INFO 7610: NLP
Naive Bayes Classifier and LLMs

Shashwat Shahi
shahi.sh@northeastern.edu

AG News Dataset:

Sample Text:

- Fears for T N pension after talks Unions representing workers at Turner Newall say they are 'disappointed' after talks with stricken parent firm Federal Mogul.
———

- The Race is On: Second Private Team Sets Launch Date for Human Spaceflight (SPACE.com) SPACE.com - TORONTO, Canada -- A second\team of rocketeers competing for the #36;10 million Ansari X Prize, a contest for\privately funded suborbital space flight, has officially announced the first\launch date for its manned rocket.
———

- Ky. Company Wins Grant to Study Peptides (AP) AP - A company founded by a chemistry researcher at the University of Louisville won a grant to develop a method of producing better peptides, which are short chains of amino acids, the building blocks of proteins.
———

- Prediction Unit Helps Forecast Wildfires (AP) AP - It's barely dawn when Mike Fitzpatrick starts his shift with a blur of colorful maps, figures and endless charts, but already he knows what the day will bring. Lightning will strike in places he expects. Winds will pick up, moist places will dry and flames will roar
———

- Calif. Aims to Limit Farm-Related Smog (AP) AP - Southern California's smog-fighting agency went after emissions of the bovine variety Friday, adopting the nation's first rules to reduce air pollution from dairy cow manure.

Data Statistics

- Length of training samples: 12000
- Length of testing samples: 7600

Naive Bayes Classifier Results

Model Accuracy: 90.4%

```
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.90      0.91      1900
           1       0.95      0.98      0.97      1900
           2       0.87      0.85      0.86      1900
           3       0.88      0.89      0.88      1900

    accuracy                           0.90      7600
   macro avg       0.90      0.90      0.90      7600
weighted avg       0.90      0.90      0.90      7600
```

LLM Results: with llama 3-1b

Zero Shot Learning

Article Title: Oil prices rise due to increased demand
Predicted Category: World

Article Title: Local team wins championship
Predicted Category: World: Yes

Article Title: New study reveals health benefits of exercise
Predicted Category: World

Article Title: Global leaders meet for climate summit
Predicted Category: World

LLM Results: with llama 3-1b

Example 1:

Article Title: Oil prices rise due to increased demand

Article Body: The global oil market is seeing an increase in prices due to rising demand and limited supply.

Category: Business

Example 2:

Article Title: Local team wins championship

Article Body: The local football team secured a thrilling victory in the national championship.

Category: Sports

Example 3:

Article Title: New study reveals health benefits of exercise

Article Body: Researchers have found that regular exercise can significantly improve mental health.

Category: Science/Technology

Results:

Article Title: Oil prices rise due to increased demand
Predicted Category: Business

Article Title: Local team wins championship
Predicted Category: World

Article Title: New study reveals health benefits of exercise
Predicted Category: Business

Article Title: Global leaders meet for climate summit
Predicted Category: World

Thank You