Final Project Report

Multi Criteria Recommendation System using SVD and Lasso Regression

Submitted By:

Kunal Sharma

2K18/MBA/911

Under the Guidance of:

Dr. Gaganmeet Kaur Awal

UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP

Bawana Road Delhi 110042

2018-2020

DELHI TECHNOLOGICAL UNIVERSITY

CERTIFICATE

This certificate verifies that the project report titled "**Multi Criteria Recommendation System using SVD and Lasso Regression**" is an original record of work done by Kunal Sharma, a student of University School of Management & Entrepreneurship, Delhi Technological University and is permitted to submit the dissertation towards the partial fulfilment of requirements of completion of the degree of Masters Of Business Administration In Business Analytics at Delhi Technological University.

Date:

Place: Delhi

DECLARATION

I, **Kunal Sharma**, student of **MBA Business Analytics, University School of Management & Entrepreneurship, Delhi Technological University** hereby declare that the project titled "**Multi Criteria Recommendation System using SVD and Lasso Regression**" submitted by me towards **University School of Management & Entrepreneurship, Delhi Technological University** in partial fulfilment of requirement for the award of the degree of Master of Business Administration in 2018-20, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition. The Author attests that appropriate citations as well as mentions where ever required have been made in this report towards the Dissertation / Project report; other than brief excerpts requiring only proper acknowledgement in scholarly writing and all such use is acknowledged.

NAME –Kunal Sharma

SIGNATURE

ACKNOWLEDGEMENT

I would like to extend my gratitude to my guide Dr. Gaganmeet Kaur Awal, who is also my concerned faculty from the esteemed management department of USME. The project report became possible under their sheer support and guidance throughout the project duration.

In preparation of report, I had to take the help and guidance of my guide and other classmates and friends, who deserve my deepest gratitude. As the completion of this assignment gave me much pleasure, I would like to show my gratitude to them, who guided me throughout numerous consultations. I would also like to expand my gratitude to all those who have directly and indirectly guided me in writing this assignment.

I also feel thankful towards my teachers at USME who helped me understand the opportunity have made valuable suggestions on my report which gave me an inspiration to improve the quality of the project.

Kunal Sharma

2K18/MBA/911

Table of Contents

List of Figures:

List of Tables:

1. INTRODUCTION

This section gives an overview of recommendation system techniques and how the organisation uses these techniques in their business functionality. This section also defines the objective of this project and gives a glimpse of the techniques that already exist in this domain. Later this section would expand the understanding of recommendation systems into multi criteria systems, which are implemented through a proposed algorithm in this project.

1.1 Recommendation systems

The evolution of multiple data collection techniques and storage systems has encouraged various platforms to collect and store data into their servers more efficiently. With this data they are now capable of suggesting more user friendly techniques and relatable content to various users. Companies are using the user generated content to give the users a more personalised experience on their platforms.

The users like choices o choose from for their next lifestyle, and often get confused to choose from among the vast variety of choices available to them. Using the research discipline based on recommendation systems, various companies can give more importance to conveying information to the users rather than raw data. The users would get recommended much more precisely relevant information made available to them further than a huge inflow of all the random data available on the website.

Typical recommendation systems work on the model of user- item matrix, where set of all possible user are paired against set of all possible items. The model then calculates a utility function that measures the likelihood of recommending a particular item to a particular user. An accurate model has a utility function that gives more fine tuned results to the user or the item. The more personalised the recommendations the user gets, the better is the overall performance of the recommendation system and better does it work for the organisation implementing it.

Considering the below case of terms,

$u \in Users$ (set of all possible Users)

$i \in Items$ (set of all possible Items)

R: $Users\ x\ Items -> R0$ (is the utility function), where

R (u,i) is the user u's rating for a particular item i and is a non-negative integer based on particular scale.

In most of the readings, the recommendation systems are based on only one criterion to recommend the ratings to a new user. In recent studies based on previous implementations of the techniques, the concept of taking into account more than one criterion is used. Users might base their overall utility function on not just the overall rating of a particular item but sub category wise distribution the ratings. This enables the system to recommend more personalised content to the users.

The uses of incorporating multiple ratings [3, 4, 5] for the user enable us to improve the overall performance of a recommendation system. This not only enables to generate more users' friendly content, but also enables the users to utilise their preference of choices on a particular platform more suited to their personality. Hence solving more complex preference of each user and item pair makes the platform more users friendly.

As an illustrative example we can look into the case of restaurants on some platform. The users might go to a particular restraint and then like a particular aspect of that restaurants more than the other aspects. The same user's hen visits other restraints might be willing to look for place where those aspects are more enlightened. Hence, the users might be more interested in the ambiance of the place rather than just the taste of the food. They might want to go to a place where there is music rather than random noises. These criteria that form a user's personality also form a basis for their choice of restaurants to visit more often. The same users would want to be in a company of likeminded individuals or groups. If two individuals been to a particular set of three restaurants and have rated them 7 on 10 for a set group of criteria. Then these two users u and u' are considers to be in a same group of users or neighbours.

In a three criteria rating setting, suppose the user us rates a particular restaurant as:

$$R(u, i) = (7, 2, 9)$$

In-case the criteria were ‘hygiene’, ‘location’, ‘quality’.

The user may have like the hygiene and the quality of food served at the restaurant, but did not appreciate the location of the restraint. A multi criteria recommendation system would take in the subjective ratings of the various criteria of the user. The system would then give the individual rating a particular weight based on the user’s choices. For example if the weights assigned are 0.3, 0.3 and 0.4, then the overall rating for that particular item for the user would be 6.3.

Now another user u’ might have different ratings patterns for the restaurants’ rating. Even though if the user u’ gets the similar overall score the recommendation systems would take into account the preference of criteria for the various ratings and then classify the user u’ into same or a different group than user u.

The use of multi criteria recommendation systems s proposed for a wide range of applications such as movies ratings, hotel reviews, e commerce buying of goods etc. Other than these areas, domains such as books, supply chain, marketing, financial stock movements and travel use such recommendation techniques to improve the services provided to their customers. Multiple research papers talk about the [5,6] utilising various dimension from the dataset. The dimension can be keywords, authors, links or any other useful information related to the domain of the document.

The general classification of recommendation systems is as:

- Content based recommendation systems
- Collaborative filtering based recommendation systems
- Knowledge based recommendation systems
- User based recommendation systems
- Hybrid recommendation systems, that combine one or of the above criteria.

1.2 Objective of the Study

From a research perspective in the domain of recommendation systems, the improvement of overall accuracy of a particular recommendation system has been the primary objective of any new study of algorithms.

From the recent research, many recommendation systems that take into account the multiple criteria settings of the users and then expand that idea to recommend better rating to the users for a more personalised experience on any business platform have taken more importance.

This project explores the traditional techniques that form the basis of this generalised recommendation system technique and then builds upon a multi criteria recommendation system with improved accuracy and precision. Hence creating an efficient recommendation system that can also act as a web personalisation tool, to recommend appropriate items to multiple users based on their preference.

Most multi criteria recommendation system work on collaborative filtering techniques, but this project aims at building an improved version of singular value decomposition algorithm to generate the individual rating. The overall implementation would be broken down into below steps.

- Dividing the multi criteria in to individual criteria and predicting the ratings using SVD techniques.
- Then generating an overall aggregation function to collate those ratings into one overall rating by allocating the weights to individual rating criteria, using regularised regression techniques like lasso or ridge regression.
- Fine tuning the parameters for better accuracy for the user prediction based on available data.

The aim of the implementation of the above stated objectives is to improve upon the accuracy of the existing recommendation systems. This project would use the existing implementation of single criteria recommendation systems and extend it to multi criteria recommendation systems. The regression techniques of machine learning would be used to learn the weights associated with each criterion and then estimate the overall criteria of the recommendation systems.

The figure below describes the general classification of single criteria recommendation systems which would then be further extended to multiple criteria systems.

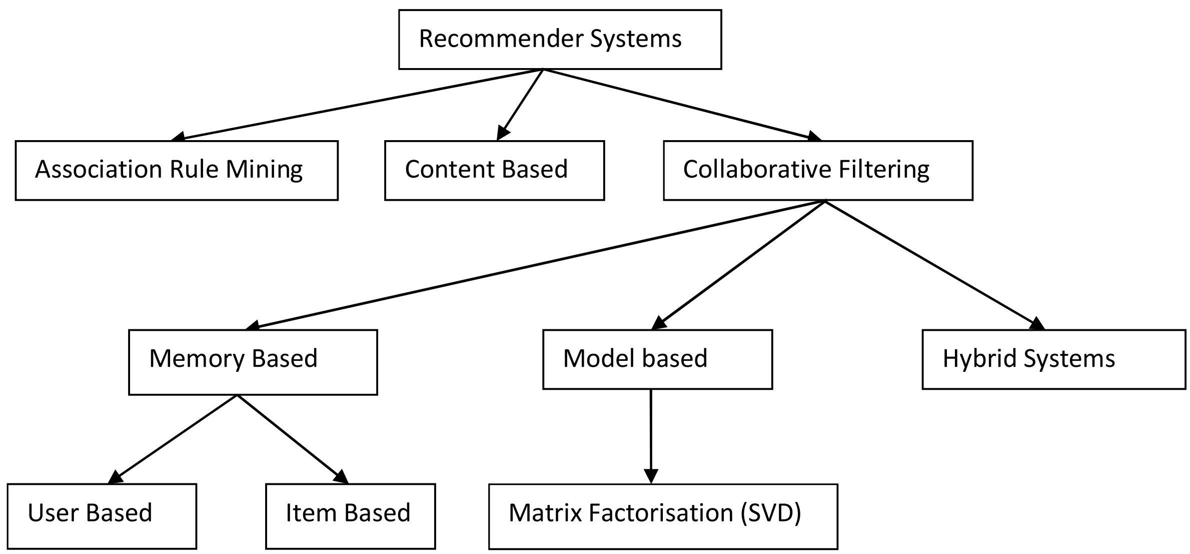

Fig 1.1 Techniques used in Recommendation systems

1.3 Organization of the report

The report begins at this section of introducing the overall domain of Recommendation systems. It then explores the various existing research done in this domain as well as techniques that form the base of algorithms to be used in this project. Also the literature review contains the study of different existing technologies that are used in this domain to study the recommendation systems. The net section is research methodology which discusses the implemented technique t obtain the desired results in this section. The major SVD algorithm for recommendation step and Lasso regression algorithm for the aggregation step are then broadly discussed. Following the implementation of algorithms the next section discusses the experiments performed and the dataset used to perform the said experiments. These experiments were mainly used to compare against the existing results. Lastly, the report discusses the limitations and future scope of this topic and finally concludes the topic.

2. LITERATURE REVIEW

This section asses the existing techniques like multi criteria decision making and single criteria recommendation systems that would eventually provide the base framework to build the final multi criteria models upon. These techniques are really important to be understood so to be implemented in the final form of proposed algorithm. Also, a better understanding of these principles would help us in improving the parameters of the proposed model, making it more robust and adaptive to various datasets.

2.1 Multi criteria Decision making process.

Multi criteria decision making methods have long existed in the research domain and have followed the existence of multi criteria recommendation system. Various papers have discussed the formation of a multi criteria decision making process [7]. The various steps to solve a multi criteria decision making process is through the discussion made by Bernard Roy [7].

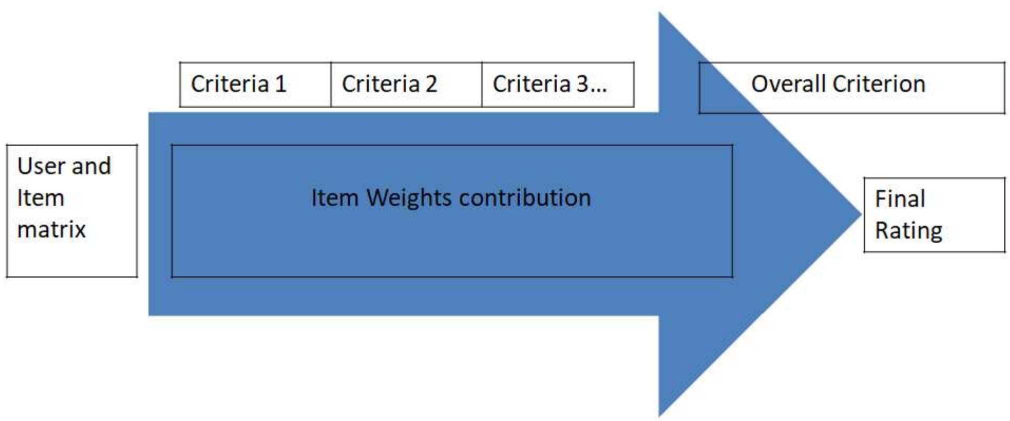

Fig. 2.1 Multi criteria Decision Making process

Figure 2.1 depicts the overall functionality of the multi dimensional decision making process. The use of this statistical technique has been extended to the domain of recommendation systems and has provided a huge scope for implementation of statistical as well as machine learning based approaches. This project has used this general functionality of multi criteria decision making along with other machine learning techniques to make an improved recommendation system.

2.2 Extending the MCDM framework for Recommendation Systems

The main categories that are used to classify the recommendation system into different classes are Content-based, Collaborative filtering, Knowledge based and Hybrid approaches.

Content-based recommendation systems take input of the user's past preferences and then make respective recommendations on how to generate future content. A collaborative filtering system computes a user's neighbourhood and the similar users to the users, and then predicts the user's data based on the other user's choices.

Multi criteria decision making systems can be in-depth analyzed and then divided into three broad categories:

- Content preference modeling based on Multiple-attribute.
    - Such systems typically use single criteria ratings given by any user in the past to compute the rating for future for that user.
    - As an example in case of movie recommender system these ratings can be movie genre or acting cast or directors etc.
- Content search and filtering based on Multiple attributes:
    - User typically record their general preference across multiple items, the model understands the commonalities for multi attribute content.
    - The underlying concept revolves around the idea that a user's past preferences would influence her future preferences as well.
- Rating-based preference on Multiple criteria
    - These systems take input of the user's rating of individual criteria that a user rates and then computes the overall ratings.
    - The individual ratings are used along with specific weight value to come up with an aggregated overall rating value.
    - This rating value is hence a reflection of user's preferences measured across multiple criteria.
    - Then the ratings are adjusted to compute or predict the future ratings for the same user in that domain.

2.3 Content preference modeling based on Multiple-attribute:

To model the user preferences of items we have to analyse the past history of items that she purchased. Many recommendation systems uses an approach call the content based recommendation system approach. This approach can be used as a standalone approach or along with the collaborative filtering approach. Such systems allow the user to input single criterion ratings.

Once the rating based on single criterion pints is made available to the users, similar items that the user might purchase or are linked to the previously available items are made available to the user. The content based techniques use common points among various items to understand the difference that the user may come across.

As a result the recommendations are made to the user based on the common points among various item points that exist on the item list.

Many traditional recommendation systems use these approaches along with a combination of similar or other set of approaches to understand the items that can be suggested to a particular user.

Several utility functions along with their respective evaluation criterion have been developed to understand the various combinations that agree with the quality of data available at hand. Content based user recommendation system uses user preferences and scores. The score may include information retrieval-based systems along with the user based modeling techniques applied to the recommendation systems, such as Bayesian classifiers.

<u>Rating-based preference on Multiple criteria</u>

Under this category of Rating based recommender systems, users can input ratings of multiple criteria. These ratings that are input by the user are an extension to the collaborative filtering technique used in traditional recommendation systems. The input fed to the systems is arranged such that it shows the user's subjective preference to various components of the item set.

As an example consider a user watches a movie and is then prompted to rate the movie. Now, instead of just one rating given to the movie, the user has to attribute multiple ratings to the movie's individual criterion. She may rate the genre of the movies in a particular fashion and the content is a different way. She may have a different rating for the movie's art direction. Hence, in this case the person's overall rating of the movie can be influenced by multiple factors of the movie.

These systems are different from the stated systems of this section, from the point that the user input does not indicate directly whether a user like an particular aspect of the movie better than the other, as in there is no choice of preferences given to the movie ratings. The ratings data set would consist just of the ratings that the users have given to individual criteria.

As an example of these kind of systems is the Smart Journey Recommendation system. In these systems users are given a travel bag and are asked to rate multiple items from that travel bag. The system would then take into consideration the individual items from the bag as criterion. The user may rate location of travel some way and the activities of travel in some other way.

Then the systems rate the travel plans of the user in a particular way that can be considered as more personal to the user's choices of travel and also aligns in a better way to the user's choices that were filled earlier.

Interest	User 1	User 2	User 3	User 4
Criteria 1	2	3	5	2
Criteria 2	4	1	3	3
Criteria 3	2	5	1	3
Criteria 4	4	1	3	2

Rate using ordinal scale: 5- Good; 1-Poor

Fig 2.2 User based preference rating on multiple criteria

To conclude, as evident from the above examples, the recommender systems employ traditional techniques as well as newer techniques like knowledge-based and content based system technique.

A certain mix of theses traditional as well as the model techniques can be seen in the implementation of the improved hybrid based recommendation system techniques.

Such example where these techniques are used across multiple criteria of recommendation systems, are classified under the umbrella of multi criteria recommender systems.

Since they model user preferences based on multiple attribute content of items that users preferred in the previous data that makes predictions for the users to point out the specific user generated content or other related choices, search or filtering conditions for multi attribute content of items.

Multi Criteria Rating Recommender systems:

This section would deal with the multi criteria recommender system techniques. We can firstly look into rating the techniques for a single criterion recommender systems and then extend it to implementation on multiple attribute oriented based recommender systems.

The section would also amplify the use of single rating recommender system, its advantage and disadvantages and the use of these systems that can be extent to multiple criteria systems.

Traditionally used techniques of recommendation system consume a two dimensional user item matrix. The utility function is to generally obtain a generalised set R0. The aim of the recommender systems is to predict the utility that an item exhibits for a particular user u.

The utility function is determined as:

$$R: \text{Items} \times \text{Users} \rightarrow R_0$$

Table 2.1: User Item matrix in the recommendation systems

User/ Movie	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User 1	5	7	5	7	?
User 2	5	7	5	7	9
User 3	6	6	6	6	5
…	…	…	…	…	…

The utility function that is used n traditional rating algorithm is based on user inputs, such as numeric data required. The numeric data can be ratings or the values that the user feeds to the system to evaluate their choice of that particular item.

Mostly used recommendation system take in the value that the user assigns a particular item in the item set. For example in the Table 1, shown above, consider that the user 1 has rated the movie 1 as shown in the ratings table.

Hence, $R(User1, Movie1) \; = 5$
Considering the ratings are assigned out of a rating scale of 10.

Let's take an example of collaborative filtering techniques to establish the rating the User1 would give to the item Movie 5. This method makes systematic estimations over the criteria set about the preference of a user by utilising the inputs of ratings provided. The assumed thought of the collaborative filtering approach is that if a user 1 has the grated an item 1 in particular way and so has the user B rated that item in a similar fashion on a set of items, the user 1 is more likely to have 2's choice for those items than that of another user that has been chosen randomly.

2.4 Implementation of traditional recommender systems to include the multi criteria Settings:

The growing appreciation of recommender systems and the growing demand of better personalisation techniques that users want on their content creators, has allowed the expansion of extending the technology of taking further the traditional recommendation techniques to next level.

The new generation of recommender systems tend to ask for a technology that has to be implemented to accomplish the personalisation of single rated techniques over traditional techniques.

Examples include:

- Yahoo movies
    - Criteria include story, action, direction and visuals
- Tripadvisor dataset,
    - Includes criteria like location, cost, ambience etc.

Content sources like the above stated example help to capture good information about user choices. This information is very helpful in making the performance of the recommender systems very efficient.

Considering the following convention:

The overall rating for a criterion can be taken as R_0

The individual rating for the user's criterion can be taken as:

R_1 R_2 R_3 R_4 …. R_k

For each individual criterion $c(c = 1, 2, 3 \dots k)$

Therefore the utility function can be modified as below:

R: $Users\ x\ Items$ $->$ R_1 x R_2 x R_3 x R_4 …. R_k.

Thus given the above modified version of utility function, the table 1 and tale 2 would tell a completely different story. Where in table 1 it seems that the user 2 was the closest neighbour to user 1 based on the rating criteria given on only one criterion. In table 2 we can very well estimate that the user 3 and user 1 have similar rating pattern over multiple criteria.

Thus taking into multiple criteria into account we could be computing User1's overall rating for item 5 based on the user 3's overall rating scheme for that item since they get more closely related to each other.

Table 2.2 User Item Matrix of a multi criteria Recommendation system

User/ Movie	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User 1	5,2,2,8,8	9,1,3,4,7	5,3,4,5,6	3,5,6,7,8	?????
User 2	5,2,2,8,8	5,8,3,4,5	6,4,5,6,7	7,5,6,7,3	9,1,2,4,6
User 3	6,2,4,7,8	6,4,6,7,8	2,3,5,3,5	4,5,6,4,6	5,3,5,6,7
…	…	…	…	…	…

This example implies that if we are just considering one overall rating we might be overlooking much information in the criteria values that the user may be following. Many aspects of an item come into picture when the users are given the ratings across multiple criteria of an item set to choose from. Also in such scenario the rating set is subjected to various algorithms which help the users to understand the role of implementing a particular idea behind making choices.

This does not only improve the overall efficiency of the recommendation systems, but also illustrates how including multiple criteria include more powerful and better focussed recommendations.

Hence we can see that new recommendation systems, better algorithms as well as newer techniques are required that can utilise the multiple criteria of data rating from the dataset to make more personalised user predictions.

2.5 Processes used for Multi Criteria Rating Recommender systems:

The process for recommendation system to work is divided into two phases:

- Prediction
  - Here a user's preference is calculated from the inputs supplied by the user.
  - In this phase, the utility function estimates the some part R of the user item matrix where the dataset maps to.
- Recommendation
  - In this step the test set data is used to predict unknown ratings for the users that have missing rating.
  - The users are recommended top N set of items that maximises his/her utility of the dataset.

Multi criteria ratings are user input dependent and can be used in both the above discussed scenario in different ways. Many approaches have been developed overtime to estimate the user ratings.

2.5.1 Process used in the Prediction step

This section of the report will deal with the use of algorithms that are used in the prediction step of recommendation systems. The prediction of overall ration as well as prediction of rating of individual criterion is one step where the algorithms would mean to be implemented.

In general the recommendation techniques can be classified into two classes [4]:

- Heuristic based approaches
    - These techniques make decision of prediction on the fly
    - They work on the observed data on the run time
    - They compute the utility of each item for a user.
    - An example is neighbourhood based approach.
    - In these approaches the overall similarity in the choices of two users is figured out and assumed that the users will exhibit similar rating pattern across multiple areas.
- Model based approaches.
    - These model based techniques work on a predictive model implemented on the test data set.
    - Typically these techniques use machine learning or a statistical model to observe the data.
    - These techniques can be expanded to a multi criteria problem.

2.5.2 Formulae used in Heuristic approaches:

The main concept behind using heuristic approaches is the use of formulae or the different objective function defined. These formulae are used as per the data that is available at hand and the dimensions that can be explored from the available data.
These approaches use similarity measures to understand the user behaviour from the dataset. The observed similarity values are then used over traditional similarity values on individual criteria. These individual similarity values are then aggregated into a common aggregation function.

Neighbourhood based collaborative filtering techniques computes the similarity between a test user and all the other users in the data set. It then uses that similarity matrix to understand the preference or say taste of the user in question. Hence the first step in such computation is the estimation of a similarity computation method.

Assuming,

$$R(u,i) \rightarrow rating\ given\ by\ user\ u\ to\ item\ i.$$

$$I(u,u') \rightarrow items\ that\ exists\ as\ common\ between\ user\ u\ and\ u'$$

Similarity measures that are mostly used:

- *Pearson correlation-based*:

$$sim(u,u') = \frac{\sum_{i\in I(u,u')}(R(u,i)-\overline{R(u)})(R(u',i)-\overline{R(u')})}{\sqrt{\sum_{i\in I(u,u')}(R(u,i)-\overline{R(u)})^2}\sqrt{\sum_{i\in I(u,u')}(R(u',i)-\overline{R(u')})^2}}$$

…………… (2.1)

- *Cosine-based*:

$$sim(u,u') = \frac{\sum_{i\in I(u,u')} R(u,i)R(u',i)}{\sqrt{\sum_{i\in I(u,u')} R(u,i)^2}\sqrt{\sum_{i\in I(u,u')} R(u',i)^2}}$$

…………… (2.2)

To extend the similarity computation methods to multi criteria recommendation system modification have to be made to the existing set of techniques.

These modifications would make the system understand the traditional methods to be implemented on a multi criteria dataset.

In multi criteria datasets,

$R(u,i)$ contains ratings r_0 which is the overall rating and

r_1 *to* r_k individual k criteria ratings.

I.e. $R(u,i) = (r_0\ r_1\ r_2\ r_3\ \ldots..\ r_k)^1$

Some research papers[8] discuss the need of an aggregation function and the associated techniques that would be used to come up with a personalised solution. As a general approach some papers [8], compute the aggregated criterion value as a general weighted sum of individual criterion values.

Hence, overall similarity values are a weighted sum of individual similarity coefficients that were used in various criterion of the dataset. Thus there are k+1 ratings in total for each pair of (u,i) instead of a single rating value.

Multiple aggregation techniques can be used to pair the individual rating values to come up with aggregated weight for each criteria and then come up to a single point of aggregation of values to rate a particular item in the item set[3].

Assuming:

The weight of each criterion 'c' is denoted by w_c

It represents how important or useful the given criteria is.

Below sets a given set of aggregation function that can be used to aggregate the individual criteria into the overall criterion value:

- *Average similarity*:

$$sim_{avg}(u,u') = \frac{1}{k+1} \sum_{c=0}^{k} sim_c(u,u')$$

- *Worst-case(smallest) similarity*:

$$sim_{min}(u,u') = \min_{c=0,\ldots,k} sim_c(u,u')$$

- *Aggregate similarity*:

$$sim_{aggregate}(u,u') = \sum_{c=0}^{k} w_c sim_c(u,u')$$

…………… (2.3)

Calculating Similarities using multi dimensional distance metrics:

Considering two users u and u'

The item that is to be considered is i from the item dataset

The distance d between user u and user u' on the item i:

d(R(u,i), R(u',i)) , can be calculated using following given distance metrics:

Manhattan distance:

$$\sum_{c=0}^{k} |R_c(u,i) - |R_c(u',i)|$$

Euclidean distance:

$$\sqrt{\sum_{c=0}^{k} |R_c(u,i) - R_c(u',i)|^2}$$

Chebyshev (or maximal value) distance:

$$\max_{c=0,\ldots,k} |R_c(u,i) - R_c(u',i)|$$

…………… (2.4)

The overall distance between the two set of users from the user data set column can be calculated as the average common distance of the rated item sets that are common to the two users. It can be used to calculate the similarity score among those two users on the basis of following formula:

$$dist(u,u') = \frac{1}{|I(u,u')|} \sum_{i \in I(u,u')} d(R(u,i), R(u',i))$$

…………… (2.5)

The better the similarity score among the two users, the better the similarity between them, this would mean that the users would have equivalently smaller distance among them for a better similarity score.

It is observed that the distance and similarity score are inversely related, hence to come up with a similarity score in accordance to the distance score compute among every set of two users from the user dataset, we can take an inverse of the distance calculated. In formula terms it can be represented as above given above.

2.5.3 Multi Criteria Ratings during Recommendation step:

As a part of recommendation step multiple criteria settings can be enhanced to provide recommendation on the predicted test dataset results as pr the models formulated in the training step.

The models may use user’s rating over both individual as well the overall ratings obtained by the user to calculate his preference for a given item to figure out the user’s utility function eventually.

If the overall rating criterion is chosen as the question point the whole process becomes quite straight forward in this scenario. The user’s preference of the item is judges based on the calculated score received for the overall rating criterion of the rating dataset. The most relevant items from the item dataset are recommended to the user to enhance his experience based on the score received for the overall item score of rating.

This process where the overall rating is considered to be the main focus point of decision making the process of recommendation of most relevant item to the user is done using the traditional recommendation system techniques.

In other cases the overall criteria rating values becomes less apparently useful to make the recommendation choice for the user. It may be due to various numbers of reasons, where the training data itself might not be very useful for the training process. Also the overall ratings might not be available for the users in the training set.

For example, we consider a set of movie recommendation system being implemented. We take the user input of rating the movies as ratings across two broader fields on a scale of 1 to 10. The ratings points can be story line and visual effects.

Suppose a person rates a particular movie as 5 on story line and 7 on visual effects, and then another user rates a movie 6 on story line and 6 on visual effects.

In the above scenario there is no overall movie rating criteria available to us that makes us understand which movie can be rated above or below than the other movie. This case makes it complex to understand the user's rating criteria value for a particular item set. Unless there is a modelling ay to map the user's individual criteria ratings to an overall criterion values using some numerical scale data, it will not be easy to understand which movie is better to be judged in a particular way.

2.6 Multi criteria Optimisation:

Such problems are widely studied in the operation research literature, although not in the context of mainly recommender systems, but focus on developing techniques that can be extended further to the recommendation system domain.

As an example we can consider the decision maker's choice of implementing a decision at an organisation. In this case he may have to take into consideration various criteria to be implemented to rate his/her employees. Such criteria as various points of conflicts or various recommendations of employees, human resource objectives or financial or socio-economic objectives or maybe environmental objectives met by the company can serve the purpose partially but not completely.

There are some approaches that can be fitted to the multi criteria recommendation system profiles and can be borrowed from the operation research domain into this domain. These approaches can really help make the whole system move forward. They can help the decision makers to better optimise the problems in various solution formats.

Optimising the operational research format in the multi criteria decision making formats would come across the main helping point in this research domain eventually. The following steps can be performed to understand the implementation of the optimisation techniques into the research domain.

- Looking out for the possible Pareto optimal points of solution that exists near to the original global maximum of the problem solution points.
- Taking some aggregation based on a machine learning or statistical model and then coming up with the optimised solution for the entire problem by splitting the k dimensional problem into k individual one dimensional problems and then aggregating based on the model that was trained.
- We can also figure out the most optimal solutions in case the domain expertise exists within the team and then optimise those criteria basically giving them more importance based on the domain expertise available.
- We can try to optimise on solution at a given iteration of time then figure out the possible trade off and then eventually make the decision of whether to keep the solution of to neglect it completely.

2.6.1 Finding the set of Pareto optima solution for item based recommendations.

In this technique a large number of possible candidate solution are suggested for an item set rather than implementing a single point solution for the item. The global optimised maximum is not taken into consideration in this problem rather a local set of optimisation solutions are taken into this problem.

The data is allocated to units called as DMU; Decision making units and frontier analysis criteria of decision making is then applied to those units to finally come up with a solution to form Data envelopment analysis or DEA. The best set of weight are defined for each DEU by the DEA and for this computation it does not take into consideration the implementation of regular A-priori instead of traditional weights measured throughout the report.

To take an example discussed in some other published papers, we can take a look at the works of Lee and Teng [8]. Their work focuses on the optimisation of multiple queries in the hotel data set to come up with a single point query to search for the best available hotel skyline to look at the required skyline point.

3. RESEARCH METHODOLOGY

This section proposes a newer algorithm to solve the problem of multi criteria recommendation systems. The algorithm uses techniques such as matrix factorisation steps from the SVD algorithms to predict the ratings into each criterion. Then the aggregation function has to be developed using the machine learning based egression algorithms. The section will then explain the overall frame of these algorithms used in the proposed implementation process and would then also explain the framework design. The implementation was performed in a python based environment for this section and the algorithm followed is depicted in a flow chart format at the starting of this section.

3.1 The Proposed algorithm

In broader sense the multi criteria proposed recommendation system frame work would consist of 4 major steps.

- The first step is to divide the overall k-dimensional multi criteria recommendation system to k 1-dimensional problem statements where each of the problem statement would carry its own recommendation systems iteratively.
- The second step is to predict the unknown ratings of the individual independent item set recommendation systems. These ratings would be predicted by the implementation of regular SVD recommendation technique.
- The third step would be to collate the predicted and the original rating dataset so as to formulate an aggregation function next.
- The fourth step would be to train the aggregation function and then predict the rating for the test set users.

Following this, as an evaluation step, appropriate evaluation metrics can be implemented to compare the results of the base and already published algorithms with the proposed algorithms.

In this project the SVD recommendation technique and Lasso based aggregation technique are used. The results obtained are compared against the entire major publish artificial neural network techniques and the other variants of SVD algorithms available.

The algorithms implemented in this process were:

- SVD for the recommendation steps
- Lasso and Ridge regression for the aggregation function steps
- As a base to compare the evaluation metrics:
    - SVD is used for recommendation steps
    - Ridge regression techniques were used for aggregation function.

The purpose of using the ridge regression techniques as a base and Lasso regression technique as the proposed algorithm is that the lasso techniques have an added advantage of making feature selection over ridge regression technique. Although both these techniques have better regularisation features over the traditional multiple linear regression techniques and is comparable in nature to general neural network methods available.

The algorithms stated below are implemented in a python jupyter notebook after following the data cleaning and the required pre-processing steps. The data sets used were Yahoo movies data set which is later described in this report.

Fig 3.1 Implementation of Proposed MCRS Algorithm

3.2 SVD algorithm for Recommendation Step

The singular value decomposition is a matrix factorisation as well as decomposition step that is traditionally used technique in machine learning. This technique is also a feature extraction technique used in various domains to extract out or pre-process various features from among a set of many features available to the users. This technique is sending inter changeability with the principle component analysis technique to understand the utilisation of features that can come up in a pre-processing step to reduce the dimensionality of the features availed at large.

The recommendation system uses the matrix factorisation part of the Singular value decomposition algorithm mainly and do not focus on the implementation of dimensionality reduction phase of the algorithm.

Matrix factorisation is technique where a higher order matrix is decomposed into a matrix of lower order. In such computations the lower order matrix when supplied with the appropriate operators can combine to for the higher order matrix. This technique can be represented as per the figure 11 shown below. The factorisation techniques are for example of a user based recommendation system implemented on a user item matrix. The user item matrix is factored into multiple matrices of lower order that represents important information from the recommendation systems into view. Hence Singular value decomposition algorithm decomposes the user item matrix into a set of matrices that contain vital information to be used further in the recommendation system.

The equation 3.1 below represents the factorisation step of SVD algorithm.

$$A \approx U \sum V^T = \sum_i \sigma_i \; u_i \; O \; v_i^T \text{ ....} \quad (3.1)$$

The above formula describes the factorisation of a bigger matrix to smaller units

A is the input matrix, can be a user item matrix or a user–user matrix or an item-item matrix depending on the type of recommendation system being implemented.

The matrix represented by u is the matrix of left singular vectors being taken into consideration which is an Eigen vector form of the original matrix features but in a lower dimension. The figure 3.2 describes the process of factorisation contained a a part of the SVD algorithm in a pictorial form.

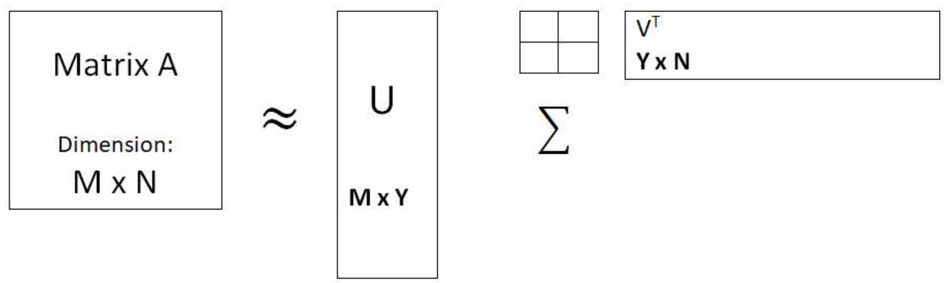

Fig 3.2 Matrix factorisation in SVD algorithm

The sigma matrix is a matrix of mostly singular matrices to account for the operators and reproduction of the original matrix in the form it is really required.

The matrix represented by v is the singular matrix on the right and is the other set of features decomposed from the user item matrix. This matrix is the transposed matrix of the features extracted to account or the rank o the original matrix.

The Singular value decomposition algorithm works on a theorem that says it is always possible to decompose a bigger matrix to smaller matrix units.

Commonly used online platforms like YouTube and Netflix use the implementation of recommendation systems to recommend movies or videos to its users to watch next. They consume the user's part data and the movies or the videos that users may have liked or dislike in the overall quality to implement their recommendations for the next video that they may be interested in watching. As described in the mage 3.3;

The image 3.3 taken from the acknowledged sources, published as a part of the book mining of massive datasets at Stanford university professors is a depiction of the matrix factorisation technique on an example dataset. It represents how a rating matrix is factored into three matrix based on the patterns observed.

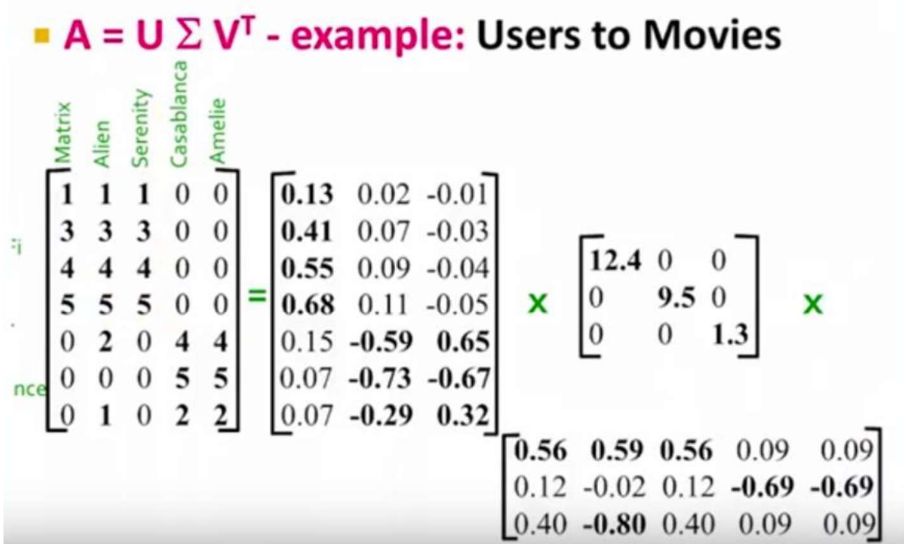

Fig 3.3 Matrix factorisation in SVD algorithm
[Image source. Mining of Massive Datasets by Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman, url: http://infolab.stanford.edu/~ullman/mmds/book.pdf]

Computationally extensive techniques are used consider the amount of data required to understand the implementation of these algorithms takes account of the understanding based on general human behaviour, how do the humans behave generally to account for the overall performance parameters implemented in these algorithms.

Considering the user item rating matrix shown below.

Table 3.1: User Item matrix, for SVD factorisation

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User 1	1	3	2	5	4
User 2	2	1	1	1	5
User 3	3	2	3	1	5
User 4	2	4	1	5	2

In general the Netflix or any such platforms databases of such rating matrix are very sparse in nature. The ratings do look very random in nature and are not exhibiting any predefined patters. In an ideal scenario if all the users rate all the items as 3, the next user's rating for the next item can be easily predicted. This is not the real case thought the real world data look sparer and scattered than we would traditionally expect.

The data hence would consist of some patterns if we examine closely. The sum of two rows can be a third row. This would mean that some users may rate the set of movies in a particular extremism fashion and moderate fashion that their rating criteria of the movies would eventually be same but the overall rating scores of these movies can be different.

Then comes a scenario where the rating scores of two columns may be an average of a third and fourth columns scores. This might be the case when items rated in similar fashion implemented to exist in a normal distribution f rating values. The items might be of some extreme importance to some people but be of a moderate value to others hence these rating patrons emerge out of such scenarios to make the ratings matrix factors able.

Such pattern are recognised from the original user item matrix and then used to identify features to compute the missing or the queried values. Another benefit of using these criteria is storage of the factorised matrix in multiple smaller matrixes also saves up a lot of information for the users to process the original matrix.

To decompose the bigger matrix into smaller decomposable matrix SVD algorithms uses iteration steps to compute the factorised model. The initial solutions are far from the original model values because the non-tuning of parameters that has been performed till that point. Once the first set of iteration values are compared to the original training set the values are then pointed towards the optimised error function. The error function is used to compare the predicted inerrable values of the matrixes and the original matrix values and then return a common result. The result can then be penalised or rewarded based on the applicability of the newer terms and then can be finally implemented into the desired domain.

A user matrix and item matrix are formed to be then taken as dot products for the combination into the bigger original matrix. Each item and each user can be represented as vector formats and then they can eventually be combined as dot products to finally reach the user item original rating value.

The expected rating values can be calculated as the first step of implementing the user item matrix as shown in the table 4. The objective is to organise the table 3 into two factors of user rating values and item rating values rating across specific genres of the movies taken as movie features from the factorisation algorithm applied on the domain data.

Table 3.2: Decomposed matrixes as Factors of Table 3.1

	Feature 1	Feature 2
User 1	1	2
User 2	2	3
User 3	2	1
User 4	3	2

	Item 1	Item 2	Item 3	Item 4	Item 5
F 1	3	4	4	3	2
F 2	4	2	2	3	1

The expected ratings are calculated by the predefined parameters of the SVD algorithm to generate the overall criteria value and the ratings can be compared against the original rating value from the matrix of training data to compute the error in the computation of the result.

The results are then regularised against the error function defined to come u with optimised parameters and when the expected results match the original value the training of the matrix can terminate to output the best known results.

The regularisation also stops the model from over fitting the parameters of SVD algorithm to fit the training data or to under fit the training data. The penalty term in the equation can be added to generate a noise that can be also optimised so that the overall model performs better on the testing data.

The use of bias term is often very important in the implementation of the SVD algorithm. For the rating value among use item pair, the bias terms calculated can actually predict the use of error function data to finally reduce the error between original and predicted values of the rating matrix. Hence the final optimised solution of the overall recommendation systems can be made available using the matrix factorisation part of the SVD algorithm. For the implementation of a regularised error function stochastic gradient descent algorithm can be used.

3.3 Regularisation based Regression techniques for Aggregation Function

A multiple linear regression technique has an optimised linear function that fits the training dataset set according to an objective function. The objective function would be either minimising the RMSE or minimising the MAE value of the overall criteria. Moreover the regression coefficients would try to fit the training data in such a way that the errors defined in the objective function would be reduced for as long as possible.

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2 \qquad \text{...............} (3.2)$$

Cost function for Simple Linear Regression

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}|w_j| \qquad \text{...............} (3.3)$$

Cost function for Lasso Regression

$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}w_j^2 \qquad \text{...............} (3.4)$$

Cost function for Ridge Regression

This optimisation of the error variable in the overall fitting of data in linear regression tends to over fit the training data sometimes. This is a very frequently observed nature of liner regression coefficients that once the training data is over fitted, the other data i.e. the testing data is not always aligned with the training data. The machine learning or statistical models must be so trained that they exhibit a nature of understanding the features of the training data and then replicating those details on the testing data.

The cost function as evident in the linear regression problem is the function that optimises the regression equation so as it fits the overall regression line as perfectly as possible. It is due to the optimisation of this cost function that the linear regression coefficient tends to over fit the training data.

The changing of the cost function in the L1 and L2 regularised regression techniques is the key advantage to overcome this problem of linear regression. Ridge and Lasso regression techniques are hence often called the techniques that do not over fit the data of the training part and hence tend to perform better predictions on the testing part of the data. This technique not only increases the performance accuracy of the statistical or the machine learning model that was implemented but also reduces the time it takes for the more complex computations being performed on the testing dataset. The overall model complexity is also gravely reduced by this implementation.

3.3.1 Ridge Regression technique:

As evident in the equation 3.1 to 3.3, the ridge regression part is changing the cost function of the multiple linear regressions by adding a noise function to it. The noise function would comprise of the squared terms so as to not decrease the coefficients. This noise is added as a penalty value calculated from among the coefficients. This is purposefully done to activate the noise function in such a way that it does not adds up random noises to the function but instead adds such points to the function that not only reduces the model complexity but also adds value to the model.

The cost function noise is the added squared magnitude of the difference of coefficients that update the overall ridge regression coefficients so as to boost the regression model's accuracy on the testing data. This model is also subject to some added incentives by optimising the noise function so that it does not deviates much far away from the actual model.

As we can notice in the equation 3.1 to 3.3 the coefficient w is optimised in this ridge regression technique. The constrain put are such that optimises the noise terms effectively enough to compute the best overall model performance. The lambda terms is chosen iteratively through the use of various values to magnify or dim the effect of the constraint term 'w'.

This choosing of various lambda terms and optimising the 'w' constraint value not only does reduces the complexity of the regression model coefficients but also reduces the multi co linearity of the different terms in the coefficients.

It is quite evident if the lambda constraints on the ridge regression coefficients are reduced to zero the cost function would eventually resemble the linear regression cost function. Hence this analogy shows the use of properly optimising the lambda values in the implementation of the regression coefficients.

The lambda function very close to zero would make the ridge regression coefficients almost behave similar to the multiple linear regression coefficients. If the lambda function values increases close to one it would tend to form rippers in the regression line fitting the training data.

Due to this lambda parameter of the ridge regression, the training accuracy would not be as high as the training accuracy of the multiple linear regression models. This is quite obvious as in case of multiple linear regressions the regression line would fit the data points as close as possible. In case the ridge regression coefficients are chosen at a higher lambda value, the regression line tends to shift away from the training data points. This training data accuracy loss comes with a reward of testing data accuracy gains. As due to accounting for such precedence patterns in the data distribution the testing data might not behave exactly like the training data. Ridge regression would still give good accuracy results in such cases.

3.3.2 Lasso Regression technique:

As evident in the figure 14, the lasso regression part is changing the cost function of the multiple linear regressions by adding computed noise function to it. The noise function would comprise of the magnitude directly from the coefficients into the previously formulated model. This noise is added as a penalty value calculated from among the coefficients. This is purposefully done to activate the noise function in such a way that it does not adds up random noises to the function but instead adds such points to the function that not only reduces the model complexity but also adds value to the model.

The cost function noise is the added magnitude of the coefficients that update the overall lasso regression coefficients so as to boost the regression model's accuracy on the testing data. This model is also subject to some added incentives by optimising the noise function so that it does not deviates much far away from the actual model.

This choosing of various lambda terms and optimising the 'w' constraint value not only does reduces the complexity of the regression model coefficients but also reduces the multi co linearity of the different terms in the coefficients.

It is quite evident if the lambda constraints on the ridge regression coefficients are reduced to zero the cost function would eventually resemble the linear regression cost function. Hence this analogy shows the use of properly optimising the lambda values in the implementation of the regression coefficients.

As we can notice in the figure 14 the coefficient w is optimised in this lasso regression technique, in a similar fashion as was done for the ridge regression part. The constrain put are such that optimises the noise terms effectively enough to compute the best overall model performance. The lambda terms is chosen iteratively through the use of various values to magnify or dim the effect of the constraint term 'w'.

The lambda function very close to zero would make the lasso regression coefficients almost behave similar to the multiple linear regression coefficients. If the lambda function values increases close to one it would tend to form rippers in the regression line fitting the training data.

One important feature to be notices here is that the coefficients can also go to zero in case of lasso regression but not in the case of ridge regression. This would mean that the model here would not only implement the a better performing model, but also implement a model which can extract out only the required features out of the set of available features.

In a similar fashion as was done for the ridge regression part, lambda parameter of the lasso based regression; the training accuracy would not be as high as the training accuracy of the multiple linear regression models.

This is quite obvious as in case of multiple linear regressions the regression line would fit the data points as close as possible. In case the lasso regression coefficients are chosen at a higher lambda value, the regression line tends to shift away from the training data points. This training data accuracy loss comes with a reward of testing data accuracy gains. As due to accounting for such precedence patterns in the data distribution the testing data might not behave exactly like the training data.

3.3.2.1 Using Lasso Regression model for feature selection

One of the main reasons in this project to implement lasso based regularised regression technique is to implement this very feature of the lasso regression where, the lasso regression techniques are also able to select features from as set of features available as a part of multi criteria regression techniques available to the users.

A multi criteria dataset contains a number of criteria available to the users to rate the item set in. Often these users do not care about a particular criterion, may leave it blank or may give general or some extreme rating into that pattern. This often comes in a situation when any kind of data, be it if the user has intentionally left the column blank would come in handy while creating a model for this user's preference.

A null value often suggests that the user is not very concerned about a criterion. Another technique to find out the pattern in the user's rating methodology to figure out whether the user is at all concerned about a criteria a lot or he is just putting random numbers into the criteria but that criteria is not at all affecting his rate of getting recommended that very item.

Lasso regression's feature selection capability gives an added advantage when used as an aggregation function step in such cases. The coefficients which are reduced to zero for a particular user would simply represent than the particular criteria in the rating matrix is not very useful for the overall decision meaning process. This is also very valuable information to the organisation who wants to fine tune their data collection methodology.

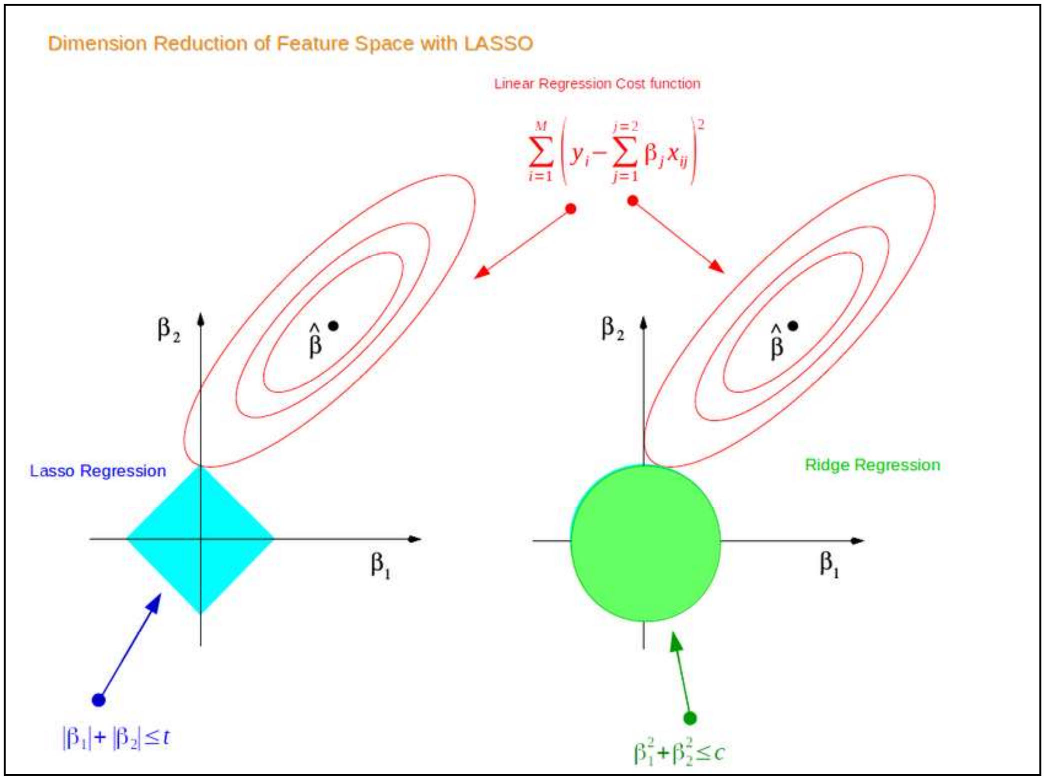

Fig 3.4 Exploring Feature selection process of Lasso regression
[Source: The Elements of Statistical Learning by Hastie, Tibshirani and Friedman, second edition, https://web.stanford.edu/~hastie/Papers/ESLII.pdf]

As evident from the figure 3.4, in a two dimensional space both ridge and lasso regression coefficients on getting generous amount of lambda relaxation would come close to the elliptical boundaries. Both the method finds the global maximal point of coefficient by making a constraint on the feature boundary. It can be visually identified that when the ellipse and the coloured spaces of the coefficients meet in case of Lasso regression, one of the coefficients from the diamond boundary will eventually become zero, this formulates the feature selection capability of lasso regression.

3.4 Model based aggregation function Approach

Aggregation function is a key aspect of getting the overall rating criterion from the individual rating criteria. In this project a lot of techniques that would eventually form an aggregation function were discussed and tried for implementation. To further add to the representation of the aggregation function implemented in this project we have formulated this section. In this section the main process of utilising the individual criterion are understood and then the next step is taken to formulate a strategy to implement the existing criteria value to an overall criteria values.

Many such techniques like regression based aggregation function techniques or probability based aggregation function techniques or neural network based aggregation function techniques exists overall in the domain of machine learning. Some research has also been formulated across the implementation of these techniques into the broader spectrum of recommendation system related research.

The main objective of this aggregation function technique is to find a utility function 'f' that maps all the individual criteria values to an overall criterion value. Using machine learning and statistical analysis tools the pattern in the data can be studied to make a model to learn this mapping of individual criteria value to the overall functional criterion value as evident in the equation shown below in equation 3.5.

$$r_0 = f\,(r_1, \ldots, r_k) \qquad \ldots\ldots\ldots\ldots\ldots (3.5)$$

As an example, considering a book recommendation system as a problem statement. The aspects that the users rate the book recommendation systems may be the width of the book, story line and characterisation. If we consider the storyline of the book to be really the highest priority among the book ratings then the aggregation function must give the highest priority to the story criteria. In case the width of the book is not exactly a recommended criteria, aggregation function 'f' must be made in such a manner that overall facilities should neglect the use of implementing the use of that criteria while recommending books to the user.

The aggregation function formulation can be divided into three broader steps.

The first step would be to divide the k criteria recommendation problem into smaller parts. These parts can be understood as individual recommendation systems and can be iteratively dealt for the recommendation system problems identically. Hence, as a result of this step the overall problem would be broken down to k- recommendation system problems, with each problem having its set of solution.

The second step would be the formulation of the aggregation function. In this step, we can understand that the overall criteria are an aggregation or a weighted sum of different techniques. These techniques can then represent the overall criteria of recommendation. The choice of aggregation function can be made based on:

- Domain expertise knowledge to weight each criterion accordingly as per the importance
- Statistical analysis on some simple data that can represent the whole situation
- Machine learning based techniques on the entire data that understands the coefficients associated with each criterion really well.

Finally the final step in this aggregation function technique is to compute the overall rating criterion based on the individual criterion and the developed utility function 'f'. The figure 3.4 is taken from a published and acknowledged source and represents the aggregation function technique adapted in this project.

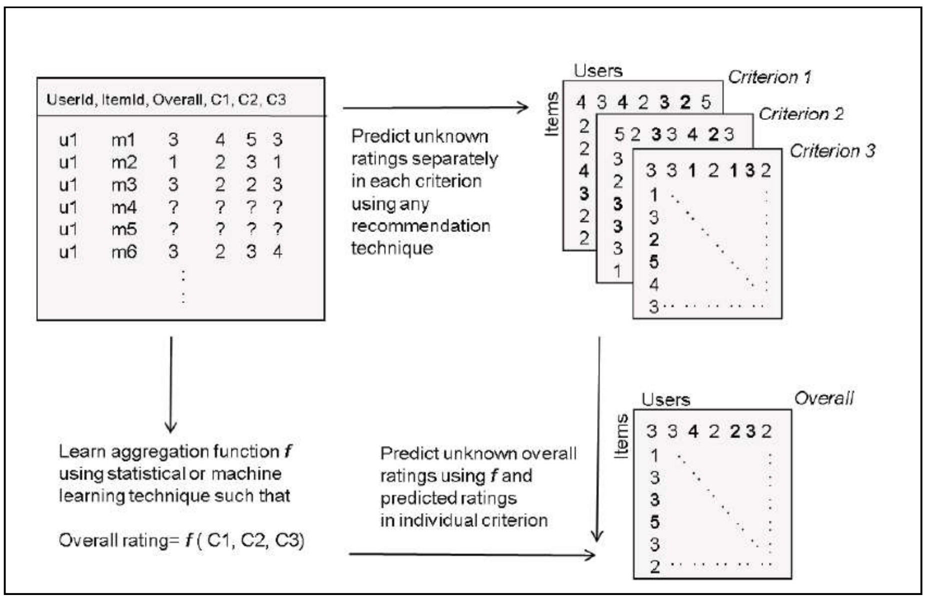

Fig 3.5 Overall aggregation function
[Source: Multi-Criteria Recommender Systems by Adomavicius, Manouselis and Kwon]

Other techniques that are discussed in the source paper [9] of image 3.4 are the similarity based techniques that use heuristic methods that can be developed such that the collaborative based recommendation stems can benefit hugely from them. As we have discussed that collaborative recommendation systems are not always useful, most of the time we use other model based or knowledge based recommendation systems as well to understand the applications of recommendation system on various domains. In those cases the recommendation system are then used along with the recommended aggregator based technique. These recommendation systems do not fit very well with the heuristic based methods.

Other techniques that are discussed for optimising the rating matrix versus the optimising the user item matrix to then be aggregated to get the final modelled value of recommendation system. These cases are very peculiar in the sense that the recommendation techniques are implemented so as to get higher accuracy in the overall performance of the recommendation systems.

In some research papers [2] the use of linear regression are implemented as a model to lean coefficients on the individual criteria values to learn the preference of those criteria values for the overall rating value. As discussed earlier in this project report the overall rating values are better recommended using the capabilities of lasso regression technique. That is the main objective also of this project to understand and check whether the recommended techniques of lasso and ridge regression are in case better performing than the other published techniques of multiple linear regressions. Also some papers have used the implementation of neural networks to estimate the regression coefficients, in this project report we would be comparing our results with those published results as well.

4 RESULTS

This section begins with the exploration of the data set used in this project to implement the proposed algorithm and compare the results against the standardized benchmarks. One such publicly available dataset in this domain is the yahoo movies dataset described in detail in this section. Following to the data description experiments to validate the applicability of the proposed algorithms are defined and the results obtained are recorded. These results are then compared against the published results of research from this domain. It is observed that the proposed results were competitive and in most of the cases better than the existing results.

4.1 Dataset used for Experiments

This section of the project report discusses the summary of the publically available datasets that are commonly standardised and used across the domain to express and compute the task at hand. The research that has been published in this domain has worked across these datasets and hence the datasets have also been used to identify the various evaluation metrics so as to analyse the research is really valuable or not in its true nature.

4.1.1 Yahoo movies data set:

This dataset is a multi criteria dataset publically available from the yahoo servers. In this domain yahoo has collected users' rating across multiple criteria such as direction of the movie, action used, story line and visual effects. The ratings of the movie were given across a 13 point scale; this scale depicted the distribution of user ratings to the movies across the domain. In this scale the least rating was F and the highest rating that any movies could be rated by any user was A+. This was the raw form of the data available from the website. This data had to be cleaned for any anomalies that would be present.

	user_id	criterion1	criterion2	criterion3	criterion4	overall	movie_id	number
0	1	6	6	8	12	8	2	1
1	1	9	11	10	9	10	26	2
2	1	6	10	9	8	7	61	3
3	1	6	6	6	5	5	86	4
4	1	10	11	10	9	10	132	5

Fig 4.1 Overview of the data

The rating data set also consisted of an overall rating criterion which was the key aspect as to why this dataset fits the whole purpose of the research domain. This overall rating dataset has to be eventually modelled through the various machine learning and statistical models. As shown above, figure 4.1 represents the description about the data set after data was cleaned and pre-processed.

As seen in the figure 4.1 the dataset has multiple criteria ratings available as a part of the raw data. This data was pre-processed for numerical machine learning models. The rating was kept at a nominal scale of 13 point itself. The rating of F was mapped to 1 and rating of A+ was mapped to 13.

In case any rating was missing it was replaced by zero. It was very important to preserve the zero ratings of the movies in the dataset. In most of the pre processing steps it is recommended to delete the rows containing the zero values. In this scenario the zero rating would contain a very valuable information as to the dataset would be sparse in nature. Mostly the data sets that a recommender system processes are sparse in nature and the missing value conveys the preconceived information that not all the users rate all the movies they watch. Hence we would train and then test the data only for that user item airs where there exists a rating value between the users and the respective items.

The dataset was already converted to a numeric scale to make it standardised across the platform of research and acceptable for the various numerical machine learning techniques to be able to process the data more effectively.

```
data_full.shape

(62156, 8)

print( "The data has ", len(data_full.movie_id.unique
<

The data has  976  uniques id and 6078 unique users
```

4.1.2 Descriptive statistics of the data

After the pre- processing the above table as shown in figure 4.2 was obtained representing the exploratory details of the cleaned dataset.

Apart from the above stated exploratory steps, data cleaning is also required to be performed. To clean this data set we had to remove the users from the data set that had rated less than 5 movies in total. This not only increased the uniformity in the entire data set but it also made sure that to predict the ratings for a user in the test set we should have enough ratings for that user at the training level to understand the overall requirements for the user. This step was also important to maintain the standardised format of data used in this research domain.

To further explore the data set the following frequency distribution table was formulated by retrieving queried information from the data.

Table 4.1: Frequency distribution statistics of the data

Value	Frequency	Percentage	Cumulative
1	3395	6	6
2	1340	2	8
3	1522	2	10
4	1329	2	12
5	2051	3	15
6	2428	4	19
7	2489	4	23
8	3251	5	28
9	5586	9	37
10	7006	11	48

11	6702	11	59
12	12153	20	79
13	12904	21	100

The exploratory analysis of the frequency distribution gave the above table as the outcome. This shows that most of the users have rated the movies in the higher quadrants. Also it shows that movies are rated evenly in the poorer sections while are rated highly in the higher sections. This not only tells us about the rating patterns of the users but also makes a list of items that would be rather higher rated than the other items. These exploratory steps give very vital information about the distribution of the data and should be kept in mind while making the final model.

4.2 Experiments performed:

Out of the various evaluation metrics available to be used, this project has used to evaluate the proposed algorithm's results with the actual results that are published across similar researches [10].

- Root Mean square error
  - This was used in published documents
  - This method predicts the rating accuracy
- F1 score of the usage prediction
  - Precision and recall both give a better sense of the overall performance of any machine learning model
  - The f1 score is a harmonic mean of precision and recall values, gives a representative score of the ratings.

The following tables are recorded from the accuracy results of this project's implemented algorithms and the same the corresponding published results from various research papers published in this domain.

4.2.1 Experiment to measure Accuracy

The results were cross validated using 5-fold cross validation technique and multiple experiments were done to evaluate the final ratings on to the evaluation metrics. The results were compared against the set of published results [10].

Table 4.2 Accuracy comparison of various algorithms

	MAE	R_squared
Single criterion CF	2.2406	0.8477
Multi-criteria CF-A and Overall	2.2367	0.8483
MCCF-Average similarity	2.2191	0.8811
MCCF-Minimum similarity	2.2434	0.8458
MCCF-A	2.2227	0.8786
MCCF- Multiple Regression	2.1995	0.9108
SVD - Lasso Regression	2.1981	0.8617
SVD - Ridge Regression	2.2109	0.8428

As evident from the results above, the proposed algorithm does provide some better results as compared to the existing algorithm results in this domain. Subjected to parameter tuning the results can even be further corrected so as to better suite the training data patterns and perform better on the testing data.

4.2.2 Experiment on different size of datasets

Another experiment was conducted with different set of users taken for the training data in the first step of the algorithm. This experiment was done to check the robustness of the proposed frame work whether this frame works would sustain user databases of different sizes. This was also done using 5 fold cross validation techniques and was hence formulated to be in the correct framework of published results [10] for comparison with other techniques that are already implemented.

Table 4.3: Accuracy comparison over varying size of data

	Performance Measures	Y 1000	Y 2000	Y 3000	Y 4000	Y 5000	Y 6078
	MAE	2.3819	2.2.2803	2.2939	2.2539	2.2343	2.2406
SCCF	R_sq	0.8132	0.828	0.88271	0.837	0.8403	0.8377
	MAE	2.3861	2.2952	2.2982	2.2466	2.239	2.2367
MCCF-AO	R_sq	0.812	0.8291	0.8283	0.8369	0.8405	0.8386
	MAE	2.2396	2.2596	2.2692	2.2217	2.2164	2.22191
MCCF-A	R_sq	0.8165	0.8344	0.8354	0.845	0.8455	0.8452
	MAE	2.3842	2.2927	2.2934	2.2506	2.2387	2.2434
MCCF-MO	R_sq	0.8088	0.8267	0.829	0.8354	0.8404	0.8367
MCCF-MA	MAE	2.3521	2.2622	2.2586	2.2293	2.2183	2.2227

	R_sq	0.8137	0.8336	0.8364	0.8447	0.8454	0.844
CF- Multiple Linear Regression	MAE	2.2212	2.2352	2.1987	2.1731	2.203	2.1995
	R_sq	0.8457	0.8394	0.8435	0.8509	0.8484	0.85
SVD - Lasso Regression	MAE	2.2201	2.2341	2.1976	2.172	2.2019	2.1984
	R_sq	0.8077	0.8256	0.8279	0.8343	0.8392	0.8356
SVD - Ridge Regression	MAE	2.2493	2.2633	2.2268	2.2012	2.2311	2.2276
	R_sq	0.8166	0.8345	0.8368	0.8432	0.8481	0.8445

As evident from the results above, the proposed algorithm does provide some better results as compared to the existing algorithm results in this domain. Subjected to parameter tuning the results can even be further corrected so as to better suited towards the training data patterns and perform better on the testing data.

5 FINDINGS AND RECOMMENDATIONS

This section deals with the analysis of several findings that were observed during the course of this curse and the recommendations towards future research directions that can shape well with the study.

The field of recommendation systems is a continuously evolving one. In the recent times this field is expanded to incorporate multi criteria stings as well as various other dimensionality constraints. These are new systems which offer many new opportunities of development to be enlisted in this section. Traditional recommendation systems have implemented newer techniques like matrix factorisation and have produced very impressive results. The MCRSs use those techniques with the aggregation function techniques to construct a predictive model. The experiments showed that the proposed results were competitive and in most of the cases better than the existing results.

It is also recommended to try alternatively available techniques like SVDpp and Bayesian reasoning, tree based regression techniques to assess their performances with respect to the proposed algorithms. Moreover the social media datasets could also provide great amount of opportunities using such recommendation systems, additional frameworks can be incorporated from the platform to collect data from the users and create hybrid model for implementation.

Multi criteria recommendation as filtering techniques:
Taking an example of a user wants just to understand the set of hotels that were rated more than 9 in a particular city or an area, he can just use the criteria based recommendation systems to filter out eh results that don't fit his query.

This approach is similar to context aware recommendation systems that can be developed gravely based on the recommendation techniques of content based recommendation systems but instead of the content optimisation the context of the recommendation query becomes the key to understand and perform the iteration of selection the best optimal user item pair based on the user queries.
After the implementation of the recommended algorithms in this project report it was observed that tuning of the parameter values would affect the performance of the algorithms. The techniques discussed in some papers [11], [12] also discuss the

newer techniques of multi criteria decision making. These techniques can also act in addition to the machine learning algorithms to even improve the accuracy.

6 LIMITATIONS OF THE STUDY

In this section, several possible limitations that the recommendation systems would still face in today's implementation and can be taken as opportunities of development of these systems into better performing predictive models. It is believed that research during this area is just in its preliminary stages, and there is style of possible additional topics that might be explored to advance multi-criteria recommender systems.

- The first limitation that affects the performance of a multi criteria recommendation system is the proper utilization of additional information received through multi-criteria ratings. This can lead to an evident issue of indiscretion. For better performance of recommender system in terms of output, users have to provide the system with a certain amount of feedback about their preferences.
- It is crucial to understand the cost and benefit of adopting multi-criteria ratings and reach an optimal solution where both the users and system designer's needs are met.
- Performing user studies on multi-criteria recommender systems would help in further understanding of the impact of submitting more ratings on the overall user satisfaction.
- Multi-criteria recommender systems generally require the users to offer more data to such systems as compared to their single-rating counterparts, hence enhancing the likelihood of obtaining missing or incomplete data.
- The usability of other existing techniques in this scenario should be explored, and novel techniques could be developed in view of the specifics of multi-criteria information, for instance the likely relationships between different criteria.
- A user's preference for an item in multi-criteria rating settings are often predicted by combining the preferences supported different rating criteria. Additionally, there may well be multiple goals for aggregating individual preferences, like maximizing average user satisfaction, reducing

misery (i.e., high user dissatisfaction), and providing a specific level of fairness (e.g., low variance with the identical average user satisfaction). Multi-criteria rating recommenders could explore the usage of a number of these approaches for aggregating preferences from multiple criteria.

- Multi-criteria rating datasets that will be used for algorithm testing and parameterization are rare. For this new area of recommender systems to realize success, it's crucial to possess style of standardized real-world multi-criteria rating datasets available to the research community.

- Some initial steps towards a more standardized representation, reusability, and interoperability of multi-criteria rating datasets are taken in other application domains, like e-learning.

7 REFERENCES

1 Meo, P.D.; Musial-Gabrys, K.; Rosaci, D.; Sarnè, G.M.; Aroyo, L. Using centrality measures to predict helpfulness-based reputation in trust networks. ACM Trans. Internet Technol. (TOIT) 2017, 17, 8.

2 Kant, V., Bharadwaj, K.: Integrating collaborative and reclusive methods for effective recommendations: a fuzzy bayesian approach. Int. J. Int. Syst. 28(11), 1099– 1123 (2013)

3 Adomavicius, G., Kwon, Y. New Recommendation Techniques for Multi-Criteria Rating Systems. IEEE Intelligent Systems, 22(3):48–55, 2007.

4 Adomavicius, G., Tuzhilin, A. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6):734–749, 2005.

5 Manouselis, N., Costopoulou, C. Overview of design options for neighborhood-based collaborative filtering systems. Personalized Information Retrieval and Access: Concepts, Methods and Practices, pages 30–54, 2008.

6 Mukherjee, R., Dutta, P.S., Jonsdottir, G., Sen, S MOVIES2GO: An Online Voting Based Movie Recommender System. In Proc. of the 5th International Conference on Autonomous Agents, pages 114–115. Montreal, Canada, 2001.

7 Wang, P. Recommendation Based on Personal Preference. Computational Web Intelligence: Intelligent Technology for Web Applications, in Zhang Y., Kandel A., Lin T., and Yao Y.(Eds.), World Scientific Publishing Company, 2004.

8 Roy, B., McCord, M.R. Multicriteria Methodology for Decision Aiding. Springer, 1996.

9 Lee, H.H., Teng,W.G. Incorporating Multi-Criteria Ratings in Recommendation Systems. In IEEE International Conference on Information Reuse and Integration, pages 273–278, 2007

10 A Neural Networks Approach for Improving the Accuracy of Multi-Criteria Recommender Systems; Mohammed Hassan and Mohamed Hamada Applied Sciences MDPI, 2017.

11 A Linear Regression Approach to Multi-criteria Recommender System; Tanisha Jhalani1, Vibhor Kant(B), and Pragya Dwivedi, International Conference on Data Mining and Big Data, 2016.

12 The Elements of Statistical Learning by Hastie, Tibshirani and Friedman, second edition.

13 Mining of Massive Datasets by Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman.

14 Masthoff, J. Modeling the Multiple People That Are Me. In Proc. of the International Conference on User Modelling (UM2003), pages 258–262. Johnstown, USA, 2003.

15 Smyth, B. Case-Based Recommendation. The Adaptive Web: Methods and Strategies of Web Personalization, 2007.

8 PLAGIARISM REPORT

Multi Criteria Recommendation System using SVD and Lasso Regression

ORIGINALITY REPORT

14%	10%	11%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.cse.iitk.ac.in Internet Source	3%
2	link.springer.com Internet Source	2%
3	"Data Mining and Big Data", Springer Science and Business Media LLC, 2016 Publication	1%
4	ids.csom.umn.edu Internet Source	1%
5	mdpi.com Internet Source	1%
6	www.bartleby.com Internet Source	1%
7	Submitted to Universiti Teknologi MARA Student Paper	<1%
8	"Recommender Systems Handbook", Springer Science and Business Media LLC, 2015 Publication	<1%

Executive Summary

In this report, we aimed to provide an implementation of multi-criteria recommender system. The report first used and explained the recommendation problem as an Mutli criteria decision making concept from statistics. Then it visited the problems associated and reviewed the MCDM methods and algorithms that can be extended to the domain of multi-criteria recommendation systems. Then, the report presented the category of multi-criteria rating recommendation systems, i.e., techniques that provide recommendations by modelling a user's utility for an item as a vector of ratings along various criteria. After which the study of existing methods and algorithms in the area was conducted to understand the area of opportunities where a newer system can be implemented. This report then implemented and proposed an improved version of a newer technique that can solve the multi criteria problems more efficiently than the existing set of techniques. Finally it can be understood that with the technologies changing so rapidly and the domain of recommendation systems being an adaptive domain, exploring this problem-rich area can be very fruitful in terms of newer research. More research and development into various algorithms and techniques are required to understand the whole potential of multi-criteria recommendation systems.