Fertilizer Recommendation System

Rishav Karmahapatra

18 May, 2023

Contents

1 Introduction

Agriculture plays a crucial role in India, serving as the main occupation for a majority of the population and covering a significant portion of the nation's land. With the responsibility of catering to the basic needs of 1.2 billion people, the agricultural sector continually seeks ways to modernize its procedures. The success of fertilizer prediction is heavily influenced by various factors, including weather conditions, environmental changes, uncertain rainfall patterns, water management, and the utilization of pesticides. However, farmers often struggle to achieve the expected crop yield due to these challenges.

To address these issues and benefit farmers, researchers have turned to data mining, machine learning, and deep learning approaches. By leveraging these technologies, they aim to enhance and improve crop yield and quality. Machine learning, in particular, enables machines to learn without explicitly programmed instructions, allowing them to improve their performance by identifying and characterizing patterns and trends in input data.

In this project, various machine learning approaches such as Decision Tree classifier,KNN are employed for fertilizer prediction. These approaches are applied to crop yield datasets from different states and encompassing various crops. The goal is to develop a system that can recommend the most suitable fertilizer for a particular land based on weather parameters and soil content, including rainfall, temperature, humidity, and pH.

2 Problem Statement

In the agricultural industry, one of the key challenges faced by farmers and agronomists is determining the optimal type and quantity of fertilizer to apply for different crops and soil conditions. Applying too little fertilizer can result in nutrient deficiencies and lower yields, while excessive fertilizer application can lead to environmental pollution and economic losses. Therefore, there is a need for an accurate and reliable system that can predict the appropriate type and amount of fertilizer required for specific crops and soil characteristics.

Traditional methods of fertilizer recommendation often rely on manual analysis of soil samples, which can be time-consuming and may not provide real-time information. Additionally, these methods may not consider various factors such as crop types, weather patterns, and market trends, which can impact the nutrient requirements of plants. Therefore, there is a demand for a predictive solution that leverages advanced technologies such as artificial intelligence and machine learning to provide precise and data-driven fertilizer recommendations.

3 Customer and Business Need Assessment

To develop an effective fertilizer prediction solution, it is essential to conduct a customer and business need assessment to understand the specific requirements and challenges faced by stakeholders in the agricultural industry. Here are some key aspects to consider during the assessment:

3.1 Customer Needs

- Conduct surveys, and focus groups with farmers, agronomists, and agricultural companies to gather insights on their fertilizer prediction needs.
- Determine the specific pain points and challenges faced by customers, such as difficulties in determining optimal fertilizer type and dosage, time-consuming manual analysis, and lack of real-time recommendations.
- Identify customer expectations regarding accuracy, reliability, ease of use, and integration with existing farm management systems.

3.2 Business Needs

- Assess the potential benefits and value proposition for businesses, including fertilizer manufacturers, distributors, and agricultural service providers.
- Understand the impact of accurate fertilizer prediction on key business factors, such as improved customer satisfaction, increased sales, reduced environmental impact, and enhanced brand reputation.
- the cost implications, including development, implementation, and maintenance costs, to determine the viability and return on investment of the fertilizer prediction solution.

4 Benchmarking alternate products

When benchmarking a fertilizer recommendation system against existing products and services, it is essential to consider several factors that contribute to their effectiveness and value for farmers. Some of the existing products and services in the field of fertilizer recommendation include:

- **Traditional Knowledge and Expertise:** Farmers have traditionally relied on their experience, local knowledge, and advice from agricultural experts to determine fertilizer application rates. This approach is based on years of observation and may vary regionally.
- **Soil Testing Services:** Soil testing services provide farmers with detailed analysis of their soil composition, nutrient levels, pH, and other relevant factors. Based on the results, they provide fertilizer recommendations tailored to specific crops and soil conditions. These services may involve collecting soil samples and sending them to a laboratory for analysis.

- **Commercial Fertilizer Products:** Commercial fertilizer manufacturers produce a wide range of fertilizers with different nutrient compositions. These products are formulated based on general crop nutrient requirements and may include instructions for application rates and timing.

  When benchmarking a fertilizer recommendation system, it is crucial to evaluate how the system compares to these existing products and services in terms of the following factors:

- **Accuracy:** The system should aim to provide accurate and reliable fertilizer recommendations that align with the specific needs of crops and soil conditions. It should be benchmarked against the accuracy of soil testing services and traditional knowledge.

- **Ease of Use:** The system should be user-friendly and accessible to farmers, considering factors such as user interface, data input requirements, and compatibility with different devices and platforms. It should be compared to the ease of use of traditional knowledge and advisory services.

- **Cost-Effectiveness:** The cost of implementing the fertilizer recommendation system, including any required technology, data collection, and maintenance, should be compared to the cost of existing products and services. It should be evaluated in terms of the potential return on investment for farmers.

- **Speed and Timeliness:** The system's ability to provide real-time or timely recommendations can be compared to the turnaround time of soil testing services and the availability of expert advice through agricultural extension services.

5 Applicable Patents

The potential patents that could be applicable to our fertilizer recommendation system, depending on the technology, software, or framework used are:

1. Patents related to Data Analytics and Machine Learning in Agriculture: These patents could cover methods, algorithms, or systems for analyzing agricultural data, applying machine learning techniques, and generating insights or recommendations for optimizing fertilizer application.

2. Patents related to Data Integration and API Technologies: If the fertilizer recommendation system integrates with external data sources, such as weather data providers, soil databases, or farm management systems, there may be patents protecting the integration methods, data processing techniques, or API implementations.

3. Patents related to Precision Agriculture Technologies: Precision agriculture involves the use of advanced technologies for site-specific crop management. Patents in this area could cover innovations in data collection

devices, sensors, drones, satellite imagery, or other hardware components used for gathering agricultural data.

4. Patents related to User Interfaces and Visualization in Agriculture: Innovations in user interface design, data visualization techniques, or user experience improvements specific to agricultural applications may be protected by relevant patents.

6 Applicable Regulations

1. **Agricultural Regulations:** Agricultural practices and products are often regulated by specific agricultural authorities or government agencies. These regulations may cover aspects such as fertilizer usage, nutrient management, environmental impact, and labeling requirements for fertilizers. Compliance with these regulations ensures that the fertilizer recommendation system aligns with agricultural standards and guidelines.

2. **Data Protection and Privacy Regulations:** The collection, storage, and processing of agricultural data, including sensitive information about farmers' practices and land, may be subject to data protection and privacy regulations.

3. **Environmental Regulations:** The use of fertilizers can have environmental impacts, such as water pollution or soil degradation. Compliance with environmental regulations ensures that the fertilizer recommendation system takes into account sustainable practices, minimizing adverse effects on the environment and promoting responsible fertilizer use.

4. **Health and Safety Regulations:** Depending on the jurisdiction, there may be health and safety regulations related to the production, handling, and application of fertilizers. Compliance with these regulations ensures that the recommended fertilizers are safe for human health, animal welfare, and the environment.

7 Applicable Constraints

1. **Data Availability and Quality:** Developing an accurate fertilizer prediction system requires access to comprehensive and high-quality data related to soil conditions, crop characteristics, weather patterns, and historical yield data. Obtaining such data can be challenging, especially in regions with limited data collection infrastructure or data privacy restrictions.

2. **Expertise in Agriculture and Agronomy:** Building a robust fertilizer prediction system requires expertise in agriculture, soil science, and

agronomy. Understanding the complex relationships between soil nutrients, crop requirements, and environmental factors is crucial for accurate predictions. Acquiring this expertise and collaborating with domain experts can be essential but may also present challenges.

3. **Computational Resources:** Fertilizer prediction systems may require significant computational resources, especially when handling large datasets or employing complex machine learning algorithms. Ensuring access to sufficient computing power, storage, and processing capabilities can be a constraint, particularly for resource-constrained environments.

4. **Cost:** Developing and deploying a fertilizer prediction system can involve significant costs, including data acquisition, software development, infrastructure, and ongoing maintenance. Budget constraints may limit the resources available for research, development of the system.

8 Business Model

To monetize the fertilizer prediction solution, I consider the following business model:

8.1 Pay-per-Use Model:

Implement a pay-per-use model where users are charged based on the number of predictions or recommendations they generate using the fertilizer prediction system. This model allows customers to pay for the exact amount of service they require and provides flexibility in terms of usage.

8.2 Partnerships Business Model:

Collaborate with fertilizer manufacturers, distributors, or farm management software providers to integrate the fertilizer prediction solution into their existing platforms. Through partnerships, you can generate revenue through licensing fees, revenue sharing, or by leveraging their customer base for cross-promotion.

9 Concept Generation

Concept generation is an iterative process. Continuously gather feedback, iterate on the prototypes, and refine the concept based on user needs, market dynamics, and technological advancements. This iterative approach will help you develop a robust and commercially viable fertilizer prediction system.

9.1 Problem Clarification:

To clarify the problem, we will utilize the Power Flow Model for Design Concepts. The model helps us understand the flow of information, energy, and materials within the fertilizer prediction system. By analyzing the inputs, transformations, and outputs, we can identify the key functions and subsystems involved.

9.2 Generation Process

The concept generation process for the fertilizer prediction system involved the following steps:

- **Brainstorming:**We conducted brainstorming sessions involving a diverse group of stakeholders, including farmers, agronomists, and data scientists. We encouraged open thinking and generated a large number of ideas related to data collection, analysis techniques, and personalized recommendation generation.

- **C-Sketch:**We used the C-Sketch technique to visually represent different concept ideas. This involved sketching rough designs and diagrams that depicted the overall system architecture, user interactions, and data flow. These sketches served as a starting point for further exploration and refinement.

- **Morphological Chart:**We created a morphological chart to organize subsystem concepts for each function. The chart included functions such as data collection, data analysis, recommendation generation, and user interface. For each function, we generated multiple alternative concepts, exploring different approaches, algorithms, and technologies.

- **Delighters:** As part of the concept generation process, we considered features that could serve as "delighters." These are unique or unexpected features that could distinguish our fertilizer prediction system from existing solutions. Examples include integration with farm management systems, real-time weather updates, intuitive visualizations, and personalized agronomic advice.

9.3 Initial Screening for Feasibility and Effectiveness:

To screen the generated concepts for feasibility and effectiveness, we employed the following method:

- **Evaluation Criteria:** We defined evaluation criteria based on the specifications and requirements of the fertilizer prediction system. These criteria included accuracy of predictions, usability, scalability, data availability, computational requirements, and integration potential.

- **Feasibility Assessment:** We conducted a preliminary feasibility assessment of the feasible alternatives. This involved analyzing the required technologies, data sources, computational resources, and potential implementation challenges. We considered factors such as data collection methods, algorithm complexity, computational power, and scalability.

10 Concept Devlopement

This is an automatic fertilizer recommendation system based on the K-Nearest Neighbors (KNN) and Decision Tree algorithms. It helps farmers make informed decisions about the types and quantities of fertilizers to apply to their crops.

- **Data Collection:** The system collects data on various factors such as crop type, soil composition, historical yield data, weather conditions, and nutrient requirements for different crops. This data serves as the input for the recommendation system.

- **Data Preprocessing:** The collected data undergoes preprocessing to clean and normalize it. Missing values are handled, outliers are identified and treated, and data is transformed if necessary. This step ensures the data is in a suitable format for analysis.

- **Feature Extraction:** Relevant features are extracted from the preprocessed data. These features may include soil nutrient levels, pH, organic matter content, rainfall patterns, temperature, and crop-specific requirements. The feature extraction process aims to capture the essential information needed for accurate fertilizer recommendations.

- **Training Data Preparation:** The preprocessed data is divided into training and testing datasets. The training data consists of labeled samples, where the label represents the optimal fertilizer recommendation based on historical data or expert knowledge.

- **Evaluation and Validation:** The trained models are evaluated using suitable metrics such as accuracy, precision, recall, and F1-score. The performance of the models is assessed on a separate testing dataset to measure their effectiveness in fertilizer recommendation.

- **Fertilizer Recommendation:** The trained and validated models are used to predict the optimal fertilizer recommendation for a given set of input features. The system provides personalized recommendations, considering the specific crop, soil conditions, and environmental factors.

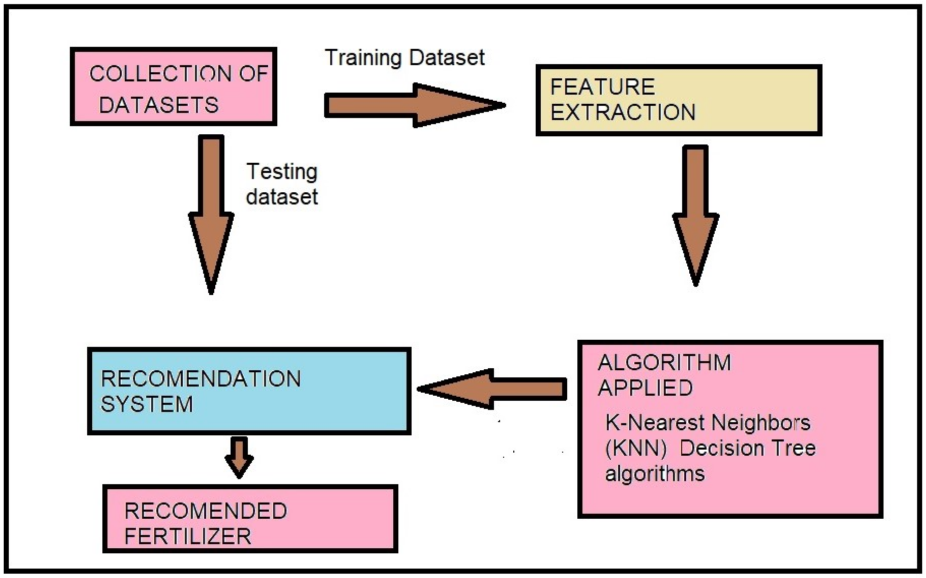

Figure 1: Block Diagram of Methodology of Proposed System

11 Product Prototype Description:

The fertilizer prediction product prototype consists of a cloud-based platform with an intuitive user interface accessible through a web application. The platform integrates advanced AI algorithms for fertilizer prediction and recommendation, leveraging machine learning techniques and data analysis.

- Users create an account on the platform, providing their basic information and payment details. Upon registration, users have access to the basic features of the platform but are required to purchase credits for fertilizer predictions.
- The pay-per-use model allows users to select the desired number of predictions or acreage coverage and pay accordingly.
- Users give input relevant data, such as soil nutrient levels, crop type, geographical location, and other parameters.
- The platform integrates with external data sources, such as weather data and historical crop performance, to enhance the prediction accuracy.
- The AI algorithms process the data and generate personalized fertilizer recommendations, including the type of fertilizer, optimal dosage, and application schedule.

- The platform presents the recommendations to the user, providing explanations and insights based on the underlying data analysis.

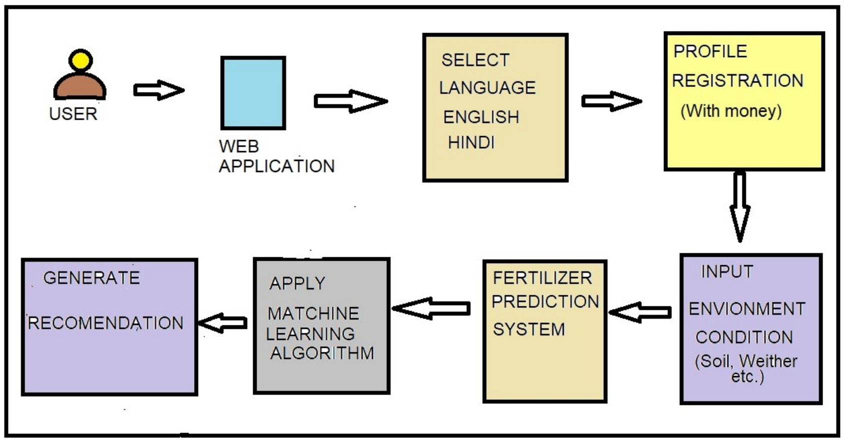

Figure 2: Schematic Diagram Of Product Prototype

12 Product Details

12.1 How Does It Works?

The fertilizer recommendation web application works by utilizing various data sources, algorithms, and user inputs to generate personalized fertilizer recommendations for farmers. Here is an overview of how the product works:

- **Model Training:** The machine learning models are trained using historical data and known fertilizer application practices to learn patterns and relationships between input variables and optimal fertilizer choices.
- **Recommendation Generation:** Based on the trained models and the user's inputs, the application generates personalized fertilizer recommendations, including the type of fertilizer, nutrient composition, application rate, and timing
- **User Interface:** The web application presents the fertilizer recommendations to farmers through a user-friendly interface. It may display the recommendations as textual information, charts, or graphs for easy comprehension.

- **User Feedback:** Farmers can provide feedback on the recommendations or update their input data, such as changes in soil test results or crop type.
- **Continuous Learning:** The application may utilize user feedback and updated data to continuously improve the recommendation models over time, enhancing the accuracy and relevance of the fertilizer recommendations.

12.2 Data Source

The dataset has been taken from Kaggle. This dataset has 100 rows and 9 attribute . The attributes are Temparature,Humidity,Moisture,Soil Type,Crop Type,Nitrogen,Potassium,Phosphorous,Fertilizer Name.

	Temparature	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fertilizer Name
0	26	52	38	Sandy	Maize	37	0	0	Urea
1	29	52	45	Loamy	Sugarcane	12	0	36	DAP
2	34	65	62	Black	Cotton	7	9	30	14-35-14
3	32	62	34	Red	Tobacco	22	0	20	28-28
4	28	54	46	Clayey	Paddy	35	0	0	Urea

Figure 3: Dataset

12.3 Algorithms, framework, software, etc. needed:

1. **KNN:**KNN stands for K-Nearest Neighbors which is one of the simplest supervised machine learning algorithms, it is a non parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learning algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

   In KNN, the choice of K, the number of nearest neighbors, is a crucial parameter that determines the decision boundary and affects the model's performance. A smaller value of K may lead to more flexible decision boundaries, but it can also make the model more sensitive to noise in the data. On the other hand, a larger value of K may lead to smoother decision boundaries but can potentially overlook local patterns.

   Additionally, while the Euclidean distance is commonly used in KNN, other distance metrics such as Manhattan distance or Minkowski distance can be utilized based on the specific characteristics of the data.

Overall, KNN is a simple yet effective algorithm that can be applied to various classification tasks, including fertilizer recommendation, by considering the nearest neighbors' information and making predictions based on the majority class.

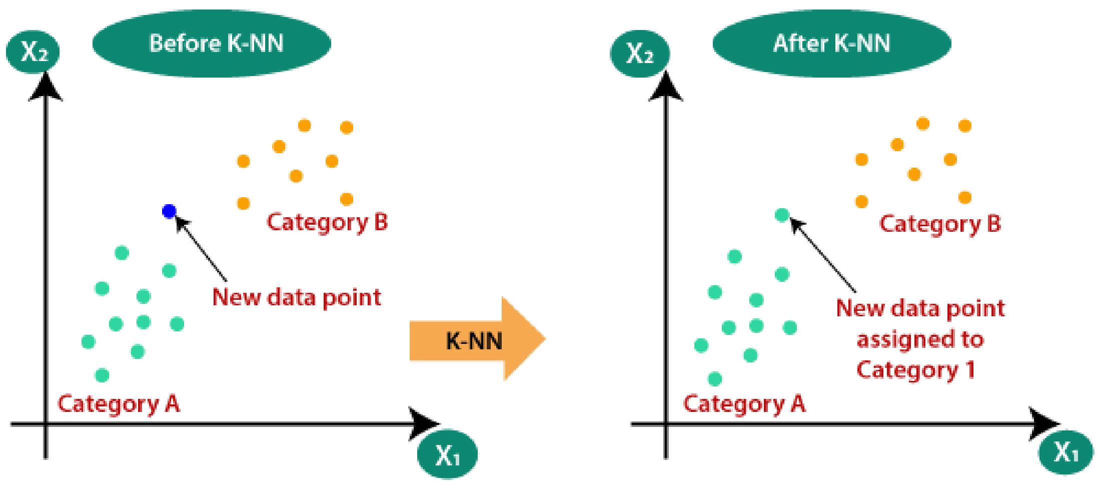

Figure 4: KNN Algorithm

2. **Decision Tree Classifier:**

- Decision Tree Classifier is a supervised machine learning algorithm that is used for both classification and regression tasks. It is a non-parametric algorithm that builds a tree-like model of decisions based on features of the dataset. The tree consists of internal nodes representing features, branches representing decision rules, and leaf nodes representing the class labels or regression values.Here is a general overview of how the Decision Tree Classifier works:

- Data Preparation: Start by collecting and preparing the dataset that includes features (such as soil conditions, weather parameters, nutrient levels) and corresponding class labels (fertilizer recommendations). Ensure that the dataset is labeled appropriately for supervised learning.

- Feature Selection: Identify the relevant features that are most informative for fertilizer recommendation. This may involve analyzing the importance of features using techniques like information gain, Gini index, or entropy.

- Model Training: Build the decision tree by recursively partitioning the data based on the selected features. The goal is to split the data in a way that maximizes the separation between different classes or reduces the impurity within each partition. This process continues until a stopping criterion is met, such as reaching a maximum tree depth or minimum number of samples per leaf.

- Prediction: Given a new set of input features, the decision tree traverses down the tree based on the feature values until it reaches a leaf node. The class label associated with that leaf node is then assigned as the predicted fertilizer recommendation.

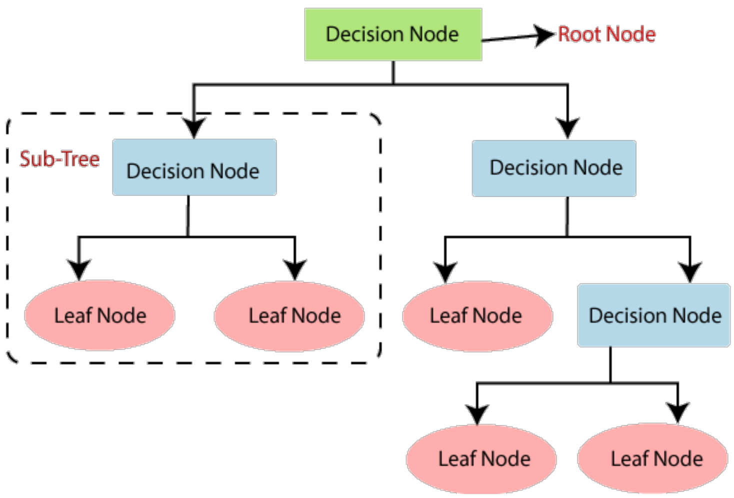

Figure 5: Decision Tree Classifier Algorithm

12.4 Flask:

Flask is a lightweight web application framework for Python that provides a simple and flexible way to build web applications.It gives us a variety of choices for developing web applications and it gives us the necessary tools and libraries that allow us to build a web application.

12.5 Python Libraries:

1. **NumPy:** NumPy stands for Numerical Python, it is an open-source library for Python programming. It adds support for large, multi-dimensional matrices and arrays, along with a gigantic collection of top-end mathematical functions to operate on these arrays and matrices. It is so much faster than using in- built Python functions.

   Syntax: import numpy as np

2. **Pandas:** It is an open-source library that gives you a highly useful set of tools to do data analysis. It can help you load, prepare, join, reshape,

analyze, process and adjust data. High-performance merging and joining of data sets, data alignment, and integrated handling of missing data are some of the fancy things of Pandas. Syntax: import pandas as pd

3. **Matplotlib:** It is a cross-platform, data visualization and graphical plotting library (histograms, scatter plots, bar plots, etc.) for Python and its numerical extension NumPy which helps us to understand trends, and patterns, and to make correlations. Syntax:

```
import matplotlib.pyplot as plt
```

4. **Scikit learn:** Scikit learn also known as sklearn, is an open-source data analysis library, and the gold standard for Machine Learning in the Python ecosystem. This library features various classification, regression, and clustering algorithms and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

   Syntax for Train-test split

```
from sklearn.model_selection import train_test_split
```

   Syntax for KNN:

```
from sklearn.neighbors import KNeighborsClassifier
```

   Syntex for Decision Tree Classifier

```
from sklearn.tree import DecisionTreeClassifier
```

12.6 Team required to develop:

- **Data Scientist:** Expertise in machine learning and deep learning can develop and train the algorithms that power the brain tumor detection system.
- **Software developer/Web developer:** Developing the web application and integrating the Python code into the app.
- **Quality assurance experts:** Ensuring that the system meets regulatory and quality standards and it performs reliably and consistently.

12.7 What does it cost?

The cost will be finalized after the product is built. The cost factor comes into play when the model has reached the deployment stage.

13 Code Implementation:

- Some Information About The Dataset

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Temparature     99 non-null     int64
 1   Humidity        99 non-null     int64
 2   Moisture        99 non-null     int64
 3   Soil Type       99 non-null     object
 4   Crop Type       99 non-null     object
 5   Nitrogen        99 non-null     int64
 6   Potassium       99 non-null     int64
 7   Phosphorous     99 non-null     int64
 8   Fertilizer Name 99 non-null     object
dtypes: int64(6), object(3)
memory usage: 7.1+ KB
```

```
25] df.describe()
```

	Temparature	Humidity	Moisture	Nitrogen	Potassium	Phosphorous
count	99.000000	99.000000	99.000000	99.000000	99.000000	99.000000
mean	30.282828	59.151515	43.181818	18.909091	3.383838	18.606061
std	3.502304	5.840331	11.271568	11.599693	5.814667	13.476978
min	25.000000	50.000000	25.000000	4.000000	0.000000	0.000000
25%	28.000000	54.000000	34.000000	10.000000	0.000000	9.000000
50%	30.000000	60.000000	41.000000	13.000000	0.000000	19.000000
75%	33.000000	64.000000	50.500000	24.000000	7.500000	30.000000
max	38.000000	72.000000	65.000000	42.000000	19.000000	42.000000

Figure 6: Information About Dataset

13.1 Simple EDA:

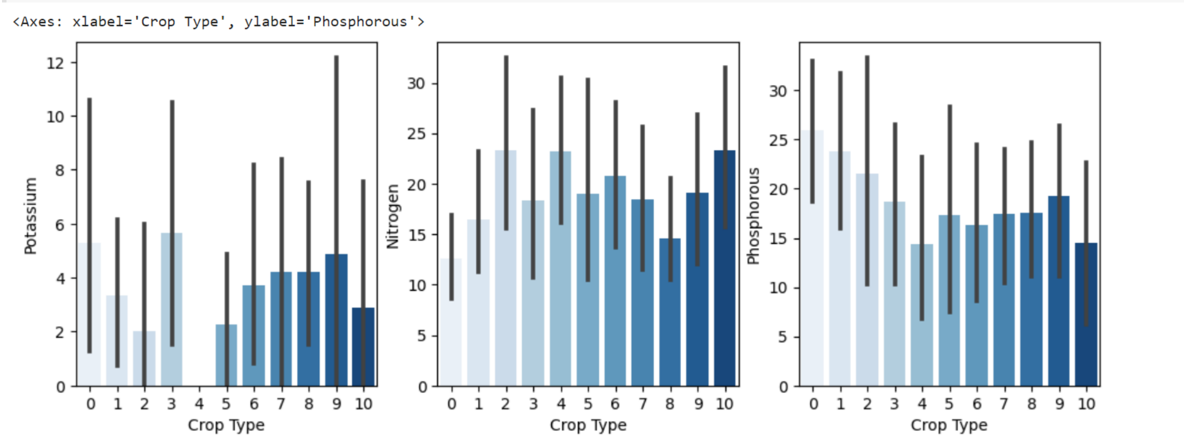

Figure 7: Barplot Between Crop Type ,Nitrogen, Potassium,Phosphorous

- **Observation:** From above figure we say that Tobacco,Maize,Barley grow well with potassium fertilizer.Wheat, Millets and groundnut grow well with nitrogen fertilizer.Barley, ground nut grow well with Phosphorous fertilizer.

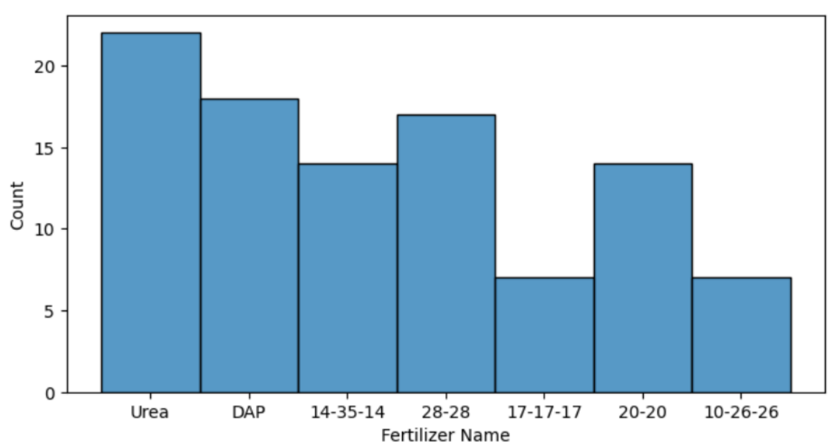

Figure 8: Histogram of various Fertilizer

- **Observation:** From above histogram we say that Urea fertilizer is widely used than other fertilizers

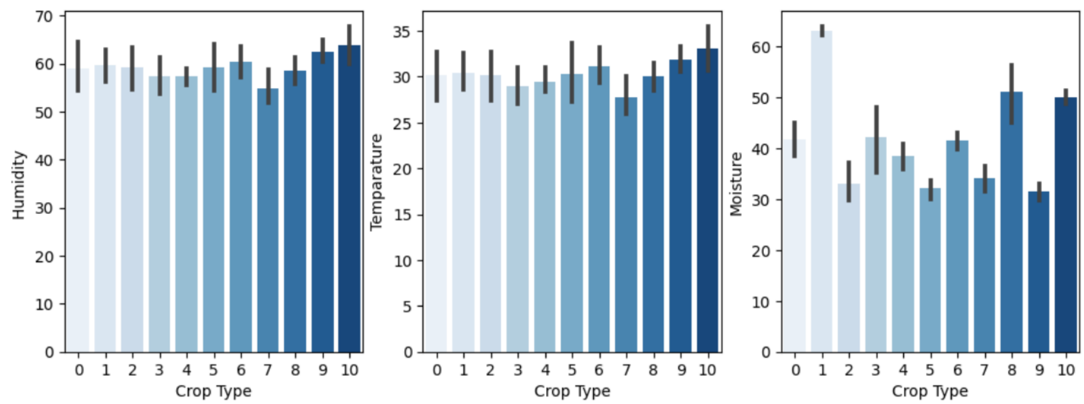

Figure 9: Barplot Between Crop Type and Humidity,Temparature,Moisture

- **Observation:** From above figure we say that Wheat requires more humidity as well as more temperature than other crop types and cotton requires more moisture.
- **Correlation matrix:**

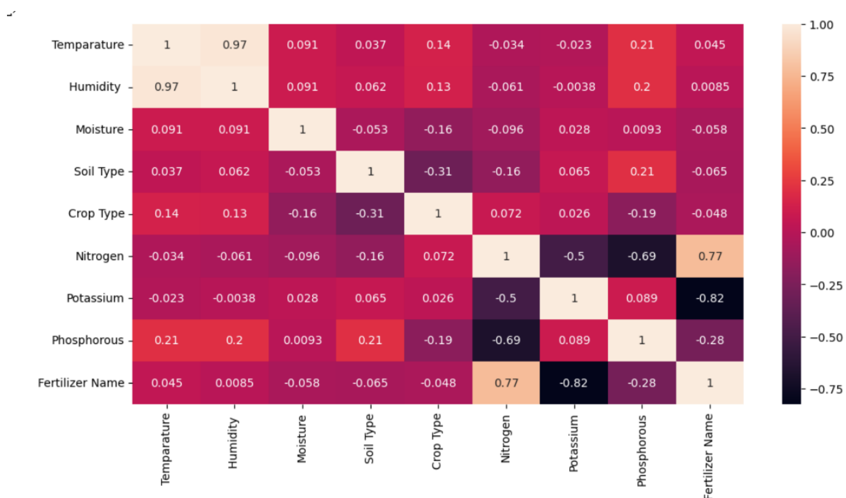

Figure 10: Correlation Between Columns

- In this, above fig., we find the same Correlation of all the columns. I use the matplotlib to resize the output of the image and using seaborn heatmap find a correlation between each of the columns.

13.2 ML Modelling:

We will use two different models and we will finalize the model which will give good accuracy.

```
y_pred_knn=knn.predict(X_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred_knn)
```

```
0.8
```

```
cm=confusion_matrix(y_test,y_pred_knn)
sns.heatmap(cm,annot=True,fmt='g')
```

```
<Axes: >
```

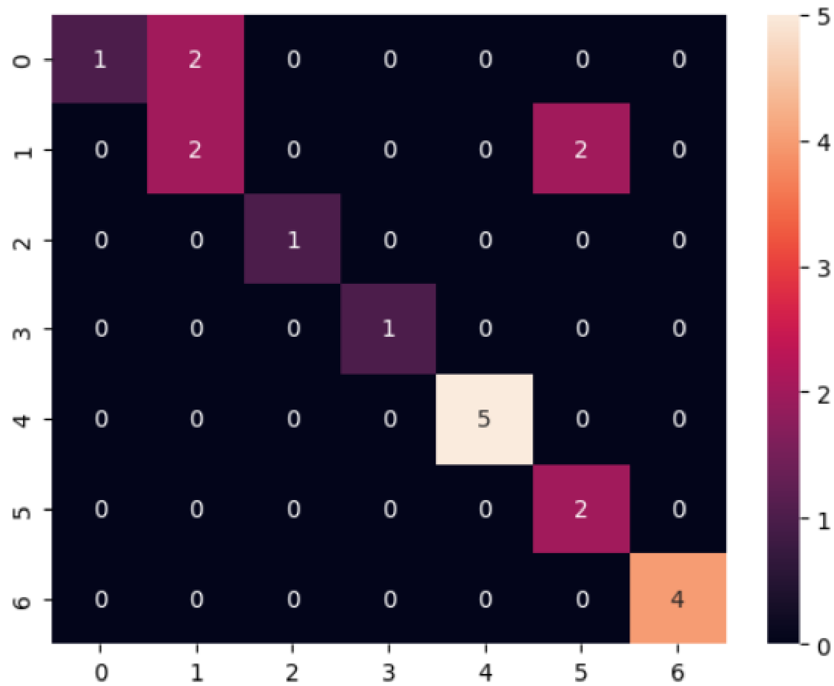

Figure 11: KNN Model Accuracy Score And Heatmap of Confusion Matrix

- KNN gives 80 percent accuracy when k=9 on test data.

```
] y_pred_DT=model.predict(X_test)
  from sklearn.metrics import accuracy_score
  accuracy_score(y_test,y_pred_DT)
```

```
0.9
```

```
] cm=confusion_matrix(y_test,y_pred_DT)
  sns.heatmap(cm,annot=True,fmt='g')
```

```
<Axes: >
```

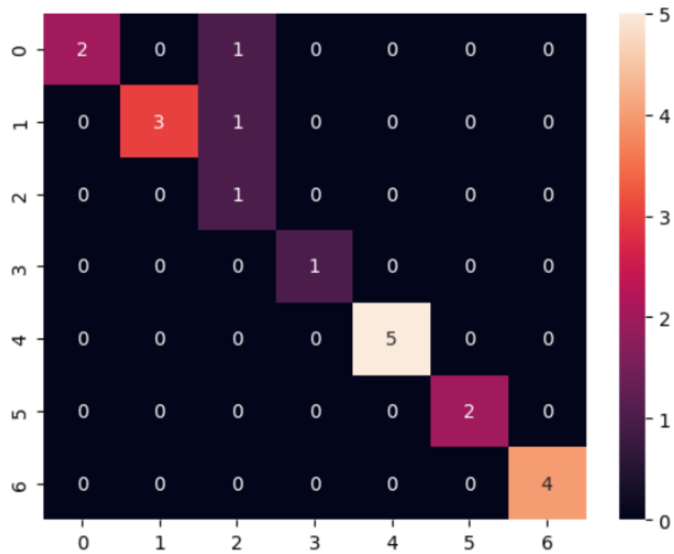

Figure 12: Decision Tree Model Accuracy Score And Heatmap of Confusion Matrix

- Decision Tree Model gives 90 percent accuracy on test data. So we finalize Decision Tree model.

13.3 Github link to the code implementation

Github link: https://github.com/Rishav19962/Feynn-Lab-First-Project-On-FertiliZer-Recomendation-System/blob/main/FeynnlabProject.ipynb

14 Conclusion:

Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of fertilizer for cultivation that will reduce their income. To reduce those type of loses we have developed a farmer friendly system with GUI, that will predict which would be the best suitable fertilizer for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the farmers to take right decision in selecting the fertilizer for cultivation such that agricultural sector will be developed by innovative idea.

References

[1] D. Manoj Kumar, N. Malyadri, M. Srikanth, *et al.*, "A machine learning model for crop and fertilizer recommendation," *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal— NVEO*, pp. 10531–10539, 2021.

[2] D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, "Crop recommendation system using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 558–569, 2021.

[3] A. Palaniraj, A. Balamurugan, R. D. Prasad, and P. Pradeep, "Crop and fertilizer recommendation system using machine learning," *IRJET*, vol. 8, 2021.

[4] D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 5, pp. 371–376, 2019.