

NAME: DARSHINI J

ROLL.NO: 2021503011

NAAN MUDHALVAN - GENERATIVE AI PROJECT

NEXT WORD GENERATION

ABSTRACT:

This project aims to generate text using a deep learning approach in Python, specifically focusing on next-word prediction. Leveraging NLTK for text pre-processing and TensorFlow/Keras for model construction, the system utilizes a dataset of news articles to train a Long Short-Term Memory (LSTM) network. The trained model is then capable of predicting the next word given a sequence of words from the input text. Key steps involve tokenization, creating input-output pairs, model training, and finally, generating text based on user-defined input. The generated text can potentially serve various applications, including text completion, creative writing assistance, and content generation.

OBJECTIVE:

- Develop a text generation system in Python.
- Utilize deep learning techniques, specifically LSTM networks.
- Train the LSTM model on a dataset of news articles.
- Pre-process the text data to prepare it for model training.
- Construct the LSTM model for next-word prediction.
- Train the model using the pre-processed text data.
- Implement functionality to generate text predictions based on given input sequences.
- Demonstrate the capabilities of deep learning in language modelling and text generation tasks.
- Provide a foundation for building more advanced natural language processing applications.

NOVELTY:

This project innovatively combines NLTK for text pre-processing and TensorFlow/Keras for deep learning model development to create a text generation system. While text generation using deep learning techniques is not new, this project's novelty lies in its focus on news article data for training and its application-oriented approach. Additionally, the project showcases the flexibility and creativity of LSTM networks in generating coherent text based on input sequences. The system's ability to generate text predictions can be valuable for various applications, from assisting writers in content creation to enhancing natural language understanding in chatbots and virtual assistants.

KEYWORDS:

Python, Natural Language Processing, NLTK, TensorFlow, Keras, Deep Learning, LSTM, Text Generation, Next-Word Prediction, News Articles, Text Pre-processing, Model Training