Social Media HW1

Link Prediction

Practice of Social Media Analytics

M11115060 廖昭翔

Code explain

In [183…

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
```

Import library

```
train_data = pd.read_csv('new_train_data.csv')
```

```
# split the data
X_train, X_val, y_train, y_val = train_test_split(train_data[['node1', 'node2']], train_data['label'], test_size=

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_val_scaled = scaler.transform(X_val)
```

這裡被切掉 test_size 用0.2去切

Dataset Import and split

Code explain

```
# random forest classifier
model = RandomForestClassifier(n_estimators=1258, random_state=0)
model.fit(X_train_scaled, y_train)
```

Import model(Random Forest)

Set decision tree number to 1258
Initial random_state(random number generator) to 0

```
# predict
val_preds = model.predict(X_val_scaled)

# calculate accuracy
accuracy = accuracy_score(y_val, val_preds)
print(f"Validation accuracy: {accuracy:.4f}")
```

Prediction and print out accuracy in float

Code explain

```
test_data = pd.read_csv('new_test_data.csv')

test_data_scaled = scaler.transform(test_data[['node1', 'node2']])

predictions = model.predict(test_data_scaled)

# add a new column in the test_data
test_data['label'] = predictions
```

Import test data for prediction

```
# Calculate precision, recall, and F1-score
precision = precision_score(test_data['label'], predictions)
recall = recall_score(test_data['label'], predictions)
f1 = f1_score(test_data['label'], predictions)

print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1-score: {f1:.4f}")
```

Print out F1-score and other accuracy pointer

Code explain

```
85…
ans = test_data.drop(['node1', 'node2'], axis=1)
ans = ans.rename(columns={'label': 'ans'})

ans.to_csv('predicted_labels.csv', index=False)
```

Close Index

Save the prediction result into CSV

as the accepted format

How to run my Code

- Install and Import library in Anaconda
    - Installation
        - conda install pandas
        - conda install scikit-learn
    -

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
```

- Open Jupyter Notebook and load HW1 – Link Prediction.ipynb