SET - 1

import pandas as pd

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

from sklearn.linear\_model import LinearRegression

data=pd.read\_csv(r"C:\Users\tp234\Downloads\movie\_dataset.csv")

data

data.head(10)

data['IMDbRating'].max()

data['IMDbRating'].min()

data['BoxOffice'].mean()

X = data[['Budget', 'Duration', 'BoxOffice']]

y = data['IMDbRating']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=1)

model = LinearRegression()

model.fit(X\_train, y\_train)

model.predict(X\_test)

SET – 2

from numpy import \*

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

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

#1

data = pd.read\_csv('Salary\_Data.csv')

data.head(10)

#2

print("Maximum :",data['Salary'].max())

print("Minimum :",data['Salary'].min())

#3

print("Average Salary : ",data['Salary'].mean())

#4

x = data[['YearsExperience']]

y = data['Salary']

#5

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3)

model = LinearRegression()

model.fit(x\_train,y\_train)

#6

model.predict(x\_test)

------------------------------------------------------------------

SET-3

#1

data=read\_csv('movies.csv')

data.tail(10)

#2

data.dropna()

#3

md=data['Profitability'].median()

md

#4

x=data.iloc[:,4:7]

y=data.iloc[:,3]

x

#5

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

#6

model=LogisticRegression()

model.fit(x,y)

model.predict(x\_test)

--------------------------------------------------------------

SET -4

from numpy import \*

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

from sklearn.linear\_model import LogisticRegression

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

#1

data = pd.read\_csv('salaries.csv')

data[['experience\_level','salary']]

#2

plt.scatter(data.experience\_level,data.salary)

from sklearn.preprocessing import LabelEncoder

#3

x = data[['salary']]

y = data['experience\_level']

#4

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

#5

model = LogisticRegression()

model.fit(x\_train,y\_train)

#6

model.predict(x\_test)

#logistioc regression will work with string and int both but linear will only work with int values