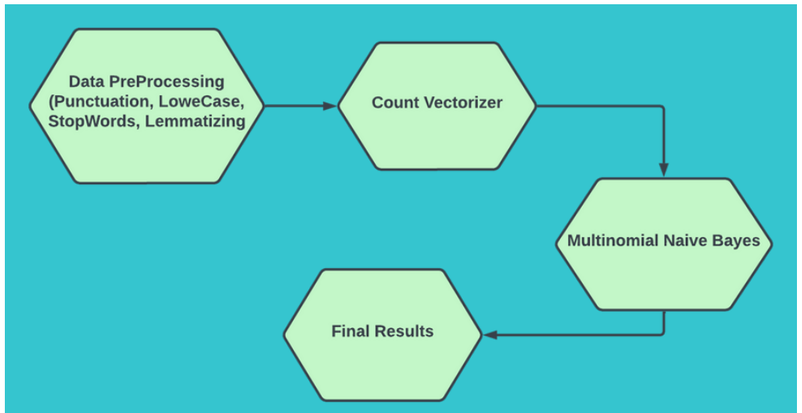
TEXT CLASSIFICATION

MNB MODEL

The Multinomial Naive Bayes (MNB) algorithm uses CountVectorizer to convert a set of text documents and their corresponding categories into numerical format.

MNB APPROACH



ABSTRACT

Text classification is indeed a crucial task in natural language processing (NLP) that involves categorizing text data into predefined categories or labels. In this project we have used different models to classify our data to their categories.

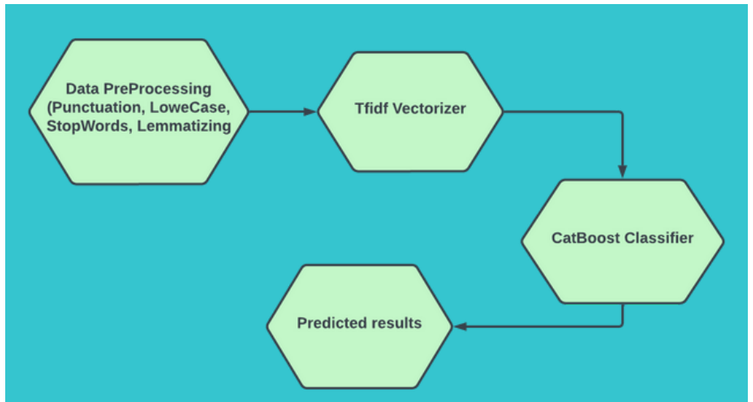
PROBLEM STATEMENT

Our goal is to develop a machine learning model that can accurately assign a label to each of the text from a pre defined set of categories.

CATBOOST MODEL

The CatBoost algorithm offers a strong method of classifying texts by utilising techniques for gradient boosting to increase the model's precision and interpretability.

CATBOOST




```
Best hyperparameters: {'alpha': 0.1, 'class_prior': None, 'fit_prior': True}
BestAccuracy: 0.98
              precision    recall  f1-score   support

    business       0.99      0.99      0.99       111
entertainment       1.00      0.95      0.97        82
    politics       1.00      1.00      1.00        85
       sport       1.00      1.00      1.00        98
        tech       0.95      1.00      0.97        69

    accuracy                           0.99       445
   macro avg       0.99      0.99      0.99       445
weighted avg       0.99      0.99      0.99       445
```

RESULT

By implementing we got to know that MNB gave us the best model, so the second picture shows the accuracy of MNB model and also the best hyper parameters

GROUP 17

SIMRAN KHOJA - AU1910606
KHUSHI PATEL - AU2040068
DEVYASH SHAH - AU2040152
AASTHA GAUDANI - AU2040032