Expert system -employee performance

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import LabelEncoder, StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import confusion\_matrix

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv(R"C:\Users\Hp\Downloads\Extended\_Employee\_Performance\_and\_Productivity\_Data.csv")

df

df.info()

df.columns

df["Education\_Level"].unique()

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

df['Education\_Level'] = le.fit\_transform(df['Education\_Level'])

print(df['Education\_Level'].unique())

df['High\_Performer'] = (df['Performance\_Score'] >= 4).astype(int)

df

plt.figure(figsize=(8, 6))

sns.boxplot(x='High\_Performer', y='Monthly\_Salary', data=df,

palette='Set2')

plt.title("Monthly Salary vs. High Performers")

plt.xlabel("High Performer (1 = Yes, 0 = No)")

plt.ylabel("Monthly Salary")

plt.show()

X = df[['Performance\_Score','Education\_Level',

'Employee\_Satisfaction\_Score', 'Promotions']]

Y = df['High\_Performer']

X\_train, X\_test, Y\_train, Y\_test =train\_test\_split(X, Y,test\_size=

0.2, random\_state= 42)

X\_test

lr = LogisticRegression()

lr.fit(X\_train, Y\_train)

y\_pred = lr.predict(X\_test)

print("Confusion Matrix:\n", confusion\_matrix(Y\_test, y\_pred))

from sklearn.metrics import accuracy\_score

accuracy = accuracy\_score(Y\_test, y\_pred)

print("Accuracy: ",accuracy)

prediction = lr.predict([[5, 1, 1.72, 2]])

print(prediction)

df.hist("Monthly\_Salary" )

sns.boxplot(x='High\_Performer', y='Monthly\_Salary', data=df, palette='Set2')

plt.title("Monthly Salary vs. High Performers")

plt.xlabel("High Performer (1 = Yes, 0 = No)")

plt.ylabel("Monthly Salary")

plt.show()

plt.hist(df['Monthly\_Salary'], bins=30, color='blue', alpha=0.7)

plt.title("Distribution of Monthly Salary")

plt.xlabel("Monthly Salary")

plt.ylabel("Performance")

plt.show()