Artificial Intelligence

Forest Type Mapping

Prepared by

Shaista Noureen (2023-CS-676)

Mishal Fatima(2023-CS-679)

Shahzaib Sadiq (2023-CS-690)

Submitted to

Ms. Shanfa

Department of Computer Science

University of Engineering and Technology, New Campus

1. Objective

The goal of this project is to create a mapping system that classifies forest types based on features from the Forest Cover Type dataset. The system uses machine learning models for classification, and the results are displayed as human-readable forest types, enabling a better understanding of the terrain and vegetation in different regions.

2. Technologies & Libraries Used

Languages: Python 3
Environment: Google Colab

Libraries:

- **pandas, numpy**: Data handling
- **matplotlib, seaborn**: Data visualization
- **scikit-learn**: Machine learning algorithms and metrics
- **ipywidgets**: Interactive feature inputs
- **PCA, KMeans**: Dimensionality reduction and clustering

3. Dataset

Name: Forest CoverType Dataset
Source: UCI Machine Learning Repository / Kaggle
Format: CSV
Instances: ~581,000
Features: 54 numeric and binary features
Target: `Cover_Type` (7 classes)

4. Data Preprocessing

- **Upload & Load Dataset**: Dataset is uploaded via `files.upload()` in Google Colab.
- **Feature/Target Split**: Features are split into `X`, and the target `Cover_Type` is assigned to `y`.
- **Data Reduction**: 20% of data is sampled for faster experimentation.
- **Scaling**: Standardization of features using `StandardScaler`.
- **Train-Test Split**: 70% for training, 30% for testing.

5. ML Algorithms Used

Algorithm	Library	Purpose
K-Nearest Neighbors	`sklearn.neighbors`	Classification
Naive Bayes	`sklearn.naive_bayes`	Classification
Random Forest	`sklearn.ensemble`	Classification
Decision Tree	`sklearn.tree`	Classification
K-Means	`sklearn.cluster`	Clustering
PCA	`sklearn.decomposition`	Dimensionality Reduction

6. Model Evaluation

Accuracy Scores:

After training the models on a reduced dataset:

Model	Accuracy (Test Set)
K-Nearest Neighbors	~0.94
Naive Bayes	~0.75
Random Forest	~0.96
Decision Tree	~0.90

Classification Reports:

Each model outputs precision, recall, f1-score, and support for each class to assess performance more granularly.

7. K-Means Clustering + PCA

Goal:

To visualize clusters in 2D space using PCA, with `n_clusters=7` to match the 7 forest cover types.

Result:

A scatter plot is generated, with clusters colored by their labels, showing separation in the reduced PCA space.

8. Interactive Prediction Tool

Features:

- Users can input the first 10 features using widgets.
- The prediction is made using the **Random Forest** model.

Output:

- Top 3 probable forest types with corresponding percentages.
- The final predicted cover type mapped to human-readable names.

Example Output:

```
matlab
CopyEdit
 Top 3 Predicted Cover Types:
1. Cover_Type 2 - Lodgepole Pine (82.47%)
2. Cover_Type 1 - Spruce/Fir (13.23%)
3. Cover_Type 3 - Ponderosa Pine (4.12%)

Final Predicted Cover_Type: 2 - Lodgepole Pine
```

9. Visualization

- **KMeans PCA Scatter Plot**: Displays the colored clusters from unsupervised KMeans overlaid on a 2D PCA projection.
- **Interactive Widgets**: Provides an easy-to-use UI for real-time forest type prediction based on user inputs.

10. Conclusion

- **Model Performance**: The **Random Forest** model consistently outperforms other algorithms in terms of accuracy.
- **Clustering**: KMeans shows some separability, but the supervised models provide more reliable results.

- **User Interaction**: The interactive widget allows users to input features and receive predictions in real time, making the model more accessible for non-technical users.

11. Future Improvements

- **Include all 54 features** in the interactive tool (currently only 10).
- Implement **hyperparameter tuning** using **GridSearchCV**.
- Add **confusion matrix** and **ROC/AUC** visualizations.
- Deploy the model as a **web app** using Flask or Streamlit for broader use.

12. Appendix

Cover Type Classes Mapping:

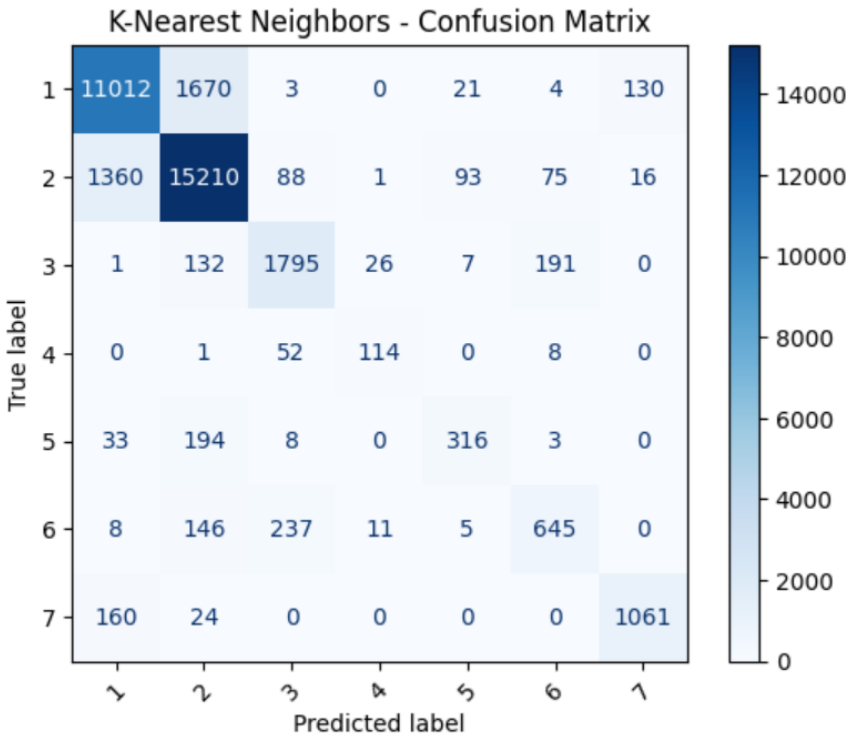
Class	Forest Type Name
1	Spruce/Fir
2	Lodgepole Pine
3	Ponderosa Pine
4	Cottonwood/Willow
5	Aspen
6	Douglas-fir
7	Krummholz

Screen Shots

1:K-Nearest Neighbour

◆ K-Nearest Neighbors ◆
Accuracy: 0.8649

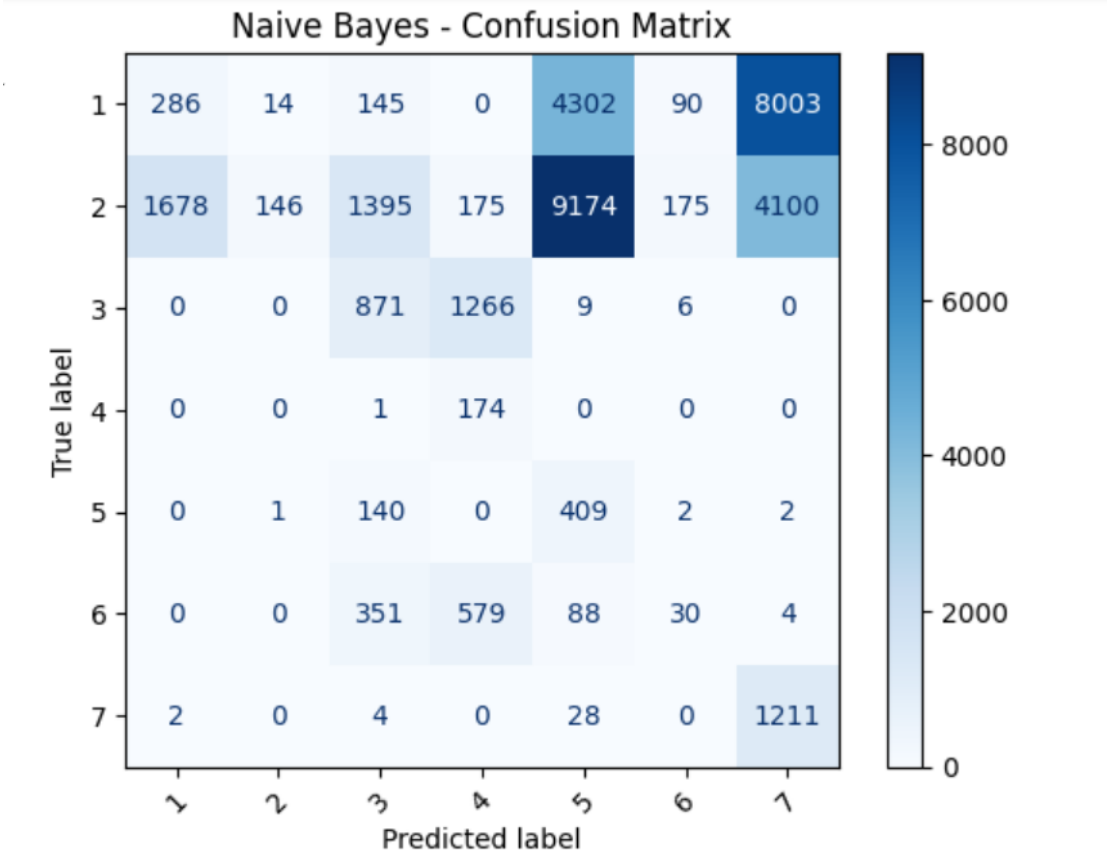
	precision	recall	f1-score	support
1	0.88	0.86	0.87	12840
2	0.88	0.90	0.89	16843
3	0.82	0.83	0.83	2152
4	0.75	0.65	0.70	175
5	0.71	0.57	0.63	554
6	0.70	0.61	0.65	1052
7	0.88	0.85	0.87	1245
accuracy			0.86	34861
macro avg	0.80	0.75	0.78	34861
weighted avg	0.86	0.86	0.86	34861



2:Naïve Bayes:

◆ Naive Bayes ◆
Accuracy: 0.0897

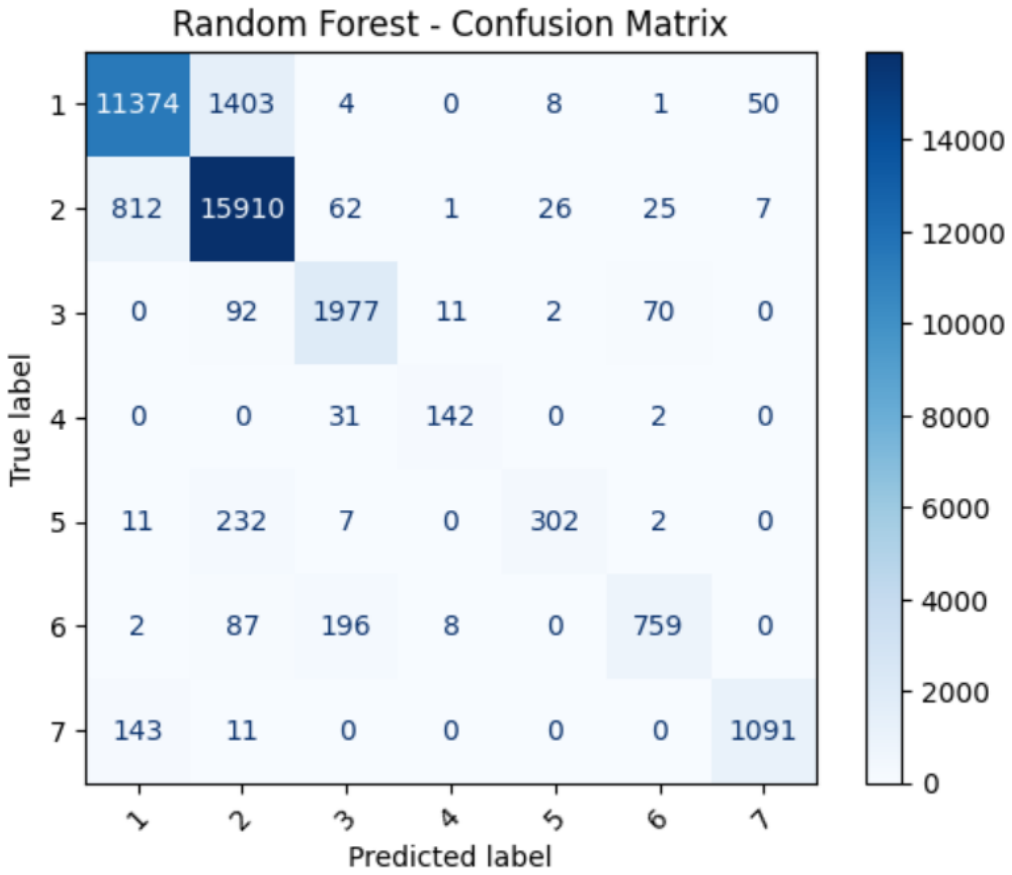
	precision	recall	f1-score	support
1	0.15	0.02	0.04	12840
2	0.91	0.01	0.02	16843
3	0.30	0.40	0.34	2152
4	0.08	0.99	0.15	175
5	0.03	0.74	0.06	554
6	0.10	0.03	0.04	1052
7	0.09	0.97	0.17	1245
accuracy			0.09	34861
macro avg	0.24	0.45	0.12	34861
weighted avg	0.52	0.09	0.05	34861



3:Random Forest:

◆ Random Forest ◆
Accuracy: 0.9057

	precision	recall	f1-score	support
1	0.92	0.89	0.91	12840
2	0.90	0.94	0.92	16843
3	0.87	0.92	0.89	2152
4	0.88	0.79	0.83	175
5	0.90	0.54	0.67	554
6	0.87	0.71	0.78	1052
7	0.95	0.88	0.91	1245
accuracy			0.91	34861
macro avg	0.90	0.81	0.85	34861
weighted avg	0.91	0.91	0.90	34861

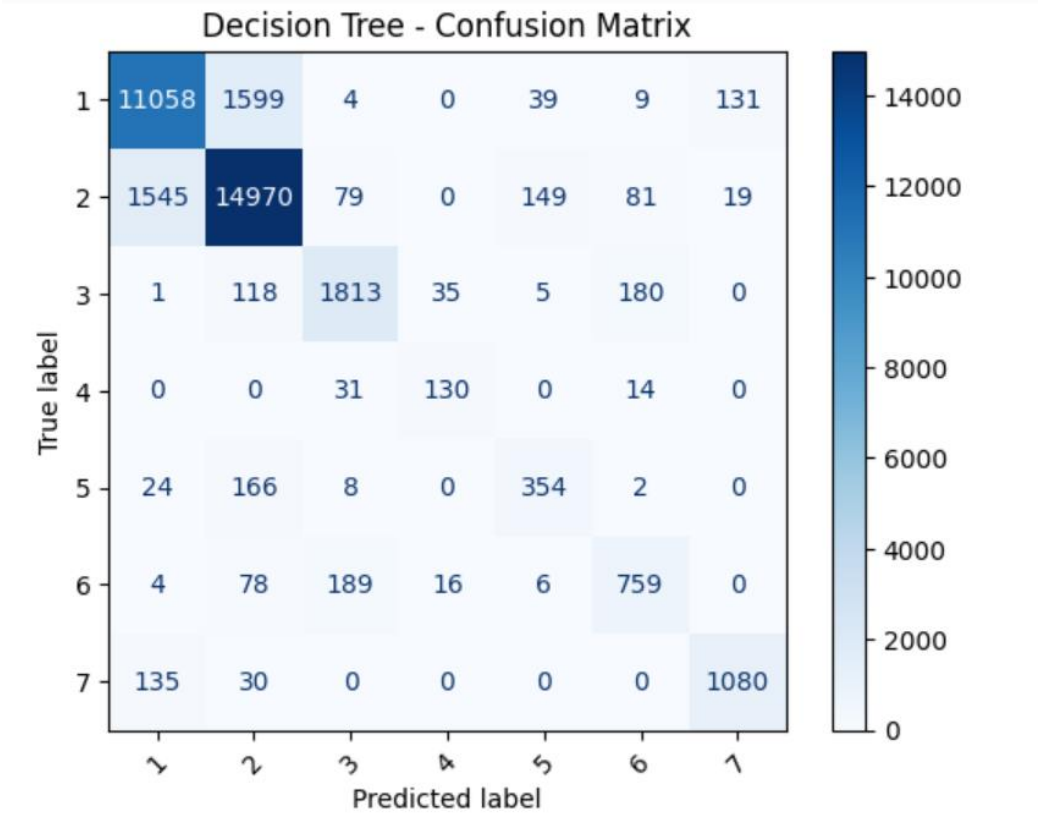


4:Decision Tree:

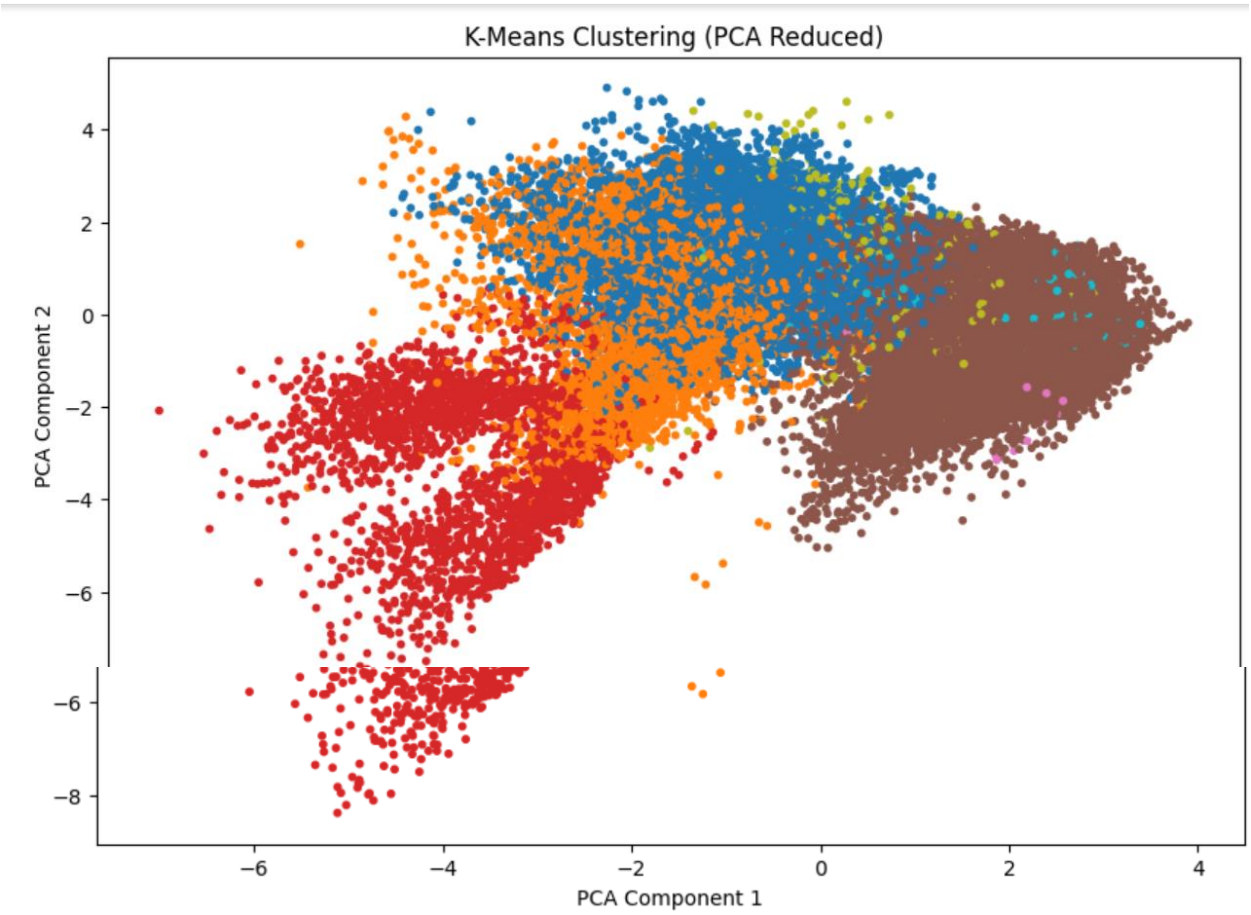
◆ Decision Tree ◆

Accuracy: 0.8645

	precision	recall	f1-score	support
1	0.86	0.86	0.86	12840
2	0.88	0.89	0.88	16843
3	0.86	0.84	0.85	2152
4	0.75	0.76	0.75	175
5	0.65	0.66	0.65	554
6	0.74	0.72	0.73	1052
7	0.88	0.86	0.87	1245
accuracy			0.86	34861
macro avg	0.80	0.80	0.80	34861
weighted avg	0.86	0.86	0.86	34861

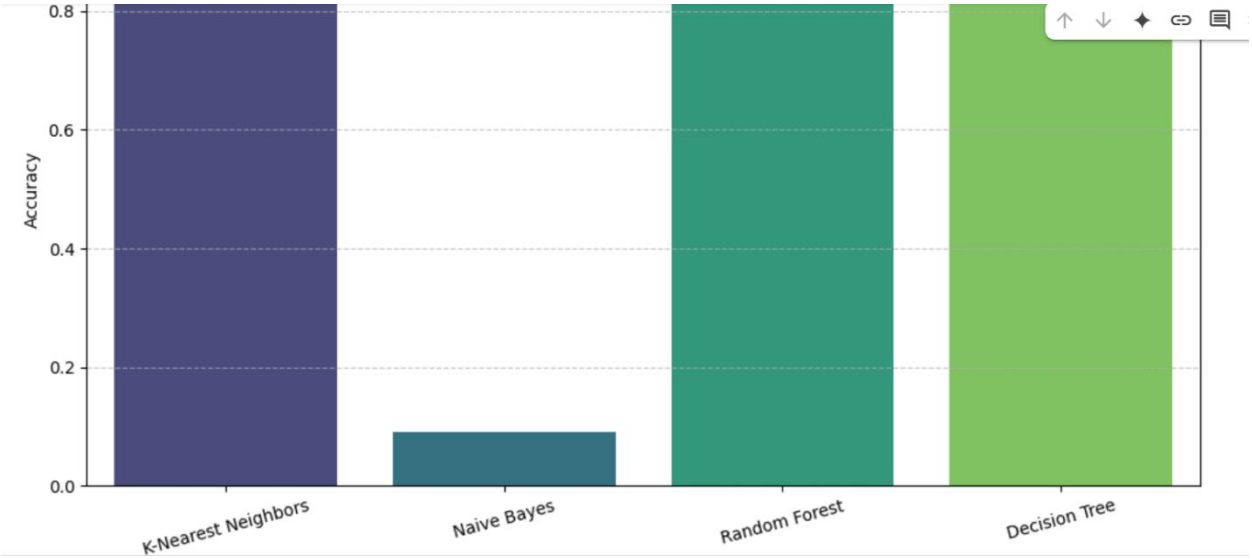


5:K Mean clustering graph:



6:Final Comparison:

✅ Final Model Comparison:
K-Nearest Neighbors: 0.8649
Naive Bayes: 0.0897
Random Forest: 0.9057
Decision Tree: 0.8645



7:Input Panel:

Enter feature values to predict Cover_Type (first 10 features):

Elevation	0
Aspect	0
Slope	0
Horizontal_Distance_To_Hydrology	0
Vertical_Distance_To_Hydrology	0
Horizontal_Distance_To_Roadways	0
Hillshade_9am	0
Hillshade_Noon	0
Hillshade_3pm	0
Horizontal_Distance_To_Fire_Points	0

Predict Cover_Type

an

8:Data Entry:

Enter feature values to predict Cover_Type (first 10 features):

Elevation	043
Aspect	045
Slope	04
Horizontal_Distance_To_Hydrology	034
Vertical_Distance_To_Hydrology	056
Horizontal_Distance_To_Roadways	03
Hillshade_9am	074
Hillshade_Noon	064
Hillshade_3pm	034
Horizontal_Distance_To_Fire_Points	45

Predict Cover_Type

9:Predicted Result: