure, and Cloud Coverage retain a cyclic correlation over time, that suggest for these properties past data retain quite significant influence in affecting future events. In particular the period for each cycle is exactly 24 hours for Temperature, Wind Speed, and Cloud Coverage, and is about 12 hours for Air Pressure. This emphasize the origin of the cyclic correlation is rooted in daily weather cycle.

Our second correlation study unravels that there exists moderately rather strong correlation between Temperature, Dew Point and Number of Cloud Layers (e.g., Cloud Coverage). Importantly we also notice that cloud type is not correlated with any obvious weather features available in the dataset, which also suggest an explanation why it is challenging to predict. A correlation heatmap visualizing the cross-correlation between different properties are shown in Fig. 8.

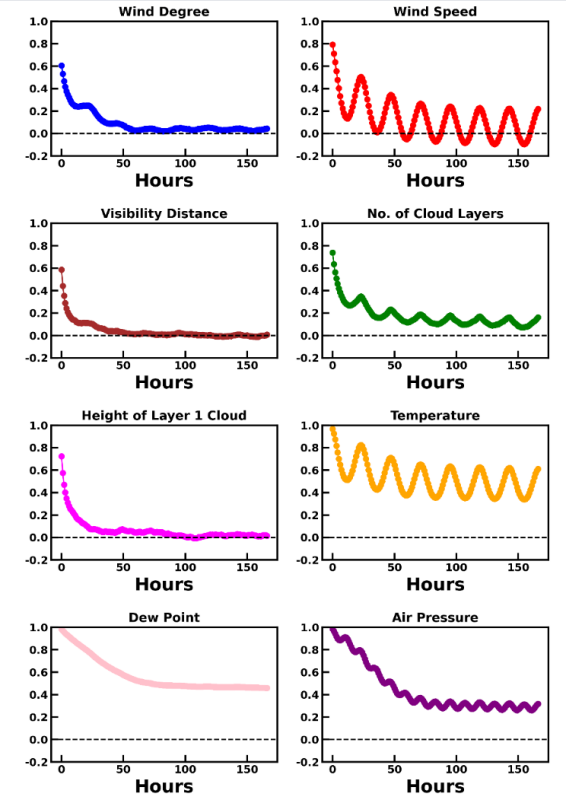


Figure 6. Autocorrelation Function analysis across timespans of 168 Hours (or 1 Week) for eight weather features. Dotted line on zero is provided to indicates points where no correlation exists. Positive value indicates positive correlation in such that if the previous value increases, the current data is also more likely to increase. Negative value thus indicates the opposite effect, in which if previous value increases its more likely to decrease at current time.

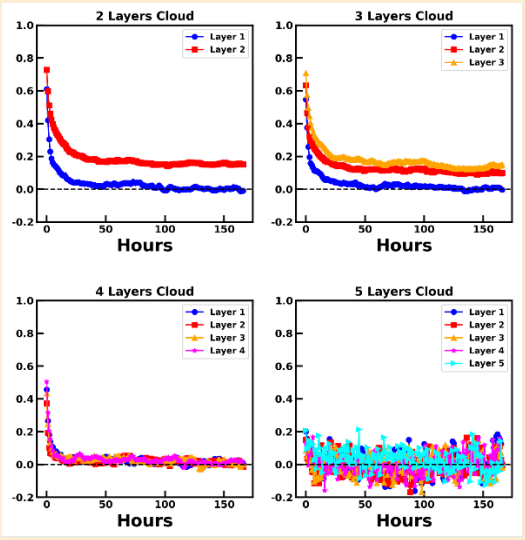


Figure 7. Auto-Correlation functions for height of the cloud on each respective higher altitude layers.

Moreover, we also deliberately analyze the data for specific weather phenomena such as rain, mist, fog, haze and so on. There are 4916 and 453 reported instances of Rain and Mist, respectively. The full tabulated result for all reported weather phenomena is shown in Table 1. The data indicate that in general over the years reports containing specific weather phenomena requiring extra attention account only about 5%. There are also several instance of Tornado and Water Sprout and several other uncanny phenomena, which here treated as outlier to prevent skewing the overall data quality and profile. We also looked into extra reported caution with runway and found out only 174 instances of minor visibility or range issue were reported, which is less than 0.2% of the dataset. Thus, in general the weather and runway situation around the Miami airport across the years are pretty much within operational and good conditions for most of the time.

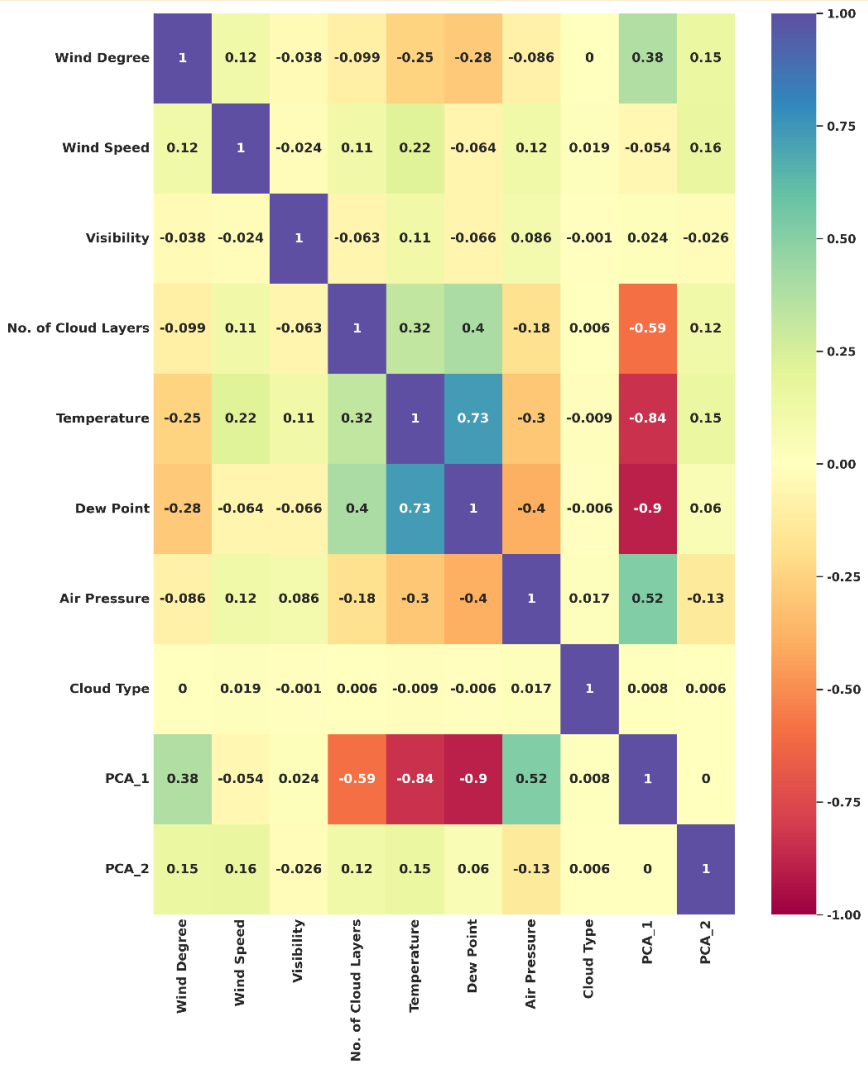


Figure 8. Pearson Correlation Heatmap displaying cross-correlation between different weather features. Here PCA_1 and PCA_2 is engineered features based on dimension reduction protocol using principal component analysis shown to exhibit second-degree correlation of different weather factors. The heatmap is diagonally symmetric and reveals either positive or negative correlations, with 1 and -1 representing complete positive and negative correlations, respectively.

Table 1. Weather Phenomena Reported in the METAR Dataset for KMIA (Miami International Airport) from the period of 1st January 2014 to 31st December 2023

Weather Phenomena	Occurrence Count
Rain	4916
Mist	453
Fog	72
Haze	39
Smoke	15
Drizzle	2
Funnel Cloud	3

Methodology

To further explains and describe techniques used here in performing exploratory data analysis and model developments in this section we provide rigorous details on the methodology, approach, and methods utilized throughout the work. Non-technical reader can freely skim through and skip these sections directly to the conclusion. For interested technical reader this section may be of value in understanding the procedures under the hood in performing analysis and ML model development presented in this work.

Data Preprocessing Protocol and strategy.

First, we begin with preprocessing of the METAR data downloaded from OCEAN data marketplace and parse each line of METAR data into a respective list. We then carefully parsed the data based on the standard METAR interpretations. To this end we leveraged upon the use of Open-source Python-based METAR Parsing package hosted here. Out of the 89,590 entries of METAR data, we identify 139 outliers that didn't met the strict criteria of the standard METAR or because it contains weird symbols non-general description, or extreme numbers. These data are put aside from the dataset leaving us with 89,451 entries of METAR data for further analysis and ML model development. Data are further cleaned by scanning over *NaN* or *-inf* values, replacing them with zero mask whenever appropriate. For categorical data we also performed numerical encoding to allows for its use throughout ML pipeline. For instance, in determining whether a certain range of wind variation exist a Boolean operator with 0s and 1s were assigned based on each METAR data. Numerical encoding for cloud type is performed by converting them using octal unit such that NSC = 0, FEW = 1, SCT =2, BKN = 5 and OVC = 8. Illustration of the parsed data is shown in Fig. 9.

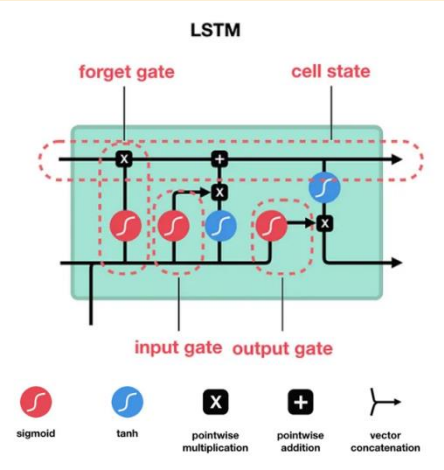
```
[45]: {'_wind': Wind[speed=5, direction=ESE, gust=None, degrees=120, unit=KT, min_variation=None, max_variation=None],
       '_visibility': Visibility[distance=10SM, min_distance=None, min_direction=None],
       '_vertical_visibility': None,
       '_wind_shear': None,
       '_cavok': None,
       '_remark': 'automated station with a precipitation discriminator sea level pressure of 1014.1 HPa  hourly temperature of 27.2°C and dew point o
      f 22.2°C  $',
       '_remarks': ['automated station with a precipitation discriminator',
        'sea level pressure of 1014.1 HPa',
        '',
        'hourly temperature of 27.2°C and dew point of 22.2°C',
        '',
        '$'],
       '_clouds': [Cloud[height=1500, quantity=CloudQuantity.FEW, type=None],
        Cloud[height=17000, quantity=CloudQuantity.OVC, type=None]],
       '_weather_conditions': [],
       '_day': 4,
       '_time': datetime.time(21, 53),
       '_message': 'KMIA 042153Z 12005KT 10SM FEW015 OVC170 27/22 A2995 RMK AO2 SLP141 T02720222 $',
       '_station': 'KMIA',
       '_flags': set(),
       '_trends': [],
       '_temperature': 27,
       '_dew_point': 22,
       '_altimeter': 1014,
       '_nosig': False,
       '_runways_info': []}
```

[347]:

	year	month	day	hour	minutes	wind_speed	wind_degrees	wind_min_var	wind_max_var	vis_distance	...	air_pressure	n_cloud_layers	clou
0	2014	3	11	8	53	0	0.0	0.0	0.0	10.0	...	1014	1	
1	2014	3	11	9	53	0	0.0	0.0	0.0	10.0	...	1014	1	
2	2014	3	11	10	53	3	40.0	0.0	0.0	10.0	...	1015	1	
3	2014	3	11	11	53	0	0.0	0.0	0.0	10.0	...	1015	2	
4	2014	3	11	12	53	0	0.0	0.0	0.0	10.0	...	1015	2	
5	2014	3	11	13	53	0	0.0	0.0	0.0	10.0	...	1015	1	
6	2014	3	11	14	53	3	57.0	0.0	0.0	10.0	...	1015	1	
7	2014	3	11	15	53	6	103.0	0.0	0.0	10.0	...	1015	1	
8	2014	3	11	16	53	9	170.0	140.0	210.0	10.0	...	1014	1	

Figure 9. Visualization of the single instance of parsed data entry (upper panel) used to curate the dataset of this study, and a random sequence clip from the produced curated dataset with each row containing an entry of weather features or information.

Machine Learning Architecture & Hyperparameter Space



In our ML development pipeline as described above we opt in using a specialized form of recurrent neural network deep learning architecture called the Long-Short Term memory. The LSTM architecture have memory cells and three type of gates that regulate flow of information passing through each memory cell. In particular the architecture also allows for the cell state to be continuously maintained and therefore exhibit good performance with retaining information over long sequences of data. In LSTM network, there are several hyperparameters that we can tune for optimizing performance. Namely, Depth of the LSTM layers, Number of (neural network) units in each layer, Batch size, Sequence length, Dropout rate, Optimizer scheme, Loss functions type, and more.

Figure 10. Illustration of LSTM model Architecture. ***Image Courtesy:*** Illustrated Guide to LSTM's and GRU's.

In this study we first build a general basis LSTM architecture using the Keras and TensorFlow module with a set of initial starting hyperparameters. We then define a sizable Hyperparameter to search for optimized combinations of parameters for each given target properties, such as Wind Speed, Temperature, and so on. The hyperparameter space defined in this work is given in Table 2.

Table 2. Hyperparameter Space defined for LSTM model development in this work.

Parameter Name	Values
LSTM Depth	{2, 3, 4}
Number of Units	{50, 80}
Sequence Length	{3, 6, 12, 24, 48, 96}
Dropout Rate	{0.1, 0.2, 0.3}
Epoch Length	{50, 100}
Number of Training Data	{1000,2000,5000,10000, 50000}
Train-Test Ratio	{80:20, 90:10}
Optimizer Type	Adam
Loss Type	Mean Squared Error

Feature Engineering and Correlation Studies

In performing our analysis and featurization engineering we have made use of the following correlation functions and statistical analysis for dimension reduction.

The AutoCorrelation Function (ACF) allows us to measures the correlation of a signal with itself at different lags or time delay. It is a useful tool in time series analysis to identify repeating patterns or seasonal effects in our weather dataset. The ACF at lag or delay time (in our context in the unit of hours) k is defined as the correlation between time series observations that are k time periods (hours) apart. Mathematically, the ACF at a given time lag is given as:

$$ACF(k) = \frac{\sum_{t=1}^{N-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{N}(y_t - \bar{y})^2}$$

Where y_t is the value at time t and $\bar{y}$ is the mean of the series, N is our total number of observations.

The Pearson Correlation Coefficient measures the linear relationship between two datasets. Ranging from -1 to 1, a value of 1 implies a perfect positive linear relationship, -1 implies a perfect negative linear relationship, and 0 implies no linear relationship. The Pearson Correlation Coefficient between two variables X and Y is defined as:

$$r_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

In the above $\bar{X}$ and $\bar{Y}$ are the means of the X and Y variables, respectively.

Principal Component Analysis is a statistical technique commonly employed in machine learning protocol for performing dimensionality reduction while preserving as much of the variance in the dataset as possible. PCA allows us to transforms the original variables into a new set of uncorrelated variables, known as principal components, which then ordered by the amount of variance they capture from the original dataset. The first principal component captures the most variance, the second captures the second most, and so on.
Most of the time the first few principal components capture best the characteristics features of the included

original variables. In mathematical terms, when performing the PCA we found the principal components by solving the eigenvalue and eigenvector of the covariance matrix of the original data. This transformation can be done by solving for $\mathrm{Z} = \mathrm{X}\,\mathrm{W}$; with X being the original data matrix, W is eigenvectors we are trying to find and Z is the finalized transformed data based on principal components. Typically, this is performed using iteration method or specific matrix operations algorithms. In our case these mathematical operations are performed using the scikit-learn module.

Model Performance Analysis and Benchmarking

After obtaining the models for each targeted weather property we want to be able assessing them and know which one is preferred to be used in the overall ML model being developed. In the previous sections we have mentioned the necessity of exploring the hyperparameter space, practically this is done by comparing the mean average error (MAE) and root mean square error (RMSE) metrics produced by each model for specific hyperparameters combination. Since the hyperparameters space are huge, here we specifically print out the optimization steps metadata containing RMSE and MAE comparisons of the different target properties, shown in Fig. 11 below.

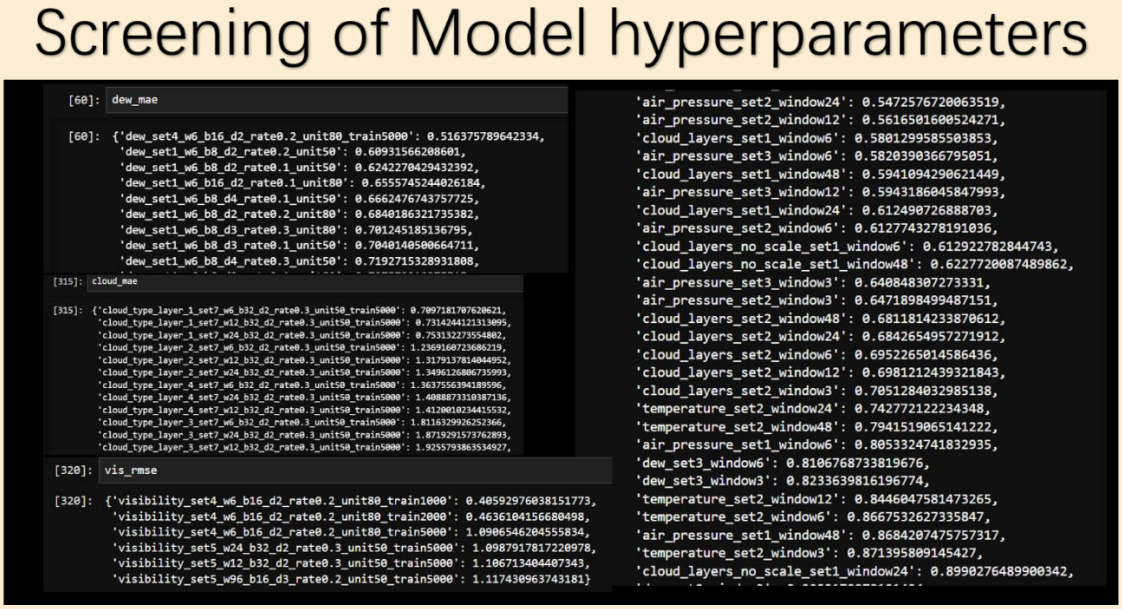


Figure 11. Screenshot of the metadata of MAE and RMSE used to judge the preferred optimized hyperparameter combination for each target properties.

On top of the hyperparameter performance analysis, we also developed a benchmarking strategy in order to evaluate the overall performance of the Model Set that is used to predict the final METAR string data output in contrast to just assessing the property specific model. To this end, we define eight dimensions of assessment, namely: 'Wind Degree', 'Wind Speed', 'Visibility', 'Cloud Coverage', 'Temperature', 'Dew', 'Air Pressure', 'String Similarity'. For each dimension we calculate a *Score* based on the deviation from true value that the model supposes to be predict. The deviation is calculated based on the MAE approach and the score is simply (1-MAE %) x 100, thereby giving maximum score of 100 and minimum of close to Zero. For categorical data such as cloud coverage, we aggregate the prediction based on True or False, assigning value of 100 to True answer and 50 to False answer. For cases with more than one layers of cloud score is averaged across number of

layers. The height of each cloud layer also is weighted into the 'Cloud Coverage' dimension by means of calculating scores for independent cloud height at each layer. For the 'String Similarity' dimension we utilize the average of the *Levenshtein Distance (Edit Distance)* and the *Jaccard Similarity* of the Predicted and Reported METAR data output string (See Fig. 2). The overall Model Set assessment is then produced as an array of scores to then be plotted as shown in the Fig. 1.

Conclusions

In this report, we have presented a detailed exploration of employing machine learning techniques for advancing weather forecasting in aviation, with a particular emphasis on analyzing and predicting weather patterns around Miami International Airport (KMIA). Our work here has been centered on developing a model capable of generating real-time, next-hour METAR predictions for aviation-grade weather data at KMIA, highlighting the efficacy of models with shorter memory spans of about 6hrs. Through this study, we have delved into various factors critical for enhancing the predictive accuracy of ML models in weather forecasting contexts. We have also comprehensively explored the yearly, seasonal, and daily weather patterns, revealing interconnectedness of different weather factors. Furthermore, we performed analysis of temperature trends, examining time-correlations and cross-correlation among different weather variables, and tackling the complexities inherent in forecasting cloud formations.

The insights gained from our investigation not only deepen our understanding of the meteorological dynamics specific to the Miami International Airport but also helps us to develop machine learning pipeline suitable for processing weather related data and prediction task. Looking ahead, we plan to extend this study to refine the model's efficiency further by using other state-of-the-art ML architecture and explore its applicability generalization to other airports, aiming to broaden the scope of our findings and their utility in weather prediction, crucial for planning and operational decision-making in aviation and related fields.

Appendix

METAR DATA FORMAT

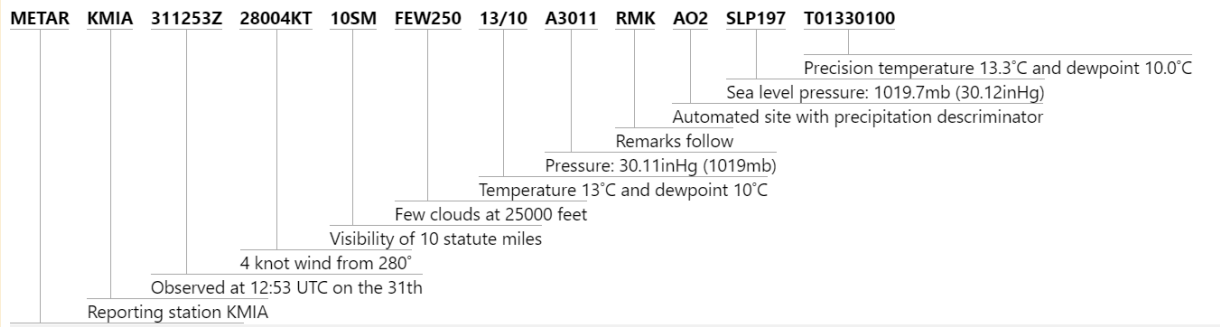


Figure. A1. Illustration of the prototypical standard METAR formatted aviation weather data. Image Generated via https://flightplandatabase.com/METAR

The term "METAR" is short for "Meteorological Aerodrome Report." These reports are typically generated at hour-by-hour basis, and sometimes updated several times within an hour in case of significant weather changes. A typical METAR report is shown in Fig. A1 which includes information on various weather elements that are critical for aviation industry and flight operations, such as: Wind speed and direction, Visibility, Weather phenomena, Sky condition, Temperature, dew point, and Air pressure. Sometimes, when necessary, information regarding runway condition or how the measurement data obtained is also included as remarks in specific formats. METAR reports are concise and encoded in a standardized format to ensure consistency and readability across the global aviation community.

References

[1] GALBOKKA H, PRADEEP. R. P., BEHERA, A., TROVATI, M., & PEREIRA, E. (2019). Long-Short Term Memory for an Effective Short-Term Weather Forecasting Model Using Surface Weather Data. *IFIP Advances in Information and Communication Technology*, *559*, 382-390. https://doi.org/10.1007/978-3-030-19823-7_32

[2] A. Srivastava and A. S, "Weather Prediction Using LSTM Neural Networks," *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, Mumbai, India, 2022, pp. 1-4, https://doi.org/10.1109/I2CT54291.2022.9824268

[3] Chen, L.; Han, B.; Wang, X.;Zhao, J.; Yang, W.; Yang, Z. Machine Learning Methods in Weather and Climate Applications: A Survey. Appl. Sci. 2023, 13, 12019. https://doi.org/10.3390/app132112019

[4] Amir Abecassis, Daniel Delahaye, Moshe Idan. A Machine Learning Framework to Predict General Aviation Traffic Counts A Case Study for Nice Cote D'Azur Terminal Control Center. *SESAR Innovation Days*, Dec 2022, Budapest, Hungary. ⟨hal-03907368⟩