

# MACHINE LEARNING Project

## Amazon Sagemaker Studio

The screenshot displays the Amazon Sagemaker Studio interface. On the left, a file explorer shows a project named 'sagemaker-studiolab-nordbois' containing files like 'Getting Started.ipynb', 'm1model.pkl', 'SalaryData.csv', 'Untitled.ipynb', and 'Untitled1.ipynb'. A purple arrow points from the text 'sample data' to the 'SalaryData.csv' file. The main area shows a Jupyter notebook with the following code and output:

```
[1]: !pip install numpy pandas scikit-learn pickle-mixin
```

Requirement already satisfied: numpy in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (1.26.2)  
Requirement already satisfied: pandas in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (2.2.14)  
Requirement already satisfied: scikit-learn in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (1.3.2)  
Requirement already satisfied: pickle-mixin in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (1.0.2)  
Requirement already satisfied: pytz=2020.1 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from pandas) (2023.3)  
Requirement already satisfied: python-dateutil=2.8.2 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from pandas) (2.8.2)  
Requirement already satisfied: tzdata=2022.1 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from pandas) (2023.3)  
Requirement already satisfied: scipy=1.5.0 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from scikit-learn) (1.11.4)  
Requirement already satisfied: threadpoolctl=2.0.0 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from scikit-learn) (3.2.0)  
Requirement already satisfied: joblib=1.1.1 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from scikit-learn) (1.3.2)  
Requirement already satisfied: six=1.5 in /home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages (from python-dateutil=2.8.2->pandas) (1.16.0)

```
[2]: # Import the libraries and dataset
import numpy
import pandas
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import pickle
# Import the dataset using pandas
dataset = pandas.read_csv('SalaryData.csv')
```

Amazon Sagemaker Studio

```
[4]: year = dataset.iloc[:, 1:-1].values
salary = dataset.iloc[:, 1].values
# splitting the dataset in half into the training dataset (used to train model) # and test dataset (used to evaluate how well model does)
year_train, year_test, salary_train, salary_test = train_test_split(year, salary, test_size = 1/2, random_state = 0)
# fitting simple Linear Regression to the training dataset
regressor = LinearRegression()
regressor.fit(year_train, salary_train)
coefficients = regressor.coef_
# predicting the salary test dataset results
salary_pred = regressor.predict(year_test)
# save our object regressor to the file named m1model.pkl
pickle.dump(regressor, open('m1model.pkl', 'wb'))
# loading ml model to compare the outcome
m1model = pickle.load(open('m1model.pkl', 'rb'))
```

output #1


```
[5]: # Computes and displaying testing error, which is on average how far off our
# predictions are from the actual salaries in the testing data: error is high since we # have a small dataset.
testing_error = mean_squared_error(salary_test, salary_pred, squared = False)
print("Testing Error of Model: {:.2f}%".format(testing_error))
```

Testing Error of Model: 44989.11

```
[6]: # Use model to predict one's salary if they have 20.8 years of experience # Final step is to run and display the forecasting for salary
prediction = m1model.predict([[20.8]])
print("Coefficient : {}".format(coefficients[0]))
print("Predicted Salary: $ {:,2f}%".format(prediction[0]))
```

output #2

Coefficient : 3499.9999999999999  
Predicted Salary : \$ 97,300.00

amazon  
**SageMaker Studio Lab**

tcygang

### My project


Runtime status  
**Running**


Runtime remaining ⓘ  
Session: **3 h 32 m**  
Today: **6 h 47 m**

Compute type ⓘ  
☒ CPU ☐ GPU

Stop runtime

Open project

IN COLLABORATION WITH  
DeepLearning.AI

  
New to machine learning?

SalaryData.csv — boot

EXPLORER

> OPEN EDITORS

> BOOT

SalaryData.csv

Users > ro > Downloads > machinelang > SalaryData.csv

1 yearsexperience, salary

2 1.0 , 28000

3 2.0 , 31500

4 5.5 , 78950

5 10.0 , 112500

6

7

