



DEPARTMENT OF APEX INSTITUTE OF TECHNOLOGY

PROJECT PROPOSAL

- **Project Title:**

i To study and compare different word embedding learning techniques.

- **Project Scope:**

The scope of the project to study and compare different word embedding learning techniques includes the following:

- 1. Data Collection: The project will collect a large corpus of text data from various sources such as news articles, social media posts, and books. The data will be diverse and representative of different domains and languages.*
- 2. Preprocessing: The collected data will be preprocessed to remove stop words, punctuation, and special characters, convert text to lowercase, and tokenize the text.*
- 3. Word Embedding Techniques: The project will explore and compare different word embedding techniques, including traditional techniques such as word2vec and GloVe and more recent techniques such as BERT and Transformer-based models.*
- 4. NLP Tasks: The project will evaluate the performance of each word embedding technique on various NLP tasks such as sentiment analysis, text classification, and machine translation.*
- 5. Performance Evaluation: The project will evaluate the performance of each word embedding technique based on their accuracy, precision, recall, and F1 score.*

The project will not cover the implementation details of each word embedding technique. Instead, it will focus on the high-level overview of each technique and compare their performance on various NLP tasks. The project will also not cover the training process of each technique but will use pre-trained models to evaluate their performance.

Overall, the project aims to provide insights into the effectiveness of different word embedding techniques for various NLP tasks and their suitability for different types of data and applications.

3. Requirements:

- i** Python Version 3.8 or above
- i** Jupyter Notebook
- i** NLP libraries like nltk, gensim

STUDENTS DETAILS

Name	UID
ABHIRAJ THAKUR	20BCS6802

Approval and Authority to Proceed:

We approve the project as described above, and authorize the team to proceed.

Name	Title	
------	-------	--

SHIKHA GUPTA