EMPLOYEE ATTRITION PREDICTION

Using Machine Learning

Developed by:ASMA,BODAGAM PRAGATHI,NEELA NAVEEN, CHUNCHU OM MAHESHWERI,ALUKAPELLI SUMANTH KUMAR

Smart Bridge-Remote Summer Internship Program

1.INTRODUCTION:

Employee attrition can be defined as the loss of the employee due to any of the following reasons:personal reasons,low salary,and bad business environment.Employee attrition can be involuntary attrition occurs when employee are terminated by their employer for different reasons,such as low employee performance or business requirements.In voluntary attrition,on the other hand,high performing employees decide to leave the company of their own volition despite the company's attempt to retain them.From early retirement or joboffers from other firms .Although companies that realise the importance of their employee usually invest in their workforce by providing substantial training and a great working environment,they too suffer from voluntary attrition and the loss of the talented employees.Another issue,hiring replacements,imposes high cost on the company,including the cost of interviewing,hiring and training.

Employee attrition predicting at a company will help management act faster by enhancing their internal policies and stratgies.Where talented employee with a risk of leaving can be offered several propositions,such as a salary increase or proper training,to reduce their likelihood of leaving.Using machine learning models can help companies predict employees attrition.Using the historical data kept in the human resources(HR) departements,analysts can bulid and train a machine learning model can predict the employees who areleaving the company. Suchmodels are trained to examine the correlation between the features of the both active and terminated employees.

1.1OVERVIEW:

Nowaday,Employee Attrition Prediction become a major problem in the organizations.Employee Attrition is a big issue for the organizations specially when trained,technical and key employees leave for the better opportunity from the oraganization.this result in financial loss to replace a trained employee.therfore,we use the current and past employee data to analyse the common reasons for employee attrition.for the prevention of employee attrition,we applied a well known Regression,Classsification SVM,Random forest,KNN.For this we implement feature selection method on data and analysis the result to prevent employee attrition.This is help to compaines to predict employee attrition prediction,and also helpful to their economic growth.

1.2 Purpose:

Our aim from the project is to make use of pandas,numpy,matplotib, libraries from the python to extract the libraies fro machine learning for the employee attrition prediction.

And in the end,to predict whether the employee stay in company or not using voting ensemble techniques of combining the prediction from the multiple machine learning algorithms and the conclusions.

2.LITERATURE SURVEY :

Employee attrition refers to the gradual loss of employees over time.Most literature on employee attrition cateogorizes either voluntary or involuntary.Organizations try to prevent employee attrition by usin machine learning to predict the risk of an employee leaving,and then takr pro-active steps for preventing such an incident.

2.1EXISTING PROBLEM:

Attrition in human resoures refers to the gradual loss of the employee over time i.e.,resignation ang retirement.A major problem in high employee attrition is its cost to an organization,Job postings,hiring processes, lack of feedback and reecognition ,little opportunity for decision making and paperwork and new hire training are some ofthe common expenses of losing the employees and replacing them.Additionally,regular employee

turnover prohibits your organization from increasing its collective knowledge base and experience over the time.Human Resource analytics (HR analytics) should be implemented do as to overcome this problem.

2.2 PROPOSED SOLUTION:

Machine Learning:

Machine learning method which efficiently performs both classification and regression.Here we use regression for employee attrition prediction.

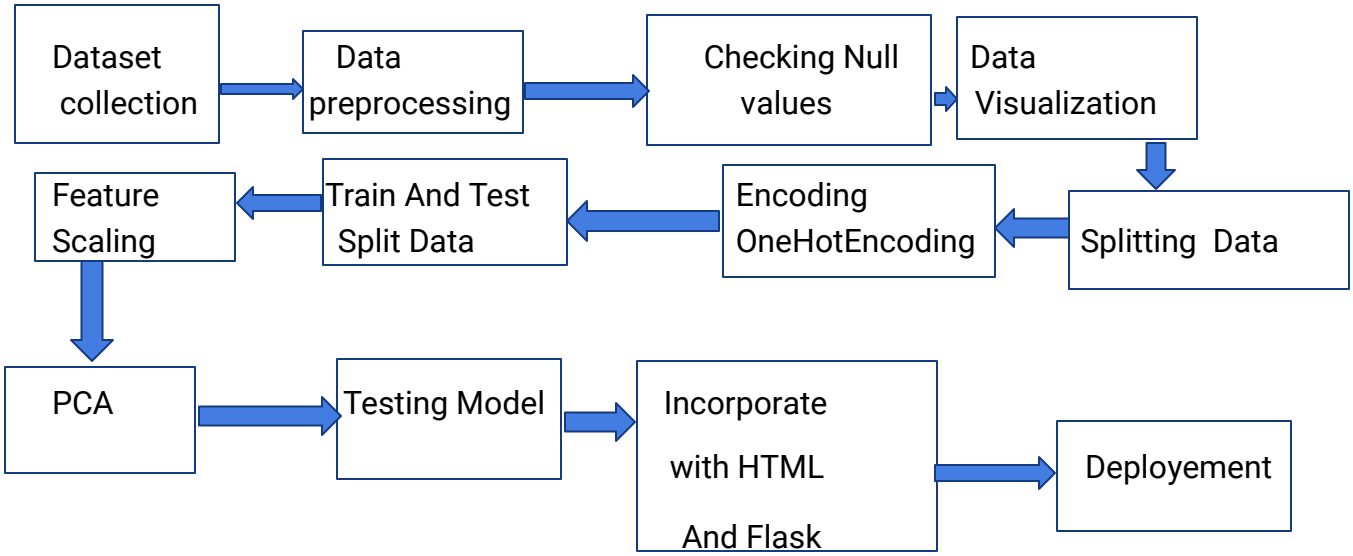
we have created an UI using the Flask for the employee attrition prediction, this UI will allow the users to predict the employee status very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details into the UI in real time it'll give the predicted value like if the continuous in job or they want to resign.

Byusing this machine learning model we are try to predict which valuable employees are probable to leave the organization is lagging behind.This is model cn be used by the Human Resoure departments of the organization to form efficient strategies to retain the valuable employees before they start lokking for new jobs like by providing a hike in their salary,offering promotions if necessary travel and stay abroad or start the hiring process. Basically this model will give the predicted by providing the encouragement and recognition and hiring the right person,providing the allow flexible work schedules.Establish the clear communication channels and enhance employee satisfaction focused moniting,providing the higher level opportunities.

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 Regression alogrithm, it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. thats how the prediction work great with the Regession Algorithm.

3.1 Block Diagram



3.2 SOFTWARE DESIGNING:

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

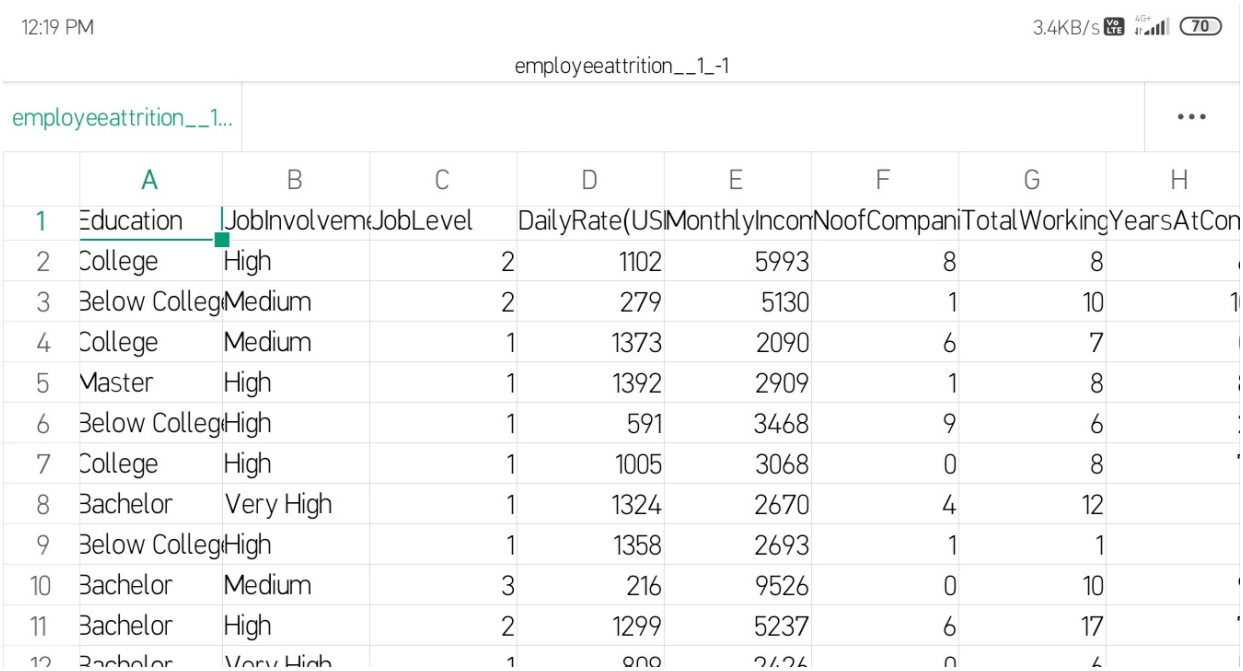
We developed this loan status prediction by using the Python language which is a interpreted and high level programming language and usng the Machine Learning algorithms. for coding we used the Jupyter Notebook environment of the Anaconda distributions and the Spyder, it is an integrated scientific programming 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 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 is HTML by creating the templates to use in th functions of the Flask and HTML.

4. EXPERIMENTAL INVESTIGATION:

In this paper, the dataset we used is derived from https://drive.google.com/drive/recent . Those attributes were shown below in the screenshot of the data set we used.



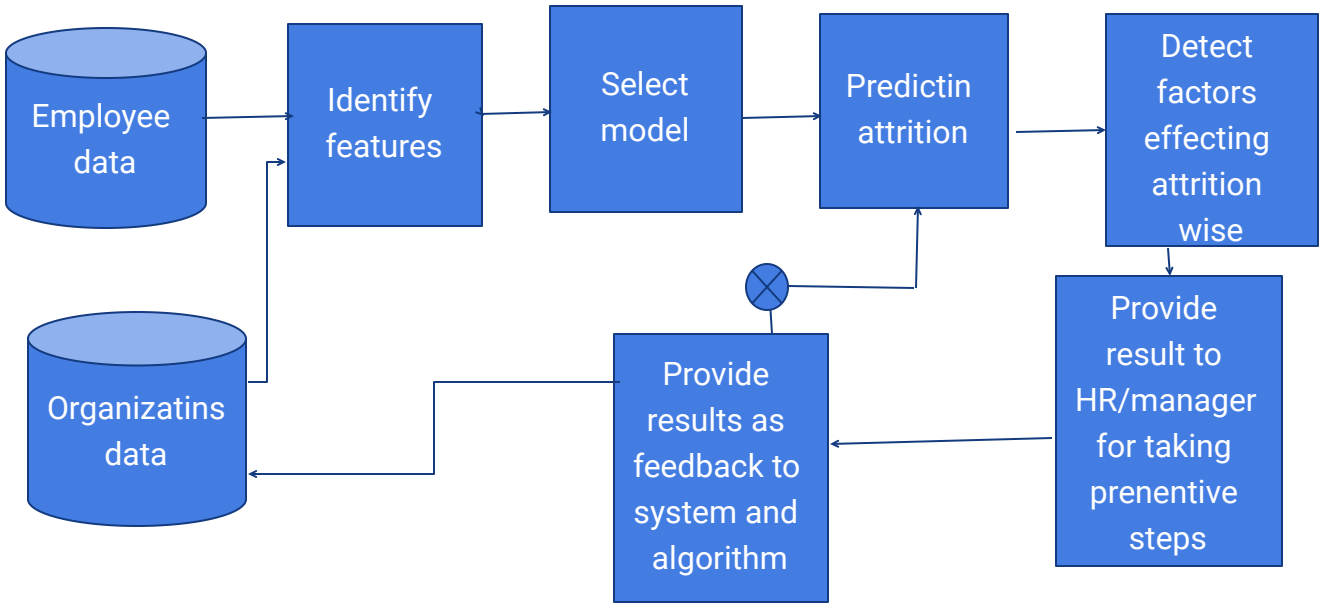
12:19 PM 3.4KB/s

employeeattrition__1_-1

employeeattrition__1...

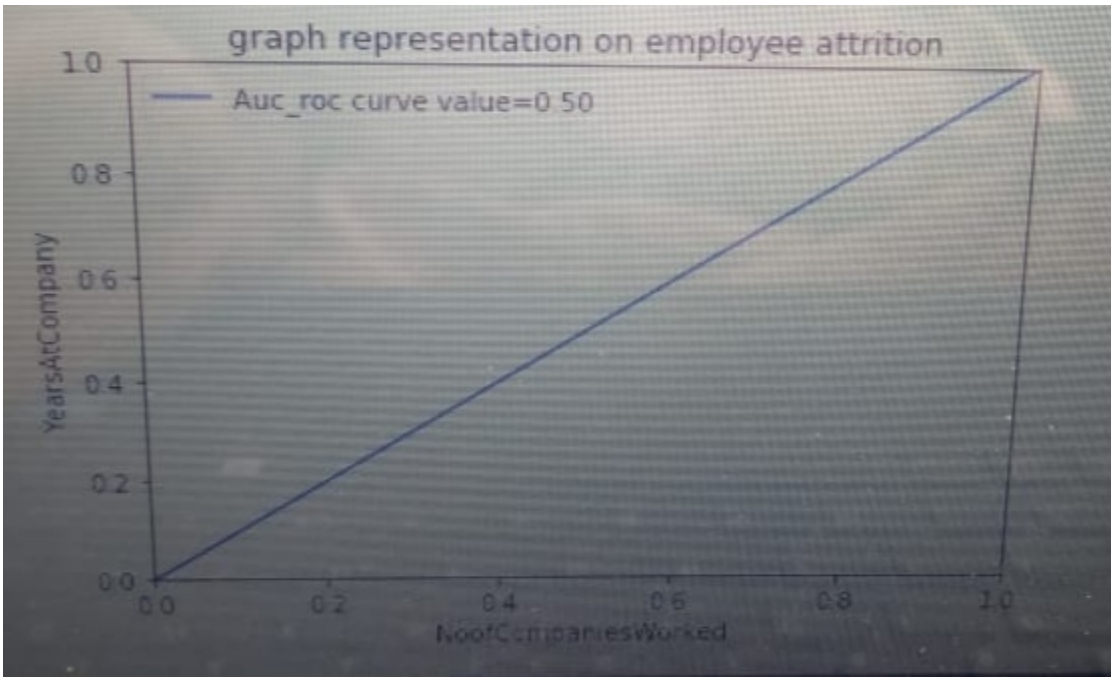
	A	B	C	D	E	F	G	H
1	Education	JobInvolveme	JobLevel	DailyRate(US	MonthlyIncon	NoofCompani	TotalWorking	YearsAtCon
2	College	High	2	1102	5993	8	8	
3	Below Colleg	Medium	2	279	5130	1	10	1
4	College	Medium	1	1373	2090	6	7	
5	Master	High	1	1392	2909	1	8	
6	Below Colleg	High	1	591	3468	9	6	
7	College	High	1	1005	3068	0	8	
8	Bachelor	Very High	1	1324	2670	4	12	
9	Below Colleg	High	1	1358	2693	1	1	
10	Bachelor	Medium	3	216	9526	0	10	
11	Bachelor	High	2	1299	5237	6	17	
12	Bachelor	Very High	1	809	2426	0	6	

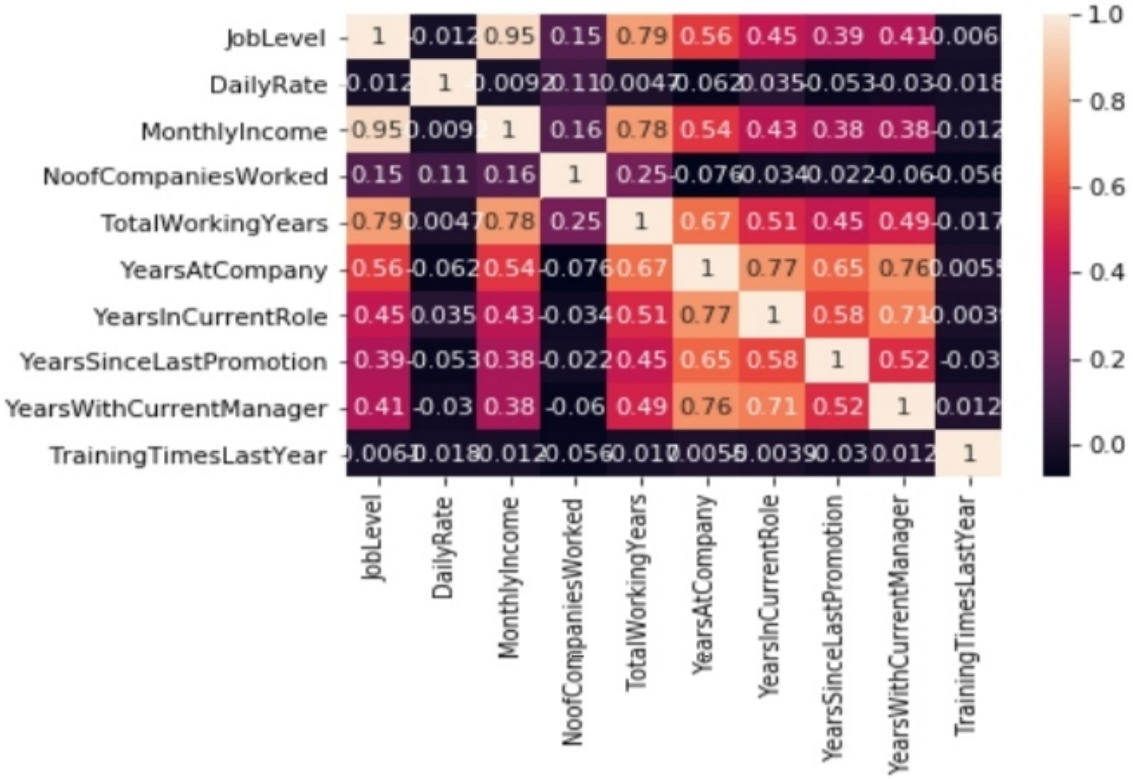
5.FLOWCHART:

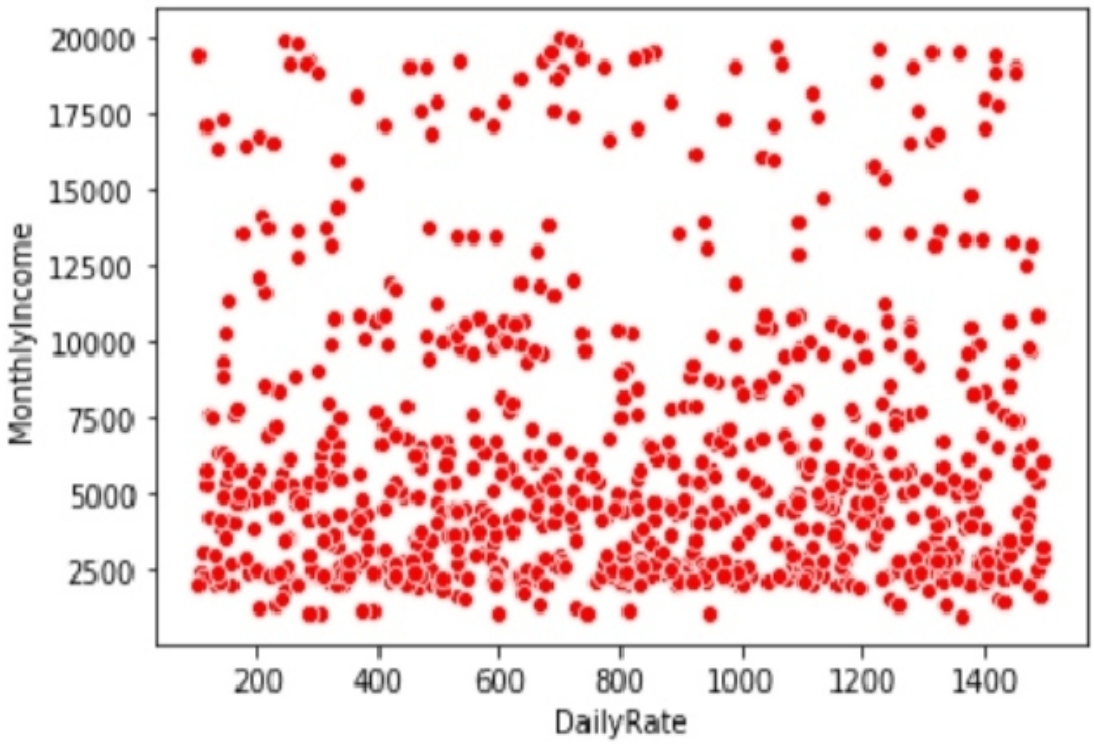


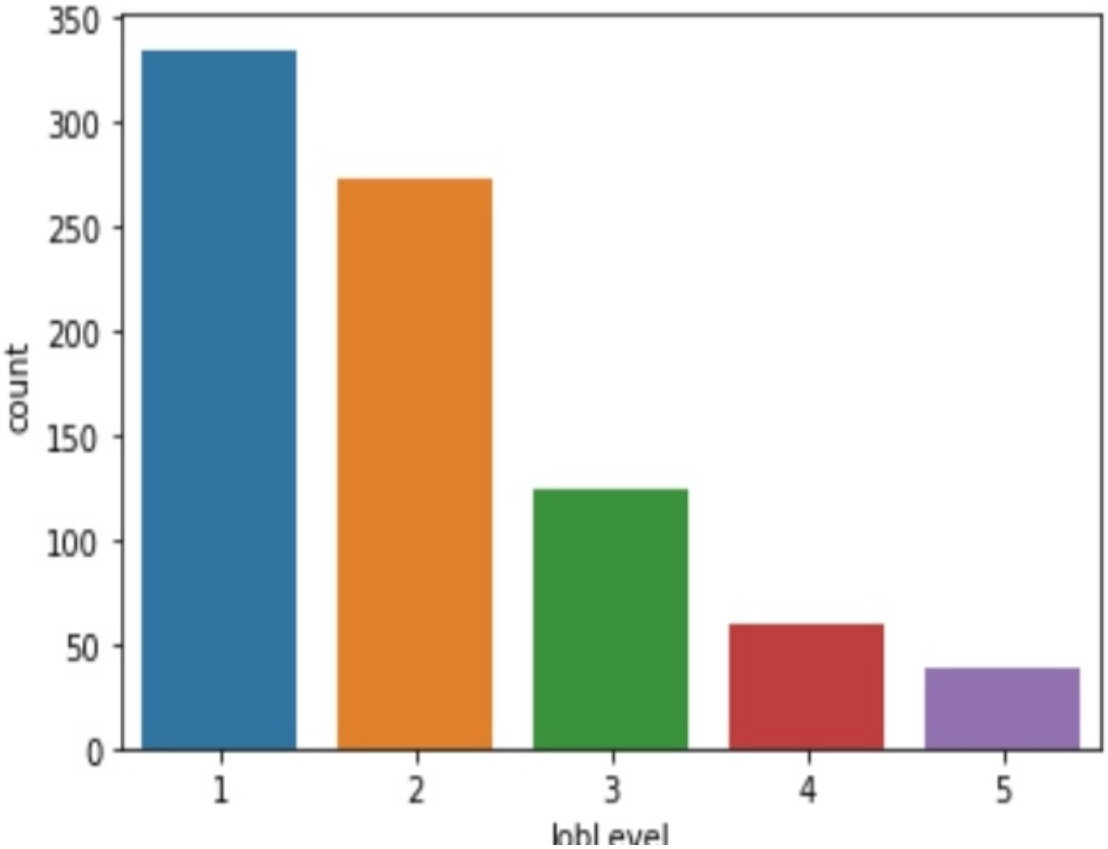
6.RESULT:

This section presents the result of the various regression used for the prediction.we have selected number of emplyees works in company.
In this paper, the Regression algorithm is used to predict its performance, and compared with another machine learning methods . The obtained results are displayed in Table below. The results show that, the performance Regression have comparable performance than that of logistic regression, random forest, SVM and decision tree, but the Regression still performs the best, with an accuracy of 82%. The ROC curve of the prediction model based on Regression are all above 0.50, indicating that the model has strong ability of generalization.
The given pie shows whos the employees likely how they work and their deatils from the data set we've took for the prediction.The ROC curve show the no of compaines worked and years at company.


graph representation on employee attrition
Auc_roc curve value=0 50
YearsAtCompany
NoofCompaniesWorked
0.0
0.2
0.4
0.6
0.8
1.0


JobLevel
DailyRate
MonthlyIncome
NoofCompaniesWorked
TotalWorkingYears
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrentManager
TrainingTimesLastYear
1.0
0.8
0.6
0.4
0.2
0.0


MonthlyIncome
20000
17500
15000
12500
10000
7500
5000
2500
200
400
600
800
1000
1200
1400
DailyRate


count
350
300
250
200
150
100
50
0
1
2
3
4
5

7.ADVANTAGES AND DISADVANTAGES:

Advantages:

●There are many employee stay with the organization long,which might mean that they are getting top of their pay scale.

●There are employees who just have been working on a slow place for years within an organization.

●If we can accurately predict which employee will leave their current company or organization,the it will save much time, effort,and cost of the employer and the help them to hire or acquire substituties in advance,and it would not create a problem in the ongoing progress of an organization.

●Many employees when some people leave an organization they open gates for new talent and new ideas.

● Provides good understanding of workforce supply and demand.

Disadvantages:

●Decreased overall performance.

●Lack of knowledgeable employees.

●Increased cost.

●Daily task management.

● Gives only less accuracy for the employees attrition status.

8.APPLICATIONS:

●Predictive attrition model helps not only taking preventive measures but also into making better hiring decisions.

●Derving trends in the candidates performance out of past data is important in order to predict the future trenda,as well as to board new employees.

● So we use Machine Learning Algorithms to analyze the data and propose what employees need to achieve their needs.

9.CONCLUSION:

In employee attrition prediction problem,an estimation can be framed for either the employee will leave the company or not.In this paper, the regression algorithm is adopted to build a UI model for predicting employee attrition perdiction default in the lending club and the results are compared with other six algorithms of logistic regression, KNN, random forest, decision tree and support vector machine. The experiment shows that the regression algorithm performs outstanding than the other six algorithms in the prediction of employee attrition prediction default and has strong ability of generalization. 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. HR to take necessaey action for the relatior of employees predicting to be at risk of leaving.

10. FUTURE SCOPE:

In futurethe Regression algorithm can be applied on other data sets available for employee attrition prediction to furtherinvestigate its accuracy. A rigorous analysis of other machine learning algorithms other than these six can also be done in future to investigate the power of machine algorithms other than these six can also be done in future to investigate the power of machine experiments on larger data sets or try to tune the model so as to achieve the state -of-art performance of the model and a great UI support system making it complete web application model.

APPENDIX:

HTML:

```
<html>
<body bgcolor ="green">
<form action ="/login" method ="post">
```

```
<p>Education</p>
<p><input type ="number" name ="Education"/></p>
<p>JobInvolvement</p>
<p><input type ="number" name ="JobInvolvement"/></p>
<p>JobLevel</p>
<p><input type ="number" name ="JobLevel"/></p>
<p>DailyRate</p>
<p><input type ="number" name ="DailyRate"/></p>
<p>MonthlyIncome</p>
<p><input type ="number" name ="MonthlyIncome"/></p>
<p>NoofCompaniesWorked</p>
<p><input type ="number" name ="NoofCompaniesWorked"/></p>
<p>TotalWorkingYears</p>
<p><input type ="number" name ="TotalWorkingYears"/></p>
<p>YearsAtCompany</p>
<p><input type ="number" name ="YearsAtCompany"/></p>
<p>YearsInCurrentRole</p>
<p><input type ="number" name ="YearsInCurrentRole"/></p>
<p>YearsSinceLastPromotion</p>
<p><input type ="number" name ="YearsSinceLastPromotion"/></p>
<p>YearsWithCurrentManager</p>
<p><input type ="number" name ="YearsWithCurrentManager"/></p>
<p>TrainingTimesLastYear</p>
<p><input type ="number" name ="TrainingTimesLastYear"/></p>
<p>PerformanceRating</p>
<p><input type ="text" name ="PerformanceRating"/></p>
<p><input type ="submit" value ="click"/></p>
<b>{{y}}</b>
</form>
</body>
</html>
```

APP.PY:

```
from flask import Flask,render_template,request
import pickle
import pickle,joblib
model=pickle.load(open('Attrition.pkl','rb'))
trans=joblib.load('transform.save')
app=Flask(__name__)
@app.route('/')
def hello_world():
    return render_template("index.html")
@app.route('/login',methods=["POST"])
def func2():
    Education=request.form['Education']
    JobInvolvement=request.form['JobInvolvement']
    JobLevel=request.form['JobLevel']
    DailyRate=request.form['DailyRate']
    MonthlyIncome=request.form['MonthlyIncome']
    NoofCompaniesWorked=request.form['NoofCompaniesWorked']
    TotalWorkingYears=request.form['TotalWorkingYears']
    YearsAtCompany=request.form['YearsAtCompany']
    YearsInCurrentRole=request.form['YearsInCurrentRole']
    YearsSinceLastPromotion=request.form['YearsSinceLastPromotion']
    YearsWithCurrentManager=request.form['YearsWithCurrentManager']
    TrainingTimesLastYear=request.form['TrainingTimesLastYear']
    PerformanceRating=request.form['PerformanceRating']

data=[[int(Education),int(JobInvolvement),int(JobLevel),int(DailyRate),int(MonthlyIncome),int(
NoofCompaniesWorked),int(TotalWorkingYears),int(YearsAtCompany),int(YearsInCurrentRole
),int(YearsSinceLastPromotion),int(YearsWithCurrentManager),int(TrainingTimesLastYear),int(
PerformanceRating)]]
```

```
    output= trans.transform(data)
    pred=model.predict(output)
    print(pred[0])
    if(pred==0):
        ans='Employee will stay'
    else:
        ans ='Employee will be terminated'
    return render_template("index.html",y=ans)
if __name__=='__main__':
   app.run(debug= True)
```