26-06-2023

Predictive Modeling For H1b Visa Approval Using IBM Watson

SMARTINTERNZ EXTERNSHIP

- PRANIT FARKADE-20BEE0373
  (VELLORE CAMPUS VIT)
- YASH KHETAN-20BEE0381
  (VELLORE CAMPUS VIT)
- AMAN YADAV-20BEE0350
  (VELLORE CAMPUS VIT)
- ADITYA RAJ-20BEE0209
  (VELLORE CAMPUS VIT)

1. INTRODUCTION

1.1 OVERVIEW

Our objective in the Guided Project is to forecast the outcome of the numerous professional overseas applicants who submit H-1B visa applications each year. In this instance, we employed it to frame the issue as a classification problem and to produce a forecast case status for the application. The applicant's characteristics serve as the input for our system. In the United States, a non-immigrant visa called an H-1B enables foreign people to work in professions that call for specialised knowledge and a bachelor's degree or higher in the relevant field.

1.2 PURPOSE

Before submitting an application to the USCIS for this visa, the applicant must already have an employment offer from a company in the US. We think that this prediction algorithm could be a helpful tool for prospective H-1B visa candidates as well as potential employers.

We will feed the model with the dataset including the necessary fields by which the machine may classify the case status as certified or refused in order to predict the case status of the applicants.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

The uncertainty and lack of openness in the selection process are two current issues with the H1B visa prediction procedure. There is a yearly cap on the amount of H1B visas given due to the program's high demand. However, the random lottery mechanism used to choose H1B visa beneficiaries can be annoying for both companies and potential employees.

The existing approach does not consider elements like the requirements, abilities, or demand for particular employment roles. As a result, highly qualified persons who would considerably boost the American economy might not be given visas, while people with less impressive credentials might get them by accident.

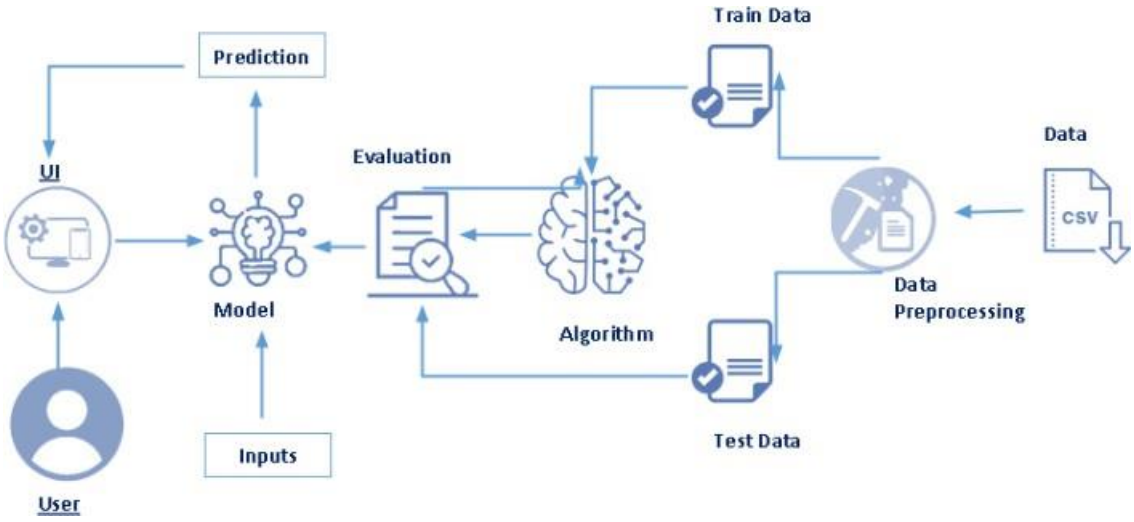
Employers who need to plan their staff and find the best employees face hurdles as a result of this uncertainty. Additionally, it causes aggravation for those who put forth time, energy, and money into the visa application procedure just to have a random drawing determine their prospects.

2.2 PROPOSED SOLUTION

We developed this Visa Approval status prediction by using the Python language which is a interpreted and high level programming language and using the Machine Learning algorithms. for coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming in the python language. For creating an user interface for the prediction we used the Flask. It is a micro web framework written in Python. It is classified as a micro frame work because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions, and a scripting language to create a webpage is HTML by creating the templates to use in th functions of the Flask and HTML..

3. THEORITICAL ANALYSIS

3.1 Block diagram



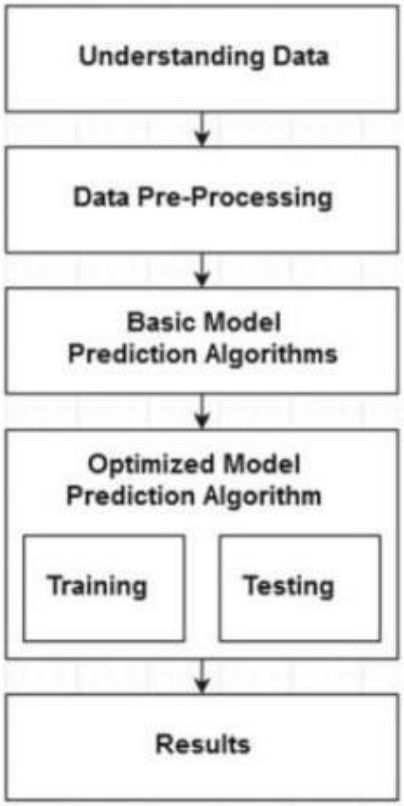
3.2 SOFTWARE REQUIRED

1. Jupyter Notebook Environment
2. Spyder Ide
3. Machine Learning Algorithms
4. Python (pandas, numpy, matplotlib, seaborn, sklearn)
5. HTML
6. Flask

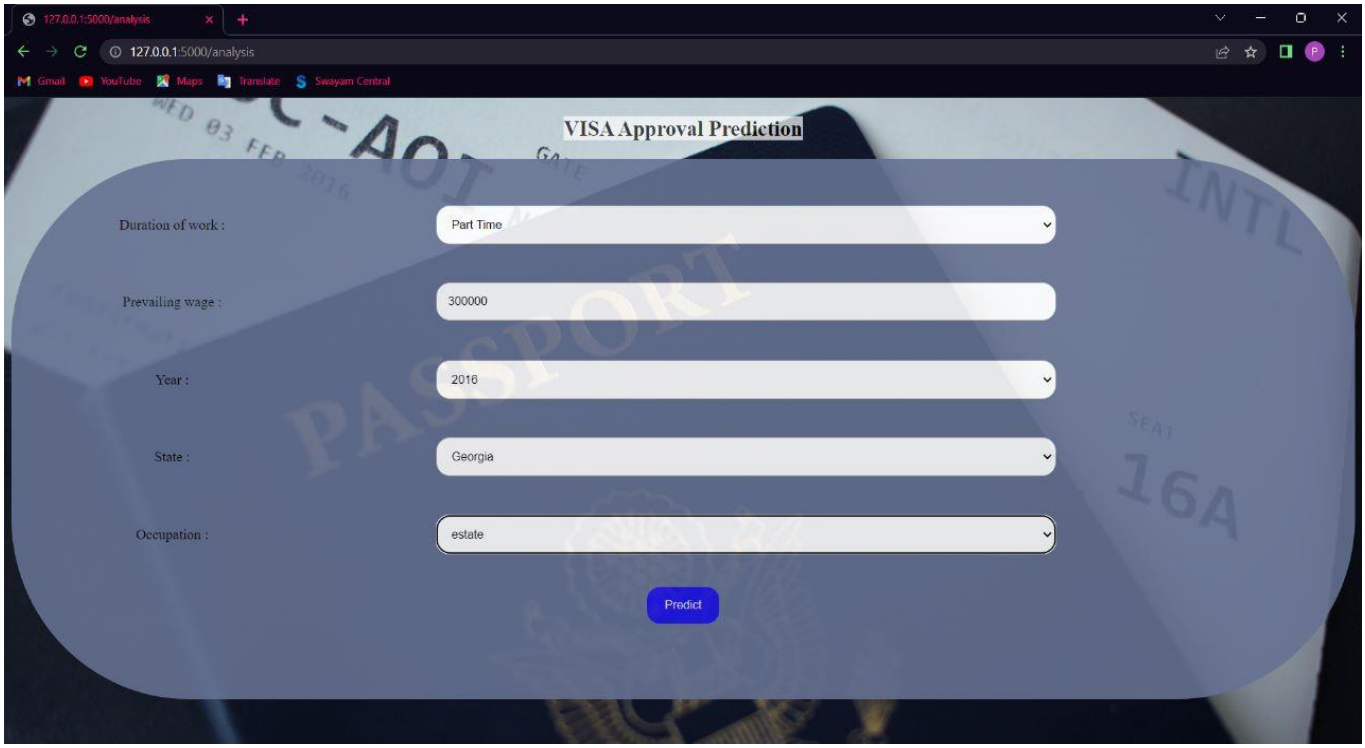
4. EXPERIMENTAL INVESTIGATIONS

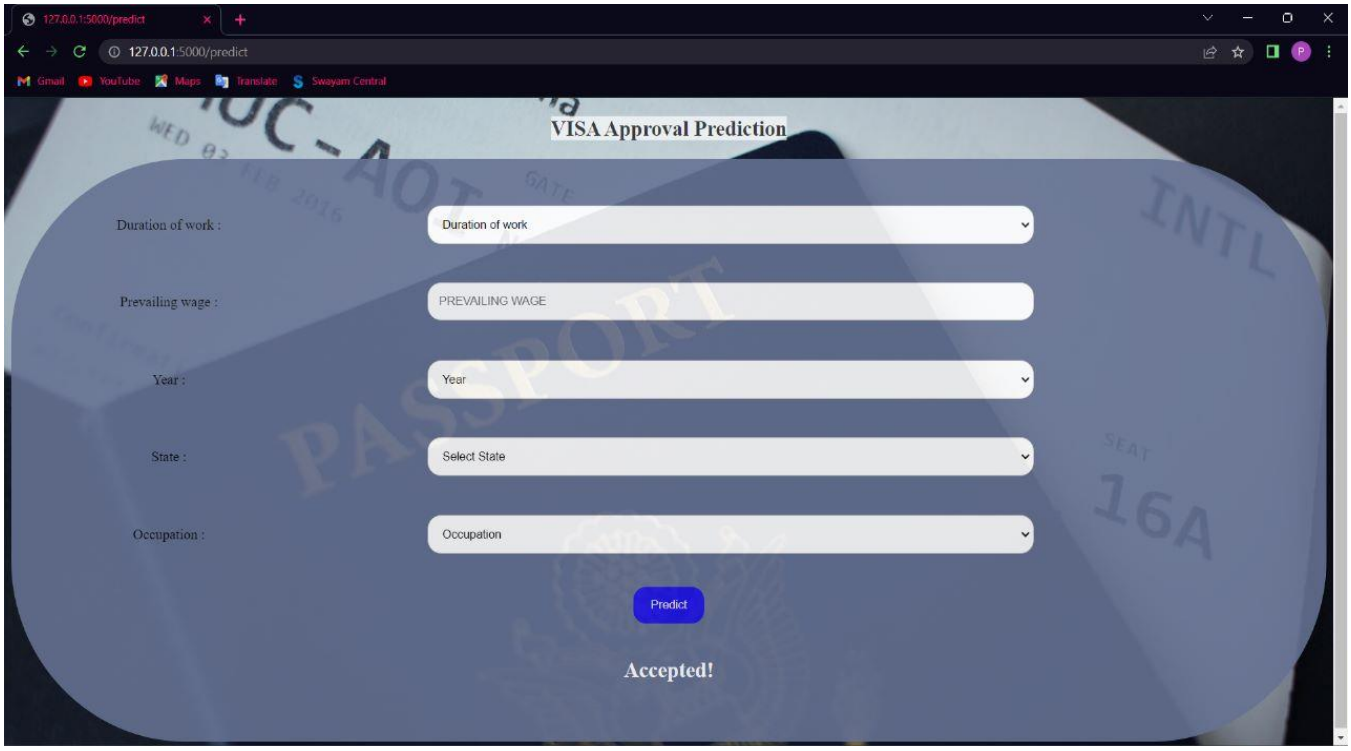
The dataset we used for this study was taken from H-1B_Kaggle.More than 10L H-1B Visa user records are contained there.It had one label and seven traits that could be analysed as attributes. The screenshot of the data set we utilised below illustrates these characteristics. CONTEXT STATUS: We did not include the cases 'CERTIFIED-WITHDRAWN' and 'WITHDRAWN' since 'WITHDRAWN' decisions are either made by the petitioning employer or the applicant, therefore they are not indicative of USCIS's future actions. 'CERTIFIED' cases received a label of 1 while 'DENIED' cases received a label of 0. Positions are listed with the notation "Full Time Position = Y; Part Time Position = N". They were formatted as "Full Time Position = 1; Part Time Position = 0". YEAR: The year that the application was submitted. The data was transformed into a one-hot-k format. PREVAILING WAGE: Prevailing wage is the typical salary given to workers with comparable credentials in the desired field of employment. The remainder of the data was used without the outlier terms. APPLICATIONS PER EMPLOYER_NAME: We built a feature to count the number of H-1B applications submitted by each employer. We eliminated data points from petitions submitted by employers with fewer than four applications. This processing phase unavoidably discards applications submitted by small businesses, however it greatly aids in correcting misspelt company names. For the success rate by employer, a functionality was developed. Apps by Social Name: The federal occupational classification system known as SOC stands for Standard Occupational Classification System. We removed data points with SOC kinds that appear less than four times in the data and built a feature for the quantity of H-1B applications by SOC type. Undesirably, this processing stage eliminates applications for jobs that are unusual, although it aids in cleanup. WORKSITE: The format for the data is "City, State". We reduced the data to a one-hot-k representation and only included "State". Following the aforementioned pre-processing processes, we divided the training, A total of 1.2 million examples made up the training set. We established two versions and test sets in order to make sense of the error analysis later on because our dataset has a built-in bias towards the "CERTIFIED" label. The initial development and test collections, which each had 400K samples, were also unbalanced. More specifically, to mimic the original dataset, almost 90% of the samples had a "CERTIFIED" label. By manually picking an equal number of "CERTIFIED" labelled examples as "DENIED" labelled cases, the second version of the development and test sets were balanced.

5. FLOWCHART



6. RESULT





7. ADVANTAGES AND DISADVANTAGES

Advantages:

Efficiency: Machine learning models are capable of swiftly and efficiently processing and analysing enormous amounts of data. This can expedite the visa

prediction process considerably and reduce waiting times for both applicants and immigration officials.

Objectivity: Machine learning algorithms are made to be unbiased and base predictions on data patterns rather than human prejudice. By doing this, possible biases in the decision-making process may be reduced or eliminated.

Increased precision: Machine learning models can learn from past visa outcomes and spot trends that human analysts might miss. This may result in more precise forecasts, which might raise the overall success rate of visa applications.

Scalability: Machine learning models may be quickly expanded to process many visa applications at once, enabling effective case processing in high volume settings.

Disadvantages:

Data restrictions: The quality and availability of training data have a significant impact on how accurate machine learning models are. The predictions could be less accurate if the previous data that was used to train the model is inaccurate, biassed, or out of date.

Dynamic policy shifts: Immigration laws and regulations are subject to frequent modification. In circumstances when the policy landscape has dramatically changed, machine learning models based on historical data may not account for these policy changes, leading to erroneous predictions.

Lack of transparency: Interpreting certain machine learning algorithms, such as intricate deep learning models, can be difficult. This lack of transparency can make it challenging to comprehend how the model came to a specific forecast, which may raise questions about fairness and accountability.

Ethics: Machine learning models may unintentionally reinforce biases found in training data, resulting in biassed results. Fairness, accountability, and openness must be carefully considered while designing and implementing these approaches.

Context and human judgement: Machine learning algorithms might not fully capture the specifics and context of each individual visa application. They frequently only evaluate quantitative data, ignoring qualitative elements that human reviewers may think are crucial when deciding whether to grant a visa.

8. APPLICATIONS

- We will be able to predict whether our application will be approved or not.
- We will be able to predict using different prediction model.
- Different prediction models accuracy will be checked and implemented according to the need.

9. CONCLUSION

The selection for H visas has increased every year over the past ten years, so the goal of this challenge is to develop a tool that will inspire every person going through the H1B visa application process and accurately forecast whether or not their application will be accepted. Additional details about the Standard Occupational

Classification (SOC) can be combined and put into practise in accordance with the H-1B Visa preference procedure. The income feature on this data set can be accurately mounted to a range of incomes using the income opinions and levels under SOC. From there, it is possible to classify visa applications according to career positions and geographic preference. Additionally, beautiful beauty algorithms that are not discriminative models can be tested with this method, and their results can also be evaluated. In comparison to all of the fantastic algorithms that can be gifted to carry out the evaluation activities, the Random Forest classifier works right right right here with the superior accuracy. Here, the Random Forest Algorithm was used to enhance and enrich the data, and as a result, we were given an accuracy of 83.06 percent. As a result, this set of recommendations is the

best available for the prediction of H1B visa approval. It has become necessary to build a tool to accurately track down H1B visa approval due to the rising trend of H1B visa applicants. Therefore, we are able to seek for the H1B visa approval rate using the useful assistance of a number of device learning sophistication algorithms. For international workers coming to the US, this can be highly advantageous.

10. FUTURE SCOPES

In in addition Random Forest set of suggestions may be implemented on wonderful information gadgets to be had for visa approvals to in addition take a look at its accuracy. A rigorous evaluation of diverse gadget learning algorithms other than those six moreover can be completed in destiny to research the energy of gadget learning algorithms for visa reputation prediction. In in addition take a look at, we can attempt to behavior experiments on big information gadgets or attempt to tune the version which will benefit the kingdom -of-paintings regular basic overall performance of the version and a great UI assist tool making it entire internet software In the destiny we can attempt to use a few greater strategies and strategies in-order to are looking for the recognition of H1B visa we can look at several possible answers to predict the candidate's Finally, we can attempt to achieve the awesome feasible answer.

11. BIBLIOGRAPHY

[1].“H-1B Fiscal Year (FY) 2018 Cap Season,” USCIS. [Online]. Available: https://www. uscis.gov/going for walksunited-states/temporary-employees/h-1b-specialty-occupationsand-fashion-fashions/h-1b-fiscal-365 days-fy-2018- cap-season. [2].The highly professional visa programs had an immeasurable impact on the relationship,” CNNMoneysaid. [In line]. Available: http://money.cnn.com/2016/04/12/technology/h1b-cap-visa-fy-2017/index.html. [3]."Use Text Analytics to Predict H1B Salaries", BigML.com Official Blog, October 1, 2013. [Online]. Available: https://blog.bigml.com/2013/10/01/the use of-textual content-analysisto-predicth1-b-wages/. [4].Predicting the Case Status of H1B Visa Applications". [In line]. Available: https://cseweb. ucsd.edu/commands/wi17/cse258- a/reports/a054.pdf. [5]. Mallikarjuna Reddy, V. Venkata Krishna, L.Sumalatha, "Facial recognition based primarily on Full Diagonal Cross Matrix", International Journal of Image, Graphics and Signal Processing (IJIGSP),Vol. 10,No.3, pp. 59-66, 2018.DOI: 10.5815/ijigsp.2018.03.07. [6]. Ayaluri MR, K. SR, Konda SR, Chidirala SR. 2021. Efficient analysis of the use of a community of vehicular convolutional cellular encoders to ensure adequate image quality. Peer J Informatica 7: e356 https://doi.org/10.7717/peerjcs.356 [7] A Mallikarjuna Reddy, Vakulabharanam Venkata Krishna, Lingamgunta Sumalatha and AvukuObulesh, “Age Classification Using Motif and Statistical Features Derived on Gradient Facial Images”, Recent Advances in Computer Science and

Communications (2020) 13: 965. https://doi.org/10. 2174/2213275912666190417151247. [8].Predicting the Case Status of H1B Visa Applications". [In line]. Available: https://cseweb. ucsd.edu/commands/wi17/cse258-a/reports/a054.pdf. [9].gov/going for walks-unitedstates/temporary-employees/h-1b-specialty-occupationsand-fashionmodels/h-1b-fiscal-365 days-fy-2018-cap-season. [Accessed: 20-Oct-2017]. [10] Swarajya Lakshmi V Papineni, SnigdhaYarlagadda.