

# Logistic regression model using text data vectorized with TF-IDF

## importing data and packages

```
In [ ]: #import data

%store -r x_text_train_tf
%store -r x_text_test_tf
%store -r y_text_train_tf
%store -r y_text_test_tf

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

x_train = x_text_train_tf
x_test = x_text_test_tf
y_train = y_text_train_tf
y_test = y_text_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: 0.9420778091719192
Precision: 0.9356775300171527
Recall: 0.933846906334067
Recall: 0.9347613219094246
```