VISA APPROVAL PREDICTION

Using Classification Algorithms

Developed by: Kotha Nithya, Poshala Amulya, Sharath Kumar Gongalla,Podila Madhu

Smart Bridge-Remote Summer Internship Program

1.INTRODUCTION

The H-1B is an employment-based visa in the United States, which allows U.S. employers to temporarily employ foreign workers in specialty occupations. To apply for H-1B visa, an U.S employer must offer an job and petition for H-1B visa with the U.S. immigration department. This is the most common and legal visa status and for international students who complete their college / higher education (Master, PhD) and work in a full-time position. Every year, the U.S. Citizenship and Immigration Services (USCIS) will officially begin accepting petitions for the H-1B Visa in April for the next fiscal year. The status of H-1B visa will definitely influence the life and work, and even the career of the international students.Since the number of applicants is very large than the number of selections and as the selection process is claimed to be as lottery there is no insight of how the attributes have influence over the outcome.

The Office of Foreign Labor Certification (OFLC) generates program data, including data about H1-B visas. The disclosure data is updated annually and is available online. The first step of the H1B application process is for the U.S. employer to file the H1B petition on behalf of the foreign worker. In second step, the prevailing and actual wages should be confirmed by the State Employment Security Agency. If the prevailing wage exceeds the offer made by the prospective employer then a wage determination will be sought. The third step of the H1B application process is to file the Labor Condition Application. The next step is to prepare the petition and file it at the proper USCIS office. Processing times for H1B application petitions are subject to vary from location to location. If you would like your petition expedited you may elect for premium processing. The final step of the H1B application process is to check the status of your H1B visa petition by entering your receipt number. Once USCIS has your application on file, they will update your status on their system.

1.1 Overview

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.Each fiscal year the United States government grants a total of 85,000 new H-1B visas; this includes 65,000 visas for overseas workers with at least a bachelor's degree and 20,000 visas available for those employees with an advanced degree in a specialty field from an accredited United States academic institution. The allotment of these visas is said to be a lottery system not depending on any specific criteria. Thus, a foreign national is not guaranteed selection for an H- 1B visa merely because he or she fulfills all the qualifying criteria.This work aims to use machine learning on known factors to make predictions under this uncertainty, with accuracy better than that achieved using the prior probablility.Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated datasets.The model has been built using data from Office of Foreign Labor Certification to predict the case status of visa.Six algorithms have been used to build the proposed model:LogisticRegression,RandomForest,DecisionTree,KNN,NaiveBayes,SVM.By using the algorithm a Flask model has been implemented and tested.The result has been discussed and a full comparison between algorithms has conducted.DecisionTree was selected as best algorithm based on accuracy and performance.

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 in order to output a predicted case status of the application. The input to our algorithm is the attributes of the applicant.

Our aim from the project is to make use of pandas,numpy,seaborn libraries from python to extract the libraries for machine learning for the visa approval prediction.

Secondly, to learn how to hyper tune the parameters for the DecisionTree machine learning algorithm.

And in the end,the objective of this phase of the project is to develop a model which can tell us whether an individual is eligible to file an H1B visa based on various characteristics in the data.The objective is achieved through creating different machine learning models to predict the outcome-certified or denied of visa.

2. LITERATURE SURVEY

A common scenario in machine learning tasks involves building a classification model that is trained on a training data set, producing a model that is then used to predict the probability of data belonging to one of the classes of the test data set. The trained model solves a classification task, producing predictions of a target variable over previously unseen inputs, which represent new cases. Since historical data is used to train the model this falls under the supervised learning techniques. This classification methodology can predict the classes in two kinds of situations; one, when predicting the decisions made by a third party is involved. For example, some common applications of classification would be loan approval prediction, medical diagnosis, etc. And the second case is when classification is based on some historical ground truths. Example spam email filtering.

The main aim of this project is to predict the certification status of the visa applications. For this purpose, the concept of classification is used. That is, when all the attributes are provided, the classifying model is trained on historical data of this application domain and probability of rows belonging to one class or status over the other is determined. This aids in predicting decisions made by USCIS and to recognize if there is any specific pattern behind their decisions.

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it.It is the core of knowledge discovery process.The various steps involved in extracting knowledge from raw data.Different data mining techniques include classification,clustering,neural networks,regression etc.Classification is the most commonly applied data mining technique,which employs a set of preclassified examples to develop a model that can classify the population of records at large.Detection and prediction applications are particularly well suited to classification technique.In classification,a training set is used to build the model as the classifier which can classify the data items into its appropriate classes.A test set is used to validate the model.

2.1 Existing Problem

The previous models have high time complexity and space complexity whereas this model is constrained with the lot of advantages and with a higher accuracy than any other model already proposed.In this model we used Machine learning algorithm named DecisionTree which give an accuracy more then previously predicted problem.

Although this area currently doesn’t seem to be well-studied, we were able to find a report that is relevant. Th reports used the same dataset we had used. The report conducted a detailed data analysis and visualization for H-1B application distribution based on different input features such as location, salary, year and job type . Although they had a prediction algorithm based on K-means clustering , it provided prediction accuracies for only a small subset of job types instead of an average one. Overall, this report gave us good insight on the distribution of our data.

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 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

Machine Learning (DecisionTree):

Decision trees model sequential decision problems under uncertainity.A decision tree describes graphically the decisions to be made,the events that may occur,and the outcomes associated with combinations of decisions and events.Probabilities are assigned to the events,and the values are determined for each outcome.A major goal of the analysis is to determine the best decisions.

Decision tree calculation is a standout between the most vital classification measures in information mining.Decision tree classifier as one sort of classifier is a stream diagram like a tree structure,where each inside hub indicates a test on a characteristic,each branch speaks to a result of the test,and each leaf hub speaks to a class.The technique that a decision tree demonstarte is utilized to group a record is to discover a way that from root to a leaf by estimating the characteristics test,and the trait on the leaf is classification result.i.e;In this method a set of training examples is broken down into smaller and smaller subsets while at the same time an associated decision tree get incrementally developed. At the end of the learning process, a decision tree covering the training set is returned.

The key idea is to use a decision tree to partition the data space into cluster (or dense) regions and empty (or sparse) regions.

In Decision Tree Classification a new example is classified by submitting it to a series of tests that determine the class label of the example. These tests are organized in a hierarchical structure called a decision tree. Decision Trees follow Divide-and-Conquer Algorithm.

And also we have created an UI using the Flask for the case status prediction,this UI will allow the users to predict the case status very easily and the User interface is user friendly not at least one complication in using the interface,and it can be used just by enterning some necessary details into the UI in real time it'll give the predicted value like if the person's visa is certified or denied.

3. THEORETICAL ANALYSIS

While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only one algorithm for the prediction problem that is DecisionTree.DecisionTree and RandomForest accuracies are very close but based on performance and accuracy we have considered DecisionTree as the best algorith for this dataset.

The pecularity of this problem is collecting the real time details and working with the prediction at the same time,so we developed an user interface for the people who'll be accessing for the visa status prediction.Accuracy is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set.The formula is as follows

Accuracy=TP+TN/TP+TN+FT+FN

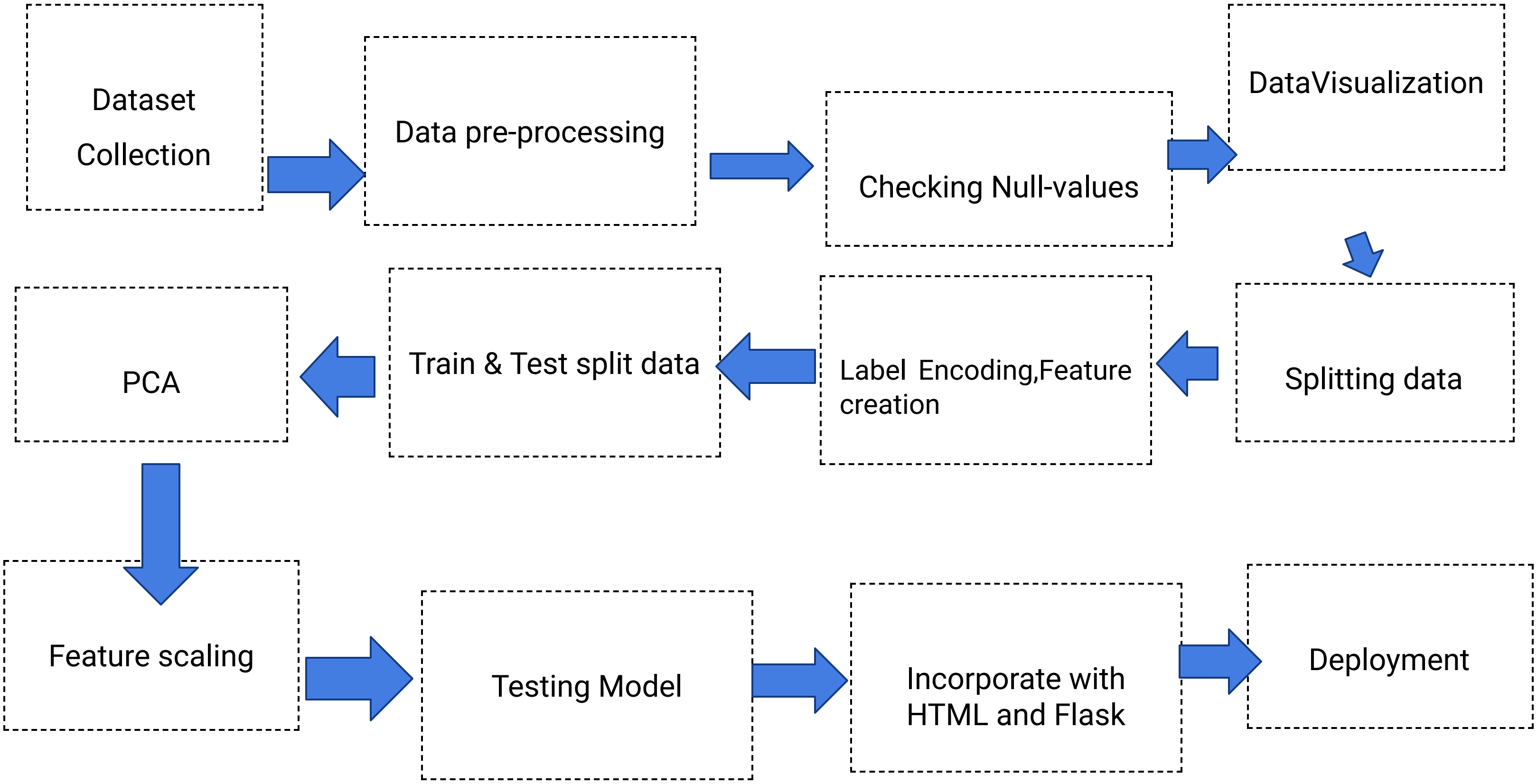
At first we got like lot of accuracies because we tried lot of algorithms for the best accurate algorithm,finally after all of that we tried the best suitable algorithm which gives the prediction accuractely is DecisionTree.And developed it to use as a real time prediction probelm for the visa case status i.e; visa approval prediction.

In statistics, a receiver operating characteristic(ROC), is a two dimensional graphical plot that illustrates the performance of a binary classifier system.The curve is created by plotting the true positive rate(TPR) against the false positive rate(FPR) at various threshold settongs.ROC curve can intuitively represent the performance of classifier.

FPR=FP/FP+TN TPR=TP/TP+FP

3.1 Block Diagram

The data was then combined. Data cleaning and preprocessing steps were followed on the data set. Depending on the task at hand, the features were divided into subsets. Python was used as the major programming language for this program and using in-built libraries, the data set was split into training and testing data in 70 and 30 percentages respectively. Scikit-learn was used to train the inductive learners and apply them to the data sets. The training data was fed to the learning models and the prediction scores of the classification task were obtained. These models were then tested on the testing data. The results are compared to find the best suited classifying model.



3.2 Software Designing

- Jupyter Notebook Environment
- Spyder Ide
- Machine Learning Algorithms
- Python(pandas,numpy,matplotlib,seaborn,sklearn)
- HTML
- Flask

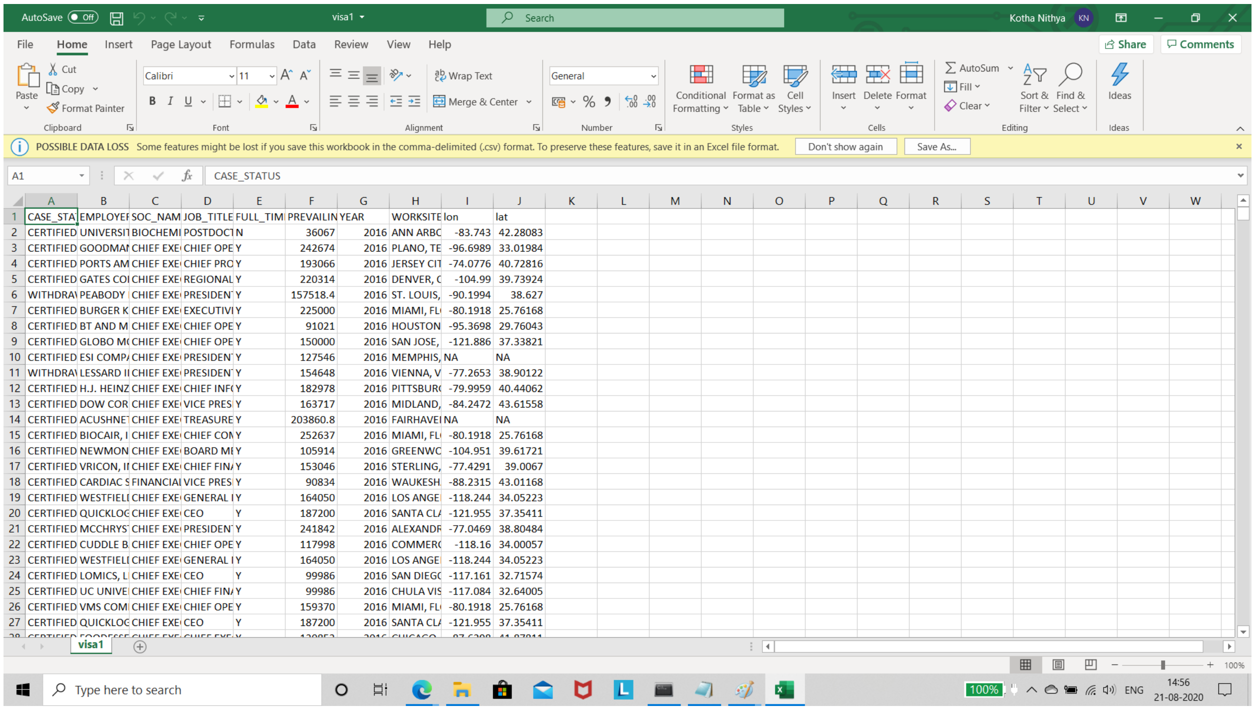
We developed this visa approval prediction by using the Python language which is a interpreted,dynamically typed and high level programming language and using the Machine Learning algorithms.

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 webframework written in Python. It is classified as a microframework 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 in HTML by creating the templates to use in th functions of the Flask and HTML.

4. EXPERIMENTAL INVESTIGATION

Our dataset is from Kaggle listed under the name “H-1B Visa Petitions 2011-2016 dataset” , and it initially included about 3 lakh data points. It contained 9 features and 1 class attribute which can be examined in Figure 1.



POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

A1 CASE_STATUS

	A	B	C	D	E	F	G	H	I	J
1	CASE_STA	EMPLOYEF	SOC_NAM	JOB_TITLE	FULL_TIM	PREVAILIN	YEAR	WORKSITE	lon	lat
2	CERTIFIED	UNIVERSIT	BIOCHEMI	POSTDOCT	N	36067	2016	ANN ARBC	-83.743	42.28083
3	CERTIFIED	GOODMAI	CHIEF EXE	CHIEF OPE	Y	242674	2016	PLANO, TE	-96.6989	33.01984
4	CERTIFIED	PORTS AM	CHIEF EXE	CHIEF PRO	Y	193066	2016	JERSEY CIT	-74.0776	40.72816
5	CERTIFIED	GATES COI	CHIEF EXE	REGIONAL	Y	220314	2016	DENVER, C	-104.99	39.73924
6	WITHDRAV	PEABODY	CHIEF EXE	PRESIDENT	Y	157518.4	2016	ST. LOUIS,	-90.1994	38.627
7	CERTIFIED	BURGER K	CHIEF EXE	EXECUTIVI	Y	225000	2016	MIAMI, FL	-80.1918	25.76168
8	CERTIFIED	BT AND M	CHIEF EXE	CHIEF OPE	Y	91021	2016	HOUSTON	-95.3698	29.76043
9	CERTIFIED	GLOBO MC	CHIEF EXE	CHIEF OPE	Y	150000	2016	SAN JOSE,	-121.886	37.33821
10	CERTIFIED	ESI COMPA	CHIEF EXE	PRESIDENT	Y	127546	2016	MEMPHIS,	NA	NA
11	WITHDRAV	LESSARD II	CHIEF EXE	PRESIDENT	Y	154648	2016	VIENNA, V	-77.2653	38.90122
12	CERTIFIED	H.J. HEINZ	CHIEF EXE	CHIEF INFC	Y	182978	2016	PITTSBURC	-79.9959	40.44062
13	CERTIFIED	DOW COR	CHIEF EXE	VICE PRESI	Y	163717	2016	MIDLAND,	-84.2472	43.61558
14	CERTIFIED	ACUSHNET	CHIEF EXE	TREASURE	Y	203860.8	2016	FAIRHAVEI	NA	NA
15	CERTIFIED	BIOCAIR, I	CHIEF EXE	CHIEF CON	Y	252637	2016	MIAMI, FL	-80.1918	25.76168
16	CERTIFIED	NEWMON	CHIEF EXE	BOARD ME	Y	105914	2016	GREENWC	-104.951	39.61721
17	CERTIFIED	VRICON, II	CHIEF EXE	CHIEF FIN/	Y	153046	2016	STERLING,	-77.4291	39.0067
18	CERTIFIED	CARDIAC S	FINANCIAI	VICE PRESI	Y	90834	2016	WAUKESH.	-88.2315	43.01168
19	CERTIFIED	WESTFIELI	CHIEF EXE	GENERAL I	Y	164050	2016	LOS ANGEI	-118.244	34.05223
20	CERTIFIED	QUICKLOC	CHIEF EXE	CEO	Y	187200	2016	SANTA CL/	-121.955	37.35411
21	CERTIFIED	MCCHRYST	CHIEF EXE	PRESIDENT	Y	241842	2016	ALEXANDR	-77.0469	38.80484
22	CERTIFIED	CUDDLE B.	CHIEF EXE	CHIEF OPE	Y	117998	2016	COMMERC	-118.16	34.00057
23	CERTIFIED	WESTFIELI	CHIEF EXE	GENERAL I	Y	164050	2016	LOS ANGEI	-118.244	34.05223
24	CERTIFIED	LOMICS, LI	CHIEF EXE	CEO	Y	99986	2016	SAN DIEGC	-117.161	32.71574
25	CERTIFIED	UC UNIVEI	CHIEF EXE	CHIEF FIN/	Y	99986	2016	CHULA VIS	-117.084	32.64005
26	CERTIFIED	VMS COMI	CHIEF EXE	CHIEF OPE	Y	159370	2016	MIAMI, FL	-80.1918	25.76168
27	CERTIFIED	QUICKLOC	CHIEF EXE	CEO	Y	187200	2016	SANTA CL/	-121.955	37.35411

Figure 1:Data points from the unprocessed Dataset

It is highly probable that a dataset in its raw form can include missing values,inconsistent entries and noise.These data entries affect the quality of the results.Thus,to improve the performance efficiency of the machine learning models,it is important to train and test the system on clean and consistent data.Redundant and unncessary fields and values were also removed.Some of the steps followed in this project for this purpose are: Data Cleaning,Feature Selection.

We processed some of the existing features, created new features that we thought could be useful for prediction and discarded some features using the library Pandas . In particular, we noticed that the JOB_TITLE feature represents highly redundant information with the SOC_NAME feature, therefore we discarded JOB_TITLE from the beginning. Also, we transformed both SOC_NAME and COMPANY_NAME features into the corresponding forms of success rate and total number of applications. Finally, we normalized all the features such that they had zero mean and unit variance. Our final list of features is further described in the following:

CASE_STATUS: We excluded the cases ’CERTIFIED-WITHDRAWN’ and ’WITHDRAWN’, since ’WITHDRAWN’ decisions are either made by the petitioning employer or the applicant, therefore not predictive of USCIS’s future behavior. We labeled ’CERTIFIED’ cases as 0 and ’DENIED’ cases as 1.

FULL_TIME_POSITION:: Positions are given in ”Full Time Position = Y; Part Time Position = N” format. We converted them to ”Full Time Position = 1; Part Time Position = 0” format

PREVAILING_WAGE:: Prevailing wage is the average wage paid to employees with similar qualifications in the intended area of employment. We discarded the outlier terms and used the rest of the data as it was.

YEAR: Year in which application was filed.

NEW_EMPLOYER:EMPLOYER_NAME contains the names of the employers and there are lot of unique employers. It is the company which submits the application for its employee. We cannot use EMPLOYER_NAME directly in the model because it has got many unique string values or categories; more than 500 employers. These employers act as factors or levels. It is not advisable to use this many factors in a single column.

We have created a new feature called NEW_EMPLOYER: If the employer name contains the string 'University' (for instance if a US university is filing a visa petition, then it has more chances of approval for the employee).So, if the EMPLOYER_NAME contains 'University', then NEW_EMPLOYER contains the university value.All the strings in EMPLOYER_NAME containing the keyword university will have 'university' as value in the NEW_EMPLOYER column. All the remaining empty rows will be filled with 'non university'.

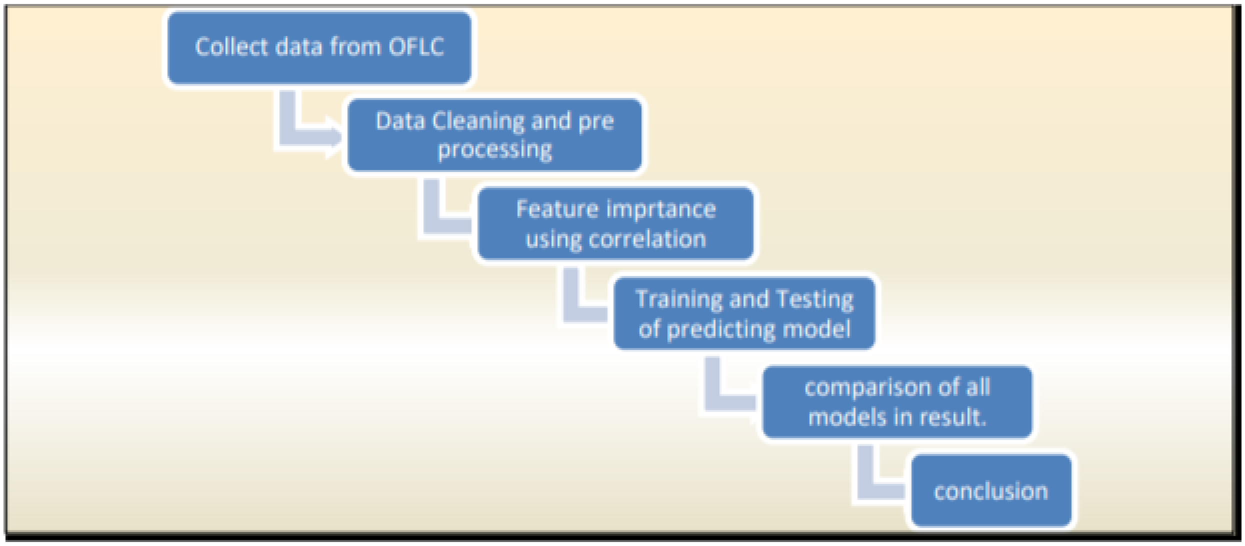
OCCUPATION: SOC_NAME, it consists of an occupation name. There are lot of values associated with SOC_NAME, so we have created a new feature that will contain the important occupation of the applicant, mapping it with the SOC_NAME value. We created a new variable called OCCUPATION. For example computer, programmer and software are all computer ocupations. This will cover the top 80% of the occupations, and minor and remaining occupations will be categorized as others.

state: Since visa applications majorly depend on State location, you should split the state information from the WORKSITE variable.

lat: This field denotes the latitude coordinates of the worksite.since it is not useful for prediction we have dropped that column.

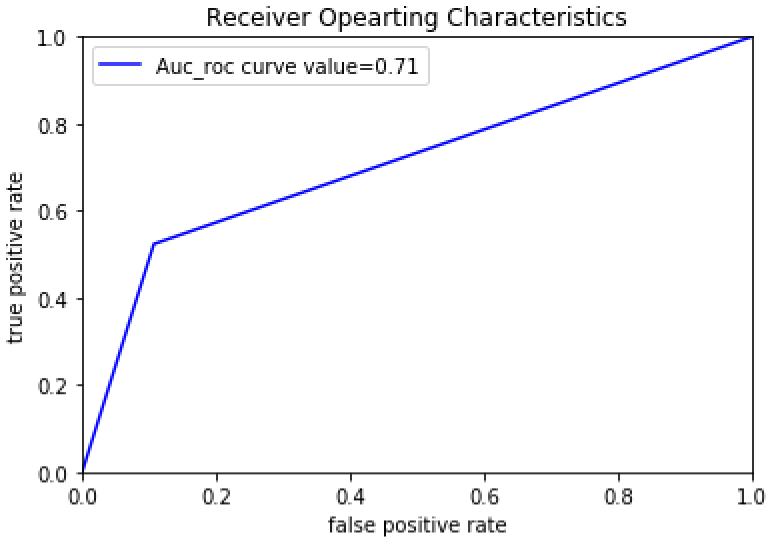
lon: This field denotes the longitude coordinates of the worksite.since it is not useful for prediction we have dropped that column.

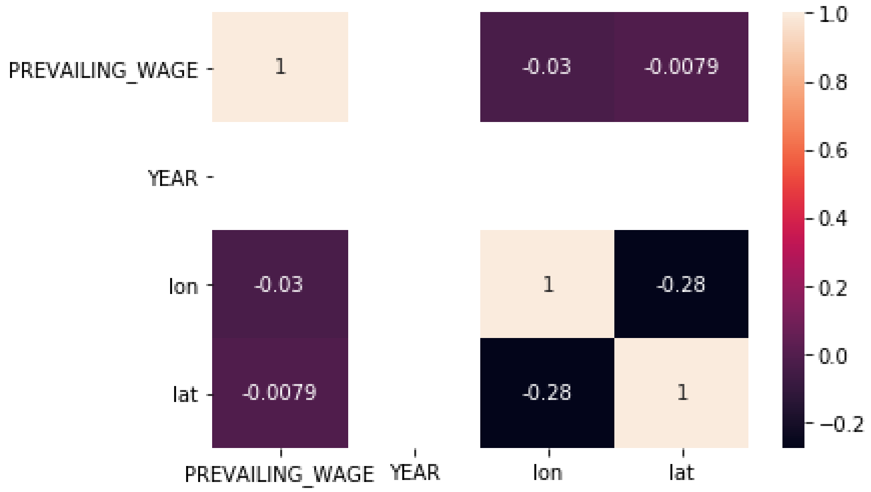
5.FLOWCHART


Collect data from OFLC
Data Cleaning and pre processing
Feature imprtance using correlation
Training and Testing of predicting model
comparison of all models in result.
conclusion

6. RESULT

In this paper, the Decision Tree algorithm is used to predict its performance, and compared with another five machine learning methods namely the Logistic Regression,KNN,Naive Bayes, SVM and the Random Forest. The obtained results are displayed in Table below. The results show that, the performance of DecisionTree and Random Forest have comparable performance than that of Naive Bayes, KNN,Logistic Regression and SVM, but the Decision Tree still performs the best, than the RandomForest .The ROC curve of the prediction model based on DecisionTree are all above 0.71, it means there is 71% chance that model will be able to distinguish between positive class and negative class,indicating that the model has ability of generalization.





S.NO	ALGORITHMS USED	ACCURACY	AUC_ROC
1	Logistic Regression	60%	0.52
2	KNN	63%	0.59
3	DecisionTree	74%	0.71

4	NaiveBayes	62%	0.57
5	RandomForest	74%	0.70
6	SVM	62%	0.55

7.ADVANTAGES AND DISADVANTAGES

Advantages:

- Easy and simple User Interface for the people to evaluate the visa approval status.
- DecisionTree gives the accurate result of the prediction which is the algorithm we used for prediction.
- It is composed using the HTML and Python for the web usage in real time.

It can work in real time and predict as soon as the necessary details for prediction are given to the model.

- Extremely fast at classifying unknown records.
- Excludes unimportant features.
- By applying Decision tree model this project has became very intuitive and easy .

Disadvantages:

- A small change in the data can cause a large change in the structure.
- Needs more than a single value for the prediction.
- Not robust to big-influentials.
- Requires higher time to train the model.
- Easy to overfit.

8. APPLICATIONS

- It is widely used for predicting visa status in the OFLC and USCIS. This can also be applied to other visa's like F1,B1,L1 etc. using the attributes related to them.
- It is one of the most widely used areas of data mining in the OFLC and USCIS for

maintaining the records and predicting the status.

● Due to tremendous growth in applications of visa the OFLC and USCIS deals with, analysis and transformation of the data into useful knowledge has become a task beyond human ability.

● So we use Machine Learning Algorithms to analyze the data and propose the visa status based on the given information for the attributes.

9. CONCLUSION

In this work LogisticRegression,KNN,DecisionTree,NaiveBayes,RandomForestand SVM were considered for determining the status of visa.DecisionTree Classifier performed the best in terms of accuracy and prediction.In the end, it is indeed possible to predict the outcome of H-1B visa applications based on the attributes of the applicant using machine learning

The utilization of machine learning related algorithm enabled us to analyze the data based on training and learning steps. Therefore, it helped us to predict the approval and deny rate of H-1B visa of the testing year by DecisionTree.These data provides the situation of the H-1B visa during the recent years. The analysis of H-1B visa application might help to guard students, individuals to accomplish their American Dreams.There is no definitive guide of which algorithms to use given any situation. What may work on some data sets may not necessarily work on others. Therefore, always evaluate methods using cross validation to get a reliable estimates.

10. FUTURE SCOPE

This dataset presented interesting challenges for dealing with complex dataset. The data still needed to cleaned, and more Machine Learning algorithm,could exploited and implemented.

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 tested and their performances can also be analyzed.A rigorous analysis of other machine learning algorithms other than these six can also be done in future to investigate the power of machine learning algorithms for visa approval prediction.

11. BIBLIOGRAPHY

- Dhanasekar Sundararaman , Nabarun Pal , Aashish Kumar Misraa ,(2017),” An analysis of nonimmigrant work visas in the USA using Machine Learning” ”, International Journal of Computer Science and Security(IJCSS), Vol. 6,

- American Immigration Council. (2018). The H-1B Visa Program: A Primer on the Program and its Impact on Job Wages, and the Economy. Retrieved from American Immigration Council: https://www.americanimmigrationcouncil.org/sites/default/files/research/the_h1b_visa_program_a_primer_on_the_program_and_its_impact_on_jobs_wages_and_the_economy.pdf

- F. Harrell. (2017, January 15). Classification vs. Prediction. Retrieved April 01, 2018, from Statistical Thinking: http://www.fharrell.com/post/classification/

CTomM.Mitchel, McGrawHil, Decision Tree Learning, Lec-ture slides for textbook Machine Learning,, 197

- 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

- N. J. Nilson. (2005). Intoduction to Machine Learning. The MIT Press.

- S. Naribole. (2017). H-1B Visa Petitions 2011-2016. Retrieved from Kaggle: https://www.kaggle.com/nsharan/h-1b-visa

- http://cs229.stanford.edu/proj2017/final-reports/5208701.pdf

- https://www.datacamp.com

- Dan, Lucas, Samuel Kabue, and Sarah Neff. Project Alien Worker. PDF. Berkeley: UC Berkeley School of Information, April 2016. https://www.ischool.berkeley.edu/sites/default/files/projects/project_alien_worker_-_final.pdf.

APPENDIX

HTML:

```
<html>
<style>
*{
  margin:0;
  padding:0;
  }
  body {
background-image:url("http://blogs.adobe.com/dreamweaver/files/2015/06/raindrops-nofilter.png?fit=539%2C320&ssl=1");
  background-size:cover;
   background-position:center;
  margin-top:40px;
  font-family:sans-serif;
  }
  .header{
            width:800px;
            background-color:rgb(0,0,0,6);
            margin:auto;
            color:#FFFFFF;
            padding:10px 0px 10px 0px;
            text-align:center;
            border-radius:15px 15px 0px 0px;
         }
 .MAIN button
{
height:30px;
width:200px;
margin-left:60px;
background-color:rgb;
}
 .MAIN {
        background-color:rgb(0,0,0,0.5);
        width:800px;
        margin:auto;
        text-align:center;
        }
    form{
         padding:100px;
         }



</style>
<body>
<div class="header">
<h1>VISA APPROVAL PREDICTION</h1>
</div>
<div class="MAIN">
<form action = "{{url_for('prediction')}}" method = "post">
<br>
<label for = "FULL_TIME_POSITION">FULL_TIME_POSITION</label>
<select name = "FULL_TIME_POSITION">
<option value = "Yes">Yes</option>
<option value = "No">No</option>
```

```
</select>
<br>
<br>
<p> PREVAILING_WAGE <span><input type = "text" name = "PREVAILING_WAGE"/></span></p>
</br>
<br>
<p> YEAR <span><input type = "text" name = "YEAR"/></span></p>
</br>
<br>
<label for = "NEW_EMPLOYER">NEW_EMPLOYER</label>
<select name = "NEW_EMPLOYER">
<option value = "-Select-">-Select</option>
<option value = "University">University</option>
<option value = "Non University">Non University</option>
</select>
</br>
<br>
<label for = "OCCUPATION">OCCUPATION</label>
<select name = "OCCUPATION">
<option value = "-Select-">-Select-</option>
<option value = "Advance Sciences">Advance Sciences</option>
<option value = "Architecture & Engineering">Architecture & Engineering</option>
<option value = "Business Occupation">Business Occupation</option>
<option value = "computer occupations">computer occupations</option>
<option value = "Financial Occupation">Financial Occupation</option>
<option value = "Management Occupation">Management Occupation</option>
```

```
<option value = "Marketing Occupation">Marketing Occupation</option>
<option value = "Medical Occupations">Medical Occupations</option>
<option value = "Others">Others</option>
</select>
</br>

<br>
<label for = "STATE">STATE</label>
<select name = "STATE">
<option value = "-Select-">-Select-</option>
<option value = "MICHIGAN">MICHIGAN</option>
<option value = "TEXAS">TEXAS</option>
<option value = "JERSEY">JERSEY</option>
<option value = "COLORADO">COLORADO</option>
<option value = "FLORIDA">FLORIDA</option>
<option value = "CALIFORNIA">CALIFORNIA</option>
<option value = "PENNSYLVANIA">PENNSYLVANIA</option>
<option value = "MASSACHUSETTS">MASSACHUSETTS</option>
<option value = "VIRGINIA">VIRGINIA</option>
<option value = "WISCONSIN">WISCONSIN</option>
</select>
</br>

<br>
<button type="submit" >SUBMIT</button></br>
</form>
<b>{{showcase}}</b>
```

```
</div>
</body>
</html>
```

APP.PY

```
from flask import Flask,render_template,request


import pickle
import numpy as np


model = pickle.load(open('case.pkl','rb'))
app = Flask(__name__)


@app.route('/')
def home():
    return render_template('base.html')




@app.route('/prediction',methods=['POST'])
def prediction():
     '''    #
    For rendering results on HTML GUI
    '''
    #x_test = [[str(x) for x in request.form.values()]]
    a = request.form['FULL_TIME_POSITION']
    if (a == "Yes"):
```

```
    a = 0
if (a == "No"):
    a = 1
 m = request.form['PREVAILING_WAGE']
n = request.form['YEAR']

b = request.form['NEW_EMPLOYER']
if (b == "University"):
    b = 1
if (b == "Non University"):
    b = 0

c = request.form['OCCUPATION']
if (c == "Advance Sciences"):
    c = 0
if (c == "Architecture & Engineering"):
      c = 1
 if (c == "Business Occupation"):
    c = 2
if (c == "Financial Occupation"):
    c = 3
if (c == "Management Occupation"):
    c = 4
if (c == "Marketing Occupation"):
    c = 5
if (c == "Medical Occupations"):
    c = 6
```

```
if (c == "Others"):
    c = 7
if (c == "computer occupations"):
    c = 8
d = request.form['STATE']
if (d == "MICHIGAN"):
    d = 28
if (d == "TEXAS"):
    d = 42
if (d == "JERSEY"):
    d = 20
if (d == "COLORADO"):
    d = 6
if (d == "FLORIDA"):
    d = 11
if (d == "CALIFORNIA"):
    d = 4
if (d == "PENNSYLVANIA"):
    d = 39
if (d == "MASSACHUSETTS"):
    d = 26
if (d == "VIRGINIA"):
    d = 45
if (d == "WISCONSIN"):
    d = 47

total = [[int(a),int(b),int(m),int(n),int(c),int(d)]]
```

```
    output=prediction_1[0]
    if(output==0):
        pred="Visa CERTIFIED"
    else:
        pred="Visa DENIED"
    return render_template('base.html', showcase=pred)

'''@app.route('/predict_api',methods=['POST'])
def predict_api():

    For direct API calls trought request

    data = request.get_json(force=True)
    prediction = model.y_predict([np.array(list(data.values()))])

    output = prediction[0]
    return jsonify(output)
    '''

if __name__ == "__main__":
    app.run(debug=True)
```