Comparative Analysis and the Progression of Different Machine Learning Models in the field of Meteorology

Sriram Yerramsetty

Abstract.

Understanding upcoming weather patterns is extremely important for several different reasons in a community. For instance, weather patterns can hint at upcoming threats to a community such as hurricanes or tornados and give people a proper amount of time to prepare or evacuate the area. Also, weather patterns can give members of the agricultural field notice of rainfall the following day, week, or even month. This allows them to make better grounded decisions in order to produce better produce for their community. Today, numerical weather predictions models serve as the basis for weather predictions for several different weather news sources. However, the recent influx of Machine Learning has made few question whether machine learning models might be a better fit for weather predictions. Machine Learning consists of several different models and algorithms which all perform in different ways. The aim of this research is to compare three different machine learning models in both their accuracy and their progressive learning given more training data by using the field of meteorology as a testbed. The models were given several points of data and expected to return a value predicting whether it would rain the following day with varying amounts of training data. The three different models used were a Logistic Regression, Multilayer Perceptron, and Support Vector Machine models. After

conducting the tests, it was found that the Multilayer Perceptron model predicted wielding the greatest accuracy. Furthermore, it was similarly found that the Multilayer Perceptron model had the greatest positive progression as more training data was inputted.

1 Introduction.

1.1 Background and Motivation

Machine learning models have shown tremendous potential in various fields, including weather forecasting. These models can analyze patterns and relationships within weather data to make accurate predictions. However, with the increasing availability of data and advancements in computational power, it is crucial to assess the effectiveness of different machine learning models and understand how their performance improves with the addition of more data. Weather prediction serves as an ideal testbed for this analysis due to the abundance of existing historical weather data and the importance of accurate forecasts.

1.2 Research Objectives

The primary objective for this paper is to analyze and compare the effectiveness of different machine learning models in predicting weather patterns. Moreover, analyze the speed at which these models improve with the addition of more learning data. Also discover any flaws or strengths which different models might have in predicting weather patterns. Primarily, I would like to analyze the compatibility of different machine learning platforms, specifically in a meteorological context.

1.3 Significance

This research can inform decision-makers about the optimal utilization of the correct machine learning models. It can also uncover the learning behaviors of each machine. I hope the findings from this study enhance weather forecasting techniques and contribute to advancements in the field of machine learning and meteorology alike.

2 Literature Review.

2.1 Machine Learning and Meteorology

Most weather stations today use a process known as numerical weather predictions (NWP) to predict weather patterns. NWP uses a large set of equations in order to accurately predict weather patterns [1]. These equations are based on a plethora of factors picked up by weather stations and output results for temperature, humidity, wind speed, but more importantly, precipitation chances. This existing model became widely used for its accuracy in the 1950s due to the introduction of computer simulations [2].

The recent influx of machine learning has shaped discussions for whether machine learning should be introduced to the field of meteorology. Given Machine Learning’s fast-paced development over the past few years it only seems like a matter of time before Artificial Intelligence and Machine Learning will take over the meteorology field. Although simple to say, the issue presents itself as being far more complex than simply a matter of time.

Researchers at Colorado State University [3] have explored the integration of machine learning into forecasts predictions. Although the progress presented by machine learning models is promising, the university iterates that currently, it is not enough to be able to replace existing NWP models.

However, the same model trained by the University was indeed better at predicting severe weather patterns than the NWP models [3]. Furthermore, the CSU- Colorado State University- Weather Machine Learning model is capable of predicting other severe weather events such as tornadoes [3] [4]. These findings contrast as they show that ML processes are indeed adequate enough to be implemented into the field of meteorology.

2.2 Previous Findings

Sancho Salcedo-Sanz and his team, most of which were from the University of Alcala, in their study [5] they researched the accuracy of a machine learning algorithm in making temperature predictions. They utilized a Support Vector Regression machine learning model and trained it using preexisting weather data. The study found that the SVM is able to give an accurate prediction for the maximum temperature for the following day.

I. Gad.., [6] presented a comparative study of different machine learning methods, including decision trees, AdaBoost, linear regression, and support vector machine, using NCDC weather data as a comparison metric. The study investigated the differences between the models through accuracy of predicted wind speed, humidity, and temperature. The researchers found that AdaBoost model performed with the greatest accuracy when compared to the other models. However, this study did not investigate the differences in learning processes between the models, it primarily looked at the prediction accuracy of the models.

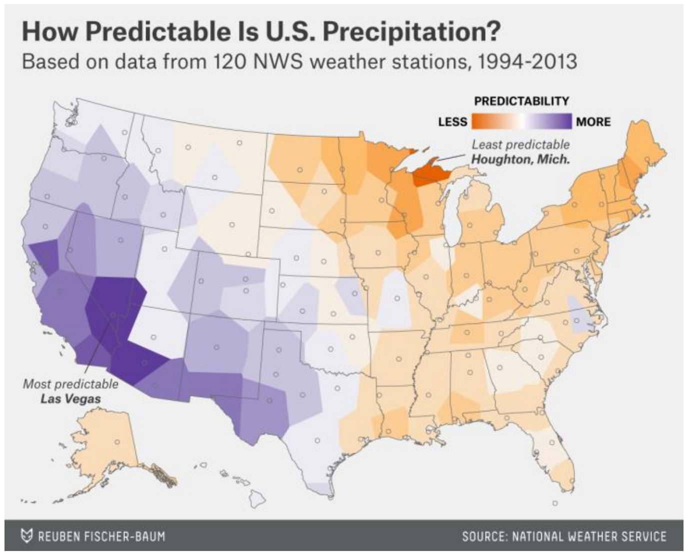
In T. Anjali et al [7], researchers studied the accuracy of three different machine learning models- Multiple Linear Regression (MLR), Artificial Neural Network (ANN) and Support Vector Machine (SVM). As a baseline, the team used temperature weather data obtained from Kerela from 2007 to 2015. To evaluate the results, Mean Error, Mean Absolute Error, Root Mean Square Error, and Correlational Coefficients were used. The study did not focus on the progressive nature of the models, rather the accuracy of which they predicted the temperature data. The team found that Multiple Linear Regression performed with the greatest accuracy overall, being evaluated as more precise than the ANN and SVM models.

In D. Cho et al [8], researchers utilized weather data featuring extreme temperatures in urban areas as a basis for comparison for machine learning models. Rather than having the machine learning model directly output a predicted temperature then comparing the accuracy of which with the NWP temperature prediction accuracy, D. Cho et al offers a combination of ML and NWP. In the study, researchers aimed to reduce the large bias present in NWP models when predicting extreme temperatures by using ML models- specifically, random forest (RF), support vector regression (SVR), artificial neural network (ANN) and a multi-model ensemble (MME). They found that the multi-model ensemble model was best at accurately reducing the bias of the NWP temperature prediction.

3 Methodology.

3.1 Data Collection

Before beginning to collect data, we need to choose where to collect data from. This needs to be kept consistent to not provide any advantages for one learning model as some environments are easier to predict patterns than other more unpredictable environments (see Figure 1). More specifically, we need to narrow data to be from only one weather station which has remained in a fixed location for at least the last decade. This provides us with ample amounts of data to feed into the models as learning data and also enough data to test with. I decided to narrow my search further to cities in the United States with unpredictable weather, as this provides more of a challenge and an opportunity to see the progressive learning of machine learning models. After all these factors considered, I chose to use weather data from Boston, Massachusetts.



[9] [10] (FiveThirtyEight, National Weather Service)

As for the data, Weather Underground [11] provided detailed and accurate historical information of a specific weather station, the Logan Airport Weather Station, for several decades.

I intend to input the Machine Learning models with the following data:

- Temperature (Max, Average, Min)
- Dew Point (Max, Average, Min)
- Humidity (Max, Average, Min)
- Wind Speed (Max, Average, Min)
- Pressure (Max, Average, Min)

I hid the precipitation values as those are what will be used as the output which the models will need to accurately predict during testing. I also hid the Day from the models as it could be used to detect trends occurring throughout the month which would greatly reduce the need for the other variables. Importantly, the precipitation values are offset by a day therefore they correlate to the precipitation for the following day rather than the same day as the other values. This makes it so the models need to predict the precipitation for the next day, if it were set to the same day then the precipitation would have already happened when the other values had been taken. Therefore, our models would not really be predicting precipitation.

I intend to feed the models data from 2015-2019 for learning then use data from 2020 for learning. In total, there are over 150 days, each consisting of over 16 different data points, that will be used in this research.

3.2 Constructing Machine Learning Models

In this study, we will use and compare three different machine learning/artificial intelligence models.

- Logistic Regression

- Multilayer Perceptron

- Support Vector Machines

To perform the study, each one of these five models needs to be constructed inside of Python using the pandas database [12].

3.2.1 Constructing the Machine Learning Models

The constructed models coded in Python looks like:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

# Load the data into the panda dataframe
# Input the file name of the sheet being used. Different sheets will
```

```
# have different amounts of training data.
data = pd.read_csv('FILE NAME')


X = data[['Max_Temp', 'Avg_Temp', 'Min_Temp', 'Max_Dew', 'Avg_Dew',
'Min_Dew', 'Max_Humid', 'Avg_Humid', 'Min_Humid', 'Max_Wind', 'Avg_Wind',
'Min_Wind', 'Max_Pressure', 'Avg_Pressure', 'Min_Pressure']]
y = data['Next_Percep?']


# Initialize the model which is being tested on by uncommenting it.
# model = LogisticRegression()
# model = MLPClassifier(hidden_layer_sizes=(100, 100), activation='relu',
# random_state=42)
# model = SVC(probability=True)


# Train the model
model.fit(X, y)


# Input testing values
test1 = [[42, 39, 36, 32, 24, 21, 73, 55.2, 47, 21, 15.5, 9, 29.8, 29.6, 29.5]]
test2 = [[48, 45.2, 41, 42, 39, 36, 90, 78.9, 63, 12, 5.3, 0, 29.8, 29.6, 29.5]]
test3 = [[40, 37.4, 33, 35, 25.2, 17, 83, 63.3, 45, 26, 16.6, 8, 29.9, 29.6, 29.4]]
test4 = [[42, 38, 34, 25, 20.8, 17, 61, 50.4, 37, 15, 9.3, 3, 30.1, 30, 29.8]]
test5 = [[44, 37.3, 31, 31, 23.2, 5, 82, 59, 32, 26, 14.2, 5, 30.1, 29.8, 29.7]]
test6 = [[69, 59.4, 49, 54, 48.1, 41, 77, 66.9, 51, 35, 17.5, 8, 30.4, 30.1, 29.9]]
test7 = [[73, 62.5, 42, 59, 46.9, 26, 81, 58.9, 31, 31, 21.3, 7, 30.4, 29.9, 29.7]]
test8 = [[43, 39, 36, 35, 30.7, 24, 82, 72.6, 53, 17, 10, 3, 30.5, 30.4, 30.3]]
test9 = [[42, 39.8, 37, 34, 32.8, 32, 83, 76.1, 67, 8, 5.9, 0, 30.4, 30.3, 30.1]]
test10 = [[47, 40.3, 32, 38, 28.6, 12, 82, 64.4, 41, 28, 13.3, 0, 30.1, 29.8,
29.6]]
```

```
# Make predictions on the testing values
pred1 = model.predict_proba(test1)
pred2 = model.predict_proba(test2)
pred3 = model.predict_proba(test3)
pred4 = model.predict_proba(test4)
pred5 = model.predict_proba(test5)
pred6 = model.predict_proba(test6)
pred7 = model.predict_proba(test7)
pred8 = model.predict_proba(test8)
pred9 = model.predict_proba(test9)
pred10 = model.predict_proba(test10)

# Print the predicted probability
prediction1 = pred1[0][1]
print(round(prediction1, 2))
prediction2 = pred2[0][1]
print(round(prediction2, 2))
prediction3 = pred3[0][1]
print( round(prediction3, 2))
prediction4 = pred4[0][1]
print( round(prediction4, 2))
prediction5 = pred5[0][1]
print( round(prediction5, 2))
prediction6 = pred6[0][1]
print( round(prediction6, 2))
prediction7 = pred7[0][1]
print(round(prediction7, 2))
```

```
prediction8 = pred8[0][1]
print( round(prediction8, 2))
prediction9 = pred9[0][1]
print(round(prediction9, 2))
prediction10 = pred10[0][1]
print( round(prediction10, 2))
```

3.3 Experimental Design

1. Input the following data points, from January 2015-2019, organized in the spreadsheets

- This will be the learning data; the model will identify trends and patterns from this data.

2. After 10 days inputted, pull only the input data values from a day in 2020. Have the model return a precipitation decimal value (0.0 – 1.0) and note down this value.

3. Repeat step 2 3 times, so there are a total of 3 predictions for 10 days of learning data.

4. Repeat steps 2-3 for every 10 days of learning data inputted. This should happen 15 times as there are 150 days of learning data.

6. Repeat steps 1-5, for all 5 models.

7. Organize the data into a chart: Inclusion of the date of the day used is recommended. Include the actual outcome in parentheses: (1 or 0).

8. Repeat for each model

3.4 Learning and Progressive Testing

Measuring the models progressive testing is important as it gives us insight into how the models perform given more data. This helps us understand the model's learning patterns and the progressive nature of the model's accuracy. A concern of this process would be the fact that we are inputting the same days of the same values and asking the machine to predict what is essentially the same precipitation factor for all of the progressive testing done. However, this is not significant because for every progressive test a new model is inherently being created which has no access to the previous model's progressive testing data. Thus, for each progressive test the tests are new and never seen by the respective computer model.

3.5 Evaluation Metrics

To express the accuracy of a precipitation prediction in a single metric the following equations will be used to represent accuracy:

If the actual precipitation value is 0

$$1 + (0 - x) \rightarrow 1 - x$$

This allows for values closer to 0 to receive a score closer to 1, meaning higher accuracy. On the other hand, scores closer to 1 receive an accuracy score closer to 0.

If the actual precipitation value is 1

$$1 + (0 - x) \rightarrow x$$

This allows for values closer to 1 to receive a score closer to 1, meaning higher accuracy. On the other hand, scores closer to 0 receive a subsequent accuracy score closer to 0.

These equations allow us to see the accuracy of the models in correctly predicting whether the following day will have any levels of precipitation. The equations will be used to address the accuracy for all of the machine learning models.

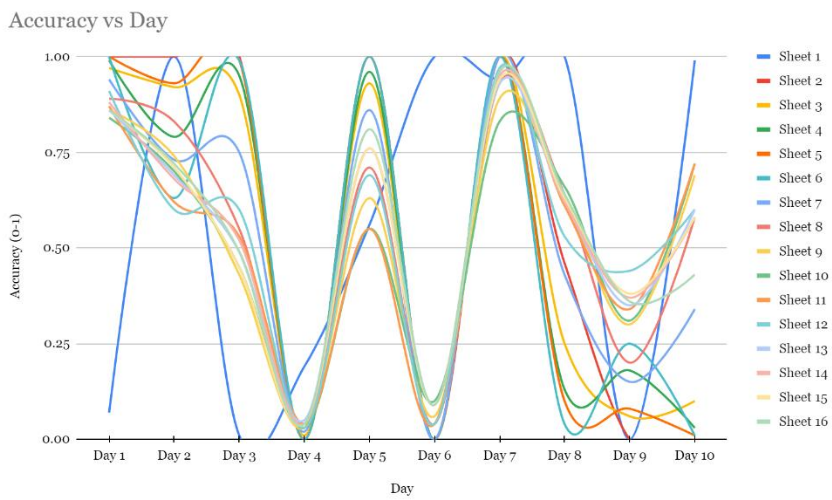
After collecting the data and determining the accuracy, the slope will be determined by using a linear line of best fit.

Using a linear line of best bit for each of the days allows us to understand the data better, specifically its trend throughout the 16 sheets. We are able to roughly identify which days responded in a greater accuracy and which ones responded with less accuracy for more data provided.

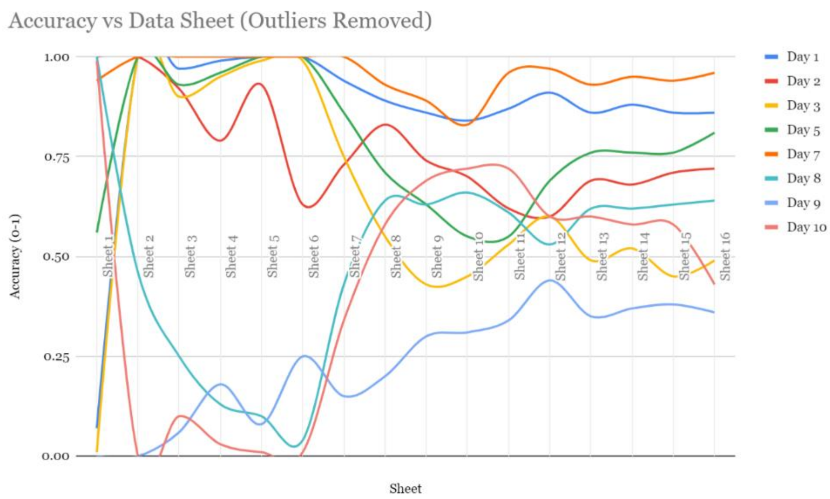
Also, by using a Accuracy vs Day chart we can check for any outlier days which served to be a challenge for the machine to predict. I will eliminate the two days with the lowest accuracy, thus leaving eight days to remain for comparative analysis.

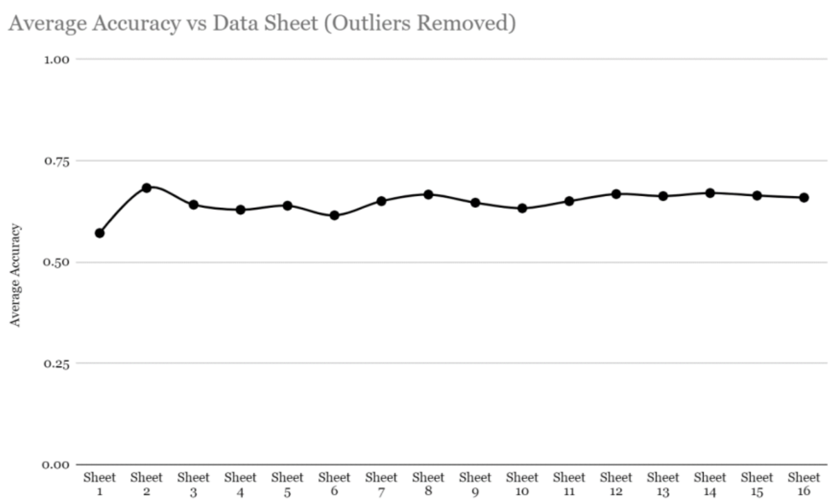
4 Experimental Results.

4.1 Logistic Regression Data (LR)



Based on the results from the Accuracy vs Day graph, it can be determined that Days 4 and 6 were extremely difficult for the logistic regression model to predict accurately. Therefore, in alignment with the experimental methods I will cut out the two days with the worst overall accuracy.

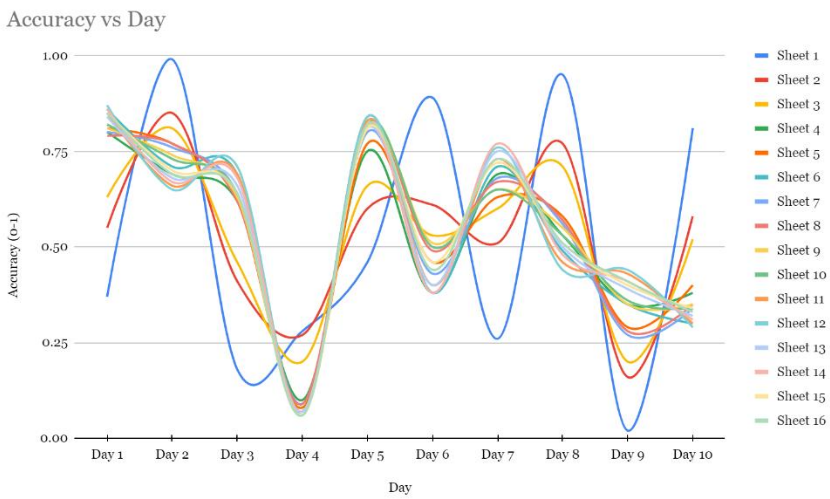




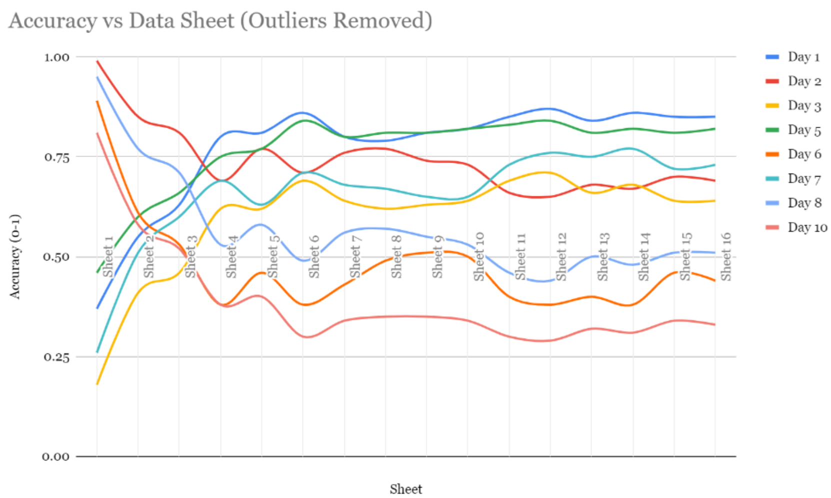
With the outliers removed, we can make our final data points which we will use to compare the progressive learning of the different machine learning models.

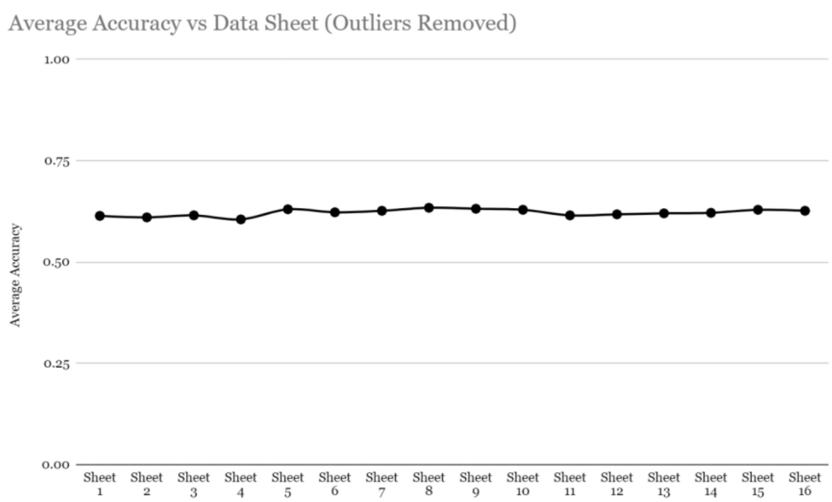
- The average across all sheets and all days was an accuracy of 0.6465625 or 64.65625%
- Sheet 2 had the highest accuracy of 0.6825 or 68.25%
- Sheet 1 had the least accuracy of 0.57125 or 57.125%

4.2 Support Vector Machine Data (SVM)



Based on the results from the Accuracy vs Day graph, it can be determined that Days 4 and 9 were extremely difficult for the logistic regression model to predict accurately. Therefore, in alignment with the experimental methods I will cut out the two days with the worst overall accuracy.

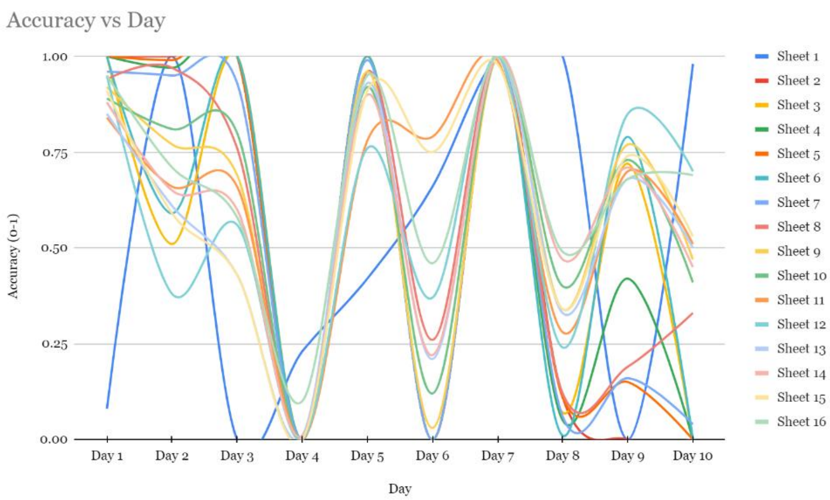




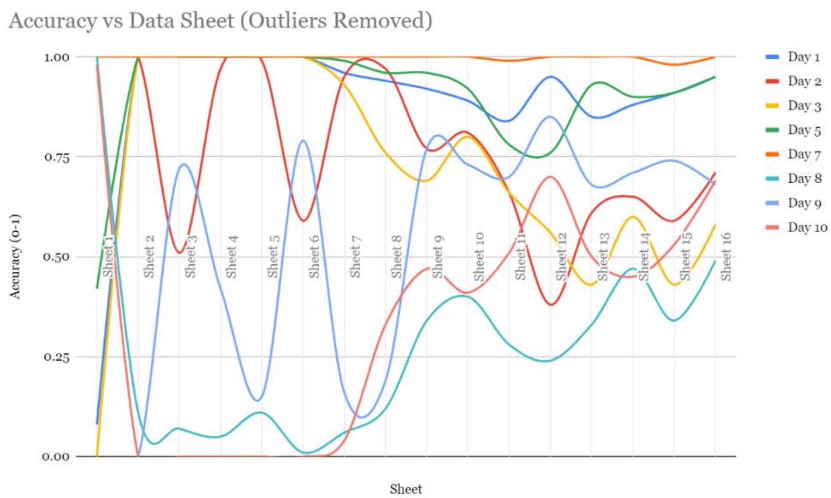
With the outliers removed, we can make our final data points which we will use to compare the progressive learning of the different machine learning models.

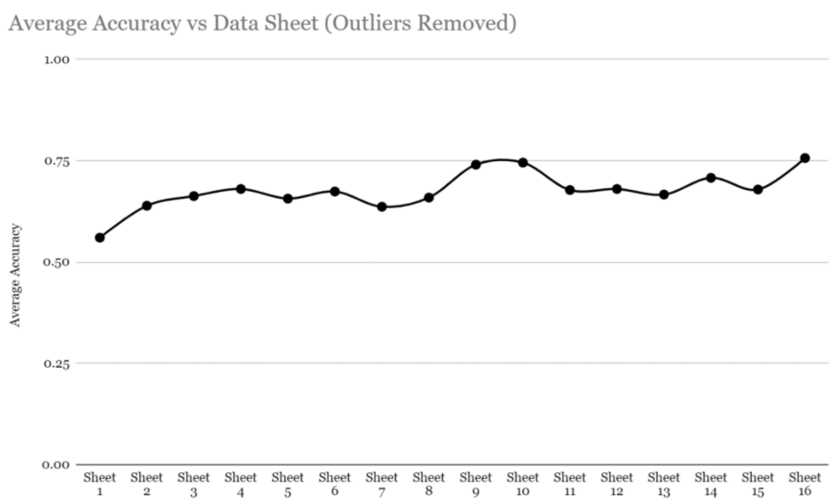
- The average across all sheets and all days was an accuracy of 0.6215625 or 62.15625%
- Sheet 2 had the highest accuracy of 0.63375 or 63.375%
- Sheet 1 had the least accuracy of 0.605 or 60.5%

4.3 Multilayer Perceptron Mode (MLP)



Based on the results from the Accuracy vs Day graph, it can be determined that Days 4 and 6 were extremely difficult for the logistic regression model to predict accurately. Therefore, in alignment with the experimental methods I will cut out the two days with the worst overall accuracy.



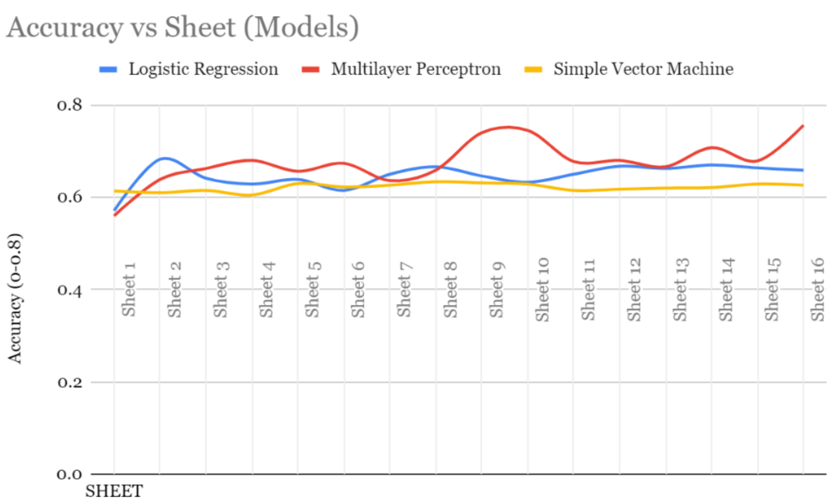


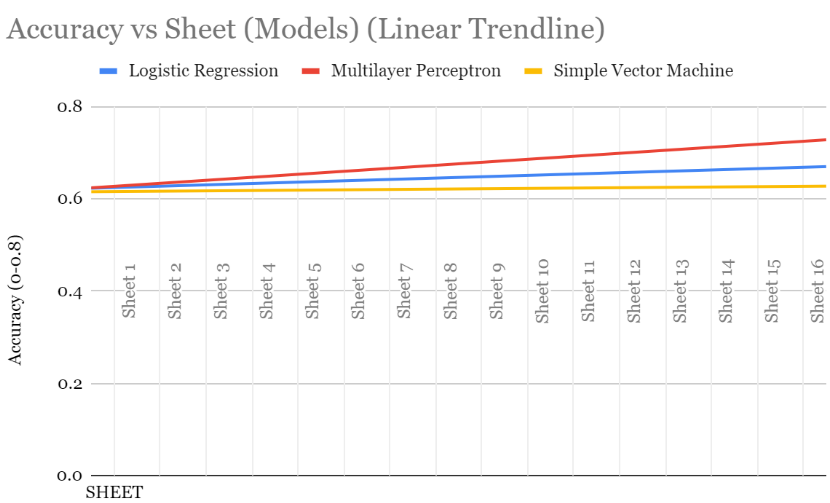
With the outliers removed, we can make our final data points which we will use to compare the progressive learning of the different machine learning models.

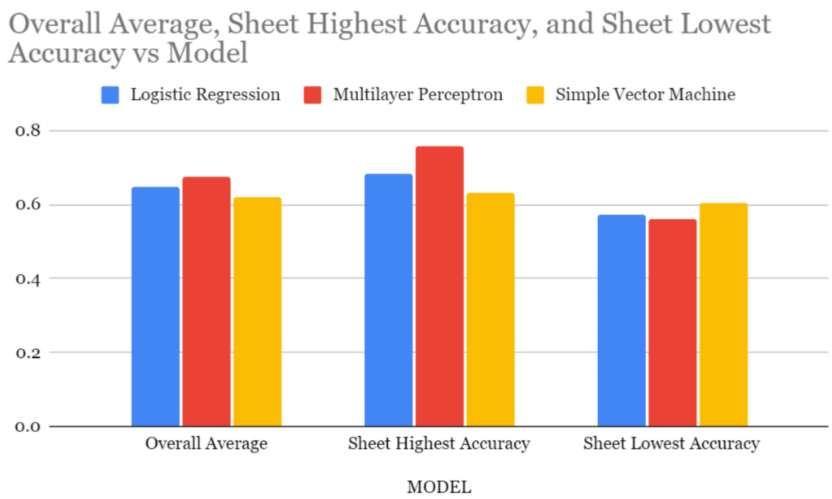
- The average across all sheets and all days was an accuracy of 0.67609375 or 67.609375%
- Sheet 16 had the highest accuracy of 0.75625 or 75.625%
- Sheet 1 had the least accuracy of 0.56 or 56.0%

5 Discussion.

5.1 Model Comparison







5.2 Model Effectiveness Analysis

In the realm of the machine's overall accuracy, it is evident that the Multilayer Perceptron model outperformed the other models. This is as the MLP model had the higher overall accuracy of 67.6%, on the other hand the SVM exhibited the lowest overall accuracy of 62.2%.

When looking at the highest accuracy for an individual sheet, MLP, again, had the greatest accuracy with an accuracy of 75.6%. This is more than 16% greater than SVM, which had the lowest sheet accuracy, at 63.4%, and even a significant percentage above the LR model. Notably, the MLP had its greatest sheet accuracy on the final sheet which had the most training data.

Overall, all three models generally experienced an increase in accuracy as the amount of training data increased. However, it is apparent that the MLP model increased to the greatest degree compared to the SVM which improved to a much more minimal degree. In between then, the LR model remained relatively consistent, being almost always in the middle of the two models.

In conclusion, based on all the considered metrics, the Multilayer Perceptron Model performed the strongest compared to the Logistic Regression and the Support Vector Machine models in the field of meteorology.

5.2 Model Improvement Analysis

I also wanted to observe how the different machines evolved while they were given more and more training data. As I expected the models to perform better with all the data (Sheet 16) than compared to the minimum amount of data (Sheet 1), thus I compared each machine's Sheet 16 versus their Sheet 1 accuracy. I also compared each machine's slope of their respective linear trendline.

The model with the least improvement over the 16 sheets was the Support Vector Machine which had a difference of +1.25%. Its slope was also the lowest of the three machines at 0.000757 which puts it very close to a slope of 0, which would mean no change. The Logistic Regression machine was next with a difference of +8.75% and a slope of 0.00293. Finally, the machine with the most improvement was the Multilayer Perceptron model with an improvement of 19.63% and a slope of 0.00651 .

6 Conclusion.

6.1 Summary and Reflection

I used three different machine learning models to test their effectiveness for making weather predictions and their evolution given more training data. The weather data was taken from the Boston Airport weather station for multiple consecutive years in the month of January. This was done so as not to have seasons make a significant difference in the model's accuracy. After compiling the data, the Multilayer Perceptron model performed the strongest for most of the categories. The Logistic Regression model performed consistently between the other two models, while the Support Vector Machine ended up performing the lowest compared to the other two models. Despite this, the SVM still ended up with an average accuracy of 62.15% which is impressive considering that the data of only five months across five years was given. The best-performing model, the Multilayer Perceptron model, had an average accuracy of 67.61%. In the scope of the improvement of the machines, the Multilayer Perceptron machine again showed the most improvement. The MLP gained 19.63% while the LR and SVM machines gained 8.75% and 1.25% respectively.

Importantly, these observation do come with some important implications and setbacks. For example, I believe it would be optimal to use even more data and introduce more data points and observe the accuracy of the model then. I also need to mention that the model had to simply predict whether or not it rained that day. I believe it would be interesting to experiment with varying rain amounts and see how accurately the model could predict the amount of rain that rained the next day given the appropriate parameters. Furthermore, it is important to mention that a few of the days I used for testing could have been outliers compared to the rest of the data. Despite attempting to reduce the impact of these days by removing two outliers, there could still be remaining outliers which could negatively impact the data. It is important to consider that weather is not always predictable and always has a level of uncertainty to it.

Despite these setbacks, I personally find these findings very exciting and impressive. With more data, I believe the machine learning models would perform at even higher accuracies given the improvement observed. I believe it would be most effective to use the MLP model as it performed the best all-around, however, the other models also provide a solid option. Given how much more data is available, I do believe that with access to even more training data, the machine learning models might even challenge the accuracy of the well-established Numerical Weather Prediction (NWP) models.

6.2 AI's Role in Meteorology

This potential of AI and Machine Learning models has not gone unnoticed, several different companies have begun to produce more complex models to predict weather patterns. The MIT Technology Review elaborates that prominent tech companies including “Nvidia, Google, DeepMind, and Huawei” have shown interest or even begun development of their own “machine-learning methods” that have the capability to perform at a level of even the conventional methods [13]. However, despite these promising results, a few complications remain that keep AI from making a proper appearance into meteorology. The Washington Post highlights a few of these complications in an article, including constantly changing weather, a potential for bias, difficulties in bringing AI to an operational status, and a lack of trust and transparency [13] [14]. However, given the relatively recent introduction of AI, it could just be a matter of time until these issues are resolved, which would allow for the integration of AI into weather prediction models.

References

[1] Numerical weather prediction (weather models) - National Weather Service. (n.d.). https://www.weather.gov/media/ajk/brochures/NumericalWeatherPrediction.pdf

[2] How has weather forecasting changed over the past two hundred years?. American Geosciences Institute. (2016, November 17). https://www.americangeosciences.org/education/k5geosource/content/weather/how-has-weather-forecasting-changed-over-the-past-two-hundred-years

[3] DeLoss, J. (2021, October 28). From research to real world: CSU atmospheric scientists develop heavy rainfall forecast tool used nationwide. Walter Scott, Jr. College of Engineering. https://engr.source.colostate.edu/from-research-to-the-real-world-csu-atmospheric-scientists-develop-heavy-rainfall-forecast-tool-used-nationwide/

[4] Manning, A. (2023, March 1). CSU machine learning model helps forecasters improve confidence in storm prediction. Walter Scott, Jr. College of Engineering. https://engr.source.colostate.edu/csu-machine-learning-model-helps-forecasters-improve-confidence-in-storm-prediction/

[5] Paniagua-Tineo, A. & Salcedo-Sanz, Sancho & Casanova, Carlos & Ortiz-García, Emilio & Cony, Marco & Hernández-Martín, E.. (2011). Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy. 36. 3054-3060. 10.1016/j.renene.2011.03.030.

[6] Ibrahim Gad & Doreswamy Hosahalli (2022) A comparative study of prediction and classification models on NCDC weather data, International Journal of Computers and Applications, 44:5, 414-425, DOI: 10.1080/1206212X.2020.1766769

[7] Anjali, T., Chandini, K., Kadan, A., & Lajish, V.L. (2019). Temperature Prediction using Machine Learning Approaches. 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 1, 1264-1268.

[8] Cho, D., Yoo, C., Im, J., & Cha, D.-H. (2020). Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Science, 7, e2019EA000740. https://doi.org/10.1029/2019EA000740

[9] natesilver538. (2014, December 4). Which city has the most unpredictable weather?. FiveThirtyEight. https://fivethirtyeight.com/features/which-city-has-the-most-unpredictable-weather/

[10] Carylsue. (2017, May 31). What cities have the most unpredictable weather in the U.S.?. National Geographic Education Blog. https://blog.education.nationalgeographic.org/2017/05/31/what-cities-have-the-most-unpredictable-weather-in-the-u-s/

[11] East Boston Weather History. Weather Underground. (n.d.). https://www.wunderground.com/history/daily/us/ma/east-boston/KBOS

[12] McKinney, W. (2008). Pandas. pandas. https://pandas.pydata.org/

[13] Heikkilä, M. (2023, July 27). Weather forecasting is having an AI moment. MIT Technology Review. https://www.technologyreview.com/2023/07/11/1076067/weather-forecasting-is-having-an-ai-moment/#:~:text=AI%20is%20proving%20increasingly%20helpful,methods%2C%20and%20much%20more%20quickly.

[14] Stillman, D. (2023, July 5). Should we trust ai to predict natural disasters?. The Washington Post. https://www.washingtonpost.com/weather/2023/07/04/ai-weather-forecasts-hurricanes-tornadoes/