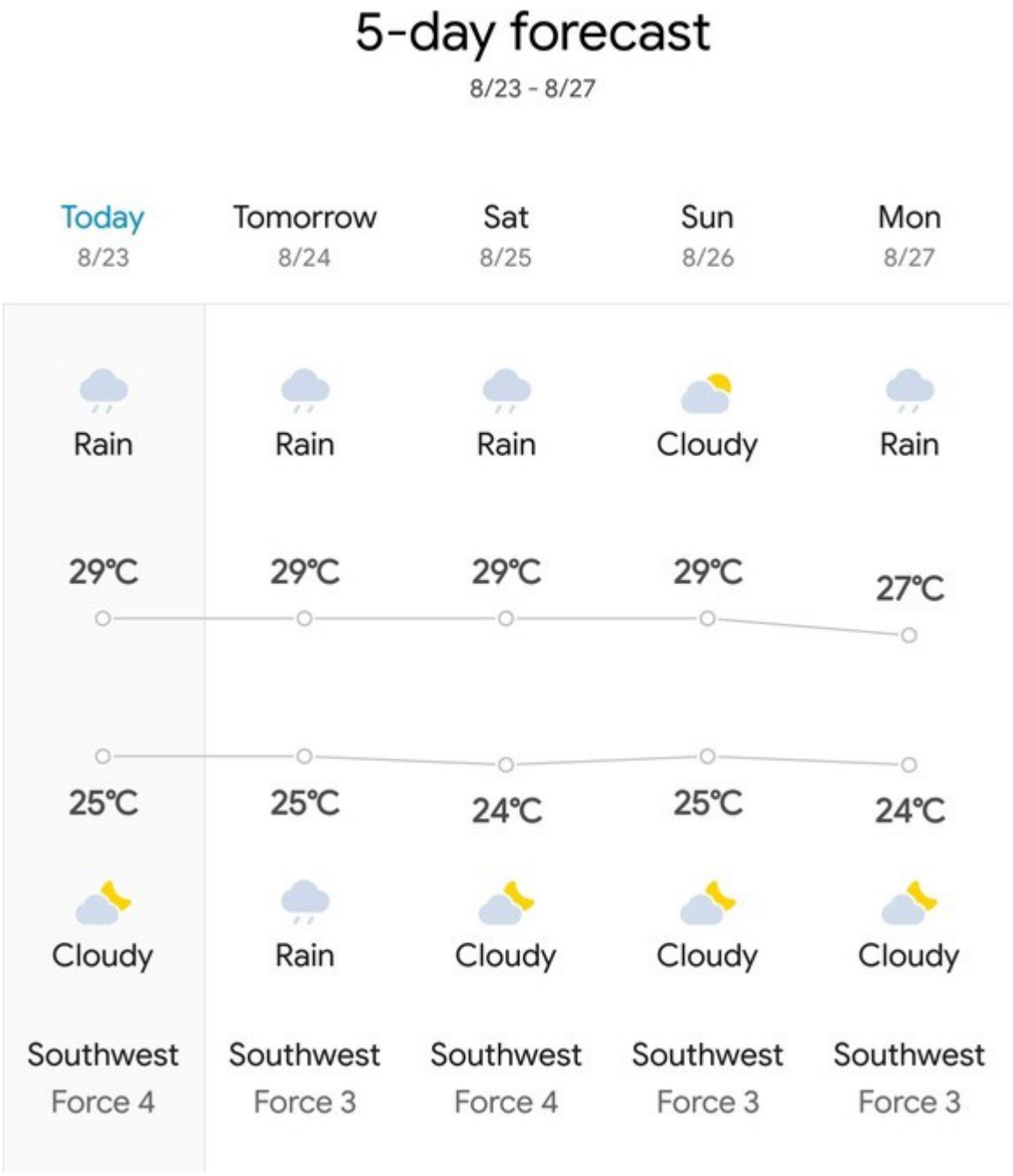
What is the weather prediction algorithm? How it works? What is the future?

Shivam Trivedi Follow
Aug 23, 2018 · 5 min read



Weather prediction aka weather forecasting is done by powerful supercomputers which process hundreds of thousands of observations of current weather conditions. To observe the current weather we have launched thousands of satellites into space. Those satellites travel across the earth and capture the data which, then, send to the data centers established on earth.

The data which is received from the satellite is in RAW format which doesn't provide any kind of information. Therefore, to get knowledge, we need to process it using various mathematical models.

To predict the weather from the RAW data we need to make data suitable to the input of the mathematical model. This processed data, then, stored in the data warehouse. Those stored data can directly take it as an input to the mathematical model which eventually give us the desired information.

The process of converting RAW data into the cleaned data is known as Knowledge Discovery Process in other words we've known as Data Mining. Further, to predict the information various methods of Data Mining is used such as;

- Decision Tree
- Rule-based Methods
- Neural Networks
- Naive Bayes
- Bayesian Belief Network
- Support Vector Machine

Among this methods, the most famous method is the Decision Tree method.

What is a Decision Tree?

The decision tree is a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. A graphical representation of the decision tree could be:

Using the following Decision tree algorithms decision tree is created.

- ID3 (Iterative dichotomiser3)
- C4.5 (successor of ID3)
- CART (Classification and Regression Tree)
- CHAID (CHI-squared Automatic Interaction Detector)
- MARS (extends decision trees to better handle numerical data)]

Example using ID3 algorithm:

Let me explain, how the decision tree is generated and how to make a prediction using it. For the calculation, I've used the ID3 algorithm. Using the following dataset we are going to predict the weather of 18th November.

To create a decision tree from any data-set we need to perform a few calculations such as *Entropy* which is a measure of the uncertainty about a source of messages. It gives us the degree of disorganization in our data. The other one is *Information Gain* which measures the expected reduction in entropy. It decides which attribute goes into a decision node. After performing the calculation the decision tree would look like this.

After observing the generated decision tree you can predict the weather of 18th November.

The sliding window algorithm

The major disadvantage of the ID3 algorithm is the small size of the weather data. As you've seen that we've taken only 5 parameters [Max. Temp, Min. Temp, Humidity, Rainfall, Outlook], it may produce the wrong prediction. Solution to this is to take several years of historical data instead of a single week. In addition, if the historical trends don't relate to the current one, there is a higher risk of hitting a failure.

That problem is solved by a sliding window algorithm:

> Weekly weather trends may not align perfectly with historical data. There may be varying conditions that could offset a historical trend that is relatively close to current conditions. The Sliding Window Algorithm accounts for this potential offset by sampling from a two-week data set. By dividing up this sample data, the algorithm can determine the *best fit* trend and predict the following day's weather.
>
> **An advantage** of this algorithm is that it uses data observed from the same time period in a previous year. The sample data can also be adjusted; data from multiple years can be observed in addition to increasing the number of historically observed weeks. Increasing the number of observed weeks should be limited to ensure that the predictions are based around the same time of year.
>
> **A disadvantage** of this algorithm is that the historical data that has been sampled may not reference the current trend whatsoever. This could happen due to impactful weather conditions that had not been observed in previous years. Occurrences like this would result in a skewed prediction due to the lack of similar data.

. . .

For experiment point-of-view, you can perform prediction on your own utilizing data set and mining tools. I would like to suggest you some of the tools and data set repository for the experiment.

- **Tools**

1. Weka [widely used]

2. Rapid Miner

- **Data set repository**

1. UCI Machine Learning Repository — Data Sets

. . .

A modern approach using machine learning

Until now, whatever we've seen is known as traditional methods of weather forecasting. These traditional methods are quite unstable and prone to error in prediction. Moreover, we can't predict for large periods of time using them because they become inaccurate at that stage.

Most commonly there are only two *machine learning techniques* are used for weather forecasting.

- Neural Network
- Bayesian Network

Neural Networks are highly dynamic in nature and therefore it seems to be the most popular machine learning model choice for the weather forecasting. The main reason for this popularity is the ability to capture the non-linear dependencies of past weather trends and future weather conditions, unlike the linear regression and functional regression models.

The remaining machine learning technique is a *Bayesian Network* which ultimately uses machine learning algorithms to find the most optimal Bayesian Network and parameters [temperature, humidity, outlook, etc.]. The computation cost of the Bayesian Network is very expensive because of a large number of different dependencies.

Implemented Machine learning algorithms:

- Linear regression

- Variation of functional regression

Machine Learning Predictions Data Science Algorithms Predictive Analytics

About Help Legal