INDEX

1. INTRODUCTION

1.1 Overview:

      H-1B is a type of non-immigrant visa in the United States that allows foreign nationals to work in occupations that require specialized knowledge and a bachelor's degree or higher in the specific specialty. This visa requires the applicant to have a job offer from an employer in the US before they can file an application to the US immigration service (USCIS). USCIS grants 85,000 H-1B visas every year, even though the number of applicants far exceed that number. The selection process is claimed to be based on a lottery, hence how the attributes of the applicants affect the final outcome is unclear. We believe that this prediction algorithm could be a useful resource both for the future H-1B visa applicants and the employers who are considering to sponsor them.

1.2 Purpose:

      In our project, we aim to predict the outcome of H-1B visa applications that are filed by many high-skilled foreign nationals every year. We framed the problem as a classification problem and applied Naive Bayes, Logistic Regression, SVM and Neural Network in order to output a predicted case status of the application. The input to our algorithm is the attributes of the applicant.

2. LITERATURE SURVEY

2.1 Existing problem:

There has not been noticeable work done in the field of predicting the case status of H1-B visa petitions. A similar project has been done at UC, San Diego for predicting the decision of the file H1- B visa petition. A project done by the students of UC Berkley [5] tried to predict the waiting time to get a work visa for a given job title and for a given employer. They used K-Nearest Neighbors as the primary model to predict 'Quickest Certification Rate' across both occupations and companies

2.2 Proposed solution:

Inorder to predict the outcome of H-1B visa applications based on the attributes of the applicant,several machine learning models like SVM, Naive Bayes can be used.Finally,this can be integrated to a web application.

3.THEORITICAL ANALYSIS:

3.1 Block diagram:
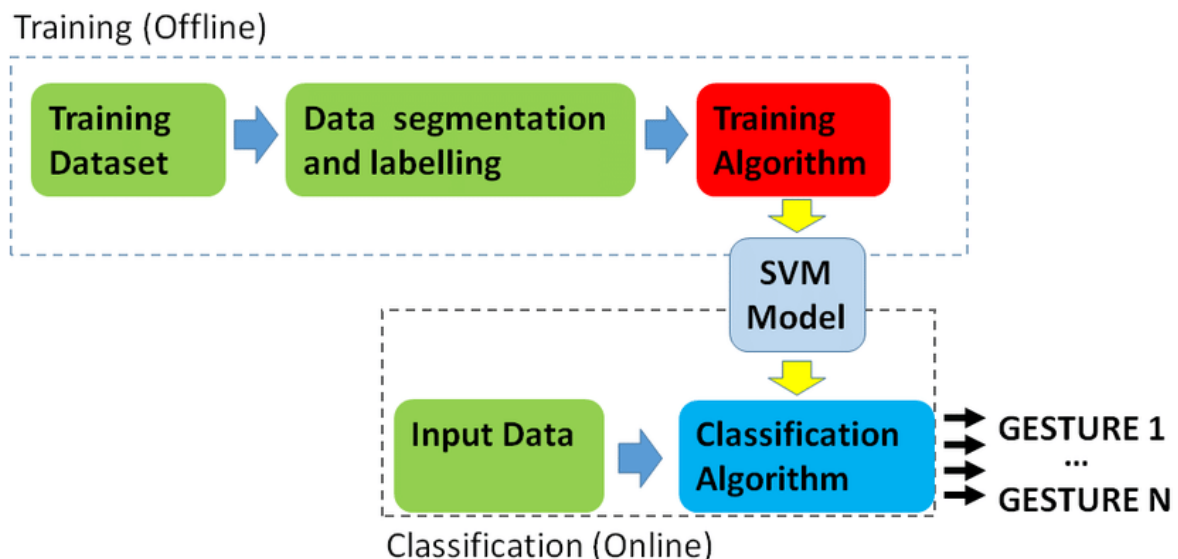
3.2 Hardware / Software designing:

The model is designed using Jupyter notebook in Anaconda which is a open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment.The application building is done using flask which is a web framework that provides tools, libraries and technologies that allow the developer to build a web application. This web application can be some web pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website.

4. EXPE RIMENTAL INVESTIGATIONS:

Support Vector Machine tries to find a hyperplane that separates two different classes such that the distance from the closest data point to that hyperplane is maximized.

SVM is another type of the most important switching methods.SVM is implemented in the inverter controlling to extract sinusoidal AC waveforms from the VSI.The real-time implementation of the SVM via a DSP-TMS320F28335 controller, which includes C280x/C28x3x ADC and C280x/C28x2833x ePWM,is shown in fig.3.Two ADC blocks are used;the first ADC block is used to read the V_{abc} voltage,and the second ADC block is used to read the V_{dc} voltage.The V_{abc} voltage is measured through pins 2,3 and 4 from port 5(ADCINB1,ADCINB2 and ADCINB3),whereas the V_{dc} voltage is measured through pin 2from port 9.(ADCINA1).The bias should be experimentally removed from the ADC_{output} signal,which is achieved through subtracting from the half of the highest value.In the SVM,the bias is found to be 2056
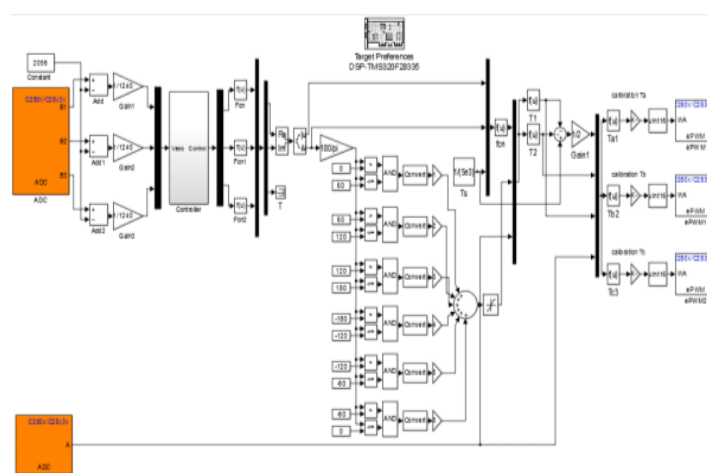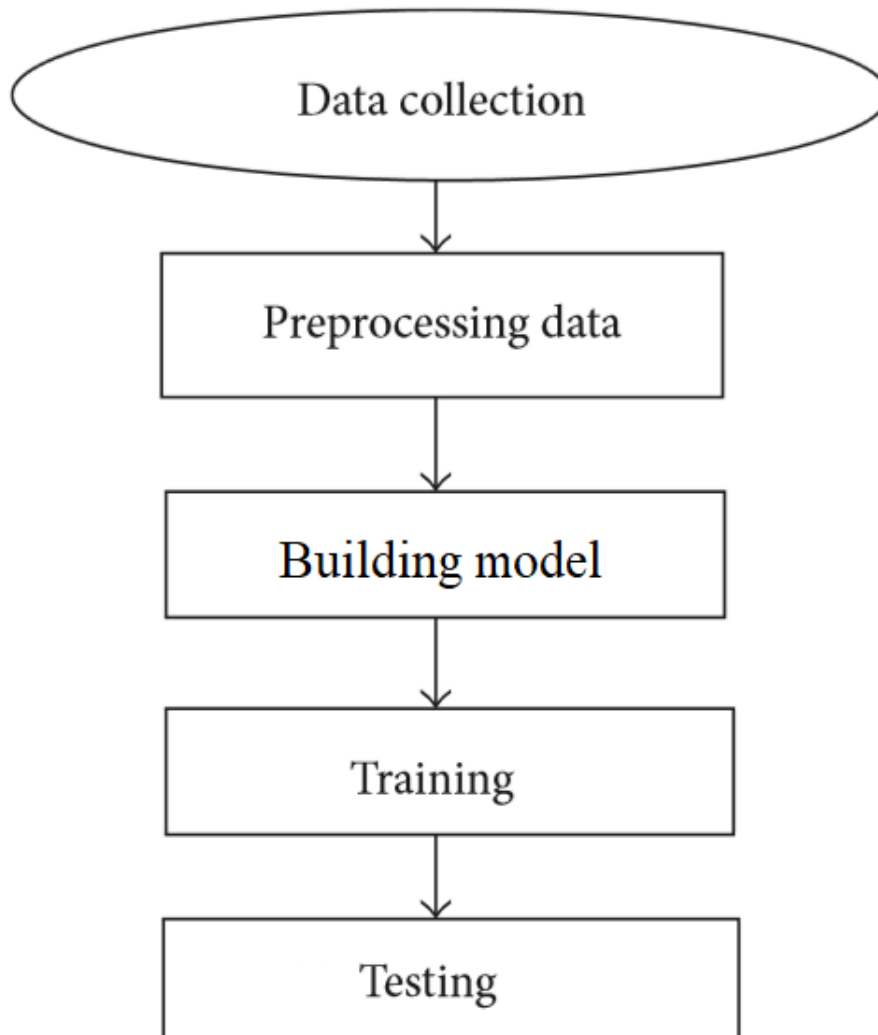

Fig.3

5.FLOWCHART:

```
        ┌─────────────────────────────────┐
        │         Data collection          │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │        Preprocessing data        │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │         Building model           │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │            Training              │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │            Testing               │
        └─────────────────────────────────┘
```

6. RESULT:

This model is built using ML model SVM. The model was integrated with a flask web application which predicts the approval of H-1B VISA applicants based on the attributes of the applicant.

7. ADVANTAGES & DISADVANTAGES:

Advantages:

- SVM are accurate in high dimensional spaces
- Moreover SVM use a subset of training points in the decision function (called support vectors), so it's also memory efficient.
- Since it is lottery based, it reduces the human work
- It acts as a useful resource for future H-1B applicants
- It is also helpful for the employers who are considering to sponsor them

Disadvantages:

The main disadvantage of the SVM algorithm is that it has several key parameters that need to be set correctly to achieve the best classification results for any given problem. Parameters that may result in an excellent classification accuracy for problem *A*, may result in a poor classification accuracy for problem *B*. The user may, therefore, have to experiment with a number of different parameter settings in order to achieve a satisfactory result.

8. APPLICATIONS:

Some real time applications of SVM are:
- Inverse Geosounding Problem
- Seismic Liquefacation Potential
- Protein Fold and Remote Homology detection
- Facial Expression Classification
- Text Classification
- Speech Recognition

9. CONCLUSION:

H1B VISA category is one of the most applied categories among other visas categories. It is designed to overcome the shortage of skilled workers in America but it affects the hiring of American workers and no. of foreign workers increased day by day. So in current work, the machine learning SVM Classifier were considered for determining the status of H1-B visa applications. SVM Classifier performed the best in terms of accuracy, precision and F1 score over others. We achieved a best of 95% classification accuracy. Our results showed that the most predictive features are Annual Salary and Occupation. One can infer from these results that the chance of being certified increases with the amount of salary and how successful your sponsor was in the previous H1B application.

10. FUTURE SCOPE:

Supplemental data concerning the Standard Occupational Classification (SOC) can be gathered and used in coordination with this data set to obtain a more comprehensive analysis of how the H-1B Visa selection process works. By using the wage evaluations and ranges under SOC, the wage attribute in this data set can be correctly put in to a range of salaries which can then be used to classify the visa petitions based on occupation roles rather than location wise. In addition, other classification algorithms other than the discriminative models can be experimented with this testbed and their performances can also be analyzed.

11. BIBILOGRAPHY:

- https://www.kaggle.com/trivedicharmi/h1b-disclosure-dataset
- https://towardsdatascience.com/
- https://www.w3schools.com/bootstrap/bootstrap_forms_inputs.asp
- https://www.quora.com/