Final Project Topic Proposal

Team Global: Sajan Kumar Kar, Sandhya Karki, Sai Pavan Mekala, Neelima Puthanveetil

We continue our analysis on the "Employability Classification of Over 70,000 Job Applicants" dataset, employing advanced machine learning techniques to construct a binary classification model and fine tune it to optimum accuracy in order to help us predict whether an individual will be employed or not in today's highly competitive job market. We experiment with a comprehensive suite of distinct machine learning algorithms such as, logistic regression, decision tree, random forest, and k-nearest neighbors model. We use the receiver operating characteristic curve and its AUC value as our model selection criteria.

SMART Questions:

1. Is there a significant difference between employed males and non-males in their respective education levels?
2. Does mental health influence previous salary?
3. Can we predict with at least 60% accuracy whether an individual will be employed or not, based on just their age, education level and number skills?
4. Is there a notable difference between distance based models and tree based models?

Dataset Link:

https://www.kaggle.com/datasets/ayushtankha/70k-job-applicants-data-human-resource/data

Github:

https://github.com/NemaPuthanveetil/DATS6101-Team-Global-Project