

**Project Initialization and Planning Phase**

|  |  |
| --- | --- |
| Date | 10 July 2024 |
| Team ID | 739836 |
| Project Title | Predictive Modelling for H1b Visa Approval Using Machine Learning |
| Maximum Marks | 3 Marks |

**Project Proposal (Proposed Solution) template**

Our project aims to leverage machine learning to predict the approval likelihood of H-1B visa applications, addressing the critical need for transparency and efficiency in the visa approval process.

|  |  |
| --- | --- |
| **Project Overview** |  |
| Objective | To predict the likelihood of visa approval based on application data, identify key factors influencing the decision, and provide insights to improve the application process. |
| Scope | This includes collecting and preprocessing H-1B visa application data, selecting relevant features, training and evaluating predictive models, analysing key factors influencing approvals, and deploying the model for practical use. |
| **Problem Statement** |  |
| Description | The H-1B visa is a non-immigrant visa that allows U.S. companies to employ foreign workers in specialty occupations that require theoretical or technical expertise, such as in IT, engineering, and science. The visa is typically issued for up to three years, with the possibility of extension. This project involves developing a machine learning model to predict H-1B visa approval outcomes. It includes data collection and preprocessing, feature selection, model training and evaluation, and deployment. The aim is to identify key factors affecting approval decisions and provide actionable insights to improve application success rates. |
| Impact | It includes increased H-1B visa approval rates, data-driven insights for better decision-making, a more efficient application process, and informed policy development. |
| **Proposed Solution** |  |
| Approach | The approach includes gathering and cleaning H-1B visa application data, selecting relevant features, training machine learning models, evaluating their performance, and deploying them for practical use. |
| Key Feature | Application Attributes: Features such as job title, salary, employer details, and applicant qualifications.  Historical Data: Utilizing past visa application outcomes to train models.  Model Selection: Choosing appropriate machine learning algorithms like logistic regression, random forests, or neural networks.  Evaluation Metrics: Using accuracy, precision, recall, and F1-score to assess model performance.  Interpretability: Ensuring models provide understandable insights into factors influencing visa approval. |

**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 , sklearn , metrics |
| Libraries | Additional libraries | e.g., scikit-learn, pandas,  numpy |
| Development Environment | IDE, version control | e.g., Jupyter Notebook, Git , Google collab ,visual studio |
| **Data** |  |  |
| Data | Source, size, format | e.g., Kaggle dataset, 500 images , CSV |