BIA 678
Team Project Proposal

H-1B Labor Condition Application Approval Prediction

Team No: CD-10

Team Members:

1. Akshat Bharatkumar Jain *(CWID: 2000 8931)*
2. Jaykumar Dineshkumar Patel *(CWID: 2000 8512)*
3. Juilee Sameer Thakur *(CWID: 2000 8598)*
4. Prathmesh Desai *(CWID: 1047 4461)*

Description:

Labor Condition Application is a required document that the H1B Sponsor/Employer need to file with the US Department of Labor (DOL) before they file the H1B petition with USCIS for any non-immigrant worker. One of the first stages to obtaining an H1B work visa in the US is having an LCA approved. This project shows how different categorization models are implemented and their performance is compared to forecast the outcome of an LCA application. The study also discusses how different models perform when used on local machines and the cloud.

We are planning to use the dataset from Data World, which includes 528000 observations and 40 columns regarding H-1B Labor Condition Approval applications submitted in 2017. Our aim is to identify correlations between labels and other features in the dataset and then draw conclusions using various models.

Dataset link: https://data.world/ian/h-1-b-disclosure-data-fy-17/workspace/file?filename=H-1B_Disclosure_Data_FY17.csv

Our objective is to implement Decision Tree, Logistic Regression and Naïve Bayes Classifier and their performance will be evaluated to make a final decision based on the same.

We are going to implement the same using PySpark.