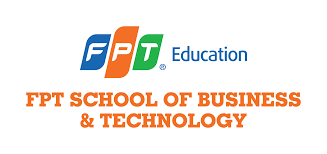
VIỆN QUẢN TRỊ & CÔNG NGHỆ FSB



BÁO CÁO kNN VÀ NAIVE BAYES

Giảng viên hướng dẫn: **TS. QUẢN THÀNH THƠ**

Thành viên: Nguyễn Hoài Phương

Lớp: MSE09

[**Dataset**](#_2f173u5f6kb9) **3**

[Training data](#_qwcitnnomez9) 3

[Test data](#_uld0bfqlpzlq) 3

[**Kết quả sử dụng k-NN**](#_rtwvc2dx5zei) **4**

[**Kết quả sử dụng Bayes (không balancing)**](#_2bg5j4ddh177) **5**

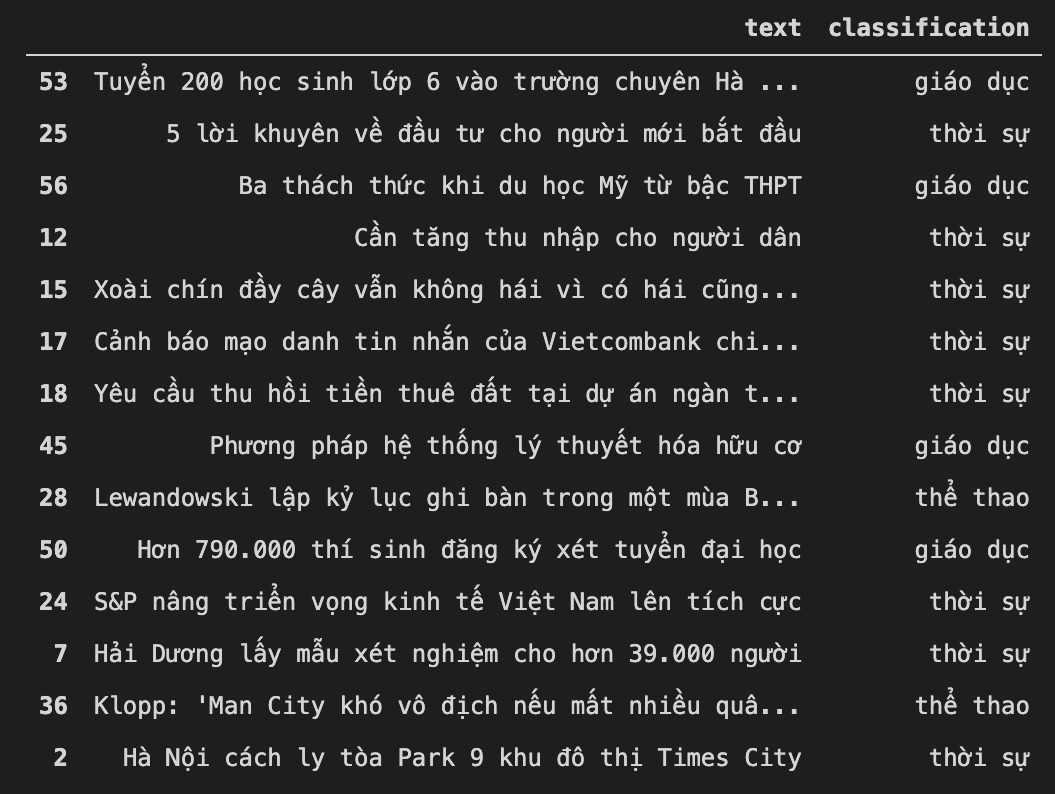
[**Kết quả sử dụng Bayes (có dùng balancing)**](#_odhjkca2x919) **5**

# Dataset

Tập training and test data được lấy từ trang web <https://vnexpress.net/>. Tiêu đề và phân loại bài báo sẽ được lưu vào cột text và classification tương ứng. Classification có thể là “thời sự", “thể thao" hoặc “giáo dục".

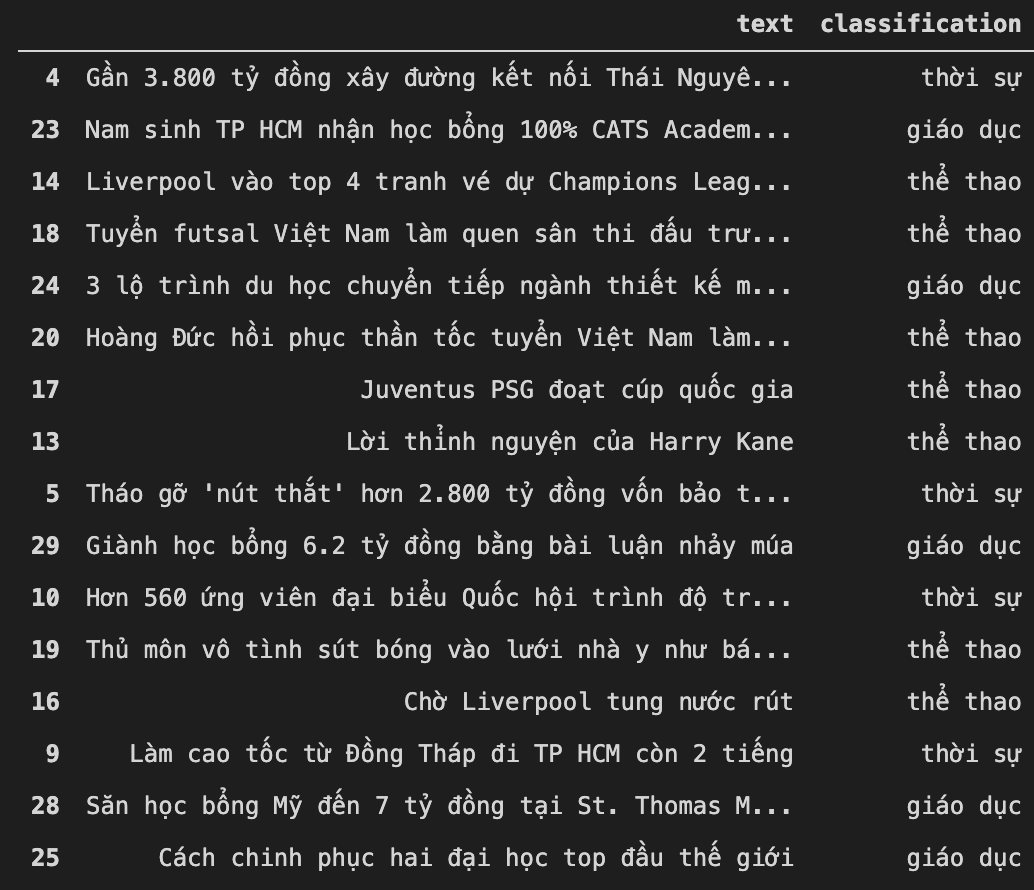
## Training data

Có tổng cộng 27 tiêu đề liên quan tới thời sự, 15 tiêu đề liên quan tới thể thao và 16 tiêu đề liên quan tới giáo dục.



## Test data

Có tổng cộng 10 tiêu đề liên quan tới thời sự, 10 tiêu đề liên quan tới thể thao và 10 tiêu đề liên quan tới giáo dục.



# Kết quả sử dụng k-NN

Accuracy : 0.533333

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| thời sự | 4 | 0 | 6 |
| thể thao | 3 | 2 | 5 |
| giáo dục | 0 | 0 | 10 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| recall | 4/10 = 0.4 | 2/10 = 0 | 10/10 = 1 |
| precision | 4/7 = 0.5714 | 2/2 = 1 | 10/21 = 0.4762 |
| f-measure | 0.4706 | 0 | 0.6452 |

# Kết quả sử dụng Bayes (không balancing)

Accuracy : 0.533333

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| thời sự | 9 | 0 | 1 |
| thể thao | 10 | 0 | 0 |
| giáo dục | 3 | 0 | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| recall | 9/10 = 0.9 | 0/10 = 0 | 7/10 = 0.7 |
| precision | 9/22 = 0.41 | 0/0 | 7/8 = 0.875 |
| f-measure | 0.5621 | 0 | 0.7778 |

# Kết quả sử dụng Bayes (có dùng balancing)

Accuracy : 0.633333

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| thời sự | 5 | 0 | 5 |
| thể thao | 4 | 4 | 2 |
| giáo dục | 0 | 0 | 10 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | thời sự | thể thao | giáo dục |
| recall | 5/10 = 0.5 | 4/10 = 0.4 | 10/10 = 1 |
| precision | 5/9 = 0.5556 | 4/4 = 1 | 10/17 = 0.5882 |
| f-measure | 0.5263 | 0.5714 | 0.7407 |