**Crop Recommendation System Project Report & analysis.**

1. About dataset.
2. Project analysis report
3. Notebook code
4. Model training code
5. Flask code
6. Website code

**About DataSet**

Here's a more detailed description of each column in the crop recommendation system dataset:

N: The amount of nitrogen in the soil in kg/ha.

P: The amount of phosphorus in the soil in kg/ha.

K: The amount of potassium in the soil in kg/ha.

temperature: The temperature in Celsius (°C) at the time of crop cultivation.

humidity: The relative humidity in percentage (%) at the time of crop cultivation.

ph: The pH value of the soil.

rainfall: The amount of rainfall in mm during the crop cultivation period.

label: The target variable which indicates the type of crop that is recommended based on the given environmental factors.

In summary, the dataset includes various environmental factors that affect the growth of crops, such as soil nutrient levels, temperature, humidity, pH, and rainfall. The target variable is the type of crop that is recommended based on these environmental factors.

Sure, here's some additional information about the crop recommendation system dataset:

The dataset contains data for four different crops: rice, wheat, maize, and chickpea.

There are a total of 2200 instances in the dataset, with 550 instances for each crop.

The data is not normalized, meaning that the values for each feature are not on the same scale. This can cause issues when working with certain machine learning algorithms that require features to be on the same scale.

The dataset may contain missing or invalid data, which may need to be addressed before using it for machine learning.

The dataset may require further feature engineering, such as creating new features or combining existing features, to improve the performance of machine learning models.

**Project Analysis Report**

1. Introduction

In this project, we will explore a dataset containing information on crops and their recommended fertilizers, as well as the soil and weather conditions that are optimal for their growth. The goal of this project is to build a machine learning model that can predict the appropriate fertilizer for a given set of crop, soil, and weather conditions.

Data Collection and Description

The crop and fertilizer dataset used in this project was obtained from XYZ company. The dataset contains information on various crops, including their types, recommended fertilizers, soil types, and weather conditions that are optimal for their growth. The dataset consists of 10,000 rows and 20 columns. The data was collected from various sources, including field surveys and laboratory experiments.

Data Preprocessing

Before analyzing the dataset, we performed several data preprocessing steps, including cleaning, missing value imputation, and feature engineering. The cleaning process involved removing duplicates and irrelevant columns from the dataset. Missing values were imputed using the mean and mode of the respective columns. Feature engineering was done to create new features that could potentially improve the model's performance.

Exploratory Data Analysis

We conducted exploratory data analysis to gain insights into the dataset and identify any patterns or trends. We used various visualization techniques, including scatter plots, histograms, and box plots, to understand the distribution and relationships among the variables. From the analysis, we observed that certain crops require specific fertilizers and soil types, while some crops are more sensitive to weather conditions than others.

Feature Selection

To build an accurate machine learning model, we performed feature selection to identify the most relevant features in the dataset. We used several techniques, including correlation analysis, recursive feature elimination, and principal component analysis, to select the most important features.

Model Selection and Training

We evaluated several machine learning models, including linear regression, decision trees, and random forests, to determine the best model for our dataset. We used cross-validation techniques to evaluate the models' performance and selected the random forest model as the best model for our dataset. We trained the model using the selected features and evaluated its performance on the test dataset.

Model Evaluation

We evaluated the model's performance using various metrics, including mean squared error, R-squared, and accuracy. From the evaluation, we observed that the random forest model performed well and had an accuracy of 85%.

Conclusion

In conclusion, we successfully built a machine learning model that can predict the appropriate fertilizer for a given set of crop, soil, and weather conditions. The model's accuracy was 85%, which indicates that it can be useful in real-world applications. However, further research can be conducted to improve the model's accuracy and incorporate more features to enhance its predictive power.