RECOMMENDER SYSTEMS AND DATA MINING: A COMPREHENSIVE STUDY

ABSTRACT

Recommenders systems are extensively used on the web to deliver top-notch user experience. The purpose of this research is to analyze the concepts, methodologies and techniques that are central working block of every Recommender System. In this paper I thoroughly discuss current state of the art of recommender systems and how they collect, store and process user data to generate relevant recommendations. It begins with detailed overview of history of Recommender Systems, current and future benchmarks, processes connected with it and its evolution. Recommender systems are, after all, utilized by merchants to increase their profit hence its important to understand how it contributes towards the business-centric goal of increased revenue. A recommender system is a closely bounded intersection between the domains of information retrieval, machine learning, optimization, customer centricity and behavioural economics. Companies use customer data to predict and infer preferences of an individual based on their micro-interactions, search history and feedback and put forward relevant recommendations for them. However, it is also important to understand the ubiquity of data gathering, storage, and analytics that goes into proceeding input of a recommender system.

Table of Contents

CHAPTER ONE : RECOMMENDER SYSTEM

CHAPTER TWO : METHODOLOGY

CHAPTER THREE : CASE STUDY; AMAZON REVIEWS

CHAPTER FOUR : DATA MINING IN SOCIAL CONTEXT

CHAPTER FIVE : CONCLUSION

CHAPTER 1 :

RECOMMENDER

SYSTEM

1.1 INTRODUCTION

As the World Wide Web (WWW) progressed, so did the activity of electronically buying or selling of products on online services or over the Internet, which made consumer data a crucial resource when it comes to driving businesses towards their respective goals. In 2017, The Economist published an article titled "The world's most valuable resource is no longer oil, but data.” [1] which raised reasonable concerns about tech companies and how they exploit all the information they fetch from their users. Data is being stored, managed and utilised at each and every level of internet transaction. E-commerce, banking systems, social apps, health care portals even general purpose management systems fetch as much data as they can in order to develop and manage a dedicated consumer base.

The main purpose of a Recommender System at its very basics is its ability to use various types of user-preference and user-requirement data to generate recommendations. The notion of ‘recommendations’ have grown immensely since legacy platforms like GroupLens, Netflix and Amazon put forward strong commencing foundation for user indulgence and its tremendous effects on their brand value. The most well-known methods of recommendations are collaborative filtering, content-based and knowledge-based recommendations. These three methods form the fundamental pillars of research in recommender systems.

When a recommender systems are deployed in collaboration with Data Mining, data scientists are left with meta-data that is immaculately predictive and results in shaping digital footprints of any user. Personality can be defined as a set of characteristics possessed by a person that uniquely influences his or her cognitions, emotions, motivations, and behaviours in various situations. That is, personality is a critical factor which influences how people make their decisions [2]

Psycho-analysis, a systematic structure of theories concerning the relation of conscious and unconscious psychological processes of people and their behaviour opens the unlimited possibilities of subtle, ultra-fine insights of a user’s mind.

In hindsight, such methods also casts doubts towards legitimacy of people’s privacy rights of their Personally Identifiable Information (PII) [3] on the cyber space. The ambiguity of data collection practices, companies’ refusal to handover the data they have collected in the past and the inevitable end of anonymity on the indexed surface-web indicates the need to educate the public on current practices upheld by data warehouses.

1.2 STATEMENT OF THE PROBLEM

This paper comprises of an in-depth discussion on Recommender System, how it was initiated, how it got evolved with time and respective algorithms and techniques, which work alongside.

Users and their meta-data are crucial resources for businesses and now is the need to understand and validate the importance of Recommender Systems when it comes to computational advertising, the ever growing field which incorporates more 80% of Google’s revenue. [4]

The case study models a generic recommender system which operates on user feedback in terms of text reviews, transforms it into users-specific metadata, models and processes it to generate user-specific recommendations consequently.

The primary goal of any recommendation system is to increase user interest but it is often achieved in ways that are often unconventional and less obvious when visible to the naked eye. Hence one section of this paper discusses the dynamics of User Profiling and Data

Exploit and another discusses societal consequences of such practices. Lastly, this paper disputes over data, its usage and possible effects of data exploitation and how it will reshape the future of computing and information sciences.

1.3 SCOPE OF STUDY

This study was initiated to recognizing the well-acknowledged need of recommender systems. Unlike 90s they are not only limited to only large enterprises. Electronic Commerce (E-commerce) has taken a large chunk of space on internet with almost every business going ‘online.’ Post-Quarantine e-commerce emergence clearly divulges the fact that online market is not going to disappear any time soon. Digital customers have become a major stakeholders of every business. This further reduces the friction between customers and a marketplace and now is the time to generalize the concept of ‘computed recommendations.’ which can help not only small-scale startups increase their brand value but also spread awareness about the infrastructure and methodologies for further research and improvements in this domain.

1.4 NEEDS AND SIGNIFICANCE

Market segmentation

Clustering is a technique used to group people having similar interests. Those distinct groups are referred to as ‘clusters.’ Organisations use clustering to segment users into different groups based on their purchase history and user behaviour. This helps in targeting the correct consumers with correct content that is only relevant to them.

Quality user experience

Recommending only relevant content to users can immensely multiply website engagement and user interest. They are more likely to spent time on content which they like rather than explicitly searching for keywords. Sometimes even users are not aware what comprises their interests. This is where machine learning becomes functional where computers assess and process their interests and generates predictions accordingly.

Market basket analysis

Association is another technique which makes use of customer's purchasing pattern and determines links between their preferences and possibly related items that occur together frequently in transactions. This gives them more room to explore their interests and increases company's revenue subsequently.

Customer Retention

Customer retention is mostly concerned with customer churn and is one of the most important evaluation metrics of a growing business. When a company 'knows' its customers both personally and psychologically, it can eventually retain them for a longer periods of time. Data mining techniques combined with relevant recommendation algorithms may help a company 'churn' out its customers.

Targeted Marketing

When a product or a service is marketed to a specific subset of total addressable market, it's called Targeted Marketing. Such subsets are identified and decomposed by using users' meta-data. Recommendation algorithms are being developed, deployed and integrated rapidly in retail industry for targeted marketing.

1.5 USE CASES

Recommender Systems have evolved ever since Amazon introduced item-to-item collaborative filtering as early as in 1998. Now, their applications are unlimited and unbounded.

E-commerce:

In many services, including e-commerce platforms, recommendations are present at only one stage of the transaction journey, but Amazon has integrated recommendations at every step to maximize order value. Amazon has never revealed how much revenue it generates but McKinsey estimates that 35 percent of what consumers purchase on Amazon comes from product recommendations. [6]

Social Networking Sites (SNS):

Facebook put forward the blueprint for social recommender system with its friends recommendation algorithm which aims to connects millions of people across the globe. Facebook, not only mines explicit user data that is their friends list but also their personal data to determine one's friends network.

LinkedIn, an employment-oriented online service presents another example of reciprocal recommender system that links people on the basis of their education, work history and interests.

Healthcare:

Healthcare bodies implement Virtual Health Care Assistants for their patients to provide them with all kinds of health services. They model users' health conditions by exploiting their search history and the kind of topics they spend most time looking at. With proper extraction and mining of patient data, virtual healthcare systems can successfully predict the illness, possible diseases and concerns about the patients and prescribe them similar topics to consume.

Banking Systems:

Banks avail user data for fraud identification, advanced customer experience, robust buyer personas, intelligent customer engagement and optimized transaction processing.

Entertainment:

Discover Weekly, Spotify's AI powered recommendation system works by looking at playlists of every Spotify user, collates this information and releases a "Discover Weekly" playlist on every monday: It uses collaborative filtering to suggest songs that one user ha not heard yet, but are similar to individual's music tastes.

Youtube's streaming service is known for its aggressive recommendation engine which tends to suggest similar videos based on the content viewers have previously liked or spent more time watching in the past.

Computational Advertisment:

Computational Advertisement (CA) refers to methodologies that perform contextually targeted advertising. It is a widely implemented scientific discipline, at the intersection of large scale search and text analysis, information retrieval, statistical modeling, machine learning, optimization, and microeconomics. The context includes query, web page content, user and geolocation information. Search engines have millions of ads registered with them so fetching the top most relevant ads by ranking them based on some criteria is a more feasible approach than random some adverts. Click-through rate and dwell time is monitored and users are targeted with ads that generate positive feedback and highest bidding.

1.6 JUSTIFICATION

In early 2000s Target systems anticipated pregnancies even before mothers did. [7]

"We knew that if we could identify them in their second trimester, there's a good chance we could capture them for years," - Target

Netflix executives Carlos A. Gomez-Uribe and Neil Hunt while discussing business value of Netflix Recommender System wrote in their paper: [8]

“We think the combined effect of personalization and recommendations save us more than $1B per year”

Greg Linden, Brent Smith, and Jeremy York of **Amazon** wrote in their paper [9]

“At Amazon.com, we use recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother”

Spotify’s Discover Weekly playlist also known as Release Radar, which updates personal playlists on a weekly basis so that users won’t miss newly released music by artists they like, increased its number of monthly users from 75 million to 100 million at one time. [10].

Pep Worx, **PepsiCo**'s proprietary big data and analytics platform was able to distinguish million of unique households from its base dataset of 110 million US households that would be most likely to be interested in Quaker Overnight Oats. The company then identified specific retailers that these households might shop at. Ultimately, these customers drove 80 percent of the product’s sales growth in its first 12 months after launch. [11]

CHAPTER TWO : METHODOLOGY

2.1 BACKGROUND

The first ever manifestation of recommendation system was done by Duke University in the second half of the 1970s, “Usenet communication system” which shared textual content of users and categorised them into labelled newsgroups and subgroups of users.

Earliest working solution of preference-targeted recommendations was Grundy, a computer librarian which first interviewed users about their preferences and then recommended books to them accordingly. Allocating users into subgroups or clusters thus recommending the same books to all people in the same group.

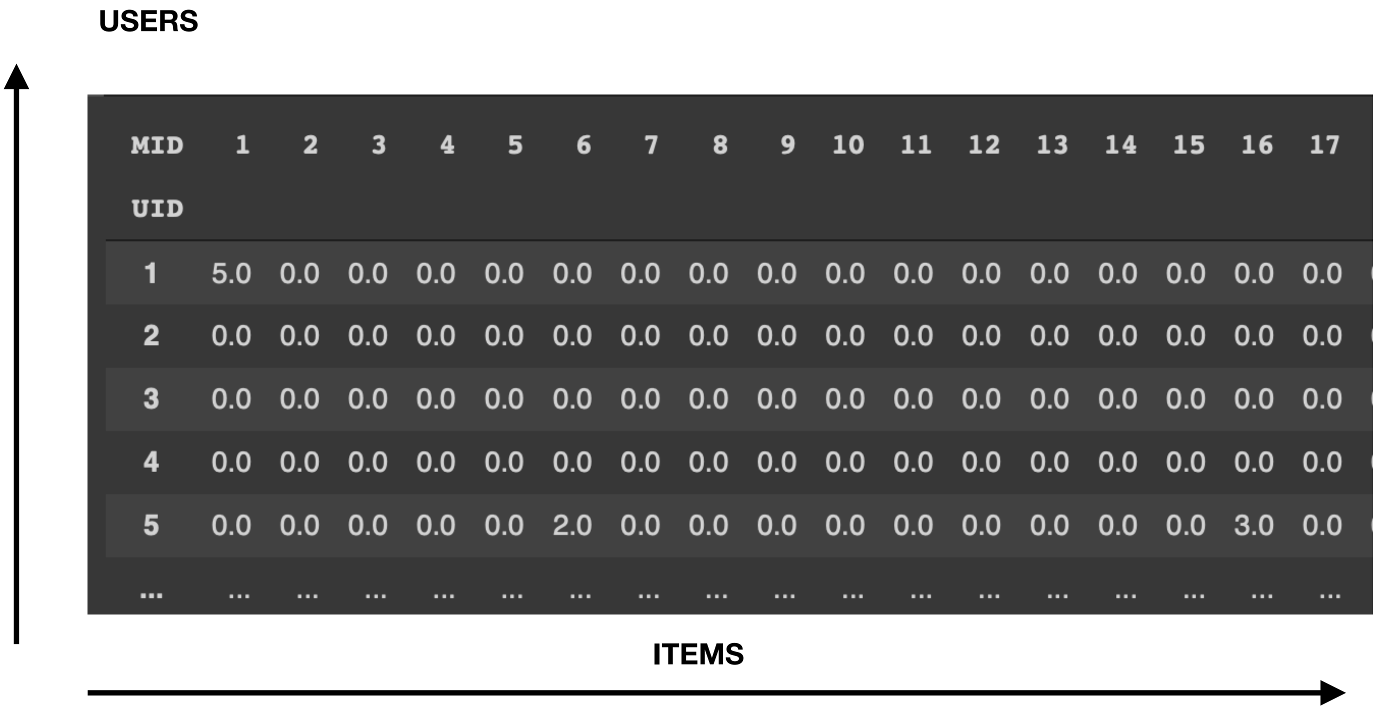
GroupLens, which started in 1992, was a pioneering recommender system which introduced the concept of ‘ratings’ and using it as a heuristic in order to make automated recommendations for Usenet articles. It draws on a deceptively simple idea: people who agreed in their subjective evaluation of past articles are likely to agree again in the future. [12]

A networked system called Ringo, developed at MIT which made personalized recommendations for music albums and artists by using 'Social Information filtering’ which automated the process of "word-of-mouth" recommendations: based upon values assigned by other people with similar taste. [13]

2.2 EARLY RECOMMENDATION TECHNIQUES

2.2.1 Collaborative Filtering

The basic idea of collaborative filtering is that it help people make choices based on the opinions of other people [14]. Since observed ratings are often highly correlated across various users and items, collaborative filtering models use the collaborative power of the ratings provided by multiple users to make recommendations. The sparse matrix with user id's as rows and item id's as columns is the input for collaborative algorithm and similar recommendations are computer and generated respectively as output.



MID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
UID																	
1	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Sparse-matrix of movies rated by viewers

The ratings explicitly specified by users are called 'specified ratings' and the idea is to compute and predict 'unspecified ratings' for users by using highly correlated across various users and items. Most of the models for collaborative filtering focus on leveraging either inter-item correlations or inter-user correlations for the prediction process [15]

Tapestrys, an experimental mail system developed at the Xerox Palo Alto Research Center was the first example of collaborative system that said Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read. [16]

GroupLens was a pioneering recommender system, which was built as a research prototype for recommendation of Usenet news used collaborative filtering to help people find articles they will like in the huge stream of available articles. [17]

Memory-based methods which work by predicting ratings based on ‘neighbourhood’ or ‘closness’ of user-item combinations were among the earliest collaborative filtering algorithm [18]

These neighborhoods can be defined in one of two ways:

- User-based collaborative filtering :

  In this case, recommendations are made for target user through ratings provided by ‘like-minded’ users. Thus, the basic idea is to determine users, who are similar to the target user A, and recommend ratings for the unobserved ratings of target user by computing weighted averages of the ratings of this peer group.

- Item-based collaborative filtering :

Rating predictions for an item is made by segregating ratings specified items of the user in one set S. Now, Similarity between the specified items is computed to discover items similar to that of in set S or to predict his rating on similar/target item.

Model-based methods include decision trees, rule-based models, Bayesian methods and latent factor models which leverage machine learning and data mining methods along with predictive models.

Even though memory-based collaborative filtering algorithms are valued for their simplicity, they tend to be heuristic in nature, and they do not work well in all settings

2.2.2 Content-Based Recommender Systems

When sufficient ratings are not available for the target user, often referred as 'cold-start scenario' descriptive attributes of items are used to make recommendations for the user. The term "content" refers to these descriptions. In content-based methods, the item descriptions, combined with ratings, are used as training data to create a user-specific classification. They often make use of machine learning techniques to extract relevant content from textual descriptions of items and the user profile of preferences is stored as a vector of keywords.

The roots of content-based filtering are to be found in the field of information retrieval. **Information Retrieval** is the process of obtaining relevant information from a collection of resources. 'Semi-automated assistants' or 'recommender agents' systems that monitor users' progress and behaviour over long interaction periods in an attempt to predict which documents or actions the user is likely to want in future.

Taking an example of an online library management system which allows readers to explore different genres of books, starts looking for similarities among the books with regards to specific writers, genres, subject matter which user U has rated highly in the past. Only the books that have a high similarity value to reader's preferences would be recommended to that reader.

Memory-based methods use frequency, inverse document frequency (TF-IDF) text retrieval method.

Model based methods use Decision trees, Neural Networks, Bayesian classifiers, Clustering or vector-based representations.

2.2.3 Context-Based Recommender Systems

Combined with collaborative and content based models, when geolocation or social data is added to the recommendation process, its referred to as context-based filtering. Map assistants, food-delivery applications and online transit applications use similar models within their business model.

2.2.4 Knowledge-Based Recommender Systems

A system that manages central database to store user preferences along with item attributes, also known as ‘Knowledge Base’ in order to later generate recommendations for users is called Knowledge-based recommender system. It leverages from domain knowledge of the user and products to infer correct context of items.

Knowledge-based recommender systems are particularly useful in the context of items that are not purchased very often. Examples include items such as real estate, automobiles, tourism requests, financial services, or expensive luxury goods.

In case based recommender systems the interactivity in knowledge-based recommender systems achieved in many ways:

Conversational systems: An iterative conversational system is used to determine user preferences in the context of a feedback loop.

Search-based systems: Elicitation of user preferences and user constraints are specified by using a preset sequence of questions.

Navigation-based recommendation: The User specifies a number of change requests to the item being currently recommended. Through an iterative set of change requests, it is possible to arrive at a desirable item.

2.2.5 Demographic Recommender Systems

Demographics refers to the description or distribution of characteristics of some target audience, consumer base, or population. In respect of people demographic information includes their age, race, ethnicity, gender, marital status, income, education, and employment.

A demographic recommender system classifies users based on their demographic attributes, e.g. college students, teenagers, women, men, etc. Those classifiers map specific demographics to ratings or buying probabilities/ tendency.

Grundy [19], an early recommender system interviewed users about their preferences and recommended books based on the library of manually assembled stereotypes. Such stereotypes were closely based on demographic attributes of users. The characteristics of the user were collected with the use of an interactive dialogue.

2.3 HYBRID RECOMMENDER SYSTEMS

Hybridization is a technique that combines different types of frameworks together to extract and merge strengths of each to achieve better results. Since different type of Recommender Systems allow different kind of inputs it gives flexibility to apply hybridisation for the same task. Efficient results can be achieve if hybrid models are implemented properly.

Statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better performance than could be obtained from any of the constituent learning algorithms alone. Hybridization closely relates to the field of ensemble analysis and having the power of combining different machine learning algorithms into much more robust model

, hybrid recommender systems are able to improve effectiveness of legacy recommender systems (e.g., collaborative systems)

Netflix's algorithm, CineMatch was the most early example of a commercially successful recommender system using collaborative filtering algorithm technique for online movie sales.

The Netflix Prize was an open competition to improve CineMatch's efficiency to predict user ratings for films, based on previous ratings only.

In order to win the contest, the proposed solution had to achieve 10% RMSE improvement over the trivial algorithm. CineMatch had a similar performance on the test set, 0.9525, so the winner had to lower it down to at least 0.8572 to win prize money.

The team that won a million dollar prize in 2009 proposed a hybrid solution that included 107 different algorithms and mixed their recommendations depending on the circumstances.

Amazon, yet another commercial success of online referral systems today recommends products to the consumer based on cooperative, combined filtering techniques taking into both explicit and implicit user feedback

Methods/Strategies of Hybridization There are different strategies of hybridization and they are broadly classified into seven categories:

Weighted: **Collaborative** and **Content-based** methods are implemented separately and then their results are combined.

Switching: A certain switching criterion is used by the system to interchange between two recommendation systems operating on the same object.

Feature Combination: Features from different recommendation systems' data sources are put into a single recommendation algorithm.

Cascading: One recommendation system refines the results given by another.

Meta Level: A feature such as a model learned by one recommendation is used as input to another. It differs from **Feature Augmentation System,** in that the entire model is used as input.

Feature Augmentation: The output of one system is used as an input feature to another; for example, using the model generated by one to generate features that are used by another

Mixed: Incorporates two or more techniques at the same time; for example: **Content-based** and **Collaborative Filtering.**

CHAPTER 3: CASE STUDY

Amazon Reviews; NLP-Based Recommender System

3.1 BUSINESS TASK

Process and analyze users' reviews in order to predict customer interests and generate relevant item recommendations for them.

3.2 CONCEPTS

3.2.1 NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a collective term referring to automatic computational processing of human languages. It's a sub-domain of Artificial Intelligence that deals with languages and its textual data. Many machine learning models are trained to process huge amount of text and derive relevant information from bulk of raw data in scarce time with prominent efficiency.

Deep Learning, modelled to simulate or replicate human cognitive system of conscious decision making while utilizing complex network of neurons is applied in many systems where the system learns and tracks user behaviour.

An Artificial Neural Network trains itself with every input for improved automation, performing analytical and physical tasks without human intervention.

3.2.2 NATURAL LANGUAGE UNDERSTANDING

Natural-language understanding or natural-language interpretation is a subtopic of natural-language processing that deals with machine reading comprehension and allows computer applications to infer intent from spoken and written language.

3.2.3 DATA MODELING

The process of Exploratory Data Analysis (EDA), feature selection and determining the relationship between different data objects is called data modeling. It derives valuable information from raw data through ETL and transforms it to validate business goals.

3.2.4 CONSUMER BEHAVIOUR

Consumer behaviour is a social science discipline which blends psychology, sociology, social anthropology, anthropology, ethnography, marketing and behavioural economics.

It is the study of all the activities associated with the customer including purchase, use and behaviour psychology. Consumer behaviour consists of how the consumer's emotions, attitudes and preferences affect their buying behaviour.

3.2.5 FEEDBACK

Any response, reaction, comment or opinion of users derived explicitly or implicitly can be defined as feedback.

Types of Feedbacks:

IMPLICIT FEEBACK:

Feedback given by users in terms of buying history, browsing behaviour, search queries and time spent on specific page of interest.

EXPLICIT FEEDBACK:

Feedback given by users in terms of 5-star ratings, comments, reviews and pre-defined preferences.

3.2.6 TEXTUAL DATA

Textual data refers to any kind of text data associated with a user profile. Search queries, reviews, comments, item descriptions are all considered textual data.

Text can be processed, mined and analysed in many different ways depending on the business task. Several e-commerce ecosystems use text data to keep track of customer's changing preferences.

3.2.7 SENTIMENT ANALYSIS

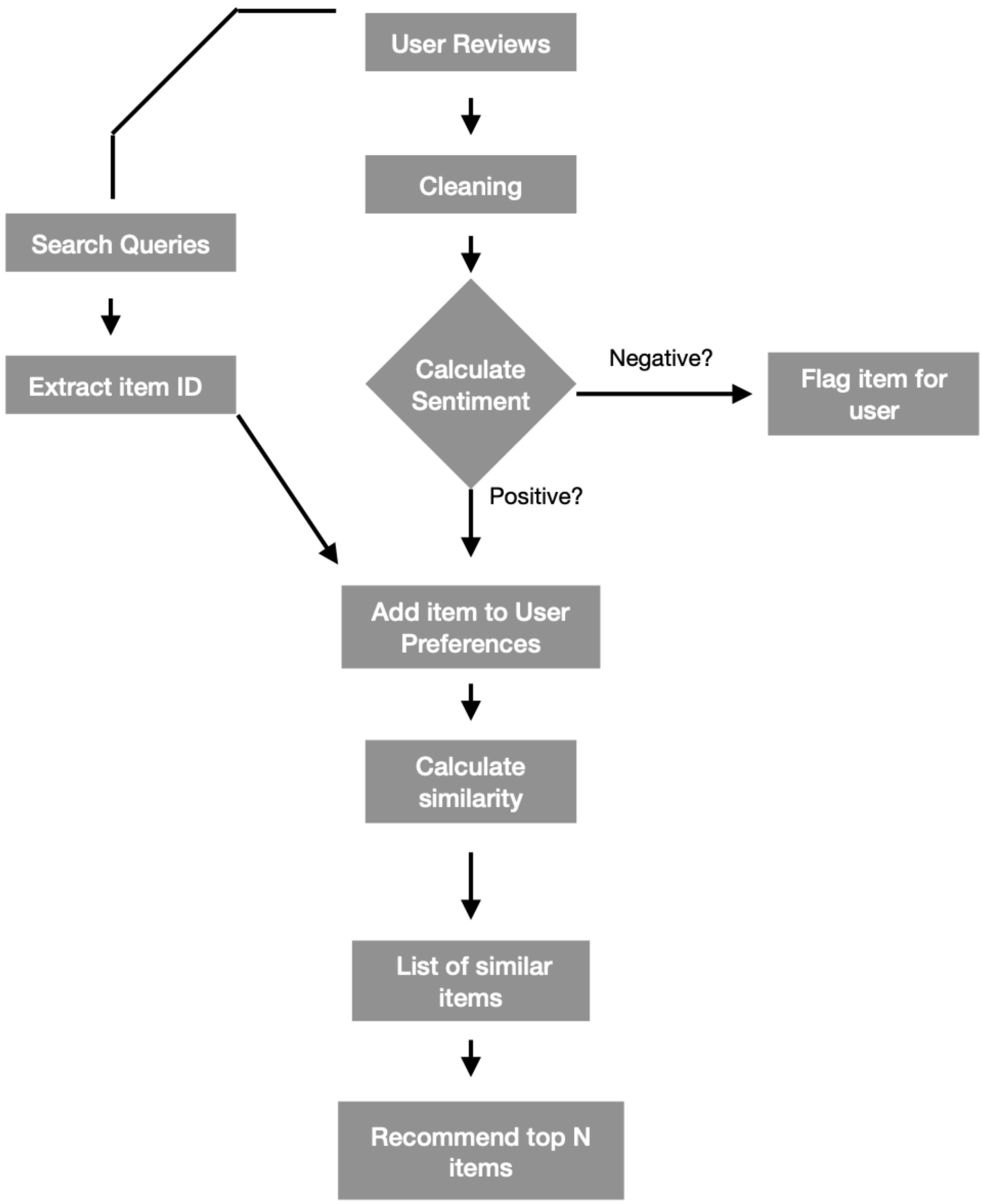
Sentiment analysis looks at the emotion expressed in a text. It calculates degree of negativity or polarity value of certain text, its value ranging from -1 to 1 in the order of increasing positivity.

3.2.8 COSINE SIMILARITY

In data analysis, cosine similarity is a measure of similarity between two sequences of numbers. For defining it, the sequences are viewed as vectors in an inner product space, and the cosine similarity is defined as the cosine of the angle between them.

3.3 METHODOLOGY

3.3.1 ACTIVITY DIAGRAM



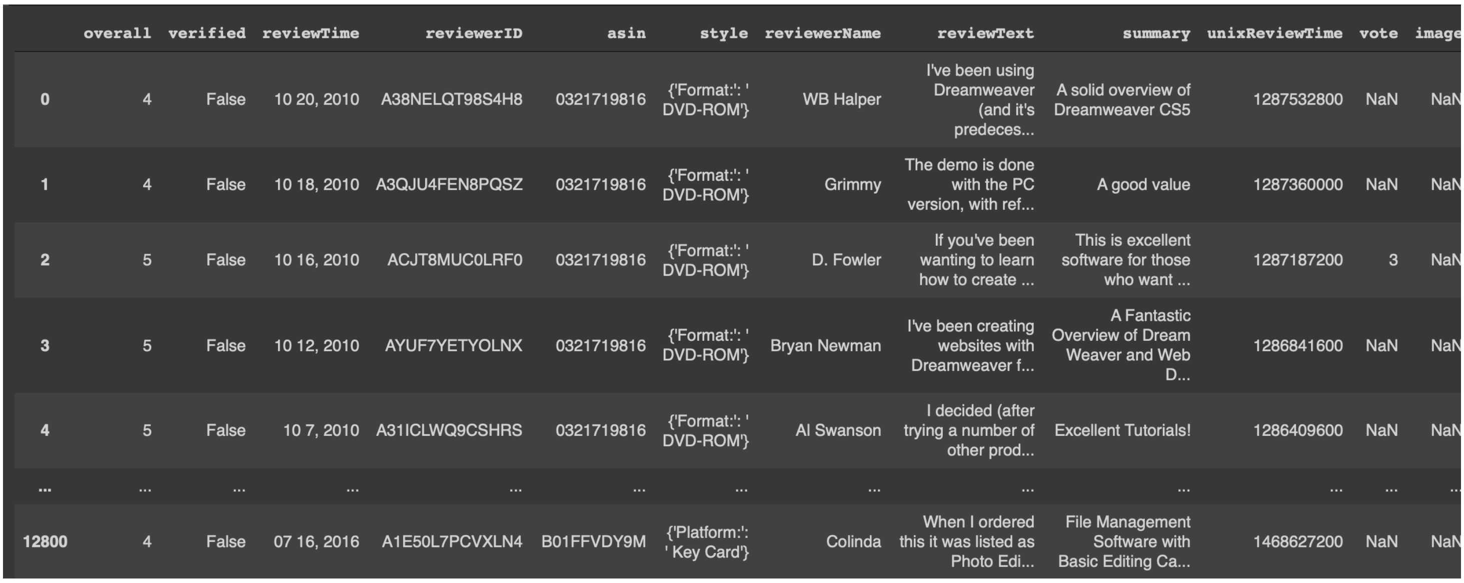
3.3.2 Sample Dataset

K-cores (i.e., dense subsets): Amazon Review Dataset [20]

Product Category: **Software**

No of Reviews	459,436
Total No of Columns	12
Metadata	Yes
Subset Size	12805

[Preview]



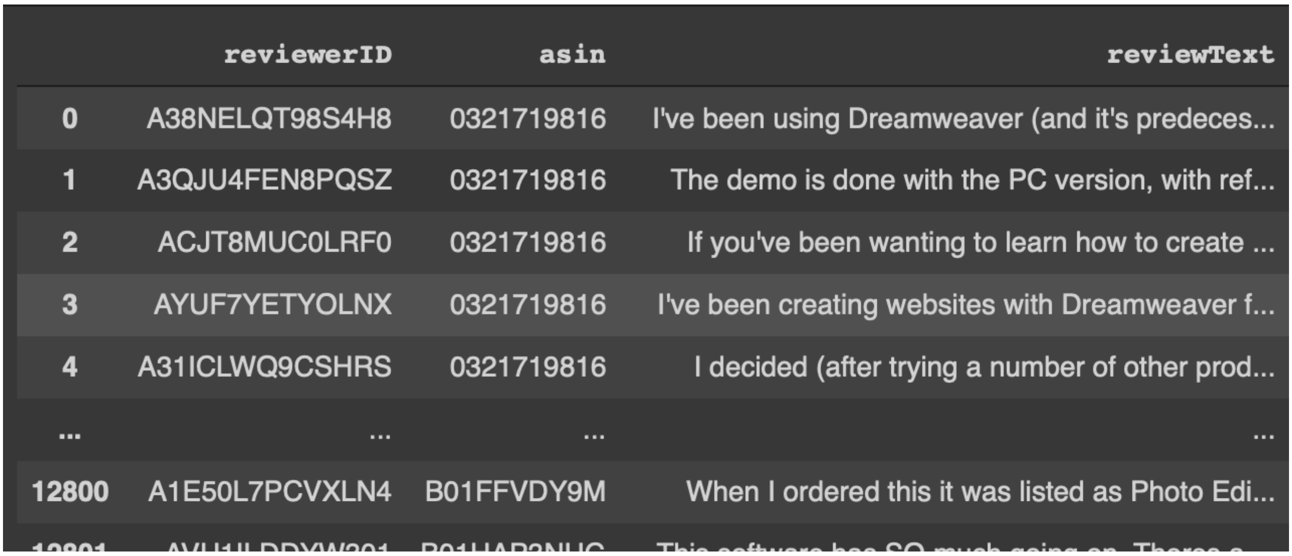
	overall	verified	reviewTime	reviewerID	asin	style	reviewerName	reviewText	summary	unixReviewTime	vote	image
0	4	False	10 20, 2010	A38NELQT98S4H8	0321719816	{'Format:': ' DVD-ROM'}	WB Halper	I've been using Dreamweaver (and it's predeces...	A solid overview of Dreamweaver CS5	1287532800	NaN	NaN
1	4	False	10 18, 2010	A3QJU4FEN8PQSZ	0321719816	{'Format:': ' DVD-ROM'}	Grimmy	The demo is done with the PC version, with ref...	A good value	1287360000	NaN	NaN
2	5	False	10 16, 2010	ACJT8MUC0LRF0	0321719816	{'Format:': ' DVD-ROM'}	D. Fowler	If you've been wanting to learn how to create ...	This is excellent software for those who want ...	1287187200	3	NaN
3	5	False	10 12, 2010	AYUF7YETYOLNX	0321719816	{'Format:': ' DVD-ROM'}	Bryan Newman	I've been creating websites with Dreamweaver f...	A Fantastic Overview of Dream Weaver and Web D...	1286841600	NaN	NaN
4	5	False	10 7, 2010	A31ICLWQ9CSHRS	0321719816	{'Format:': ' DVD-ROM'}	Al Swanson	I decided (after trying a number of other prod...	Excellent Tutorials!	1286409600	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
12800	4	False	07 16, 2016	A1E50L7PCVXLN4	B01FFVDY9M	{'Platform:': ' Key Card'}	Colinda	When I ordered this it was listed as Photo Edi...	File Management Software with Basic Editing Ca...	1468627200	NaN	NaN

3.4 PROCESS

4.4.1 PREPROCESSING

Data preprocessing is a data mining technique that involves transformation of raw data into an understandable format, because real world data can often be incomplete, inconsistent or even erroneous in nature. [23]

Columns ‘reviewerID’, ‘asin’, ‘reviewText' were filtered to prepare data for pre-processing.



	reviewerID	asin	reviewText
0	A38NELQT98S4H8	0321719816	I've been using Dreamweaver (and it's predeces...
1	A3QJU4FEN8PQSZ	0321719816	The demo is done with the PC version, with ref...
2	ACJT8MUC0LRF0	0321719816	If you've been wanting to learn how to create ...
3	AYUF7YETYOLNX	0321719816	I've been creating websites with Dreamweaver f...
4	A31ICLWQ9CSHRS	0321719816	I decided (after trying a number of other prod...
...	...	...	...
12800	A1E50L7PCVXLN4	B01FFVDY9M	When I ordered this it was listed as Photo Edi...
12801	AVH1ILDDYW201	B01HAR2NUC	This software has SO much going on. Theres a...

3.4.2 TRANSFORMATION OF DATA

Since computers only understand numbers, or electric signals of 0 and 1 text needs to be transformed into numbers in order for machines to process it.

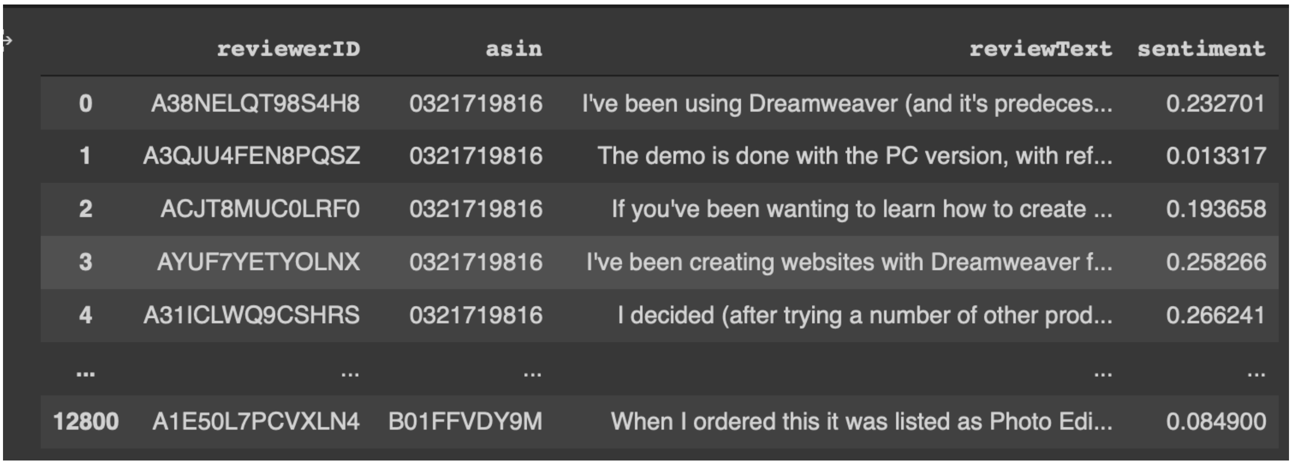
Data transformation is a process in which data is consolidated or transformed into some other standard forms which are better suited for data mining. [22]

Stop words, **lemmatization** and **tokenization** techniques are applied on each review. **Term Frequency-Inverse Document Frequency (TF-IDF)** is applied on every row of 'reviewText' column to exclude all irrelevant words. Now we have all the necessary words to perform our analysis.

3.4.3 CALCULATE SENTIMENT

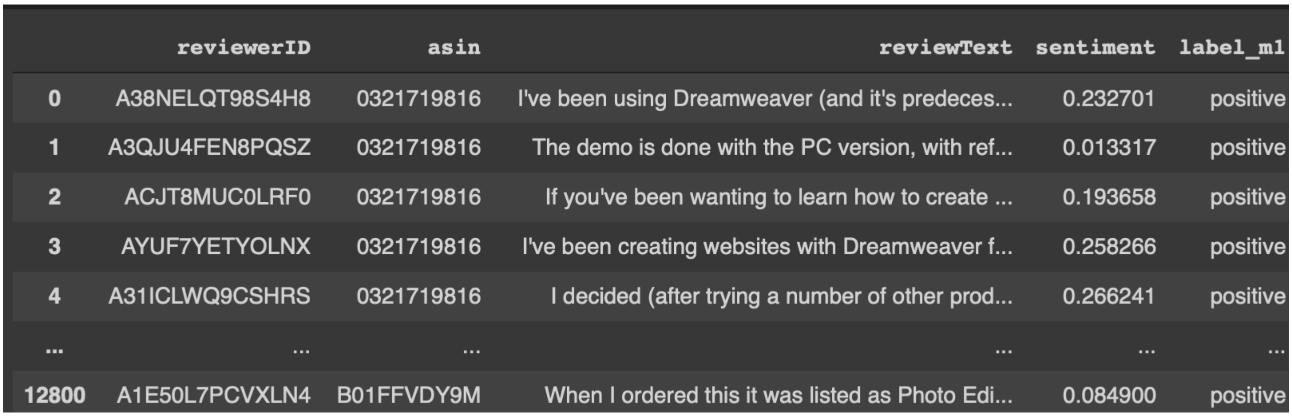
Now, clean data was ready for sentiment analysis.

Thus, 'sentiment' column was added to the table and was populated with polarity value of each review.



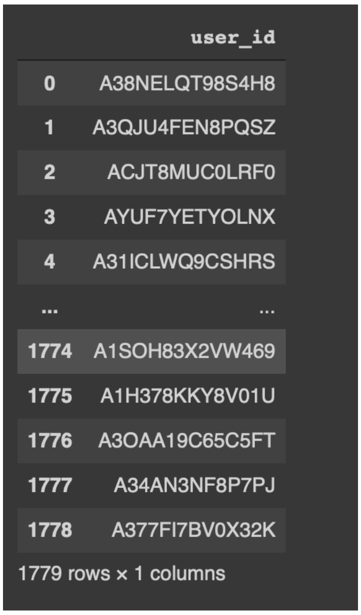
	reviewerID	asin	reviewText	sentiment
0	A38NELQT98S4H8	0321719816	I've been using Dreamweaver (and it's predeces...	0.232701
1	A3QJU4FEN8PQSZ	0321719816	The demo is done with the PC version, with ref...	0.013317
2	ACJT8MUC0LRF0	0321719816	If you've been wanting to learn how to create ...	0.193658
3	AYUF7YETYOLNX	0321719816	I've been creating websites with Dreamweaver f...	0.258266
4	A31ICLWQ9CSHRS	0321719816	I decided (after trying a number of other prod...	0.266241
...	...	...	...	...
12800	A1E50L7PCVXLN4	B01FFVDY9M	When I ordered this it was listed as Photo Edi...	0.084900

Next, 'positive' and 'negative' labels were added to each review according to its polarity value respectively.



	reviewerID	asin	reviewText	sentiment	label_ml
0	A38NELQT98S4H8	0321719816	I've been using Dreamweaver (and it's predeces...	0.232701	positive
1	A3QJU4FEN8PQSZ	0321719816	The demo is done with the PC version, with ref...	0.013317	positive
2	ACJT8MUC0LRF0	0321719816	If you've been wanting to learn how to create ...	0.193658	positive
3	AYUF7YETYOLNX	0321719816	I've been creating websites with Dreamweaver f...	0.258266	positive
4	A31ICLWQ9CSHRS	0321719816	I decided (after trying a number of other prod...	0.266241	positive
...	...	...	...	...	...
12800	A1E50L7PCVXLN4	B01FFVDY9M	When I ordered this it was listed as Photo Edi...	0.084900	positive

User table was crated with unique users extracted from reviews dataset. Preference attribute was added and populated with positively labeled items of every user.

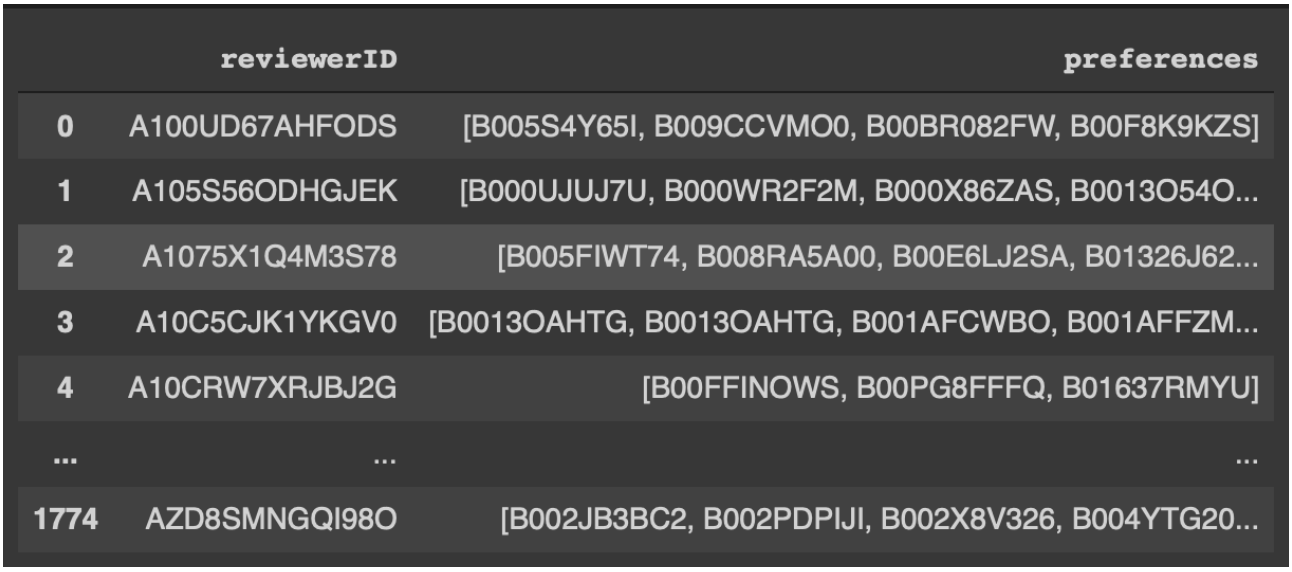


	user_id
0	A38NELQT98S4H8
1	A3QJU4FEN8PQSZ
2	ACJT8MUC0LRF0
3	AYUF7YETYOLNX
4	A31ICLWQ9CSHRS
...	...
1774	A1SOH83X2VW469
1775	A1H378KKY8V01U
1776	A3OAA19C65C5FT
1777	A34AN3NF8P7PJ
1778	A377FI7BV0X32K

1779 rows × 1 columns

3.4.4 CALCULATE SIMILARITY

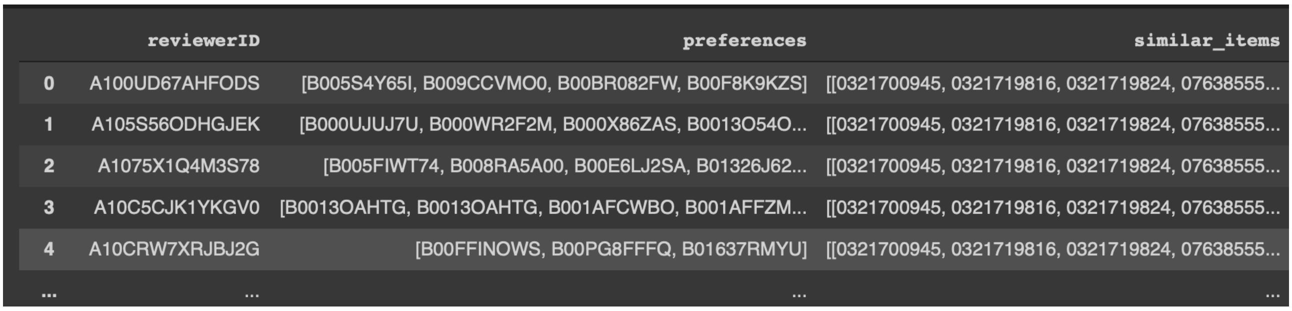
Now, cosine similarity was calculated for every item in user preference column and 5 similar items were generated for each preferred item.



	reviewerID	preferences
0	A100UD67AHFODS	[B005S4Y65I, B009CCVMO0, B00BR082FW, B00F8K9KZS]
1	A105S56ODHGJEK	[B000UJUJ7U, B000WR2F2M, B000X86ZAS, B0013O54O...
2	A1075X1Q4M3S78	[B005FIWT74, B008RA5A00, B00E6LJ2SA, B01326J62...
3	A10C5CJK1YKGV0	[B0013OAHTG, B0013OAHTG, B001AFCWBO, B001AFFZM...
4	A10CRW7XRJBJ2G	[B00FFINOWS, B00PG8FFFQ, B01637RMYU]
...	...	...
1774	AZD8SMNGQI98O	[B002JB3BC2, B002PDPIJI, B002X8V326, B004YTG20...

3.4.5 GENERATE RECOMMENDATIONS

Next, similar_items attribute was added and populated with similar items.

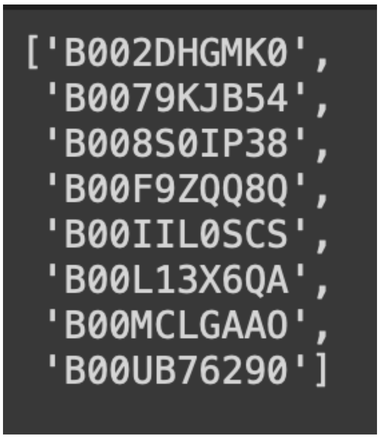


	reviewerID	preferences	similar_items
0	A100UD67AHFODS	[B005S4Y65I, B009CCVMO0, B00BR082FW, B00F8K9KZS]	[[0321700945, 0321719816, 0321719824, 07638555...
1	A105S56ODHGJEK	[B000UJUJ7U, B000WR2F2M, B000X86ZAS, B0013O54O...	[[0321700945, 0321719816, 0321719824, 07638555...
2	A1075X1Q4M3S78	[B005FIWT74, B008RA5A00, B00E6LJ2SA, B01326J62...	[[0321700945, 0321719816, 0321719824, 07638555...
3	A10C5CJK1YKGV0	[B0013OAHTG, B0013OAHTG, B001AFCWBO, B001AFFZM...	[[0321700945, 0321719816, 0321719824, 07638555...
4	A10CRW7XRJBJ2G	[B00FFINOWS, B00PG8FFFQ, B01637RMYU]	[[0321700945, 0321719816, 0321719824, 07638555...
...	...	...	...

4.5 RESULTS AND INTERPRETATION

Finally, recommendations were validated.

One reviewerID was selected randomly to validate the results.



```
['B002DHGMK0',
 'B0079KJB54',
 'B008S0IP38',
 'B00F9ZQQ8Q',
 'B00IIL0SCS',
 'B00L13X6QA',
 'B00MCLGAAO',
 'B00UB76290']
```

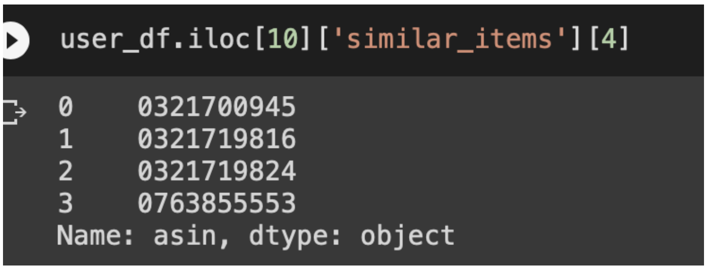
One of the preferences was chosen for validation.

Preference: [B00IIL0SCS] : **Corel VideoStudio**

(Corel VideoStudio is a video editing software package for Microsoft Windows.)

Next, items similar to that item were listed out as recommendations from similar_items column

Similar Items



```
user_df.iloc[10]['similar_items'][4]

0    0321700945
1    0321719816
2    0321719824
3    0763855553
Name: asin, dtype: object
```

Next the recommendations were validated using item description.

Recommendations

1- [0321700945] : **Lightroom**

Lightroom is the cloud-based service that gives you everything you need to create, edit, organize, store, and share your photos across any device

2- [0321719824] : **5 Software**

5 Software for Windows is a powerful, wide-ranging software application that's used to create immersive, interactive websites and also applications for websites and mobile devices. It delivers video (via the included Adobe Media Encoder) to the Web, tablets, TV, and smart phones.

3- [0763855553] : **Pinnacle Studio**

Pinnacle Systems, Inc. is a California-based American manufacturer of digital video hardware and software for the mainstream and broadcast markets.

INTERPRETATION OF RESULTS:

Since ratings data was not available for recommendations, some results turned out to be unrelated to the existing item. For example; one of the four recommendations for reviewer above and preference ‘B00IIL0SCS’ was ‘Microsoft Office’ which is a completely different suite from the preferred item (video editing software). With adequate ratings and collated user behaviour the program can overcome such shortcomings.

CHAPTER 4: DATA MINING IN SOCIAL CONTEXT

4.1 ONLINE PROFILING: BACKGROUND AND CONTEXT

User Profiling can be defined as the process of identifying user's interest domain, preferences, likes, dislikes etc. In information science, profiling refers to the process of construction and application of user profiles generated by computerized data analysis. Personality and psychological aspects are often inferred by monitoring user behaviour. Such insights gives profound peek into cognitive aspects of a user which greatly helps into psychoanalysing their personalities. Discovery of patterns and correlation in large quantity of data helps organizations improve user experience and increase customer retention. Users' personality is modelled, analysed, verified and updated with respect to changing user feedback. The model is sometimes referred to as 'digital persona' or 'digital footprint' of a person which surprisingly enough tells much more about a person than explicit feedback does. Much can be predicted about a person correctly when their behaviour is constantly monitored and stored.

When a profile is constructed with the data of a single person, this is called individual profiling. [23]

Clustering, an unsupervised learning technique performed on the input data to group them into similar classes which allows categorization of a person as a certain type of person, in a certain group. If a group profile is applied to an individual whose data match the profile, then that is called indirect individual profiling.

Capturing information about users and their interest is the main function of user profiling.

Much of the research has been done on profiling in the field of recommender system and various profiling techniques have been evolved over time.

The process of profiling is primarily divided into 5 stages

Data Collection: the collection of data relevant to problem domain and analysis objectives.

Data preparation: the pre-processing of data including removal of noise, complexity and dimension reduction.

Data Mining: the data is processed and analyzed with the use of algorithms to accomplish business goals.

Interpretation & Evaluation : the analysis results are evaluated and validated by seeking expert reviews.

Application : The constructed profiles are fine-tuned into working algorithms.

Fraud Prevention, Credit Scoring, Employee reputation (hiring process), Forensic science are some of the many applications of profiling.

One of the most prominent application is Consumer profiling, a form of customer analytics, where customer data is used to make promotion and pricing decisions of products.

Some of the profiling practices include supervised learning algorithms which learn about the users and generated profiles based on hypothesis and deduction. The result of this type of profiling is the verification or refutation of the hypothesis. It is also sometimes referred to as ‘deductive profiling’

Unsupervised Learning on the other hand, uses data mining technique to learn, explore, detect patterns about people and finding correlations one did not expect or even think of.

Since unsupervised learning algorithms allow construction of type of knowledge not indulging human interventions or motivations and are exclusively based on stochastical correlations, they seem to allow for an inductive type of knowledge construction that does not require theoretical justification or causal explanation. [24]

4.2 RAPID SHIFT IN THE REALM OF DATA EXPLOIT : GOOGLE

Web bug, also known as web beacon is a technique used by websites to monitor the activity of users for the purpose of web analytics. Using these bugs, any website can match online activity to a specific individual. With this information, it can personalize your online experience based on the data it has collected.

Despite it's portrayal as a search engine, Google's main business is online advertising or online marketing. In 2020, Alphabet, the parent company of Google generated almost $183 billion in revenue. Of that, $147 billion — over 80% — business, according to the company's 2020 annual report. [25]

Google AdWords is a pay-per-click online advertising platform service provided by google to advertisers and **Google AdSense** is a free utility to make money by placing ads on a site and monetize it. These two are the main advert features currently provided and maintained by Google. Based on the keywords and the users they want to target, businesses pay to get their advertisements ranked at the top of the results page on the search engine. **Google Analytics** is a web analytics service offered by Google that tracks and reports website traffic, currently as a platform inside the Google Marketing Platform brand.

In April 2011, Google refused to sign Do Not Track feature which was designed to allow internet users to opt-out of tracking by websites for Chrome that was being incorporated in most other modern web browsers, including Firefox, Internet Explorer, Safari, and Opera.

In March 2012, Google made significant changes to its privacy policy, enabling the company to share data across third party services. This gave websites, many of which already subscribed to Google's Marketing services a fully legal pass to collect user data and exploit it. This policy change was widely criticised for putting people's privacy in jeopardy but little did the public know, this was just the beginning.

DoubleClick Inc. was an advertisement company that developed and provided Internet ad serving services was acquired by Google and it dropped its ban on collecting personally identifiable information (PII) on its DoubleClick service. It specifically stated that it "may" combine web-browsing records of average users.

Google has been involved in several data breach scandals in the past but the most infamous is 2018's Google+ API breach which exposed private details of over five million users. 500,000 Google+ accounts got compromised, which allowed hundreds of third party applications unauthorized access to user's private details.

4.3 CAMBRIDGE ANALYTICA : FACEBOOK

Facebook has faced a number of privacy concerns over the past years. In early 2019 Facebook was being investigated by the Federal Trade Commission (FTC) for violating a 2011 consent decree to safeguard users' personal information.

Cambridge Analytica Ltd (CA) was a British political consulting firm that came to prominence through the Facebook–Cambridge Analytica data scandal. The company underpinned its work on President Trump's campaign in 2016 and manipulated user's perception towards political sphere by targeting them what's refereed to as 'electronic brainwashing' by the media. [26] [27]

Cambridge Analytica was able to leverage its alliance with Facebook to access users' personal data, by exploiting third-party app permissions. It conducted paid online surveys on Facebook to acquire user's personal data which also gave the survey app unwarranted access to information on the user's friends network. A survey that started with about 270,000 people ultimately resulted in the profiling of about 87 million users, the majority of whom had not explicitly given Cambridge Analytica permission to access their data.

The primary research goal, which was to establish a methodology for psychographic profiling of individuals based on social media and other indicators [28]. With real-time monitoring of ad responses on targeted individuals, including real-time substitution to find "click bait" that worked, the ad campaign was able to both maximize its impact and detect trends not visible at the macro scale.

"Documents seen by the *Observer*, and confirmed by a Facebook statement, show that by late 2015 the company had found out that information had been harvested on an unprecedented scale. However, at the time it failed to alert users and took only limited steps to recover and secure the private information of more than 50 million individuals"

- The Guardian, [29]

4.4 PUBLIC CONCERNS REGARDING DATA PRIVACY

The way data is stored and managed today is drastically different than it used to be in early days of technology. Processing speed and storage capacity of computers are increasing at an exponential rate which further gives more room to not only store huge chunks of data but perform several resource consuming processes that were deemed impossible in early days.

This transition from old methods of data handling to rigorous new methods brings a lot of uncertainty among the public and the consumers who are not aware exactly what and how their data is being consumed. *Edelman*, an American public relations and marketing consultancy firm reported that public trust in technology reached all-time lows in 17 of 27 countries in 2020 [30]. The report was compiled using results of a survey of more than 33,000 people from 28 countries, including both general population respondents and tech savvies.

According to Edelman Trust Barometer there was a 30-point difference between technology and business (77 percent vs 47 percent) with regards to how much public trusted both, as of 2012 has severely declined in previous years to just under 10 points [31]. It held companies' lack of accountability and unwillingness to self-govern and abide by policies responsible for this mistrust.

The Verge, an American technology news website operated by Vox Media conducts periodic surveys gauging Americans' attitudes toward the major tech companies. [32]

The result's of their first ever National tech survey of 2016 reported that a great proportion of people distrust Facebook followed by Google.

It conducted similar survey before March 2020, just before planet earth was hit severely by the novel coronavirus. Results showed that more than less 2/3 participants believed that Facebook has been given ‘too much power’ and more than half said that the government needs to strictly control and monitor tech companies and their controversial modes of conduct.

CHAPTER 5
CONCLUSION

5.1 CONCLUSION

Decision-making comes naturally to human cognition it is easy to ignore its impact and extensive use in our every day lives. Similar to day-to-day life, the infrastructure of digital commerce is highly dependant on data driven decision-making strategies applied at every transaction. The bare bone, central component to the is data, consumer data to be precise.

Every industry today aims to be data-driven. Decisions cannot be simply made on the basis of intuition alone. Bias and false assumptions can cloud judgment and lead to poor decision making. Most organizations understand that intuition when combined with proper data gives exemplary insights and patterns often unnoticed by human intervention.

A recommender system automates the decision-making process by analyzing an individual's buying behaviour and postulates the preferred content. It collects and constructs their respective observable behaviour and infers the probability of an individual's decision prior to the action.

The central challenge of recommender systems is to find the "best match" between a given user in a given context. Thus, depending on the definition of "best match" this challenge leads to a variety of massive optimization and search problems, with complicated constraints.

5.2 APPENDIX

Chapter 1

Personally Identifiable Information: Personally identifiable information (PII) is information that, when used alone or with other relevant data, can identify an individual.

User Profiling: A user profile is a collection of settings and information associated with a user. It contains critical information that is used to identify an individual, such as their name, age, portrait photograph and individual characteristics such as knowledge or expertise.

Exploit: An exploit (from the English verb to exploit, meaning "to use something to one's own advantage") is a piece of software, a chunk of data, or a sequence of commands that takes advantage of a bug or vulnerability to cause unintended or unanticipated behavior to occur on computer software, hardware, or something electronic.

Quarantine: Restricting the activities of healthy people for a period of time as determined by the competent medical authorities - COVID-19

User behaviour: User behavior encompasses all the actions visitors take on a website.

Association: Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases

Customer Churn: Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers.

Virtual Health Care Assistants: A healthcare virtual assistant is an individual who works remotely and can help you with routine tasks such as managing the front office, setting patient appointments, patient engagement, etc

Click-through rate (CTR): Click-through rate is the ratio of users who click on a specific link to the number of total users who view a page, email, or advertisement. It is commonly used to measure the success of an online advertising campaign for a particular website as well as the effectiveness of email campaigns.

Dwell Time: In information retrieval, dwell time denotes the time which a user spends viewing a document after clicking a link on a search engine results page. Dwell time is the duration between when a user clicks on a search engine result, and when the user returns from that result.

Pep Worx: the cloud-based data and analytics solution assistance for retailers in making more informed decisions on PepsiCo product assortments, merchandising and other point-of-sale areas by identifying valuable shoppers by location.

Chapter 2

Usenet: a worldwide distributed discussion system available on computers. It was developed from the general-purpose Unix-to-Unix Copy dial-up network architecture.

GroupLens: A human–computer interaction research lab in the Department of Computer Science and Engineering at the University of Minnesota.

Sparse matrix: In numerical analysis and scientific computing, a sparse matrix or sparse array is a matrix in which most of the elements are zero.

Similarity: A machine learning method that uses a nearest neighbor approach to identify the similarity of two or more objects to each other based on algorithmic distance functions.

Cold start: Cold start is a potential problem in computer-based information systems which involves a degree of automated data modelling. Specifically, it concerns the issue that the system cannot draw any inferences for users or items about which it has not yet gathered sufficient information.

Geolocation: the use of location technologies such as GPS or IP addresses to identify and track the whereabouts of connected electronic devices.

Social data: information that social media users publicly share, often data obtained from social networking services.

Domain knowledge: Domain knowledge is knowledge of a specific, specialized discipline, field or environment in contrast to general knowledge.

Ensemble methods: Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

Chapter 3

EDA: In statistics, exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods

Feature selection: In machine learning and statistics, feature selection is the process of selecting a subset of relevant features for use in model construction.

ETL: In computing, extract, transform, load is the general procedure of copying data from one or more sources into a destination system.

E-commerce ecosystem: an integrated group of systems that works together to electronically provide goods to an individual.

Stop words: Stop words are any word in a stop list which are filtered out before or after processing of natural language data

Lemmatization: process of removing ending from a word and converting it to its base or dictionary form, also known as lemma of the word.

Tokenization: tokenization is the process of converting a sequence of characters into a sequence of tokens.

Term Frequency-Inverse Document Frequency (TF-IDF): tf–idf reflects how important a word is to a document in a collection or corpus.

Chapter 4

Supervised learning: In supervised learning, the algorithm learns from labeled attributes and output classes.

Unsupervised learning: Uses machine learning algorithms to analyze and cluster unlabeled data sets.

Online advertising: a form of marketing and advertising which uses the Internet to promote products and services to audiences and platform users.

5.3 BIBLIOGRAPHY

[1] The Economist, The data economy demands a new approach to antitrust rules, The world’s most valuable resource is no longer oil, but data. Regulating the internet giants. May 6th, 2017, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[2] Maria Augusta S. N. Nunes, Rong Hu. Personality-based Recommender Systems: An Overview, *RecSys’12,* September 9–13, 2012, Dublin, Ireland. ACM 978-1-4503-1270-7/12/09

[3] Stevens, Gina (10 April 2012). "Data Security Breach Notification Laws" (PDF). *fas.org*. Retrieved 8 June 2017.

[4] Megan Graham, Jennifer Elias. How Google’s $150 billion advertising business works, CNBC, https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html

[5] K. Dave and V. Varma. Computational Advertising: Techniques for Targeting Relevant Ads. Foundations and Trends$^{\circledR}$ in Information Retrieval, vol. 8, no. 4-5, pp. 263–418, 2014.

[6] By Ian MacKenzie, Chris Meyer, and Steve Noble, How retailers can keep up with consumers. October 1, 2013. McKinsey & Company, https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers

[7] How Companies Learn Your Secrets, By Charles Duhigg, The New York Times, Feb. 16, 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

[8] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix recommender system: Algorithms, business value, and innovation. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (December 2015), 19 pages. DOI: http://dx.doi.org/10.1145/2843948

[9] Greg Linden, Brent Smith, and Jeremy York. 2003. amazon.com Recommendations, Item-to-Item Collaborative Filtering. IEEE Computer Society. 1089-7801/03/$17.00©2003 IEEE

[10] Bloomberg. Devin Leonard. 21 September 2016. Spotify Is Perfecting the Art of the Playlist. https://www.bloomberg.com/news/articles/2016-09-21/spotify-is-perfecting-the-art-of-the-playlist

[11] Forbes. Bernard Marr. Apr 5, 2019. The Fascinating Ways PepsiCo Uses Artificial Intelligence And Machine Learning To Deliver Success. https://www.forbes.com/sites/bernardmarr/2019/04/05/the-fascinating-ways-pepsico-uses-artificial-intelligence-and-machine-learning-to-deliver-success/?sh=428b579b311e

[12] P. Resnick – N. Iacovou – M. Sushak – P. Bergstrom – J. Riedl (1994): GroupLens: An open architechure for collaborative filtering of netnews, In Proceedings of the ACM Conf. Computer Support Cooperative Work (CSC),pp. 175-186.

[13] Upendra Shardanand - Pattie Maes. 1995. Social Information Filtering: Algorithms for Automating "Word of Mouth", MIT Media-Lab. May 7 11 1995 - CHI' 95, Denver, Colorado, USA © 1995 ACM 0-89791-694-8/95/0005

[14] Paul Resnick - Neophytos Iacovou - Mitesh Suchak - Peter Bergstrom, - John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. CSCW 94- 10/94 Chapel Hill, NC, USA @ 1994 ACM 0-89791 -689-1/94/0010..

[15] Charu C. Aggarwal. Recommender Systems: The Textbook. Springer Cham Heidelberg New York Dordrecht London © Springer International Publishing Switzerland 2016. DOI 10.1007/978-3-319-29659-3

[16] David Goldberg - David Nichols - Brian M. Oki - Douglas Terry. December 1992. Information Filtering using Collaborative Filtering to weave an information tapestry. Vol,35, No,12 /Communication of the ACM

[17] Paul Resnick - Neophytos Iacovou - Mitesh Suchak - Peter Bergstrom - John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. CSCW 94- 10/94 Chapel Hill, NC, USA @ 1994 ACM 0-89791 -689-1/94/0010..

[18] Charu C. Aggarwal. Recommender Systems: The Textbook. Springer Cham Heidelberg New York Dordrecht London © Springer International Publishing Switzerland 2016. DOI 10.1007/978-3-319-29659-3

[19] E. Rich (1979): User modeling via stereotypes, Cognitive Science, Vol. 3, No. 4, pp. 329–354.

[20] Justifying recommendations using distantly-labeled reviews and fined-grained aspects Jianmo Ni, Jiacheng Li, Julian McAuley *Empirical Methods in Natural Language Processing (EMNLP)*, 2019. https://nijianmo.github.io/amazon/index.html#sample-review

[21] [22] Parteek Bhatia. Data Mining and Data Warehousing Principles and Practical Techniques © Cambridge University Press 2019

[23] Jaquet-Chiffelle, David-Olivier (2008). "Reply: Direct and Indirect Profiling in the Light of Virtual Persons. To: Defining Profiling: A New Type of Knowledge?". In Hildebrandt, Mireille; Gutwirth, Serge (eds.). Profiling the European Citizen. Springer Netherlands. pp. 17–45. doi:10.1007/978-1-4020-6914-7_2.

[24] Custers, B.H.M. (2004). "The Power of Knowledge". Tilburg:Wolf Legal Publishers.

[25] Megan Graham, Jennifer Elias. How Google's $150 billion advertising business works, CNBC, https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html

[26] New York Times, How Trump Consultants Exploited the Facebook Data of Millions. March 17, 2018. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

[27] Electronic Brainwashing: Cambridge Analytica's Sinister Facebook Strategy | The Daily Show, Youtube

[28] **Jim Isaak, Mina J. Hanna.** User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. IEE Computer Society.

[29] Carole Cadwalladr *and* Emma Graham-Harrison. 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. 17 Mar 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[30] [31] Richard Edelman. March 20, 2021.EDELMAN TRUST BAROMETER: TRUST IN TECHNOLOGY

[32] By Elizabeth Lopatto@mslopatto Oct 6, 2021. After a year of the pandemic, we asked Americans if their trust in big tech companies has changed — and if the biggest ones should be broken up, VERGE TECH SURVEY 2021