A Comparative Study of Recommender Systems

Machine Learning Project
By:-
Abhinav, Aryan, Ashmik, Dheeraj, Dhruv and Diwakar

Why do we need Application Recommendation ?

- The Google Play Store offers millions of apps. More apps are added daily. How can users find new and engaging content ? Yes, you can use search to access your content.

- App Recommendation system helps in addressing the information overload problem by retrieving the information desired by user based on his or similar user’s preferences or interests

Is keyword search enough for efficient app browsing on app store?

- Keyword search finds app based on the keyword entered by the app.
- Recommender Systems recommend apps based on users preferences.

- For example, if a user has recently used Uber App, then she may also find the Rapido very useful. Such suggestions can only be made through a recommendation system.

- There are more than 2,714,499 applications available in the Google Play store. In addition, there are currently 250 million application downloads every day.

- Therefore, there is an urgent need to provide effective app recommendation services.

Related Works

App Recommendation: A Contest between Satisfaction and Temptation

1. Formulated dataset as a tuple containing user id, app id, user’s current app collection and the action that user takes on the said app.
2. Actual-Tempting Model captures factors that invoke a user to replace an old app with a new one.
3. They have used the data logged from an iPhone App, namely, *Limited-time Free*8, which collects information of iPhone apps that become available for free for a limited time in the Apple App Store.

Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data

1. Aher et al have built a course recommendation system using simple k-means algorithm and association rule algorithm – Apriori.
2. They have considered 13 course categories and 82 courses that were related to two branches i.e. Computer Science & Engineering (CSE) and Information Technology (IT) to collect data from students.

A Mobile Application Recommendation Framework by Exploiting Personal Preference With Constraints

App recommendation parameters

- Popularity of apps
- User preference
- Device constraints

Recommendation of Mobile Application based on Social and Contextual User Information

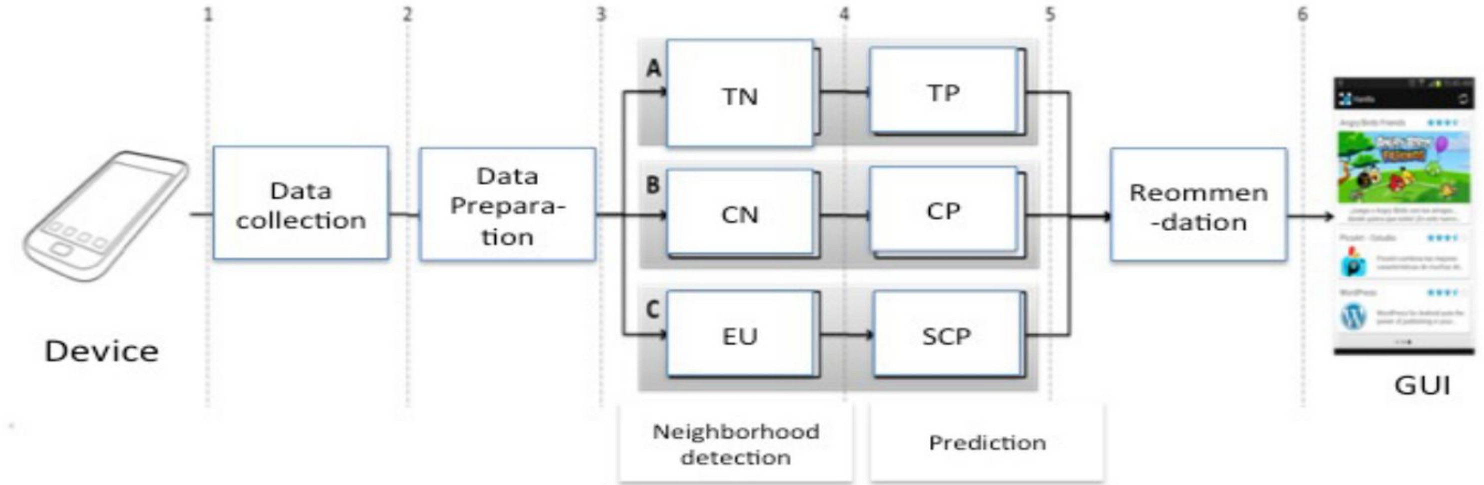


Figure 1. Vanilla overall process

Mobile Applications Recommendation Based on User Ratings and Permissions

- A dataset of 100 android applications along with their user ratings and permissions is used.
- Recommendation parameters used:
    - Popularity of the app.
    - Permissions of the app

- K-means clustering algorithm is used.

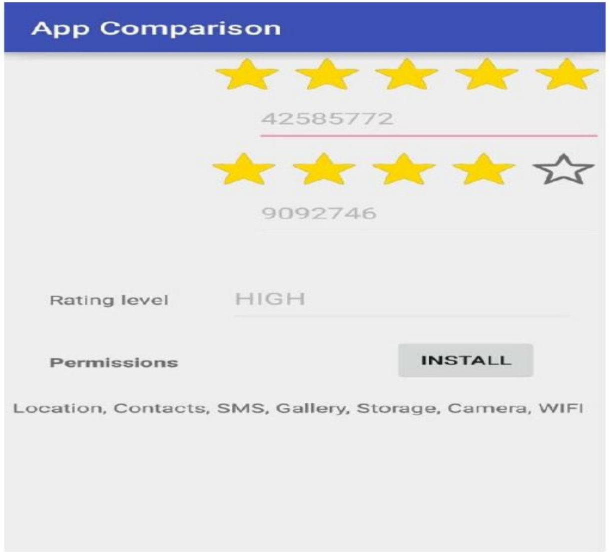


Fig. 3. Rating and Permissions Details

TABLE I. DATASET AFTER CRAWLING

Appname	Overall-Rating	5-Star Rating Count	4-Star Rating Count
YouTube	4.3	15101432	3112255
WhatsApp	4.4	43355932	9224952
Facebook	4.1	43599840	11351024
Messenger	4.0	29658015	7565690
Skype	4.1	6192942	1665217
Xender	4.4	757706	187167
Hike	4.3	1732438	584664
Youcam perfect	4.5	33325194	4291020
Clashof clans	4.6	1052755	220279

TABLE III. APPLICATIONS ACCESS PERMISSIONS

Appname	Location	Contacts	Camera	SMS
YouTube	Y	Y	Y	N
WhatsApp	Y	Y	Y	N
Facebook	Y	Y	Y	Y
Messenger	Y	Y	Y	Y
Skype	Y	Y	Y	N
Xender	N	N	N	N
Hike	Y	Y	Y	Y
Youcam perfect	N	N	Y	N
Clashof clans	N	N	N	N

Multi-objective mobile app recommendation: A system-level collaboration approach

- **The need for a multi-objective approach**: All the methods before focused Primarily only on Similarity. However other attributes like diversity, utility and robustness have also proven to be effective according to recent studies.

- **Leveraging the multiobjective approach:** This is done in order to satisfy the variety of app needs of the users by capturing different objectives in the generation of recommendations.

- **Utilizing the system collaborative strategy:** Taking source recommendations from various source RS's and then employing it's own optimizations and techniques.

Multi-objective mobile app recommendation for mobile apps

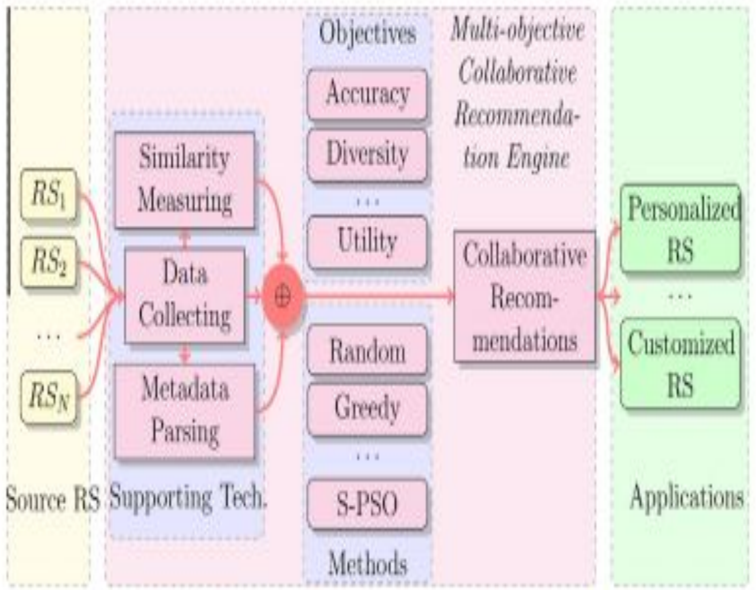


Fig. 1. Multi-objective collaborative recommendation for mobile apps.

To investigate the efficiency, effectiveness of the collaborative recommendations we set the number of iterations in our algorithm to be fixed and observe that on an average it needs only thirty iterations to generate better collaborative recommendations for 80% of the apps.

A New Collaborative Filtering Algorithm Using K-means Clustering and Neighbors’ Voting

- This is combination of two algorithms
- Collaborative filtering and K-means
- It divides data into groups or “clusters” such that similar points are in the same and dissimilar points are in different clusters.

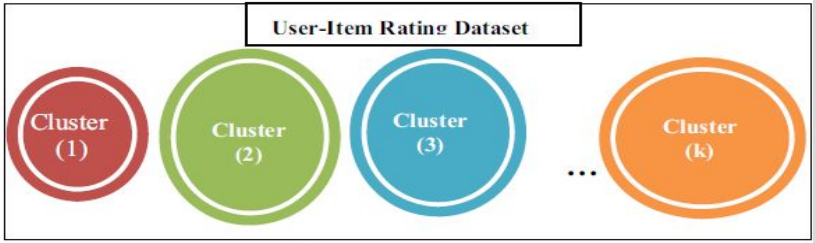


Figure 2. User clustering

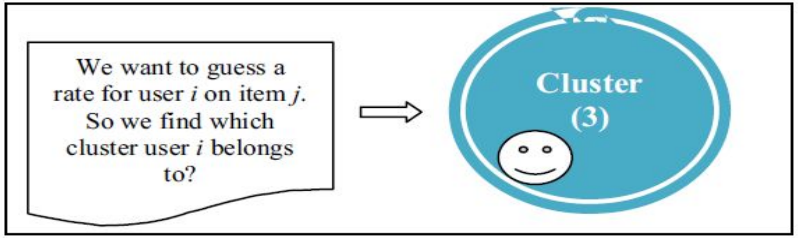


Figure 4. Finding neighbors of user i

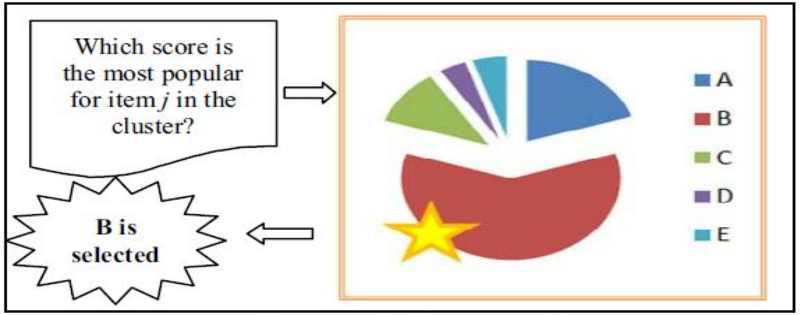


Figure 5. Producing prediction

TABLE I. A SAMPLE OF USER-ITEM DATABASE

	Item$_1$	**Item$_2$**	**Item$_3$**	**…**	**Item$_n$**
User$_1$	0	4	3	…	0
User$_2$	5	0	2	…	4
User$_3$	4	5	1	…	5
…	…	…	…	…	…
User$_m$	0	4	5	…	3

Mobile App Recommendation System Using Machine learning Classification

Here researchers use apkpure.com which is one of the famous android application markets for collecting the dataset by crawling.

After Content Extraction, He takes three parameters i.e., size, user ratings, and permissions. Here, he is using the ID3 Algorithm which is a classifying algorithm.

Here he classify it into three part A, B, C

So the total scores can be classified as Class A: 0-5 Class B: 5.1-10 Class C : more than 10.1A cloud is used for storing information during classification. This paper gives a simple outlook on how ratings, permissions, and size of applications can be used for an application recommendation system.

Google Play Store App Dataset

App Dataset

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.48	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.0	2.2 and up
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varies with device	Varies with device
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

10841 rows × 13 columns

User Review Dataset

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
2	10 Best Foods for You	NaN	NaN	NaN	NaN
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000
...	...	...	...	...	...
64290	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64291	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64292	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64293	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64294	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN

64295 rows × 5 columns

- We are provided with an app dataset having 10840 samples and 13 attributes which are *app name, category, rating, reviews, size, installs, type, price, content rating, genres, last updated, current version, and android version*.

- Also, there is a user review dataset having 64294 samples and 5 attributes which are *app name, user review, sentiment, sentiment polarity, sentiment subjectivity score*.

- The given app dataset contains lots of missing values, duplicates and anomalies. We remove all such types of impurities from the dataset.

- Moreover, we have dropped unimportant attributes such as Size, Genre, Current Ver, Android Version, and Last Updated.

- Next, we have scaled all the numerical valued attributes using a MinMax Scaler.

App Dataset after cleaning

	Category	Rating	Reviews	Installs	Type	Price	Content Rating
0	ENTERTAINMENT	4.7	11661	1,000,000+	Free	0	Everyone
1	TOOLS	4.0	19	10,000+	Free	0	Everyone
2	SOCIAL	4.6	22098	1,000,000+	Free	0	Everyone
3	LIFESTYLE	3.8	718	10,000+	Paid	$399.99	Everyone
4	COMICS	4.4	190	10,000+	Free	0	Everyone
...	...	...	...	...	...	...	...
8190	NEWS_AND_MAGAZINES	4.4	27	100+	Free	0	Everyone
8191	COMMUNICATION	4.7	573	10,000+	Free	0	Mature 17+
8192	TOOLS	4.5	259	10,000+	Free	0	Everyone
8193	COMICS	3.5	115	10,000+	Free	0	Mature 17+
8194	SOCIAL	4.5	40467	1,000,000+	Free	0	Everyone

8195 rows × 7 columns

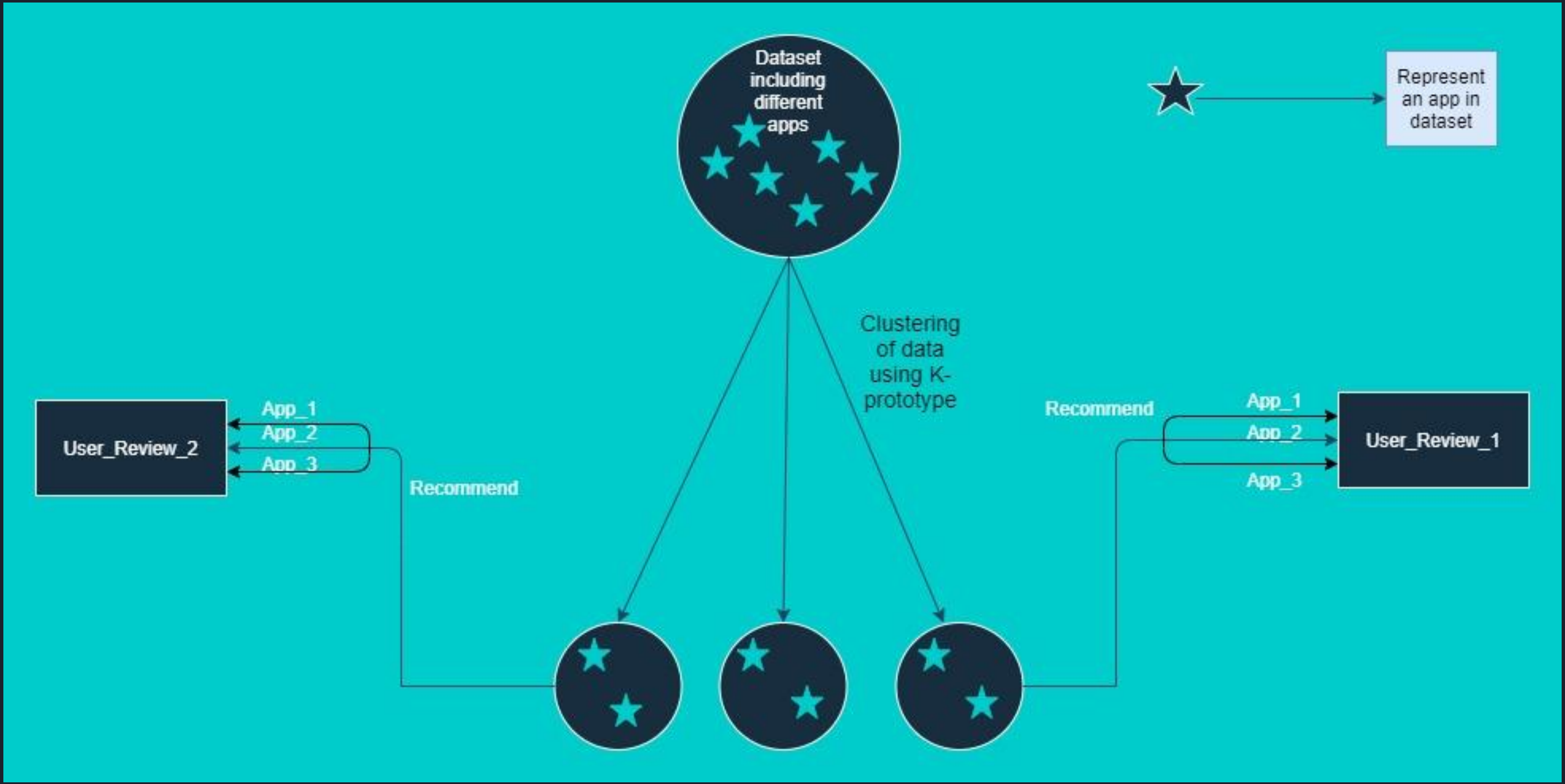
Methodology

K-Prototype Clustering Algorithm

K-Prototype is an combination of K-means algorithm (for numerical-type data) and K-modes algorithm (for categorical data) that performs clustering on a mixed-type dataset.

If a user gives a positive review about an app, she must also prefer similar apps that are equally popular. We will recommend the user better apps (with respect to the popularity metric) from the same cluster as the app that was reviewed by her. (Popularity metric = rt + rv + ins)

If she gives a negative review about the app, then she must require apps from the same category which should perform better than the rated app. For this, we recommend her better apps (with respect to the overall rating of the app) that belong to the same category and the same cluster as the reviewed app.


Dataset including different apps
Represent an app in dataset
Clustering of data using K-prototype
User_Review_2
App_1
App_2
App_3
Recommend
Recommend
App_1
App_2
App_3
User_Review_1

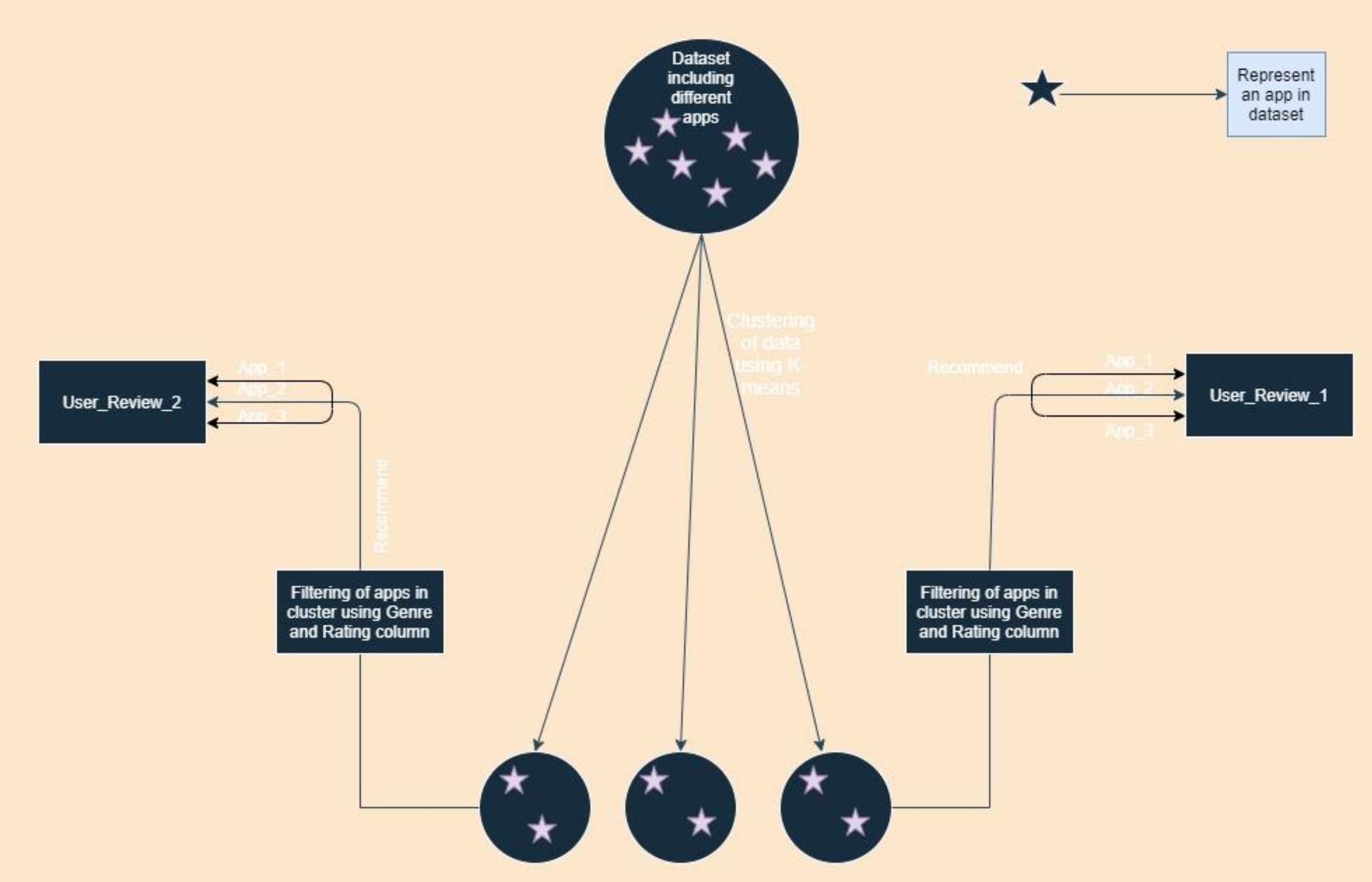
K-Means and the Genre clusters

In this method, we utilise the K-means clustering method to cluster the numerical attributes of the dataset, while leaving out the categorical attributes :- **KMeans Cluster**

We use a k-modes like approach to cluster the apps on the 'Genre' attribute. :- **Genre Cluster**

If a user gives a positive review about an app, she must also prefer similar apps that are equally popular. In this case, we will recommend apps from the same k-means cluster and same genre cluster.

If she gives a negative review about the app, then she must require apps from the same category which should perform better than the rated app. For this, we recommend her better apps (with respect to the overall rating of the app) that belong to the same genre cluster and the same k-means cluster as the reviewed app.



Dataset including different apps
Represent an app in dataset
Clustering of data using K-means
User_Review_2
App_1
App_2
App_3
Recommend
Recommend
App_1
App_2
App_3
User_Review_1
Filtering of apps in cluster using Genre and Rating column
Filtering of apps in cluster using Genre and Rating column

Evaluation

Metrics

- **Coverage** is the percent of items in the training data the model is able to recommend on a test set.

- **Inta-List Similarity** is the average cosine similarity of all items in a list of recommendations. This calculation uses features of the recommended items to calculate the similarity.

Results

Method	Coverage	ILS
K-Prototype	92.9%	14%
K-Means and Genre cluster	93.2%	16%

Conclusion

- As can be seen, both the methods perform almost the same on this dataset.
- This is quite an untraditional way of designing a recommender system as the literature has not seen any work that has targeted any dataset like this.
- This recommender system might be useful when the recommendation has to be made but no information on the user is available. This might be the case when one has just logged in to an app store or any online store for that matter.

Contribution

Abhinav Singh :- Surveyed Yin2013 and Aher2013 papers. Formulated and Implemented KMeans and genre cluster method.

Aryan Sharma :- Surveyed Zhu2017 and Chamorro2017 papers. Formulated and Implemented Gaussian mixture cluster method.

Ashmik Harinkhede :- Surveyed Xia2014 paper. Formulated and Implemented KPrototype cluster method.

Dheeraj Kumar :- Surveyed Huang2018 paper. Formulated and Implemented DBScan Algorithm

Dhruv Narayan Gupta :- Surveyed Jisha2018 and Omar2019 papers. Performed dataset cleaning and normalisation.

Diwakar Bharti :- Surveyed Jisha2020 and Dakhel2011 papers. Formulated and Implemented Agglomerative algorithm.