

Requirements for Sports-Related Sentiment Analysis Code

1. Data Requirements

- **Dataset:** Sports-related text data with labeled sentiment categories (e.g., positive, negative, neutral).
 - **File Format:** Excel file (all_data.xlsx) containing at least two columns:
 - Cleaned_Text: Preprocessed sports-related comments.
 - Category: Corresponding sentiment labels.
-

2. Libraries and Frameworks

Core Libraries

- **Pandas:** Data manipulation (pd.read_excel, df.dropna).
- **NumPy:** Numerical operations (np.array, np.isfinite).

Machine Learning & Preprocessing

- **Scikit-Learn:**
 - Model selection (train_test_split)
 - Preprocessing (CountVectorizer, LabelEncoder)
 - Models (MultinomialNB, RandomForestClassifier, KNeighborsClassifier)
 - Evaluation (accuracy_score, classification_report)
 - Model stacking (StackingClassifier)

Deep Learning

- **Keras / TensorFlow:**
 - Model building (Sequential, Model)
 - Layers (Embedding, Conv1D, GlobalMaxPooling1D, Dense, Dropout)
 - Custom layer creation (SimpleTransformerBlock with MultiHeadAttention, LayerNormalization).

3. Model Requirements

Models Used:

1. **Multinomial Naive Bayes (MNB):** For Bag-of-Words features.
2. **CNN Model:** For deep learning text classification using embeddings.
3. **Transformer Neural Network (TNN):** Simplified Transformer-based model.
4. **Stacking Classifier:** Combines multiple models into a stronger ensemble.

4. Preprocessing Steps

- **Data Cleaning:** Remove missing values from the Cleaned_Text column.
- **Text Tokenization:** Convert text into sequences using Tokenizer.
- **Padding:** Ensure consistent input size using pad_sequences.
- **Label Encoding:** Convert sentiment labels to numeric format using LabelEncoder.

5. Evaluation Metrics

- **Classification Report:** Detailed metrics (precision, recall, F1-score).
- **Accuracy Score:** Model performance evaluation.

6. Additional Requirements

- **Hardware Requirements:** GPU-enabled system (Google Colab recommended).
- **Software Environment:** Python 3.7+, Google Colab, or a local machine with required libraries installed.

These requirements ensure a robust implementation of sentiment analysis on sports-related data using advanced ML and DL models.