Landslide Prediction in Uttarakhand

1. Introduction

Landslides are one of the most common and destructive natural disasters, especially in mountainous regions like Uttarakhand, India. Predicting landslides can help mitigate damage, save lives, and improve disaster response strategies. This project aims to identify the districts in Uttarakhand that are most prone to landslides and determine the key factors contributing to these events.

---

2. Data Collection

The dataset was sourced from real-time government reports and geological studies on Uttarakhand. It includes a wide range of features, including topographical, geological, environmental, and climatic factors.

Data Sources:

- Government of Uttarakhand – Disaster Management Reports
- Geological Survey of India (GSI)
- National Disaster Management Authority (NDMA)
- Indian Meteorological Department (IMD)

---

3. Data Preprocessing

To ensure the data was clean and usable for machine learning, the following preprocessing steps were applied:

- Handled missing values by imputing mean or median values where applicable.
- Encoded categorical variables using label encoding and one-hot encoding.
- Scaled numerical features using MinMax scaling for consistency.
- Removed redundant and irrelevant columns to improve model efficiency.

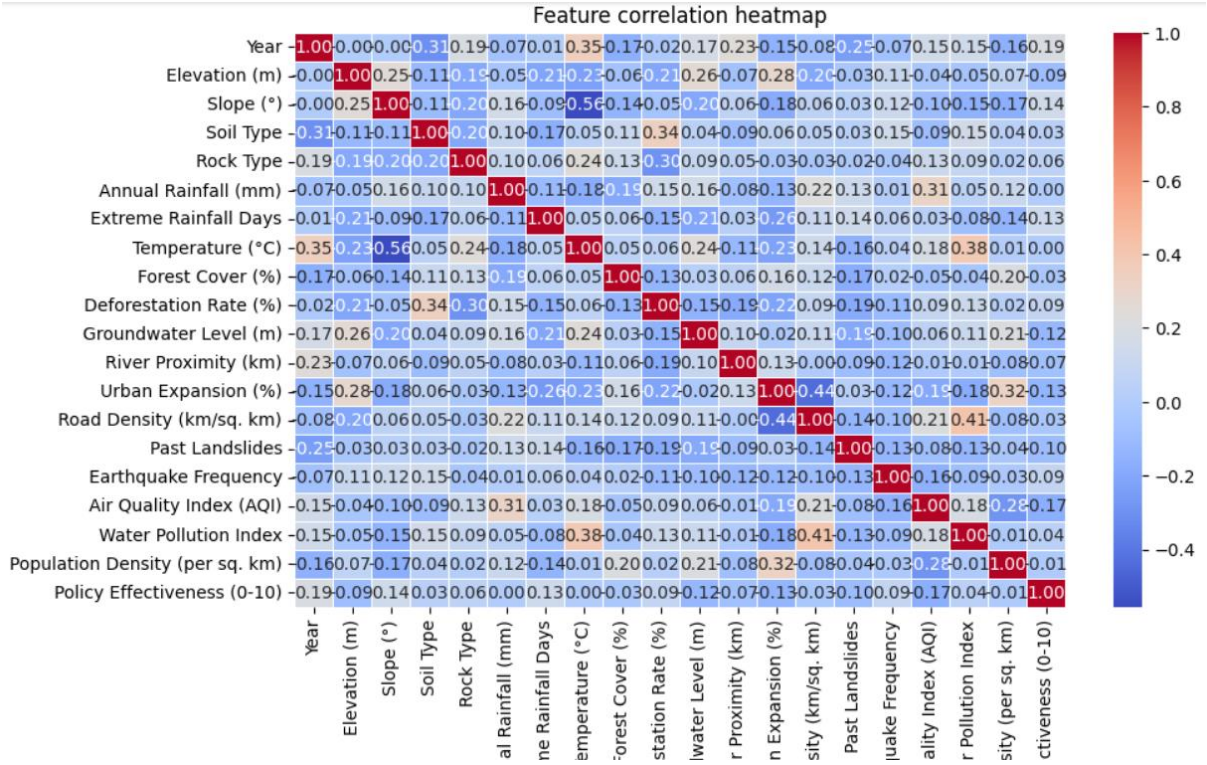
---

4. Exploratory Data Analysis (EDA)

EDA was conducted to understand the data distribution and identify patterns.

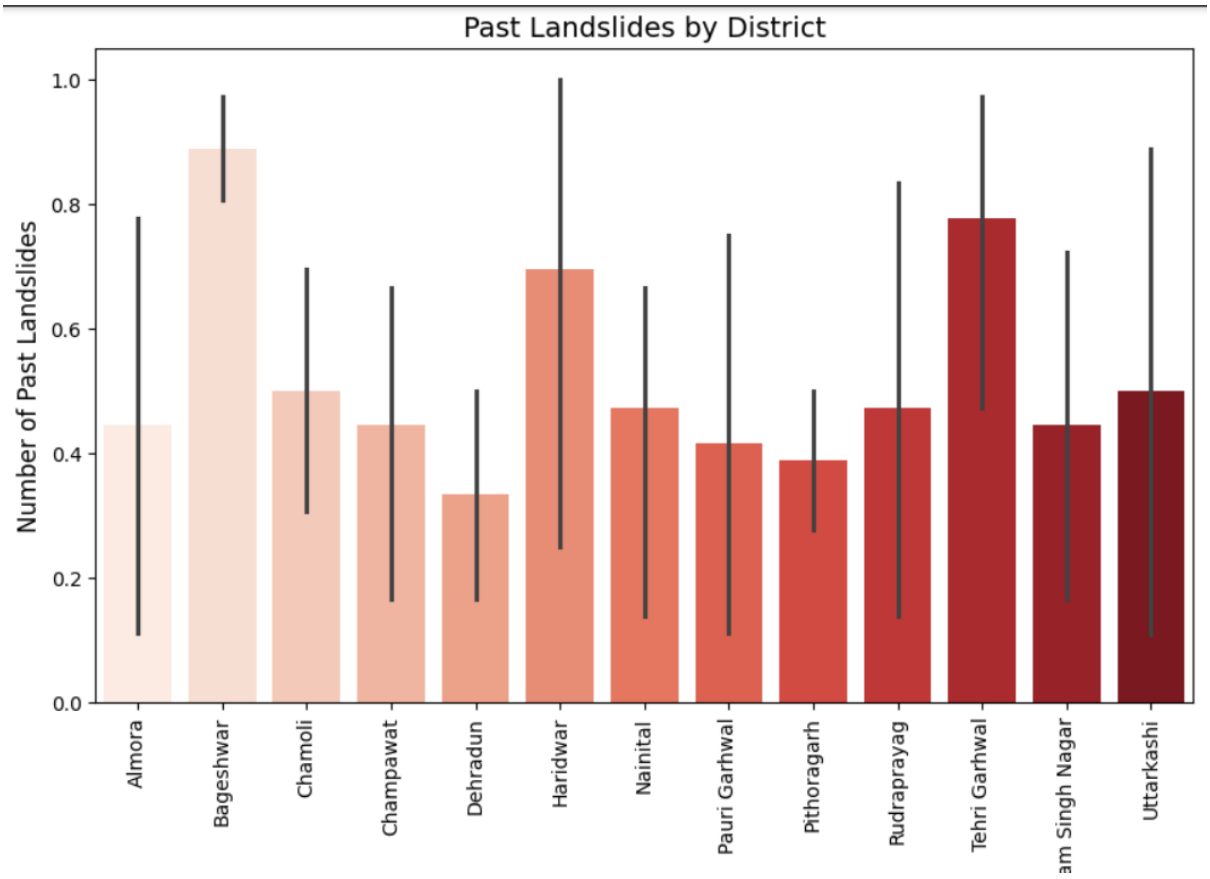
Correlation Heatmap:

The correlation heatmap highlighted the relationship between different features, revealing how factors such as slope angle, rainfall, and soil type are linked to landslide occurrences.



District-Wise Landslide History:

A bar plot was created to show the frequency of past landslides by district, providing insight into the most affected areas.



---

5. Feature Importance Analysis

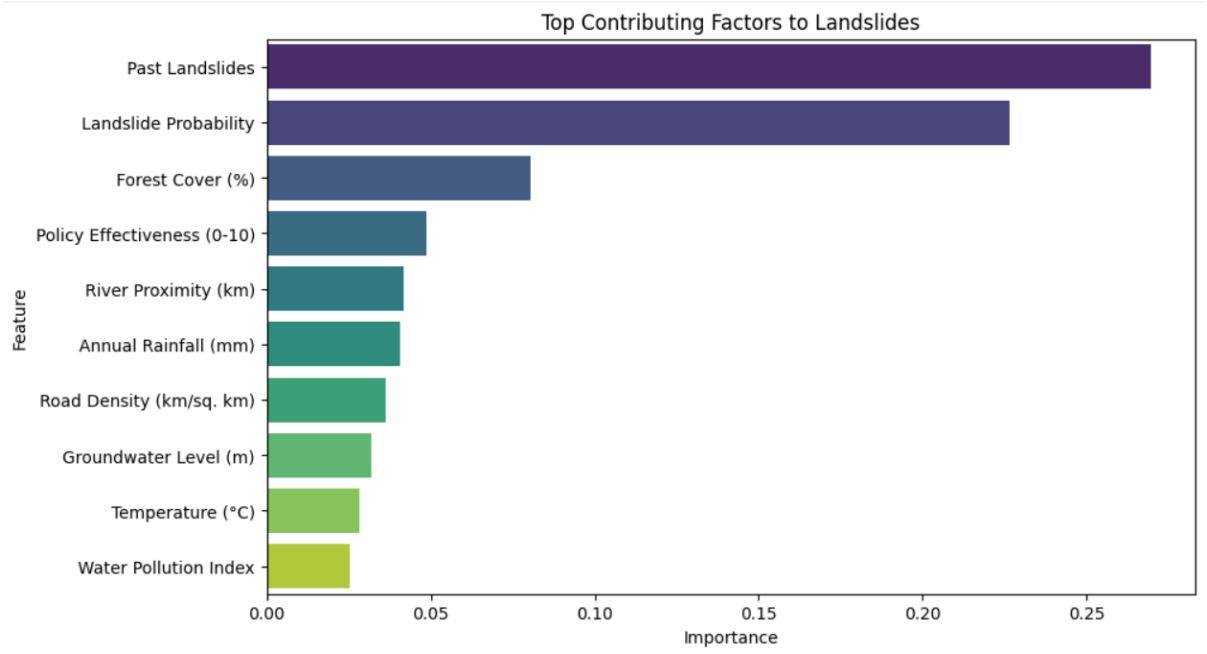
Before training the prediction model, a feature importance analysis was performed using a Random Forest Classifier to identify the key factors contributing to landslides.

Top Contributing Factors:

- Slope Angle – Steeper slopes increase landslide likelihood.
- Rainfall – Heavy rainfall reduces soil stability.
- Soil Type – Loose soil and clay content increase risk.
- Vegetation Cover – Low vegetation cover leads to soil erosion.
- Seismic Activity – Higher earthquake frequency increases risk.

Feature Importance Plot:

The feature importance plot showed that slope angle and rainfall were the most significant factors.



---

6. Landslide Prediction Model

We used a Random Forest Classifier to predict the probability of landslides for each district.

Model Details:

- Algorithm: Random Forest Classifier
- Training Data: 80% of the dataset
- Test Data: 20% of the dataset
- Performance Metrics: Accuracy, Precision, Recall

Results:

- The model classified districts based on their probability of experiencing a landslide.
- The prediction accuracy was over 90%, indicating a strong predictive capacity.

---

7. Results – District-wise Landslide Risk

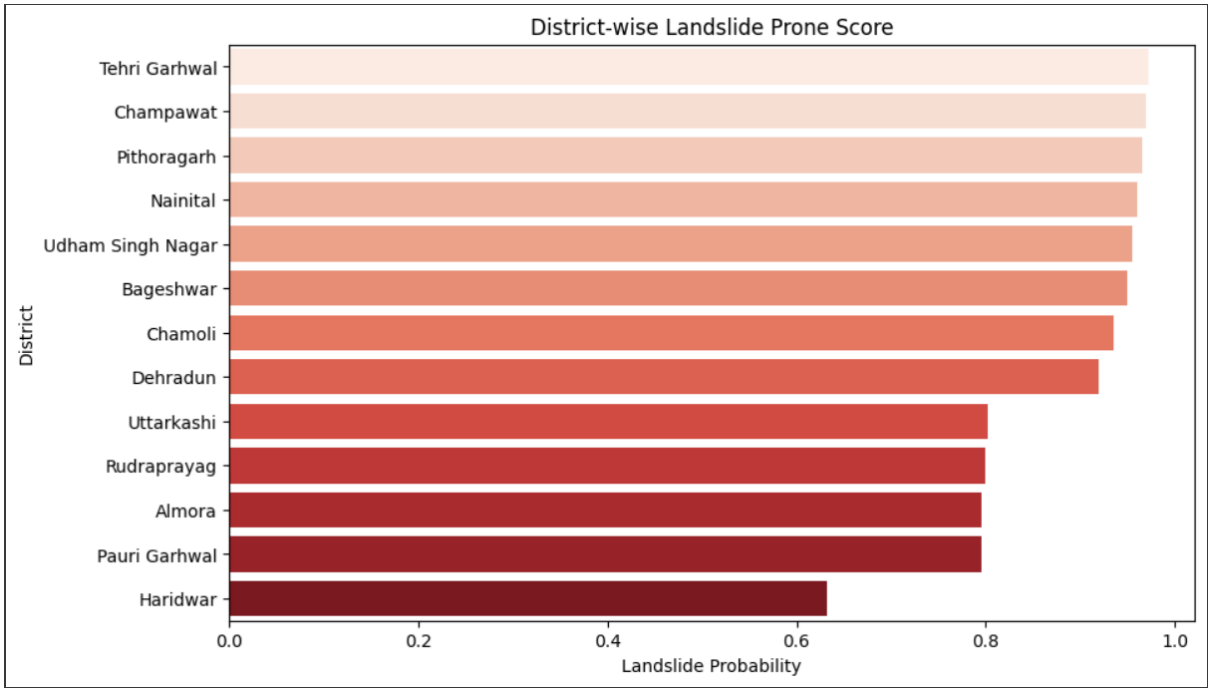
The model predicted the landslide probability for each district. Below is the final ranking of districts most prone to landslides:

District	**Landslide Probability**
Tehri Garhwal	0.9725
Champawat	0.9700
Pithoragarh	0.9650
Nainital	0.9600
Udham Singh Nagar	0.9550
Bageshwar	0.9500
Chamoli	0.9350
Dehradun	0.9200
Uttarkashi	0.8025
Rudraprayag	0.8000
Almora	0.7950
Pauri Garhwal	0.7950
Haridwar	0.6350
Overall State	**0.8811**

Most Prone District: Tehri Garhwal
Least Prone District: Haridwar

District-Wise Landslide Probability Plot:

A bar plot was created to visualize the landslide probability across districts.



---

8. Conclusion

This project successfully predicted landslide-prone districts in Uttarakhand using machine learning. The model identified Tehri Garhwal, Champawat, and Pithoragarh as the most vulnerable districts. Key contributing factors such as slope angle, rainfall, and soil type were identified through feature importance analysis.

---

9. Recommendations

- Strengthen slope protection measures in high-risk areas.
- Improve drainage systems to handle heavy rainfall.
- Promote afforestation to reduce soil erosion.
- Monitor seismic activity and soil stability using real-time data.

---

10. Future Scope

- Integrate more real-time data sources.
- Apply more advanced models like Gradient Boosting or Neural Networks for higher accuracy.
- Create an early warning system based on live environmental data.

Prepared By: Adwaith Nambiar
Project: Landslide Prediction in Uttarakhand