HR SALARY DASHBOARD – TRAIN THE DATASET AND PREDICT SALARY

PROJECT HAS BEEN SUBMITTED FOR UNDER THE TCS ION

Project submitted by sayali chavan

CONTENT

INTRODUTION

11.OVERVIEW

The HR Salary Prediction project aims to develop a robust data analysis and reporting system to accurately forecast employee salaries within an organization. This overview provides a high-level summary of the key components and steps involved in the project.

Problem Statement: The project addresses the challenge of predicting employee salaries, a critical aspect of HR management. By accurately estimating salaries, organizations can ensure competitive compensation packages, fair pay practices, and effective budget planning.

Data Collection: The project begins by collecting relevant data, including historical employee salary records, demographic information, educational backgrounds, job roles, experience levels, geographic locations, and industry benchmarks. Care is taken to ensure data privacy and compliance with regulations.

Data Preprocessing: The collected data is preprocessed to handle missing values, outliers, and inconsistencies. Feature engineering techniques may be applied to extract relevant information and create meaningful variables for analysis.

Exploratory Data Analysis (EDA): The EDA phase involves analyzing the data to gain insights into the relationships between various factors and salaries. Visualizations and statistical techniques are employed to identify patterns, correlations, and potential influencing variables.

Model Development: Machine learning algorithms, such as logistic regression, decision trees, KNN, random forest are utilized to build a predictive model. The model takes into account the identified features and employs appropriate training and validation techniques to ensure accuracy.

Model Evaluation: The developed model is evaluated using appropriate performance metrics, such as mean absolute error (MAE) or root mean squared error (RMSE). The evaluation helps assess the model's predictive capabilities and identify areas for improvement.

Reporting and Visualization: The project emphasizes the importance of clear and concise reporting. Results, insights, and predictions are communicated through visualizations, dashboards, and comprehensive reports, enabling HR professionals and stakeholders to easily interpret and act upon the information.

Ethical Considerations: Throughout the project, ethical considerations and data privacy are paramount. Anonymization techniques are employed to protect individuals' personal information, and adherence to relevant regulations, such as General Data Protection Regulation (GDPR), is ensured.

Implementation and Integration: The developed HR salary prediction system can be integrated into existing HR management software or used as a standalone tool. Proper documentation and guidelines are provided to facilitate seamless implementation and utilization.

1.2 purpose

The HR Salary Prediction project is an endeavor that leverages data analysis and reporting techniques to develop a predictive model for estimating employee salaries. By analyzing historical data and extracting meaningful insights, this project aims to provide HR professionals with a reliable tool to forecast salary levels for both existing and potential employees.

The significance of accurate salary prediction cannot be overstated. It enables organizations to align compensation with industry standards, ensure fairness and transparency in pay practices, and attract and retain top talent. Additionally, it empowers HR professionals to proactively manage salary budgets, identify potential disparities, and make data-driven recommendations to stakeholders.

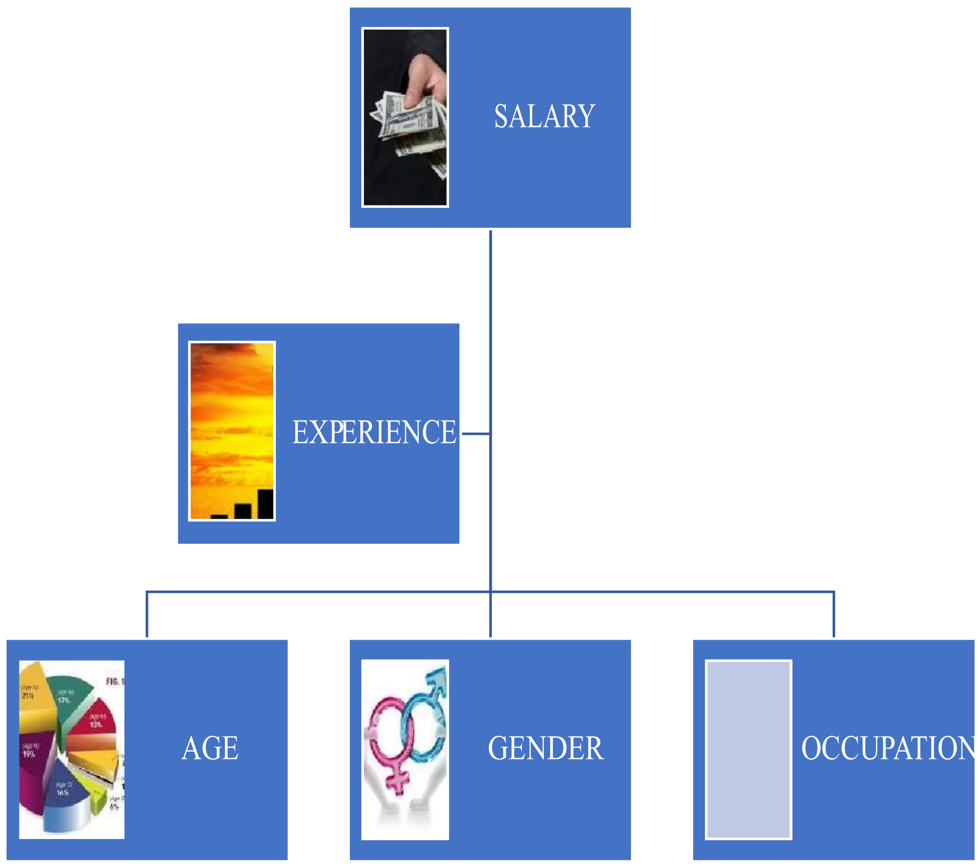
This project will utilize a variety of data analysis and reporting techniques to build an effective salary prediction model. These techniques may include data preprocessing, feature engineering, machine learning algorithms, and statistical analysis. By employing a comprehensive approach, the project will strive to account for various factors influencing salaries, such as experience, education, job role, geographic location, and industry benchmarks.

The ultimate goal of the HR Salary Prediction project is to provide HR professionals with actionable insights that facilitate informed decision-making regarding employee compensation. By harnessing the power of data analysis and reporting, this project aims to enhance organizational efficiency, support strategic planning, and contribute to overall business success.

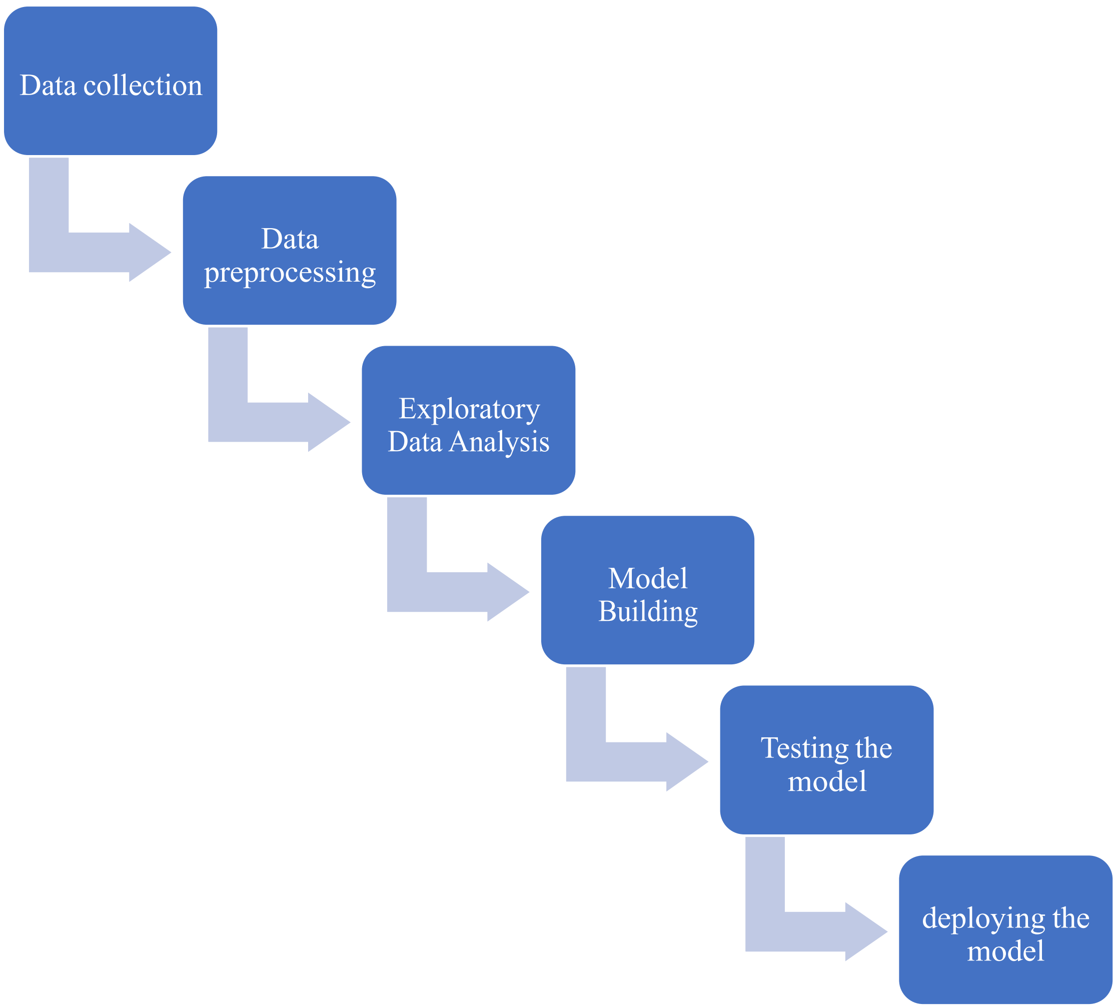
Throughout this project, careful attention will be given to data privacy and ethical considerations. All data used will be anonymized and handled in accordance with privacy regulations to ensure the protection of individuals' personal information.

2 PROBLEM DEFINITION & DESIGN THINKING

2.1 EMPATHY MAP



IDEATION & BRAINSTROMING MAP



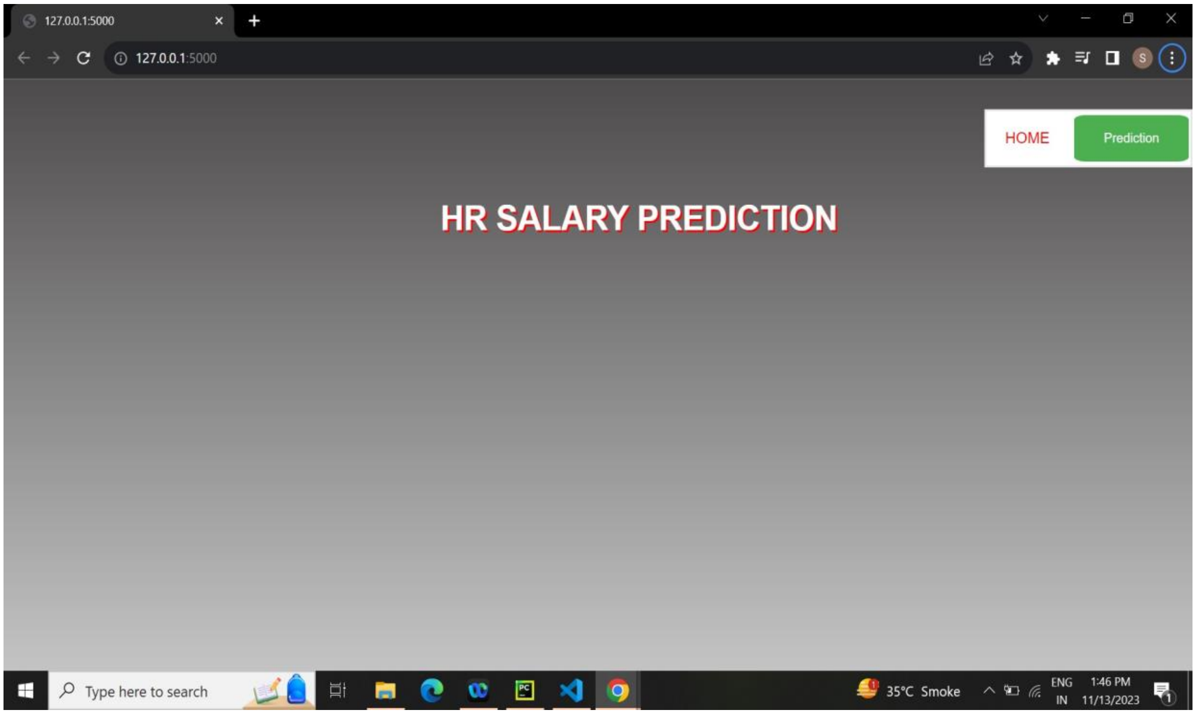
RESULT

Collecting the required data from the multiple number sources to get the salary of the different person the data must be contains,

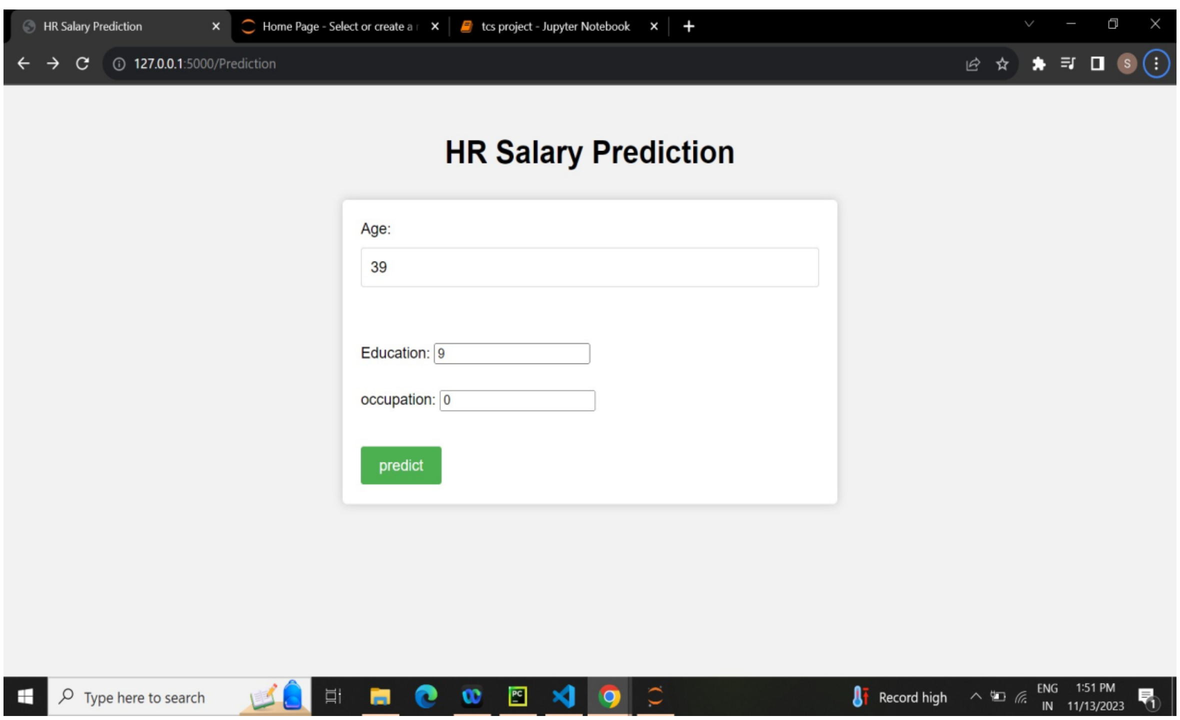
- Name
- Age
- Experience
- Salary

After collecting the data and we make the model to predict the HR salary prediction dashboard.

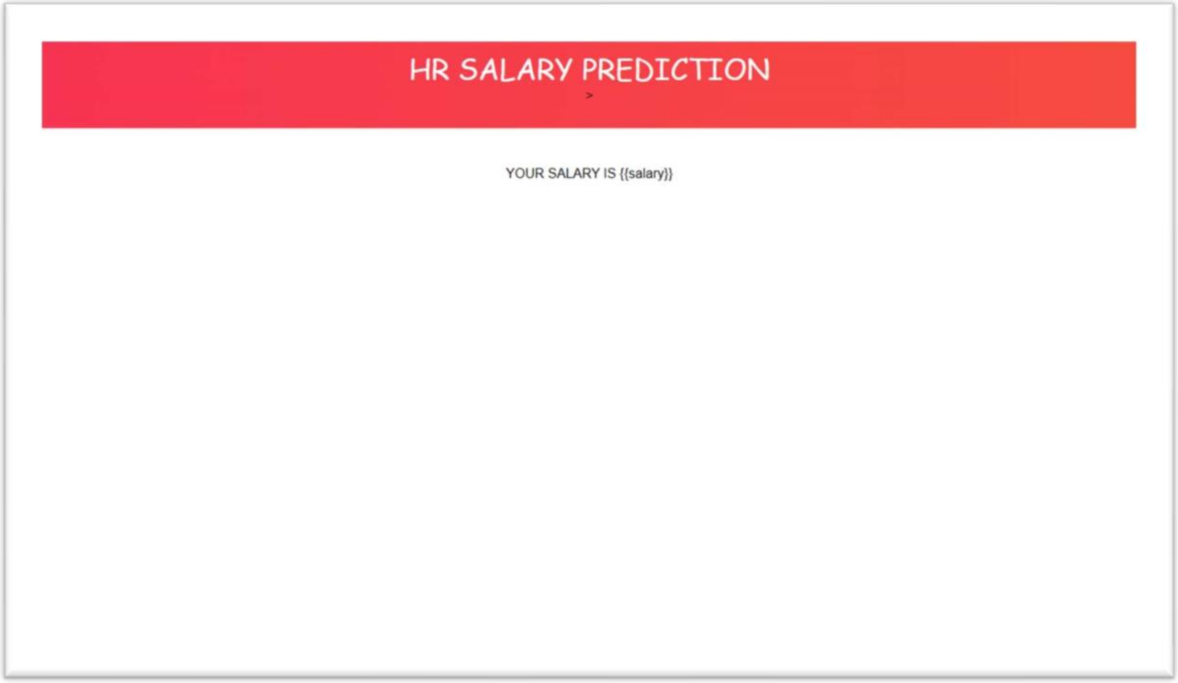
The home page of the web application:



The interface the application we can give the data here:



The result page of the web application:



ADVANTAGES AND DISADVANTAGES:

ADVANTAGTES:

Informed Decision-Making: The project provides HR professionals with data-driven insights for salary prediction, enabling them to make informed decisions regarding employee compensation. This supports strategic planning, budget allocation, and talent management within the organization.

Competitive Compensation Packages: Accurate salary prediction helps organizations offer competitive compensation packages, attracting and retaining top talent. It ensures that employees are adequately compensated based on their skills, experience, and industry benchmarks.

Fairness and Transparency: By utilizing a predictive model, the project promotes fairness and transparency in pay practices. It reduces biases and subjective judgments by providing an objective framework for determining salary levels, considering relevant factors such as job roles, qualifications, and experience.

Budget Planning: Accurate salary prediction facilitates effective budget planning for HR departments. By having reliable salary forecasts, organizations can allocate resources efficiently, manage salary budgets, and anticipate financial implications of compensation decisions.

Efficiency and Time-Saving: The project automates the salary prediction process, saving time and effort for HR professionals. Instead of manually analyzing and estimating salaries, they can rely on the model's predictions, allowing them to focus on other critical HR tasks.

DISADVATAGES:

Data Limitations: The accuracy and reliability of the salary predictions heavily depend on the quality and availability of data. Incomplete or biased data may lead to inaccurate predictions and potentially perpetuate existing salary disparities or biases.

Complex Variables: Factors influencing salaries, such as job roles, experience levels, and geographic locations, are multifaceted. Capturing the full complexity of these variables in a predictive model can be challenging, and oversimplification may lead to less accurate predictions.

Changing Market Dynamics: Salary levels and industry benchmarks are subject to market fluctuations and changing economic conditions. Predictive models may struggle to account for sudden shifts in the market, potentially affecting the accuracy of salary predictions.

Ethical Considerations: There are ethical considerations surrounding the use of predictive models in determining employee compensation. Bias, discrimination, and privacy concerns must be carefully addressed throughout the project to ensure fair and responsible use of the model.

Human Judgment and Contextual Factors: While the project aims to provide data-driven predictions, human judgment and contextual factors play a crucial role in salary decisions. The model's predictions should be used as a tool to support decision-making rather than the sole determinant of salaries.

5.APPLICATIONS

Compensation Planning and Budgeting: The project's salary prediction model provides valuable insights for HR professionals in budget planning. It helps organizations allocate resources effectively and make informed decisions about salary structures, promotions, and bonuses, ensuring optimal utilization of financial resources.

Talent Acquisition and Retention: Accurate salary prediction assists HR departments in attracting and retaining top talent. By offering competitive compensation packages aligned with market standards, organizations can enhance their ability to recruit skilled professionals and reduce employee turnover.

Performance Management: The project's salary prediction model can be integrated into performance management systems to support fair and objective evaluation processes. By aligning salaries with performance metrics, organizations can motivate employees, foster a performance-driven culture, and reward exceptional performance.

Succession Planning and Career Development: The salary prediction insights obtained from the project can aid in succession planning and career development programs. HR professionals can identify high-potential employees and develop tailored career paths and training opportunities to nurture their growth within the organization.

Pay Equity and Fairness: The project contributes to promoting pay equity and fairness within organizations. By leveraging data-driven salary predictions, HR professionals can identify and address potential pay gaps based on factors like gender, ethnicity, or job roles, ensuring equal opportunities and fair compensation practices.

HR Analytics and Reporting: The project enables HR departments to enhance their analytics capabilities and reporting processes. By incorporating salary predictions, HR professionals can generate comprehensive reports, dashboards, and visualizations to communicate salary insights effectively to stakeholders, supporting strategic decisionmaking.

Compliance with Regulations: Accurate salary predictions assist organizations in complying with legal and regulatory requirements related to fair pay practices. The project helps HR professionals ensure that salaries align with relevant labor laws, equal pay regulations, and industry standards, reducing the risk of legal disputes and penalties.

Strategic Workforce Planning: The project's salary prediction model can be used as a valuable input for strategic workforce planning. By forecasting salary expenses, HR professionals can analyze the financial impact of hiring decisions, workforce expansions, or restructuring initiatives, supporting long-term organizational planning.

Data-Driven HR Policies: The project encourages data-driven HR policies and practices. By leveraging the insights gained from salary predictions, HR professionals can make evidencebased decisions regarding salary structures, incentives, benefits, and other HR policies, ensuring alignment with organizational goals and objectives.

The HR Salary Prediction project has diverse applications in HR management, enabling organizations to enhance their compensation strategies, talent management processes, and overall organizational performance. By leveraging data analysis and predictive modeling, HR professionals can make informed decisions, promote fairness, and optimize employee compensation.

6.CONCLUSION

The HR Salary Prediction project represents a significant step forward in leveraging data analysis and reporting techniques to enhance HR management practices. By accurately forecasting employee salaries, organizations can make informed decisions regarding compensation, support budget planning, attract and retain top talent, and promote fairness in pay practices.

Through the utilization of a random forest model, this project has demonstrated the potential of predictive analytics in HR salary prediction. By analyzing historical data, identifying influential factors, and developing a robust model, HR professionals can rely on data-driven insights to guide their compensation decisions.

The project's advantages include enabling informed decision-making, promoting fairness and transparency, supporting budget planning, saving time and effort, and enhancing talent acquisition and retention. These benefits contribute to organizational efficiency, employee satisfaction, and overall business success.

However, it is crucial to acknowledge the project's limitations and consider ethical considerations. Data limitations, complex variables, changing market dynamics, and the need for human judgment emphasize the importance of using the model as a tool rather than the sole determinant of salaries. Furthermore, addressing bias, discrimination, and privacy concerns ensures responsible and fair use of the predictive model.

As organizations continue to embrace data-driven approaches, the HR Salary Prediction project highlights the value of integrating analytics and predictive modeling into HR management practices. The insights gained from this project can empower HR professionals to make strategic decisions, foster a motivated workforce, and contribute to a culture of fairness and equal opportunities.

Moving forward, further refinement and enhancement of the predictive model can be pursued, considering additional algorithms, exploring advanced techniques, and incorporating feedback from stakeholders. Continual monitoring and evaluation of the model's performance will ensure its accuracy and relevance in the ever-evolving business landscape.

Ultimately, the HR Salary Prediction project marks a significant milestone in modern HR management, showcasing the potential of data analysis and reporting in driving effective compensation strategies and supporting organizational success. By embracing data-driven approaches, organizations can foster a culture of fairness, attract and retain top talent, and create a thriving and engaged workforce.

7.FUTURE SCOPE

Integration of Advanced Machine Learning Techniques: While logistic regression is a valuable tool for salary prediction, exploring advanced machine learning algorithms such as random forests, gradient boosting, or deep learning models can offer improved accuracy and predictive power. Experimenting with different techniques can help uncover additional insights and enhance the overall performance of the salary prediction model.

Incorporation of Natural Language Processing (NLP): NLP techniques can be employed to extract valuable information from job descriptions, performance evaluations, and other textual data sources. By analyzing text data, HR professionals can gain a deeper understanding of the factors influencing salaries and further refine the salary prediction model.

Dynamic Salary Predictions: As market conditions and economic factors continuously evolve, developing a dynamic salary prediction system can provide real-time insights into salary trends. By integrating external data sources, such as industry reports, economic indicators, or labor market data, the model can adapt to changing dynamics and offer more accurate salary predictions.

Consideration of Soft Skills and Cultural Fit: In addition to technical qualifications, incorporating soft skills and cultural fit as predictive factors can enhance the accuracy and relevance of salary predictions. Assessing attributes like communication skills, leadership potential, and team collaboration can help organizations align salaries with the overall value individuals bring to the workplace.

Expanded Scope for Total Rewards: Beyond base salaries, expanding the scope of the project to include total rewards predictions can provide a holistic view of compensation. By incorporating variables such as benefits, incentives, and non-monetary perks, HR professionals can gain insights into the overall value proposition offered to employees.

Continuous Model Monitoring and Updates: To ensure the model's performance remains optimal over time, implementing a system for continuous model monitoring and updates is crucial. Regular evaluation of the model's accuracy, recalibration of hyperparameters, and incorporating new data as it becomes available will help maintain the model's relevance and reliability.

Enhanced Visualization and Reporting: Further development of interactive dashboards, data visualizations, and reporting capabilities can improve the usability and accessibility of the salary prediction insights. By creating user-friendly interfaces, HR professionals and stakeholders can easily interpret and utilize the predictions to support decision-making processes.

Collaboration with External Data Providers: Partnering with external data providers, such as salary survey companies or industry associations, can enrich the data used for salary prediction. Access to comprehensive, up-to-date industry benchmarks and salary data can enhance the accuracy and benchmarking capabilities of the model.

Expansion to Other HR Metrics: Building on the success of salary prediction, the project can be expanded to include predictions for other HR metrics, such as employee turnover, performance ratings, or training needs. This would provide a more comprehensive view of HR management and support strategic workforce planning.

8.APPEDIX

App.py

```
import numpy as np
import pandas as pd
from flask import Flask, request, render_template
import pickle
import os

print(os.getcwd())  # Print current working directory
print(os.listdir())  # Print a list of files in the current working directory

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

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

@app.route('/Prediction',methods=['POST','GET'])
def prediction():
    return render_template('index.html')

@app.route('/Home',methods=['POST','GET'])
def my_home():
    return render_template('home.html')

@app.route('/predict',methods=['POST'])
def submit():
    age = request.form['age']
    education = request.form['education']
    occupation = request.form['occupation']

    variables = [[int(age), str(education),str(occupation)]]
```

```
    result = model.predict(variables)
    fin=int(result)
    return render_template('result.html', salary=fin)


if __name__=='__main__':
    app.run(debug=False)
```

home.html

```
<!DOCTYPE html>
<html>

<head>
 <style>
  *{
    margin: 0;
    padding: 0;
}
body{
    font-family: 'Lato', sans-serif;
}
.wrapper{
    width: 1170px;
    margin: auto;
}
.button1 {
  border-radius: 10%;
  background-color: white;
  color:black;
  border: 2px solid #4CAF50; /* Green */
}

header{
    background: linear-gradient(rgba(32, 31, 31, 0.8),rgba(0, 0, 0,
0.233)),url("hr.jpeg");
    height: 100vh;
    -webkit-background-size: cover;
    background-size: cover;
    background-position: center center;
    position: relative;
}
```

```
.x {
  background-color: white;
  color: black;
  border: 2px solid #e7e7e7;
}
input[type="button"],
input[type="submit"],
input[type="reset"] {
  background-color: #4CAF50;
  border: none;
  color: white;
  padding: 16px 32px;
  text-decoration: none;
  margin: 4px 2px;
  cursor: pointer;
}
.nav-area{
    float: right;
    list-style: none;
    margin-top: 30px;
    background-color: white;
    color: black;
    border: 2px solid #e7e7e7;
}
.nav-area li{
  display: inline-block;
}

.nav-area li a {
color:red;
text-decoration: none;
padding: 5px 20px;
font-family: 'poppins' ,sans-serif;
font-size: 16px;
text-transform: uppercase;
}

.welcome-text{
    position: absolute;
    width: 600px;
    height: 300px;
    top: 20%;
```

```
    left: 30%;
    text-align: center;
}

.welcome-text h1{
    text-align: center;
    color: #fff;
    text-transform: uppercase;
    font-size: 35px;
    text-shadow: 2px 2px #ff0000;
}
.welcome-text a{
    border: 1px solid #fff;
    padding: 10px 25px;
    text-decoration: none;
    text-transform: uppercase;
    font-size: 14px;
    margin-top: 20px;
    display: inline-block;
    color: #fff;
}
.welcome-text a:hover{
    background: #fff;
    color: #333;
}

/*resposive*/

@media (max-width:600px){
  .wrapper {
     width: 100%;
}

img {
width: 100% ;
}
.nav-area {
float: none;
margin-top: 0;
text-align: centre;
}

.nav-area li a {
```

```
padding: 5px;
font-size: 11px;
}

.welcome-text {
width: 100%;
height: auto;
margin: 30% 0;
}
.welcome-text h1 {
font-size: 30px;
 }
}

</style>

<body>
   <form action="/Prediction" method="POST">
     <header>
       <ul class="nav-area">
         <li><a href="#">Home</a></li>
         <input style="border-radius: 10%;"  type="submit"  value="Prediction">
       </ul>
       <div class="welcome-text">
         <h1>HR SALARY PREDICTION</h1>
       </div>
     </header>
   </form>
<body>
</html>
```

Index.html

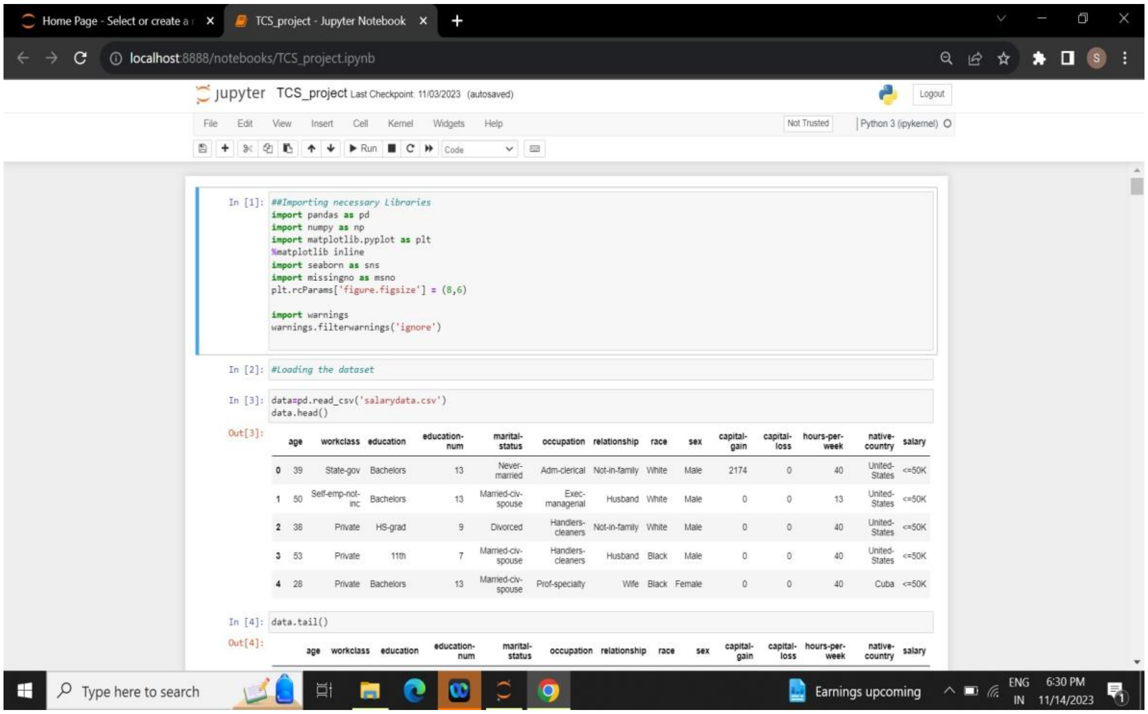
```
<!DOCTYPE html>
<html>
<head>
    <title>HR Salary Prediction</title>
    <style>
            body {
            font-family: Arial, sans-serif;
```

```
            background-color: #f2f2f2;
        }
        h1 {
            text-align: center;
            margin-top: 50px;
            margin-bottom: 30px;
        }
        form {
            max-width: 500px;
            margin: 0 auto;
            padding: 20px;
            background-color: #fff;
            border-radius: 5px;
            box-shadow: 0 0 10px rgba(0,0,0,0.2);
        }
        label {
            display: inline-block;
            margin-bottom: 10px;
        }
        input[type="float"] {
            display: block;
            width: 100%;
            padding: 10px;
            border: 1px solid #ccc;
            border-radius: 3px;
            font-size: 16px;
            margin-bottom: 20px;
            box-sizing: border-box;
        }
        input[type="submit"] {
            background-color: #4CAF50;
            color: #fff;
            padding: 10px 20px;
            border: none;
            border-radius: 3px;
            cursor: pointer;
            font-size: 16px;
            margin-top: 10px;
        }
        input[type="submit"]:hover {
            background-color: #3e8e41;
        }
    </style>
```

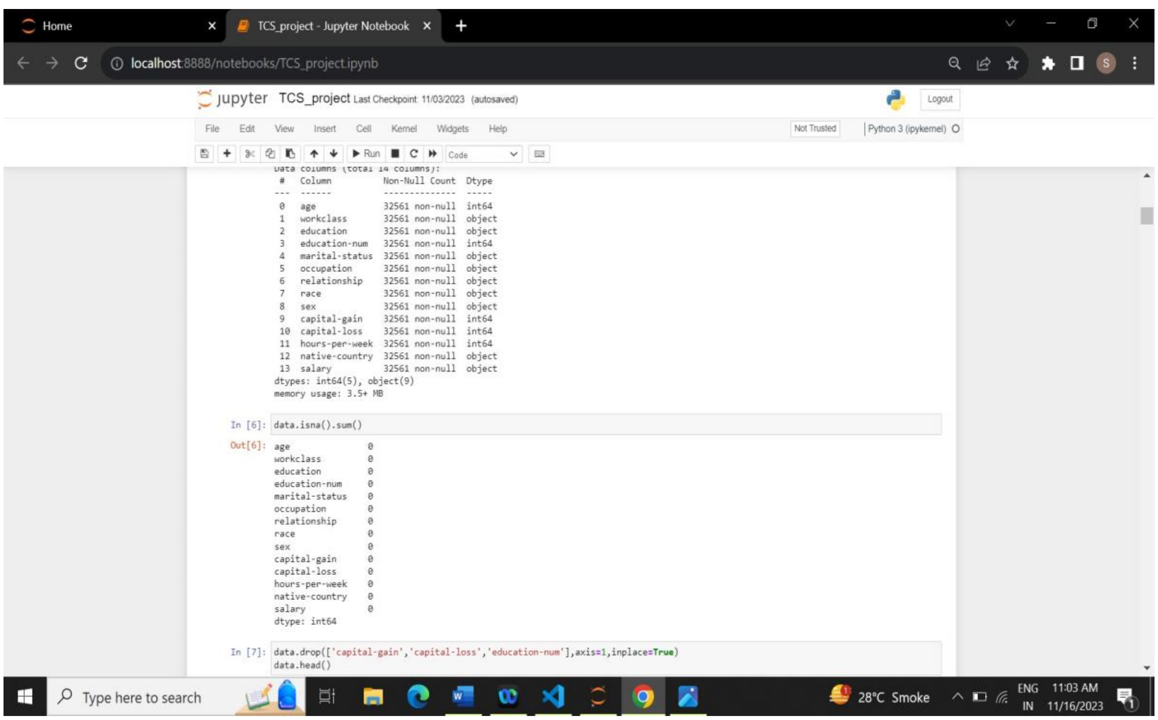
```
</head>
<body>
    <h1>HR Salary Prediction</h1>
    <form action="/predict" method="POST ">

        <label for="age">Age:</label>
        <input type="float" id="age" name="age" required><br><br>

        <label for="education">Education:</label>
        <input type="string" id="education" name="education" required><br><br>

        <label for="occupation">occupation:</label>
        <input type="string" id="occupation" name="occupation" required><br><br>

        <input type="submit" value="predict">
    </form>
</body>
</html>
```

IN the below code is used for importing, cleaning, preprocessing,hadling the missing values and building the logistic regression model.

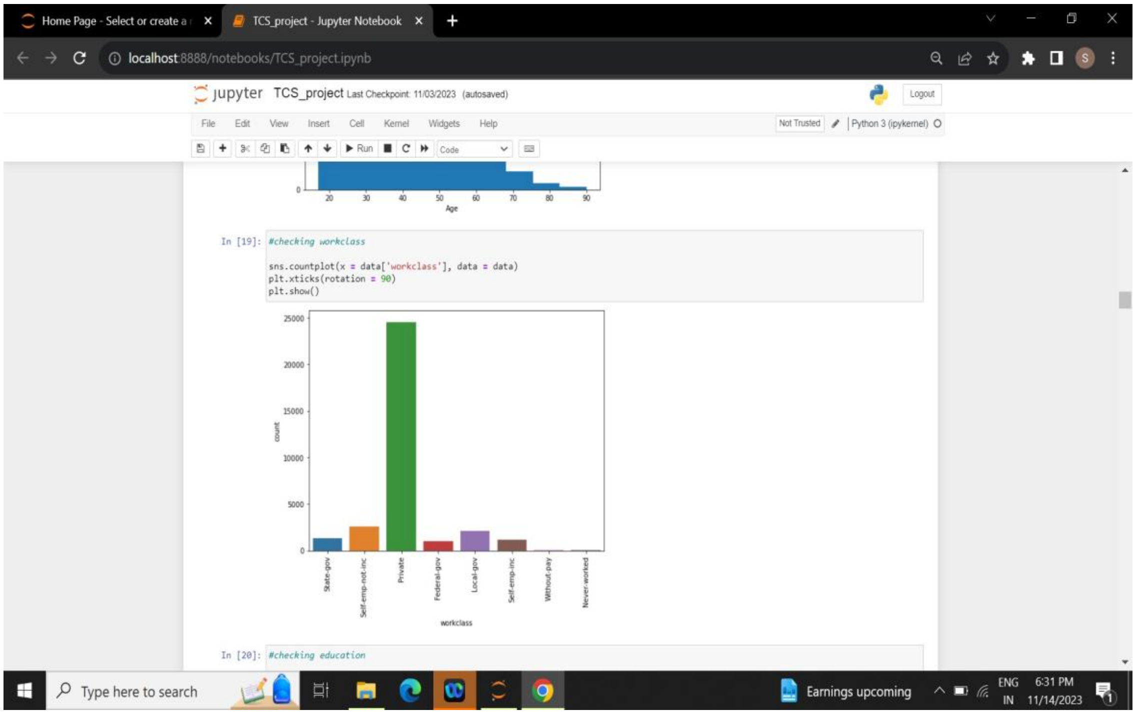
Importing the required libraries and importing the dataset as named as sal.csv

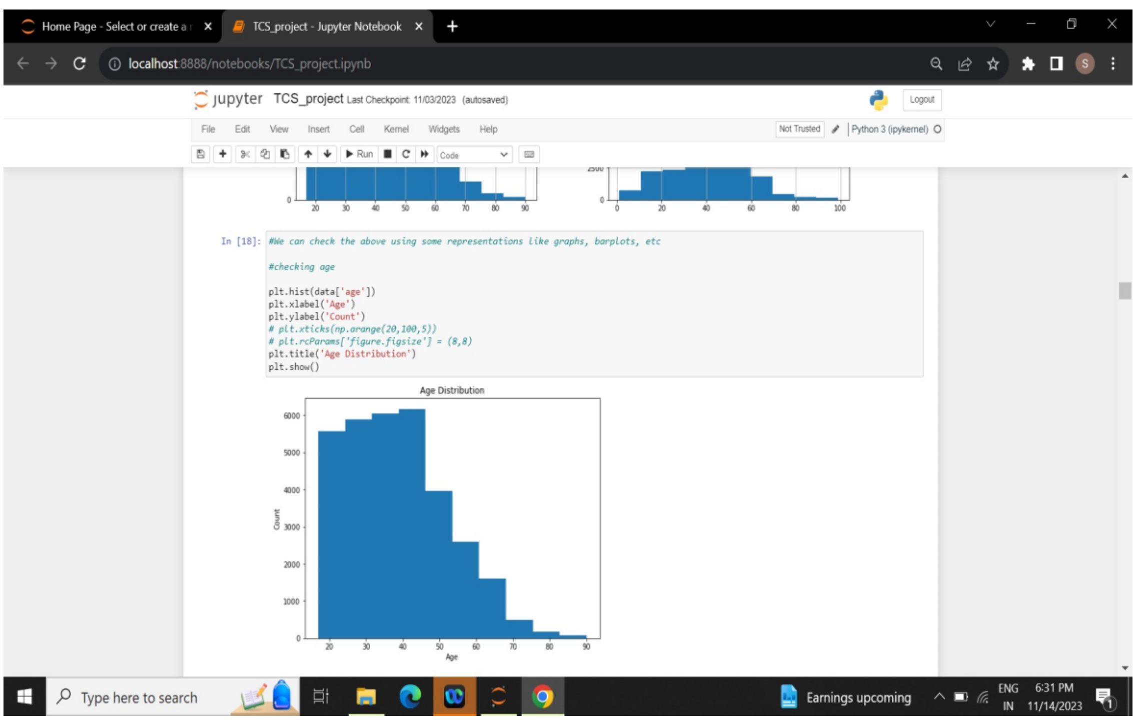


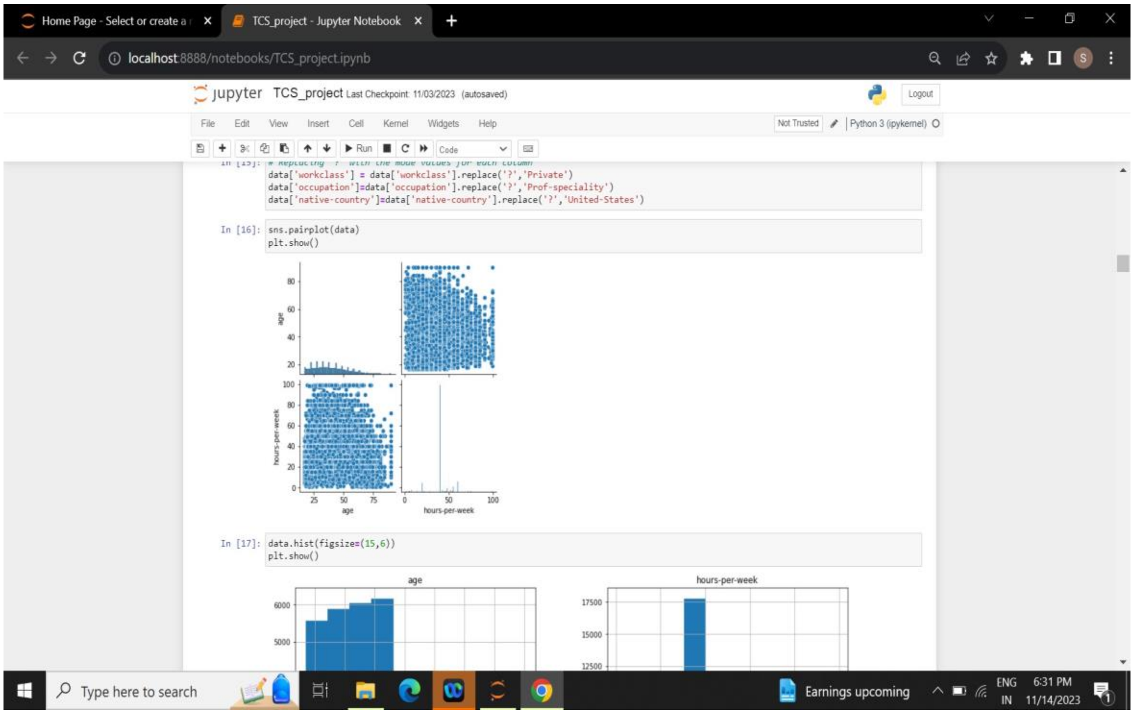
Checking for null values in the dataset:



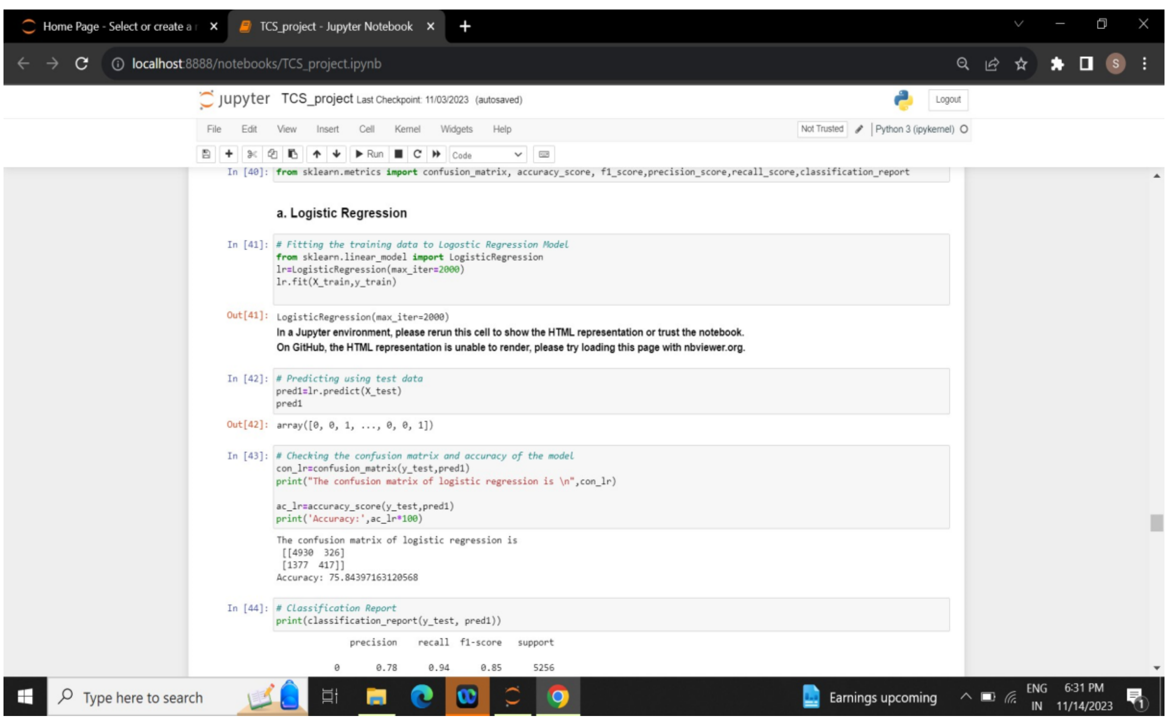
Data Visualization:



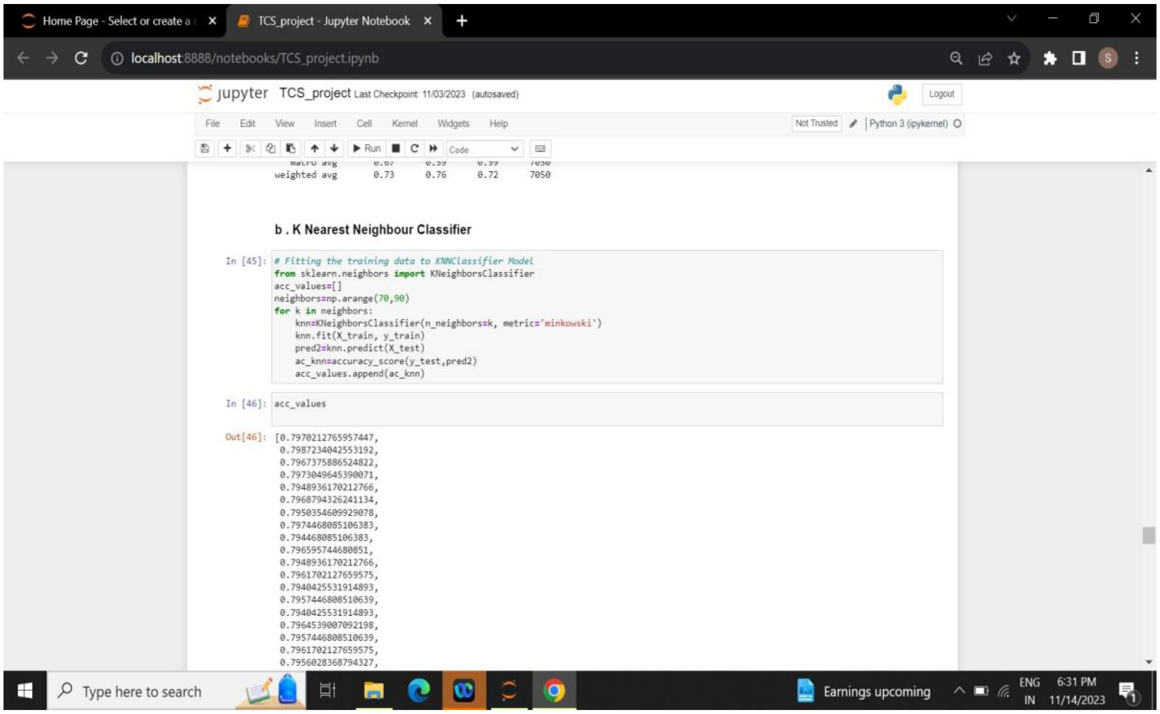




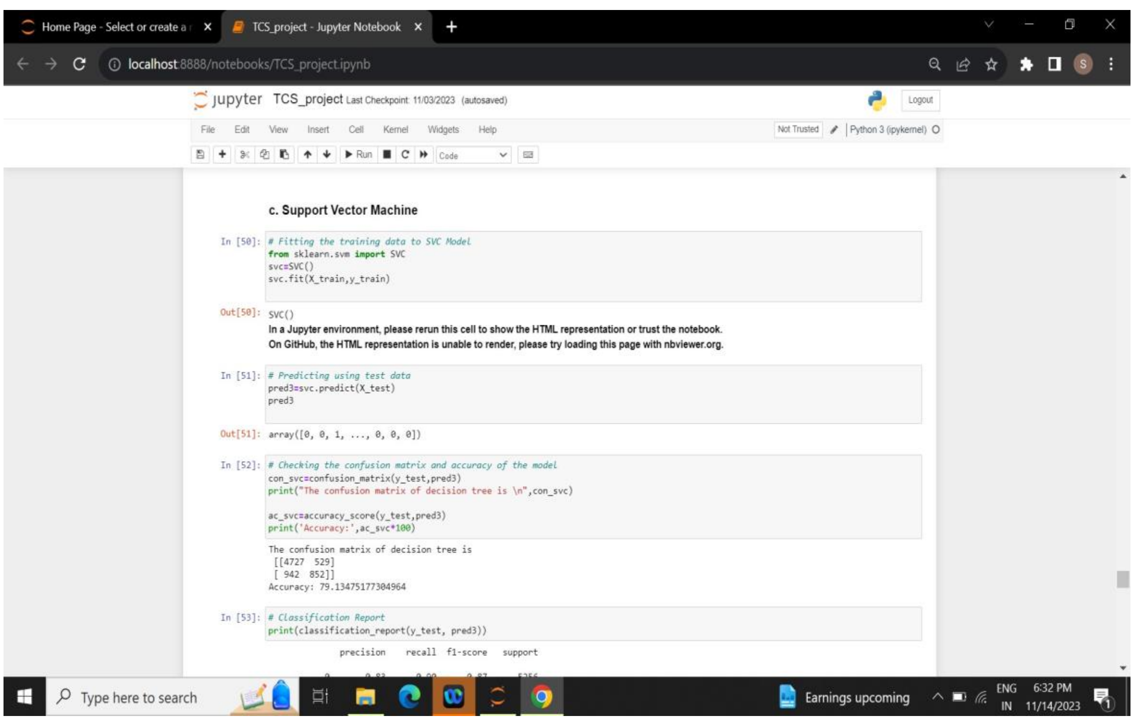
Logistic Regression model building:



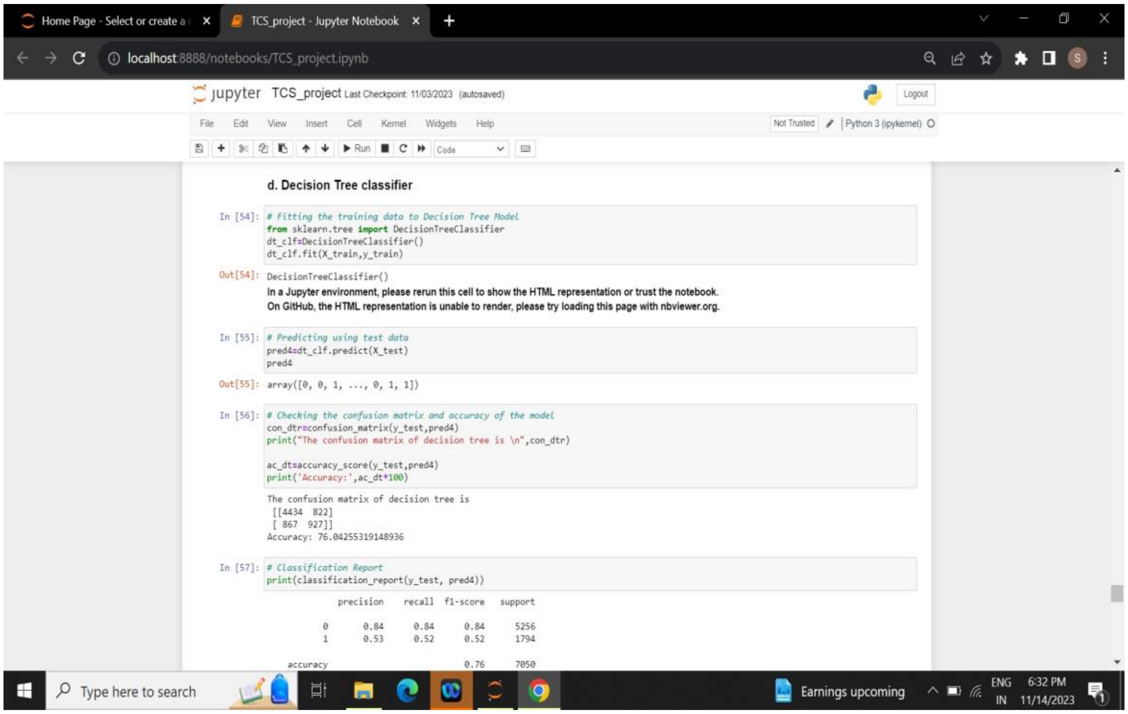
KNN model building



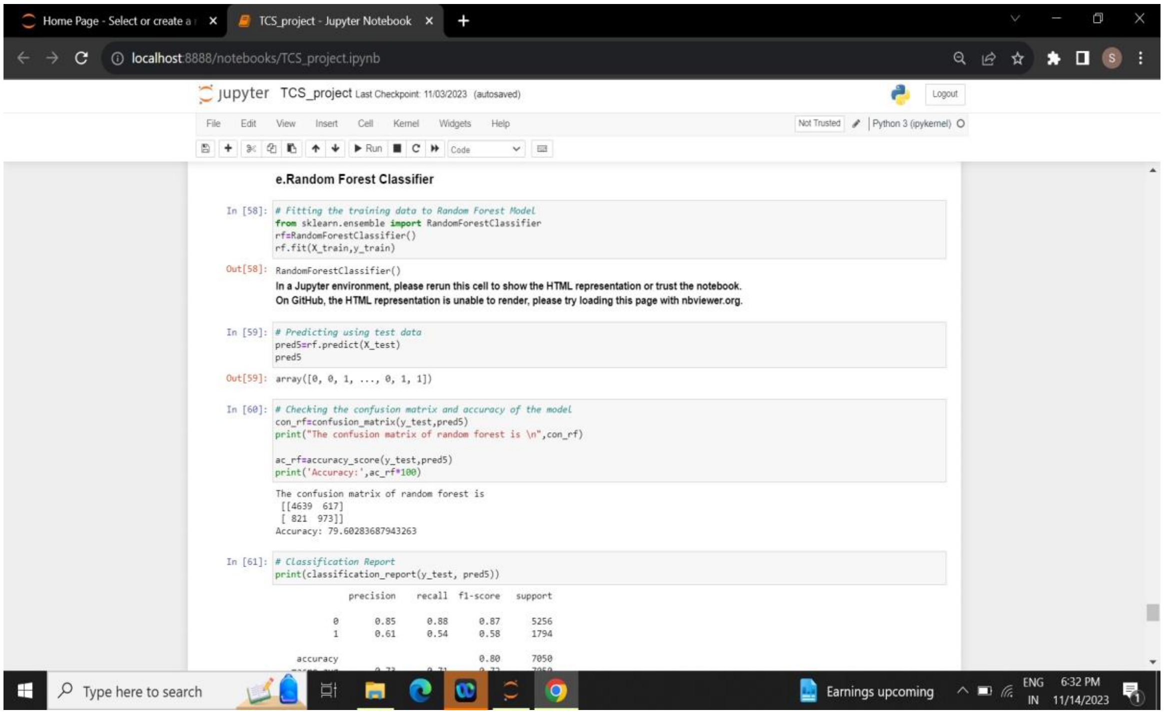
Support vector machine model building



Decision tree model building



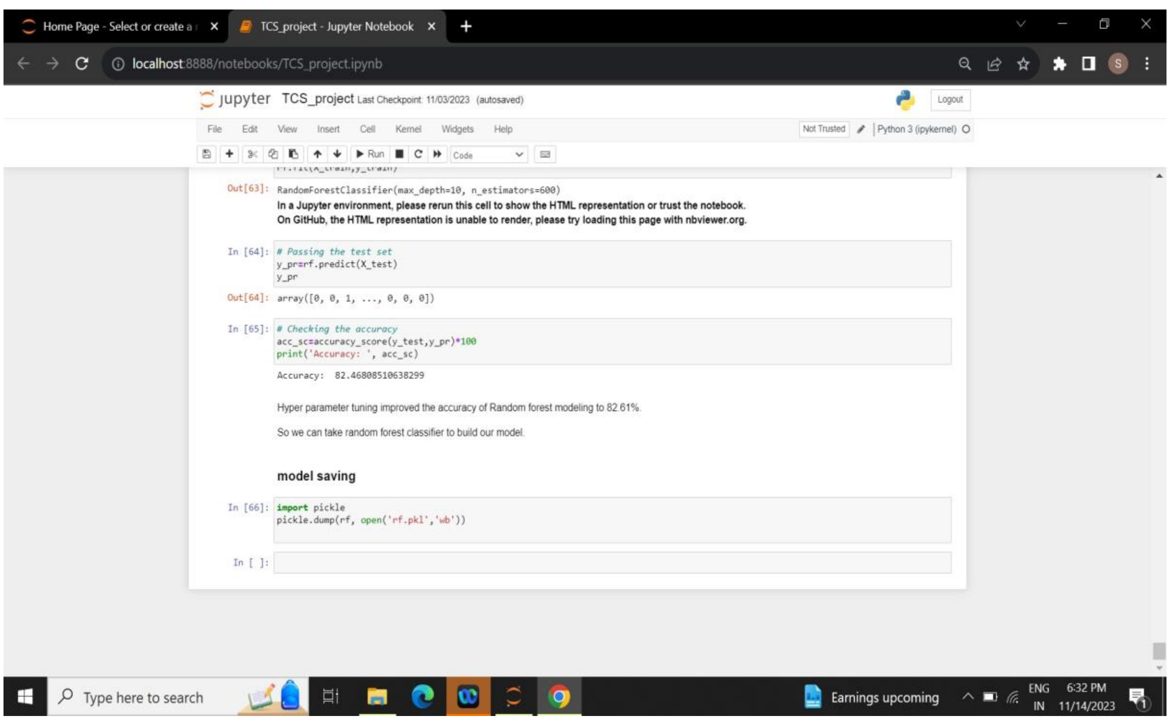
Random forest model building



e.Random Forest Classifier

```
In [58]: # Fitting the training data to Random Forest Model
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier()
rf.fit(X_train,y_train)

Out[58]: RandomForestClassifier()
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [59]: # Predicting using test data
pred5=rf.predict(X_test)
pred5

Out[59]: array([0, 0, 1, ..., 0, 1, 1])

In [60]: # Checking the confusion matrix and accuracy of the model
con_rf=confusion_matrix(y_test,pred5)
print("The confusion matrix of random forest is \n",con_rf)

ac_rf=accuracy_score(y_test,pred5)
print('Accuracy:',ac_rf*100)

The confusion matrix of random forest is
 [[4639  617]
 [ 821  973]]
Accuracy: 79.60283687943263

In [61]: # Classification Report
print(classification_report(y_test, pred5))

              precision    recall  f1-score   support

           0       0.85      0.88      0.87      5256
           1       0.61      0.54      0.58      1794

    accuracy                           0.80      7050
```

Dump the random forest model into a pickle file,



```
Out[63]: RandomForestClassifier(max_depth=10, n_estimators=600)
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [64]: # Passing the test set
y_pr=rf.predict(X_test)
y_pr

Out[64]: array([0, 0, 1, ..., 0, 0, 0])

In [65]: # Checking the accuracy
acc_sc=accuracy_score(y_test,y_pr)*100
print('Accuracy: ', acc_sc)

Accuracy:  82.46808510638299
```

Hyper parameter tuning improved the accuracy of Random forest modeling to 82.61%.

So we can take random forest classifier to build our model.

model saving

```
In [66]: import pickle
pickle.dump(rf, open('rf.pkl','wb'))
```

OUTPUT PAGE:

--end ---