Model Development Phase Template

Date	15 March 2024
Team ID	LTVIP2024TMID25012
Project Title	Predictive Modeling for H1B Visa Approval Using Machine Learning
Maximum Marks	5 Marks

Feature Selection Report Template

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

Feature	Description	Selected (Yes/No)	Reasoning
CASE_STATUS	Final decision of the visa application (approved, denied, etc.)	**No**	This is the target variable and not a feature.
EMPLOYER_NAME	Name of the employer sponsoring the visa	**No**	Employer names are too specific and do not provide predictive value for visa approval. The category is too large and can introduce noise into the model.

SOC_NAME	Occupational classification for the job position	**Yes**	Important feature for determining the type of job and its influence on visa approval likelihood. Different occupations can have varying approval rates.
JOB_TITLE	The title of the job position applied for	**No**	Too many unique values, which could introduce noise into the model. Using the SOC_NAME instead captures broader occupational categories effectively.
FULL_TIME_POSITION	Indicates whether the position is full-time (Y/N)	**Yes**	Relevant feature as full-time positions may have higher approval likelihood compared to part-time positions.
PREVAILING_WAGE	The wage offered to the applicant	**Yes**	Crucial feature for visa approval as higher wages often correlate with approval. Helps determine if wage competitiveness plays a role in approval.

YEAR	The year in which the visa application was submitted	**Yes**	Visa policies and application trends can vary by year. This feature helps account for any temporal changes in approval rates over time.
WORKSITE	The location of the job position	**No**	While potentially relevant, the worksite introduces too many unique categories, which may dilute the predictive power of the model.
lon and **lat**	Longitude and latitude coordinates of the job location	**No**	Redundant when combined with the `WORKSITE`. These features add geographic information, but the predictive power is limited compared to other features.