**TF-IDF** is a widely used method in Natural Language Processing (NLP) and text mining [[17], [18]](#re17). It helps assess how significant a word is within a document, based on both the frequency of that word in the document and the frequency of the word in the entire set of documents. In other words, TF-IDF allows us to determine which words are more prominent in a document compared to other documents in the same set.

In this study, we use TF-IDF as a **method to supplement information about the frequency of word occurrences, which helps to enhance the ability to distinguish between important and unimportant words. This contributes to improving the accuracy of fake news classification without interrupting the context learning process of Transformer models.**

**Specifically, TF-IDF is integrated as an additional layer into the features of the [CLS] token generated by Transformer models. This [CLS] token is the result of the hidden state from the final layer and represents the entire sentence, containing context information synthesized from all the words in the sentence. This approach will keep the process of learning context and the relationships between words in the sentence completely handled by the Transformer without being affected by TF-IDF.**

**TF-IDF** là một phương pháp được sử dụng rộng rãi trong xử lý ngôn ngữ tự nhiên (NLP) và khai thác văn bản [17], [18]. Nó giúp đánh giá mức độ quan trọng của một từ trong một tài liệu, dựa trên tần suất của từ đó trong tài liệu và tần suất của từ trong toàn bộ tập hợp tài liệu. Nói cách khác, TF-IDF cho phép chúng ta xác định từ nào nổi bật hơn trong một tài liệu so với các tài liệu khác trong cùng tập hợp

Trong nghiên cứu này, chúng tôi sử dụng TF-IDF như một phương pháp để **bổ sung các thông tin về tần suất xuất hiện của các từ, giúp tăng cường khả năng phân biệt giữa các từ quan trọng và không quan trọng. Việc này góp phần tăng độ chính xác trong việc phân loại tin giả mà không làm gián đoạn quá trình học ngữ cảnh của các mô hình Transformer.**

**Cụ thể, TF-IDF được tích hợp như một tầng bổ sung vào các đặc trưng của token [CLS] đã được các mô hình Transformer tạo ra. Token [CLS] này là kết quả từ hidden state của layer cuối cùng và đại diện cho toàn bộ câu, chứa thông tin ngữ cảnh tổng hợp từ tất cả các từ trong câu. Cách tiếp cận này sẽ giữ cho quá trình học ngữ cảnh và quan hệ giữa các từ trong câu được xử lý hoàn toàn bởi Transformer mà không bị ảnh hưởng bởi TF-IDF**

**Prolem 1:**

"Tác giả trình bày một phương pháp sử dụng các mô hình Transformer để phát hiện tin giả trên các nền tảng mạng xã hội, với ứng dụng cho tiếng Việt. Họ đã xử lý dữ liệu mẫu không cân bằng, sau đó tích hợp kỹ thuật tiền xử lý dữ liệu TF-IDF, và so sánh nhiều mô hình khác nhau.

Tại sao bạn lại chọn TF-IDF trong bước tiền xử lý? Bạn có điều tra và so sánh các phương pháp khác để xử lý tập dữ liệu không?"

=> Chúng tôi chọn TF-IDF vì phương pháp này giúp bổ sung thông tin về tần suất xuất hiện của các từ, giúp tăng cường khả năng phân biệt giữa các từ quan trọng và không quan trọng. Việc này góp phần tăng độ chính xác trong việc phân loại tin giả mà không làm gián đoạn quá trình học ngữ cảnh của các mô hình Transformer.

Chúng tôi cũng đã thêm thông tin này vào mục “2.2.6 TF-IDF” để làm rõ lý do và cách kết hợp TF-IDF cho mô hình của mình.

Sau khi nộp bản thảo, chúng tôi cũng đã triển khai thêm kết hợp với một số phương pháp khác như Word2Vec và đã mang lại kết quả tốt, chúng tôi cũng đã bổ sung thêm phương pháp Word2Vec vào bản thảo của mình để làm đa dạng hơn các kỹ thuật được kết hợp cùng Transformer thay vì chỉ dùng TF-IDF

Chúng tôi đã thêm mục “2.2.7 Word2Vec” gồm các kiến thức về Word2Vec, đồng cũng thời thêm kết quả đánh giá của PhoBERT + Word2Vec trong “Bảng 1” và các nhận xét tương ứng trong mục “4.3 Results”.

**Data processing**

**Sau bước thu thập chúng tôi đã thu được 1100 tin tức, trong đó có 812 tin đã được gán nhãn dựa trên nguồn thu thập từ các trang uy tín của quốc gia hoặc là các thông tin, kiến thức hoặc sự kiện thực tế đã được chứng minh là đúng hoặc sai. Còn lại 288 tin đã được chúng tôi gán nhãn thủ công dựa trên suy luận từ các kiến thức có sẵn, hoặc dựa theo quan điểm của cộng đồng tại thời điểm thông tin được đăng tải. Ngoài ra, các tin tức về mê tín dị đoan hay tuyên truyền chống phá nhà nước, xã hội đều được mặc định là tin giả trong quá trình gán nhãn. Cuối cùng chúng tôi thu được 703 tin thật và 397 tin giả cho bộ dữ liệu của mình.**

**Sau đó,** chúng tôi thực hiện làm sạch dữ liệu qua các bước sau: loại bỏ các mục trống, không hợp lệ hoặc bị trùng lặp, chuyển tất cả văn bản thành chữ thường và loại bỏ các ký tự đặc biệt và URL. Sau đó, chúng tôi chọn các trường dữ liệu sẽ được sử dụng và dữ liệu còn lại sẽ được hiển thị như trong Hình 3. Ở giai đoạn này, chúng tôi chỉ tập trung vào nội dung của các bài đăng và nhãn phân loại của chúng, nhưng chúng tôi dự định mở rộng nghiên cứu để bao gồm phân tích các bình luận trong tương lai.

Rõ ràng, số lượng mẫu tin giả thấp hơn nhiều so với tin thật (được thể hiện trong biểu đồ ở Hình 4), điều này có thể dẫn đến sự thiên lệch trong việc huấn luyện mô hình và kết quả không chính xác. Để giải quyết vấn đề này, chúng tôi đã thực hiện hai giải pháp:

● Thêm Dữ liệu từ VFND: Chúng tôi đã thêm các bài báo từ bộ dữ liệu VFND, được mô tả trong luận văn của Ho Quang Thanh, "VNFD - Vietnamese Fake News Datasets" [19]. Tuy nhiên, vì bộ dữ liệu này được thu thập từ năm 2019-2020, chúng tôi chỉ chọn những tin tức không thay đổi theo thời gian, chẳng hạn như kiến thức khoa học đã bị chứng minh là sai, và những tin tức không thay đổi theo thời gian như tin đồn về mê tín, hoặc lối sống lệch chuẩn. Dữ liệu bổ sung này chiếm không quá 20% tổng số tin giả mà chúng tôi hiện đang thu thập.

● Sử dụng Kỹ thuật Random Oversampling: Chúng tôi đã sử dụng kỹ thuật Random Oversampling từ thư viện "imbalanced-learn" để cân bằng lại dữ liệu. Kỹ thuật này sao chép ngẫu nhiên các mẫu hiện có của nhãn có số lượng ít hơn cho đến khi số lượng mẫu của các nhãn được cân bằng. **Tuy nhiên chúng tôi không sao chép những mẫu tin được chúng tôi gán nhãn thủ công.**

**Prolem 3:**

Chúng tôi đã chỉnh sửa các câu và cách diễn đạt sai được bạn nhắc đến trong bản thảo, chúng tôi rất biết ơn những góp ý của bạn đã giúp bản thảo của chúng tôi mượt mà hơn.

Chúng tôi cũng đã chỉnh sửa lại số lượng mẫu trong bộ dữ liệu là 1406 mẫu ở mục “4.1 Corpus” và cung cấp URL để truy cập đến bộ dữ liệu của chúng tôi. Đồng thời chúng tôi cũng bổ sung thêm thông tin trong mục “3.3 Data processing” để làm rõ trong bộ dữ liệu có bao nhiêu mẫu là được chúng tôi gán nhãn thủ công và cách mà chúng tôi làm điều đó như thế nào.

Cuối cùng chúng tôi cũng đã thay đổi three numbers after the decimal, instead six numbers ở trong bảng 1.