Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Anbar
College of Computer Science and Information Technology
Department of Computer Science

Recommender Systems for Market Predictions

A Thesis submitted to Department of Computer Science, College of Computer Science and Information Technology - University of Anbar. As a Partial Fulfillment of Requirements for Degree of Master of Science in Computer Science

By

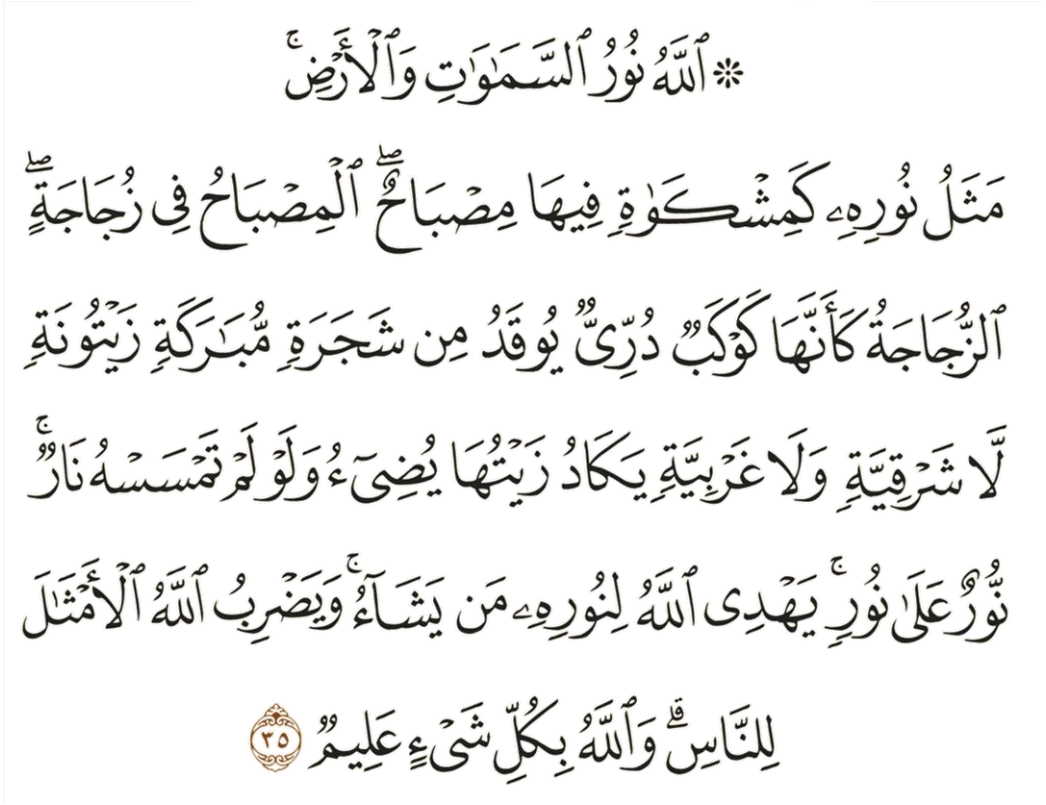
Lamees Yousif Abd

Supervised by

Prof. Dr.
Murtadha M. Hamad

Dr.
Ahmed J. Aljaaf

2022 A.C **1443 A.H**



سورة النور : 35

الاسم : لميس يوسف عبد

الكلية: كلية علوم الحاسبات وتكنولوجيا المعلومات – قسم علوم الحاسبات

عنوان الرسالة:

أنظمة التوصية لتنبؤات السوق

طبقا لقانون حماية المؤلف رقم 3 لسنة 1971 المعدل العراقي فأن للمؤلف حق منع أي حذف او تغيير للرسالة او الاطروحة بعد اقرارها وهي الحقوق الخاصة بالمؤلف وحده والتي لا يجوز الاعتداء عليها. فلا يحق لاحد ان يقرر نشر مصنف احجم مؤلفه عن نشره او اعادة نشر مؤلف لم يقر مؤلفه بذلك، فإذا قام بذلك اعتبر عمله غير مشروع لأنه استعمل سلطة لا يملكها قانونا.

عنوان البحث			
A Review of Marketing Recommendation Systems			
جهة النشر	Al-Kadhum 2nd International Conference on Modern Applications(MAICT_2021) (AIP)	نوع المجلة	Scopus
حالة النشر	Accepted	رقم المجلة	
رابط البحث			

عنوان البحث			
Building a Product Recomender Systems with Sales Data of marketing			
جهة النشر	4th International Scientific Conference of Engineering Sciences and Advances Technologies (IICESAT) (AIP)	نوع المجلة	Scopus
حالة النشر	Accepted	رقم المجلة	
رابط البحث			

اسم وتوقيع المشرف	اسم وتوقيع المشرف	اسم وتوقيع رئيس القسم
أ. د مرتضى محمد حمد	د. أحمد جاسم محمد	أ.م.د وسام محمد جاسم

Supervisor Certificate

*We certify that this thesis entitled (**Recommender Systems for Market Predictions**) was prepared under my supervision at the Department of Computer Science and information technology University of Anbar, by (**Lamees Yousif Abd**) and that, in my opinion, it meets the standards of a thesis for the degree of Master of Science in Computer Science.*

Signature:
Name:**Prof .Dr. Murtadha M. Hamad**
(Supervisor I)
Date: / / 2022

Signature:
Name: **Dr. Ahmed J. Aljaaf**
(Supervisor II)
Date: / / 2022

Examination Committee Certification

*We certify that, we have read this thesis (**Recommender Systems for Market Predictions**), and as an Examining Committee the student **Lamees Yousif Abd** in its contents and that in our opinion it is adequate to fulfill the requirements for the degree of Master of Science in Computer Science.*

Signature:
Name: **Dr.**
Date: / / 2022
Member

Signature:
Name: **Dr.**
Date: / / 2022
Member

Signature:
Name: **Dr.**
Date: / / 2022
Chairman

Signature:
Name:**Prof .Dr. Murtadha M. Hamad**
(Supervisor I)
Date: / / 2022

Signature:
Name: **Dr. Ahmed J. Aljaaf**
(Supervisor II)
Date: / / 2022

Approved by, the Computer Science Department, Computer Sciences and Information Technology College, University of Al-Anbar.

Signature:
Name: **Prof. Dr. Salah A. Salman** (Dean of College)
Date: / / 2022

Acknowledgements

First and foremost, I must thank Allah for His unlimited bounties and graces for helping me finishing this thesis to its best form.

It is a pleasure to thank those who make this thesis possible. First of all, I would like to express my gratitude and respect to my supervisor ***Dr. Murtadha M. Hamad*** *, For his unlimited supports, guidance and advice throughout this study; His positive outlook and confidence at work inspired me and gave me confidence. His careful editing contributed greatly to the production of this treatise.*

I would like to express my deep thanks and gratitude to my supervisor ***Dr. Ahmed J. Aljaf*** *for his advice, assistance and supervision throughout this work to be in the best way possible.*

I wish to thank the members of the College of Computer Sciences and information technology, University of Al-Anbar, for their care, help, support, and guidance during my work.

Deepest thanks go to all my friends who took part in making this work come to reality and supported me morally in overcoming the hard and stress times during this research and preparation of this thesis . Last but not least,

I would like to thank all those who had participated in one way or another in the achievement of this work.

Dedication

- *To my parents, without their never ending love and support, might not be the person I am today;*
- *My husband …My ever reliable partner, the greatest treasure that I have ever had.*
- *To my sisters and brothers who never left my side ;*
- *To my friends…*
- *While studying... one of the greatest gifts I got is my beautiful child... Yusuf*
- *I say: Without your encouragement, assistance; work would not have been fulfilled.*

With love

Lamees Yousif

ABSTRACT

The Internet's growth has resulted in a significant dispersion of data resources. Specialized recommendations on various types of information, products, and services are needed to support users in overcoming the problems of information overload. The recommendation system is one of the secrets ways utilized by successful companies, and this system considers a magical marketer for services and products by observing customers and understanding their behavior to help them making the right decision.

This thesis contains machine learning techniques combined with deep learning techniques. In addition, the apriori algorithm is used to create association rules because array based , large memory and it scans DB multiple times in order to improve the system efficiency to make accurate predictions and recommend suitable products. One of the most important techniques that was used in addition to apriori is a proposed model, this technique was applied by combining machine learning techniques and deep learning techniques by the application of a proposed model between GMM & KNN, GMM & SVM, GMM & LSTM. calculate (precision recall, f1-score, accuracy).

The proposed approach utilized the Modcloth dataset sold Amazon containing of 998,94 transaction records. The obtained results were compared using assessment measures to determine which model is the best ,the results showed that the best classifier was K-means-SVM where achieved 0.999 rate of accuracy then the K-means-KNN achieved 0.998 rate of accuracy, based on the above results, the best classifier GMM-SVM where achieved the accuracy 0.996 , then the GMM-KNN achieved 0.992 rate . The proposed system was implemented using the Python programming language and imported some libraries to get high performance mode.

Keywords: Recommender System(RS) , Gaussian Mixture Modelling(GMM) , K-means , Market Basket Analysis(MBA); Support Vector Machine (SVM) ; Data Mining .

List of Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

BI	Business Intelligence
BPR	Bayesian Personalized Ranking
CBF	Content-Based Filtering
CF	Collaborative Filtering
CMFH	Collective Matrix Factorization Hashing
DM	Data Mining
DNN	Deep Neural Network
EM	Expectation Maximisation
ETL	Extract Transform and Load
GMM	Gaussian Mixture Modelling
HCI	Human-Computer Interaction
IR	Information Retrieval
LDA	Latent Dirichlet Allocation
LSTM	Long Short Term Memory
MBA	Market Basket Analysis

MPUM	Multi-Product Benefit Maximization
NCTR	Next Campaign To Run

OAPP	Open Architecture Product Platform
OEMs	Original Equipment Manufacturers
PSRS	Price-Sensitive Recommendation System
QM	Quality Management
RBF	Radial Basis function
RNN	Recurrent neural network
RS	Recommender System
SMEs	Small and Medium-Sized Enterprises
SVD	Singular Value Decomposition
SVM	Support Vector Machine
RSs	Recommender Systems
LR	Logistic Regression
PCA	Principal Component Analysis
RBF	Radial Basis Function
NB	Naive Bayes

Chapter One
General Introduction

Chapter One

1.1 Introduction

The amount of data on the Internet has grown exponentially in the recent years . E-commerce has been driving research on consumer decision support systems by providing customers with an access to vast volumes of products knowledge. In particular, recommendation systems (RSs) have showed their benefit in helping customers [1, 2]. RS can be called here an information filtering engine or platform that seeks to predict the "rating" or "preference" that a customer might give to an item [3]. On virtually every e-commerce platform, RS is used, benefiting millions of users. E- Commerce websites such as Netflix and Movie lens for moves, Amazon for books,CDs and many other items [4]. Companies chose to provide an RS in an effort to improve revenue upon the advent of the internet and the age of e-commerce[5]. RSs provide forecasts of goods that the consumer will find interesting to buy. It has been shown that RSs are beneficial for users and businesses [6]. In addition, RS is a web-based application in the e-commerce sense that specifically or indirectly collects the preferences of a customer and recommends the goods or services of personalized e-sellers accordingly [7]. RSs can also benefit users, since successful predictions can minimize the user's search space, [8]. The recommendation systems is to provide each consumer with appropriate recommendations based on their preferences and behaviors [9]. A general recommendation system's main purpose is to predict items that potential customers will want to buy based on their specified preferences, online shopping options, and purchases of people with similar tastes or demographics[10]. It is possible to classify recommendation solutions in elements of: filtering method (collaborative, content-based, and hybrid), technique of graphs, ontologies, and rules are examples of knowledge representation Cold start are not enough data to accurately suggest new items or new users and data (just a small percentage of items have been evaluated) , variety (the same elements that are identified appear to be recommended over and over again), consistency(to recommend items that are actually applicable to the user) [11].

In understanding consumers and evaluations of products to be recommended, the recommended digital marketing process has a strong function and influence in understand the business rules defined by the domain rule builder; understand user buying style and create custom preferences. The process is illustrated in figure (1.1) [12].

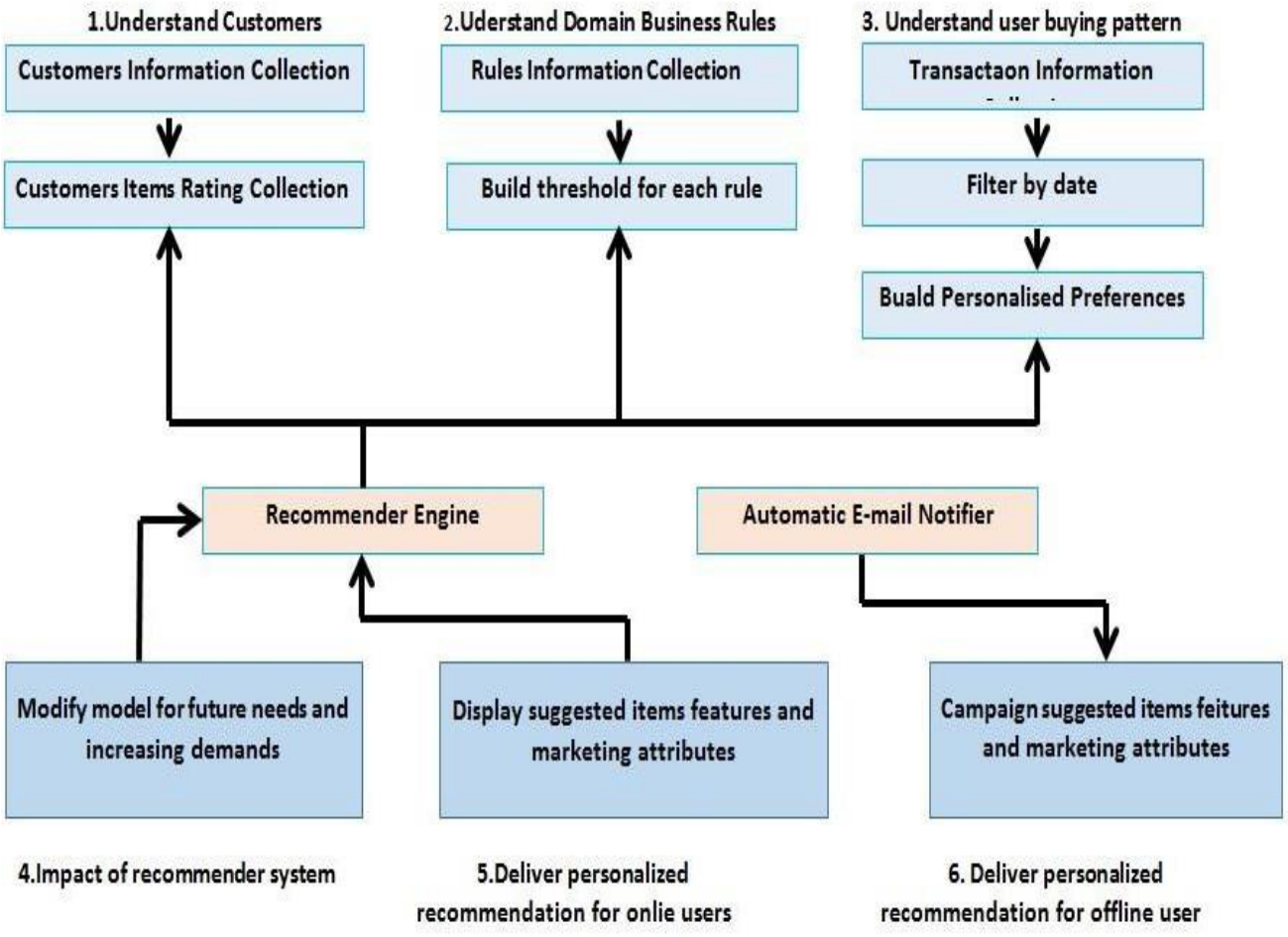


Figure (1.1)Stages of digital marketing recommended process

1.2 Stages of Recommendation Process

The Recommender Framework digs out from the vast amount of overload data that is collected according to the needs and likes of the consumer. RS follows three main stages for the recommendation process including the information-gathering stage, the learning stage, and the prediction/recommendation stage, for the recommendation of goods or objects to the customers. The following are the steps of recommendation system[13].

1 - **Information Collection stage**: In this stage, data about the user is collected from various sources such as the Internet / social network, media and databases. The data is extracted to find relevant and appropriate information [14]. This process collects t h e user's facts for the build of a prediction functioncustomer profile or model, including customer attributes, activity, or resources accessed by the customer [15]. Inputs obtained in various forms like implicit evaluation ,explicit evaluation and hybrid evaluation, are the basis of a recommendation method.[16].

2- **Learning stage:** This stage uses the techniques and algorithms of filtering the information collected indirectly or explicitly to extract useful insights from the information needed in the next stage.[14]

3- **Prediction\recommendation stage:** Recommends or predicts the items that users may prefer based on the dataset or user behaviors observed through the systems [13]. For users in this step, preferred items are suggested. Review of input obtained in the information collection process.[17].

1.3 Literature Review

A number of researchers have studied recommendation systems in various fields, and we have searched in many scientific sites and libraries such as Google Scholar, Springer Link, IEEE and many other scientific journals. These searches included the following keywords: "Recommendation Systems, Customers, Products, E-commerce, Collaborative filtering, Personalization". We found many sources using recommendation systems, some of them within the research strategy and others outside the scope of the research, so 10 selected studies related to the research were used.

1- Panniello Umberto, 2015," **Developing a price-sensitive recommender system to improve accuracy and business performance of ecommerce applications"** [18] in this work, How to include the price in the recommendation systems and how it positively impacts customers' purchasing decisions, Research results include price in the recommendation system to improve the accuracy of recommendations and business performance through the use of a price-sensitive recommendation system, or PSRS.

2- Masahiro Sato, Hidetaka Izumo and Takashi Sonoda,2015**," Discount Sensitive Recommender System for Retail Business"**,[19]. this work, include

the recommendation algorithm with tailored discount sensitivity, The findings indicate that discount sensitivity is a key component of retail domain recommendation systems.

3- Dietmar Jannach& Gediminas Adomavicius, 2017," **Price and Profit Awareness in Recommender Systems"**.[20]. He focuses on designing a system that helps to maximize the users' utility by relating each user to a specific element. Research results focus on including price and profit information for the recommendation system in order to balance the interests of customers and the priorities of service providers.

4-Qi Zhaoy, Yongfeng Zhangz_, Yi Zhangy & Daniel Friedman, 2017, **"Maximizing the Multiple Product Benefits of Economic Recommendation"** [21].Multi-Product Benefit Maximization (MPUM) has been proposed as a general approach to economic principles-driven recommendation, MPUM compared to common e-commerce recommendation algorithms, And results showed that MPUM significantly outperforms algorithms using Top-K rating scales.

5- Jingyuan Yang, Chuanren Liu, Mingfei Teng, Ji Chen and Hui Xiong,2017," **A Unified View of Social and Temporal Modeling for B2B Marketing Campaign Recommendation"**.[22]. Focus this research a coherent view of social and temporal modeling for the recommendation of B2B marketing campaigns. In the B2B purchasing processes, manipulate temporal activity trends and build a marketing campaign recommendation framework, Results are shown to improve the quality of recommendations through the proposed method for the to challenge B2B marketing tasks.

6- Rahul S. Gaikwad, Sandeep S. Udmale and Vijay K. Sambhe,2018**," E-commerce Recommendation System Using Improved Probabilistic Model "**.[23]. A collaborative filtering recommendation system was proposed using NB algorithm optimized bigram language for improved analysis, Theresults showed that the proposed model has 14% more efficiency compared tothe simple Naive Bayes model.

7- Yan Guo, Chengxin Yin, Mingfu Li, Xiaoting Ren and Ping Liu,2018," **Mobile e-Commerce Recommendation System Based on Multi-Source Information Fusion for Sustainable e-Business"**. [24].The proposes a method for analyzing consumer requirements in e-commerce for recommendationsystems by using multi-source information, The results of the

research that the mobile e-commerce recommendation system provides consumers with timely information and provides a more comfortable, more accurate and effective shopping experience than the traditional method.

8- Rishabh Misra, Mengting Wan and Julian McAuley,2018, **"Decomposing Fit Semantics for Product Size Recommendation in Metric Spaces"** [25]. The proposes a predictive framework for solving the product suitability problem, which uses metric learning to address label imbalance problems, results obtained Models with Kdimensional latent variables outperform methods with only one latent variable

9- Hailin Li, Yenchun Jim Wu and Yewang Chen,2020, "**Time is money: Dynamic-model-based time series data-mining for correlation analysis of commodity sales**" [26]. To investigate the sales correlations among various commodities, a method of conducting dynamic-model-based time series data mining was presented. Results of a retail sales analysis based on Goods that have a sales correlation in different markets are included in the dynamic model. can be recognized Cross-marketing decisions, merchandise promotion, merchandise warehousing, and planning can all be aided by keeping track of time windows.

10- Sandra Rizkallah, Amir F. Atiya and Samir Shaheen,2021" **New Vector-Space Embeddings for Recommender Systems**" [27] presented a new type of vector embedding adaptable in the manner that may be used to forecast ratings as well as to recommend top things that are likely to pique your interest consumers. The presented models have been successfully applied to real-world datasets such as "MovieLens," "ModCloth," "Amazon: Magazine Subscriptions," and "Online Retail," with competitive results.

1.3.1 Summary of Literature review

The topic of this thesis is based on extracting aspects from a literature review of researchers on RSs. Therefore the Table1.1 introduces a brief literature review.

Table1.1 Summary of literature review

No	Authors	Issues name	Types of Filtering	Technique used	Results and Dataset used
1	Panniello Umberto (2015)	Recommendation Accuracy	Collaborative Filtering	Compared between item-based ,customer-based	Research results include price in the recommendation system to improve the accuracy ,The design for MD recommendation system use these simulated purchase transactions as implicit ratings
2	Masahiro Sato, Hidetaka Izumo and Takashi Sonoda(2015)	the real-world prediction problem	Collaborative Filtering	using the matrix factorization for Bayesian Personalized Ranking (BPR) with to include a discount effect.	The findings indicate that discount sensitivity is a key component of recommendation systems, used Ta-Feng dataset was verified a public retail dataset.
3	Dietmar Jannach & Gediminas Adomavicius (2017)	Attempting to increase the customers' utility by to determine the most relevant items for each customers	Collaborative Filtering	A factorization algorithm, Rearrangement objects for each consumer greedily.	Research results focus on including price and profit information for the recommendation system. used took the MovieLens 1M dataset and values for each of the movies
4	Qi Zhaoy, Yongfeng Zhangz_, Yi Zhangy & Daniel	One major problem is that the top ranked recommendations might be very similar or even duplicated and	collaborative ltering, content-based ltering, or hybrid algorithms	Precision and recall at top-K are the most widely used adopted a greedy method to generate top-K products	results showed that MPUM significantly outperforms algorithms using Top-K rating scales, used two real-world datasets are used like Shop.com and Amazon

No	Authors	Issues name	Types of Filtering	Technique used	Results and Dataset used
	Friedman (2017)	the diversication problem			
5	Jingyuan Yang, Chuanren Liu, Mingfei Teng, Ji Chen and Hui Xiong (2017)	cold-start problems and (NCTR) problem	collaborative filtering approach	take out dynamic properties by time-series analysis and the proposed NCTR framework works very well for predicting the B2B customers campaign preferences	Results are shown to improve the quality of recommendations .applied on real-world B2B marketing data sets
6	Rahul S. Gaikwad, Sandeep S. Udmale and Vijay K. Sambhe,(2018)	Recommender dation Accuracy	Content-Based Recommendation and Collaborative Filtering Recommendation	the Model using enhanced probabilistic approach and commodity prediction and top-K prediction	The results showed that model more efficiency compared to the simple Naive Bayes model.Datasets for the include customer Conduct on an e-commerce site
7	Yan Guo, Chengxin Yin, Mingfu Li, Xiaoting Ren and Ping Liu (2018)	Recommendation Accuracy	collaborative filtering, content filtering and data-mining techniques	improved radial basis function (RBF) network to to determine the weights of recommendations	The results of the research that the mobile e-commerce recommendation system more comfortable, more accurate. collect data from the shopping platform

No	Authors	Issues name	Types of Filtering	Technique used	Results and Dataset used
8	Rishabh Misra, Mengting Wan and Julian McAuley,(2018)	The problem of product size recommendation, label imbalance.	collaborative filtering was combined with metric learning.	For each client and product, the technique assumes a single latent variable. For final classification, the learned features are used in Logistic Regression (LR)	results obtained Models with Kdimensional latent variables outperform methods with only one latent variable ,Applied to two datasets from ModCloth websites and RentTheRunWay
9	Hailin Li, Yenchun Jim Wu and Yewang Chen,(2020)	The Apriori method requires significant time costs to identify recurring elements in the transaction Database .	Time series data mining entails a collection of intelligence approaches such as time series clustering, classification	K-nearest neighbors similarity search and Apriori algorithm for working out association rules.	Results of a retail sales analysis based on Goods that have a sales correlation in different markets .dataset of 4,070 items (commodities) and 25,900 transactions The two "minimum support" and "minimal confidence" levels.
10	Sandra Rizkallah, Amir F. Atiya and Samir Shaheen ,(2021)	optimization problem for recommendering improve the accuracy of prediction	collaborative filtering	Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) and Neighborhood-Based Collaborative Filter	Apply datasets, for example "MovieLens," "ModCloth," "Amazon: Magazine Subscriptions," and "Online retail price",

1.3.2 Justification of Literature Review

By examining the literature review consisting of the research cited by researchers in Section 1.3, we concluded the following:

1. RS is a framework to provide visitors with customized recommendations for products on the basis of their preferences.
2. The use of data analysis techniques and tools such as extraction, cleaning and transformation processes consider very beneficial such as (collaborative filtering, content filtering, hybrid algorithms, and data-mining techniques) contribute to the advancement of different businesses and lives, which are important and influenced in decision-making.
3. Poor performance of the accuracy of recommendations, due to the problem of scattered data entered in the recommendation systems and the problem of cold start . Alack of familiarity or knowledge for buyer or seller that results in inaccurate predictions of recommendations, as similarity calculations cannot be performed through a small number of evaluations, it is one of the most limitations of recommendation systems. In addition to synonyms and scalability, some problems have been solved through the use of some techniques such as improving radial basis function (RBF), Used to recommend deep learning methods, Collective Matrix Factorization Hashing (CMFH), and other methods
4. Data mining is a method of capturing and discovering knowledge.

1.4 Problem Statement

A recommendation system has an essential role that would by recommending users to buy or sell products. But there are many problems with RS that need to be addressed, including inaccuracy of recommendations and inaccurate predictions of recommendations due to a cold start problem, when a new product is added to a supermarket, or when a new piece of content is uploaded to a media platform that, at first, does not Nobody knows about it and there are no interactions or ratings, the collaborative filtering algorithms recommend each item (product or piece of content) based on user actions such as views, ratings or purchases. The more user actions an item performs, the easier it is to see who else might be interested in it. This is called cold starting of the product or component.Item Cold Start Resolved Using Deep Learning.

1.5 The Contribution

RS used to cope with the problem of information overload and data may recommendation methods to until know , no one is best for all users in all situations , to give a method to help customers in product selection, with the goal of predicting or filtering preferences based on the user's choices , applied by combining machine learning techniques and deep learning techniques by the application of a proposed model between GMM & KNN, GMM & SVM, GMM & LSTM. calculate (precision recall, f1-score, accuracy).

1.6 Aims of Thesis

This thesis aims to achieve the following main objectives:

1. Predict and discover information using the data mining methods to help the decision-making process and leverage business resources.
2. Implement algorithms and enhance the Apriori algorithm of the association rule mining.
3. Recommendation marketing is an important marketing tool, such as providing a high-quality product, building trust in the brand and increasing customer loyalty.

1.7 Thesis Outline

This thesis consists of the following chapters:

Chapter two : "Theoretical Background". Discusses the concepts of the thesis' theoretical section, which serves as a framework for proposed models based on recommendation system concepts and kinds.

Chapter three : "Research Methodology". Explains the framework that will be utilized in the system, which comprises the schema, interfaces, application, and other design and analysis tools.

Chapter four: " System Implementation & Results Discussion" . Discuss the system's implementation and the results obtained as a result of the system's implementation in this chapter.

Chapter five: "Conclusions and Recommendations for Future Works". The conclusions of this work and proposals for future work are discussed.

Chapter Two
Theoretical Background

Chapter Two

2.1 Introduction

This chapter introduces the theoretical background of many basic concepts and definitions of the recommendation system such as types and the basic stages technologies of RSs. Furthermore, the most important part in building a recommendation system is the use of collaborative filtering techniques. The most important strategies and challenges of the custom market and methods of data mining.

2.2 Major Issues in recommender system

RS is a method used to deal with large and composite databases of information. Based on their interest, it recommends the information or products to the customer but faces many problems in the recommendation process. The figure (2.1) contains some problems. [13]

Figure 2.1 Issues in Recommendation System

1. **Cold Start:** occurs result not enough classification information or a new customer joining the system, so it is hard obtain the likeness between

customers and the Elements [28]. Solve the problem, like asking the new customer to evaluate those things at the beginning, Inquire specifically about the status of new users [29].

2. **Synonym:** This problem arises when more than one name for the same element or similar meanings represents the same element [4]. By using these approaches, Singular Value Indexing Decomposition (SVD), Construction of a thesaurus, and Latent Semantic the synonymous problem can be solved[29].
3. **Privacy:** Recommendation systems significantly depend on comments that may be implicit or explicit received from users. These comments contain the user's personal preferences, directions and opinions. Privacy Fearshinder release of data needed to advance recommendation algorithms [4].
4. **Scalability:** This occurs when the number of products and users increases dramatically [13]. The system aims to react to immediately to the customers which requires higher scalability.[30].This is done by using some measures to develop the system,including:-
   - **a- Training time:** Recommendationsystems need a separate training phase from the testing phase, where the total needed training as one of the steps, model is used. Often of the time.
   - **b- Prediction time:** Here it determines the response time in which the user receives responses to determine the most important recommendations for aspecific customer.

   c- Memory requirements: An algorithm is designed to reduce the very high memory requirements, as it is difficult to keep large classifications in the main memory, as systems are difficult to use in a widerange [31].
5. **Data sparsity:** This issue stems from the phenomenon that users only rate a small number of things in general [32]. It is difficult to infer user taste and may be associated with a pseudo-next [33]. The table shows popular sites that use the recommender system

Table 2.1: Popular sites that use recommendation systems[4]

Site	What is recommended
Amazon	Books
Spotify	Songs
LinkedIn	Jobs
E-Commerce Sites	Items
Movie Lens	Movies
Facebook	Friends
Netflix	Movies\DVDs

2.3 Niche marketing strategies and challenges

As a result of the rapid development of technology and the emergence of new products, market requirements and marketing conditions changed. With increasing competition in the markets and changes in consumer behavior, it is imperative that companies be dynamic and constantly develop themselves. By developing new strategies that distinguish it from others due to the intense competition in today's marketing approach [34]. A niche market is a small portion of the market consisting of a small community of consumers who need a niche product [35]. Although a clear concept of specialist marketing is hard to find, there are some features that can explain specialized activities[36]:

- Thinking and acting small by providing small quantities of production, concentrating on a few customers and avoiding a market with several rivals or a dominant rival (such as Hezar et al 2017)
- Focusing on client needs (Dalgic & Leeuw)
- Reputation of companies and using word-of - mouth references (Dalgic & Leeuw)
- Charging a premium prices (Dalgic & Leeuw) Most managers put in place the necessary procedures for marketing planning in order to think in an orderly way to clarify their economic models for business. Figure 2 .2 shows the stages that must be gone through to reach the marketing plan (marketing planning). [37].

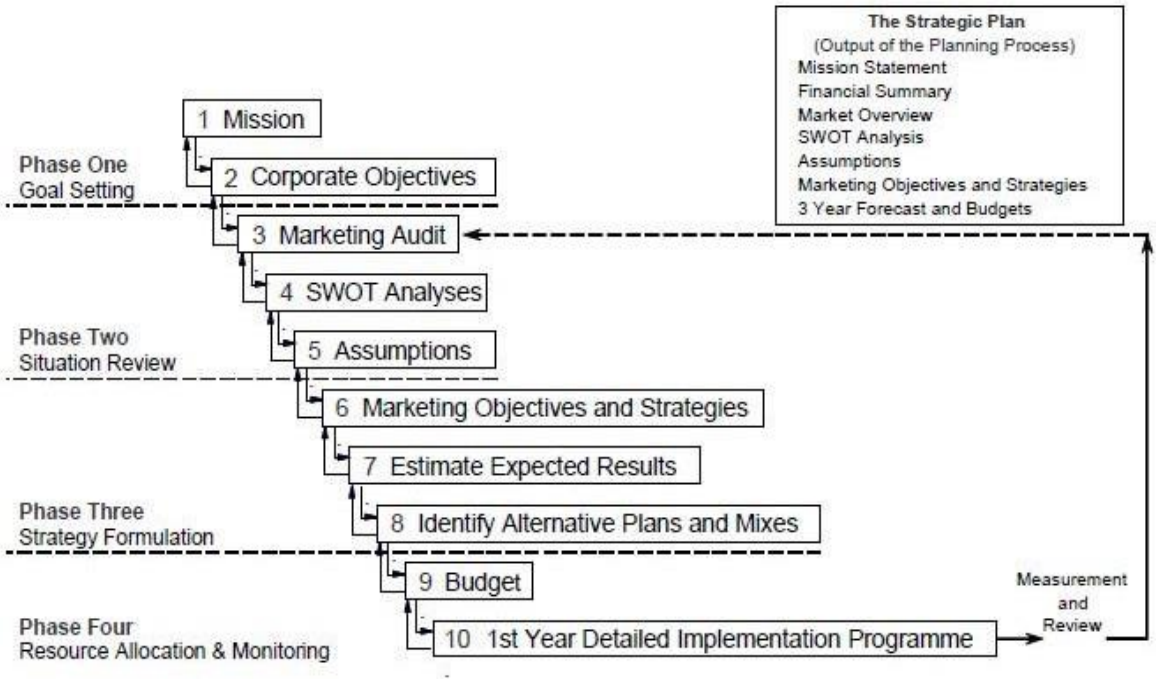


Figure 2.2 The Ten Steps of the Strategic Marketing Planning Process[37].

Figure includes phase One (Goal Setting),phase Two (Situation Review),phase Three(Stratagy Formulation)and phase Four(Resource Allocation & Monitoring) .The role of marketing is not only to provide consumers with goods or services, but also to ensure that these products or services are able to satisfy customers, resulting in repeated consumer purchases. This principle also focuses on producing a product better than its rivals[38].

2.4 Recominder systems

2.4.1 Types of Recominder systems

Recominder systems can be divided based on user need and the type of data used in the recominder.

1. **Collaborative Filtering:** Collaborative filtering methods construct the system by taking into account the past actions of the user (rating is given to certain items, previously similar decisions taken by different users, then using the system to measure the item or other rating that might be of interest to the user[39]. The collaborative filtering algorithm is primarily divided into, such as memory-based and model-based

- ✓ **Memory-based filtering technique** primarily divided into two categories, such as user-based and item-based filtering techniques
  a. **User-based filtering:** The user-based filtering method calculates the uniformity between users and not between objects.
  b. **Item-based filtering:** Using the similarity between objects and not between users, the item-based filtering technique computes recominder

  It determines how similar the retrieved items are to the target item fromthe user-item matrix [33].
- ✓ **Model-based** filtering **technique** The obscure ratings are predicted by model-based CF methods Using machine learning or statistical methods after learning a model from the underlying knowledge. [40].

2. **Content-based filtering:** Recominder for content-based filtering depend on users' previous decisions. In content-based filtering, item definition and a profile of the user's orientation play an important role.
3. **Demographic Filtering:** recommender on demographic filtering are based on the user's demographic profile. such as nationality, age, gender etc., the recommender is based on the information given by the user.
4. **Hybrid filtering:** Hybrid filtering is a blend of more than one approach to filtering. The hybrid filtering strategy is implemented to solve some common problems associated with above-mentioned filtering methods as an example problem of cold start, problems of overspecialization and problem of sparsity. Improving the consistency and efficacy of the recommender process is another motivation behind the introduction of hybrid filtering[30].
5. **Knowledge-Based recommender system** Knowledge based recommender system. In terms of how useful the recommended item is to the user, the accuracy of the model is judged[39].

2.4.2 Functional Architecture of Recommender systems

It proposes a conceptual structure based on an adaptable open architecture product platform (OAPP) for the customized distributed product configuration process., Koren et al (2013) first proposed an open architecture product. It is characterized as "one with a framework that allows modules from various sources to be incorporated in order to adjust product functionality exactly to the needs of the consumer." Large businesses, such as original equipment manufacturers

(OEMs).The standard platform tends to be built and the interface defined, small and medium-sized enterprises (SMEs) create add-on modules (both tailored and personalized) that can be linked to the OAPP[1]. OAP is regarded as providing a framework and open interfaces from which it is possible to connect various add-on modules from different sources to meet customer requirement

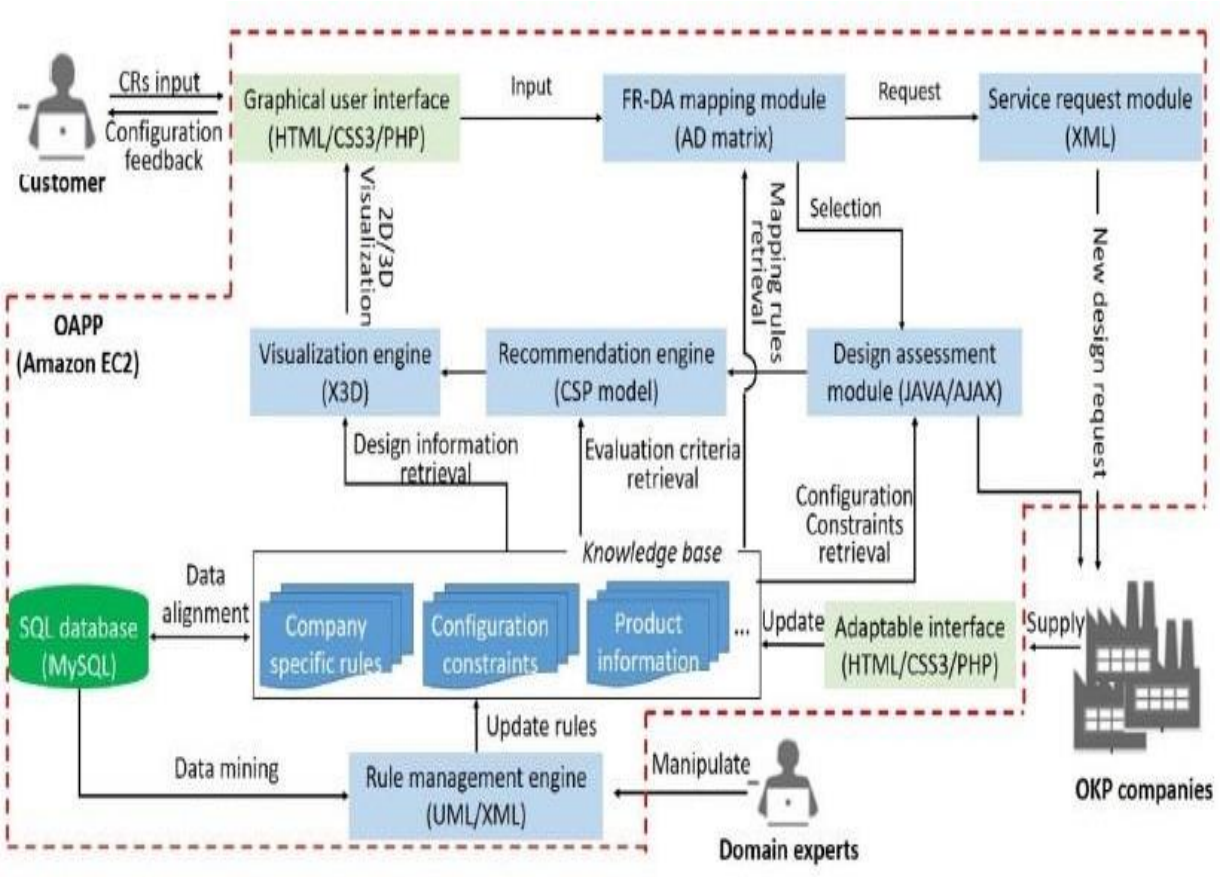


Figure 2.3 Technical configuration system in an OAPP [1]

The architecture of the technical configuration system in an OAPP is shown in Figure2.3 in order to allow the proposed two-stage configuration procedure. It consists mainly of 8 main components, i.e. graphical interface (user interface and adaptable interface of OKP companies), FR-DA mapping module, service request module, specification review module, recommendation engine, visualization engine, rule management engine, and knowledge base[1].

2.5 Data mining with Recommender System

The inevitable trend of e-commerce service development is the implementation of big data mining technology to introduce intelligent, customized, and initiative smart services and further encourage business model innovation. The intelligent service, such as knowledge recommendation service, has become a new model of e-commerce service development with the help of large data mining[41]. The data mining process usually consists of three steps including: Data Preprocessing, Data Analysis, and Interpretation of Findings. The growth trend of e-commerce services is broad data mining technologies to provide intelligent service to customers[42]. The most popular methods of data mining used in RS are grouping, clustering, and association rules discovery.

2.5.1 Association Rules Discovery

In data mining, the function of association is to find attributes that appear at a time. In the business world, the study of the shopping analysis is more generally called (market basket analysis). [43] It is a rules of data mining that attempts to find the relationship between common object patterns within a data set [44].

2.5.2 Clustering

The clustering process can be stated as following: if you have a set of data points with unique qualities and a degree of similarity, they should be clustered where the data points in that cluster are very similar to each other. Separate clusters of data are likely to be distinct to one another. To determine how close or far apart two clusters are. Marketing segmentation is a beneficial application of clustering, in which different groups of clients are created in the market and different marketing methods are applied to each of the subsets. [45]. To find outliersin a given set of values is an unsupervised technique. In finding noisy data, it is helpful. K-Means, Density-based clustering, Distribution-based and Centroid- based clustering are different clustering techniques. A descriptive strategy is clustering[46].

2.5.3 Classification

It is a method of classifying new tuples into groups based on training data. It is a predictive process. Unknown values of several other similar variables can be predicted using the known values of one variable. It is a methodology under supervision. Decision tree, Support Vector Machine (SVM), Bayesian classification, neural networks, induction rules and so on are numerous techniques used for classification[46].

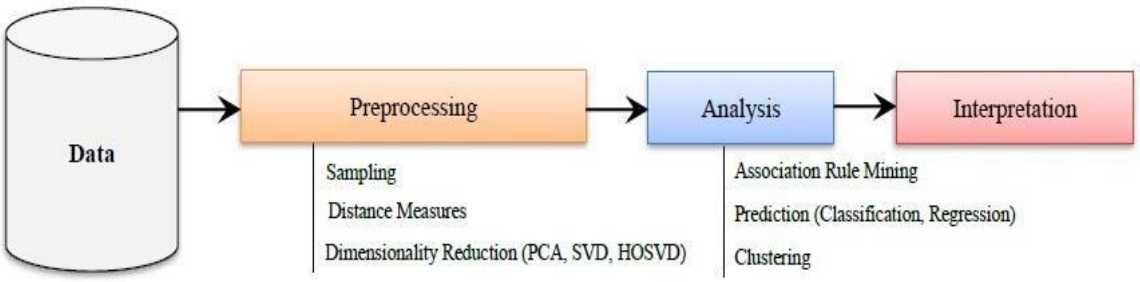


Figure 2.4 Main steps and methods in a data mining problem. [47]

2.6 Recommender System in Commercial Use

Recommendation systems are important for e-commerce sites to increase sales revenue by recommending preferred goods and products to customers over the Internet and by building recommendation algorithms with high achievement, high speed and low cost and the goal of recommendation algorithms is to achieve high sales profit [48]. This evolution allows e-commerce sites to observe rich user data.Here, to illustrate how to use commercial recommendation solutions in business, tasks and components [49]. The goal of RSs is to provide personalised services product recommendations to website customers. They gain knowledge from the customer and make recommendations, the respective products to the user. These systems make the experience more personal by achieving user interest. There are three methods in which a product is recommended as a result of the top total sellers on the site, the demographics of the customer, or the customer's history of prior purchases [50].

The company, that provides the recommendation, can be described as the seller, and the company that applies it to websites can be described as the customer, and the customer who interacts with thesites to obtain customer service or obtain the product, and the figure 2.5 representsa platform Resonance Recommendations for Certona, a template for commercial recommendation systems[51]. These include testing and interpretation of data patterns, pre-processing and data extraction so that the marketing analyst can understand the behavior of customers and the products purchased [52]. Market basket analysis determines the buying patterns of customers by finding the relationships in sales that appear simultaneously by the sales companies / supermarkets to find out the items that are bought by consumers at the same time through the correlation rule or correlation analysis to reveal the interconnectednessbetween the elements, which is considered one of the exploration a data mining techniques [53].

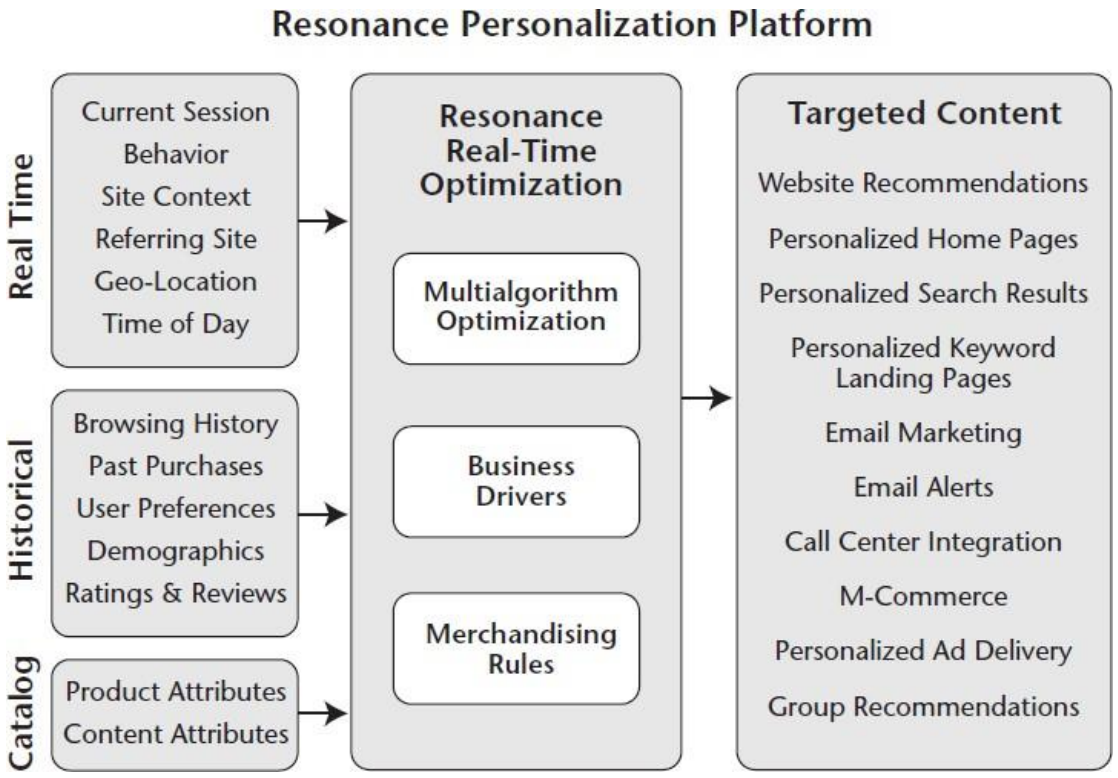


Figure 2.5 Example of a Commercial Recommender System. [54].

The figure shows the recommendation engine data sources that include (current session context, historical data, product catalog data) and the variables used by the recommendation engine such as (business, rules and algorithms) as well as the recommended content types. To clarify the recommendations work, the seller must prepare a custom algorithm for the customer and be prepared by the solution providers by integrating an additional element in the algorithm or customer attributes,the behavioral data may be mouse movement, search terms, clicks, or time that Spending

2.7 Market analysis using Association Rules Mining

Major retailers have used a technique to uncover correlations between items. This achieves by looking for groups of items that occur together frequently in transactions. Correlation analysis in which it attempts to find common patterns foritems in large data sets. One of the specific applications is often called a market basket analysis [52].

2.7.1 Market Basket Analysis

Market basket analysis (MBA) is a method of extracting data in various fields such as marketing, bioinformatics, nuclear science, and fields of education, also known as association rule learning or affinity analysis. Its main purpose is to manage business in marketing is to understand the buying behavior of the buyer and to provide information to the retailer in order to make the appropriate and correct decision and the useful information is extracted by exploring this data and Knowledge Discovery and Data Mining (KDD) [52]. Predictive analyzes use several statistical and analytical techniquesto extract recommendations are made meaning market basket analysis allows determining the relationships between products that customers buy. To start an MBA, need a data set of transactions, and each transaction represents a group of purchased products or items such as one group of items (paper, pencil, rubber ,pins). All items are purchased in one transaction.

In MBA, the rules of associationare defined by analyzing the transactions, for example “One of the rules: (pencil, paper) greater than or equal to (rubber). Here, this means that when the customer purchases pencil and paper, he is likely to buy rubber as well[53].

2.7.2 Association Rules

Association rules mining among large sets of data items finds interesting associations and relationships. This rules demonstrates how often a group of items can repeat in a transaction [55].

Association Rules is an implicit function that can be expressed as $X \rightarrow Y$, where X, Y denotes two separate groups, i.e. $X \cap Y = \emptyset$. The strength of the association rule depends on two key factors: support and confidence .

Several algorithms are available in association rule mining, including: Apriori, Apriori – TID, Eclat, FP – Growth, LCM. [55]

Apriori Algorithm With Examples

It is an important supervised algorithm used to find frequent sets of productsduring a transaction. It is used with an association rule. In this algorithm, if the support level exceeds the bare minimum, the repeated (large) item sets will be calledand the groups of elements will be small if the support level is lower.[56]. The way for algorithm works is that the algorithm from the frequent itemset will produce new candidates from the k-itemset and determine the support value of the k-itemset.An object that has a support value below Minsup will be removed. When no new generated item frequencies are generated, the algorithm ceases. The next step is calculated minconf from the results of frequent itemset. Support does not need to be seen again, as from viewing its minsup, the produced frequent itemset isobtained. If the rule that meets the limits and constraints defined is high, then the rule is classified as strong. The Apriori algorithm is iteratively processed and one object is first recognized by frequentitemset. [43]

Example of apriori algorithm:

The a transactional data containing product items that are purchased with various parameters as shown in the Table(2.2). The database is checked to classify all the frequent 1 item sets by counting each of them and collecting those that satisfy the minimum support threshold. Each frequent element set requires the entire database to be scanned until no more frequent k-itemsets can be identified.[57]

Table 2.2 Sample of transactional data[57]

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

The threshold for minimum support used is 2. Therefore, in the next step of algorithm processing, only records that reach a minimum support count of 2 will be used. According to Figure 2.6

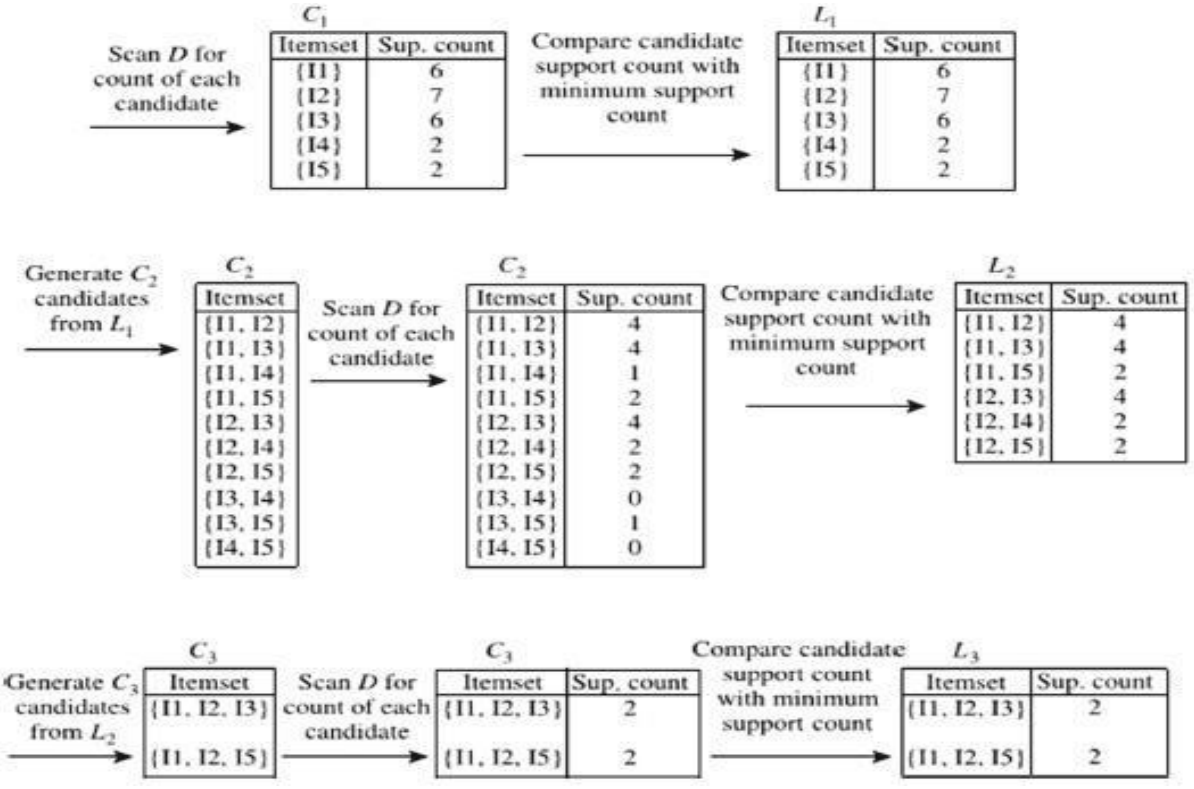


Figure 2.6 Generation of candidate itemsets and frequent itemsets[57]

The methodology for correlation analysis can be divided into two phases according to Kusrini and Luthfi:

1.High frequency pattern analysis :- At this phase , the search for theelements that meet the basic requirements value of the database support value,as shown in the formula:

$$\text{Support}(A) = \frac{Amount\ of\ transaction\ Contains\ A}{Amount\ of\ transactions} \quad \ldots\ldots\ldots (2.1)$$

The support value of 2 items, as shown in the formula:

$$\text{Support}(A \cap B) = \frac{Amount\ of\ transaction\ Contains\ A\ and\ B}{Amount\ of\ transactions} \quad \ldots\ldots\ldots (2.2)$$

2. **The establishment of associative rules:** After finding all the patterns with a high frequency, the associative rule that satisfies the minimum confidence is applied by computing the correlative rule A → B. And also the confidence value of the base A → B by the following formula [43]: -

$$\text{Confidence} = P(B|A) = \frac{Amount\ of\ transaction\ ContainsA\ and\ B}{Amount\ of\ transaction\ Contains\ A} \quad \ldots\ldots (2.3)$$

The code of the algorithm: we can establish an improved Apriori algorithm is as follows **[58]:**

Algorithm (2.1): The Apriori algorithm

***Input**: Database D, Mini Support ϵ, Mini Confidence ε*

***Output**: Rt All association rules*

1- L1 = large 1-itemsets;

2- for($k=2$; $L_{k-1} \neq \emptyset$; $k++$) do begin

3- Ck =apriori-gen(L_{k-1}); //generate new candidates from L_{k-1}

4- for all transactions $T \in D$ do begin

5- Ct=subset(C_k,T); //candidates contained in T.

6 - for all candidates $C \in C_t$ do

7- Count(C)=Count(C)+1; // increase support count of C by 1

8 - end

9- $L_k=\{C \in C_t \mid Count(C) \geq \epsilon \times |D|\}$

10 - end

11- $L_f = \cup k\ Lk$

12- Rt=GenerateRules(L_f, ε)

The Apriori algorithm include input Database D, Mini Support ϵ, Mini Confidence ε and output for determining frequent itemsets is used to produce Ck, does not exceed the support threshold all elements in Cn are ignored,Increase support by identifying candidate itemsets from CK.

2.8 Data Mining Stages for Recommender Systems

Recommendation systems use data mining and machine learning techniques that are given to retrieve information[59], and it can involve three stages: preprocessing stage, analysis stage and interpretation of results

2.8.1 Preprocessing Stage

Data pre-processing such as data cleaning, filtering, and transformation to obtain clean data suitable for later analyzes. The most widely used techniques are similarity measures or mathematical distance (such as Euclidean distance, cosine similarity and Pearson correlation coefficient), and sampling in addition to dimensionality reduction technique such as (Principal Component Analysis (PCA) and remove redundant information or Singular value decomposition (SVD)).[60]

2.8.2 Data Analysis

Data analysis includes two main objectives, which is predictive and descriptive. The predictive includes several techniques such as (k-Nearest Neighbors (KNN), decision trees, rules, Bayesian networks, Support Vector Machines (SVM)or artificial Neural Networks (ANN), while descriptive includes association rules as well as (k-means, density-based, message-passing based or using hierarchical approaches). [59].

(i) k-means clustering

The essence of today's business world is swift and competitive. It entails a large amount of data collected from various sources. K-Means Clustering – What is it, and how does it work. K-means clustering utilizes “centroids”, K different points are awarded at random in the data, and can be created data all point to the closest centroid. After each point is set, the centroid is transferred to the average of every of the points that have been allocated to it.

The k-means clustering technique tries to divide a given unknown data set (one that has no information about class identity) into a collection of k clusters.

A total of k centroids is picked at the start. A centroid is an imaginary or actual data point in the center of a cluster.

Each centroid in Praat is an existing datapoint in the supplied input data set, chosen at random, such that all centroids are unique (that is, for all centroids ci and cj, ci ≠ cj) and following that, each centroidis arranged to the arithmetic mean of a cluster it defines.

The classification and centroid correction process is repeated until the centroids' values stabilize. The finalized centroids will be utilized to construct the input data's final classification/clustering. [61].

D = √(((x1-y1)^2+(x2-y2)^2) .....(2.4)

Where

X1 = x-axis

Y1 = y-axis

Algorithm (2.2): K-means algorithm [62].

Steps for K-Means Clustering

Input: *Data set and the number of cluster k*

Output: *Solution to the k partition problem*

Step-1: *Determine the number of clusters K, which is an initial initialization step.*

Step-2: *Select k random points from the data as centroids.*

Step-3: *Determine all the points to the closest cluster centroid.*

Step-4: *Recomputed the centroids of newly formed clusters.*

Step 5: *Repeat steps 3 and 4.*

(ii) K-Nearest Neighbor Algorithm

K-Nearest Neighbor (KNN) is a method for classifying objects that involves learning data that is the most similar to the object and comparing it to prior and current data. In the learning process, KNN uses the Euclidean distance formula to determine the distance to the nearest neighbor, whereas other approaches optimize the distance formula by comparing it to other comparable formulas in order to get optimal results. [63]

Algorithm (2.3): K Nearest Neighbors [64]

Input: *Data set and the consists of the k closest data set.*

Output: *Class*

Start

Step1: The initial stage in the KNN process is to load the training and test data..

Step2: The value of K, i.e. the closest data points, must be chosen. Any integer can be used as K..

Step3: Using Euclidean distance, calculate the distance between test data and each row of training data.

End.

(iii) Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a type of model that can be used to solve classification and regression issues. It can solve both linear and nonlinear problems which is useful for a wide range of applications. A support vector machine is a machine learning model that can generalize across two classes if the algorithmis given in the training set, there is a set of labelled data. The SVM's primary role is to look for a hyperplane that can distinguish between the two classes. [65].SVM considers all of the data points and generates a line known as the 'Hyperplane,' which divides the two groups. The 'decision border' is the name given to this line. There are various hyperplanes to choose from, but the optimum hyperplane for separating the two classes is one with a big distance between the hyperplanes of both classes. The main goal of SVM is to locate the optimal hyperplanes. The hyperplane might be linear classifier or nonlinear classifier ,[66]. A straight line (hyperplane) can be used to classify data points from distinct classes in linear SVM. Soft margin SVM is beneficial when the training datasets are not totally linearly separable.When the data is not linearly separable, SVM employs kernel methods to make it linearly separable. Given a set of non-linearly separable training data, it can almost certainly be turned by projecting it into a higher-dimensional space via some non-linear transformation into a linearly separable training set. Kernel tricks are a method of determining the to make a dot product of two vectors , how much they influence each other. The probability of linearly non-separable data sets being linearly separable rise in increasing dimensions, according to Cover's theorem. Polynomial kernel, Gaussian kernel, sigmoid kernel, Radial basis function (RBF) kernel, and others are examples of well-known kernel functions.[67]. Figure (2.7) shows the optimal hyperplane and support vectors in SVM

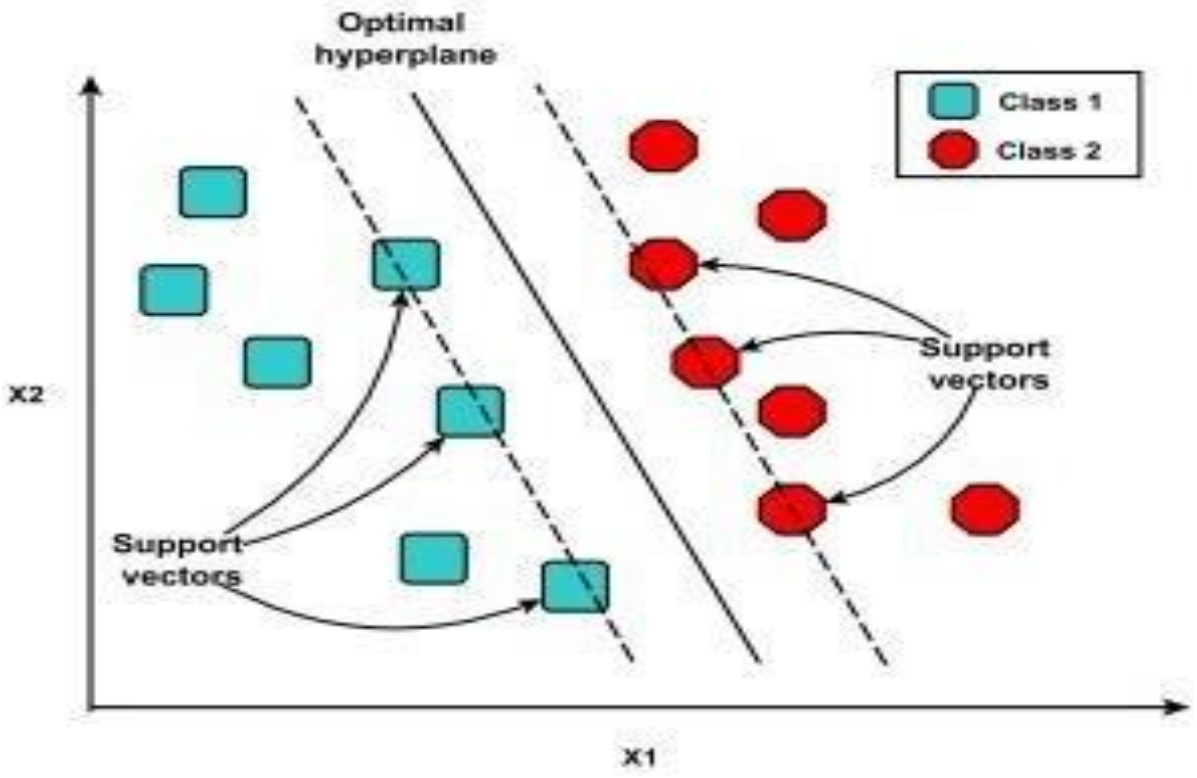


Figure 2.7 shows the optimal hyperplane and support vectors in SVM [68]

Algorithm (2.4): Support Vector Machine [69]

Input: *Data set*

Output: *Class*

Start

Step1: Load the data . for training the support vector machine classifier

Step2: The training data is separated into input and output

Step3: Shifting the input data . The shiftmean value is determined first before shifting the input data.

Step4: The input data is scaled. The scalefactor is computed by dividing one by each column's standard deviation.

Step5: The Kernel matrix is calculated after the Kernel function is chosen.

Step6: Calculation of the Support vectors.

End.

(iv) Long Short Term Memory (LSTM)

The LSTM is a type of recurrent neural network that can deal with long-term dependencies. Deep learning neural networks with long short-term memory (LSTM) are one sort of deep learning model. dependencies and is thus commonly used for time-series predictions.

The LSTM is excellent tool for solving sequence prediction problems and are well-suited to capturing user taste progression. Long Short Term Memory network is an advanced RNN, or sequential network, that permits information to persist in order to be applied to recommendation .RNNs work in a similar way [70].The LSTM is made up of three sections, each of which serves a distinct purpose

The Forget gate is the first step in the process, and it determines whether the preceding timestamp's information should be remembered or ignored. The second part of the cell is the input gate, which tries to learn new information from the input. In the third section, known as the Output gate, the cell delivers updated information from the current timestamp to the next timestamp. An LSTM cell is made up of three parts: gates, switches, and transistors. An LSTM, like a standard RNN, has a hidden state, with H(t-1) representing the hidden state of the previous timestamp and Ht representing the hidden state of the current timestamp. LSTMs also have a cell state, which is represented by C(t-1) and C(t) for past and current timestamps, respectively.

Short term memory refers to the hidden state, whereas Long term memory refers to the cell state.[71]. Figure (2.8) shows LSTM (Long Short Term Memory Network) cells. It primarily contains a memory state for storing information as well as methods for adding and rejecting data during the learning process.[72]

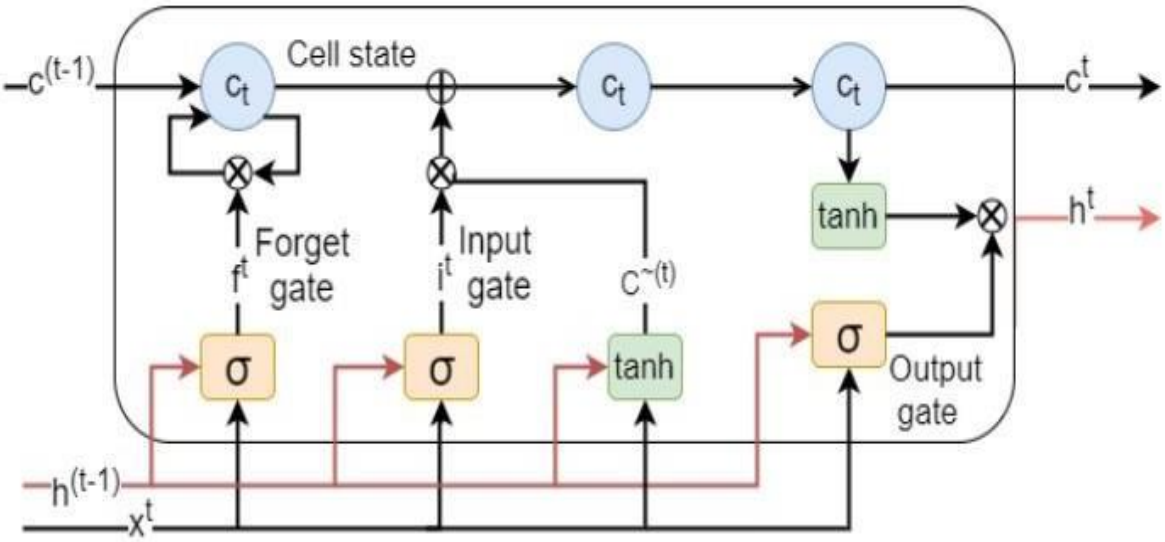


Figure 2.8 The basic elements of Long Short Term Memory (LSTM) cell [72]

RNNs can use the LSTM structure to keep memory over longer periods of time. By implementing new gates, input and forget gates, it overcomes the problem of gradient vanishing (or gradient explosion). These extra gates might give more control over the gradient, allowing to choose what information to keep and what to discard. These gates are sigmoid functions with outputs in the range [0,1], and they can pass either limited or all information. A value of zero means the information is totally filtered out, whereas a value of one means the information is completely passed through. Because it uses short-term memory processes to build longer memory, the structure is dubbed Long Short-Term Memory. In addition to the hidden layers [73].

(LSTM)

Algorithm (2.5): Long Short Term Memory (LSTM) [74]

Input: Dataset

Output: Class

Start

Step1: Data Preparation, Split the dataset for training stages and testing stages.

Step2 Training LSTM Network , Design of LSTM Choice of parameters such as

hidden layer, output layer with activation, choice of parameters

function and optimizer , choice batch size and epochs.

Step3: LSTM Prediction , to predict the test data, use a trained model.

Step4: Calculation of, Precision, Recall, F1-score, and Support are calculated using an impalement model.

Step5: Evaluation by calcte the accurcy of recommender systems .

End.

2.9 Gaussian Mixture Modelling (GMM)

A Gaussian Mixture Model (GMM) is a probabilistic data clustering technique in which all data points are drawn from a mixture of finite Gaussian distributions with unknown parameters. Data clustering, image coding tracking, and activity classification are all possible uses of GMM-based modeling. GMM is a parametric probability density function that can be expressed mathematically as a weighted sum of M Gaussian components.[75]

The GMM is a weighted sum of Gaussian component densities that represents a parametric probability density function. GMMs are often employed in biometric systems as a parametric model of the probability distribution of continuous measurements or characteristics.

At its most basic case, GMM is a clustering technique in which each cluster is simulated using a separate Gaussian distribution. Rather than having hard assignments into clusters like k-means, it is have soft assignments with flexible and probabilistic approach to data modeling [76] .A Gaussian distribution's probability density function is calculated by:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{.....(2.5)}$$

Where: μ is represent the mean and σ2 is the variance.

The iterative Expectation-Maximization (EM) technique or Maximum A Posteriori (MAP) estimation from a well-trained prior model are used to estimate GMM parameters from training data will explain later.[77].

The Expectation Maximisation (EM) Algorithm: The Expectation-Maximization (EM) algorithm is a standard inferential method and a potent tool for fitting Gaussian mixture models to observed data. Its purpose is to optimize the log-likelihood function in relation to the parameters of the model (mean, covariance, and prior probability). The EM algorithm is an iterative procedure thatconverges to the maximum likelihood estimation after two key steps: expectation and maximization. [75] **There are two steps in the Expectation-Maximization algorithm:-**

a. **Expectation step (E – step):** In this stage, we try to estimate the values of the missing data using the dataset's observed accessible data. Finally, we get complete data with no missing values as a result of this phase.
    1. Give some random values for the mean μk, covariance matrix Σk and the mixing coefficients πk
    2. After that, estimate the value of the latent variables for the provided parameter values(γk)
b. **Maximization Step:** The complete data is utilized to update the parameters based the estimated values generated in the E-step i.e, update the hypothesis.

Many algorithms, like Gaussian Mixture Models, are based on Expectation-Maximization.) Update the parameters' values (μk, Σk, and πk).[78]

The code of the algorithm [78]

Algorithm (2.6): Gaussian Mixture Modelling (GMM)

Input:Dataset

Output:Classifier

1.Initialize the mean $\boldsymbol{\mu k}$*, the covariance matrix* $\boldsymbol{\Sigma k}$ *and the mixing coefficients* $\boldsymbol{\pi k}$ *by some random values(or other values)*

2.Compute the $\boldsymbol{\gamma k}$ *value for all* $\boldsymbol{k}$*.*

3.Again Estimate all the parameters using the current γk *values.*

4. Compute log-likelihood funcation **.**

5.Put some convergence criterion

6. If the log-likelihood value converges to some value (or if all the parameters converge to some values) then ***Stop ,***

Else return ***Step 2 .***

2.10 Performance Measuring of the Recommender systems

The statistical accuracy tests are used to determine how well an RS can predict the ratings of all user item pairs[79] . Precision, recall, and F1-measure are just a few of the metrics used to determine the efficacy of classification systems. The confusion matrix is used to estimate a variety of measures[80] .Some of the decision support measures arelisted below:- [79] .

1. **Precision:-** it evaluates the algorithm's ability to produce accurate recommendations. As seen in the equation below, the user consumed fraction of the recommended products.

$$Precision = \frac{TP}{TP+FP} \quad \text{.....(2.6)}$$

2. **Recall :-** it is a statistic that's commonly used in recommender systems to evaluate the top-number of recommendations. It calculates what percentage of the things enjoyed by users were recommended by the algorithm, as shown in equation.

$$RECALL = \frac{TP}{TP+FN} \quad \text{.....(2.7)}$$

3. **F1 measure:** which includes precision and recall ,the accuracy is measured using the F1-Measure as shown in the following equation.

$$F\_MEASURE = \frac{2*recall*precision}{recall+precision} \quad \text{.....(2.8)}$$

4. **Accuracy:** The number of correct predictions is divided by the total number of predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots\dots (2.9)$$

2.11 Summary

The chapter presents a study of the combination of machine learning and deep learning, which has been researched as a method to achieve recommender systems.

This chapter explains some techniques and types of recommendation systems and an algorithm to find out how much the user likes a particular product. Through the use of classification techniques and building blocks. Therefore, some details of the algorithms used in this paper are presented. Also, the apriori algorithm and association rules between products are described.

Chapter Three

Implementation of the Proposed System

Chapter Three

3.1 Introduction

After we explained the concept and structure of recommendation systems in the first chapter, as well as the concept and structure of market strategies and the most important challenges in the second chapter. This chapter will illustrate the most important algorithms that were used in marketing. Additionally, the machine learning algorithms have used with deep learning algorithms in addition to the apriori algorithm which is used to mine recursive elements to extract data, and association rules., as well as combining techniques K-means with KNN, K-means with SVM, K- means with LSTM, as well as GMM with knn, GMM With SVM, GMM With LSTM .The purpose of combining algorithms is to achieve higher performance and optimal results

3.2 The Proposed System

This section describes several algorithms for the recommendation system used in the market, where the phases of the system are illustrated with algorithms, to find similarities, the prediction process, and finally the recommendations. The results of the stages and the applied algorithms will be presented in the fourth chapter,where several algorithms were used, including mining of association rules to generate data in addition to other algorithms. The Dataset has been used and loaded from the Kaggle site, which is a group of products sold in Amazon for Modcloth. This dataset consists of 998,94 of transaction records. Then the preprocessing stage, such as Cleaning, Reduction, Integration, add to Normalization and these methods will be explained in detail. As well as dividing the data into two groups as a Training data set and a testing data Set Passing through the analysis by applying the proposed algorithms and then arriving at the results. The proposed algorithms then access the results. To determine which of these algorithms will give higher accuracy. Figure (3-1) shows a flowchart with the stages (steps) of the proposed RS system in detail.

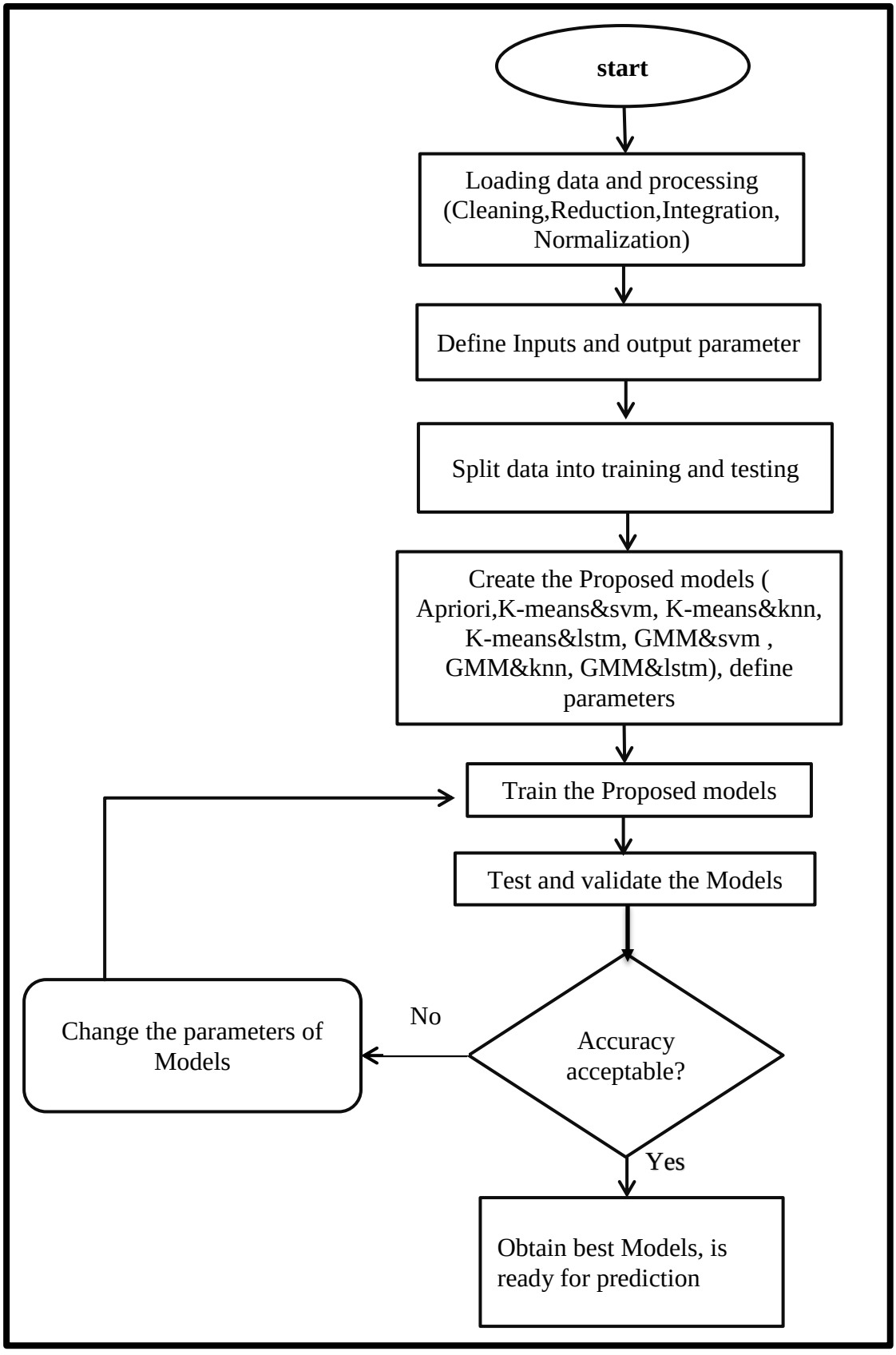


Figure 3.1 Flowchart of the Proposed System

3.3 Dataset Information

These datasets contain details products offered on ModCloth Amazon that can lead to recommendations that are biased (in particular, characteristics relating to the marketing of the products). Interactions between users and items are included in the data. The data was collected from Kaggle and consisted of 998,94 of transaction records ,handled as a CSV file(Comma Separated Values)for efficient use in the Python framework, dataset features include user_attr , user_id, item_id , size, fit , rating ,timestamp , category,brand ,year ,split. Some of these data such as category ,rating, user_attr and other that were used in the model may be of a text data type that cannot be deal with, so it is converted to be used in the model. Use paradigms, methodologies, andalgorithms to recommend the appropriate product to the appropriate users at the appropriate time. That is, encoding it with 0 or 1, and null in order to deal with it and apply the proposed algorithms .

3.4 Dataset preprocessing

This process paved the way for project management to be used later in the data analysis step, where many operations are performed on the data including splitting the data set, removing duplicates, extracting unique elements, splitting the data into training data and test data, standardizing data Feature Scaling etc. The training data takes 70% of the original data and 30% of the test data. Data preprocessing is one of the cornerstones to achieve quality, this section explains some of the concepts that were used includes four stages:-

a- **Data Cleaning:** clean and prepare the data, by processing incomplete data, duplicate data, outliers, mismatched data.

b- **Data reduction:** Data reduction help in analyzing the low representation of a data set without compromising the integrity of the original data and producing qualitative knowledge such as data cube aggregation technique, data compression, differentiation and hierarchy creation, in reducing the number data integration. Data compression is used.

c- Data Integration: involved in the data analysis task that collects as is the case with data warehousing, data from many sources is combined in a single data warehouse. Multiple databases, data cubes, and static files are examples of these sources.

d-Data Normalization: Any data assertion when there are complex, repetitive, random data that is converted from a complex structure to a simple matrix. The method of normalization or encoding is to get rid of useless data such as removing duplicates and others.

Algorithm (3.1) Explains the steps of the Dataset Preparation

Algorithm (3.1): Dataset Preparation

Input: dataset

Output: preprocessing dataset

Start

Step1: loading modcolth of amazon dataset.

Step2: define the data type and read the data sets.

df = pd.read_csv('df_modcloth1.csv')

Step3: Remove duplicate and empty values and encode values for some fields

Step4: The dataset's features are extracted.

Step5: For dataset classification, run the Scikit-learn program.

Step6: split the dataset into train data and test data.

Step7: apply the usedmodels on these variables in training dataset.

End.

3.5 Cross Validation

It is a method used to make the data fit and accurately with the model that is built in machine learning., In other words, it is not to overfitting by dividing the data into training data that is applied to it and test data to calculate the resulting error, in addition to a third group investigation (validation).It is used to adjust inputs such as the number of hidden layers or the activation function in the LTSM algorithm, as well as the kernel used in SVM and also used in classification.

Discrete validation is used when we have a lot of data where the best decisions can be reached for the algorithms used or the data used It is a powerful tool about the data used as well as gives additional information about thealgorithms.

3.6 Feature Selection

The main purpose of market recommendation systems is to analyze the market basket by identifying the right advantage When we have a huge amount of data, identifying the features is a basic approach by selecting the most suitable data, because excess data affects decision-making, in short, less data helps algorithms train very quickly. Here the data set of ModCloth company was used in the Amazon, which contains 12 features of the data set, including (category, user_attr, item_id, split , user_id, model_attr, brand, year timestamp, rating, , fit, size), among the reasons that help us to feature selection:

1. Reduces overfitting
2. Train Algorithms Faster
3. Reduce model complexity and improve model accuracy.

3.7 Scikit- Learn

The dataset can be divided and classified into training data and test data through an important library in Python "Scikit-Learn" which is used in the classification, node and regression algorithms. In addition to being used in the stage of data processing and evaluation of models through the use of (sklearn.model). It is one of the most important offices used in machine learning, as it is used with other offices such as Scipy, Numpy and Matplotlib, as we mentioned earlier, through this library, the data is divided into training data and test data.

3.7.1 Training Phase

The data is divided into training data and test data. In our work, the data is divided into 70% for training and = 30% for testing Because of getting the best results , where take the training data set the bulk of the division to build the model, then the model evaluates the data repeatedly to learn more about the behavior of the data .

3.7.2 Testing Phase

Testing the data once the model has been developed checks its capacity to make correct predictions. Test data is an invisible data set that serves as a last real-world check to ensure that the algorithms have been properly trained.

3.8 Confusion Matrix

After loading the data, the data cleaning and processing were performed to build an excellent model, the confusion matrix is a metric for measuring recall, specificity, and accuracy.

Get the probability output. The confusion matrix is important and necessary in machine learning. The confusion matrix contains four possibilities: True Positive(TP), True Negative(TN), False Positive(FP), and False Negative(FN). The classifier creates a set of predictions, the confusion matrix makes a tabular summary of the number of correct and incorrect predictions, in addition, it evaluates the performance of the classifier model by calculating performance measures such as accuracy, recall and F1 score.

3.9 Proposed Algorithms for Markets Recommendation

In this thesis , several techniques were used to measure the accuracy of the recommendations (Associative principles and a clustering algorithm), where the implemented to do so, the a priori algorithm was utilized find the association rules, in addition to using a proposed model such as K-MEANS or Gaussian Mixture Models(GMM) with the proposed algorithms, as explained in detail. The purpose of combining algorithms is to achieve higher performance and optimal results.

3.9.1 The Apriori Algorithm

The Apriori algorithm was implemented to thesis for import an tassociation rules from a given dataset in df_modcloth1 datasets as csv file. This algorithm has three main elements: support, confidence and lift. It first detects groups of repeated elements that have support and confidence values above a predefined threshold value. In this algorithm we import pandas DataFrame reading the dataset and pre-proccessing the data, Transaction Encoder will convert the transactions into an encrypted DataFrame, where each column consists of TRUE and FALSE values that indicate whether an item is included in a transaction or not,And we do a training process. Then apply the Apriori algorithm and take min_support and on the basis of it determine the frequency items. Its purpose is to find important links in data.

Pre-requisites: pandas, from mlxtend.frequent_patterns import apriori, association_rules, TransactionEncoder

Algorithm (3.2) Explains the steps of the Apriori algorithm

Algorithm (3.2): Apply Apriori algorithm

Input: Dataset

Output: All Association Rules

Start

Step1: Importing the necessary libraries to the systemImport pandas as pd

from mlxtend.frequent_patterns import apriori, association_rules

from mlxtend.preprocessing import TransactionEncoder.

Step2: Data Loading and exploration After downloading the data, it is read

df=pd.read_csv('df_modcloth1.csv')

Step3: Data Proprocessing dataset must be in the form of a list of lists in order to use the apriori ,through a series of procedures

Step4: Splitting the data according to transaction

Step5: Encoding the Data

Step6: Building the models and applying AprioriModel Execution min_support = 0.00005, Frq_items = apriori(df, min_support = 0.00005, use_colnames = True)

Step7: Analysing the results

End.

In this model, we load the data and read it by importing pandas from mlxtend to deal with apriori, association_rules, and TransactionEncoder, then do the preprocessing and encode the values such as missing values, convert them to zero, etc., building the model The first step is to do the TransactionEncoder process and do a training process after which the algorithm is applied apriori we extract min_support and on its basis we note the repeated values for association_rules.

3.9.2Proposed Recommender Systems

Here we will used two types of logistic regressions for determine which number of variables achieve best accuracy.

A. The Classifier KNN Model

The model is built by applying the Proposed approach by collecting K-Means with KNN or combining GMM with KNN after data processing and performing a series of transformations where the number of neighbors 50 is applied.

Algorithm (3.3) Explains the steps of the K-Nearest Neighbor(KNN)

Algorithm (3.3): Apply GMM and K-Means on KNN

Input: Dataset after preprocessing

Output: Class

Start

Step1: Apply the GMM and K-Means according algorithm(2.3)where theGMM cluster the dataset to three groups .

Step2: Split the dataset to training and testing based on cross validation, rate 30 and 70 .

Step3: Prepare model KNN detrmine the number of neighbors 50

Step4: Build the model according the training phase .

Step5: Test phase pridect the x test based on model build in step 4 .
Step6: Evaluation by calcte the accurcy of recommender systems .

End.

In this model, we load and read the data by importing Pandas and Numpy and from Sklearn.mixture to handle Gaussian Mixture and from Sklearn.metrics to handle measurement evaluation accuracy_score, f1_score, precision_score recall_score, classification_report, confusion_matrix then read modcloth1.csv and Transaction Encoder values and then do the operation As well as pre-treatment. Thendo training and testing based on cross validation, rate 30 and 70, then we apply K-means, and we also apply GMM to the model by taking the number of clusters at three and taking KNN detrmine the number of neighbors 50. We make a performance scale for the model with confusion_matrix, classification_report, and accuracy_score.

B. The Classifier Support Vector Machine (SVM)

The dataset is used to build the second model and evaluate its performance and accuracy using the support vector machine algorithm. In this work, a model was prepared to increase the accuracy of the recommendations by combining SVM and K-means methods as well as combining with the used GMM model with three clusters. The data was divided and the training data was in SVM using Linear Kernel, where a series of mathematical operations are used to process the data. To implement Kernel functions, the 'scikit-learn' library must be called. The train the classifier using our training data. Before training, will need to import df_modcloth1 datasets as csv file where we will train five features out of all features.

Pre-requisites: Numpy, Pandas, Scikit-learn, from Sklearn.mixture import GaussianMixture and from sklearn.cluster import KMeans

Algorithm (3.4) Explains the steps of the Support Vector Machine (SVM).

Algorithm (3.4): Apply GMM and K-Means on SVM

Input: Dataset after preprocessing

Output: Class

Start

Step1: Apply the GMM and K-Means where the GMM cluster the dataset tothree groups according algorithm(2.4) .

Step2: Split the dataset to training and testing based on cross validation, rate 30 and 70 .

Step3: Prepare model SVM using Linear Kernel

Step4: Build the modelaccording the training phase .

Step5: Test phase pridect the x test based on model build in step 4 .

Step6: Impalement model to calcte the Confusion Matrix ,Precision, Recall, F1-score and Support .

Step7: Evaluation by calcte the accurcy of recommender systems .

End.

In this model, we load and read the data by importing pandas and numpy and from sklearn.mixture to handle GaussianMixture and from sklearn.metrics to handle measurement evaluation accuracy_score, f1_score, precision_score recall_score, classification_report, confusion_matrix then read modcloth1.csv and TransactionEncod values and then do the operation As well as pre-processing. And from sklearn we import SVM and then do the data partition

Then we apply K-means, and also apply GMM to the model by taking the number of clusters three, we implement SVM and take Kernel type is Linear, we do data fitting and training to build the model.

C. The Classifier Long-Short Term Memory (LSTM)

The dataset is used to build the model, and the calculation and evaluation of model performance and accuracy was performed using the Long-Short Term Memory algorithm, that field of deep learning. In this work, a model was prepared to increase the accuracy of the recommendations by combining LSTM and K-MEANS methods as well as combining with the used GMM model with three clusters. The data is divided into training data and test data using cross validation.

The dataset is read and preproccessing is done. In this work, LSTM was dealt with on three layers, and applied to five features. This model is sequential and the basis of the LSTM's work. The common activation functions are log-sigmoid and hyperbolic tangent.

Pre-requisites: Numpy, Pandas, Matplotlib, Tensorflow, from Sklearn.mixture import GaussianMixture and from sklearn.cluster import KMeans.

Algorithm (3.5) Explains the steps of the Long-Short Term Memory(LSTM)

Algorithm (3.5): Apply GMM and K-Means on LSTM

Input: Dataset

Output: Class

Start

Step1: Apply the GMM and K-Means algorithm, where the GMM divides thedataset into three groups algorithm(2.5)

Step2: Split the dataset to training and testing based on cross validation, where rate 30 and 70 .

Step3: Prepare model Long-Short Term Memory (LSTM) ,Where LSTM was dealt with on three layers, and applied to five features.

Step4: Design the model in accordance with the training phase.

Step5: Test phase pridect the x test based on model build in step 4 .

Step6: Confusion Matrix, Precision, Recall, F1-score, and Support are calculated using an impalement model.

Step7: Evaluation by calcte the accurcy of recommender systems .

End.

In this model, we load and read the data by importing pandas and numpy and from sklearn.mixture to handle GaussianMixture, K-means and from sklearn.metrics to handle measurement evaluation accuracy and accuracy_score, f1_score, precision_score, recall_score, classification_report, confusion_matrix and modelsorflow.keras. The form is Sequential, then reads the modcloth1.csv and TransactionEncod values as the values are encoded, after which we preprocess. The model is built once with GMM and again with K-means, here the model is applied to three layers.

3.10 Evaluation of Model

To evaluate the model, the dataset must be divided into two parts: the training data and the test data. Our dataset was divided into 70% of the data for training and 30% of the data for testing. Special scales have been used to evaluate the special model, while the result of each model is calculated and the confusion matrix is also calculated. Accuracy recall, f1 score, support, accuracy. Proposed methods such as K-mean-SVM, K-mean-KNN and K-MEANS-LSTM were used.

For classification predictions, different types of results can occur:

(1) True positives occur when we predict that an observation belongs to a class,and it turns out that it does.
(2) True negatives happen when we predict that an observation does not belong toa class, and it does not belong to that class.
(3) False positives happen when we incorrectly guess that an observation belongs to a class when it does not.
(4) False negatives happen when we incorrectly predict that an observation does not belong to a class when it does.

On a confusion matrix, these four results are plotted. Accuracy, precision, and recall are the three primary metrics used to evaluate a classification model. These are illustrated in Chapter Two (2.10) . Figure 3.2 represents the Evaluation model.

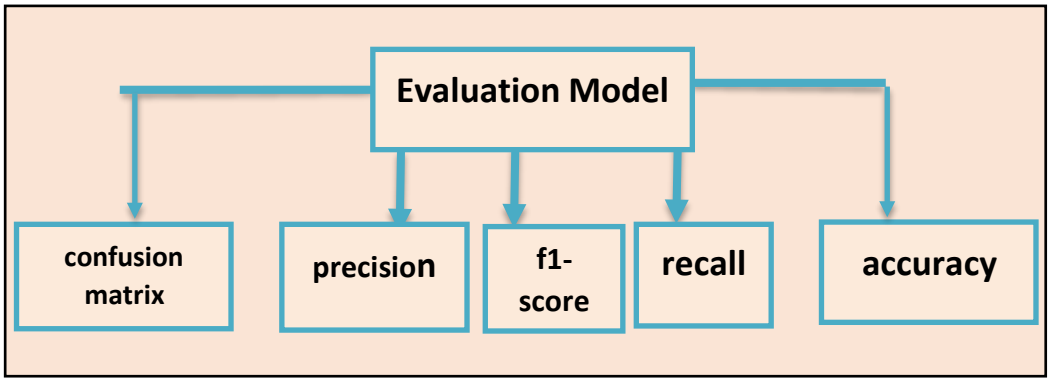


Figure 3.2 Evaluation Model

Chapter Four

Results Comparison and Discussion

Chapter Four

4.1 Introduction

In this chapter, we will present the results and discuss the evaluation and performance of ourwork based on the experimental results. The experiment was conducted ona dataset that was previously used in other related works as standard case study.

4.2Hardware Specifications

software that have been used to implement the results successfully, and sufficient memory of more than 16 GB due to consume time in the CPU, a method that uses Compute Unified Device Architecture (CUDA) to call up GPU resources for the training process acceleration. The platform specifications and all details are illustrated in Table (4.1).

Table 4.1: Hardware and software platform.

Name	Specification
CPU	Intel®Core™i7-9750H CPU@ 2.60GHz
RAM	DDR 4 16GB
OS	Windows 10 Home 64-bit
Framework	Python, minianconda, pycharm
Library	numpy, pandas, sklearn, apriori, tensorflow, keras, matplotlib

4.3 Experimental Data

One of the most common applications of association rules is the market basket analysis. When a customer buys a thing, applying association rules can predict what the additional products he or she will place in the basket with a high degree of certainty.When store's managers know which products are frequently purchased together, theycan order the shelves accordingly. Customers will be able to access these products more easily as a result.

As a result, there is a rise in sales rates and the development of effective sales strategies. The following is the dataset that will be utilized for the Marketing Modcloth Thesis:

- The dataset contains a total of 99894 samples.
- The dataset contains a total of 12 different features.

Multiple variables from products sold on both Amazon and Modcloth are included in the dataset, which can be used to demonstrate Amazon's marketing bias.

It will be retrieved from the following source: Another focus will be on the Collaborative-based strategy, which will be applied to a specific group of items throughout the dataset collected from the previously described source. Table 4.2 below shows some of the dataset's fundamental statistics.

Table 4.2: show the basic statistics of the Modcloth dataset

Description	ModCloth
Total number of ratings	99,893
Number of users	44,783
Number of items	1,020
Maximum number of ratings by one user	250
Minimum number of ratings by one user	1
Median number of ratings by one user	1

Data on the Marketing Modcloth: databases contain characteristics about products sold on ModCloth Amazon that could bias in recommendations (in particular, attributes about how the products are marketed). For recommendation purposes, data also includes user/item interactions. (item id, user id, rating, timestamp, size, fit, user attr, model attr, category, brand year, split) Metadata is (category ,item id, model attr, user attr, user id, split, rating, timestamp, brand ,size, fit, year,)

Table 4.3: show the sample of the Modcloth Amazon dataset

item_id	user_id	Rating	…	model_attr	category
7443	Alex	4	…	Small	Dresses
7443	carolyn.agan	3	…	Small	Dresses
7443	Robyn	4	...	Small	Dresses
.					
.					
.					
.					
71607	Amy	3	…	Large	Outerwear
119732	Sarah	3	…	Small	Dakota

4.4 Preprocessing

The initial stage in data processing is to see if the dataset contains any missing values, outliers, or unknown variables, and if so, try to remove them from the dataset so that the exploratory data analysis and modeling phases are not hampered. To prepare the dataset, take the following steps:

• Load the library and tools: For this stage, following libraries will be imported from Python to begin working on the data available inside the set plannedto be used for this specific study.

• Missing values: The following input displays any missing values; however, when this dataset was examined for missing values, it turned out to have none, which was a reassuring factor that both ensured that the dataset values were fine to go and that there was no need to filter out the missing data.as shown in table (4.4)

Table 4.4: The number fill missing value before and after fill missing

Colum Name	No. null value	No. null after fill missing value
item_id	0	0
category	0	0
user_id	1	0
rating	0	0
timestamp	0	0
size	21760	0
fit	18506	0
user_attr	8367	0
model_attr	0	0
brand	73980	0
year	0	0
Split	0	0

4.5Visualize Data

Checking the distribution of the ratings attribute with the following code input reveals that customers typically give purchased products a positive rating, while users seldom give products a negative rating, as demonstrated in Figure 4.1.

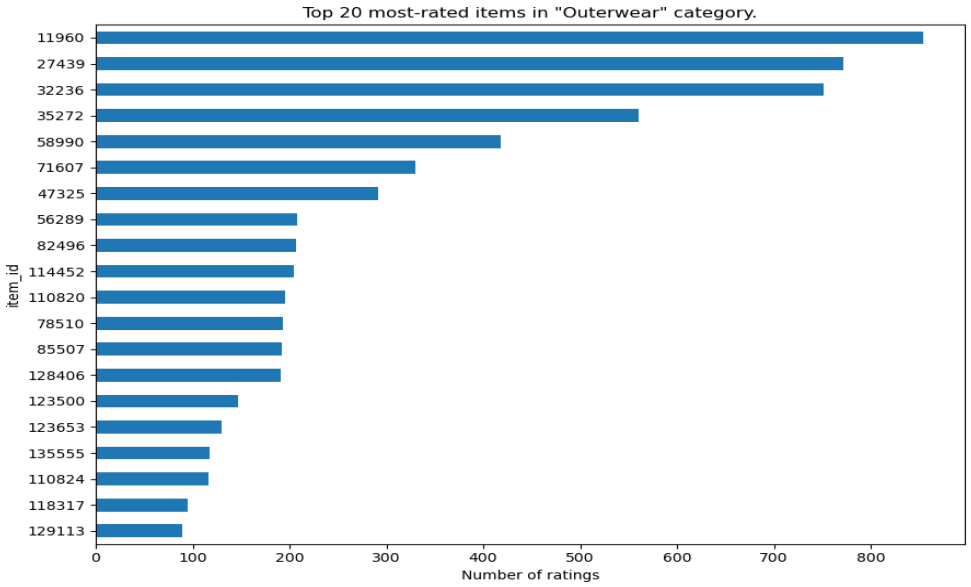


Figure 4.1: Top 20 highest-rated items in "Outerwear" category

Checking the distribution of the ratings attribute with user-id the following Top 20 users with most ratings of all time that customers typically give purchased products a positive rating, , as demonstrated in Figure 4.2.

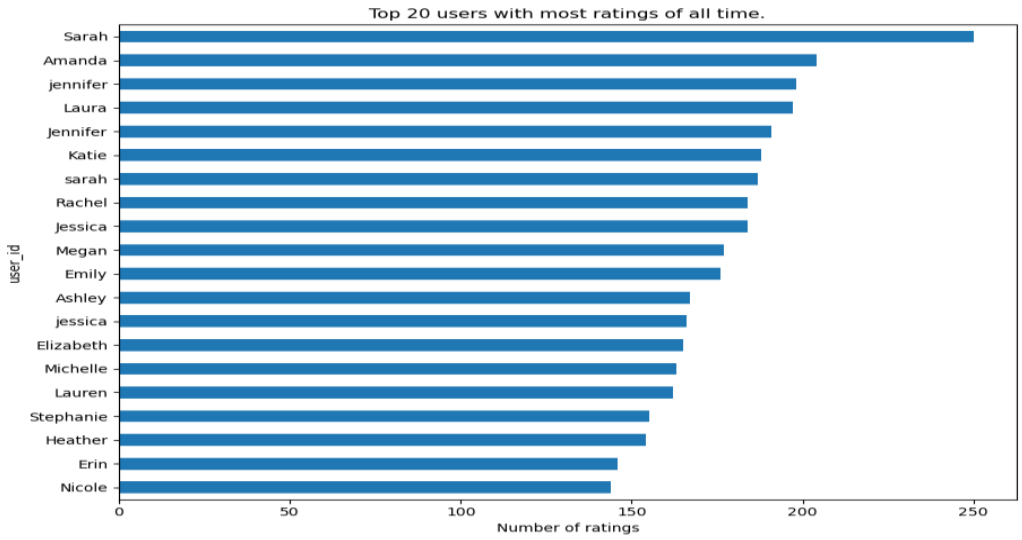


Figure 4.2: Top 20 users with most ratings of all time

Checking the distribution of the ratings attribute with item-id the following Sales percentages by brand in 2016, that customers typically give purchased products a positive rating, as demonstrated in Figure 4.3.

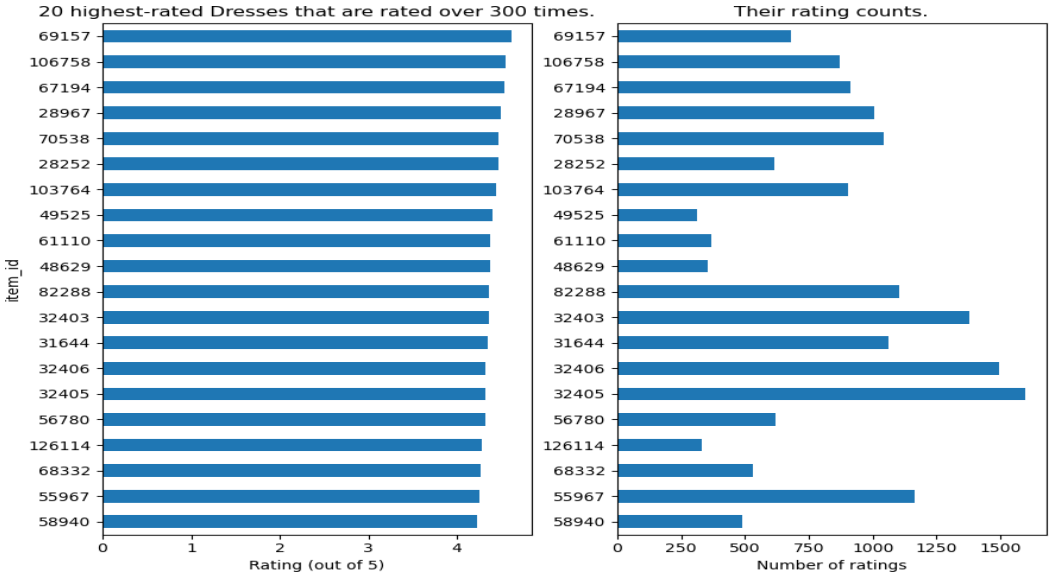


Figure 4.3: Sales percentages by brand in 2016

Checking the distribution of the size with frequency, the following All-time size distribution of items that mostly fitted customers , that customers typically give purchased products a positive rating, , as demonstrated in Figure 4.4.

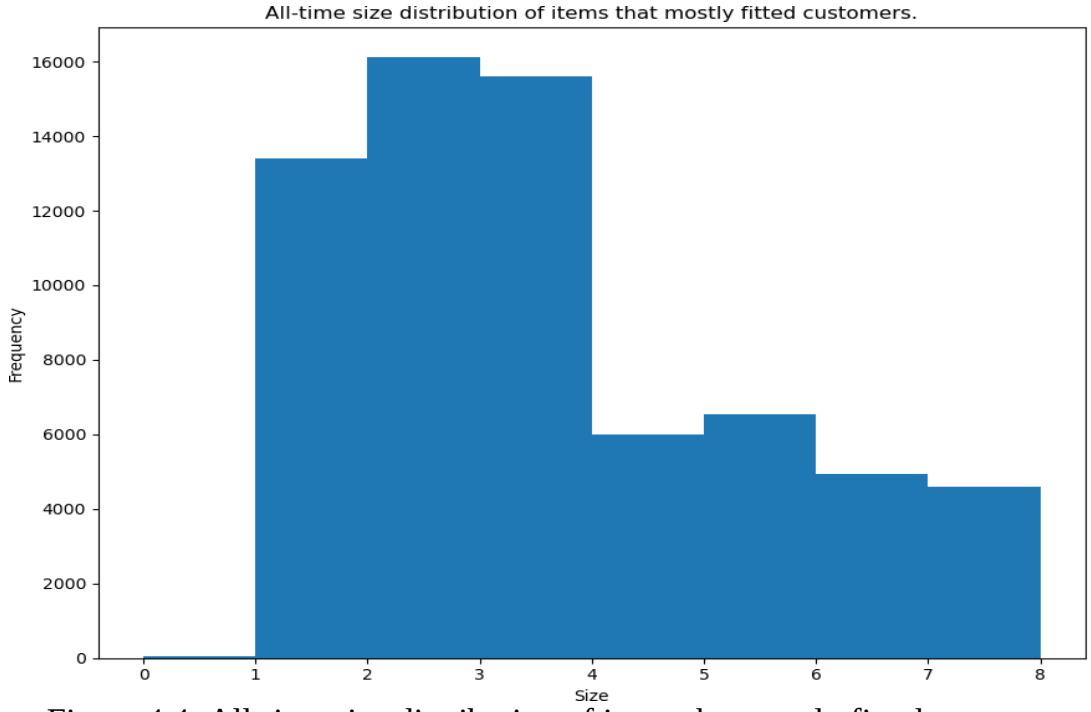


Figure 4.4: All-time size distribution of items that mostly fitted customers.

Checking the distribution of the Time with Number of sales, the following Sales trends of 3 biggest competitors between Jan. 2011 and Jun. 2019 , that customers typically give purchased products a positive rating, , as demonstrated in Figure 4.5.

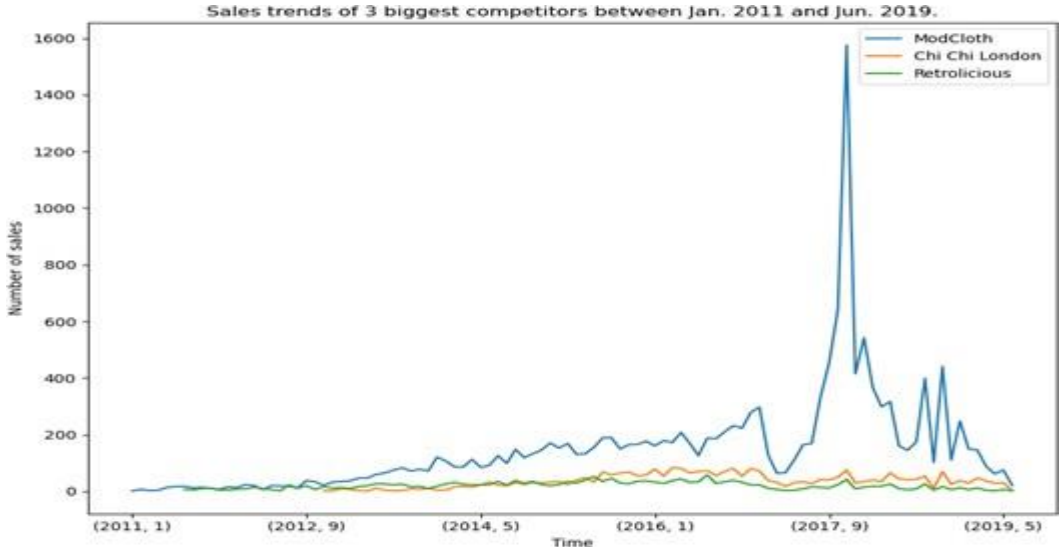


Figure4.5: Sales trends of 3 biggest competitors between Jan. 2011 and Jun. 2019.

Sales percentages by brand in 2016, Such as sales of ModCloth, Chi Chi London, Retrolicious, Steve Madden, ...other as demonstrated in Figure 4.6

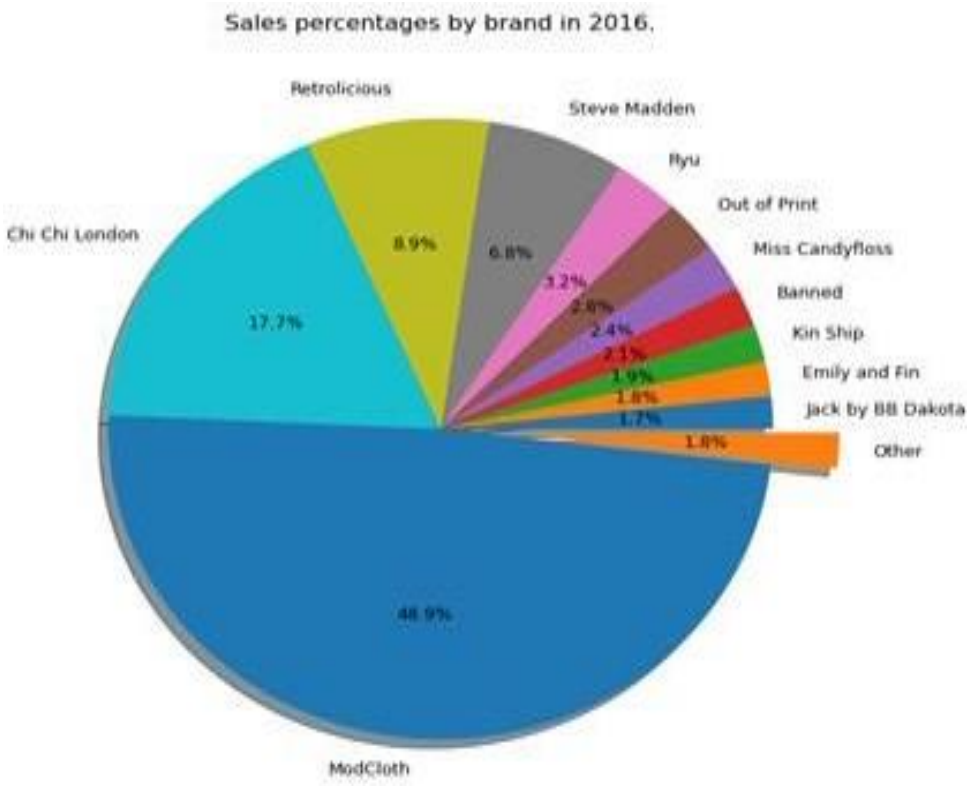


Figure4.6: Sales percentages by brand in 2016

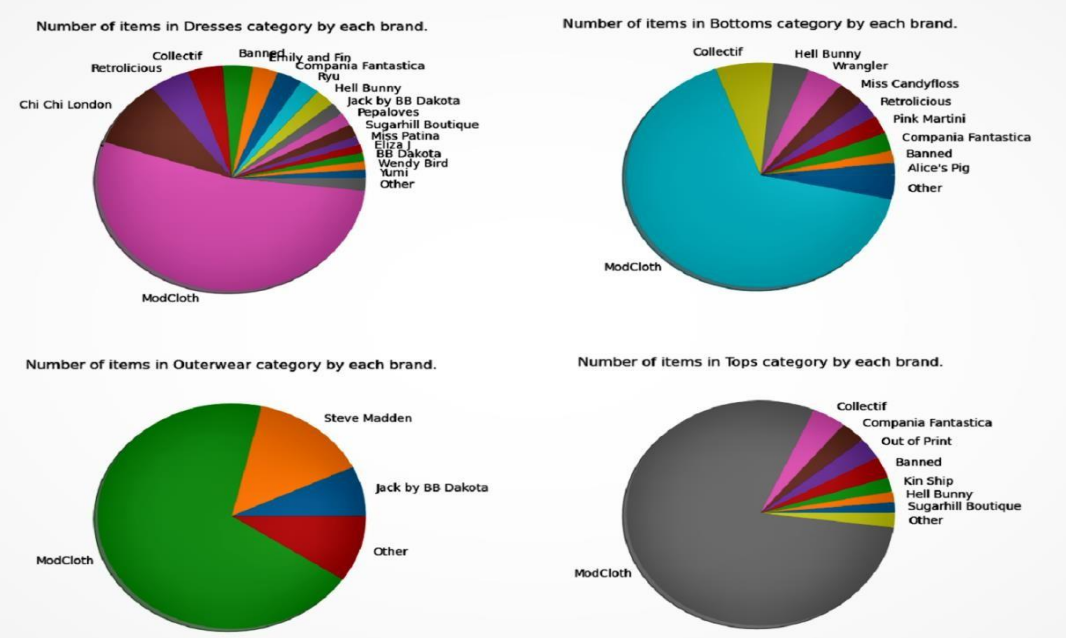


Figure 4.7: Number of items in category by each brand

The figure 4.7 selects the Number of items in category by each brand , Such as Tops ,Bottoms,Dresses,Outerwear.

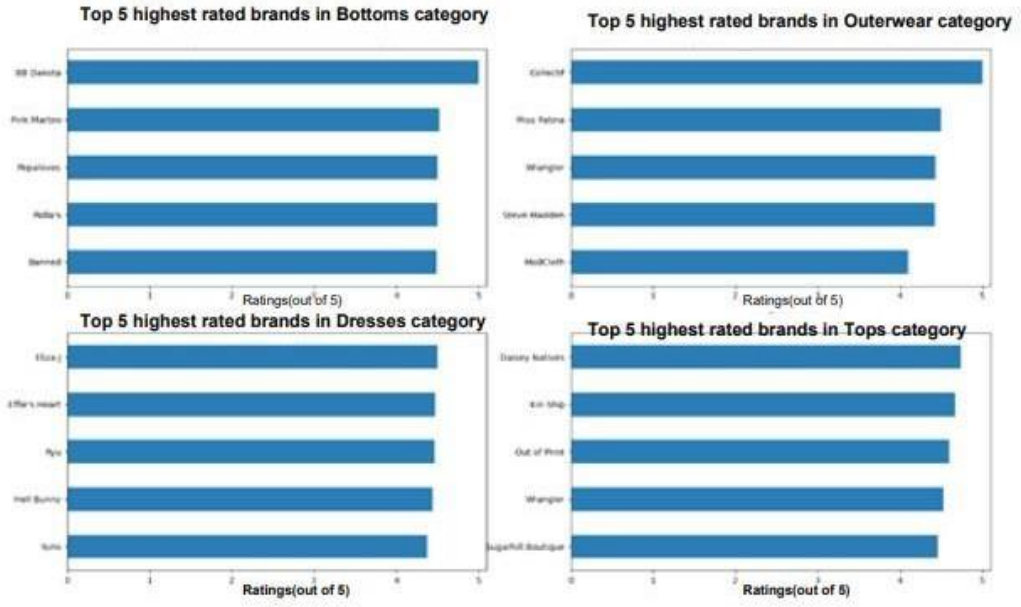


Figure 4.8: Top 5 highest-rated brands in category.

The figure above selects the Top 5 highest-rated brands in category, Such as Tops ,Bottoms,Dresses,Outerwear.

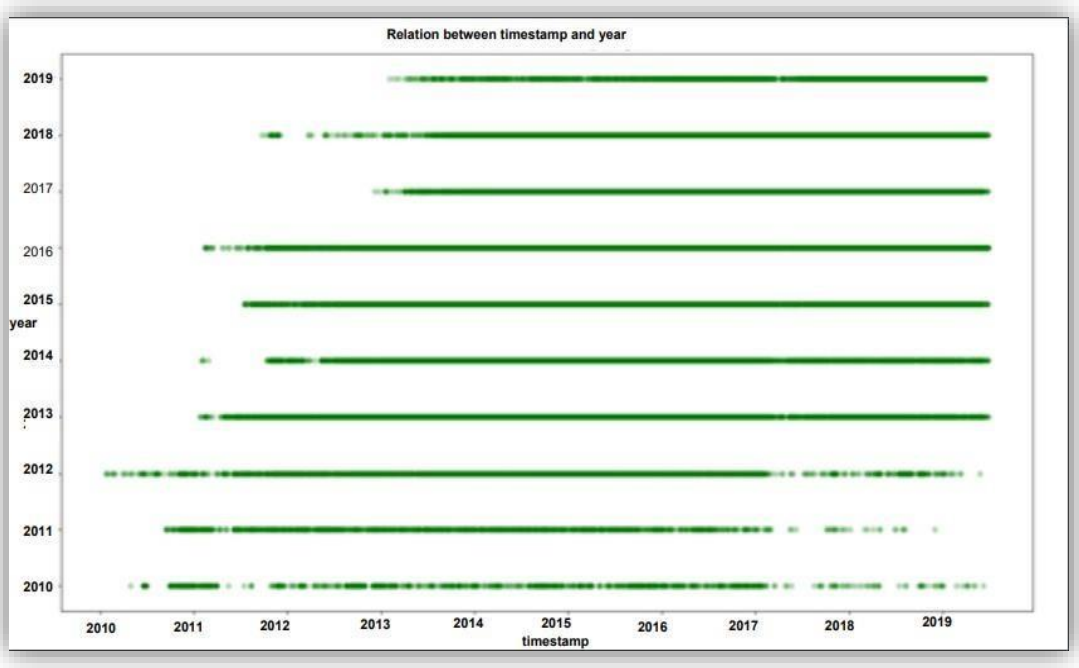


Figure 4.9: Relation between timestamp and year

The figure above represent Checking the distribution of the timestamp with year, Relation between timestamp and year , as demonstrated in Figure 4.9.

4.6 Results of Proposed Models

This work presents a model for improving the accuracy of recommender systems by combining two widely used methods (associative rules and clustering algorithm).

4.6.1 Apriori Algorithm

The Apriori technique was used in this study to extract meaningful association rules from a dataset. This approach identifies the most common item sets with support and confidence values greater than a pre-determined threshold. The association rules were applied to those rules as shownin figure 4.10 and table 4.5.

Table 4.5: show the result of the performance of associative rule based on apriori algorithm (min_support = 0.00005)

id	antecedent support	consequent support	Support	confidence	lift	leverage	conviction
0	7.01E-05	7.01E-05	6.01E-05	0.857142857	12231.8	6.01E-05	6.99950948
1	7.01E-05	7.01E-05	6.01E-05	0.857142857	12231.8	6.01E-05	6.99950948
2	7.01E-05	8.01E-05	5.01E-05	0.714285714	8919.018	5.00E-05	3.4997197
3	7.01E-05	8.01E-05	5.01E-05	0.714285714	8919.018	5.00E-05	3.4997197
4	7.01E-05	8.01E-05	5.01E-05	0.714285714	8919.018	5.00E-05	3.4997197
5	7.01E-05	8.01E-05	5.01E-05	0.714285714	8919.018	5.00E-05	3.4997197
6	8.01E-05	7.01E-05	5.01E-05	0.625	8919.018	5.00E-05	2.6664798
7	8.01E-05	7.01E-05	5.01E-05	0.625	8919.018	5.00E-05	2.6664798
8	8.01E-05	7.01E-05	5.01E-05	0.625	8919.018	5.00E-05	2.6664798
9	8.01E-05	7.01E-05	5.01E-05	0.625	8919.018	5.00E-05	2.6664798
10	8.01E-05	8.01E-05	5.01E-05	0.625	7804.141	5.00E-05	2.6664531

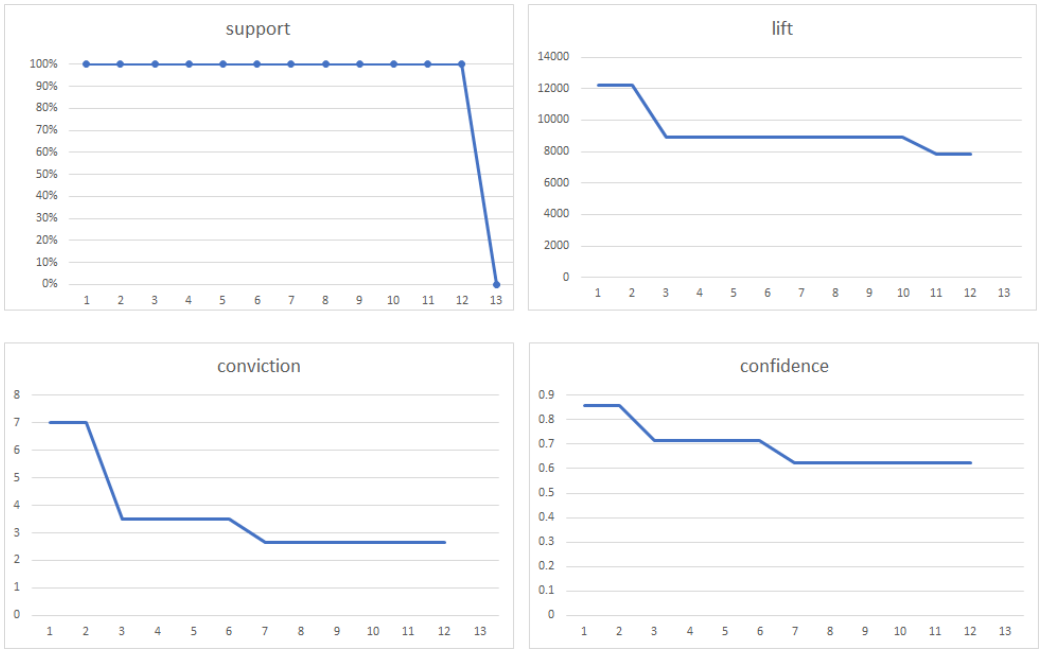


Figure 4.10: show the performance of associative rule based on apriori algorithm

Here the apriori algorithm works to obtain the extent of association between the items, and here the top 10 values were selected where each item was represented by id and min_support was compared and knowing the extent of the association between items From Figure4.10 above and Table 4.5 the maximum value in antecedent is 8.01E-05, Consequent is 8.01E-05, the support is 6.01E-05, the confidence is 0.857143, the lift is 12231.8, the leverage is 6.01E-05, and the conviction is 6.999509

4.6.2 The Classifier LSTM, SVM, KNN model the GMM

This work proposes a model for improving the accuracy of recommender systems based on the possible similarities of two regularly used approaches in recommender systems (clustering algorithm and associative rules). This step will utlise the Proposed methods such as GMM-SVM, GMM-KNN, GMM-LSTM according algorithm in chapter three, where apply the GMM to cluster the dataset in two three group and the using the classifier as described in chapter three, and the result of each model and compute the confusion matrix and calculate the precision recall, f1-score, support, accuracy as shown in Table 4.6 and Table 4.7.

Table 4.6: Confusion Matrix of proposed model

Confusion Matrix GMM-KNN			
	C1	C2	C3
C1	15506	47	9
C2	0	10213	0
C3	14	55	4024
Confusion Matrix GMM-SVM			
	C1	C2	C3
C1	15508	0	54
C2	0	10213	0
C3	64	0	4129
Confusion Matrix GMM-LSTM			
	C1	C2	C3
C1	11320	4242	0
C2	302	9911	0
C3	1	4192	0

From table 4.6 result above observed the best result according true state in confusion matrix GMM-SVM and GMM-KNN

Table 4.7: compute the classification report of classifier model(LSTM,KNN,SVM) with GMM

Classification Report GMM-KNN				
	precision	recall	f1-score	Support
C1	0.99	1.00	0.99	15562
C2	0.99	1.00	1.00	10213
C3	1.00	0.96	0.98	4193
Accuracy = 0.99249199145				
Classification Report GMM-SVM				
	precision	recall	f1-score	Support
C1	1.00	1.00	1.00	15562
C2	1.00	1.00	1.00	10213
C3	0.99	0.98	0.99	4193
Accuracy = 0.9964919914575547				
Classification Report GMM-LSTM				
	precision	recall	f1-score	Support
C1	1.00	1.00	1.00	15562
C2	1.00	1.00	1.00	10213
C3	0.99	0.98	0.99	4193
Accuracy = 0.5562253				

From Table 4.7 result above observed the best classifier GMM-SVM whereachieved the accuracy 0.996, then the GMM-KNN achieved 0.992 rate of accuracy.

4.6.3The Classifier LSTM, SVM, KNN model the K-MEANS

This work provides a classifier which is a method for improving the precision of recommender systems. It will utilize the Proposed method such as K-means-SVM, K-means-KNN, K-MEANS-LSTM according algorithm in chapter three, where apply the K-MEANS to cluster the dataset in two three group and the using the classifier as described in chapter three, and the result of each model and compute the confusion matrix and calculate the precision recall, f1-score, support, accuracy as shown in Table 4.8 and Table 4.9

Table 4.8: Confusion Matrix of classifier model (LSTM,KNN,SVM) with k-means

Confusion Matrix K-MEANS-KNN			
	C1	C2	C3
C1	13081	2	1
C2	17	10957	0
C3	15	0	5895
Confusion Matrix K-MEANS -SVM			
	C1	C2	C3
C1	13080	4	0
C2	0	10974	0
C3	4	0	5906
Confusion Matrix K-MEANS -LSTM			
	C1	C2	C3
C1	6009	7075	0
C2	533	10441	0
C3	0	5910	0.99

From table 4.8 result above observed the best result according true state in confusion matrix k-means-SVM and k-means-KNN

Table 4.9 show the classification repot the classifier model (LSTM,KNN,SVM)with k-means

Classification Report K-MEANS -KNN				
	precision	recall	f1-score	Support
C1	0.99	0.99	0.99	13084
C2	0.99	0.99	0.99	10974
C3	0.99	0.99	0.99	5910
Accuracy = 0.998832				
Classification Report K-MEANS-SVM				
	precision	recall	f1-score	Support
C1	0.99	0.99	0.99	13084
C2	0.99	0.99	0.99	10974
C3	0.99	0.99	0.99	5910
Accuracy 0.99907				
Classification Report K-MEANS -LSTM				
	precision	recall	f1-score	Support
C1	0.92	0.46	0.61	13084
C2	0.45	0.95	0.61	10974
C3	0.1	0.45	0.1	5910
Accuracy = 0.55489188				

From result above observed the best classifier K-means-SVM where achieved the accuracy 0.999 then the K-means-KNN achieved 0.998 rate of accuracy.

4.7 Statistical Analysis

The following is a statistical analysis of Table (4.7) with BOX PLOT and INTERVAL PLOT. The analysis was conducted using Minitab V.18 software.

Table 4.10 (A,B) Analysis of Variance for table 4.7

(A)

Classifier model	Precision	Recall	fl-score	Support
KNN	0.993	0.987	0.990 a	9989
SVM	0.997	0.993	0.990 a	9989
LSTM	0.997	0.993	0.440 b	9989
P-Value¥	0.729 N.S	0.850 N.S	0.492 N.S	1.00 N.S
lSD	-	-	-	-

¥ : One-way ANOVA were used, N.S : Not significant (P >0.05), * : significant (P<0.05)

(B)

classifier model	Accuracy
KNN	**0.9925**
SVM	**0.9965**
LSTM	**0.5562**
SE±	**0.146**

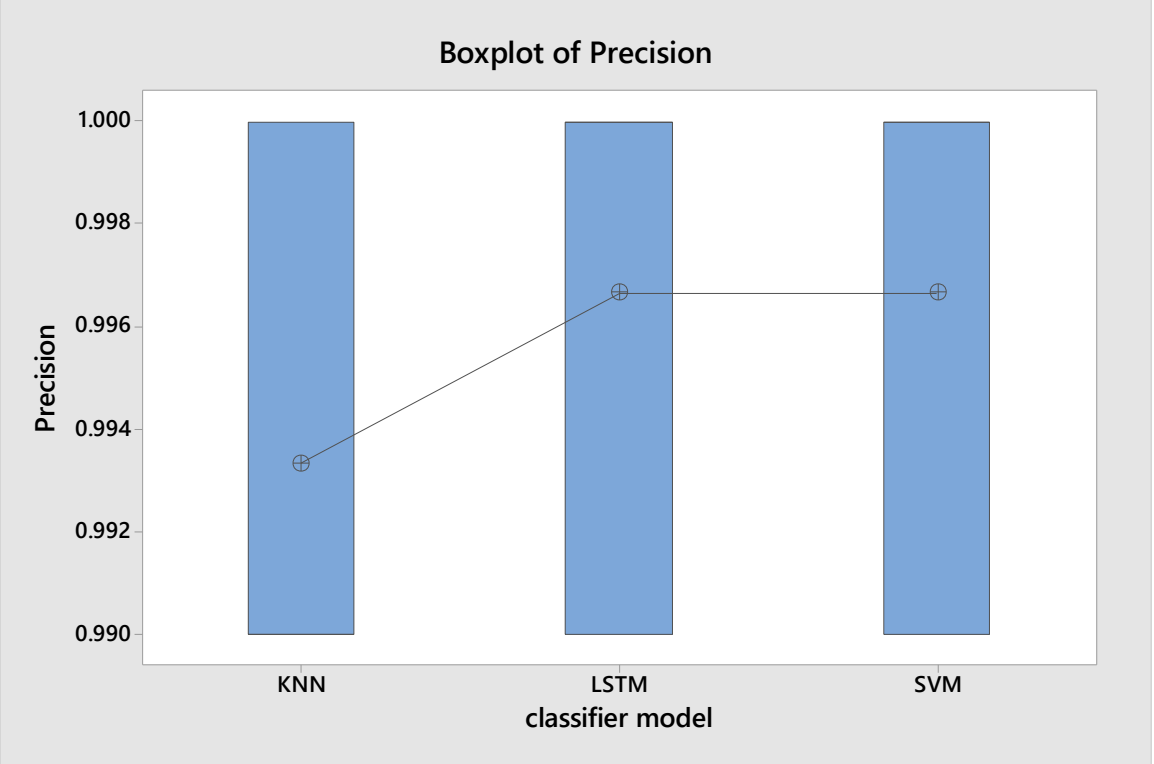


Figure 4.11: Show Boxplot of Preision

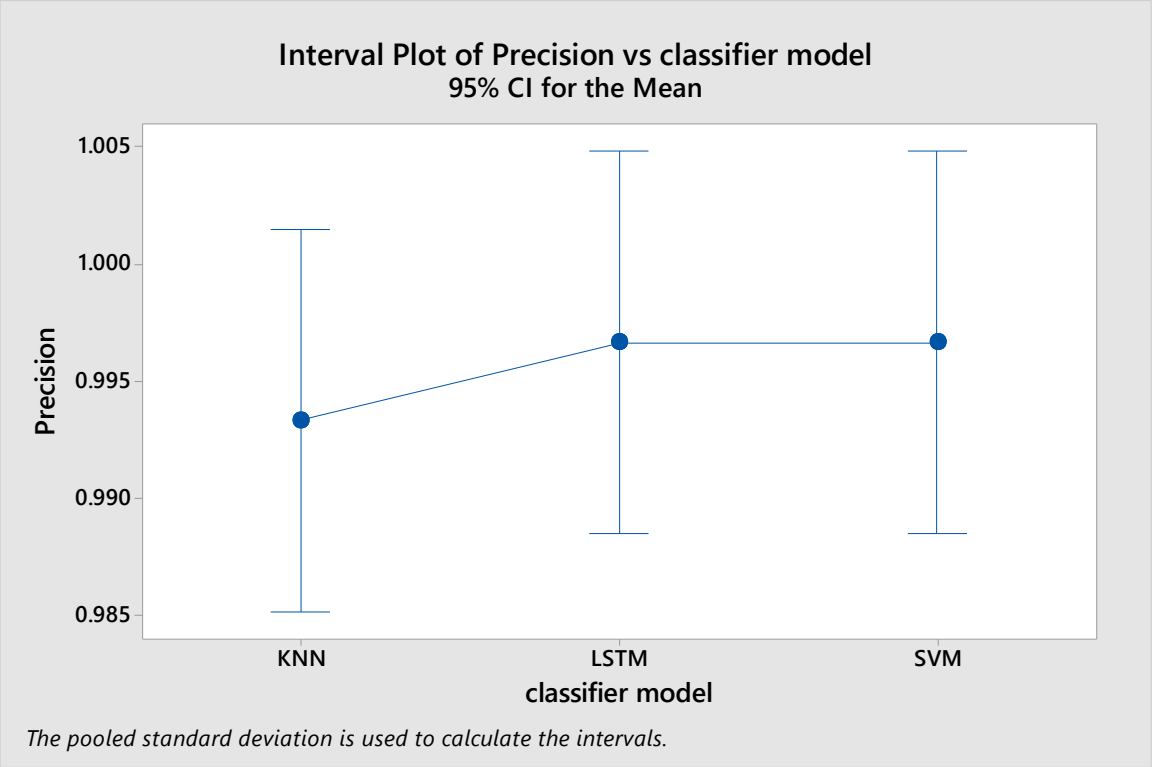


Figure 4.12 : Show Interval Plot of Precision &Classifier Model 95% C1 for the Mean

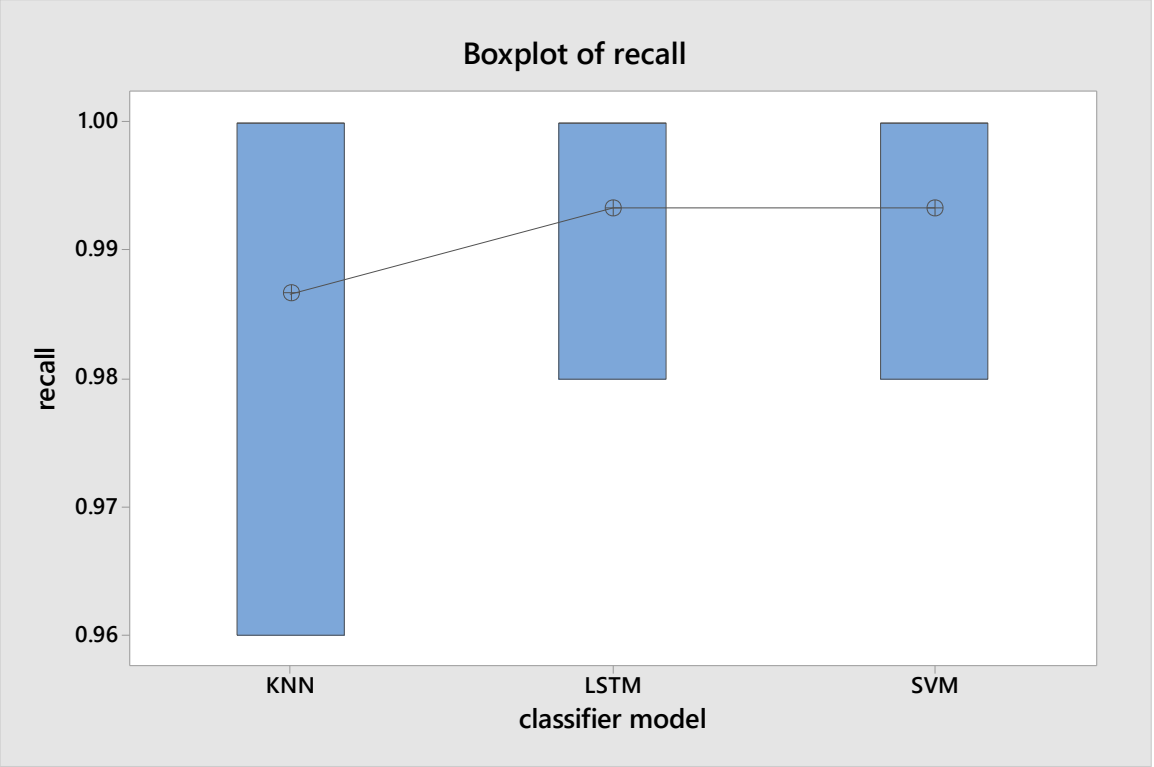


Figure 4.13: Show Boxplot of Recall .

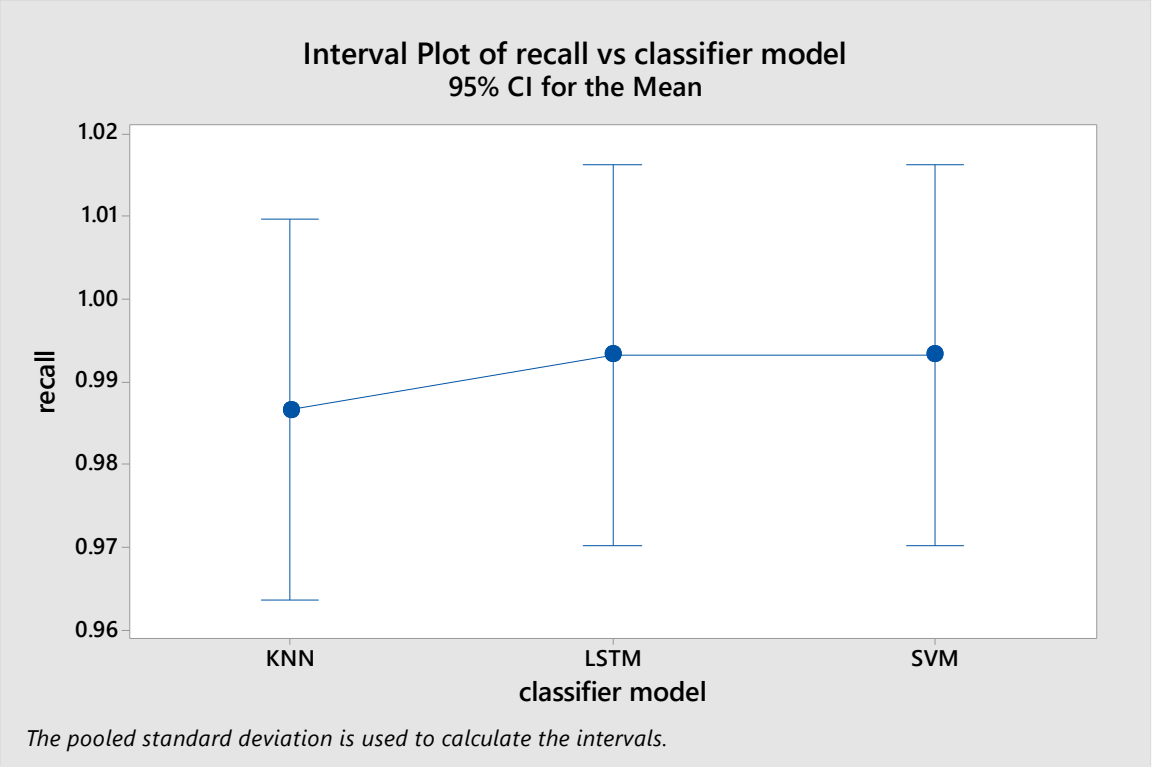


Figure 4.14 : Show Interval Plot of Recall &Classifier Model 95% C1 for the Mean

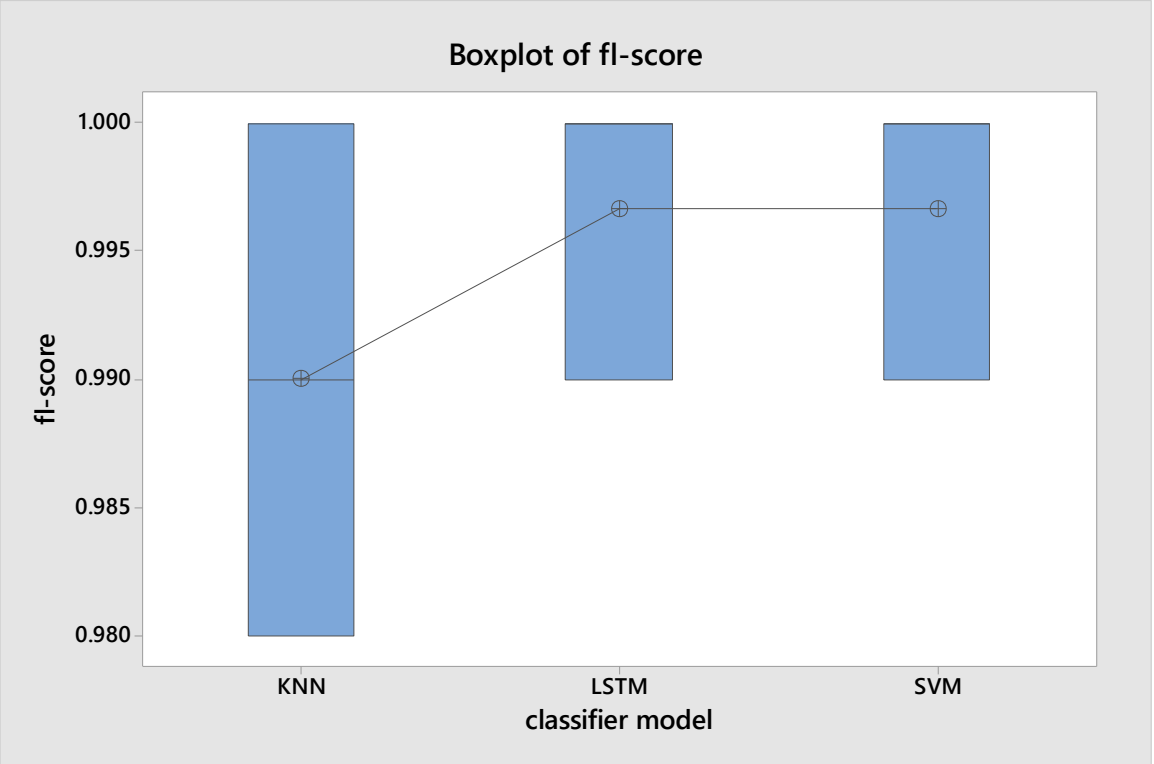


Figure 4.15: Show Boxplot of f1-score .

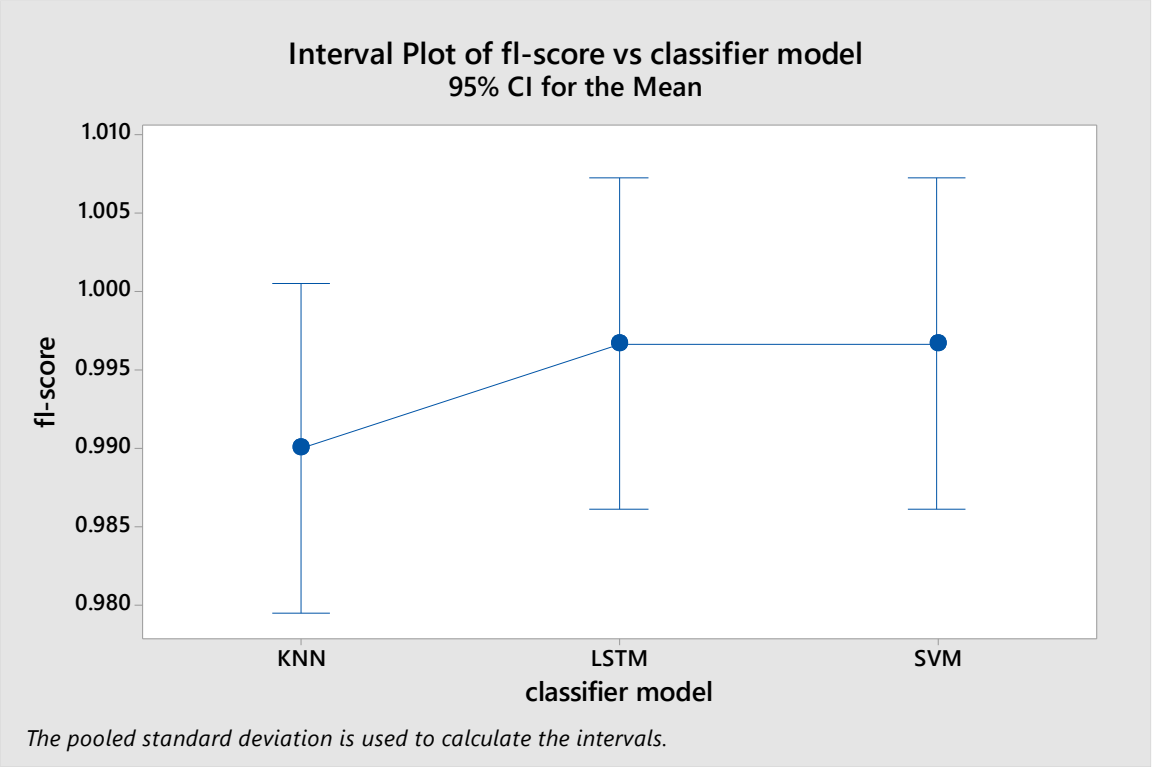


Figure 4.16 : Show Interval Plot of f1-Score &Classifier Model 95% C1 for the Mean

The following is a statistical analysis of Table (4.9) with BOX PLOT and INTERVAL PLOT. The analysis was conducted using Minitab V.18 software.

Table 4.11 (A,B) Analysis of Variance for table 4.9

(A)

classifier model	Precision	recall	fl-score	Support
KNN	0.990	0.990	0.990 a	9989
SVM	0.990	0.990	0.990 a	9989
LSTM	0.490	0.620	0.440 b	9989
P-Value¥	0.066 N.S	0.052 N.S	0.011*	1.00 N.S
lSD	-	-	0.34	-

¥ : One-way ANOVA were used, N.S : Not significant (P >0.05), * : significant (P<0.05)

(B)

classifier model	Accuracy
KNN	**0.9988**
SVM	**0.9991**
LSTM	**0.5549**
SE±	**0.148**

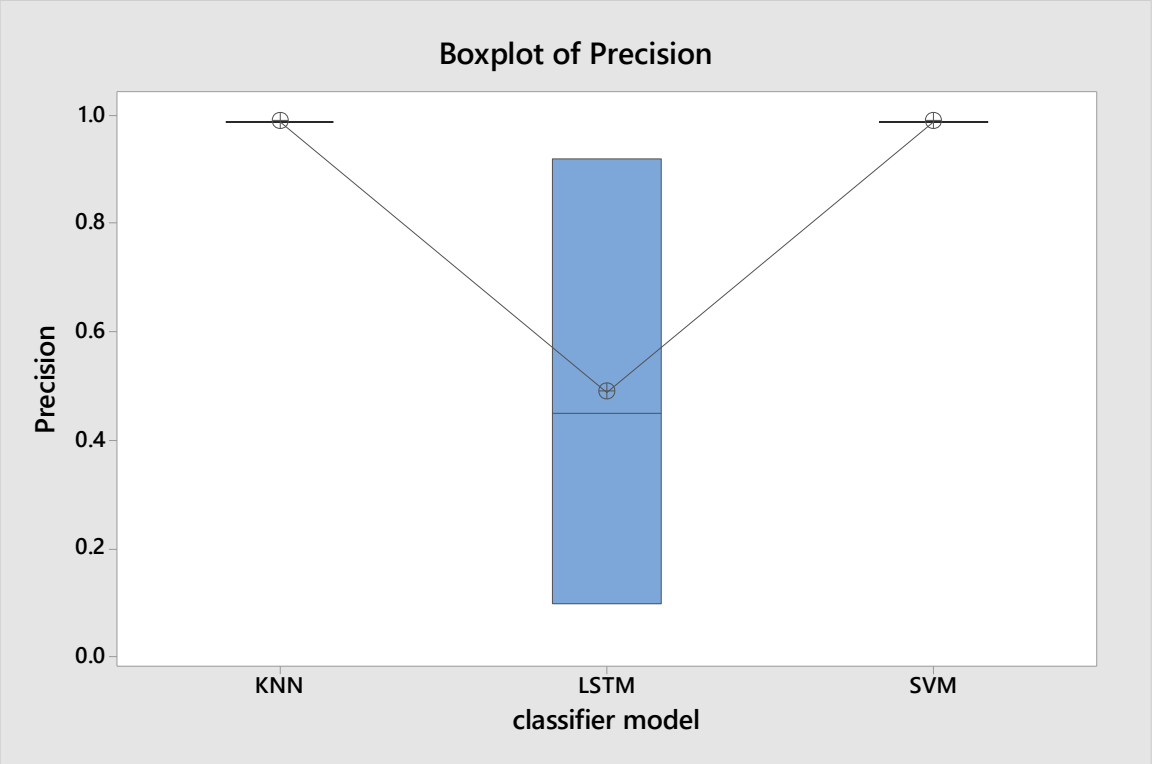


Figure 4.17 : Show Boxplot of Precision

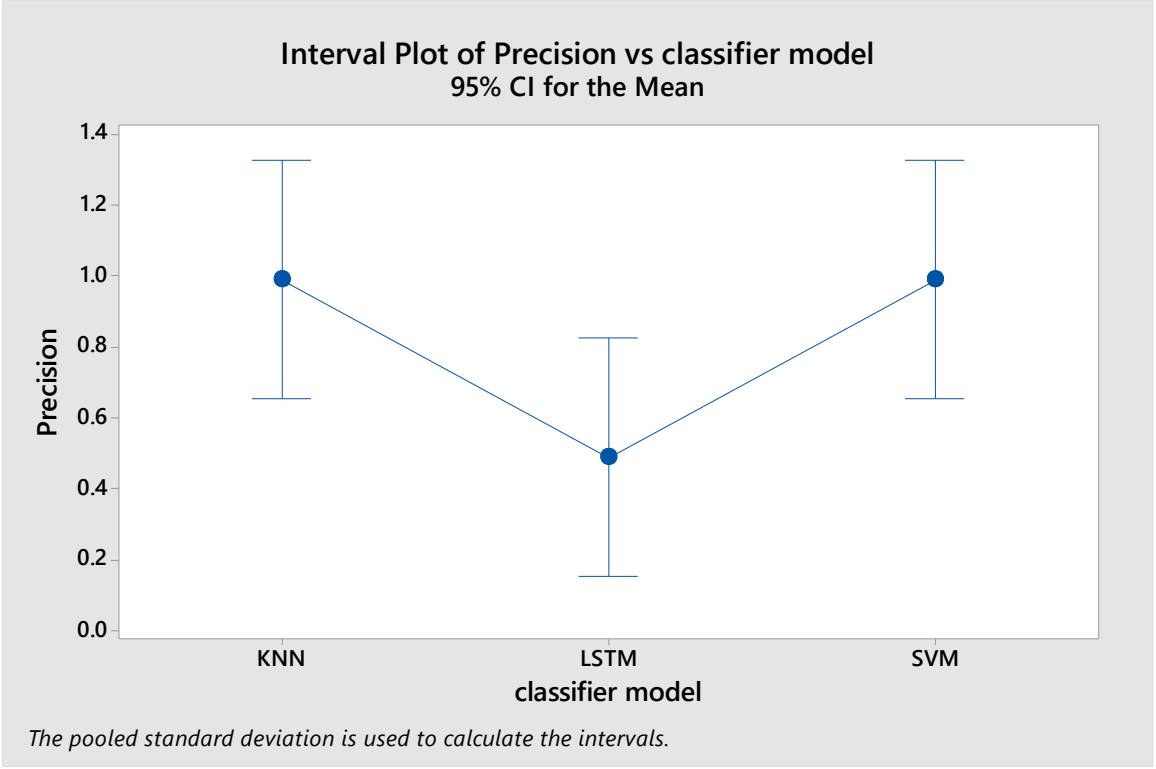


Figure 4.18 : Show Interval Plot of Precision &Classifier Model 95% C1 for the Mean

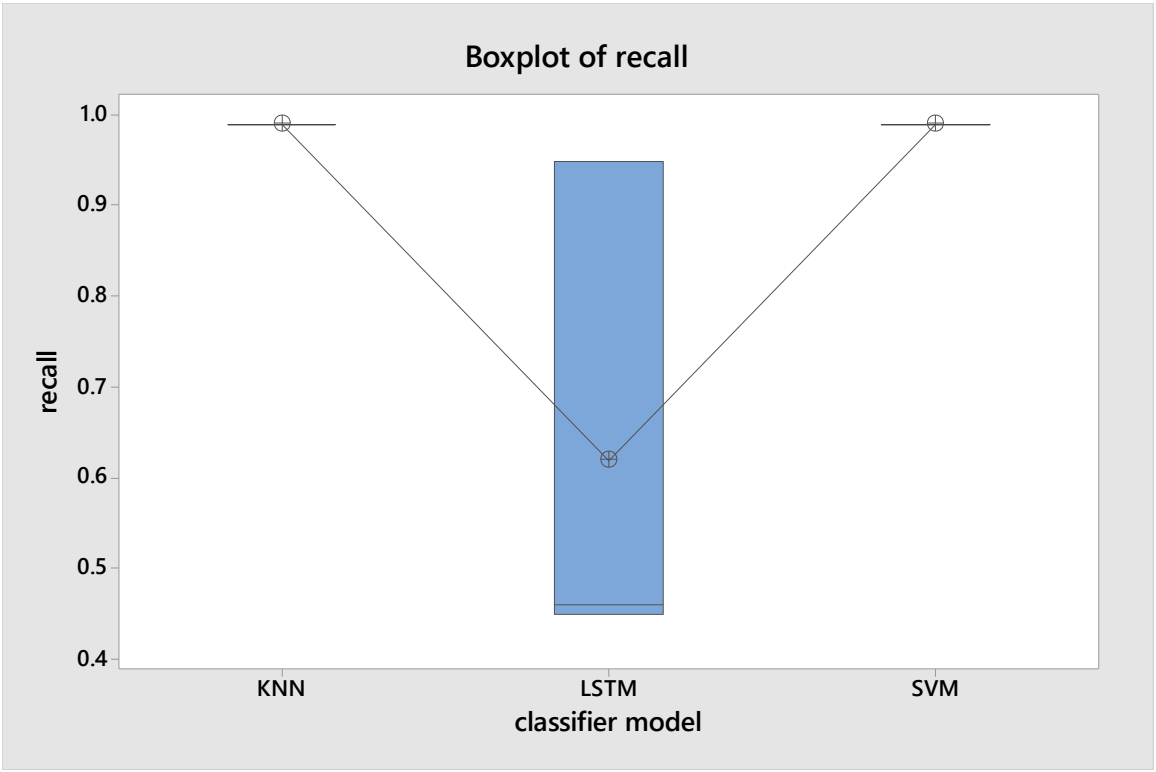


Figure 4.19: Show Boxplot of Recall

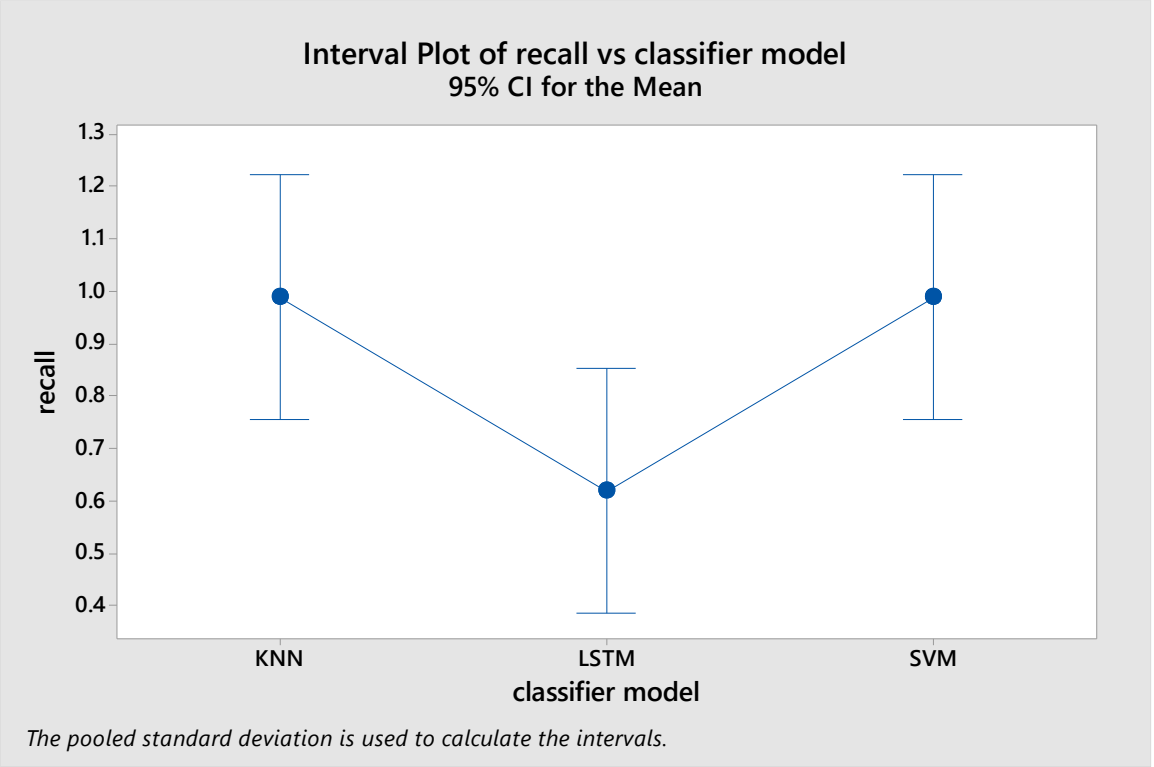


Figure 4.20 : Show Interval Plot of Recall &Classifier Model 95% C1 for the Mean

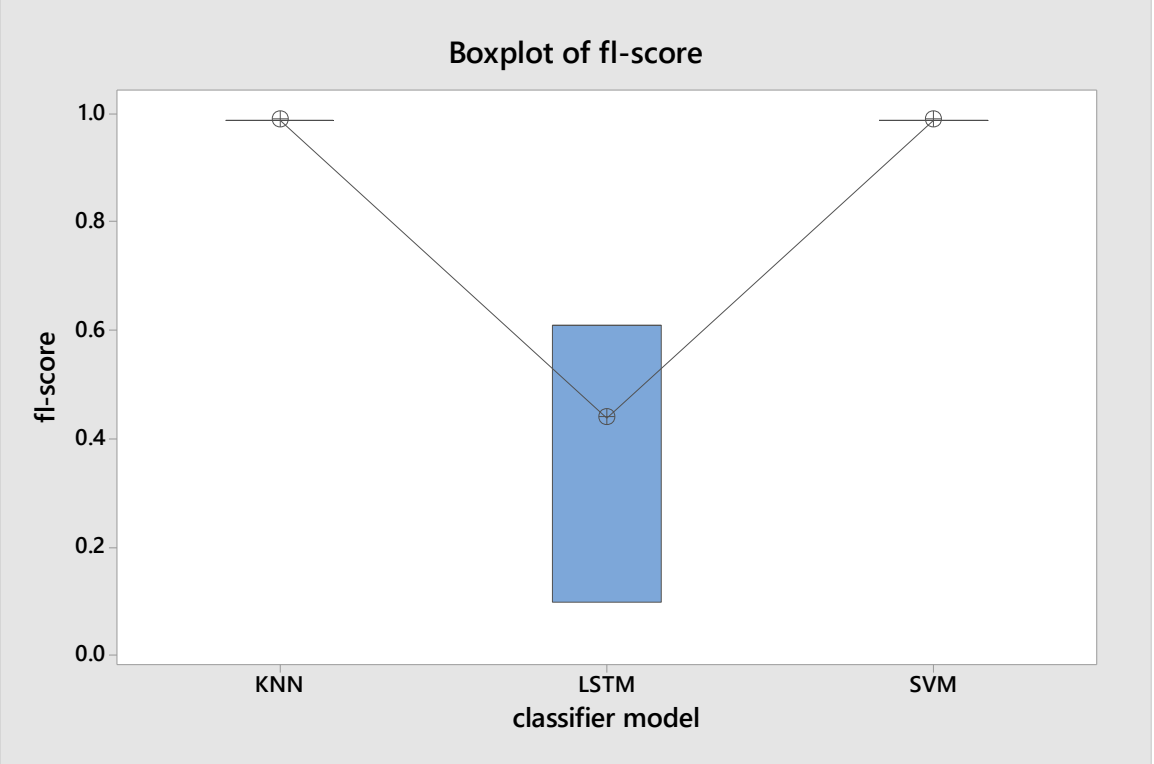


Figure 4.21: Show Boxplot of f1-score.

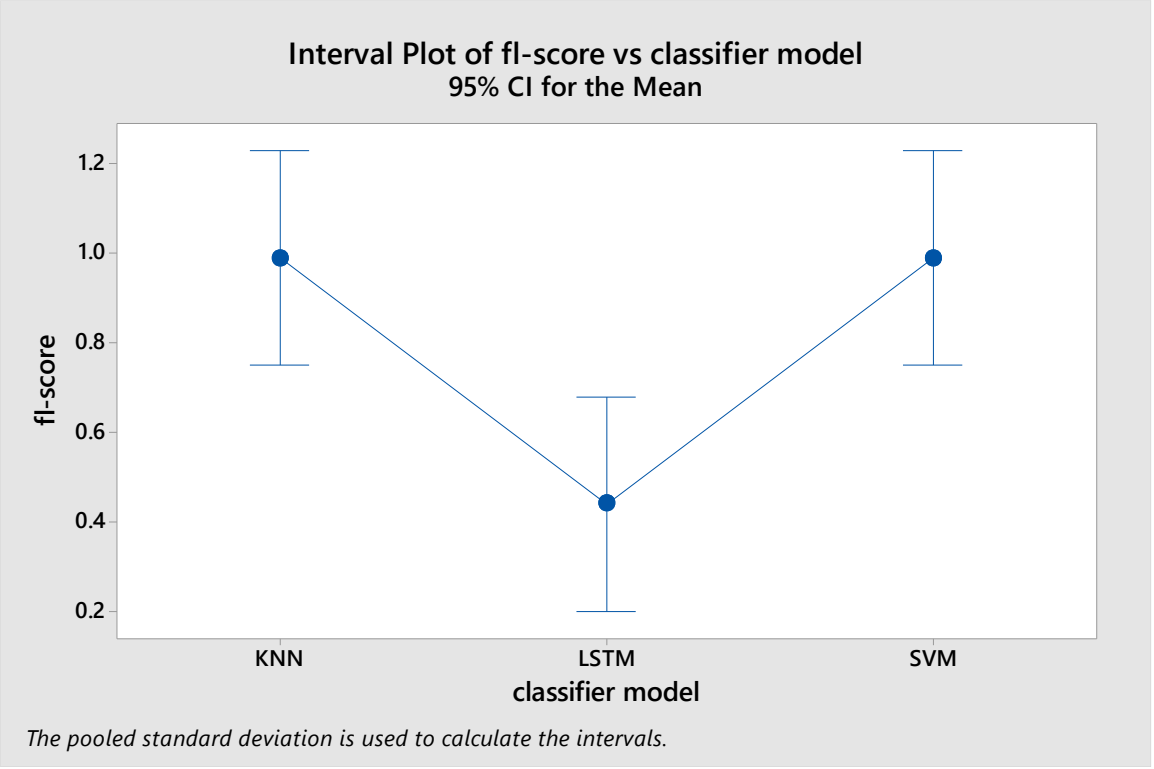


Figure 4.22 : Show Interval Plot of f1-score &Classifier Model 95% C1 for the Mean

4.8 Discussion

The result of proposed system has been compared with the three previous related works and showed that our system more accurate than the compared works, as shown in Table (4.12)

Table 4.12: Comparison with some related work

Authors	Tool	accuracy
Rizkallah,2021	Non-negative Matrix Factorization(NMF)	0.948
A. Nechaev,2020	SVD	0.9547
R. Misra,2018	K-LF-ML	0.657
Proposed	Proposed G-KNN	0.992
	Proposed G-SVM	0.996
	Proposed G-LSTM	0.555
	Proposed K-KNN	0.998
	Proposed K -SVM	0.999
	Proposed K-LSTM	0.554

The above table(4.12) shows the comparison with previous works, where it was also applied to goods, products and data mining . My work was compared with these studies. The first study relied on multiple datasets, namely ModCloth, MovieLens 100 K and others. Non-negative Matrix Factorization (NMF) techniques were applied to it. Reaching an accuracy of 0.948, the second study was applied to multiple datasets such as Modcloth and Amazon Phones and Movielens 1M by applying SVD to the data and obtained 0.9547, and the third study was applied to ModCloth and used K-LF-ML and the accuracy was reached 0.657. In our work, machine learning algorithms were applied with deep learning algorithms and the highest accuracy proposed was reached. Kmeans-SVM is 0.999 and Proposed Kmeans-KNN has reached 0.998

Chapter Five

Conclusions & Future Works

Chapter Five

5.1 Introduction

This chapter gathers the final findings of the results of suggested method, which were extracted from the previous chapters' results in order to improve the RS's preference. In addition, this chapter discusses future work and ideas that might be applied to the dataset to improve the performance of recommender systems.

5.2 Conclusions

The conclusions that could be drawn from this work are listed as following:

1. Apply machine learning techniques (SVM, KNN) with deep learning (LSTM) and combine them with K-means, GMM algorithms to predict products.

2. The results obtained show that the proposed machine learning algorithms give better results and performance to predict products than deep learning algorithms because deep learning needs huge data and is continuous.

3. This thesis uses Modcloth Amazon dataset, where the some of t basic statistics are (Total number of ratings (99,893), Number of users (44,783), Number of items (1,020), and the main metadata is (Item Id, User Id, Rating,Timestamp, Size, Fit, User Attribute, Model Attribute, Category,Brand Year, Split).

4. Then visual dataset to using the following code input to examine the distribution of the rating property , it is clear that the most consumers give purchased products a positive rating ,while negative ratings are uncommon.

5. Apriori algorithm implemented to get a significant association rules from a given the dataset. This algorithm first finds out the frequent item sets that have support and confidence values above a pre-specified threshold value, when suppose the min support equal 0.00005 .

6. Then Proposed cluster GMM with Classifier such as GMM-SVM, GMM-KNN,GMM-LSTM and obtain result above observed the best classifier GMM- SVM where achieved the accuracy 0.996, then the GMM-KNN achieved 0.992 rate, then the GMM-LSTM achieved 0.554 rate of accuracy.

7. Then Proposed cluster k-mean with Classifier for increase the accuracy of modelsuch as K-means-SVM, K-means-KNN, K-MEANS-LSTM, observed the best classifier K-means-SVM where achieved the accuracy 0.999 then the K-means-KNN achieved 0.998 rate of accuracy, then the K-means-LSTMachieved 0.558 rate of accuracy

5.3 Future Works

The results of this study point to the following future projects:

1. Use more than one dataset and apply the proposed model to know theeffectiveness of the model
2. Building the method as a web application to be online and real time
3. Use new and diverse methods in hybrid cluster strategy
4. Develop the stage of interpretation of the association rules through the use of: Genetic algorithm (GA) or fuzzy logic

References

References

[1] P. Zheng, X. Xu, S. Yu and C. Liu, "Personalized product configuration framework in an adaptable open architecture product platform", Journal of Manufacturing Systems, vol. 43, pp. 422-435, 2017.

[2] Khan, B.M., Mansha, A., Khan, F.H., and Bashir, S.: 'Collaborative filtering based online recommendation systems: A survey', in Editor (Ed.)^(Eds.): 'Book Collaborative filtering based online recommendation systems: A survey' (IEEE, edn.), pp. 125-130, 2017

[3] Sridevi, M., Rao, R.R., and Rao, M.V.: 'A survey on recommender system', International Journal of Computer Science and Information Security, 14, (5), pp. 265, 2016

[4] P. Kumar and R. Thakur, "Recommendation system techniques and related issues: a survey", International Journal of Information Technology, vol. 10, no. 4, pp. 495-501, 2018.

[5] M. Scholz, V. Dorner, G. Schryen and A. Benlian, "A configuration-based recommender system for supporting e-commerce decisions", European Journal of Operational Research, vol. 259, no. 1, pp. 205-215, 2017.

[6] T. Silveira, M. Zhang, X. Lin, Y. Liu and S. Ma, "How good your recommender system is? A survey on evaluations in recommendation", International Journal of Machine Learning and Cybernetics, vol. 10, no. 5, pp. 813-831, 2017.

[7] S. Li and E. Karahanna, "Online Recommendation Systems in a B2C E-Commerce Context: A Review and Future Directions", Journal of the Association for Information Systems, vol. 16, no. 2, pp. 72-107, 2015.

[8] Y. Huang, H. Liu, W. Li, Z. Wang, X. Hu and W. Wang, "Lifestyles in Amazon: Evidence from online reviews enhanced recommender system", International Journal of Market Research, vol. 62, no. 6, pp. 689-706, 2019.

[9]H. Tahmasebi, R. Ravanmehr and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data", Neural Computing and Applications, vol. 33, no. 5, pp. 1607-1623, 2020.

[10] C. Guan, S. Qin, W. Ling and G. Ding, "Apparel recommendation system evolution: an empirical review", International Journal of Clothing Science and Technology, vol. 28, no. 6, pp. 854-879, 2016.

[11] F. García-Sánchez, R. Colomo-Palacios and R. Valencia-García, "A social-semantic recommender system for advertisements", Information Processing & Management, vol. 57, no. 2, p. 102153, 2020.

[12] R. Behera, A. Gunasekaran, S. Gupta, S. Kamboj and P. Bala, "Personalized digital marketing recommender engine", Journal of Retailing and Consumer Services, vol. 53, p. 101799, 2020.

[13] Malik, S., Rana, A., and Bansal, M.: ‘A Survey of Recommendation Systems’, Information Resources Management Journal (IRMJ), , 33, (4), pp. 53-73, 2020

[14] Narke, L., and Nasreen, A.: ‘A Comprehensive Review of Approaches and Challenges of a Recommendation System’

[15] Isinkaye, F.O., Folajimi, Y., and Ojokoh, B.A.: ‘Recommendation systems: Principles, methods and evaluation’, Egyptian Informatics Journal, 16,(3), pp. 261-273, 2015

[16]Sahoo, A.K., Mallik, S., Pradhan, C., Mishra, B.S.P., Barik, R.K., and Das, H.:‘Intelligence-based health recommendation system using big data analytics’:‘Big data analytics for intelligent healthcare management’ (Elsevier), pp. 227-246, 2019

[17] Sahoo, A.K., Pradhan, C., and Mishra, B.S.P.: 'SVD based privacy preserving recommendation model using optimized hybrid item-based collaborative filtering', in Editor (Ed.)^(Eds.): 'Book SVD based privacy preserving recommendation model using optimized hybrid item-based collaborativefiltering' (IEEE, edn.), pp. 0294-0298,2019

[18] Umberto, P.: 'Developing a price-sensitive recommender system to improve accuracy and business performance of ecommerce applications', " International Journal of Electronic Commerce Studies, 6, (1), pp. 1-18", 2015

[19]Sato, M., Izumo, H., and Sonoda, T.: 'Discount sensitive recommender systemfor retail business', in Editor (Ed.)^(Eds.): 'Book Discount sensitive recommender system for retail business' (edn.), pp. 33-40, 2015

[20]Jannach, D., and Adomavicius, G.: 'Price and profit awareness in recommender systems', arXiv preprint arXiv:1707.08029, 2017

[21] Zhao, Q., Zhang, Y., Zhang, Y., and Friedman, D.: 'Multi-product utility maximization for economic recommendation', 'Book Multi-product utility maximization for economic recommendation' (edn.), pp. 435-443, 2017

[22]Yang, J., Liu, C., Teng, M., Chen, J., and Xiong, H.: 'A unified view of socialand temporal modeling for B2B marketing campaign recommendation', IEEE Transactions on Knowledge and Data Engineering, 30, (5), pp. 810-823, 2017

[23] Gaikwad, R.S., Udmale, S.S., and Sambhe, V.K.: 'E-commerce Recommendation System Using Improved Probabilistic Model': 'Information and Communication Technology for Sustainable Development'(Springer), pp. 277-284, 2018

[24]Guo, Y., Yin, C., Li, M., Ren, X., and Liu, P.: 'Mobile e-commerce recommendation system based on multi-source information fusion for sustainable e-business', Sustainability, 10, (1), pp. 147, 2018

[25] Misra, R., Wan, M., and McAuley, J.: 'Decomposing fit semantics for product size recommendation in metric spaces', 'Book Decomposing fit semantics for product size recommendation in metric spaces' (edn), pp. 422-426,2018

[26] Li, H., Wu, Y.J., and Chen, Y.: 'Time is money: Dynamic-model-based time series data-mining for correlation analysis of commodity sales', Journal of Computational and Applied Mathematics, 370, pp. 112659, 2020

[27] Rizkallah, S., Atiya, A.F., and Shaheen, S.: 'New Vector-Space Embeddings for Recommender Systems', Applied Sciences, 11, (14), pp. 6477, 2021

[28] Fu, W., Peng, Z., Wang, S., Xu, Y., and Li, J.: 'Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems', in Editor (Ed.)^(Eds.): 'Book Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems' (edn.), pp. 94-101,2019

[29] Natarajan, S., Vairavasundaram, S., Natarajan, S., and Gandomi, A.H.: 'Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data', Expert Systems with Applications, 149, pp. 113248, 2020

[30] Thorat, P.B., Goudar, R., and Barve, S.: 'Survey on collaborative filtering, content-based filtering and hybrid recommendation system', International Journal of Computer Applications, 110, (4), pp. 31-36, 2015

[31] Aggarwal, C.C.: 'Recommender systems' (Springer) , 2016

[32] Kashef, R.: 'Enhancing the Role of Large-Scale Recommendation Systems inThe IoT Context', IEEE Access, 2020

[33] Patel, B., Desai, P., and Panchal, U.: 'Methods of recommender system: a review', 'Book Methods of recommender system: a review' (IEEE, edn.), pp. 1-4, 2017

[34] Kavoura, A., Kefallonitis, E., and Giovanis, A.: 'Strategic innovative marketing and tourism', Springer Proceedings in Business and Economics, seen at https://doi. org/10.1007/978-3-030-12453-3_101, 2019

[35] Liu, Y., Xiong, Q., Sun, J., Jiang, Y., Silva, T., and Ling, H.: 'Topic-based hierarchical Bayesian linear regression models for niche items recommendation', Journal of Information Science, 45, (1), pp. 92-104, 2019

[36] Akbar, F., Omar, A., and Wadood, F.: 'The Niche Marketing Strategy Constructs (Elements) and its Characteristics-A Review of the Relevant Literature', Galore international journal of applied sciences & humanities, 1, (1), pp. 73-80,2017

[37] McDONALD, M.: 'Strategic marketing planning: theory and practice', Themarketing book, pp. 87, 2016

[38] Saputra, I.F.: 'Analysis Segmentation, Targeting, and Positioning (STP) Toward The Development of Halal Hanwoo Beef in South Korea',Universitas Muhammadiyah Surakarta, 2020

[39] Virk, H.K., Singh, E.M., and Singh, A.: 'Analysis and design of hybrid onlinemovie recommender system', International Journal of Innovations in Engineering and Technology (IJIET) Volume, 5, 2015

[40] Mustafa, N., Ibrahim, A.O., Ahmed, A., and Abdullah, A.: 'Collaborative filtering: Techniques and applications', 'Book Collaborative filtering: Techniques and applications' (IEEE, edn.), pp.1-6, 2017

[41] Zhang, J., Zhang, C., and Yu, H.: 'Research on e-commerce intelligent service based on Data Mining', 'Book Research on e- commerce intelligent service based on Data Mining' (EDP Sciences, edn.), pp. 03012, 2018

[42] Najafabadi, M.K., Mohamed, A.H., and Mahrin, M.N.r.: 'A survey on data mining techniques in recommender systems', Soft Computing, 23, (2),pp. 627-654, 2019

[43] Putra, P.B.I.S., Suryani, N.P.S.M., and Aryani, S.: 'Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket', International Journal of Engineering and Emerging Technology, 3, (1), pp. 13-17, 2018

[44] Bayer, H., Aksogan, M., Celik, E., and Kondiloglu, A.: 'Big data mining and business intelligence trends', Journal of Asian Business Strategy, 7, (1),pp. 23, 2017

[45]Mishra, B.K., Hazra, D., Tarannum, K., and Kumar, M.: 'Business intelligenceusing data mining techniques and business analytics', 'Book Business intelligence using data mining techniques and business analytics' (IEEE, edn.), pp. 84-89, 2016

[46]Sindhu, D., and Sangwan, A.: 'Optimization of Business Intelligence using Data Digitalization and Various Data Mining Techniques', International Journal of Computational Intelligence Research, 13, (8), pp. 1991-1997, 2017

[47]Nilashi, M.: 'An overview of data mining techniques in recommender systems', Journal of Soft Computing and Decision Support Systems, 3,(6), pp. 16-44, 2016

[48]Nguyen, L.: 'Introduction to a Framework of E-commercial Recommendation Algorithms', American Journal of Computer Science and Information Engineering, 2, (4), pp. 33-44, 2015

[49]Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., and Kashef, R.: 'Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities', Applied Sciences, 10, (21), pp. 7748, 2020

[50]Zhou, M., Ding, Z., Tang, J., and Yin, D.: 'Micro behaviors: A new perspective in e-commerce recommender systems','Book Micro behaviors: A new perspective in e-commerce recommender systems' (edn.), pp. 727-735, 2018

[51]Aldrich, S.E.: 'Recommender systems in commercial use', AI Magazine,32, (3), pp. 28-34, 2011

[52]Kaur, M., and Kang, S.: 'Market Basket Analysis: Identify the changing trendsof market data using association rule mining', Procedia computer science, 85, (Cms), pp. 78-85, 2016

[53]Setiawan, A., Budhi, G.S., Setiabudi, D.H., and Djunaidy, R.: 'Data mining applications for sales information system using market basket analysis on stationery company', 'Book Data mining applications for sales information system using market basket analysis on stationery company' (IEEE, edn.), pp. 337-340, 2017

[54]Gangurde, R., Kumar, B., and Gore, S.: 'Building prediction model using market basket analysis', Int. J. Innov. Res. Comput. Commun. Eng, 5,(2), pp. 1302-1309, 2017

[55]Aggarwal, C.C.: 'Data mining: the textbook' (Springer) , 2015

[56]Doomra, N., and Verma, R.: 'Web Mining Algorithms', 2019

[57]Chee, C.-H., Jaafar, J., Aziz, I.A., Hasan, M.H., and Yeoh, W.: 'Algorithms for frequent itemset mining: a literature review', Artificial Intelligence Review, 52, (4), pp. 2603-2621, 2019

[58]Fournier-Viger, P., Lin, J.C.W., Vo, B., Chi, T.T., Zhang, J., and Le, H.B.: 'A survey of itemset mining', Wiley Interdisciplinary Reviews: Data Mining andKnowledge Discovery, 7, (4), pp. e1207, 2017

[59]Negre, E.: 'Recommender Systems', Information and Recommender Systems, 4, pp. 7-27,2015

[60]Ricci, F., Rokach, L., and Shapira, B.: 'Recommender systems: introduction and challenges': 'Recommender systems handbook' (Springer), pp. 1-34, 2015

[61]Ali, H.H., and Kadhum, L.E.: 'K-Means Clustering Algorithm Applications inData Mining and Pattern Recognition', International Journal of Science and Research, 6, (8), pp. 1577-1584, 2017

[62]Sharma, P.: 'The Most Comprehensive Guide to K-Means Clustering You'll Ever Need' ,URL:https://www.analyticsvidhya. com/blog/2019/08/comprehensiveguide-k-means-clustering, 2019

[63]Lubis, A.R., and Lubis, M.: 'Optimization of distance formula in K-Nearest Neighbor method', Bulletin of Electrical Engineering and Informatics,9, (1), pp. 326-338, 2020

[64] Zhang, S.: 'Cost-sensitive KNN classification', Neurocomputing, 391, pp. 234-242 ,2020

[65] Vasanth, T., PeriyaKaruppan, C., and PoornaKumar, M.: 'Multi-domain recommendation system using hybrid fil tering and support vector machine classification', 2020

[66] Suthaharan, S.: 'Support vector machine': 'Machine learning models and algorithms for big data classification' (Springer), pp. 207-235, 2016

[67] Wang, H., Zheng, B., Yoon, S.W., and Ko, H.S.: 'A support vector machine-based ensemble algorithm for breast cancer diagnosis', European Journal of Operational Research, 267, (2), pp. 687-699 ,2018

[68] Aljarah, I., Ala'M, A.-Z., Faris, H., Hassonah, M.A., Mirjalili, S., and Saadeh, H.: 'Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm', Cognitive Computation, 10, (3), pp. 478-495, 2018

[69] Devikanniga, D., Ramu, A., and Haldorai, A.: 'Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm', EAI Endorsed Transactions on Energy Web, 7, 2020

[70] Devooght, R., and Bersini, H.: 'Collaborative filtering with recurrent neural networks', arXiv preprint arXiv:1608.07400, 2016

[71] Sunny, M.A.I., Maswood, M.M.S., and Alharbi, A.G.: 'Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model', in Editor (Ed.)^(Eds.): 'Book Deep Learning-Based Stock Price PredictionUsing LSTM and Bi-Directional LSTM Model' (IEEE, edn.), pp. 87- 92, 2020

[72]Jenkins, I.R., Gee, L.O., Knauss, A., Yin, H., and Schroeder, J.: 'Accident scenario generation with recurrent neural networks', in Editor (Ed.)^(Eds.): 'Book Accident scenario generation with recurrent neural networks' (IEEE, edn.), pp. 3340-3345, 2018

[73] Moghar, A., and Hamiche, M.: 'Stock market prediction using LSTM recurrent neural network', Procedia Computer Science, 170, pp. 1168-1173, 2020

[74] Verma, H., and Kumar, S.: 'An accurate missing data prediction method using LSTM based deep learning for health care', 'Book An accurate missing data prediction method using LSTM based deep learning for health care' (edn.), pp. 371-376,2019

[75] Najar, F., Bourouis, S., Bouguila, N., and Belghith, S.: 'A comparison between different gaussian-based mixture models', 'Book A comparison between different gaussian-based mixture models' (IEEE,edn.), pp. 704-708, 2017

[76] Reddy, A., Ordway-West, M., Lee, M., Dugan, M., Whitney, J., Kahana, R., Ford, B., Muedsam, J., Henslee, A., and Rao, M.: 'Using gaussian mixture models to detect outliers in seasonal univariate network traffic', in Editor (Ed.)^(Eds.): 'Book Using gaussian mixture models to detect outliers in seasonal univariate network traffic' (IEEE, edn.), pp. 229-234, 2017

[77] Kolouri, S., Rohde, G.K., and Hoffmann, H.: 'Sliced wasserstein distance for learning gaussian mixture models', 'Book Sliced wasserstein distance for learning gaussian mixture models' (edn.), pp. 3427-3436,2018

[78]Azimbagirad, M., and Junior, L.O.M.: 'Tsallis generalized entropy for Gaussian mixture model parameter estimation on brain segmentation application', Neuroscience Informatics, 1, (1-2), pp. 100002, 2021

[79] Datta, S., Das, J., Gupta, P., and Majumder, S.: 'SCARS: A scalable context-aware recommendation system', 'Book SCARS: A scalable context-aware recommendation system' (IEEE, edn.), pp. 1-6, 2015

[80] Abuhaiba, I.S., and Dawoud, H.M.: 'Combining different approaches to improve arabic text documents classification', International Journal of Intelligent Systems and Applications, 9, (4), pp. 39, 2017

الخلاصة :

أدى نمو الإنترنت إلى تشتت كبير لمصادر المعلومات. ونتيجة لذلك ، هناك حاجة إلى توصيات متخصصة بشأن أنواع مختلفة من المعلومات والمنتجات والخدمات لدعم المستخدمين في التغلب على مشكلة الحمل الزائد للمعلومات ، وتميل الطبيعة البشرية إلى اتباع أنماط معينة في مقارنة الخيارات المتاحة ، ونظام التوصية هو أحد الأسرار من نجاح العديد من الشركات ، وهذا النظام هو المسوق السحري للخدمات والمنتجات ، ويلاحظ العملاء ويفهم سلوكهم لمساعدتهم على اتخاذ قراراتهم.

تحتوي هذه الرساله على تقنيات التعلم الآلي جنبًا إلى جنب مع تقنيات التعلم العميق. بالإضافة إلى ذلك يتم استخدام خوارزمية Apriori لتقييم الأداء وإنشاء قواعد الارتباط من أجل تحسين كفاءة النظام لعمل تنبؤات دقيقة والتوصية بالمنتجات المناسبة. من أهم التقنيات المستخدمة بالإضافة إلى Apriori تم تطبيق نموذج مقترح من خلال الدمج مع تقنيات مثل K-means &KNN , K-means &SVM ,K-means&LSTM بالاضافه تطبيق نموذج هجين من خلال الدمج GMM &KNN ,GMM&SVM ,GMM&LSTM حساب استدعاء الدقة ، درجة F1والاستدعاء والدقة ، يحتوي النهج المقترح على بيانات Modcloth المباعة تم تطبيق أمازون التي تحتوي على 998،94 سجل معاملات.

تمت مقارنة النتائج التي تم الحصول عليها باستخدام مقاييس التقييم لتحديد النموذج الأفضل ، لوحظ أفضل مصنف K-means & SVM وقد حقق دقة 0.999 وبعدها K-means & KNN وقد حقق دقة تصل الى 0.998 وبعدها GMM&SVM حيث دقه 0.996 ثم حقق GMM&KNN حققت دقة 0.992.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة الانبار
كلية علوم الحاسبات وتكنولوجيا المعلومات
قسم علوم الحاسبات

أنظمة التوصية لتنبؤات السوق

رسالة مقدمة الى قسم علوم الحاسبات – كلية علوم الحاسوب وتكنولوجيا المعلومات – جامعة الأنبار، كجزء من متطلبات نيل شهادة الماجستير في علوم الحاسبات

تقدم بها

لميس يوسف عبد

بإشراف

الأستاذ الدكتور
مرتضى محمد حمد

الدكتور
أحمد جاسم محمد

1443 هجري

2021 ميلادي