

# Multilingual Agricultural Query Classification System

Group Members:

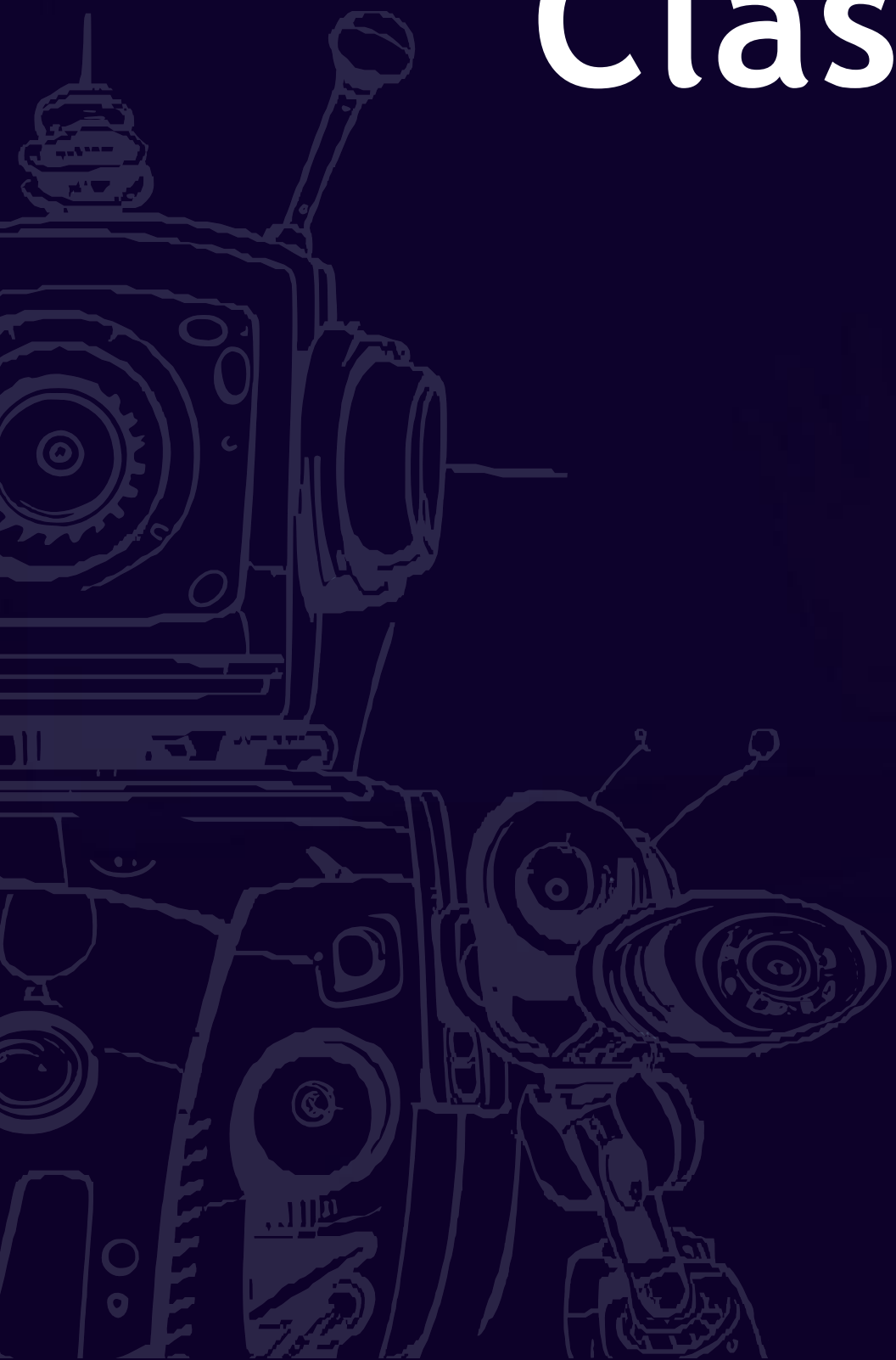
Yash Joshi – U23AI119

Harshit Shah – U23AI075

Diwyanshu – U23AI089

Pawan – U23AI130

Ashish – U23AI129



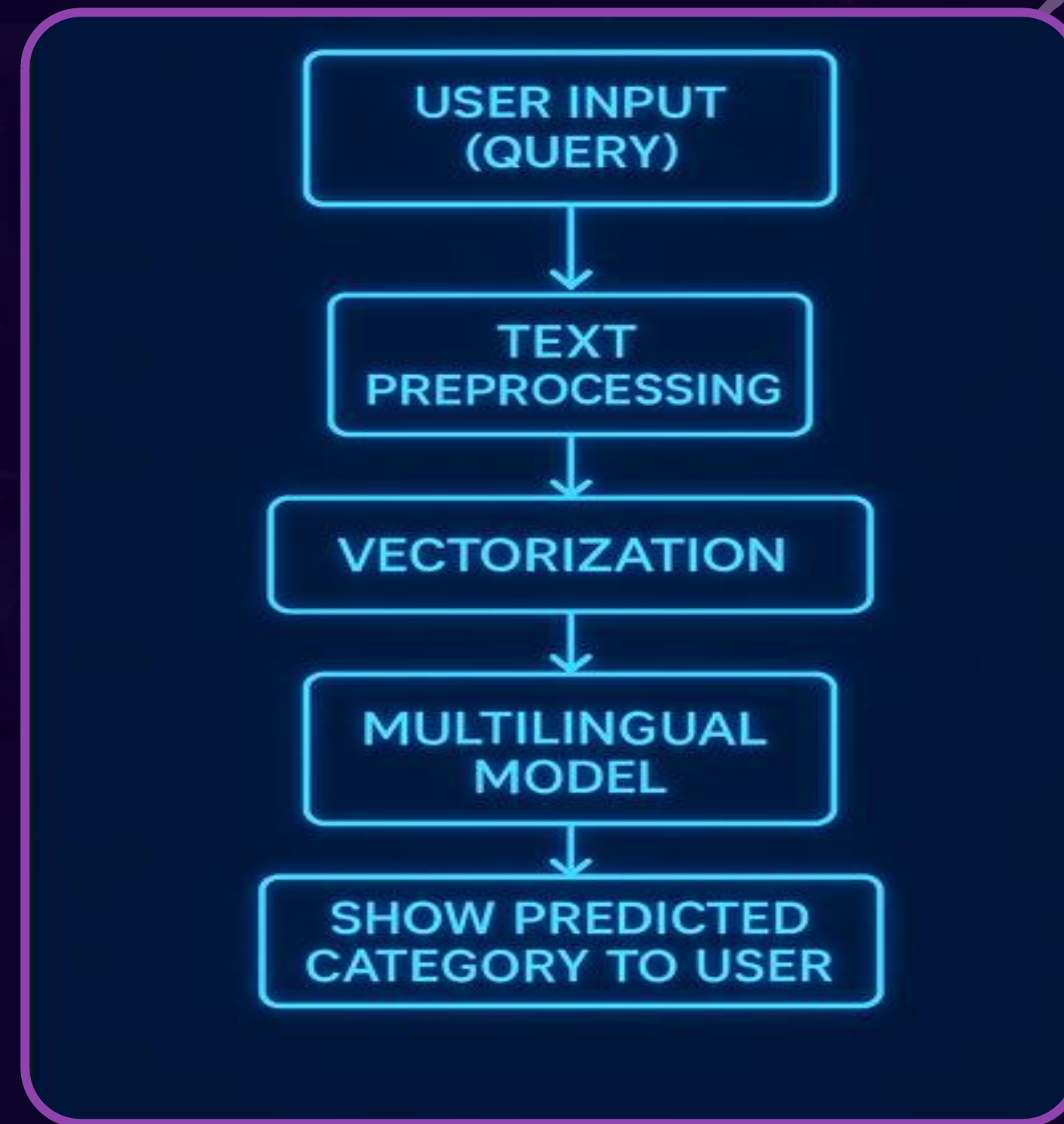
# Our project

This project focuses on intelligently categorizing agriculture-related queries into ten predefined classes by leveraging multilingual classification models.

LABEL	DESCRIPTION
0	CULTURAL PRACTICES
1	FERTILIZER USE AND AVAILABILITY
2	FIELD PREPARATION
3	GOVERNMENT SCHEMES
4	MARKET INFORMATION
5	NUTRIENT MANAGEMENT
6	PLANT PROTECTION
7	VARIETIES
8	WEED MANAGEMENT



# Approach



# Data Preprocessing

## 1. Text Cleaning and Normalization

- Removed special characters, numbers, and punctuations
- Lowercased all text
- Normalized white spaces

## 2. Stemming and Lemmatization

- Stemming: Reduced words to root form (e.g., farming → farm)
- Lemmatization: Converted words to dictionary form (e.g., better → good)

## 3. Feature Extraction

- Applied TF-IDF Vectorizer to transform text into numerical features
- Captured word importance across queries

## Purpose:

- Improve model performance
- Reduce noise and redundancy
- Enhance classification accuracy

# Model Training

## Data Splitting

The Training Set comprises 80% of the data for model training

The Test Set consists of 20% of the data for final evaluation after training.

## Algorithms Used in Hard Voting Ensemble

Logistic Regression

Support Vector Machine (SVM)

Random Forest

Naive Bayes

XGBoost

Stochastic

Gradient

Descent (SGD)

## Voting Mechanism

The Hard Voting mechanism is used, where the final prediction is based on the majority vote, which is the most common class predicted by the individual models in the ensemble.

# Model Testing

The model testing process involves evaluating the performance of the trained ensemble model on the test set.

Apply the trained ensemble model to the test set.

Measure the accuracy and other evaluation metrics.

Compare the performance between different ensemble models to determine the most effective approach.

# Model Testing

**Accuracy:** The proportion of correct predictions.

**Precision:** How many of the predicted positives are actually correct.

**Recall:** How many of the actual positives were predicted correctly.

**F1-Score:** The harmonic mean of precision and recall.

**Confusion Matrix:** Analyzes true positives, true negatives, false positives, and false negatives.

# English Model Testing

	precision	recall	f1-score	support
Cultural Practices	0.47	0.36	0.41	842
Fertilizer Use and Availability	0.48	0.65	0.55	520
Field Preparation	0.14	0.24	0.18	256
Government Schemes	0.48	0.77	0.59	404
Market Information	0.81	0.78	0.80	408
Nutrient Management	0.51	0.51	0.51	885
Plant Protection	0.88	0.77	0.82	3792
Varieties	0.50	0.79	0.62	270
Weather	0.90	0.81	0.85	2323
Weed Management	0.63	0.84	0.72	300
accuracy			0.71	10000
macro avg	0.58	0.65	0.60	10000
weighted avg	0.74	0.71	0.72	10000



# Hindi Model Testing

	precision	recall	f1-score	support
Cultural Practices	0.48	0.34	0.40	842
Fertilizer Use and Availability	0.47	0.62	0.53	520
Field Preparation	0.12	0.25	0.16	256
Government Schemes	0.43	0.72	0.54	404
Market Information	0.75	0.77	0.76	408
Nutrient Management	0.49	0.53	0.51	885
Plant Protection	0.87	0.74	0.80	3792
Varieties	0.49	0.80	0.61	270
Weather	0.91	0.80	0.85	2323
Weed Management	0.60	0.83	0.70	300
accuracy			0.69	10000
macro avg	0.56	0.64	0.59	10000
weighted avg	0.73	0.69	0.70	10000

# Gujrati Model Testing

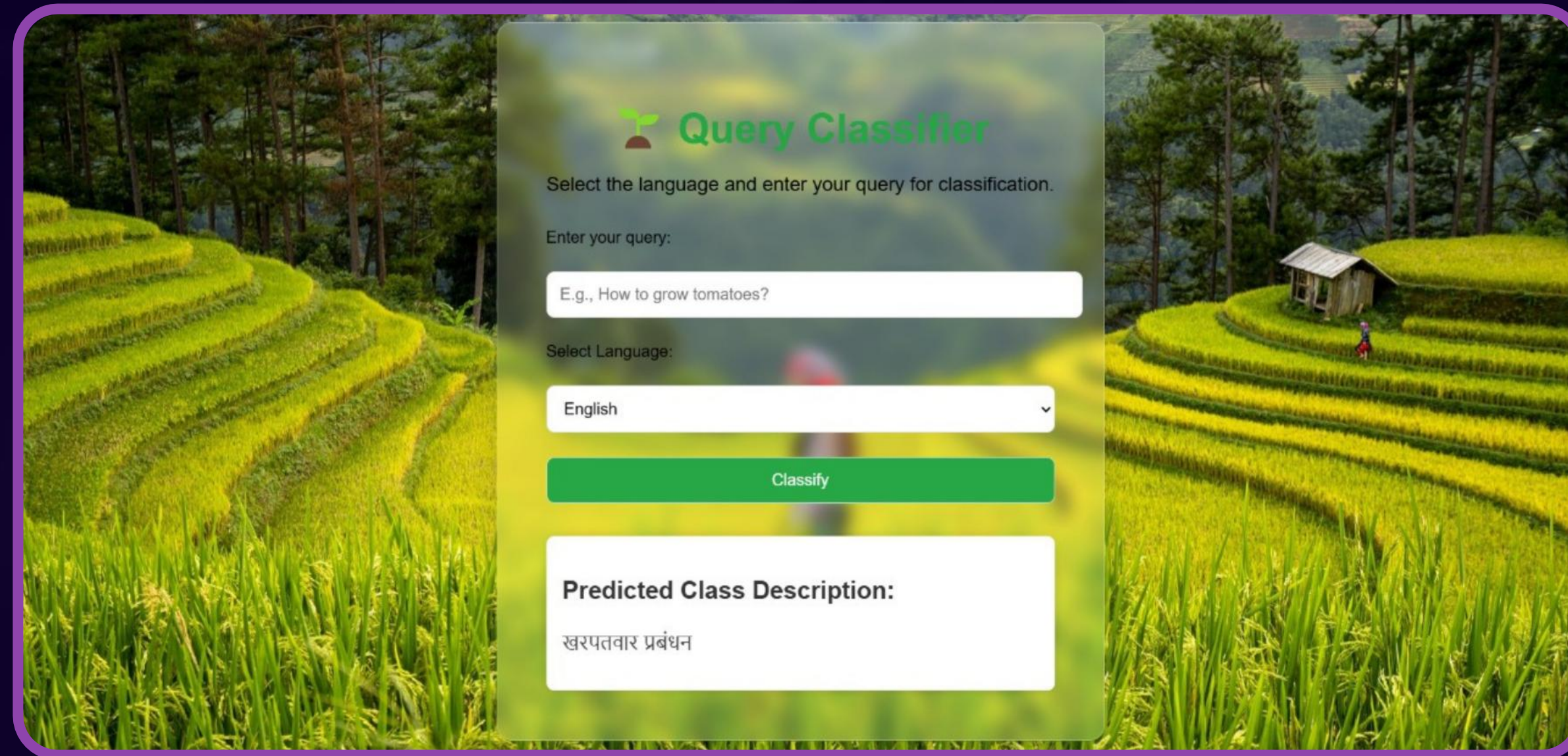
	precision	recall	f1-score	support
Cultural Practices	0.45	0.34	0.39	842
Fertilizer Use and Availability	0.46	0.63	0.53	520
Field Preparation	0.12	0.24	0.16	256
Government Schemes	0.47	0.74	0.57	404
Market Information	0.79	0.78	0.79	408
Nutrient Management	0.47	0.53	0.50	885
Plant Protection	0.88	0.74	0.80	3792
Varieties	0.48	0.79	0.60	270
Weather	0.90	0.80	0.85	2323
Weed Management	0.62	0.83	0.71	300
accuracy			0.69	10000
macro avg	0.56	0.64	0.59	10000
weighted avg	0.73	0.69	0.70	10000

# Multilingual Model Testing


	precision	recall	f1-score	support
Cultural Practices	0.44	0.16	0.23	2526
Fertilizer Use and Availability	0.44	0.23	0.30	1561
Field Preparation	0.15	0.12	0.13	768
Government Schemes	0.31	0.36	0.34	1211
Market Information	0.74	0.57	0.64	1223
Nutrient Management	0.45	0.19	0.27	2654
Plant Protection	0.86	0.29	0.44	11376
Varieties	0.34	0.32	0.33	810
Weather	0.34	0.90	0.50	6969
Weed Management	0.25	0.35	0.29	902
accuracy			0.42	30000
macro avg	0.43	0.35	0.35	30000
weighted avg	0.57	0.42	0.40	30000



# User Interface



The image shows a user interface for a 'Query Classifier' application. The background is a scenic view of terraced rice fields in a mountainous area, with a small wooden hut visible on one of the terraces. The interface is centered on the screen and consists of a semi-transparent white box with rounded corners. At the top of the box is a green plant icon followed by the title 'Query Classifier'. Below the title, there is a prompt 'Select the language and enter your query for classification.' followed by 'Enter your query:'. A text input field contains the placeholder text 'E.g., How to grow tomatoes?'. Below this is a 'Select Language:' label and a dropdown menu currently showing 'English'. A green 'Classify' button is positioned below the language selection. At the bottom of the interface, there is a section titled 'Predicted Class Description:' which displays the text 'खरपतवार प्रबंधन' (Weed Management) in Devanagari script.

 **Query Classifier**

Select the language and enter your query for classification.

Enter your query:

E.g., How to grow tomatoes?

Select Language:

English

**Classify**

**Predicted Class Description:**

खरपतवार प्रबंधन