SmartCrop: Integrating Machine Learning for Precision Agriculture and Yield Prediction

Sayali Hatwar

Date : 15th May 2024

Abstract

Accurate crop prediction is paramount for effective agricultural planning and sustainable business practices. This study presents a comprehensive model for crop prediction, integrating soil elements (potassium(K), phosphorous(P), nitrogen(N)), environmental factors (temperature, pH, rainfall, humidity), crop price, and state-specific data. Utilizing advanced machine learning techniques, we aim to forecast crop yields accurately. By analysing historical datasets, our model provides actionable insights into crop cultivation, aiding stakeholders in strategic decision-making. This predictive framework offers valuable guidance for businesses involved in crop seed distribution, enabling informed resource allocation and enhancing agricultural sustainability.

1.0 Problem Statement

In today's agricultural landscape, small-scale farmers face significant challenges in maximizing crop yields while minimizing costs. With limited resources and access to proper guidance, these farmers struggle to optimize their farming practices effectively. Additionally, the global perception of agriculture often overlooks the plight of small farmers, further exacerbating their challenges.

To address this issue, there is a critical need to provide small-scale farmers with the necessary knowledge and tools to enhance crop yields efficiently. Understanding the specific soil requirements based on factors such as temperature, rainfall, and humidity, as well as geographical location (State), is paramount in achieving this goal. By leveraging this information, agricultural science centres (Krushi Kendra) and large companies can develop tailored fertilizer solutions that optimize crop growth while minimizing resource inputs with the crop prediction system.

This approach not only benefits small farmers by enabling them to achieve higher yields with minimal resources but also supports the broader agricultural ecosystem by promoting sustainable and efficient farming practices. Ultimately, empowering small-scale farmers with proper knowledge and access to the Crop Prediction System helps tailor fertilizer solutions and has the potential to revolutionize agriculture, saving money, time, effort, and manpower while ensuring food security and livelihood sustainability.

2.0 Market/Business Need Assessment

Given the distribution of farmers just in India, with approximately 48.16% classified as small farmers, 43.22% as medium farmers, and 8% as large farmers, there is a pressing need to address the financial constraints and knowledge gaps faced by small and medium-scale farmers. These farmers often struggle with limited resources and lack access to accurate information on crop selection and fertilizer usage.

A potential business opportunity lies in leveraging machine learning and AI technologies to develop predictive models that assist farmers in making informed decisions about crop cultivation and soil management. By analysing various variables that affect crop growth, such as soil composition, climate data, and historical yield patterns, machine learning algorithms, particularly ensemble techniques, can offer precise predictions with high accuracy.

The proposed business model involves the development of a user-friendly application that provides small and medium-scale farmers with access to these predictive models and expert guidance. The application would offer personalized recommendations on crop selection, fertilizer application, and optimal planting times, tailored to the specific needs and conditions of each farmer's land.

Furthermore, this solution could be implemented through a government initiative aimed at supporting agricultural development and improving farmer livelihoods. Alternatively, the business model could be franchised to entrepreneurs and agribusinesses interested in providing value-added services to farmers in their communities.

Key aspects of the business assessment include:

1. Market Opportunity: There is a significant market opportunity to cater to the needs of small and medium-scale farmers who require cost-effective solutions for crop management and soil fertility.

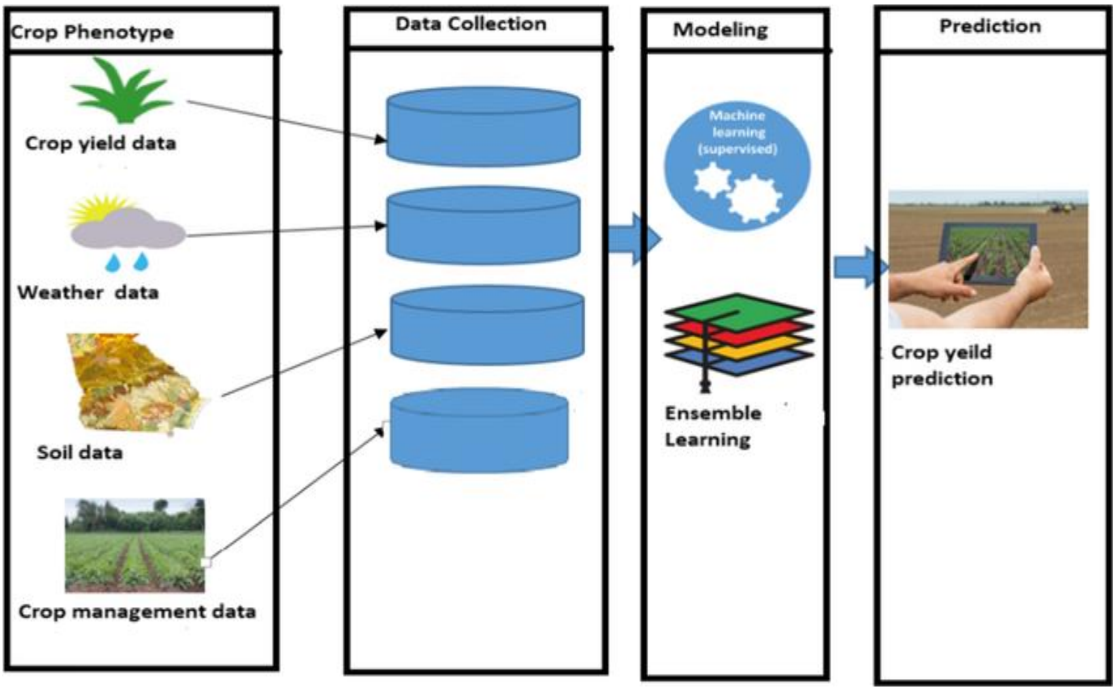
2. Technology Adoption: The adoption of machine learning and AI technologies enables precise predictions and recommendations, improving the efficiency and profitability of agricultural operations.

3. Accessibility and Affordability: The application-based approach ensures accessibility to farmers across diverse geographical locations, while government support or franchising arrangements can make the solution affordable and sustainable.

4. Scalability and Impact: The scalability of the business model allows for widespread adoption and impact, potentially transforming agricultural practices and enhancing farmer incomes nationwide.

Overall, the proposed business model addresses critical challenges faced by small and medium-scale farmers all over the world, offering a viable solution to improve agricultural productivity, profitability, and sustainability.

1.1 Architecture diagram of Crop yield prediction



3.0 Target Specification and Characterization

1. **Small-scale and Medium-scale Farmers with low-budget**
2. **Poor Climatic Conditions**
3. **Low Water resources**
4. **Illiteracy among people**

To overcome all these challenges and make large revenue from small farms with minimum cost, we need the help of predictions and analysis.

A. **Predictive Analysis:** Crop Prediction System can analyze historical data on soil composition, climate patterns, crop yields, and agricultural practices to generate predictive insights. These insights can help farmers make informed decisions about crop selection, planting schedules, and fertilizer usage, leading to improved yields and profitability.

B. **Decision Support Systems:** AI-driven decision support systems provide farmers with interactive tools and dashboards that visualize data, forecast trends, and simulate scenarios. These systems empower farmers to experiment with different agronomic practices virtually, assess potential outcomes, and identify the most cost-effective strategies for their farm operations.

C. **Customized Recommendations:** Machine Learning algorithms can generate customized recommendations for farmers based on their specific location, soil type, crop preferences, and resource constraints. These recommendations may include optimal crop varieties, fertilizer formulations, and planting strategies tailored to local conditions, helping farmers achieve better results with fewer inputs.

4.0 External Search

The dataset utilized in the project is the **Crop Prediction Dataset**, which is available on Kaggle. (link given below)

mailto:https://www.kaggle.com/datasets/jiteshmd/crop-prediction-data

This dataset was built by augmenting datasets of rainfall, climate and Soil Elements data available for India. There are over 2200 data rows for 10 columns containing soil elements (Potassium(K), Phosphorus (P), Nitrogen(N)), soil pH, climatic conditions like humidity, temperature, rainfall, state location, recommended crop and crop prices.

```
df = pd.read_csv('/kaggle/input/crop-prediction-data/indiancrop_dataset.csv')
df.head()
```

[8]:

	N_SOIL	P_SOIL	K_SOIL	TEMPERATURE	HUMIDITY	ph	RAINFALL	STATE	CROP_PRICE	CROP
0	90	42	43	20.879744	82.002744	6.502985	202.935536	Andaman and Nicobar	7000	Rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	Andaman and Nicobar	5000	Rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	Andaman and Nicobar	7000	Rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	Andaman and Nicobar	7000	Rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	Andaman and Nicobar	120000	Rice

+ Code + Markdown

[4]:

```
df.shape
```

[4]: (2200, 10)

Information about our variables considered are given below,

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   N_SOIL       2200 non-null   int64  
 1   P_SOIL       2200 non-null   int64  
 2   K_SOIL       2200 non-null   int64  
 3   TEMPERATURE  2200 non-null   float64
 4   HUMIDITY     2200 non-null   float64
 5   ph           2200 non-null   float64
 6   RAINFALL     2200 non-null   float64
 7   STATE        2200 non-null   object 
 8   CROP_PRICE   2200 non-null   int64  
 9   CROP         2200 non-null   object 
dtypes: float64(4), int64(4), object(2)
memory usage: 172.0+ KB
```

+ Code + Markdown

A Case Study that I considered for this model is

1. Crop Prediction based on soil and environmental characteristics using a feature selection technique [Web Link]

2. Crop Prediction based on Soil classification using Machine Learning with classifier ensembling.[Web Link]

source:

- Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India;
- Department of Computer Science and Engineering, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India [Web Link]
- t, M. E. S. College of Engineering Pune, Maharashtra, India 5Professor, Dept. of Computer Engineering, M. E. S. College of Engineering Pune, Maharashtra, India [Web Link]

5.0 Benchmark

Benchmarking agricultural crop prediction methods with and without artificial intelligence (AI) involves comparing their accuracy, efficiency, and scalability in forecasting crop yields and optimizing agricultural practices. When utilizing machine learning and AI techniques, such as predictive modelling, neural networks, and data analytics, crop prediction can benefit from advanced algorithms that analyze large datasets, identify complex patterns, and generate accurate predictions based on diverse variables such as soil composition, climate conditions, and historical yield data. In contrast, conventional methods of crop prediction may rely on simpler statistical models or expert knowledge, which could be less accurate and time-consuming. By benchmarking these approaches, researchers and practitioners can assess the performance and effectiveness of AI-driven crop prediction models in improving agricultural productivity, resource management, and decision-making, thereby informing the adoption of innovative technologies for sustainable farming practices.

6.0 Applicable Regulations by Governments

In India, agriculture is governed by a range of regulations, acts, and policies aimed at ensuring the sustainable development of the sector and the welfare of farmers. These include the Agricultural Produce Market Committee (APMC) Act and Model APMC Rules, which regulate agricultural marketing to safeguard farmers' interests; the Fertilizer Control Order (FCO) and Soil Health Card Scheme, which oversee integrated nutrient management for balanced fertilization; and the National Food Security Mission (NFSM) and Horticulture Mission, which promote crop diversification and horticultural development for food security and income generation.[web]

Additionally, laws such as the Official Languages Act and Seeds Act provide frameworks for language use in government communications and quality control in seed production, respectively. Through these regulations, the government aims to foster cooperation among farmers, encourage the adoption of mechanization and modern technologies, ensure sustainable natural resource management, and facilitate farmer welfare and agricultural extension services, thus advancing the nation's agricultural goals and objectives.

7.0 Applicable Constraint

- Continuous Data Collection from farmers and vendors
- Maintenance of Large amounts of data
- Lack of technical knowledge and illiteracy among farmers
- Convincing Farmers to use and implement new technology at their farms.

8.0 Business Model

1. **Precision Agricultural Services:** The Crop prediction System can provide precise recommendations to farmers on crop selection, planting schedules, and fertilizer application based on real-time data. Offering precision agriculture services as a business can help farmers optimize their yields while minimizing input costs, thereby increasing their profitability.
2. **Customized Fertilizer Formulations:** By analyzing soil composition and environmental conditions, the system can generate recommendations for customized fertilizer formulations tailored to the specific needs of each farm. Partnering with fertilizer companies to develop and market this specialized product can create a lucrative niche in the market.
3. **Data Analytics and Insights:** Collect and analyze data generated by our crop prediction model to identify trends, patterns, and opportunities in the agricultural market. Offer data and analytics services to stakeholders such as seed companies, commodity traders, and agricultural insurers to help them make informed decisions and mitigate risks.

Farmers in different countries often come from poor or middle-class backgrounds and typically own small plots of land with limited budgets and resources. By utilizing machine learning technology for crop yield prediction, provided by business people, they can effectively address these challenges. This approach can help them save money and increase their revenue.

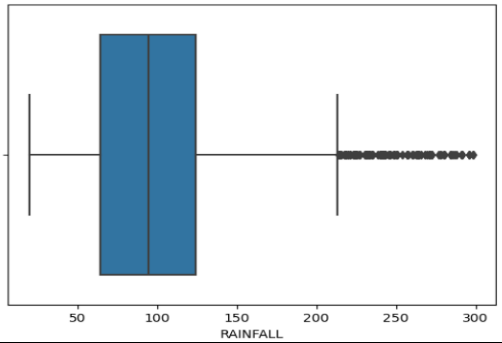
9.0 Concept Generation

Concept generation for crop prediction systems involves integrating diverse data sources, such as soil composition, climate patterns, and historical yield data into a comprehensive machine-learning framework. The concept revolves around developing predictive models that leverage advanced algorithms, including ensemble methods, to identify complex relations and patterns within the data. The customer might want to spend the least amount of time giving input data. This accuracy will take a little effort to nail because it’s imprudent to rely purely on the Classic Machine Learning algorithm.

a) Clean up the outliers in the data

```
[11]: import seaborn as sns
      sns.boxplot(x = df["RAINFALL"])
```

[11]: <Axes: xlabel='RAINFALL'>



localhost:8888/notebooks/Downloads/crop-prediction.ipynb

pynative python exerc... Python Tutorial Dashboard | ExcelR Tableau Desktop Solve Python | Hacker... 1000 Python MCQ (M... SQL CREATE TABLE, A... All Bookmarks

File Edit View Run Kernel Settings Help Trusted

Markdown Open in... Python 3 (ipykernel)

```
[12]: Q3=df["RAINFALL"].quantile(0.75)
      Q1=df["RAINFALL"].quantile(0.25)
      IQR = Q3-Q1
      print("IQR -",IQR)
      Lower_bound= Q1-1.5*IQR
      print("LowerBound -",Lower_bound)
      Upper_bound= Q1+1.5*IQR
      print("UpperBound -",Upper_bound)

      IQR - 59.715821800000015
      LowerBound - -25.022046700000033
      UpperBound - 154.1254187
```

```
[13]: df_c1= df[(df['RAINFALL'] >= Lower_bound)&(df['RAINFALL'] <= Upper_bound)]
      df_c1.head()
```

[13]:

	N_SOIL	P_SOIL	K_SOIL	TEMPERATURE	HUMIDITY	ph	RAINFALL	STATE	CROP_PRICE	CROP
100	71	54	16	22.613600	63.690706	5.749914	87.759539	Gujarat	2600	Maize
101	61	44	17	26.100184	71.574769	6.931757	102.266244	Gujarat	825	Maize
102	80	43	16	23.558821	71.593514	6.657965	66.719955	Gujarat	1100	Maize
103	73	58	21	19.972160	57.682729	6.596061	60.651715	Gujarat	825	Maize
104	61	38	20	18.478913	62.695039	5.970458	65.438354	Gujarat	3425	Maize

b) Encoding the data to convert categorical values to numerical values

c) Splitting Data into input and output

```
[16]: X = df_c1_encoded.drop(['CROP'],axis=1)
      X.shape

[16]: (1799, 29)

[17]: y = df_c1_encoded['CROP']
      y.shape

[17]: (1799,)
```

d) Split data into train test data

```
[18]: from sklearn.model_selection import train_test_split
      X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.45,random_state=30)

[19]: print('Shape of X_train',X_train.shape)
      print('Shape of X_test',X_test.shape)
      print('Shape of y_train',y_train.shape)
      print('Shape of y_test',y_test.shape)

      Shape of X_train (989, 29)
      Shape of X_test (810, 29)
      Shape of y_train (989,)
      Shape of y_test (810,)
```

Now, we will apply different ensembling models to achieve the highest accuracy scoring model.

I. Decision Forest Classifier

```
[21]: from sklearn.tree import DecisionTreeClassifier
      classifier1 = DecisionTreeClassifier(criterion='gini',max_depth=5)

[22]: classifier1.fit(X_train,y_train)
      classifier1.score(X_test,y_test)

[22]: 0.4283950617283951
```

II. Random Forest Classifier

```
[23]: from sklearn.ensemble import RandomForestClassifier
      classifier2 = RandomForestClassifier(n_estimators=500,criterion='gini')

[24]: classifier2.fit(X_train,y_train)
      classifier2.score(X_test,y_test)

[24]: 0.9962962962962963
```

III. KNeighborsClassifier

```
[25]: from sklearn.neighbors import KNeighborsClassifier
      classifier3 = KNeighborsClassifier()

[26]: classifier3.fit(X_train,y_train)
      classifier3.score(X_test,y_test)

[26]: 0.43950617283950616
```

IV. Naïve Bayes

```
GaussianNB

[29]: from sklearn.naive_bayes import GaussianNB
      GNB = GaussianNB()

[30]: GNB.fit(X_train,y_train)
      GNB.score(X_test,y_test)

[30]: 0.9938271604938271

MultinomialNB

[31]: from sklearn.naive_bayes import MultinomialNB
      MNB = MultinomialNB()

[32]: MNB.fit(X_train,y_train)
      MNB.score(X_test,y_test)

[32]: 0.38765432098765434

BernoulliNB

[33]: from sklearn.naive_bayes import BernoulliNB
      BNB = BernoulliNB()

[34]: BNB.fit(X_train,y_train)
      BNB.score(X_test,y_test)

[34]: 0.5481481481481482
```

V. Support Vector Machine Classifier

Kerner = 'rbf'

```
[37]: from sklearn.svm import SVC
      classifier_rbf = SVC(kernel='rbf')

[38]: classifier_rbf.fit(X_train,y_train)

[38]: ▾ SVC
      SVC()

[39]: classifier_rbf.score(X_test,y_test)

[39]: 0.09012345679012346
```

Kernel = 'linear'

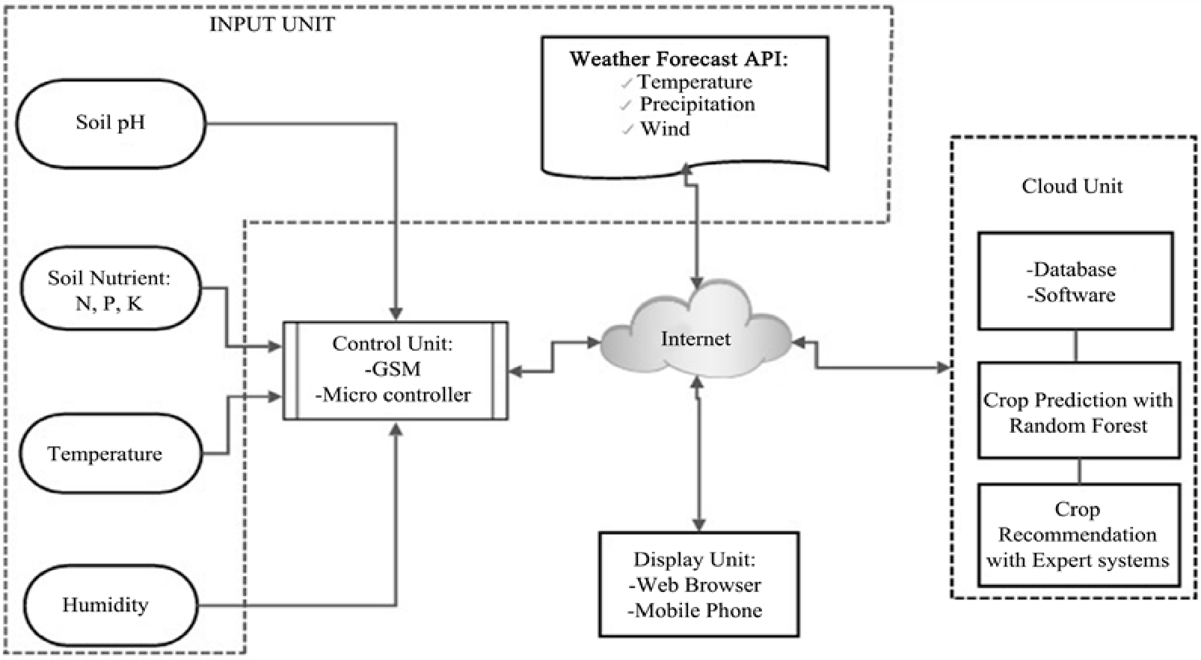
```
[40]: from sklearn.svm import SVC
      classifier_lin = SVC(kernel='linear')

[41]: classifier_lin.fit(X_train,y_train)

[41]: ▾ SVC
      SVC(kernel='linear')

[42]: classifier_lin.score(X_test,y_test)

[42]: 0.9950617283950617
```

By Analyzing all models, we concluded that **the Random Forest Classifier** gives the highest accuracy score of **0.9962.**

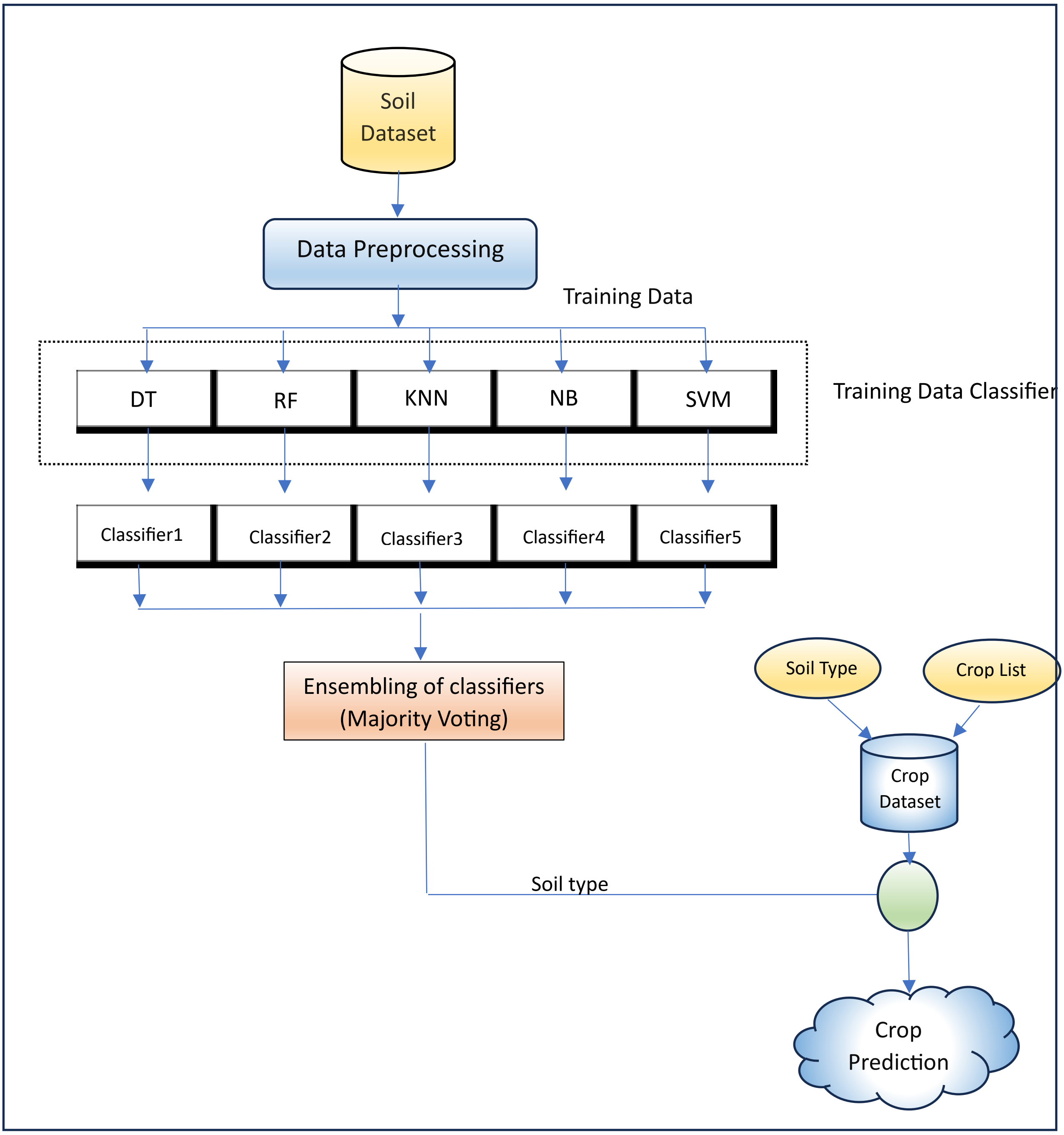
10.0 Concept Development

The concept involves designing user-friendly interfaces and mobile applications that deliver personalized insights and recommendations to farmers, extension workers, and agribusinesses, facilitating adoption and scalability. Through continuous refinement and validation, the concept seeks to revolutionize crop prediction, empower stakeholders with actionable intelligence, and drive innovation in sustainable agriculture.

11.0 Final Product Prototype



<u>**Schematic Diagram showing steps from collection of data to prediction of crop**</u>



<u>Application of various ensembling methods on a dataset for better prediction</u>

This schematic diagram outlines the flow of data from various sources through the prototype, including data collection, processing, and analysis using machine learning techniques, and finally, presentation to users through a user interface.

12.0 Conclusion

Agriculture involves a lot of risk, as farmers often encounter challenges such as climate changes, unseasonal rainfall, soil quality, and access to quality fertilizers. These factors increase the risk of achieving good yields and high-quality products. One way to mitigate these challenges is through the use of technology, such as a Crop Prediction System. This system utilizes machine learning models to provide reliable information that can assist both small and large businesses in supporting farmers in various aspects, including seeding and fertilization. By predicting the best crop to be grown in a specific area based on factors such as soil structure, elements, and climate, this technology can be a valuable resource for farmers.