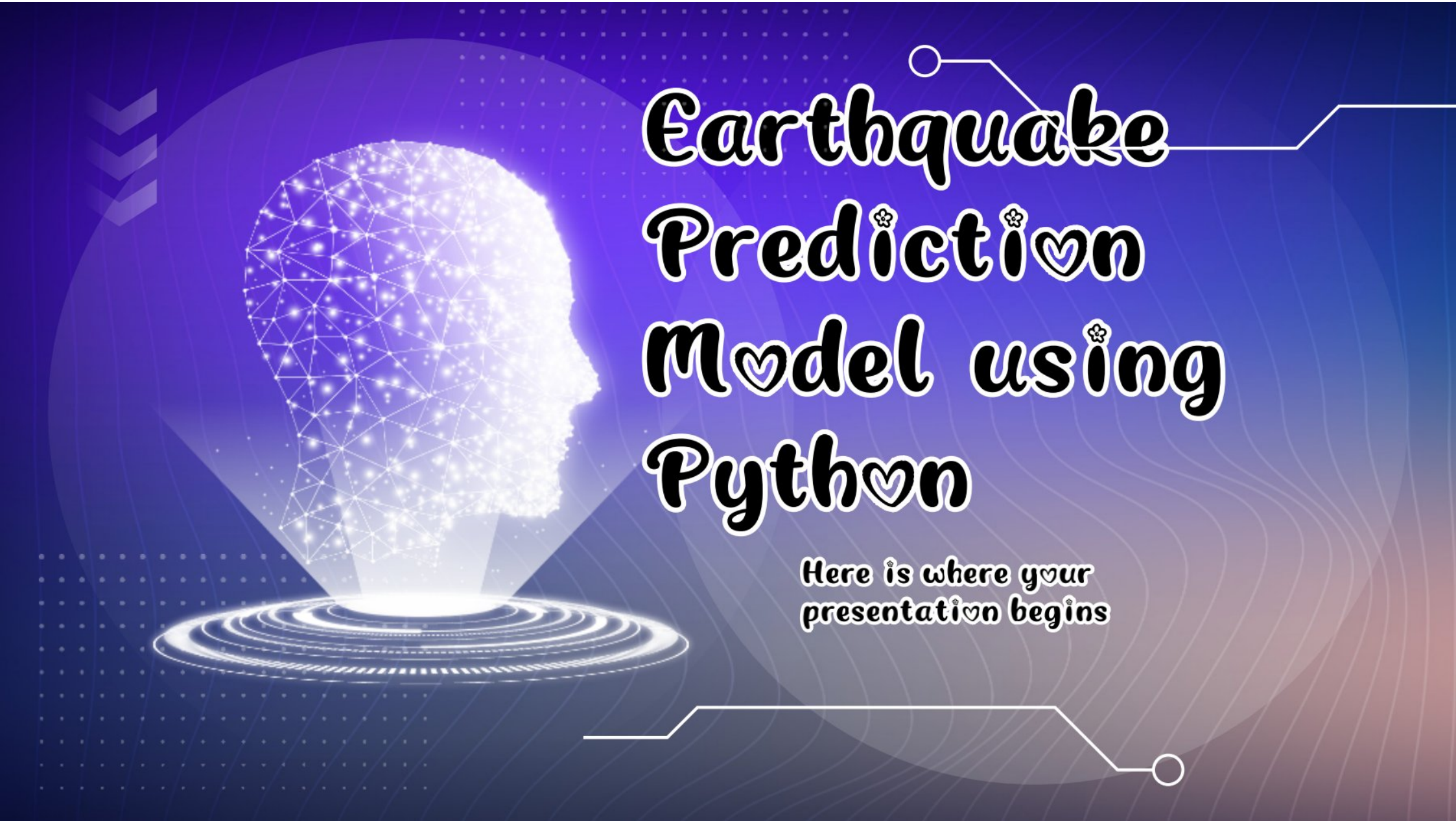
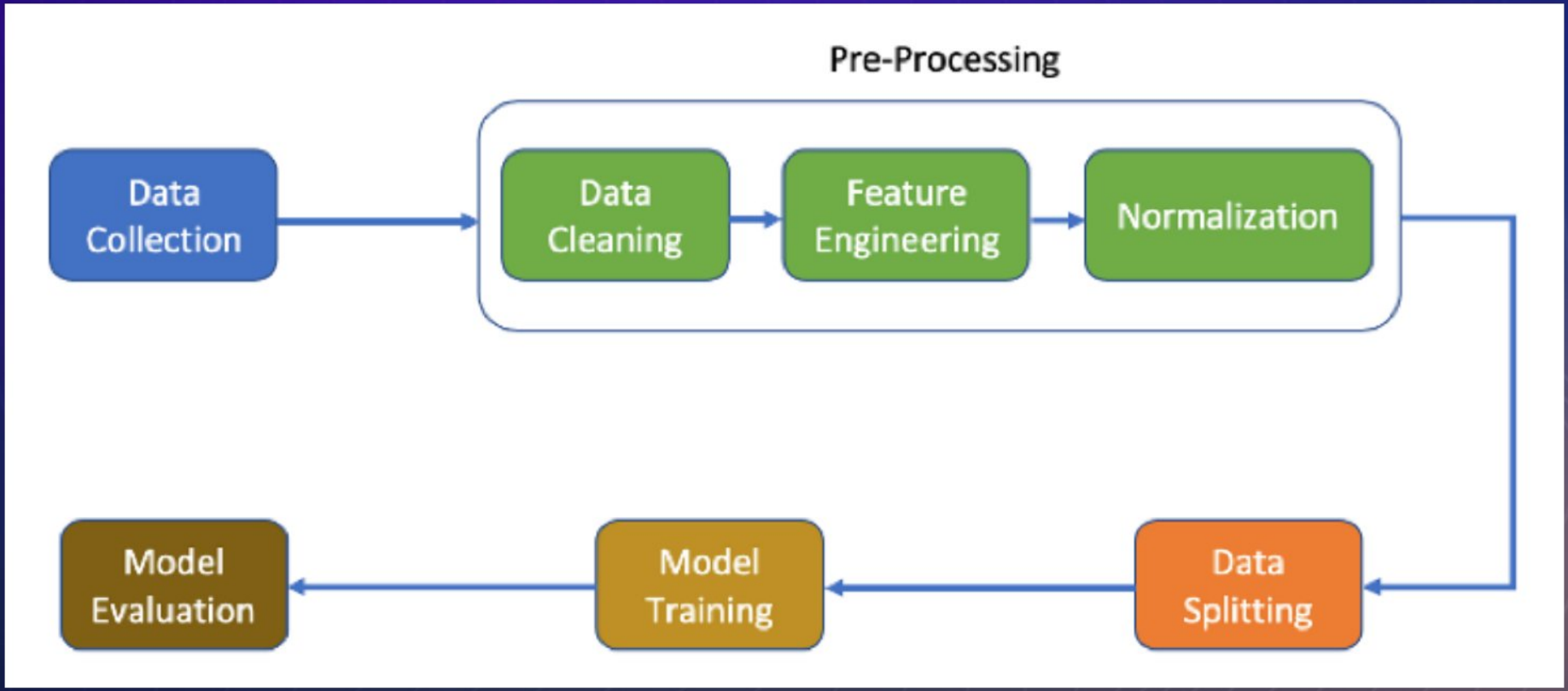


Problem Statement

Explore the key features of earthquake data and design an object for those features, such as date, time, latitude, longitude, depth, and magnitude. Before developing the prediction model, visualize the data on a world map to display a complete overview of where the earthquake frequency will be higher. Split the data into a training set and a test set for validation. Lastly, build a neural network to fit the data from the training set.

Flow of the Methodology



Data Preprocessing

- Several methods to clean the data were implemented to preprocess the data before use.
- The data was preprocessed by handling missing values, then scaling and normalization were done.
- The dates in the data were parsed, character encoding was done and inconsistent data entry was handled.

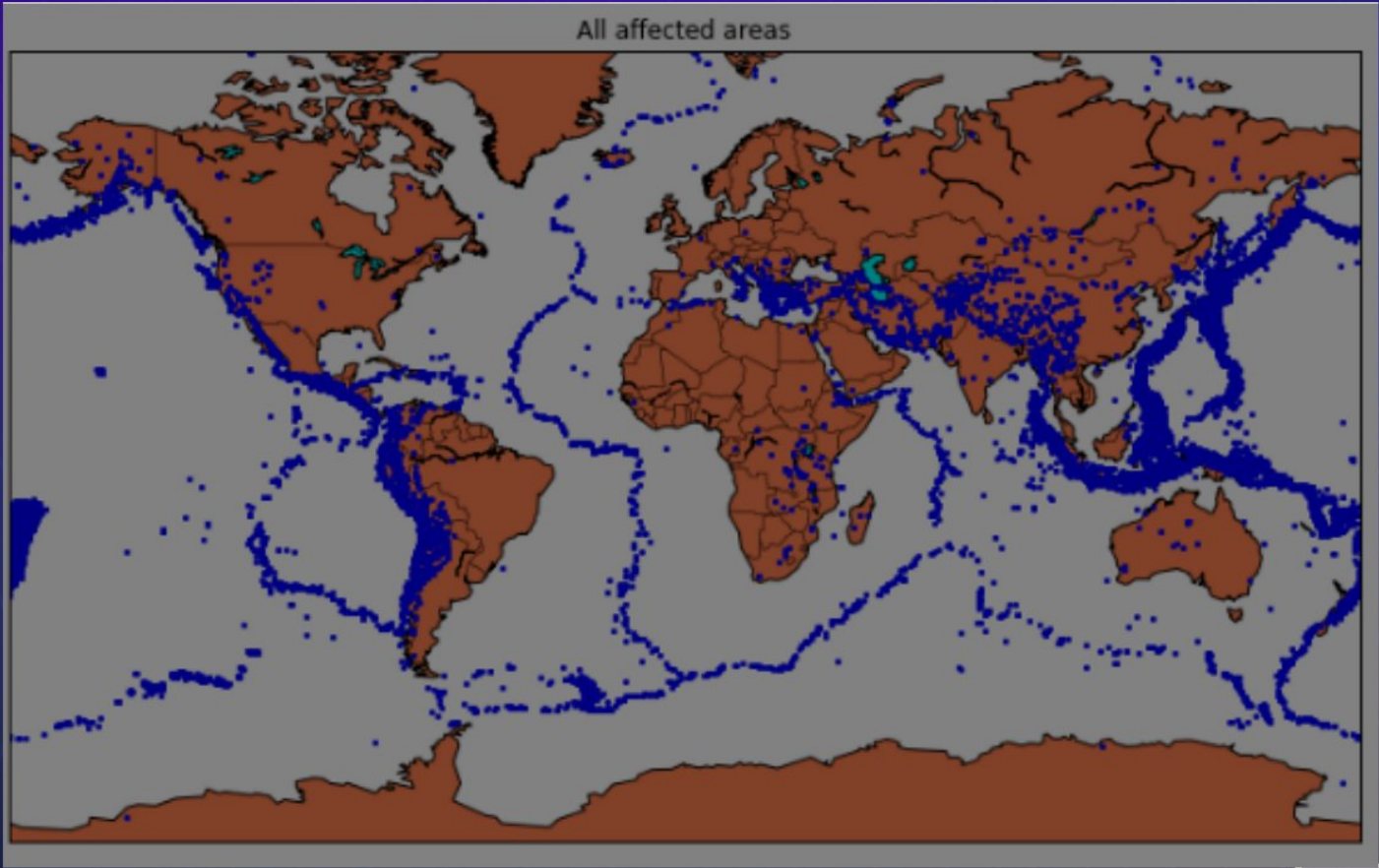
Methodology

- The data was divided (after randomly shuffling it) into two separate sets:
- A training set (80%) - Used for constructing the classification model
- A testing set (20%) - Used for testing the model
- The model was constructed using only the training set, and tested on the testing set

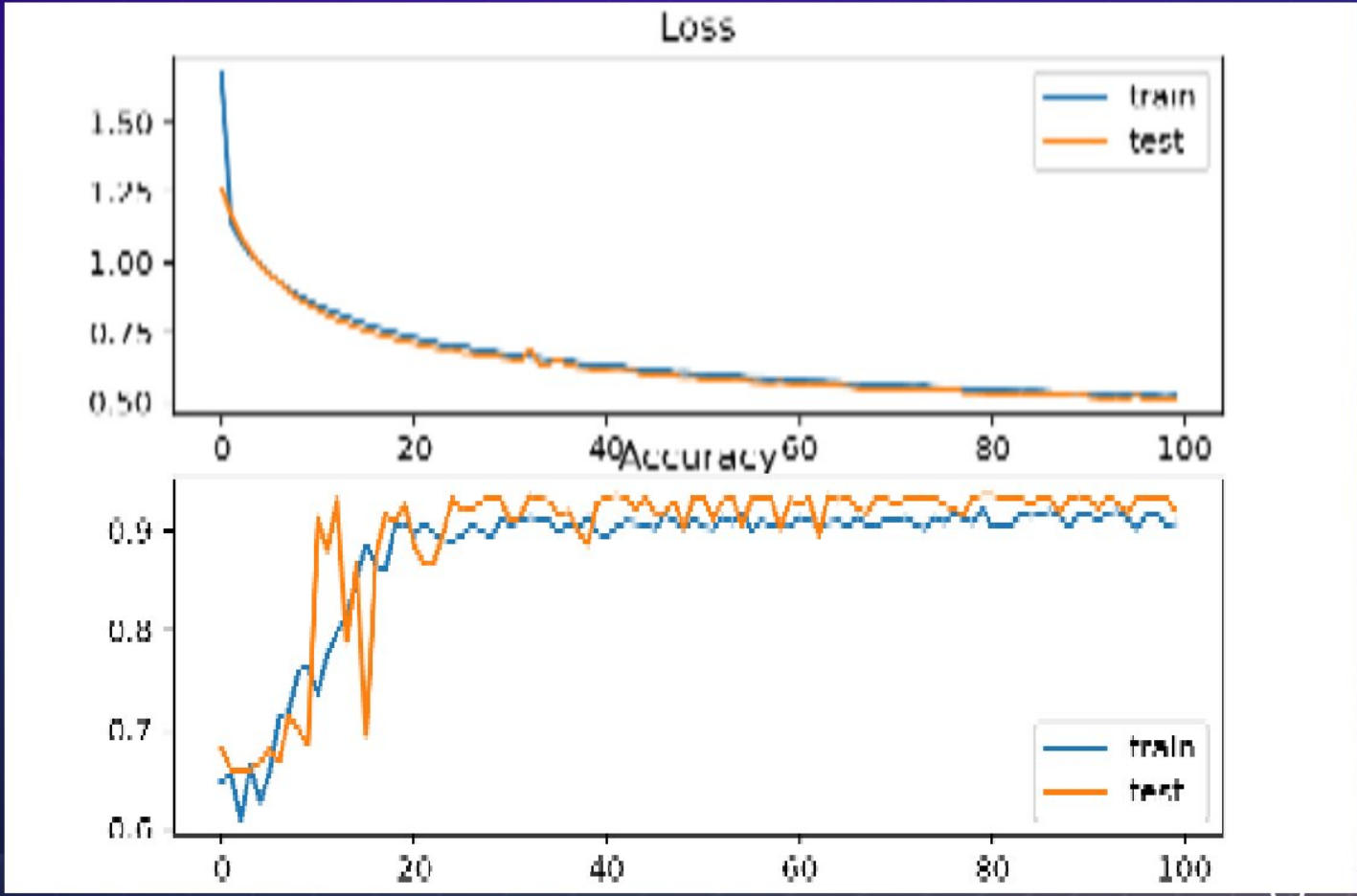
Methodology

- 2. Build a decision tree from the bootstrap sample. On each node:
- a. Randomly select d features without replacement
- b. Split the nodes using the feature that provides the best split according to the objective function.
- 3. Repeat steps 1) and 2) k
- 4. Combine predictions based on each tree to assign a class label based on the most votes.

Data Visualization



Data Model



Conclusion

- We have successfully built a data model for the prediction of earthquakes in a given area with the latitudinal and longitudinal data of the area.
- The model performs well in our test with an accuracy of about 92.41%.
- This could prove to be helpful for people deciding what cities to move into knowing the safety issues in certain areas.