|  |  |
| --- | --- |
| **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
   1. Create a baseline model
   2. 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:

<https://drive.google.com/file/d/1ZKVzTnCE-U5uMkopcBsPNj0LFtPTX3z4/view?usp=sharing>

**Data Set Explanation:**

(WELFake) is a dataset of 72,134 news articles with 35,028 real and 37,106 fake news. For this, authors merged four popular news datasets (i.e. Kaggle, McIntire, Reuters, BuzzFeed Political) to prevent over-fitting of classifiers and to provide more text data for better ML training. Dataset contains four columns: Serial number (starting from 0); Title (about the text news heading); Text (about the news content); and Label (0 = fake and 1 = real). There are 78098 data entries in csv file out of which only 72134 entries are accessed as per the data frame.

The title column has the heading of the news article and text has the complete article. Label is the target column.   
You need to convert the title and text columns into mathematical embeddings to train a model.

**Project Deliverables:**

You need to submit a well commented jupyter notebook including predictions on the test data set and F1 score for the same.

**Project Guidelines:**

Best Practices

Coding Standards: Standard code standard for Python code.

Version Control: Use Git for version control and regularly commit changes.

Documentation: Comment your code and provide clear explanations for your logic.

Collaboration: Use collaborative tools like GitHub or GitLab for team projects.

**Timeline:**

Define the project timeline, including milestones and deadlines.

**PROJECT DOUBT CLARIFICATION SESSION ( PROJECT AND CLASS DOUBTS)**

**About Session:** The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

**Note: Book the slot at least before 12:00 Pm on the same day**

**Timing: Tuesday, Thursday, Saturday (5:00PM to 7:00PM)**

**Booking link :<https://forms.gle/XC553oSbMJ2Gcfug9>**

**LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)**

**About Session:** The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

**Note: This form will Open on Saturday and Sunday Only on Every Week**

**Timing: Monday-Saturday (11:30PM to 12:30PM)**

**Booking link :** [**https://forms.gle/1m2Gsro41fLtZurRA**](https://forms.gle/1m2Gsro41fLtZurRA)