Machine Learning CS539 - Fall'19

Project Proposal- Group 7

H1B Case Status Prediction Model

Kashyap Bhuva, Varun Eranki, Vishaal Prabhakar, Arjun Rao

Project Objectives:

The objective of this project is to predict the status of an H1B petition and provide insights to convert a denial into an acceptance. Our objective is to implement a machine learning model that will be useful for employers seeking foreign talent in order to make their case stronger and avoid denials. The insights from this project would also enable foreign nationals to target the right companies and locations so as to increase the chances of H1B acceptance.

Background:

The US economy thrives on the immigrants and foreign workers from across different parts of the world. The H1B Visa Programme is one of the major sources of labor pool for the specialist occupations in which there is a talent shortage. The Office of Foreign Labor Certification (OFLC) discloses the H1B petitions data publicly and is updated annually. H1B visas fall in a category of employment-based, non-immigrant visa for temporary foreign workers in the United States. A US employer has to offer a job to a foreign worker before applying for the H1B petition to the immigration department.

Dataset Description:

The dataset consists of 6,54,348 rows and 52 columns. The target variable is case status which consists of 4 distinct values namely Certified, Certified-Withdrawn, Denied and Withdrawn. Source: https://www.kaggle.com/abishekanbarasan1995/h1b-case-status-prediction#H-1B_Disclosure_RAW_Data_FY18.csv

Functionality and Beneficiaries of this Tool:

The main functionality of this tool is to not just predict the status of H1B petition of an individual but also give further insights if a denial of H1B is predicted. The idea is to implement a classification machine learning algorithm to predict the outcome given a few input parameters. As a part of the implementation, these parameters can be tweaked to change the status and assist the applicant in the visa process.

Just like most of the students in our class, our sole aim after graduation is to land in a good company with H1B sponsorship. Many students and graduates from other universities across USA strive for the same. By using a tool like this, one can note the trends of candidates whose visa has been certified and make necessary changes to their application profile to get H1B certification. Thus, this tool will be very helpful for candidates filing for H1B and also for companies looking to hire prospective candidates.

The Uniqueness of this Tool:

Although solutions for this problem statement have been proposed previously, they lack accuracy due to the ever-changing immigration visa climate and the existing solutions do not provide

substantial insight into the final decision given to an applicant. Our model can be used to suggest changes to the applicant's profile by zeroing in on the reasons behind an applicant's H1B visa rejection. It can also be told that this ever-changing immigration scenario is a key reason why building an accurate version of this tool is a difficult challenge.

The Methodology for Building the Tool:

We plan to build the tool using supervised machine learning algorithms for the binary classification. We would clean the data, engineer some features before giving the input to the algorithm. The data we are using is the publicly available H1B case status dataset for 2018. We will try out different machine learning algorithms and implement the one which will give the best performing model.

Machine Learning Model Validation:

The tool will predict the case status for an H1B petition and the probability of rejection. We would test the model on the new data to demonstrate how accurately it could predict a rejection. Once the model is implemented, we presume a significant upsurge in the number of accepted H1B petitions.

Literature:

International students have the opportunity to work under F1 visa by applying for the Optional Training Program(OPT jobs), but those who want to work in the USA require to file for an H1B petition and find companies willing to sponsor them to work in the USA.

When does an H1B petition get rejected? There are quite a few reasons for this. Some of them are:

1. When petitioning employers' fail to meet the requirement
2. The position is not the position the petitioner claims for
3. The petitioning employer will not be the true employer

There are few other reasons but these are some of the important ones. What we aim to do is look at these reasons and by ranking the reasons according to their importance, predict how a person can get their H1B petition approved. We plan on ranking these reasons by assigning weights to each factor. The importance of a factor is determined by checking how crucial a role it plays in determining the status of the petition.

Our dataset is highly imbalanced since we have 88% of the petitions as approved and 12% as rejected. We must, therefore, be strict while assigning weights to our parameters and mainly focus on the rejection data.

Resources:

The most basic resource is **Python3.** We will be using multiple python libraries such as:

1. **scikit-learn** for machine learning algorithms
2. **pandas** for importing csv files and preprocessing our data
3. **NumPy** for scientific computing
4. **TensorFlow**, If we implement neural networks
5. **matplotlib** to analyze the data and plot our results
6. Apart from this, we will be looking at research papers that focus on classification in order to yield results that are up to date with the existing state of the art technology.