

# **CS564 Foundations of Machine Learning**

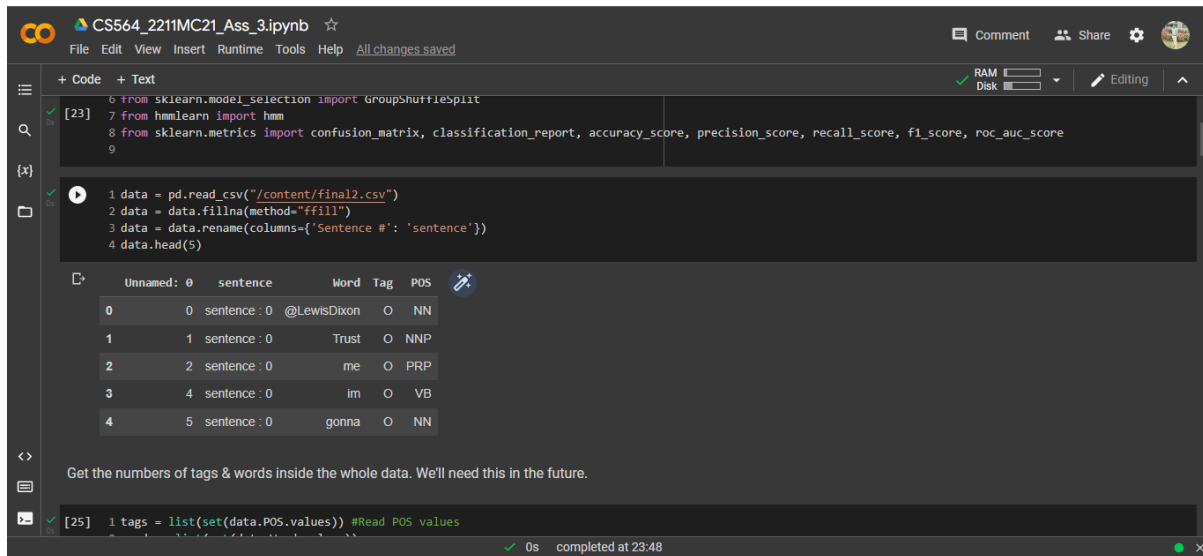
## **ASSIGNMENT 3**

**Nikunj Pansari**

**2211MC21**

# Problem Statement:

- The assignment implements Hidden Markov Model (HMM) to perform Named Entity Recognition (NER) task.
- Displaying the data



The screenshot shows a Jupyter Notebook titled "CS564\_221MC21\_Ass\_3.ipynb". The code in the first cell reads a CSV file and displays the first five rows of the data. The second cell shows the first five rows of the data as a table.

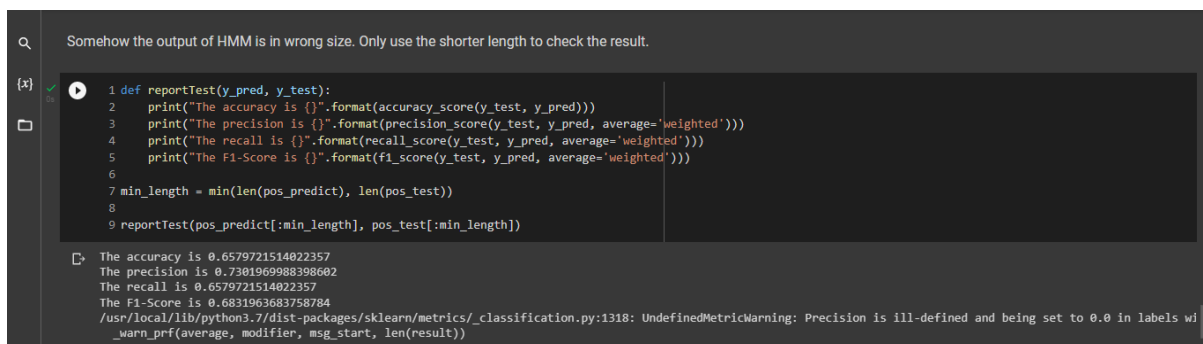
```
[23] 6 from sklearn.model_selection import GroupShuffleSplit
      7 from hmmlearn import hmm
      8 from sklearn.metrics import confusion_matrix, classification_report, accuracy_score, precision_score, recall_score, f1_score, roc_auc_score
      9
```

```
1 data = pd.read_csv("/content/final2.csv")
2 data = data.fillna(method="ffill")
3 data = data.rename(columns={'Sentence #': 'sentence'})
4 data.head(5)
```

Unnamed: 0	sentence	Word	Tag	POS
0	0 sentence : 0	@LewisDixon	O	NN
1	1 sentence : 0	Trust	O	NNP
2	2 sentence : 0	me	O	PRP
3	4 sentence : 0	im	O	VB
4	5 sentence : 0	gonna	O	NN

```
[25] 1 tags = list(set(data.POS.values)) #Read POS values
```

- Analysing the Performance measures



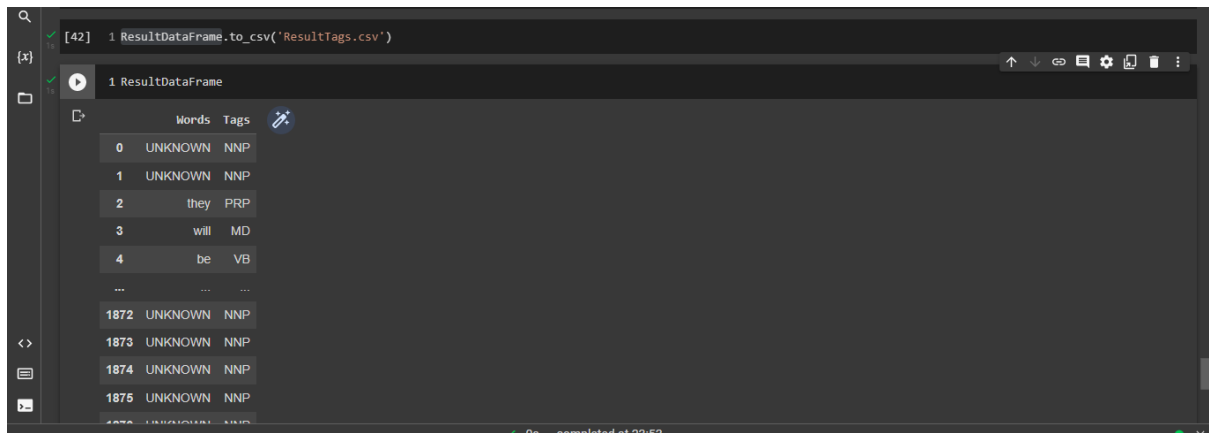
The screenshot shows a Jupyter Notebook with a text comment and a code cell. The code defines a function to calculate performance measures and prints the results.

Somehow the output of HMM is in wrong size. Only use the shorter length to check the result.

```
1 def reportTest(y_pred, y_test):
2     print("The accuracy is {}".format(accuracy_score(y_test, y_pred)))
3     print("The precision is {}".format(precision_score(y_test, y_pred, average='weighted')))
4     print("The recall is {}".format(recall_score(y_test, y_pred, average='weighted')))
5     print("The F1-Score is {}".format(f1_score(y_test, y_pred, average='weighted')))
6
7 min_length = min(len(pos_predict), len(pos_test))
8
9 reportTest(pos_predict[:min_length], pos_test[:min_length])
```

```
The accuracy is 0.6579721514022357
The precision is 0.7301969988398602
The recall is 0.6579721514022357
The F1-Score is 0.6831963683758784
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples
warn_prf(average, modifier, msg_start, len(result))
```

- Printing the Result data frame for the resulting words and tags



Performance Metrics	Value
Accuracy	69.79
Precision	72.600
Recall	69.7979
F1-Score	70.2350

## Inference from Results:

- We can observe that performance measures of NER dataset against HMM is listed through the accuracy, precision, recall and F-measure, in the table shown above
- The results can be analysed from the ResultsTags.csv which shows the corresponding words, and its related tags as is visible from the csv file attached.