1. Introduction

1.1. Introduction to supervised learning, regression, and logistic regression

Supervised learning, often known as supervised machine learning, is a machine learning and artificial intelligence subcategory. It is distinguished using labeled datasets to train algorithms that properly categorize data or predict outcomes. As input data is fed into the model, the weights are adjusted until the model is well fitted, which occurs as part of the cross-validation process. Supervised learning assists enterprises in solving a wide range of real-world issues on a large scale, such as categorizing spam in a distinct folder from your email. (Education, 2020)

In the discipline of machine learning, regression analysis is a key concept. It is classified as supervised learning since the algorithm is taught using both input characteristics and output labels. It aids in the establishment of a link between variables by estimating how one influences the other. (Kurama, 2020)

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). The logistic regression, like other regression studies, is a predictive analysis. Logistic regression is a data analysis technique that is used to define and explain the connection between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. (What is Logistic Regression? - Statistics Solutions, n.d.)

1.2. Introduction to prediction system

The term "prediction" refers to the output of an algorithm after it has been trained on a previous dataset and applied to new data in order to estimate the likelihood of a specific result, such as whether or not the employee will leave his/her job or not. For each record in the new data, the algorithm will provide probable values for an unknown variable, allowing the model builder to determine what that value will most likely be.
(Prediction, n.d.)

1.3. Introduction of the chosen problem domain

Employee attrition is a significant expense to a business, and forecasting such attritions is the most critical necessity of many firms' Human Resources departments. Many employees leave their job due to various factors in a short notice. This causes the company a big loss in monetary as well good employee aspect. Employee retention is only possible when the HR manager knows if the employee is leaving the job. It is very difficult to predict if any of the employee might leave the job or not. The HR can apply his employee retention measures only when he properly predicts if the employee wants to leave the job. It is very hard for a company if the employee who they rely on leaves the company. So, the solution to the problem is to apply employee retention measures but to apply retention measures firstly the prediction needs to be done to identify which employee is more prone to leave the job. Employee attrition must be decreased for any company or business as it increases cost and time of the organizations. So, proper system which will help the company to solve this problem should be made.

2. Background

2.1. Research on chosen topic

Logistic regression is a classification system that is used to determine the likelihood of event success and failure. It is used when the dependent variable is binary in nature (0/1, True/False, Yes/No). It allows you to categorize data into distinct classes by investigating the link between a collection of labelled data. It first learns a linear connection from the provided dataset before introducing a non-linearity in the form of the sigmoid function. (Rout, 2020)

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical.

For example,

To predict whether an email is spam (1) or (0)

Whether the tumor is malignant (1) or not (0)

Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time. From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1. (Swaminathan, 2018)

To make a prediction system, firstly the data needs to be collected. Then, the filtration or cleaning of the data needs to be done. After which the data will be divided into two parts, which are “Training” and “Testing” data. Then the model is built using the “Training Data set”. Then we can finally get the prediction or probability after applying the logistic regression algorithm. After all this we can also find the accuracy of the model by testing it. Logistic Regression is widely used in projects to make

recommendation systems like stock market price will pump or not, will there be a rainfall or not, will the employee leave the job or not and many other projects can be done using logistic regression.

2.2. Review and analysis of similar work in the problem domain:

2.2.1. Employee Attrition Analysis using Logistic Regression with R by Analytics Vidhya

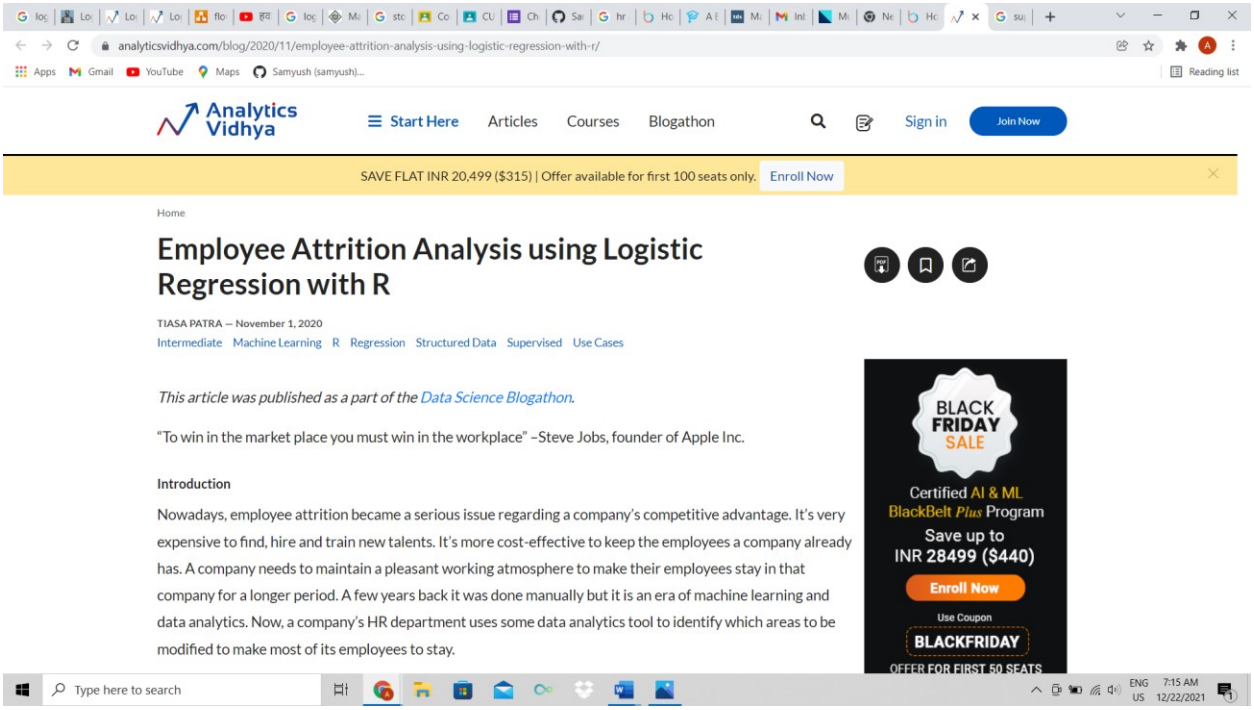


Figure 1: Research work1

(Patra, 2020)

This screenshot was taken from an article in analyticsvidhya.com. This article consists of the full documentation of a very similar project to mine. In the article they have explained why they are using logistic regression for the project and why it is best suited. This part made me realize that I have also chosen the best possible algorithms for my project. This article contains detailed graphical review and how the dataset they used is being properly used for prediction. This will help me a lot in my future work as well as this is very similar to my project and this article mostly answers many difficulties that I might face during the development of the project. The steps to get

the final result will also be similar and has been stated above in the first part of background section in this report.

2.2.2. HR Analytics- Why Do Employees Leave Prematurely by Mirdula

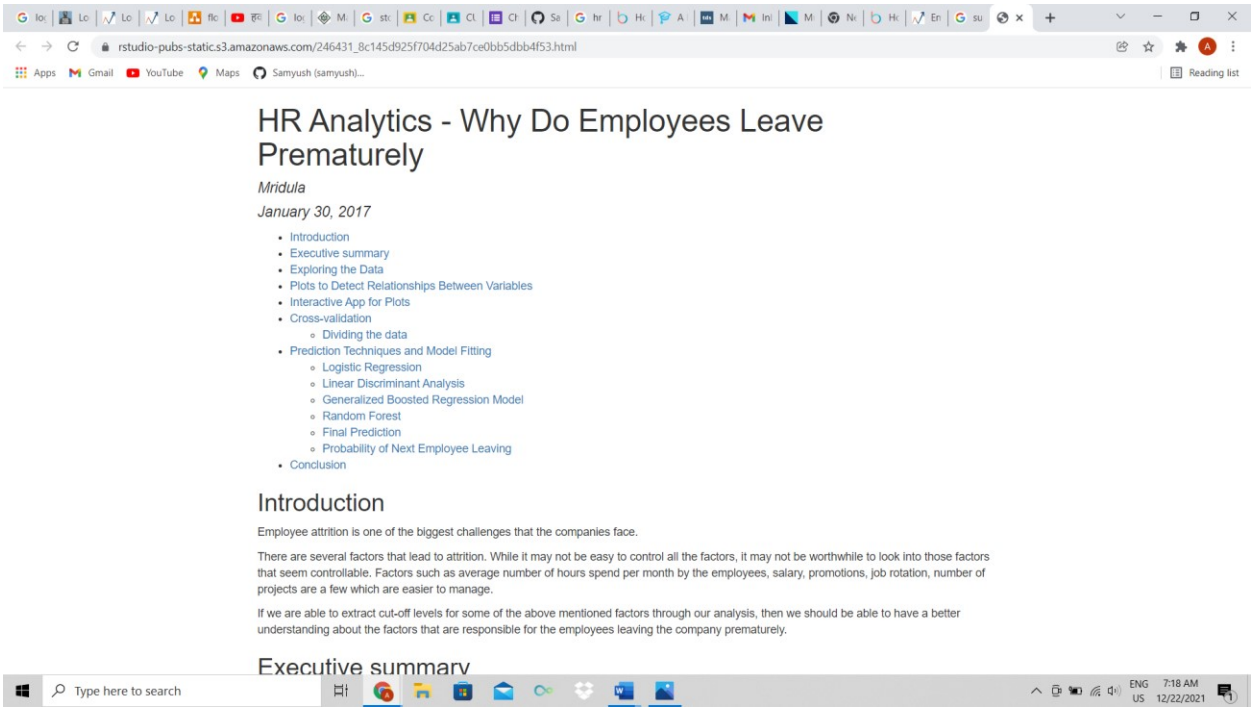


Figure 2: Research Work2

(HR Analytics - Why Do Employees Leave Prematurely, 2017)

The above-mentioned screenshot is also a similar project that does similar work just like the project that I am doing. The article says that employee attrition is one of the biggest challenges that the companies face. Thus, making this a very serious matter to solve. They have shown a detailed solution of how they have done the prediction of probability of the employee leaving. This article consists of all the detailed review of the work they have done for helping in employee retention. This article concludes by telling us why the best and most experienced employees leave prematurely with a set of predictions they made using the system they built.

2.3. Advantages/Disadvantages of the topic/problem domain:

The advantages of this project are that it helps the companies to make prediction of who is going to leave and who will stay in the job. This project will assist both small and big companies in predicting employee attrition. In some case scenarios when the company wants lesser number of employees than this kind of prediction will help them make their team shorter after knowing removing whom won't affect the employees lives as well.

In terms of the drawbacks, this project may not be 100% accurate which might cause the companies to take actions to stop someone who might be willing to stay in the job. Some employees who are newbies and have a very low salary might be predicted as most likely to leave which is not true as they are the ones who have just started and want to stay in the job. If the employees come to know about such project being used to know if they want to leave or not, then they can fake their data in order to show that they want to leave to gain employee retention policy benefits.

3. Solution (proposed solution to the chosen problem)

3.1. Explanation of the proposed solution/approach to solving the Problem

To apply company employee retention program. Firstly, it is required to predict which employees will leave the job. So, to get the prediction logistic regression is being used. Employee retention management entails taking intentional steps to keep employees engaged and focused so that they choose to remain employed and fully productive for the benefit of the firm. A thorough employee retention program may help to recruit and retain essential personnel, as well as reduce turnover and its associated expenses. All these factors contribute to an organization's overall productivity and business effectiveness. Retaining a quality employee is more efficient than recruiting, training, and orienting a substitute employee of the same caliber. (Managing for Employee Retention, n.d.)

By using logistic regression, at first, I will use a dataset with required features. Then, model will be created. The data is trained for proper efficiency. After all this the prediction is done which gives us the prediction whether the employee will leave or not. This way using the prediction system the company can predict whether the employee will leave the job and use employee retention measures which might help them in retaining the employee. In this way, the problem that was addressed before will be solved using logical regression which gives us good prediction.

3.2. Explanation of the AI algorithm/algorithms used

3.2.1. Logistic function:

Logistic function:

$$\frac{e^x}{1+e^x}$$

In logistic regression, we use the right-hand side of our logistic regression model results to give us the beta weights β (and ultimately the summed values) we need to plug into the logistic function and generate our prediction.

$$\frac{e^{\beta 0+\beta 1x1+\beta 2x2\ldots}}{1+e^{\beta 0+\beta 1x1+\beta 2x2\ldots}}$$

If you look carefully, you'll see that in this equation, we still have our series of input values and beta weights just as we did before in our logistic equation above.

The top piece of the logistic function $e^{\beta 0+\beta 1x1+\beta 2x2\ldots}$ gives us the odds of the event happening.

The bottom piece $1+e^{\beta 0+\beta 1x1+\beta 2x2\ldots}$ is just 1 + those odds.

Putting this all together, we have the the following relationship and can generate the predicted probability p of the outcome:

p=odds1+odds=eβ0+β1x1+β2x2…1+eβ0+β1x1+β2x2…

The upshot of the whole process, then, is that the result of the basic logistic formulation ex1+ex is equal to the probability of the "1" outcome that we are trying to predict for each observation in our data. We can therefore use the results of our logistic regression model to calculate the probability of the outcome for an individual given their predictor values. (Predictive HR Analytics: What is Logistic Regression? - HR Analytics 101, 2021)

3.2.2. Multi-variate logistic regression:

Logistic regression is a classification algorithm that uses supervised learning to predict the probability of a target variable. Multi-variate logistic regression has more than one input variable. This figure shows the classification with two independent variables, x_1 and x_2:

Logistic regression determines the weights b_0, b_1, and b_2 that maximize the LLF. Once you have b_0, b_1, and b_2, you can get:

The logit $f(x_1, x_2) = b_0 + b_1x_1 + b_2x_2$

The probabilities $p(x_1, x_2) = 1 / (1 + \exp(-f(x_1, x_2)))$

The dash-dotted black line linearly separates the two classes. This line corresponds to $p(x_1, x_2) = 0.5$ and $f(x_1, x_2) = 0$. (Stojiljkovic, n.d.)

3.3. Pseudocode of the solution:

```
Start
    Import Libraries
    Import Dataset
        Read CSV
    Clean Data
        Drop [Column_name]
    Identify Independent and Dependent Variables
        Initialize dependent and independent variable
    Create A Training Model
        Initialize training data and testing data
    Fit training data in the Model
    Train the Model
    Check the accuracy of the model
        If accuracy >= threshold value Then Model accepted
        Else Make New Model
    Input required data from the user
        Employee_detail=Input (“Enter Employee Details”)
    Perform Prediction
    Display Prediction
        Print Prediction
End
```

3.4. Diagrammatic representations of the solution:

3.4.1. Flowchart:

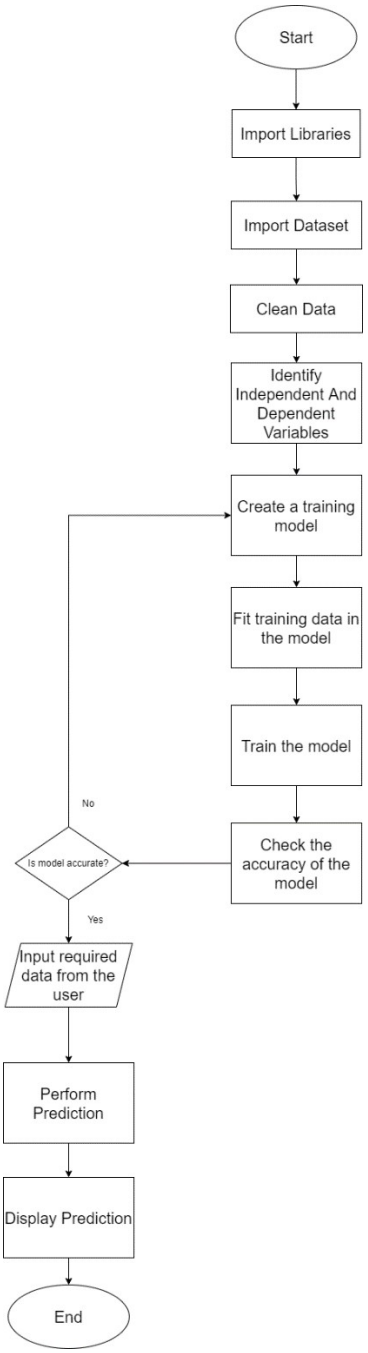


Figure 3: Flowchart

The above figure shows the graphical representation of the actual system which is going to be developed. It represents the summarization of the working mechanism of the whole system.

4. Development Process

The project has been done here is Employee attrition prediction system. Data can be used in many ways for predicting and analyzing different things in today's world which was not possible before. HR analysis can play a huge role for making decisions regarding the human resources of a business. In this project I have tried to make a system which predicts if an employee wants to leave the job or not. Using a dataset based on employees wanting to leave or not according to different factors the prediction system is built.

I developed a model to predict the probability of an employee leaving the job by taking in inputs such as satisfaction level, total monthly hours of work, salary level and if the employee has received any promotion in the last five years. Finally, based on these factors the prediction is given of what percentage the employee will stay in the job. So, based on the ability of the model to make proper prediction, it can be used by different companies for HR analysis and prevent HT attrition and be fast with employee retention measures to stop their valuable employees from leaving the job.

4.1. Libraries Used

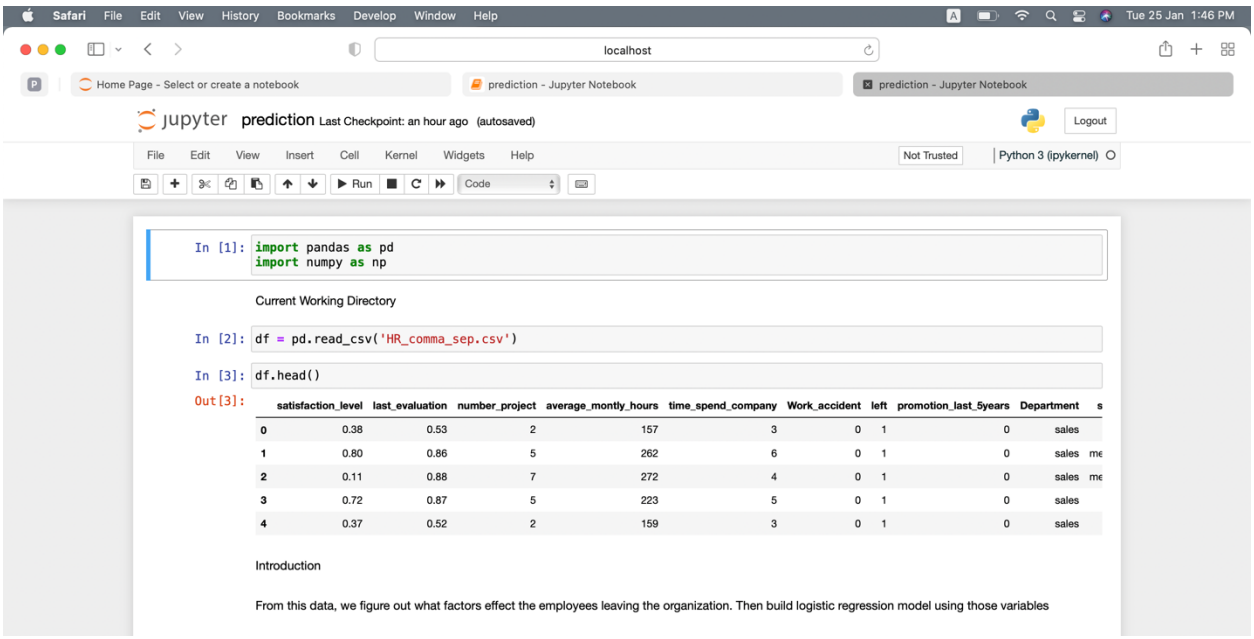
Pandas:

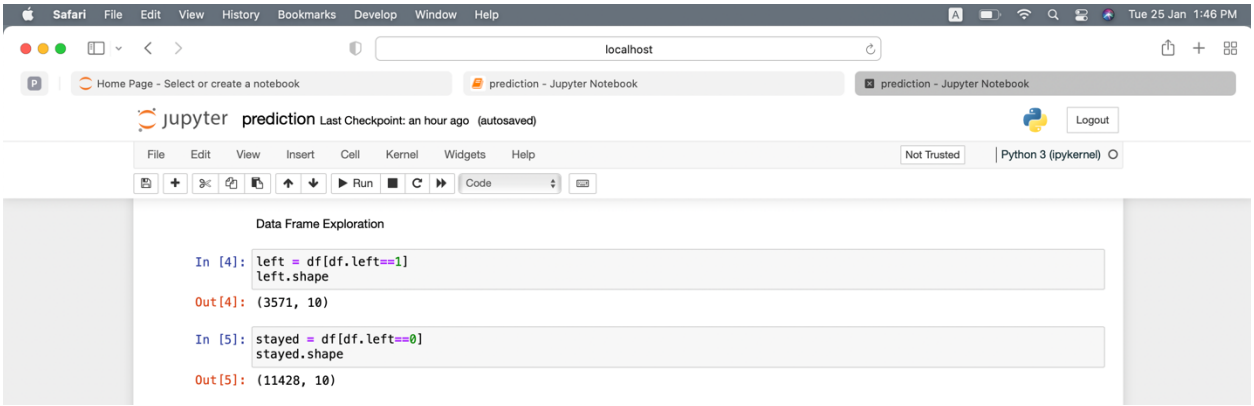
Numpy:

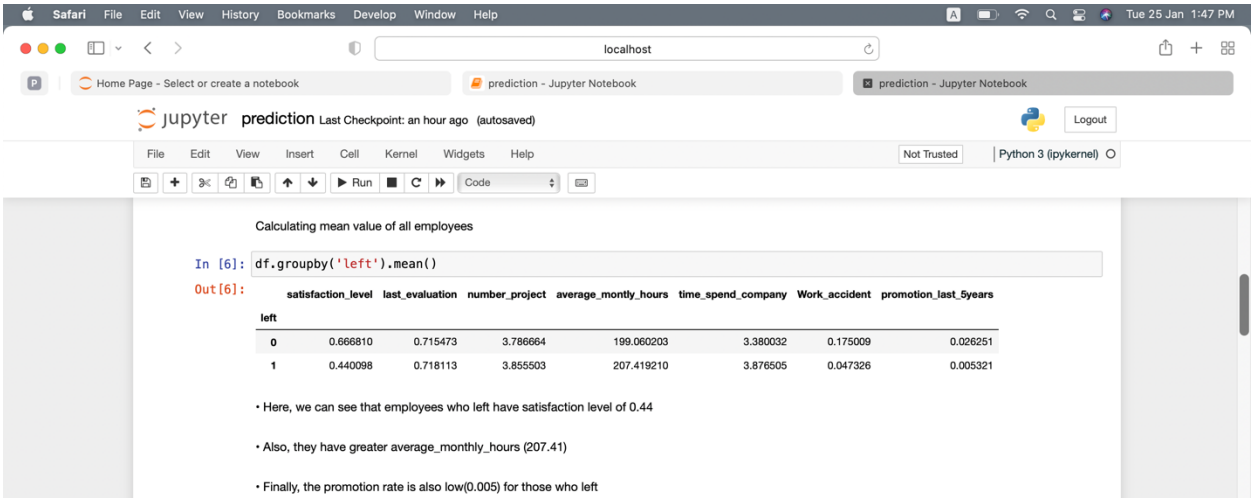
4.2. Tools Used

Jupyter Notebook:

5. Achieved Results


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:46 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [1]:
import pandas as pd
import numpy as np
Current Working Directory
In [2]:
df = pd.read_csv('HR_comma_sep.csv')
In [3]:
df.head()
Out[3]:
satisfaction_level last_evaluation number_project average_montly_hours time_spend_company Work_accident left promotion_last_5years Department s
0 0.38 0.53 2 157 3 0 1 0 sales
1 0.80 0.86 5 262 6 0 1 0 sales me
2 0.11 0.88 7 272 4 0 1 0 sales me
3 0.72 0.87 5 223 5 0 1 0 sales
4 0.37 0.52 2 159 3 0 1 0 sales
Introduction
From this data, we figure out what factors effect the employees leaving the organization. Then build logistic regression model using those variables


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:46 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Data Frame Exploration
In [4]:
left = df[df.left==1]
left.shape
Out[4]:
(3571, 10)
In [5]:
stayed = df[df.left==0]
stayed.shape
Out[5]:
(11428, 10)


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:47 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Calculating mean value of all employees
In [6]:
df.groupby('left').mean()
Out[6]:
satisfaction_level last_evaluation number_project average_montly_hours time_spend_company Work_accident promotion_last_5years
left
0 0.666810 0.715473 3.786664 199.060203 3.380032 0.175009 0.026251
1 0.440098 0.718113 3.855503 207.419210 3.876505 0.047326 0.005321
• Here, we can see that employees who left have satisfaction level of 0.44
• Also, they have greater average_monthly_hours (207.41)
• Finally, the promotion rate is also low(0.005) for those who left

Safari File Edit View History Bookmarks Develop Window Help Tue 25 Jan 1:47 PM

localhost

Home Page - Select or create a notebook | prediction - Jupyter Notebook | prediction - Jupyter Notebook

jupyter **prediction** Last Checkpoint: an hour ago (autosaved) Logout

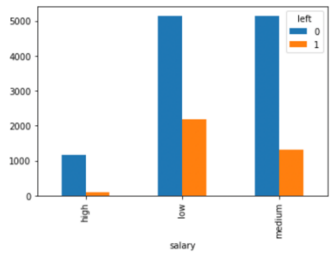
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

Check if salary effects the employees leaving

```
In [7]: pd.crosstab(df.salary,df.left).plot(kind='bar')
```

```
Out[7]: <AxesSubplot:xlabel='salary'>
```



So it seems employees with higher salaries tend to stay

Safari File Edit View History Bookmarks Develop Window Help Tue 25 Jan 1:48 PM

localhost

Home Page - Select or create a notebook | prediction - Jupyter Notebook | prediction - Jupyter Notebook

jupyter **prediction** Last Checkpoint: an hour ago (autosaved) Logout

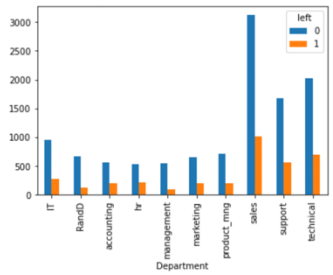
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

Check if department effects the employees leaving

```
In [8]: pd.crosstab(df.Department,df.left).plot(kind='bar')
```

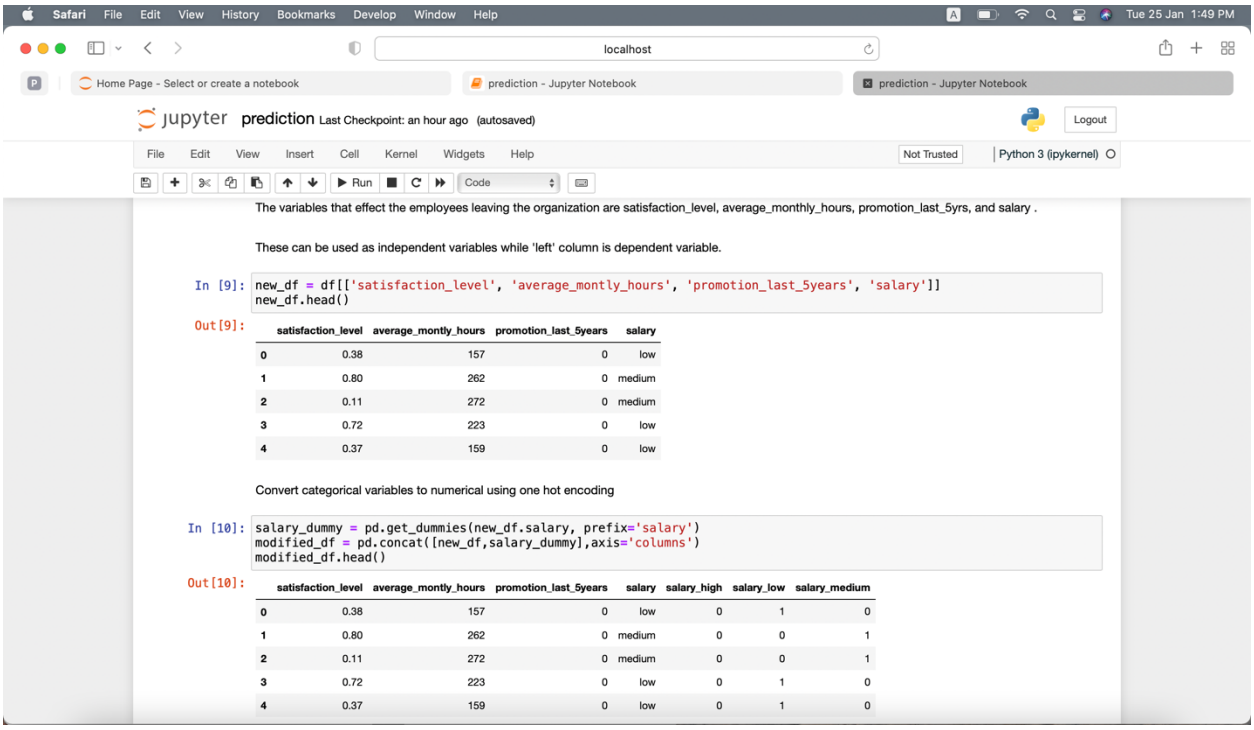
```
Out[8]: <AxesSubplot:xlabel='Department'>
```

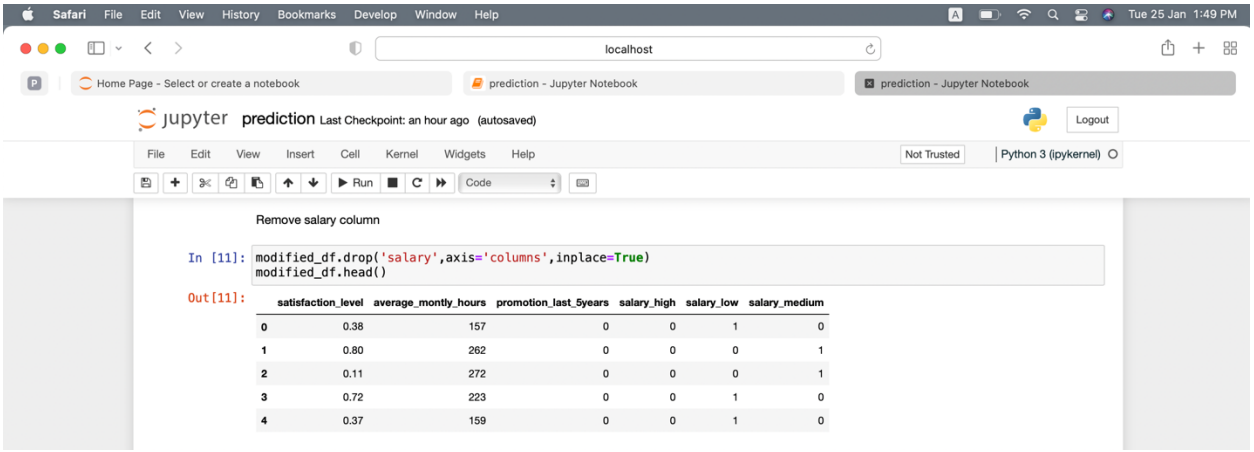


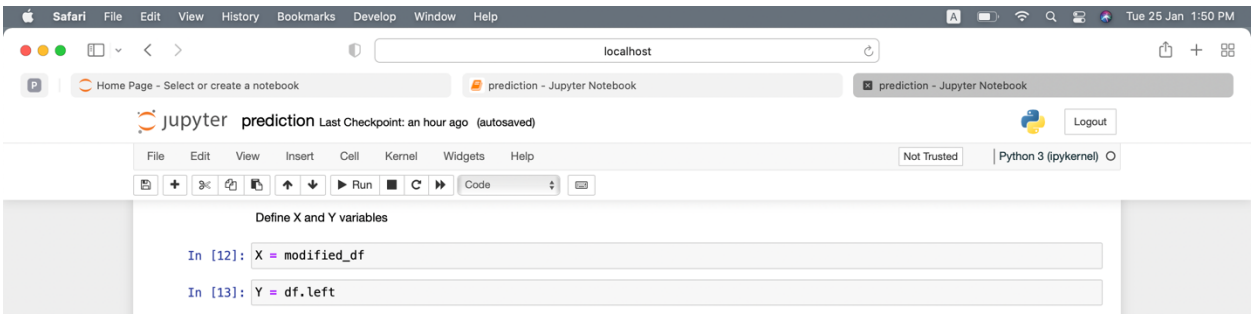
It shows that sales department has high leaving rate, but its number is high as well. So not much of a deciding factor.

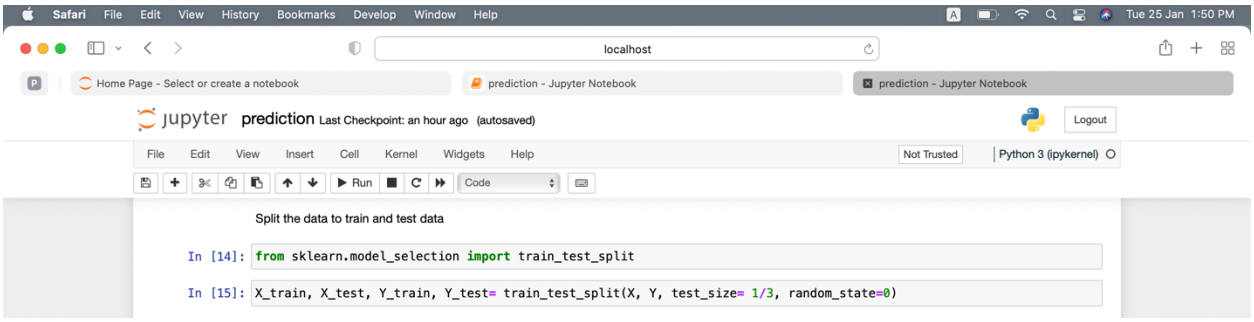
The variables that effect the employees leaving the organization are satisfaction_level, average_monthly_hours, promotion_last_5yrs, and salary .

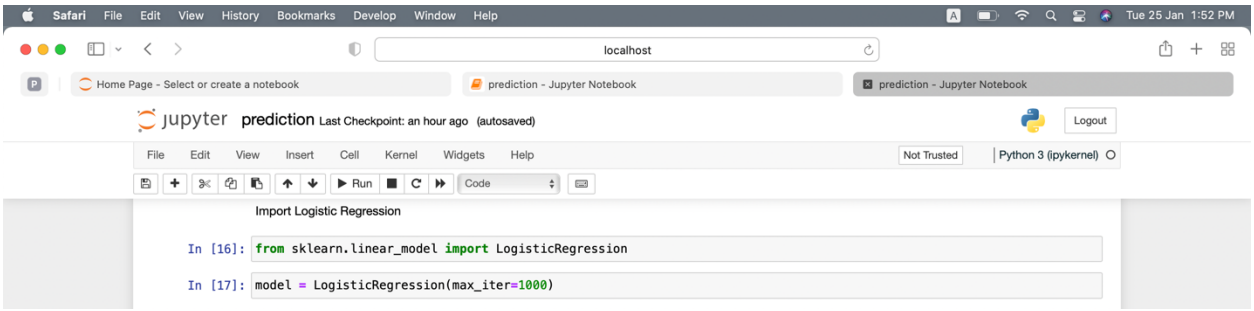
These can be used as independent variables while 'left' column is dependent variable.

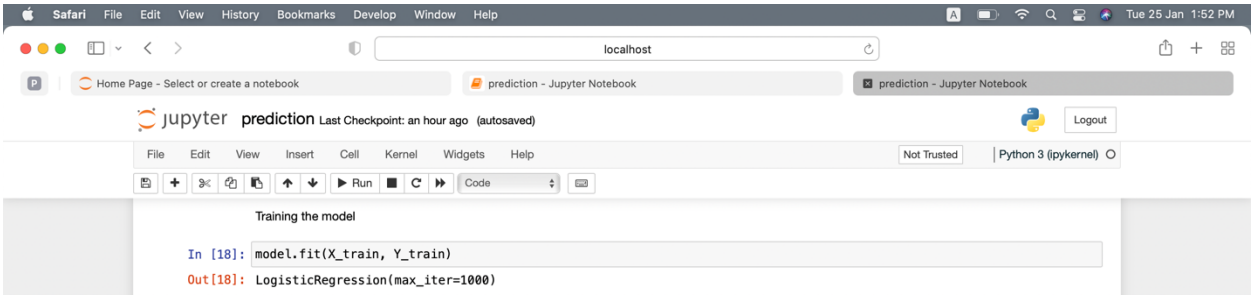

Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:49 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
The variables that effect the employees leaving the organization are satisfaction_level, average_monthly_hours, promotion_last_5yrs, and salary .
These can be used as independent variables while 'left' column is dependent variable.
In [9]:
new_df = df[['satisfaction_level', 'average_montly_hours', 'promotion_last_5years', 'salary']]
new_df.head()
Out[9]:
satisfaction_level average_montly_hours promotion_last_5years salary
0 0.38 157 0 low
1 0.80 262 0 medium
2 0.11 272 0 medium
3 0.72 223 0 low
4 0.37 159 0 low
Convert categorical variables to numerical using one hot encoding
In [10]:
salary_dummy = pd.get_dummies(new_df.salary, prefix='salary')
modified_df = pd.concat([new_df,salary_dummy],axis='columns')
modified_df.head()
Out[10]:
satisfaction_level average_montly_hours promotion_last_5years salary salary_high salary_low salary_medium
0 0.38 157 0 low 0 1 0
1 0.80 262 0 medium 0 0 1
2 0.11 272 0 medium 0 0 1
3 0.72 223 0 low 0 1 0
4 0.37 159 0 low 0 1 0

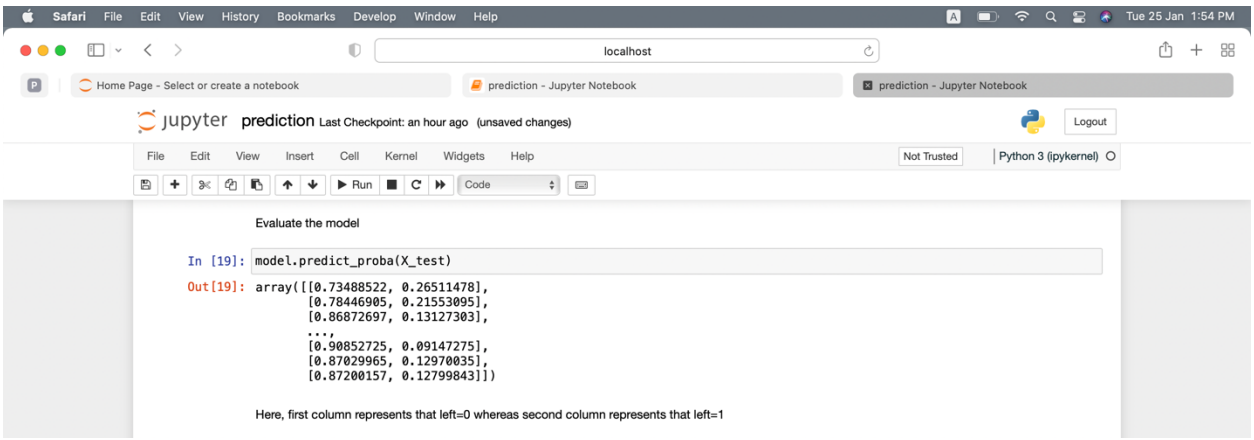

Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:49 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Remove salary column
In [11]:
modified_df.drop('salary',axis='columns',inplace=True)
modified_df.head()
Out[11]:
satisfaction_level average_montly_hours promotion_last_5years salary_high salary_low salary_medium
0 0.38 157 0 0 1 0
1 0.80 262 0 0 0 1
2 0.11 272 0 0 0 1
3 0.72 223 0 0 1 0
4 0.37 159 0 0 1 0

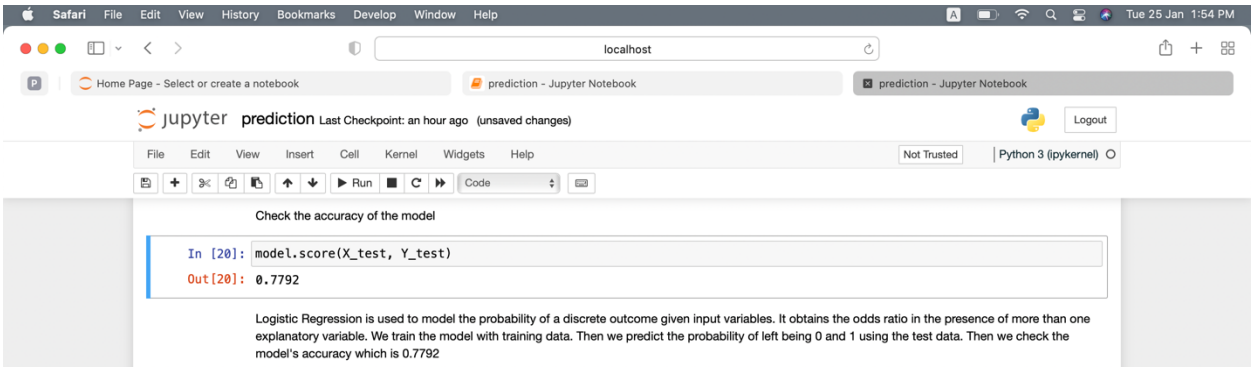

Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:50 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Define X and Y variables
In [12]:
X = modified_df
In [13]:
Y = df.left

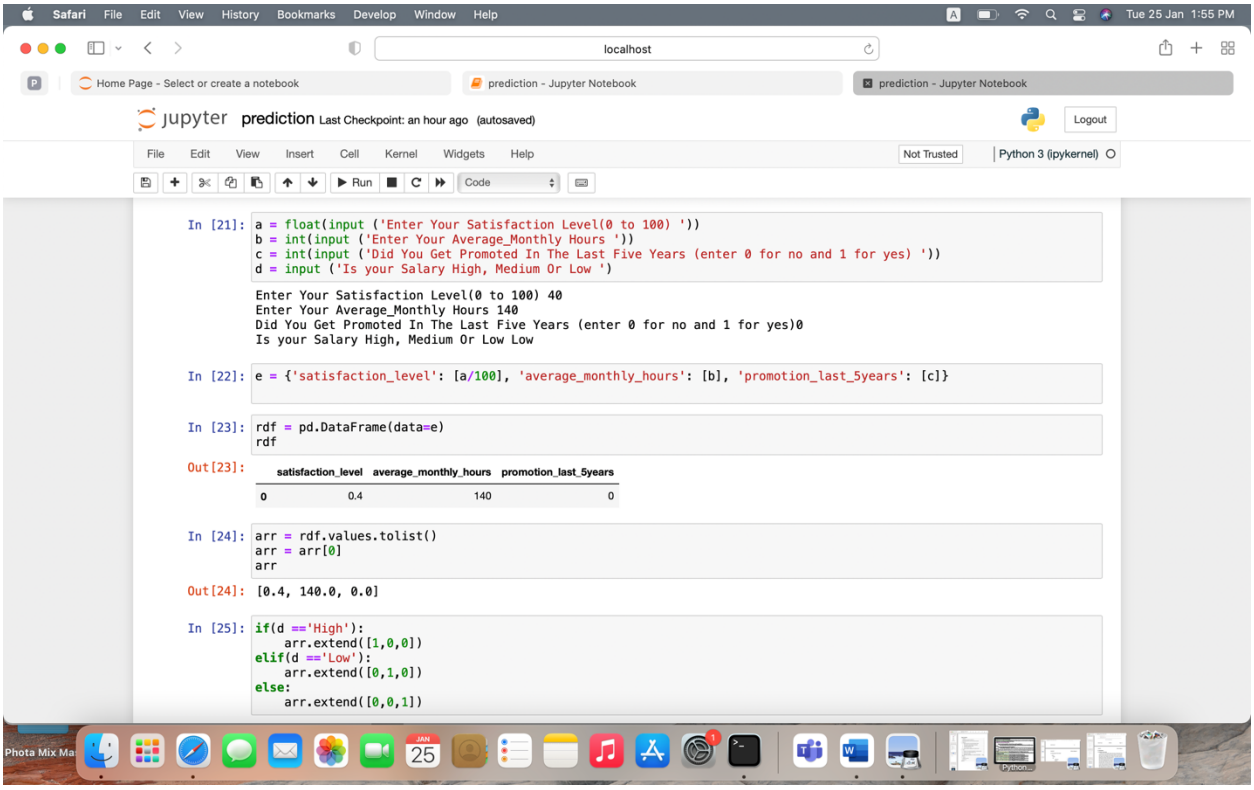

Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:50 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Split the data to train and test data
In [14]: from sklearn.model_selection import train_test_split
In [15]: X_train, X_test, Y_train, Y_test= train_test_split(X, Y, test_size= 1/3, random_state=0)

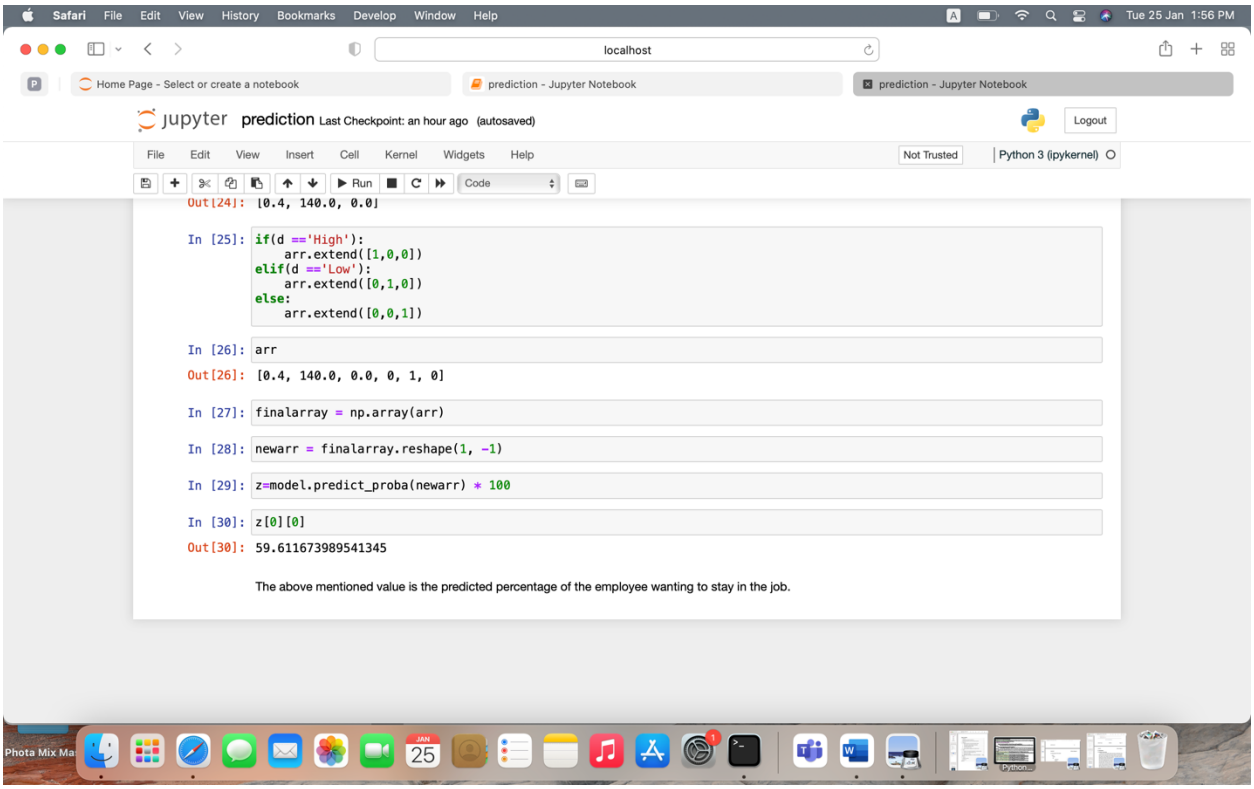

Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:52 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Import Logistic Regression
In [16]: from sklearn.linear_model import LogisticRegression
In [17]: model = LogisticRegression(max_iter=1000)


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:52 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Training the model
In [18]: model.fit(X_train, Y_train)
Out[18]: LogisticRegression(max_iter=1000)


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:54 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Evaluate the model
In [19]: model.predict_proba(X_test)
Out[19]: array([[0.73488522, 0.26511478],
[0.78446905, 0.21553095],
[0.86872697, 0.13127303],
...,
[0.90852725, 0.09147275],
[0.87029965, 0.12970035],
[0.87200157, 0.12799843]])
Here, first column represents that left=0 whereas second column represents that left=1


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:54 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Check the accuracy of the model
In [20]:
model.score(X_test, Y_test)
Out[20]:
0.7792
Logistic Regression is used to model the probability of a discrete outcome given input variables. It obtains the odds ratio in the presence of more than one explanatory variable. We train the model with training data. Then we predict the probability of left being 0 and 1 using the test data. Then we check the model's accuracy which is 0.7792


Safari File Edit View History Bookmarks Develop Window Help
Tue 25 Jan 1:55 PM
localhost
Home Page - Select or create a notebook
prediction - Jupyter Notebook
prediction - Jupyter Notebook
jupyter prediction Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [21]:
a = float(input ('Enter Your Satisfaction Level(0 to 100) '))
b = int(input ('Enter Your Average_Monthly Hours '))
c = int(input ('Did You Get Promoted In The Last Five Years (enter 0 for no and 1 for yes) '))
d = input ('Is your Salary High, Medium Or Low ')
Enter Your Satisfaction Level(0 to 100) 40
Enter Your Average_Monthly Hours 140
Did You Get Promoted In The Last Five Years (enter 0 for no and 1 for yes)0
Is your Salary High, Medium Or Low Low
In [22]:
e = {'satisfaction_level': [a/100], 'average_monthly_hours': [b], 'promotion_last_5years': [c]}
In [23]:
rdf = pd.DataFrame(data=e)
rdf
Out[23]:
satisfaction_level average_monthly_hours promotion_last_5years
0 0.4 140 0
In [24]:
arr = rdf.values.tolist()
arr = arr[0]
arr
Out[24]:
[0.4, 140.0, 0.0]
In [25]:
if(d =='High'):
arr.extend([1,0,0])
elif(d =='Low'):
arr.extend([0,1,0])
else:
arr.extend([0,0,1])



6. Conclusion

6.1. Analysis of the work done

Until now the initial documentation has been done where mostly the research have been done for how to solve the problem that has been stated at the start of the project. Firstly, the problem which requires a solution using AI is chosen and described with the required AI topics that will be used to complete the project. Then, review and analysis of existing work in the problem domain was done. Later, the proposed solution to the problem was explained with the explanation of the AI algorithm that was used. Then the pseudocode of the proposed solution was written. After all this, flowchart and state transition diagram were made which are the diagrammatic representations of the solution. These are the things that were done for the first coursework and with the completion of this I have a clear idea of how the second coursework should be done.

6.2. How the solution addresses real world problems

This project addresses the problems faced by many companies in the real world when their employee suddenly leaves the job with only a short notice and the company, or the HR department of the company cannot apply employee retention measures to keep the employee from not leaving the job. The solution of the project is to give the company a prediction of which of the employees might leave the job and the company can use the prediction to stop the employee from leaving.

5. References:

Education, I., 2020. What is Supervised Learning?. [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/supervised-learning> [Accessed 22 December 2021].

Kurama, V., 2020. Regression in Machine Learning: What it is and Examples of Different Models. [online] Built In. Available at: <https://builtin.com/datascience/regression-machine-learning> [Accessed 22 December 2021].

Statistics Solutions. n.d. What is Logistic Regression? - Statistics Solutions. [online] Available at: <https://www.statisticssolutions.com/free-resources/directory-ofstatistical-analyses/what-is-logistic-regression/> [Accessed 22 December 2021].

DataRobot AI Cloud. n.d. Prediction. [online] Available at: <https://www.datarobot.com/wiki/prediction/> [Accessed 22 December 2021].

Rout, A., 2020. *Advantages and Disadvantages of Logistic Regression - GeeksforGeeks*. [online] GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logisticregression/> [Accessed 22 December 2021].

Swaminathan, S., 2018. *Logistic Regression — Detailed Overview*. [online] Medium. Available at: <https://towardsdatascience.com/logistic-regression-detailed-overview46c4da4303bc> [Accessed 22 December 2021].

Patra, T., 2020. *Employee Attrition Analysis using Logistic Regression with R*. [online]

Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2020/11/employee-attrition-analysis-usinglogistic-regression-with-r/> [Accessed 22 December 2021].

Rstudio-pubs-static.s3.amazonaws.com. 2017. HR Analytics - Why Do Employees Leave Prematurely. [online] Available at: <https://rstudio-pubsstatic.s3.amazonaws.com/246431_8c145d925f704d25ab7ce0bb5dbb4f53.html> [Accessed 22 December 2021].

Shrm.org. n.d. Managing for Employee Retention. [online] Available at: <https://www.shrm.org/resourcesandtools/tools-andsamples/toolkits/pages/managingforemployeeretention.aspx> [Accessed 22 December 2021].

HR Analytics 101. 2021. Predictive HR Analytics: What is Logistic Regression? - HR Analytics 101. [online] Available at: <https://hranalytics101.com/predictive-hranalytics-what-is-logistic-regression/> [Accessed 22 December 2021].

Stojiljkovic, M., n.d. Logistic Regression in Python – Real Python. [online] Realpython.com. Available at: <https://realpython.com/logistic-regressionpython/?fbclid=IwAR1PxxD-EFUAF0SfgmQABKo63P9J0AwyDFeYLFYnDyFejBwtydHeFQtI3JY#classification> [Accessed 22 December 2021].