AI Product/Service Business & Financial Modelling
Crop Recommendation App using Machine Learning

Umang Tiwari
28/08/2023
Task-2

INTRODUCTION

This project can be broken down into 4 parts:

1. Prototype Selection:

a. Feasibility: Product/Service can be developed in the short term future. (2-3 years)

b. Viability: Product/Service should be relevant or able to survive in the long term future. (20-30 years)

c. Monetization: Product/Service should be monetizable directly.

2. **Prototype Development:** Small scale code implementation/model building of the Prototype. Prototype development involves creating a small-scale implementation or model of the AI product or service to demonstrate its functionality and feasibility. This stage focuses on converting the conceptual ideas into a tangible prototype that can be tested and iterated upon. This step is optional if we implement a Machine Learning model.

3. **Business Modelling:** Developing a business model for the AI Product/Service.It will also include the monetization plans of the business.

4. **Financial Modelling (equation**) with Machine Learning & Data Analysis

a. Identify which Market your product/service will be launched into by performing market segmentation.

b. Collect some data /statistics regarding that Market Online.

c. Perform forecasts/predictions on that Market using regression models or time series

d. Design Financial Equation corresponding to that Market Trend.

Prototype Selection

The farming sector faces multiple risks and uncertainties due to changing climatic conditions and market trends, which results in significant production losses and wasted resources. With the ever increasing population as well as limited land resources there is a need to establish a system where maximum yield can be generated without over-utilizing the land in use.Therefore there is the requirement of knowing the soil type,weather conditions, previous cropping patterns beforehand so that our farmers are able to maximize their yield and minimize their losses which would be caused due to crop diseases,climate change and insect and pest attacks.

The aim is to make a ML model which takes data(soil analysis,weather report etc.) and trains itself using various Machine Learning techniques and Algorithms(Random Forest, Decision Tree) and predict the yield and best fertilizer that suits for the crops in a virtual environment by considering the overall factors that contribute in his overall yield.
Evaluating the following criteria for our project:
1. Feasibility:
The idea of a crop recommendation system is feasible in the short term future. The technology required for data collection, analysis, and machine learning is already available and can be integrated into a website or mobile application. The development of the system may take some time, but it is achievable within 2-3 years.
2. Viability:
The crop recommendation system is relevant and can survive in the long-term future. Agriculture is a vital sector for the economy and food security of any country, and the demand for optimized agricultural practices will only increase with time. The system can adapt to changing environmental conditions, crop varieties, and user requirements, making it a sustainable solution for the agricultural sector.
3. Monetization:
The crop recommendation system can be monetized directly through various revenue streams. The website can offer premium consulting services to farmers, charge a subscription fee for advanced features, and earn a commission from affiliate marketing or collaborations with agriculture companies. The system can also provide data analytics services to research institutions and generate revenue from advertising. Therefore, the product/service is monetizable directly.

Prototype Development

Exploratory Data Analysis

In [41]:
```
df.head()
```

Out[41]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

In [42]:
```
df.size
```

Out[42]: 17600

Types of Crops and their count

In [46]:
```
df['label'].unique()
```

Out[46]:
```
array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
       'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
       'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
       'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
      dtype=object)
```

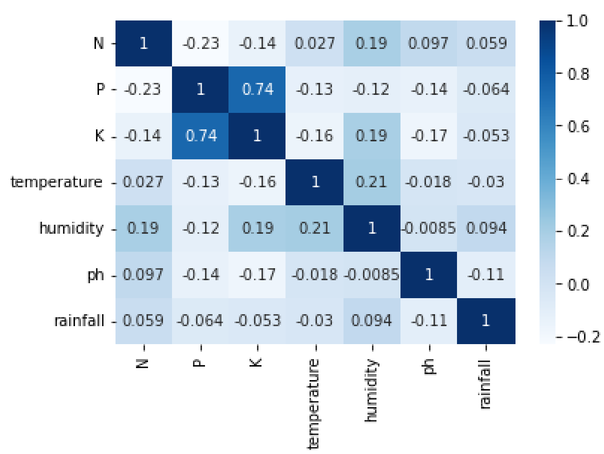
In [47]:
```
df['label'].value_counts()
```

Out[47]:
```
rice           100
maize          100
jute           100
cotton         100
coconut        100
papaya         100
orange         100
apple          100
muskmelon      100
watermelon     100
grapes         100
mango          100
banana         100
pomegranate    100
lentil         100
blackgram      100
mungbean       100
mothbeans      100
pigeonpeas     100
kidneybeans    100
chickpea       100
coffee         100
Name: label, dtype: int64
```

Heat Map:

In [49]:
```
sns.heatmap(df.corr(),cmap="Blues",annot=True)
```

Out[49]: <AxesSubplot:>



Market Segmentation:

The dataset used is crop_production.csv , the shape of the dataset is (246091,7). The column
values are State_Name, District_Name, Crop_year, Season, Crop, Area, Production

Out[2]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

Data Description

In [5]:

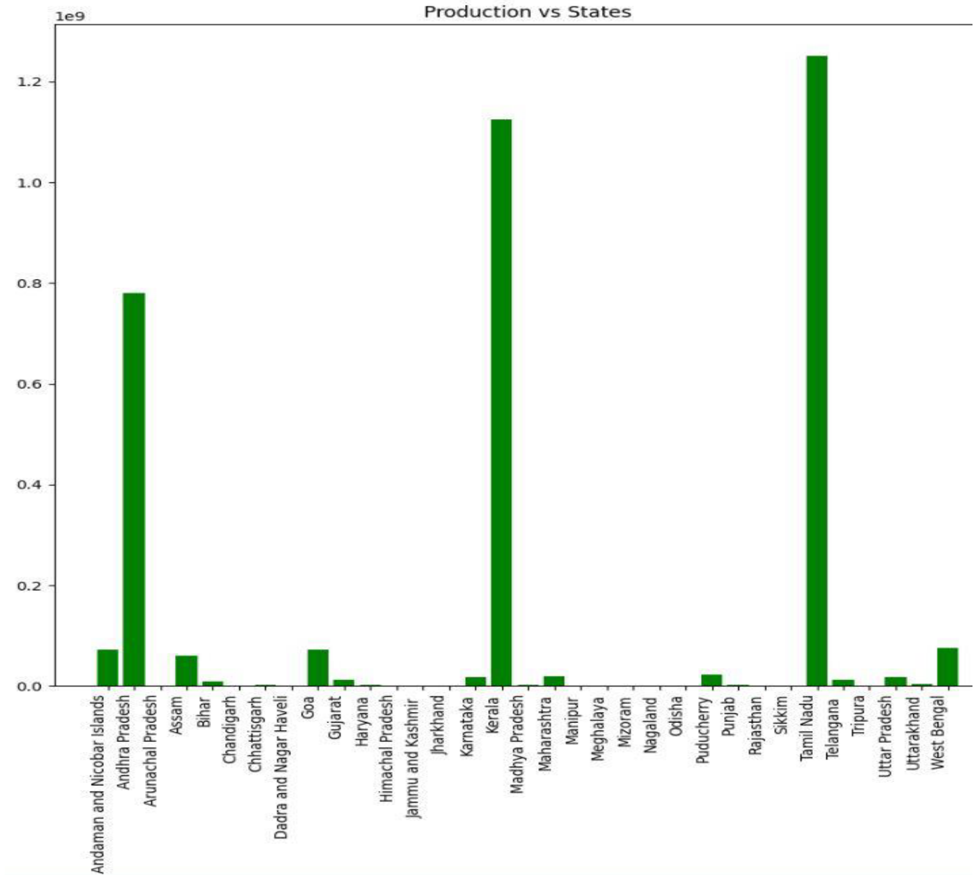
```
# statistical inference of the dataset

crop_data.describe()
```

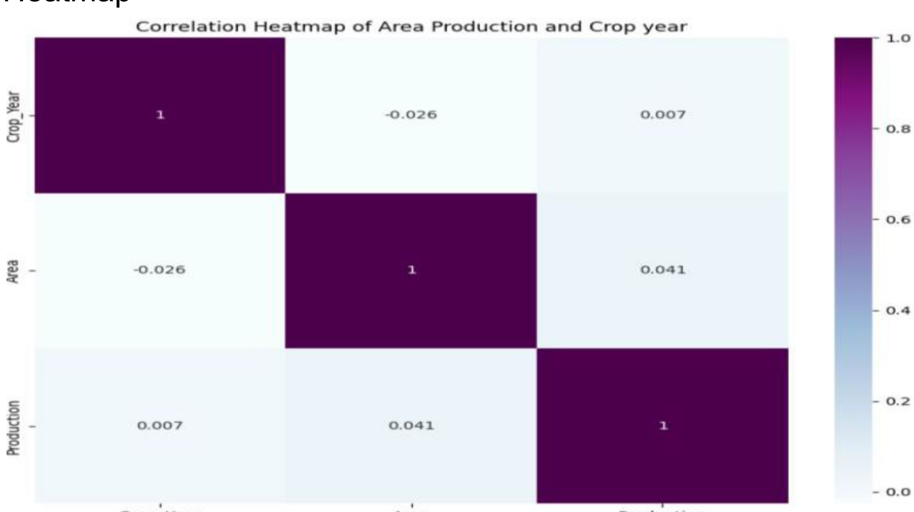
Out[5]:

	Crop_Year	Area	Production
count	246091.000000	2.460910e+05	2.423610e+05
mean	2005.643018	1.200282e+04	5.825034e+05
std	4.952164	5.052340e+04	1.706581e+07
min	1997.000000	4.000000e-02	0.000000e+00
25%	2002.000000	8.000000e+01	8.800000e+01
50%	2006.000000	5.820000e+02	7.290000e+02
75%	2010.000000	4.392000e+03	7.023000e+03
max	2015.000000	8.580100e+06	1.250800e+09

Crop Production by State



Heatmap



Decision Tree Algorithm

In [53]:
```
from sklearn.tree import DecisionTreeClassifier

DecisionTree = DecisionTreeClassifier(criterion="entropy",random_state=2,max_depth=5)

DecisionTree.fit(Xtrain,Ytrain)

predicted_values = DecisionTree.predict(Xtest)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Decision Tree')
print("DecisionTrees's Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))
```

```
DecisionTrees's Accuracy is:  90.0
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       0.59      1.00      0.74        16
    chickpea       1.00      1.00      1.00        21
     coconut       0.91      1.00      0.95        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.74      0.93      0.83        28
 kidneybeans       0.00      0.00      0.00        14
      lentil       0.68      1.00      0.81        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
   mothbeans       0.00      0.00      0.00        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      0.84      0.91        19
  pigeonpeas       0.62      1.00      0.77        18
 pomegranate       1.00      1.00      1.00        17
        rice       1.00      0.62      0.77        16
  watermelon       1.00      1.00      1.00        15

    accuracy                           0.90       440
   macro avg       0.84      0.88      0.85       440
weighted avg       0.86      0.90      0.87       440
```

Random Forest Algorithm

In [56]:
```python
from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)

print(classification_report(Ytest,predicted_values))
```

```
RF's Accuracy is:  0.990909090909091
              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       1.00      1.00      1.00        17
   blackgram       0.94      1.00      0.97        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      1.00      1.00        21
      coffee       1.00      1.00      1.00        22
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        18
        jute       0.90      1.00      0.95        28
 kidneybeans       1.00      1.00      1.00        14
      lentil       1.00      1.00      1.00        23
       maize       1.00      1.00      1.00        21
       mango       1.00      1.00      1.00        26
   mothbeans       1.00      0.95      0.97        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       1.00      1.00      1.00        19
  pigeonpeas       1.00      1.00      1.00        18
 pomegranate       1.00      1.00      1.00        17
        rice       1.00      0.81      0.90        16
  watermelon       1.00      1.00      1.00        15

    accuracy                           0.99       440
   macro avg       0.99      0.99      0.99       440
weighted avg       0.99      0.99      0.99       440
```

Prediction

```
In [60]: #making a prediction
         data = np.array([[104,18, 30, 23.603016, 60.3, 6.7, 140.91]])
         prediction = RF.predict(data)
         print(prediction)

['coffee']
```

Business Model

The monetization idea of the product would consist of the following:

- A free version of the app will provide basic features like just recommending the farmer the crop for their required land.For more advanced features subscription fee will be charged

- The app will be based on two monetization models: a **subscription model** as well as a **pay-per-use** basis(provide features based on need of the user)

- **Advertising:** For the free version of the app, advertisements will be displayed that will consist of agricultural products,manures,fertilizers,etc. which will be done in partnership with the brands that manufacture these products.Partnership can also be done using Google Ad-Sense.

- **Subscriptions:** A significant mode for income generation will be the subscription services that the product provides.This will include in-depth analysis of the agricultural land,recommendation of the manure,fertilizers that will increase the yield,monthly update checking on the farmers' land and personalized help for the farmer with a partner agricultural consultant.

FINANCIAL MODELING

The financial model for a crop recommendation system would typically include revenue projections, cost projections, and cash flow projections.
The revenue projections would be based on the subscription-based revenue model, consulting fees, and affiliate marketing and advertising revenue.
The cost projections would include the cost of developing and maintaining the system, marketing and advertising costs, and other operating expenses. To create a financial model for a crop recommendation system, we would need to

estimate the number of subscribers, consulting clients, and affiliate marketing and advertising revenue. This would require market research to determine the potential market size and demand for the system. The cost projections would be based on the cost of developing and maintaining the system, marketing and advertising costs, and other operating expenses. The financial model would also need to take into account any potential risks and uncertainties, such as changes in market demand, competition, and regulatory changes. Considering all of this, following financial equation can be used:

For advertising model:

Total Revenue = (Number of Subscribers x Subscription Fee) + (Number of Premium Consulting customers x Consulting Fee) + (Revenue from Advertising) + (Revenue from Affiliate Marketing)

Total Cost = (Cost of Website Development and Maintenance) + (Cost of Data Collection and Analysis) + (Cost of Expert Consultations) + (Marketing and Advertising Expenses) + (Cost of Affiliate Commissions)

Profit = Total Revenue - Total Cost 6/10 farmers in rural India, have access to a smartphone with an active internet connection which is 60%. India officially has anywhere from 90 million-plus to almost 150 million farmers. Taking an average of this figure, India roughly has 120 million farmers. So, roughly 72 million of them have an internet connection. The aim would be for around 10% of them to join the services, 7.2 million farmers would be using the app. If we aim for around 7% of them to go for consultations, 5.04 million farmers will be availing that facility. If we were to offer premium services with a subscription model of Rs.100 for 3 months, the revenue generated would be more than 1 billion.

Affiliate marketing and advertising account for around 15% of the total revenue. The one-time consultation fees could be as nominal as Rs. 15.

If total revenue is Rs x, x = (Number of Subscribers x Subscription Fee) + (Number of Premium Consulting customers x Consulting Fee) + 0.15*(x)
Supposing the market is growing linearly, we have,

$$y=mx(t)+c$$

where y = total profit, m=price of subscription, x(t)=total sales (market as a function of time) c=production, maintenance costs.