Preface

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 which needs to be addressed and explained in one book to understand the importance of the data mining applications. The implications of data mining can be understood by the facts 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 being 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 bulk of the data. The new and emerging areas of data mining techniques have surprised many a researchers and business persons who are actually gaining a lot of hidden and unknown information for increasing their ROI. This will widen the application area and more interest will be created in budding researchers to pursue their research in data mining. The data mining is all about the revealing of hidden and unknown information from the huge available data. The techniques of data mining which are primarily making it happen are:

1. Classification
2. Clustering
3. Association rule mining
4. Neural network
5. Genetic algorithms

These techniques are able to classify the given data on the basis of whether it is supervised or unsupervised learning methodology. In case of supervised learning, the dependent and independent variables are considered. There are set of independent variables on the basis of which the value of the dependent variable is predicted while in the case of unsupervised learning, the useful information is searched by forming the clusters or group. The variables in both the case can be nominal, ordinal, categorical or continuous variables depending upon the available data which enable us to apply the various algorithms of the different techniques discussed above. Considering the vast application area of data mining, book is targeted towards finding the data mining applications in emerging areas. These areas are already hot topics in the research. By the inclusion of the data mining in such areas, the application and usability of all the said areas will be widened.

With the proliferated use of social networks in today’s modern era, data mining has found its significant place in social network analysis and its security. Privacy preservation of social networks is a brewing topic of research these days. With large amounts of content being posted on social networks, privacy breach has become one of the prime issues of social networking. Data mining techniques like classifica-
tion, clustering, and association rule mining has been used extensively for social network analysis and data mining techniques like heuristics based, reconstruction based, and cryptographic techniques that are being applied on social networks for providing desired security. There are also various anonymization techniques like clustering and clustering with constraints that make use of data mining to provide privacy preservation of released social network data. Much less work has been done in the area of dynamic social network security and is a focus of study for various researchers and practitioners. Social networks contain a vast amount of information and such high dimensional data are difficult to be handled by the traditional systems like OLAP. For this, data mining has proved to be a blessing in disguise. Various data mining and statistical techniques find its use in analyzing large amounts of online (Dynamic) social network data, where the interactions among the users of the network are studied to find out interesting patterns and also to find out various outliers in the data. The chapters in the first part of the book aim towards fulfilling the above said aim and finding the wide and emerging current development in the application of data mining in dynamic social network.

The use of fuzzy logic can be intervened with data mining so as to give another dimension to the concept of data mining. The fuzzy logic is used in data mining to create a new concept called the fuzzy data mining which makes data mining more flexible and extends its utility by a large extent in fields such as intrusion detection, approximations of missing values, power plant optimizations, human resource management, cross-selling, detection of quality of water, decision making or medical image processing. Fuzzy theory is useful for data mining systems performing rule based classification. The future prospect of data mining in the field of fuzzy data mining is revealing applications on fuzzy sets. The research is currently focused on finding the utilization of other techniques of data mining in fuzzy systems like genetic algorithms. The focus of the study is also as how more effective results can be drawn on application of data mining on fuzzy systems. The chapters in the second part of the book aim towards fulfilling the above said aim and finding the wide and emerging current development in the application of data mining in fuzzy systems.

The objective of the this edited book is 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 which affect the common men so that a better system can be developed with improved and modern techniques of data mining. The mission of the proposed publication was to come up with an edited book which aims at latest and most advanced topic inclusion and simultaneously discussion of contributions of renowned researchers whose work created a revolution in the area.

The development in the area of application of data mining in Dynamic social network and fuzzy system will help and target the researchers and academicians targeting their Doctoral and post Doctoral research and this edited book will act as a stepping stone for grooming of the budding researchers who intend to work in the area of data mining and its applications. The book will target the audiences which would like to work on latest and advanced concepts of data mining inclusions in the applicative areas of Dynamic social network and fuzzy system for the benefit of the society and country. The specific beneficiary will be:

1. The researchers will be able to know the latest application area of the data mining.
2. The business users will get to know how inclusion of data mining can provide added advantages for them in their professions.
3. The common people will get the secure edge for their data which is revealed in the social network analysis.
Intended Audiences:

1. Business users
2. Researchers
3. Common peoples

In my May 2012 call for chapters, I urged and sought for contribution to this book from researchers, IT savvys and young Engineers, across the globe with an aim to extract and accumulate the whole modern day research in the field of data mining in dynamic social network and fuzzy system and gradually I started getting quality and very conceptual, basic and advanced contributions that too from various contributors from different countries. 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 an great response with authors started to respond which encouraged me and motivated me by showing that this area is gaining its importance. After screening through them, my aim and objective was clear, which aimed and concentrated on getting chapters which focused on elementary issues, needs, demand of the data mining in said area and finally on application areas of the data mining in dynamic social networks and fuzzy systems.

The book is a collection of the fifteen chapters which are written by eminent professors, researchers and Industry people from different countries. The chapters were initially peer reviewed by the Editorial board members, reviewers and industry people who themselves span over many countries. The whole book is divided into two sections, namely Section I Data Mining in Dynamic Social Networks and Section II Data Mining in Fuzzy Systems.

SECTION 1: DATA MINING IN DYNAMIC SOCIAL NETWORKS

Chapter 1 by Gurdeep S Hura urged us to felt that there is really an urgent need for the dynamicity in social networking sites considering the data mining perspective. This chapter presents this new emerging technology of social media and networking with a detailed discussion on: basic definitions and applications, how this technology evolved in the last few years, the need for dynamicity under data mining environment. It also provides a comprehensive design and analysis of popular social networking media and sites available for the users. A brief discussion on the data mining methodologies for implementing the variety of new applications dealing with huge/big data in data science is presented. Further, an attempt is being made in this chapter to present a new emerging perspective of data mining methodologies with its dynamicity for social networking media and sites as a new trend and needed framework for dealing with huge amount of data for its collection, analysis and interpretation for a number of real world applications.

In Chapter 2, Preeti Gupta and Dr. Vishal Bhatnagar discussed about the social network analysis which is of significant interest in various application domains due to its inherent richness. Social network analysis like any other data analysis is limited by the quality and quantity of data and for which data preprocessing plays the key role. Before the discovery of useful information or pattern from the social network data set, the original data set must be converted to a suitable format. In this chapter authors present various phases of social network data preprocessing. In this context they discuss various challenges in each phase. The goal of this chapter is to illustrate the importance of data preprocessing for social network analysis.
The analysis of the dynamic network data is a challenge in modern era. In Chapter 3, Dhiraj Murthy, Alexander Gross and Alex Takata identifies a number of the most common data mining toolkits and evaluates their utility in the extraction of data from heterogeneous online social networks. It introduces not only the complexities of scraping data from the diverse forms of data manifested in these sources, but also critically evaluates currently available tools. This analysis is followed by a presentation and discussion on the development of a hybrid system, which builds upon the work of the open-source Web-Harvest framework, for the collection of information from online social networks. This tool, VoyeurServer, attempts to address the weaknesses of tools identified in earlier sections, as well as prototype the implementation of key functionalities thought to be missing from commonly available data extraction toolkits. Authors concluded the chapter with a case study and subsequent evaluation of the VoyeurServer system itself. This evaluation presents future directions, remaining challenges, and additional extensions thought to be important to the effective development of data mining tools for the study of online social networks.

The need of privacy and security in the social network data has increased with more and more dynamincity in social network data. The urge and demand of the growing users of the social network had resulted in contribution from Sanur Sharma and Dr. Vishal Bhatnagar in Chapter 4 that in recent times there has been a tremendous increase in the number of social networking sites and their users. With the amount of information posted on the public forums, it becomes essential for the service providers to maintain the privacy of an individual. Anonymization as a technique to secure social network data has gained popularity but there are challenges in implementing it effectively. In this chapter, authors have presented a conceptual framework to secure the social network data effectively by using data mining techniques to perform in-depth social network analysis before carrying out the actual anonymization process. The framework in the first step defines the role of community analysis in social network and its various features and temporal metrics. In the next step authors proposes the application of those data mining techniques that can deal with the dynamic nature of social network and discover important attributes of the social network. Finally, authors mapped security requirements and findings of the network properties which provide an appropriate base for selection and application of the anonymization technique to protect privacy of social network data.

Chapter 5 by Luca Cagliero, Luigi Grimaudo and Alessandro Fiori argued that User-generated content (UGC) coming from social networks and online communities continuously grows and changes. By analyzing relevant patterns from the UGC, analysts may discover peculiar user behaviors and interests which can be used to personalize Web-oriented applications. In the last several years, the use of dynamic mining techniques has captured the interest of the research community. They are focus on analyzing the temporal evolution of most significant correlations hidden in the analyzed data. However, keeping track of all temporal data correlations relevant for user behaviors, community interests, and topic trend analysts may become a challenging task due to the sparseness of the analyzed data. The authors in this Chapter presents a novel data mining system that performs dynamic itemset mining from both the content and the contextual features of the messages posted on Twitter. Dynamic itemsets represent the evolution of data correlations over time. The framework exploits a dynamic itemset mining algorithm, named HiGen Miner, to discover relevant temporal data correlations from a stream of tweet collections. In particular, it extracts compact patterns, namely the HiGens that represent the evolution of the most relevant itemsets over consecutive time periods at different abstraction levels. Taxonomy is used to drive the mining process and prevent the discarding of knowledge that becomes infrequent in a certain time period. Experiments, performed on real Twitter posts, show the effectiveness and the usability of the proposed system in supporting Twitter user behavior and topic trend analysis.
Chapter 6 by Manish Kumar focused on the application of data mining in dynamic social network analysis. Social Networks are nodes consisting of people, groups and organizations growing dynamically. The growth is horizontal as well as vertical in terms of size and number. Social network analysis has gained success due to online social networking and sharing sites. The accessibility of online social sites such as MySpace, Facebook, Twitter, Hi5, Friendster, SkyRock and Beb offer sharing and maintaining large amount of different data. Social network analysis is focused on mining such data i.e. generating pattern of people’s interaction. The analysis involves the knowledge discovery that helps the sites as well as users in terms of usage and business goals respectively. Further it is desired that the process must be privacy preserving. This chapter describes the various mining techniques applicable on social networks data.

Chapter 7 by Luca Cagliero and Alessandro Fiori reviews statistics about past few years which have witnessed the rapid proliferation of Web communities such as social networking sites, wikis, blogs, and media sharing communities. The published social content is commonly characterized by a high dynamicity and reflects the most recent trends and common user behaviors. The Data mining and Knowledge Discovery (KDD) process focuses on discovering and analyzing relevant information hidden in large data collections to support expert decision making. Hence, the application of data mining techniques to data coming from social networks and online communities is definitely an appealing research topic. This chapter presents overviews on most recent data mining approaches proposed in the context of social network analysis. In particular, it aims at classifying the proposed approaches based on both the adopted mining strategies and their suitability for supporting knowledge discovery in a dynamic context. To provide a thorough insight into the proposed approaches, main work issues and prospects in dynamic social network analysis are also outlined.

The demand and essentiality of the mobile device is not hidden from any one of us. The growth of both fields simultaneously that is social network and mobile devices had presented a great potential for the researchers across the world. In Chapter 8 Gebeeyehu Belay Gebremeskel, Zhongshi He and Hua-zheng Zhu explores and argued for Data mining is a key paradigm for Mobile Social Networks (MSNs) in which growing and exciting area of research that has in front of itself a long way to go across many fields, including social networks. With the evolution of the social networking and the rapid adoption of a mobile device’s large and unprecedented amount of data generated, which gained significant attention to improve social network’s applications easily, and accessible platform for users. However, the traditional social-networking system could not capable and scalable to provide these dynamic applications. The challenges are unable to handle and manage large-scale data, which includes noisiness, unstructured, diversified sources and dynamic in nature, and other functionality challenges. Unable to accommodating new technologies, including social technology, mobile devices and computing are other potential problems, which are significant challenges to social-networking service. The very broad range of such social-networking challenges and problems are demanding advanced and dynamic tools. Therefore, in this chapter, authors introduced and discussed data mining prospects to overcome the traditional social-networking challenges and problems, which led to optimization of MSNs application and performances. Data mining prospects in MSN has an enormous potential to handle, manage and extract actionable patterns, which can be beneficial for transformation of the conventional social networks to advance one (MSN). The proposed method infers defining and investigating social-networking problems using data mining techniques and algorithms based on the large-scale data. The approach is also exploring the possible potential of users and systems contexts, which leads to mine the personal contexts such as the users’ locations and situations from the mobile logs. As a result, data mining playing an essential and dynamic
role to developing appropriate solutions, visualized the social contexts, users’ contexts and others in section 3, 4 and 5. In these sections, they discussed and introduced new ideas on social technologies, data mining techniques and algorithm’s prospects, social technology’s key functional and performances, which include social analysis, security and fraud detections by presenting a brief analysis, and modeling based descriptions. The approach also empirically discussed using the real survey data, which the result showed how data mining vitally significant to explore MSNs performance and its crosscutting impacts. Finally, this chapter provides fundamental insight to researchers and practitioners who need to know data mining prospects and techniques to analyze large, complex and frequently changing data. This chapter is also providing a state-of-the-art of data mining techniques and algorithm’s dynamic prospects. In addition, authors provide insights on future research directions.

Chapter 9 by Gebeyehu Belay Gebremeskel, Zhongshi He and Xuan Jing discusses the Semantic integrating with intelligent cloud data mining platform for optimizing of the Mobile social networks (MSNs). With the evolution of the social-networking in relation to computing development (DM and BI) and the rapid adoption of mobile devices such as cell phones and other handheld devices, social networks, which began as web-based applications, and migrated onto the semantic cloud platform. It is the fundamental knowledge and optimal features of social networks, which helps to manage the human relationship and interaction using semantic intelligent cloud. It revolutionizes how social relationships develop, understand, and secure real value through a social cloud that involving semantic intelligent agents and mobile clouds of the dynamic social networking. However, the current social-networking service has many challenges, including the lack of integrated semantic intelligent agents cloud and mobile devices. The challenges are including less consolidation of social-networking, unable to defining the problem of characterization of social-networking structures towards large-scale data handling and analysis missed the opportunity to bind social context tightly with the intelligent agents cloud system and local context of interacting users. In this chapter, authors focused on optimization of MSNs based on integrating for intelligent DM and BI platforms, which involving mobile devices. The approach is defining the challenges based social network trends and current situation explorations, and then applying the techniques to exploring the social media towards social cloud technology, which focused on creating a scalable, adaptable and optimal social cloud as the users’ contexts and IT technologies. The newly proposed method is vigorously significant to develop flexible social networking in relation to the development of IT, which facilitates data/information access, distributions, high availability and a large amount of data analysis and others. Therefore, the techniques this chapter is vitally crucial to improve the performance and use of social networking in a comprehensive and powerful way. Nutshell, this chapter overviews the impetus for the development of intelligent semantic cloud and diversified social-networking in both physical and wireless sectors, which representing a wide aspect of social cloud change, and increasingly appropriate service providing a platform for innovative ideas and technological innovation in the business environment.

Chapter 10 by Dr. Sunil Kr Pandey and Dr. Vineet Kansal presents an in-depth online survey on the growing demand of the social media analytics. In present context, online social media demonstrates a fundamental shift in the way information is being produced, transferred and consumed. Large volume of user generated contents in the form of posting blogs, comments, and tweets establishes a connection between the producers and the consumers of information. It has been observed that sensing and tracking the pulse of social media channels may make possible for companies to gain feedback and detailed insight in how to improve, promote and market their services & products better. For consumers, the wealth of information and opinions from various diversified sources facilitates them tap into the wisdom
of crowds, to aid in making more informed decisions. The advent of online social networks has been one of the most exciting events in this decade. It has established itself in such a way that in a business organization, search engine optimization, search engine marketing, social media and e-commerce is now a very small piece of an overall online marketing business plan. More and more companies today are focused on brand protection and understanding how people discuss, comment and share information about their company online. Companies like the New Jersey Devils and Gatorade have created social intelligence command centers that listen and mine data within social media platforms such as Google+, Facebook, LinkedIn and Twitter (Li, Surendran, & Shen, 2005). Companies are mining massive data sets to get a competitive edge on their competition. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. This type of information that can be used to find out what people are seeking, what type of branding might be working and then be able to react accordingly.

Many popular online social networks such as Twitter, LinkedIn, and Facebook have become increasingly popular (based on Facebook statistics by Country, 2011 and other social media reports, 2011, 2012). In addition, a number of multimedia networks such as Flickr have also seen an increasing level of popularity in recent years. Many such social networks are extremely rich in content, and contain tremendous amount of content and linkage data which can be leveraged for analysis. The linkage data is essentially the graph structure of the social network and the communications between entities; whereas the content data contains the text, images and other multimedia data in the network. The richness of network provides unprecedented opportunities for data analytics in the context of social networks. This significant increase in its usage and increased number of users, there has been trend of a substantial increase in the volume of information generated by users of social media. Irrespective of primary domain in which organization is operating in to, whether it is insurance sector, social media (including facebook, twitter etc), medical science, banking etc. Virtually a large number of varying nature and services of organizations are making significant investments in social media. But it is also true that many are not systematically analyzing the valuable information that is resulting from their investments. This chapter aims at providing a data-centric view of online social networks; a topic which has been missing from much of the literature and to draw unanswered research issues which can be further explored to strengthen this area.

SECTION 2: DATA MINING IN FUZZY SYSTEMS

Chapter 11 by Sinchan Bhattacharya and Dr. Vishal Bhatnagar explored critical parameters for application of fuzzy data mining. Research on data mining is increasing at an incessant rate and to improve its effectiveness other techniques have been applied such as fuzzy sets, rough set theory, knowledge representation, inductive logic programming, or high-performance computing. Fuzzy logic due to its proficiency in handling uncertainty has gained its importance in a variety of applications in combination with the use of data mining techniques. In this chapter we take this association a notch further by examining the parameters which allow fuzzy sets and data mining to be combined into what has come to be known as fuzzy data mining. Analyzing and understanding these critical parameters is the main purpose of this chapter, so as to acquire maximum efficiency in applying the same which impelled the authors to work extensively and find out the crucial parameters essential to the application of fuzzy data mining.

Chapter 12 by Zekai Sen provides a clear idea to trends developing in Fuzzy Clustering. Fuzzy methodologies show progress day by day towards better explanation of various natural, social, engineer-
ing and information problem solutions in the best, economic, fast and effective manner. This chapter provides cluster analyses from probabilistic, statistical and especially fuzzy methodology points of view by consideration of various classical and innovative cluster modeling and inference systems. After the conceptual assessment explanation of fuzzy logic thinking fundamentals various clustering methodologies are presented with brief revisions but innovative trend analyses as k-mean-standard deviation, cluster regression, relative clustering for depiction of trend components that fall within different clusters. The application of fuzzy clustering methodology is presented for lake time series and earthquake modeling for rapid hazard assessment of existing buildings.

Chapter 13 by Sara Moridpour analyses the Performance of a Fuzzy Lane Changing Model Using Data Mining. Heavy vehicles have substantial impact on traffic flow particularly during heavy traffic conditions. Large amount of heavy vehicle lane changing manoeuvres may increase the number of traffic accidents and therefore reduce the freeway safety. Improving road capacity and enhancing traffic safety on freeways has been the motivation to establish heavy vehicle lane restriction strategies to reduce the interaction between heavy vehicles and passenger cars. In previous studies, different heavy vehicle lane restriction strategies have been evaluated using microscopic traffic simulation packages. Microscopic traffic simulation packages generally use a common model to estimate the lane changing of heavy vehicles and passenger cars. The common lane changing model ignores the differences exist in the lane changing behaviour of heavy vehicle and passenger car drivers. An exclusive fuzzy lane changing model for heavy vehicles is developed and presented in this chapter. This fuzzy model can increase the accuracy of simulation models in estimating the macroscopic and microscopic traffic characteristics. The results of this chapter shows that using an exclusive lane changing model for heavy vehicles, results in more reliable evaluation of lane restriction strategies.

Chapter 14 by Başar Öztayşi and Sezi Çevik Onar showed User Segmentation Based on Twitter Data Using Fuzzy Clustering. Social Networking Sites, which create platform for social interactions and sharing are the mostly used internet websites, thus are very important in today’s world. The vast usage of social networking sites (SNSs) has affected the business world, new business models are proposed, business process are renewed and companies try to create benefit from these sites. Besides the functional usage of SNSs such as marketing and customer relations, companies can create value by analyzing and mining the data on SNSs. In this paper, a new segmentation approach, using Text Mining and Fuzzy Clustering techniques. Text mining is process of extracting knowledge from large amounts of unstructured data source such as content generated by the SNSs users. Fuzzy clustering is an algorithm for cluster analysis in which the allocation of data points to clusters is fuzzy. In the proposed approach, users self description text are used as an input to the Text Mining process, and Fuzzy Clustering is used to extract knowledge from data. Using the proposed approach in this chapter, companies can segment their customers based on their comments, ideas or any kind of other unstructured data on SNSs.

Chapter 15 by Başar Öztayşi and Sezi Çevik Onar defined the factors that Effect User Interest on Social Network News Feeds via Fuzzy Association Rule Mining through the Case of Sports News. Social networking became one of the main marketing tools in the recent years since it’s a faster and cheaper way to reach the customers. Companies can use social networks for efficient communication with their current and potential customers but the value created through the usage of social networks depends on how well the organizations use these tools. Therefore a support system which will enhance the usage of these tools is necessary. Fuzzy Association rule mining (FARM) is a commonly used data mining technique which focuses on discovering the frequent items and association rules in a data set and can be a powerful tool for enhancing the usage of social networks. Therefore the aim of the chapter is to propose
a fuzzy association rule mining based methodology which will present the potential of using the FARM techniques in the field of social network analysis. In order to reveal the applicability, an experimental evaluation of the proposed methodology in a sports portal will be presented.

The applications of data mining in dynamic social network and fuzzy system are so vast that it cannot be covered in a 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 ever-stopping growth in the field of dynamic social network and fuzzy system.

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