**TEAM NAME : CSE-024**

**STUDTENT MEMBER’S:**

SRI HARSHITHA YANNAM   
CHANDANA SANGA

SOWMYA KANDAGTLA

SAI REVANTH METTELA

FAIYAZUDDIN MOHAMMED

**[PREDICTING EMPLOYEE ATTRITION USING RANDOM FOREST]**

1. **INTRODUCTION** 
   1. Overview

Building a webpage and a machine learning model with flask integration, which demonstrates how to predict the employee attrition using random forest.

* 1. Purpose

To predict the employee attrition for company’s future enhancement.

* + 1. **LITERATURE SURVEY:**
  1. Existing Problem

Gaussian Naive Bayes –

Training Accuracy: 0.782

Testing Accuracy: 0.825

Naive Bayes classifier for multivariate Bernoulli models –

Training Accuracy: 0.831

Testing Accuracy: 0.845

Logistic Regression classifier –

Training Accuracy: 0.865

Testing Accuracy: 0.875

K-nearest neighbor’s (K-NN) -

Training Accuracy: 0.842

Testing Accuracy: 0.852

Decision tree classifier -

Training Accuracy: 0.792

Testing Accuracy: 0.823

Support Vector Machines (SVM) classification -

Training Accuracy: 0.851

Testing Accuracy: 0.859

Linear Support Vector Machines (LSVM) classification -

Training Accuracy: 0.858

Testing Accuracy: 0.879

* 1. Proposed Solution

Random Forest Classifier –

Training Accuracy: 0.963

Testing Accuracy: 0.8935

Here we found that Random Forest can give the best accuracy compared to other machine learning models.

* + 1. **THEORETICAL ANALYSIS**
  1. Block Diagram

Import Libraries

Remove null values

Import Dataset

Plotting the dataset for predicting the best range for the column values

Dividing the dataset into categorical and numerical

Remove Unnecessary columns

Train and test the dataset Using the Random Forest

Using One Hot Encoding for converting Categorical data into numerical data

Split the dataset into training and testing

Save the model using pickle module

Predict the values

Application Building with Flask and HTML

1. Hardware/ Software Designing

Hardware – PC/ Laptop

Software – Anaconda Navigator, Spyder, Anaconda Prompt, Jupyter Notebook/Google Collab, Chrome/Microsoft Edge, Some modules/ libraries such as Flask, TensorFlow, keras, NumPy, pandas, matplotlib, seaborn, pickle.

1. **EXPERIMENTAL INVESTIGATIONS**

We have visited several websites regarding the project and confirmed with various models and checked the accuracy and confirmed the best model (Random Forest Classifier). We even collected the dataset.

1. **FLOWCHART**

Start

Upload Information of an employee

Test

YES

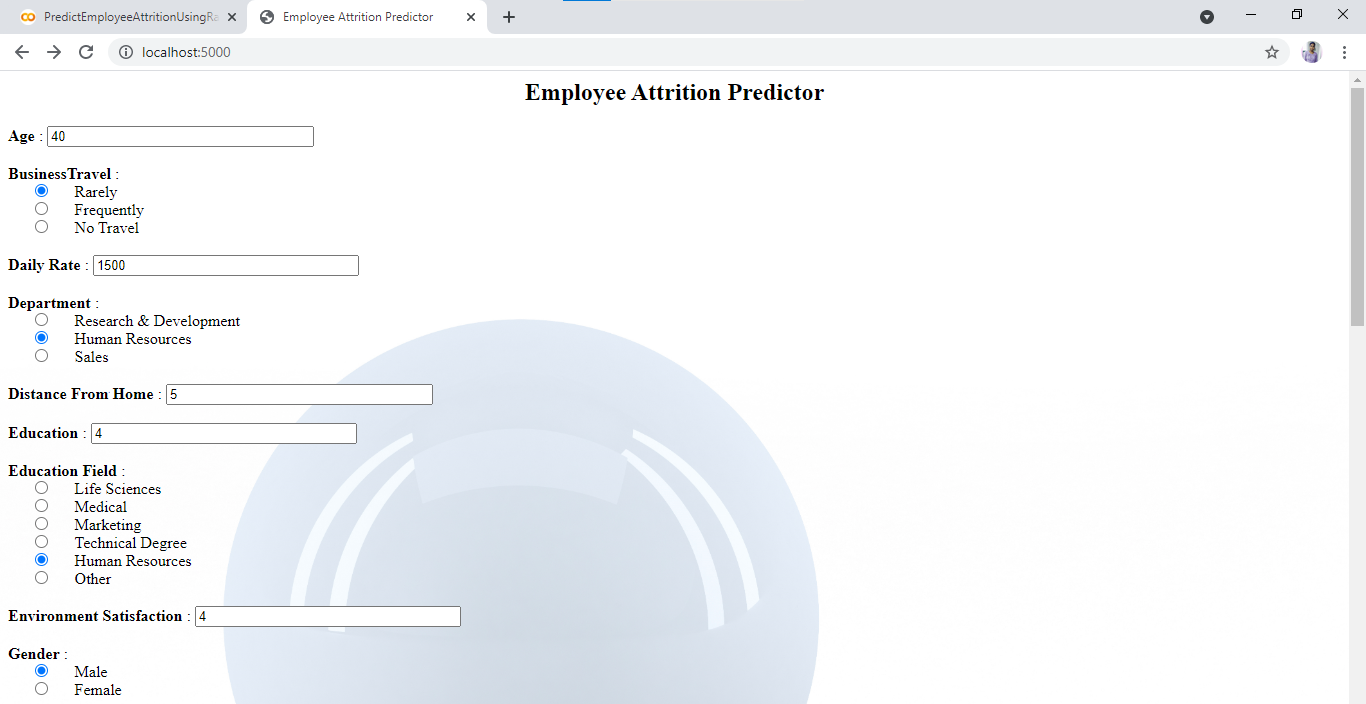
No

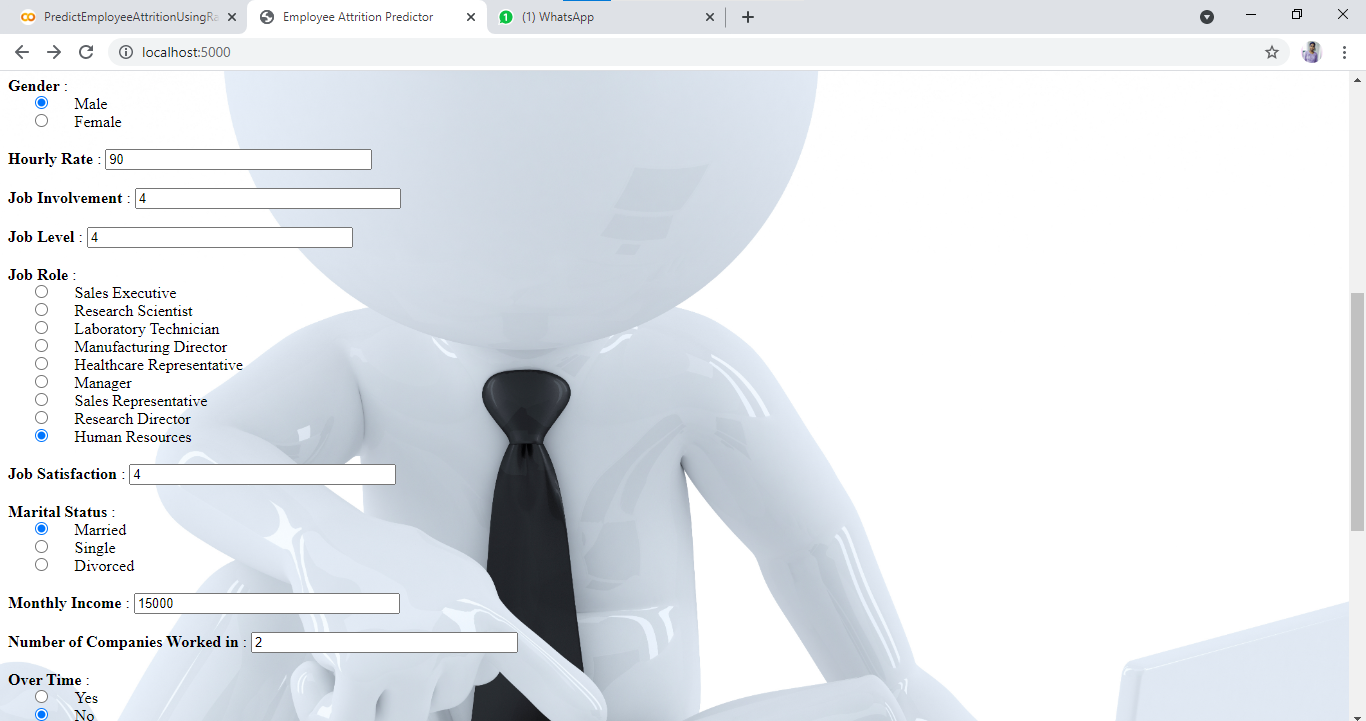
Employee might leave

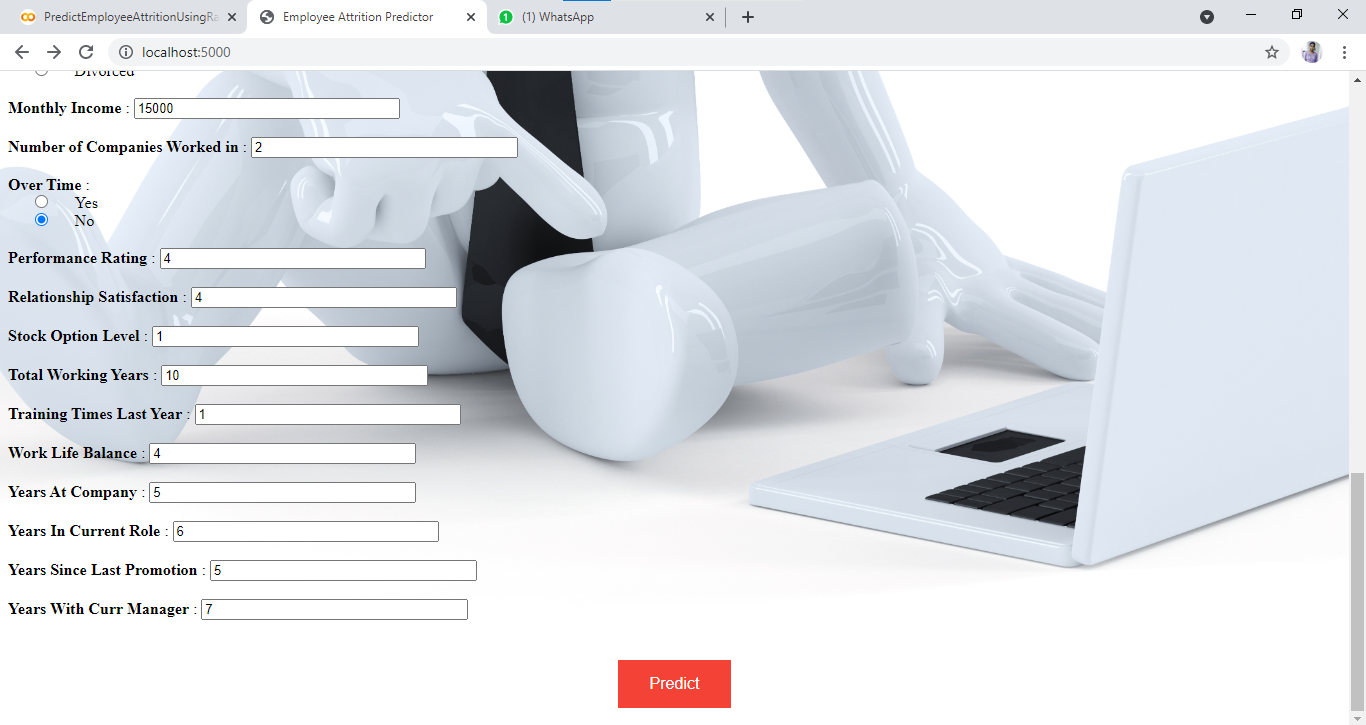
Chance of staying in the company

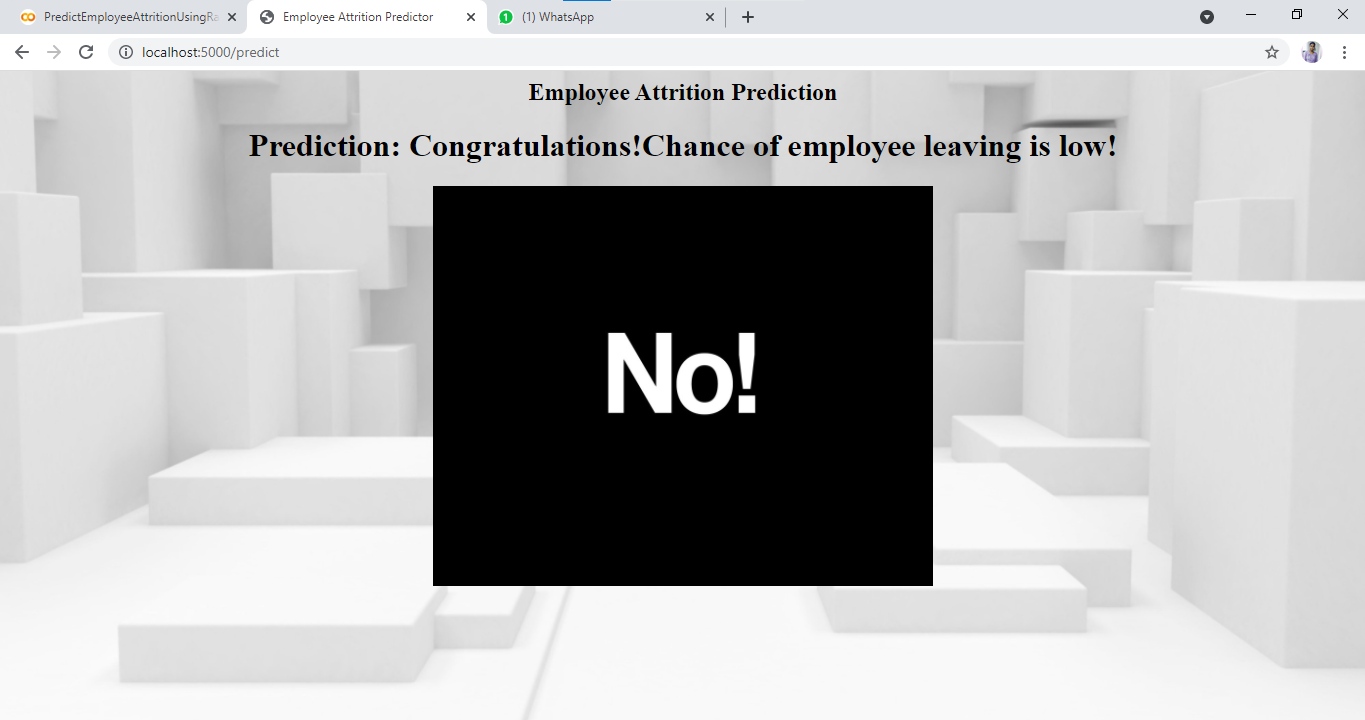
1. **RESULT**

Employee Attrition NO: Chance of staying in the company

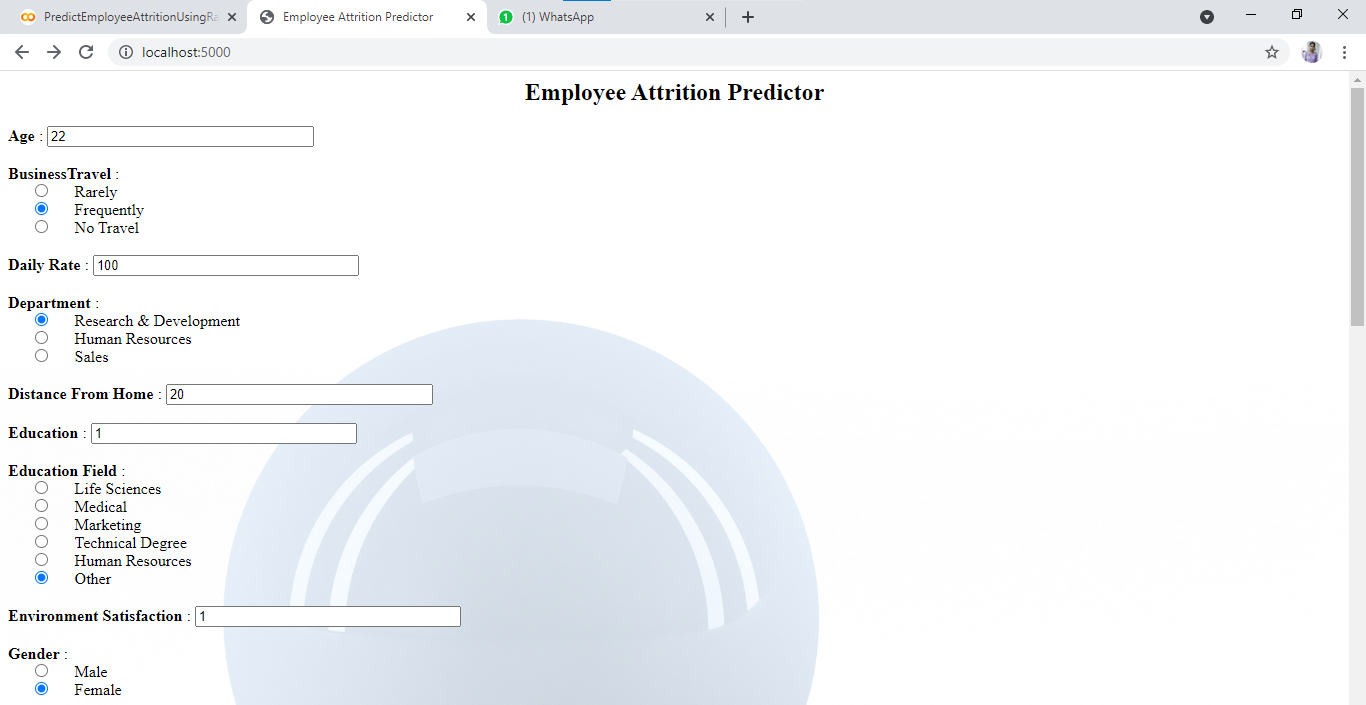


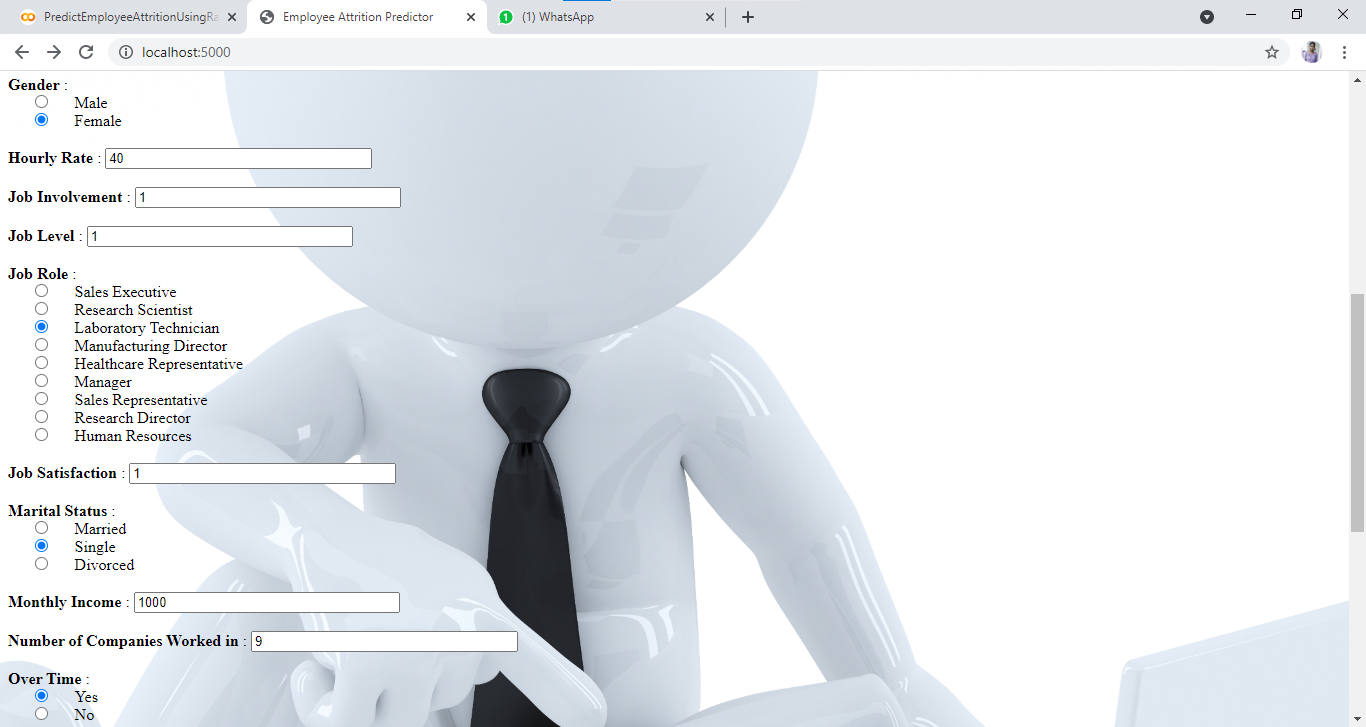


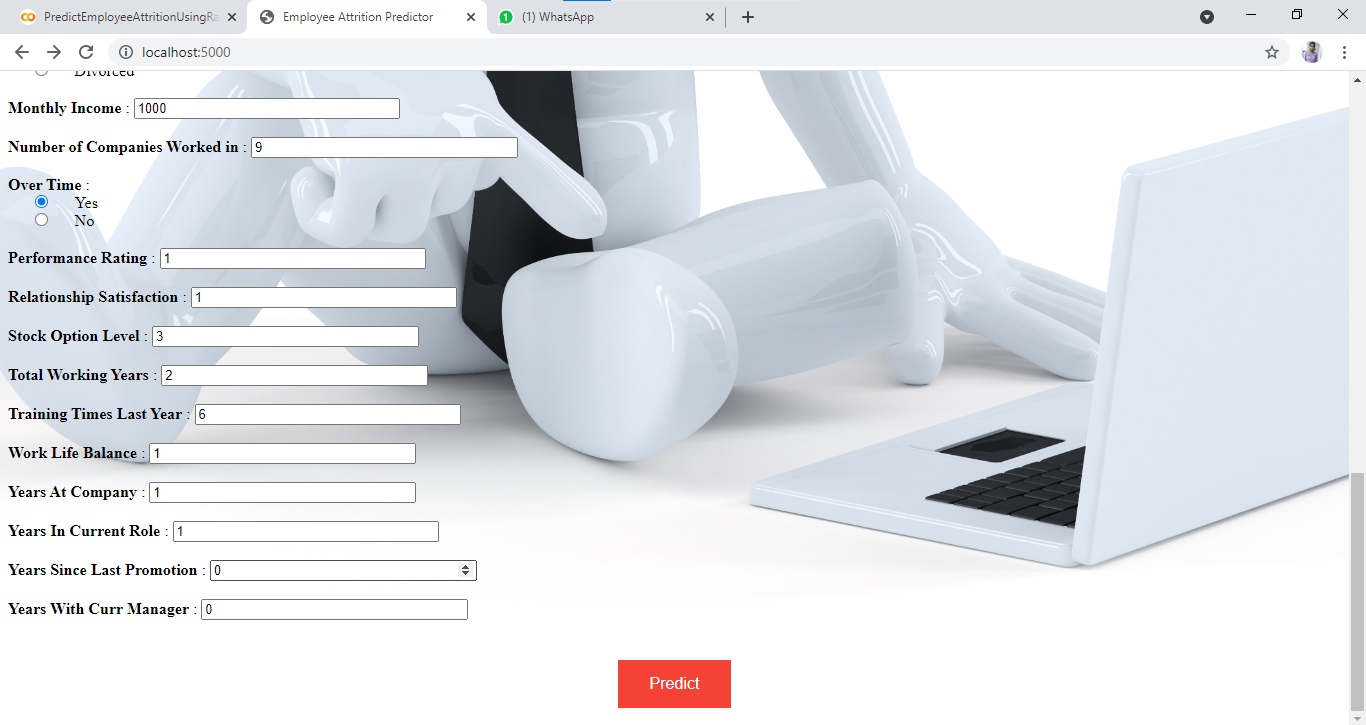


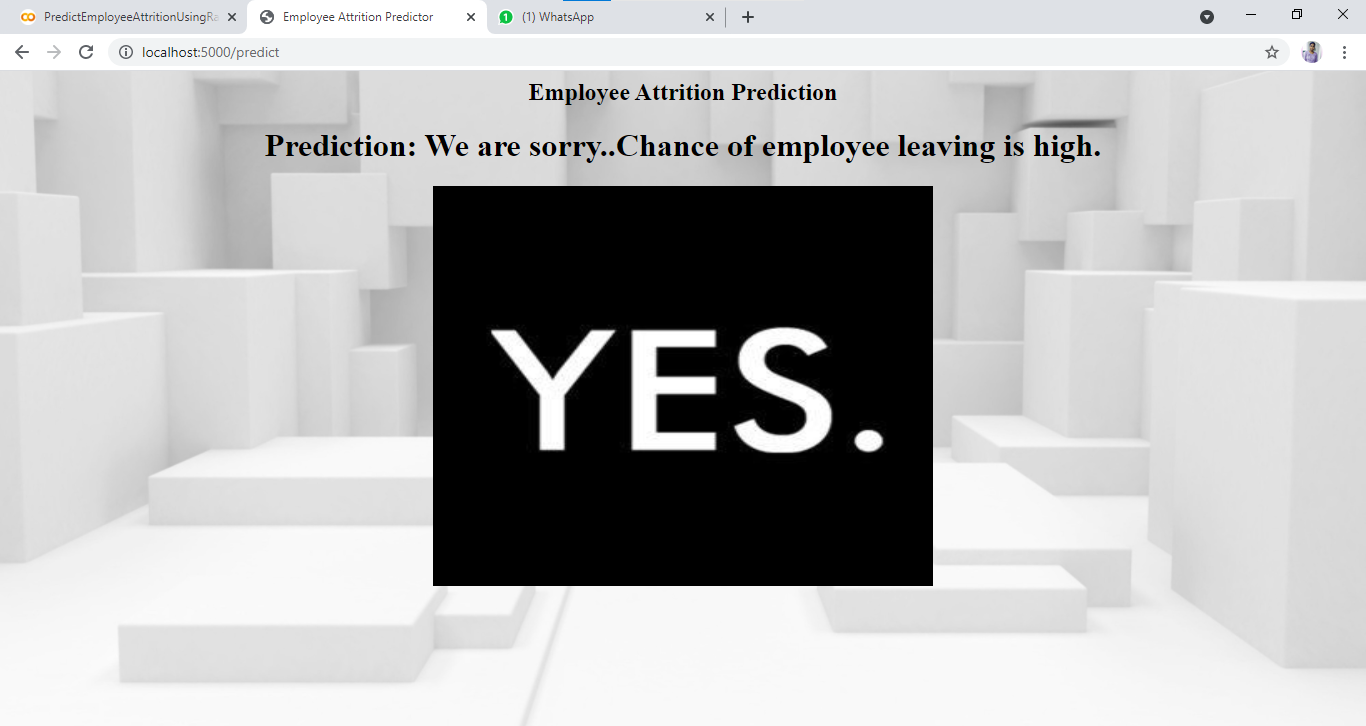


Employee Attrition YES: Chance of Leaving the company









1. **ADVANTAGES & DISADVANTAGES**

Advantages:

1. It reduces overfitting in decision trees and helps to improve the accuracy
2. It is flexible to both classification and regression problems
3. It works well with both categorical and continuous values
4. It automates missing values present in the data
5. Normalizing of data is not required as it uses a rule-based approach.

Disadvantages:

1. It requires much computational power as well as resources as it builds numerous trees to combine their outputs.
2. It also requires much time for training as it combines a lot of decision trees to determine the class.
3. Due to the ensemble of decision trees, it also suffers interpretability and fails to determine the significance of each variable.

1. **APPLICATIONS**

**Banking Sector**

Banking analysis requires a lot of effort as it contains a high risk of profit and loss. Customer analysis is one of the most used studies adopted in banking sectors. Problems such as loan default chance of a customer or for detecting any fraud transaction, random forest can be a great choice.

**Healthcare Sectors**

In pharmaceutical industries, random forest can be used to identify the potential of a certain medicine or the composition of chemicals required for medicines. It can also be used in hospitals to identify the diseases suffered by a patient, risk of cancer in a patient, and many other diseases where early analysis and research play a crucial role.

1. **CONCLUSION**

We conclude that random forest classifier is the highest accuracy model, which predicts the data at highest accuracy compared to other models, so by which we can predict the employee attrition that whether the employee will leave the company or stay.

1. **FUTURE SCOPE**

In future we can take also predict that

1. Employee Salary can be hiked or not.
2. Employee can get promotions or not.
3. **BIBILIOGRAPHY**

[**https://www.knime.com/blog/predicting-employee-attrition-with-machine-learning**](https://www.knime.com/blog/predicting-employee-attrition-with-machine-learning)

[**https://medium.com/analytics-vidhya/predict-employee-attrition-a34e2c5a972d**](https://medium.com/analytics-vidhya/predict-employee-attrition-a34e2c5a972d)

[**https://data.world/aaizemberg/hr-employee-attrition**](https://data.world/aaizemberg/hr-employee-attrition)

**APPENDIX**

* + 1. Source Code:

[**https://github.com/smartinternz02/SI-GuidedProject-4354-1626245333**](https://github.com/smartinternz02/SI-GuidedProject-4354-1626245333)