GLACIER ANALYSIS IN THE USA

HTTPS://GITHUB.COM/GENESISJRUIZ/PROJECT_4_USA_GLACIERS

PROJECT 4 TEAM

Cruzald Boholst - aradr1
Genesis Ruiz - genesisjruiz
Kyle McDaniel - Kyle-McGitHub
Parth Pisolkar – parthpisolkar2002
Stephen Markovic - BC_SM
Zahra El Oula Kirathi - KZahraeloula

OBJECTIVE

This project aims to analyze glacier data from the USA by leveraging diverse data sources and applying machine learning techniques. The goals are to:

1. **Predict Glacier Types**: Utilize machine learning models to classify and predict different types of glaciers based on available data.
2. **Visualize Trends**: Create visualizations to illustrate key trends related to glaciers in the USA, including their locations, areas, elevations, and temperature variations.

PURPOSE

Our research aims to enhance our understanding of environmental changes by examining glacier dynamics. By analyzing glacier data and employing machine learning techniques, this study seeks to uncover insights into glacier types, trends in their location, area, elevation, and temperature.

This understanding will contribute to broader climate studies and inform strategies for addressing environmental challenges related to glacier changes.

OVERVIEW

Data Collection

Data Preprocessing

Dashboard

- Exploratory Data Analysis (EDA)
- Predictive Model

Limitations & Recommendations

Conclusion

DATA COLLECTION

Glacier Dataset: NASA - NSIDC Data (https://nsidc.org/home)

Temperature Dataset: NCEI Climate at a Glance (https://www.ncei.noaa.gov)

Location Data: Simple Maps US Cities (https://simplemaps.com/data/us-cities)

DATA PROCESSING

Data Acquisition: Data about glaciers was acquired from NASA in .kml format and supporting data related to temperature and location.

Cleaning: Loaded into Jupyter Notebook, removed unnecessary columns, and cleaned data.

Saving: Cleaned data was saved to a CSV file, loaded into a database, and then into a dataframe.

DATA MODELING

Separation: Numerical data was separated from categorical data.

Dummy Variables: Created dummy variables from categorical data and merged them back.

Splitting: Data was split into testing and training sets.

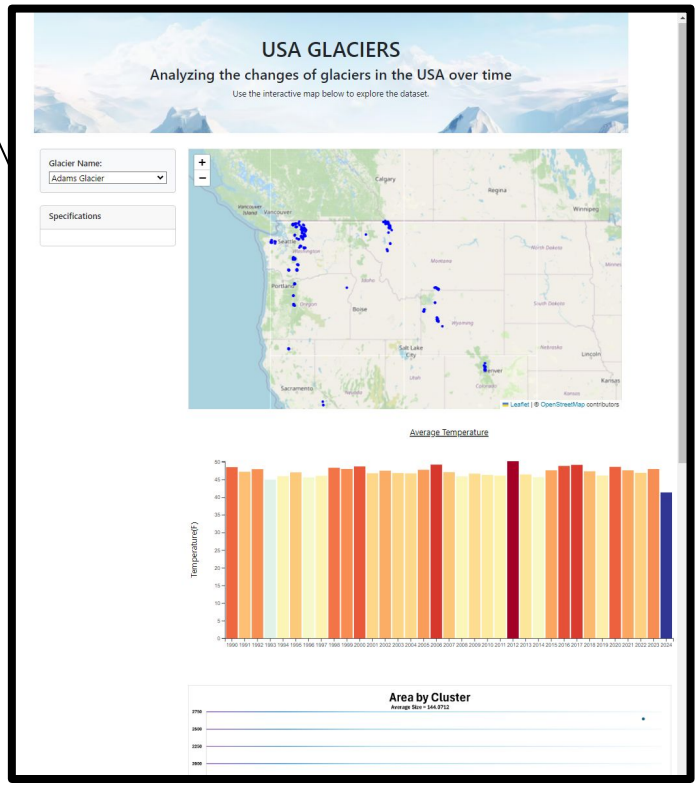
Model Initialization: Initialized RandomForestClassifier and trained the model.

Prediction: Initially predicted glacier existence, but shifted to predicting glacier types due to poor initial results.

Performance: Achieved an R-square score of 93% for predicting glacier types.

DASHBOARD

Interactive dashboard showcasing data analysis performed on the data sets.



LIMITATIONS

Geographical Scope: Data is limited to the USA, primarily Alaska and northern parts.

Feature Imbalance: Uneven amount of features compared to documentation.

Recommendation: To further enrich these findings, integrating higher-quality data, additional environmental factors, and expanding the analysis to include global data over extended periods could deepen our understanding of glacier dynamics and enhance our predictive capabilities regarding their formation and evolution.

CONCLUSION

In conclusion, our research applied machine learning to analyze US glacier data, revealing key insights into glacier types and trends. Notably, the model demonstrated high accuracy in predicting glacier types. Despite certain limitations, our findings contribute to climate studies and strategies for environmental challenges.

THANK YOU