Project Title	Fake News Classifier
Skills take away From This Project	Data cleaning, Data augmentation, Text pre-processing, Feature extraction, Sentiment Analysis, Model Selection, Model training and evaluation, Hyper-parameter tuning, Natural Language Processing
Domain	Media and Entertainment

Problem Statement:

Fake News classifier - create a neural network based deep learning model to classify fake news.

Business Use Cases:

The proliferation of fake news on digital platforms poses significant risks to society by spreading misinformation and influencing public opinion. Developing an effective fake news classification system using deep learning can help mitigate these risks by automatically identifying and flagging misleading content. This project aims to build a robust and scalable deep learning model to classify news articles as either genuine or fake.

Approach:

Tasks and Methodology:

1. Data collection and preprocessing: Download the data from the given link. Clean and preprocess the data to handle missing values, outliers, and categorical variables.
2. Exploratory data analysis: Perform EDA to understand data distributions and relationships between variables.
3. Feature Engineering: Create new features that could enhance the predictive power of the model.
4. Model development
    a. Create a baseline model
    b. Deep learning models: RNNs and Transformer-based models
5. Model training and evaluation: Evaluate the performance of various models developed
6. Model Selection: Compare various models using an appropriate evaluation metrics
7. Hyper-parameter tuning

Results:

The project aims to develop a high-performing fake news classifier capable of:

- Achieving a high F1-score to balance precision and recall.
- Handling diverse and large-scale datasets efficiently.
- Providing reliable predictions that can be integrated into news platforms to flag potentially fake news articles.

Project Evaluation metrics:

You need to achieve a high F1-score to balance precision and recall.

Technical Tags:

Natural Language Processing (NLP), Text Preprocessing, Tokenization, Stemming, Lemmatization, Stop Word Removal, TF-IDF (Term Frequency-Inverse Document Frequency), Word Embeddings, Word2Vec, GloVe (Global Vectors for Word Representation), BERT (Bidirectional Encoder Representations from Transformers) LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units)

Data Set:

You can find the dataset here: