Preface

Recommender systems have developed in parallel with the web. With the development of web, the information available online increased at an exponential rate. This information overload required a system which could remove redundant information and provide the most valuable information to a user in minimum time. Collaborative Filtering is one the most accurate and widely adopted approaches for providing such information. It has found its application in domains ranging from e-commerce and e-learning to social networks and web search. Owing to its vast field, techniques, and challenges pertaining to collaborative filtering requires it to be conglomerated at one place to understand its underlying principle, working and application in its entirety. Collaborative filtering finds its roots in data-mining.

Data mining is finding hidden and unknown information from inside large databases. Data mining tools and techniques are finding its immense applications in the modern day. Collaborative filtering using data mining will widen the application area and more interest will be created in budding researchers to pursue their research in the same. The implications of data mining can be understood by the fact that whether it's a public or private sector organization, all are taking the advantage of the data mining tools and techniques to reveal the hidden and unknown information from the available data. This has been widened primarily because of the large or can we say terabyte of data which is collected by all the organizations over the year and they are confused as how to use such a bulk of data. The new and emerging areas of data mining techniques have surprised many researchers and business persons who are gaining a lot of hidden and unknown information for increasing their ROI. Collaborative Filtering is one the most accurate and widely-adopted approaches for providing such information. It has found its application in domains ranging from e-commerce and e-learning to social networks and web search. The primarily techniques of data mining are:

1. **Classification:** A supervised learning-based technique in which different items are classified into target classes. This technique is used in the cases where the exact prediction is required. In this, a training set is prepared that finds the association between the values of predictors and the target. The target is the value assigned to the class and the predictor is the value associated with the domain whose target class needs to be found. The major classification techniques employed are Naïve Bayes Algorithm, Decision Tree and Support Vector Machines (SVM). This technique finds the significant application in the detection of credit card fraud, and suspicious emails.
2. **Clustering:** Cluster analysis, or clustering, is the exercise of taking a set of objects and dividing them into groups in such a way that the objects in the same groups are more similar to each other according to a set of parameters than to those in other groups. These groups are known as clusters. Cluster analysis is one of the main tasks in the field of data mining and is a commonly used

technique for the statistical analysis of data. Cluster analysis does not refer to an algorithm but an exercise that has to be undertaken on the given data set. Various algorithms can be used for cluster analysis. The algorithms are divided into various categories and they differ significantly in their idea of what a cluster is constituted of and how the clusters are identified. The most popular ideas on the basis of which clusters are defined and identified include groups with small distances among the constituent members, areas of the data space which are highly dense, intervals or particular distributions. Clustering is a multi-objective problem that it is a mathematical optimization problem. A clustering algorithm consists of parameter settings such as a distance function, a density threshold (the number of clusters expected to be formed). Based on the available data set and the use of result as intended by the user, apt clustering algorithm may be used.

3. **Association Rule Mining:** In association rule mining, the association between item sets are considered or found with the help of Support and Confidence. The Rule are framed according to the data values and corresponding relationship between them.
4. **Neural Network:** A Neural Network (NN) is used to recognize patterns in data. The data can be specified according to the different domains like Financial Fraud including Credit Card Fraud detection and phishing, etc. NNs are used for those problems where the exact solution is not required, such that this technique is not sensitive to errors. Some common types of NNs are Artificial Neural Networks (ANNs) and Multilayer Artificial Neural Networks (MNNs).
5. **Genetic Algorithms:** Genetic Algorithms (GAs) predict using generated logic rules and fitness functions in order to detect financial fraud and suspicious e-mails. The major steps used are Mutation, Inheritance, Selection and Crossover. GAs and NNs can be used in combination to solve a complex problem. Every model inherits traits from previous models and compares it with the other models to more accurately model remains. It is based on the theory of the survival of the fittest, which means that the model which is fit will survive to the next generation and the others will not be applied to the next level.

These techniques are able to classify the given data on the basis of whether it is supervised or unsupervised learning methodologies. In case of supervised learning, the dependent and independent variables are considered. There are a set of independent variables based on which the value of the dependent variable is predicted, while in the case of unsupervised learning, the useful information is searched by forming clusters or groups. The variables in both cases can be nominal, ordinal, categorical or continuous variables depending upon the available data which enables us to apply the various algorithms of the different techniques discussed above. Collaborative filtering finds its roots in data-mining. Data mining is finding hidden and unknown information from large databases. The data mining tools and techniques are finding its immense applications in the modern day. Such an application is being proposed by the editor of this book which aims to find the data mining applications in emerging areas. These areas are already hot topics in the research. By including data mining in such areas, the application and usability of all said areas will be widened. The researchers are already working in the area of Collaborative filtering using the traditional methodologies. The editors are finding the data mining applications in this field with a motive of developing an effective recommendation system with accurate and precise information at the disposal of the users.

Collaborative filtering is defined as a technique that filters the information sought by the user and patterns by collaborating multiple data sets, such as viewpoints, multiple agents and pre-existing data

about the users' behavior stored in matrices. Collaborative filtering is required when a huge data set is present. The collaborative filtering methods are used to create recommender systems for a wide variety of fields with lots of data having varied formats, such as sensing and monitoring of data in battlefields, line of controls and mineral exploration; financial data of institutions that provide financial services, such as banks and stock markets; sensing of large geographical areas from which data is received from all kinds of sensors and activities; ecommerce and websites where the focus is to recommend products to users to increase sales, to name a few.

A definition of collaborative filtering, which is somewhat newer and a bit narrow in sense states that it is a way of automating the process of making predictions, a process which is known as filtering, about the preferences and dislikes of a user by collecting data from as big a number of users as possible, a process which is known as collaborating, hence the name collaborative filtering. The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion of an issue as a person B, then A is more likely to have an opinion similar to B's opinion on a related but different issue. It is noteworthy that such predictions are specific to the user, but they are formed by using data from a number of users. The personal information of the user such as age, gender and location are generally not used in collaborative filtering (CF) but a partially observed matrix of ratings is used. The rating matrix may be binary or ordinal. The binary matrix contains the ratings by the users in columns in the form of likes or dislikes while the user' name or id is in the rows. The ordinal matrix contains ratings in form of a number of responses from the user such as excellent, very good, good, average, poor or simply in form of stars out of five or ten, a system that is used frequently in this day and age. The rating matrix can easily be gathered implicitly by the website's server, for example using click stream logging. Clicks on links to pages of goods or services provided can be considered to be a positive review of the user. While the rating matrices can prove to be useful, one major drawback is that they are extremely sparse, so it is very difficult to clump similar users together into classes. This is due to each and every user does not give reviews about each and every product. Thus, collaborative filtering consists of storing this sparse data and analyzing it to create a recommendation system.

The objective of the proposed publication was to make aware researchers and other prospective readers with latest trends and patterns in the inclusion of the data mining tools and techniques in the areas of Collaborative filtering which helps to develop a system with precise knowledge and accuracy for helping the users of the system. The inclusion of improved and proven algorithms of the data mining helps to extract the nuggets of hidden and unknown information which helps to frame an effective recommendation system using Collaborative filtering. The mission of the proposed publication was to come up with an edited book which aims at being the latest and most advanced topic inclusion and simultaneously acts as a discussion of the contributions of renowned researchers whose work has created a revolution in this area. The contributions by eminent researchers in fields of data mining, opinion mining, sentiment analysis and Collaborative filtering will be part of book in emerging e-areas like retail, financial institutions and social networks. The objective would be to cover each and every aspect of Collaborative filtering, such as memory-based, model-based and Hybrid methodologies. The unique characteristics of the publication were:

1. The proposed work of eminent researchers in the aspect of Collaborative filtering—like memory-based, model-based and Hybrid methodologies—in areas such as retail, financial institutions and social networks which are current focuses of research will be part of the proposed publication.

2. The proposed publication will be targeted towards providing the highest quality, most accurate and latest research by eminent researchers considering the facts of how such research affects and influences common people in their everyday lives with effective and precise recommendation systems.
3. The area which will be part of published work will have a significant influence on business users, common people and have a great impact on society.

In August 2015, in the call for chapters, I urged and sought contributions to this book from researchers, IT savvy's, and young Engineers across the globe with an aim to extract and accumulate the modern day research in the field of Collaborative Filtering Using Data Mining and Analysis, and gradually I started receiving quality and very conceptual, basic and advanced contributions from different contributors from across the globe. Initially, I thought as whether I will be getting any chapters on this topic as it is very new and emerging area, but surprisingly I saw a great response with authors started to respond, which encouraged me and motivated me by showing that this area is gaining importance. After screening through them, my objective was clear, this aimed and concentrated on getting chapters which focused on elementary issues, needs, and the demand for Collaborative Filtering.

The book is a collection of the fourteen chapters which have been written by eminent professors, researchers, and industry people from different countries. These chapters were initially peer-reviewed by the Editorial board members, reviewers, and industry people who themselves span over many countries. The book is divided into three sections: Section 1, Data Mining techniques and analysis: An Overview; Section 2, Collaborative filtering: An Introduction; and Section 3, Applications of data mining techniques and data analysis in collaborative filtering.

SECTION 1: DATA MINING TECHNIQUES AND ANALYSIS: AN OVERVIEW

Chapter 1 by Dr. Renuka Mahajan, revolves around the synthesis of three research areas- data mining, personalization, recommendation systems and adaptive e-Learning systems. It also introduces a comprehensive list of parameters, extricated by reviewing the existing research intensity during the period of 2000 to October 2014, for understanding what should be essential parameters for adapting an e-learning. In general, we can consider and answer few questions to answer this body of literature 'what' can be adapted? What can we adapt to? How do we adapt? This review tries to answer on 'what' can be adapted. Thus, it advances earlier personalization studies. The gaps in the previous studies in building adaptive e-learning systems were also reviewed. It can help in designing new models for adaptation and formulating novel recommender system techniques. This will provide a foundation to industry experts and scientists for future research in adaptive e-learning.

Chapter 2 by Mamta Mittal, Dr. R.K. Sharma, Dr. V.P. Singh and Lalit Mohan Goyal enlightened that Clustering is one of the data mining techniques that investigates these data resources for hidden patterns. Many clustering algorithms are available in literature. This chapter emphasizes on partitioning based methods and is an attempt towards developing clustering algorithms that can efficiently detect clusters. In partitioning based methods, k-means and single pass clustering are popular clustering algorithms but they have several limitations. To overcome the limitations of these algorithms, a Modified Single Pass Clustering (MSPC) algorithm has been proposed in this work. It revolves around the proposition of a threshold similarity value. This is not a user defined parameter; instead, it is a function of data objects left to be clustered. In our experiments, this threshold similarity value is taken as median of the paired

distance of all data objects left to be clustered. To assess the performance of MSPC algorithm, five experiments for k-means, SPC and MSPC algorithms have been carried out on artificial and real datasets.

In Chapter 3 by Neethu Akkarapatty, Anjaly Muralidharan, Nisha S. Raj and Dr. Vinod P underlined that Sentiment analysis is an emerging field, concerned with the analysis and understanding of human emotions from sentences. Sentiment analysis is the process used to determine the attitude/opinion/emotions expressed by a person about a specific topic based on Natural Language Processing (NLP). Proliferation of social media such as blogs, Twitter, Facebook and LinkedIn has fuelled interest in Sentiment analysis. As the real time data is dynamic, the main focus of the chapter is to extract different categories of features and to analyze which category of attribute performs better. Moreover, classifying the document into positive and negative category with fewer misclassifications is the primary investigation performed. The various approaches employed for feature selection involves TF-IDF, WET, Chi-Square and mRMR on benchmark dataset pertaining diverse domains.

SECTION 2: COLLABORATIVE FILTERING: AN INTRODUCTION

Chapter 4 by Venkatesan M and Dr. Thangadurai K analyzes the recommender systems, their history and its framework in brief. The current generation of filtering techniques in recommendation methods can be broadly classified into the following five categories. Techniques used in these categories are discussed in detail. Data mining algorithms techniques are implemented in recommender systems to filters user data ratings. Area of application of Recommender Systems gives broad idea and such as how it gives impact and why it is used in the e-commerce, Online Social Networks (OSN), and so on. It has shifted the core of Internet applications from devices to users. In this chapter, issues and recent research in recommender system are also discussed.

In Chapter 5 by Neeti Sangwan and Naveen Dahiya urged that Recommendation making is an important part of the information and e-commerce ecosystem. Recommendation represent a powerful method that filter large amount of information to provide relevant choice to end users. To provide recommendations to the users, efficient and cost effective methods needs to be introduced. Collaborative filtering is an emerging technique used in making recommendations which makes use of filtering by data mining. This chapter presents a classification framework on the use of data mining techniques in collaborative filtering to extract the best recommendations to the users on the basis of their interests.

Chapter 6 by Amrit Pal and Dr. Manish Kumar describes that Size of data is increasing; it is creating challenges for its processing and storage. There are cluster based techniques available for storage and processing of this huge amount of data. Map Reduce provides an effective programming framework for developing distributed program for performing tasks which results in terms of key value pair. Collaborative filtering is the process of performing recommendation based on the previous rating of the user for a particular item or service. There are challenges while implementing collaborative filtering techniques using these distributed models. Some techniques are available for implementing collaborative filtering techniques using these models. Cluster based collaborative filtering, map reduce based collaborative filtering are some of these techniques. Chapter addresses these techniques and some basics of collaborative filtering

In Chapter 7 by Anu Saini focused that today every big company, like Google, Flipkart, Yahoo, Amazon etc., is dealing with the Big Data. This big data can be used to predict the recommendation for the user on the basis of their past behaviour. Recommendation systems are used to provide the recom-

mendation to the users. The author presents an overview of various types of recommendation systems and how these systems give recommendation by using various approaches of Collaborative Filtering. Various research works that employ collaborative filtering for recommendations systems are reviewed and classified by the authors. Finally this chapter focuses on the framework of recommendation system of big data along with the detailed survey on the use of the Big Data mining in collaborative filtering.

SECTION 3: APPLICATIONS OF DATA MINING TECHNIQUES AND DATA ANALYSIS IN COLLABORATIVE FILTERING

Arushi Jain, Dr. Vishal Bhatnagar and Pulkit Sharma in Chapter 8 canvass that there is a proliferation in the amount of data generated and its volume, which is going to persevere for many coming years. Big data clustering is the exercise of taking a set of objects and dividing them into groups in such a way that the objects in the same groups are more similar to each other according to a certain set of parameters than to those in other groups. These groups are known as clusters. Cluster analysis is one of the main tasks in the field of data mining and is a commonly used technique for statistical analysis of data. While big data collaborative filtering defined as a technique that filters the information sought by the user and patterns by collaborating multiple data sets such as viewpoints, multiple agents and pre-existing data about the users' behaviour stored in matrices. Collaborative filtering is especially required when a huge data set is present.

In chapter 9 Prof. Carson K. Leung, Fan Jiang, Edson M. Dela Cruz and Vijay Sekar Elango presents that Collaborative filtering uses data mining and analysis to develop a system that helps users make appropriate decisions in real-life applications by removing redundant information and providing valuable to information users. Data mining aims to extract from data the implicit, previously unknown and potentially useful information such as association rules that reveals relationships between frequently co-occurring patterns in antecedent and consequent parts of association rules. This chapter presents an algorithm called CF-Miner for collaborative filtering with association rule miner. The CF-Miner algorithm first constructs bitwise data structures to capture important contents in the data. It then finds frequent patterns from the bitwise structures. Based on the mined frequent patterns, the algorithm forms association rules. Finally, the algorithm ranks the mined association rules to recommend appropriate merchandise products, goods or services to users. Evaluation results show the effectiveness of CF-Miner in using association rule mining in collaborative filtering.

Chapter 10 by Mahima Goyal and Dr. Vishal Bhatnagar discusses that the recent trend of expressing opinions on the social media platforms like Twitter, Blogs, Reviews etc., a large amount of data is available for the analysis in the form of opinion mining. This analysis plays pivotal role in providing recommendation for ecommerce products, services and social networks, forecasting market movements and competition among businesses, etc. The authors present a literature review about the different techniques and applications of this field. The primary techniques can be classified into Data Mining methods, Natural Language Processing (NLP) and Machine learning algorithms. A classification framework is designed to depict the three levels of opinion mining –document level, Sentence Level and Aspect Level along with the methods involved in it. A system can be recommended on the basis of content based and collaborative filtering.

Sheng-Jhe Ke and Wei-Po Lee in Chapter 11 emphasise that Traditional collaborative filtering recommendation methods calculate similarity between users to find the most similar neighbours and take into account their opinions to predict item ratings. Though these methods have some advantages, however, they encounter difficulties in dealing with the problems of cold start users and data sparsity. To overcome these difficulties, researchers have proposed to consider social context information in the process of determining similar neighbours. In this chapter, we present a data analytics approach that combines user preference and social trust. This approach regards the collaborative recommendation as a classification task. It includes a data analysis procedure to explore the target dataset in terms of user similarity and trust relationship, and a data classification procedure to extract data features and build up a model accordingly. A series of experiments are conducted for performance evaluation. The results show that this approach can enhance the recommendation performance in an adaptive way without an iterative parameter-tuning process.

In Chapter 12 Dr. Marenglen Biba, Dr. Narasimha Rao Rao Vajjhala and Lediona Nishani provides a state-of-the-art survey of visual data mining techniques used for collaborative filtering. The chapter will begin with a discussion on various visual data mining techniques along with an analysis of the state-of-the-art visual data mining techniques used by researchers as well as in the industry. Collaborative filtering approaches will be presented along with an analysis of the state-of-the-art collaborative filtering approaches currently in use in the industry. The chapter will also include the key section of the discussion on the latest trends in visual data mining for collaborative mining.

Chapter 13 by Snehalata Sewakdas Dongre and Dr. Latesh Malik explored that A data stream is giant amount of data which is generated uncontrollably at a rapid rate from many applications like call detail records, log records, sensors applications etc. Data stream mining has grasped the attention of so many researchers. A rising problem in Data Streams is the handling of concept drift. To be a good algorithm it should adapt the changes and handle the concept drift properly. Ensemble classification method is the group of classifiers which works in collaborative manner. Overall this chapter will cover all the aspects of the data stream classification. The mission of this chapter is to discuss various techniques which use collaborative filtering for the data stream mining. The main concern of this chapter is to make reader familiar with the data stream domain and data stream mining. Instead of single classifier the group of classifiers is used to enhance the accuracy of classification. The collaborative filtering will play important role here how the different classifiers work collaborative within the ensemble to achieve a goal.

Lediona Nishani and Prof. Marenglen Biba in Chapter 14 presents that people nowadays base their behaviour by making choices through word of mouth, media, public opinion, surveys, etc. One of the most prominent techniques of recommender systems is Collaborative filtering (CF), which utilizes the known preferences of several users to develop recommendation for other users. CF can introduce limitations like new-item problem, new-user problem or data sparsity, which can be mitigated by employing Statistical Relational Learning (SRLs). This review chapter presents a comprehensive scientific survey from the basic and traditional techniques to the-state-of-the-art of SRL algorithms implemented for collaborative filtering issues. Authors provide a comprehensive review of SRL for CF tasks and demonstrate strong evidence that SRL can be successfully implemented in the recommender systems domain. Finally, the chapter is concluded with a summarization of the key issues that SRLs tackle in the collaborative filtering area and suggest further open issues in order to advance in this field of research.

The applications of Collaborative Filtering Using Data Mining and Analysis are so vast that it cannot be covered in single book. However with the encouraging research contribution by the researchers in

this book, we (contributors) tried to sum the latest development and work in the area. This edited book will serve as the stepping stone and a factor of motivation for those young Researchers and Budding Engineers who are witnessing the every stopping growth in the field of Collaborative Filtering Using Data Mining and Analysis.

Vishal Bhatnagar
Ambedkar Institute of Advanced Communication Technologies and Research, India