iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** VI **Month of publication:** June 2020

**DOI:** http://doi.org/10.22214/ijraset.2020.6218

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

# Prediction of Solar Eclipses using Extreme Gradient Boost Algorithm

Anurag S. Pingalkar
*Big Data Analytics Department, Sunbeam Institute of Information Technology, Pune, Maharashtra, India.*

***Abstract: An eclipse is an event that obscures the light of the Moon when it is in Earth's shadow or the light of the Sun when the Moon is between it and the Earth. An eclipse occurs when there is a full moon or a new moon. To make exact eclipse predictions, you would need to calculate the precise motions of the sun and moon, and that requires a computer and proper software. You can predict eclipses by understanding the conditions that make them possible. There are two conditions for an eclipse: The sun must be crossing a node, and the moon must be crossing either the same node (solar eclipse) or the other node (lunar eclipse). That means, of course, that solar eclipses can occur only when the moon is new, and lunar eclipses can occur only when the moon is full.***
***Keywords: Solar, Eclipse, prediction, machine learning, supervised learning.***

## I. INTRODUCTION

Eclipses are not just simply brief spectacles for us to gaze at in awe, but important events used to study the Earth and the moon. Centuries ago, people discovered the shape of the Earth was round by observing the moon during an eclipse, and scientists are still using eclipses to study the moon's surface in detail. A total solar eclipse is an ideal opportunity to study the Sun under conditions impossible at any other time, as the crisp lunar shadow reveals the corona's inner and middle parts in visible light. Ground-based observations along with satellite data offer the most complete view of the Sun's atmosphere ever seen. The minutes of totality also give scientists an excellent opportunity to measure the temperature of Mercury, the closest planet to the Sun. With the Sun hidden, Mercury can be studied in infrared light (which is used to measure the temperature of objects that are not hot enough to radiate light). Eclipses do not happen at every new moon, of course. This is because the moon's orbit is tilted just over 5 degrees relative to Earth's orbit around the sun. For this reason, the moon's shadow usually passes either above or below Earth, so a solar eclipse does not occur. However, as a rule, at least twice each year (and sometimes as many as five times in a year), a new moon will align itself in just such a way to eclipse the sun. That alignment point is called a node. Depending on how closely the new moon approaches a node will determine whether a particular eclipse is central or partial. In addition, of course, the moon's distance from the Earth — and to a lesser degree, Earth's distance from the sun — will ultimately determine whether a central eclipse is total, annular or a hybrid. Moreover, these alignments do not happen haphazardly, for after a specific interval of time, an eclipse will repeat itself or return. This interval is known as the Saros cycle and was known as far back as the days of the early Chaldean astronomers some 28 centuries ago. The word Saros means "repetition" and is equal to 18 years, 11⅓ days (or a day less or more depending on the number of leap years that have intervened). After this interval, the relative positions of the sun and moon relative to a node are nearly the same as before. That third of a day in the interval causes the path of each eclipse of a series to be displaced in longitude a third of the way around Earth to the west with respect to its predecessor. For example, on March 29, 2006, a total eclipse swept across parts of western and northern Africa and then across southern Asia. One Saros later, on April 8, 2024, this eclipse will recur, except instead of Africa and Asia, it will track across northern Mexico, the central and eastern United States and the Maritime provinces of Canada. Apart from being fascinating for their own sake, these Eclipse phenomena are important to understand because they have an impact on the rest of the solar system, including Earth.

## II. LITERATURE SURVEY

You can predict eclipses by understanding the conditions that make them possible. As you begin to think about these conditions, be sure you are aware of your point of view. Later you will change your point of view, but to begin you must imagine that you can look up into the sky from your home on Earth and see the sun moving along the ecliptic and the moon moving along its orbit. The project will be indirectly useful for the scientists to study the objects in our solar system. Since, due to eclipse, there is a shadow of moon the earth's surface, it is easy for scientists to observe the objects in our solar system. Using this project, scientists can predict the eclipse date and eclipse time so that they can be ready for the research they are trying to do in the time of eclipse. The goal was to build a system capable of the following tasks:

*A. Collecting Fundamental and Technical Data from the Internet*

We have to find a dataset that have Date, time, and location of every eclipse in five thousand years. Using this, we can use those previous dates to predict occurrence of eclipse in the future.

*B. Simulating Strategies*

The system should offer ways to specify and simulate fundamental and technical trading strategies. Additionally, combining the two approaches should be possible.

*C. Evaluating Strategies*

The system should evaluate and visualize the prediction performance of the simulated strategies. This allows a comparison to be made between technical, fundamental and the combined approaches.

*D. Purpose*

The aims of this project are as follows:

*1)* To identify factors causing solar eclipse.
*2)* To generate the pattern from large set of data of previously occurred eclipses for prediction of eclipse that will occur in future.
*3)* To predict where on the planet earth the eclipse can be seen.
*4)* To provide analysis for users through Desktop application

*E. Modern Times*

"The method we use now is based on something that people came up with in the 19th century", Ernie Wright, a visualization expert at NASA says. The people who started using more modern calculations to predict the eclipse paths were Friedrich Bessel and William Chauvenet. Today, we are able to get even more specific, thanks to our understanding of the shape of the moon. The moon—contrary to every elementary school drawing you ever laboured over—is not in fact, shaped like a banana or a perfect sphere. Like the Earth, it has mountains and valleys that make its shape a little rough around the edges, and that means that its shadow is uneven as well.

## III. CONCEPTION STUDY AND REQUIREMENTS

*A. Dataset Context*

Eclipses of the sun can only occur when the moon is near one of its two orbital nodes during the new moon phase. It is then possible for the Moon's penumbral, umbral, or antumbral shadows to sweep across Earth's surface thereby producing an eclipse. There are four types of solar eclipses: a partial eclipse, during which the moon's penumbral shadow traverses Earth and umbral and antumbral shadows completely miss Earth; an annular eclipse, during which the moon's antumbral shadow traverses Earth but does not completely cover the sun; a total eclipse, during which the moon's umbral shadow traverses Earth and completely covers the sun; and a hybrid eclipse, during which the moon's umbral and antumbral shadows traverse Earth and annular and total eclipses are visible in different locations. Earth will experience 11898 solar eclipses during the five-millennium period -1999 to +3000 (2000 BCE to 3000 CE). Eclipses of the moon can occur when the moon is near one of its two orbital nodes during the full moon phase. It is then possible for the moon to pass through Earth's penumbral or umbral shadows thereby producing an eclipse. There are three types of lunar eclipses: a penumbral eclipse, during which the moon traverses Earth's penumbral shadow but misses its umbral shadow; a partial eclipse, during which the moon traverses Earth's penumbral and umbral shadows; and a total eclipse, during which the moon traverses Earth's penumbral and umbral shadows and passes completely into Earth's umbra. Earth will experience 12064 lunar eclipses during the five-millennium period -1999 to +3000 (2000 BCE to 3000 CE).

*B. Proposed Algorithm*

XGBoost (Extreme Gradient Boosting) is a powerful, and lightning fast machine learning library. XGBoost, short for "Extreme Gradient Boosting", was introduced by Chen in 2014. Since its introduction, XGBoost has become one of the most popular machine learning algorithm. If you are already familiar with Random Forests, the Gradient Boosting algorithm implemented in XGBoost is also an ensemble of decision trees. Those trees are poor models individually, but when they are grouped they can be really performant. However, it's an intimidating algorithm to approach, especially because of the number of parameters — and it's not clear what all of them do. Although many posts already exist explaining what XGBoost does, many confuse gradient boosting, gradient boosted trees and XGBoost. The purpose of this post is to clarify these concepts. Also, to make XGBoost's hyperparameters less intimidating, this post explores (in a little more detail than the documentation) exactly what the hyperparameters exposed in the scikit-learn API do. XGBoost is one of the implementations of Gradient Boosting concept, but what makes XGBoost unique is that it uses "a more regularized model formalization to control over-fitting, which gives it better performance," according to the author of the algorithm, Tianqi Chen. Therefore, it helps to reduce overfitting.

XGBoost is one of the fastest implementations of gradient boosted trees. It does this by tackling one of the major inefficiencies of gradient boosted trees: considering the potential loss for all possible splits to create a new branch (especially if you consider the case where there are thousands of features, and therefore thousands of possible splits). XGBoost tackles this inefficiency by looking at the distribution of features across all data points in a leaf and using this information to reduce the search space of possible feature splits. Although XGBoost implements a few regularization tricks, this speed up is by far the most useful feature of the library, allowing many hyperparameter settings to be investigated quickly. This is helpful because there are many, many hyperparameters to tune. Nearly all of them are designed to limit overfitting (no matter how simple your base models are, if you stick thousands of them together they will overfit). The list of hyperparameters was super intimidating to me when I started working with XGBoost, so I am going to discuss the 4 parameters I have found most important when training my models so far. Before moving forward, it is important for us to be cleared with the concepts and mechanism of Solar Eclipses.

### *C. What is a Solar Eclipse?*

Sometimes when the moon orbits Earth, it moves between the sun and Earth. When this happens, the moon blocks the light of the sun from reaching Earth. This causes an eclipse of the sun, or solar eclipse. During a solar eclipse, the moon casts a shadow onto Earth.

There are three types of solar eclipses.

1) *Total Solar Eclipse:* A total solar eclipse is only visible from a small area on Earth. The people who see the total eclipse are in the centre of the moon's shadow when it hits Earth. The sky becomes very dark, as if it were night. For a total eclipse to take place, the sun, moon and Earth must be in a direct line.
2) *Partial Solar Eclipse:* This happens when the sun, moon and Earth are not exactly lined up. The sun appears to have a dark shadow on only a small part of its surface.
3) *Annular Solar Eclipse:* An annular eclipse happens when the moon is farthest from Earth. Because the moon is farther away from Earth, it seems smaller. It does not block the entire view of the sun. The moon in front of the sun looks like a dark disk on top of a larger sun-coloured disk. This creates what looks like a ring around the moon.

During a solar eclipse, the moon casts two shadows on Earth. The first shadow is called the umbra. This shadow gets smaller as it reaches Earth. It is the dark centre of the moon's shadow. The second shadow is called the penumbra. The penumbra gets larger as it reaches Earth. People standing in the penumbra will see a partial eclipse. People standing in the umbra will see a total eclipse. Solar eclipses happen once every 18 months. Unlike lunar eclipses, solar eclipses only last for a few minutes.

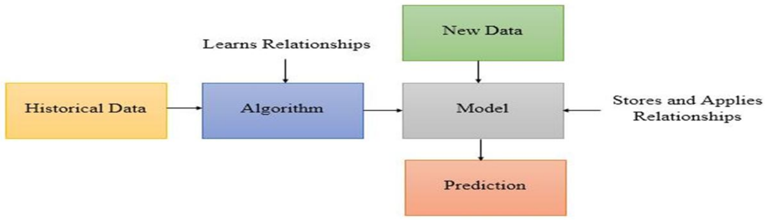
### *D. Requirement Analysis:*

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements. These requirements are listed below:

1) *Functional Requirements:* Functional requirement are the functions or features that must be included in any system to satisfy the needs and be acceptable to the users. The system should be able to generate a prediction with good accuracy. The system should collect accurate data from the provided dataset in consistent manner. Prior to application of eclipse's occurrence, the database is updated by the latest values. The user can look previous data Information that was collected. The user can also study the concepts based on the variables used to predict the eclipse such as the date, time, location (Longitudes and latitudes) and so on.
2) *Non-Functional Requirements:* Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non- functional requirements are essentially based on the performance, information, concepts and security efficiency and services. The system should provide better accuracy. It should have simple interface for users to use and it should perform efficiently in short amount of time. The reliability of the product will be dependent on the accuracy of the dataset. In addition, the past data of the eclipse's occurrence used in the training would determine the reliability of the software. The user will only be able to access the website using his

login details and will not be able to access the computations happening at the back end. The maintenance of the product would require training of the software by recent data so that there commendations are up to date. The database has to be updated with recent values. The website is completely portable and the recommendations completely trustworthy as the data is dynamically updated. The interoperability of the website is very high because it synchronize the entire database with the wamp server.

## IV. PROPOSED ARCHITECTURE

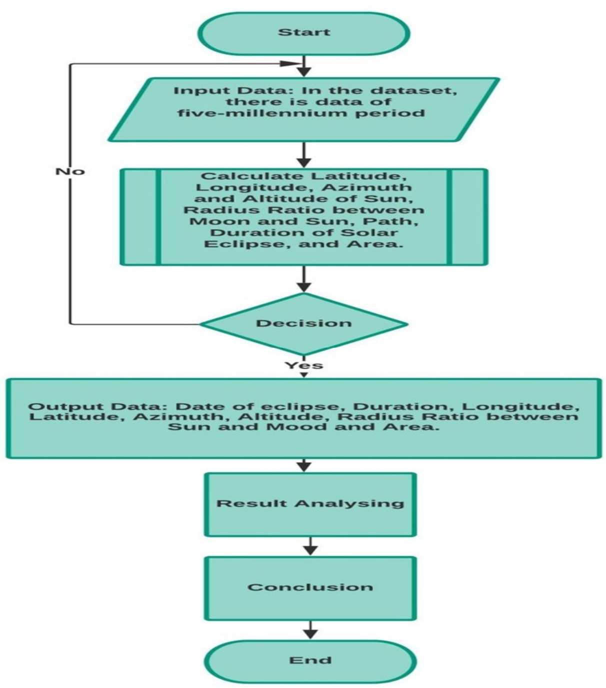



In the dataset, as explained in the Dataset Context, there is a data of five-millennium period. Prediction result will be handled though the algorithm written by us and generated by System. The system will be built, through which the result of prediction and system performance will be analysed. The algorithm learns the relationship between dependent and independent variables in the dataset. The dependent and Independent variables are decided by the programmer from the dataset in order to achieve the required prediction. During the phase of learning relationship, the new data is created for the better prediction. We can say it in term of the Dimensionality Reduction paradox for the machine learning algorithms. Thus, using above use cases, the model of the machine learning algorithm can be created. This model further does the prediction and the result can be studied by the user. In this step, the data is fetched and stored in RAM and the relationship is applied between the dependent and independent variables from the dataset. This use case is must important in whole project. The key feature of this project is to predict the date of the eclipse. Thus, this will be available in user interface and viewer can observe them.

### *A. Julian to Gregorian date reconciliation:*

For some considerations it might be useful to replace the dates with a linear time format. In Astronomy the Julian Day (JD) does exactly this - it counts the number of days elapsed since Jan 1, 4713 BC (conventional). This conversion also circumvents the issue of the change of calendars in 1582, namely the switch from Julian to Gregorian calendar. This in particular accounts for 10 missing days in the calendars: The dates from October 5 to 14, 1582 do not exist in any calendar.

The function takes a given date (either Gregorian or pre-Gregorian) and converts it into the corresponding Julian date.



## V. FINDINGS AND CONCLUSIONS

The system evaluation on the solar eclipse prediction is carried out. For given prediction the various factor causing solar eclipse taking into consideration and the algorithm predicts the forecasting using all the independent variables from the dataset. To make exact eclipse predictions, you would need to calculate the precise motions of the sun and moon, and that requires a computer and proper software. Such software is available for desktop computers, but it is not necessary if you are satisfied with making less exact predictions. In fact, many primitive peoples, such as the builders of Stonehenge and the ancient Maya, are believed to have made eclipse predictions. In the very beginning, collecting the information and learning the about the solar eclipse was very essential task. Gathering all the necessary information about the mechanism and science behind the happening of solar eclipse was a challenge. Getting the help of popular search engine https://www.google.co.in, I found all the case studies and facts about solar eclipse as well as the mechanism about the same. My predictive model is evaluated on solar eclipse on the history of solar eclipses occurrence data stored for five-millennium period from -1999 to 3000. The result of the algorithm to be founded on the console is the eclipse date had been calculated in the dataset and the eclipse date, which is calculated by the algorithm. As discussed above, the accuracy of the algorithm is about 96%. We can observe that the accuracy decreases as the date moves in the future from the date entered.

## REFERENCES

[1] Freeth, Tony. (2014). Eclipse Prediction on the Ancient Greek Astronomical Calculating Machine Known as the Antikythera Mechanism. PloS one. 9. e103275. 10.1371/journal.pone.0103275.

[2] Miguel Querejeta, 'On the Eclipse of Thales, Cycles and Probabilities', Culture And Cosmos, Vol. 15, no. 1, Spring/Summer 2011, pp. 5–16, www.cultureandcosmos.org.

[3] Kochhar R, 2010, Rahu and Ketu in mythological and 'astronomological' context, Indian Journal of History of Science, vol 45, no 2, June 2010, p 287-297

[4] Subbarayappa, B V, 2008, Traditions of Astronomy in India and Jyotishshastra, Centre for Studies of Civilizations, Vol IV part 4, Centre for Studies of Civilisations, Viva Books.

[5] Yabuuti, Kiyoshi. "Astronomical Tables in China, from the Han to the T'ang Dynasties', in ChÂgoku chÂsei Kagaku gijutsushi no kenkyÂ (Studies in the history of medieval Chinese science and technology), 1963, pp. 445-492.

[6] Yabuuti, Kiyoshi. "Researches on the Chiu-chih Li -Indian Astronomy under the T'ang Dynasty", Acta Asiatica, vol. 36, 1979, pp. 7-48.`