Explainable Recommender Systems

Yashaswini Viswanath, DR. Rashmi Agarwal, Suresh Lokiah

Reva Academy of Corporate Excellence, Reva Academy of Corporate Excellence, Business School of AI,
yashaswiniviswanath.ai01@race.reva.edu.in,
rashmi.agarwal@reva.edu.in,
sureshlokiah@race.reva.edu.in

***Abstract*— Recommenders systems add business value and help in better monetization for multiple stakeholders. The objectives of the recommender system vary from domain to domain and in this paper movie recommender system is the focus. The advent of powerful mobile devices have led to proliferation of many mobile based movie recommenders whereas earlier they were confined to desktop or web based applications. The future generation of users expect the recommender systems to explain itself as they want to make informed choices rather than consuming just anything the AI offers to them. The veil of AI as magic is to be lowered as the consumers seek meaningful explanations for the recommendations that are laid in front of them. This paper builds a content based recommender system based on genre data of MovieLens dataset by GroupLens. Clustering technique k-means is used to find hidden patterns to form clusters of similar movies. Surrogate models are used to provide explanations of k-means clustering using decision trees. The decision path along with the rules form the explanation of cluster assignments and reasoning behind recommended movies.**
Traditionally recommender systems are built on similarity measures based on user preferences which leads to filter bubbles and echo chambers. Serendipitous recommendations can be generated by coming up with novel and unexpected movies using cosine similarity between centroid vectors of k-means clustering. These serendipitous movies can be explored by user based on the explanations given by the decision paths of surrogate model. Last but not the least, the explanations can be delivered by multiple interfaces. The prevalent ones are textual and visual methods. This paper explores multimodal explainability through an additional audio medium.

***Keywords*— Recommender system, Content based recommendations, explainable AI, Serendipitous recommendations, Explanation interfaces**

I. INTRODUCTION

This Recommender systems can be defined as applications that suggest items which the recommender system thinks will be useful to the user. Recommendations help the users to decide to choose an item from a collection such as which doctor to consult, what food to try, who is a potential friend or a job. What is recommended by the recommender systems form the collection of items which is an open pool for the user to select from. Earlier recommender systems used to focus on one type of item but now we have multiple types of items in the same application, for example, LinkedIn suggests jobs and also suggests people to follow or connect.

A recommendation engine provides the items to suggest and is supported by the user interface such as a Graphical User Interface in the form of a website or mobile application. In recent years we have seen recommendation systems in conversational bots like Alexa from Amazon which listen to the search term by the user and recommend products from the Amazon e-commerce catalogue and receives the order for products through audio itself using AI techniques such as Automatic Speech Recognition. Recent study conducted as shown the challenges in delivery of the recommendation and the challenges faced when the interface is purely audio [21]. Also, recommender systems exist as part of certain devices such as Kindle Readers where the user sees the recommendation engine as part of the device itself.

Search, recommendations and personalization are counterparts which are closely related in applications which have items to offer. Generally a search query is placed such as a book to read on a Kindle device and the recommendation engine churns out suggestions which are then ranked and filtered based on personalization techniques. This results in a unique set of recommendations to each user along with a generic set of recommendations which may be based on trends and seasonality. These types of generic recommendations have their roots in social sciences where people with domain knowledge with prior experience were consulted for suggestions and are called knowledge based recommender systems.

The most prevalent types of commanders are content based and collaborative filtering techniques. This paper focuses on content based recommenders where the features of an item are pivotal to come up with suggestions. This calls for mechanisms to have feature data to be available and is a challenge when applying recommendation principles to new domains. A recent research work conducted explored applying a media recommender based on content based techniques for mental health patients which required information about the triggers in each movie in the catalogue.

Explanations on why a particular item was recommended by a content based recommender system can be obtained by explicitly the features of the item which influenced the decision to recommend. This information about the features enables the model developers to access the fairness and biases if any in coming up with the recommendation. Also, the trust and transparency of the recommendations are improved by providing these types of explanations.

One of the main advantages of content based recommender systems is in addressing cold start problems where the users or items are new to the recommender system. This problem is seen in collaborative filtering based recommenders which are solely based on user ratings and the system can start recommending only after garnering a substantial number of ratings. Thus by using information that is initially available the recommendations can be suggested and this information needs to be legit from a verified source. Hence, This paper chooses MovieLens dataset which is a seasoned dataset made specifically for research on algorithmic advances of recommender systems [22].

MovieLens dataset has information about the movie category called genre which is enumerated by experts. Earlier research has used genre correlations which are based on a set of genres for each movie. The literature talks about a mechanism to collect user preference of genre and using this to personalise the recommendation so obtained in the previous step. Also it takes into account the average rating of the movies based on the correlations of genre feature [17].

Another body of research talks about the usage of cluster analysis using vectors in multi-dimensional space to find patterns in the catalogue and come up with recommendations. The items in one cluster have similar features when compared with items in all other clusters [18]. This idea forms the basis for this paper where clustering technique is used along with genre information. Similarity of movies is a relevant strategy used as a basis for recommending in the prevalent movie recommendations. But in recent years there has been traction in a new strategy where novel and unexpected items are used as an recommendation strategy to learn more about users and also to improve user satisfaction in movie recommendations by Amazon and Netflix [12] [13] [14].

Recommendations tend to become stale, entering filter bubble and echo chambers [23] [24]. This calls for diversity in the recommendations to maintain user engagement and improve business outcomes. Serendipitous recommendations is a mechanism used by users to learn more about areas not yet traversed, in this case it relates to movies not yet watched before. Here the user initially does not know potential genre items or item names or information about items and relies on the recommender for discovery.

The problem of over-specialization of the recommender system is the motivation for this paper to research on serendipitous recommendations which are called by multiple names such as serendipity, tangent, surprise [12] [13] [14]. Novelty is characterised by items which the user would have discovered autonomously but serendipitous recommendations are hard to discover items. There is a possibility that users might like this serendipitous item or not, it is a gamble the recommender system takes [25]. All serendipitous items are novel but not all items are serendipitous. For example, a new movie that the user has not watched is novel but a movie from a director who has never watched it before is serendipitous.

II. LITERATURE REVIEW

An The traditional approaches of recommender systems involve collaborative filtering using techniques of matrix factorization where ratings from multiple users are taken into consideration. The recent advancement in this type of recommenders is the introduction of tensor factorization with users, topics and tasks matrices. Latent variable models are probabilistic models, efficient, scalable and parallelizable models which are advantageous in large scale recommenders.

Neural Embeddings using deep learning techniques were introduced where user embeddings are used to come up with product recommendations and recent advancements use a joint user-item embedding. Also Neural Collaborative Frameworks used user latent vectors and item latent vectors which are fed into neural networks. The common thread connecting these traditional methods is the user centric approach nature where user needs, user interests, user behaviour and interactions along with personalization is given importance.

In recent advances, multi-stakeholders are identified and how they can benefit individually in marketplaces is the crux of improvement. Earlier we had offline marketplaces which evolved to online ecommerce sites, tasks such as taxi booking, professional services like

Practo. Vertical marketplaces focus on one problem/thing whereas horizontal has a larger customer base and also facilitates exchanges in different categories such as Amazon. They may offer various types of items such as goods, information, services and banking products. Examples of multi stakeholders are advertisers, drivers, campaigners, hosts and platform providers. The various metrics for gauging the success of the recommender system particular to each application such as exposure, audience growth, revenue, diversity, engagement levels, retention, depth and other proxies for user satisfaction [28].

The cold start problem of collaborative filtering strategy can be solved by content based recommender systems which take into account the features of the items up for recommendation. Also, in the research conducted about media recommenders for the mental health domain, it was found that having the ratings of other users may not always be available due to privacy concerns. The scalability of content based recommenders is notable given it doesn't depend on the number of users and the algorithm runs the same for any number of users added. Personalization is easy in content based recommenders as the user's preference can be used to filter the candidates generated for recommendation. Niche items which are consumed by less number of users have a good advantage in content based recommenders [29].

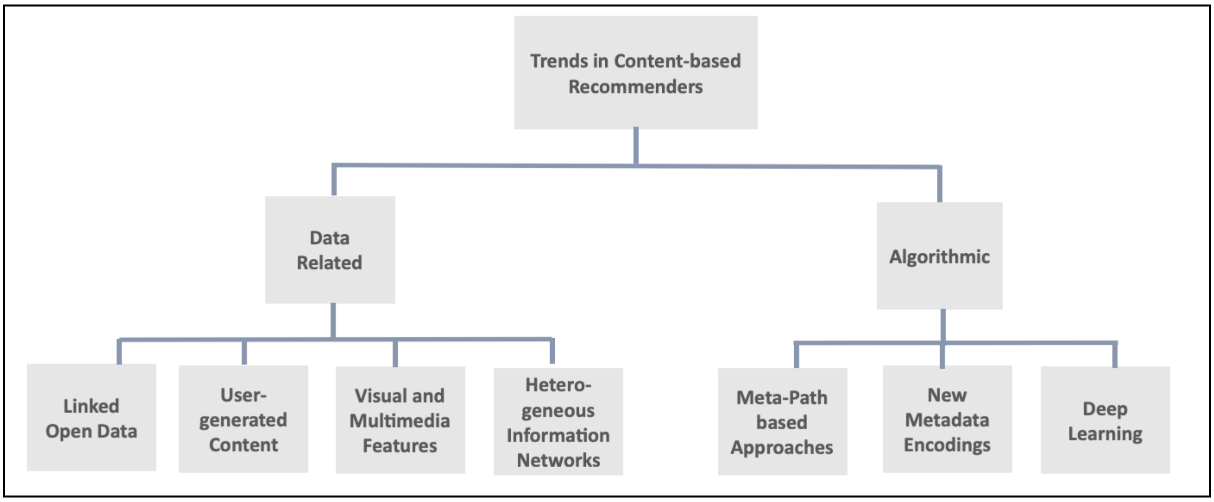

Figure No. 2.1: Advances in Content based recommenders [30]

Metadata of description of items or textual indexing are used in content based recommenders. The availability of Linked Open Data (LOD) offers new methods for usage of this external knowledge to come up with recommendations. For example the MovieLens data considered in this example had additional information added through link information to other movie databases such as IMDB and TMDB . This information is useful in bias mitigation techniques which are aimed at improving the diversity of movies recommended. Also, these linked data are useful in generating explanations for the content based recommendations which enable better trust and transparency of the recommender system.

The proliferation of affordable internet has enabled the accumulation of user generated content for example product reviews, discussion in forums and tags for the items which are used to generate explanations and to enhance the recommendations. The recent advances in the analytics space of image data and video data has enabled parsing or studying of the object itself to come up with metadata. In a recent research conducted the media to be consumed by mental health patients was involving recognition of triggers in the media using automated software. In another research Convolution Neural Network was used to under the features of the image in fashion retail recommender system [30].

Explanations cater to the "why" of reasoning of recommendations. This enables persuasiveness, transparency, effectiveness, user satisfaction and trustworthiness of recommendation systems from a user's perspective. Another added advantage of explanations is the facilitation of debugging and refinement of the system by the modellers and designers. Model agnostic explainers and model intrinsic explainers have different merits and are based on observation of human cognition where sometimes we take decisions first and then explain why we chose that path or otherwise explain what are the deciding factors and choose an item [31].

There is a trade-off between explainability and effectiveness of model and here comes the need for surrogate models for explainability [9]. Explanations can be obtained from different sources of information and mode of delivery can be different based on the domain. For example, a relevant item or radar chart or sentence or an image can be one of the display styles. Content based recommenders go along with feature based explanations and are intuitive in nature.

Recent Clustering techniques find patterns in data by quantization of data points which are unlabelled. The users of clustering algorithms want to know why a data point was assigned to a particular cluster. The challenge is due to consideration of all the features to form cluster assignments. The initial introduction of decision trees for solving this explainability problem was successful [9]. An unsupervised decision tree was a canonical example of a clustering model with explainability.

The improvement in explainability can be achieved by usage of a small decision tree to partition the data set into clusters. This is a straightforward approach as we can characterise each of the clusters using the decision tree . This approach is independent of input size and dimensions and provides novelty compared to older approaches. The characteristics of a tree by restricting to k leaves where k is the number of clusters, the solution is independent of data dimension. This threshold tree uses k-1 features and any dataset can be quantized. Also, any new data point can be assigned a cluster using the tree.

The general strategy of recommendations is similarity of item related to historic or recent consumption of items by the users. This narrows down the horizon of the users' and also may lead to less user engagement and resulting in decrement in business revenue. There is a value add in showing items which have not been explored and which are unexpected by the user. This is a win-win situation where users get help in item discovery and the other stake holders benefit by increase in user satisfaction and user engagement. Generally the accuracy metrics of mean average error, precision and recall can only partially evaluate the effectiveness of a recommender system. More than just persuading users to buy items, they can be helped with discovery and also gain their trust by providing explanation to the serendipitous items. Explanations enable users to conclude and choose which items match their needs [6].

In recent years, the question of how to automatically generate and present explanations has attracted increased interest in research. Today some basic explanation facilities are already incorporated in e-commerce Web sites such as Amazon.com. In this work, we continue this line of recent research and address the question of how explanations can be communicated to the user in a more effective way [15].

III. PROBLEM STATEMENT

Responsible Recommenders are pivotal to garnering trust by the users which enables deeper proliferation into the society. Recommender systems for various domains enable easy access to data and the learnings from one domain are applicable to other domains as well. This paper tries to tackle the problem of building a content-based recommender system by using genre data using k-means clustering and providing explanation using decision trees as surrogate model.

Recommender systems have evolved into multiple forms and one of them is serendipity, tangent, surprise recommendation which are recommender systems that come up with novel and unexpected, diverse items. This paper tries to tackle the problem of showing novel items to the user by employing similarity measures.

If the explanation generation for the recommendations is a challenge, then the mode of delivery of these explanations is a far more important one. This paper tries to tackle this problem of delivery of explanation after the generation of the explanation.

The primary objective of this study is to build a movie recommender system using MovieLens dataset which is a research dataset. Recommender systems have lot of scope in the future and conducting research in this area is a good opportunity. Among the many types of recommender systems such as content based, collaborative filtering, memory based, model based, hybrid recommenders, etc., this paper focuses on content-based recommender system based on genre using k-means clustering.

The secondary objective is to bring in a component of novelty in the recommender system. Generally, recommender systems suggest movies based on similarity, similar to movies that have been watched before. This paper tries to suggest movies belonging to category which has not been watched before adding an element of surprise to the user of the recommender system.

The tertiary objective is to provide explanations to the recommendations as to why a particular movie was recommended. Explainable recommender systems are the need of the hour, and this paper tries to conduct research on that front. The novelty brought in this space is the multi modal nature of the explanation delivery. Textual and visual explanations are the norm and here an audio agent conveys the explanation for the recommendation.

This paper is purely implementation based and acts as a small lamp shining new light with novel ideas on already well researched topics.

IV. DATASET

GroupLens by University of Minnesota publishes the MovieLens dataset which are widely used in industry, academia, research, and education. They are used in popular books, educational courses, and application software and are downloaded thousands of times a year. These datasets are a product of member activity in the MovieLens movie recommendation system, an active research platform that has hosted many experiments since its launch in 1997. MovieLens datasets are actively used to develop and test the recommender system's core algorithmic advancements across the world.

The dataset used in this research paper is ml-latest-small dataset downloaded from MovieLens website. It has 9764 movies with their titles and genres. It also consists of users and the ratings they have provided along with the timestamp as shown in Figure No. 6.1.

The Genre field is a set of categories which are "|" separated. The total number of categories is 18 and. A subset of them is provided for each movie. These datasets are the result of actual people interacting with MovieLens, a steadily evolving system.required for this paper is sourced from National Institute of Diabetes and Digestive and Kidney Diseases. This research institution is instrumental in conduction of research on common and chronic health conditions. The mission of National Institute of Diabetes and Digestive and Kidney Diseases is dissemination of science-based information for improvement of quality of life and health of general public. As part of the Institution's research activities, this dataset was captured and released to further study of researchers. The significance of this dataset is the population under study is Indian citizens. This gives us pride to conduct research on this dataset.

V. METHODOLOGY

The MovieLens dataset needs to be prepared for usage in this paper. The dataset needs to be cleaned to be usable for further processing. Processing the dataset such that the models can find patterns is a crucial step in the paper methodology. Exploratory data analysis is to be conducted to understand the MovieLens dataset where the features of the dataset are the predictor variables. Once the data is ready, k-means model needs to be built and cluster assignment is to be performed. The important step is to build a decision tree classifier with cluster segments as labels. The so created decision tree classifier will provide explanations where the leaf nodes are the clusters and the non-leaf nodes are the features and their ranges. The accuracy of the decision tree is the percentage of explainability. Watching movies on the Recommender System and Exploration of movie by the recommender system are simulated using audio. The paper methodology is pictorially shown in Figure No. 5.1.Pima Indian Diabetes dataset needs to be prepared for usage in this paper. The dataset needs to be cleaned to be usable for further processing. Processing the dataset such that the models can find patterns is a crucial step in the paper methodology. Exploratory data analysis is to be conducted to understand the diabetes dataset where the features of the dataset are the predictor variables. Once the data is ready, kmeans model needs to be built and cluster assignment is to be performed. The important step is to build a decision tree classifier with cluster segments as labels. The so created decision tree classifier will provide explanations where the leaf nodes are the clusters and the non-leaf nodes are the features and their ranges. The accuracy of the decision tree is the percentage of explainability. The paper methodology is pictorially shown in Figure No. 5.1 MovieLens Dataset Preparation: The MovieLens dataset is downloaded as a comma separated values file which is loaded into a pandas dataframe for usage. The features of the dataset are captured for future processing. Genre data is explored as shown in Figure No. 7.1. The main feature which this paper will be utilizing is genre. Total 951 unique genres are available in the dataset.

```
movies_data.genres.value_counts()

Drama                                               1053
Comedy                                               946
Comedy|Drama                                         435
Comedy|Romance                                       363
Drama|Romance                                        349
                                                    ...
Action|Crime|Horror|Mystery|Thriller                   1
Adventure|Animation|Children|Comedy|Musical|Romance    1
Action|Adventure|Animation|Comedy|Crime|Mystery        1
Children|Comedy|Fantasy|Sci-Fi                         1
Action|Animation|Comedy|Fantasy                        1
Name: genres, Length: 951, dtype: int64
```

Figure No. 7.1: Genre data in MovieLens

One hot encoding mechanism is applied to the entire dataframe to convert categorical into numeric data which can used for computation in the algorithms. The genre which exits is given a 1 value and which is absent is given a 0 value as shown in Figure No. 7.2.

	movieId	title	genres	Adventure	Comedy	Action	Drama	Crime	Children	Mystery	...	Horror	Fantasy	Western	Film-Noir	Romance	Sci-Fi	Musical	War
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1	1	0	0	0	1	0	...	0	1	0	0	0	0	0	0
1	2	Jumanji (1995)	Adventure|Children|Fantasy	1	0	0	0	0	1	0	...	0	1	0	0	0	0	0	0
2	3	Grumpier Old Men (1995)	Comedy|Romance	0	1	0	0	0	0	0	...	0	0	0	0	1	0	0	0
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance	0	1	0	1	0	0	0	...	0	0	0	0	1	0	0	0
4	5	Father of the Bride Part II (1995)	Comedy	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0

Figure No. 7.2: One hot encoding of Genre data

The list of unique Genres is given below:

- Adventure
- Comedy
- Action
- Drama
- Crime
- Children
- Mystery
- Animation
- Documentary
- Thriller
- Horror
- Fantasy
- Western

- Film-Noir
- Romance
- Sci-Fi
- Musical
- War

The dataset is cleaned to drop those movies which do not have any genre listed. They are specified by “no genres listed”. Thus, the dataset contains 9708 movies with one hot encoding of 18 genres.

Model building – k-means: The already implemented k-means algorithm from the well-known library of sci-kit learn which is used in data science and AI research, is used to build the model.

The number of clusters is given as input to build the k-means model. The distance measure used in k-means for this implementation is Euclidean distance as the genre data is converted to numeric using one hot encoding. k-means fit method from sci-kit learn library is used to activate the unsupervised learning algorithm to find patterns and form clusters of movies.

The number of clusters is determined by elbow method by running multiple iterations of k-means model as shown in Figure No. 7.3:

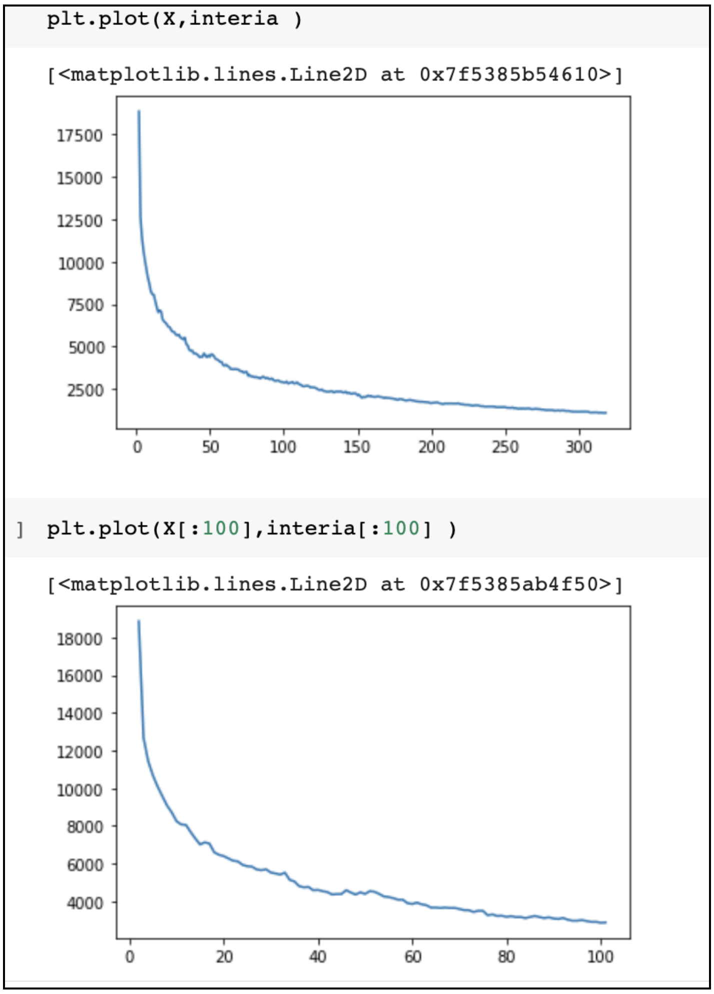

Figure No. 7.3: Elbow diagram for k-means model

The optimal number of number of clusters for k-means model is fixed at 40 from the elbow diagram.

Surrogate model building – Decision tree: The type of decision tree classifier implemented in sci-kit learn is an optimized version of the CART algorithm. It supports only numeric data and the processed dataset consists of only numeric data. The clustering using k-means will result in cluster assignments to each row of the dataframe. These cluster assignments are the target for the Decision Tree Classifier. The Classifier when used to classify builds a decision tree where the leaf nodes are the k-means cluster and the non-leaf nodes explain the impact of the features on the cluster assignment. The accuracy of decision tree provides the percentage of explainability of the clustering. The accuracy of the classifier is given by metrics module of sci-kit learn.

The depth of the decision tree to be built is fixed at 4 which results in 16 leaves. This gives an accuracy of 0.59 for explanations of k-means cluster assignments.

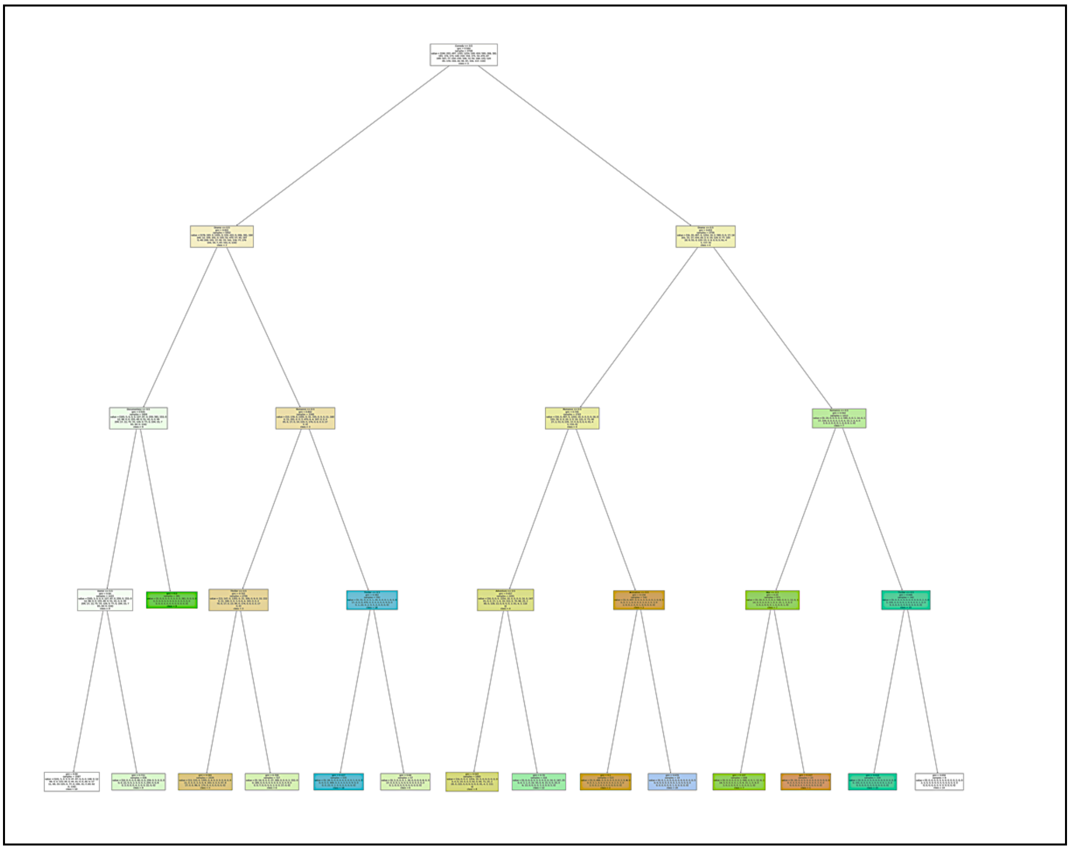

Figure No. 7.4: Decision tree for explanation

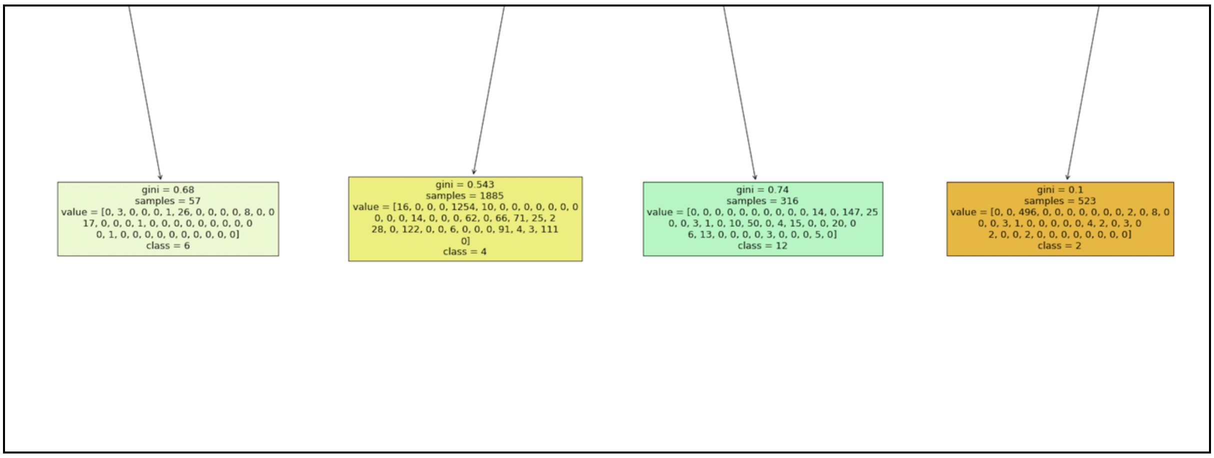

Figure No. 7.5: Decision tree nodes of surrogate model

Build Recommender system: Content based recommender system is built by using Genre data. K-means clustering creates clusters of similar movies based on Genre. Decision tree surrogate model provides explanation of why a particular movie was in a particular cluster.

```
#Random shuffle movies and show 10 movies
import random
random_list = random.sample(range(0, 9708), 10)
i = 1
for x in random_list:
  print(str(i)+".  " + movies_data_clean.iloc[x,1])
  i = i+1

1.  Daddy Long Legs (1919)
2.  Pat Garrett and Billy the Kid (1973)
3.  Golden Compass, The (2007)
4.  Fort Tilden (2014)
5.  Woman Under the Influence, A (1974)
6.  Justice League: Crisis on Two Earths (2010)
7.  Walk Hard: The Dewey Cox Story (2007)
8.  Naked (1993)
9.  Knock Knock (2015)
10.  Ferngully: The Last Rainforest (1992)

sel_mov = input("Select any movie: ")

Select any movie: 3
```

Figure No. 7.6: Movie recommender system

Random package is used randomly select 10 movies from the MovieLens dataset. The user can select any of those 10 movies to watch.

Recommendation technique: The cluster to which the user selected movie belongs is obtained. The cosine similarity between the centroid vectors for that cluster and all other clusters are computed as shown in Figure No. 8.5 using scipy library. The cluster with minimum cosine similarity is used for recommendation. K-means cluster with the novel cluster is obtained. Random is invoked to select 10 movies from the novel cluster.

Multi -modal explanation: Jack in the box selects one movie out of the 10 recommended movies. The selected movie is found in the decision tree which is used as surrogate model. The decision tree path to the leaf node of that movie is extracted. The rules indicate the reasoning behind cluster assignment. This information is used by Jack in the box to provide reasoning for the novel and unexpected movie being recommended. Apart from the textual explanations an audio explanation is implemented to bring in novelty to this paper. Google text to speech library is used to provide audio explanations.

The recommendations are also available as csv file with < Movie, Recommendation, Reason> which can be used to build out different interfaces to display the recommendation across different strategies.

VI. RESULTS AND FUTURE SCOPE

Selection of the movie “Three Burials of Melquiades Estrada” to simulate watch, leads to following recommendations and explanations as shown in Figure No. 8.1.

	Movie Title	Serendipitous Genre	Explanation of already seen
0	Seabiscuit (2003)	['Drama']	['Comedy', 'Romance', 'Thriller']
1	Washington Square (1997)	['Drama']	['Comedy', 'Romance', 'Thriller']
2	My Dinner with Andr√© (1981)	['Drama']	['Comedy', 'Romance', 'Thriller']
3	All the King's Men (2006)	['Drama']	['Comedy', 'Romance', 'Thriller']
4	Mouchette (1967)	['Drama']	['Comedy', 'Romance', 'Thriller']
5	Three Colors: Red (Trois couleurs: Rouge) (1994)	['Drama']	['Comedy', 'Romance', 'Thriller']
6	Eddie and the Cruisers (1983)	['Drama']	['Comedy', 'Romance', 'Thriller']
7	Man Who Planted Trees, The (Homme qui plantait des arbres, L') (1987)	['Drama']	['Comedy', 'Romance', 'Thriller']
8	Talent for the Game (1991)	['Drama']	['Comedy', 'Romance', 'Thriller']
9	Jump In! (2007)	['Comedy', 'Drama', 'Romance']	['Thriller']
10	Mo' Better Blues (1990)	['Drama']	['Comedy', 'Romance', 'Thriller']
11	Mulholland Dr. (1999)	['Drama', 'Romance']	['Comedy', 'Thriller']
12	Company, The (2003)	['Drama']	['Comedy', 'Romance', 'Thriller']
13	Jet Li's Fearless (Huo Yuan Jia) (2006)	['Drama']	['Comedy', 'Romance', 'Thriller']
14	On a Clear Day (2005)	['Drama']	['Comedy', 'Romance', 'Thriller']
15	One True Thing (1998)	['Drama']	['Comedy', 'Romance', 'Thriller']
16	World Traveler (2001)	['Drama']	['Comedy', 'Romance', 'Thriller']
17	To Each His Own (1946)	['Drama']	['Comedy', 'Romance', 'Thriller']
18	Ordinary People (1980)	['Drama']	['Comedy', 'Romance', 'Thriller']
19	Jersey Boys (2014)	['Drama']	['Comedy', 'Romance', 'Thriller']

Figure No. 8.1: Explanations for serendipitous recommendations

Analysis: The movie watched belongs to the Genre “Adventure and Crime” and the movies recommended are in the Genre “Drama” for exploration. There are two movies in Drama and Romance category. This shows the serendipitous nature of recommender system which appropriate explanations.

Movie posters communicate the most of essence of the movie and here are the posters for the watched movie and the recommended movies:

Movie Watched:

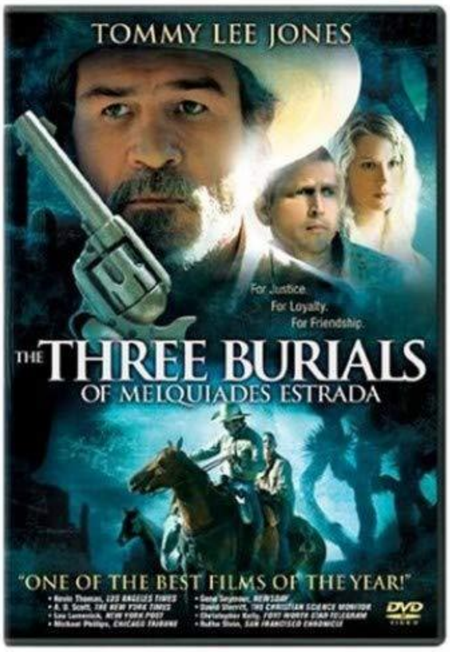

Figure No. 8.2: Movie poster

Movies Recommended for watching:

Figure No. 8.3. Recommended movies

Explanation for Recommendation says these movies were recommended as it wanted the user to explore Drama different from Comedy, Romance and Thriller.

Here is another run of the recommender system where some movies are shown randomly to simulate watch as shown in Figure No. 8.4. The user selects a movie called “Wow! A Talking Fish” to watch.

```
1.  Stripes (1981)
2.  Parasite (1982)
3.  Robin Hood: Prince of Thieves (1991)
4.  Helvetica (2007)
5.  Cabaret (1972)
6.  Wow! A Talking Fish! (1983)
7.  Dunston Checks In (1996)
8.  Sabrina (1954)
9.  Little Boxes (2017)
10.  Our Town (1940)
```

```python
sel_mov = input("Select any movie: ")
```

```
Select any movie: 6
```

Figure No. 8.4: Selection of movies to simulate watch

The serendipitous recommendations and explanations as shown in Figure No. 8.5 are provided to the user.

	Movie Title	Serendipitous Genre	Explanation of already seen movies
0	Cruise, The (1998)	['Documentary']	['Comedy', 'Drama', 'Musical']
1	Mortdecai (2015)	['Comedy', 'Romance']	['Drama', 'Action']
2	I Am Trying to Break Your Heart (2002)	['Documentary']	['Comedy', 'Drama', 'Musical']
3	GLOW: The Story of the Gorgeous Ladies of Wres	['Documentary']	['Comedy', 'Drama', 'Musical']
4	Hoop Dreams (1994)	['Documentary']	['Comedy', 'Drama', 'Musical']
5	Kid Stays in the Picture, The (2002)	['Documentary']	['Comedy', 'Drama', 'Musical']
6	28 Up (1985)	['Documentary']	['Comedy', 'Drama', 'Musical']
7	My Architect: A Son's Journey (2003)	['Documentary']	['Comedy', 'Drama', 'Musical']
8	Last Days, The (1998)	['Documentary']	['Comedy', 'Drama', 'Musical']
9	When We First Met (2018)	['Comedy']	['Drama', 'Romance', 'Action']
10	Rock School (2005)	['Documentary']	['Comedy', 'Drama', 'Musical']
11	Night and Fog (Nuit et brouillard) (1955)	['Documentary']	['Comedy', 'Drama', 'Musical']
12	Blackfish (2013)	['Documentary']	['Comedy', 'Drama', 'Musical']
13	49 Up (2005)	['Documentary']	['Comedy', 'Drama', 'Musical']
14	Wild Parrots of Telegraph Hill, The (2003)	['Documentary']	['Comedy', 'Drama', 'Musical']
15	Human Planet (2011)	['Documentary']	['Comedy', 'Drama', 'Musical']
16	Iron Man (1931)	['Drama']	['Comedy', 'Romance', 'Thriller']
17	Eyes of Tammy Faye, The (2000)	['Documentary']	['Comedy', 'Drama', 'Musical']
18	The Thinning (2016)	[]	['Comedy', 'Drama', 'Documentary', 'Horror']
19	Life and Debt (2001)	['Documentary']	['Comedy', 'Drama', 'Musical']

Figure No. 8.5: Recommendations of movies with explanation

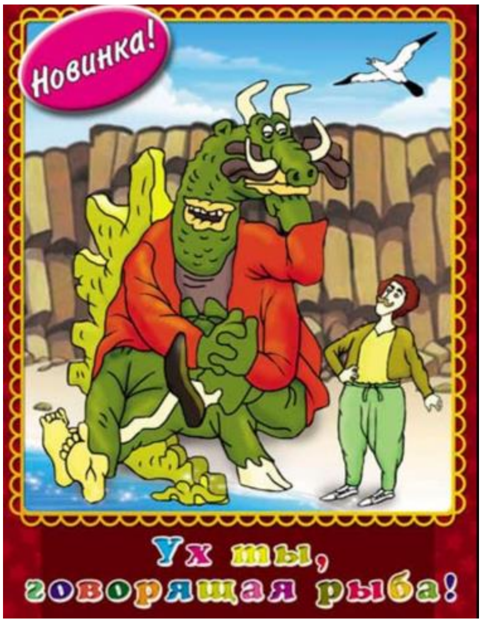

Figure No. 8.6: Movie poster of “Wow! A talking fish!”

Figure No. 8.7: Movies that were recommended

Here is the audio file with personification as Jack in the Box which provides audio explanation:
When played, the audio explanation interface utters the following sentence which has the movie recommended and the explanation why that movie was recommended”

” Hi, I am Jack in the box , The movie recommended for you is The Wild Parrots of Telegraph Hill which is a Documentary movie. I recommended this movie as you have seen Comedy, Drama, Thriller”

The recommender systems with serendipitous recommendations is successfully built with audio explanations. Interactive Demo is created in Jupyter notebook. In future we can explore collaborative filtering using k-means and provide majority voting ensemble recommender. Also, other strategies for serendipity can be explored.

VII. BIBLIOGRAPHY

1. Harper, F. M., & Konstan, J. A. (2015). The MovieLens datasets: History and context. Acm transactions on interactive intelligent systems (tiis), 5(4), 1-19.
2. Ahuja, R., Solanki, A., & Nayyar, A. (2019, January). Movie recommender system using K-Means clustering and K-Nearest Neighbor. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 263-268). IEEE.
3. Rahul, M., Pal, P., Yadav, V., Dellwar, D. K., & Singh, S. (2021). Impact of similarity measures in K-means clustering method used in movie recommender systems. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012101). IOP Publishing.
4. Zarzour, H., Maazouzi, F., Al–Zinati, M., Jararweh, Y., & Baker, T. (2021, December). An Efficient Recommender System Based on Collaborative Filtering Recommendation and Cluster Ensemble. In 2021 Eighth International Conference on Social Network Analysis, Management and Security (SNAMS) (pp. 01-06). IEEE.
5. Wikipedia contributors. (2022, June 26). Netflix Prize. In Wikipedia, The Free Encyclopedia. Retrieved 09:30, July 2, 2022, from https://en.wikipedia.org/w/index.php?title=Netflix_Prize&oldid=1095044995
6. Tintarev, N., & Masthoff, J. (2007, April). A survey of explanations in recommender systems. In 2007 IEEE 23rd international conference on data engineering workshop (pp. 801-810). IEEE.
7. Tintarev, N., & Masthoff, J. (2011). Designing and evaluating explanations for recommender systems. In Recommender systems handbook (pp. 479-510). Springer, Boston, MA.
8. Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In Recommender systems handbook (pp. 353-382). Springer, Boston, MA.
9. Dasgupta, S., Frost, N., Moshkovitz, M., & Rashtchian, C. (2020). Explainable k-means and k-medians clustering. arXiv preprint arXiv:2002.12538.
10. Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58.
11. Melville, P., & Sindhwani, V. (2010). Recommender systems. Encyclopedia of machine learning, 1, 829-838.
12. Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008, September). Introducing serendipity in a content-based recommender system. In 2008 eighth international conference on hybrid intelligent systems (pp. 168-173). IEEE.
13. Onuma, K., Tong, H., & Faloutsos, C. (2009, June). Tangent: a novel,'surprise me', recommendation algorithm. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 657-666).
14. Kotkov, D., Veijalainen, J., & Wang, S. (2020). How does serendipity affect diversity in recommender systems? A serendipity-oriented greedy algorithm. Computing, 102(2), 393-411.
15. Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. International Journal of Human-Computer Studies, 72(4), 367-382.
16. Ko SK. et al. (2011) A Smart Movie Recommendation System. In: Smith M.J., Salvendy G. (eds) Human Interface and the Management of Information. Interacting with Information. Human Interface 2011. Lecture Notes in Computer Science, vol 6771. Springer, Berlin, Heidelberg
17. Ko, S. K., Choi, S. M., Eom, H. S., Cha, J. W., Cho, H., Kim, L., & Han, Y. S. (2011, July). A smart movie recommendation system. In Symposium on Human Interface (pp. 558-566). Springer, Berlin, Heidelberg.
18. Kużelewska, Urszula. (2014). Clustering Algorithms in Hybrid Recommender System on MovieLens Data. Studies in Logic, Grammar and Rhetoric. 37. 10.2478/slgr-2014-0021.
19. Molnar, C. Interpretable Machine Learning. Lulu. com, 2019. https://christophm.github.io/interpretable-ml-book/.
20. W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, και B. Yu, ‘Definitions, methods, and applications in interpretable machine learning’, Proceedings of the National Academy of Sciences, τ. 116, τχ. 44, σσ. 22071–22080, Οκτωβρίου 2019.
21. Dash, A., Chakraborty, A., Ghosh, S., Mukherjee, A., & Gummadi, K. P. (2022, April). Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers. In Proceedings of the ACM Web Conference 2022 (pp. 3695-3705).
22. “MovieLens,” MovieLens. https://MovieLens.org/ (accessed Aug. 23, 2022).

23. Leysen, J., Michiels, L., Goethals, B., & Smets, A. (2022). What Are Filter Bubbles Really? A Review of the Conceptual and Empirical Work. In Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization.

24. Cookson, J. A., Engelberg, J., & Mullins, W. (2022). Echo chambers. Available at SSRN 3603107.

25. Barraza-Urbina, A. (2017, August). The exploration-exploitation trade-off in interactive recommender systems. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 431-435).

26. Asanov, D. (2011). Algorithms and methods in recommender systems. Berlin Institute of Technology, Berlin, Germany.

27. Vasile, F., Smirnova, E., & Conneau, A. (2016, September). Meta-prod2vec: Product embeddings using side-information for recommendation. In Proceedings of the 10th ACM conference on recommender systems (pp. 225-232).

28. "KDD 2020 Tutorial: Advances in Recommender Systems," KDD 2020 Tutorial: Advances in Recommender Systems. https://sites.google.com/view/kdd20-marketplace-autorecsys/ (accessed Aug. 24, 2022).

29. "Content-based Filtering Advantages & Disadvantages | Machine Learning | Google Developers," Google Developers. https://developers.google.com/machine-learning/recommendation/content-based/summary (accessed Aug. 24, 2022).

30. Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. User Modeling and User-Adapted Interaction, 29(2), 239-249.

31. Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends® in Information Retrieval, 14(1), 1-101