AI-powered product search system for an e-commerce website

By Team **Code Wizards**

First Phase evaluation Report

Team Leader: Yash Agrawal

Team Members: Parth Madan, Pallav Sharma, Nitin Kumar Das, Keshav Garg

Project Description: Our project aims to improve search query for e-commerce website by incorporating semantic search rather than relying on old school key word search method, which improves customers retention time which grow online revenue, increases average order value and allows personalization pipelined for specific customer's need. Even googles journey since the bought semantic database Freebase as the basis for the knowledge graph in 2010 to their introduction of MUM is quite remarkable.

Work Done till now:

1. We have created the interface for the search engine using a dummy API of Wikipedia. It extracts the data from the Wikipedia API and results are shown to the users in a very interactive manner.
2. We extracted a part of the flip kart dataset and done data preprocessing on it which includes NLP methods. These methods include following steps to perform:
    a. Tokenization:
        i. Split the text into sentences using sentence segmentation
        ii. Split sentences into words using word tokenization
    b. Text Normalization:
        i. Convert text to lowercase
        ii. Remove punctuation, special characters, and stop words.
        iii. Perform stemming or lemmatization to convert words to their base form
    c. Parts-of-speech Tagging:
        i. Tag each word in a sentence with its grammatical role (e.g., noun, verb, adjective, etc.)
3. In the next step we embed the data using the Hugging face transformers library. We followed the following steps:
    a. We install the library and loaded a pre-trained model.
    b. Then we pass the tokenized text to the model to get the embeddings as the output.

    c. We now extracted the embeddings.
4. We have done the work till extracting these embeddings.

Now we have the embeddings with us and now we will use them in the vector space model to get the cosine similarity and approximate nearest neighbors then we will rank them to show the most precise and accurate results to the users.

By this we will develop a fully function AI based search system for the e-commerce website which is both efficient and incorporate semantic search.