📄 Short Explanation of Project

In this project, the goal was to classify land cover types (like water, forest, and grass) using time-series NDVI (Normalized Difference Vegetation Index) data.

I started by cleaning the dataset and using **mean imputation** to handle missing NDVI values. The NDVI timepoints spanned 27 satellite image dates.

I used **Label Encoding** to convert land cover classes into numeric labels and trained a **multinomial Logistic Regression** model using `scikit-learn`.

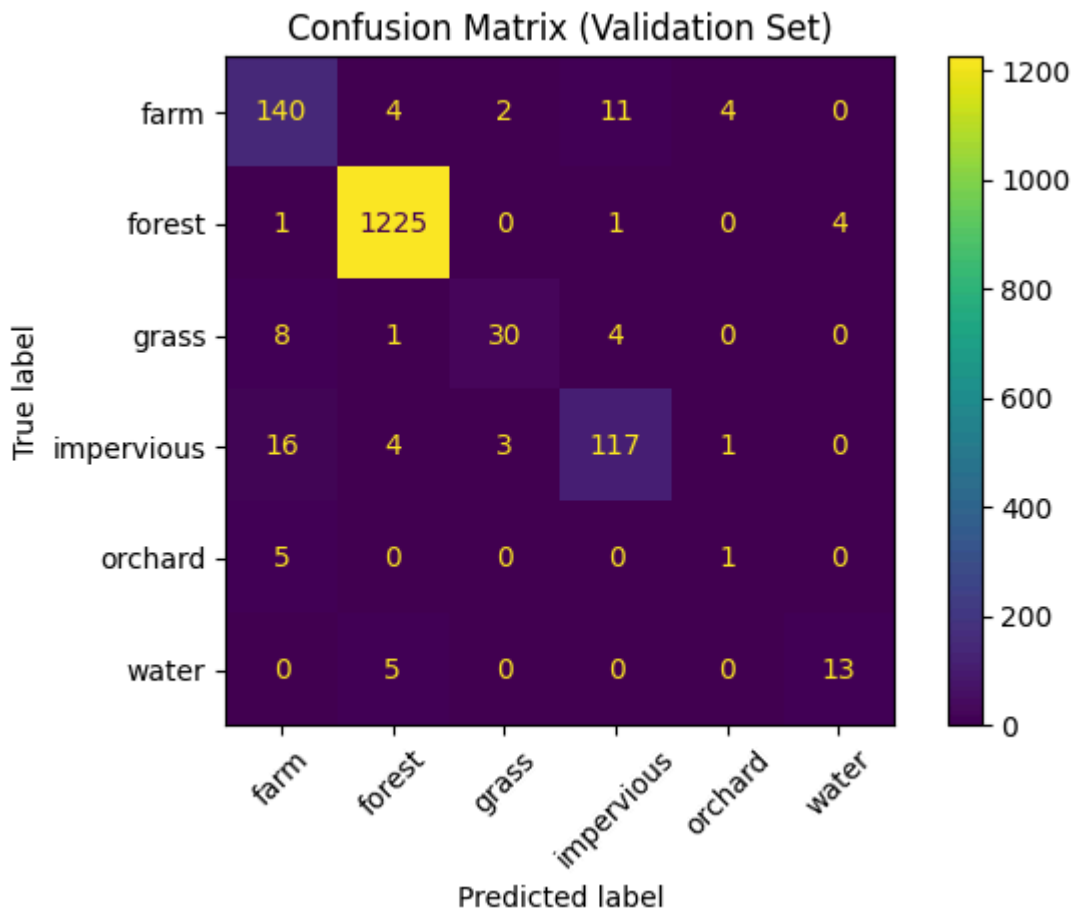
The model achieved an accuracy of **95.3% on the validation set**, with especially strong performance on dominant classes like **forest** and **farm**.

I then predicted on the test data and generated a submission file in the required format. The entire pipeline—from preprocessing to submission—is included in the attached notebook/PDF

.

Validation Results

Class	Precision	Recall	F1-Score	Support
Farm	0.82	0.87	0.85	161
Forest	0.99	1.00	0.99	1231
Grass	0.86	0.70	0.77	43
Impervious	0.88	0.83	0.85	141
Orchard	0.17	0.17	0.17	6
Water	0.76	0.72	0.74	18
Overall Accuracy	–	–	95.3%	1600 samples




Class Distribution in Training Data
Number of Samples
6000
5000
4000
3000
2000
1000
0
forest
farm
impervious
grass
water
orchard
Land Cover Type