**Deep Learning**

**Project Submission Document**

**Name**: SuryaPrakash P  
**Register Number**: 23CSEG34  
**Google Colab Notebook**: https://colab.research.google.com/drive/15WMm4tEJvCqYvg0JBRn0bvP3soeXk3GW?usp=sharing  
**GitHub Repository**: https://github.com/Sury5000/NewsArticle

**Problem Statement**

In this project, we aimed to classify news articles into various genres using text data. The goal was to build a machine learning model that accurately categorizes articles into predefined genres, such as sports, technology, politics, and more. The problem involves text classification, which is essential for organizing and retrieving news articles efficiently.

**Dataset Source and Description**

**Dataset Source**: The dataset was sourced from the News API.  
**Dataset Description**: The dataset consists of news articles collected from various sources across different genres. Each article includes fields such as title, description, and genre. The genres represent the category of the news article, such as sports, technology, politics, etc.

**Model Training and Regularization Techniques**

**Model**: A Long Short-Term Memory (LSTM) network was used for text classification.

**Regularization Techniques**:

* **L2 Regularization**: Applied to the LSTM layer to penalize large weights and prevent overfitting.
* **Dropout**: Used with a rate of 0.5 to randomly drop units during training to improve model generalization.

**Training Details**:

* **Tokenizer**: Tokenized the text data and padded sequences to a fixed length.
* **Epochs**: Trained for 10 epochs.
* **Batch Size**: Set to 32.
* **Loss Function**: Categorical Crossentropy.
* **Optimizer**: Adam.

**Results**

* **Validation Accuracy**: [Validation Accuracy Value]
* **Validation Loss**: [Validation Loss Value]

**Plots:**

1. **Loss Curve**: The plot shows the training and validation loss over epochs, indicating how well the model is learning.
2. **Accuracy Curve**: The plot displays training and validation accuracy, highlighting the model’s performance.
3. **Confusion Matrix**: Visual representation of the model’s performance across different genres, showing the distribution of true vs. predicted classifications.