

Logistic regression model using title data vectorized with TF-IDF

importing data and packages

```
In [ ]: #import data

%store -r x_title_train_tf
%store -r x_title_test_tf
%store -r y_title_train_tf
%store -r y_title_test_tf

In [ ]: #rename variables more simply for easier interpretation

x_train = x_title_train_tf
x_test = x_title_test_tf
y_train = y_title_train_tf
y_test = y_title_test_tf

In [ ]: from sklearn.metrics import (accuracy_score, recall_score, precision_score, confusion_matrix, ConfusionMatrixDisplay, f1_score)
        from sklearn.linear_model import LogisticRegression
```

Logistic regression

```
In [ ]: lr = LogisticRegression()

lr.fit(x_train, y_train)

Out[ ]: ▾ LogisticRegression
LogisticRegression()

In [ ]: y_pred = lr.predict(x_test)

In [ ]: print("Accuracy:", accuracy_score(y_test, y_pred))

print("Precision:", precision_score(y_test, y_pred))

print("Recall:", recall_score(y_test, y_pred))

print("Recall:", f1_score(y_test, y_pred))

Accuracy: 0.8854053466637688
Precision: 0.8716472749540722
Recall: 0.8702616776718024
Recall: 0.8709539252279263
```