VoyageVista: H-1B Visa Prediction
A Machine Learning Approach

By: Aditya Raj (229310470)

Guided by: Ms. Rishika Singh

Problem Statement

H-1B visa applications face high unpredictability.

Applicants lack a reliable way to estimate their approval chances.

A data-driven predictive model can provide useful decision support.

Project Overview

- **VoyageVista**: A predictive system for H-1B visa approval likelihood.
- Uses machine learning algorithms to analyze historical visa data.
- Web application implementation using Flask.
- Practical tool for applicants to assess their approval chances.

Project Features

Uses Kaggle H-1B dataset (2011–2016) with over 3M records.

Trains multiple ML models: Logistic Regression, Decision Tree, Random Forest.

Deploys best model via Flask-based web app.

Clean, easy-to-use web form for inputs and predictions.

Methodology: Dataset & Preprocessing

- **Dataset:** H-1B visa applications (2011-2016) from Kaggle
- **Key Features:** CASE_STATUS, SOC_NAME, JOB_TITLE, PREVAILING_WAGE, WORKSITE
- **Preprocessing:**
  - ❑ Handling missing values and outliers
  - ❑ Encoding categorical variables
  - ❑ Feature selection based on correlation

Dataset Overview

- Source: Kaggle H-1B Visa Data (2011–2016).
- Records: 3 million+.
- Features: Employer Name, Job Title, Wage, Work Location, Full-time/Part-time.
- Label: Case Status (Certified/Denied).

Machine Learning Models

Three algorithms implemented:

Decision Tree

- Tree-structured classifier with nodes, branches, and leaves
- Handles non-linear relationships

Logistic Regression

- Predicts binary outcome probabilities
- Efficient with linear decision boundaries

Random Forest

- Ensemble of decision trees
- Reduces overfitting through aggregation of predictions

Model Evaluation Results

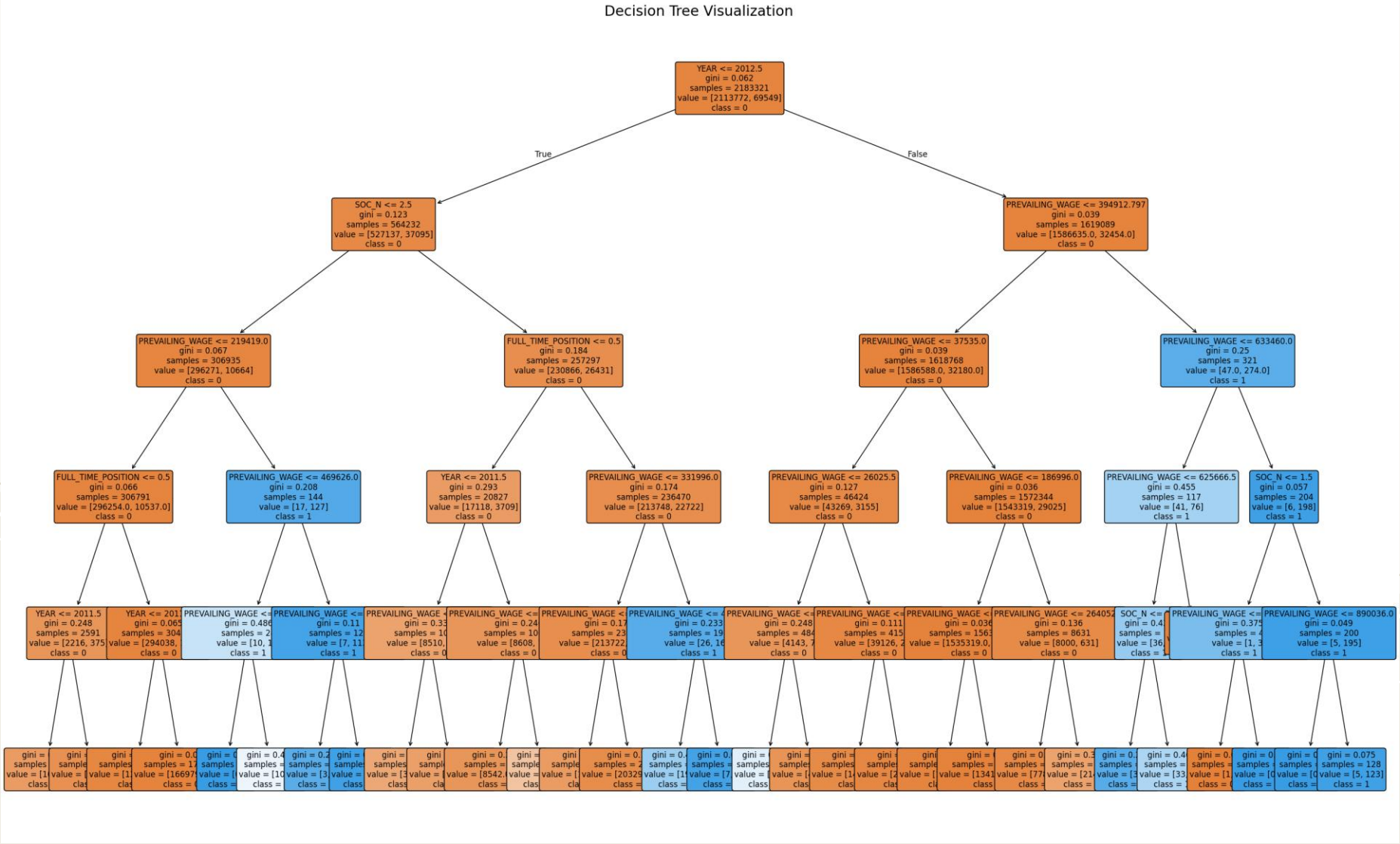
Model Performance Comparison:

- Random Forest: 96.4% accuracy
- Decision Tree: 96.5% accuracy (Selected for deployment)
- Logistic Regression: 39.4% accuracy

Key Metrics for Decision Tree :

- Precision: 95%
- Recall: 97%
- F1-Score: 95%

Decision Tree Visualization



System Architecture

FRONTEND: HTML, CSS, JAVASCRIPT FOR USER INTERFACE

BACKEND: FLASK WEB SERVER FOR HANDLING REQUESTS

ML COMPONENT: SCIKIT-LEARN MODELS FOR PREDICTION

DATA FLOW: USER INPUT → FLASK → ML MODEL → PREDICTION → USER

Limitations & Future Work

- Data restricted to 2011–2016 applications.
- Prediction accuracy depends on available features.
- Advanced ML/DL algorithms for higher accuracy
- Incorporating recent visa application data
- Expanding to other visa categories
- Multi-country visa prediction system
- Explainable AI integration for transparency

Conclusion

- VoyageVista successfully predicts H-1B visa approval with 89.8% accuracy
- Web application deployed with user-friendly interface
- Practical tool for visa applicants to improve application strategy
- Foundation established for future enhancements

Thank You