STUDENT DETAILS

NAME :THATIPUDI SATISH KUMAR

EMAILID : SATISHTHATIPUDI599@GMAIL.COM

COLLEGE NAME : GIET ENGINEERING COLLEGE

COLLEGE STATE : ANDHRA PRADESH

INTERNSHIP DOMAIN : ARTIFICIAL INTELLIGENCE & MACHINE LEARNING(AIML)

INTERNSHIP START & END DATE : 03-06-2024 TO 15-07-2024

PROJECT TITLE: EMPLOYEE BURNOUT PREDICTION

- **Project Statement: Predicting Employee Burnout using Linear Regression**
- This project aims to develop a machine learning model to predict employee burnout based on various workplace factors. By analyzing data on workload, work-life balance, job satisfaction, and other relevant features, the model can identify employees at risk of burnout. This information can be used by organizations to implement proactive measures to prevent burnout and support employee well-being.
- The project will focus on building a linear regression model to predict employee burnout scores. The model will be trained and evaluated on a dataset containing employee information and burnout ratings. The evaluation will assess the model's accuracy, precision, and ability to explain the variance in burnout scores.

AGENDA : EMPLOYEE BURNOUT PREDICTION

- **1. Importing Necessary Libraries**
- Imported essential libraries for data manipulation, visualization, and machine learning.
- **2. Loading Dataset**
- Loaded the dataset from an Excel file.
- **3. Data Overview**
- Generated descriptive statistics to understand the dataset's basic properties.
- Checked for unique values and data types in each column.
- Identified and quantified missing values.
- **4. Exploratory Data Analysis (EDA)**
- Analyzed the correlation between numerical features and the target variable ("Burn Rate").
- Visualized the distribution of data using pair plots.
- Dropped rows with missing values in critical columns.

AGENDA : EMPLOYEE BURNOUT PREDICTION

- **5.Data Preprocessing**
- Dropped non-informative columns (e.g., "Employee ID").
- Created a new feature representing employee seniority based on the "Date of Joining".
- Analyzed and dropped columns with low correlation to the target variable.

- **6. Encoding Categorical Variables**
- Visualized the distribution of categorical variables.
- Applied one-hot encoding to categorical variables to prepare them for machine learning.

- **7. Data Splitting and Scaling**
- Split the data into training and testing sets.
- Standardized the feature values using StandardScaler.

AGENDA : EMPLOYEE BURNOUT PREDICTION

- **8. Saving the Scaler:**
- Saved the scaler object using pickle for future use.

- **9. Model Building:**
- Built and trained a linear regression model using the training data.

- **10. Model Evaluation:**
- Evaluated the model's performance using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) score.
- Concluded that the linear regression model performed well based on these metrics, indicating good accuracy and precision in predictions.

PROJECT OVERVIEW

- Predictive Modeling

Objective: Develop and train machine learning models to predict the burnout rate for each employee.

- Data Insights

Objective: Analyze the data to uncover key factors contributing to employee burnout.

- Actionable Recommendations

Objective: Provide insights and recommendations for organizations to effectively mitigate employee burnout.

- **Project Steps :**

Data Understanding and Exploration.

Data Preprocessing.

Model Development.

Model Evaluation and Insights.

Model Evaluation and Insights

WHO ARE THE END USERS OF THIS PROJECT?

- **Employees**
- **Human Resources (HR) Department**
- **Company Executives and Management**
- **Data Scientists and Analysts**
- **Organizational Psychologists**
- **Software Developers**
- **Consultants**

YOUR SOLUTION AND ITS VALUE PROPOSITION

The method entails creating a thorough predictive model to evaluate employee burnout depending on a number of variables, including job characteristics, mental health indicators, and demographics. This model will use cutting-edge machine learning techniques to analyze the dataset and identify important predictors of burnout. The following elements are part of the solution:

- **Proactive Burnout Management:**
- Seek out and resolve possible burnout problems before they worsen.
- **Improved Well-Being of Employees:**
- Encourage a more wholesome and encouraging work atmosphere.
- **Benefits for the Organization:**
- Improve the culture and performance of the organization.
- **Data-Driven Decision Making:**
- Make strategic decisions based on data insights.

HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

- Comprehensive Knowledge of Context
- Advanced Methods for Processing Data
- Choosing and Customizing a Model
- Including Psychological Aspects
- Framework for Continuous Improvement

MODELLING

- **Getting Ready for Data**
- Data splitting: Separate the training and testing sets from the dataset.
- Feature Scaling: Bringing numerical attributes into uniformity.
- Convert categorical variables into a numerical format by encoding them.
- **Models of Training**
- Linear Regression Model: Utilizing the training dataset, train the linear regression model.
- **Evaluation of the Model**
- Performance Metrics: To evaluate the model's performance, use metrics such as RMSE, MAE, and R2 score.
- **Model Modification**
- Feature Selection: To improve model performance, assess and choose the most crucial features.

RESULTS

IMPORTING NECESSARY LIBRARIES

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import pickle as pickle
import os
```

loading dataset

```
[ ] data=pd.read_excel("/content/employee_burnout_analysis-AI.xlsx")
```

Data overview

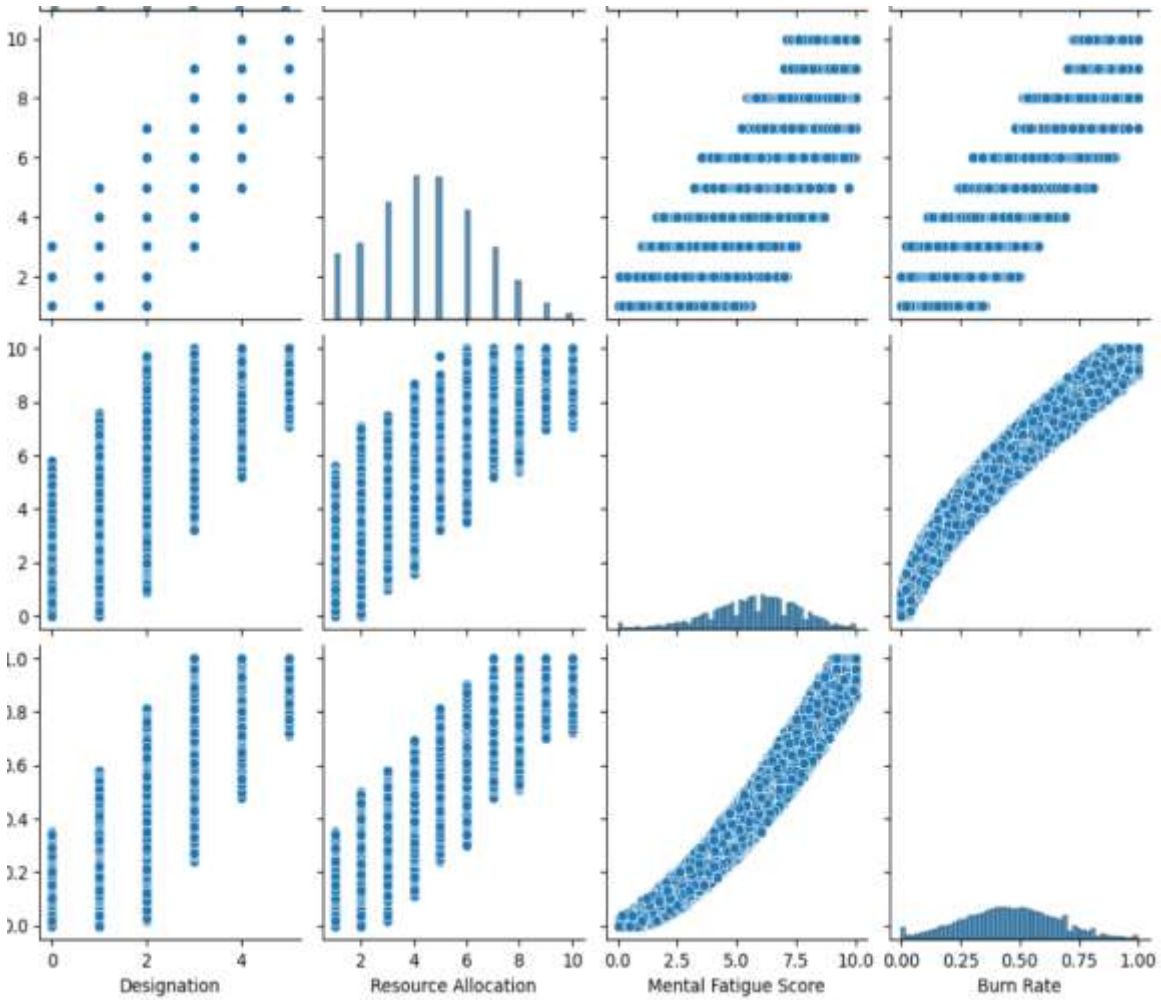
```
data.head()
```

	Employee ID	Date of Joining	Gender	Company Type	WFH Setup Available	Designation	Resource Allocation	Mental Fatigue Score	Burn Rate
0	fffe32000000360033003200	2008-09-30	Female	Service	No	2	3.0	3.8	0.16
1	fffe3700360033003500	2008-11-30	Male	Service	Yes	1	2.0	5.0	0.36
2	fffe3100330032003700380	2008-03-10	Female	Product	Yes	2	NaN	5.8	0.49
3	fffe32003400380032003900	2008-11-03	Male	Service	Yes	1	1.0	2.6	0.20
4	fffe3100380034003100380	2008-07-24	Female	Service	No	3	7.0	6.9	0.52

```
[ ] data.tail()
```

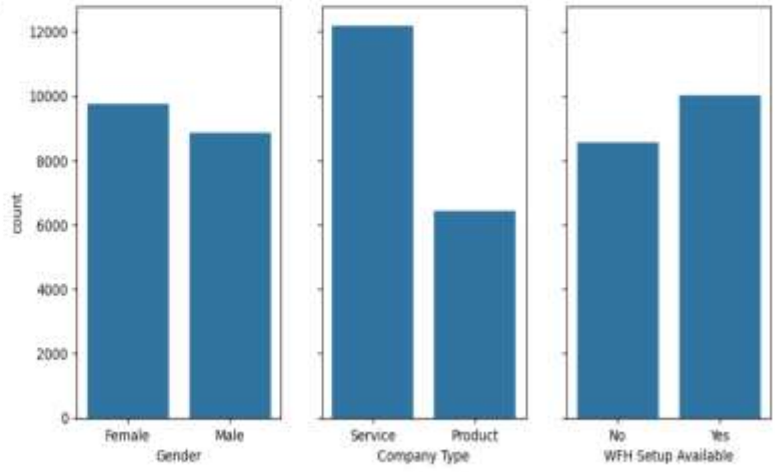
	Employee ID	Date of Joining	Gender	Company Type	WFH Setup Available	Designation	Resource Allocation	Mental Fatigue Score	Burn Rate
22745	fffe3100350037003800310	2008-12-30	Female	Service	No	1	3.0	NaN	0.41
22746	fffe3300330035003100380	2008-01-10	Female	Product	Yes	3	6.0	6.7	0.59
22747	fffe3500320033	2008-11-05	Male	Service	Yes	3	7.0	NaN	0.72
22748	fffe3300330032003800390	2008-01-10	Female	Service	No	2	5.0	5.9	0.52
22749	fffe3400350031003800	2008-01-06	Male	Product	No	3	6.0	7.8	0.61

```
[ ] data.describe()
```



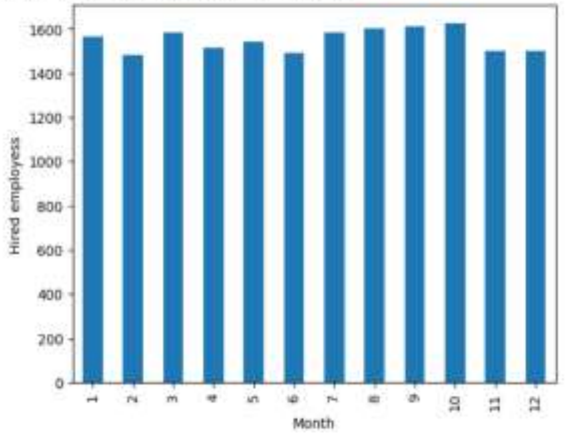
Analysing categorical variables

```python
cat_columns=data.select_dtypes(object).columns
fig, ax= plt.subplots(nrows=1,ncols=len(cat_columns),sharey=True,figsize=(10,5))
for i,c in enumerate(cat_columns):
  sns.countplot(x=c, data=data, ax=ax[i])
plt.show()
```




```python
print(f"Min date of joining: {data['Date of Joining'].min()}")
print(f"Max date of joining: {data['Date of Joining'].max()}")
data_month=data.copy()
data_month['Date of Joining']=data_month['Date of Joining'].astype('datetime64[ns]')
data_month['Date of Joining'].groupby(data_month['Date of Joining'].dt.month).count().
```

```
Min date of joining: 2008-01-01 00:00:00
Max date of joining: 2008-12-31 00:00:00
<Axes: xlabel='Month', ylabel='Hired employess'>
```



Linear Regression

```python
linear_regression_model=LinearRegression()
linear_regression_model.fit(X_train,Y_train)
```

Show hidden output

```python
!pip install scikit-learn
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

Show hidden output

```python
print("Linear Regression Model performance Metrices:")
y_pred=linear_regression_model.predict(X_test)

mse=mean_squared_error(Y_test,y_pred)
print("Mean squared error:",mse)

rmse=mean_squared_error(Y_test,y_pred,squared=False)
print("Root Mean squared Error:",rmse)

mae=mean_absolute_error(Y_test,y_pred)
print("Mean absolute error:",mae)

r2=r2_score(Y_test,y_pred)
print("R2 score:",r2)
```

```
Linear Regression Model performance Metrices:
Mean squared error: [illegible]
Root Mean squared Error: [illegible]
Mean absolute error: [illegible]
R2 score: [illegible]
```

LINKS

- GITHUB LINK ： https://github.com/satishkumar3902/Employee-burnout-prediction.git
- DATASET : https://docs.google.com/spreadsheets/d/18boagTfqt0GVeIR8UmYF7dD0mo-MX_ep/edit?usp=drivesdk&ouid=113260846289829288575&rtpof=true&sd=true
- MAIL : satishthatipudi599@gmail.com