Hybrid Recommender System

(A B2C expert system)

Project Report submitted in partial fulfillment of the requirement for the degree of

Bachelor of Technology

In

Computer Science & Engineering

Under the Supervision of

Ms. Ramanpreet Kaur

By

Akanksha Gupta

Roll No. :111314

To

Jaypee University of Information and Technology

Waknaghat, Solan – 173234, Himachal Pradesh

Certificate

This is to certify that project report entitled "**Hybrid Recommender System**", submitted by Akanksha Gupta (111314) in partial fulfillment for the award of degree of Bachelor of Technology in Computer Science & Engineering to Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision.

This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Date:

Ms. Ramanpreet Kaur
Assistant Professor
Department of Computer Science & Engineering
Jaypee University of Information Technology

Acknowledgement

First of all I would like to thank my supervisor Ms. Ramanpreet Kaur for helping me in all queries that occurred during the tenure of the project. It was her support and stimulating suggestions that helped in achieving the goal as well as her encouragement to maintain the progress in track.

I would also like to appreciate the guidance given by other few other faculty members that has also contributed in the successful implementation of my project.

In addition, I want to express gratitude to all my fellow students and friends who always lend a sympathetic ear. I am grateful for their moral support, which was an essential motivation for keep going with my project.

Date:

Akanksha Gupta
111314
Computer Science & Engineering

Abstract

Today, the concept of recommender systems have become a fundamental application in electronic commerce and information access, providing suggestions that effectively prune large information spaces so that users are directed toward those items that best meet their needs and preferences. . This has helped users in finding items that they would like to buy or consider based on huge amounts of data present on the websites. Recommender Systems have evolved to fulfill the natural dual need of buyers and sellers by automating the generation of recommendations based on data analysis. A variety of techniques have been proposed for performing recommendation, including content-based, collaborative, knowledge-based and other techniques. These recommendations techniques are divided so based on varying inputs to the system, such as items popular on the company's Website; user characteristics such as geographical location or product preference or other personal information; or past buying behavior of top customers. Since each of these techniques has its own pros and cons, thus so as to improve performance level and reduce the chances of flaws, these methods are generally combined and used which are known as hybrid recommenders. As my final year project, I also intend to implement one such model of a hybrid recommender system which will combine the knowledge based and collaborative filtering techniques for a shopping website where in any customer according to his needs, taste and physical appearance shall get a filtered list of available options.

The project implementation involves some of the concepts of Artificial Intelligence and also requires maintaining a huge database of dresses along with brief description of each dress that shall contribute in better recommendation level.

In short, this site will behave like a virtual fashion designer for the customers who will interact with the customers through forms and depending on the backend computation, shall try to recommend best options out of the whole lot of dresses.

Table of Contents

List of Figures

List of Tables

Chapter 1
INTRODUCTION

Today, almost all the e-commerce sites offer millions of products every day from which a customer can choose as per his/ her needs and desire. Selecting the best among so many equally good options thus turns out to be a really challenging task for the customers. In such a situation, the researchers came up with a concept which today forms the base for almost all the e-commerce websites i.e. the idea of recommender systems. These recommender systems are also an outcome of the intense research and hard work done in combining the principles and formulations of various computer disciplines. Applications have been pursued in diverse domains ranging from recommending web pages to music, books, movies and other consumer products.

1.1 Overview

Initial formulations for recommender systems were based on straightforward correlation statistics and predictive modeling, not engaging the wider range of practices in statistics and machine learning literature. But as there was burgeoning consumerism buoyed by the emergence of the web, buyers were being presented with an increasing range of choices while sellers are being faced with the challenge of personalizing their advertising efforts so that neither the customers feel burdened by their products nor the customers feel deprived of relevant items.

Computers are changing our lives and emerging technologies and techniques are paving the way for this rapid transformation. One such rapid transformation was in the e-commerce sector due to concepts of artificial intelligence.

Artificial intelligence came up as the branch of computer science concerned with making computers behave like humans. This discipline aims to understand the nature of human intelligence through the construction of computer programs. These programs are so written that they are able to imitate intelligent behavior. The techniques of artificial intelligence are now-a-days successfully used in most of the areas of science, engineering, education, business, etc. Knowledge engineering and Machine Learning are a core part of AI research.

Going two decades back, such a concept of e-commerce was unknown to many as this field was like a small seed which was not fully developed. E-commerce was just the use of computing and communication technologies in commerce between some or all parts of a business and its customers. This was the name given to the business which grew over Internet and "from home-at home" services were provided to the customers. Due to several factors, people were a bit reluctant to shop online. But as the technology improved and the issues regarding security and privacy were overcome, this sector became dominant player in the retail sector and thus today rules about 60% of the retail market. As the field of e-commerce expanded, it was divide as B2C (Business to Consumer) e-commerce and B2B (Business to Business) e-commerce. B2B e-commerce has around 80% share of total e-commerce market, and B2C captures the rest. In B2C e-commerce, AI is used primarily for product selection and recommendation, negotiation, auctions, etc.

One factor that heavily contributed in growth in e-commerce sector was the extensive implementation of the AI techniques into the development of the expert systems namely B2C and B2B e-commerce systems. This combination of AI and e-commerce has helped companies and organizations carry out a variety of operations and transactions. Through this combination, organizations have been able to simulate skills that are considered innate in people like the capability to memorize the information and reuse it in other transactions. It has also helped in automation or "computerization" of the tasks.

Coming to the concept of recommendations, it is natural human behavior to depend on "word-of-mouth". Customers need a system to advice to them about items they might wish to purchase or examine. The past ratings for a product play a very important role in setting the initial inclination level of the customer. Also, in the present scenario when there is a bulk of options to satisfy our need, finding the perfect product out of all can be very time consuming. The users expect a solution which can help them navigate through large information spaces of product descriptions. Here is when the need of a recommendation system was felt. Recommendations thus are a particular form of information filtering, which exploits past behaviors and user similarities to generate a list of information items that is personally tailored to an end-user's preferences.

The features basic recommendation system has are:

- Assisted and augmented the natural social process of relying on recommendations.
- Addressed the problem of filtering information from burgeon information databases.
- Helped retailers in coping with the rising demand for personalized services in e-commerce.
- Helped Internet stores accelerate profits.

Recommender systems are now popular both commercially and in the research community, where many approaches have been suggested for providing recommendations. In many cases a system designer that wishes to employ a recommendation system must choose between a set of candidate approaches.

Fig 1.1: Applications of Recommender Systems

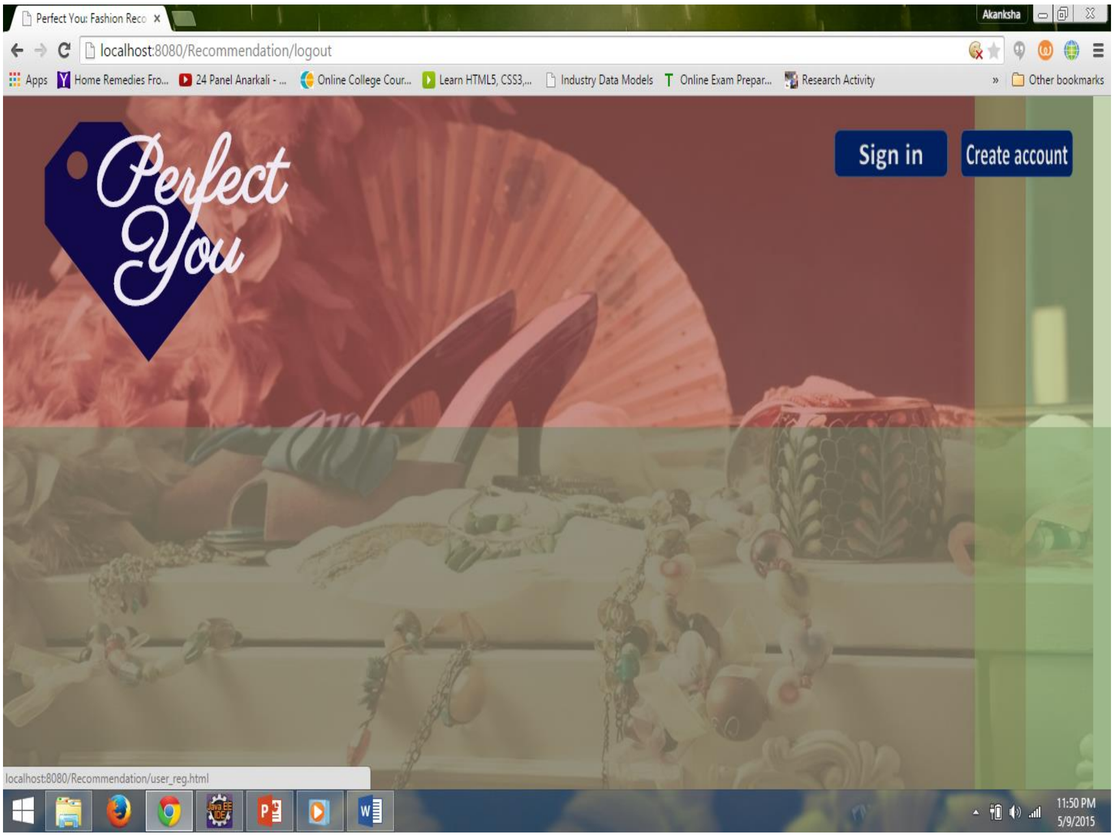
The image above is just a brief overview to summarize the current implementation scenario of recommender systems. This recommendation technology which was initially just an effort to understand human preferences has today become an inseparable part of the e-commerce websites.

1.2 Motivation

The motivation to take up this project came up from my eagerness to learn the concepts of artificial intelligence and getting a deep understanding of recommendation systems that has given a boost to the e-commerce sector not only for purchasing items but also in various other activities like online selling of products and other virtual environment services provided to attract the customers. The reason to take these two techniques for the hybrid model is that they overcome each other’s weaknesses thus making the system highly efficient. Knowledge based recommender system prove to be useful, if there is not a big purchase history which can be analyzed. Collaborative filtering recommender system will be advantageous because of its ability to learn thus adding on to the recommendation capacity of our system. A prototype system of our novel hybrid recommender will be implemented in Java programming language.

1.3 Objective

With the advancement in technology, the personalized recommendation has become the core technology for E-commerce online video and other Internet applications also. In this project, I aim to design a website “Perfect You- a fashion.com” that uses a hybrid model for recommendations. Every technique has its own way of predicting the user preference for a new item based on existing users’ data and their current inclination.
This website will be a logical agent- a virtual fashion designer who would suggest the customers’ with the clothes according to the user requirement and her physical appearance from the whole database of hundreds of dresses.



Snapshot: Website Layout

Through this project, I aim to design a recommender system which uses the fundamentals of both knowledge based technique and collaborative filtering technique. It shall take user input as the key to initiate the recommendation and this input will be through a customized form which will focus on the key factors contributing to the appearance of an individual.

This form will then direct the input to the backend java classes which after performing the required algorithm on them will return the recommendations to the user. I have also tried to display a level comment that shall tell the user in which category the recommended dress falls i.e. Best, Good or Average.

1.4 Approach

As mentioned above, the recommender system will have the features of both-knowledge based and collaborative filtering technique. The core functionality of the knowledge-based systems is built with the help of special-purpose programming environment known as rule engines or inference engines which with the help of rules fed into them through programming will help the customers in choosing the best for them out of the whole lot present. Other parts of the application– like the Web interface– are however built with standard technology i.e. Java which are the means of providing interface between the user and the actual algorithm running at the back end. These interfaces are a source of data exchange as they take user requirements as input and output the relevant recommendations based on the computations performed on the queries and the knowledge base present with the application. The heart of collaborative filtering technique lies in the ratings to the items provided by customers. These ratings will help the system in generating a view regarding customers taste and preference. Knowledge-based and collaborative-filtering recommender systems when combined into a hybrid model ease the burden of electronic commerce websites by helping users find appropriate products from large catalogs.

The reason to choose these two techniques for the hybrid model lies in the fact that having a huge purchase history for every dress and customer is practically not always true. Also the issues of cold start and static behavior of any recommendation which are the drawbacks of collaborative method and knowledge based method need to be overcome so as to make the system always useful in every scenario i.e. for either a new customer or a new entry of a product (dress in our case).

Chapter 2

AI AND E-COMMERCE

2.1 Introduction

Artificial Intelligence is a discipline that aims to understand the nature of human intelligence through the construction of computer programs that try to imitate intelligent behavior. AI techniques are successfully developed and used in most of the areas of science, engineering, education, business, etc.

E-commerce is the use of computing and communication technologies in commerce between some or all parts of a business and its customers. AI techniques are extensively used in the development of e-commerce systems also. The field of e-commerce can be classified as B2C e-commerce and B2B e-commerce, in terms of AI techniques involved in this field. In this paper, we present some important AI techniques that are useful in the design and development of e-commerce systems.

Table 2.1: Comparison of B2B and B2C e-commerce

B2B Marketing	B2C Marketing
✓ Relationship driven	✓ Product driven
✓ Maximize the value of the relationship	✓ Maximize the value of the transaction
✓ Small, focused target market	✓ Large target market
✓ Multi-step buying process, longer sales cycle	✓ Single step buying process, shorter sales cycle
✓ Brand identity created on personal relationship	✓ Brand identity created through repetition and imagery
✓ Educational and awareness building activities	✓ Merchandising and point of purchase activities
✓ Rational buying decision based on business value	✓ Emotional buying decision based on status, desire, or price

2.2 AI in B2C e-commerce

Electronic commerce is the buying and selling of goods and services over Internet. These services can include auctions, negotiations, and contracts, brokering, promotions, advertising, and many more. The artificial Intelligence techniques, abbreviated as AI techniques can be incorporated into e-commerce sector in many ways. The main categorization includes three groups: AI techniques in B2C (Business to Customer) e-commerce, AI techniques in B2B (Business to Business) e-commerce and third category as a mix of both i.e. AI techniques in both B2B and B2C e-commerce. AI approaches have proven to be useful in the development and implementation of B2C and B2B e-commerce systems. B2B e-commerce captures around 80% share of total e-commerce market, and B2C the rest. In B2C e-commerce, AI is used primarily for product selection and recommendation, negotiation, auctions, solving real-world scheduling problems and enhancing servers' scalability, generating automated responses, and decisions on bundling and pricing of goods, etc. In B2B e-commerce, AI is used mainly for supply chain management**.**

2.2.1 Product selection and Recommendation

AI has been extensively used at the backend of e-commerce websites to advise the users on the items they can to examine or would definitely like to purchase through the Internet. This kind of advice is analogous to the ones provided by the salesperson and is extremely necessary because there are no real persons to advice the customers in the Internet. This advice proves to be helpful in navigating a large range of product catalogs. There methodologies implemented at the backend so as to provide the best of product selection and recommendation advices vary from situation to situation and organization to organization. The basic recommendation techniques include Knowledge based recommendation, collaborative filtering recommendation, content based recommendation and demographic based recommendation. These techniques can also be used and implemented in a hybrid model.

2.2.2 Online Negotiation

Negotiation is a situation which comes into play when a user is willing to purchase a product but due to some mismatch of interests, the deal comes on hold for some time.

Negotiation is a process which is done with the aim to benefit both the buyer and seller, in which both the parties bargain on parameters like product price, features etc. User can specify product attribute values, constraints between the attributes, negotiation strategic rules, etc., for negotiation. In case the customer demands are over specified with, multiple constraints, then finding even a single product that satisfies all of the constraints becomes extremely difficult or even impossible. In such cases the demands or constraint need to be relaxed to some extend eventually. Contrary to this, if the situation turns totally opposite where in the demands are under specified or the constraints aren't properly defined or not at all defined, then the retrieval of large number of products turns out to be of no use. In such case the system needs to impose more number of demands during the course of negotiation.

2.2.3 Online Auctions

In today's world of technology, there are approximately 150 auction websites providing the user to not only buy but also purchase second hand products at ease. Most of these online auctions are common value auctions providing the facility for daily use home products. For example the auction sites for computers, cars, houses, home furniture or any usable electronic item. For such sites, in order to evaluate any product accurately, the user needs to find or at least approximate the market price of the item. This is done so that the customer can evaluate the product more correctly. User can bid for any item at the optimal price if he/she is aware of the actual market price or he/she is able to predict the correct or at least the approximate market price of the same item. The information which is gathered from the various websites including many of other auction sites does directly affect the market price of the item. But working with multiple auction sites parallel is an extremely tedious task. This is where the concepts of AI come into action. The proposed AI help customers to gather information and accordingly predict for the optimum bidding price. For such a purpose, a concept involving agents has been found successful. In such a scenario, there are a number of bidder agents involved with a master agent supervising and coordinating their activities. To cope with the large number of websites, different bidder agents are assigned a set of different auction sites. These agents then simultaneously monitor the prices of an item over several bidding websites and then communicate and corporate among themselves through the master agent designed to arrive at an estimated value of the desired item. In real time scenario also, the user approximates the market price of any item based on the information gathered by various agents.

2.2.4 Pricing Goods

Making a system which can intelligently and automatically answer the queries of a customer without the need of a skilled person to do so. Also in order to attract more customers, the system should be able to do dynamic pricing i.e. which goods can be offered in which rate so that the customers get interested in buying and the organization also doesn't suffer any loss. Proper pricing and bundling decisions are vital to expand the customer-base and also to increase profits. Traditionally human beings take these decisions. Automated decision-making ease human beings from this complex task of decision-making.

2.2.5 Solving real world problems and for enhancing server scalability

E-commerce data handling servers should primarily serve the feature of availability and should be capable of solving any real time problem. For example, consider the following situation of travelling domain where finding the flights with the constraints of required schedule, price range, comfort facilities, route taken etc. This is a real time problem, Similar are other online booking sites like that of theatre show booking site or other reservation sites where continuous monitoring is required so as to show availability at regular intervals. In such a situation, the servers need to be scalable so that it can be accessed by any number of users without the issue of system failure or server crash. The algorithms involving CSP (Constraint Solving Problem) are used at backend in such a situation. These techniques have proved to be efficient autonomous problem solvers.

2.3 AI in B2B ecommerce

In successful establishment of B2B markets, SCM have evolved as a key factor thus becoming the key to B2B e-commerce also. As there is a rise witnessed in the e-commerce sector, SCM has gained more importance as more companies are reengineering their methodologies and diverting towards the online markets. “A supply chain is a network of autonomous or semiautonomous business entities responsible for procurement, manufacturing and distribution activities associated with one or more families of related products”. A well-integrated and equipped supply chain helps the businesses to share the real time information and thus significantly reduce the cost overheads to a large extend. This is extremely important for any organization and also for the B2B e-commerce. Again, these SCM techniques involve the concept of agents which are focused on the traditional knowledge involving price, manufacturing and delivery dates, quantity of the products etc.

All the approaches involved in B2B e-commerce further treat these problems as a centralized CSP. The reason behind this lies in the fact that now a days, the communication in any supply chain is not only between the pair of members but between all the members involved in that given supply chain. When considering a centralized CSP, the SCM problem is initially mapped onto a CSP and then a suitable CSP algorithm is decided and used to solve the given problem. There are also systems designed which can utilize agents for the purpose of sub-contacting where the supply chain coordination is transformed into a virtual supply chain involving multiple agents in the process and negotiation taking place between those software agents.

The basis of both the e-commerce methodologies is the exchange of information for which ontologies are also put into use.

2.4 AI in hybrid model

Now a days, a number of e-commerce sites are coming up every day which are being developed, configured and used every day by various organizations for varied business activities. And to serve this purpose, rich format of data needs to be exchanged between the sites in both B2B and B2C e-commerce purposes. In such a case, standardization of the business models, protocols and knowledge structures is very important in achieving the expected and aimed ROI (return on investment) by both the organizations and the customers. The main aim to be achieved by any e-commerce site is to design systems that can share information meaningfully.

Ontologies play a very important role in achieving this goal. It basically stands for the basic terminologies and the meanings between the members. The purpose of ontologies is to enable a hassle free communication between the computer systems and/ or people in such a way that it is independent of the individual system technologies, information architectures and application domain. Ontologies are also vital for the collaboration, agent-to-agent communication, knowledge management, and even for the different database systems to interoperate. Although there are no domain independent ontologies existing as of today, we observe that the following knowledge representation methods are used in constructing ontologies for many domains: *concepts, relations, instances*, and *axioms*. A concept represents a set of entities or ‘things’ of a domain. For example, the set of desktop computers forms a concept in the domain of electronic items. Relations describe the interactions between concepts or a concept's properties. For example, a relation may specify how the prices of desktop computers vary with their processing speed. Instances are specific instantiations of concepts. For example, a desktop computer in a lab is an instance of the computer “concept”. Since an axiom is a conceptualization of the domain, ontologies should not contain any instances. Finally, axioms are used to specify constraints on the values for classes or instances. The upper limit on the processing speed of a desktop computer is an example for axioms.

One of the basic aims of AI is to build systems that can mimic the behavior of human beings. By adding components, an e-commerce system should behave more “natural” to its users. For example, recommender systems. Like any other technology, AI should also enhance the efficiency/performance of an e-commerce system. Efficiency is in terms of producing fast and elegant solutions and consumption of less system resources. Development of domain independent ontologies and/or development of ontologies for complex domains are very important for successful communications with e-commerce systems. As a result, any AI research that can enhance this dimension is useful. In conclusion, naturalist, efficiency, and providing better interface (i.e., better interaction facilities) are some important dimensions that can be used in assessing and evaluating future AI applications to e -commerce.

Chapter 3

RECOMMENDER SYSTEMS

3.1 Introduction

These systems are built so as to support the user in finding and selecting products, services or information when there are too many items to consider or the user has a lack of knowledge about the topic or domain. These are the systems that “have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options”. These engines are a platform that seeks to predict the ‘rating’ and/or ‘preference’ that user would give to an item. It uses knowledge about the user and products to give suitable recommendations. This system has many approaches and every approach uses a different data source.

Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user[4]. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read. “Item” is the general term used to denote what the system recommends to users. These recommender systems normally focuses on a specific type of item (e.g., CDs, or news) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item. RSs are primarily directed towards individuals who lack sufficient personal experience or competence to evaluate the potentially overwhelming number of alter

3.2 Classification

Fig. 1 gives an overview of the recommendation process, the variety of inputs fed into the system and the computations done in the black box just so as to provide the best of recommendations to the customers. Different inputs divide the recommender systems into various categories.

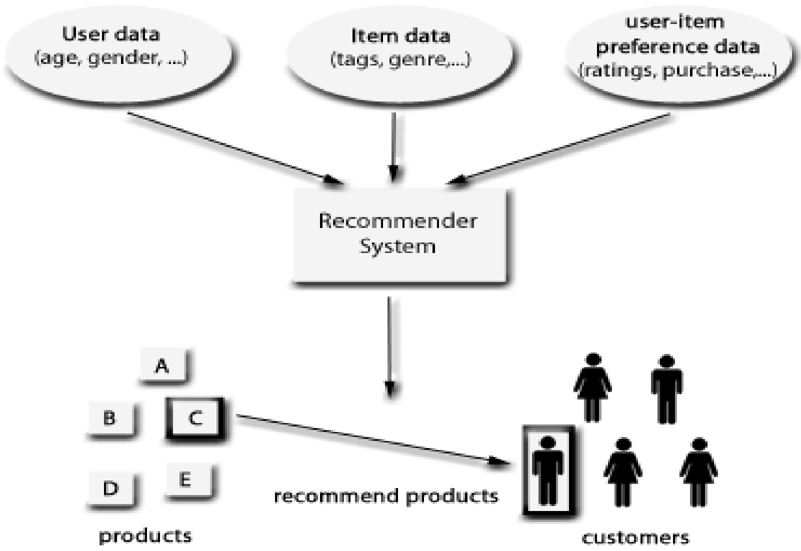


Fig 3.1: Recommendation Process

Thus, the recommender systems can be categorized as follows:

3.2.1 Collaborative Filtering: User Ratings

Collaborative filtering algorithm works according to the opinion of other customers. It uses the similarities between users and items to recommend next items to the users. The algorithm basically calculates the similarity between users. If the similarity is high, it will recommend the items which one user bought but the other hasn't. It assumes that those individuals who agreed in the past tend to agree again in the future.

Often a large amount of dataset is required for an appreciable efficiency of this algorithm[3]. This recommendation technique basically exploits the very common human nature whereby there are many individuals with common opinion many a times. The advantages of this method are that the system does not need to know anything about the items itself; it is only concerned about its preference inclination for a customer. Since the system every time takes into account the ratings provided by the customer for a particular item, this makes the system dynamic hence popular for implementation.

There are many companies having designed their own recommendation system to support their Web applications, such as the Google news recommendation, Friends of Friends system of Facebook and the music recommender of Yahoo!, etc. Amazon, Facebook, LinkedIn, and other commercial and social networking websites use these systems. Parsing a huge amount of data to predict a user’s preference or his or her similarity with other group of users is the core of a recommender system. Among all these typical systems, it has been noted that the collaborative filtering (CF) is the commonly used recommender technology due to its dynamic capabilities.

Now there still exist some drawbacks which compel the programmers to use this technique in hybrid form.

3.2.2 Content based Filtering: Product Features

In content based filtering, the algorithm recommends items according to product feature similarity. For example: if a person is looking for a raincoat, the system will recommend the customer with few options of umbrellas also as both serve the same purpose of protecting a person from rain[1]. Thus, the recommendations depend on how well defined are the features for a product. Here, the user needs to define what he is interested in as well as every item needs to be labeled by hand what requires a huge effort. But for items where information can be extracted automatically (for example newspaper articles) this method can be very good.

The system generates recommendations from two sources: the features associated with products and the requirement that a user has provided to the system. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on product features.

3.2.3 Knowledge based recommender: User Preference

Knowledge based recommender systems turn out to be beneficial, if there is not a big purchase history which can be analyzed[3]. This kind of recommender system offers a dialog that effectively walks the user down a discrimination tree of product features. It does not have a ramp-up or cold start problem since its recommendations do not depend on a base of user ratings and past preferences. It has a conversational style of approach whereby the system knows about the customer requirements through form-filling approach[1]. The user in this case can specify, modify and even provide explicit feedback. Based on this knowledge, the system generates appropriate recommendations. If no such item exists satisfying all the constraints, products satisfying a maximal set of constraints are showed in a rank wise manner with proper explanations as to why the recommendation was done.

3.2.4 Demographic based Recommender: User Demographics

A demographic recommender provides recommendations based on a demographic profile of the user. Recommended products can be produced for different demographic niches, by combining the ratings of users in those niches. Demographic recommender systems aim to categorize the user based on personal demographic. Demographic techniques form “people-to-people” correlations like collaborative ones, but use different data. The benefit of a demographic approach is that it may not require a history of user ratings of the type needed by collaborative techniques.

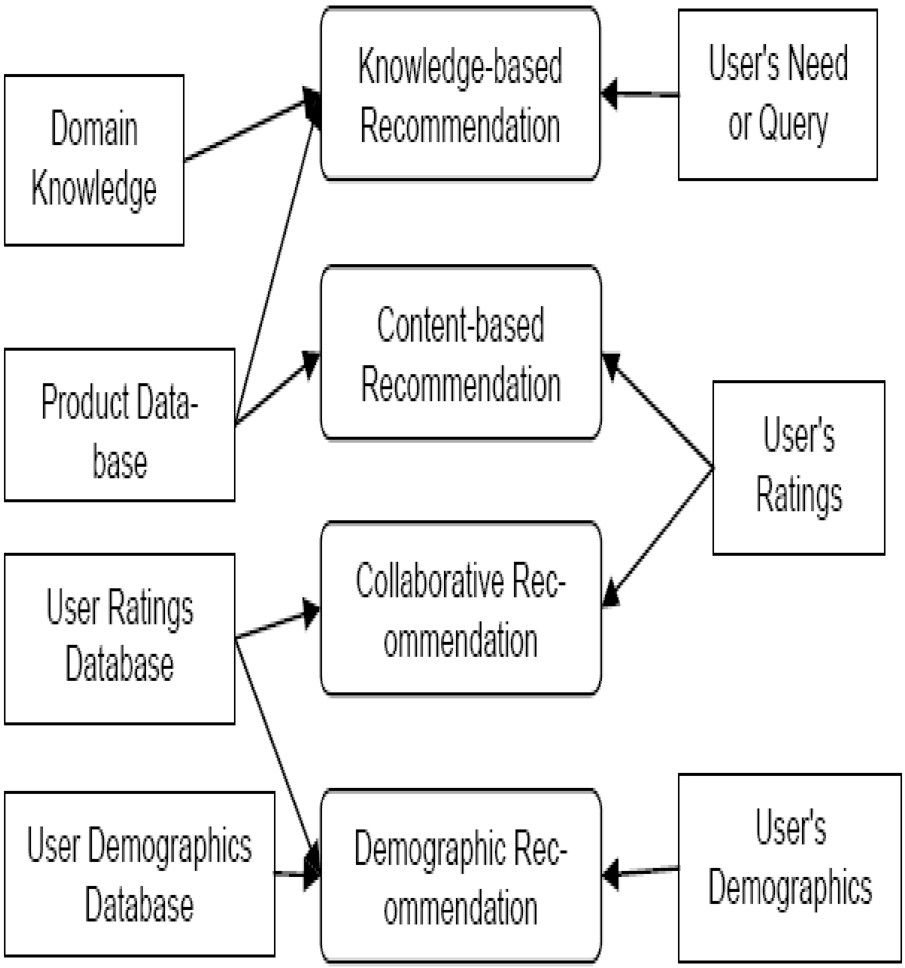


Fig 3.2: Different Recommendation Techniques

The figure above is an overview of the variety of implicit and explicit inputs required by each recommendation technique.

3.3 Comparison

After getting a brief overview of the various recommendation techniques through which suggestions as per the domain and user requirement can be generated, it is very important to be familiar of the[1] various types of problems which become a disadvantage for one or the other recommender system.

(i) Gray Sheep

Gray Sheep problem refers to the problem where by the users whose opinions do not consistently agree or disagree with any group of people and thus do not benefit from collaborative filtering.

(ii) Shilling Attacks

In cases where any user can provide recommendations, people may give a great deal of positive recommendations for their own materials and negative recommendations for their competitors.

(iii) Cold-start

It is difficult to give recommendations to new users as their profile is almost empty and he hasn't rated any items yet so his/her taste is unknown to the system. This is called the cold start problem. Items can also have a cold-start when they are new in the system since they haven't been rated before.

(iv) External Knowledge

Maintaining external databases for items and user preference makes the system more information loaded and the recommendation process complex. Though more knowledge may at times prove to be advantageous as it may suggest more accurate items, but this increases the memory consumption of the application. This external knowledge varies from one technique to another as in Collaborative Filtering, it is Ratings data whereas in Content based technique, it the data pertaining to item features.

Based on the above mentioned some very common issues or problems that can affect the efficiency of the system, the following table summarizes the differences/drawbacks that each recommendation technique when compared to another.

Table 3.1 Comparison between various techniques

	Collaborative filtering	Knowledge based	Content based	Demographic based
Shilling attacks	✓			
Grey sheep problem	✓			
Cold Start	✓			
Domain Knowledge		✓		
Demographic knowledge				✓
Product knowledge			✓	
User’s taste or preference	✓			
Static (No learning capability)		✓	✓	✓

3.4 Hybrid Recommendation Techniques

As we have seen, every method has its advantages and disadvantages. Techniques to be implemented under the hybrid model should be chosen in such a way that they try to overcome the disadvantages of the other and provide a combine list of features as advantages of multi-technique recommender system.

3.4.1 Weighted Hybrid Recommender

A weighted hybrid recommender is a hybrid recommendation technique where in the score of a recommended item is computed from the results of all of the available

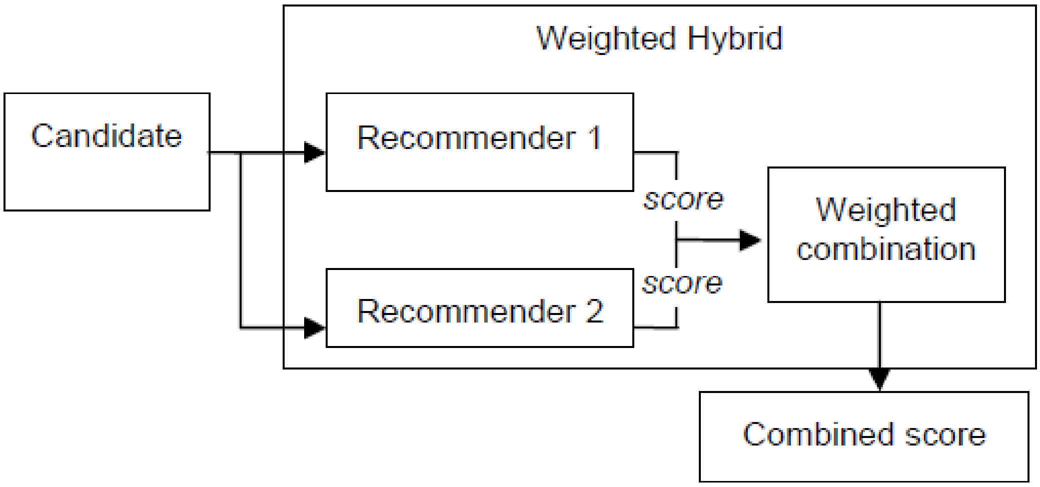


Fig 3.3: Weighted Hybrid Recommender System

recommendation techniques present in the system. For example, the simplest combined hybrid would be a linear combination of recommendation scores. The benefit of such blend is that all of the system's capabilities are brought to bear on the recommendation process in a straightforward way and it is easy to perform post-hoc credit assignment and adjust the hybrid accordingly. However, the implicit assumption in this technique is that the relative value of the different techniques is more or less uniform across the space of possible items. From such discussion above, we can analyze and thereby conclude that a collaborative recommender will be weaker for those items with a small number of raters.

3.4.2 Switching Hybrid Recommender

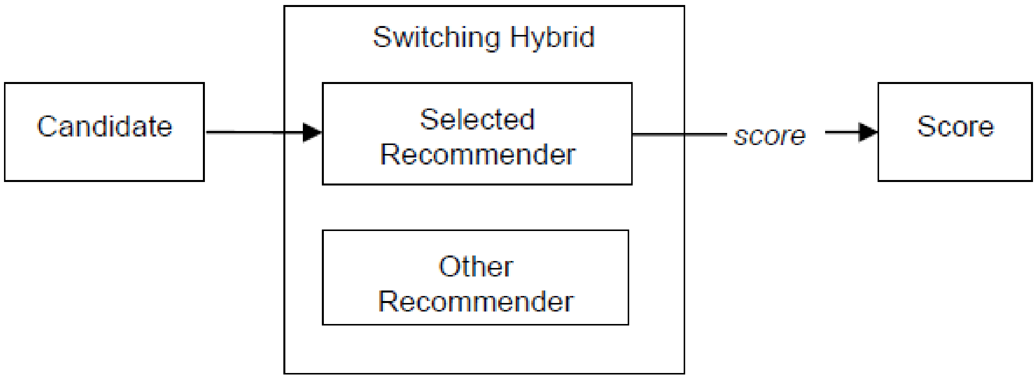


Fig 3.4: Switching Hybrid Recommender System

A switching hybrid technique builds in item-level sensitivity to the hybridization strategy i.e. the system uses some criterion to switch between recommendation techniques. Switching hybrids introduce additional complexity into the recommendation process since the switching criteria must be determined, and this introduces another level of parameterization. However, the benefit is that the system can be sensitive to the strengths and weaknesses of its constituent recommenders.

3.4.3 Mixed Hybrid Recommender

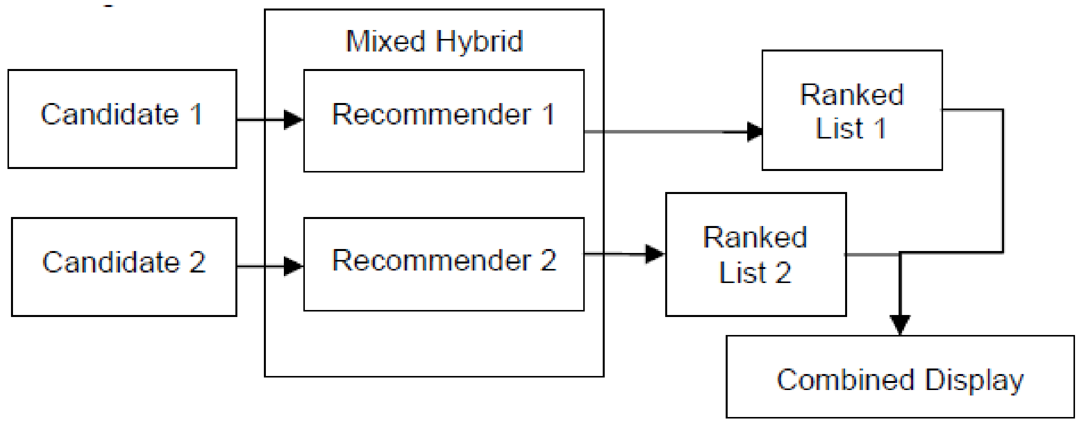


Fig 3.5: Mixed Hybrid Recommender System

Where it is practical to make large number of recommendations simultaneously, it may be possible to use a "mixed" hybrid, where recommendations from more than one technique are presented together. Recommendations from the two techniques are combined together in the final suggested program. The mixed hybrid avoids the "new item" start-up problem: the content-based component can be relied on to recommend new shows on the basis of their descriptions even if they have not been rated by anyone. It does not get around the "new user" start-up problem, since both the content and collaborative methods need some data about user preferences to get off the ground, but if such a system is integrated into a digital television, it can track what shows are watched (and for how long) and build its profiles accordingly. Like the fallback hybrid, this technique has the desirable "niche-finding" property in that it can bring in new items that a strict focus on content would eliminate. Usually, recommendation requires ranking of items or selection of a single best recommendation, at which point some kind of combination technique must be employed.

3.4.4 Feature Combination Hybrid Recommender

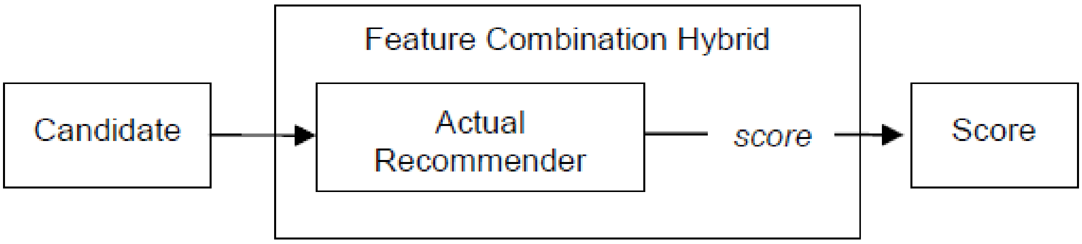


Fig 3.6: Feature Combination Hybrid Recommender System

Another way to achieve the content/collaborative merger is to treat collaborative information as simply additional feature data associated with each example and use Content-based techniques over this augmented data set. The feature combination hybrid lets the system consider collaborative data without relying on it exclusively, so it reduces the sensitivity of the system to the number of users who have rated an item. Conversely, it lets the system have information about the inherent similarity of items that are otherwise opaque to a collaborative system.

3.4.5 Cascade Hybrid Recommender

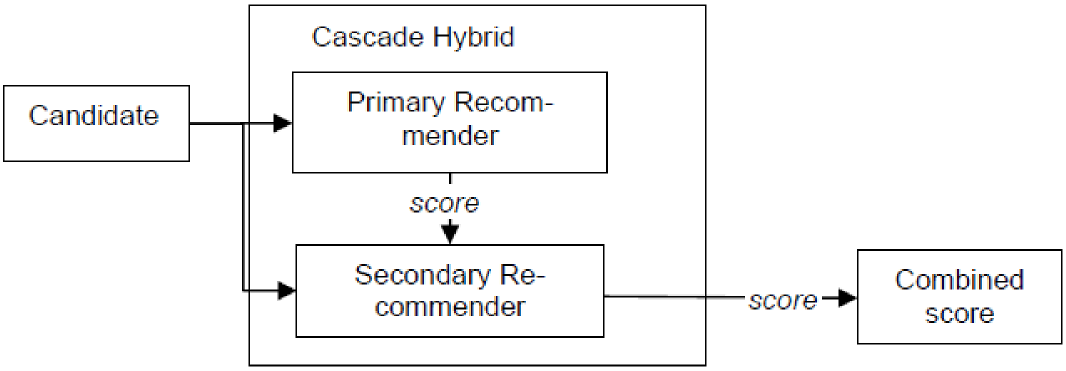


Fig 3.7: Cascade Hybrid Recommender System

Unlike the previous hybridization methods, the cascade hybrid involves a staged process. In this technique, one recommendation technique is employed first to produce a coarse ranking of candidates and a second technique refines the recommendation from among the candidate set. The recommendations are placed in buckets of equal preference, and the collaborative technique is employed to break ties, further ranking the suggestions in each bucket. Cascading allows the system to avoid employing the second, lower-priority, technique on items that are already well-differentiated by the first or that are sufficiently poorly-rated that they will never be recommended. Because the cascade's second step focuses only on those items for which additional discrimination is needed, it is more efficient than, for example, a weighted hybrid that applies all of its techniques to all items. In addition, the cascade is by its nature tolerant of noise in the operation of a low-priority technique, since ratings given by the high-priority recommender can only be refined, not overturned.

3.4.6 Feature Augmentation Hybrid Recommender

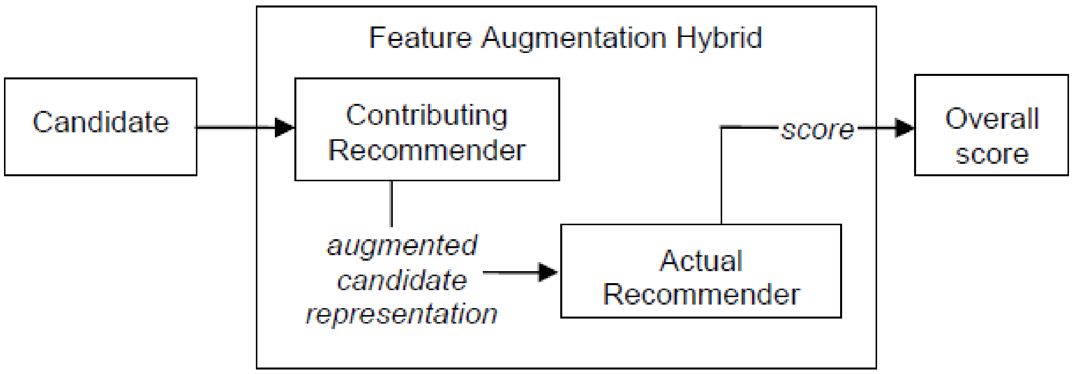


Fig 3.8: Feature Augmentation Hybrid Recommender System

One technique is employed to produce a rating or classification of an item and that information is then incorporated into the processing of the next recommendation technique. Additional functionality is added by intermediaries who can use other techniques to augment the data itself. This is different from feature combination in which raw data from different sources is combined. While both the cascade and augmentation techniques sequence two recommenders, with the first recommender having an influence over the second, they are fundamentally quite different. In an augmentation hybrid, the features used by the second recommender include the output of the first one. In a cascaded hybrid, the second recommender does not use any output from the first recommender in producing its rankings, but the results of the two recommenders are combined in a prioritized manner.

3.4.7 Meta-level Hybrid Recommender

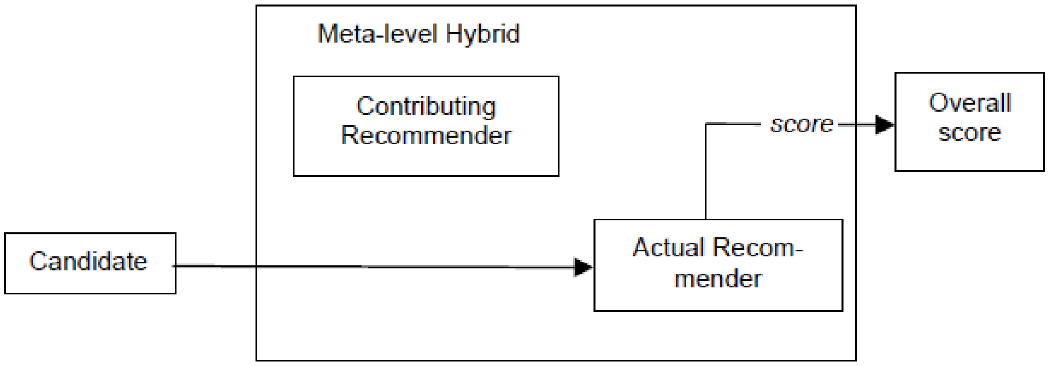


Fig 3.9: Meta Level Hybrid Recommender System

Another way that two recommendation techniques can be combined is by using the model generated by one as the input for another. This differs from feature augmentation: in an augmentation hybrid, we use a learned model to generate features for input to a second algorithm; in a meta-level hybrid, the entire model becomes the input. The benefit of the meta-level method, especially for the content/collaborative hybrid is that the learned model is a compressed representation of a user's interest, and a collaborative mechanism that follows can operate on this information-dense representation more easily than on raw rating data.

The following table summarizes the various hybrid methodologies that can be adopted to implement an efficient autonomous system.

Table 3.2 Various Hybrid Techniques

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another.

Chapter 4

SYSTEM ANALYSIS AND DESIGN

4.1 Use Case Diagram

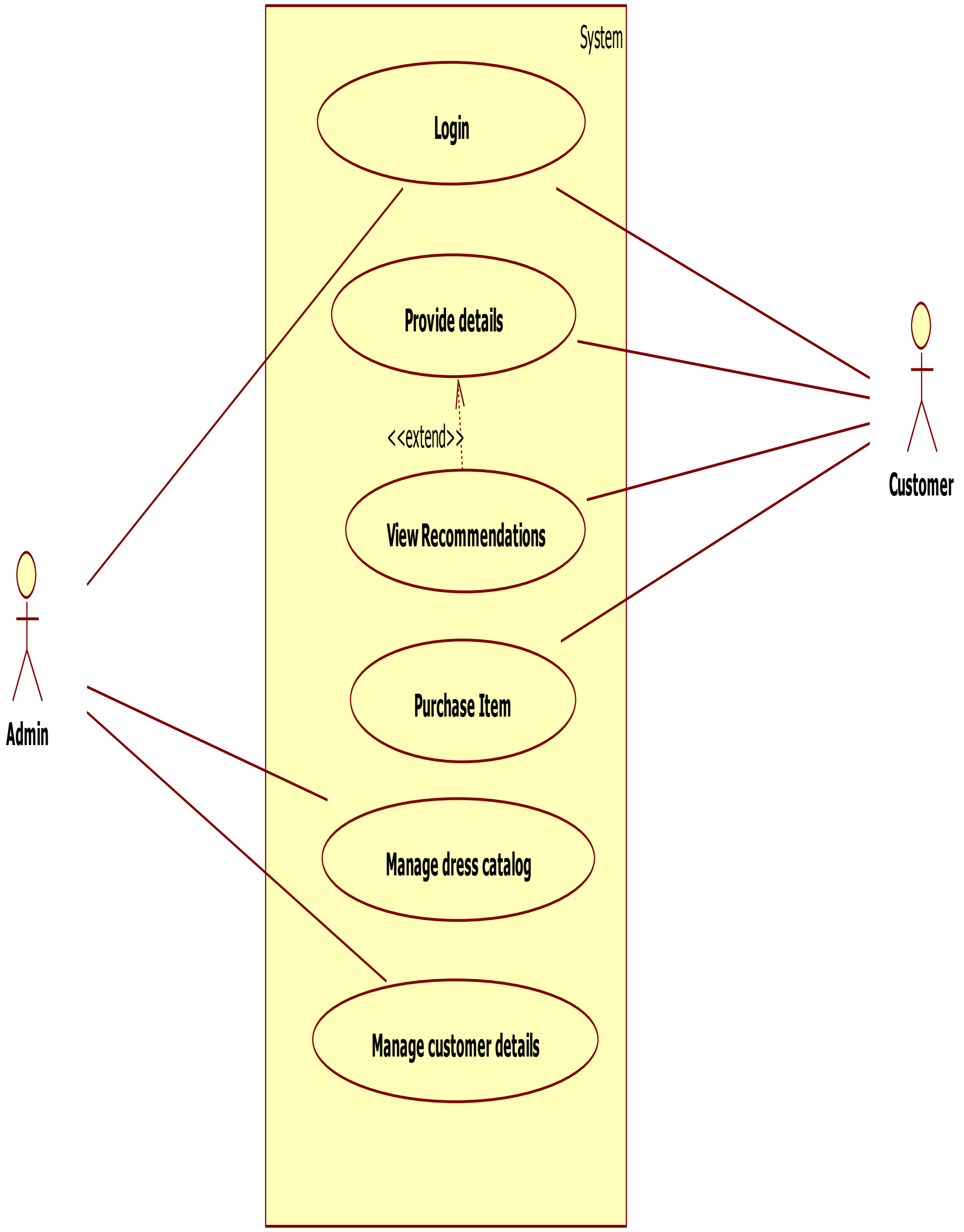


Fig 4.1: Use Case Diagram

4.2 System Architecture

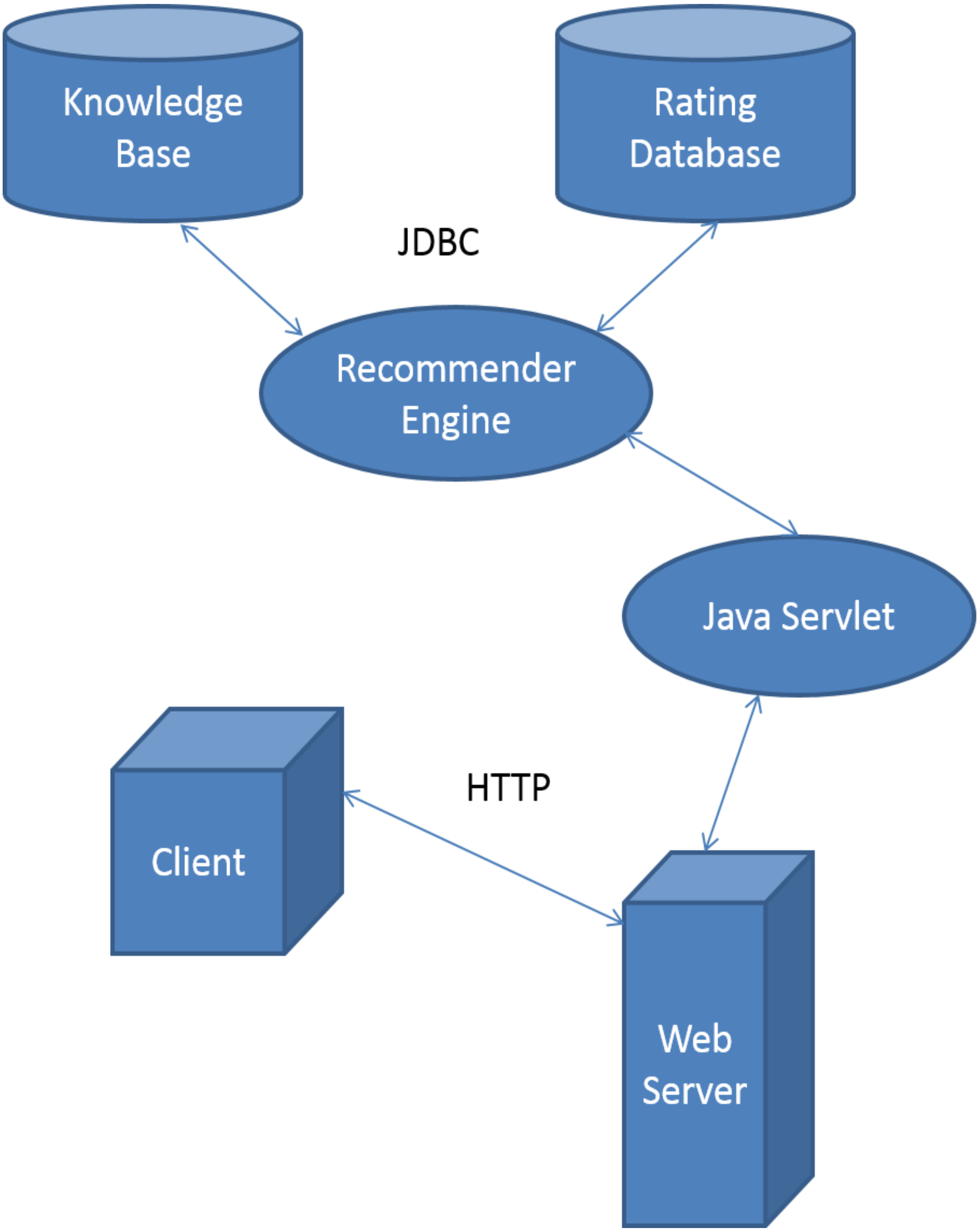


Fig 4.2: System Architecture

4.3 Flow Chart

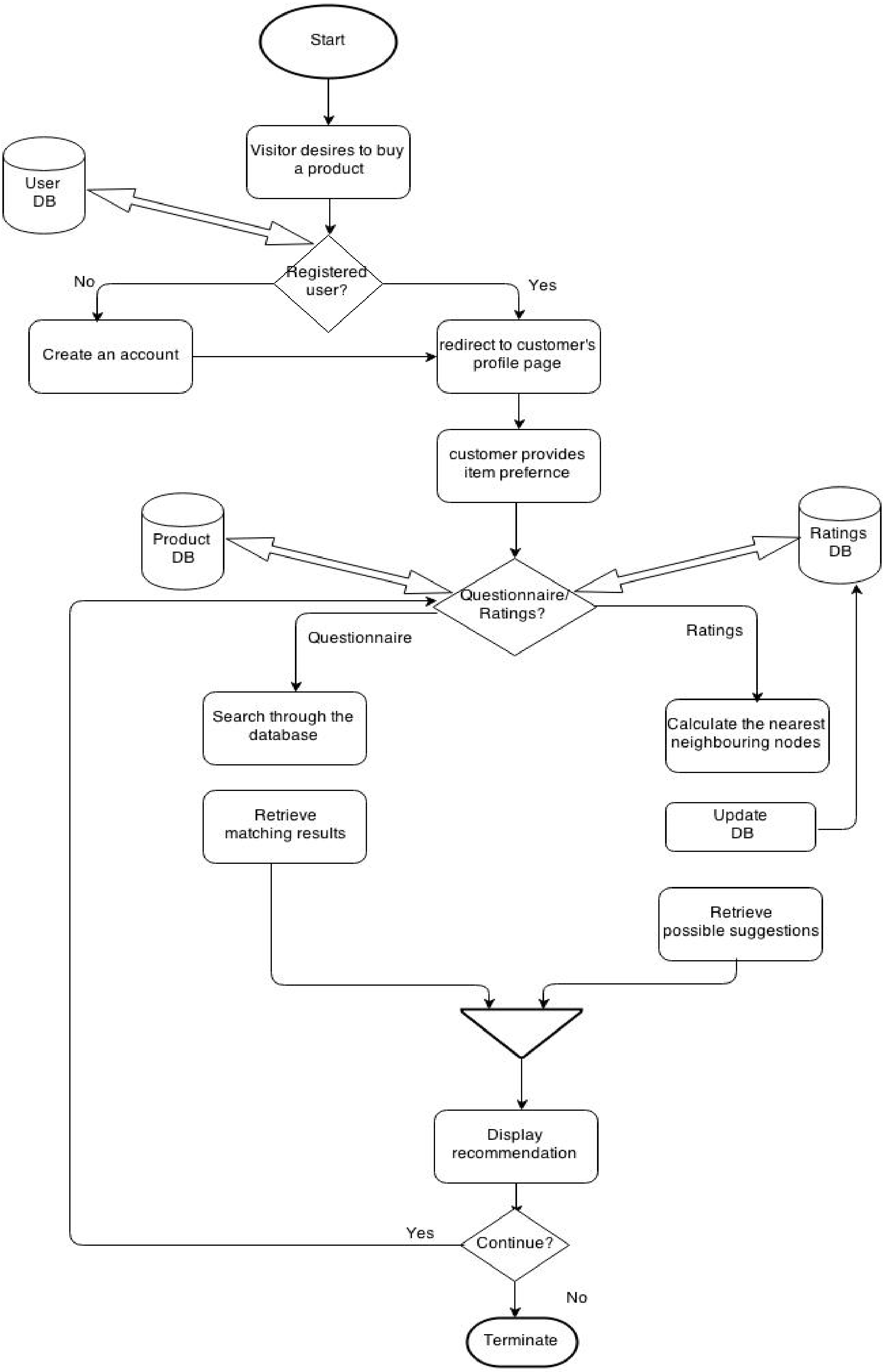


Fig 4.3: Flow Chart

4.4 Data Flow Diagram

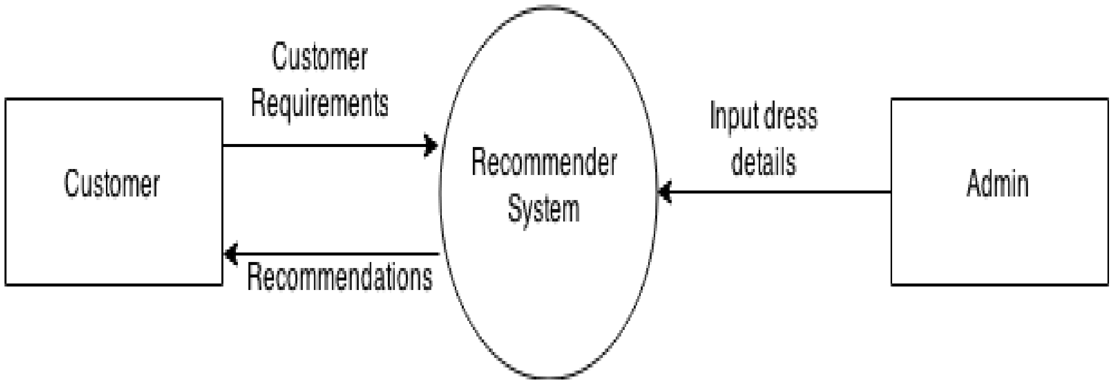


Fig 4.4 Data Flow Diagram(Level 0)

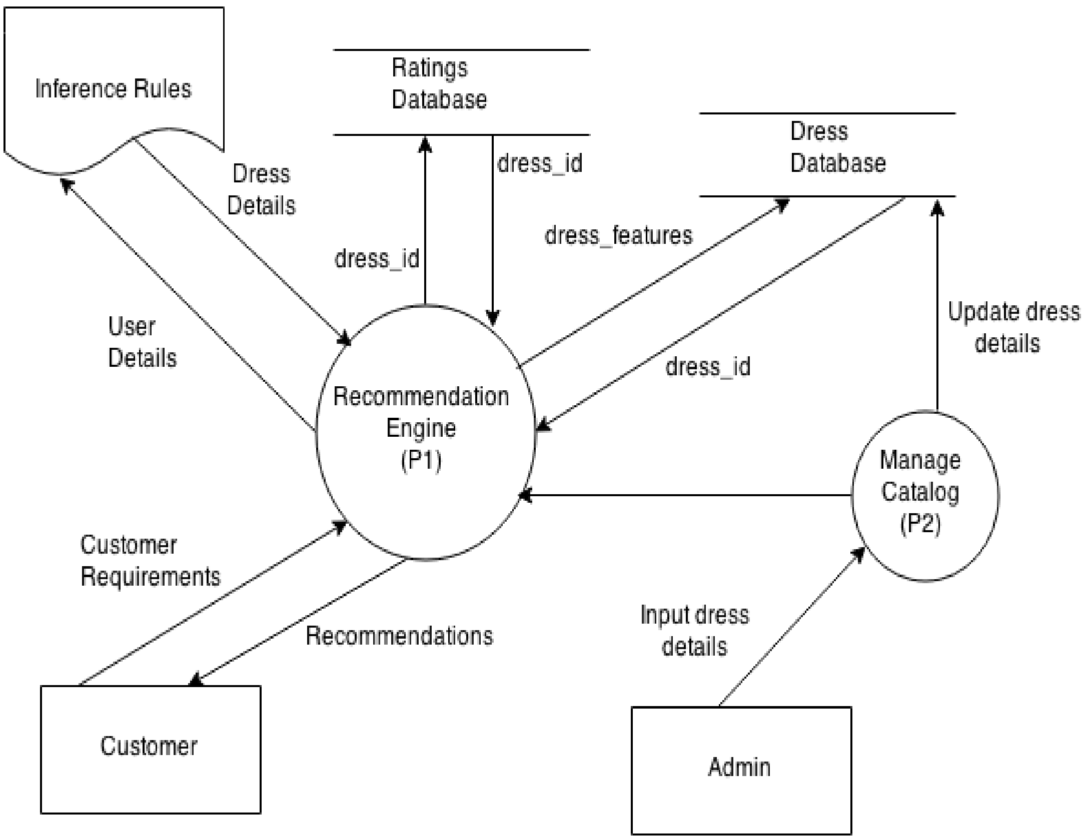


Fig 4.5 Data Flow Diagram(Level 1)

Chapter 5
IMPLEMENTATION

5.1 Algorithm

The recommender system implemented in our project is a blend of two recommendation techniques, i.e. Knowledge based recommendation technique and Collaborative filtering recommendation technique, and hence the implementation also involved two different procedures. These concepts have been discussed in the section below.

5.1.1 Knowledge Based Recommender System

The two key concepts behind the implementation of any knowledge based recommender systems are forward chaining and backward chaining.

(i)Forward-chaining

These rules are somewhat like if… then statements in a procedural language. Here the RHS is executed only when the if.. part is satisfied.

(ii) Backward-chaining

These rules on the other hand, don’t have a clear analogy in procedural programming. They are also similar to if … then statements, but a backward-chaining rule actively tries to satisfy the conditions of its if part.

The inference engine used in the implementation of the recommender system in the project is based on Rete algorithm[5,6]. The Rete algorithm is implemented by building a network of nodes, each of which represents one or more tests found on a rule LHS. Facts that are being added to or removed from the working memory are processed by this network of nodes. At the bottom of the network are nodes representing individual rules. When a set of facts filters all the way down to the bottom of the network, it has passed all the tests on the LHS of a particular rule and this set becomes an *activation*. The associated rule may have its RHS executed (*fired*) if the activation is not invalidated first by the removal of one or more facts from its activation set.

The Rete Algorithm is intended to implement the concept of forward-chained rules. When a query arrives to the root a copy of that token is sent to each "kind" node where a SELECT operation is carried out that selects only the tokens of its kind. When the token(query) arrives to the destined node a PROJECT operation extracts from the token tuple's the components that match the variables of the pattern. The resulting tuple is returned as the output.

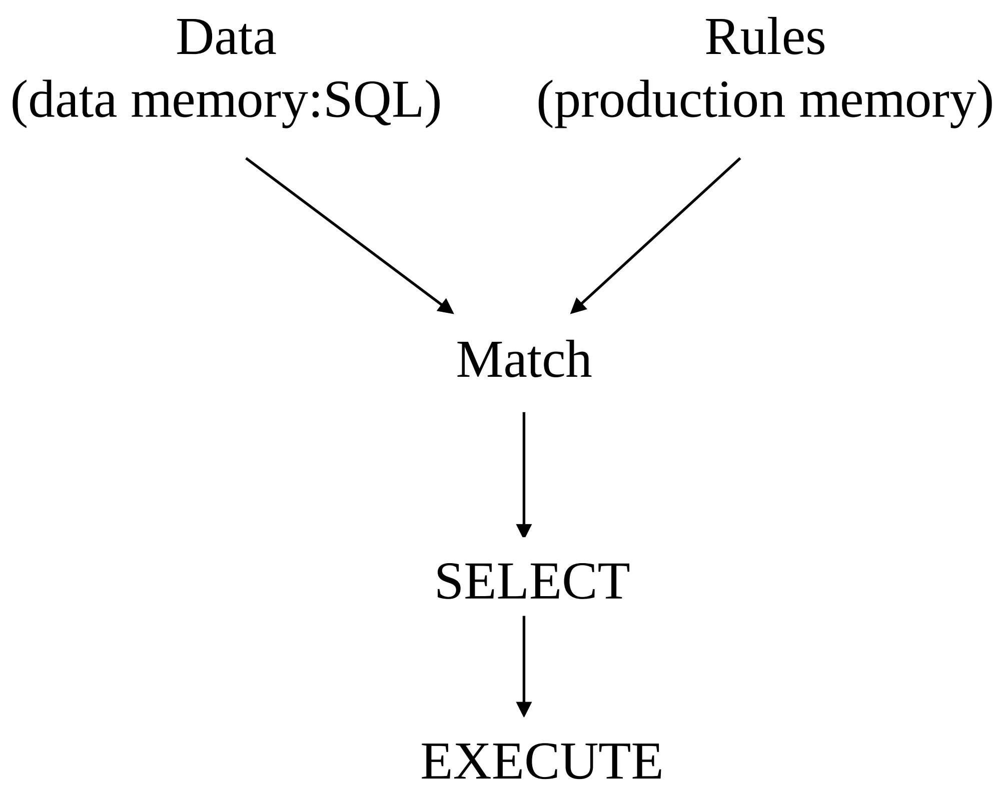


Fig 5.1: Forward Chaining Mechanism

5.1.2 Collaborative Filtering Recommender System

Collaborative filtering algorithm is the most popular recommendation algorithm as it is one such recommendation technique which helps making system dynamic i.e. takes into account user likes and dislikes and then recommends the best suited items. Similar in this project, I have implemented on of the methodologies to generate a similarity index between the items and user likings. This index is known as Pearson Correlation Coefficient which finds the correlation between the items liked by the current user and similar items rated by neighboring users to see which other items can be suggested to the current user based upon his taste or preference. This taste is evaluated based upon the rating a user provides to the items they view or even purchase.

Thus, the pearson correlation coefficient is as mentioned below

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (y)^2]}}$$

Where :

N =number of pairs of scores

$\sum xy$ = Sum of the products of paired scores

$\sum x$ =Sum of x scores

$\sum y$ =Sum of y-scores

$\sum x^2$ =Sum of squared x-scores

$\sum y^2$ =Sum of squared y-scores

When talking about the customer feedback, it can be explicit like as in the case of user ratings, or it can be implicit as when the user purchases the item or at least clicks on so as to view the item out of the whole list of recommendations generated.

In this project also, I have tried to consider both the forms of feedback to a very large extend so as to enhance the efficiency of the system.

5.1.3 Switched Hybrid Recommendation Process

Using switched hybrid technique in this project ensures that when I don’t have a large purchase history of any user, or there are less number of recommendations generated by collaborative filtering technique, then the knowledge based recommender supports the recommendation process by covering up for the deficiency of preceding recommender.

Considering a threshold of minimum 8-10 dresses to recommend to a user for each category, then in such a situation the correlation index generated for the user-item matrix may not result in good number of recommendations to offer to the customer. Such a situation is backed by the recommendations generated by though static but highly precise knowledge based recommender system.

The steps followed the implementation of a KB/CF hybrid recommender are as follows:

1. User A is asked to login with his credentials if she is an existing user else redirected to a page where she can create a new account for herself by entering the personal details and physical specifications.
2. Once the user has logged in, she is asked to specify the type of dress type she desires to purchase end even the occasion for which the purchase is being made.
3. After the requirements are fed through the questionnaire, the classes at backend compute the features suited for that body type, height and skin color.
4. The features extracted by the inference engine (JESS) are then used to retrieve a list of dresses having ratings>0 i.e. the dresses which have a viewing or purchase history.
5. Simultaneously the system retrieves the purchase history of the current user.
6. Based on the above two data tables, Pearson Coefficient Correlation is calculated for every item rated by the user I past to every item rated by any user in the same category.
7. The list is then filtered for all the unique values in arranged in descending order to look for the best suited dresses.
8. If the recommendations produced are less than the threshold number, then based on the features extracted, knowledge based recommender extracts the deficit number of dresses.
9. The user can then view or purchase the product which is liked by the user and the actions are monitored and stored in the database for analysis in future predictions.

5.2 Tree Representation

Now, as the forward chaining mechanism essentially requires production rules so as to select the corresponding data from the data memory , thus for this project the base for production rules is their requirement and physique. The tree representation below shows the division of women clothing I have considered for this project. The women clothing can be of either traditional category which may include kurti, suits or saaris and when the same division is considered for western clothing, it may constitute of gowns, dresses or shorts at a broader level.

Now this tree alone cannot do anything until we have the appearance specifications of the individual for whom the suggestions are to be generated. The physical features taken into account for the implementation of the project are the body type of the individual along with her skin complexion and height.

The features of clothes taken into account for deciding which dress would suit the customer the most cover the cloth fabric and other descriptive features like neck cut, dress length or dress color based on the users appearance input.

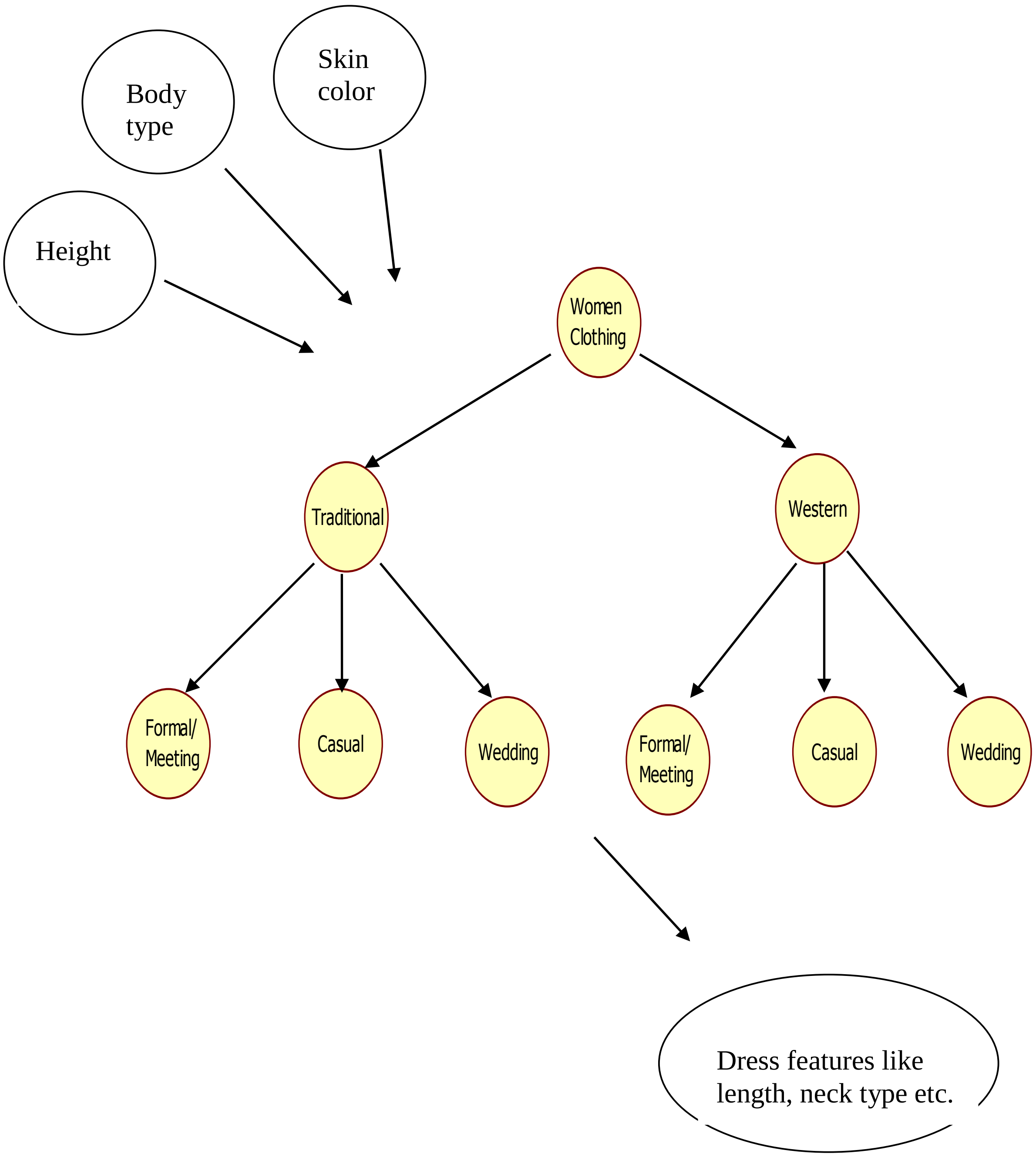


Fig 5.2: Features' Overview

Based on these inputs from the user as per the requirement, the rules have been formed which are maintained in the working memory of the inference engine.

5.3 Tools Used

As this project involves around successful implementation of both knowledge based and collaborative filtering recommender system, thus the various tools which have been the support system of the application include the following:

5.3.1 JESS

In many real-world or commercial applications, the part constituting of the intelligent systems in generally developed with the help of special programming environments such as LISP or Prolog or by use of a rule-engines [5]. During the implementation of the first phase of this project i.e. Knowledge based recommender system, even I have used on such rule engine i.e. JESS (Java Expert System Shell). Jess is a general-purpose rule engine, developed at Sandia National Laboratories. Written in the Java programming language, Jess offers easy integration with other Java-based software.
It is a superset of the CLIPS programming language. The language provides rule-based programming for the automation of an expert system, and is frequently termed as an expert system shell.[1] In recent years, intelligent agent systems have also developed, which depend on a similar capability.
Rather than a procedural paradigm, where a single program has a loop that is activated only one time, the declarative paradigm used by Jess continuously applies a collection of rules to a collection of facts by a process called pattern matching. Rules can modify the collection of facts, or they can execute any Java code. The Jess rules engine utilizes the Rete algorithm. The feature of JESS mostly used during the implementation of rules in the project was the “defrule” functionality. This keyword helps the programmer define rules which automatically get executed when their IF…. Part satisfies.

5.3.2 Java Servlets

The web interface or the database layer is probably developed with the help of various standard technologies. One such very common and stand technology that has been implemented is Java Classes and Servlets.

A Java servlet is a Java programming language program that extends the capabilities of a server. Although servlets can respond to any types of requests, they most commonly implement applications hosted on Web servers. Servlets provide a component-based, platform-independent method for building Web-based applications, without the performance limitations of CGI programs. Servlets have access to the entire family of Java APIs, including the JDBC API to access enterprise databases.

5.3.3 MySQL database

In order to store the whole chunk of customer data and the data of the dresses to be shopped, the most popular open source database has been used i.e., MySQL, the interface to which is bridged through JAVA classes.

5.3.4 Web Technology

The attractive UI design that initiates the process by interacting with the customers through the forms clicks involves the use of most common web technologies i.e. HTML (Hypertext Markup Language) and CSS (Cascading Style Sheet). These interfaces provide the data exchange channel between the customer and the software application.

5.4 Website

5.4.1 Database Structure

The database "recommend" consists of the following tables which play a vital role in recommendation process. The more refined and structured the data is in these table, the simpler the recommendation process will become.

Table 4.1: dress

```
mysql> use recommend;
Database changed
mysql> describe dress;
+--------------+-------------+------+-----+---------+-------+
| Field        | Type        | Null | Key | Default | Extra |
+--------------+-------------+------+-----+---------+-------+
| dress_name   | char(25)    | YES  |     | NULL    |       |
| dress_id     | varchar(20) | NO   | PRI |         |       |
| dress_img    | longblob    | YES  |     | NULL    |       |
| dress_price  | float       | YES  |     | NULL    |       |
| dress_type1  | char(25)    | YES  |     | NULL    |       |
| dress_type2  | char(25)    | YES  |     | NULL    |       |
| dress_cat    | char(25)    | YES  |     | NULL    |       |
| dress_rview  | float       | YES  |     | NULL    |       |
| dress_fabric | char(20)    | YES  |     | NULL    |       |
| dress_color1 | char(20)    | YES  |     | NULL    |       |
| dress_color2 | char(20)    | YES  |     | NULL    |       |
| dress_neck   | char(20)    | YES  |     | NULL    |       |
+--------------+-------------+------+-----+---------+-------+
12 rows in set (0.80 sec)
```

This table stores the complete information of the dresses available for the customer alongwith the images for display to the customer.

Table 4.2 customer

```
mysql> describe customer;
+-------------+-------------+------+-----+---------+-------+
| Field       | Type        | Null | Key | Default | Extra |
+-------------+-------------+------+-----+---------+-------+
| cust_name   | char(25)    | NO   |     | NULL    |       |
| cust_id     | varchar(20) | NO   | PRI | NULL    |       |
| cust_pass   | varchar(20) | NO   |     | NULL    |       |
| cust_btype  | varchar(20) | YES  |     | NULL    |       |
| cust_scolor | char(20)    | YES  |     | NULL    |       |
| cust_height | char(10)    | YES  |     | NULL    |       |
+-------------+-------------+------+-----+---------+-------+
6 rows in set (0.09 sec)
```

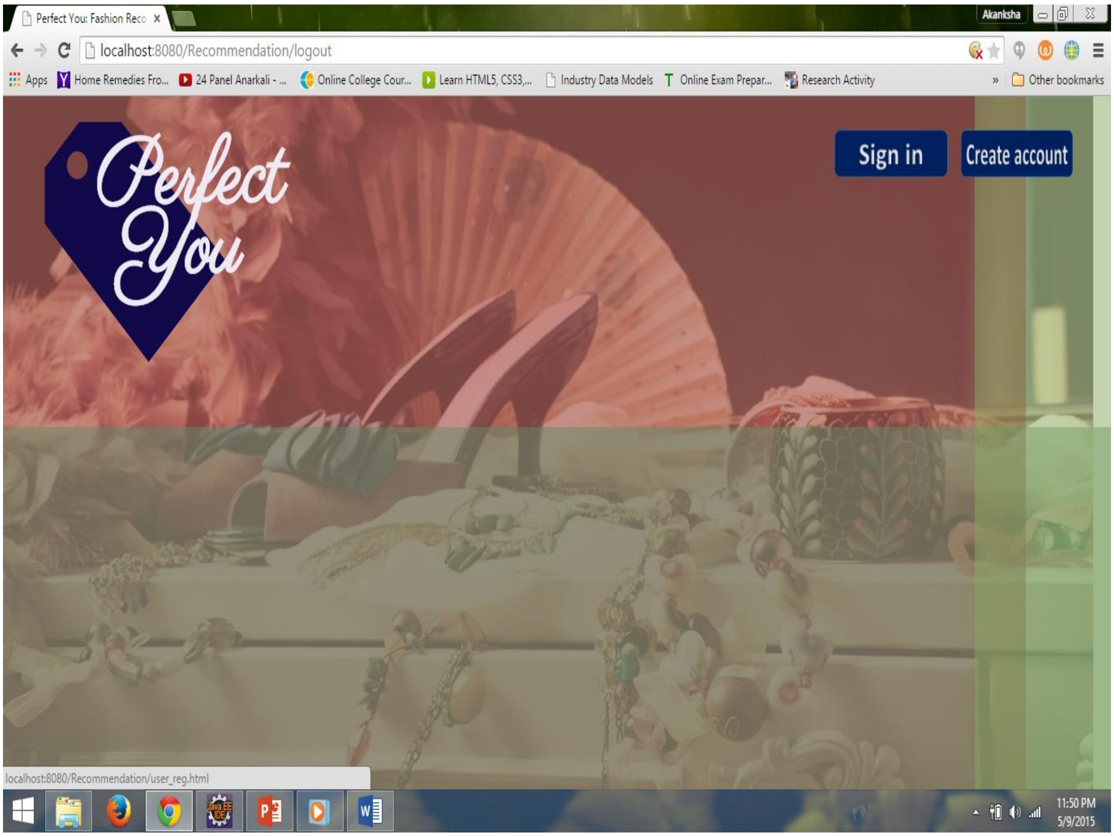
This table stores the complete information of the customers associated with our website.

Table 4.3 ratings

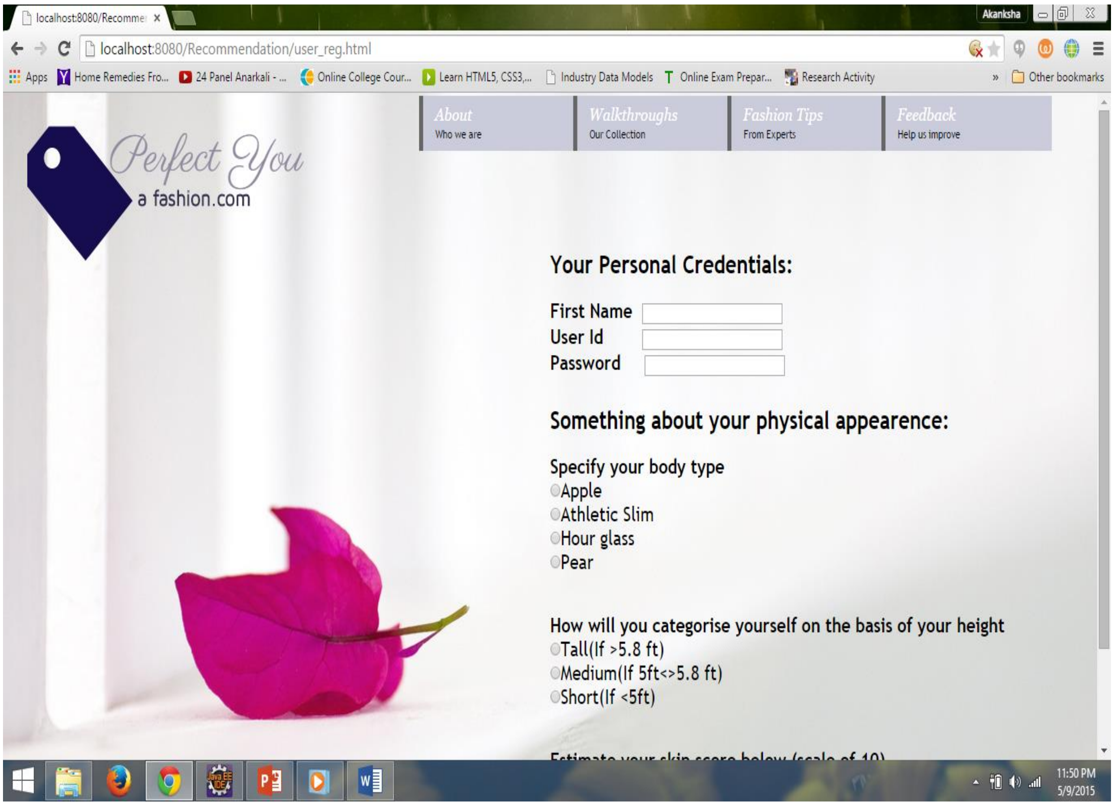
```
mysql> describe ratings;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| cust_id  | varchar(20) | NO   |     |         |       |
| dress_id | varchar(20) | NO   |     |         |       |
| ratings  | float       | YES  |     | NULL    |       |
| phistory | int(11)     | YES  |     | NULL    |       |
+----------+-------------+------+-----+---------+-------+
4 rows in set (0.05 sec)
```

This table stores the complete information of the customer feedback in the form of ratings and purchase history.

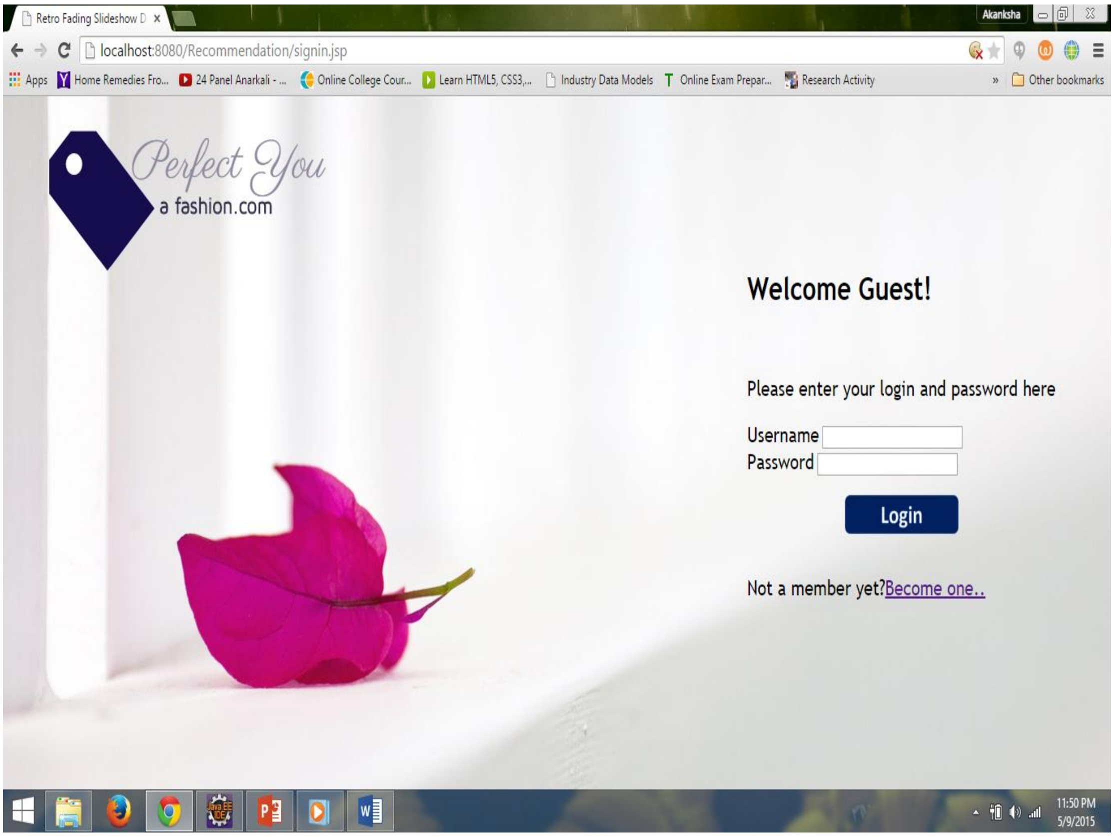
5.4.2 User Interface



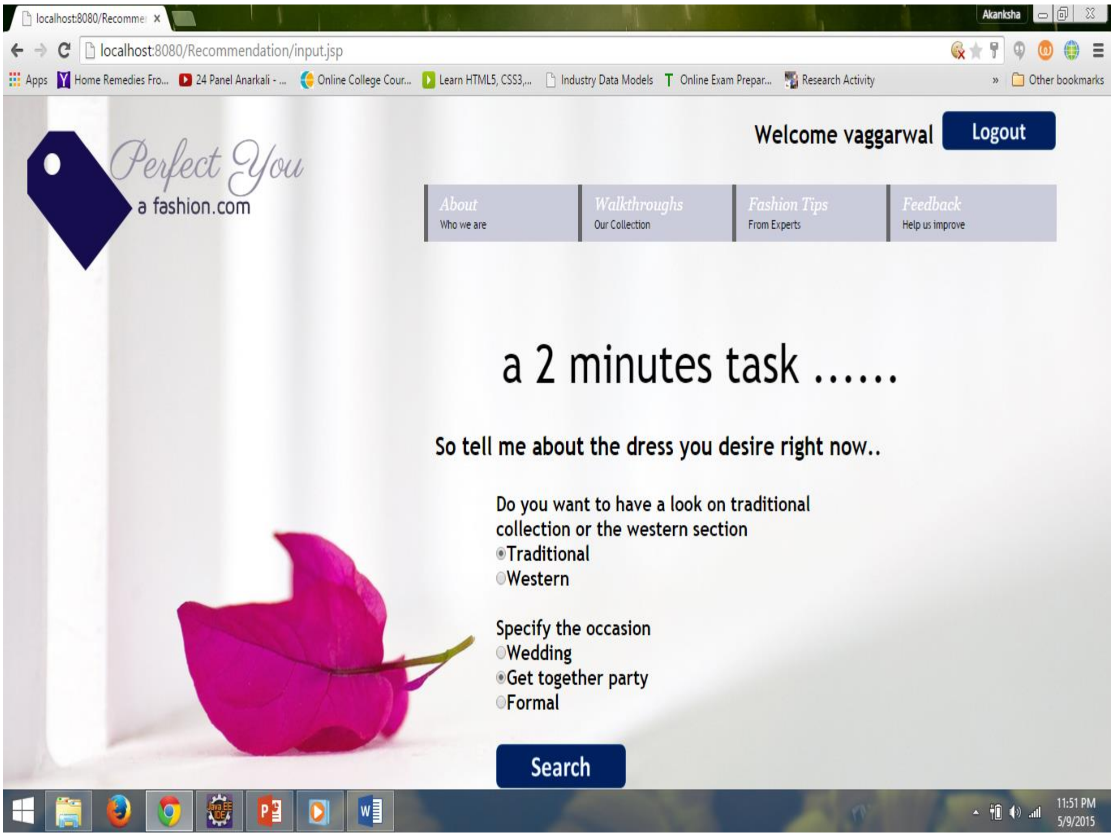
Snapshot: Home Page



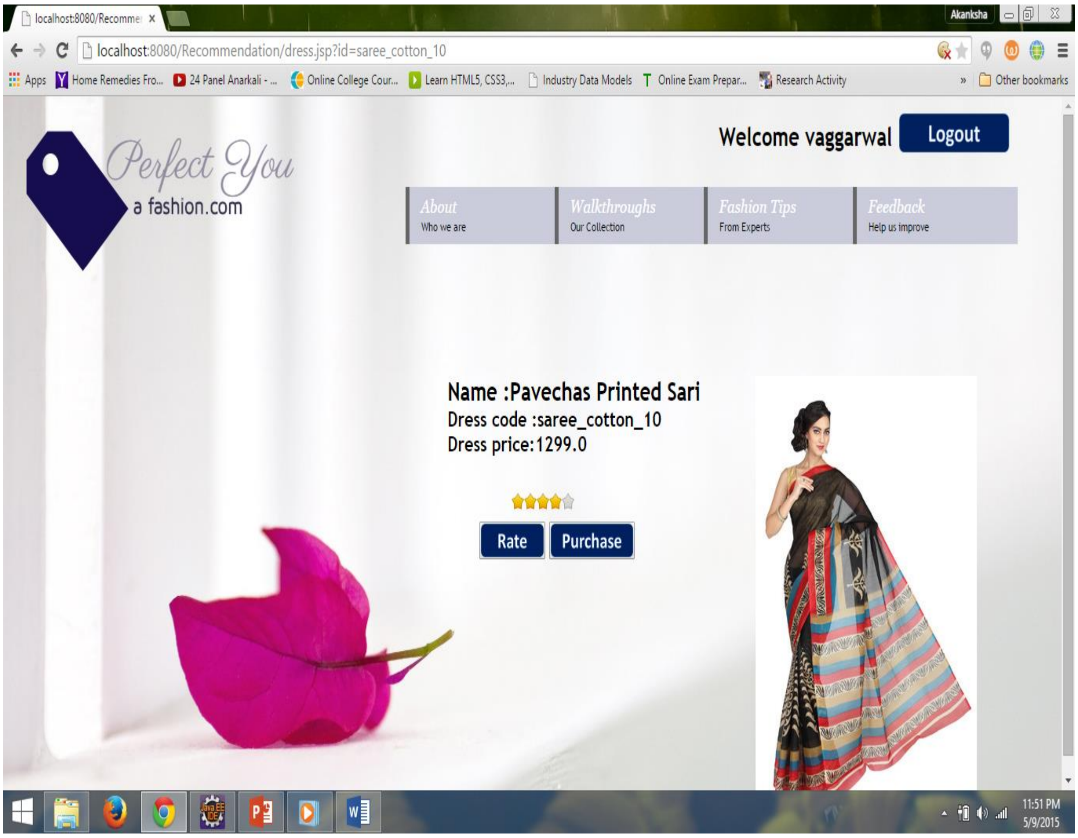
Snapshot: New User Login



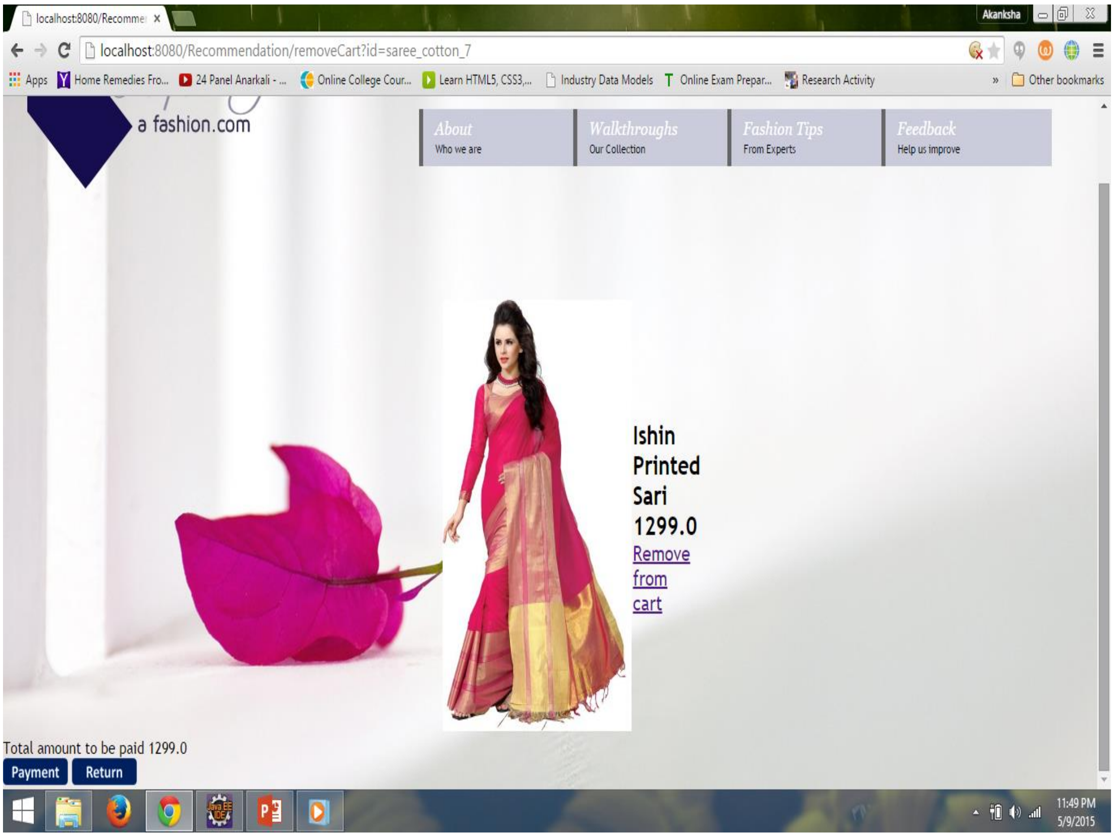
Snapshot: Registered User Login



Snapshot: Purchase Questionnaire

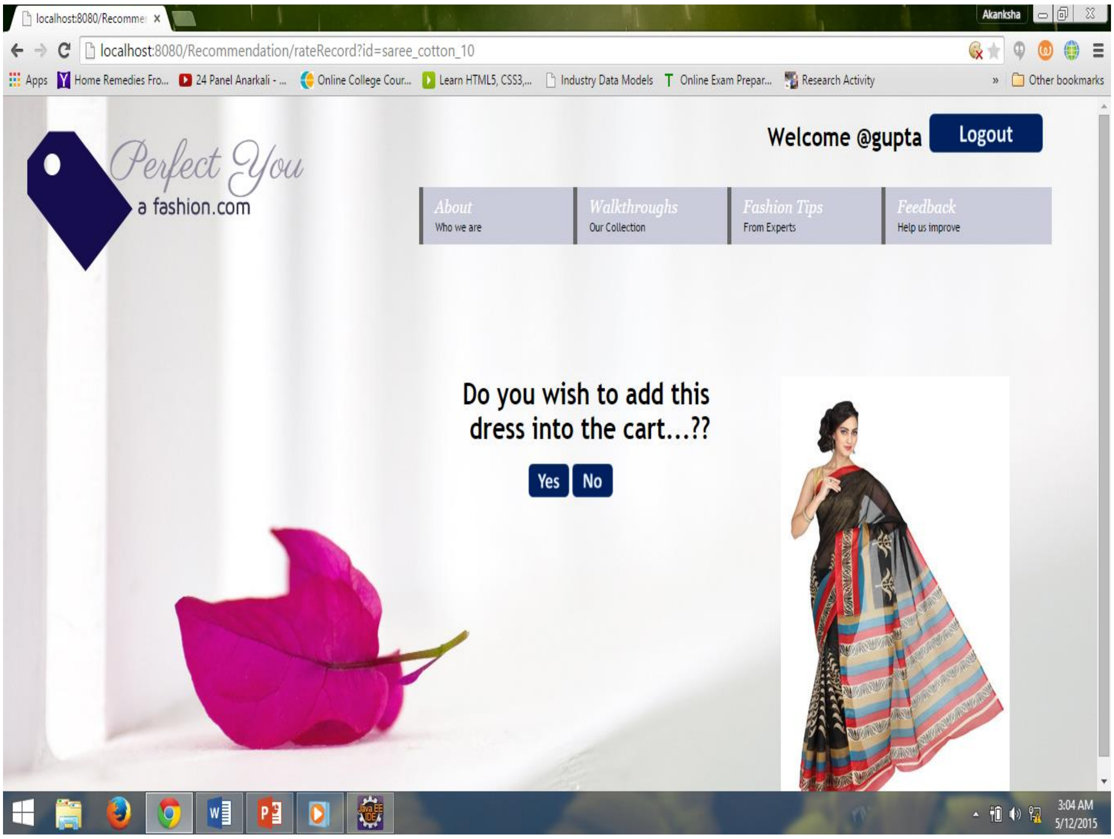


Snapshot: User Feedback



Snapshot: Payment Page

Snapshot: Dress Recommendations



Snapshot: Purchase Confirmation

Chapter 6

RESULTS AND EXPERIMENTATION

The various parameters that can be considered to analyze the performance of a recommendation engine are:

6.1 Recall or Clicks

$$recall@M = \frac{no. of\ repos\ the\ user\ watches\ in\ the\ top\ M}{total\ number\ of\ repos\ the\ user\ watches}$$

In our project when the same was applied, following were the results deduced(the value of M in our project is fixed as 5)

Table 6.1 Recall factor of various users

User A(@gupta)	User B(daggarwal)	User C(vaggarwal)
(2/10) = 0.2	(1/8) = 0.125	(2/7) = 0.28

6.2 Item Space Coverage

The term coverage refers to the proportion of items that the recommendation system can recommend. This is often referred to as catalog coverage.

This analysis was performed on one of the users registered with our site. @gupta when tried to look for traditional formal dresses, was offered with 12 out of 29 dresses present under the same category. This difference was due to the constraints applied which information retrieval in the form of physical specifications.

6.3 User Space Coverage

Coverage can also be the proportion of users or user interactions for which the system can recommend items. It can be measured by the richness of the user profile required to make a recommendation.

When this accuracy factor was evaluated in our case, following observations were made

Table 6.2 User Space Coverage with variations

No. of parameters provided as input	No. of dresses recommended
5 parameters namely Height, Body Type, Skin Color, Occasion and Dress Type	14
3 parameters namely Height, Body Type and Skin Color	33
None of the personalized feature provided as parameter	52

From the above table it can be observed that as more and more constraints are applied as filters to the recommender engine, lesser number of outputs were generated but those were of high correlation unlike the bulk of dresses presented as recommendation where the confidence level is low.

6.4 Conclusion

Though many accuracy parameters have not been evaluated primarily due to the constraint of limited dataset, but still from the above analysis, following observations can be made:

- The recall factor, which is appreciable shows that the user views approximately 2 out of the top 5 dresses recommended.
- The item space coverage shows that the since is capable of generating sufficient number of recommendations as the results are summation of both collaborative filtering and knowledge based recommendation techniques
- The user space coverage reflects the need of user input, otherwise the recommendation process will become inefficient.

Chapter 7

CONCLUSION

The recommender system has been implemented using the concepts of Artificial Intelligence and the major aim of the system has been achieved as the system is capable of generating recommendations of dresses for the users as per their past preferences and current requirement. This personalized system is capable of generating recommendations based on the specific physical characteristics of every individual viewing the site. These customers are shown only those set of dresses which are best fitted in their appearance and requirement.

- To cope with the changing requirements of few of the customers, it is very important to know about their requirement and then guide them accordingly. And this task is achieved by the knowledge based recommender as it asks the user about her requirement through questionnaire and then lists the suitable dresses with the company.
- Understanding the customer choice through the feedback provided and providing the dresses as recommendation accordingly is the main job of the any recommender system. And this task is achieved by the collaborative filtering recommender system working in sync with knowledge based recommender.

Chapter 8

FUTURE ENHANCEMENTS

As future work, I would try to overcome the short coming of the recommender.

- This project can be expended to recommend male customers in the same manner and also improve the overall recommendation process by also considering more personalized details.
- Another drawback which this system is currently facing is limited database of dresses and customer feedback.
- The recommendation process can be made more accurate by considering the personal details of any customer in a more precise manner. For example, considering the age factor of any customer can help the recommender recommend dresses in a better manner.

REFERENCES

[1] S.A. Fatumo, Emmanuel Adetiba, J.O.Onaolapo: **Implementation of XpertMalTyph: An Expert System for Medical Diagnosis of the Complications of Malaria and Typhoid.** IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 8, Issue 5 (Jan. - Feb. 2013), PP 34-40

[2] Nana Yaw Asabere: **Towards a Perspective of Hybrid Approaches and Methodologies in Recommender Systems.** International Journal of Computer and Telecommunications [Volume 3, Issue 11, November 2012]

[3] Joonseok Lee ,Mingxuan Sun and Guy Lebanon: **PREA: Personalized Recommendation Algorithms Toolkit.** Journal of Machine Learning Research 13 (2012)

[4] Maddali Surendra Prasad. Babu Boddu Raja Sarath Kumar**: An Implementation of the User-based Collaborative Filtering Algorithm.** Maddali Surendra Prasad Babu et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1283-1286

[5] Thomas Tran: **Combining Collaborative Filtering and Knowledge-Based Approaches for Better Recommendation Systems.** (2007)

[6] Janusz Sobecki: **Implementations of Web-based Recommender Systems Using Hybrid Methods.** International Journal of Computer Science & Applications Vol. 3 Issue 3, pp 52 – 64. (2006)

[7] Ernest Friedman-Hill: **JESS in Action- Rule based System in Java** Manning Publications Co. (2003)

[8] Burke, R.: **Hybrid Recommender Systems: Survey and Experiments**. UMUAI 12 (4), 378-405. (2002)

[9] In-Gook Chun, In-Sik Hong: **The Implementation Of Knowledge-Based Recommender System For Electronic Commerce Using Java Expert System Library,** ISIE 2001, Pusan, KOREA, 1766-1770. (2001)

[10] Burke, R.: **Knowledge-based Recommender Systems.** A. Kent (ed.): Encyclopedia of Library and Information Systems, 69, Sup. 32. (2000)

[11] http://www.jessrules.com/

[12] http://synotix.home.xs4all.nl/robbieng/docs/jess43doc.html