Predictive Analysis of Weather in Trinidad and Tobago

Jordhan Couteau   
*Department of Computing and Information Technology  
University of the West Indies, St. Augustine*The Republic of Trinidad and Tobago, West Indies  
jordhan.couteau@my.uwi.edu

Lillian Charles  
*Department of Computing and Information Technology  
University of the West Indies, St. Augustine*The Republic of Trinidad and Tobago, West Indies  
lillian.charles@my.uwi.edu

Caleb Ramdath  
*Department of Computing and Information Technology  
University of the West Indies, St. Augustine*The Republic of Trinidad and Toabgo, West Indies  
caleb.ramdath@my.uwi.edu

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—Weather Forecasting, Pre-Processing, Data, Weather Model, (key words)

# Introduction

Weather forecasting is an essential aspect of daily life, shaping decision-making across critical sectors such as agriculture, transportation, disaster management, and energy planning. The ability to predict weather conditions accurately not only influences individual activities but also drives economic efficiency and safeguards communities against adverse weather events. However, despite advancements in technology, traditional weather prediction models often fall short in providing precise forecasts especially in Trinidad and Tobago. These limitations hinder proactive planning and preparedness, especially in regions prone to severe weather phenomena.

In response to these challenges, this paper seeks to enhance the accuracy and granularity of weather predictions through the application of big data analytics and exploratory analysis of historical weather datasets. By identifying specific parameters—such as temperature thresholds, humidity levels, wind speed patterns, and atmospheric pressure changes—we aim to detect critical conditions that precede severe weather events. Through this approach, our analysis strives to bridge the gaps in traditional forecasting methods by offering data-driven insights tailored to localized needs.

Accurate and reliable weather predictions are essential for mitigating the devastating effects of natural disasters, optimizing operations within weather-dependent industries, and safeguarding public safety. Improved forecasting capabilities can empower farmers with actionable data for planting and harvesting cycles, enable more efficient and cost-effective transportation logistics, and bolster the preparedness of island communities for extreme weather events. Beyond disaster prevention, our efforts also focus on enhancing the reliability of everyday weather forecasts, making them an indispensable tool for routine decision-making.

# Methodology

## Background

In this section, the preliminary approach that involves the leveraging of machine learning algorithms to develop a predictive model for weather forecasting will be highlighted. Further, factors such as data considerations, feature engineering and challenges faced are detailed.

## Approach

Our approach consisted of gathering relevant data sources, discerning which dataset would yield the most information, selecting which variables would be the most beneficial to Trinidad and Tobago and deciding which models would be best suited to train the selected dataset.

* Data Collection

Three data sources were identified to be best suited to be utilized. Meteostat was considered as it provides historical data for Piarco International Airport, facilitating pattern and trend analysis, this dataset also offered visualization tools that could enhance data exploration and interpretation. On the other hand, Climate Change Model Portal offered an extensive historical climate data for Trinidad and Tobago, including temperature, precipitation and extreme weather events. This dataset aids in trend analysis and future climate predictions using various models and scenarios. Finally, Visual Crossing Weather API was selected as the dataset that would utilized as it provided real-time and historical data for regions across Trinidad and Tobago with metrics such as temperature, humidity, wind speed and rainfall.

* Machine Learning Models

Ridge Regression Model was not selected as it assumes a linear relationship between features in the dataset and the target variable, which would affect predictions as weather patterns are non-linear, influenced by connections between for example, temperature, pressure and humidity.

AutoRegressive Integrated Moving Average (ARIMA) Model was not chosen due to its limitation of being an univariate model [1]. This means that the model forecasts based on a singular variable, this with weather forecasting involving multiple variable, the ARIMA would not be able to handle the complex data manipulation needed for the dataset.

Extreme Gradient Boosting (XGBoost) was selected as the main model for this project. This model was selected as it handles non-linear patterns, processes large datasets, resistant to overfitting and it works well for both short-term and long-term weather forecasting [2].

# Analysis

Data from Visual Crossing API pertaining to regions in Trinidad and Tobago between the years 2000 and 2025 were downloaded, cleaned and merged to create a combined dataset with weather information pertaining to Trinidad and Tobago.

Observations were made to ensure that the data being utilized would provide valuable insights. Features were then selected and engineered to enhance the predictability of chosen model. Visualizations were also made to help identify weather patterns as well as hidden relationships between the data to make more informed decisions.

## Data Collection and Preprocessing

Figure 2 depicts the records that includes observations of the regions downloaded based on Figure 1, showing precipitation, humidity, average temperature (avgtemp c), maximum temperature (tempmax c) and minimum temperature (tempmin c).

Fig. 1. Picture showcasing regions in Trinidad and Tobago

Fig. 2. Weather Dataset showcasing its features

## Data Loading and Cleaning

This step involved loading the raw dataset into an human-readable structured format utilizing the pandas library in Python. Preliminary data exploration was conducted to identify missing values, inconsistencies, outliers and to remove any irrelevant variables such as snow that has no correlation to the region as showcased in Figure 3 and 4.

Fig. 3. Preprocessing of data

Fig. 4. Preprocessing of data

Identify applicable funding agency here. If none, delete this text box.

## Feature Selection and Engineering

To enhance the predictive capabilities of the XGBoost model, various key features were engineered.

* Heat Index

Average temperature and Humidity were used to calculate the Heat Index. Also known as the apparent temperature, heat index is a measurement of how hot it feels when air temperature and relative humidity are combined [3]. Calculations made were compared against the National Oceanic and Atmospheric Administration (NOAA) chart and their heat index calculator to ensure accuracy of results.

* Dew Point

Equation (1) depicts the formula needed to calculate dew point, with Ts depicting dew point, T for temperature and RH representing relative humidity. In this case, average temperature and humidity was used to engineer this feature.

Ts = (b × a (T,RH)) / (a - α(T,RH)) 

* Feature 3
* Feature 4

Fig. 5. Feature Engineering

Fig. 6. Feature Engineering

## Visual Analysis

Visual analysis allowed for the data to be analysed more thoroughly to derive trends, outliers and relationships. It also plays a crucial role in both the development and validation of the weather forecasting model in various ways such as feature engineering as highlighted in Figure 5 and 6, parameter tuning and understanding the model’s limitations.

Fig. 7. Name of Screenshot

Fig. 8. Name of Screenshot

Fig. 9. Name of Screenshot

# Model Training

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Model Selection

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Feature Scaling

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Model Evaluation Metrics

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

# Model Results

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

# Web Application

The web application developed provides a user-friendly interface for accessing and visualizing the output of the advanced weather model. It integrates historical data, model predictions and interactive visualizations, enabling a comprehensive forecasting of the weather time period selected.

Fig. ?. Screenshot

# Discussion

# Conclusion

##### References

1. “ARIMA Modeling,” CORP-MIDS1 (MDS). https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling/
2. Analytics Vidhya, “Understanding the Math behind the XGBoost Algorithm,” Analytics Vidhya, Sep. 06, 2018. https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/
3. N. US Department of Commerce, “What is the heat index?,” www.weather.gov, 2024. https://www.weather.gov/ama/heatindex
4. “What is Dew Point? Camlab,” www.camlab.co.uk. https://www.camlab.co.uk/blog/what-is-dew-point
5. K. Elissa, “Title of paper if known,” unpublished.
6. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
7. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
8. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
9. K. Eves and J. Valasek, “Adaptive control for singularly perturbed systems examples,” Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
10. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>
11. S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC
12. “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**