


# Fashion Recommendation



## Team Members:-

1. Parshv Patel
  2. Varun Patwari
  3. Ankit Dhokariya
- 

# Introduction



- Recommendation engines are the core of the businesses that are customer focused.
- They are responsible for bringing in a major chunk of revenue to the company.
- Amazon's 35% revenue is from product recommendations. So we believe by implementing Fashion Recommendation, we can learn more about the in-demand use case of recommendation engines.
- The recommendation engine we build in this project is content based recommendation as collaborative filtering needs item/user information which might be tough to find because of privacy concerns.

# Introduction

## Product



Brand: DULUTEE

### DULUTEE Women's T Shirts V Neck Tops Shorts Sleeve Tops Casual Tshirts

★★★★☆ 4,272 ratings | 15 answered questions

6 Price Changes

Was: \$14.89 Details

Price: **\$12.89** ✓prime & FREE Returns

You Save: **\$2.00** (13%)

Fit: True to size. Order usual size. ▾

Size:

Large ▾

Color: Red1



Size Chart ▾

• 80% Cotton, 20% Polyester

**\$12.89**

✓prime & FREE Returns ▾

FREE delivery: **Saturday, Oct 23**

Order within 4 hrs and 52 mins

[Details](#)

📍 Deliver to Swaroop - San Jose  
95113

**In Stock.**

Qty: 1 ▾

Add to Cart

Buy Now

🔒 Secure transaction

Ships from Amazon

Sold by DULUTEE

Packaging Shows what's inside. T...

[Details](#)

Return policy: Extended holiday

return window till Jan 31, 2022

▾

1-Click ordering is not available for  
this item.

## Recommendations

### Products related to this item

Sponsored



SAMPEEL Women T-Shirts Short Sleeve V Neck Tunic Tops Loose Fitting Clothing Burgun...  
★★★★☆ 15,611  
**\$21.99** ✓prime



Amoretu Womens V Neck Short Sleeve Summer T Shirts Curved Hem Tee Tops(Yellow,S)  
★★★★☆ 15,356  
**\$18.99** ✓prime



Biucly Women's Summer Casual Sexy Short Sleeve Crewneck Color Block Tee Loose Tops ...  
★★★★☆ 7,510  
**\$22.99** ✓prime



Hount Tees for Women Summer Short Sleeve Casual Tops Round Neck Loose T Shirts Coral S  
★★★★☆ 84  
**\$18.99** ✓prime



Buyears Women Summer T Shirts Short Sleeve Rounded V Neck Pocket Tee Tops Pink X-Large  
★★★★☆ 113  
**\$19.99** ✓prime



YunJey short sleeve round neck triple color block stripe T-shirt casual blouse.Red,XX-  
Large  
★★★★☆ 42,464  
**\$16.99** ✓prime



imesrun Womens V Neck Tshirts Short Sleeve Loose Casual Summer Tops with Pokcet...  
★★★★☆ 261  
**\$18.98** ✓prime




Page 2 of 305 | [Start over](#)

# System Design & Implementation Details



- We have built 2 types of recommendation engines:
  - Text based recommendation engines
    - BoW based
    - TF-IDF based
    - IDF based
    - Average Word2Vec (Text Semantics) based
    - IDF weighted Word2Vec (Text Semantics) based
    - Weighted similarity using brand and color
  - Visual features based
- We have built these 2 types of recommendation engines because in some cases text based engines might give better results (for books) and in some cases visual features based engines might give better results (for apparel).

- 
- We have considered the basic techniques of text pre-processing for Text based recommendation engines with the exception of Word2Vec, we did this because our main goal was to build a robust recommendation engine and not complicate things by applying more advanced techniques.
  - We have used libraries like sklearn and gensim.models for text preprocessing and Word2Vec.
  - We have also used keras library to extract features from the product images.



# Dataset

- The dataset consists of a total 183,138 number of records of different products. Each record is defined using 19 features, some of which are ASIN which stands for Amazon Standard Identification Number, the Brand of the product, the Type of the product, Availability of the product, Reviews given by users for the products, 3 different sizes of the product image and Manufacturer of the product and many more.
- Of these 19 features, we used only 7 features:
  1. ASIN ( Amazon standard identification number)
  2. Brand ( brand to which the product belongs)
  3. Color ( Color information of apparel, it can contain many colors as a value ex: red and black stripes )
  4. product\_type\_name (type of the apparel, ex: SHIRT/T-SHIRT)
  5. medium\_image\_url ( URL of the image )
  6. title (title of the product.)
  7. formatted\_price (the price of the product)

# Data Preprocessing

- We have 2325 products that have the same title but different colors. These shirts are exactly the same except that they are different in size.

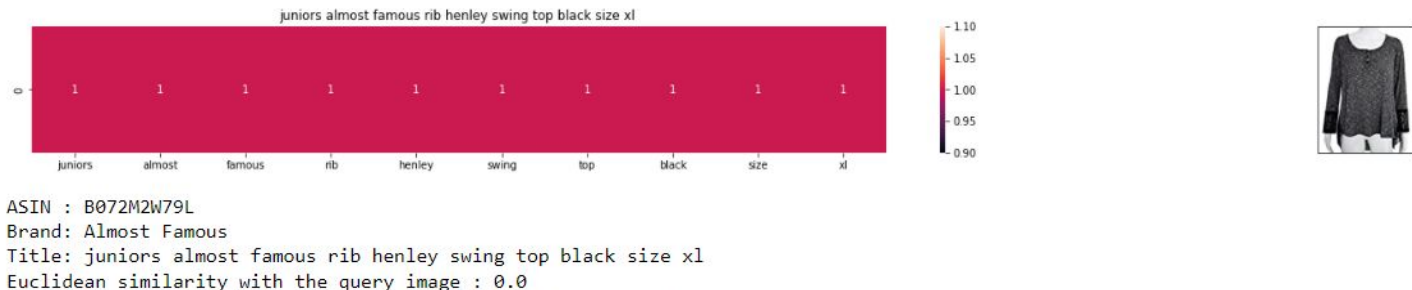


- These shirts are exactly the same except in color.
- We removed the duplicates from the dataset.

# Algorithms

## 1. Bag of words on product titles

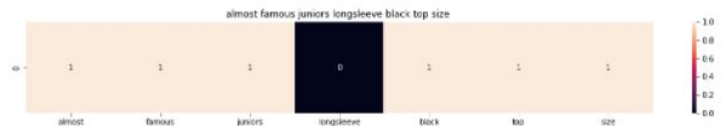
- We separated words from all the product titles and formed vectors of each title. We used these vectors to find similarity between other product title vectors and recommended the ones with the most similar titles.



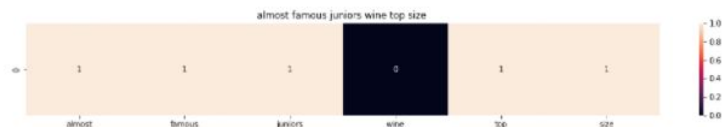




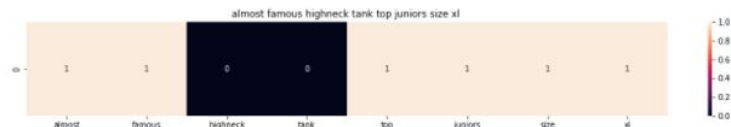
ASIN : B01MAZCL6C  
 Brand: Almost Famous  
 Title: almost famous swing henley top size l  
 Euclidean similarity with the query image : 2.0



ASIN : B06XRPPVQC  
 Brand: Almost Famous  
 Title: almost famous juniors long sleeve black top size  
 Euclidean similarity with the query image : 2.23606797749979



ASIN : B01M3RLP9B  
 Brand: Almost Famous  
 Title: almost famous juniors wine top size  
 Euclidean similarity with the query image : 2.449489742783178

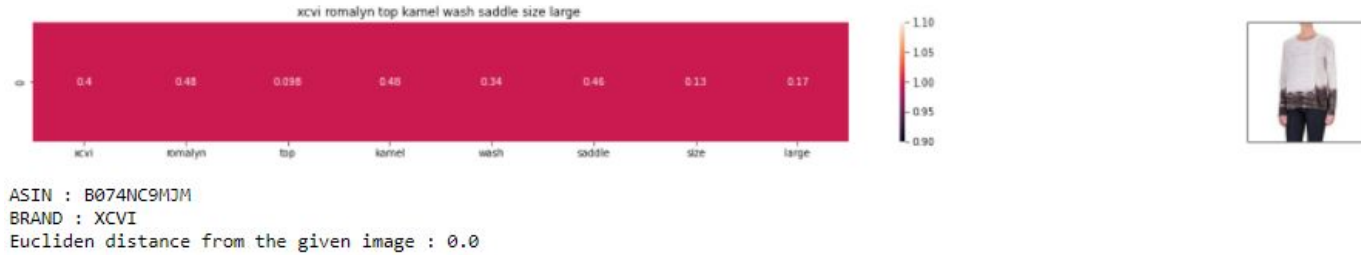


ASIN : B01MR9VMSE  
 Brand: Almost Famous  
 Title: almost famous highneck tank top juniors size xl  
 Euclidean similarity with the query image : 2.449489742783178

Fig. Recommended Products

## 2. TF-IDF based product similarity

- Term frequency times Inverse document frequency is used to scale down the impact of words that occur more frequently and do not contribute in distinguishing a title.



*Fig. Input for TF-IDF Model*



ASIN : B071VY48VC

BRAND : XCVI

Eucliden distance from the given image : 1.15704694867847



ASIN : B01N7V7SIG

BRAND : XCVI

Eucliden distance from the given image : 1.1718496553150264



ASIN : B01M31Q4Z0

BRAND : XCVI

Eucliden distance from the given image : 1.2098060863159672



ASIN : B01M7U5BHH

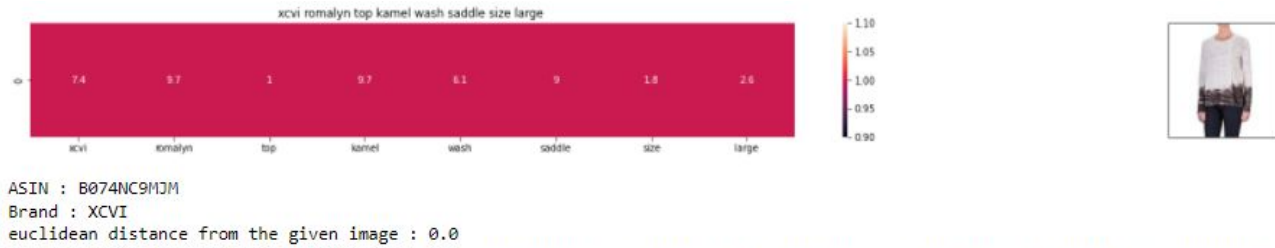
BRAND : XCVI

Eucliden distance from the given image : 1.22960337478935

Fig. Recommended Products

### 3. IDF based product similarity

- Inverse document frequency is used to find similarity between the vectors to recommend products.



*Fig. Input for IDF Model*

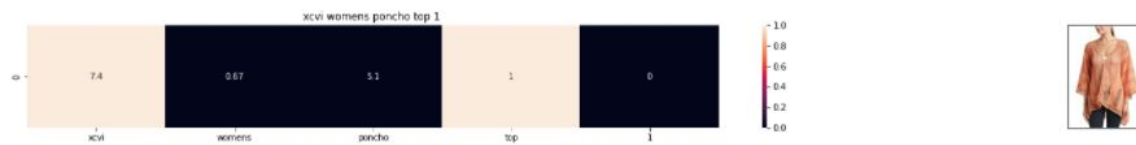


Fig. Recommended Products

## 4. Average Word2Vec product similarity

- Used Word2Vec to generate a vector for the product titles.
- Used these vectors to recommend similar products.

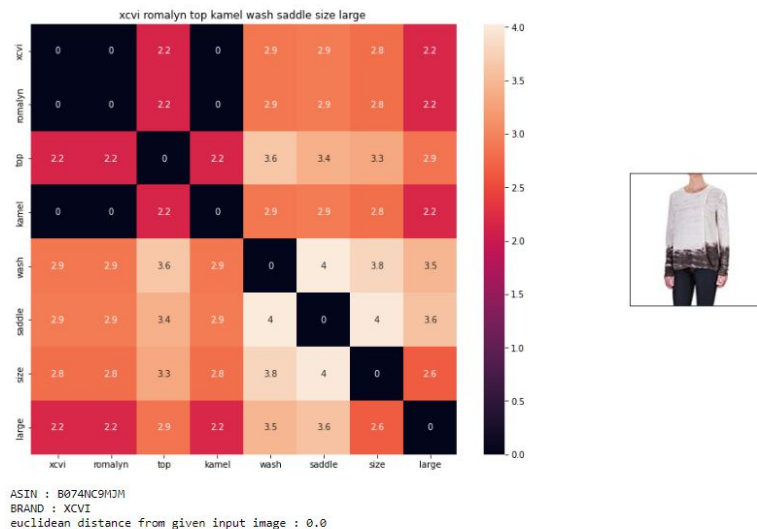
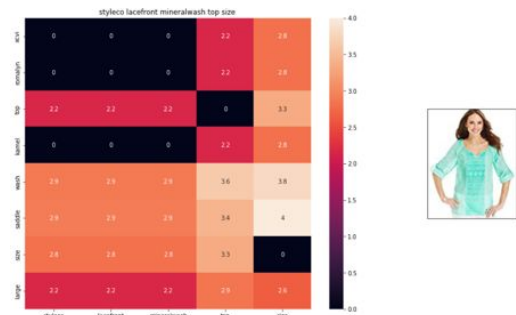
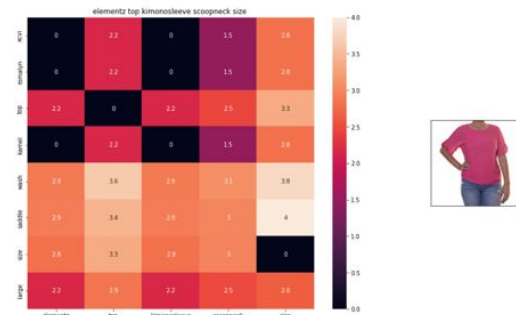


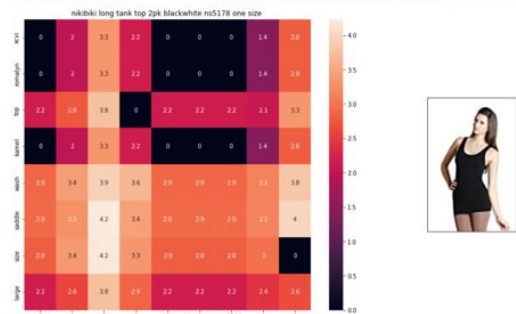
Fig. Input for Average Word2Vec Model



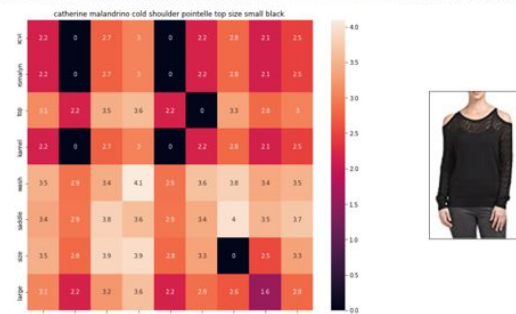
ASIN : B001EV662Q  
 BRAND : hotel Collections  
 euclidean distance from given input image : 0.5927989



ASIN : B014W6I22H  
 BRAND : Elementz  
 euclidean distance from given input image : 0.6163741



ASIN : B00F2905DC  
 BRAND : NIKIBIKI  
 euclidean distance from given input image : 0.6557259



ASIN : B0739280KL  
 BRAND : Catherine Malandrino  
 euclidean distance from given input image : 0.7018532

Fig. Recommended Products

## 5. IDF Weighted Word2Vec for product similarity

- Used Word2Vec to generate a vector for the product titles.
- While generating vector from Word2Vec multiply the vector provided by Word2Vec with the inverse document frequency of the title.
- Used these vectors to recommend similar products.

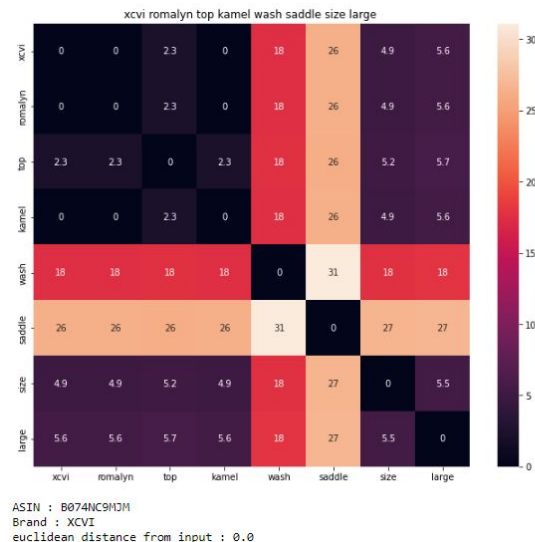
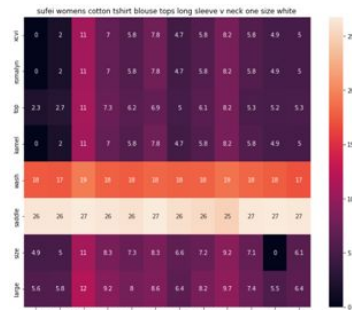
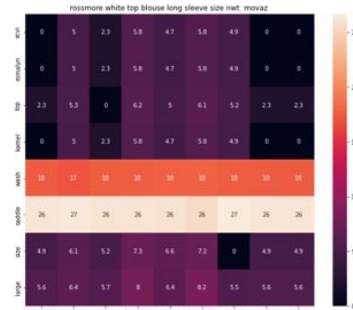


Fig. Input for IDF Weighted Word2Vec Model

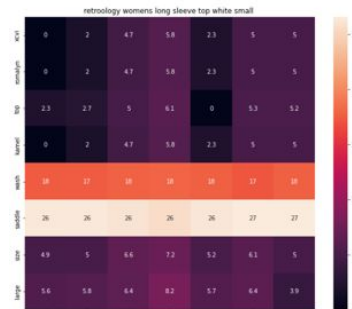




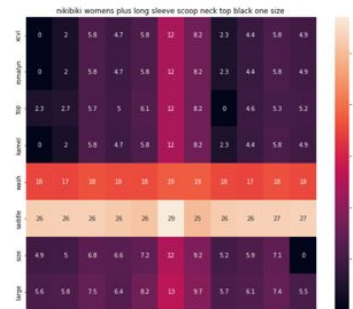
ASIN : B0746QDQ2  
Brand : suifei  
euclidean distance from input : 3.858672



ASIN : B0K12ETD4  
Brand : Rossmore  
euclidean distance from input : 3.862223



ASIN : B06XN2H956  
Brand : Retro-logy  
euclidean distance from input : 3.868939



ASIN : B0151XWGLV  
Brand : Nikibiki  
euclidean distance from input : 3.9017062

Fig. Recommended Products

## 6. Weighted similarity using brand and color

- Made use of brand and color feature to recommend products.
- Used weighted importance of brand and color while calculating similarity of two products based on the title.



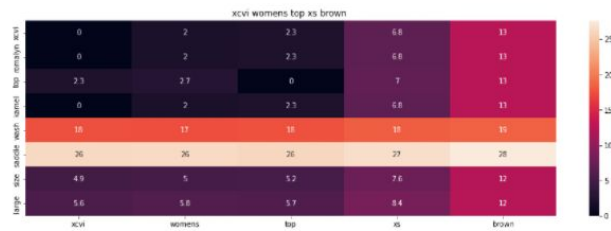
ASIN : B074NC9M1M

Brand : XCVI

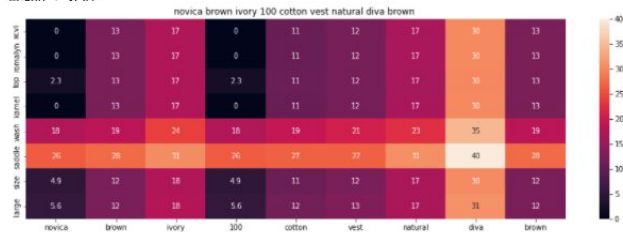
euclidean distance from input : 0.0



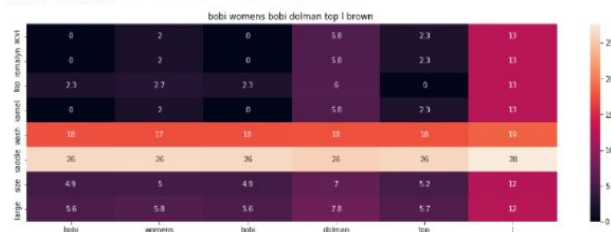
*Fig. Input for Weighted Similarity using brand and color*



ASIN : B01N7V7SIG  
Brand : VYUT



ASIN : B06Y41Y5TC  
Brand : NOVICA



ASIN : B06ZZX46Z6  
Brand : Bobi  
euclidean distance from input : 1.9460667936128642

Fig. Recommended Products

# Visual features based product similarity

- Used images to recommend similar products.



Product Title: xcvi romalyn top kamel wash saddle size large  
Euclidean Distance from input image: 6.180516e-06  
Amazon Url: [www.amazon.com/dp/B074NC9M3M](http://www.amazon.com/dp/B074NC9M3M)



Product Title: cable gauge large dark blush peasant top  
Euclidean Distance from input image: 35.34564  
Amazon Url: [www.amazon.com/dp/B074XPJM7T](http://www.amazon.com/dp/B074XPJM7T)



Product Title: style co womens metallic scoopneck pullover top gray  
Euclidean Distance from input image: 36.92667  
Amazon Url: [www.amazon.com/dp/B01N5K6P4E](http://www.amazon.com/dp/B01N5K6P4E)



Product Title: ideology raglan spaceddyed longsleeve top size xs  
Euclidean Distance from input image: 37.799213  
Amazon Url: [www.amazon.com/dp/B01MFDNJ28](http://www.amazon.com/dp/B01MFDNJ28)



# Analysis of Results

- We used A/B testing to find out which recommendation engine gave the best results.
- We took 10 test subjects (consisting of friends and family) and we divided them into 2 groups, namely, Group A and Group B.
- Group A was shown the results from the “Text based recommendation engines” and Group B was shown the results from the “Visual based recommendation engines”.
- Since we can measure the “goodness” through user (test subjects feedback), we took the feedback from 2 groups.

Group A (shown the results from Text based recommendation engines)



| Person   | Feedback                                     |
|----------|--|
| Person 1 | “Not so relevant results”                    |
| Person 2 | “Decent, not so great”                       |
| Person 4 | “I searched for shirt, recommended tank top” |
| Person 7 | “I liked the recommendations”                |
| Person 8 | “Somewhat relevant”                          |

Group B (shown the results from Visual based recommendation engines)

| Person    | Feedback  |
|-----------|---|
| Person 3  | “brilliant”   |
| Person 5  | “Relevant to my product”  |
| Person 6  | “This is quite relevant and helped me to understand my options” |
| Person 9  | “Its okay”  |
| Person 10 | “I like the fact that my recommendations were relevant”         |



# Conclusion

- Conclusion was that the **Visual based recommendation engines** performed better than the **Text based recommendation engines** and it makes sense because in case of apparel the “Visual” part is the core of recommendations as its the same way a human would recommend an apparel.
- In case of books, we feel **Text based recommendation engines** will perform better because there is no “Visual” part required to recommend books, book titles, genres would be the core of book recommendation as its the same way a human would recommend a book.
- Our engines are built on simple yet very effective techniques which give very good recommendations.

## Decisions made/Things that worked:

- We decided to make multiple recommendation engines so that our outputs (recommendations) are more robust and this helped us a lot.

## Difficulties faced/Things that didn't work well

- We tried stemming and it gave us really bad recommendations, our test subjects did not find the recommendations relevant.
- The data for the recommendation was taken from an archived page of amazon since the Amazon API was giving technical issues.





# Link to Dataset and Code

- **Dataset Link**

<https://drive.google.com/file/d/1w3j8M4Oh7Ogt3sPNUmSwlmG4umYkXSxo/view?usp=sharing>

- **Code Link**

<https://github.com/ankit-dhokariya/CMPE-256-Recommender-System-Project>