Machine Learning Based Algorithms for Crop Prediction System

Sahil Singh
Dept. of CSE & IT, Amity University Madhya Pradesh, Gwalior, India
e-mail: sahilsingh20121998@gmail.com

Abstract – In India most of the people's livelihood is based on agriculture and it is their main occupation. ***Generally, the farmers plant the various crops based on the choice of people and they also use more fertilizers to plant them. Due to this, there is an enormous improvement in the usage of machine learning in the field of research and the different industries over the recent few years. This is the reason that we want to use machine learning extensively to build an advanced system in the field of agriculture and this is all done to improve the adaptability of the farmers in doing the agricultural practices. Over the years the research papers have just used only one attribute to showcase the idea about the usage of ML in the various systems. Further we want to increase the overall efficiency of the system and the results by adding more attributes to our ML based system, this indirectly can help us to figure out the patterns and yield more effective results based on prediction analysis. After which the system can be used to predict which crop can be grown in the different regions.***

***Key Words*: Machine Learning in Agriculture, Classification Algorithms, Decision Tree, KNN**

1. INTRODUCTION

The complete production of crops is depended upon the conditions of the soil, environmental factors and the input parameters influenced by the farmers. The production of the crops in the field of agriculture varies from the various fields used and the farmers associated with it. It would be very difficult to collect the information about this on a very large scale as it would be a very complicated task. By the way the Indian Meteoric Department collects the information on the various environmental conditions and has presented it in a tabulated format about the different components in various districts per 1sq.m space. The information provided by the meteoric department can be used to predict the growth of the crops in the different districts and the places. Numerous different technologies are employed by the various researches across the globe in the field of agriculture related to the evaluation and the association of the science. The various researches related to it are present in the alternative countries where the usage of chemicals is hugely influenced for the production of large amounts of crops in consideration with the state policies by the agricultural researchers. Hence a relation between the yield of crops and the usage of chemical [1] is defined. Over the last few years the agricultural sector has benefitted the most by the usage of the various technologies and the techniques of the machine learning(ML), knowledge related to the science and detection technology. In this era of ever increasing population and temperature on the planet the recent developments in the field of agriculture has helped us to sustain with the demand of food to be produced according to the society at the international stage. The knowledge of sensing the field remotely helps per cubic centimeter to better yield according to the area. In order to support the human existence, the agricultural field is trying to innovate at a large scale for the crop yield and quality of crops produced.

There is an increase in the short form of jobs in the agricultural fields over a period of time. Only a handful of people are involved in the cultivation process. Due to this right crop at the right area, time, according to the need of the weather conditions, humidity, temperature, becomes an important step to feed the large population effectively in this fast paced world and ever changing weather patterns before the manufacture of the crops takes place. To increase the production capacity in the agricultural field the adaptation of the new technologies should take place according to the existing climatic conditions and the soil structure, air patterns in order to fight against the food insecurity. In order to solve the agricultural issues, the usage of the computer through coding in multiple languages over the platform to provide solutions and bridging the traditional practices with the present system through rainfall prediction, heat and water distributions forms can be taken into account at different areas and in different phases of the crop yield process. It will also help the policy makers and the distributors as well as the farmers to review and model the existing production processes effectively through computer code and also to catch upon the temperature factors and the changing diversity patterns of the climate to further increase the efficiency of crop production through proper planning and implementation so that we can avoid the devastating effects of the crop failure. The monitoring on the basis of the experimentation of the weather and water conditions is taken into account to better yield the crop across various regions and also to predict the changes in the field. Proper management of the changes in the soil and the weather over the field and also from pests should take place time to time.

The management of the information related to various aspects with one another should take place. An overall assessment on the impact of the temperature in the field of agriculture will take place. By following the various developing models according to the theory and proper management of the crop, breeding programs that are novel in nature, new crop rotation patterns, the agriculture should adapt themselves. The recording of the climatic changes in the various seasons can be done in a timely manner so that we can maximize the growth of crops. After this by the usage of the computer code the assessment in the water distribution and the conditions of the climate based on machine learning will be done and also various alternative situations will be monitored. The information is obtained from the various trends, patterns, experimental analysis across the different fields by the help of data processing and the analysis of the information available over a period of time so that we can provide it to the users. Hence the users will be able to summarize the knowledge that is collected and can create the relations among them accordingly. Then they will be able to predict the right thing to expect. The patterns are found among the data sets that is available in the form of massive databases and the characteristic relations of the techniques are evaluated on the basis of Machine Learning techniques. The observed patterns are shared to the users that are the trends and the relationships so that the information can be improved and associations be arranged. According to this the prediction of the crop losses can take place based on the machine learning so that farmers can improve the overall condition of crop production.

2. Review of Literature

Automated farming is being used everywhere for agricultural practices. In the area of the generalized prediction model the Decision Tree algorithm is applied for the prediction purpose that is a supervised learning algorithm. To increase the productivity of the Soyabean according to the climatic conditions a research is being made to see the effect of the technique that is based on the decision tree. Several rules were framed accordingly for the users that is based on the decision tree. The attributes such as the depth of soil, classification capability of the land, drainage, permeability, erosion, and the texture are selected based upon the paper from Md. Tahmid Shakoor & co [4]. In our study we have implemented the two supervised machine learning algorithms. According to the right source the properties related to the soil and weather are considered and their data is being taken automatically by our system. The data of the soil and the information from it is obtained at high resolutions that are aligned with the most effective input data set and their system has an advantage that it worked on the large regions. It has provided us with the authorized ability of forecasting the pattern of crop before the

beginning of the season. Hence it helps us to take more generic action towards the plantation of the crop and also to change the strategies of implementation in this changing types of crop, so that we can predict the crop cycle pattern in advance and accommodate variations in climates in advance [2].

The algorithm created implements a data-driven model to predict and estimate crop yield using mutual dependencies of soil and climate features. Although there are several techniques existing to obtain rainfall forecasts, the algorithm described in this paper succeeded in focusing on Rainfall along with the crop yield prediction. This developed model took into account both the most important environment and soil parameters influencing crop growth, so that in the final prediction each of those parameters was given equal weight. Even before the sowing season starts, the results of this research will help farmers by knowing the investment capital on the crop to be sown. The algorithm's predictive pattern will support local self-government and financial institutions by allocating appropriate funds or tax credits to farmers. Naive Bayes can also be useful for the broad dataset. Use of naïve Bayes and decision tree makes the model very computationally efficient. The machine is versatile, as it can be used to test various crops. The best time of sowing, plant production, and plant harvesting can be calculated from the yield graphs. However, it may also incur the best and worst environmental situation. The model focuses on all types of farms and it can also help smaller farmers. This model can be further enhanced to determine the yield of each crop, and recommended for pesticides. In addition, it may be modified to suggest the need for fertilizers and crop irrigation.

3. Work Done

1. System Based on Scalable Machine Learning In Agricultural Yield For Pre- Seasonal Forecast:

The structure and system predicted during this work is generated by a neural network wherever the area unit of the inputs is handled individually. Static soil information in handled by completely connected layers while continuous LSTM layers handle dynamic meteorological information. This specific concept was trained against historical yield labels at county level, with historical information for many soil properties, precipitation, minimum and most temperature. The model was tested during training in an extremely separate collection of information and showed comparable results with established methods of yield prognostication that make use of in-depth remote sensing data. The most important lesson learned from our experiments is that as a result of the projected neural network model, it is possible to obtain an ascendable yield forecast and exploit redundant information within the soil and in the environment. To boot, the model could learn AN implicit illustration of the crop cycles evaluated during this paper, taking into account the atmospherically seasonal information used as input.

2. Machine learning technique to crop yield forecasting based on climatic parameters

The current study gives the potential use of information mining techniques to predict the crop yield that aided input parameters for the environmental condition. The developed webpage is user-friendly and hence the accuracy of square predictions measure higher than 75 per cent all told the crops and districts designated within the study indicating higher predictive accuracy. The user-friendly web content developed to predict crop yield can be used by any user for their crop alternative by providing knowledge of that place about environmental condition.

3. Crop Prediction on Indian region belts: A Naïve Bayes MapReduce Agricultural Precision Model

The planned work presents efficient degree recommendation system for economic crops. The use of naïve mathematicians makes the model in terms of computation poorly economical. The device is versatile since looking at totally different crops may be normal. The simplest time of sowing, plant growth and plant collection can be understood from the yield graphs. Also the best and worst condition may be borne jointly. The model sheds light on all farm types and it may also help smaller farmers.
This model can be increased to figure out each crop's yield, and for chemical recommendation. Together, it can be changed to recommend fertilizers and crop irrigation requirements.

4.Evaluation of Predictive Data Mining Algorithms for Automated and Optimized Crop Recommendation in Soil Data Classification.

In this review, we gave potential analyzes for soil classification through mistreatment well-known classification algorithms such as J48, BF Tree, and OneR and Naïve Bayes; in data processing. The analysis

was carried out on instances of knowledge from the district of Kasur, Pakistan. We also ascertained that the comparative analysis of these algorithms has the specific level of accuracy to assess predictions 'efficacy and potency. The advantages of a higher understanding of soil groups, however, would enhance farm productivity, reduce reliance on fertilizers, and build higher prognostic rules for advice on yield increase. We have a propensity in the future to establish a soil conservation management.

5. Prediction of agricultural production using supervised machine learning techniques

During this analysis two supervised classification machine learning algorithm was applied. The preference Tree Learning-ID3 (Iterative Dichotomiser 3) and KNNR discover patterns within the information set containing the average temperature and precipitation value of six major crops over the last twelve years in 10 major cities of Bangladesh, and provide the prediction. ID3 uses the tree table of choice which consists of precipitation ranges, temperature and knowledge of yields. The research provides an answer to the current downside in the People's Republic of Bangladesh, which was much needed for farmers. Although research is limited to some mounted dataset, the long-term prospect promises to add a lot of knowledge that will be analyzed with more machine learning techniques to produce higher accuracy crop predictions. Moreover, the research will result in income and innovation of advanced farming techniques which will improve our economy and can encourage United States stand out as a technologically advanced nation.

4. EXISTING SYSTEM

For its economic development an agro-based country is dependent on agriculture. As a country's population increases reliance on agriculture, it also rises, and affects the country's subsequent economic development. In this situation the rate of crop yield plays an important role in the country's economic development. Therefore, crop yield levels need to increase. Many biological approaches (e.g. crop seed quality, crop hybridization, strong pesticides) and other chemical solutions (e.g. fertilizer usage, urea, potash) are applied to solve this issue. Among these methods, a crop sequencing technique is required in order to improve the crop's net yield rate over the season. One of the current methods that we have defined is the crop selection method (CSM) for achieving a net crop yield rate over the season. We have taken CSM as an example to show how it helps farmers to achieve more yield.

Crop is classifiable as:

a) Seasonal crops— Seasonal crops may be planted during a season. E.g. cotton wheat.

b)Crops throughout the year or Whole year crops — crops can be planted throughout the year. For instance, vegetables, paddy, toor.

c)Short time span plantation crops— Growing crops which take a short time. For example, potato, vegetables and the ratio. d) Long-term planting crops — These crops can take a long time to grow. Onion, e.g. sugar cane. You can select a combination of these crops in a series dependent on yield rate per day. Illustrates crop sequences with seasonally average yield performance. The CSM method shown in can boost the crop net yield rate using the limited land resource and also increase the land reusability.

In crop selection process, it essentially uses technique in which it recommends different crop set over the years for the same field. For farmers there are different choices available to choose from. They can pick one of the choices, and look at the results. The combination which will give the same area a high yield is produced as production for that field. In this way the CSM method attempts to predict the correct crops for a given region. In India, farming systems are used strategically, depending on the locations where they are most suitable. The agricultural systems that affect India's agriculture substantially are subsistence farming, organic farming, and commercial farming. Regions in India vary in the forms of agriculture they use; some are focused on horticulture, ley farming , agroforestry, among many others. The research papers surveyed provided a rough idea with only one attribute about using ML. We plan to add more attributes to our method and boost the performance, which will increase the yields and we can identify some prediction trends. This method should help to explain what crop can be grown in a given area.

1. PROPOSED SYSTEM

In our system we are making use of a classification algorithms to improvise the crop yields. The proposed system has been shown in Figure 1.

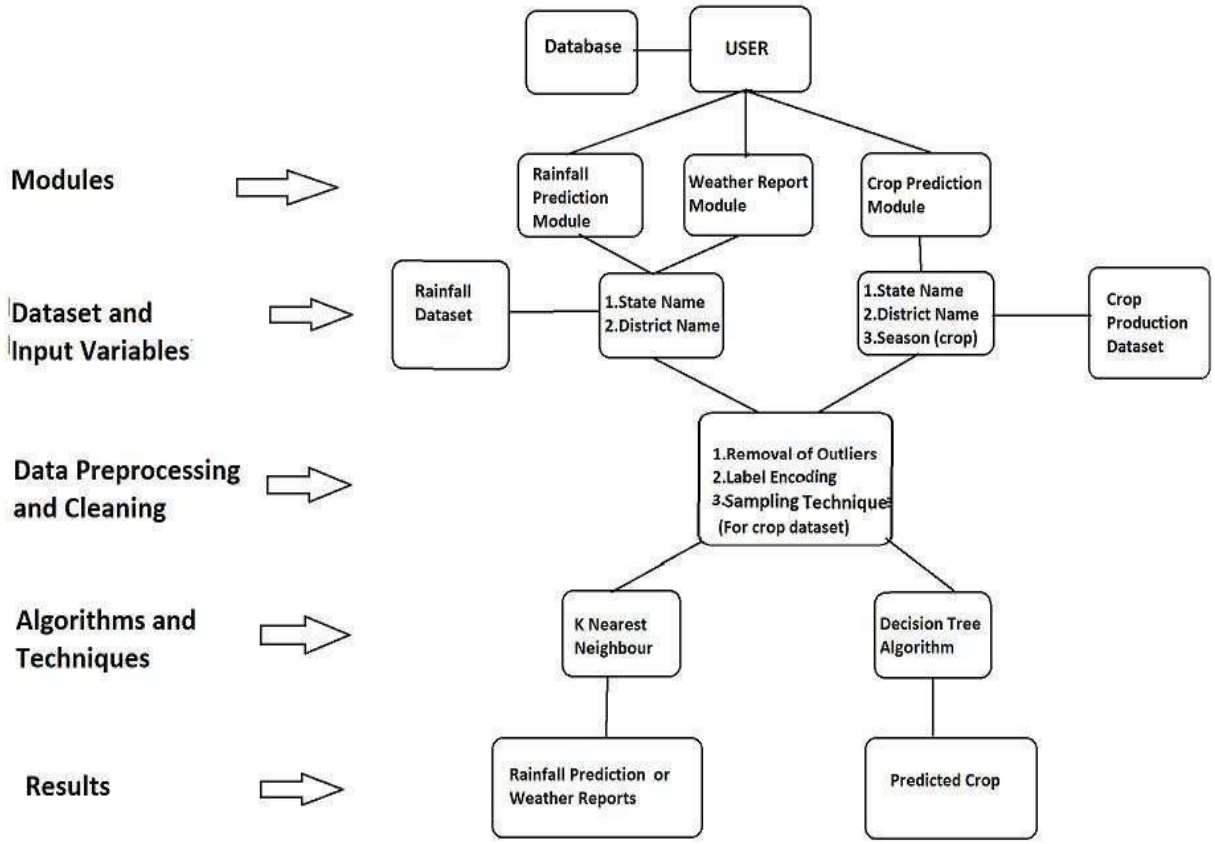

Figure 1: Proposed System

5.1. Data Acquisition:

Dataset must have following attributes

- Soil parameters: Soil Type, Soil ph Value
- Climatic Parameters: Humidity Temperature Wind, Rainfall
- Production Cost of cultivation, Previous year yield details for that region

Planning of the crop production is being done for the district level in this project. We here have our main aim as to find the patterns of the rainfall, climatic conditions, moisture, contents of the soil, temperature, of the production details from the datasets over the past 10 to 12 years and keeping in mind all the parameters. The usage of the classifiers is being done from the dataset that will act as a major factor in the prediction of the crops in the long run. In this everything is assessed and then the prediction of the crops is being done accordingly.

5.2 Preprocessing:

Due to the presence of the noisy data, attributes that are redundant in nature the dataset needs to pre-processed before using it. During the prediction process the redundant data is being removed by the process of the data cleaning methods as they are not fit for the prediction of the datasets. The unrelated data and the Over18 which have the same values for all the employees are not included in the prediction task. The values are assigned as 0 and 1 according to the factor which is present or not as it is split on the basis of data analysis which have categorical factors that are

exploratory in nature. The further classification is done on the basis of the assistance of the values assigned to the factor.

5.3 Classifier Models:

5.3.1 Decision Tree Classifier:

We split the tree according to the attributes until the classes that are same in nature does not appear and the best root nodes are being selected from the method of decision tree. In the area of classification problems, algorithms based on supervised learning is generally used for the continuous and categorical actualization of the variables The main aim of this algorithm is to make the groups as distinct as possible by the help of algorithm in which we split the homogeneous sets in more than two sets according to the most significant attributes. Amongst the population the most suitable crop is being selected from the basis of the decision tree as it will give us the best spilt of sets. Prediction of the crops is more accurate according to the decision tree classifier as it has the feature of selection that has an effective methodology.

5.3.1.1 Gini Index

It says, the random selection of the two items should consist of the same class and if the population is sure then the probability of it is 1 which is based on the random selection. In this impurity is calculated from the given class features.

5.3.1.2 Entropy

Top-down approach from the root node is taken up by the decision tree and also the partitioning into subsets from the data that contain homogeneous instances which are similar in nature. If the sample content is equally divided, then the value of entropy is one and if it is completely homogeneous then the value of entropy is zero

5.3.1.3 Information Gain

After the dataset is split according to the attribute, prior to this the decrease in the entropy results in the gain of information. The foundation of the attribute that provides us with the highest gain in information is based on the decision tree that is being constructed (i.e., branches that are homogeneous in nature). For the prediction of the crop vital role is played by the attribute selection methods.

5.3.1.4 Algorithm

The ripping criteria is used by the C4.5 algorithmic program to gain info. It will handle everything similarly such as the missing values, categorical and numerical information. In order to handle the continuous values at first the threshold is generated and then the division of attributes with the prices takes place, according to the edge value and the prices. It also offers the edges that are subsequent in nature. We have to give a class to the instance for the recorder according to the input for the different classifiers which are explicable in nature. The tree graphs are also known as the call tress in which the nodes and the branches work upon the learnt classification and the categorizations are ultimate in nature are denoted by the leaves.

5.3.3 KNN

KNN is based on the instance based approach for learning, it is used when the performance approximation takes place regionally and also the delayed computation of the processes takes place till it's end. The weights are assigned to the various contributions of the neighbors, as a technique that helps both the regression and the classification, so that the weights nearer contribute more in comparison with the distant ones.

6. RESULTS AND ANALYSIS

The dataset containing various attributes like the name of crop, cultivation cost, production cost, irrigation cost that are independent in nature and the yield per hectare that is a dependent variable is tested with the KNN classifier, naïve bayes classifier, and decision tree.

Confusion matrix is used to represent the result that is obtained from the various samples of training datasets that is tested to the level which also gives us the relationship among the actual values and the prediction of algorithms

.

1. Decision Tree Classifier:

Confusion matrix for decision tree classifier:
[[221, 49],
[67, 163]]
Accuracy for Decision-Tree: 76.8%
Precision for Decision-Tree:0.767
Specificity for Decision-Tree:0.708

2. KNN Classifier: Confusion matrix for KNN: [[269, 30], [23,178]]

Accuracy for KNN: 89.4% Precision for KNN: 0.921 Specificity for KNN:0.8825 Accuracy Comparison:

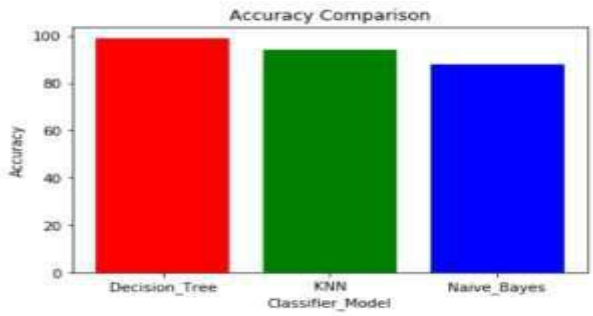

Figure 2: Accuracy Comparision

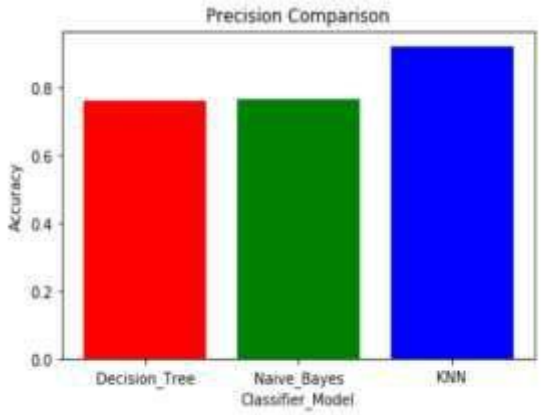

Figure 2: Accuracy Comparision

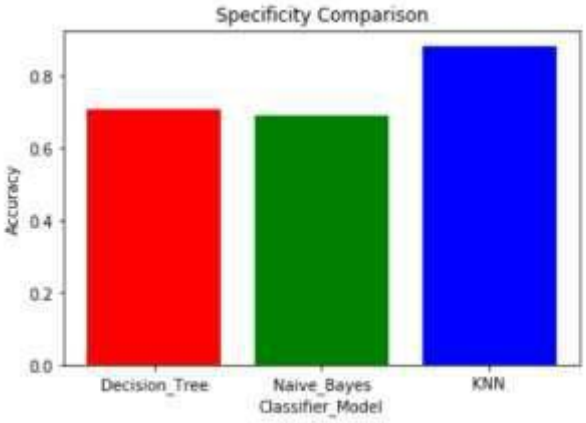

Figure 4: **Specificity Comparison:**

7. CONCLUSION

In this the classifier models are used to introduce the crop recommendation system so that the project can be efficient. The crops of various fields can be tested by the present system and it can also be scalable. The prediction analysis of the crops, sowing time, harvesting, growth of plants can be done through the graph yields. For large datasets the naïve bayes are more efficient in comparison with the decision trees and also the results are more accurate for naïve bayes. The combination of the naïve bayes and the decision trees classifier as a whole that is termed as the algorithm of combination classification perform better in comparison with the classifier model that is single in nature.

REFERENCES

[1.] S. Veenadhari, Bharat Misra, C.D. Singh, "Machine Learning Approach for Forecasting Crop Yield Based on Climate Parameters", International Conference on Computer Communication and Informatics, pp.1-5, 2014.

[2.] Renato L.F. Cunha; Bruno Silva; Marco A.S. Netto, "A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast", IEEE 14th International Conference on E-Science, pp.1-6, 2018.

[3.] Neha Rale, Raxi Kumar Solanki, Doina Bein, James Andro Vasko, wolfgang Bein, "Prediction of Crop Cultivation", pp.1-5, 2019.

[4.] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques", 1st International Conference on Next Generation Computing Applications, pp.1-5, 2017.

[5.] G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi, "Predictive Analysis to

Improve Crop Yield Using a Neural Network Model", IEEE International Conference , pp.1-4, 2018.

[6.] Rashmi Priya, Dharavath Ramesh, "Crop Prediction on the Region Belts of India: A Naïve Bayes Map Reduce Precision Agricultural Model" International Conference on Advances in Computing, Communications and Informatics, pp.1-6, 2018.

[7.] Talha Siddique,Dipro Barus, Zanntual Fredous, Amitabh Chakravarti. "Automated Farming Prediction" Intelligent Systems Conference (IntelliSys), pp.1-5, 2017.

[8.] Takeshi Yoshida Noriyuki Murakami and Hiroyuki Tauiji, "A Hybrid Machine Learning Approach to Automatic Plant Phenotyping for Smart Agriculture", International Joint Conference on Neural Networks (IJCNN), pp.1-5, 2017.

[9.] Rakesh Kumar ; M.P. Singh ; Prabhat Kumar ; J.P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp.1-5, 2015.

[10.] Ramalakshmi.M, K.Thulasimani, "Automatic Plant Phenotyping for Smart Agriculture Using Machine Learning Approach", Journal of Applied Science and Computations, Vol. VI, Issue IV, pp. 2680- 2688, 2019.

[11.]