

ABSTRACT

In Indian economy and employment agriculture plays major role. The most common problem faced by the Indian farmers is they do not opt crop based on the necessity of soil, as a result they face serious setback in productivity. This problem can be addressed through precision agriculture. This method takes three parameters into consideration, viz: soil characteristics, soil types and crop yield data collection based on these parameters suggesting the farmer suitable crop to be cultivated. Precision agriculture helps in reduction of non-suitable crop which indeed increases productivity, apart from the following advantages like efficacy in input as well as output and better decision making for farming. Crop yield prediction incorporates forecasting the yield of the crop from past historical data which includes factors such as temperature, relative humidity, ph., rainfall and area (Hectares). This method gives solutions like proposing a recommendation system through an ensemble model with majority voting techniques using Random Forest and K Nearest Neighbor as learner to recommend suitable crop based on soil parameters with high specific accuracy and efficiency.

LIST OF ABBREVIATIONS

SR. NO	ABBREVIATION	EXPANSION
1	KNN	K Nearest Neighbor

LIST OF TABLES

SR. NO	NAME OF TABLE	PAGE NUMBER
1	ALGORITHM COMPARISION	31

LIST OF FIGURES

FIGURE. NO.	NAME OF FIGURE	PAGE NUMBER
3.1	Classification vs Regression	10
3.2	System Architecture	12
4.1	Random Forest	13
4.2	KNN	14
5.1	Density Diagram	24
5.2	Confusion Matrix for Random Forest	25
5.3	Confusion Matrix For KNN	26
5.4	Correlation Diagram	27
5.5	Correlation Matrix Plot Diagram	28
5.6	Accuracy Formula	29
5.7	Precision Formula	29
5.8	Recall Formula	30
5.9	F1-score Formula	30

5.10	Algorithm Comparison	31
5.11	Hist Diagram	33
5.12	State Prediction	34
5.13	Crop Prediction	35

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	i
	LIST OF ABBREVIATIONS	ii
	LIST OF TABLES	iii
	LIST OF FIGURES	iv
1	INTRODUCTION	1
	1.1 OUTLINE	1
	1.2 MODEL IDE	1
	1.3 PROBLEM STATEMENT	2
	1.4 PROPOSED SYSTEM	4
	1.5 OBJECTIVE	4
2	LITERATURE SURVEY	5
3	METHODOLOGY	8
	3.1 INTRODUCTION TO MACHINE LEARNING	8
	3.2 TRAINING THE DATA	9
	3.2.1 SUPERVISED LEARNING	9
	3.2.2 UNSUPERVISED LEARNING	9
	3.3 METHODS IN SUPERVISED LEARNING	10
	3.3.1 CLASSIFICATION	10
	3.3.2 REGRESSION	11
	3.4 SYSTEM ARCHITECTURE	11

4	ALGORITHMS	13
	4.1 RANDOM FOREST	13
	4.2 K NEAREST NEIGHBOUR	14
	4.3 DATASETS	14
	4.4 PACKAGES	15
	4.4.1 DATA MANIPULATION PACKAGES	16
	4.4.2 MODEL BUILDING PACKAGE	17
	4.4.3 DATA VISUALIZATION PACKAGES	18
	4.5 SYSTEM REQUIREMENTS	19
	4.6 BACKGROUND STUDY	20
	4.6.1 OVERVIEW	20
	4.6.2 YIELD PREDICTION	20
5	RESULTS AND DISCUSSION	23
	5.1 RESULTS	23
	5.2 ANALYSIS	33
6	CONCLUSION AND FUTUREWORK	37
	6.1 CONCLUSION	37
	6.2 FUTURE SCOPE	37
	REFERENCES	38
	APPENDIX	41
	A. SAMPLE CODE	41
	B. SCREEN SHOTS	46
	C. PUBLICATION WITH PLAGIARISM REPORT	50

CHAPTER 1

INTRODUCTION

1.1 OUTLINE

Agriculture is the one amongst the substantial area of interest to society since a large portion of food is produced by them. Currently, many countries still experience hunger because of the shortfall or absence of food with a growing population. Expanding food production is a compelling process to annihilate famine. Developing food security and declining hunger by 2030 are beneficial critical objectives for the United Nations. Hence crop protection; land assessment and crop yield prediction are of more considerable significance to global food production.

This project uses Python 3.6 for the programming in a scientific development environment called the PyCharm. Various data manipulation, machine learning and visualization packages are used to create and analyze the dataset using a traditional machine learning model. A Data Visualization tool called Tableau is used to interpret the results provided by the model after analysis to represent and act as a proof for the intended result.

1.2 MODEL IDE

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python programming language. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda.

1.3 PROBLEM STATEMENT

Problem # 1. Instability:

Agriculture in India is largely depending on monsoon. As a result, production of food-grains fluctuates year after year. A year of abundant output of cereals is often followed by a year of acute shortage. This, in its turn, leads to price income and employment fluctuations. However, for the thirteen years, in successive (1987-88 to 1999-00) a normal monsoon has been observed.

Problem # 2. Cropping Pattern:

The crops that are grown in India are divided into two broad categories: food crops and non-food crops. While the former comprises food-grains, sugarcane and other beverages, the latter includes different kinds of fibers and oilseeds. In recent years there has occurred a fall in agricultural production mainly due to fall in the output of non-food articles. Moreover, rabi production has become as important as kharif production in the late 1990s. In 1999-2000, for example, of the total grain production of 209 mn. tones, rabi accounted for 104 mn. tones. This indicates a structural change in agricultural production.

Problem # 3. Land Ownership:

Although the ownership of agricultural land in India is fairly widely distributed, there is some degree of concentration of land holding. Inequality in land distribution is also due to the fact that there are frequent changes in land ownership in India. It is believed that large parcels of land in India are owned by a- relatively small section of the rich farmers, landlords and money-lenders, while the vast majority of farmers own very little amount of land, or no land at all. Moreover, most holdings are small and uneconomic. So, the advantages of large-scale farming cannot be derived and cost per unit with 'uneco- nomic' holdings is high, output per hectare is hectare is low. As a result, peasants cannot generate sufficient marketable surplus. So, they are not only poor but are often in debt.

Problem # 4. Sub-Division and Fragmentation of Holding:

Due to the growth of population and breakdown of the joint family system, there has occurred continuous sub-division of agricultural land into smaller and smaller plots. At times small farmers are forced to sell a portion of their land to repay their debt. This creates further sub-division of land. Sub-division, in its turn, leads to fragmentation of holdings. When the size of holdings become smaller and smaller, cultivation becomes uneconomic. As a result, a major portion of land is not brought under the plough. Such sub-division and fragmentation make the efficient use of land virtually impossible and add to the difficulties of increasing capital equipment on the farm. All these factors account for the low productivity of Indian agriculture.

Problem # 5. Land Tenure:

The land tenure system of India is also far from perfect. In the pre-independence period, most tenants suffered from insecurity of tenancy. They could be evicted any time. However, various steps have been taken after Independence to provide security of tenancy.

Problem # 6. Conditions of Agricultural Labourers:

The conditions of most agricultural labourers in India are far from satisfactory. There is also the problem of surplus labour or disguised unemployment. This pushes the wage rates below the subsistence levels.

Problem # 7. Other Problems:

There are various other problems of Indian agriculture.

These are related to:

- (i) The systems and techniques of farming,
- (ii) The marketing of agricultural products and
- (iii) The indebtedness of the farmers.

1.4 PROPOSED SYSTEM

- Our proposed system is an application which predicts name of the crop as well as calculate its corresponding yield.
- Name of the crop is determined by several features like temperature, humidity, wind-speed, rainfall etc. and yield is determined by the area and production.
- In this project KNN and Random Forest is used for prediction. It will attain the crop prediction with best accurate values.

1.5 OBJECTIVE

In Indian economy and employment agriculture plays major role. The most common problem faced by the Indian farmers is they do not opt crop based on the necessity of soil, as a result they face serious setback in productivity.

This problem can be addressed through precision agriculture. This method takes three parameters into consideration, viz: soil characteristics, soil types and crop yield data collection based on these parameters suggesting the farmer suitable crop to be cultivated.

Precision agriculture helps in reduction of non-suitable crop which indeed increases productivity, apart from the following advantages like efficacy in input as well as output and better decision making for farming.

This method gives solutions like proposing a recommendation system through an ensemble model with majority voting techniques using Random Forest and K - Nearest Neighbor as learner to recommend suitable crop based on soil parameters with high specific accuracy and efficiency.