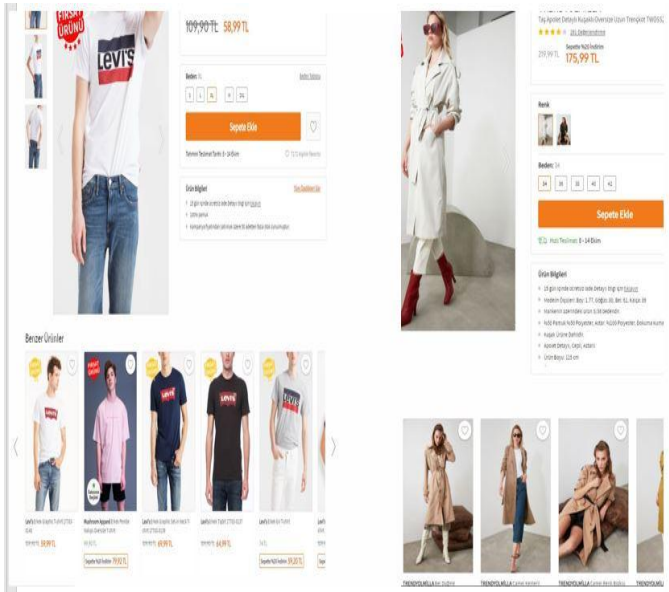
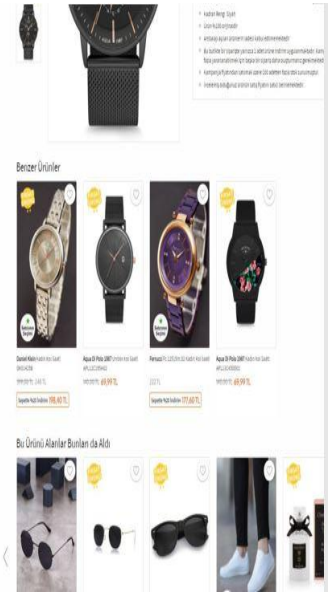
Trendyol Recommendation Engine

A recommendation system based on users former purchases and product similarity

Up School Capstone Project

Recommendation Examples





- “Similar products”
- “Customers who bought this also bought these”
- “Because you watch this”
- “Recommended videos”
- “Mostly sold this month”

Applied methods to project

- Category Based Ranking System
- Content Based Filtering System
- Collaborative Filtering System

Category Based Ranking System

```
df.tail(10)
```

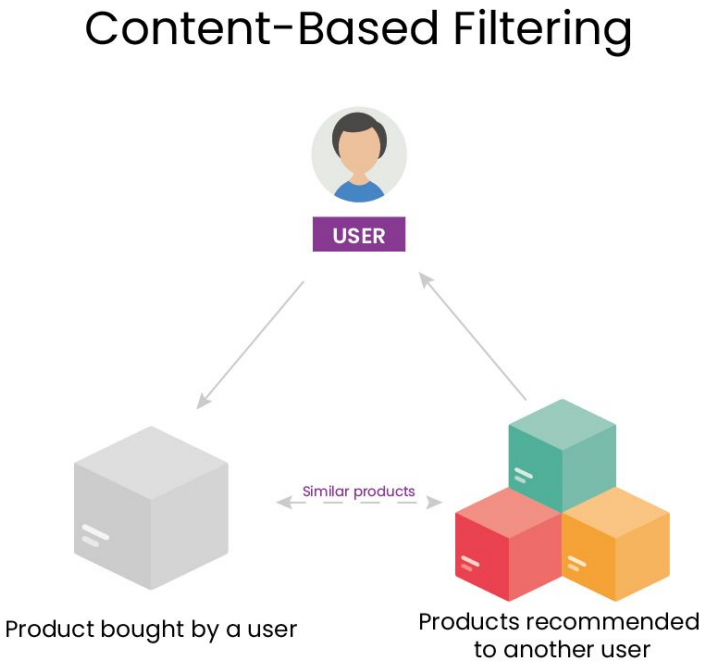
	partition_date	orderparentid	user_id	productcontentid	brand_id	category_id	category_name	gender	price	color_id	business_unit
508218	1.08.2020 19:00	325912035	19547	31076500	953433	934	Avize	Unisex	259.00	4.0	Aydınlatma
508219	12.08.2020 08:00	330211854	69402	32919087	13025	525	Eşofman Altı	Kadın	49.99	14.0	Kadın B
508220	28.07.2020 00:00	324345089	55671	3052680	5085	2247	Tek Kişilik Nevresim	Unisex	79.90	12.0	Ev Tekstil
508221	7.08.2020 13:00	328355157	87535	1762168	15663	3452	Fırça	Unisex	52.75	NaN	Kırtasiye & Ofis
508222	2.08.2020 10:00	326029540	59112	36420255	835853	1516	Yaka Kartı & Aksesuarları	Unisex	29.00	NaN	Kırtasiye & Ofis
508223	5.08.2020 14:00	327329628	82960	40176648	8820	419	Terlik	Kadın	49.99	3.0	Branded Shoes B

Steps Followed

- Chose ‘category_name’ and ‘business_unit’ as parameters
- Estimated weights separately
- Merged weight values with main dataframe
- Provided single weight
- Showed the results into webapp(streamlit)

	user_id	category_name	business_unit	weight
32980	56062	T-Shirt	PL Woman	0.075117
21648	76223	T-Shirt	PL Woman	0.075117
21609	73007	T-Shirt	PL Woman	0.075117
21631	2289	T-Shirt	PL Woman	0.075117
21633	54092	T-Shirt	PL Woman	0.075117
21657	16169	T-Shirt	PL Woman	0.075117
21646	104006	T-Shirt	PL Woman	0.075117
21656	42060	T-Shirt	PL Woman	0.075117
21655	109762	T-Shirt	PL Woman	0.075117
21654	94171	T-Shirt	PL Woman	0.075117

Content-Based Filtering System



Steps Followed

- Created a single column with all the words related to each user
- Applied stopwords to avoid miscalculations
- Applied countvectorizer
- Fit the model into cosine similarity
- Showed the results into web app with images

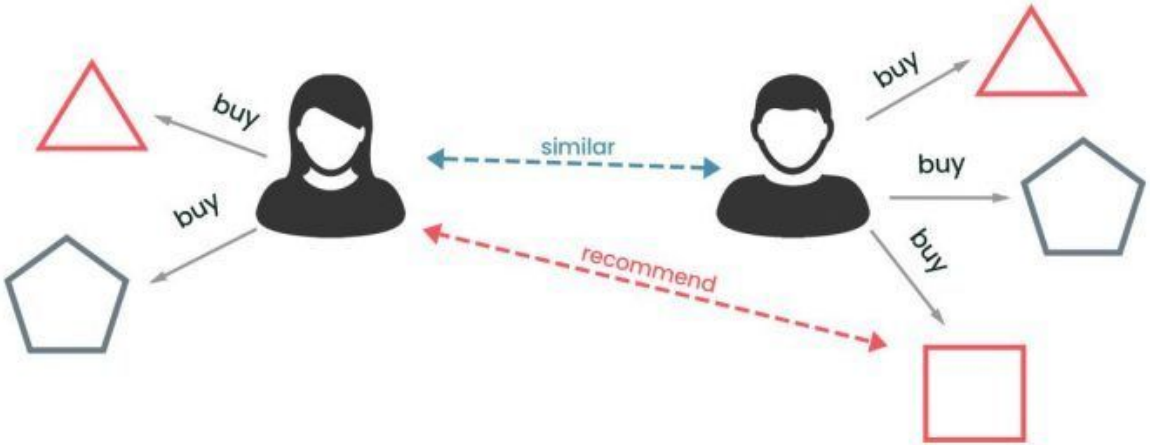
Most Seen Words Based on WordCloud

Cosine Similarity Matrix

```
cosine_sim = cosine_similarity(count_matrix, count_matrix)
print(cosine_sim)
```

```
[[1.   0.75 0.75 ... 0.25 0.25 0.25]
 [0.75 1.   0.75 ... 0.25 0.25 0.25]
 [0.75 0.75 1.   ... 0.25 0.25 0.25]
 ...
 [0.25 0.25 0.25 ... 1.   0.5  0.5 ]
 [0.25 0.25 0.25 ... 0.5  1.   0.5 ]
 [0.25 0.25 0.25 ... 0.5  0.5  1.  ]]
```

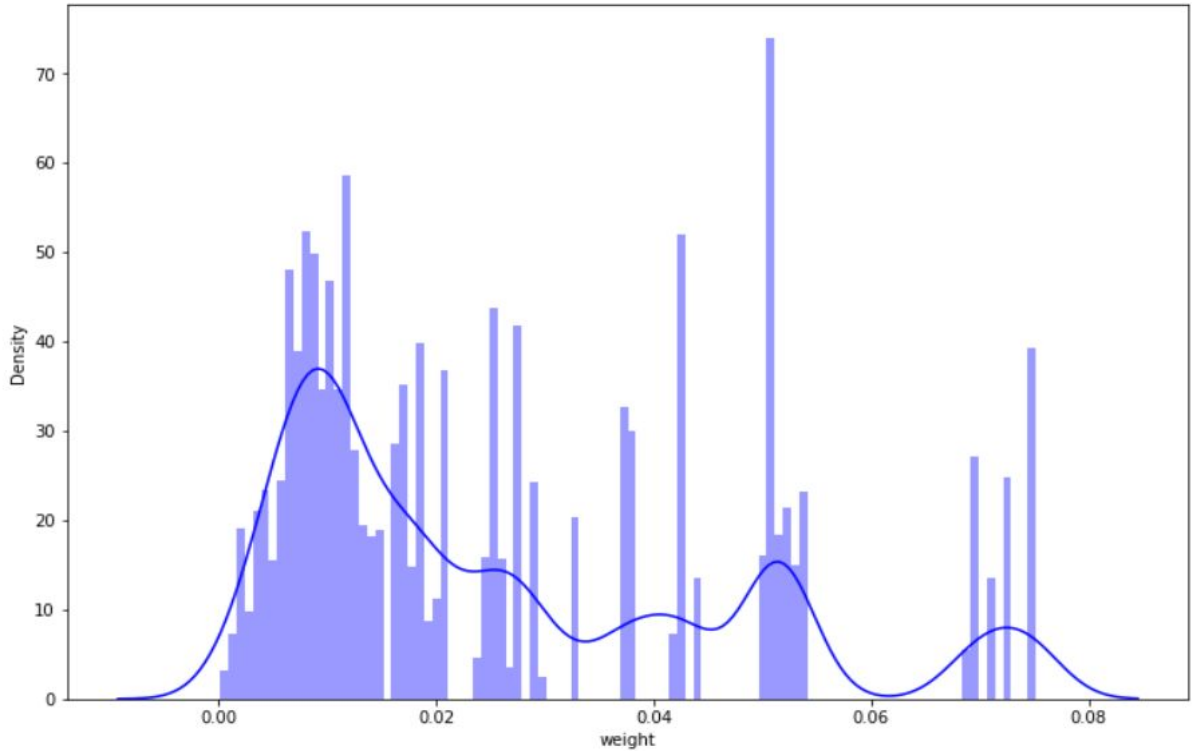
Collaborative Filtering System



Steps Followed

- Created a sample dataframe with size of 12500
- Chose surprise library from python to apply predictions
- Benchmarked algorithms within the library
- Chose KNNBaseline() as algorithm
- Split train-test data from sample
- Had predictions with a function
- Showed the results into web app with images

Weights Distribution in the Sample Data



Benchmark for to Choose Best Algorithm

Algorithm	test_rmse	fit_time	test_time
KNNBaseline	0.019721	2.415495	0.016818
BaselineOnly	0.019734	0.053152	0.014345
KNNBasic	0.020618	2.202554	0.019631
KNNWithMeans	0.020763	2.281601	0.019117
SlopeOne	0.020788	2.191420	0.018591
KNNWithZScore	0.020805	2.591727	0.020052
CoClustering	0.021605	1.653928	0.013235
NMF	0.021649	2.215785	0.018170
SVDpp	0.022675	1.234359	0.016984
SVD	0.025402	0.673845	0.033508
NormalPredictor	0.027922	0.016192	0.045313

Possibilities to Improve

Estimating weights for each customer for collaborative filtering

Improving cosine similarity estimation with clustering methods

Trying hybrid recommendation models

Adding more information about customer interactions

Final Notes and my experience

A bunch of different models were chosen to create a better recommendation. Not all results worked properly but it gave me an exposure to a great variety of tools. With this start, I am planning on having more success in the future. My own experience as a consumer and exposure to these recommendation systems were actually the original reason that I had an interest in machine learning tools.

Human Skills

- Improve together
- Genuine appreciation
- Learning how to learn

FEYZA JOHNSON

https://www.linkedin.com/in/feyzapamukoglujohnson/

Thank You!