Crop Prediction

Aanshi Patwari (AU1841004), Miracle Rindani (AU1841017), Bhumiti Gohel (AU1841051), Dipika Pawar (AU1841052)
School of Engineering and Applied Science (SEAS), Ahmedabad University
Machine Learning (CSE523) Project

***Abstract*—About 70% of the people do farming as their occupation. Predicting the crop according to soil and weather is a major task for them. It is based on a idea of classification problem from a Machine Learning point of view. To predict the crop that should be grown under particular set of conditions, we have implemented several algorithms like KNN, SVM Classifiers.The crop prediction system based on the input parameters like soil type and weather conditions can predict the crops that can be grown which can be helpful to the farmers.**

***Index Terms*—Classification, Crop Prediction, SVM, KNN, Naive Bayes classifier**

I. Introduction

India is a primarily an agricultural country. Hence, along with a farmer's intuition and market demand details; a good knowledge of which crop to produce in specific weather and soil conditions is highly necessary for farmers so that they can produce a good yield. In today's world, predicting about crop is becoming uncertain due to problem of sudden climate change which in turn leads to poor economy as well as poor yield production in the future crops. So to design a recommendation system which can predict the crop that can be grown in a particular piece of land becomes a necessity. By considering the soil and weather related values and considering other phenomenon, the crop is predicted. This can help in reducing the pressure from the farmers about predicting the crop to be grown.

II. Literature Survey

A. Crop recommendation system for precision agriculture [1]

Precision Agriculture principle is applied on small, open farms at individual farmer and crop level. Using data preprocessing techniques(Sammon's mapping method), high dimensional agricultural data is reduced to small size data to acquire the useful parameters. Crop Selection Method(CSM) helps to solve problem of crop selection as well as improve the net yield rate of the crop. The method provides output in the form of crop recommendations which can be grown over various seasons considering the factors of weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. K–Nearest Neighbor, random tree , CHAID(Chi-squared Automatic Interaction Detection), Naïve Bayes algorithm are used to improve the accuracy of the prediction to 80 to 88. The dataset used contains 5 environmental variables, 3 biotic variables and 2 area related variables to determine the crop yield in different districts.

B. The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms [2]

The dataset used is dependent on the soil quality which is determined based on soil's NPK values, soil-PH value, crop disease and pesticides, seasonal parameters such as Kharif, rabbi, and summer crops. Data is collected from National Oceanic and Atmospheric Administration (NOAA) which contains climate and weather-related parameters and the Indian government agricultural portal www.data.gov.in. Algorithms like The neural network, SVM, Naive Bayes, KNN are implemented.

C. Crop Prediction Using Machine Learning [3]

Dataset containing 1850 number of records and 5 number of features is used. It contains soil parameters as the main attributes like Nitrogen, Phosphorus, Potassium and pH values of soil. KNN algorithm is used to train the model with a value of k=10 (best value) and accuracy of 85%. Also KNN with Cross Validation, Decision Tree, Naive Bayes and SVM are used for the comparison of performance in which kNN with Cross Validation performed the best with 88% accuracy.

D. Crop Prediction System using Machine Learning Algorithms [4]

The dataset consisting of soil, weather and production parameters is used to predict the best suitable crop. The preprocessing steps are well defined. Some methods to find most correlated features for the target column like Gini index, Entropy and Information Gain are discussed. KNN algorithm is used to predict the rainfall and weather report. Decision tree, KNN classifier and Naive Bayesian algorithms are used for crop prediction out of which the Decision tree gives poor results and the Naive Bayes gives the best results. The combination of both the classifiers also gives the good result.

III. Implementation

For our crop prediction system we have used the dataset that mainly consists of features related to weather conditions and soil nutrient and pH levels at a particular place. The dataset contains 22 categories of crop labels on which we have trained the model. For the preprocessing part, we have calculated the correlation between the feature values with the crop label and with each other to make sure that the columns were i.i.d in nature. The major steps involved in preprocessing are:

1) Removing columns with high correlation values.
2) Deleting all the rows with one or more null values.

3) Equalising the number of data points for each crop label so that the final dataset we get is not biased towards any crop.
4) Extracting all the independent variables (soil and weather data) in a separate matrix and the labels (crop labels) in an array.
5) Dividing the dataset into train (70%) and test (30%) sets.

The main algorithms we implemented on our dataset for the classification task were:

1) **SVM (Support Vector Machines)**
   For SVM, the polynomial kernel, gave the accurate results.
2) **KNN (K-Nearest Neighbours)**
   For KNN, we have tried it against various k value = 8 or 10, n value = 22, and leaf size = 30; the algorithm gave it's best result.
3) **Naïve Bayes Classifier**
   For Naive Bayes, we have used the in-built GaussianNB() classifier and it gave the best accuracy results on the dataset.

The implementation of these algorithms is done in python using the scikit-learn library.

IV. Results

A. Comparison of Various Algorithms

The following section shows the accuracy results of various algorithms employed in by us:

Algorithm	Train Accuracy	Test Accuracy
SVM (scratch)	94.74%	92.87%
SVM (in-built)	98.92%	97.95%
Naive Bayes	99.61%	99.09%
KNN	95.15%	93.63%

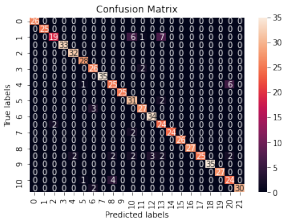
Here, we can see that KNN gives the lowest accuracy since it is a lazy algorithm. From the given algorithms, we chose to employ SVM from scratch mainly because it gave a comparable accuracy to Naive Bayes and it handles the interactions between the features in a better way as compared to Naive Bayes classifier.
The following table shows the space and time complexities of these algorithms.

Algorithm	Space Complexity	Time Complexity
SVM	*O(n*d)*	*O(n*d)*
Naive Bayes	*O(d*m)*	*O(n*d*m)*
KNN	*O(n*d)*	*O(n*k*d)*

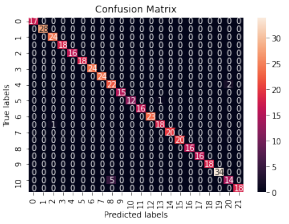
where,
n = number of training examples,
d = number of features,
m = number of classes, and
k = number of nearest neighbours.

B. Graphs Generated

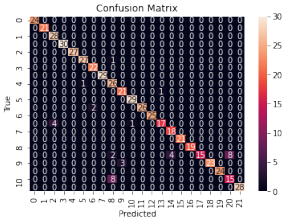
The following are the confusion matrices generated by the algorithms employed.



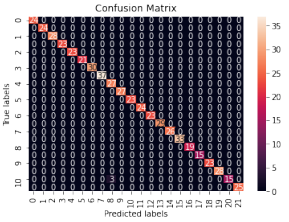
(a) SVM (Scratch)



(b) SVM (In-built)



(c) KNN



(d) Naive Bayes

Fig. 1: Confusion matrices results

We also did plots of training loss on the SVM algorithm built from scratch. Below are the results of the same:

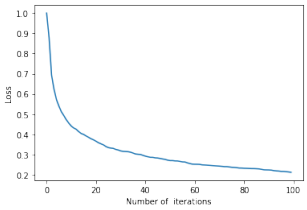


Fig. 2: Train Loss results of SVM (scratch)

V. Conclusion

The SVM gives almost similar results like that of Naive Bayes algorithm. Any of the two algorithms can be used for crop prediction based on the soil and the weather conditions which can help to determine which crop can be grown in which season. Work on improvising the KNN algorithm needs to be done.

References

[1] R. H. R. C. K. T. K. S. Pudumalar, E. Ramanujam and J. Nisha, "Crop recommendation system for precision agriculture."
[2] B. Viviliya and V. Vaidhehi, "The design of hybrid crop recommendation system using machine learning algorithms."
[3] M. M. S. L. V. Kevin Tom Thomas, Varsha S and E. J. Thomas, "Crop prediction using machine learning."
[4] V. Pavan and Shrikant, "Crop prediction system using machine learning algorithms."