

Final Report: Fake News Classification in Palestine Using ML and NLP

Students Name :
SALMA T.S. SHAHEEN
RASHA A.M. ABURKAB

Supervised by:
DR TAREQ ALTALMAS

Group C



17 JULY 2025

Abstract :

This project aims to build a machine learning model that classifies Arabic news as either real or fake. We used **Natural Language Processing (NLP)** techniques to clean and prepare the text data. After testing different models, the **XGBoost** classifier gave the best performance with **95%** accuracy. The results show that the model is effective in detecting fake news in Arabic and can be improved further in the future.

Introduction :

Fake news spreads quickly on social media and can cause confusion and harm. Detecting and stopping false information is very important. In this project, we used machine learning and NLP to build a model that can classify Arabic news articles as real or fake. We followed a clear process: analyzing the data, cleaning it, extracting features, training the model, and testing the results.

Dataset Analysis:

The dataset contains columns like id, title, text, platform, and label. The data was unbalanced at first (more real articles), so we fixed this using balancing techniques.

We also found:

- Some platform names written in different formats (**AFP and AFP Fact Check**)
- A few rows had incorrect values in the platform column these were removed ('1', 'معبّر كيسوفيم يستقبل 300 شاحنة')
- Some articles had a very similar title and content.

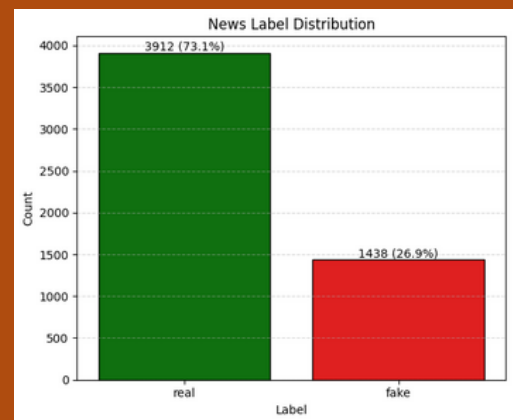
Cleaning steps:

- Checked data shape and unique values.
- Removed invalid rows.
- Fixed platform names (standardization).
- Removed symbols, unwanted spaces, and unclear words.

Analysis :

The dataset structure is focused on both data visualization as bar plots, word clouds, key patterns and insights.

as bar plots, word clouds, and key patterns and in-



Feature Engineering :

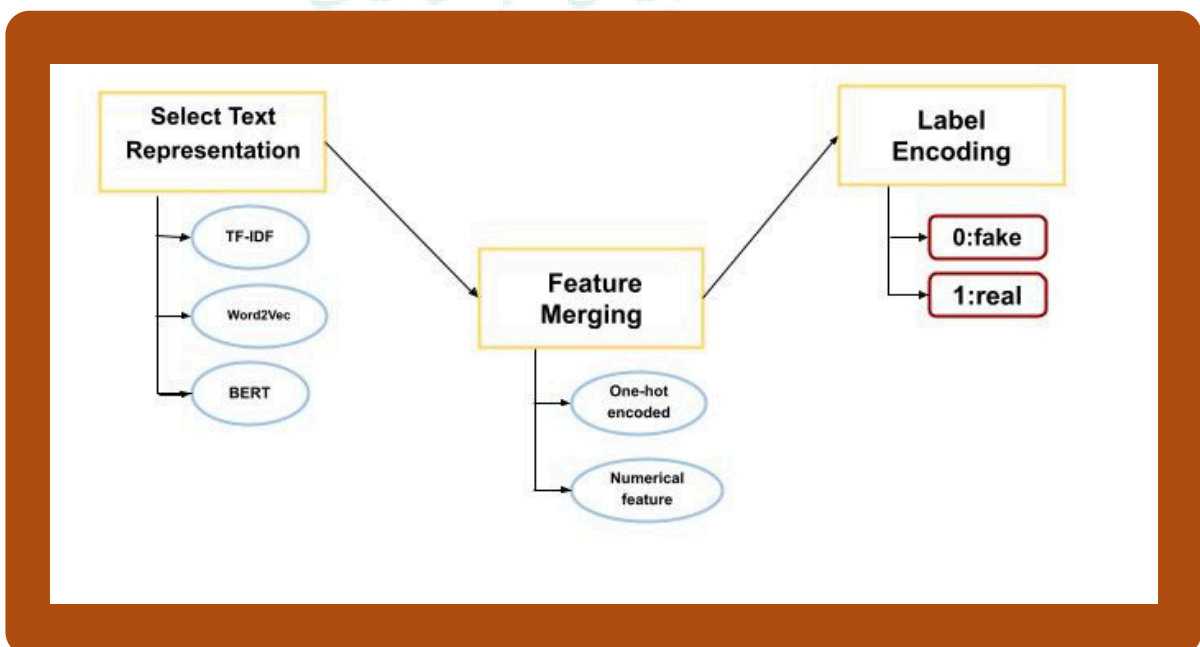
In addition to traditional text preprocessing techniques, we created several custom features to enrich the model's understanding of the content. These include:

- **has_emoji**: Count of emojis in the text
- **has_number**: Count of numerical digits
- **has_punctuation**: Count of punctuation marks
- **has_hashtag**: Count of hashtags
- **char_count**: Total number of characters in the article
- **word_count**: Total number of words in the article

These features helped capture structural and stylistic differences between real and fake news articles, improving the model's performance.

Feature Extraction :

1. Converted text into vector representations using:
 - **TF-IDF**
 - **Word2Vec (CBOW & Skip-gram)**
 - **AraBERT embeddings**
2. Merged text features with:
 - **One-hot encoded platform data**
 - **Numerical features**
3. Encoded labels by converting fake/real categories into binary values (**0** and **1**).



Dataset Quality – Our Opinion

- 1.The dataset is good and useful for fake news classification in Arabic.
- 2.Some news titles and contents were the same content .
- 3.The data was imbalanced, which could cause bias toward certain classes or platforms

Suggestions for better data:

- Collect from different and trusted sources.
- Avoid collecting all data manually to reduce errors.
- Include more topics (not just politics).
- Avoid repeating the same topic from many platforms during the same time.

Discussion – Key Questions ?

1.Should we use lemmatization or stemming?

- Is it better to remove them if the model works well without them?

2.How should we handle numbers in the text?

- Remove them? Replace with "number"? Classify them as small/large? Or keep them?

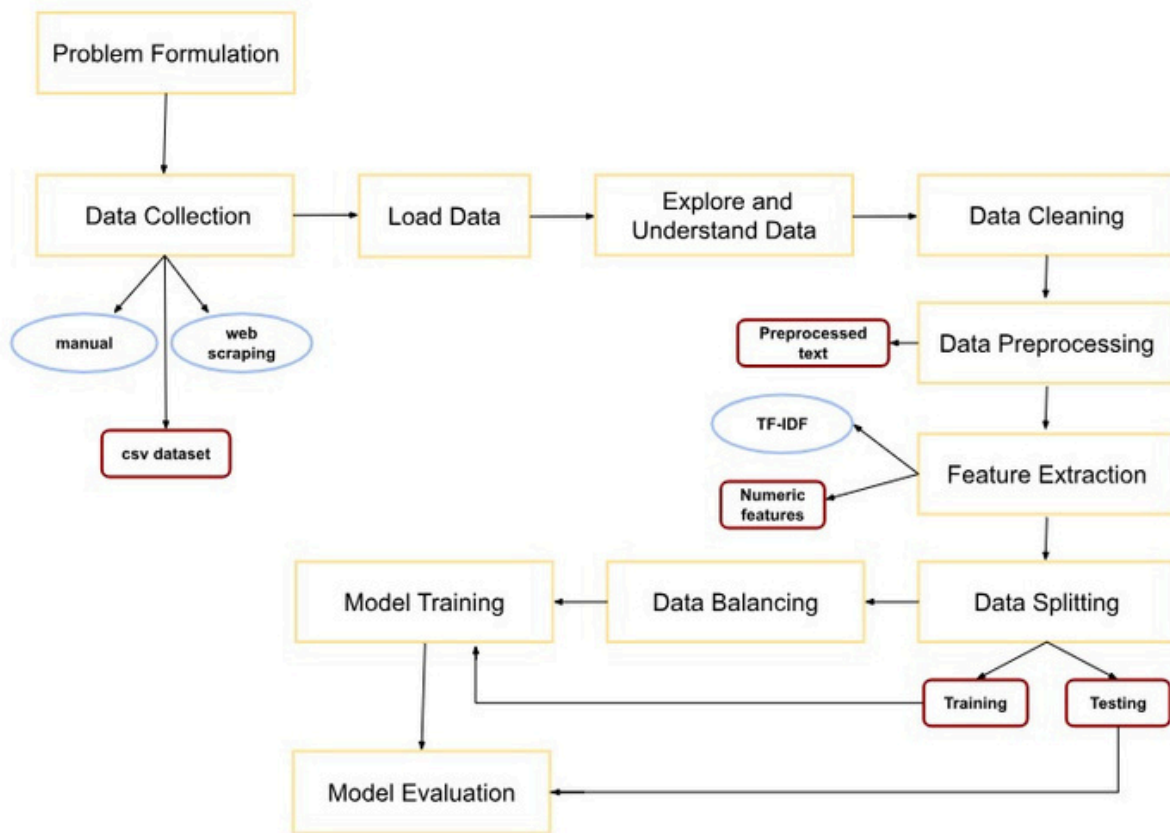
3.What to do with emojis?

- Delete them or convert them to words like "laugh" or "angry"? Do they help detect fake news?

4.How to handle English words in Arabic text?

- Keep them? Translate? Or delete if not important?

Flowchart of Methods



Model Selection

We tested several models for binary classification, including **XGBoost**, **Logistic Regression**, **SVM**, and **AraBERT**. Each model was evaluated on different feature sets (**TF-IDF**, **Word2Vec**, **AraBERT embeddings**) combined with metadata like platform and numerical features.

- XGBoost gave the highest accuracy (**95.05%**) using TF-IDF + Text + Platform.
- It showed balanced performance for both fake and real news.
- Logistic Regression with Word2Vec Skip-gram + Platform reached **92.99%**.
- AraBERT + XGBoost achieved **92.80%**, slightly lower than TF-IDF features.

Final choice: XGBoost with TF-IDF (Text + Platform).

Results :

1. Performance Comparison of Models and Feature Sets

The table below shows the performance of different models using various feature combinations, highlighting the best model for each setup based on accuracy and F1-scores.

Feature Set	Best Model	Accuracy	F1-Score (Fake)	F1-Score (Real)
TF-IDF (Text Only)	XGBoost	92.06%	0.85	0.95
TF-IDF (Text + Platform)	XGBoost	95.05%	0.91	0.97
TF-IDF (Text + Platform + Numerical)	XGBoost	95.05%	0.91	0.97
Word2Vec CBOW (Text Only)	XGBoost	87.76%	0.77	0.92
Word2Vec CBOW (Text + Platform)	XGBoost, Logistic Regression	92.62%	0.88	0.95
Word2Vec CBOW (Text + Platform + Numerical)	Logistic Regression	92.62%	0.88	0.95
Word2Vec Skip-gram (Text Only)	XGBoost	89.16%	0.80	0.93
Word2Vec Skip-gram (Text + Platform)	Logistic Regression	92.99%	0.88	0.95
Word2Vec Skip-gram (Text + Platform + Num.)	XGBoost	93.36%	0.89	0.95
AraBERT (Text + Platform + Numerical)	XGBoost	92.80%	0.88	0.95

2. The model was tested on two news articles from external sources, one fake and one real.

a. Sample 1

```
sample_text = ""
الفيديو لمظاهرة معارضة لتصدير النفط العراقي إلى الأردن بعد أزمة مباراة فلسطين
، الجنوب العراقي يشتعل غضباً ويحرق على الحقول النفطية
. عازماً على منع تصدير النفط إلى الأردن اللوطة.
""
sample_platform = "تويتر"
```

b. Sample 2

```
sample_text = ""
مسؤولة إغاثة: وضع مستشفيات قطاع غزة كارثي
:المتحدثة باسم الاتحاد الدولي لجمعيات الهلال والصليب الأحمر للجزيرة
.وضع مستشفيات قطاع غزة كارثي
.نقص كبير بعدد سيارات الإسعاف العاملة في غزة
.انعدام الوقود يعني غياب الحياة تماماً بغزة
.انعدام الوقود بمستشفيات غزة يعني انعدام فرص الحياة
.فقدنا العديد من موظفينا بسبب القصف على غزة
""
sample_platform = "الجزيرة"
```


Results :

Sample / test case	Platform	Real labeled	Predicted label
sample 1	تويتر	fake	fake
sample 2	الجزيرة	real	real

Based on the results in the table, **the model achieved high performance across multiple feature sets**. It also successfully classified two external examples correctly, demonstrating its ability to generalize beyond the training data.

Conclusion

We successfully built a strong model for classifying Arabic news as real or fake. The results show that machine learning and NLP can help fight misinformation.

Recommendation :

- Use more and newer data.
- Try deep learning models.
- Analyze emotional tone and emojis.
- Classify news into more detailed fake news categories.

References :

- <https://github.com/NNLP-IL/Arabic-Resources>
- <https://github.com/GhiathAjam/Arabic-Tweets-Stance-And-Classification-NLP/blob/main/README.md>
- https://github.com/ArwaData/Arabic-Text-Classification/blob/main/NLP_Project.ipynb

THANK YOU !

