CASE STUDY

HOME DEPOT

Bhakti Ukey

PROBLEM STATEMENT

Developing a content recommendation system that delivers highly relevant content (articles) by measuring the similarity between user search terms and content titles, ensuring precise and meaningful suggestions.

INFORMATION PROVIDED

Dataset

- **content-data.csv :** set of content titles and slug
- **Label.csv** : search term and article pairs (manually evaluated for relevancy)
- **Test.csv** : a set of search terms for which you are required to provide the content recommendations

Additional Info

- We ignore user's **past search behavior**

METHODOLOGY USED

Core Objective

- **Match user search terms** with **relevant content titles** effectively.
- System returns **highly relevant** articles and minimizes **irrelevant results**.

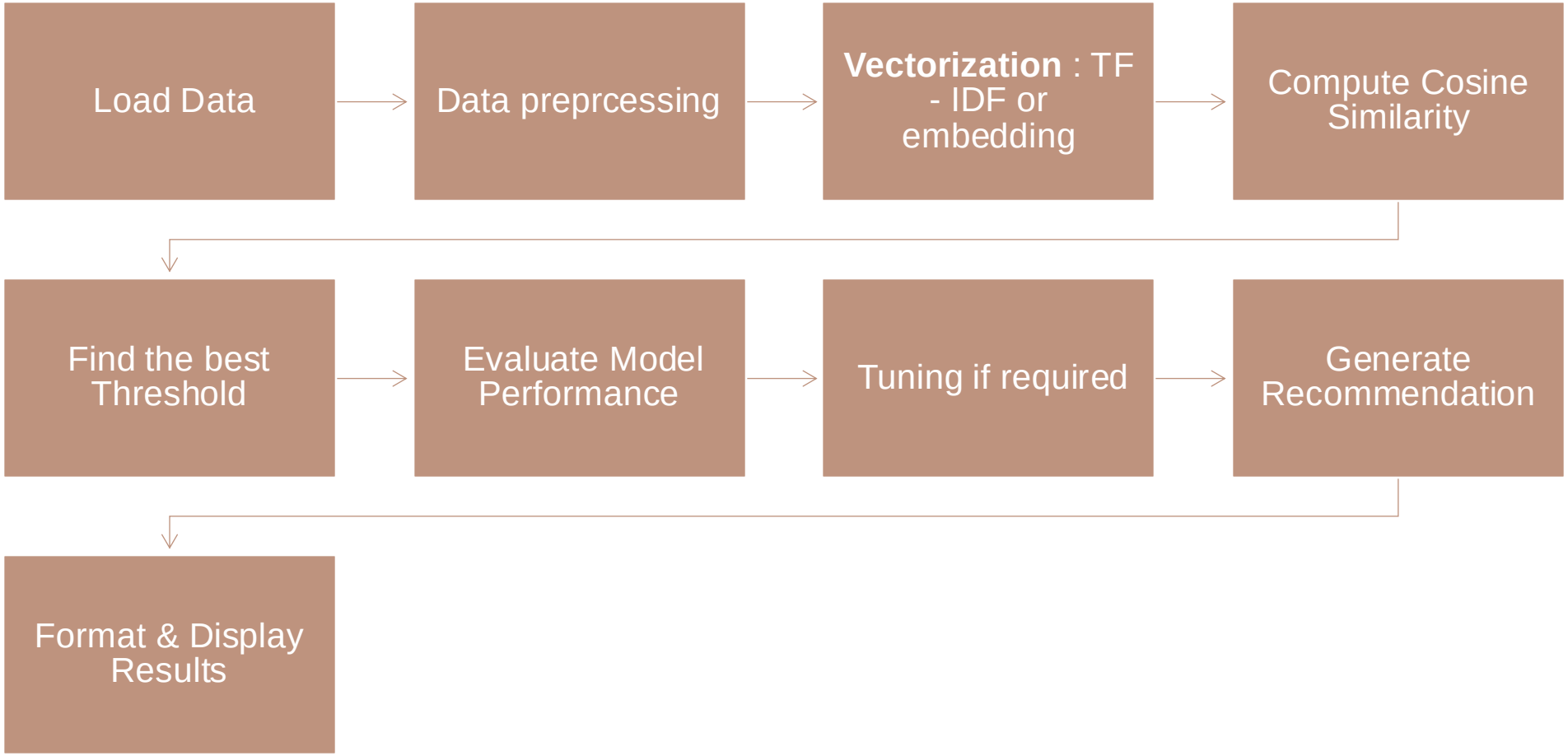
Key Components

- User search term
- Compute similarity between the search term and content titles.

Approach for Similarity Computation

- TF-IDF + Cosine Similarity
- Sentence Embeddings + Cosine Similarity
- Hybrid Approach : Combine both scores (e.g., **0.3 * TF-IDF + 0.7 * Embeddings Similarity**).
- Sentence Embeddings+ FAISS + Rag

FLOW


Load Data
Data preprcessing
Vectorization : TF - IDF or embedding
Compute Cosine Similarity
Find the best Threshold
Evaluate Model Performance
Tuning if required
Generate Recommendation
Format & Display Results

DETAILED FLOW

Data Cleaning

1.Missing values
2.Renamed columns
3. Merged content.csv and labels.csv

Data Preprocessing

1.Convert lowercase
2. Remove stop-words (spaces, punctuations)
3. Merged content.csv and

Vectorization

1.TF-IDF : converts text into numeric values - Used for keyword search
2. Sentence Transformer:
Uses transformer all-MiniLM-L6-v2 .
Semantic meaning vectors

COSINE SIMILARITY

Term	Search Query	Title 1	Title 2
how	0.3	0.2	0.0
install	0.4	0.5	0.0
kitchen	0.5	0.5	0.4

Title 1 Similarity: 0.92 ✅
(Highly Relevant)

How Cosine Similarity is used in Recommendation System

1. After text is converted into Vector Representation
2. Compute Cosine Similarity : Calculates similarity between user search from test and each content tile in labels data
3. Apply a Similarity Threshold : If **cosine similarity > threshold**, consider it **relevant**.

$$\cos(\theta) = \frac{A \cdot B}{||A|| \times ||B||}$$

THRESHOD SELECTION FOR CLASSIFICATION

Why a threshold?

- Classifies search -terms & article pairs as Relevant/ Not relevant
- Tested multiple threshold (0.0 - 1.0)
- Evaluated using F1 score for optimal performance

MODEL EVALUATION METRICS

TF-IDF

```
Best threshold: 0.3500000000000003 with macro F1: 0.6078363583981561
              precision    recall  f1-score   support

NOT RELEVANT       0.55      0.28      0.37        43
    RELEVANT       0.78      0.92      0.85       123

    accuracy                           0.75       166
   macro avg       0.67      0.60      0.61       166
weighted avg       0.72      0.75      0.72       166
```

Sentence Embedding

```
Best threshold: 0.4 with macro F1: 0.7350781998084903
              precision    recall  f1-score   support

NOT RELEVANT       0.58      0.65      0.62        43
    RELEVANT       0.87      0.84      0.85       123

    accuracy                           0.79       166
   macro avg       0.73      0.74      0.74       166
weighted avg       0.80      0.79      0.79       166
```

MODEL EVALUALTION METRICS

Hybrid- Model

```
Best threshold: 0.33 with macro F1: 0.7502165883908622
              precision    recall  f1-score   support

NOT RELEVANT       0.60      0.67      0.64        43
    RELEVANT       0.88      0.85      0.86       123

    accuracy                           0.80       166
   macro avg       0.74      0.76      0.75       166
weighted avg       0.81      0.80      0.80       166


Search Term: **34 in. to 36 in. x 72 in. shower door**
```

RECOMMENDATION TF-IDF

Search Term: **34 in. to 36 in. x 72 in. shower door**

Slug	Recommended Title
how-to-screen-in-a-porch	How to Screen in a Porch
how-to-care-for-roses-in-summer	How to Care for Roses in Summer
how-to-kill-weeds-in-your-lawn	How to Kill Weeds in Your Lawn
how-to-set-a-pattern-in-floor-tile	How to Set a Pattern in Floor Tile
how-to-repair-cracks-in-a-concrete-driveway	How to Repair Cracks in a Concrete Driveway

Search Term: **outdoor prelit christmas tree**

Slug	Recommended Title
how-to-decorate-a-christmas-tree	How to Decorate a Christmas Tree
traditional-nostalgic-christmas-tree	Traditional Nostalgic Christmas Tree
how-to-make-a-christmas-tree-card-holder	How to Make a Christmas Tree Card Holder
how-to-fix-christmas-tree-lights	How to Fix Christmas Lights
outdoor-decor-ideas	Outdoor Decor Ideas

Search Term: **respirators**

Result not found

Search Term: **1 gal #hdgb28 paint with primer**

Slug	Recommended Title
how-to-repair-cracks-in-a-concrete-driveway	How to Repair Cracks in a Concrete Driveway

Search Term: **outdoor prelit christmas tree**

Slug	Recommended Title
how-to-decorate-a-christmas-tree	How to Decorate a Christmas Tree
traditional-nostalgic-christmas-tree	Traditional Nostalgic Christmas Tree
how-to-make-a-christmas-tree-card-holder	How to Make a Christmas Tree Card Holder
how-to-fix-christmas-tree-lights	How to Fix Christmas Lights
outdoor-decor-ideas	Outdoor Decor Ideas

Search Term: **respirators**

Result not found

Search Term: **1 gal #hdgb28 paint with primer**

Slug	Recommended Title
how-to-paint-a-room	How to Paint a Room

Search Term: **hydrangea**

Result not found

RECOMMENDATION SENTENCE EMBEDDING

Search Term: **34 in. to 36 in. x 72 in. shower door**

Slug	Recommended Title
tips-for-selecting-shower-doors	Tips for Selecting Shower Doors
rug-sizes-for-your-space	Rug Sizes for Your Space

Search Term: **outdoor prelit christmas tree**

Slug	Recommended Title
traditional-nostalgic-christmas-tree	Traditional Nostalgic Christmas Tree
how-to-decorate-a-christmas-tree	How to Decorate a Christmas Tree
best-artificial-christmas-trees-for-the-season	Best Artificial Christmas Trees for the Season
how-to-make-a-christmas-tree-card-holder	How to Make a Christmas Tree Card Holder
how-to-decorate-for-christmas-with-lights	How to Decorate for Christmas with Lights

Search Term: **respirators**

Slug	Recommended Title
types-of-respirator-masks	Types of Respirator Masks
best-air-filters-for-your-home	Best Air Filters for Your Home
best-air-compressors-for-your-project	Best Air Compressors for Your Project

Search Term: **1 gal #hdgb28 paint with primer**

Slug	Recommended Title
how-to-prime-a-wall	How to Prime a Wall
how-to-choose-a-paint-color	How To Choose a Paint Color

Search Term: **hydrangea**

Result not found

RECOMMENDATION HYBRID MODEL

Search Term: **respirators**

Slug	Recommended Title
types-of-respirator-masks	Types of Respirator Masks
best-air-filters-for-your-home	Best Air Filters for Your Home
best-air-compressors-for-your-project	Best Air Compressors for Your Project

Search Term: **1 gal #hdgb28 paint with primer**

Slug	Recommended Title
how-to-paint-a-room	How to Paint a Room

Search Term: **hydrangea**

Result not found

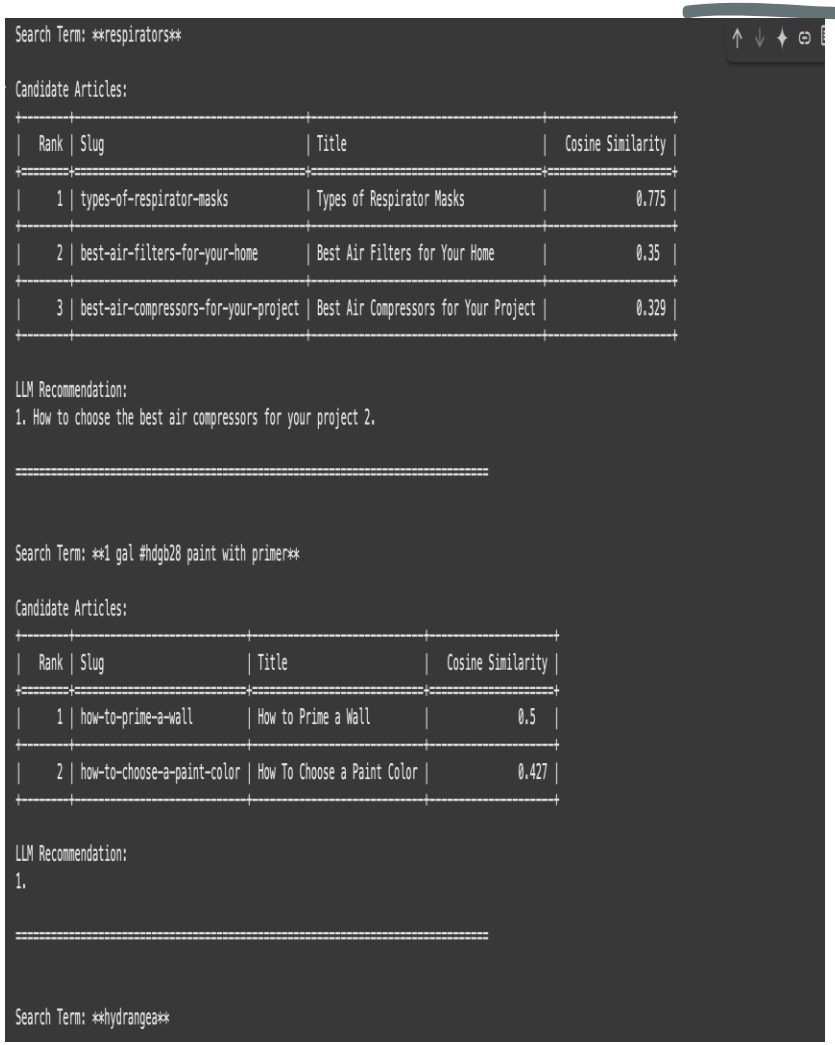
Search Term: **34 in. to 36 in. x 72 in. shower door**

Slug	Recommended Title
tips-for-selecting-shower-doors	Tips for Selecting Shower Doors
how-to-install-a-pivot-shower-door	How To Install a Pivot Shower Door

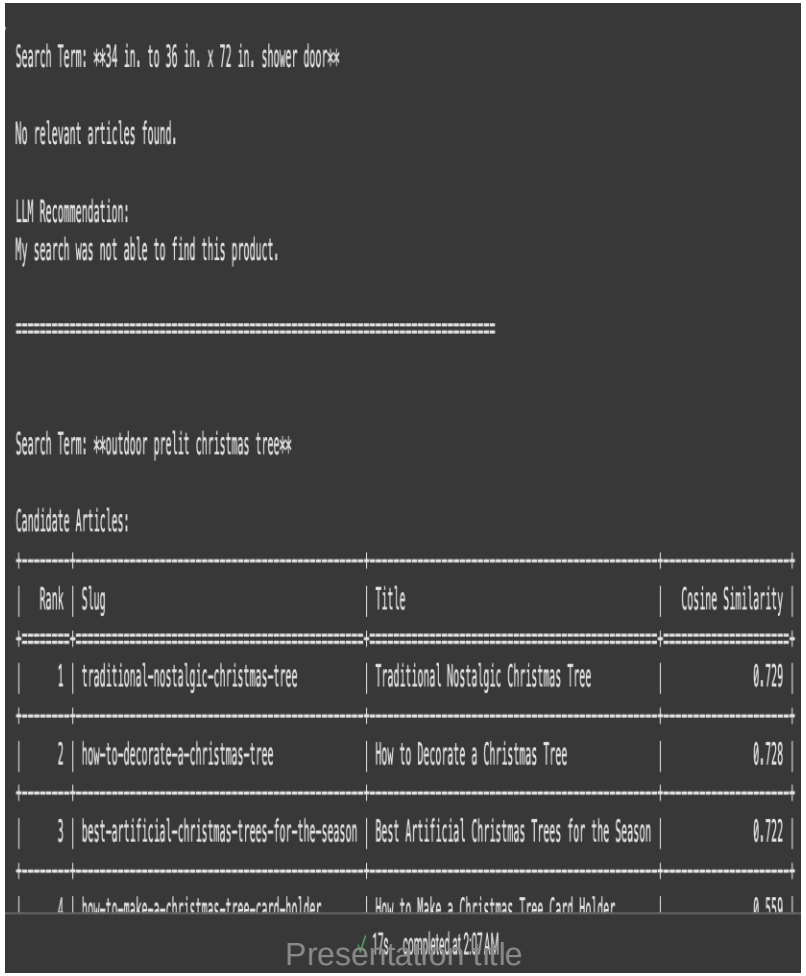
Search Term: **outdoor prelit christmas tree**

Slug	Recommended Title
how-to-decorate-a-christmas-tree	How to Decorate a Christmas Tree
traditional-nostalgic-christmas-tree	Traditional Nostalgic Christmas Tree
best-artificial-christmas-trees-for-the-season	Best Artificial Christmas Trees for the Season
how-to-make-a-christmas-tree-card-holder	How to Make a Christmas Tree Card Holder
how-to-decorate-for-christmas-with-lights	How to Decorate for Christmas with Lights

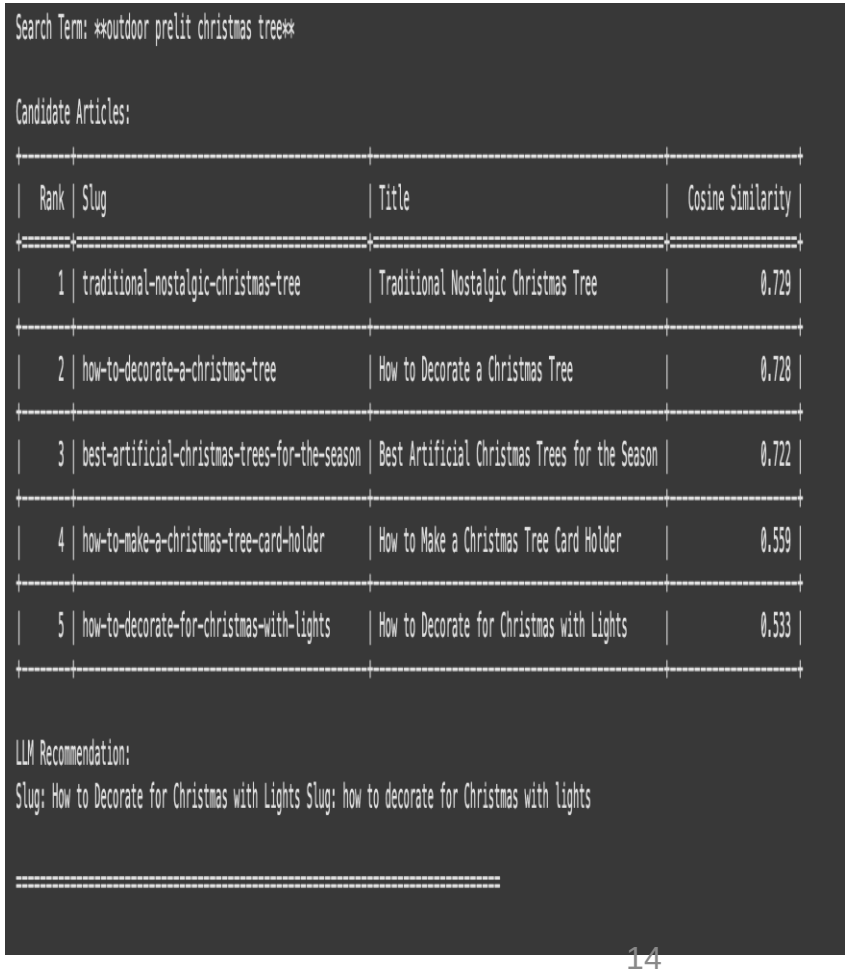
RECOMMENDATION FAISS+ RAG+ LLM



```
Search Term: **respirators**

Candidate Articles:
+--------+-----------------------------------+-----------------------------------+--------------------+
|   Rank | Slug                              | Title                             |   Cosine Similarity |
+========+===================================+===================================+====================+
|      1 | types-of-respirator-masks         | Types of Respirator Masks         |              0.775 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      2 | best-air-filters-for-your-home    | Best Air Filters for Your Home    |              0.35  |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      3 | best-air-compressors-for-your-project | Best Air Compressors for Your Project |          0.329 |
+--------+-----------------------------------+-----------------------------------+--------------------+

LLM Recommendation:
1. How to choose the best air compressors for your project 2.

================================================================


Search Term: **1 gal #hdgb28 paint with primer**

Candidate Articles:
+--------+-------------------------+-------------------------+--------------------+
|   Rank | Slug                    | Title                   |   Cosine Similarity |
+========+=========================+=========================+====================+
|      1 | how-to-prime-a-wall     | How to Prime a Wall     |              0.5   |
+--------+-------------------------+-------------------------+--------------------+
|      2 | how-to-choose-a-paint-color | How To Choose a Paint Color |          0.427 |
+--------+-------------------------+-------------------------+--------------------+

LLM Recommendation:
1.

================================================================


Search Term: **hydrangea**
```



```
Search Term: **34 in. to 36 in. x 72 in. shower door**

No relevant articles found.

LLM Recommendation:
My search was not able to find this product.

================================================================


Search Term: **outdoor prelit christmas tree**

Candidate Articles:
+--------+-----------------------------------+-----------------------------------+--------------------+
|   Rank | Slug                              | Title                             |   Cosine Similarity |
+========+===================================+===================================+====================+
|      1 | traditional-nostalgic-christmas-tree | Traditional Nostalgic Christmas Tree |          0.729 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      2 | how-to-decorate-a-christmas-tree  | How to Decorate a Christmas Tree  |              0.728 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      3 | best-artificial-christmas-trees-for-the-season | Best Artificial Christmas Trees for the Season |  0.722 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      4 | how-to-make-a-christmas-tree-card-holder | How to Make a Christmas Tree Card Holder |  0.559 |
+--------+-----------------------------------+-----------------------------------+--------------------+

✓ 17s  completed at 2:07 AM
```

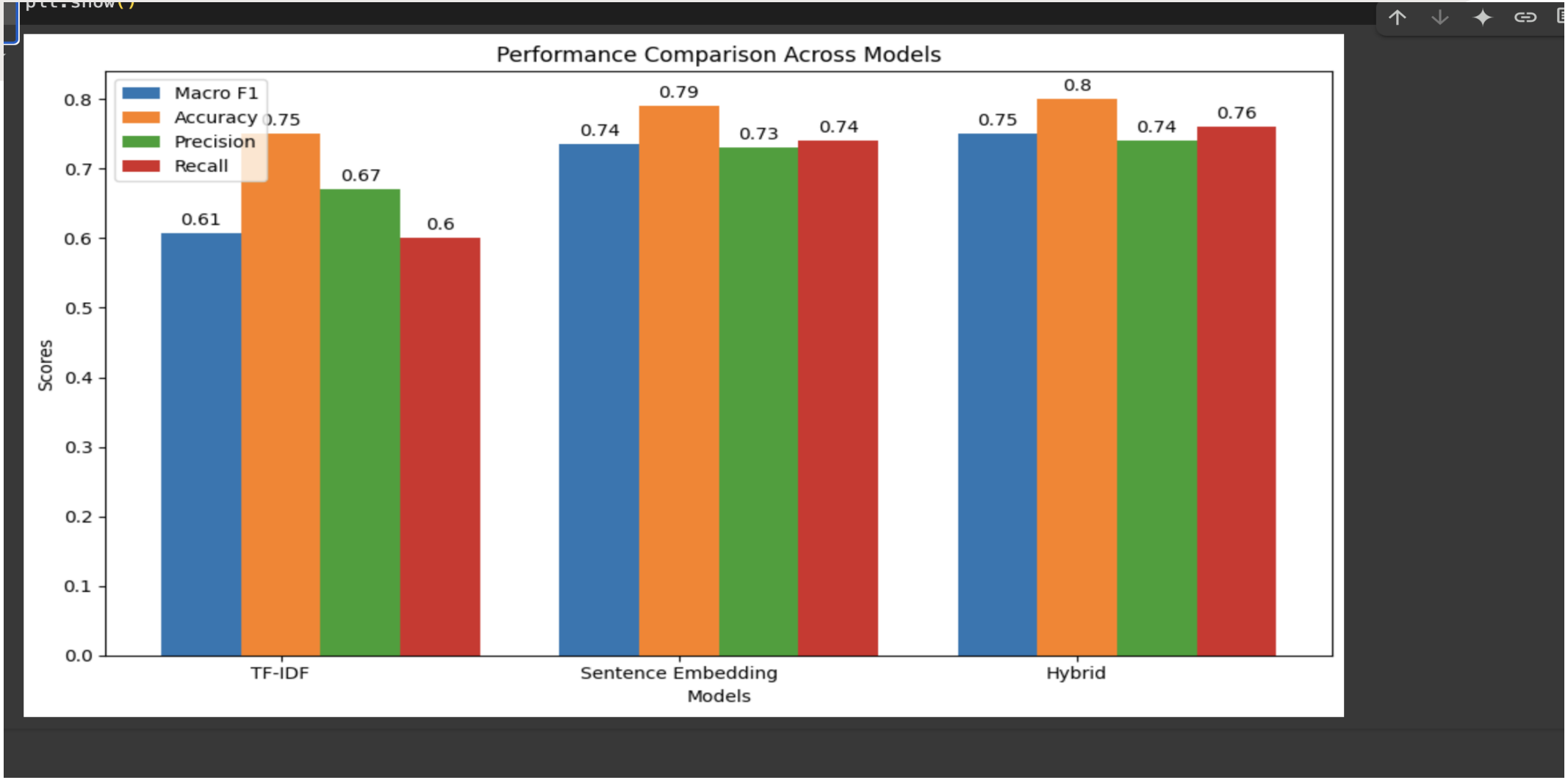


```
Search Term: **outdoor prelit christmas tree**

Candidate Articles:
+--------+-----------------------------------+-----------------------------------+--------------------+
|   Rank | Slug                              | Title                             |   Cosine Similarity |
+========+===================================+===================================+====================+
|      1 | traditional-nostalgic-christmas-tree | Traditional Nostalgic Christmas Tree |          0.729 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      2 | how-to-decorate-a-christmas-tree  | How to Decorate a Christmas Tree  |              0.728 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      3 | best-artificial-christmas-trees-for-the-season | Best Artificial Christmas Trees for the Season |  0.722 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      4 | how-to-make-a-christmas-tree-card-holder | How to Make a Christmas Tree Card Holder |  0.559 |
+--------+-----------------------------------+-----------------------------------+--------------------+
|      5 | how-to-decorate-for-christmas-with-lights | How to Decorate for Christmas with Lights |  0.533 |
+--------+-----------------------------------+-----------------------------------+--------------------+

LLM Recommendation:
Slug: How to Decorate for Christmas with Lights Slug: how to decorate for Christmas with lights

================================================================
```

COMPARISION

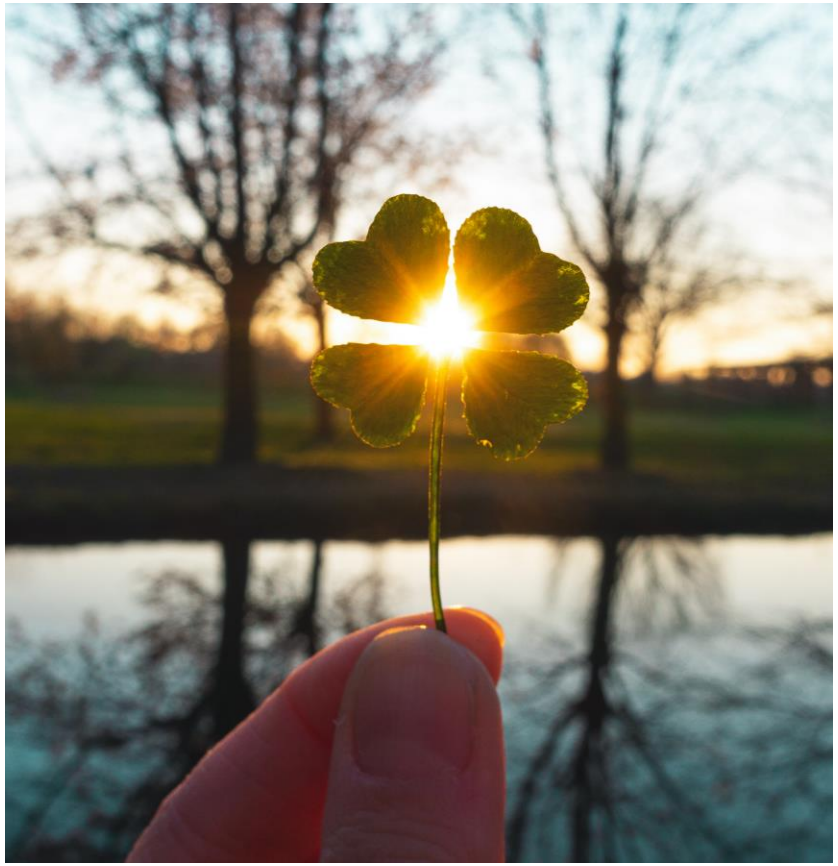
Feature	TF-IDF	Sentence Embedding	Hybrid (TF-IDF + Embedding)	FAISS + LLM
Speed	Fastest	Slower (due to model inference)	Moderate	Fast (FAISS Optimized)
Semantic Understanding	Limited	Stronger semantic similarity	Best	Best with LLM support
Threshold Optimization	Required	Required	Required	Not needed
Performance	Decent for exact matches	Better for paraphrased quieries	Best overall	Best with contextual reasoning

VISUAL


Performance Comparison Across Models
Macro F1
Accuracy
Precision
Recall
Scores
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.61
0.75
0.67
0.6
0.74
0.79
0.73
0.74
0.75
0.8
0.74
0.76
TF-IDF
Sentence Embedding
Hybrid
Models

KEY FINDINGS

- **Key Findings:**
  - TF-IDF is efficient but lacks deep semantic understanding.
  - Sentence Embeddings provide more context-aware recommendations.
  - Hybrid Model balances performance and accuracy.
  - FAISS-Based Retrieval + LLM achieves the best combination of speed, accuracy, and relevance.
- **Future Improvements:**
  - Experiment with more advanced transformer models (e.g., BERT, SBERT).
  - Implement dynamic weight tuning for hybrid models.
  - Enhance real-time search performance with optimized embeddings and FAISS clustering.
  - Improve Data volume



CHALLENGES

1. Duplicate Results

- **Issue:** The same article (identified by its slug) appears multiple times in the recommendations.
- **Solution:**
  - Normalize slugs (lowercase, remove extra spaces).
  - Use a set (seen_slugs) to track and filter out duplicates.

2. Weak Matches (Low Similarity Scores)

- **Issue:** Some recommendations have low similarity scores, leading to irrelevant or weak matches.
- **Solution:**
  - Apply a stricter similarity threshold (e.g., > 0.6).
  - Improve preprocessing and model tuning to enhance relevance.

3. Low Recall (No Results Found)

- **Issue:** Some search terms return no recommendations due to insufficient matches.
- **Solution:**
  - Increase top_k (number of recommendations retrieved).
  - Use query expansion techniques (e.g., synonyms) to broaden the search.

4. Dynamic Weight Tuning

- **Issue:** Hybrid models require manual tuning of weights for combining TF-IDF and embeddings.
- **Solution:**
  - Implement dynamic weight tuning based on dataset characteristics.
  - Experiment with different weighting schemes for optimal performance.

THANK YOU

Bhakti Ukey

Ukey.b@northeastern.edu