Predictive Analysis of Biomedical Data Using Clustering Data Mining Techniques

Ms. Richa Mehta[1], Mr. Rinal Doshi[2], Dr. Aashish N. Jani[3]

[1]Research Scholar, Singhania University, Rajasthan, SKPIMCS, MCA Dept., KSV University, Gandhinagar, Gujarat, India
[2]Research Scholar, School of Engineering, R.K.University, Rajkot, L.C.Institute of Technology, GTU, Bhandu, Gujarat, India
[3]Research Guide, SKPIMCS, MCA Department, KSV University, Gandhinagar, Gujarat, India
[1] triumphricha@gmail.com, [2] rinal.doshi@lcit.org, [3] aashishjani@yahoo.com

Abstract: The constant improvement in the Biomedical Instrumentation Technologies and generation of large datasets as a result of measurement of health parameters of patients in widespread distributed sites has greatly increased the accumulation of Biomedical Data in the repositories indicating exponential trends of increase of data volume. These datasets in the repository are awaiting great deal of Predictive Analysis to manage the effective Analytic Operations on such huge datasets that requires suitable application of Data Mining Techniques. These techniques result in extraction of intermediate patterns that can be analyzed to derive predictive conclusions. These outcomes can be fruitful in supporting decision making as well as planning for remedial measures for various chronic diseases. The work undertaken and documented in this paper is of Predictive Analytic nature subjected to Diabetic Diagnosis dataset for analytical operations in Open Source Tool Environment of Weka [15]. Likewise Medical Databases can be coupled with Data Mining Techniques to explore the Interdisciplinary field of research named as Medicalinformatics. The working has applied selected Clustering Data Mining Techniques and analyzed the datasets leading upto Comparative Analysis followed by Predictive Analysis.

Keyword: *Biomedical Instrumentation Technologies, Biomedical Data, Predictive Analysis, Analytic Operations, Data Mining Techniques, Medical Databases, Medicalinformatics, Clustering Data Mining Techniques, Comparative Analysis.*

I. INTRODUCTION

The accumulation of Biomedical Data at any point of time is voluminous and resulting data repository volume increases significantly. Such a huge data requires the use of Data Mining Techniques to extract unknown or hidden information from targeted dataset to have Predictive Analysis on the dataset. The work included in this research paper aims to focus on the process of analysing biomedical web data repository and selecting targeted dataset from the secondary sources. The targeted dataset has a considerable volume of data which needs the Application of Data Mining Techniques [10]. Among the various available data mining techniques, this research work has selected to apply clustering techniques to achieve the purpose of finding hidden patterns and analysing them to derive fruitful conclusions [6]. This will lead to achieve predictive analysis which may be suitable to cure a disease for which the dataset has been selected.

The analytical tech. coupled with data mining and widespread access to public Biomedical Data set has increased attention to Biomedical Data mining and predictive analysis using data mining techniques [3]. This direction of research is associated with Bioinformatics giving solution to the problems Biomedical Data analysis. The question that needs researched issues in designing and implementing Bioinformatics solutions to aid Biomedical Researchers in carrying out predictive analysis [4].

II. BIOMEDICAL AND BIOLOGICAL DATA

The Biomedical Instrumentation Technologies has improved to the stage that generates large datasets as a result of measurement of health parameters of the patients. The analysis of these varieties of large volume of data sets does require the application of data mining techniques and statistical techniques to arrive at important conclusions. These Biomedical data sets are also the secondary source of Biological data related to cell, tissues etc [4]. Further the Techniques of using mixed data from genome, proteome and transcriptome can be used for detection of chronic diseases. These techniques require not only intelligent processing but also high performance computing platform

along with analysis tools for data mining and statistical analysis [7]. The modelling of Biological and Biomedical data requires the tuning of different formats to unified formats and thereafter the data can be subjected for data mining algorithm like data Classification, Clustering etc. The outcomes of the predictive analysis lead to predict the Diagnostic Genes for the application of suitable cure therapy.

The diversity characteristics of life has been noticed by us with the evidences of great differences among living creatures but the molecular details of these living organisms are almost universal. Protein, which is called as a complex family of molecular is also a smallest units of organism and even the main structural and functional unit of cell, therefore it can be called as the reason behind every biological activity of living organism. One can look at the example of a protein which is Enzymes which speeds up the chemical processes.

There are enormous amount of Biological Data which are represented in tabular format in the below Table 1.

Table 1: Biological Data categories

Biological Data categories	Description Outline
Experimental Data	The collected datasets obtained directly from laboratory experiments such as digital images, notes, observations, etc.
Raw data	The collected datasets that is awaiting the initial processing to derive meaningful information out of it.
Phylogenetic data	The collected datasets depicting evolutionary relationships between variegated groups of organisms.
Metabolic data	The collected datasets of system biology and enzymatic reactions (metabolic pathways)
Sequence data	The collected datasets from MSA (multiple sequence alignment) inclusive of DNA or protein sequencing process.
Structure data	The collected datasets containing linear notations of small molecules like 3D protein structures, DNA or RNA.

The diversity characteristics of life has been noticed by us with the evidences of great differences among living creatures but the molecular details of these living organisms are almost universal. Every living organism depends upon biological activity of complex family of molecules called proteins. Proteins are main structural and functional units of a cell and the smallest unit of organism [14]. Enzymes are typical example of protein that accelerates chemical reaction.

There are four levels of protein structural arrangement as listed as below:

- Primary Structure: Amino Acids Sequence forming a chain called polypeptide.
- Secondary Structure: Polypeptide structure after folding.
- Tertiary Structure: Stable 3D structure forming a polypeptide.
- Quaternary Structure: 3D structure formed by the conjugation of two or more polypeptides.

Observation of sets of proteins that have same sets and functions creates the similarity among organisms that are most unlike and different from each other. Second Family of molecules that are known as "Nucleic Acids" that creates another similar characteristic among living organism. The work of this Nucleic Acid is to contain information that codes life [1].

Both Proteins and Nucleic Acids are defined as macro molecules as they are large in size in comparison to other molecules and which have a responsibility to carry that information that quotes life [14]. Molecular Biology is a branch that offers the study of structures and functions of these macro molecules in order to understand the life.

Monomers are the smaller molecules which combine in a linear way and make polymers. These Proteins and Nucleic acids are the example of these polymers. Macro molecules that are made of monomers and the order of monomers in that macro molecule are referred by a term sequence which is a string of symbols one for each monomer. There are twenty protein monomer called amino acids. There are two nucleic acids- DNA (Dioxy Ribo Nucleic Acid) and RNA (Ribo Nucleic Acid). DNA and RNA monomer are nucleotides. There are five different nucleotides depending on nitrogen based they contain. They are namely A(Adenine), C(Cytocine), G(Guanine), T(Thymine), U(Uracil). DNA doesn't contain U whereas RNA contains U instead of T.

DNA is genetic material of almost every living organism. RNA is genetic material for some virus such as HIV. RNA has many functions inside a cell and plays an important role in protein synthesis. Some generalised types of RNA are mentioned below in the Table 2.

Table 2: Generalised types of RNA

RNA basic types	Description Outline
mRNA (Messenger RNA)	Information carrier from DNA to protein.
rRNA (Ribosomal RNA)	Location of protein synthesis and Constituent of the cellular ribosomes.
tRNA (Transfer RNA)	transferring of AA's to Ribosomes during protein synthesis
snRNA (Small nuclear RNA)	Located in nucleus and associated with processes like RNA splicing.

Broadly organisms can be bifurcated into two categories, one is Eukaryotes which includes Animals, plants fungi and protists and second is single cell type Prokaryotes which include bacteria and archea.

An organism may contain one or more Chromosomes, that are Long double standard DNA molecules, the form in which the genetic material of an organism. A gene is a DNA sequence located in a particular chromosome and encodes the information for the synthesis of proteins (RNA molecules).

All the genetic material of a particular organism constitutes its genome. The molecular biology describes the flow of sequence information – DNA is transcribed into RNA and then RNA is translated into protein. The flow of biological sequence information is depicted in the Fig. 1.

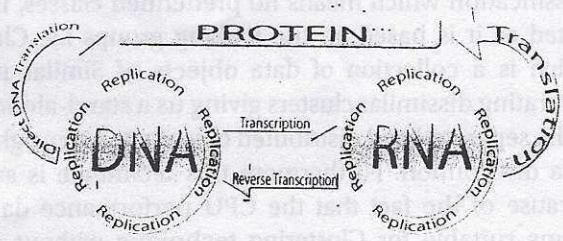

Fig. 1: The flow of biological sequence information

The circular arrow around DNA denotes its replication ability. There are some specific transfers say in retroviruses, RNA is reverse transcribed into DNA also some viruses can be replicated in to RNA.

III. DATA MINING IN BIOINFORMATICS

Data Mining is used to extract the knowledge from the data, the process is known as Knowledge Discovery in Databases (KDD)[11]. KDD has some process like selection, cleaning, and transformation visualization and evaluation. The main tasks that take place in KDD are Classification, Regression, Clustering, Association Analysis, and Sequence analysis etc[7]. All these processes are done according to the need of Information and data that are available. The process follows the Algorithms and these can be a predictive or descriptive that classifies Common Data Mining task as below in Table 3.

Supervised learning machine algorithms followed by predictive algorithms and unsupervised followed by descriptive. The selection is done according to the effect and purpose as well as data mining algorithm is chosen according to the efficiency and scalability [9].

Illustration data cleansing, disease diagnosis, disease treatment optimization, gene finding, protein function interference and domain detection, function motif

Table 3: Common data mining tasks

Predictive		Descriptive		
Classification: Classifies data into predefined classes/ groups.	Regression: Maps data into a real valued prediction variable.	Association Analysis: The production of rules that describe relationships among data.	Sequence Analysis: Same as association, but sequence of events is also considered.	Clustering: Groups similar input patterns together. .....

detection, prediction of protein sub-cellular location as well as protein and gene interaction network reconstruction are some of the Areas of Bioinformatics that can also be called as an Application areas of Data mining[8][10].

Various experiments based on the microarray technologies are performed to predict a patient's genotypic microarray data outcome which is helpful to estimate their survival time and risk of malignant disease like tumor metastasis or its recurrence. The other technique is Machine learning that can be used for peptide identification availing mass spectroscopy. The Correlation among fragment ions in a tandem mass spectrum is crucial in reducing stochastic mismatches for peptide identification by database searching. An efficient scoring algorithm that considers the correlative information in a tuneable and comprehensive manner is highly desirable.

Many general data mining systems such as SAS Enterprise Miner, SPSS, S-Plus, IBM Intelligent Miner, Microsoft SQL Server 2000, SGI Mine Set, and Inxight Viz Server can be used for Biological Data Mining [5]. However, some Biological Data mining tools such as Gene Spring, Spot Fire, Vector NTI, COMPASS, Statistics for Microarray Analysis, and Affymetrix Data Mining Tool have been developed [12].

The rapid progress in Biomedical Research has led to a dramatic increase in the amount of available information, in terms of published articles, journals, books and conference proceedings. Today, Pub Med alone provides access to more than 14 million citations from MEDLINE and additional life sciences journals. In total, more than 4,800 journals are currently indexed by Pub Med. Although Pub Med is by far the richest database of abstracts, citations and full text articles, there is a plethora of such sources of scientific publications on Biology such as NCBI Book Shelf for e-books and a large number of on-line resources. The researchers' need to exploit this enormous volume of available information, along with the avail of high performance and efficiency data mining, natural language processing and information retrieval tools has given birth to a new field of research and application, called Bioinformatics Text Mining (BTM). Other terms for BTM

are Biological Text Mining and Biomedical Text Mining[11]. From a data miner's point of view, Biomedical literature has certain characteristics that require special attention, such as heavy use of domain-specific terminology, polysemic words (word sense disambiguation), low frequency words (data sparseness), creation of new names and terms and different writing styles.

IV. CHALLENGES AND FUTURE TRENDS

Data mining in Bioinformatics with Biological Databases has brought Challenging issues related to dataset size, number of instances in the dataset and diversity in database structures and lack of standard ontology to query heterogenous data[5][6][13]. The scattered nature of the related data does bring additional challenge in the process of integration of distributed dataset. Analysis of microarrays presents a number of unique challenges for data mining. Typical data mining applications in domains like banking or web, have a large number of records thousands and sometimes millions), while the number of fields is much smaller (at most several hundred). In contrast, a typical microarray data analysis study may have only a small number of records (less than a hundred), while the number of fields, corresponding to the number of genes, is typically in thousands. Given the difficulty of collecting microarray samples, the number of samples is likely to remain small in many interesting cases.

The special characteristics of Biological Data raise variety of new problems that are of extremely high importance in the area of Bioinformatics research [8]. A large number of critical issues is still open and demands active and collaborative research by the academia as well as the industry. Moreover, new technologies such as the microarrays led to a constantly increasing number of new questions on new data. Examples of major problems in Bioinformatics are the accurate prediction of protein structure and gene behaviour analysis in microarrays. Bioinformatics demands and provides the opportunities for novel and improved data mining methods development. As Houle et al. [2] mention, these improvements will enable the prediction of protein function in the context of higher order processes such as the regulation of gene expression, metabolic pathways (series of chemical reactions within a cell, catalyzed by enzymes) and signalling cascades (series of reactions which occur as a result of a single stimulus). The final objective of such analysis will be the illumination of the way conveying from genotype to phenotype.

The exploration of the area of Bioinformatics supported with data mining and Analytics using software tools can help the investigators to lead to the localization of the reduced data set where deeper understanding can be developed with regard to the Biological issues hidden in the voluminous data set[11]. To approaches under mining and analysis of voluminous data with extraction of data with publics databases are presented to give the understanding of flow of data mining and analysis.

V. PROPOSED WORK BACKGROUND

The proposed Work is aimed at the study and analysis of various data mining techniques to get the ultimate mined data from the voluminous data. Among the various data mining techniques i.e. Classification, Regression, Association Rule Mining and Clustering.

Clustering technique is used to get the proposed outcome. Clustering technique is an unsupervised classification which means no predefined classes, is best suited as it is based on the making groups i.e. Clusters which is a collection of data objects of Similar nature separating dissimilar clusters giving us a stand-alone view of the separated and distributed clusters to get insight into data distribution. Furthermore this technique is availed because of the fact that the CPU performance data set seems suitable for Clustering technique without much preprocessing with regard to attribute category, noisy and outliers' data.

Out of various clustering techniques this work has selected four clustering techniques giving almost near to one another output structure facilitating analysis of the results to arrive at the optimum performance value of selected attribute. These selected clustering techniques are to be applied on the same dataset to get targeted output of each technique in the same processing environment.

The dataset will have the Diabetes Diagnosis data - Pregnancies, PG Concentration, Diastolic BP, Tri Fold Thick, Serum Ins, BMI, DP Function, and Age.

The resulting outcome of the implemented clustering techniques is compared and analyzed to arrive at optimum predictive results of the analysis of instances of Diabetic Dataset.

There are various clustering Data Mining techniques in WEKA [15] like Cobweb, EM, Farthest First, Filtered Clusterer, Hierarchical Clusterer, Make Density based Clusterer, Simple K Means, OPTICS, sIB, XMeans. Among various clustering techniques four clustering techniques- Simple K Means(SKM), Farther First(FF), Filtered Clusterer (FC) and Exception Maximization(EM) are availed for generating results that are suitable comparative analysis and result optimization are considered for implementation.

Simple K Means (SKM) Clustering technique produces computationally faster and tighter cluster as compare to hierarchical clustering technique. SKM algorithm is able

to identify the non linear structure. It is best suit for the real life dataset or web log data. Farthest First (SKM) is SKM Clustering algorithm variant that places each cluster centre in turn at the point furthest from the existing cluster centers. This point must lie within the data area. This precisely means the steps in sequences as: Pick first center randomly, next is the point furthest from the first center, next is the point furthest from both previous centers and so on till convergence. This greatly speeds up the clustering because of the less need of relocation and tuning. Filtered Clusterer (FC) is a variant of Simple K Means cluster technique but takes more processing time then SKM. FC algorithm is based on the techniques of filtering the similar and dissimilar clusters using Filtered Clusterer class.

Exception Maximization (EM) clustering algorithm is distance-based algorithm whose dataset modeling assumption is linear combination of multi variant normal distributions. This algorithm finds distribution parameters that maximize log likelihood. This algorithm is chosen to cluster data because it can handle high dimensionality because of its linearity and converges fast having given appropriate initialization. There are certain cluster analysis situation with real world dataset where the need we to perform cluster analysis a small region of interest. EM Clustering Algorithm gives optimized results even compare to result given by SKM algorithm.

VI. CLUSTERING TECHNIQUES IMPLEMENTATION ON DIABETES DATASET

1. Dataset Attributes Specification

This research work takes the data set from the UCI Web Repository for the purpose of aimed research. The dataset attributes descriptions are as under in Table 4:

Table 4: Diabetes_diagnosis Dataset Attributes

Attribute Name	Description
Pregnancies	Number of pregnancies
PG Concentration	Plasma glucose at 2 hours in an oral glucose tolerance test
Diastolic BP	Diastolic Blood Pressure (mm Hg)
Tri Fold Thick	Triceps Skin Fold Thickness (mm)
Serum Ins	2-Hour Serum Insulin (mu U/ml)
BMI	Body Mass Index: (weight in kg/ (height in m)^2)
DP Function	Diabetes Pedigree Function
Age	Age (years)

Table 5: Diabetes_Diagnosis Dataset Class

Class name	Description
Diagnosis	sick, healthy

2. Implementation using Clustering

Implementation using SKM

The Cluster central values are tabulated in Table 6.

Table 6: Cluster Central Values

Attribute	Full data (768)	Cluster 0 (515)	Cluster 1 (253)
Pregnancies	3.8451	2.0835	7.4308
PG Concentration	120.8945	115.3282	132.2253
Diastolic BP	69.1055	65.9903	75.4466
Tri Fold Thick	20.5365	21.8194	17.9249
Serum Ins	79.7995	85.0194	69.1739
BMI	31.9926	31.7751	32.4352
DP Function	0.4719	0.4708	0.4741
Age	33.2409	26.7728	46.4071

ALL* (768) = healthy (515) + sick (253)

This process took 7 iterations and to build the clustered datasets it took 0.02 second. Number of clusters selected by cross validation is 2 (0 through 1). The model and evaluation of clustered instance are tabulated in the below Table 7.

Table 7: Clustered Instances

Cluster no	Instances	Percentage	Diagnosis
0	515	67 %	Healthy
1	253	33 %	Sick

Implementation using FF

This research work used FF method for discovering pattern form datasets. In this technique incorrect instances are 263.

Cluster Central Values obtained by Weka tool are tabulated in below Table 8.

Table 8: Cluster Central values

Cluster	Central Values
0	2.0 83.0 66.0 23.0 50.0 32.2 0.497 22.0
1	4.0 197.0 70.0 39.0 744.0 36.7 2.329 31.0

The model and evaluation of clustered instance are tabulated in below Table 9.

Table 9: Cluster Instances

Cluster	Instances	Percentage	Diagnosis
0	759	99%	Healthy
1	9	1%	Sick

This FF technique resulted in 2 clusters comprising of healthy and sick diagnosis for Diabetes.

Implementation using FC

In Filtered Clusterer Data Mining technique number of iteration is seven. In this techniques cluster central values are tabulated in below Table 10.

Table 10: Cluster Central values

Attribute	Full Data (768)	Cluster 0 (515)	Cluster 1 (253)
Pregnancies	3.8451	2.0835	7.4308
PG Concentration	120.8945	115.3282	132.2253
Diastolic BP	69.1055	65.9903	75.4466
Tri Fold Thick	20.5365	21.8194	17.9249
Serum Ins	79.7995	85.0194	69.1739
BMI	31.9926	31.7751	32.4352
DP Function	0.4719	0.4708	0.4741
Age	33.2409	26.7728	46.4071

Incorrect instance are 255 and to build the clustered datasets it took 0.03 second. The model and evaluation of clustered instance are tabulated in below Table 11.

Table 11: Cluster Instance

Cluster	Instance	Percentage	Supplier
0	515	67 %	Healthy
1	253	33 %	Sick

This FC Technique resulted in 2 clusters comprising of healthy and sick diagnosis for Diabetes.

Implementation using EM

Number of clusters selected by cross validation is 7 (0 through 6). This research work finds the means and standard deviation of different parameter tabulated in Table. This process took 69.64 second to build to model the clustered the datasets. The model and evaluation of clustered instance are tabulated in below Table 12.

This EM Technique resulted in 7 clusters comprising of No class, Healthy, and Sick.

Table 12: Clustered Instances

Cluster no	Instances	Percentage	Diagnosis
0	42	5 %	No class
1	235	31 %	Healthy
2	84	11 %	No class
3	35	4 %	No class
4	190	25 %	No class
5	23	3 %	No class
6	159	21 %	Sick

VII. PREDICTIVE ANALYSIS WITH RESULTS

The Comparative Predictive results of Diabetes Diagnosis are tabulated in below Table 13.

Table 13: Predictive Results of Diabetes_Diagnosis

Clustering Technique	Healthy	Sick	No classes
SKM	67%	33%	-
FF	99%	1%	-
FC	67%	33%	-
EM	31%	21%	48%

The comparative result analysis which has used 4 clustering techniques has indicated the two techniques SKM and FC are giving identical results to classify healthy and sick clusters of the Diabetic datasets. The FF technique is not in tune with the analyzed techniques. The result given by the EM technique has indicated third category of cluster under the head of "no classes" in addition to healthy and sick clusters. Considering the proportion of Healthy and sick clusters instances if the no class instances are directed to healthy and sick clusters the results of EM technique also matches with SKM and FC.

Instead of repeating the technique with multiple datasets if multiple techniques is used with same datasets the suitability of technique to classify desired clusters can be achieved with much better accuracy for Predictive Analysis.

VIII. CONCLUSION WITH RESEARCH EXTENTION

Four clustering techniques have been applied in this paper, namely: Simple K Means(SKM), Farther First(FF), Filtered Clusterer(FC) and Exception Maximization(EM) which solve the issues of categorizing data into a number of clusters based on similarity attributes displaying similarity

within clusters and variations among clusters. The four Algorithms and techniques have been implemented, Compared, and finally predicted against a dataset for medical diagnosis of Diabetic Diagnosis.

The comparative result analysis of clustered Data results applying multiple clustering techniques to the same data has indicated that the appropriateness of the selection of clustering techniques can easily be asserting for predictive data analysis. The clustering techniques implemented here alone can be considered as a stand-alone approach but need to conjunction with various Clustering as well as Data Mining Techniques to Since there are more Clustering and Data Mining Techniques used in this work, the current work can be extended to conjunction the left out Clustering and Data Mining Techniques to arrive at a Optimum Comparative Accuracy level in Predictive Analysis.

REFERENCES

Digital Reference

[1] Hunter, L., "Life and Its Molecules: A Brief Introduction". AI Magazine, 25(1), 9-22, 2004.

[2] Houle, J.L., Cadigan, W., Henry, S., Pinnamaneni, A. & Lundahl, S., "Database Mining in the Human Genome Initiative. Whitepaper, Bio-databases.com", . Amita Corporation. Available:http://www.biodatabases.com/whitepaper.html, 2004.

Books

[1] Biomedical Science Practice: Experimental and Professional Skills (Paperback) by Hedley Glencross (shelved 1 time as *biomedical*) avg rating 0.0 — 2 ratings — published 2010.

[2] Biomedical Ethics by Tamara L. Roleff (Editor) (shelved 1 time as *biomedical*) avg rating 3.60 — 9 ratings — published 1998.

[3] Biological database modeling Jake Chen (Editor), Amandeep S. Sidhu (Editor).

[4] Data Mining for Bioinformatics Published October 26, 2012 by CRC Press By Sumeet Dua, Lousiana Tech University, USA; Pradeep Chowriappa, LA.

[5] Analysis of Biological Data, The By Michael Whitlock , Dolph Schluter.

National & International Journals

[1] Data Mining and Bioinformatics: First International Workshop by Mehmet M Dalkilic, Sun Kim, Jiong Yang.

[2] INTERNATIONAL JOURNAL OF DATA MINING AND BIOINFORMATICS ISSN ONLINE : 1748-5681 , PRINT : 1748-5673.

[3] Vol 1 No 2, 114-118 APPLICATION OF DATA MINING IN BIOINFORMATICS.

[4] International Journal of Data Mining and Bioinformatics Publication type: Journals. ISSN: 1748568.

[1] Proceedings Of The 11th International Workshop Data Mining In Bioinformatics China — August 12 , 2012.

[2] Proceedings of the third international workshop on Data and text mining in bioinformatics November 02 - 06, 2009 ACM New York, NY, USA ©2009.

[3] Gribskov, M., McLachlan, A.D., Eisenberg, D.: Profile analysis: detection of distantly related proteins. PNAS 84(13), 4355-4358 (1987).

[4] Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

Ms. Richa Mehta, MCA, is working as Lecturer in MCA Department of S K Patel Institute of Management & Computer Studies-a constituent college of Kadi Sarva Vishwavidyalaya, Gandhinagar. She is pursuing Doctor of Philosophy in Biological Data Visualization in the area of Bioinformatics and Medicalinformatics. She has total teaching experience of approximately four years. Her areas of Interest Bioinformatics, SPM, Software Architecture for Visualization and knowledge Discovery. He has published two research papers in peer reviewed Journals.

Mr. Rinal H. Doshi, is working as Assistant Professor-MCA Department at L.C.Institute of Technology, Bhandu affiliated to Gujarat Technological University. He has four and half years of teaching experience. He has completed MCA from Gujarat Vidyapith, Ahmedabad in the year 2008. In the year 2012, he joined PhD program at RK University at Rajkot. He has published three Research Papers in peer review Journals. His Research Interest includes Data Mining, Data Warehouse, Web Mining, Bioinformatics, information retrieval and E-Commerce etc.

Dr. Ashish N. Jani, Ph. D. is working as Assistant Professor in MCA Department of S K Patel Institute of Management & Computer Studies-a constituent college of Kadi Sarva Vishwavidyalaya, Gandhinagar. He has total teaching experience of 6 years. He has got funded project from GUJCOST. He is actively involved in consultancy work. His areas of Interest are Embedded System with RTOS, Mobile Computing and Computer Vision. He has published seven research papers in peer reviewed Journals and he is author of four books. Two research Scholars are working under his guidance leading to Pd. D. Degree. Recently he has been selected by Florida Atlantic University, USA to work as Post Doctoral Fellow at their site for collaborative research from October 01, 2012.