**REPORT**

**PROJECT PROPOSAL:**

The project aims to address the problem of rice variety classification using artificial intelligence techniques. The dataset used for this project is the "Rice\_Image\_Dataset," which contains images of different rice varieties. The dataset meets the requirements with more than 100 samples, at least 4 features, and a minimum of 3 classes. The AI techniques employed in this project include Random Forest Classifier, Multilayer Perceptron (MLP) Classifier, Deep Convolutional Neural Network (CNN), and K-means Clustering.

**DATA PRE-PROCESSING AND VISUALIZATION:**

The provided code includes resizing the images to a consistent size of 180x180 pixels. To improve accuracy, the image data is converted into ‘numpy’ arrays and then scaling is performed by dividing the image pixel values by 255. This step ensures that the pixel values are within the range of 0 to 1, which can enhance the training process for certain algorithms. Also we flatten the data so that the image data will reshaped as one dimension array which is given to the model to get the maximum accuracy.

**IMPLEMENTATION & EVALUATION:**

* **Task 1: Random Forest Classifier:**

The code implements a basic Random Forest Classifier first then we have used Grid Search CV to get the optimal results using scikit-learn. At first we have implemented basic Random Forest with only parameter ‘random\_state = 42’. And then chosen parameters for the classifier, such as the number of estimators, maximum depth, and minimum samples split for the optimized Random Forest Model, are set in a parameter grid which is:

param\_grid\_rf = {

'n\_estimators': [50, 100, 150],

'max\_depth': [None, 5, 10],

'min\_samples\_split': [2, 5, 10]

}.

Grid search is performed to find the best parameters for the model using cross-validation. The accuracy of the resulting basic Random Forest model is reported as 97% on the test data and for the optimized Random Forest model we got accuracy of 96.6%.

* **Task 2: Multilayer Perceptron (MLP) Classifier:**

The code implements a basic MLP Classifier using scikit-learn. The MLP model has 3 hidden layers with 20, 15, and 10 neurons respectively, and uses the logistic activation function and then we have implemented optimized model by using Grid Search CV on the following parameters which is placed in parameter grid which is:

param\_grid\_mlp = {

'hidden\_layer\_sizes': [(20,), (20, 15), (20, 15, 10)],

'alpha': [0.0001, 0.001, 0.01]

}.

The model is trained using the provided training data. The accuracy of the basic resulting model is reported as 75.6% on the test data and for the optimized model is 96.2%.

* **Task 3: Deep Convolutional Neural Network (CNN):**

The code implements a basic CNN using the Keras library. The CNN model consists of multiple convolutional and pooling layers followed by fully connected layers. The model is compiled with the Adam optimizer, loss = 'sparse\_categorical\_crossentropy' and metrics=['accuracy']. While fitting our basic CNN model we have used epochs=10, batch\_size=32 and trained using the training data. The accuracy of the resulting model is reported as 95.9% on the test data. To improve the accuracy, we have implemented a optimized CNN model which uses Data Augmentation layer to perform random zoom to 30%, random rotation to 30%, random contrast to 30% and random flip to horizontal. After doing this technique we have got 93.9% accuracy.

* **Task 4: Clustering:**

The code applies basic K-means clustering to the dataset. The number of clusters is set to 5, corresponding to the number of classes. The accuracy of the clustering results is measured by comparing the predicted labels with the true labels from the test data. The resulting accuracy is reported as 14%, indicating that the clusters do not correspond well to the classes in the dataset. This happens sometimes based on the dataset we are using.

**ETHICAL AND SOCIAL IMPACT:**

A number of ethical and social concerns are raised by the use of artificial intelligence in the classification of rice varieties. The privacy of the people depicted in the photos needs to be respected, hence data privacy is a major problem when working with image datasets. To make sure that the classification models do not discriminate against specific rice types or favour particular traits, bias and fairness should be carefully evaluated. The environment, consumers, and farmers may be affected by the outcomes of the AI models. Accurate classification helps farmers optimise their farming techniques, while rice variety features help buyers make wise decisions. To win the trust of stakeholders, it is imperative to make sure that the AI models are transparent, comprehensible, and reliable.

Fairness, accountability, and openness should be given top priority when implementing AI responsibly, and any recognised ethical or social concerns should be given consideration for potential mitigations.

The research proposal's overall goal is to classify rice varieties using AI methods. The implementation includes a variety of techniques, including CNN, K-means clustering, Random Forest Classifier, and MLP Classifier. The code implementation shows how to use various algorithms and reports on how accurate each one is. There are recommendations for enhancing each algorithm's precision. The section on ethics and societal impact highlights the significance of ethical AI implementation and tackles potential issues with data privacy, bias, fairness, and stakeholder implications.