Crop Recommendation and Price Prediction using Machine Learning

Vishnu Sadasivan
Department Of Computer Application
Amal Jyothi College Of Engineering
Kanjirappally,India
vishnusadasivan2023b@mca.ajce.in

Sr. Elsin Chakkalackal S.H
Department Of Computer Application
Amal Jyothi College Of Engineering
Kanjirappally,India
elsinchakkalackal@amaljyothi.ac.in

***Abstract*— In rural places, agriculture is the main source of income. The vast majority of farmers in India rely on their instincts to choose which crop to grow during a certain season. The most frequent issue that a farmers facing is that they don't choose the right crop based on the needs of their soil and climate condition in that area. Therefore, productivity is affected. The KNN algorithm and linear regression are used in this paper to propose a machine learning technique for crop recommendation and cost prediction. The goal is to accurately inform farmers about the best crop to plant in their area and calculate the selling price of crop. The suggested method makes use of PH of the soil, nitrogen in the soil, phosphorous in soil, potassium in soil and weather data to train the machine learning models to predict the crop suitable for cultivation in a specific area and uses the sales dataset of a e-commerce website for predicting the selling price of a crop.**

***Keywords*- Price, crop, Prediction, agriculture, Linear regression**

KNN algorithm, Machine Learning

I. INTRODUCTION

The farmers faces a frequent problem that they don't choose the right crop based on the needs of their soil and climate in their area. Productivity suffers as a result. A single farmer, however, cannot be expected to take into account all of the various factors that affect crop development when determining which crop to plant. Increased productivity will enable farmers to make more money from the same amount of land while using less labour. The peculiarities of the soil and other environmental inputs might make crop production challenging. Estimates of season crop yields are inconceivably acknowledged as a key contribution for investigating food accounting reports and yield deficits.

One of the most foundational supervised learning-based machine learning algorithms is K-Nearest Neighbour. A new data point is classified using the algorithm known as K-NN based on similarity after storing all the prior data. Based on the closeness of weather patterns between historical data and the current season, the KNN algorithm is utilized here to suggest crops suitable for the area. This method considers the weather patterns that have resulted in the highest yield for a particular crop in a particular area and soil in that area. The target variable (crop) is encoded into numerical labels using the LabelEncoder class from the sklearn.preprocessing module.

The recommended crop's production cost is estimated using linear regression. The model is trained using sales for similar type of crop on a e-commerce website. The expected crop cost entered by the user is used by the program to predict selling price of the crop. The results of the studies show how well the suggested method works at suggesting crops and estimating their selling price of the crop. Overall, the suggested method can assist farmers in selecting crops to plant and determining the selling price of their crop. This can raise agricultural output and cut down on resource waste, which will ultimately increase farmer profitability.

Challenges:

i) Data quality: The reliability of crop forecasting models is highly dependent on the efficiency of the data we are providing. The effectiveness of prediction models can also be impacted by missing or incomplete data.

ii)Feature selection: Choosing pertinent features that can affect crop production is another difficulty. Information about the climate, soil, and other relevant features may be included.

iii)Model selection: To recommend the crops, a variety of machine learning algorithms, including decision trees, the KNN Algorithm, and others, can be used

.

II. Literature Review

Manish Kumar, Dr. (Prof.) Deva Prakash conducted a study on Recommending the crop using Machine Learning technique for Big Data [1]. The dataset was gathered by the authors from open data sources like www.data.gov.in and www.data.world. They develop a machine learning model that recommends crops as a component of a tech-based crop recommendation system. The model is intended to be used by farmers as a reference tool, helping them choose which crop to plant in order to lower risk and boost agricultural output. Numerous techniques, including SVM Linear, SVM Radial, SVM Polynomial, naive Bayes, k-nearest neighbour, LDA, QDA, RDA, MDA, Decision Tree, and Random Forest model, are compared in the study. The accuracy of the algorithms was compared after training with 0.8 percent of the input data and testing with the remaining 0.2 percent of the test dataset.

S. Veenadhari, Dr Bharat Misra, Dr CD Singh proposed approach of Machine learning for crop yield forecasting based

on climatic features[2]. They showed how data mining techniques may be used to forecast agricultural yields based on climate input characteristics. The website that was created is user-friendly, and all of the crops and study districts that were chosen had predictions that were more than 75% accurate. Anyone can use the user-friendly website created for crop yield

DOI: 10.5281/zenodo.7957012
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

prediction by entering the local climate data for the crop of their choice.

Mr A Suresh ,Dr. P. Ganesh kumarDr.M.Ramalatha.Pediction of major crop yields of Tamilnadu using K-means and Modified KNN [3]. This study used K-Means and Modified KNN to estimate the primary crop outputs in Tamilnadu.Three different types of algorithms—fuzzy, KNN, and Modified KNN—were used in the analysis, with Modified KNN proving to be the best of the bunch. Future research would compare different bio-inspired techniques depending on how accurately each algorithm performed.

Monali Paul, Santosh K. Vishwakarma, Ashok Verma analyse the behaviour of the soil and crop yield prediction using Data Mining Approach[4].Data mining techniques are used in this work to classify soil into low, medium, and high categories in order to predict crop yield using the dataset that is currently available. This study can help landowners and soil researchers choose the best planting locations to increase the yield of agriculture.

Cover TM, Hart PE, K Nearest Neighbor pattern classification[5]. They suggested Since a training set is used each time a classification is made, K-Nearest Neighbour does not have a learning phase. The k-nearest neighbour algorithm is predicated on the idea that similar samples will yield similar classifications. The parameter K describes the similar known samples that are used to classify an unknown sample.

Rakesh Kumar , M.P. Singh, Prabhat and J.P. Singh found a Crop Selection Method Using Machine Learning for Boosting Crop Production Rate[6]. The research presented by them offers a method called CSM for choosing the order of harvests to be plant throughout the growing season. Crops that will be grown this season may yield more essentially thanks to the CSM approach. The recommended strategy deals with crop selection based on predicted yield rates influenced by several factors, such as weather, features of soil ,density of water density in that area and crop type.

III. METHODOLOGY

The purpose of this work is to create a machine learning model using the dataset offered that can precisely predict a crop that will grow appropriately given the input parameters provided by the user. Two sets of training and testing data are included in the dataset. The model must be trained on more data in order to increase its accuracy. This model employs the KNN algorithm to forecast crops based on variables like soil composition and climate. This paper also works to predict the selling cost of a product in an e-commerce website. The model uses the sales database of the e-commerce website to predict the cost. Linear regression is used in this model which train a model on the training data and evaluate the model based on the test data. The Linear Regression model is imported from the scikit-learn library.

A. Data Collection: For the crop prediction, Kaggle provided the training and testing datasets, and for the selling price

prediction, we use the sales database from an e-commerce website. These datasets contain crucial details about various variables that affect crop prediction, such as the pH of soil, nitrogen content, phosphorus content, potassium content, and weather feature, among others. To predict the selling price which uses cost of the similar product which is previously sold

B. Data Pre-processing: This is the first stage in any machine learning algorithm. Data cleaning, data transformation, and data minimization are all part of this process. All of this is done to improve the data's effectiveness. The data can be analyzed to improve the accuracy of our model. In order for the categorization to be correct.

a. Cleaning Data – Any null values from the training dataset were removed because they are not needed for the feature selection method. A few columns in the dataset were also removed. New columns with numerical values were created and stored for prediction after the data was processed. Additionally, the dataset's categorical data columns were eliminated. As a result, a training dataset that was appropriate for the study purposes was obtained.

b. Splitting of Data – The data is separated into a training dataset and a test dataset after formatting to create two independent datasets. After the algorithm for machine learning has been created using the training dataset, its effectiveness has been assessed using the testing dataset.

C. Machine Learning: Users can use this to predict which crop will be suitable for their situation as well as the selling price of crop. The effectiveness of machine learning algorithms depends on how they are trained. The kind of problem that needs to be solved, the resources that are available on the computer, and the type of data all have an impact on the best algorithm to use.

Fig 1: Overview of Methodology

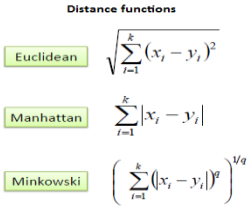
1. k-nearest neighbours Algorithm- With k-NN, it postpone the computation until following the function has been evaluated and the function is only in the area approximated. Because this algorithm relies on distance for classification, it may substantially improve accuracy if its characteristics reflect distinct physical units or have different scales.

Giving weights to neighbour their contributions can be helpful in both classification and regression since it enables the near neighbours to contribute more to the norm than the distant neighbours. A typical weighing method, for instance, assigns

each neighbour a weight of 1/d, where d is the distance that

DOI: 10.5281/zenodo.7957012
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

exists between the neighbours. The k nearest neighbour is measured by using distance function.



2. Linear Regression - A linear relationship exists between one or more unrelated variables and the variable that is dependent. It is frequently employed in the field of machine learning for both classification and regression tasks. Regression uses features from the input data to predict a continuous target variable. Using the Ordinary Least Squares (OLS) method, the model determines the line of best fit Which decreases the total errors between the determined and actual values. For a variety of tasks, such as estimating consumer demand, housing prices, and stock prices, the linear regression algorithm is clear and easy to use. when x is the independent variable, b0 and b1 are the intercepts, and y is the dependent variable and slope of the line, respectively

IV. IMPLEMENTATION

A.Crop Prediction:

The model building is the main step in the Crop Prediction. While building the model user use the algorithms

1. Import the packages that are necessary.

```
import pyttsx3
import pandas as pd
from sklearn import preprocessing
from sklearn.neighbors import KNeighborsClassifier
import numpy as np
import PySimpleGUI as sg
```

Fig 2-Packages imported

2. Add the data into a Data Frame [Fig3] then get the shape of data.

NITROGEN	PHOSPHORUS	POTASSIUM	TEMPERATURE	HUMIDITY	PH	RAINFALL	CROP
90	42	43	21	82	6.5	203	rice
85	58	41	22	80	7.0	227	rice
60	55	44	23	82	7.8	264	rice
74	35	40	26	80	7.0	243	rice
78	42	42	20	82	7.6	263	rice
69	37	42	23	83	7.1	251	rice
69	55	38	23	83	5.7	271	rice
94	53	40	20	83	5.7	242	rice
89	54	38	25	84	6.7	230	rice
68	58	38	23	83	6.3	221	rice
91	53	40	27	81	5.4	265	rice
90	46	42	24	81	7.5	250	rice
78	58	44	27	81	5.1	284	rice
93	56	36	24	82	7.0	185	rice

Fig 3-csv used for prediction

3. This code [Fig4] appears to be a part of a data preprocessing step where LabourEncoder from the scikit-learn library's preprocessing module is used to encode categorical data in an crop column of Excel file.

```
le = preprocessing.LabelEncoder()
crop = le.fit_transform(list(excel["CROP"]))
```

Fig 4-using LabelEncoder for data pre-processing

4.The values for each of the other columns (Nitrogen, Phosphorous, Potassium, Temperature, Humidity, pH, and Rainfall) are extracted from the Excel file and saved as lists after

```
NITROGEN = list(excel["NITROGEN"])  # Making the whole row consisting of nitrogen values to come into nitro
PHOSPHORUS = list(excel["PHOSPHORUS"])  # Making the whole row consisting of phosphorus values to come into phosphorus.
POTASSIUM = list(excel["POTASSIUM"])  # Making the whole row consisting of potassium values to come into potassium.
TEMPERATURE = list(
    excel["TEMPERATURE"])  # Making the whole row consisting of temperature values to come into temperature.
HUMIDITY = list(excel["HUMIDITY"])  # Making the whole row consisting of humidity values to come into humidity.
PH = list(excel["PH"])  # Making the whole row consisting of ph values to come into ph.
RAINFALL = list(excel["RAINFALL"])  # Making the whole row consisting of rainfall values to come into rainfall.

features = list(
    zip(NITROGEN, PHOSPHORUS, POTASSIUM, TEMPERATURE, HUMIDITY, PH, RAINFALL))  # Zipping all the features together
features = np.array([NITROGEN, PHOSPHORUS, POTASSIUM, TEMPERATURE, HUMIDITY, PH,
                     RAINFALL])  # Converting all the features into a array form
```

the CROP column has been encoded.

Fig 5-Extracting Excel file

5.By using the GUI the input of the user is accepted and predict1 array generate a prediction based on the input given by the user

```
event, values = window.read()
if event == sg.WINDOW_CLOSED or event == 'Quit':  # If the user will press the quit button then the program will end up.
    break
print(values[0])
nitrogen_content = values[0]  # Taking input from the user about nitrogen content in the soil.
phosphorus_content = values[1]  # Taking input from the user about phosphorus content in the soil.
potassium_content = values[2]  # Taking input from the user about potassium content in the soil.
temperature_content = values[3]  # Taking input from the user about the surrounding temperature.
humidity_content = values[4]  # Taking input from the user about the surrounding humidity.
ph_content = values[5]  # Taking input from the user about the ph level of the soil.
rainfall = values[6]  # Taking input from the user about the rainfall.
predict1 = np.array([nitrogen_content, phosphorus_content, potassium_content, temperature_content, humidity_content, ph_co
print(predict1)  # Printing the data after being converted into a array form.
predict1 = predict1.reshape(1,-1)  # Reshaping the input data so that it can be applied in the model for getting accurate
print(predict1)  # Printing the input data value after being reshaped.
predict1 = predict1.astype(float)
predict1 = model.predict(predict1)  # Applying the user input data into the model.
```

Fig 6-Accepting the input from GUI

B. Selling Price Prediction:

1. Import the packages that are necessary.

```
import pandas as pd
from sklearn.linear_model import LinearRegression
from home.models import MyProduct,MyAnalysisProduct
from sklearn.model_selection import train_test_split
```

Fig 7-Packages imported

DOI: 10.5281/zenodo.7957012
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

2.Accessing the database and convert it into a queryset. Then converting the query set to a pandas DataFrame.

```
queryset = MyAnalysisProduct.objects.filter(Product_name=Product_name)

# Convert the queryset to a pandas dataframe
data = pd.DataFrame(list(queryset.values()))
```

Fig 8-converting database to DataFrame

3. Then Create datasets for testing and training from the dataset.

```
X_train, X_test, y_train, y_test = train_test_split(data.drop('price', axis=1), data['price'], test_size=0.8,
                                                    random_state=42)

# Train a linear regression model on the training data
model = LinearRegression()
model.fit(X_train, y_train)

# Evaluate the model on the test data
score = model.score(X_test, y_test)
```

Fig 9-splitting the dataset

4. Predicting the Price

```
input_data = pd.DataFrame({'Product_name_' + Product_name: [1], 'soldprice': [1],'quantity': [1], 'Addedquantity':
input_data['soldprice'] = expected_cost
input_data = input_data[X_train.columns]  # Ensure column order is the same as in training data
predicted_cost = model.predict(input_data)[0]
```

Fig 10-Prediction using the input values

V. RESULT

The result shows that Predicted crop based on the input entered by the user and selling price based on the expected cost entered by the user. The KNN algorithm is utilized here to suggest crops and data is preprocessed using LabourEncoder from the scikit-learn. KNN helps to predict a highly accurate prediction. Upon the data entered it should Predict the values we entered is low , high or medium and suggest a crop which is suitable in that area.
For selling price Prediction we use Linear regression which provides accuracy. It reduces the estimation procedure and, even more importantly, the modular level comprehending of these linear equations is unambiguous.

Crop Recommendation

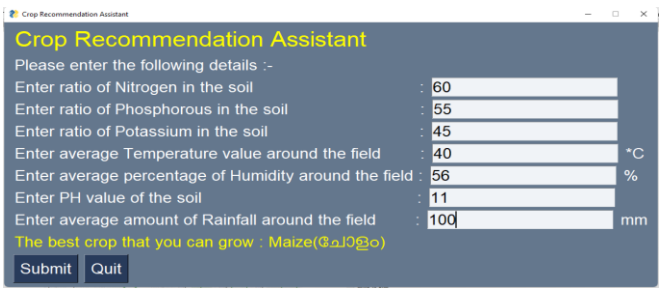


Fig 11-Prediction of crop.

Upon the values which is entered by the user it predict the humidity, potassium in soil, phosphorous in soil ,Nitrogen in soil and Rainfall is less or not and Nature of soil is acidic or alkaline.

```
medium humid
hot
less
not to less but also not to high
not to less but also not to high
less
alkaline
```

Fig 11-Prediction of crop.

The model predict the selling price of the crop[Fig 3] Based on the similar crop sold in a e-commerce website and expected value the user expecting for his crop.

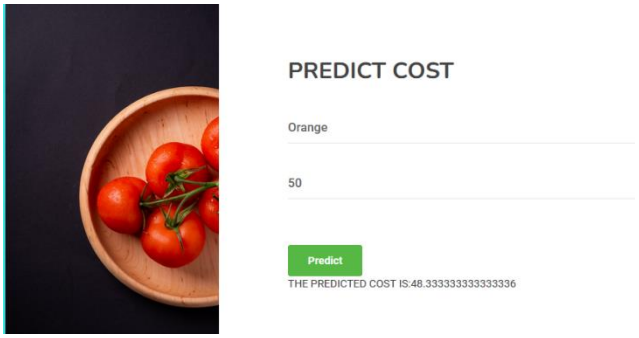


Fig 12 -Predicting the selling price of crop

VI. CONCLUSION AND FUTURE WORK

This paper explains how to Recommend a crop based on the geographical features in that area such as nitrogen, pH, phosphorous, potassium in soil and climate data using a set of collected data, which is pre-processed, modeled, and investigated to test the algorithmic rule. Machine learning methods are used to predict suitable crop accurately and predict the selling price of the crop which help the sellers to know what is the suitable price to sold their crop.
By swapping more efficient data sets, future work might try to build models that are more effective. Due to the presence of some difficulties, this experiment only employs an inadequate dataset. Consequently, a wider set of data of a minimum of 1 Gigabyte could possibly be used in subsequent work.

REFERENCES

[1] Manish Kumar, Dr. (Prof.) Deva Prakash 2022.” Crop Recommendation Predictor using machine learning for big data . International Journal of Engineering Research in Computer Science and Engineering (IJERCSE).” ISSN (Online) 2394-2320

DOI: 10.5281/zenodo.7957012
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

[2] S. Veenadhari, Dr Bharat Misra, Dr CSingh.2019.”Machine learning approach for forecasting crop yield based on climatic

parameters.”.978-1-4799-2352- 6/14/$31.00 ©2014 IEEE.

[3] Mr A Suresh, Dr. P. Ganesh kumar , Dr.M.Ramalatha.
Prediction of major crop yields of Tamilnadu using K-means and Modified KNN Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018) IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-4765-3

[4] Monali Paul, Santosh K. Vishwakarma, Ashok Verma
(2015), ‘Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach’, International Conference on Computational Intelligence and Communication Network

[5] Cover TM, Hart PE,”K Nearest Neighbor pattern classification”, IEEE Trans Info Theory 13(1) : 21-27, 1967.

[6] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), ,International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICST) conference on Computational Intelligence and Communication Network

[5] Cover TM, Hart PE,”K Nearest Neighbor pattern classification”, IEEE Trans Info Theory 13(1) : 21-27, 1967.

[6] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), ,International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICST)

DOI: 10.5281/zenodo.7957012
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam