1.0 INTRODUCTION

1.1 RESEARCH BACKGROUND

The Internet is a massive resource pool where versatile collections of data are available. Since the huge volume of information is available in the World Wide Web, it is quite difficult for users to take the right decision and arrive at the decisive solution on the web in order to retrieve their required information. Due to the heterogeneous nature of web data, users lack in confidence and find it difficult to select the essential information on the web. Therefore, any concrete system needs to be developed to address this issue.

It is required to propose an authenticated recommendation that can offer and provide the required guidance to users on the web to select required information. The recommendation with relevant information about various services like the selection of right products, obtaining career guidance information, movies and books which are required by the users. The earlier version of the recommendation system is spread through ‘word of mouth’ which is prominently used by many users to buy a new product or to select the information in the online by analyzing the opinions and feedbacks of the various users.

In recent time, the online recommendation system is enhanced to the new heights by choosing social media as a primary source of data to provide the more interactive and useful recommendation for the user. The evolution of social media has also brought the massive volume of data which include reviews, comments, posts, tweets, tag and opinions from various social media networks such as Twitter, Facebook, LinkedIn and other e-commerce sites like Amazon and Flipkart.

The extraction of key features like tweets, comments, posts etc; from social media data needs logical interpretation, therefore it is the responsibility of the users to collect, process and analyze the data based on users' behaviors, activities, reviews, ratings, features and preferences. It is a
challenging process to develop an online recommendation system with user personalization and authentication. It is the social responsibility of the researcher to provide the authenticated and suitable recommendations system for the benefit of the users. Hence, this research work focuses on developing a generic framework for online recommendation system in heterogeneous social media environment to satisfy the need of the various users. Debashis Das et al., (2017) describe the possible approaches and algorithms used in the recommendation system. The existing online recommendation system is driven through approaches such as content-based and collaborative-filtering approach in which the recommendation is based on items that are similar in content to items the user has preferred. Both the approaches have its own merit and demerit, but the fact that, reliability and flexibility in terms of recommendations are not consistent, therefore the hybrid approach is used in this research which combines both content-based and collaborative filtering technique in a semantic way.

The existing online recommendation system in the social media environment is also examined which brings the various challenges that as to be addressed. Jian Wei et al.,(2017) has highlighted the issue of cold start problem in recommendation systems such as complete cold start problem (CCSP) and the incomplete cold start problem (ICSP). Complete cold start problem in which no rating records are available for the user to choose the product and incomplete cold start problem in which a small number of rating records is available for users in the system. The proposed Personalized and Optimal Ranking System (PORS) Framework for Recommendation in heterogeneous social media environment is implemented in such a way that it overcomes the common challenges in usual recommendation system such as sparsity, cold-start, lack of data and other issues in the existing system. In the social media environment, the data are in different patterns such as comments, posts, reviews, features, tweets. Therefore, streamlining these data into an interoperable one is the challenging task.
The Personalized and Optimal Ranking System (PORS) for a recommendation system in the social media environment overcomes the issues involved in the existing system. The data which are used in this process are (i) social media like Twitter and Facebook (ii) e-commerce sites like Amazon and Flipkart. The system is designed in such a way that, the involved process strikes out most of the issues discussed in the existing system. The process starts with invoking the query from the user through interface who requests for a recommendation in online. Hariton A. Efstathiades (2013) describes knowledge extraction from the social network. The data are collected from online social network mainly through a programming procedure with the use of a suitable application programming interface. However, the majority of social science researchers usually have no programming experience. Thus, the procedure from data retrieval to knowledge extraction is not an easy task. In this research work, the acquisition of data such as reviews and features are extracted from the social media such as Facebook and Twitter and e-commerce sites like Amazon and Flipkart.

The extracted data is in heterogeneous type; therefore it is necessary to convert them into interoperable one. The collected social media data are stored in the data repository from which the review data and features are fetched as per the user query. The acquired data such as tweets, reviews and ratings are warehoused in the repository which can be retrieved through query based on the keywords and necessary recommendations are offered. The Personalized and Optimal Ranking System (PORS) framework uses the linear regression analysis to generate the ranking the similar products and service information.

The challenges faced in the existing system systematically addressed in the proposed work. This system ensures that the suitable recommendations are offered which can match the expectation of the online web users to a great extent in order to fetch the right information at right time.
1.2 RECOMMENDATION SYSTEM

It is quite important to understand the concept of basic recommendation system and need of social media in the current scenario. The core part of this thesis is driven after getting complete insights into the recommendation system. The recommendation system is an online software application or a tool for helping and guiding the users to the most related items which are selected, rated, preferred and ranked by the other user, based on different criteria. These suggestions may be in terms of different intentions, such as products to buy, student to choose their career options, people who can be selected as friends in a social network, selecting the movie to be watched, choosing the institute to pursue higher education or online news to be read. In other words, the system offers the personalized recommendation to the requesting user based on information obtained such as reviews, features, ratings, user feedback and user personalized information.

1.2.1 The objective of Recommendation System (RS)

The main objective of the recommendation system is to fabricate significant suggestions to the user who finds it difficult to take wise decision online during the retrieval of information. Various scenarios can be identified where the user on the web struggles to take a right call. The users on the web need to rely on the authenticity and reliability of the data, which are offered by the recommendation system with ease. Recommendation system ensures that the information suggested is an authenticated and verified one.

Recommendation system, in general, disseminates the information from various channels such as reviews of the authenticated user, ratings of the group of people who gives genuine opinions, key features of the item, user information like user’s preferences, user’s budget, user’s favourite, user’s profession. This information strengthens the recommendation system to the next level in terms of authentication and validation. Therefore, the primary aim of the recommendation system is quite transparent, but the fact
that the designing of such recommendation engines is a hectic task which depends on the domain and the particular characteristics of the information available on the web sources

1.2.2 Types of Recommendation Techniques

The core part of this recommendation system is the uniqueness in offering the suggestion to the corresponding user. The recommendations to the user are offered using various strategy and logic. The technical support provides the necessary recommendation which is described via various techniques used in this area of research. There are different approaches to build a recommendation system:

- Collaborative filtering technique – It is a method of building automated recommendations about the interests of a user by collecting user preferences or taste information from many users. It is also based on users interaction such as likes, views, ratings, comments; this technique is extremely popular in online services, shops.

- Content-based filtering technique – Suggests items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of information, typically the words that occur in a document. It is also based on the similarity of items which is used to suggest text articles, songs etc,

- Knowledgebase technique – This techniques persist on users who explicit knowledge of the item or product to offer a most predictable recommendation, since the users specify requirements for impersonal recommendations

- Hybrid technique – a combination of collaborative and content-based technique. In which both users’ feedback and information about the item or product are considered for the recommendation.
The various techniques of the recommendation system elicit the basic need of constructing a brand new framework for recommendation system with all necessary requirements. Table 1.1 shows the pros and cons of the various techniques which emphasize the point of selecting the right approach to deliver a suitable recommendation system.

### 1.2.3 Issues and Challenges

Some of the issues and challenges involved in the recommendation system are listed in table 1.2. The listed issues reflect the challenges involved in the established recommendation system with a specific framework which needs to be addressed first.
### Table 1.2: Issues & Challenges in Recommendation System

<table>
<thead>
<tr>
<th>Problems &amp; Challenges in RS</th>
<th>Description</th>
<th>Impact/Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparsity</td>
<td>It is common in e-commerce and other domains that people usually purchase or rate relatively few items compared with the total number of items. That leads to a sparse users-items representation matrix and therefore, inability to locate neighbours or derive common behaviour patterns and the final result is low-quality recommendations.</td>
<td>High-Low</td>
</tr>
<tr>
<td>Cold-start</td>
<td>The cold start problem arises when there is a little information about content is available and information without user ratings. This issue affects every recommendation system.</td>
<td>Low–High</td>
</tr>
<tr>
<td>Loss of neighbour transitivity</td>
<td>Assume that user A is highly correlated with user B, user B is highly correlated with user C. Possibly, user C is also highly correlated with user A. Such relationships are not captured by recommendation systems.</td>
<td>High-Low</td>
</tr>
<tr>
<td>Synonymy</td>
<td>This issue arises rarely when the naming of the item is interchanged. For example, an item “Laptop HP” and an item “Bag for laptops” are both neither laptops nor bags, but both are of class “IT products” and, therefore, have a degree of commonality overseen.</td>
<td>Neutral</td>
</tr>
</tbody>
</table>
### Problems & Challenges in RS

<table>
<thead>
<tr>
<th>Problem</th>
<th>Description</th>
<th>Impact/Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>Scalability problem arrives when the size of the database grows along with the ratings of the items. It is beneficial to try to make systems, which can handle large amounts of data and produce accurate recommendations quickly.</td>
<td>High-Low</td>
</tr>
<tr>
<td>Recommendation quality</td>
<td>Among other details, the user is sensible for false negatives (incorrect recommendations, which the user does not like). Assume the user likes genre Sci-Fi and highly rated many other Sci-Fi movies. If the recommendation system will rate “The Matrix” as bad one, but the user likes it, the prediction will be a false negative. In such cases, users lose trust in the system and stop using it.</td>
<td>Low-High</td>
</tr>
<tr>
<td>Privacy</td>
<td>Privacy issue found in most of the recommendation system. Since the personalization factor is not addressed when the information is processed in bulk.</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

### 1.3 RESEARCH OBJECTIVES

The primary objective of this research is to develop a generic framework for online recommendation system for selection of products, providing career guidance information, selection of movies, books, academic tutors, Institute etc., Reviews and features are acquired from social media such as Facebook, Twitter and e-commerce sites like Amazon and Flipkart from which the end user gets the opinions and feedbacks of the various items which are analyzed and processed for personalized recommendation system.
along with ranking for the requested user. The problems which are considered:

- To compare the existing technique of recommendation system and to enumerate the need of the proposed (PORS) framework.
- To convert user query into interoperable format so that it can generate the required reviews and features from the different web source.
- To enable the heterogeneous social media data into interoperable one.
- To offer a recommendation along with the ranking with user authentication and personalization.
- To evaluate the performance of the system by comparing the recommendations, whether the solution provided is effective enough or not.

1.4 **RESEARCH CONTRIBUTIONS**

The contributions of this research work are as follows,

- The main contribution of this research work is to provide the generic framework for online recommendation system in a heterogeneous social media environment.
- This framework is applicable for various domains to offer suitable recommendations and ranking.
- The transformation of user query into interoperable one. Users post their queries in a natural language statement which is appropriately transmitted into the interoperable format.
- The social media information which flows from different sources such as Facebook, Twitter; are acquired in more dynamic way. The social media data is acquired effectively from different web source which extracts the reviews and opinions.
- The acquired reviews are classified using SVM classification. The features are derived from the reviews with the help of revised RAKE technique.
● This research focuses on the quality of the recommendation with the fullest user satisfaction. Therefore, the opinions of the users such as reviews, comments and features are analyzed using sentiment analysis.
● The analyzed information is implemented with the help of a hybrid algorithm which measures the sentiments of the reviews to offer the suggested recommendations to the users.
● Some of the other contributions of this research work include user authentication, user personalization and user satisfaction.
● User-personalized information is also gathered to provide the recommendation based on the user’s experience, preferences, interest, ability etc.
● Linear regression analysis is used for ranking along with the recommendation based on the similarity.
● This research study is foresighted to add few more features and provisions in the future. The dataset used in this study is extended to the other domain or the area where the actuality of the data is not affected much.
● The framework developed in this research work can be applied to the multimedia formats of data and other stringy kinds of data such as emoticons, gifs.

1.5 ORGANIZATION OF CHAPTERS

The outline of the thesis consists of seven major chapter which deals with the research background and basic concepts of recommendation system. The various issues and challenges involved in the recommendation system are discussed. The research objectives and contribution of this research work are identified and discussed in chapter 1. Chapter 2 presents literature review in which the scope of the review and different dimension of the literature survey is dealt with in brief.

Chapter 3 describes the constructive development of proposed framework and emphasizes the need of the framework. This chapter also
elucidates the query processor phase where the user queries are processed, tested and converted into interoperable format for further usage. The sample queries were tested and results are also discussed.

Chapter 4 describes the data acquisition from heterogeneous social media. This chapter includes the data classifier phase where the complete techniques and methods involved in extracting the data from the social media are summarized. The needy algorithms and stored procedures are also discussed in this section. The result of the data extraction is also recorded and analyzed in detail.

In Chapter 5, recommendation scenario and information required for the reliable recommendation are discussed in detail. It includes the algorithm and techniques to compare the keyword with user query to offer a trustworthy recommendation.

Chapter 6 enumerates the necessity of ranking along with recommendation. This section briefly describes the requirement of ranking scenario where the feature similarity mapping and feature analysis are carried out for the good number of available data. This chapter also emphasizes the key aspect of recommendation system which explains the importance of user personalization where the constant user feedbacks are collected and cross verified with the recommendation. In the same section, the summary is presented for the experiments and results which are obtained to measure performance of the system.

Chapter 7 concludes the thesis which illustrates the contribution of the researcher in the online recommendation system. It summarizes the scope for the further study in the same field. This chapter ends with major contribution to the study and concluding remark.
2.0 LITERATURE REVIEW

2.1 INTRODUCTION

This research work is fabricated on the prior work done in the field of recommendation system in social media environment and the various approaches used by the researchers across the globe on various domains. The literature review is structured in such a way that, the basic techniques and mechanism are followed in the recommendation system scenarios which are examined in detail. The integral part of the proposed study can be classified into several categories such as processing the user query, extracting the data from the social web source, storing the data in an organized way, retrieving the right information on right time via recommendation and offering a reliable ranking to the requested user based on feature analysis and user information. Therefore, the extensive literature survey on the existing study clarifies the need for the proposed system.

The prerequisite to addressing the challenges and issues during the construction of new framework is projected in two different dimensions. Firstly, the user query processing includes how the user query is processed using existing approaches and what are the challenges raised and how to overcome those challenges. Secondly, the related study of recommendation system as a framework is completely analyzed via conducting analytical research on the existing recommendation system.

2.1.1 User Query Processor

In the beginning of this chapter, the effort taken to understand the possible work done in the user query process is discussed. Initially, various attempts have been made to process the user query to convert it into SQL to access the data from the database. Conversion of natural language to dedicated SQL query to database is discussed. Posting query to databases in natural language is a simple task to access the data, especially for unfussy users who do not recognize the complex database query language such as
SQL. The social web sources are the combination of various aspects, therefore the flexible interface is in need. Some of the contributions made by several researchers are highlighted in the following section.

Satav, et al., (2014) describe a system that provides search interface/ NLP System for users without knowing any specific syntax or knowledge of a database language. They present a system that will provide the search interface with accuracy and efficiency for users especially for online applications, search engines and many other different databases. The analysis of the existing system shows that the user is not restricted to formulate any kind of query. The query posted by the users which invoke a random search rather than meaningful search which will be encountered in this study.

Kaur.G et al., (2013) emphasizes the usage of regular expressions in NLP to search text and it is considered as a useful technique. Regular Expressions are generic representations for a string or a collection of strings. Regular expressions (regexps) are one of the most useful tools in Computer Science. NLP as an area of computer science has greatly benefited from regexps: it is used in phonology, morphology, text analysis, information extraction, & speech recognition. It helps a reader to give a general review on the usage of regular expressions from natural language processing, but at times the clear collection of expression in a sentence will not be specified clearly which is acting as an issue which will be addressed in this research study.

Gaikwad.M.P (2013) discusses the need for natural language interfaces to database which has become increasingly acute as more and more people access information from web browsers. Yet NLI (Natural Language Interface) is used if the mapping of natural language triggers SQL queries promptly. Natural Language processing is becoming one of the most active areas in Human-Computer Interaction. The goal of NLP (Natural Language Processing) is to enable communication between people and computers without requiring to memorization of complex commands and
procedures. The main purpose of natural language query processing is for a sentence which is to be interpreted by the computer and appropriate action taken for it. The attempts made to convert the NLP into SQL are not up to the mark since the NLI uncertain to accept the complicated natural language statement from the user.

Kaur.J et al., (2013) describe the purpose of natural language query processing which is used to interpret an English sentence and hence a complementary action is taken. Querying to databases in natural language is a convenient method for data access, especially for new users who have less knowledge about complicated database query languages such as SQL. The author emphasizes the structural designing methods for translating English query into SQL using automata. A system which is capable of handling simple queries with standard join conditions are introduced, but all forms of SQL queries are not supported. The processing of complicated queries is quite a challenging task and therefore further development would be required.

Database plays a vital role in web-based system. Here are some examples of database NLP systems proposed by Anil M. Bhadgale et al., (2013) describes a system that answered questions about rock samples brought back from the Moon. LIFER/LADDER is one of the first database NLP systems. It is designed as a natural language interface to a database of information about US Navy ships. It uses a semantic grammar to parse questions and framed query for a distributed database. But, the fact is LIFER/LADDER system could only support simple one-table queries or multiple table queries with easy join conditions which definitely require upgrading in the future work which is focused in this research work. In the online recommendation system, a user interface is very much required to get the query from the user. The query posted by the user is of normal conventional statement which must be converted into official structured query; therefore several logical attempts have been made in the proposed research study to overcome issues encountered in the existing system.
Agrawal et al., (2013) describes a method for semantic analysis of natural language queries for Natural Language Interface to Database (NLIDB) using domain Ontology. Implementation of NLIDB for serious applications like railway inquiry, airway inquiry, corporate or government call centers require higher precision. This can be achieved by increasing role of language knowledge and domain knowledge at the semantic level. Design of semantic analyzer should behave in such a way that it can easily be ported to other domains as well. The intermediate result of the system is evaluated for a corpus of natural language queries collected from casual users who are not involved in the system design. The domain ontology has the minimum level of data extraction tendency which is enhanced in this research work.

Deshpande et al., (2012) proposed “Natural Language Processing using probabilistic context-free grammar”, authors discussed a method to create a new NLDBI system using Probabilistic Context-Free Grammar (PCFG). The study highlights Natural language statement which is converted into an internal representation based on the syntactic and semantic knowledge of the natural language. This representation is then converted into queries using a representation converter, but the optimization factor is missing in finding the right grammar which is handled in this research work.

Tamrakar et al., (2012) published an article titled “Query Optimization using Natural Language Processing” which proposed the architecture for translating user query into SQL using semantic grammar. LIFER/LADDER method is used in the syntax analysis. The LIFER/LADDER system could only support simple one table queries or multiple table queries with easy join conditions which restrict the system to a large extent.

An important issue proposed by Gage. M et al., (2012) is re-emerging in the field of relational database management systems which have the ability for non-expert users to access stored data using the more powerful aspects of the Structured Query Language (SQL). The widespread use of relational database management systems in the industry as well in scientific research has increased the need for a solution to this issue. The used method will
allow non-expert users more successfully obtain data with the use of an artificial intelligence application to process natural language from the user in the form of a question or sentence into a SQL statement. The author explores the foundations of this field as well as the branches of the more recent approaches including multilingual solutions, phrase recognition and substitution, SQL keyword mapping, and fuzzy logic applications. Some of the aspects are analyzed in the current work.

Nihalani, N et al., (2011) propose about the Structured Query Language (SQL) norms which are pursued in almost all languages for relational database systems. However, it is very difficult for researchers to write SQL queries because they are not aware of the structure of the database. So this leads to the development of Intelligent Database System (IDBS). There is an overwhelming need for non-expert users to query relational databases in their natural language instead of working with the values of the attributes. As a result, many intelligent natural language interfaces to databases have been developed, which provide flexible options for manipulating queries. The idea of using natural language instead of SQL has prompted the development of new type of processing called Natural language Interface to Database. NLIDB is a step towards the development of intelligent database systems (IDBS) to enhance the users in performing flexible querying in databases. The author emphasizes the overview of NLIDB which can be enhanced for the real-time user queries with complex ingredients.

Giordani et al., (2010) proposed “semantic mapping between natural language questions and SQL queries via syntactic pairing”, the author proposed an automatic translation of natural language query into SQL query using support vector machine algorithm and kernel functions. In this algorithm, a dataset of relational pairs containing syntax trees of questions and queries need to be encoded using kernel functions. During the encoding process, the key portion of the user queries are skipped and some parts of the used functionalities are ineffective at times to convert the user query in a critical situation which is addressed in the proposed research study.
Rao et al., (2010), proposed the architecture for translating natural language query into SQL using semantic grammar. Lexicon and post preprocessor are used in the semantic analysis. Post preprocessor transforms the semantic representations of the sentence into a SQL query. This system capable of handling simple queries with standard joined conditions but the flexibility to handle the complicated query is under a scanner which will be taken into account in the proposed work.

Karande et al., (2009) describes “natural language database interface for selection of data using grammar and parsing”, the authors proposed the NLDBI system considers the selection of data and performing primitive queries on the database and join operation with some constraints. ATN (Augmented Transition Network) parser is used for generating a parse tree. The query processing is initiated at the different levels. The interface has been restricted to process the query with the complicated grammar based content and sequence of the transition is awaited at the other end of the system which makes the whole system as unequipped.

An interface module (Saravjeet Kaur et al., 2012), converts user query given in natural language into a corresponding SQL command. Asking question to the database in a natural language like English is a very convenient and easy method to access data from database system especially for normal users who do not understand complicated database query languages. The design phase of the interface is quite handy for the new users; where the posting of queries is made reasonably easy. The complicated and user-defined queries are not examined. And the complete semantic conversion is not attained due to complex sentences as a query statement which is addressed in this research work.

Thus, the issues involved and the challenges faced in the first dimension of the literature review have been examined.
2.1.2 Recommendation systems

The second dimension of the discussion in the literature review targets an integral part of the research i.e. the recommendation system, its approaches, its practices and recent contributions. The various attempts made so far in the field of the recommendation system and its related area will be discussed in the following section. The recommendation system in social media and the challenges faced in the social media environment are also reviewed and discussed below.

Gurpreet Singh et al., (2015) briefs about recommendation systems, components of recommendation systems and various approaches of recommendation systems i.e. collaborative filtering approach, content-based recommendation approach, demographic recommendation approach, social network-based recommendation approach, hybrid recommendation approach and context-based recommendation approach. He also explains various application areas of recommendation systems. After examining the approach the issues such as cold start problem, sparsity can be addressed and recommendation systems quality can be improved which will be discussed in this research work.

Daoud et al., (2015) discuss the various recommendation techniques such as collaborative filtering technique, content-based technique, knowledge-based technique and demographic technique. In collaborative filtering technique systems takes the behavior, opinions of a large group of consumers into account, whereas the content-based technique depends on product features and textual item descriptions. In the Knowledge-based technique, recommendation offered is based on explicit knowledge models from the domain. The hybrid approach combines two or more techniques. Daoud et al., also discusses the association rule mining technique which concentrates on the mining of associations over sales data and helps shop managers to analyze past transaction data and to improve their future business decisions and recommend products to a consumer on the basis of
other consumers’ ratings. The techniques implemented in the study are completely analyzed in the proposed research work.

Pathan et al., (2015) emphasizes about the product recommendation in e-commerce environment where the product based sales data is collected from typical inventory. Such systems make use of statistics and data from user behavior e.g. Purchase history, product ratings. So, the decision to display a specific product from a specific category is taken after considering such parameters. In Hyper-Local based services (Locality Based) recommendation systems operate in a challenging environment. Such as, new customers have too much-limited information associated, less purchase history, no product ratings etc. Secondly a large retailer has too many categories to choose from users’ tends to have scattered data-less patterns. In order to handle such information mainly, three methods are available: search-based methods, collaborative filtering and cluster models. These methods are more suitable for a vast user base environment. Whereas, in small-scale environments, a set of customers whose purchased and rated products overlaps with a current user's purchased and rated products which create a chaos in many occasion. As the approach used by the author has a lot of limitations, therefore the result of the system is subjected to various constraints. The improved implementation in terms of recommendation in e-commerce environment is encountered in this research work.

Kiran Kumar et al., (2016) emphasize about the personalized recommendation system. The demographic locations of the users are tracked and the suitable response is offered to the users through GUI. The user profile is created and information of the users is collected. The location of the user is identified with the help of GPS. The recommendation scenario is designed based on the user interest. The reliability of the recommendation cannot be predicted since tracking the location of the user is a quite a challenging task. Thus, the current research study examines the possibility of suggesting the user in the most accurate and appropriate way.
Chen et al., (2013) presents the preference-based system to enable the e-commerce consumers to access needy products in an optimal way. The recommendations to the users are provided by creating the clusters of favorite items in the product list. The purpose of the cluster is to segregate the content based on the demand and cost. The cluster-based analysis is carried out and the volume of the dataset used must be defined to derive the well-defined clusters. The factors involved in finding the preferences of the users are defined with the system. The user profile database is gradually processed to get the user data. The technique used in this study is examined and enhanced in the current research work.

Kabore et al., (2012) has designed and implemented recommendation system as a module for a portal in which the basic approaches of recommendation system are revisited and appropriate recommendation is offered to the users who are seeking a job. User profile data is collected from the web through API. Two major problems taken into account are (i) the implementation of the recommender engine using the Apache Mahout library, and (ii) the integration of the recommender in a portal. Prior to the design of the application, decisions are taken at the preliminary stage which shows that the analysis of the requirements for the recommendation system is not carried out properly. The decisions made at the design level are explained and the risks involved are analyzed and mitigated. The approach and analysis used in the study are observed and fine-tuning is carried out in this research work.

Esparza et al., (2012) analyzed a real-time web data to develop an online recommendation system for product buying fashion in online. The web data is extracted using a web crawler and extracted data is organized in an appropriate form. The product information is derived in a random state and processed according to the demand of the system. The RS offers the suggestions in a most predictive way based on the user history and user navigation in the web. The weblog data provides the necessary information about the user behavior on the web and pulls the product-related data for the system implementation. But, the information collected from the weblog data
has junk values and incorrect piece of content, which affects the performance overall system which is addressed in the current research work.

Xiang et al., (2012) confers about the recommendation system practices in a detailed manner. The approaches and techniques involved in the recommendation system are clearly examined and proper direction is provided for the researchers to pursue the further enhancement in this domain. The key part of the recommendation system depends on the dataset. The dataset should offer the needy attributes for finding the right suggestion for the users. The approaches such as content-based filtering technique, collaborative technique etc, will be used only when the dataset is tailor-made for the different types of testing. The best practices of RS are referred and ample amount of dataset is used in this research work.

He.J et al., (2010) proposed a model of recommendation systems which can considerably improve performance by utilizing information in social networks. In particular, the author addresses the following challenging issues in building a social network-based recommendation system. i) Investigating the existence of similarity among friends in rating items using statistical analysis on a dataset crawled from a real online social network, Yelp.com. ii) To understand the role of social relationships in a social network-based recommendation system, a Bayesian network-based recommendation system (SNRS-BN) is constructed. iii) Developing a social network-based recommendation system (SNRS) which utilizes more information in social network, including user preferences, an item's likability and similarity effects among friends, to provide better recommendations. iv) To overcome the problems of heterogeneities in social networks, the author proposes to select relevant friends for inference based on the semantics in fine-grained user ratings on their buying decisions. An experiment is carried out in a graduate student class to validate performance improvements from such as semantic filtering. The result shows that the recommendation offered to the users suffers from the sparsity and cold start issue since the less amount of review dataset is used for testing. Ample amount of review dataset is used in the proposed research work.
Zhao et al., (2010) discuss the user based collaborative recommendation system in which the user data is stored in hadoop database. Since the huge volume of the unstructured dataset is used therefore the nature of the dataset is undefined and the format of the dataset is flexible enough to adhere to various constraints. The recommendation system which is implemented by the author depends on the user opinion from the past and the present. The consistency of the feedback is cross verified and necessary suggestions are offered. The straightforward constraint found in user-based collaborative RS is the inconsistencies of the user feedback are processed. This inconsistency is addressed in this research work.

Choi.J.Y et al., (2008) discusses a collective collaborative tagging (CCT) service architecture in which both service providers and individual users can merge folksonomy data (in the form of keyword tags) stored in different sources to build a larger, unified repository. The unified repository acts as a data pool from which requests information which can be recovered as per the user’s wish. But this system is restricted to address the similar kind of user query since it tags the content and expects for the relativity in terms of the query, whereas the proposed research framework is applicable for different collections of queries.

Leimstoll et al., (2007) describes about the integration of collaborative filtering approach to recommend the products for users in online shops. The product data is collected from the web inventory and all the existing approaches are applied. The better result is obtained using the collaborative technique. The most recommendation systems use the collaborative filtering method in order to provide the personalization information. The collaborative filtering method is a very efficient and convenient way of achieving personalization as there is no need to introduce semantic information about the products. The integrating user demand and options to select the right product are the challenging aspect which is lacking in this study which is implemented in the proposed framework.
The social media analysis carried out by Rabia Batool et al., (2013) who analyzed information from tweets. The system is tested on a collection of 40,000 tweets for finding semantic content. The knowledge enhancer and synonym binder module are applied to the extracted information which increases information gain in a range from 0.1% to 55%. The proper and meaningful tweets have been gathered. The collected tweets are clustered based on the purpose of the tweets and each cluster is isolated and the corresponding tweet analysis is done. In tweet clustering, the majority of the contents are squeezed in which the key part of the information is left as volatile. The tweet analysis has to be scrutinized which will be addressed in this research work.

Vo Thi Ngoc et al., (2012) uses the technique to evaluate the faculty performance by signed and unsigned student feedback using a regression technique. It shows the significance of the student feedback in making certain decisions in learning environment which are addressed in this research work. Apriori algorithm is implemented in which the student and placement based data are mined. The outcome of this study provides a suitable analysis of student performance and placement possibilities. It is evident that the various factors are involved in determining the overall performance of a student. But the implementation of the system is incomplete since very minimal amount of data set is used in the experimental study.

Brady et al., (2010) emphasizes the significance of social network in distance learning environment. Tutor has a choice of picking the right interface to transfer the knowledge through social networks like Facebook, Myspace, and so on. Different forms are used to collect the data of a student to analyze a suitable mode of distance learning in terms of analyzing the career guidance of a student who completes education in open or distance learning environment. The data extraction process needs the revision in collecting the data in bulk rather than squeezing individual data.
Fidalgo et al., (2013) investigate the enhancement in learning through online or distance mode by analyzing a traditional online teaching and social network forum. The level of interaction between the tutor and learner are measured and analyzed in terms of frequency in which the interaction occurs. Knowledge sharing among various web users emphasizes the need of proposed framework which is discussed in this research work.

Walnuj et al., (2013) intend to offer a recommendation system in apache framework. The basic requirement to build a system on a specific framework is explained for the e-commerce domain. The recommendation framework is designed and apache platform is used to program the requirements of the e-commerce online system. The product information and user information are two important entities which are used by the framework design to implement the system. The framework is applicable only for the system where a limited amount of dataset is in practice. This limitation is encountered in the current research work.

Kuo et al., (2014) intends to investigate a degree to which interaction and other predictors contribute to student satisfaction in online learning settings. The effects of student background variables on predictors are explored. The results showed that learner-instructor interaction, learner-content interaction and Internet self-efficacy are good predictors of student’s satisfaction. Therefore, a better interaction has to be incorporated for positive e-learning environment.

DIPRO 2.0 is an educational social network for university professors to develop their training in the area of personal learning environments through collaborative learning and production of knowledge. In this, web 2.0 social network tool is used efficiently to extract the knowledge from various web sources. Members from various educational institutions interact through social network tool. The ideas and thoughts are shared among the various users. The level of knowledge transmitted through social network tool is in high-end since the collaborative environment is provided with the help of web
2.0. The knowledge sharing phenomena is also discussed in this framework (Verónica et al., 2014) which is quite handy for the proposed research work.

A study on collaborative learning system speculates the importance of online distance learning system where the students are provided with social interaction to share knowledge among the various group of people. Social interaction session among the various users is initiated by examining the challenges involved in a social network environment which is encountered in the article published by Muuro et al., (2014). The impact of social media and its data are pretty much required for the establishment of interaction which results in the knowledge sharing among the different peers. The collaborative collection of social media data is filtered and processed. The level of information collected from social media has to be revised since the volumes of the data used for testing are not enormous.

Jianming He et al.,(2010) proposes a paradigm of recommendation systems which can make use of information in social networks, as well as user preference, item general approval and influence from social links. A probabilistic model is constructed to make personalized recommendations from such information. Data from social networks is extracted and analysed. Experimental results show that the system not only improves the prediction accuracy of recommendation systems but also rectifies the data sparsity and cold-start issues. To improve the performance of the system social media data is validated for consistency which is implemented in this research work.

Khoshnood et al., (2012), describe the personalization factor in recommendation system, the information is provided to the user in a way that personalizes content in a suitable frame considering individual’s location. The idea of using social recommendation systems is to identify the user’s interests and preferences based on the user’s current place; it also offers suggestions to the end user. Users information includes personal information and interests in social network sites. Through this data, users interaction decreases and can acquire their favorite information and services. The experimental result shows that the personalization factor needs to be
addressed. The improvisation and modification are carried out in this research work.

Fijalkowski et al., (2011) discuss the data obtained from social network profiles of its users. The architecture modeling approach is developed within the project of a mashup Web application that integrates with Facebook API. The user profile and information available in the forum are taken into account for prediction, where the limited numbers of users are in the position to share their personal information, therefore the prediction system can be applied only with the minimal test cases. The efficiency of the system cannot be measured with the limited dataset which is addressed in this research work.

Rahman.M (2012) presented a systematic data mining architecture to mine intellectual knowledge from the social media source. Here, the author used the social networking site Facebook as a primary data source. He collected different attributes such as 'about me', 'comments', 'wall post' and 'age' from Facebook as raw data and used advanced data mining approach with the help of Weka tool to excavate intellectual knowledge. The extracted knowledge is used to predict human behavior, pattern recognition, decision making and product promoting; but the fact that the volume of dataset used for testing is minimal.

A computational method for actionable knowledge extraction from online media is implemented. The used approach is based on mutual bootstrapping and combined with knowledge reasoning. The approach used acquires more types of action knowledge, and needs much less human labor. However, knowledge extraction through conventional method is time-consuming which is addressed in this research (Ansheng Ge, et al., 2013).

Danyllo et al., (2013) discusses the method of collected data from the social network Twitter and compared them with data from a financial institution in order to model the network and analyze the similarities. The result reveals that the most of the users have more credit restrictions than neighbors, and users with no restrictions normally have neighborhoods with
no credit restriction as well. Here, social network analysis is done with reliable metrics on the Twitter database. However, the knowledge extraction in social network differs in large extent which is highlighted in this research work.

Liu et al., (2013) proposed a system which offers the suitable and accurate recommendation system which takes up the user preference and user opinion. In the ecommerce environment, the product buying and selling is basically depended on the user preference and opinion which can be directly controlled by the quality of the product. The set of product list with its attribute is collected and appropriate testing has been implemented for a suitable suggestion for the users. The result of the system suggests that, the volume of dataset used in the testing phase is quite less which is taken care of in the current research work.

Ojokoh et al., (2013) focuses on mining the opinions expressed on some electronic products, providing ranks or ratings for the features, with the aim of summarizing them and making recommendations to potential customers for better online shopping; but the fact that the ranking and recommendation are offered only for the single source of data and performance of the system is not measured which is encountered in this research.

Sankar et al., (2014) discussed about the data extraction in social media environment from a financial institution in order to model the network and analyze their similarities. The result of the comparative study provides the model which has too many routing channels to reach the endpoint; in other words the complicated model is developed with the less emphasize on the overall result. The method used to extract the dataset from Twitter is lasts long only for minimal period of time therefore again the volume dataset used for testing is not sufficient.
A novel Ontology-based Sentiment Analysis Process for Social Media content (OSAPS) with negative sentiments is presented. The social media content is automatically extracted from the Twitter messages (Pratik Thakor et al., 2015). The learning collaborations and communications with subject experts are highly practiced in corporate sectors by maintaining a knowledge portal for the stakeholders. The solution offered is shaped up for the various users where the requirements are dynamic in nature where the results are static; therefore the nature of the system must be addressed.

Li et al., (2015) discusses the reviews and synthesizes extant empirical studies to provide a coherent view of research on RS and identify gaps and future directions. The approaches involved in the recommendation system are analyzed with the real time examples. Understanding the need of the consumer and providing the required recommendation will be the core concept behind RS which is incorporated in this research work.

Dhawan et al., (2017) enumerates on the importance of the collaborative filtering technique in the product recommendation in social network. The opinions of the various users are collected and trend forum is created based on the topic selected among the active users in the social media. These active users register the post and comments about the product features in the social media which are summarized for opinion mining. The ratings of the products are compared and appropriate recommendation is offered. The actual nature of the collaborative filtering approach is tested and major part of the technique is implemented in this study.

Sridevi.M et al., (2016) discusses Personalized RS in e-commerce environment which includes two aspects, (1) the input of recommendation systems, such as the acquisition and presentation of customers’ interest profile as well as item profiles; (2) the typical methods of various recommendation techniques; the customer profile is collected through conventional feedback system and item profile is static collection of items with the minimal features. The personalized recommendation is provided with the limited entities. The available approaches such collaborative filtering
technique, content-based filtering techniques along hybrid techniques is used to user input data such as user demographics data, rating data, transaction data and production data. Different approaches operated with various concrete algorithms are found for effective recommendation (Example taken is to recommend Items to shopped among the e-commerce users). User input data is minimum, much user personalization RS has to be designed and cold start problem, sparsity needs to be addressed in this paper. The basic functionalities of recommendation system are enriched in this research work.

Asanov et al., (2015) propose a various traditional and modern approaches of RS. There are different algorithms of measuring similarities among items in database and those in users’s profile is identified and one of such approaches is cosine similarity. Context-aware approach is used and OWL (Ontology Web Language) is found for effective recommendation (Example is taken to recommend a most favorite movie among larger users). RS is expanded from a nascent level to most advanced level using modern techniques and algorithms, but the issues such as cold start problem, sparsity, scalability and privacy are not addressed by the author.

Thus, the existing work and strategies are discussed in detail in this section which provides the sufficient reason to expand and enable the current work. On the whole, all the existing work provides the complicated and different types of frameworks which are not suitable for many cases which holds up the reason for the extension
3.0 PERSONALIZED AND OPTIMAL RANKING SYSTEM (PORS) FRAMEWORK

3.1 INTRODUCTION

In this chapter, a construction of a personalized, optimal ranking system (PORS) framework and query processor component are discussed in detail. The construction of the framework is initialized after the detailed study on the process involved in the system. The various implied components in the framework are defined as per the demand of the system. The main objective of the framework is to offer online recommendation for the users. The designed PORS framework also offers the optimal ranking system for the users. The detailed information about the working components of the framework is elucidated in the following section.

3.2 DESIGN OF PORS FRAMEWORK.

The designed framework for recommendation system possesses the various constraints during the development of the personalized recommendation system in a social media environment. The design emphasizes the basic need of a recommendation system in a dynamic environment where the data is varying constantly.

The figure 3.1 depicts the proposed framework for Personalized and Optimal Ranking System (PORS) for Recommendation in heterogeneous Social Media Environment. This framework has the different phases of components which are explained in the subsequent sections. The PORS framework is applicable for both product based recommendation and service-based recommendation. The framework is flexible enough to collect the reviews and features from social media such as Facebook, Twitter and other ecommerce sites. Recommendation procedures are followed for both selection of product in the online and choosing the career guidance service information in the online.
The components of the framework are very broadly classified into four as follows,

- Query processor
- Data Extractor
- Review & Features Analysis
- Recommendation with ranking component

Figure 3.1: Framework for Personalized and Optimal Ranking System for Recommendation in Heterogeneous Social Media Environment

In the figure 3.1, the first phase depicts the query processor component in which the end user posts the query. The second phase depicts the data extractor component where the social media data is extracted and stored. Review and feature analysis component is represented in the third phase where the detailed analyses are carried out for the recommendation process. The fourth component of the figure 3.1 specifies the recommendation with ranking component where the tasks involved in offering the recommendation along with ranking is projected.
**Query Processor** is a component in which the user query is processed through the GUI from which the query is semantically transformed into interoperable format. The transformation of user query into the interoperable format is carried out using revised keyword extraction algorithm. The outcome of the query processor brings the interoperable format of the given user query.

**Data Extractor** component is the platform from which the social media data source such as Facebook, Twitter and E-commerce web source are explored to fetch the reviews & features. The extracted data is organized using a support vector machine in which review data are classified into the positive review, negative review and mixed review. The features are fetched using RAKE technique from which the review texts are processed step by step to extract the features of the corresponding query. Further, the extracted features and reviews are clustered based on the similarity. In the end, needy information is organized in a structured way and stored in the repository which consists of huge volume of organized social media data.

**Review & Features Analysis** component meant for analyzing the data which are extracted from the previous component. As a part of the process, the reviews and features collected from end users are compared with the data available in the stored repository using index matching algorithm. The reviews are analyzed using sentiment analysis to provide the recommendation for the user requested query.

**Recommendation with ranking component** ensures the personalization and ranking factor, the features of the other available items in the repository are listed and cross verified. The products and other service information which are related to the query are suggested for recommendation. The ranking of the other products and other service information take place by considering its reviews and features which match the features of the requested product by the user. To achieve this, the revised ranking algorithm and linear regression analysis are used which offers ranking. To authenticate the ranking and recommendation, user
personalized information such as user interest, preferences, ability are considered and personalized recommendation with ranking is offered to the requested user.

Thus, the basic objective of the proposed PORS framework is streamlined in this section. The PORS framework proves to be the generic one since the component design is quite flexible which can be implied to any domain. The following section discusses the query processor.

3.3 QUERY PROCESSOR
3.3.1 Introduction

The PORS framework satisfies the need of user and user preferences. The primary target audience of the system is the online web users or the system end user, therefore comforting them is a crucial part of the system. The user satisfaction is the primary goal of this framework. The users are different in nature; the complexity in the system in terms of access may direct the users to select some other system with flexibility. In this context, the user requirement and expectation need to be processed. In this research work, the primary role of the end user is to provide the user query to the system. The system processes the query and corresponding response is given back to the requested user. The requested query is analyzed based on the types of query, syntactic and semantic nature of the query. So, many researchers have done the ample amount of work on this subject to enhance the user with the suitable interface to meet their basic requirements. In this process, the user query from the user is received as plain text, series of words, continuous statement or sometimes a typical keyword. The challenges for the researchers in this area of study are (i) how user query is processed in the system, (ii) how the system is going to manipulate the query, (iii) what technique or methodology will be owned by the system to tackle this scenario. The detailed research has been carried out with the aim of processing the user query which makes both the system and user in the comfort zone.
3.3.2 User Query Transformation

Enormous evolution of social web data creates a peculiar myth in the field of computer and information technology for extracting the meaningful content from the web. The user query may be in any type such as structured or unstructured type. The query is in the form of natural lingual statement which has to be processed. The basic idea is to find the suitable way to convert natural language query to SQL and access the data in the database. The query is converted into SQL query with the support of query processor.

Natural Language Processing (NLP) is becoming one of the most active techniques used in Human-computer Interaction which plays a vital role since the social media is started playing its part to a large extent in the current trend. In the context of social media, the query conversion is quite important in terms of bringing out the exact data which is requested by the web users who surf the net. The query or request will be of natural statements such as blog, comment, tweets and these statements must be converted into a most reliable and acceptable form of typical query which system can understand and crack the exact data from the database. The objective of NLP is to facilitate communication between human and computers without memorization of multifaceted instructions and procedures.

NLP to SQL Transformation:

Figure 3.2: Block diagram for NLP to SQL
The user query which is triggered to the social media which requests for the specific information where the information is stored in the database. The typical social media, Facebook is examined in which the simple query is processed. For example, the query statement “What is name and comments of Facebook who likes Flipkart” and comment of a Facebook whose likes “Flipkart” is tracked then formed into a SQL query “Select comments from Facebook where Flipkart="like"; For a novice user it is not possible to form SQL query so using the system he/she can simply post a question like “What is name and comments of Facebook who likes Flipkart?”.

In the figure 3.2, the block diagram for converting the natural language into SQL is determined. A user gives an NLP sentence as input with the help NLP query converter the sentences are tokenized. The tokenizer splits the sentences into word based on whitespace character. The tokenized words are taken to extractor for stemming process. In stemming process, tokenized words are identified for extracting the documents from the web sources. In the semantic analysis, the identified set of words will be given as input. The parse tree is generated through parser and subject in which the set of words are identified for mapping. Keywords are mapped to suitable SQL query which is built already in the database. Each SQL query type is identified with the help of attribute selection. The mapped SQL query helps to retrieve the information from the repository.
NLP to SQL converter using N2S algorithm:
The step by step process of transformation from natural language query to SQL is deliberated using N2S algorithm. The steps involved in N2S algorithm are as follows,

Step1: Process Query
- Divide Query in tokens.
- Remove punctuation marks.
- Do initializations.

Step2: Generate Intermediate Analysis for query formation
- Divide Query into parts using criteria words.
- Identify column attributes and table names from user Query and remove unwanted words.
- Replace synonyms of column attributes and table names in Query with its actual names.
- Arrange parts in proper sequence.

Step 3: Formation of SQL Query
- Take intermediate Query as input
- Identify / derive 3 things from Query
  - Select keyword: - These are attributes which user wants to retrieve.
  - From keyword: - This is the table name from which user want to retrieve data.
  - Where keyword: - This is condition specified in query.

Step 4: Replace select keyword with actual table attributes.
- If there is only one from keyword then replace it with actual table name. else form following sequence table name1 JOIN table name2 ON attribute1(primary key of table1) = attribute2(attribute in table2 which is foreign key of table1)
- Replace where keywords attribute with actual table attribute and concatenate “=” following with value specified by user. Form standard template of SELECT Query and substitute above keywords i.e. select keyword, from keyword, and where keyword in their appropriate place.
The process of converting the NLP to SQL is implemented as a part of query analysis where the information from social media is analyzed. The implemented query conversion is projected in the figure 3.3. The extension of this query conversion process is applied for the different ranges of database access. The figure 3.4 illustrates the query conversion process for service recommendation in which user query is transformed into SQL; here the user is giving NLP query to fetch the placement details of the student. The NLP statement is processed step by step and the corresponding SQL is generated appropriately. The primary part of the query is converted into typical SQL to fetch the required information from the database. The query analysis reflects the usage of the user query and conversion of user query into SQL which has its own limits and delimits. The implementation of the process suggests that the user query is narrow down to the SQL which is not the case of information available on the social media which will be of different format.
Thus, the first phase of user query analysis concludes with the notifying improvement in processing the user query to a meaningful SQL query from which the relevant and required data are fetched. The need of the system is constantly changing since the system deals with heterogeneous social media web source. The social media data with more than one source is used as the resource pool. Therefore the improvement and enhancement are required in the process the user query in a lot more better way. The following section discusses the semantic query analyzer which concentrates on the improvement and enhancement in processing the user query.

3.3.3 Semantic Query Analyzer

As discussed in the previous section, the user query is the instrumental part of the recommendation system, but converting the user query into SQL which has its own limits and delimits. Moving forward with the social media environment where the enormous volume of data is dealt in which the user query needs to be handled in an effective way. This section talks about improvement and enhancement required to process the user query in a best possible way.
The requested query from a user is channelized appropriately and the query statement is transformed into the interoperable format. The query is converted into interoperable one by applying the revised keyword extraction algorithm (RKEA). The query which converts the user query into keyword goes through the series of the process such as parsing, stemming and keyword constraint.

The user query is collected and processed where the noisy part of the query is cleared. The stop words (such as is, am, to, as,) are removed from the query and remaining part of query undergoes stemming process where similar words of the family (“use” is the stem of the user, using, usage etc.) are merged. After the stop word removal and stemming, the query is processed using constraint which converts the query into the interoperable format. The parameter passes to find the similarity in the query which is left after the parsing and stemming, if the query matches these criteria then the keyword has been successfully extracted if the condition does not occur when it should start from the first iteration. Thus, the RKEA method is used to transform the user query into the interoperable format.

The Revised Keyword extraction algorithm for transforming the user query into the interoperable format is discussed below in which the user query is processed step by step. In each traversal, the key part of the user query is processed and required element of the query is collected for further process. The extraction of a keyword from the user the query is achieved after screening an integral part of the user query where the unwanted contents are removed and remaining part of the query are retrieved appropriately.

The retrieved part of the user query is examined with the keyword comparison constraint which verifies the collected content with the constraint whether the query is completely transformed into the interoperable format. The result of the keyword comparison constraint will fetch the exact transformed interoperable content. The transformed interoperable content is suitable to access the social media data repository where the vast amount of
heterogeneous collection of the information is available from which the needy information is retrieved for the recommendation process.

**Revised Keyword Extraction Algorithm:**

*Keyword Extractor (Query): User query to Interoperable one*

*Input: Query (Qi)*

*Output: (Keyword (Ki))*

**Start**

**Step 1:** Collect Input Query (Qi) where i=1, 2, 3….n;

**Step 2:** for each input Qi;

Extract Query (EQi) = Qi; //extract word process for all

For i=1, 2, 3…n in //

**Step 3:** for each EQi;

Stop Word (SQi) = EQi; // apply Stop word elimination process //

Stemming (Si) = SQi; // It create stems of each word //

**Step 4:** for each Ki;

Keyword_Comparisor (KCi) = Ki;

if

{Ki=Qi-(EQinSQinski);

return Ki ;
}

else

{
Get Query ()
}

**End**
3.3.4 Result Analysis

The experimental study conducted in view to convert the user query into the interoperable format is applied to the domain where the user posts the query on searching the products in online mode and user posts the query in search for the career guidance information. Firstly the search of the information basically initiated through the typical user query on the specific or irrelevant domain. The user query is basically driven by the user to search the product in the online for the relevant and reliable recommendation, therefore in this context; the user query is categorized into four different types such as

- Search by name.
- Search by keyword
- Search by features.
- Search by incorrect & random text.

Based on these categories, the query sampling is done and each category of the query is addressed and set of the possible output for each query is trained and tested. The results show that conversion is quite handy and needy information can be retrieved with ease when compared to the other techniques. The samplings of the queries for the entire four categories are shown in table 3.1 where the collection of queries of the entire category are trained and tested. The results are tabulated appropriately.

<table>
<thead>
<tr>
<th>Table 3.1: Different Category of User Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category of User Query</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Search by name</td>
</tr>
<tr>
<td>Should I buy the Asus Zenfone or the Lenovo K6 Power?</td>
</tr>
<tr>
<td>Show the placement probability of rocer</td>
</tr>
</tbody>
</table>
As shown in table 3.1 the numerous range of user queries are collected and the corresponding transformation is carried out. The ability to convert the query is tested with various levels of queries for the range of different domains. The results obtained are manipulated appropriately and the selected domain for the training and testing the data is mobile products. The query floats from a different direction and corresponding techniques are implicated to produce the required result. The following screenshot shown in figure 3.5 reflects the system driven query to interoperable format.
The user query is preprocessed and the resultant part of the query is tokenized, parsed and stemmed. The screened part of the query is further bisected and the result is displayed. The following screenshot shown in figure 3.6 reflects the query conversion after the series of the process involved.
As shown in the above screenshots, the sampling of user queries is trained and tested. In which more than 4800 user queries are tested and query conversion rate is recorded. The success rate of query conversion is also calculated. It is found that the probability of successful query conversion into keyword is more sensible when the search begins with a specific keyword. Based on the result achieved the corresponding category is catered for the reliable query conversion. Table 3.2 shows the results obtained from the trained and tested user query dataset for the entire four categories of the user query.

**Table 3.2: Query processor Analysis with Sampling Queries**

<table>
<thead>
<tr>
<th>Types of Query</th>
<th>Query Posted</th>
<th>Query Processed</th>
<th>keyword extracted</th>
<th>Query Conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search by Name</td>
<td>1200</td>
<td>1200</td>
<td>1115</td>
<td>92%</td>
</tr>
<tr>
<td>Search by Keyword</td>
<td>1200</td>
<td>1200</td>
<td>1200</td>
<td>100%</td>
</tr>
<tr>
<td>Search by Features</td>
<td>1200</td>
<td>1200</td>
<td>950</td>
<td>79%</td>
</tr>
<tr>
<td>Search by Random &amp; Incorrect</td>
<td>1200</td>
<td>1200</td>
<td>600</td>
<td>50%</td>
</tr>
</tbody>
</table>

Thus, the trained and tested sets of user queries for all the possible categories are implemented. The result shows that the query conversion into the interoperable format is more prominent for the search which is done by the exact keyword. The result of the trained and tested user queries for the other categories are monitored and recorded. The possible chance to address the unstructured database where the social media data which plays a key role where the different formats of data is organized and the user might post any kind of query, but the system designed to post the query on a specific domain. Thus, as the part of the feasibility study, the categories are classified and required strategies which are applied to process the query which belongs to the certain domain. Thus, in the query processor phase has been carried out effectively for the reliable recommendation system.
3.5 CONCLUSION

The outcomes of this chapter are the construction of PORS framework and the need of the query processor. The framework design is the functional attribute of the recommendation system where it specifies the actual structure and flow in which the process needs to be executed. On another hand, the query processor is the first component of the framework which administers the all-important part of the system. In the next chapter, the establishment of a data repository from the heterogeneous social media network is discussed.
4.0 DATA ACQUISITION FROM HETEROGENEOUS SOCIAL MEDIA

4.1 INTRODUCTION

This chapter deals with the data acquisition from heterogeneous social web sources. Social media plays a major role in the recommendation system where huge volumes of interactive information are processed. The data acquisition phase selects web sources in the social media such as Facebook, Twitter and e-commerce. It also focuses on the data analysis where the user query and the subsequent information from social media data are retrieved and analyzed appropriately.

The data acquisition from heterogeneous social media web sources includes the data classifier phase in which the reviews and features are extracted step by step. The acquired data is classified and warehoused appropriately. The information acquired is transformed into interoperable one where the information is organized and stored in the best possible way for access. Firstly, the data acquired from heterogeneous social media for recommending the online electronic products are discussed in the data classifier phase. Secondly, the data acquisition for the career guidance recommendation is discussed in the data acquirement phase.

4.2 DATA CLASSIFIER

The inclusion of social media in online recommendation system provides the new dimension to the overall outcome of the system since the social media deals with the more interactive and dynamic content on any domain. The data acquisition process from social media is explained through workflow which is shown in figure 4.1.
The retrieval of information from the social media and making that information accessible as per the need of the system is a challenging task. The following are the phases involved in the data acquisition process,

- Data Collector
- Review Classifier
- Feature Extractor
- Cluster & Store

### 4.2.1 Data Collector

In the Data Collector component, the social media networks such as Facebook, Twitter and other E-commerce sites like Amazon and Flipkart are selected from which the reviews and features of the different electronic products are extracted. The collected reviews and features which take up the subsequent parameter to crawl the web content and convert into corresponding text. The web content consists of tweets, posts, and comments etc; which possess noisy content such as unconventional symbol and incorrect text which are cleaned, processed and organized appropriately.
The information from the social network is retrieved which will be of raw state. The information is organized into a suitable format using review parameter. The following screenshots show the acquisition of data from the social media, the domain selected to retrieve the information is an electronic product from which the various reviews of the mobile products are collected.

**Figure 4.2: Static Review Extractor Using Import io**

Figure 4.2 shows the static review extractor from which the reviews of the product are fetched appropriately. The dynamic extractor Walmart API is used to fetch the dynamic content from the various social media networks. The selected social media for the review extraction are Facebook, Twitter and E-commerce sites like Amazon and Flipkart. The parameter is passed and corresponding reviews are fetched. The data acquisition process is continued until the data turns into an interoperable format which will be more suitable for access.

The attempt of extracting the data from the social media brings the quite a number of possible product reviews which is shown in figure 4.3. These extracted reviews are a simple textual collection of data which needs the logical interpretation. These reviews are cleaned, bisected and processed in the best possible way. The reviews of the different products need to be
targeted to provide the number of options to the user to select and rate the products. The problem of cold start is resolved since the researcher collects the good number of data which can uphold the purpose of the recommendation system.

4.2.2 Review Classifier

As emphasized in the previous section, the raw review needs to be organized. The Review Classifier classifies the collected reviews using support vector machine classification technique. SVM classification is used to bisect the data into two dimensions with a valid vector parameter which defines the kind of reviews to be segregated. The classified reviews are
analyzed based on the polarity of the words like good, bad, critical, and most helpful. Based on the polarity, the number of positive reviews, negative reviews and mixed reviews are identified. The reviews are classified into positive reviews, negative reviews and mixed reviews. The sample reviews have been collected for the various products and it is classified using SVM classification technique shown in table 4.1

Table 4.1: Classified Review Dataset Using SVM Classification

<table>
<thead>
<tr>
<th>Types of Products</th>
<th>Total No. of Reviews Extracted</th>
<th>Positive Reviews</th>
<th>Negative Reviews</th>
<th>Mixed Reviews</th>
<th>T.P</th>
<th>T.N</th>
<th>F.P</th>
<th>F.N</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>150</td>
<td>60</td>
<td>50</td>
<td>40</td>
<td>60</td>
<td>50</td>
<td>70</td>
<td>55</td>
</tr>
<tr>
<td>P2</td>
<td>120</td>
<td>35</td>
<td>45</td>
<td>40</td>
<td>35</td>
<td>45</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>P3</td>
<td>110</td>
<td>40</td>
<td>35</td>
<td>25</td>
<td>40</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>P4</td>
<td>140</td>
<td>60</td>
<td>70</td>
<td>10</td>
<td>60</td>
<td>70</td>
<td>60</td>
<td>48</td>
</tr>
<tr>
<td>P5</td>
<td>130</td>
<td>60</td>
<td>50</td>
<td>20</td>
<td>60</td>
<td>50</td>
<td>45</td>
<td>40</td>
</tr>
</tbody>
</table>

The table 4.1 showcases the classified reviews based on the following assumption,

- True positive (T.P) - correctly classified as the class of interest.
- True negative (T.N) - correctly classified as not the class of interest.
- False positive (F.P) - incorrectly classified as the class of interest.
- False negative (F.N) - incorrectly classified as not the class of interest.

Therefore, using SVM the review data is organized into true positive, true negative, false positive and false negative. This sampling of review data brings the optimal availability of information in the data repository. The accuracy of the classified review data is measured using precision and recall technique which addresses the accuracy ratio of the data retrieved. The
reviews of the five different products are cross-examined using SVM classification and accuracy of the collected data is the measured precision value of 47% and recall value 59%. The precision and recall measurement of the sample reviews are listed in figure 4.4. The accuracy measure of the dataset is manipulated; the reference data set more than 1650 reviews are measured. The corresponding results obtained will bring the optimal range of review data for the system requirement.

**Figure 4.4: Precision & Recall Measure for Extracted Reviews**

Thus, the review classifier component classifies the reviews. To enrich the recommendation system, the reviews alone are not sufficient therefore the attempt has been initiated to extract the features of the various products. The extraction of features through the review is briefly discussed in the next section.
4.2.3 Feature Extractor

In Feature Extractor component Revised RAKE (Revised Rapid Automatic keyword extraction) technique is used, once reviews are classified into various categories, the features of the corresponding products are retrieved. The feature extraction is carried out by screening the collected reviews individually. Figure 4.5 shows the workflow diagram of the extraction of the features from the reviews. The process starts with collecting the textual review data and this textual data undergoes the typical text delimitation and stemming. After the initial preprocessing, the review data is filtered and screened. Next, the filtered and screened review data is examined to compute the word score metric.

![Workflow diagram for Feature Extractor](image)

**Figure 4.5: Work-flow diagram for Transformation of Feature from Review**

By following the procedure the reviews are transformed into features. The word score metric is calculated for the screened and filtered review text, after computing the word score metric, the highest word score candidate is extracted at the end. The selected highest word score candidate is taken and it is compared with the corpus collection of a feature-based keyword which is maintained, where a huge volume of feature-based dynamic keywords are
available. The result of the comparative analysis provides the concrete evidence that the collected keyword can be tagged as a feature of the corresponding product. If the result of the comparative analysis doesn't produce the appealing output then the information which is not tagged as a feature.
### Revised RAKE Algorithm: Transforming reviews to features

**Input:** Reviews  
**Output:** Features

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1:</td>
<td>Split the document into an array of words, breaking it at word delimiters (like spaces and punctuation).</td>
</tr>
</tbody>
</table>
| Step 2: | Split the words into sequences of contiguous words, breaking each sequence at a stop word. Each sequence is now a “candidate keyword”.
| Step 3: | Calculate the “score” of each individual word in the list of candidate keywords. This is calculated using the metric: degree (word) / frequency (word). |
| Step 4: | For each candidate keyword, add the word scores of its constituent words to find the candidate keyword score. |
| Step 5: | Find the Highest word score in the list and store it for next level manipulation. |
| Step 6: | Create a Corpus collection repository-Pool of feature-based keywords is used for comparison. |
| Step 7: | Compare & store the matched keyword with the word score. |
| Step 8: | Re-Iterate the process till the last keyword. |

The highest matched words score content matches the data in the corpus collection then the tagged as feature keyword and it stores intact for the further process. The steps involved in the execution of the revised rake technique is explained via the following procedure, the step by step procedure clarify the possibility of extracting the features from the reviews. The word score calculation for the sample review “Samsung has good battery backup” is displayed in figure 4.6
After calculating the word score, the highest word score candidate is selected. In this case from the review “good battery backup” has the highest word score value with 12, which will be selected and this candidate is compared with the corpus collection of the feature-based keyword. The comparison starts with the index element of the data repository that matched the keyword which will give the status of the feature. In the example, each word is processed individually from which the keyword “battery” and “backup” are tagged as features. All the reviews are individually manipulated with the word score. More than 5000 review samples are taken into account and the word score calculation is carried out individually and the highest word score is identified and the corresponding comparative analysis is conducted for the highest word score candidate. The samples of the word score calculation for the collected reviews are shown in figure 4.7
Figure 4.7: Word Score for the Collected Reviews via RAKE

The highest word score component is measured for each and every review separately and subsequent comparative analysis are conducted. The result of the revised rake technique fulfils the actual objective of the system where both reviews and features of the various products are collected.

4.2.4 Cluster & Store

In Cluster and Store Component, the collected review and features are successfully stored. After the classification, reviews and features are clustered based on the brand name of the product. The feature data along with the classified reviews are clustered using BNP k-means clustering technique. The features are clustered based on the brand name of the product where the brand names are bubbles out of each cluster. The purpose of using the BNP k-means clustering technique is easy to interpret the clustering results, clusters will be of uniform size and clusters have a similar density which will be quite easy to calculate the variance of each cluster. The brand name based k-means clustering algorithm is explained in the following
Brand Name based Product K-means Clustering technique:

*Input:* reviews & features;

*Output:* clusters of product with reviews & features

**Start**

Given $k$, the BNP-k-means algorithm is implemented in four steps:

<table>
<thead>
<tr>
<th>Step 1:</th>
<th>Choose $k$ data points (product reviews) to be the initial centroids, cluster centres (product brand name) // compares all the attributes by visiting the whole data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2:</td>
<td>Assign each data point ($X$) to the closest centroid ($D$) (next attribute (product feature) is targeted as next cluster) // Finding the similarity in attributes which most likely matches the corresponding data for clustering</td>
</tr>
<tr>
<td>Step 3:</td>
<td>Re-compute the centroids using the current cluster memberships (until the similar type of attribute encountered).</td>
</tr>
<tr>
<td>Step 4:</td>
<td>If a convergence criterion is not met, go to 2.</td>
</tr>
</tbody>
</table>

**End**

![Figure 4.8: Clustering Pattern - Product Features & Brand](image-url)
The result of the clustered pattern of the product features and the brand name is displayed in figure 4.8.

In each cluster corresponding reviews, feature data are well organized for the further accessibility. The clustering is enhanced further to group the feature data based on the various constraints. The range is fixed based on the cost of the product. Therefore different clusters are discovered for various cost range. The clusters of BNP-K mean clustering technique for the various brands of the product is collected. The result of the k-means clustering technique provides the data repository with the sequential levels of data in an optimized and organized way.

4.3 DATA ACQUIREMENT

To process the recommendation for any other services such as career guidance, the professional data needs to be extracted. The data is collected from various web sources like Facebook and Twitter. The dataset collected for the recommendation is from the social media Facebook where the academic details of the learners are extracted from the facebook group. Data are collected from the social web community BSAU professional group which is developed and maintained to process the student professional data for determining career guidance and placement opportunities. BSAU professional group is an exclusive and private forum in which the students enroll themselves and post their professional data where various people communicate to share their valuable thoughts about career and placement opportunities. Therefore the data is collected from the BSAU forum back-end database is shown in figure 4.9.
The collected web data is in an unstructured form which is converted into a structured form. The web content is indexed with the help of Meta tag indexing method. The attributes targeted for indexing are web document id, the frequency of the term in a web document and position of a term in the document. In meta tag indexing method the collected web data from educational sites are compared with indexed based on the metadata occurrence that is, the web content which is surfed by the user most frequently is compared with the collected web content. The data is extracted by using R tool in the data extraction phase; the complete conversion from unstructured form to structured form is done here. With help of R tool, the collected data from the previous phase is extracted and organized for further analysis of data to be stored in the common repository. Figure 4.10 illustrates the data extraction phase of sample student data; the data extracted in this phase defines the organized attribute of the collected data.
The data is collected and organized at this juncture, but the complete enrichment of data is done using an unsupervised learning technique called 'data clustering'. In data clustering, a web document is collected and it is organized according to content similarities. Here K-means clustering technique is used where the extracted data is clustered step by step based on the similarity in web data.

Table 4.2: Clusters Formed Using K-Means Algorithm

<table>
<thead>
<tr>
<th>Cluster id</th>
<th>Cluster category</th>
<th>Cluster constraint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High performer</td>
<td>Overall gpa&gt;8.5 Soft skills rating = “A” grade Analytical skill rating = “Good”</td>
<td>A student who satisfies the required constraints with the deserved score for a high performer.</td>
</tr>
<tr>
<td>2</td>
<td>Intermediate performer</td>
<td>Overall gpa&gt;6.5 Soft skills rating = “C” grade Analytical skill rating = “Better”</td>
<td>A student who satisfies the required constraints with the deserved score for a high performer.</td>
</tr>
<tr>
<td>3</td>
<td>Low performer</td>
<td>Overall gpa&gt;6.0 Soft skills rating = “E” grade Analytical skill rating = “Worst”</td>
<td>A student who satisfies the required constraints with the deserved score for a high performer.</td>
</tr>
</tbody>
</table>
The cluster formed using K-means clustering technique is shown in table 4.2 reflects the group of entities with different constraints based on their performance. Thus, the data is acquired from social media to recommend the career guidance for the requested end users’.

4.4 SUMMARY

Thus, the primary objective of converting the raw social media data into interoperable one is done using classification and clustering technique respectively. The data repository consists of the necessary accessible data from which the response for requested user query can be initiated and fetched. The data repository with all the relevant data is gathered. The objective of this section is fulfilled with the building up of a common repository. In the next section, the core part of the recommendation system is discussed. The review and feature analysis are discussed in detail for processing the user opinion and other vital information pertaining to the recommendation.
5.0 OPINION MINING AND RECOMMENDATIONS

In this section, the essential need of recommendation system is established. The basic objective of any recommendation system is to provide the user comfortable in terms of all their basic requirements online. After the acquisition of the required data, the system will process and offer the best possible solution for the requested user.

5.1 INTRODUCTION

The online recommendation system relies on both the user opinion and attribute which are collectively processed in this phase. This recommendation system adopts a hybrid approach which combines a collaborative filtering technique and content-based filtering technique. In this chapter, the recommendation process is carried out for both products based on recommendation and career guidance recommendation. Initially, the process involved in the product based recommendation is discussed which is followed by the career guidance recommendation.

In this phase, the interoperable data format is compared with the information available in the data repository. Figure 5.1 shows the conceptual diagram for the retrieval of information from the repository for the recommendation. BPL index matching algorithm is used to find the similarity between the requested query and information available on the data repository. The end result of this will retrieve the relevant reviews and features of the relevant products. After collecting the corresponding reviews and features; sentiment analysis is used to provide the suitable recommendations to the requested end user.
5.2 OPINION MINING

Opinion mining focuses on the initial step of the recommendation process, the key attributes of the user query and corresponding information pertaining to the attributes are retrieved. To compare the similarities among the information, several information retrieval techniques and algorithms are studied and identified a BPL (Brand and product list) index matching algorithm which retrieves the necessary information for the recommendation process. The comparison initializes with the root element of the record and traverses to the end element to fetch the corresponding result.

If the attribute of the user query and record in the index element matches, then the reviews and features of the corresponding index element are fetched. The steps involved in the BPL index matching algorithm are shown above; Index matching algorithm checks the record individually for all the available attributes of the user query. The attribute of each user queries are verified and the information available in the repository are compared. As a result, the numerous volumes of reviews and features are fetched and analyzed simultaneously.
**BPL index matching algorithm:**

**Input:** User query;

**Output:** Reviews & Features

**Step 1:** User query is supplied as input to the system (IQ).

**Step 2:** Applying Revised Keyword Extraction Algorithm (RKEA) with IQ to transform QA (QA stands for Query Attribute).

**Step 3:** Compare QA with the Brand and product list Index (BPL) in repository.

For i=1…..n;

BPL = IRi, IRi+1…IRn. (IR = Individual Record)

Cmp (QA, BPL);

for (i=1; i<=n; i++)

{} if (QA==BPL) then

{} return BPLi (Reviews and Features) recommendation.

} else

{} Message “System requires brand name or product name “;

} End

Thus, the BPL index matching algorithm provides enough information to analyze the reviews and features for the requested user query. The fetched reviews and features for the recommendation system are analyzed using sentiment analysis which is discussed in the following section.
5.3 REVIEW ANALYSIS

In this section, the collected reviews and features are analyzed using the sentiment analysis. The hybrid approach is used where both collaborative and content-based approach are combined for analyzing the information. The review analysis starts with user initialization on the web. User search for the product in the online as per the need; initially the browsed content starts with the value 0 as the navigation process continues the value will be incremented as the product or information is identified by the user; the corresponding reviews will be fetched for the requested query, the review analysis takes place by calculating the number of positive reviews, number of negative reviews and number of mixed reviews. The query sampling and review analysis are done for the recommendation process. The novice user enters into the GUI in search of any information basically posts the query. This query targets the system and communicates to the service provider in a specific format to get the respective response. In this aspect, the content based strategy has pointed out by previous studies where it uses the description of the items that are previously glanced or purchased by the customer and evaluated by them in a positive way. The system recommends the consumers items similar to the items which they liked in the past. The hybrid algorithm to analyze the sentiments of the reviews and features are described below:
Review–based Hybrid Algorithm (Tb,Bm)

**Input:** User query/search for information (either product or any other service);

**Output:** Recommendation for the query/search

**Start**

\[ T_b = \text{No. of products by } u_{iif} \]  
//Note: uiif is total no.of search on products carried out by the user.

\[ b_m = \text{the browser matrix of all users} \]  
If \( b_m = 0 \) or \( TS = 0 \) Note: initialized value is 0

\[ s.\text{count} = \text{search count of the users} \]

**Step 1: initialize search**

\[ \text{if} \ (s.\text{count}==0) \]

\[ \{ \]

\[ \ T_e = \text{Viewed } N \text{ products} \]  
//Note: if \( n \) value is 1 it will be \( N \) reviews

otherwise

\[ \text{return } T_b; \]

End if

\[ \} \]

**Step 2: Calculate no. of User Search**

\[ U = \{n1, n2, nn\}; \]  
//no. of user search

\[ Tr = \{A,B,C\}; \]  
//types of reviews

\[ \text{Count} = 0; \]  
//first iteration

For each \( u1 \) in \( U \)  
//no. of user search with reviews

**Step 3: Review Evaluation**

\[ \text{Do} \]

\[ \{ \]

\[ N_b = u; \]  
//Both content based & collaborative filtering

\[ \text{if} \ (N_b > \_ T_e) \] then

\[ \text{Count} = \text{count} + 1; \]
End if
End for
}
if (Count <= 0) then.  //reviews and user search are evaluated
{
  \( T_r = n_1, n_2, \ldots, n_n; \)
  return \( T_r; \)
  End if
}

Step4: Recommendation Process
Do
{
  //R-Recommendation
  R = (A-B+ (c1+c2))  // A+veReview, B –veReview, C- Mixed review
  (X): A+c1;// Total no. of +ve review
  (Y): B+c2;//Total no.of –ve review
  R=(X)> (Y)
}
End

5.4 RESULT ANALYSIS

This section discusses the result obtained after the complete analysis of the user opinion and query which are posted by the end users. Firstly, the result analysis starts with the recommendations offered for the query posted by the users to select the required electronic product in the online. The reviews and features of the various products are analyzed and reliable recommendations are offered. Secondly, the opinions and reviews of the professional details of the student are analyzed to provide the recommendation for the career guidance in terms of selecting the institute, course, placement probability etc.
5.4.1 Product Recommendation

The recommendation to select the appropriate products in the online are emphasized with the hybrid approach. The hybrid approach algorithm starts with the initialization, where the total number of users’ search for the information in online is initialized to 0 and the subsequent traversal of the users are notified and their behaviours are recorded sequentially. The total number of search for the want of product and number of the request is posted to know the status of the product is listed. The corresponding reviews and features of the respective products are generated. The reviews are manipulated based on their sentiments. For example, the reviews for a particular product are collected and analyzed. The sentiments of the reviews are analyzed. The table 5.1 represents a recommendation process for a sample product which has the collection of positive, negative and mixed reviews. As discussed earlier in review-based hybrid algorithm, the recommendation process is defined and calculated. The parameters R, X and Y are used to specify the recommendation(R), total number of positive reviews in combination with positive reviews occur in the mixed review(X) and total number of negative reviews in combination with negative reviews occur in the mixed review(Y). These parameters are cross-examined and required recommendation is offered to the requested user.

**Table 5.1: Recommendation Process for a Product**

<table>
<thead>
<tr>
<th>A (positive Review)</th>
<th>B (Negative Review)</th>
<th>C(Mixed Review)</th>
<th>Total Review Count(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td>240</td>
<td>115</td>
<td>110</td>
</tr>
</tbody>
</table>

R-Recommendation
R= (A-B+ (c1+C2))
R= (245-240+ (225))
(X): Total no. of +ve review= A+c1=360
(Y): Total no. of –ve review=B+c2=350
R=(X)>(Y)
Since R=(X)>(Y), therefore the requested product is recommended for selection as shown. The sentiments of the various reviews are analyzed based on their polarity count and corresponding recommendation are offered.

Figure 5.2: Set of Stored Positive Reviews for Analysis

Figure 5.2 shows the set of positive reviews are stored and expelled out for the particular user query. These reviews are used and their polarity is predicted to offer the most suitable recommendation for the requested user. The recommendation for the requested query is analyzed individually for the various end users’. The recommendation takes place one at a time. The result of the sentiment analysis brings the most suitable and suggestion to the requested user. The typical recommendation is offered to the user on responding to the queries posted by the end user is shown in figure 5.3

Figure 5.3: Customized Recommendation to the Requested User

---

I used Yoga 2 Pro for last two years as my primary PC and bought Yoga 2 Pro to explore it. First of all it’s a premium to an illusion screen in the same body size and design construction in the same weight, I just that some buttons like volume and muting lock buttons on the sides had to be satisfied (plastic to metal body transition doesn’t come for free), although handrest volume buttons were quite small, it made a lot. Also gave it the impression "messages" touch buttons withLatin feedback. Now it looks like only way to make a mistake at the tablet goes is to lose the side power button, although problems would be power saving. I often used my Yoga 2 Pro to work, while music was playing on the table and plugged in my charging. It’s an issue possible with the new plug. I just don’t have the space to store in a set of boxes and typing these are not so (there). I just once had the right shift key.

The recommendation for the requested query is analyzed individually for the various end users. The recommendation takes place one at a time. The result of the sentiment analysis brings the most suitable and suggestion to the requested user. The typical recommendation is offered to the user on responding to the queries posted by the end user is shown in figure 5.3.
The sentiment analysis provides the required results for the requested user with the minimum offering of recommendation to the query which is posted. Thus, the product recommendation for the requested user query is discussed.

5.4.2 Service Recommendation.

To analyze the recommendation process for the academic domain, an online recommender application for the learners online that requires the recommendation in terms of their career guidance is discussed in detail. The implementation work is based on the collected web data from the educational forum referred as 'BSAU forum' where the student community shares their academic and personal data. A case study is conducted by collecting the educational data from the BSAU forum which describes various attributes in an academic institute such as the academic performance of a student, extracurricular achievement attained by the student, placement probability of a student, overall performance of an institution in terms of web rating, and so on. The recommendation for the career guidance and placement probability for a postgraduate student which is based on the current grade point average (GPA) score of a student, undergraduate score, analytical skill rating, soft skill rating, overall performance rating and various other norms followed in the recruiting corporate sectors. Users posts the query which is converted into a structured query will be directed to the database by the knowledge engine where the collected data will be cross-examined with the requested query. After a query conversion, the required data will be fetched. After the implementation, the results exhibit the compatibility of the proposed framework. The placement criteria of various corporations are listed in Figure 5.4, which is based on the user ranking in terms of different recruiting procedures by different corporations.
Figure 5.4: Ranked Placement Criteria

The student eligibility criterion which must be attained by the student for the successful participation in the placement process. The eligibility criteria are listed based on the placement standards which are maintained in all the corporate sectors which shows the statistical information about eligibility criteria of various companies which pursues the student's performance in their curriculum. The performance ranking of a student indicated in Figure 5.5 reveals the various factors such as academic performances, extracurricular activity and placement criteria. The ranking is based on the accumulation of all the performances. The performer can be of various classes such as a high performer, intermediate performer, and so on.
The overall performance is calculated which reveals the effective data which can be used by the institution for the efficient improvement in the performance of the student and placement probability in the future. Thus this case study with 1107 students reveals the factors involved in and user intervention in deciding the placement probability and academic performance of postgraduate students. Finally, the requested end users get the appropriate response through dedicated GUI with the help of the generated report which displays the placement probability of a particular student. As the result reveals that the current system is worthwhile for recommending the needy information to the end user.

5.5 CONCLUSION

Thus, the recommendation offered in the system takes place using the mechanism of capturing and sharing of knowledge with the right person at the right time. The recommendation strategy has the significant contribution in taking a query from a user and corresponding reviews and features expel out to predict the right decision. The framework is designed in such a way.
that, it accepts the various constraints to offer a recommendation for any kind of domain. Here, both product-based recommendation and service-based recommendation are offered to the end user.

The expectations of the users in the system are extended beyond the normal recommendation. The end users are not satisfied only with the recommendation, additional information than the recommendation is offered through ranking and personalization. Therefore, the enhancement is carried out by the means of ranking and suggesting the other possible options. The typical recommendation is supported by additional information to make the user completely satisfy with the system. The optimal ranking and concept of personalization will be discussed in the next section, where ranking and personalization in the social media environment plays a key role.
6.0 PERSONALIZED RECOMMENDATION WITH RANKING

6.1 INTRODUCTION

This chapter focuses on the personalized recommendation along with the optimal ranking of the other available options to the end user. The uniqueness of the RS is projected through two important criteria; firstly the optimal ranking of the suggestions with the solid evidence and justification, secondly the personalization factor which is verified for the personalization of the user by cross-examining with the user personalized information. The system should understand the constraints which are contrived by the end user on different aspects of the system usage. The PORS framework is constructed in such a way that it accepts all kinds of improvisation and changes dynamically as per the end user demand.

In this aspect, the groundwork needs to be initiated to fulfill the two additional criteria or two important elements of the whole RS developed. To attain the ranking of the other suggested product or any other services and user personalized recommendation, the system has to be thoroughly studied and the possible inclusion of the needy strategy and logical interpretation must be examined. Figure 6.1 shows the workflow diagram of the final phase where the RS is improvised and modified to showcase a ranking and personalized recommendation.

For ranking the other products or finding the additional information for any other services, the primary task is to find the similarity among the products/service information in terms of a feature of the other available product/service information in the repository. So the strategy needs to be implicated to list the other products/service information first, which is followed by the feature similarity mapping of the other products/service information. The attributes of the requested query are compared with the available information in the repository and similar item/information and its attributes are listed. The collected attributes are tested and ranking has been carried out based on the various factors. The ranking takes place after the initial
recommendation process offered by the system. Here, the ranking process for both product based recommendation and service-based recommendation is discussed.

Figure 6.1: Work-flow diagram for Ranking with Recommendation

To offer the personalized recommendation, the end user information is preserved. This information indicates the actual ability of the end user who needs the prime support from the system at any level of access. The user personalized information like user preferences, user navigation, user hobbies, user cost budget, user area of interest, user feedback, user historical details, user constant demand are gathered and appropriate recommendation along with the ranking is offered. Thus, in the next section, the key aspect of ranking phenomena and personalized recommendation factors will be discussed in detail.
6.2 FEATURE ANALYSIS

This section enumerates about the underpinning aspect of recommendation system along with the ranking. As discussed in the previous section, the feature of the other products and any other information for service available in the social media common repository desired to be analyzed. The analysis is essentially initiated through listing the product/service information from the repository and listed products/service information features which are evaluated gradually. The manipulated features of the other products/service information are compared with the product/service information which is initially requested by the end user. The result of the comparison is recorded and appropriate approach is adapted to provide the necessary outcome of the analysis. The analysis of features and feature similarity mapping are implemented with the help of Revised Rainbow Algorithm (RRA) which is meant to select the skyline of the data and generates the data with the required information. The feature similarity among the other products/service information will be also identified accordingly. The ranking processes are implemented using the Linear Regression Analysis (LRA).

Revised Rainbow Algorithm (RRA) is used to select the similar contents which are requested by the end user. The algorithm works step by step where similar contents are listed at the beginning through skyline selection and constraint is set to define the feature. The advantage of using revised rainbow algorithm is user queries are verified and features are listed in the incremental procedure. Linear Regression Analysis (LRA) is used to provide the ranking for the selected information on the web. The benefit of using the linear regression analysis is, the analysis goes through the various processes like normalizing the features, assigning weightage to the features, optimized value is measured and ranking takes place at the end. This step by step procedure is applicable for any kind of dataset since the process involved are well defined to meet the generic requirement for ranking.
6.3 REVISED RAINBOW ALGORITHM (RRA) FOR FEATURE SIMILARITY MAPPING

To initialize the ranking factor of the other products / other service information, the revised rainbow algorithm is incorporated. The main objective of this algorithm is to track the required information for the system and specify the parameters which can be used to list the products/service information effectively. The major part of this technique focuses on finding the similarity among the features of the other product with the user requested product.

RR Algorithm (Revised Rainbow Algorithm) is used for listing the product attributes such as features and reviews of the product in a suitable manner; the algorithm structured in such a way that the process starts from product/information search, other than the recommended product / service information, the corresponding flags are set to indicate whether the other product/service information and its attributes are listed; then the similarity of the products / service information is identified by comparing the features and the list is incremented in sequence to store the information. The skyline selection of data is enabled i.e. the search of the data in the repository begins as per the user demand. The area of selection depends upon the intensity of the user query where the search takes place in a random direction. The skyline search identifies the list of product/service information which is received through a wide range selection based on different criteria.

The skyline selection search is an iterative process where the lists of products/service information are generated by comparing the user query. The first step of the algorithm provides the selection of the other products whereas the consequent step of the revised rainbow algorithm is intended to showcase the more useful and handy information for the ranking. Once the skyline selection tracks the most probable set of information, the second step of the algorithm takes one more leap to identify the similarity among the features of the product/service information. The key aspect of this algorithm is projected in the second step, where the rainbow spectrum is generated i.e.
the similar product is listed down with the specific constraint. The steps involved in the skyline selection of the product/service information are described below:

**Revised Rainbow algorithm for product/service information retrieval and similarity measurement**

**Revised Rainbow (product list/service)**

*Input:* Set $i$ of responding similar items to a buyer’s query $q_1$

*Output:* A subset $S$ of $i$ (skyline items) as the finalized list of products/service information

**Step 1: Start**

**Step 2: Skyline selection of products or services**

2.1 For all products or service information $x \in i$

2.2 If ($x_i == product/service information$) then

$\{
    \text{Set the Flag} = \text{true}
\}$

Else

$\{
    \text{Set the Flag} = \text{false}
\}$

2.3 If Flag = true,

$\{
    \text{Add attributes of product/service information in the list } L_i;
\}$

Else

$\{$

78
Step 2: End if

Step 3: Generating the Rainbow spectrum (Traversing step-by-step to feature related products/Service Information)

Input: Set I of responding similar items to a user query q1 & compare

Output: A subset S of I (skyline items) as the finalized list of products/service information

3.1 Start with I = 1 //Availability of products/service information
3.2 While (I ≠0) // Item search is not empty
   {
   Add product/service information attribute Li;
   }
3.3 Select features Li products/service from set I &cmp (i)
3.4 If (Li=I)
   {
   Return (Increment index i= I + 1)
   }
   Else
   {
   Goto step 3.1;
   }
3.5 End if
The later part of the revised rainbow algorithm focuses on finding the similarity among the feature. The spectrum of features are selected, each feature of the corresponding other products are listed. The feature similarities of the products are identified through extensive comparative study conducted and corresponding results are stored. The result of the revised rainbow algorithm for the sample query “Is it safe to buy iPhone 7 plus cash back 7500 paytm” is listed in Table 6.1

Table 6.1: Product List with its Attribute

<table>
<thead>
<tr>
<th>Listed products</th>
<th>Selected Feature List (Li)</th>
<th>Reviews</th>
<th>Camera Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
<td>RAM Size</td>
<td>+ve</td>
</tr>
<tr>
<td>Swipe Konnect Plus</td>
<td>8000</td>
<td>2</td>
<td>1050</td>
</tr>
<tr>
<td>Blackberry Passport(Black)</td>
<td>8500</td>
<td>3</td>
<td>1125</td>
</tr>
<tr>
<td>Blackberry Passport(White)</td>
<td>10000</td>
<td>6</td>
<td>1350</td>
</tr>
<tr>
<td>Apple iPhone 6s Plus</td>
<td>12000</td>
<td>4</td>
<td>1280</td>
</tr>
<tr>
<td>Apple iPhone 6s Plus(Grey)</td>
<td>18000</td>
<td>8</td>
<td>1300</td>
</tr>
</tbody>
</table>

The features of the product are listed; the algorithm works in a loop where the search of the product is repetitive. For the sample query, the product features are listed in the above table. The revised rainbow algorithm uses an appropriate method to extract the product list and corresponding features based on nature of the query. The query might be a generalized one or it might be attribute specific such as cost, review, cam pixel. Therefore, the revised rainbow algorithm picks a user query and lists the product attribute where the features and other attributes which are the most matching. The listed product and its attribute need to be semantically interpreted with the help of an appropriate method for ranking. This information is streamlined and carefully examined in the next phase with the help of the technique linear
regression analysis in which the ambience for the ranking is initialized. The product with its features and the inclusion of reviews are examined in the next section.

6.4 LINEAR REGRESSION ANALYSIS (LRA) FOR PRODUCT RANKING

The ranking of other products with appropriate justification is very crucial in personalizing the recommendation as per the user request. The personalization factor does carry a good value to the overall recommendation process. Linear regression analysis is used to predict and rank the set of dependent variable and also the outcome of the supporting parameter is calculated in a meaningful way. The main objective of the linear regression analysis is to conduct a cautious inspection of the features which are collected from the previous phase. This investigation should offer the required set of crucial information to precede the process of ranking. The key aspect of the investigation brings the normalization factor and optimization factor into account. The linear regression analysis targets both multi attribute-based query and single attribute-based query. LRA-MAPR (Linear regression analysis for multi-attribute based product ranking) is applied for multi attribute-based query and LRA-SAPR (Linear regression analysis for single-attribute based product ranking) is applied for the single attribute-based query. The steps involved in the linear regression analysis are described below:
Linear Regression Analysis for Product Ranking

**Input:** Multi-attribute feature of a product

**Output:** Recommended information with ranking

**Procedure:**

1. Normalize the listed product features (pf) // optimizing feature data
2. Calculate \( N, \) where \( N = (pf \times \text{max} (pf) \times 100) \) // normalize
3. for \( i = 1 \) to \( k \), //start with cost of the product
4. Iterate \( N \) for \( i=1 \) to \( n \) // normalize all the features
5. Compute OR; // Optimized ranking of products
6. Sort OR; // Sort the ranked product based on recommendation
7. Do \( R \); // offer a recommendation with optimized ranking
8. End

The product list with its attribute is normalized and the corresponding weight is assigned for each feature and linear regression formula is derived for ranking and conventional sorting is done based on the user preference. Once the normalization process is done, the weight is assigned for each attribute based on the impact and the ranking process is implemented. The product ranking is implemented by deriving the equation of linear regression analysis where the dependent variable is interpolated and corresponding values are incremented until the endpoint is reached. The normalization steps involved are listed below:

- Normalize the feature for optimizing the values.
- Assign weightage for each feature and calculate the optimal value.
- Find the best possible product (optimized value) from the list.
- Sort the product based on the user preference.
- Recommend the product with optimized ranking.
The linear regression analysis is implemented for the sample query “Is it safe to buy iPhone 7 plus cash back 7500 paytm?” the product list with its attribute are listed through the revised rainbow algorithm. The product names are assigned with the product numbers such as p1, p2 etc; for applying the linear regression analysis. The product list with its attribute is listed in table 6.2

Table 6.2: Product List with its Attribute for Normalization

<table>
<thead>
<tr>
<th>PRODUCT LIST</th>
<th>COST</th>
<th>RAM SIZE</th>
<th>POSITIVE</th>
<th>NEGATIVE</th>
<th>CAMERA RESOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>8000</td>
<td>2</td>
<td>1050</td>
<td>975</td>
<td>8</td>
</tr>
<tr>
<td>P2</td>
<td>8500</td>
<td>3</td>
<td>1125</td>
<td>1020</td>
<td>13</td>
</tr>
<tr>
<td>P3</td>
<td>10000</td>
<td>6</td>
<td>1350</td>
<td>1260</td>
<td>13</td>
</tr>
<tr>
<td>P4</td>
<td>12000</td>
<td>4</td>
<td>1280</td>
<td>1175</td>
<td>16</td>
</tr>
<tr>
<td>P5</td>
<td>18000</td>
<td>8</td>
<td>1300</td>
<td>967</td>
<td>16</td>
</tr>
</tbody>
</table>

The normalization process is applied to the product features which are listed in table 6.2. The normalization procedure is derived by scrutinizing the features and its impact in the context of the recommendation and query posted by the user. The example query is used here is multi-attribute feature based query; therefore the normalization rules are derived and described as follows:

**DERIVED NORMALIZATION PROCESS:**

Cost Normalization (CN) = \( \frac{\text{product cost}}{\max(\text{product cost})} \) *100

Ram Size Normalization (RSN) = \( \frac{\text{ram size}}{\max(\text{ram size})} \) *100

Positive Review Normalization (PRN) = \( \frac{\text{positive review}}{\max(\text{positive review})} \) * 100

Negative Review Normalization (NRN) = \( \frac{\text{negative review}}{\max(\text{negative review})} \) * 100

Camera Resolution Normalization (CRN) = \( \frac{\text{camera resolution}}{\max(\text{camera resolution})} \) * 100
The derived normalization rules are applied for the sample query which is tested and the corresponding normalized value is deduced which is shown in the table 6.3 the normalization process takes up the maximum value of the entire attribute and equally bisected for the feature weight.

**Table 6.3: Normalized Value Product with Attribute**

<table>
<thead>
<tr>
<th>PRODUCT LIST</th>
<th>COST (CN)</th>
<th>RAM (RSN)</th>
<th>POSITIVE REVIEW (PRN)</th>
<th>NEGATIVE REVIEW (NRN)</th>
<th>REVIEW</th>
<th>CAMERA RESOLUTION (CRN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>44</td>
<td>25</td>
<td>78</td>
<td>77</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>P2</td>
<td>47</td>
<td>38</td>
<td>83</td>
<td>81</td>
<td>2</td>
<td>81</td>
</tr>
<tr>
<td>P3</td>
<td>56</td>
<td>75</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>P4</td>
<td>67</td>
<td>50</td>
<td>95</td>
<td>93</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>P5</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>77</td>
<td>19</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6.3 shows the first iteration of the normalization where all the attributes are streamlined to a certain extent. The next iteration focuses on the assigning the weightage to the attributes based on the impact of the attribute which narrows down the value furthermore suitable for ranking. The assigned weight for the product attributes are shown in table 6.4

**Table 6.4: Weightage Assigned for Product Attributes**

<table>
<thead>
<tr>
<th>Product Attributes</th>
<th>Assigned Weightage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>10%</td>
</tr>
<tr>
<td>Ram Size</td>
<td>30%</td>
</tr>
<tr>
<td>Reviews</td>
<td>40%</td>
</tr>
<tr>
<td>Camera Resolution</td>
<td>20%</td>
</tr>
</tbody>
</table>
This weightage assignment varies the complexion of the overall statistics. Since the impactful product attribute is identified, the weightage is assigned by evaluating the user interest and impact on the various features. The cost and the camera resolution weightage are assigned with its impact. The most crucial part of the attribute which decides the outcome of the recommendation is the user review of the product. After assigning the weightage to the product attribute, the optimization factor is considered where the optimization ranking equation is derived as follows:

\[
\text{OR} = \sum_{\text{MIN}}^{\text{MAX}} (X_1 \cdot W_1) + (X_2 \cdot W_2) + (X_3 \cdot W_3) + \ldots + (X_n \cdot W_n)
\]

\[
\text{OPTIMIZED RANKING} = (\text{CN} \cdot \{30/100\}) + (\text{RSN} \cdot \{30/100\}) + (\{\text{PRN} - \text{NRN}\} \cdot \{40/100\}) + (\text{CRN} \cdot \{20/100\})
\]

Table 6.5: Product List with Optimal Ranking

<table>
<thead>
<tr>
<th>WEIGHTAGE</th>
<th>10%</th>
<th>30%</th>
<th>40%</th>
<th>20%</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRODUCT LIST</td>
<td>COST (CN)</td>
<td>RAM SIZE (RSN)</td>
<td>POSITIVE (PRN)</td>
<td>NEGATIVE (NRN)</td>
<td>REVIEW (RN)</td>
</tr>
<tr>
<td>P1</td>
<td>44</td>
<td>25</td>
<td>78</td>
<td>77</td>
<td>1</td>
</tr>
<tr>
<td>P2</td>
<td>47</td>
<td>38</td>
<td>83</td>
<td>81</td>
<td>2</td>
</tr>
<tr>
<td>P3</td>
<td>56</td>
<td>75</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>P4</td>
<td>67</td>
<td>50</td>
<td>95</td>
<td>93</td>
<td>2</td>
</tr>
<tr>
<td>P5</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>77</td>
<td>19</td>
</tr>
</tbody>
</table>
Now, the subsequent modification is implicated in the product table with its attribute, the optimized ranking equation is applied and the result is calculated for all the normalized product attributes. The derived OR formula is substituted in the normalized attribute table and it shows the concrete evidence that the ranking is conducted in genuine manner. The result found in table 6.5 shows that, the optimal ranking of the other suggested products are listed and it justifies the implication of linear regression analysis.

As table 6.5 suggests that the ranking has conducted in an optimal way using linear regression analysis. The OR formula with the assignment of weightage for the sample query is shown below:

\[
\text{OR} = \left( X_1 \times W_1 \right) + \left( X_2 \times W_2 \right) + \left( X_3 \times W_3 \right) + \left( X_4 \times W_4 \right)
\]
\[
\text{OR} = \left( \text{cost} \times 0.1 \right) + \left( \text{ram size} \times 0.3 \right) + \left( \text{review} \times 0.4 \right) + \left( \text{cam resolution} \times 0.2 \right)
\]
\[
\text{OR} = 67.6 \quad \text{// Product “P5” top ranked and recommended.}
\]

Thus, the result obtained for the given sample query “Is it safe to buy iPhone 7 plus cash back 7500 paytm?” is shown in table 6.6 in which the product list is sorted as per the rank list.

**Table 6.6: Optimal Rank List with Product Recommendation for Query**

<table>
<thead>
<tr>
<th>PRODUCT LIST</th>
<th>OPTIMIZED RANKING(OR)</th>
<th>Rank List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Iphone 6s Plus(Grey)</td>
<td>67.6</td>
<td>1</td>
</tr>
<tr>
<td>Blackberry Passport(White)</td>
<td>44.3</td>
<td>2</td>
</tr>
<tr>
<td>Apple Iphone 6s Plus</td>
<td>42.5</td>
<td>3</td>
</tr>
<tr>
<td>Blackberry Passport(Black)</td>
<td>33.1</td>
<td>4</td>
</tr>
<tr>
<td>Swipe Konnect Plus</td>
<td>22.3</td>
<td>5</td>
</tr>
</tbody>
</table>

Thus, with the help of linear regression analysis, the optimal ranking of the product is established. The following section discusses the ranking process involved for the other services such as career guidance.
6.5 LINEAR REGRESSION ANALYSIS (LRA) FOR SERVICE RANKING

The ranking of other service information such as career guidance information to the end user is offered with more options. The requested end users are provided with the additional information in terms of the ranking. A case study conducted with the 1107 student professional data is revisited to track the feature component for similarity study and ranking. Here, the ranking of the students with the various performance factors are listed. Some of the features taken into account are academic performances; extracurricular activity skill set ratings and score in placement criteria. The ranking is based on the accumulation of all the performances. The performer can be of various classes such as a high performer, intermediate performer, and so on. Some of the criteria involved in ranking the students are a collection of the academic performance record, score obtained in the extracurricular activity and skill set ratings and placement criteria of the student.

The main objective of the linear regression analysis is to conduct analysis where the impactful features are listed and spotted for the reliable ranking. This investigation should offer the required set of crucial information to precede the process of ranking. The key aspect of the investigation brings the normalization factor of the features. The steps involved in the linear regression analysis are described below:
Linear Regression Analysis for Service Ranking

**Input:** Multi-attribute feature of student information

**Output:** Recommended information with ranking

**Procedure:**

1. Normalize the listed service features (sf)  // optimizing feature data
2. Calculate N, where $N = (sf \times \text{max}(pf) \times 100)$  // normalize
3. for $i = 1$ to $k$,  // start with academic score of the student
4. Iterate N for $i=1$ to $n$  // normalize all the features
5. Compute OR;  // Optimized ranking of students
6. Sort OR;  // Sort the ranked student based on recommendation
7. Do R;  // offer recommendation with optimized ranking
8. End

The student list with its attribute is normalized and the corresponding weightage is assigned for each feature and linear regression formula is derived for ranking and conventional sorting is done based on the user preference. Once the normalization process is done, the weightage is assigned for each attribute based on the impact and the ranking process is implemented. The student ranking is implemented by deriving the equation of linear regression analysis where the dependent variable is interpolated and corresponding values are incremented until the endpoint is reached. The normalization steps involved are listed below:

- Normalize the feature for optimizing the values.
- Assign weightage for each feature and calculate the optimal value.
- Find the best possible product (optimized value) from the list.
- Sort the student based on the impactful feature.
- Recommend the service with optimized ranking.

The linear regression analysis is implemented for the sample query “generate the performance ranking of Alex” the student list with its attribute are listed through the revised rainbow algorithm. The student names are
assigned as s1, s2 etc; for applying the linear regression analysis. The student list with its attribute is listed in table 6.7

Table 6.7: Student List with its Attribute for Normalization

<table>
<thead>
<tr>
<th>Student list</th>
<th>Academic score (25)</th>
<th>Skillset ratings (25)</th>
<th>Score in extracurricular activity (25)</th>
<th>The score in placement criteria (25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>21</td>
<td>20</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>S2</td>
<td>23</td>
<td>19</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>S3</td>
<td>24</td>
<td>22</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>S4</td>
<td>20</td>
<td>23</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>S5</td>
<td>22</td>
<td>21</td>
<td>23</td>
<td>20</td>
</tr>
</tbody>
</table>

The normalization process is applied to the features which are listed in the table 6.7. The normalization procedure is derived by scrutinizing the features and its impact in the context of the recommendation and query posted by the user. The normalization rules are derived and described as follows:

**DERIVED NORMALIZATION PROCESS:**

- Academic Score Normalization (ASN) = \( \frac{\text{Score}}{\text{max (score)}} \) * 100
- Skill Set Normalization (SSN) = \( \frac{\text{Skill set}}{\text{max (skill set)}} \) * 100
- Extracurricular Score Normalization (ESN) = \( \frac{\text{EScore}}{\text{max (Escore)}} \) * 100
- Placement Score Normalization (PCN) = \( \frac{\text{PScore}}{\text{max (Pscore)}} \) * 100

The derived normalization rules are applied for the sample query which is tested and the corresponding normalized value is deduced which is shown in table 6.8. The normalization process takes up the maximum value of the entire attribute and equally bisected for the feature weight.
Table 6.8: Normalized Value Student List with Attribute

<table>
<thead>
<tr>
<th>Student List</th>
<th>Academic score (ASN)</th>
<th>Skillset ratings (SSN)</th>
<th>Extracurricular Score (ESN)</th>
<th>Placement Criteria (PCN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>84</td>
<td>86</td>
<td>91</td>
<td>87</td>
</tr>
<tr>
<td>S2</td>
<td>95</td>
<td>82</td>
<td>100</td>
<td>91</td>
</tr>
<tr>
<td>S3</td>
<td>100</td>
<td>95</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>S4</td>
<td>83</td>
<td>100</td>
<td>87</td>
<td>100</td>
</tr>
<tr>
<td>S5</td>
<td>91</td>
<td>91</td>
<td>95</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 6.8 shows the first iteration of the normalization where all the attributes are streamlined to a certain extent. The next iteration focuses on assigning the weightage to the attributes based on the impact of the attribute which narrows down the value furthermore suitable for ranking. The assigned weightage for the ranking attributes are shown in the below table 6.9

Table 6.9: Weightage Assigned for Ranking Attributes

<table>
<thead>
<tr>
<th>Ranking Attributes</th>
<th>Assigned Weight-age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>20%</td>
</tr>
<tr>
<td>Skill Set</td>
<td>40%</td>
</tr>
<tr>
<td>Extracurricular</td>
<td>10%</td>
</tr>
<tr>
<td>Placement Criteria</td>
<td>30%</td>
</tr>
</tbody>
</table>

This weightage assignment varies the complexion of the overall statistics. Since the impactful ranking attribute is identified, the weightage is assigned by evaluating the impact of the various features. The skill set and the placement criteria weightage are assigned with its impact. After assigning the weightage to the ranking attribute, the optimization factor is considered where the optimization ranking equation is derived as follows:
Now, the subsequent modification is implicated in the student table with its attribute, the optimized ranking equation is applied and the result is calculated for all the normalized product attributes. The derived OR formula is substituted in the normalized attribute table and it shows the concrete evidence that the ranking is conducted in genuine manner. The result found in the table 6.10 shows that the optimal ranking of the students along with the recommendation to the requested student.

Table 6.10: Student List with Optimal Ranking

<table>
<thead>
<tr>
<th>Weightage</th>
<th>20%</th>
<th>40%</th>
<th>10%</th>
<th>30%</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student List</td>
<td>Academic score (ASN)</td>
<td>Skill set ratings (SSN)</td>
<td>Extracurricular Score (ESN)</td>
<td>Placement Criteria (PCN)</td>
<td>Optimized Ranking (OR)</td>
</tr>
<tr>
<td>S1</td>
<td>16.8</td>
<td>34.4</td>
<td>9.1</td>
<td>26.1</td>
<td>86.4</td>
</tr>
<tr>
<td>S2</td>
<td>19.0</td>
<td>32.8</td>
<td>10</td>
<td>27.3</td>
<td>89.1</td>
</tr>
<tr>
<td>S3</td>
<td>20.0</td>
<td>38.0</td>
<td>10</td>
<td>28.5</td>
<td>96.5</td>
</tr>
<tr>
<td>S4</td>
<td>16.6</td>
<td>40.0</td>
<td>8.7</td>
<td>30</td>
<td>95.3</td>
</tr>
<tr>
<td>S5</td>
<td>18.2</td>
<td>36.4</td>
<td>9.5</td>
<td>24.9</td>
<td>89.0</td>
</tr>
</tbody>
</table>

As table 6.10 suggests that the ranking has conducted in an optimal way using linear regression analysis. The OR formula with the assignment of weightage for the sample query “generate the performance ranking of keane alex” is shown in the table 6.11 in which the student list is sorted as per the

\[
\text{OR} = \sum_{\text{MIN} = 0}^{\text{MAX} = 100} (X_1 * W_1 + (X_2 * W_2) + (X_3 * W_3) + \ldots + (X_n * W_n))
\]

\[
\text{OPTIMIZED RANKING} = ((\text{ASN} * (10/100)) + (\text{SSN} * (40/100)) + ((\text{ESN} * (10/100)) + (\text{PCN} * (30/100)))
\]
rank list. The performance ranking of the student is displayed as per the impactful features.

**Table 6.11: Optimal Rank List with Service Recommendation for Query**

<table>
<thead>
<tr>
<th>STUDENT LIST</th>
<th>OPTIMIZED RANKING(OR)</th>
<th>Rank List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kane Richard</td>
<td>86.4</td>
<td>5</td>
</tr>
<tr>
<td>Farhan Nadeem</td>
<td>89.1</td>
<td>3</td>
</tr>
<tr>
<td>Keane Alex</td>
<td>96.5</td>
<td>2</td>
</tr>
<tr>
<td>Rizwan Alam</td>
<td>95.3</td>
<td>1</td>
</tr>
<tr>
<td>Arun Terry</td>
<td>89.0</td>
<td>4</td>
</tr>
</tbody>
</table>

Thus, with the help of linear regression analysis, the optimal ranking of the career guidance of the student is established. The uniqueness of the recommendation system which is constructed through PORS framework is projected through the ranking. Most of the recommendation system facilitates the provision to the end user by offering the suggestions with the user profile information. The approach used in this research clears most of the challenges and issues listed in the existing work. The result obtained in each phase of this of this research work is discussed in detail in the next section.

**6.6 RESULTS**

The implemented PORS framework proved that the flow of process had brought the significant outcome of the overall result. As the research ambience augmented, the existing research work and related study in this domain were extensively elaborated in the literature review. Out of which the standard issues in recommendation system and situational issues involved during the process of framework construction were recorded. These issues were addressed at each level of research progress gradually.
As discussed earlier, the gateway of any system was quite crucial thus, the GUI was designed in such a way that it acknowledged all the inroads made by the end user through user query. Therefore, in phase 1 of this research work, more than 4800 user queries were tested in the query processor phase and query conversion ratio was summarized which signified the transformation of user query into the interoperable format.

In phase 2, acquisitions of data from the social media discussed in which reviews and features were collected. The collected reviews were classified using SVM classification technique to filter out the key attribute of the social media data. More than 1650 sampling reviews were taken into account and the accuracy of the data was measured using precision and recall with the accuracy values of 52% and 47%. After extracting the reviews, the rapid automatic keyword extraction technique was used to extract the features from the review text. The highest word score metric was calculated and candidate keyword was compared with the corpus collection of the keyword based repository. The corresponding keyword was verified and tagged as a feature. Once the reviews and features were gathered, the k-means clustering technique was used to cluster the reviews and features. This information was stored in social media reviews and feature repository.

Once the data was stored, the next phase targeted to get the required information to the user after processing the query, BNP index matching algorithm showcased the similarity among the query from the user’s perspective and information available in the data repository to pull out the matching reviews and features. The result of index matching algorithm retrieved the information one after the other based on the user query frequency. Here, the huge volume of data was compared where the cold start problem and sparsity issue were addressed in big time. Once the comparative analysis was conducted, the typical recommendation process for the requested query was initiated.
The recommendation processes started by invoking the hybrid approach both collaborative and content-based techniques were used. The reviews and features of the particular query were flooded out to which the polarity detection was found to analyze the sentiments of the reviews. The positive review, negative review and mixed review were semantically analyzed and the customized recommendation was offered using sentiment analysis.

This PORS framework was renowned for its extravagant service i.e. after the conventional recommendation offered for product selection and other service information selection; this system indulged itself to provide the optimal ranking also along with the personalized recommendation. The sample queries tested were “any Lenovo model mobile with the cost of 7000 to 20000” and “generate the performance ranking for Keane Alex”.

### For the user query: “any lenovo model mobile with cost of 7000 to 20000”

<table>
<thead>
<tr>
<th>PRODUCT LIST</th>
<th>OPTIMIZED RANKING(OR)</th>
<th>Rank List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micromax canvas</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>Samsung s3</td>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>Redmi 4S</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>Nokia Lumia</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Karbonn</td>
<td>40</td>
<td>5</td>
</tr>
</tbody>
</table>

### Optimized Cost based Ranking with Recommendation

<table>
<thead>
<tr>
<th></th>
<th>Optimized Cost based Ranking with Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micromax canvas</td>
<td>61</td>
</tr>
<tr>
<td>Samsung s3</td>
<td>57</td>
</tr>
<tr>
<td>Redmi 4S</td>
<td>55</td>
</tr>
<tr>
<td>Nokia Lumia</td>
<td>45</td>
</tr>
<tr>
<td>Karbonn</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OPTIMIZED RANKING(OR)</th>
<th>Rank List</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 6.2: Ranking with Recommendation for Single Attribute Query
Figure 6.2 shows the ranking with a recommendation for single-attribute based query where the ranking and recommendation are projected for the attribute “cost”. Other feature based ranking and recommendation can be implemented for a multi attribute-based query where more than one attribute was analyzed. The recommendation process was improved by 61% to offer the ranking with a personalized recommendation for selecting the product.

Figure 6.3 shows the performance ranking with a recommendation for a query where the ranking and recommendation are projected for skill set attribute which has the more impact on deciding the performance of the student. Other feature based ranking and recommendation can be implemented for a multi attribute-based query where more than one attribute was analyzed. The recommendation process was improved by 96.5% to offer the ranking with a personalized recommendation for career guidance.

<table>
<thead>
<tr>
<th>STUDENT LIST</th>
<th>OPTIMIZED RANKING</th>
<th>Rank List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kane Richard</td>
<td>86.4</td>
<td>5</td>
</tr>
<tr>
<td>Farhan Nadeem</td>
<td>89.1</td>
<td>3</td>
</tr>
<tr>
<td>Keana Alex</td>
<td>96.5</td>
<td>2</td>
</tr>
<tr>
<td>Rizwan Alam</td>
<td>95.3</td>
<td>1</td>
</tr>
<tr>
<td>Arun Terry</td>
<td>89.0</td>
<td>4</td>
</tr>
</tbody>
</table>

**For the user query:** “generate the performance ranking of keana alex”

Figure 6.3: Performance Ranking with Recommendation for a Query
The ranking and recommendation offered need to be measured and user personalization factor as to be retained. Therefore the recommendation offered is measured using a metrics f-measure where the accuracy of the recommendation and performance of the overall system is measured. The final recommendation takes up the user feedback where historical details of the previous recommendation are compared with the current recommendation. With the support of F-measure the accuracy of the system and inference made. The F-measure value 1.14 reflects the threshold value for the given queries which proves that the system provides user personalized recommendation and optimal ranking with the solid justification. The significance of this research work is transparently projected with the tested results. Thus, the study reaches the point to declare that the system implemented is user-friendly and effective enough for various domains. The overall result obtained is described in the table 6.12

Table 6.12: Result Analysis - Summary of Overall Result Obtain

<table>
<thead>
<tr>
<th>RESEARCH PHASE</th>
<th>AIM/OBJECTIVE</th>
<th>ALGORITHM/TECHNIQUE USED</th>
<th>INERENCE MADE/RESULT OBTAINED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query processor</td>
<td>To convert the user query into an Interoperable format</td>
<td>Revised Keyword extraction algorithm (RKEA)</td>
<td>No. of user query tested: 4800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Query conversion ratio:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100% conversion rate via exact keyword</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>90% conversion rate via product name</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>78% conversion rate via product feature</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50% conversion rate via Random &amp; incorrect text</td>
</tr>
<tr>
<td>Acquiring reviews &amp; features from social media web source</td>
<td>To acquire, process and organize the reviews and features from a social web source</td>
<td>SVM classification Technique, Revised RAKE Technique &amp; BNP K-means Clustering Technique, K-means partitioning algorithm</td>
<td>No. of review and features acquired: 5000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inference made:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>On examining the reviews using SVM, the accuracy measure provides Precision value: 47% &amp; Recall Value: 59%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>More than 650 features have been extracted from the review text using Revised RAKE Technique</td>
</tr>
<tr>
<td>RESEARCH PHASE</td>
<td>AIM/OBJECTIVE</td>
<td>ALGORITHM/TECHNIQUE USED</td>
<td>INFERENCE MADE/RESULT OBTAINED</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>--------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Mapping of opinions &amp; customized recommendation</td>
<td>To find the keyword similarity with the user query and available dataset in the repository</td>
<td>BPL index matching algorithm &amp; Sentiment Analysis (Hybrid approach algorithm)</td>
<td>No. of the user query and review data compared: 4800 Inference made: 95% of review and features are expelled out for exact keyword search. Less than 50% of review and features expelled out for random query and incorrect search The customized recommendation is offered to the user on the requested query. (more than 150 recommendations offered in testing phase)</td>
</tr>
<tr>
<td>Feature analysis, optimal ranking &amp; Performance measure of the system</td>
<td>To analyze the feature and offer to rank with a personalized recommendation</td>
<td>Revised Rainbow algorithm &amp; Linear Regression Analysis</td>
<td>No. of query ranked after the recommendation: 225 Feature Analysis: Features are analyzed for each query sent by the user. The ranking is offered with the personal recommendation based on the user query. Recommendation process improves by 96.5%. F-measure value of 1.14 is achieved in testing performance of the overall system.</td>
</tr>
</tbody>
</table>

The overall result obtained is shown in table 6.12 reflects the uniqueness of this research study. The amount of effort involved in each phase of the research work showcases the enhancement in the system when compared to the previous investigations.

The implementation of the framework crafts the end user to experience the various elements of the recommendation system. On the whole, the construction PORS framework in view of developing an effective recommendation system is reached to a point where the system provides the solution for all kinds of user query within a prescribed domain. Thus, the
detailed result analysis proves the system is quite evident to be the established one for any domain under the required circumstances.

As the investigation on the earlier study brings out the various challenges which have been encountered in this research work, therefore it is the high time to offer a concluding remark on the current study and to suggest the areas of improvement and future enhancement. In the next chapter, the exclusive remark on the proposed research work and future enhancement of the research will be briefly discussed.
7.0 CONCLUSION AND SCOPE FOR FUTURE RESEARCH

7.1 INTRODUCTION

The purpose of this research is to construct a generic framework for personalized and optimal ranking system for a recommendation which can determine feasible suggestions or decisions to be taken in a crucial juncture. This chapter discusses the goals, strategy and contributions of the research work. At the end of this chapter, future scope and enhancement of the research work are also discussed. The primary goal of this research work is to direct the end user with the most suitable system which can guide or suggest them to take a right decision during the critical situation. To achieve the goal, few lines have been drawn to meet the basic requirements of the recommendation system. The effort begins with an extensive study of the existing strategies and approaches. The typical approaches such as collaborative filtering technique, content-based filtering technique, and knowledge-based filtering technique have been considered. The challenges and issues involved in these approaches are carefully examined by the proposed research work.

7.2 HIGHLIGHTS OF THE WORK DONE

The well-known issues in RS which is addressed in the first place are cold start problem, sparsity, scalability, privacy, quality etc. These issues are taken into account as a preliminary prerequisite for the construction of PORS framework. JianWei et al.,(2017) highlighted the issue of cold start problem and its impact which is reflected in the overall behaviour of the recommendation system. Therefore, the PORS framework is developed in such a way that the system retrieves an ample amount of real-time and dynamic data from the social media which eradicates the cold start problem completely.
The other factors such as sparsity, scalability were also considered and addressed in this research work. In the social media environment, the data was of different patterns such as comments, posts, reviews, features, tweets. Therefore, streamlining these data into an interoperable one was a challenging task which was achieved. The other issues discovered during the construction stage were an excessive flow of social media data provides too many solutions which were stabilized to a certain extent by organizing the social media data through classification technique. Only one source of social media was used for a recommendation in the existing RS which converted into a heterogeneous collection of data in this research work.

The recommendation system developed in this research work was focused on the generic end user’s perspective rather than a specific one. The framework was flexible for all the domains. The main aim of the PORS framework was to apply the strategy in any domain, based on the area of selection, the corresponding components and their purpose were fine-tuned accordingly. In this dissertation, the selection of products in the online and recommendation for a service such as a career guidance of a student was taken into account and the PORS components were built on this aspect.

This research work introduced various approaches to process the requesting end users. As a part of the strategy, the user query was processed with the help of technique revised keyword extraction algorithm (RKEA) in which the user query was converted into the interoperable format.

In the recommendation system investigated so far it is found that the recommendation was offered mostly using static storage of the data, without proper authentication and source of data used for recommendation is homogenous. But in this research study both static and dynamic storage of data from social media was emphasized. The heterogeneous collection of social media data was applied in this study.
From the social media data, the reviews and features are processed; SVM classification technique was applied to fetch the review data in an organized way. And from the review text, the features were extracted with the new approach called revised rapid automatic keyword extraction (RAKE) where the feature content was collected through step by step process. In the end, the brand name based clustering was done using k-means clustering technique where the reviews and features were stored based on the brand names.

Another important contribution of this research work was preparing the appropriate information for the opinion mining to respond to the end user. In order to fetch the information from the repository, BNP index matching algorithm was used. This index matching algorithm retrieved the information by analyzing the user query and information available in the common repository. Based on the matching constraint the respective reviews and features were retrieved accordingly for the end user's query.

One of the key contributions of this research study was initiated after a detailed analysis of social media. The collaborative and content-based approaches were individually evaluated. The combined form such as the hybrid approach was reinvented and applied to measure the sentiments of the information. The typical sentiment analysis was conducted to predict and provide a customized recommendation to the requested user.

The other contribution was the ranking ability of the system. In the evolution of RS, the ranking was quite crucial which efficiently projected in this research study. Once the customized recommendation was offered, the ranking ability of the system was tested. Two new approaches had been introduced to carry forward to the ranking process. Firstly, the revised rainbow algorithm (RRA) was used to list the product or any other information based on the similarity in terms of feature excluding the recommended information. Secondly, the linear regression analysis (LRA) which was used to process the listed information through normalization of the product/information attribute and assignment of the weightage to the various
impactful attributes. The appropriate optimization process was done using linear regression derivation where the product attributes and service related information was formalized. The optimized ranking along with the personalized recommendation was offered to the user.

The personalization factor was also preserved in the RS by collecting user personalized information such as user interest, user rating, user feedback, user history, user budget, user profession. Based on these factors, the recommendation was fine-tuned and trustworthy output was facilitated to the end user. Another massive contribution of this research work with comparison with earlier study on the same domain was performance analysis. The recommendation and ranked values were measured using an f-measure technique. In the performance analysis, the measured value f1 had met the actual threshold limit, according to f measure the result value f1 which had floated around 0 to 1 which ranged from worst to the best where \( F_1 \) score had reached its best value at 1 (perfect precision and recall) and worst at 0. The actual f1 score deduced 1.14 which was an optimal one. Thus, the performance of the system was also measured. Unlike the other recommendation system, the proposed system ensured the user personalization, data integrity, data authenticity and reliable ranking along with the recommendation.

The system proven to be an established one since a complete performance analysis and comparative analysis with the existing system were tested and results were processed effectively. This approach of recommendation can satisfy most of the needs of the user to take a wise decision at a crucial juncture. The obvious success factor of this system was directly proportional to the users’ engagement in the system on a regular basis.

7.3 FUTURE ENHANCEMENT OF THE RESEARCH

The proposed research work is narrow down into a selected set of social media data for knowledge extraction, which can be expanded in the
future as multiple selections of social media’s web source irrespective of the data format. This research work is confined for the text mode which can be expanded for multimedia web source. The dataset used in this research work is streamlined; actual nature of the data is changed as per the system need, whereas in the future, the actuality of data needs to be processed without affecting the originality of the data. User-personalized information is also used for the recommendation which is stored statically which can be improved to retrieve the live user data for the future work.

The textual part of the review and features are processed in this research work which can be extended for image, emoticons and other audio, video files too which can bring more interactive based recommendation system in the future. The proposed research work is directed to be applied only for one domain at the time which can be expanded to multiple domains simultaneously. The possible effort is required to develop a recommendation system for multiple domains are difficult, which can be ease down by finding the most fitting techniques or the strategies which take care of the integration process. The integration of different components and functionalities of the various domains are a crucial part of the task which will uplift the different dimension of the overall system on successful implementation.

This research work can also be expanded in the future by applying different strategies and algorithms which might offer an enhanced result. The improvisation and modification can be done at each phase of the proposed research work to add a few more components such as RS for different kinds of age group among the end users can be offered. The strategy used in this research work and PORS framework design can be applied to develop an application based recommendation system for a different set of users using android and ios platform. Thus, the thesis emphasizes its essential need through extensive conclusion remark and also the possible future directions of the research have been discussed in this chapter.