# SRM UNIVERSITY

# FORMAT FOR PREPARATION OF MINI PROJECT REPORT

## FOR

## B. TECH. – Information Technology

1. **ARRANGEMENT OF CONTENTS:**

The sequence in which the project report material should be arranged and bound should be as follows:

* 1. Cover Page & Title Page
  2. Bonafide Certificate
  3. Acknowledgements
  4. Abstract
  5. Table of Contents
  6. List of Tables
  7. List of Figures
  8. List of Symbols, Abbreviations and Nomenclature
  9. Chapters
  10. Appendices
  11. References

The table and figures shall be introduced in the appropriate places.

1. **PAGE DIMENSION AND BINDING SPECIFICATIONS:**

The dimension of the project report should be in A4 size. The project report should be bound using flexible cover of the thick white art paper. The cover should be **printed in black letters** and the text for printing should be identical.

1. **PREPARATION FORMAT:**

**3.1 Cover Page & Title Page** – A specimen copy of the Cover page & Title page of the project report are given in **Appendix 1.**

**3.2 Bonafide Certificate –** The Bonafide Certificate shall be in double line spacing using Font Style Times New Roman and Font Size 14, as per the format in **Appendix 2.**

The certificate shall carry the guide signature and shall be followed by the guide’s name, academic designation (not any other responsibilities of administrative nature), department and full address of the institution where the supervisor has guided the student. The term **‘GUIDE’** must be typed in capital letters between the supervisor’s name and academic designation.

**3.3 Abstract –** Abstract should be one page synopsis of the project report typed double line spacing, Font Style Times New Roman and Font Size 14.

**3.4 Table of Contents –** The table of contents should list all material following it as well as any material which precedes it. The title page and Bonafide Certificate will not find a place among the items listed in the Table of Contents but the page numbers of which are in lower case Roman letters. One and a half spacing should be adopted for typing the matter under this head. A specimen copy of the Table of Contents of the project report is given in **Appendix 3.**

**3.5 List of Tables –** The list should use exactly the same captions as they appear above the tables in the text. One and a half spacing should be adopted for typing the matter under this head.

**3.6 List of Figures –** The list should use exactly the same captions as they appear below the figures in the text. One and a half spacing should be adopted for typing the matter under this head.

**3.7 List of Symbols, Abbreviations and Nomenclature** – One and a half spacing should be adopted or typing the matter under this head. Standard symbols, abbreviations etc. should be used.

**3.8 Chapters** – The chapters may be broadly divided into (i) Introductory chapter, (ii) Requirement Analysis work (iii) Design (iv) Implementation (V) Testing and (VI) Conclusion.

The main text will be divided into several chapters and each chapter may be further divided into several divisions and sub-divisions.

* Each chapter should be given an appropriate title.
* Tables and figures in a chapter should be placed in the immediate vicinity of the reference where they are cited.
* Footnotes should be used sparingly. They should be typed single space and placed directly underneath in the very same page, which refers to the material they annotate.

**3.9 Appendices** – Appendices are provided to give supplementary information, which is included in the main text may serve as a distraction and cloud the central theme.

* Appendices should be numbered using Arabic numerals, e.g. Appendix 1, Appendix 2, etc.
* Appendices, Tables and References appearing in appendices should be numbered and referred to at appropriate places just as in the case of chapters.
* Appendices shall carry the title of the work reported and the same title shall be made in the contents page also.

**3.10 List of References** –The listing of references should be typed 4 spaces below the heading “REFERENCES” in alphabetical order in single spacing left – justified. The reference material should be listed in the alphabetical order of the first author. The name of the author/authors should be immediately followed by the year and other details.

A typical illustrative list given below relates to the citation example quoted above.

**REFERENCES**

1. Ariponnammal, S. and Natarajan, S. (1994) ‘Transport Phonomena of Sm Sel – X Asx’, Pramana – Journal of Physics Vol.42, No.1, pp.421-425.

2. Barnard, R.W. and Kellogg, C. (1980) ‘Applications of Convolution Operators to Problems in Univalent Function Theory’, Michigan Mach, J., Vol.27, pp.81–94.

3. Shin, K.G. and Mckay, N.D. (1984) ‘Open Loop Minimum Time Control of Mechanical Manipulations and its Applications’, Proc.Amer.Contr.Conf., San Diego, CA, pp. 1231-1236.

* + 1. **Table and figures -** By the word Table, is meant tabulated numerical data in the body of the project report as well as in the appendices. All other non-verbal materials used in the body of the project work and appendices such as charts, graphs, maps, photographs and diagrams may be designated as figures.

**4. TYPING INSTRUCTIONS:**

The impression on the typed copies should be black in colour.

One and a half spacing should be used for typing the general text. The general text shall be typed in the Font style ‘Times New Roman’ and Font size 14.

**\* \* \* \* \***

##### APPENDIX 1

##### RECOMMENDATION ENGINE USING MACHINE LEARNING

##### A MINI PROJECT REPORT

###### ***Submitted by***

##### URJASVIT SINHA(RA1511008010413)

##### PRAKHAR MITTAL(RA1511008010417)

##### SIDHARTH DUDHORIA(RA1511008010415)

##### KARAN RAJU(RA1511008010034)

##### SHREY AGARWAL(RA1511008010370)

***in partial fulfillment for the award of the degree***

***of***

##### B.TECH

IN

INFORMATION TECHNOLOGY



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR

APRIL & 2018

##### RECOMMENDATION ENGINE USING MACHINE LEARNING

##### A MINI PROJECT REPORT

###### ***Submitted by***

##### URJASVIT SINHA(RA1511008010413)

##### PRAKHAR MITTAL(RA1511008010417)

##### SIDHARTH DUDHORIA(RA1511008010415)

##### KARAN RAJU(RA1511008010034)

##### SHREY AGARWAL(RA1511008010370)

***in partial fulfillment for the award of the degree***

***of***

##### B.TECH

IN

INFORMATION TECHNOLOGY



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR

APRIL & 2018

APPENDIX 2

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR

**BONAFIDE CERTIFICATE**

Certified that this project report **“RECOMMENDATION ENGINE USING MACHINE LEARNING.”** is the bonafide work of “**…URJASVIT SINHA,PRAKHAR MITTAL,KARAN RAJU,SIDHARTH DUDHORIA,SHREY AGARWAL”** who carried out the project work under my supervision.

**SIGNATURE SIGNATURE**

**MS. K.NIMALA** **DR.G.VADIVU**

**HEAD OF THE DEPARTMENT Information Technology**

INTERNAL EXAMINER

**DECLARATION**

I…URJASVIT SINHA(RA1511008010413), PRAKHAR MITTAL(RA1511008010417), KARAN RAJU(RA1511008010034), SIDHARTH DUDHORIA(RA1511008010415), SHREY AGARWAL(RA1511008010370)studying in III year B.Tech Information Technology program at, SRM University, Kattankulathur, Chennai, hereby declare that this project is an original work of mine and I have not verbatim copied / duplicated any material from sources like internet or from print media, excepting some vital company information / statistics and data that is provided by the company itself.

Signature of the Student

Date: 12th April 2018

Place: Chennai

**ACKNOWLEDGEMENT**

The success and the final outcome of this project required guidance and assistance from different sources and we feel extremely fortunate to have got this all along the completion of our project. Whatever we have done is largely due to such guidance and assistance and we would not forget to thank them.

We express our sincere thanks to the Head of the Department, Department of Information Technology, **Dr.G.Vadivu**, for all the help and infrastructure provided to us to complete this project successfully and his valuable guidance.

We owe our profound gratitude to our project guide **Ms. K.Nimala**, who took keen interest in our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all the Teaching staff of the Department of Information Technology which helped us in successfully completing our minor project work. Also, we would like to extend our sincere regards to all the non-teaching staff of the department of Information Technology for their timely support.

URJASVIT SINHA(RA1511008010413)

PRAKHAR MITTAL(RA1511008010417)

KARAN RAJU(RA1511008010034)

SIDHARTH DUDHORIA(RA1511008010415)

SHREY AGARWAL(RA1511008010370)

APPENDIX 3

TABLE OF CONTENTS

**CHAPTER NO. TITLE PAGE NO.**

ABSTRACT iii

**LIST OF TABLE xvi**

LIST OF FIGURES xviii

**LIST OF SYMBOLS xxvii**

**1. INTRODUCTION 1**

1.1 MACHINE LEARNING 1

1.2 USES OF MACHINE LEARNING 2

1.2.1 5

1.2.2 . . . . . . . . . . . 12

1.2.2.1 General 19

1.2.2.2 . . . . . . . . . . 25

1.2.2.3 . . . . . . . . . . 29

1.2.3 . . . . . . . . . . . . 30

1.3 . . . . . . . . . . .. . . . . . . 45

1.4 . . . . . . . . . . . . . . . . . . 58

**2. REQUIREMENT ANALYSIS 69**

2.1 GENERAL 75

* 1. . . . . . . . . . . 99

2.2 ……………. 100

**Chapters** – The chapters may be broadly divided into (i) Introductory chapter, (ii) Requirement Analysis work (iii) Design (iv) Implementation (V) Testing and (VI) Conclusion.

**Minimum Number of pages is 45 – 50**

1. **INTRODUCTION**

**1.1 ABOUT MACHINE LEARNING**

**Machine learning** is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.[1]

The name *machine learning* was coined in 1959 by **Arthur Samuel**. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data– such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank, and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioural profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

**1.2 Uses of Machine Learning:**

1.Stopping Malware:

Kaspersky Lab reported it was detecting 325,000 new malicious files every day. At that rate, humans and even signature-based security solutions can't keep up, which is why machine learning and deep learning are necessary.

2. Medicines and Health Care:

Medecision used a machine learning platform to gain a better understanding of diabetic patients who are at risk for avoidable hospitalization or emergency room use. It trained the platform on a database of approximately 8 million patients. The model identified seven or eight independent variables that can be used to predict avoidable hospitalizations.

## 3. Understand Legalese:

Legal documents are often too complicated for the average person to easily comprehend. Some hire a lawyer. Others may skim the documents, or even ignore a document's content, hoping that somehow everything will work out. Some are overconfident about their ability to understand the documents.

Deep learning can help. In addition to translating legal language into plain language, Legal Robot can determine what's missing from a contract and whether there are elements in a contract that shouldn't be there, such as a royalty fee section in a non-disclosure agreement.

4. Prevent Fraud and Money Laundering:

PayPal is using deep learning to prevent fraud and money laundering at granular levels. By combining deep learning with machine learning and other tools, the company can precisely discern between legitimate and fraudulent buyers and sellers. According to Hui Wang, PayPal's senior director of global risk sciences, it's all about anomaly detection.

## 5. Improve Cybersecurity:

Israeli communication services provider Orange (aka Partner) has been using machine learning for the past two years to help protect its business and customer data.

The system monitors all traffic coming from and being exchanged among PCs and servers, combinations of those things and more to identify anomalous behavior and to minimize its impact. Recently, the system detected malicious code in a video file that an employee had downloaded. The security team instantly notified the employee.

6. Recommendation Engine:

Often termed as Recommender Systems, they are simple algorithms which aim to provide the most relevant and accurate items to the user by filtering useful stuff from of a huge pool of information base. Recommendation engines discovers data patterns in the data set by learning consumers choices and produces the outcomes that co-relates to their needs and interests.

### 1.2 Machine learning tasks:

Machine learning tasks are typically classified into two broad categories, depending on whether there is a learning "signal" or "feedback" available to a learning system:

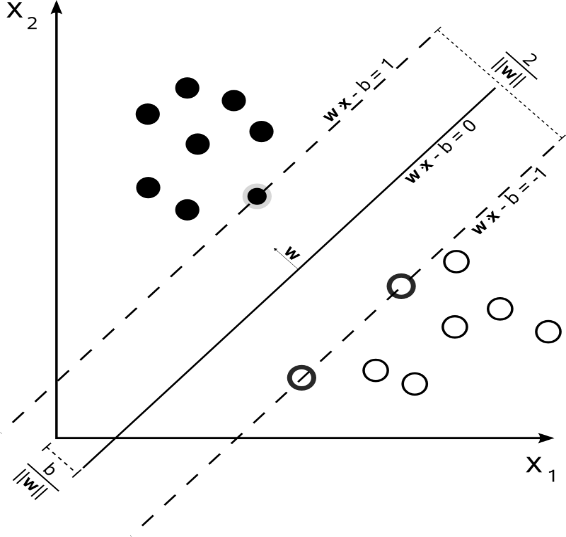
* **Supervised learning:** The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs. As special cases, the input signal can be only partially available, or restricted to special feedback:
  + Semi-supervised learning: the computer is given only an incomplete training signal: a training set with some (often many) of the target outputs missing.
  + Active learning: the computer can only obtain training labels for a limited set of instances (based on a budget), and also has to optimize its choice of objects to acquire labels for. When used interactively, these can be presented to the user for labeling.
  + Reinforcement learning: training data (in form of rewards and punishments) is given only as feedback to the program's actions in a dynamic environment, such as driving a vehicle or playing a game against an opponent.
* **Unsupervised learning:** No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

### 1.3 Machine learning applications:

Another categorization of machine learning tasks arises when one considers the desired *output* of a machine-learned system:[5]:3

* In classification, inputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are "spam" and "not spam".
* In regression, also a supervised problem, the outputs are continuous rather than discrete.
* In clustering, a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task.
* Density estimation finds the distribution of inputs in some space.
* Dimensionality reduction simplifies inputs by mapping them into a lower-dimensional space. Topic modeling is a related problem, where a program is given a list of human language documents and is tasked to find out which documents cover similar topics.

Among other categories of machine learning problems, learning to learn learns its own inductive bias based on previous experience. Developmental learning, elaborated for robot learning, generates its own sequences (also called curriculum) of learning situations to cumulatively acquire repertoires of novel skills through autonomous self-exploration and social interaction with human teachers and using guidance mechanisms such as active learning, maturation, motor synergies, and imitation.



A support vector machine is a classifier that divides its input space into two regions, separated by a linear boundary. Here, it has learned to distinguish black and white circles.

**1.4 Approaches to Machine Learning:**

### Decision tree learning

Decision tree learning uses a decision tree as a predictive model, which maps observations about an item to conclusions about the item's target value.

### Association rule learning

Association rule learning is a method for discovering interesting relations between variables in large databases.

### Artificial neural networks

An artificial neural network (ANN) learning algorithm, usually called "neural network" (NN), is a learning algorithm that is vaguely inspired by biological neural networks. Computations are structured in terms of an interconnected group of artificial neurons, processing information using a connectionist approach to computation. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs, to find patterns in data, or to capture the statistical structure in an unknown joint probability distribution between observed variables.

#### Deep learning

Falling hardware prices and the development of GPUs for personal use in the last few years have contributed to the development of the concept of deep learning which consists of multiple hidden layers in an artificial neural network. This approach tries to model the way the human brain processes light and sound into vision and hearing. Some successful applications of deep learning are computer vision and speech recognition.

### Inductive logic programming

Inductive logic programming (ILP) is an approach to rule learning using logic programming as a uniform representation for input examples, background knowledge, and hypotheses. Given an encoding of the known background knowledge and a set of examples represented as a logical database of facts, an ILP system will derive a hypothesized logic program that entails all positive and no negative examples. Inductive programming is a related field that considers any kind of programming languages for representing hypotheses (and not only logic programming), such as functional programs.

### Support vector machines

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

### Clustering

Cluster analysis is the assignment of a set of observations into subsets (called *clusters*) so that observations within the same cluster are similar according to some predesignated criterion or criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some *similarity metric* and evaluated for example by *internal compactness* (similarity between members of the same cluster) and *separation* between different clusters. Other methods are based on *estimated density* and *graph connectivity*. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis.

### Bayesian networks

A Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms exist that perform inference and learning.

### Reinforcement learning

Reinforcement learning is concerned with how an *agent* ought to take *actions* in an *environment* so as to maximize some notion of long-term *reward*. Reinforcement learning algorithms attempt to find a *policy* that maps *states* of the world to the actions the agent ought to take in those states. Reinforcement learning differs from the supervised learning problem in that correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected.

### Representation learning

Several learning algorithms, mostly unsupervised learning algorithms, aim at discovering better representations of the inputs provided during training. Classical examples include principal components analysis and cluster analysis. Representation learning algorithms often attempt to preserve the information in their input but transform it in a way that makes it useful, often as a pre-processing step before performing classification or predictions, allowing reconstruction of the inputs coming from the unknown data generating distribution, while not being necessarily faithful for configurations that are implausible under that distribution.

Manifold learning algorithms attempt to do so under the constraint that the learned representation is low-dimensional. Sparse coding algorithms attempt to do so under the constraint that the learned representation is sparse (has many zeros). Multilinear subspace learning algorithms aim to learn low-dimensional representations directly from tensor representations for multidimensional data, without reshaping them into (high-dimensional) vectors. Deep learning algorithms discover multiple levels of representation, or a hierarchy of features, with higher-level, more abstract features defined in terms of (or generating) lower-level features. It has been argued that an intelligent machine is one that learns a representation that disentangles the underlying factors of variation that explain the observed data.

### Similarity and metric learning

In this problem, the learning machine is given pairs of examples that are considered similar and pairs of less similar objects. It then needs to learn a similarity function (or a distance metric function) that can predict if new objects are similar. It is sometimes used in Recommendation systems.

### Sparse dictionary learning

In this method, a datum is represented as a linear combination of basis functions, and the coefficients are assumed to be sparse. Let *x* be a *d*-dimensional datum, *D* be a *d* by *n* matrix, where each column of *D* represents a basis function. *r* is the coefficient to represent *x* using *D*. Mathematically, sparse dictionary learning means solving x ≈ D r {\displaystyle x\approx Dr} where *r* is sparse. Generally speaking, *n* is assumed to be larger than *d* to allow the freedom for a sparse representation.

Learning a dictionary along with sparse representations is strongly NP-hard and also difficult to solve approximately. A popular heuristic method for sparse dictionary learning is K-SVD.

Sparse dictionary learning has been applied in several contexts. In classification, the problem is to determine which classes a previously unseen datum belongs to. Suppose a dictionary for each class has already been built. Then a new datum is associated with the class such that it's best sparsely represented by the corresponding dictionary. Sparse dictionary learning has also been applied in image de-noising. The key idea is that a clean image patch can be sparsely represented by an image dictionary, but the noise cannot.

### Genetic algorithms

A genetic algorithm (GA) is a search heuristic that mimics the process of natural selection, and uses methods such as mutation and crossover to generate new genotype in the hope of finding good solutions to a given problem. In machine learning, genetic algorithms found some uses in the 1980s and 1990s.Conversely, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms.

### Rule-based machine learning

Rule-based machine learning is a general term for any machine learning method that identifies, learns, or evolves `rules’ to store, manipulate or apply, knowledge. The defining characteristic of a rule-based machine learner is the identification and utilization of a set of relational rules that collectively represent the knowledge captured by the system. This is in contrast to other machine learners that commonly identify a singular model that can be universally applied to any instance in order to make a prediction. Rule-based machine learning approaches include learning classifier systems, association rule learning, and artificial immune systems.

#### Learning classifier systems

Learning classifier systems (LCS) are a family of rule-based machine learning algorithms that combine a discovery component (e.g. typically a genetic algorithm) with a learning component (performing either supervised learning, reinforcement learning, or unsupervised learning). They seek to identify a set of context-dependent rules that collectively store and apply knowledge in a piecewise manner in order to make predictions.

**1.5 Working with Recommendation Engines:**

Two types of recommendation algorithms that are also used by most of the tech giants like Google and Facebook in their advanced recommender system modules as a typical business problem.

Two Types are:

* Content Based Recommendations
* Collaborative Filtering Algorithm

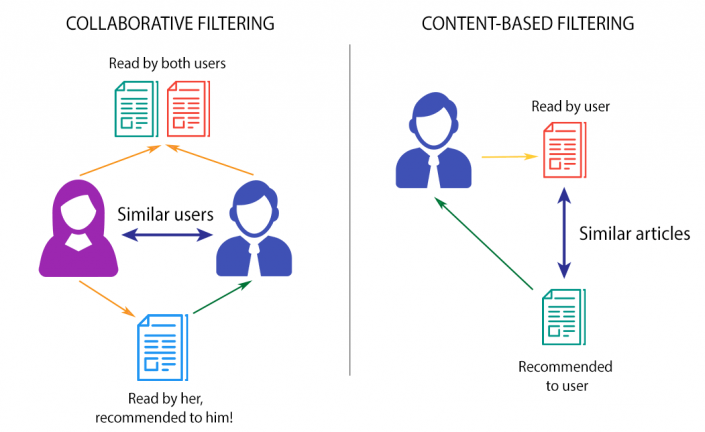
**1.5.1 Content Based Recommendations:**

Content based systems, recommends item based on a similarity comparison between the content of the items and a user’s profile. The feature of items are mapped with feature of users in order to obtain user – item similarity.

* Several issues have to be considered when implementing a content-based filtering system.
* First, terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms from items.
* Second, the terms have to be represented such that both the user profile and the items can be compared in a meaningful way.
* Third, a learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile.

**1.5.2 Collaborative Filtering Algorithm:**

* Content-based recommendation lacks in detecting inter dependencies or complex behaviours.
* For example: People might like smartphones with Good Display, only if it has retina display and wouldn’t otherwise.
* Collaborative Filtering algorithm considers “User Behaviour” for recommending items. They exploit behaviour of other users and items in terms of transaction history, ratings, selection and purchase information.
* Other users behaviour and preferences over the items are used to recommend items to the new users. In this case, features of the items are not known.



**2. REQUIREMENT ANALYSIS**

**2.1 Requirement Definition and Analysis:**

Requirements elicitation(analysis) & deﬁnition focuses on the collection of requirements from different stakeholders. Typical resulting artifacts are, for example, textual requirement descriptions, scenario descriptions, use cases, and sketches of prototypical user interfaces. The following recommendation approaches support activities related to requirements elicitation & deﬁnition:

**Collection of data:**

* The first step in creating a recommendation engine is gathering data. Data can be either explicit or implicit data. Explicit data would consist of data inputted by users such as ratings and comments on products.
* And implicit data would be the order history/return history, Cart events, Pageviews, Click thru and search log. This data set will be created for every user visiting the site.

**Storing the data:**

* The more data you can make available to your algorithms, better the recommendations will be. This means that any recommendations project can quickly turn into a big data project.
* The type of data that you use to create recommendations can help you decide the type of storage you should use. You could choose to use a standard SQL database.

**Analysing the data:**

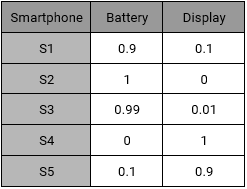
* **Real-time systems** can process data as it’s created. This type of system usually involves tools that can process and analyse streams of events. A real-time system would be required to give in-the-moment recommendations.
* **Batch analysis** demands you to process the data periodically. This approach implies that enough data needs to be created in order to make the analysis relevant, such as daily sales volume. A batch system might work fine to send an e-mail at a later date.
* **Near-real-time analysis** lets you gather data quickly so you can refresh the analytics every few minutes or seconds. A near-real-time system works best for providing recommendations during the same browsing session.

**2.2 Problems and Solution on Content Based Recommendations:**

Q.1 Consider a scenario of an e-commerce website which sells thousands of smartphones. With growing number of customers every day, the task in hand is to showcase the best choices of smartphones to the users according to their tastes and preferences.

To understand how recommendation engine works, let’s slice the data into a sample set of five smartphones with two major features “Battery and Display”. The five smartphones have following properties:

* S1 has good battery life but poor display
* S2 has an amazing battery performance but very rough display
* S3’s battery is one of the best but display lacks quality
* S4 & S5 are good in terms of display but poor in terms of battery performance.
* Using these characteristics, we can create an ***Item – Feature Matrix***. Value in the cell represents the rating of the smartphone feature out of 1.



Our sample set also consist of four active users with their preferences:

**Aman**: He prefers battery over display as an ideal smartphone feature.

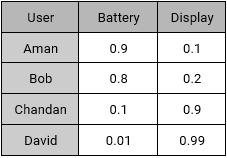
**Bob**: He likes a long lasting battery.

**Chandan**: For Chandan, display should be decent, battery should be normal.

**David**: For David, Display is extremely important but not the battery.

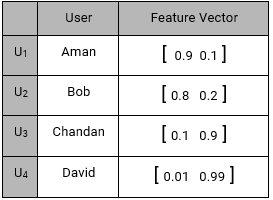
**2.2.1 Solution to Problem:**

Using their interests, we can create a *User – Feature Matrix* as follows:

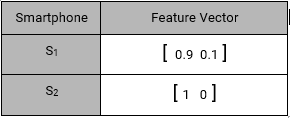
****

Two matrices: *Item – Feature* and *User – Feature.* We can create the recommendation of smartphones for our users using algorithms.

The top matched pairs are given as recommendations, as demonstrated below: Representing every user by a feature vector:



Also, every item representation as a feature vector:



Content Based ***Item – User Mapping Recommendations*** are given by the equation:

MAX ( U(j)T . I(i) )                                 i,j -> n,m

For User U1 (Aman), Smartphone recommendation is:

MAX( U1TS1, U1TS2, U1TS3, U1TS4, U1TS5)

MAX([0.9 0.1]T [0.9  0.1], [0.9  0.1]T [1  0], [0.9  0.1]T [0.99 0.01], [0.9  0.1]T [0.1 0.9],

[0.9  0.1]T [0.01  0.99])

MAX(0.82 , 0.9 , 0.89 , 0.18 , 0.10)

= S2(0.9), S3(0.89) & S1(0.82)

* Smartphones S2, S3 and S1 has the highest recommendation scores,
* Hence S2, S3 and S1 are recommended to Aman.