13. Data Science – Machine Learning – Saving model, Joblib & Pickling

Contents

13. Data Science – Machine Learning – Saving model, Joblib & Pickling

1. Save model

- ✓ Once the model got created then we can save that model.
- ✓ There are two ways to save the model
  - By using python File IO pickle concept
  - By using Joblib library

Program Name: Predict price of a home with area = 5000 sqr ft
demo1.py

```
import pandas as pd
from sklearn.linear_model import LinearRegression

df = pd.read_csv("homeprices.csv")
new_df = df.drop('price', axis='columns')

reg = LinearRegression()
reg.fit(new_df, df.price)

# Predictions
print(reg.predict([[5000]]))
```

Output

```
[859554.79452055]
```

2. Pickling

- ✓ The process of writing state of object to the file is called as pickling.

Program Name: Save the model into pickle file
demo2.py

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import pickle

df = pd.read_csv("homeprices.csv")
new_df = df.drop('price', axis='columns')

model = LinearRegression()
model.fit(new_df.values, df.price.values)

with open('model_pickle', 'wb') as file:
        pickle.dump(model, file)

with open('model_pickle', 'rb') as file:
        model1 = pickle.load(file)
        print(model1.predict([[5000]]))
```

Output

```
[859554.79452055]
```

Program Name: Doing prediction by using saved model
demo3.py

```
import pickle

with open('model_pickle', 'rb') as file:
        model1 = pickle.load(file)
        print(model1.predict([[5000]]))
```

Output

```
[859554.79452055]
```

Program Name: Doing prediction by using saved model
demo4.py

```
import pickle

with open('model_pickle', 'rb') as file:
        model1 = pickle.load(file)
        print(model1.predict([[6000]]))
```

Output

```
[995342.46575342]
```

Program Name: Save Trained Model Using Joblib
demo5.py

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import joblib

df = pd.read_csv("homeprices.csv")
new_df = df.drop('price', axis='columns')

model = LinearRegression()
model.fit(new_df.values, df.price.values)

joblib.dump(model, 'model_joblib')

print("Model got saved")
```

Output

```
Model got saved
```

Program Name: Do prediction with saved model

demo6.py

```
import joblib

mj = joblib.load('model_joblib')

print(mj.predict([[5000]]))
```

Output

```
[859554.79452055]
```