ISSN No:-2456-2165

Rainfall Prediction Using Machine Learning

Arnav Garg
Department of Computer Science Engineering, SRM IST, Kanchipuram Tamil Nadu

Himanshu Pandey
Department of Computer Science Engineering, SRM IST, Kanchipuram Tamil Nadu

Abstract:- As global warming increases detection and prediction of rainfall is becoming a major problem in countries which do not have access to proper technology and which if done accurately can help them for several purposes such as farming, health, drinking and many other. And for this purpose we predict the rainfall of coming year using SVR, SVM and KNN machine learning algorithm and compare the results inferred by each algorithm.

I. INTRODUCTION

As global warming has increased so has earth's temperature and due to which our local region's yearly rainfall patterns have also been affected and this harms the population living in the areas, as farmers and other people who heavily rely on rainfall for all their water based needs ,so in these regions accurate predictions of rainfall is of utmost importance while there are many ways of predicting them one chosen for this study is by observing and collecting the previous year rainfall data (in mm) and then predicting the rainfall(in mm) for the coming year , though it is not a full proof method and may become inefficient due to any given factor such as a "sudden increase in co2 levels " ,it is cheap and may help in counties where seasons are consistent and which are way behind the world in terms of technological advancements such as rural areas of counties.

II. MACHINE LEARNING ALGORITHMS USED

➢ *SVM*

Support-vector machines (SVMs) are supervised learning models with related learning algorithms that break down information utilized for classification and regression examination. Given a lot of preparing precedents, each set apart as having a place with either of two classes, a SVM preparing calculation fabricates a model that appoints new guides to one class or the other, making it a non-probabilistic binary linear classifier (although techniques such as Platt scaling exist to utilize SVM in a probabilistic order setting). A SVM model is a portrayal of the precedents as focuses in space, mapped with the goal that the instances of the different classes are isolated by a reasonable hole that is as wide as would be prudent. New models are then mapped into that equivalent space and anticipated to have a place with a classification dependent on which side of the hole they fall. SVMs can productively play out a non-direct characterization utilizing what is called the kernel trap, mapping their contributions to high-dimensional component spaces.

At the point when information is unlabeled, directed learning is beyond the realm of imagination, and an unsupervised learning approach is required, which endeavors to discover natural clustering of the data to gatherings, and afterward map the new information to these shaped gatherings. The support-vector clustering algorithm applies the measurements with the help of vectors, created by the vector machine's calculation, to sort unlabeled information, and is a standout among the most broadly utilized grouping calculations in modern applications

➢ *SVR*

A variant of SVM for regression was proposed in 1996 by Vladimir N. Vapnik, Harris Drucker, Christopher J. C. Burges, Linda Kaufman and Alexander J. Smola.This strategy is called Support-Vector Regression (SVR). The model created by support vector arrangement depends just on a subset of the training data, on the grounds that the cost capacity for building the model does not depend on training data that lie on the outer side of the margin. Comparably, the model created by SVR depends just on a subset of the preparation information, in light of the fact that the cost function for construction of the model disregards any training data near the model's prediction.

➢ *KNN*

In pattern recognition, the k-nearest neighbors classification (k-NN) is a non-parametric strategy utilized for classification and regression. In the two cases, the information comprises of the k nearest data in the component space. k-NN is a kind of instance-based learning where the function is just approximated locally and all calculation is conceded until grouping. The k-NN calculation is among the most straightforward of all AI calculations. Both for classification and regression, a valuable method can be utilized to assign a weight to the commitments of the neighbors, so that the closer neighbors contribute more to the normal than the ones which are far apart. For instance, a typical weighting plan comprises in giving each neighbor a weight of 1/d, where d is the separation to the neighbor.

The neighbors are taken from the data set for which the classes (for k-NN classification) or the object property estimation (for k-NN regression) is known. This can be thought of as the training data set for the calculation.

ISSN No:-2456-2165

III. DATA SET

The data refers to district wise rainfall (in mm) in India from period 1951-2015. [2]

The data set is released under: National Data Sharing and Accessibility Policy (NDSAP)[2] and is contributed by Dr. K. Somasundar (Chief Data Officer) Ministry of Earth Science ,India[1]

IV. MODELS AND THEIR RESULTS

➢ *Model formalization*

The initial steps taken for the formation is data cleaning using python as the major language in the software "Jupyter notebook"[8].

Then the data is divided into training and testing data i.e data from 1951 to 2014 acts as training data and data of the year 2015 acts as testing data. Then we train each machine learning model and obtain the results in form of a graph.

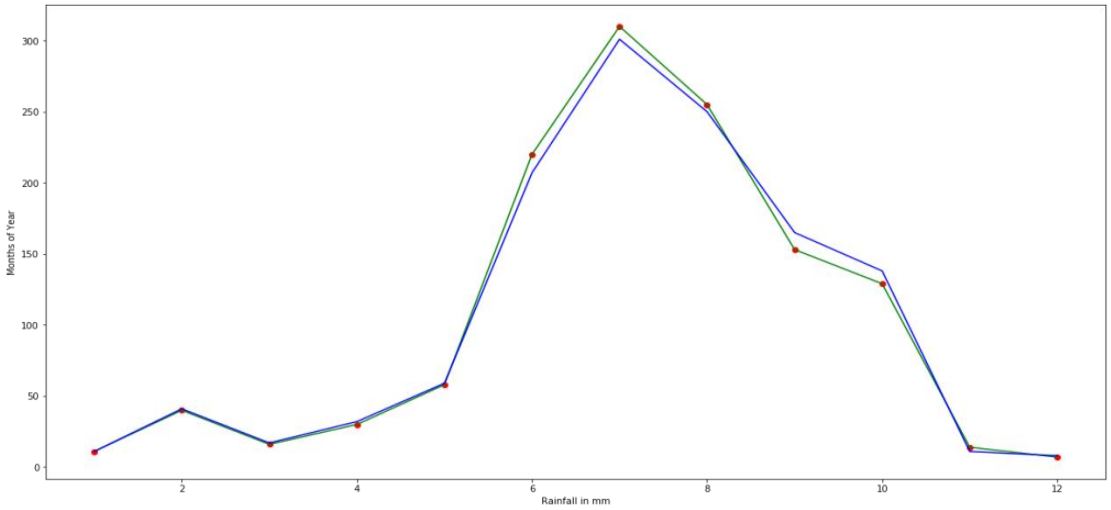
➢ *Results*

As it can be seen from the graphs below SVR is most efficient and SVM is more efficient than KNN. This is because to the data set being used is a time series based and SVR machine learning algorithm is the most accurate for the required task but for practical purpose one can argue that SVM is the best because even though SVR is the most accurate but in real life we get errors in recorded data and even though we may not know about how much but there is always a bias in natural rainfall patter and hence a near perfect model may prove to be inefficient so the best thing would be to add bias to SVM model and create a lower and upper limit.

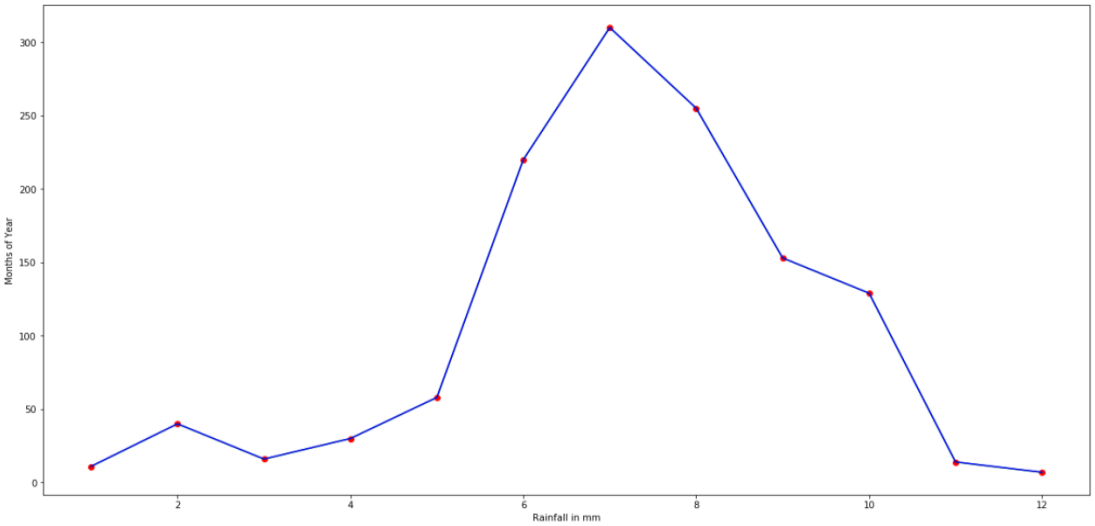
➢ *Graphs*

The graphs are scaled between rainfall (in mm) and month of the year 2015.

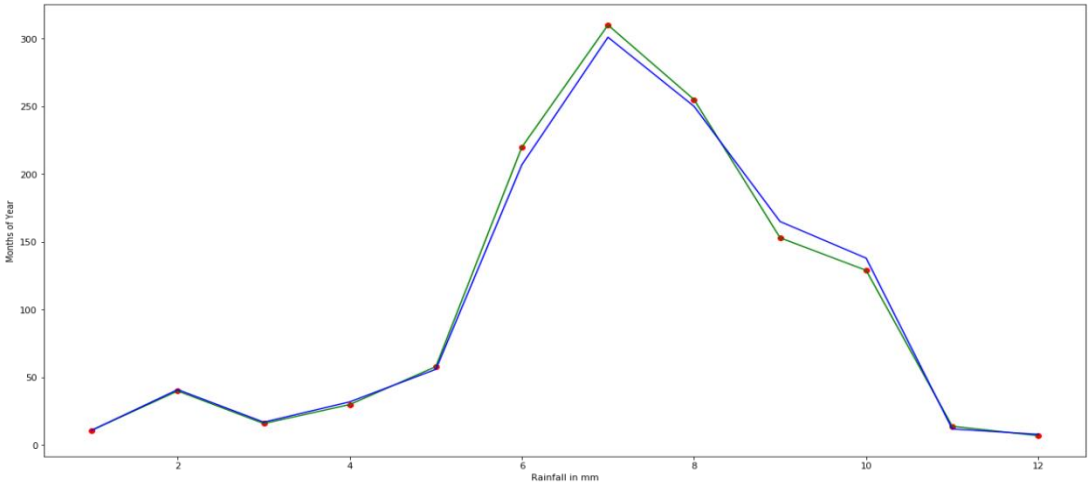
➢ Blue Line - Predicted Values
➢ Green Line/ Red Dots -Actual values



Graph 1:- (Font-10, Bold): Graph plotted using SVM Machine learning algorithm



Graph 2:- (Font-10, Bold): Graph plotted using SVR Machine learning algorithm



Graph 3:- (Font-10, Bold): Graph plotted using KNN Machine learning algorithm

V. CONCLUSION

Though all the algorithms i.e. SVM, SVR and KNN are equally useful, for rainfall prediction SVM and SVR are more accurate than KNN and even though SVR provides near perfect if we consider all the anomalies and biases found of nature we can easily conclude that for rainfall prediction SVM is the best among the three and the best way to use it is by forming a range of highest and lowest predicted values by adding bias in the model.

REFERENCES

[1]. Ministry of Earth Science, India
[2]. https://data.gov.in/catalog/all-india-area-weighted-monthly-seasonal-and-annual-rainfall-mm
[3]. https://www.kdnuggets.com/2016/05/machine-learning-key-terms-explained.html
[4]. https://en.wikipedia.org/wiki/Support-vector_machine
[5]. https://en.wikipedia.org/wiki/Logistic_regression
[6]. https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html
[7]. https://demodata.nic.in/sites/default/files/NDSAP.pdf
[8]. https://jupyter.org/
[9]. https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/
[10]. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
[11]. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
[12]. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
[13]. https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics
[14]. https://scikit-learn.org/stable/modules/classes.html#module-sklearn.svm