



REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

REVA Academy for Corporate Excellence (RACE)

Explainable Recommender Systems



**Yashaswini Viswanath,
Dr. Rashmi Agarwal,
Rajanna K
RACE, REVA University.**

01 Introduction

Background | Current status | Why this study

02 Literature Review

Seminal works | Summary | Research Gap

03 Problem Statement

Technical/Functional Problem

04 Project Objectives

Primary & Secondary Objectives | Expected Outcome

05 Project Methodology

Conceptual Framework | Research Design

06 Resources Specification

Software | Hardware

07 Software Design

High Level | Low Level Design Diagrams

08 Implementation

Demo | Application | Use cases

09 Testing and Validation

Test Results | Learnings

10 Analysis and Results

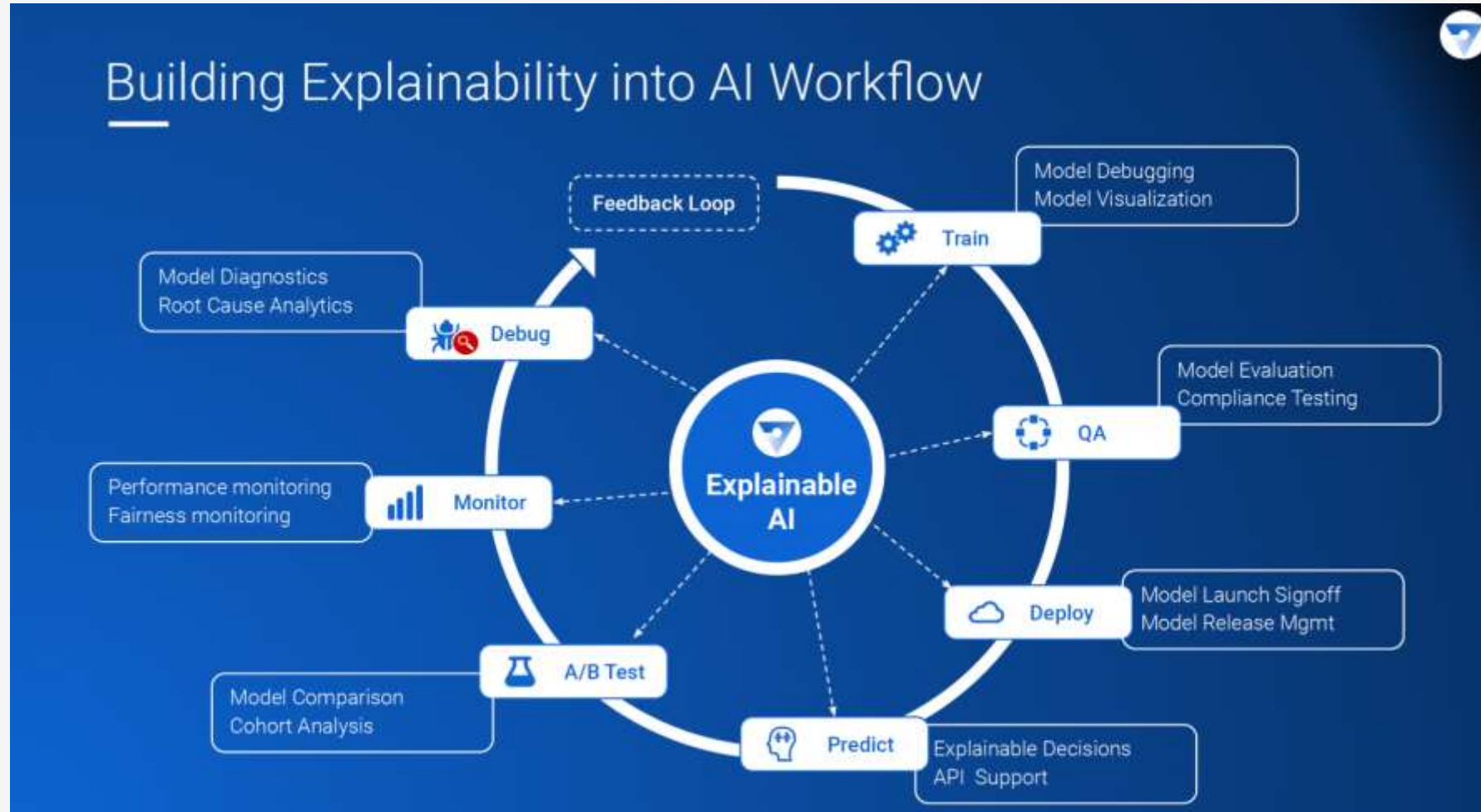
Key Findings | Insights

12 Suggestions and Conclusion

Insights | Next Step | Future Scope

13 Annexure

References | Publications | Plagiarism Score





REVA
UNIVERSITY

Bengaluru, India

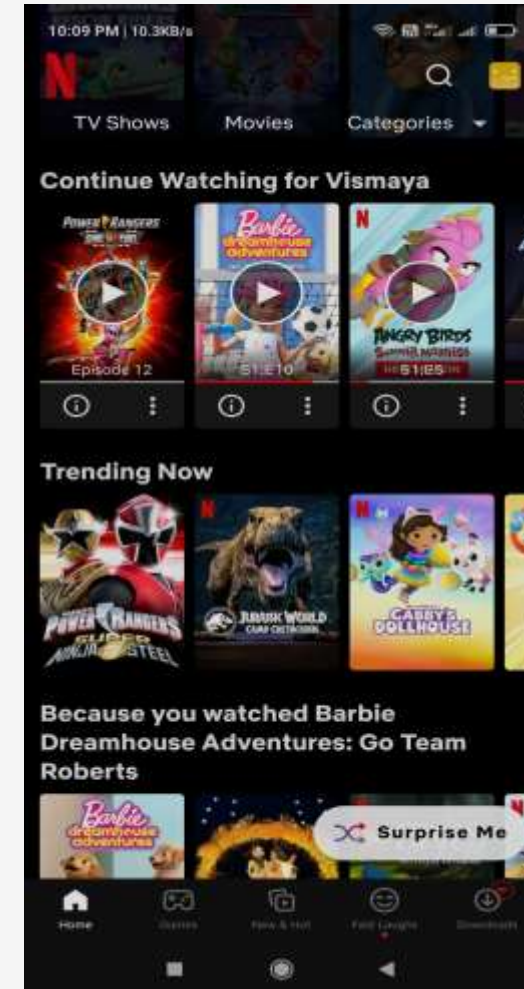
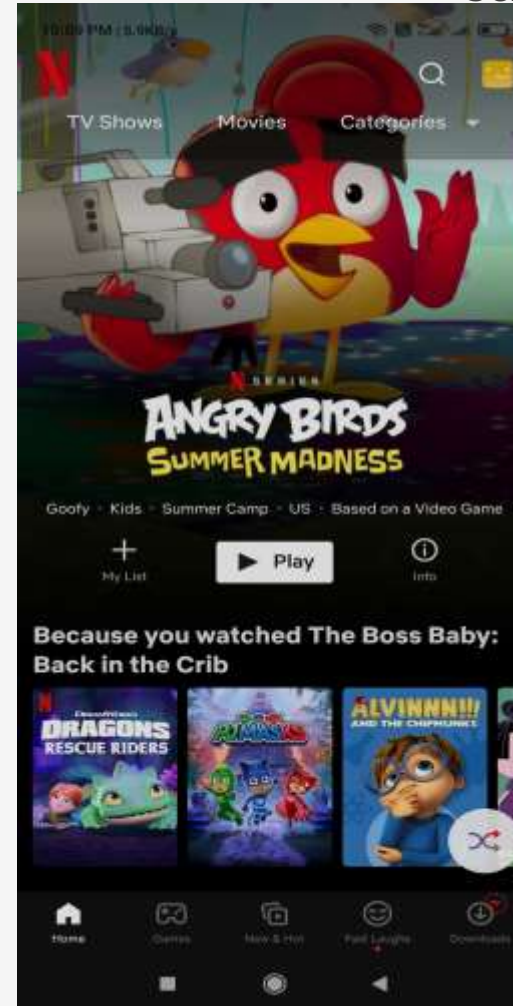
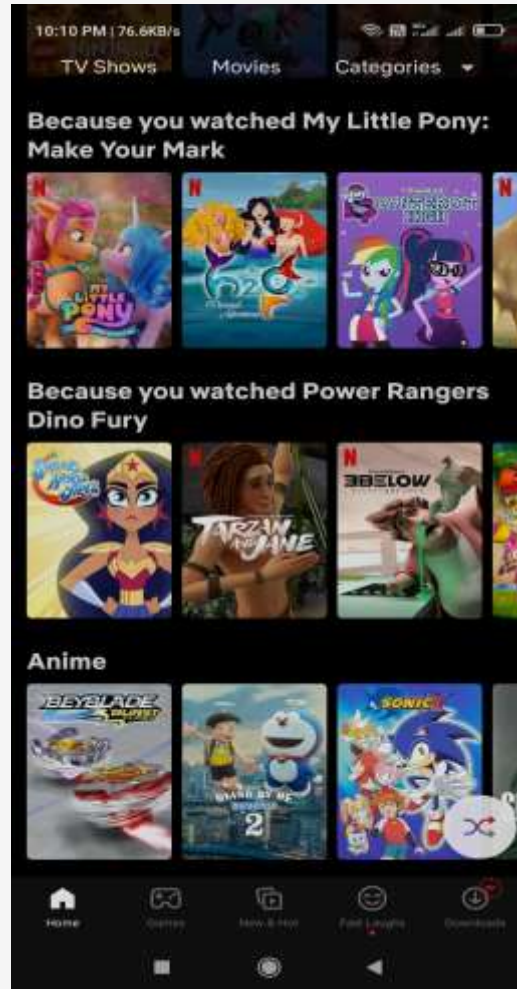
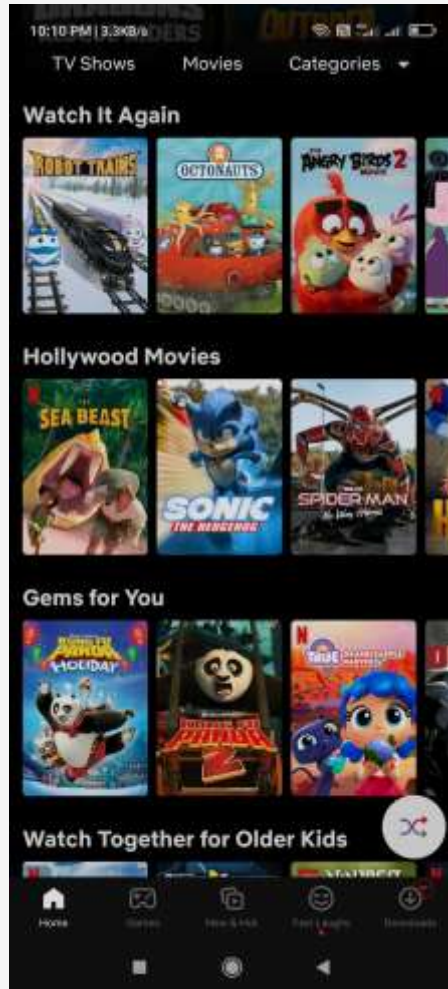
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

Literature Review

Seminal works | Summary | Research Gap





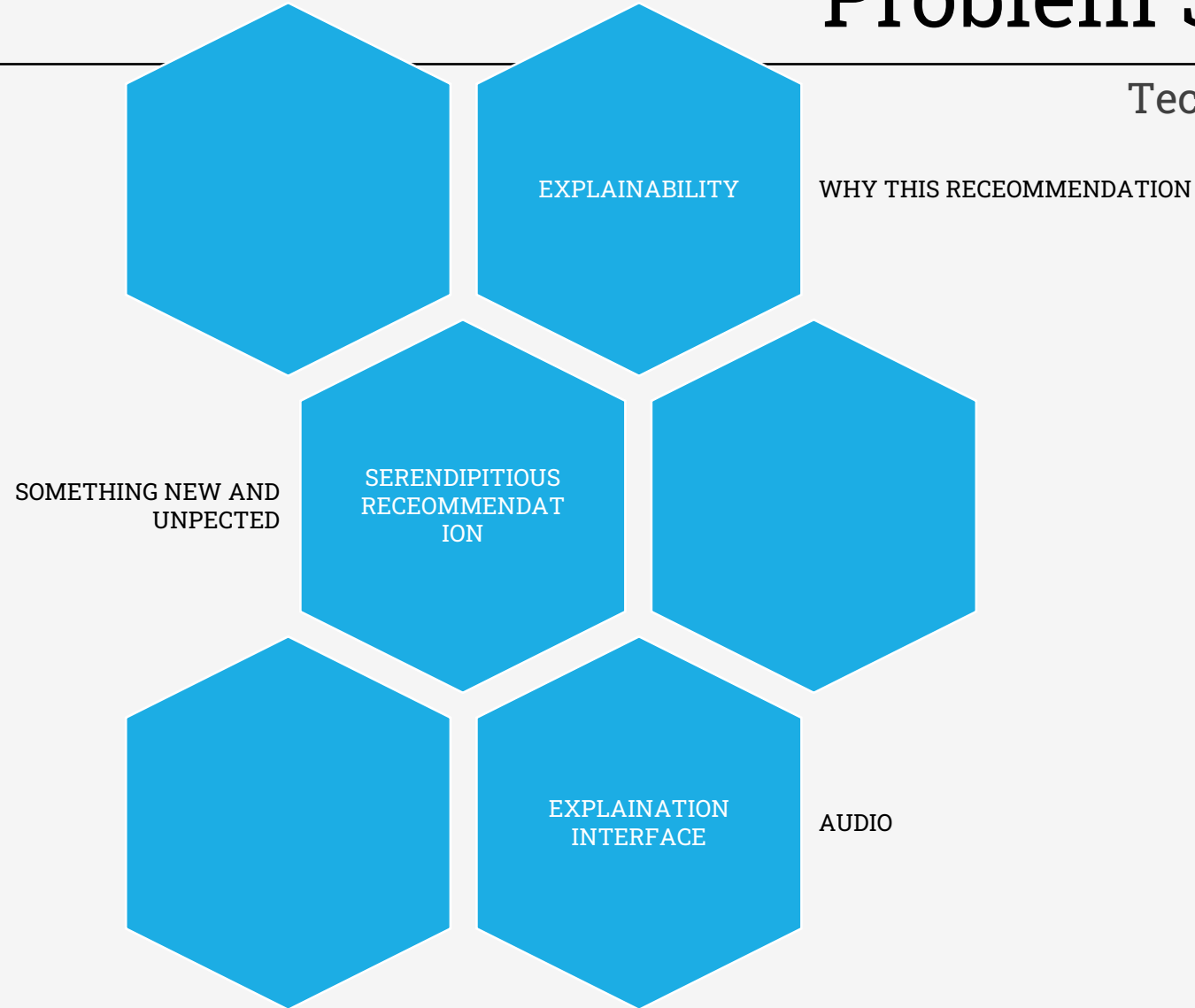
REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Problem Statement

Technical | Functional





REVA
UNIVERSITY

Bengaluru, India


Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

Project Objectives

Primary & Secondary Objectives | Expected Outcome



BUILD A MOVIE RECEOMMEDNER SYSTEM	<ul style="list-style-type: none">• MOVIELENS DATASET• CONTENT BASED RECEOMMENDATIONS
ADD EXPLAINABILITY TO THE RECOMENDATIONS	<ul style="list-style-type: none">• KMEANS CLUSTERING AS MAIN MODEL• SURROGATE MODEL DECISION TREES FOR EXPLANATIONS
MODE OF DEVLIVERY OF EXPLANATIONS	<ul style="list-style-type: none">• TEXT• AUDIO

Project Methodology

Conceptual Framework | Research Design

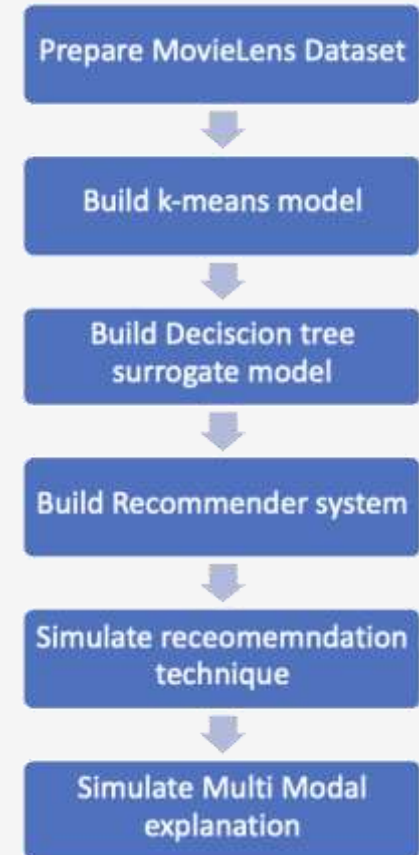
The *MovieLens* dataset needs to be prepared for usage in this project.

Processing the dataset such that the models can find patterns is a crucial step in the project 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.



MovieLens

GroupLens Research has collected and made available rating data sets from the MovieLens web site (<https://movielens.org>). The data sets were collected over various periods of time, depending on the size of the set. Before using these data sets, please review their README files for the usage licenses and other details.

Seeking permission? If you are interested in obtaining permission to use MovieLens datasets, please first read the terms of use that are included in the README file. Then, please [fill out this form](#) to request use. We typically do not permit public redistribution (see [Kaggle](#) for an alternative download location if you are concerned about availability).

recommended for new research

MovieLens 25M Dataset

MovieLens 25M movie ratings. Stable benchmark dataset. 25 million ratings and one million tag applications applied to 62,000 movies by 162,000 users. Includes tag genome data with 15 million relevance scores across 1,129 tags. Released 12/2019

- [README.txt](#)
- [ml-25m.zip](#) (size: 250 MB, [checksum](#))

Permalink: <https://grouplens.org/datasets/movielens/25m/>

MovieLens Tag Genome Dataset 2021

10.5 million computed tag-movie relevance scores from a pool of 1,084 tags applied to 9,734 movies. Released 12/2021. This dataset also contains input necessary to generate the tag genome using both the original process (Vig et al. 2012) and a more recent improvement (Kotkov et al. 2021)

- [genome_2021_readme.txt](#)
- [genome_2021.zip](#) (size: 1.8GB)

Permalink: <https://grouplens.org/datasets/movielens/tag-genome-2021>

recommended for education and development

Resource Specifications

Software | Hardware | Others

```
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	



REVA
UNIVERSITY

Bengaluru, India

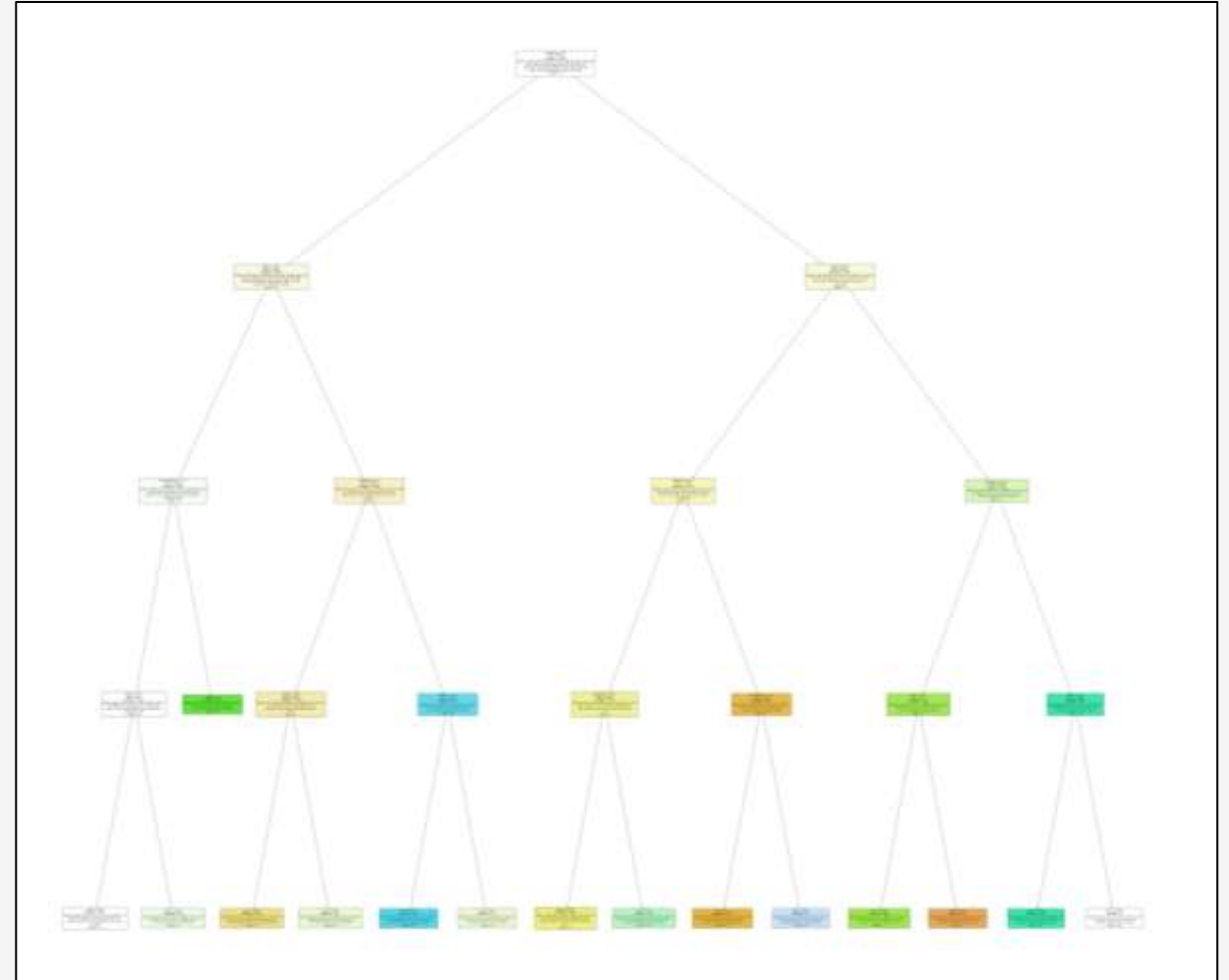
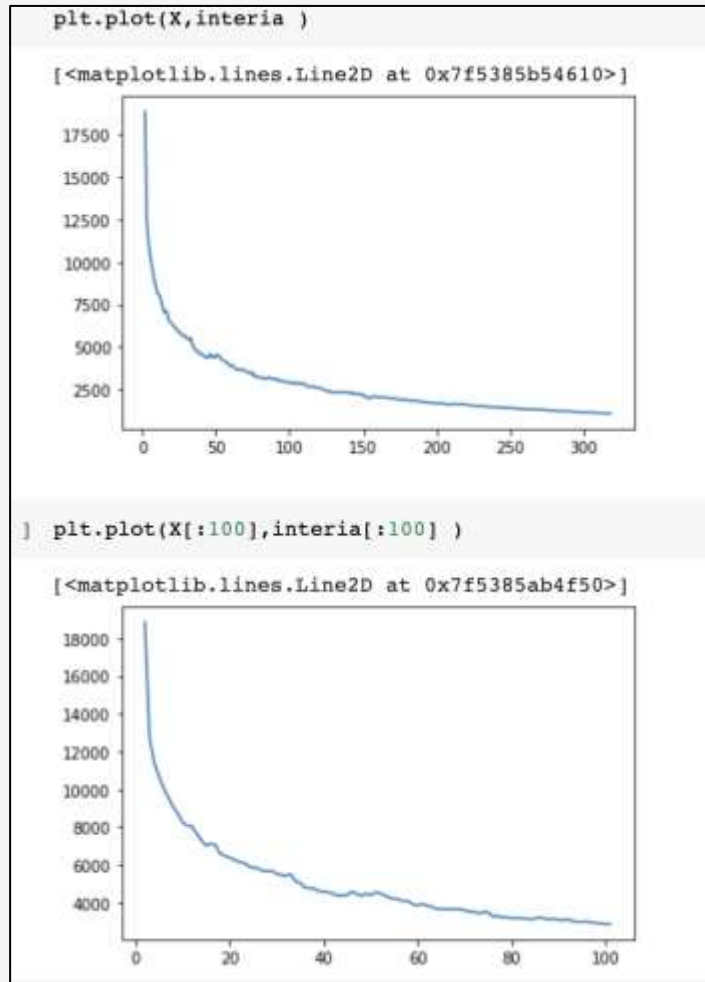
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

Implementation

Demo | Application | Use cases



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)

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

Select any movie: 6



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 Wrestling (2019)	['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']



Results

Key Findings | Insights

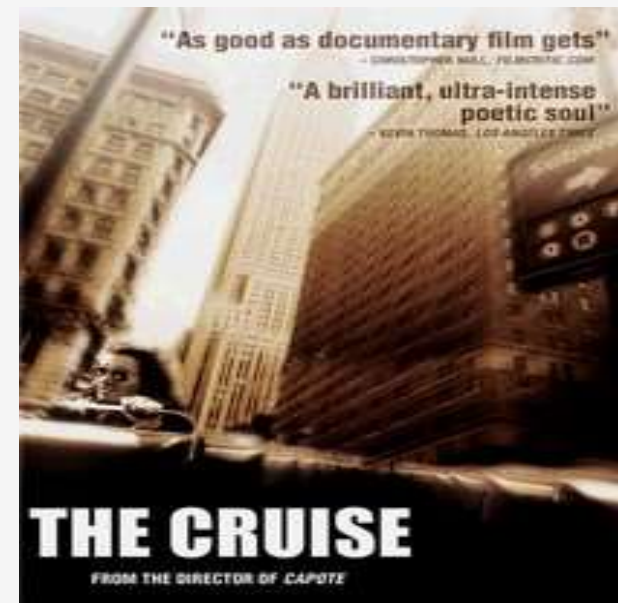
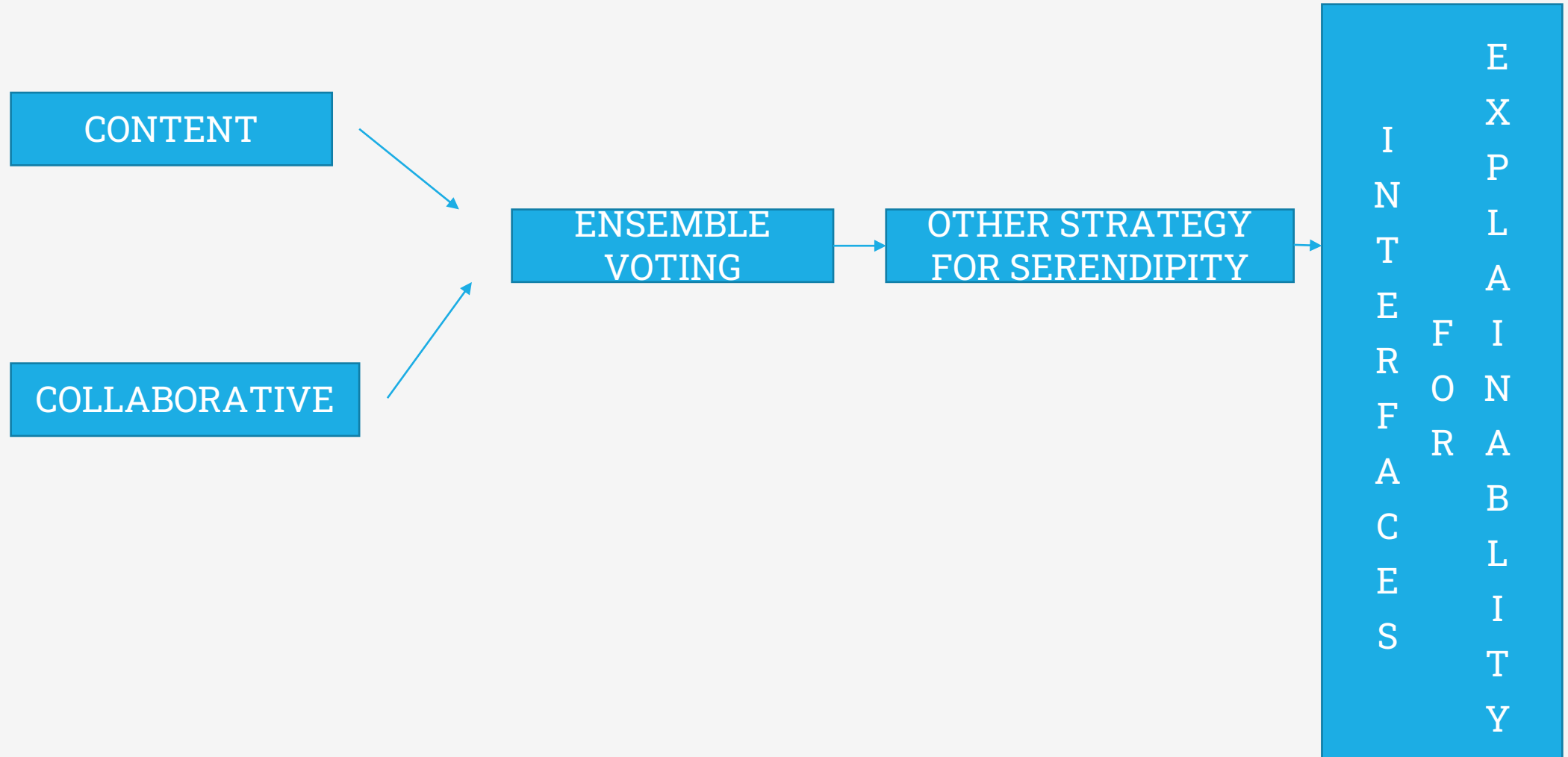


Figure No. 8.7: Movies that were recommended

Suggestions and Conclusion

Insights | Next Step | Future Scope



References

Bibliography | Webliography

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.

References

Bibliography | Webliography

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., & Sindhvani, 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.

References

Bibliography | Webliography

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. Kuźelewska, 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).

References

Bibliography | Webliography

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* (p. 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.



REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

Annexure

Additional Information | Plagiarism score

Explaining Clustering using Decision Trees

ORIGINALITY REPORT

5%	5%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ukcatalogue.oup.com Internet Source	2%
2	onezero.blog Internet Source	1%
3	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
4	Y. Brandman, A. Orlitsky, J. Hennessy. "A spectral lower bound technique for the size of decision trees and two-level AND/OR circuits", IEEE Transactions on Computers, 1990 Publication	<1%
5	www.ccis2k.org Internet Source	<1%
6	www.coursehero.com Internet Source	<1%
7	www.traumacenters.org Internet Source	<1%



REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi



*Thank
you!*

