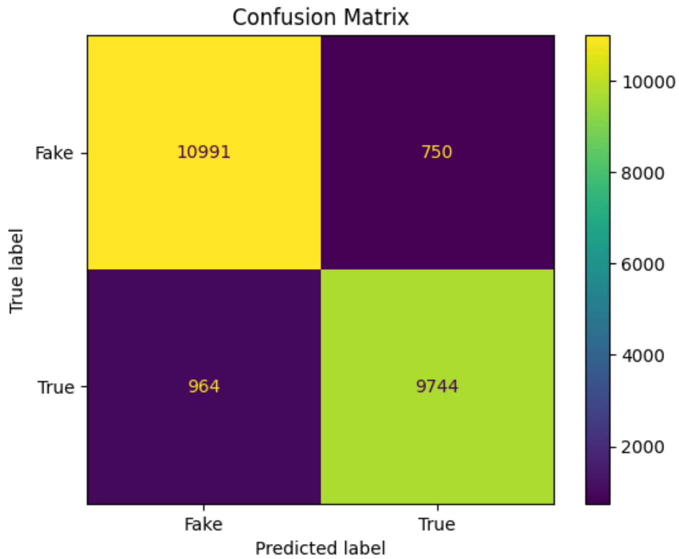
Fake News Detection: Building a Logistic Regression Classifier for Misinformation Detection

In platforms like Twitter/X the misinformation on social media spreads rapidly, influencing the elections and financial markets. We aim to create a feature for news/article's analytics based on mobile app, which will help the users to identify potentially misunderstanding the content if it a Real news or a Fake one before the engagement.

Logistic Regression was chosen due to its **simplicity, interpretability, and efficiency in binary classification tasks**. The dataset was split into **80% training and 20% testing**, allowing the model to learn effectively while ensuring it could generalize to unseen data to address this current issue, this project develops a **Logistic Regression-based fake news classifier** using **TF-IDF vectorization** to analyze news article titles.

The **TF-IDF *(Term Frequency-Inverse Document Frequency)* vectorization** method transforms the text data into numerical features by measuring word importance which will minimize the influence of the common words. It improves in differentiating between pertinent terms linked to false information, unlike basic word counts. This approach makes sure that commonly used terms like "the," "is," or "and" don't have an excessive influence on classification, while more significant words associated with patterns of misinformation are given more weight when making decisions. It is known for its ease of use and efficiency in binary classification, logistic regression offers a probability-based prediction, which makes it appropriate for real-time applications. Instead of just marking an article as true or false, this method lets readers determine the probability that it is a fake by giving predictions probabilities. This model is a starting point that can be enhanced further by fact-checking integrations, ensemble approaches, and context-aware deep learning techniques.

The accuracy of the model, which was trained and evaluated on both fake and actual news titles, was **93.92%.** However, in real-world disinformation detection, **accuracy by itself does not equate to reliability.** Additional metrics like precision, recall, and F1-score are essential for evaluating how effectively the model manages various error kinds, even though accuracy offers an overall assessment of performance. While a high recall suggests greater detection of fake news cases, a high precision means fewer false positives.



Key labeling characteristics are highlighted by the confusion matrix: The model performs well in detecting fake news **(low false positives)**, but the difficulties lie in the trouble in classifying certain legitimate news articles as fake **(false negatives)**, which might undermine public confidence. This problem implies that although the model avoids wrongly detecting an excessive number of actual news pieces, it might find it difficult to identify novel, subtle misleading strategies. Because users may believe dangerous or deceptive content, false negatives—where true misinformation is mistakenly classified as real—present a serious concern.

The high accuracy could mean that the model is overfit with historical data, which would limit its ability to adjust to new patterns of disinformation. This raises additional concerns about **possible overfitting**. When a model is overfit to the dataset it was trained on, it becomes less flexible to change disinformation tactics. If the model is not regularly updated with new training data, its performance may deteriorate as news and misinformation strategies change over time. Although it works well for identifying **pattern-based false information**, the current method has drawbacks. Since **TF-IDF** simply considers word frequency and ignores the semantic significance of news titles, it lacks contextual comprehension. Because of this shortcoming, TF-IDF is unable to comprehend the meaning of phrases, even though it can identify articles that contain frequently used misleading words. Because of this, deceptive headlines that are structured differently from previous examples could go unnoticed. Because of this, the model is susceptible to deceptive manipulation methods. Furthermore, it is challenging to identify sophisticated false news strategies because **logistic regression implies a linear decision limit**, **but disinformation spreads in complicated, non-linear ways.** The absence of an **external fact-checking mechanism** further limits the model's ability to verify flagged content against credible sources.

User feedback methods could be incorporated into the mobile app for real-world use, enabling users to contest or validate content that has been reported, thus increasing the model's reliability. To ensure decision-making openness, **the classifier** should also include explainability elements like a **confidence score** and highlighted words that affected the categorization. Giving users information about the reasons behind an article's flagging can increase trust and facilitate well-informed decision-making. By improving our methodology with a **hybrid modeling strategy** and **real-time adaption mechanisms**, this fake news detection function may more successfully fight false information on Twitter/X and greatly increase user confidence.

References

ChatGPT. (n.d.). https://chatgpt.com/

Gemini - chat to supercharge your ideas. (n.d.). Gemini. https://gemini.google.com/app