

Project Initialization and Planning Phase

Date	15 March 2024
Team ID	LTVIP2024TMID25012
Project Title	Predictive Modeling for H1b Visa Approval Using Machine Learning.
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	The primary objective of this project is to develop a machine learning model to predict the approval or denial of H1B visa applications. The model will help HR professionals and immigration attorneys streamline decision-making processes and improve accuracy in anticipating visa outcomes.
Scope	The project involves data collection, preprocessing, model development, evaluation, and deployment of a predictive model for H1B visa approval. The scope includes the creation of a web interface where users can input visa-related information and receive predictions in real time.
Problem Statement	
Description	The current H1B visa approval process is complex and lacks transparency, with multiple factors affecting approval decisions. Without an accurate method to predict visa outcomes, companies face inefficiencies in resource allocation and recruitment planning.
Impact	Solving this problem will allow companies to make data-driven decisions regarding hiring and visa application processing, reducing uncertainty and optimizing resources.

Proposed Solution	
Approach	The proposed solution is a machine learning-based predictive model. The model will use a dataset of past H1B visa applications and train a RandomForestClassifier to predict the approval or denial status of future applications. Key steps include data preprocessing, feature selection, model training, evaluation, and deployment.
Key Features	<ul style="list-style-type: none"> • Predicts visa approval or denial based on key application attributes (e.g., job role, wage, full-time position). • Uses Random Forest for classification. • Integrated with a Flask-based web application for user interaction. • Provides real-time predictions once deployed to a cloud platform.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 8 GB
Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		
Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	e.g., scikit-learn, pandas, numpy
Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset, CSV format, 3002458 rows