Agriventure – Data analytics for farming and agro-businesses

Akshita Biyani
apbivani333@gmail.com
K. J. Somaia College of Engineering,
Mumbai, Maharashtra

Bhoomi Bhanushali
bnni1@somaiya.edu
K. J. Somaiya College of Engineering,
Mumbai, Maharashtra

Charmi Savla
savla.cp@somaiya.edu
K. J. Somaiya College of Engineering,
Mumbai, Maharashtra

Drishti Jain
drishtijain@somaiya.edu
K. J. Somaiya College of Engineering,
Mumbai, Maharashtra

Sangeeta Nagpure
sangeetanagpure@somaiya.edu
K. J. Somaiya College of Engineering,
Mumbai, Maharashtra

ABSTRACT

Agricultural statistics and forecasts are a valuable resource that the government has not fully used despite their significance. Our project’s goal is to automate this process by incorporating data mining and analytics concepts. More precisely, our project aims to address the social issue of drought by analyzing data for every crop in the state of Maharashtra, including crop statistics, rainfall, temperature and strain, production data, and other factors. Efficient countermeasures and recommendations would be provided based on the detailed studies conducted as part of this initiative, which, if applied quickly, will assist in addressing the drought issue in our state. It will also include forecasts for increases or decreases in consumer demand for specific agricultural goods, which will support agro-based sectors and enterprises. Data may be analyzed to uncover different patterns, such as recommendations to farmers for growing specific crops based on soil type and weather forecasts, district-level rainfall, and increases or decreases in market demand for specific agricultural goods. The project’s final product will be research-driven publications that detail these patterns and will be based on data collected over the last three years. Drought mitigation measures would be proposed and also crop suggestions for the drought-prone regions, which will aid in the productive operation of AgriviBusinesses.

Keywords: Agriculture, Drought Prediction, Crop Suggestion, Machine Learning

1. INTRODUCTION

One of the significant occupations in our nation is Agriculture, and a significant product of an assortment of yields come from the province of Maharashtra. With a huge agricultural sector in the area, the climatic conditions winning here are to some degree conflicting to what in particular is required. Maharashtra is a drought ridden state, with one of the most obliterating dry seasons happening just in the ongoing past. Drought is a delayed time of anomalous low precipitation prompting a shortage of water for human use and cultivating.

In agricultural crop yield analysis Data Mining is an evolving field. In our project, we are concentrating on the applications of Data Mining methods in the agricultural field. In data mining various mining techniques like the Support Vector Machines (SVM), K-Nearest Neighbour (KNN), etc. are used in recent data mining applications in the field of agriculture.

We have considered the problem of yield production in this project. Yield prediction is a critical agricultural issue that has yet to be solved using current data. With the help of Data Mining Methodologies the problem of Yield Production can be solved. The aim of this study is to identify appropriate data models that can predict yields with high precision and generality. Various forms of Data Mining techniques were used for this purpose. Currently, the Maharashtra government and the Indian government gather data in its raw form, which is irrelevant to the end consumer, namely farmers. Our project's goal is to collect this raw data, standardize it, analyse it, and feed it into a framework that can generate relational patterns. Using various parameters such as amount of rainfall, production of crops, area-wise statistics, agricultural inputs etc., agricultural tendencies will be formulated and a report-based system will provide solutions to farmers for eschewing droughts and sustaining development.

The existing research has mainly concentrated on compilation and cataloguing of raw data. This data in the raw form can be standardized. Once standardization is done, it can be analyzed and we can find different kinds of trends which will help in building solutions for farmers. This project is suitable to be used in areas that have similarities. In terms of droughts and agricultural produce Indian states like Gujarat, Madhya Pradesh, Maharashtra are facing such similar situations. This methodology will help propose ways to get rid of or lessen the impact of droughts in these states by extracting reliable data and applying proper algorithms to that data.
2. LITERATURE REVIEW

The following articles were found in our analysis of related literature for the use of classification algorithms in agriculture. The following section provides a concise summary of each article: For agricultural decision-making, we make use of vast datasets of soil and temperature. The data is in a ready-to-use format for a robust soil classification system that was developed in collaboration with international partners. A significant subset of the soil database, obtained from different areas, was subjected to statistical analysis and data mining. The paper concludes that the classification algorithms, of which Naive Bayes produced the best results, can be used to classify soil profiles and verify true trends in soil data.[1]

The paper illustrates the following data mining measures: data scrubbing, data extraction, data storage, data synthesis, analysis of data, mining of data, and representation of knowledge. Further, it describes how structural data mining can be helpful in obtaining patterns from vast datasets. Also, it gives the steps included in k-means clustering. The above paper ends with visualisations of the study of clusters for parameters like rainfall and temperature, as well as with suggestion that the findings may be better by taking into account more variables.[2]

In the next paper, an architecture for predicting crop production success based on crop, soil and soil inputs, temperature, and other factors is proposed. For the generation of rules and classification of test results, the proposed scheme proposes using C4.5, Naive Bayes, ANNs, and Decision Trees. On the basis of historical data of the above parameters, the approach would help in agricultural improvement by rendering predictions as per the season in which the crop is grown and the current weather. This study's aim is to optimise farmers' profits by maximising yield of the crops.[3]

This study gives an coherent description of data mining approaches that have previously been used on agricultural datasets. Classification, Clustering, Association Rule Mining, and Regression are among the data mining strategies discussed in the article. The paper then goes on to include a few instances of these algorithms, one of which is as follows. Support Vector Machines can be used to classify crops and predict shifting weather conditions.[4]

After reviewing all of the documents, classification algorithms seem to be the best fit for implementing our suggested approach. Because of their ability to deal with agricultural results, as seen in the aforementioned documents, and in the case of evolving weather scenarios, we will be using Decision Tree – Iterative and Random Forest Classifier algorithms.[5]

3. METHODOLOGY

The project has two phases, the first phase is where we have studied various algorithms like Support Vector Machines (SVM), Iterative Dichotomiser 3 (ID3) and Random Forest and have worked on building a model with the algorithm which gives us the best accuracy. The second phase is the prediction of Drought in various regions of Maharashtra and also Crop Yield prediction in various regions of Maharashtra. This prediction can help in finding if there is possibility of drought in the region and the framers can plan things accordingly.

Following are the methods we used for data collection and implementation.

Data Collection: Data was collected through various sources mentioned below. This can be accomplished by directly importing files already in .csv or .xlsx format. The data collected consisted of parameters like rainfall, temperature and pressure and crop statistics of the last 16 years.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.indiaweteorportal.org/articles/district-wise-monthly-rainfall-all-data-list-raingaustations-india-meteorological-department">http://www.indiaweteorportal.org/articles/district-wise-monthly-rainfall-all-data-list-raingaustations-india-meteorological-department</a></td>
<td>For rainfall</td>
</tr>
<tr>
<td><a href="http://www.timeanddate.com">http://www.timeanddate.com</a></td>
<td>For temperature and pressure</td>
</tr>
<tr>
<td><a href="https://www.kaggle.com/abhinand05/op-production-in-india">https://www.kaggle.com/abhinand05/op-production-in-india</a></td>
<td>For crop production</td>
</tr>
</tbody>
</table>

Data Preprocessing: Missing data values can occur in any dataset in any data mining application. These missing values have the potential to degrade the efficiency of a data mining system and must be handled accordingly. This can be accomplished by describing the missed values with an approximation of the mean value. Missing values for parameters such as ‘Rainfall,’ ‘Temperature,’ ‘Pressure,’ and ‘Crop Productivity’ are substituted with their average rainfall. This data can then be restructured to make it more easily accessible by algorithms. We restructured our data such that it was processed year by year, month by month, like the two axes.

Data Transformation: Transformation of data refers to changing the form of data. Various classification algorithms require data as discrete values as their input and they do not work with numeric values. The algorithms in this project require discrete data hence we have assigned labels to the numeric values. This is done by dividing the data into various intervals. Here the step size of the interval is according to the crop requirements. The rainfall, temperature, pressure data is divided into 9.6, 6 intervals respectively.

Data Integration: Until being fed into a classification algorithm, various parameters and their related labelled tuples must be combined. We accomplished this by creating a special dataframe in pandas that imported data from other data warehouses.

Data Mining – Classification: The original learning algorithm runs at this point, further constructing a model that will identify any unlabeled instances. The integrated dataframe obtained in the previous step is divided into training which is 70% and the remaining 30% is equally divided into testing and validation data for this purpose. The data is then fed into the algorithms, which create a model. To implement these algorithms, we use scikit-learn, which has built-in functions such as fit(), which builds a model based on training data, and predict(), which predicts labels for unlabeled data using the model.

Visual Representation: Any data mining application's final results must be visualised in order for them to be easily interpreted and understood. Confusion matrices and accuracy parameters are used to assess the success of classification algorithms. It's a special table structure that helps you to see how an algorithm, usually a supervised learning algorithm, does. In a projected class, instances are expressed in each column of the
matrix, while instances in a real class are represented in each row (or vice versa).

4. IMPLEMENTED SYSTEM

Classifier 1:
The system builds a model for comparisons of 3 classification algorithms to find the most accurate algorithm among them. For this model, the highest accuracy was given by the SVC algorithm. Also, the RFC algorithm depicted a close accuracy value as in the SVC algorithm.

SVC Algorithm

1. Random Forest Classifier

For the prediction of drought, we are taking input as location, year, and month for which we need to predict drought.

Drought Prediction

<table>
<thead>
<tr>
<th>Location</th>
<th>Year</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>2022</td>
<td>June</td>
</tr>
</tbody>
</table>

**Fig 3.15 Input for Drought Predictor**
As we can see from the above diagram 3.13 we have taken the input for Yawatmal District for the month April in year 2022. After submitting the selected inputs we can see the output (figure 3.14) showing a high drought probability in that area at that time of the year.

**Classifier 2:**
We tested the prediction model for various crops. We discovered that the accuracy of the prediction model could be increased if we fed the data of individual crops to the dataset. The performance of the algorithms vary from crop to crop here. The accuracy of the model varied for different crops. Considering the models for all the crops, the RFC and SVM algorithms gave better accuracy. Taking into consideration 2 crops, Jowar and Sugarcane.

**Jowar**

1. **ID3 Algorithm**

   ![Fig 3.7 ID3 Algorithm Jowar](image)

   **Fig 3.7 ID3 Algorithm Jowar**

2. **SVC Algorithm**

   ![Fig 3.8 SVC Algorithm Jowar](image)

   **Fig 3.8 SVC Algorithm Jowar**

**Sugarcane**

1. **ID3 Algorithm**

   ![Fig 3.10 ID3 Algorithm Sugarcane](image)

   **Fig 3.10 ID3 Algorithm Sugarcane**

2. **SVC Algorithm**

   ![Fig 3.11 SVC Algorithm Sugarcane](image)

   **Fig 3.11 SVC Algorithm Sugarcane**

3. **Random Forest Classification**

   ![Fig 3.12 RFC Algorithm Sugarcane](image)

   **Fig 3.12 RFC Algorithm Sugarcane**

© 2021, www.IJARIIT.com All Rights Reserved
As seen from the above crops, SVC and RFC both are equally accurate for jowar whereas for Sugarcane we can see that SVC Algorithm gives the best accuracy.

**Module 2:**

I. **Crop Yield Prediction**

Here we had a dataset with the data of the crops grown in Maharashtra, their yearly production (tonnes), area (hectares), season of the crop. The temperature and rainfall data was integrated by the previous classifier data. We created prediction models for 10 major crops grown in the state. Crops: ['Arhar/Tur', 'Bajra', 'Jowar', 'Moong(Green Gram)', 'Soyabeen', 'Urad', 'Gram', 'Wheat', 'Sugarcane']

Here, we ask the user to enter the region, year for prediction, crop name and the season to be grown in as inputs. This model classified the crop productivity as, low, medium, high and very high.

As we can see in the figure 3.15 we are taking location as Nagpur for the year 2022 and considering the jowar crop in kharif season. The output in the figure 3.16 shows that in kharif season there is possibility that jowar yield would be low.

**5. CONCLUSION**

This project focuses on the use of machine learning and data mining algorithms in agriculture. Early drought forecasts and crop advice in impacted regions, if delivered to end-users in a proper manner, would greatly assist drought-affected villages and districts. Traditional agricultural practices would be augmented by decision-making based on theoretical models. Overall, the SVM – Support Vector Machines – algorithm proved to be the most successful in providing predictions for Classifier 1 in our project. Classifier 1 – Drought Classification reached up to 86.90 percent accuracy. These findings can be improved and used in further studies.

**6. FUTURE SCOPE**

The further features you add to a learning-based framework, the more reliable the outcomes would be. As a result, after examining a small set of attributes, we can extend our reach by using additional criteria such as soil quality, agricultural inputs, soil nutrients, and irrigated area.

These criteria should account for data inconsistencies while also improving precision by a factor of ten. Instead of using fixed intervals, unsupervised clustering to label data for classifiers would increase precision. The integration of the process from data collection to report generation would provide us with a dependable end-to-end framework. Our current project simply classified instances based on a number of predefined parameters, but future implementations will be able to anticipate these parameters, allowing us to predict the drought and crop viability in advance.

After effectively introducing it for various divisions in the state of Maharashtra, this scheme can be adopted in various other states and countries in the future, solely to relieve the misery of farmers. This is something that governments in all states that are dealing with similar problems can do to help them solve their issues.

**7. REFERENCES**


