**Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers**

**Alternative Title:**

Feature selection techniques and classifiers on crop prediction

**Aim:**

Predict the crop recommendation for a agriculture environment using machine learning technique and also using various feature selection techniques and classifiers.

**Abstract:**

 Agriculture is a growing field of research. In particular, crop prediction in agriculture is critical and is chiefly contingent upon soil and environment conditions, including rainfall, humidity, and temperature .In the past, farmers were able to decide on the crop to be cultivated, monitor its growth, and determine when it could be harvested.

Today, however, rapid changes in environmental conditions have made it difficult for the farming community to continue to do so. Consequently, in recent years, machine learning techniques have taken over the task of prediction, and this work has used several of these to determine crop yield. To ensure that a given machine learning (ML) model works at a high level of precision, it is imperative to employ efficient feature selection methods to preprocess the raw data into an easily computable Machine Learning friendly dataset. To reduce redundancies and make the ML model more accurate, only data features that have a significant degree of relevance in determining the final output of the model must be employed. Thus, optimal feature selection arises to ensure that only the most relevant features are accepted as a part of the model. Conglomerating every single feature from raw data without checking for their role in the process of making the model will unnecessarily complicate our model. Furthermore, additional features which contribute little to the ML model will increase its time and space complexity and affect the accuracy of the model’s output.

**Existing System:**

This work uses the Random Over-Sampling Examples (ROSE), Synthetic Minority Over-sampling Technique (SMOTE), and Majority Weighted Minority Over sampling Technique (MWMOTE) to help balance the given dataset. Feature selection is used to find salient features from the given dataset, resulting in better performance and classification techniques that help identify the target class.

**Problem Definition:**

Wrapper feature selection techniques such as the Boruta, Recursive Feature Elimination (RFE), and Modified Recursive Feature Elimination (MRFE) are used in this work to discover the
dataset’s salient features

**Proposed System:**

The RFE technique is a wrapper feature selection technique that starts with the entire dataset. The ranking method crucial to the RFE technique orders the dataset from the best to the worst, based on which salient features are selected. At each iteration, it eliminates the least important features from the dataset and updates the dataset, continuing the process until the most important ones are selected. RFE is a Wrapper-type feature selection and elimination technique that employs the greedy algorithm. The RFE algorithm recursively identifies and eliminates the least relevant features from the dataset until a sophisticated level of optimization is achieved. In the Wrapper method, the feature selection process is carried out based on a core machine learning algorithm which is fit into the dataset.SMOTETomek using oversampling on a dataset.

**Advantage:**

It is not known in advance how many features a model must keep. Therefore, to determine the optimal number of features, the RFE algorithm is cross-validated. Recursive Feature Elimination Cross-Validation (RFECV) works just like RFE but, in addition to RFE, it cross-validates the features, automatically selecting the features which give the best performance its also increase the accuracy.

**Modules:**

* Dataset Collection
* Data Preprocessing
* Algorithm Implementation
* Prediction

**Dataset Collection:**

 The Data’s are collected from various online resources. The dataset contains a more number of records, where some records are with some missing values. Those missing records have been removed from the dataset and filled the data using python package techniques (Pandas, NumPy).

**Data Preprocess:**

Sampling techniques are applied during preprocessing to balance the dataset and maximize the prediction performance. The sampling techniques used include the SMOTETomek is used for binary classification in the presence of rare classes and SMOTETomek for better classifier performance in the while SMOTETomek handles imbalanced dataset issues in crop prediction.

**Algorithm:**

After the Data Preprocessing Techniques, the model is implemented by six choose ML Algorithms are Gradient Boosting Classifier; Extra Tree Classifier. Among the six chosen models, on a feature vector created by count vector, while Logistics Regression performed poorly, the task of classifying Smishing messages is a sensitive one, and the return of false positives and false negatives should be taken into account.

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Software Requirements:**

* Operating System : Windows 10 (64 bit)
* Software : Python
* Tools : Anaconda

**Conclusion:**

Predicting crops for cultivation in agriculture is a difficult task. This paper has used a range of feature selection and classification techniques to predict yield size of plant cultivations. The results depict that an ensemble technique offers better prediction accuracy than the existing classification technique. Forecasting the area of cereals, potatoes and other energy crops can be used to plan the structure of their sowing, both on the farm and country scale. The use of modern forecasting techniques can bring measurable financial benefits.

**Future Work:**

 When we apply stacked regression, the result has been so improvised than when those models were applied individually. The output which has been shown in figure is currently a web application, but our future work would be building an application where the farmers can use it as app and converting the whole system in their regional language.

**Architecture Diagram:**

Camera

Pre-process

Dataset

Machine Learning Algorithms

Model Creation

Real time

User input

User interface

Prediction

