# Cloth-Smiths

**09113702022 Jatin Alagh**

**03613702022 Asif Ali**

**04713702022 Vansh Khandelwal**

**06421102022 Krit Barnwal**

### Abstract

The project focuses on developing a recommendation system using Natural Language Processing (NLP) techniques, machine learning models, and deep learning frameworks. The primary objective is to preprocess product descriptions, create a TF-IDF vectorizer, and compute cosine similarity to recommend similar products based on a given description. Additionally, the project employs a ResNet-50 model for image-based feature extraction to enhance recommendation accuracy. The project demonstrates data collection, preprocessing, model development, deployment, and testing processes. This system's relevance lies in its potential applications in e-commerce and retail industries, where personalized product recommendations can enhance user experience and boost sales.

### Introduction

#### Frameworks and Libraries Used

The project utilizes various frameworks and libraries, including:

* **Pandas:** For data manipulation and analysis.
* **NLTK (Natural Language Toolkit):** For text preprocessing tasks such as tokenization, stop word removal, and stemming.
* **Scikit-learn:** For implementing the TF-IDF vectorizer and calculating cosine similarity.
* **TensorFlow/Keras:** For leveraging a pre-trained ResNet-50 model to extract image features, facilitating a hybrid recommendation system that combines text and image data.

#### Relevance in Industry

Recommendation systems are crucial in the e-commerce industry as they help personalize the shopping experience, increase customer satisfaction, and drive sales. By suggesting relevant products, businesses can enhance user engagement and retention. The incorporation of both textual and visual data into the recommendation system further enhances its capability to deliver accurate and diverse suggestions, catering to different customer preferences.

### Data Collection and Preprocessing

The dataset comprises product information loaded from a CSV file (data.csv). The productDisplayName column is preprocessed using the following steps:

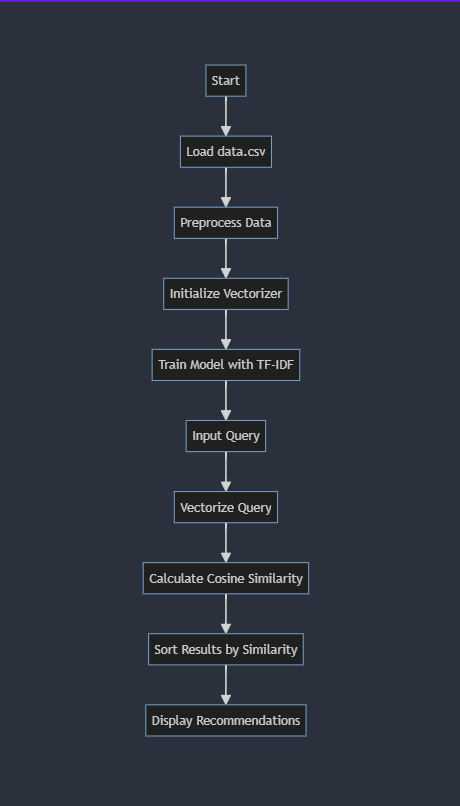
1. **Lowercasing:** Converts text to lowercase.
2. **Removing Punctuation and Special Characters:** Strips non-alphanumeric characters.
3. **Tokenization:** Splits text into individual words.
4. **Stop Words Removal and Stemming:** Filters out common stop words and reduces words to their root forms using NLTK.

Additionally, product images are processed using a pre-trained ResNet-50 model to extract visual features. The dataset contains various columns such as id, productDisplayName, productCategory, productDescription, price, brand, and imagePath, providing comprehensive information necessary for building the recommendation system.

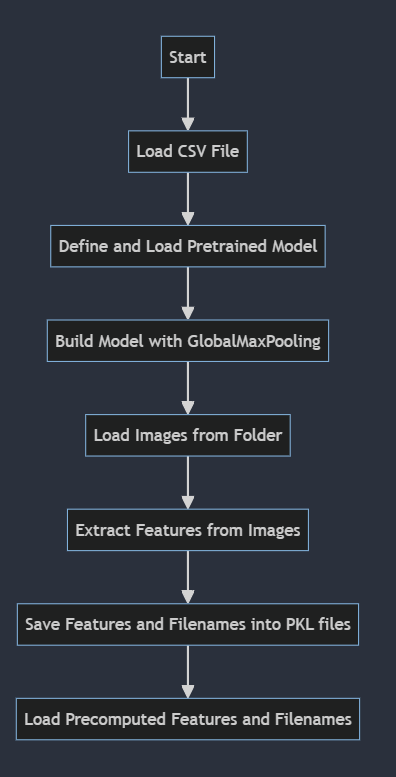
### Methodology

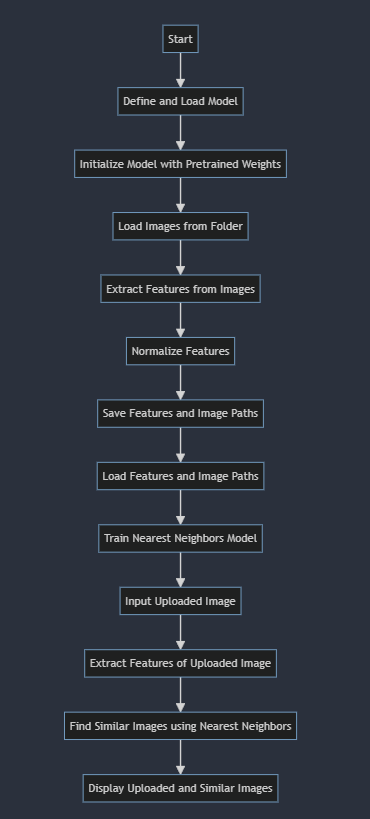
The methodology involves preprocessing the text data, vectorizing the text using TF-IDF, and computing cosine similarity to recommend outfits. The steps include:

1. **Data Preprocessing:** Cleaning and preparing text data for vectorization and extracting features from images.
2. **Vectorization:** Converting text data into numerical vectors using TF-IDF.
3. **Image Feature Extraction:** Using a ResNet-50 model to obtain feature representations from product images.
4. **Similarity Computation:** Calculating cosine similarity between text vectors and combining them with image features to find similar products.
5. **Recommendation:** Developing a function to recommend the top five similar outfits based on combined text and image similarity.

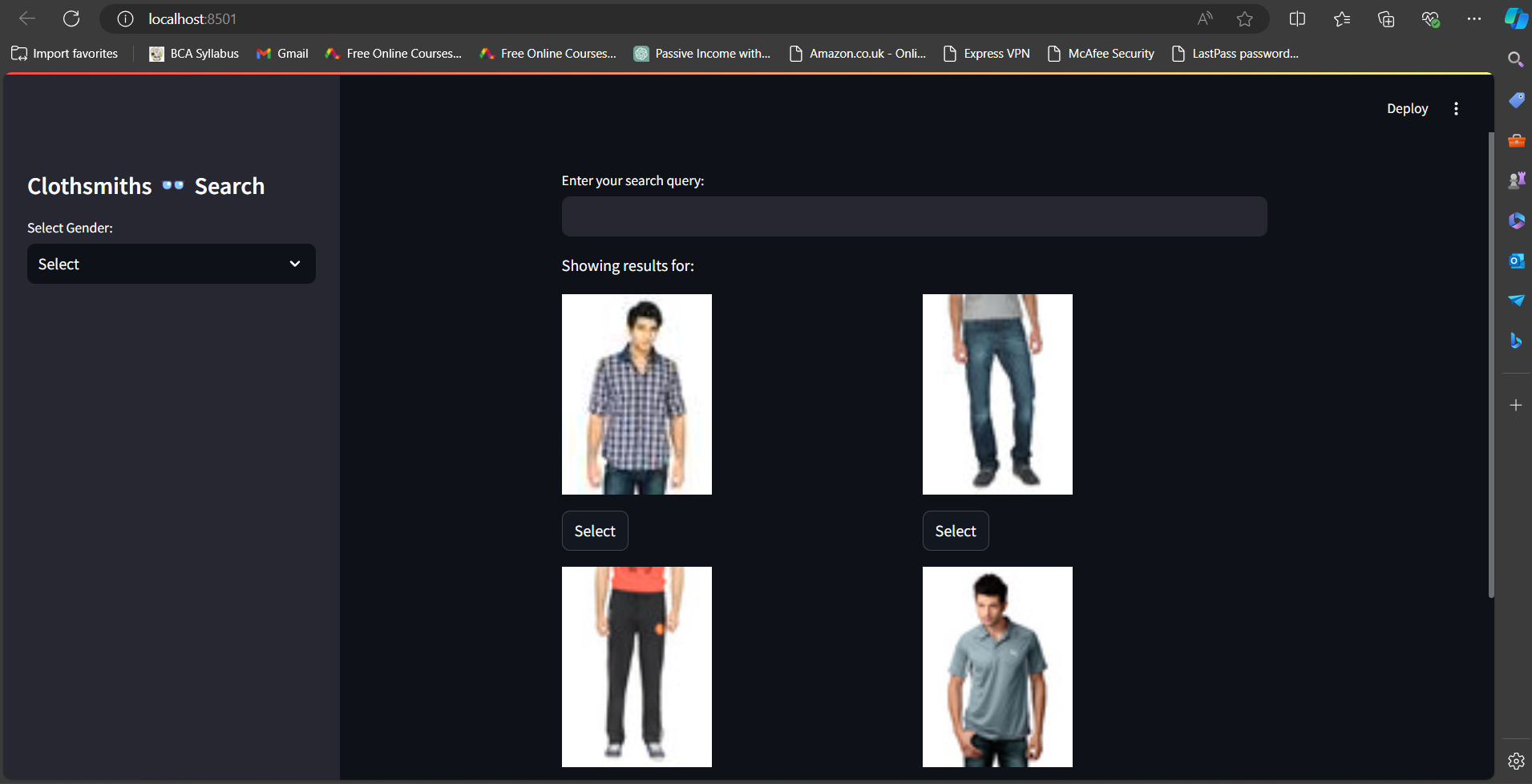
**BASIC RECOMMENDATION**

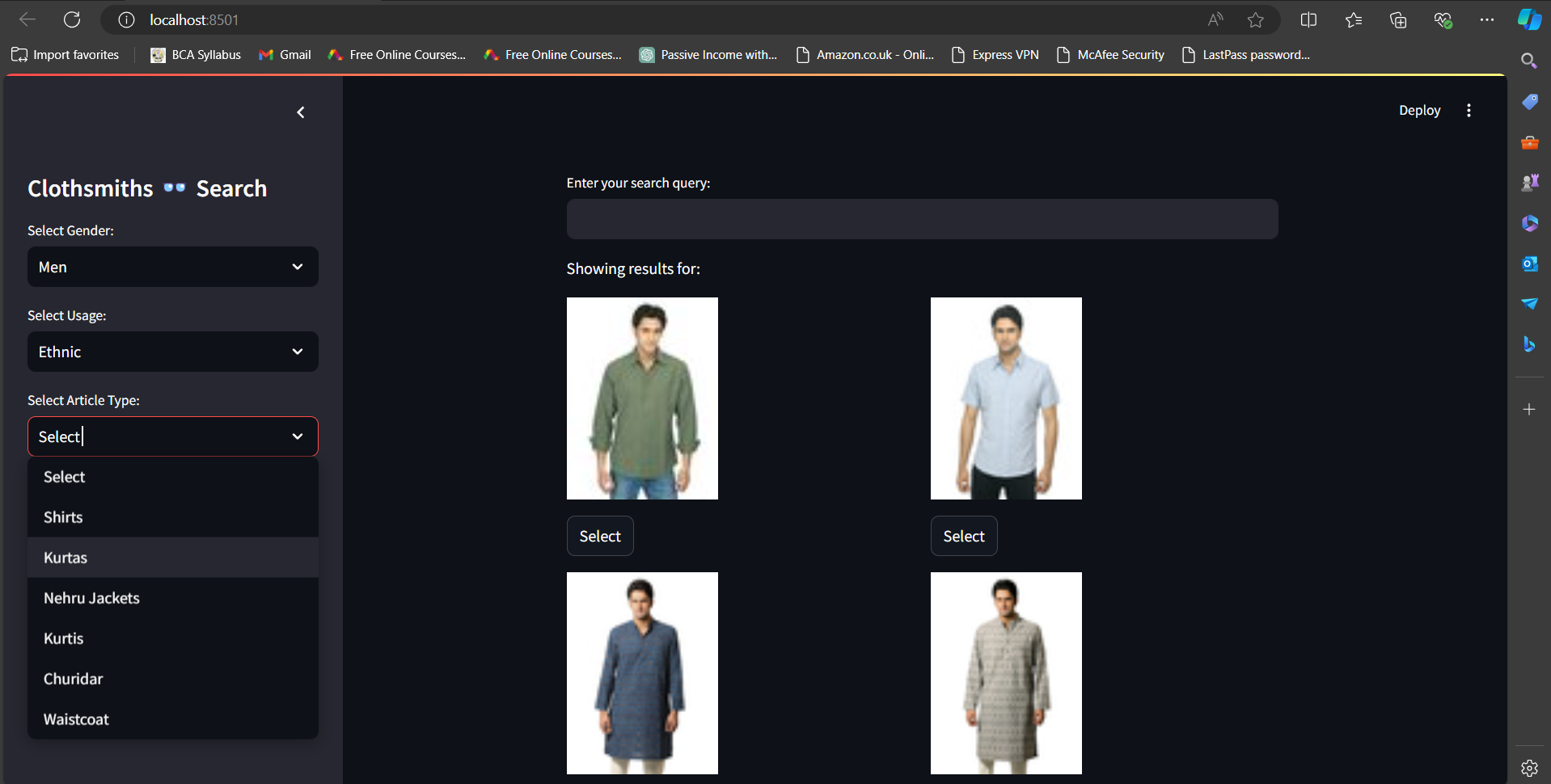
**MODEL TRANSFER LEARNING**

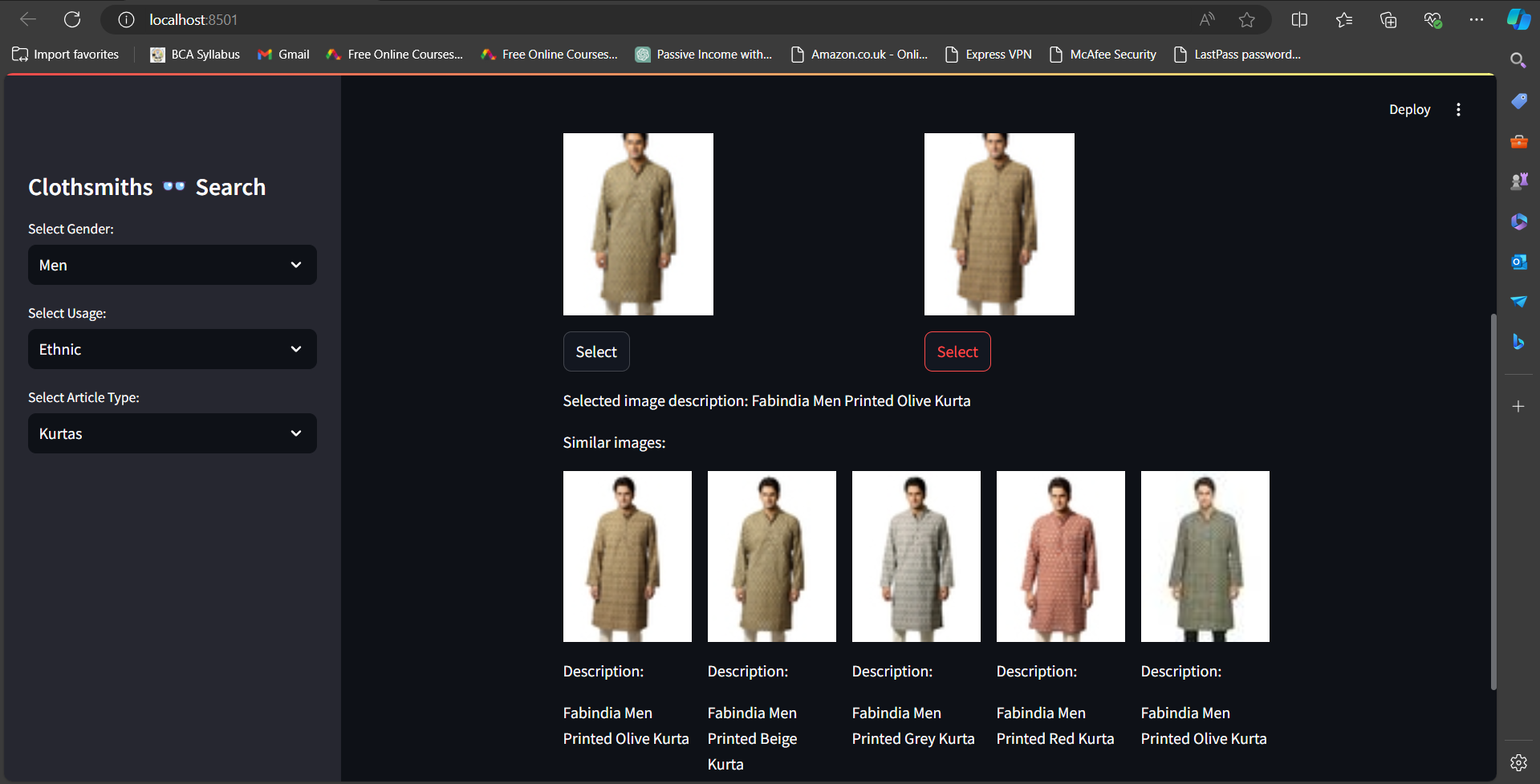
****

**** **RECOMMENDATION USING MODEL**

**Demonstration**

****

****

****

## Vision

In the context of outfit recommendation, the vision includes:  
1. Personalized Styling: Providing personalized outfit suggestions based on user preferences and past purchases.  
2. Seasonal Recommendations: Adapting recommendations to current fashion trends and seasons.  
3. Cross-Category Suggestions: Recommending complementary items across different categories (e.g., shoes with dresses).  
4. Real-Time Updates: Offering real-time outfit recommendations as new products are added to the inventory.  
5. Enhanced User Experience: Continuously improving the recommendation engine to enhance user satisfaction and engagement.

## Tools and Techniques

The project employs several tools and techniques, including:  
- Pandas: For data manipulation and analysis.  
- NLTK: For text preprocessing.  
- Scikit-learn: For TF-IDF vectorization and cosine similarity computation.

## ResNet-50 and TF-IDF

### ResNet50

### ResNet50 is a type of deep learning model designed to recognize images. It’s very deep (50 layers) and uses special shortcuts, called residual connections, to make it easier for the model to learn. ResNet50 is used to pull out important features from images without the final classification part, and then these features are turned into a simpler format using a GlobalMaxPooling layer.

### Nearest Neighbors

### Nearest Neighbors is a method for finding similar items based on their features. it helps find images that are most similar to a given query image. It does this by measuring the distance between feature vectors, with closer distances indicating higher similarity. So, if you upload an image, the Nearest Neighbors algorithm will find and list images that are closest in appearance to the one you uploaded.TF-IDF and Cosine Similarity

**TF-IDF (Term Frequency-Inverse Document Frequency)** is a statistical measure used to evaluate the importance of a word in a document relative to a corpus. It converts text data into numerical vectors, which are then used to compute cosine similarity. Cosine similarity measures the cosine of the angle between two vectors, indicating their similarity. This method is efficient for finding similar products based on text descriptions.

## Conclusion

The Cloth-Smith project successfully demonstrates the development of a recommendation system using NLP and machine learning techniques. By leveraging TF-IDF vectorization and cosine similarity, the system can provide relevant product recommendations, showcasing its potential application in the e-commerce industry.  
  
The project's methodology includes data preprocessing, vectorization using TF-IDF, and similarity computation using cosine similarity. The dataset comprises detailed product information, enabling comprehensive analysis and recommendation. Tools such as Pandas, NLTK, and Scikit-learn facilitate efficient data processing and model development.  
  
The vision for Cloth-Smith involves enhancing the recommendation engine to offer personalized styling, seasonal recommendations, cross-category suggestions, and real-time updates. These improvements aim to enhance user experience and engagement, making the shopping process more enjoyable and efficient.  
  
Incorporating advanced techniques like ResNet-50 for image feature extraction could further improve the recommendation accuracy by combining text and image data. This multi-modal approach would provide a richer context for recommendations, catering to diverse user preferences.  
  
Overall, Cloth-Smith demonstrates the significant impact of recommendation systems in the e-commerce industry. By providing personalized and relevant suggestions, businesses can enhance customer satisfaction, increase sales, and foster brand loyalty. The project's success highlights the importance of continuous innovation and adaptation to emerging technologies to maintain a competitive edge in the rapidly evolving digital marketplace.

## References

1. Vaseem Sir for project guidance.  
   2. ChatGPT, Claude.  
   3. YouTube.  
   4. GitHub.