Predictive Modeling for H1B Visa Approval Using Machine Learning

[1]C. Kavitha Santhoshi, [2]G. Srilekha, [3]B. R. Ramadevi, [4]K. Chandrika

[1,2,3,4] UG Student, Department of Computer Science Engineering,

Dr K V Subba Reddy College Of Engineering For Women, Kurnool, Andhra Pradesh, India

Abstract

Employers in the United States can temporarily hire non-immigrant workers with the H-1B visa. People with a bachelor's degree or work experience equivalent to it are eligible for this visa, which only allows specialty workers to work in the United States. The H-1B visa is valid for three years, but it can be extended to six years. Although the H-1B visa is the most sought-after in the world, its approval rate is low. In 2019, a total of 2,000 people worldwide applied for the visa, but only 85,000 were selected, resulting in a 42% approval rate. As the US economy improves, this visa battle becomes more competitive. The employer, salary, and other factors all play a role in this decision. This strategy can be used by both the individual and the employer between applying for the visa and receiving the final decision to be informed of the outcome before it occurs because the H1-B visa category is one of the most highly sought-after ones. Taking into account all relevant factors, this project aids in predicting whether or not an individual will be granted the H1B visa. Using the Random forest algorithm, the proposed system achieved high accuracy.

1. Introduction

Employers in the United States can temporarily hire non-immigrant workers with the H-1B visa. People with a bachelor's degree or work experience equivalent to it are eligible for this visa, which only allows specialty workers to work in the United States. The H-1B visa is valid for three years, but it can be extended to six years. Although the H-1B visa is the most sought-after in the world, its approval rate is low. In 2019, a total of 2,000 people worldwide applied for the visa, but only 85,000 were selected, resulting in a 42% approval rate. As the US economy improves, this visa battle becomes more competitive. The employer, salary, and other factors all play a role in this decision. This strategy can be used by both the individual and the employer between applying for the visa and receiving the final decision to be informed of the outcome before it occurs because the H1-B visa category is one of the most highly sought-after ones. Taking into account all relevant factors, this project aids in predicting whether or not an individual will be granted the H1B visa. Using the Random forest algorithm, the proposed system achieved high accuracy.
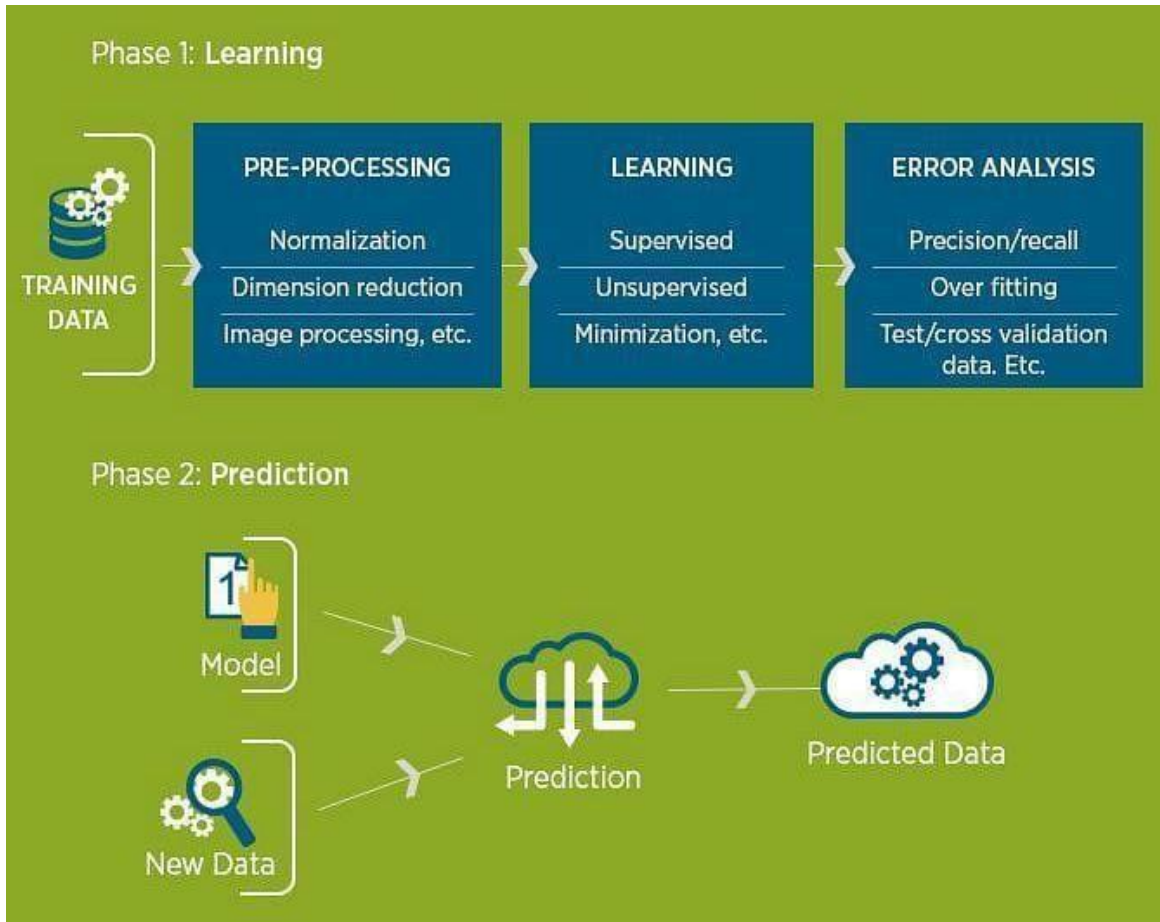
Fig.1 Machine Learning Process

2. Literature Review

The training data in unsupervised learning are unknown and unlabeled, indicating that no one has previously examined the data. The unsupervised term comes from the fact that the algorithm cannot be guided to the input without the aspect of known data. This data is used to train the model and fed into the Machine Learning algorithm. The trained model tries to find a pattern and respond in the right way. In this instance, it frequently appears as though the algorithm is attempting to crack code, much like the Enigma machine, but with a machine rather than a human brain. Here, we will use the Random Forest Classifier to construct a class model. A set of rules for supervised research is random forests. It is applicable to both class and regression. Additionally, it is the most adaptable and user-friendly set of regulations. A random wooded area is a meta estimator that uses averaging to improve predictive accuracy and control over-fitting by matching a number of selected tree classifiers to various dataset subsamples. Random forests get their prediction from each tree and vote to select the best solution, creating choice timber from randomly selected facts samples. The requirement is to provide you with novel features entirely based on the dataset's useful statistics. It is essential to remember to avoid correlated features in the not-yet-determined future of this manner. Only the data in the dataset must be decorated by each characteristic
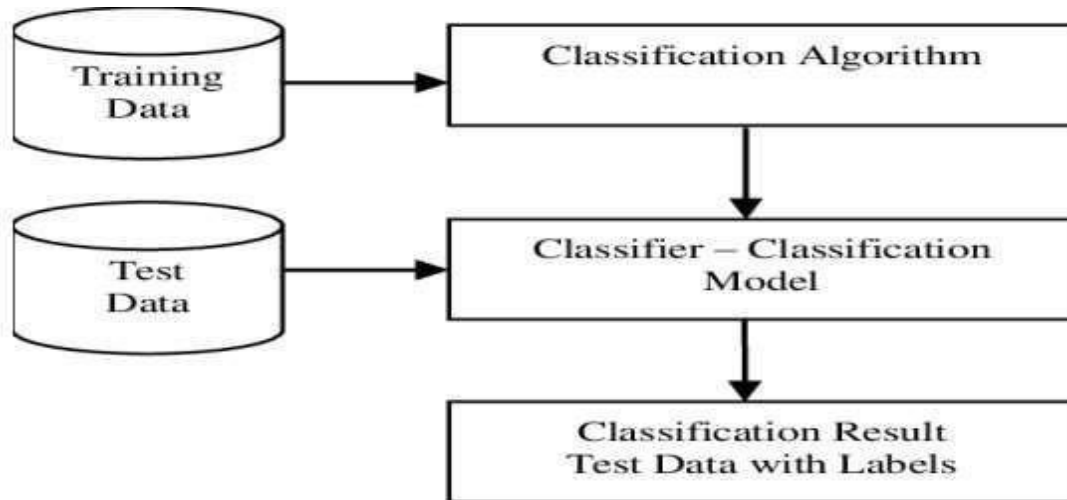
111

Fig.2 H1B visa approval system

3. Proposed System

Anaconda is a scientific computing distribution of the Python and R programming languages that aims to make package management and deployment simpler. Packages for data science that are suitable for Windows, Linux, and macOS are included in the distribution. Anaconda, Inc., which was established in 2012 by Peter Wang and Travis Oliphant, is responsible for its development and upkeep. It is also known as Anaconda Distribution or Anaconda Individual Edition because it is a product of Anaconda, Inc. Other products from the company include Anaconda Team Edition and Anaconda Enterprise Edition, neither of which are free.

The package management system conda is in charge of managing Anaconda's package versions. Because it proved to be useful on its own and for applications other than Python, this package manager was released as its own open-source package. Miniconda is a smaller, bootstrap version of Anaconda that only includes Python, conda, the packages that they depend on, and a few other packages.

In addition to the conda package and virtual environment manager and over 7,500 additional open-source packages that can be installed from PyPI, the Anaconda distribution comes pre-installed with over 250 packages. Anaconda Navigator, a graphical alternative to the command line interface (CLI), is included as well. Python 2.7 and Python 3 are included in the default installation of Anaconda2.
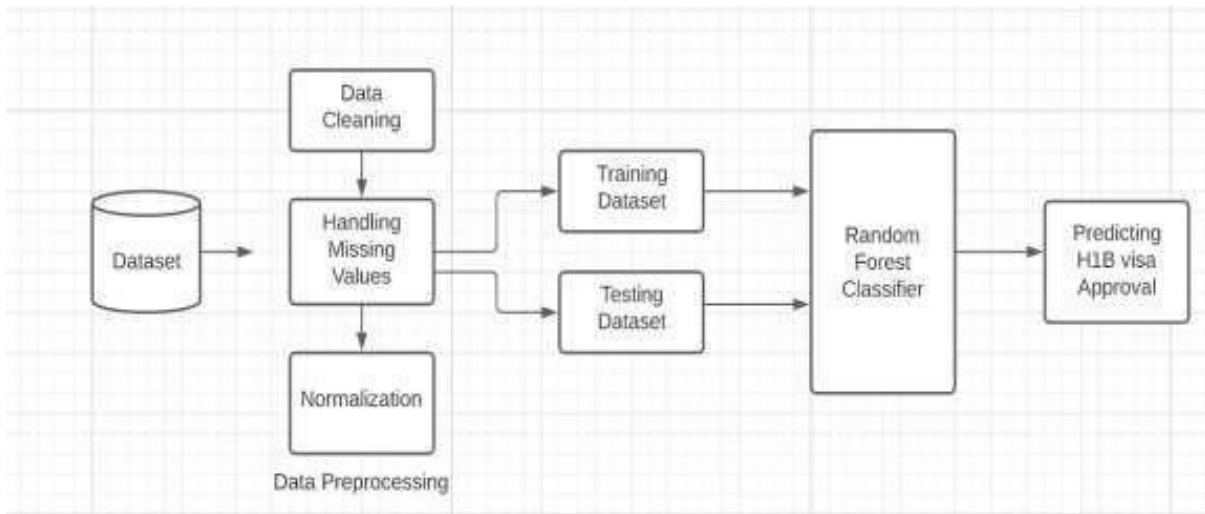
112

Fig.3 Proposed Architecture

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python. There is also a small, bootstrap version of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.
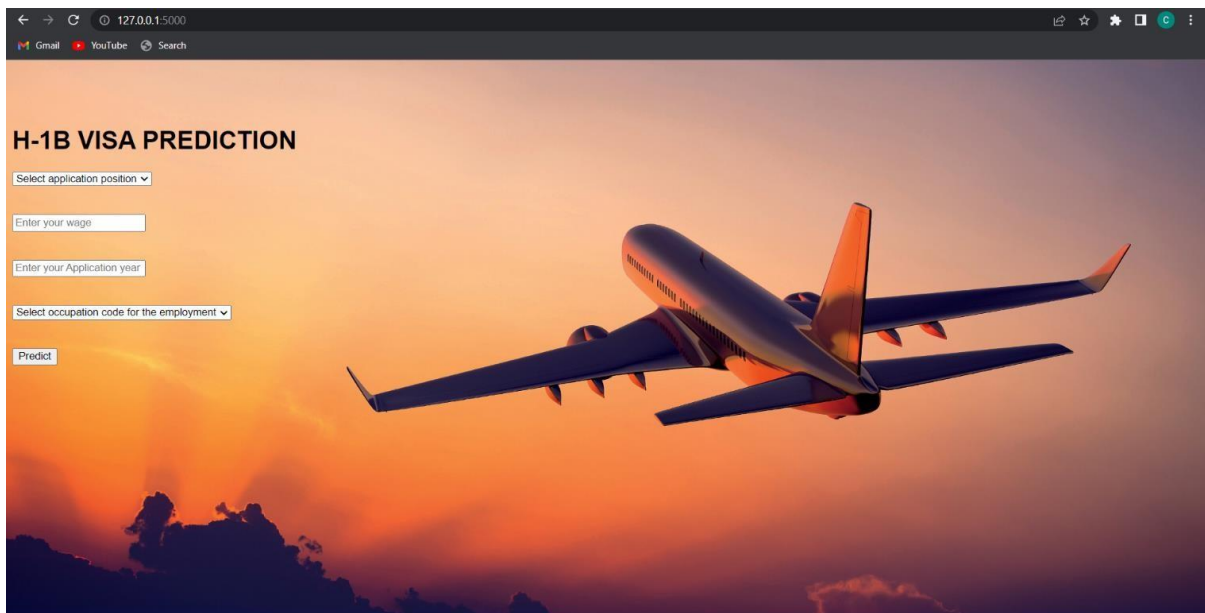

Fig.4 Output Of The Project

5.Conclusion

The scope of this challenge is to gather a tool for providing a desire to every character who is suffering for H1B visa magnificence method and predicting the recognition of software with the highest accuracy Supplemental information regarding the Standard Occupational. Since the selection for the H1B visa has increased annually over the past decade, The H-1B Visa

113

preference procedure allows for the accumulation and implementation of classification (SOC). The income feature of this data set can be correctly mounted to a variety of salaries using the income opinions and levels provided by SOC. These salaries can then be used to categorize visa petitions based solely on career roles in desire to region. In addition, great beauty algorithms other than discriminative models can be tested in this study, and their results can also be compared. The Random Forest classifier performs best right here, with the highest accuracy in comparison to all of the great algorithms that are suitable for performing the assessment operations. With the help of the Random Forest Algorithm, we were able to improve and inform the data with an accuracy of 86.88% in this case, and our results indicate that this set of recommendations is the best for predicting H1B visa approval. Due to the increasing number of H1B visa applicants, it is now mandatory to develop a tool for accurately requesting H1B visa approval. As a result, we are able to search for the H1B visa approval reputation with the helpful assistance of a number of device learning beauty algorithms. Employees from other countries who are traveling to the United States may greatly benefit from this.

References

1.   Dhanasekar Sundararaman, Nabarun Pal, Aashish Kumar Mis-raa, (2017)," An analysis of non- immigrant work visas in the USA using Machine Learning", International Journal of Computer Science and Security(IJCSS), Vol. 6,

2.   https://www.foreignlaborcert.doleta.gov/performancedata.cfm.

3.   UNITED STATES DEPARTMENT OF LABOR. (2009, January 15). OFLC Performance Data. (www.dol.gov) Retrieved September 09, 2017, from UNITED STATES DEPARTMENT OF LA-BOR Employment & Training Administration:

4.   Trim Bach, S., (2016), Giving the Market a Microphone: Solu-tions to the Ongoing Displacement of US Workers through the H1B Visa Program. Nw. J. Int'l L. & Bus., 37, p.275.

5.   Doran, K., Gelber, A. and Isen, A., 2014. The effects of high-skilled immigration policy on firms: Evidence from H-1B visa lot-teries (No. w20668). National Bureau of Economic Re- search. https://doi.org/10.3386/w20668.

6.   Bound, J., Khanna, G. and Morales, N., (2017). Understanding the Economic Impact of the H- 1B Program on the US. In High-Skilled Migration to the United States and its Economic Con- sequences. University of Chicago Press.

7.   Mithas, S. and Lucas Jr, H.C., 2010. Are foreign IT workers cheaper? US visa policies and compensation of information tech-nology professionals. Management Science, 56(5), pp.745-765. https://doi.org/10.1287/mnsc.1100.1149.

8. Kaggle H-1B dataset, https://www.kaggle.com/nsharan/h-1b-visa 13 outsourcing companies took nearly one-third of all H-1B visas in 2014, https://www.nytimes.com/interactive/2015/11/06/us/outsourcing-companies-dominate-h1b-visas.html.

9. Disney 'forced 250 of its American IT workers to train up the Indi-an workers who replaced them', http://www.dailymail.co.uk/news/article-4037392/Disney-fired-250-American-workers- replaced-Indian-staff-visas-suit-says.html. 2016

10. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), pp.2825-2830, 2011

11. Qing-yun Dai, Chun-ping Zhang, Hao Wu, "Research of Decision Tree Classification Algorithm in Data Mining", Vol.9, No.5 (2016), pp.1-8

12. J.Ross Quinlan, "C4.5: Programs for machine learning", Elsevier, 2014.