Product Rating Prediction

Rami Fisher

Agenda

- Amazon reviews dataset

- Recommendation Systems (Challenges)

- Data Preprocessing

- EDA

- Model Selection

- Evaluation Metrics

- Model Comparison

- Application Demo

Amazon Reviews Dataset

	userName	verified	itemName	description	image	brand	feature	category	price	rating	reviewTime	summary	reviewText	vote
0	Amazon Customer	True	Toblerone Swiss Milk Chocolate Bar, Crunchy Sa...	["Made from deliciously decadent ingredients, ...	['https://images-na.ssl-images-amazon.com/imag...	Toblerone	[]	Prime_Pantry	$1.63	5.0	2018-01-01	Five Stars	super smooth and yummy with crunchy bits	0

~550K reviews
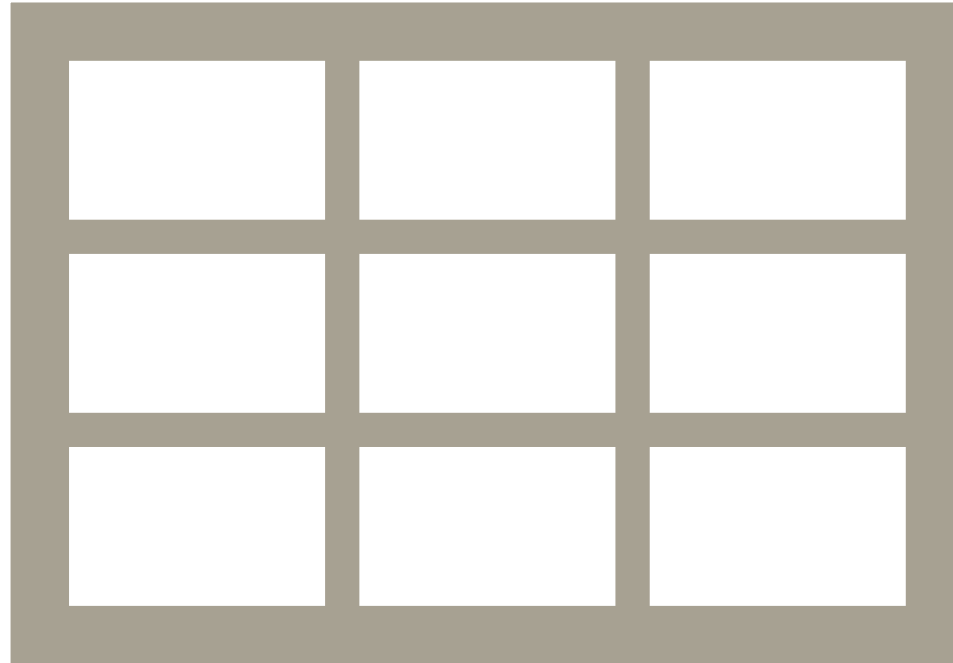
~130K unique users

~100K unique items

Recommendation Systems

Recommend items to users based on their preferences, behaviors, and patterns of similar users or items.

Challenges
- Memory limitations
- "Cold Start"
- Data Sparsity
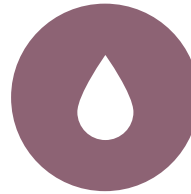- Exploration Vs Exploitation

Data Preprocessing

Filtering 95% Verified Users

Convert and Imputing 16% of Price values

Clean Textual Features

Clean nan samples less than 1%

Sampling – Most frequent items and users.

Generate users and items IDs

Dataset Preprocessed

	userName	itemName	brand	category	price	rating	vote	user_id	item_id
0	Carla	Prince Peace Organic Tea Oolong 100 Tea Bags	Prince Peace	Grocery Gourmet Food	6.4	5.0	0	2199	7943

~150K reviews   **~17K** unique users   **~11K** unique items

Distribution of category
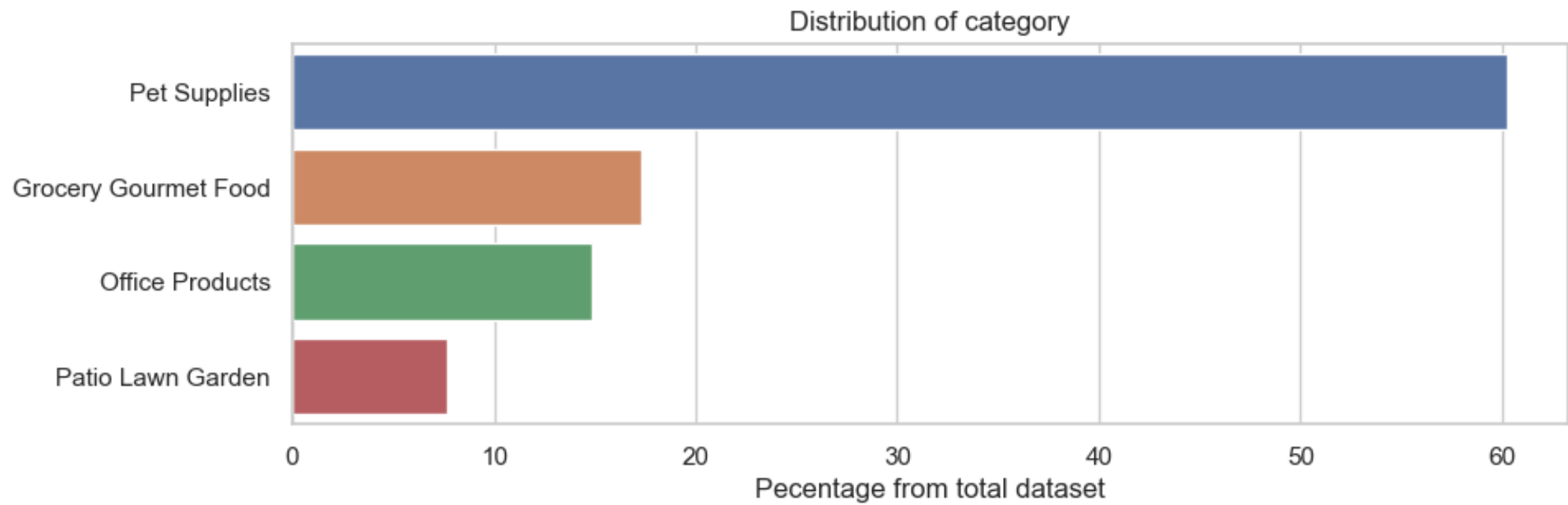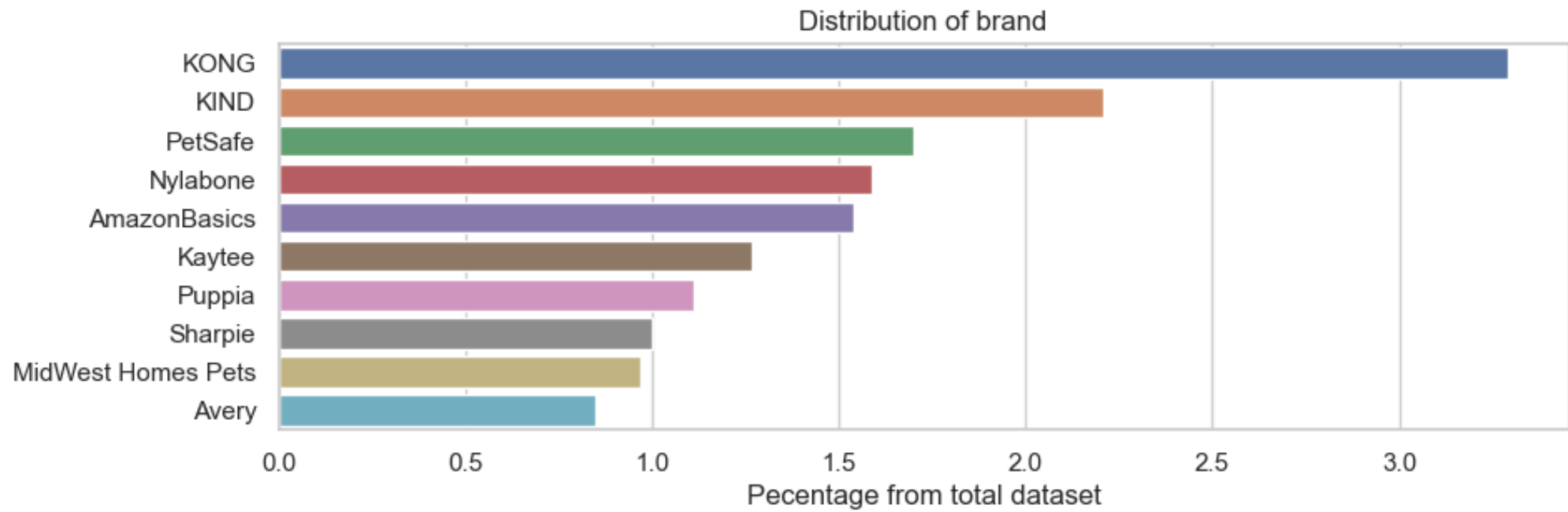
EDA

Category Distribution – Pet Supplies with more than 60%

EDA

Brand Distribution – Kong and Kind are most common
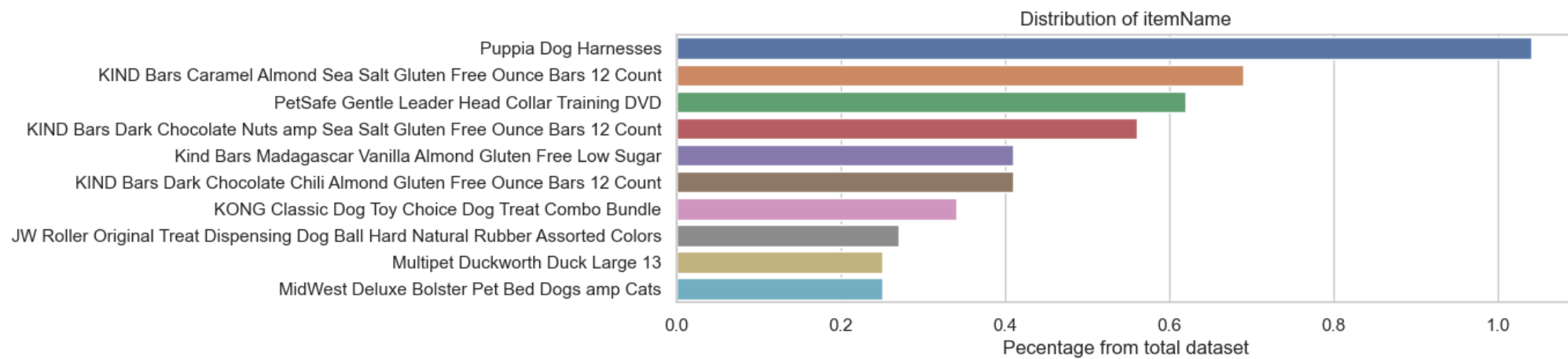
EDA

Item Distribution

Distribution of Rating

EDA

Rating Distribution – Average Rating is ~4.3
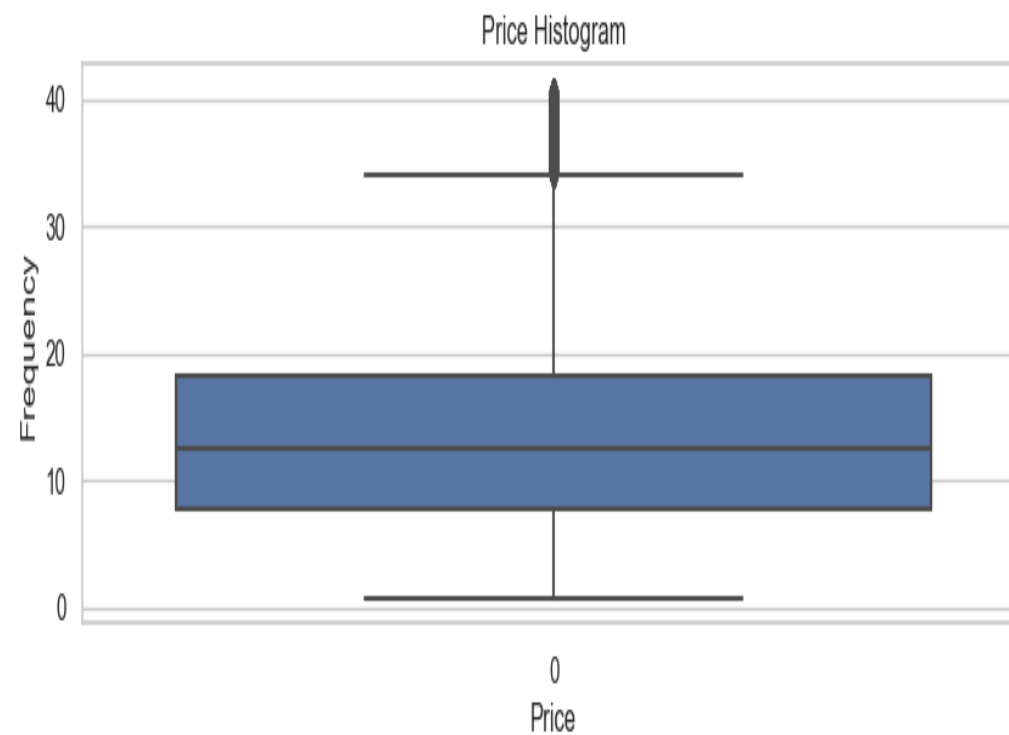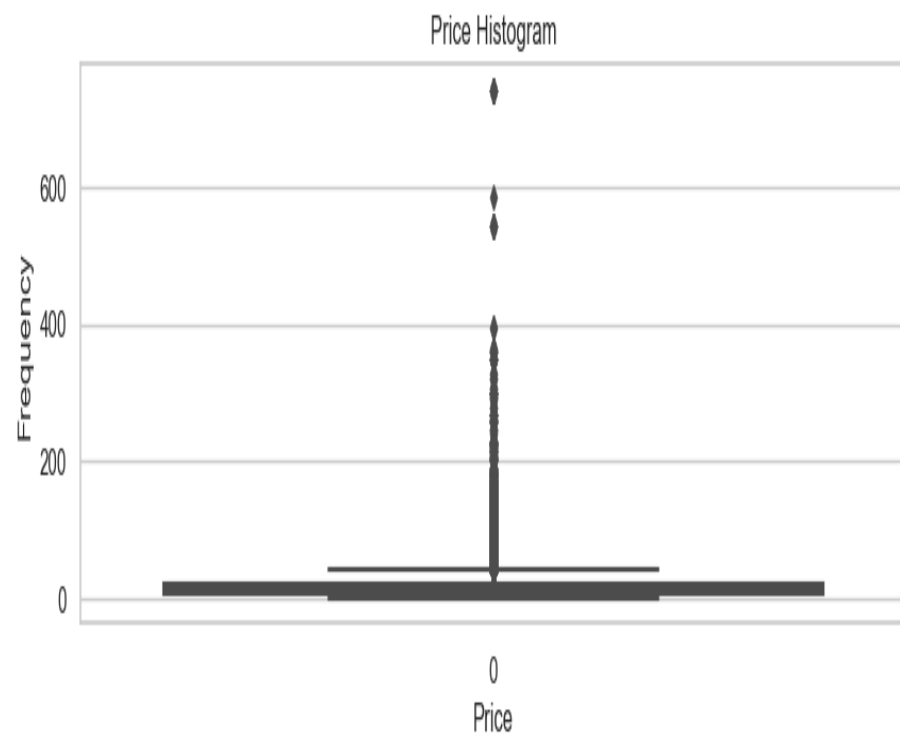
EDA

Price Box-Plot Distribution ~9K outliers , Average Price ~14$

Model Selection
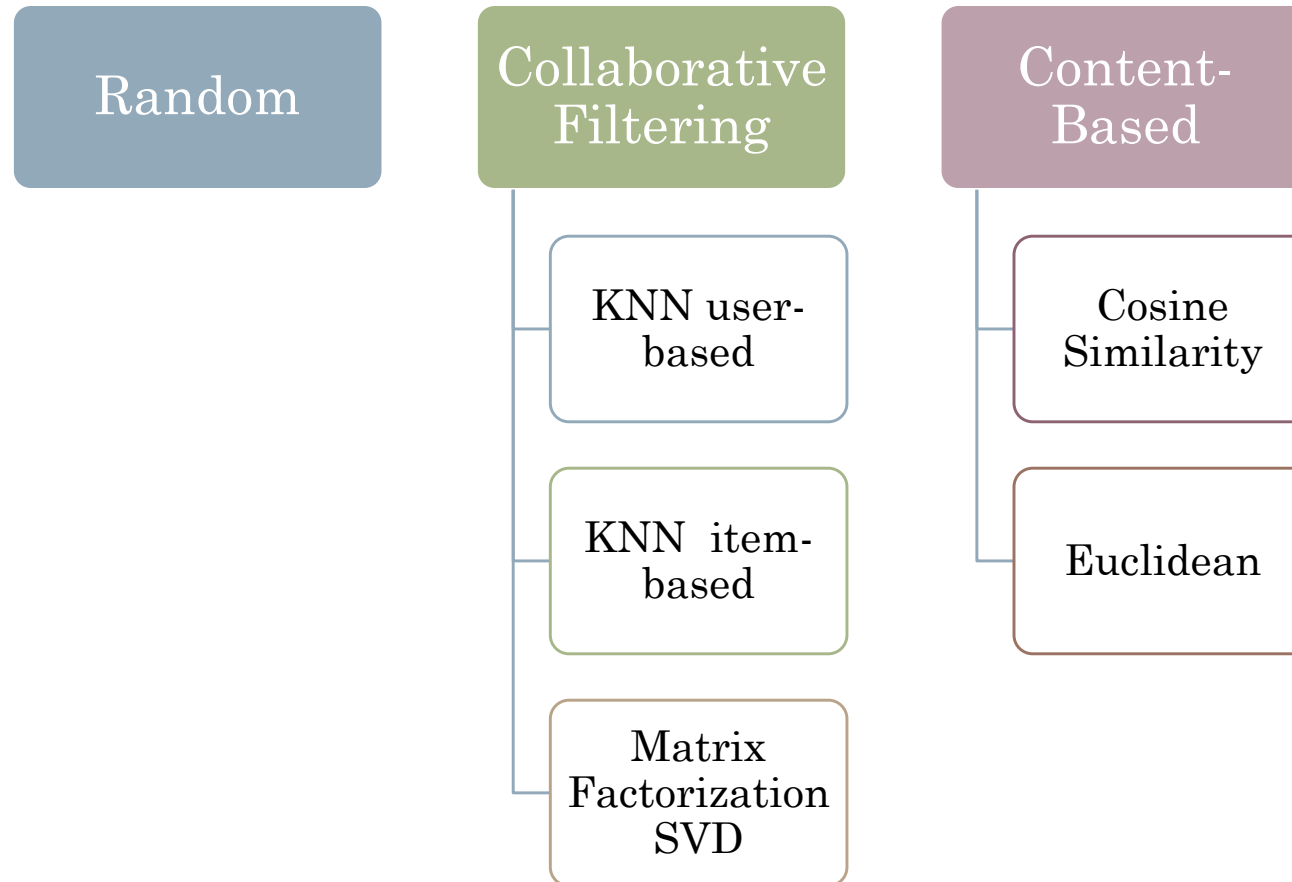
CF vs CB

Content Based Model

Concatenate Textual Features

$\xrightarrow{1}$

Text Embedding – Pretrained Sentence Transformer.

Create aggregated embed vector for each user and item.

$\xrightarrow{3}$

Calculate Top K items for a given user (User @ Items) by Cosine or Euclidean.

2

t-SNE Visualization with Category
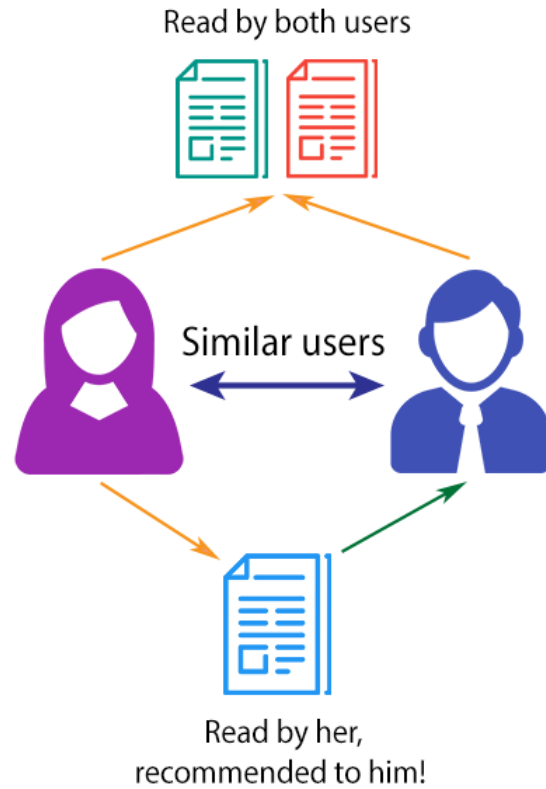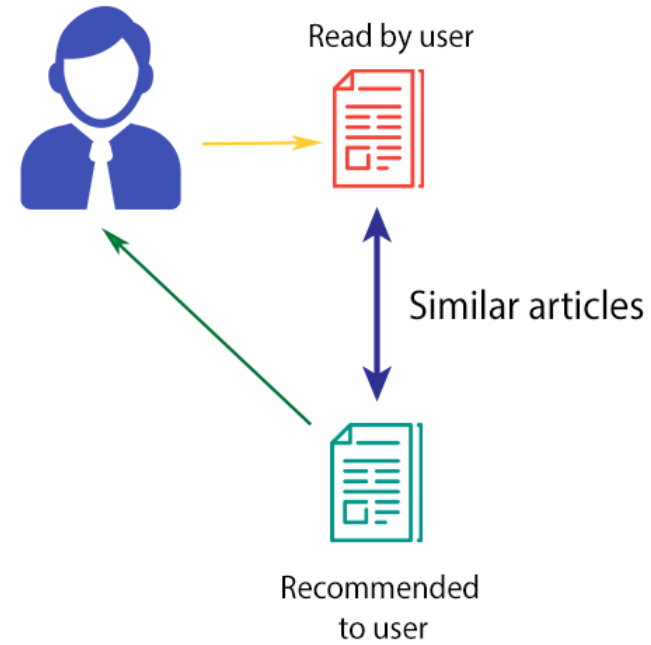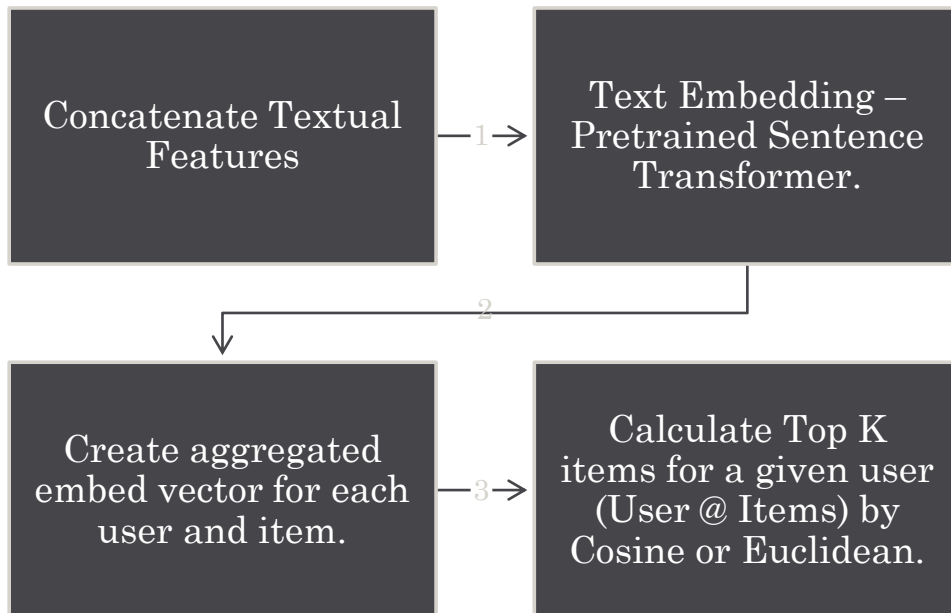
Evaluation Metrics

- RMSE
- recall@K (K ,Rating Threshold)
- precision@K (K ,Rating Threshold)

Note - AP@K , MAP@K and more.

All documents

Retrieved documents

$$precision@5 = \frac{3}{3+2} = \frac{3}{5}$$

$$recall@5 = \frac{3}{3+1} = \frac{3}{4}$$

Model Comparison

	Random	KNN Users	KNN Items	MF SVD	CB Cosine	CB Euclidean
RMSE	1.41	1.18	1.07	0.95	1.13	1.57
recall@K	0.56	0.78	0.8	0.79	0.82	0.05
precison@K	0.65	0.76	0.8	0.79	0.8	0.07

K = 10

Rating Th = 4

	userName	itemName	brand	category	price	rating	vote	user_id	item_id
128729	Kristie D	quot Guinea Habitat rdquo Guinea Pig Cage amp ...	MidWest Homes Pets	Pet Supplies	55.770000	5.0	0	7642	11046
138704	Kristie D	Kaytee Clean amp Cozy Colored Small Animal Bed...	Kaytee	Pet Supplies	8.270000	5.0	0	7642	5003
138810	Kristie D	Oxbow Animal Health Hamster Gerbil Fortified Food	Oxbow	Pet Supplies	20.178349	5.0	0	7642	6831

✓ #KNN ITEM BASED# ···

	itemName	brand	category
0	Prince Peace Organic Tea Oolong 100 Tea Bags	Prince Peace	Grocery Gourmet Food
1	Pet Champion Adjustable Harness Collar Leash M...	Pet Champion	Pet Supplies
2	Blue Buffalo Life Protection Formula Natural P...	Blue Buffalo	Pet Supplies
3	Elk Antlers Dogs Large quot Premium Antler Che...	Heartland Antlers	Pet Supplies
4	BIC Round Stic Xtra Life Ballpoint Pen Medium ...	BIC	Office Products

✓ #KNN USER BASED# ···

	itemName	brand	category
0	Prince Peace Organic Tea Oolong 100 Tea Bags	Prince Peace	Grocery Gourmet Food
1	Elk Antlers Dogs Large quot Premium Antler Che...	Heartland Antlers	Pet Supplies
2	PetSafe Nylon Dog Leash Strong Durable Traditi...	PetSafe	Pet Supplies
3	Idepet 1SET Dog Muzzles Adjustable Dog Mouth C...	Idepet	Pet Supplies
4	Seachem Ammonia Alert	Seachem	Pet Supplies

Example CF KNN

	userName	itemName	brand	category	price	rating	vote	user_id	item_id
128729	Kristie D	quot Guinea Habitat rdquo Guinea Pig Cage amp ...	MidWest Homes Pets	Pet Supplies	55.770000	5.0	0	7642	11046
138704	Kristie D	Kaytee Clean amp Cozy Colored Small Animal Bed...	Kaytee	Pet Supplies	8.270000	5.0	0	7642	5003
138810	Kristie D	Oxbow Animal Health Hamster Gerbil Fortified Food	Oxbow	Pet Supplies	20.178349	5.0	0	7642	6831

✓ #SVD# ⋯

	itemName	brand	category
0	Uni Jetstream Multi Function 4 Color Ballpoint...	uni	Office Products
1	Smartsticks Rawhide Free Dog Chew	SmartBones	Pet Supplies
2	Pilot Fountain Pen Ergo Grip Extra Fine Nib Bl...	Pilot	Office Products
3	Kaytee Timothy Complete Diet Rabbit	Kaytee	Pet Supplies
4	Millers Forge Stainless Steel Dog Nail Clipper...	Millers Forge	Pet Supplies

Example SVD

	userName	itemName	brand	category	price	rating	vote	user_id	item_id
128729	Kristie D	quot Guinea Habitat rdquo Guinea Pig Cage amp ...	MidWest Homes Pets	Pet Supplies	55.770000	5.0	0	7642	11046
138704	Kristie D	Kaytee Clean amp Cozy Colored Small Animal Bed...	Kaytee	Pet Supplies	8.270000	5.0	0	7642	5003
138810	Kristie D	Oxbow Animal Health Hamster Gerbil Fortified Food	Oxbow	Pet Supplies	20.178349	5.0	0	7642	6831

✓ #Content Based# ···

	itemName	brand	category
0	Oxbow Animal Health Cavy Cuisine Adult Guinea ...	Oxbow	Pet Supplies
1	Oxbow Animal Health Orchard Grass Hay Pets	Oxbow	Pet Supplies
2	Oxbow Carnivore Care Pet Supplement	Oxbow	Pet Supplies
3	Oxbow Simple Rewards Natural Oven Baked Treats...	Oxbow	Pet Supplies
4	Oxbow Natural Science JOINT Supplement small a...	Oxbow	Pet Supplies

Example CB

- Exploitation vs Exploration

- Cosine Similarity VS Euclidean

- RMSE vs @K metric

- Future Work
  - CF and CB Hybrid
  - RBM
  - FM
  - DSSTNE

Conclusion