Forecasting Temperature Using AI Linear Models

By Vikram Kavalipati

Abstract:

Currently, weather forecasting is done primarily using traditional computer models. Artificial intelligence (AI) would allow predictions to be made much more easily with far less effort. However, AI weather models are still less accurate than traditional computer models. The goal of this project is to explore the possibility of using artificial intelligence in forecasting and to gain an understanding of its potential and limitations while creating an accurate weather forecast. The models use a 70 MB dataset that is in json format. The dataset includes a variety of factors including min and max temperature, wind speed and direction, air pressure, and precipitation for more than 50 years. This paper uses a wide variety of regression models to forecast the next day's temperature for Roseville, CA, including Linear Regression, Ridge Regression, Lasso etc. Linear regression was consistently the best performing model, with a mean absolute error of 2.29 degrees, suggesting that even simple linear models can produce weather forecasts that are generally close to the correct value.

1.0 Introduction:

Weather forecasts are crucial for the day to day lives of many people. These forecasts are usually used to predict temperature and precipitation. Current weather forecasts are based on numerical weather forecasting (NOAA).. Numerical weather forecast models rely on physics equations describing the movement of fluids.. The models then use a variety of approximations and numerical methods to convert the equations into something that computers can use, as computers cannot perform calculus. The forecast is then created by massive supercomputers that provide an approximate calculation of the forecast.

However, numerical weather forecasting has its limitations. One problem is its reliance on approximation. The equations used in numerical weather forecasting are usually approximations of the real calculus based equations. Another limitation of numerical weather forecasting is that it is impossible to know every single detail of how fluids move, and all the details of how the atmosphere functions aren't fully known. Another limitation of traditional weather forecasting is the range at which it is viable, as it often struggles to make predictions for very short and long periods of time. By improving weather forecasts, people will be able to more accurately and efficiently predict storms, where rain will fall, and which locations will be most in danger from a disaster. They will also allow weather forecasts to extend much farther out into the future allowing people to become more efficient by being able to plan for the future.

1.1 AI Advantages:

One solution to address the limitations of numerical weather forecasting is utilizing AI in forecasts. AI weather forecasts don't need physics or equations to make their predictions. Instead, they rely on past weather data and use it to train their own models that predict the weather. These models might work better than conventional physics equations, as these models might be able to take variables into account that would never be considered by those equations, and they might be able to identify patterns that are unknown to humans. Because the models don't have the limitations of conventional equations, they might be better at forecasting the weather for very long and short time frames. Another advantage of AI forecasting is its efficiency. Traditional forecasting requires massive supercomputers to perform all of its calculations, while some AI weather forecasting models can alternatively be done very quickly on a normal laptop.

1.2 Models:

Weather forecasting can be attempted using a wide variety of possible AI models. However, this project will focus on using regression based models to predict the next day's temperature. Regression is the process in which AI models use data to produce a numerical estimate of the output. The goal of these models is to get as close to the desired output value as possible. The project will compare the effectiveness of linear regression, ridge regression, bayesian ridge, lasso, lasso lars, elastic net, and orthogonal matching pursuit in predicting the next day's weather.

2.0 Methodology:

This section outlines the models used in order to generate weather forecasts and the nature of the dataset used to train and test the models. It also discusses how the performance of the model is evaluated.

2.1 Dataset:

The data used to trian and test the model is based on the Visual Crossing API. The data is for Roseville CA, and covers every day from Jan 1, 1970, to Dec 31, 2022. The data consists of a large number of variables, many of which are used in the prediction of the temperature. The necessity of including variables related to the temperature is relatively clear. The humidity and dew point were included as humidity levels often correlate with certain types of temperatures, depending on the location. A higher humidity level can mean lower temperatures, while a higher dew point can sometimes mean a higher temperature. The pressure and wind speed and direction are included because they can indicate the arrival of a certain type of weather system, which can cause a change in temperature. The precipitation and precipitation probability are included as they may impact temperature. The dataset also includes rolling averages of the temperature from the past two to seven days.

The data is split into a train set to train the models and a test set to test their performance. In the model, 80% of the data is part of the training set and 20% was part of the test set. The data is split into chronological order, with the first 80% of values being in the training set and the last 20% being in the test set. This is done to better simulate weather forecasting, in which past data is used to predict future weather.

2.2 Models:

The following section provides a brief description of each of the AI models used to forecast temperature.

2.2.1 Linear Regression:

Linear Regression is used to create a relationship between an independent variable and a dependent variable. The Linear Regression used in the model is multivariable Linear Regression, with each variable being multiplied by a coefficient. The sum of the products of the variables and their coefficients, along with an additional constant term, is used to calculate the prediction of the model (Hastie et al.). In Linear Regression, the coefficients are estimated by minimizing the sums of the square differences between the estimated value and the actual value for each data point.

2.2.2 Ridge Regression:

Ridge Regression is a linear model, like Linear Regression. It also has one coefficient for every variable and a coefficient added at the end (Hastie et al.). It then seeks to optimize the estimate.

Ridge Regression does aim to minimize the sum of the errors squared, just like Linear Regression. However, the model also takes into account a shrinkage penalty . The shrinkage penalty adds the squares of all of the coefficients and multiplies this value by a constant, the tuning parameter. The purpose of the penalty is to penalize large coefficients. The advantage of the shrinkage penalty is that it can help reduce overfitting in the model and improve performance in the test set.

2.2.3 Lasso Regression:

Lasso Regression, a.k.a. Lasso, is similar to Ridge Regression. Lasso is also a linear model and it also uses a shrinkage penalty to shrink the values of coefficients (Hastie et al.). In Ridge Regression, the model always takes into account every single variable. This is the case even if the dataset has many variables, like the dataset in our model. However, in most datasets with many variables, only a few of the variables are actually important. Lasso models work by eliminating all of the variables except for the ones that are very important .

2.2.4 Bayesian Ridge Regression:

Bayesian Ridge Regression is very similar to ridge regression. Bayesian Ridge involves all of the same coefficients as the other types of regression, and it also has a shrinkage penalty just like Ridge Regression (Hastie et al.). However, it is different from Ridge Regression because the value of the penalty is not fixed. In Bayesian Regression, the model automatically assumes a distribution of coefficients beforehand. The coefficients, in the form of a vector, is then multiplied by a vector denoting the likelihood of that distribution . This value is then used to calculate the penalty. This form of regression is better at taking into account the likelihood of a given distribution.

2.2.5 Elastic Net:

Elastic Net is a form of regression similar to Lasso. Just like lasso, Elastic Net has the advantage of eliminating all of the unimportant variables, especially in large datasets with many variables (Wieringen). However, the Elastic Net model can often be more effective than Lasso Regression as Lasso only selects a small number of features while completely ignoring the rest, while Elastic Net takes more features into account. This model does this by taking into account the penalties of both Lasso and Ridge Regression . However, the model is still often superior to Ridge Regression as it performs feature selection instead of looking at every single feature.

2.2.6 Lasso LARS:

LARS stands for Least Angle Regression. In this model, all of the coefficients start off at zero (Efron et al.). The coefficients are treated as one large vector,and the direction of the vector determines the value of the coefficients. The model starts by finding the most highly correlated variable and taking a step in that direction. The model moves it that direction until another variable becomes equally correlated with the output, and the model then moves in a direction that is equiangular between the vectors. This process goes on for every variable that the model finds to be correlated with the output . Lasso LARS also has the same method of selecting features as Lasso Regression and it often only uses a few variables and is designed to penalize the model for having large coefficient values.

2.2.7 Orthogonal Matching Pursuit:

Orthogonal Matching pursuit, much like Lasso LARS, starts off with all variables set to zero (Blumensath and Davies). The model first calculates the residual, or the error between the predicted and actual value, based on prediction using the current dataset using the current variables. The model then finds the most highly correlated variable with the current residual value by finding the dot product between the residual and each variable. The model then updates the value of the coefficient based on a method that varies based on the model. Then, the residual value is recalculated with the change in the coefficient and the same process is repeated. This occurs until the residual value reaches a point of convergence, and then the training process is complete.

2.3 Normalization:

One thing that can have a significant impact on the ability of the models to predict temperature is normalization. Normalization is the process of adjusting taking variables in a dataset that are measured on different scales, such as temperature and precipitation, and converting them to the same scale. This is doen by taking every single value of a given variable and dividing it by the largest value of that given variable in the dataset, making all of the values for every variable fall between 0 and 1.

2.4 Mean Absolute Error:

The performance of each model is determined by its MAE, or mean absolute error. The MAE is calculatied by taking the average of the absolute value of the difference between the predicted temperature and the actual temperature for every day in the test set.

3.0 Results:

The following section shows the performances of every single model with and without normalization. It also discusses the importance of the variables for different models. Lastly, it contains an evaluation of the best performing model to check for biases in the prediction.

3.1 Average MAE of Models:

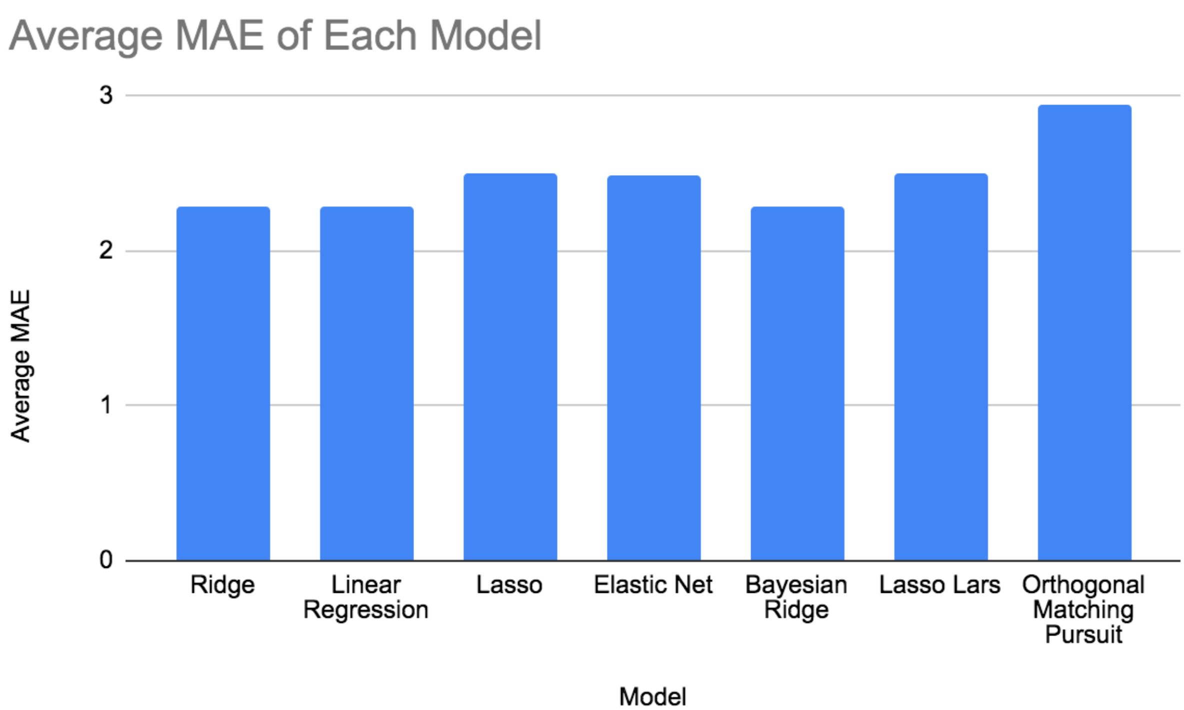


Figure 1- The average MAE of every linear model in degrees Fahrenheit

The performance of the models was measured by their mean absolute errors. The error was calculated based on a set of testing data. The testing data consists of the last 20% of the total data when the data was arranged in chronological order. The graph(Fig 1) shows the Mean Absolute Error of every single model. The model with the smallest MAE is Linear Regression, with an MAE of about 2.29 degrees Fahrenheit. Ridge Regression and Bayesian Regression also have an MAE of around 2.29, but their errors are just a small fraction of a degree above the error in Linear Regression. Lasso, Lasso LARS, and Elastic Net all have an average error of around 2.5 degrees Farherneit. The model with the highest error is Orthogonal Matching Pursuit, with an average error of around 2.94 degrees.

3.2 Average MAE of Models with Normalization:

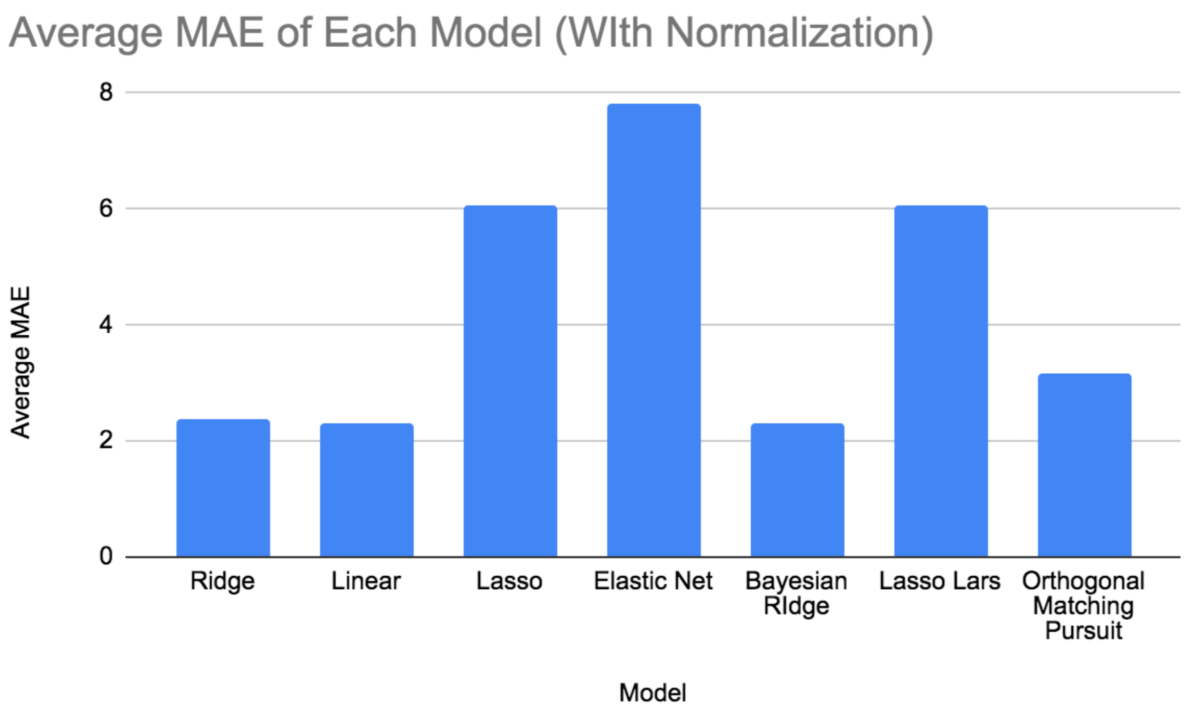


Figure 2- The average MAE of every linear model after the data was normalized

The effect of normalization on the models is shown in figure 2. Normalizing the data proves to negatively impact the performance of most of the models. The only model to not be impacted by normalization is Linear Regression, which has the exact same MAE after normalization. This is likely because Linear Regression has no penalty to prevent overfitting, so the magnitude of the values of the features doesn't affect the prediction. The MAE of the Bayesian Ridge Regression model increases only very slightly after normalization. The Ridge Regression model has its MAE increased from around 2.29 to around 2.37 degrees Fahrenheit. The Orthogonal Matching Pursuit model had an increase in its MAE from 2.95°F to 3.15°F. The three models that are most negatively impacted by normalization were Lasso, Lasso LARS, and Elastic Net. The MAEs of Lasso and Lasso LARS increase from around 2.51 to around 6.06 degrees Fahrenheit. Normalization has a very lage negative impact on the Elastic Net model, which has its MAE increase from 2.49 to 7.83 degrees Fahrenheit.

3.3 Linear Regression Feature Importance:

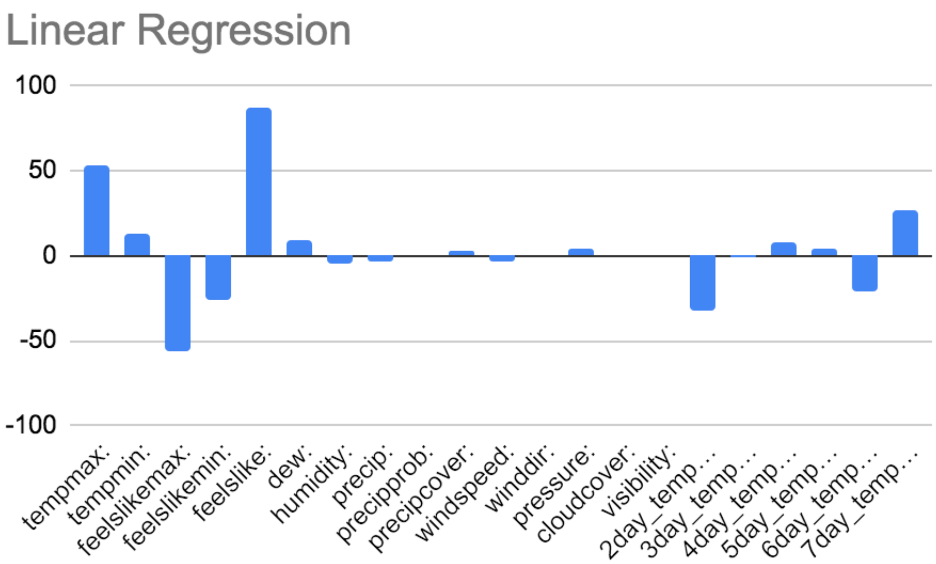


Figure 3- The feature importances of every feature in the linear regression model, as measured by their coefficients.

The feature importances of the models can tell us how important every piece of information about the weather is for the forecasting of temperature. They can also give us more insight into the inner workings of the model. In a linear model, the output is calculated by multiplying every single variable by a coefficient, taking the sum of all of these products, and adding some other terms to get the output. Therefore, the value of the coefficients of each feature is directly correlated with the importance of that feature for that particular model. The data values should also be normalized prior to extracting their coefficients to avoid bias resulting from some of the variables having larger average values.

While there were differences between models, the maximum daily temperature on the previous day (tempmax) was a key feature in many of the models. The Linear Regression model used all of the features, however, the feels like temperature, maximum temperature, and maximum feels like temperature had the greatest impact on the model. This result is shown in figure 3. Ridge Regression and Bayesian Ridge have a very similar set of coefficients to those of ridge. The Lasso model only considers maximum temperature and the average temperature over the previous 7 days in its forecast, and it places much greater weight on maximum temperature. The weight associated with each parameter in the Lasso LARS model is similar to the weight associated with each parameter in Lasso. The features have unique relative importance values in the Elastic Net model, as every single feature has a similar weight in the model, except for a few features that weren't considered.

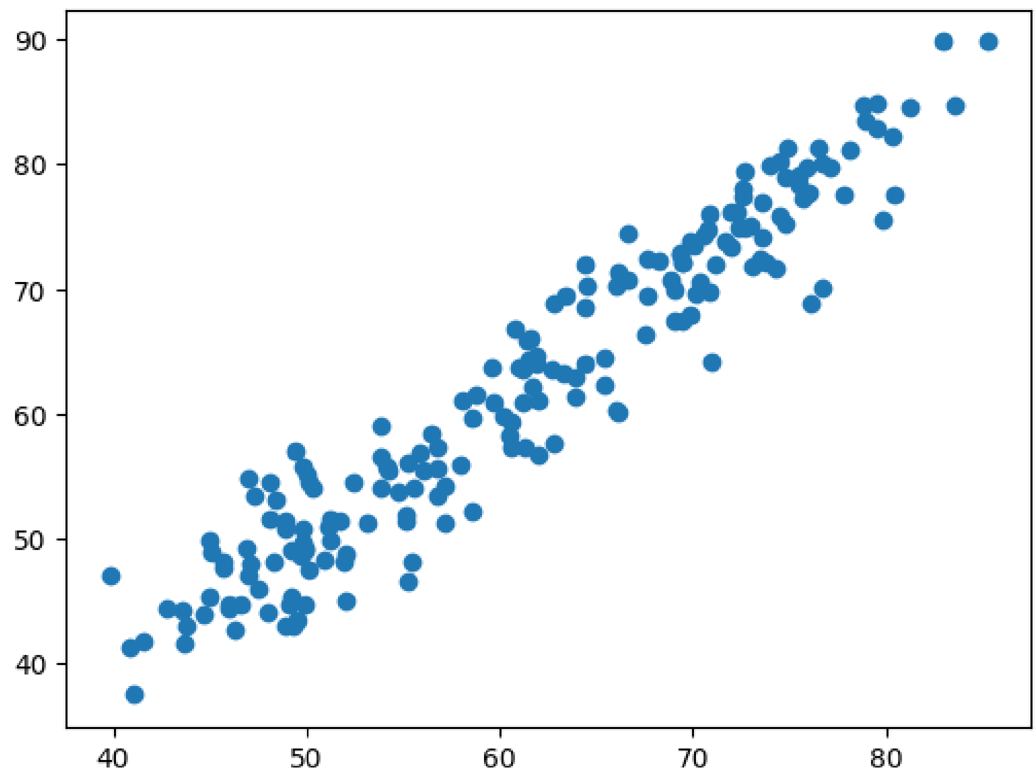
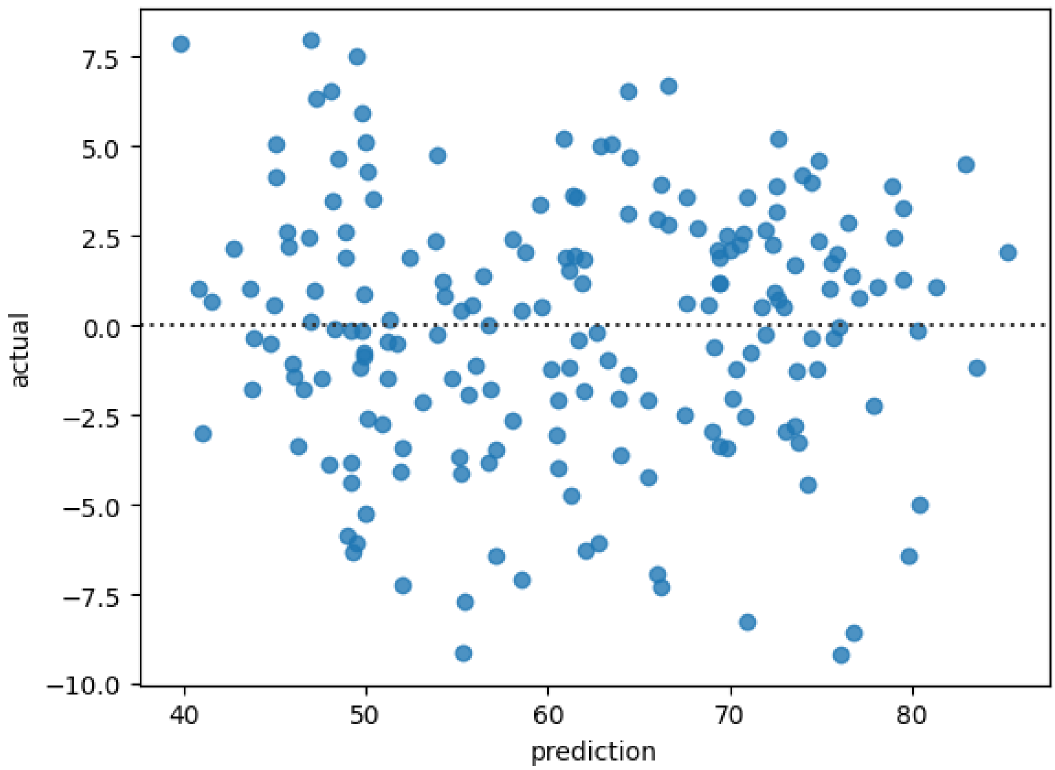


Figure 4- The predicted vs actual plot of a random sample of predictions from the test dataset, showing the predicted temperature on the x axis and the actual temperature on the y axis.

Error of Prediction(°F)



Predicted Temperature (°F)(linear regression)

Figure 5- A residual plot showing the predicted temperature and the the error of the prediction for a set of random data points in the test set.

Figures 4 and 5 can be used to demonstrate that the Linear Regression model has very little bias in its prediction. In figure 4, the predicted temperature vs actual temperature plot forms a line with a near 1 to 1 correlation. This shows that the predicted temperature is strongly correlated with the actual temperature, and this means that the model is performing well and is able to fit the data effectively. Figure 5, a residual plot showing the average error of predictions made at different temperature values. The plot shows that the model performs similarly at low and high temperatures and is not biased in this regard.

4.0 Conclusions:

Artificial Intelligence models can be used to predict the temperature of the next day with an MAE of as low as 2.29 degrees Fahrenheit. The performance of an AI based weather forecast can vary heavily based on the model that is used to create the weather forecast. Linear Regression is consistently the best performing model, with a mean absolute error of 2.29°F for temperature prediction. The performance of the forecast varies significantly depending on which model is used. The performance of some of the models also varies heavily based on whether or not the data is normalized, with some models performing much less effectively after normalization. Models like Linear Regression, are largely unaffected by normalization.

The models also have varying performances depending on what features they emphasize more in their predictions. The models that are the most successful seem to take all features into account but focus more on the few features that are most closely correlated with temperature. Many of the models that perform poorly eliminate too many features, including features that could be very helpful in determining temperature. The Elastic Net model struggles because it seems to weigh every feature almost equally, except for the few it eliminates. The data seems to suggest that a good forecast can be attained by taking many variables into account but paying more attention to a few, very important variables. The Linear Regression model relies primarily on the weather from the previous day to forecast weather. However, the model also uses the average temperature from the previous seven day to predict the next day's weather. This suggests that weather we experience today can help us determine what the weather will be one week from now. Therefore, Artificial Intelligence models may be able to use data we have today to provide insights on weather events including severe storms one week in advance.

My results suggests that even simple linear models can produce weather forecasts that are generally close to the correct value, and these weather forecasts can be produced at a fraction of the time and cost of numerical weather forecasting. However, a mean absolute error of 2.29°F suggests that the model is still far from perfect and can be off significantly on some days.

The analysis yields many other interesting results as well. All of the models other than Linear Regression have some type of penalty, whose job it is to penalize large coefficient values. The goal of this penalty is to prevent overfitting in the model, which should lead to better performance in the test data but worse performance in the training data. However, Linear Regression has the best performance in the test dataset despite not having any correction for overfitting, suggesting that overfitting was not a major concern in this case. Another reason why the results are surprising is the complexity of the models. People typically assume that complex model and algorithms are more effective. However, in this case, we see that Linear Regression, the simplest of all the models, is the best performing model, suggesting that simplicity is sometimes better in models.

There are several next steps that can be taken to attempt to improve the models. One step is to test the impact of adding data from neighboring locations to the model in order to help the model gain a more holistic view of the weather. This can include data from the ocean such as ocean temperatures. Another next step would be to attempt to use Artificial Intelligence to forecast precipitation. A similar paper could be written comparing all of the linear models to determine which model can most effectively predict the amount of precipitation that will occur the next day. Another next step would be to extend the models to predict the weather three days, a week, or two weeks from now, or to predict the weather a couple hours from now.

Citations:

1. Blumensath, Thomas. “On the Difference between Orthogonal Matching Pursuit and Orthogonal ...” *IEEE*, 2007, eprints.soton.ac.uk/142469/1/BDOMPvsOLS07.pdf.
2. Efron, Bradley. “407 - Donuts Inc..” *Stanford*, 2004, tibshirani.su.domains/ftp/lars.pdf.
3. Hastie, Trevor. *An Introduction to Statistical Learning*, 2021, hastie.su.domains/ISLR2/ISLRv2_website.pdf.
4. “Numerical Weather Prediction (Weather Models) - National Weather Service.” *National Weather Service*, www.weather.gov/media/ajk/brochures/NumericalWeatherPrediction.pdf. Accessed 17 Sept. 2023.
5. Visual Crossing API- https://www.visualcrossing.com/weather-data
6. Wieringen, Wessel. “Lecture Notes on Ridge Regression.” *Arxiv*, 2023, arxiv.org/pdf/1509.09169.pdf.