International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

An Efficient Mining for Recommendation System for Academics

Nikhat Akhtar, Devendera Agarwal

Abstract: At present time huge numbers of research articles are available on World Wide Web in any domain. The research scholar explores a research papers to get the appropriate information and it takes time and effort of the researcher. In this scenario, there is the need for a researcher to search a research based on its research article. In the present paper a method of Knowledge ablation from a collection of research articles, is presented to evolve a system research paper recommendation system (RPRS), which would generate the recommendations for research article based on researcher choice. The RPRS accumulate the knowledge ablated from the pertinent research articles in the form of semantic tree. It accumulates all the literal sub parts with their reckoning in nodes. These parts are arranged based on their types in such a way that the leaf nodes stores the words with its prospect, the higher layer gives details about dictum with its reckoning, next to it an abstract. A Bayesian network is applied to construct a verisimilitude model which would quotation the pertinent tidings from the knowledge tree to construct the recommendation and word would be scored through TF-IDF value.

***Index Terms*: *Bayesian Network, Text Clustering, Knowledge Extraction, Text Classification,* Recommender System**

I. INTRODUCTION

Recommender systems are [1] one of the most [2] common and convenient conceivable [3] applications of Big data [4]. Presently, intense growth in the amount of information is available it is time consuming to explore the information of great value from the collective information [5]. This occurrence is called information overload. The information overload actually implies the availability of too much data or information that is beyond the manageable limits of the user and underlay big nodes in all sorts of decision makings. This onerous occurs mainly, when the system is unable to manage and process this huge amount of information in an orderly manner. Because of this, in many e-commerce applications, commonly the user has a plenty of options, but with a very limited time to enucleate them all. A recommendation system, the most magnificent mechanism in this direction, effort to tackle the information overload issue, and its subclass of information filtering system [4] that try to quest to prediction 'assessment' or 'precedence' [1] that a user would give to an item for instance research paper, books, music, movies or social element they had not yet considered. On the basics of previous effectuation of recommendation systems, there are two main elemental [5] prescripts to this hypostasis prescript and immaterial prescript. The hypostasis methods are analytical in character and immensely extensively used. These perspectives rank the phrase based on some important extent to devise the result [6]. In this paper, we discuss the generic hypostasis prescript to engender recommendations for conglomeration of associated corpus [7].

It is demonstrated tentatively that time decay in perusal is highly correspond with the organization of abstract and phrase [4]. This RPRS decomposes the process into two distinct sub processes. In the first process extraction of semantic knowledge of the corpus is done. In the second process, facts are going to be applied for the recommendation via Bayesian network [6] of word and abstract. The introduced RPRS are envisaged in elaboration in the subsequent section.

II. BACKGROUND

Recommender systems are increasingly used on the Web to assist users to discover material relevant to their interests. The first research article on recommender system was published in 1998. After that an impressive number of research articles had been published. The many circumstances have been explained to progress the credibility of recommender system. The Recommendation systems endow a promising method to ranking research article according to a user's choice [8]. The most vital feature of a recommender system is its proficiency to "supposition" a user's liking and interests by inspecting the demeanor of this user and & or the demeanor of other users to originate personalized recommendations [9]. In the [10], authors have talk about the offline and online evaluation of research paper recommender model and conclude that offline evaluation in this domain does not provide potential outcome. Yang et al. [11] presented a recommendation system for sciential papers that used a class oriented collaborative filtering method. However, their system overcomes the cold-start matter by utilizing implicit behaviors extracted from a subscriber access log web usage data are noisy and not genuine generally as pointed out in. Again research paper recommender model is introduced using a Dynamic Normalized Tree of Concepts (DNTC) model and a complicated ontology [12].

Manuscript published on January 30, 2020.
* Correspondence Author
Nikhat Akhtar*, Research Scholar Ph.D (Computer Science & Engineering), Babu Banarasi Das University, Lucknow, India
Prof. (Dr.) Devendera Agarwal, Professor, Babu Banarasi Das University, Lucknow, India

The system is evaluated offline using ACM digital library papers and the outcome show that this model performs preferable than the vector of space framework. Ekstrand et al. [13] introduced the two phases by running a content-based filtering and a collaborative based filtering recommender in parallel and admixture the resulting classed lists.

The first items on the modulate recommendation list are those items which appeared on together lists, ordered by the sum of their class. An astonishingly collaborative filtering outperforms all hybrid algorithms in their assessment. MKu¨cu¨ktunc et al. [14] developed a personalized paper recommendation service, called the Advisor3, which permission a subscriber to delineate her search toward neoteric progress research papers utilize a flank conscious unsystematic itinerancy with restart algorithm. The recommended articles regress by the Advisor is miscellaneous by parameterized ease up on spatial utmost.

Wang et al. [15] proposes to include textual information to construct a topic model of the research papers and adds an additional unnoticeable variable to differentiate between the focus of a research paper and the domain it just talks about. Sholom M. al. [16] introduced the algorithm in retrieving the most pertinent documents, meanwhile precision shows the capacity of the framework in excluding not interesting documents. Instantly the model's predictive performance is adjusted, the final phase consists of presenting new and unseen data and gets the concluding result of the classification model. Caragea et al. [17] introduced the matter of quotation recommendation utilize eccentric significance putrefaction of the contiguity matrix related with the quotation graph to build an abstruse semantic stead a deficient dimensional stead where consonant in research papers can be more dispassionately identified. The authors have proposed [18] keywords based retrieval procedure in for giving an overview and a multiple arrangement of papers as a piece of the preparatory reading index.

K. El-Ariniet al. [19] proposed by returns a set of reasonable research articles by optimizing an objective function based on a fine grained notion of impact between research papers and it says that, research article recommendation, elucidate a query as a small cluster of familiar with to be appropriate research papers are superior than a string of keywords. Authors have discussed [20] the performance of stereotype and most renowned recommendations in the area of scholarly recommender frameworks. W. Huang et al. [21] proposed the vocabulary used in the citation reference and in the content of the research article is absolutely dissimilar. To address this drawback, and presented to use translation model, which can viaduct the gap between two different mother tongue. The assortment performance is the simple performance of implementing assortment models [22]. X. Liu et al. [23] proposed citation recommendation from the miscellaneous network mining perspective has attracted more scrutiny. Moreover, research papers, metadata such as authors or keywords are also to be of opinion as entities in the graph schema. Tarus et al. [24] proposed to literature review is presented on ontology based recommender frameworks in the area of e-learning.

III. PROPOSED METHODOLOGY

A. Overview

At the beginning, a corpora has generated by amassing several research periodicals concerned to dissimilar domains. A conceptual perception is fabricated by made up of weighted textual content of all the research materials discovered pertinent to the given scholar's sample abstract. The pertinent papers in other words, conglomerations of homogeneous research periodicals are chosen based on prioritize their abstract using Naive Bayes classifiers. Subsequently, the abridged is going to be procreated by enforce a suggested hypostasis procedure on the conglomerated datum derived from the scholar's abstract. Generate the researcher's thought by applying Naive Bayes NB10 method, entire set of research periodicals are prioritize and the uppermost would be chosen in the recommendation process. The research paper's abstract materials are used in the procedure of prioritize [25]. The scoring is calculated by a Bayesian network technique. Thereafter, rank of abstract is contemplating to construct the recommendation.

To keeping in mind an abstract has characteristic of representing the entire content of a research paper, a recommendation list is generated from the selected research papers abstract [26]. Lastly a recommendation list retrieves and integrate abstract alongside entire the research papers. The sectionalisation of research periodicals is needed to divide into feasible fragment for creates the conceptual perception. The greater number of the research periodicals sequence in the segments of research periodicals for instance the name of a research paper abstract, introduction, related work, proposed method, results and last conclusions. In universally a lot of research papers are not accessible freely for example (IEEE, Science Direct, and ACM) in full text only research paper abstract is accessible.

Pseudo code

```
Procedure RPRS (t, S)
sSet ← NB_10 (S)
k ← ∅
While sSet ≠ ∅ do
s ← pop (sSet)
k ← k ∩ ABS (s)
end while
K_score ← BN (k)
RC ← RankAbs(K_score)
l ← ∅
While RCSet ≠ ∅ do
pop max(RCSet)
T ← title (RC)
l ← l ∪ T
end while
return l
end procedure
```

In this scenario, we will use only abstract in our proposed idea. That will utilize in the process of construction of conceptual perception that will play the role of essential constituents for the recommendation formation.

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Through hypostasis manner [26] the significant portion of our work is that the presentation of matter of fact, research paper literal elements are being ranked. The result of previous stage, at the moment we need to be stored in the conceptual perception, so that appropriate prioritization of all research paper wording identical to dictum extant at differing schema of research paper performed. The objective of the present paper is to construct a list that will provide the title of the research paper.

B. The Knowledge Tree formation

Certain features are desirable for the perception delineation [27] in terms of apparent, emphatic, better expressiveness, dexterous, influential, and independent of the circumstances [28]. The tree much robust knowledge base [29] that can accumulate the regain intimation from the collection of research paper. It's completely arduous to split the document if the limitations of segments are not explicit. In this context, stream of words are not identical via whole research paper script. For the suitable intellect eduction, research paper script becomes split suitably at the point where word transformation is high [30]. As the research documents we are previously split into many segments, and split according to the order of headings.

The knowledge is extracted from the abstracts of all the research paper and become visible for effortlessly conceivable and able to be used in our tree structure. After that arranged in order of delineation intellect is utilized to conglomeration the similar research paper documents, on the basis of category of research paper documents be related to shown in figure 1. This amalgamates intellect framework applying for the creation of an accurate and pervasive knowledge. The objective of this circumstance arranged intellect delineation is the [31] concernment of research paper textual units is comfortably roughly calculated in bottom-top manner. Afterward the generation of a tree every research paper script is split into constituent ahead into dictum. This procedure requires recurring until vocabulary's level of Eduction accomplished triumphantly. In this framework word would be scored firstly through term frequency inverse document frequency (*tf-idf*) value.

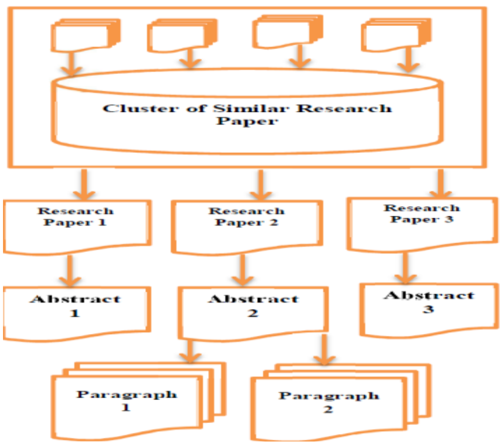

Figure 1: Knowledge Tree framework

C. Bayesian Network Based Scoring

The research paper abstract choices are done with a Bayesian network of this research paper abstract. Bayesian networks are probabilistic models that clearly take over the familiar conditional dependence with directed edges in a particular graph ideality and this ideality can demonstrate correlation between variables and used to calculate likelihood. These network models take over both conditionally dependent and conditionally independent relationships between random variables. Consequently, the connections among nodes point out the likelihood reliance among them. Basically, these networks have two components, firstly the directed acyclic graph (DAG). In this first component Bayesian network B would be defined to represent a 2 tuple $< D,\lambda >$ where D is the directed acyclic graph whose nodes are indiscriminate variable N_1, N_2, N_3, . . . , N_n and where connections indicate provisory reliance among [32] arbitrary variables. The secondly components condition, likelihood table for every variable. In this component of the pair λ indicates the conglomeration of that reliance likelihood amidst the vertices. The conglomeration includes measurable factor $\lambda\, N_i |\sigma_i = L_B (N_i |\sigma_i)$ showing likelihood of N_i conditioned on σ_i.

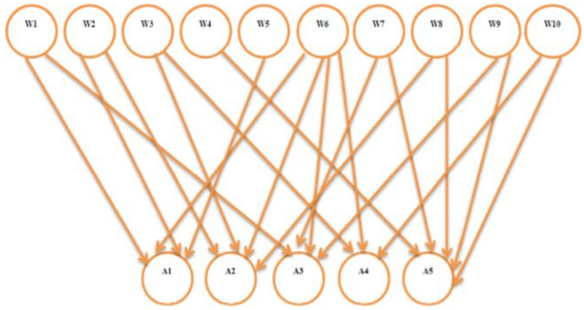

Figure 2: The Bayesian Network Construct among Word Vertices & Abstract Vertices

A Bayesian network B is procreated among abstract & word provided in figure 2. Here upon these network procreated via linking the abstract vertices to word vertices, if the particular word be suited with that abstract. Again, Tf-idf of word is applied for the calculation of the edges' weight to spruce the next factor λ of the Bayesian network. The likelihood of the abstract is count up in the given expression.

$$L_B(L_j) = \prod_{i=1}^{k} \lambda L_{j|\sigma_i}$$

In training procedure it reads the text from research paper sentence by sentence, discover the count of feature words and differentiate with the abstract stored. Again, it computes the frequency of every feature word. In the check-up procedures it discovers the counts of feature words to categories according to the research paper.

IV. RECOMMENDATION RESULTS & EVALUATION

The presented RPRS method is computed by a repository that has been made by selection some research paper. For instance research papers domain like Artificial Intelligence, Network Security, Data Mining, Big Data, and Document Recommendation. Every domain is constructed by the collection of same kind of research papers.

Based on the likeness with given research articles, appropriate research articles would be chosen from the knowledge tree and used for recommendation purpose. From the classified cluster of research paper that contain research paper from various field like Artificial Intelligence, Network Security, Data Mining, Big Data [4][33], and Document Recommendation recommender system the domain of paper is identify by using Bayesian classification. After this identification process a set of research papers abstract is used for the ranking process by using Naïve based classification NB10 approach.

The steps of this ranking process are given as follows. We have considered 5 abstract [34][35][36][37][38]. For the sake of simplicity 4 abstract are used in training the system and one abstract is used for testing the system.

Step 1 Data ablation technique is applied on all 4 training research paper abstract to consider only conceptual word means delimiters and common words like (the ,is, there, etc) are removed.

Step 2 The tables for words are create that include word count of each unique word.

Table- I: The Bag of the Words Representation Paper 1

No.	Words	Counts
1	Text Mining	1
2	Techniques	1
3	Statistical analysis	2
4	Word, term, phrase	8
5	Frequency	2
6	Document	3
7	Same	2
8	Meaning	2
9	Contributes	2
10	Appropriate	1
11	Semantics	1
12	Text	1
13	More	1
14	Importance	2
15	Concept -based mining	1
16	Analyses	1
17	Terms based	2
18	Sentence level	1
19	Document level	1
20	Corpus level	1
21	Sentence based	1
22	Concept analysis	2
23	Term frequency	2
24	Document term frequency	2
25	Document based	1
26	Corpus based	1
27	Document frequency	2
28	Concept based	1
29	Similarly	1
30	Corpus	1
31	Concept-Based Analysis Algorithm	1
32	Clusters	2
33	Web	1
34	Efficient	1
35	Quality	1
36	Model	1
37	Significantly	1
38	Traditional	1
39	Single -term-base	1
40	Approaches	1
	Vocabulary = 40	Total Counts = 61

Table- II: The Bag of the Words Representation Paper 2

No.	Words	Counts
1	Text Mining	1
2	Techniques	1
3	Statistical analysis	2
4	Terms	3
5	Enhance	1
6	Document	3
7	Sentence	2
8	Mining	2
9	Corpus based	1
10	Concept based	1
11	Clustering	1
12	Text	1
13	Model	2
14	Important	1
15	Concept based mining	1
16	Analysis	2
17	Quality	1
18	Corpus	1
19	Approach	1
20	Corpus level	1
21	Sentence based	1
22	Concept analysis	2
23	Similarity	1
24	Measure	1
25	Document based	1
	Vocabulary = 25	Total Counts = 35

Table- III: The Bag of the Words Representation Paper 3

No.	Words	Counts
1	Study	3
2	Dictionary based	1
3	Method	2
4	Extract	2
5	Expressive	1
6	Concepts	2
7	Document	2
8	Concept mining	1
9	English	1
10	Turkish	1
11	Agglutinative	1
12	Language	1
13	Immature	1
14	Dictionary	2
15	WordNet	1
16	Lexical	1
17	Database	1
18	Grouping	1
19	Word	1
20	Synsets	1
21	Synonyms	1
22	Hypernyms	1
23	Hyponyms	1
24	Relationships	1
25	Meaning	1
26	Text	1
27	Success	1
28	High	1
29	Determining	1
30	Collected	1
31	Corpora	1
	Vocabulary = 31	Total Counts = 38

Table- IV: The Bag of the Words Representation Paper 4

No.	Words	Counts
1	Quick	1
2	Growth	1
3	Information	1
4	Extract	2
5	Discover	1
6	Knowledge	1
7	Data sources	1
8	World Wide Web	1
9	Methods	2
10	Text mining	1
11	Statistical analysis	2
12	Phrase or word	1
13	Document	5
14	Bags of words	1
15	Importance	1
16	Meaning	2
17	Content	1
18	Term	5

19	Frequency	2
20	Significance	1
21	Within	1
22	Same	1
23	Pays	1
24	Sentence	1
25	Concept -based	3
26	Analyses	2
27	Planned	1
28	Model	2
29	Clustering	1
30	K-means	1
31	Similarity	1
	Vocabulary = 31	Total Counts = 48

Step 3 Simultaneously a common list that contain all unique words from all the research papers abstracts is also generated that is called vocabulary of the corpora.

Table- V: The Common List Vocabulary of the Corpora

No.	Words
1	Text Mining
2	Techniques
3	Statistical analysis
4	Word, term, phrase
5	Frequency
6	Document
7	Same
8	Meaning
9	Contributes
10	Appropriate
11	Semantics
12	Text
13	More
14	Importance
15	Concept -based mining
16	Analyses
17	Terms based
18	Sentence level
19	Document level
20	Corpus level
21	Sentence based
22	Concept analysis
23	Term frequency
24	Document term frequency
25	Document based
26	Corpus based
27	Document frequency
28	Concept based
29	Similarly
30	Corpus
31	Concept-Based Analysis Algorithm
32	Clusters
33	Web
34	Efficient
35	Quality
36	Model
37	Significantly
38	Traditional
39	Single -term-base
40	Approaches
41	Mining
42	Sentence
43	Measure
44	Enhance
45	Study
46	Dictionary based
47	Method
48	Extract
49	Expressive
50	Concepts
51	Concept mining
52	English
53	Turkish
54	Agglutinative
55	Language
56	Immature
57	Dictionary
58	WordNet
59	Lexical
60	Database
61	Grouping
62	Synsets
63	Synonyms
64	Hypernyms
65	Hyponyms
66	Relationships
67	Success
68	High
69	Determining
70	Collected
71	Quick
72	Growth
73	Information
74	Discover
75	Knowledge
76	Data sources
77	World Wide Web
78	Bags of words
79	Content
80	Within
81	Pays
82	Planned
83	K -means
	Vocabulary = 83

Step 4 The probabilities of each research paper document is computed.

$$\textbf{p of Set s} = \frac{\textbf{Entire number of training phrase belonging to Set s}}{\textbf{Entire number of phrase in the training set}}$$

Table- VI: The Training abstracts probability p of set "s"

Class c	Training abstracts probability p of set s	
D_1	10	10/(27)= 0.37037
D_2	5	5/(27)= 0.18518
D_3	5	5/(27)= 0.18518
D_4	7	7/(27)= 0.25925

Step 5 After that the verifying abstract is considered. For the sake of ingenuity we are using only one longest sentence.

> Concept Mining has become an important research area. Concept Mining is used to search or extract the concepts embedded in the text document. Concept based approach search for the informative terms based on their meaning rather than on the presence of the keyword in the text.

Step 6 After text ablation techniques only certain words will identified these words are used for the computation task.

> **Concept Mining** xxx xxxxxx xx xxxxxxxxx **research** xxxx. **Concept Mining** xx xxx xx **search** xx **extract** xxx **concepts embedded** xx xxx **text document**. **Concept based approach search** xxx xxx **informative terms** xxxx xx xxx **meaning** xxxx xxxx xx xxx **presence** xx xxx **keyword** xx xxx **text.**

Step 7 In subsequent computation we find the ranking of all abstract in these ranking the highest certain papers are recommended to the user.

Table- VII: The Enumeration of Assay Term in "k" in Set "s"

Class c	Enumeration of Assay Term in "k" in Set "s"								
	Concept Based	Approach	Search	Information	Term	Meaning	Present	Keyword	Text
D_1	2	2	1	1	9	3	1	1	2
D_2	2	2	1	1	4	1	1	1	2
D_3	1	1	1	1	2	1	1	1	1
D_4	2	1	1	2	1	2	1	1	1

Table- VIII: The Probability of Assay Term in "k" in Set "s"

Class c	Probability of Assay Term in "k" in Set "s"								
	Concept Based	Approach	Search	Information	Term	Meaning	Present	Keyword	Text
D_1	0.01379	0.01379	0.00689	0.00689	0.06200	0.02068	0.00689	0.00689	0.01379
D_2	0.01680	0.01680	0.00840	0.00840	0.03361	0.00840	0.00840	0.00840	0.01680
D_3	0.00819	0.00819	0.00819	0.00819	0.01639	0.00819	0.00819	0.00819	0.00819
D_4	0.01515	0.00757	0.00757	0.01515	0.00757	0.01515	0.00757	0.00757	0.00757

Table- IX: The Multiplication M of Assay Term in "k" in Respective Set

Class c	Multiplication [M of Assay Term in "k" in Set Propitious s] = M(Concept Based) * M (Approach) * M (Search) * M (Information) * M (Term)* M (Meaning) * M (Present) * M (Keyword) * M (Text)								Multiplication
	Concept Based	Approach	Search	Information	Term	Meaning	Present	Keyword	Text	
D_1	0.01379	0.01379	0.00689	0.00689	0.06200	0.02068	0.00689	0.00689	0.01379	$7.57725512E^{-18}$
D_2	0.01680	0.01680	0.00840	0.00840	0.03361	0.00840	0.00840	0.00840	0.01680	$6.66488696E^{-18}$
D_3	0.00819	0.00819	0.00819	0.00819	0.01639	0.00819	0.00819	0.00819	0.00819	$3.31779973E^{-19}$
D_4	0.01515	0.00757	0.00757	0.01515	0.00757	0.01515	0.00757	0.00757	0.00757	$6.54354624E^{-19}$

Table- X: The r related to Set "s"

Class c	M [r related to Set "s"] = Multiplication [M of Assay Term in "k" in Set s] * M of Set s									Multiplication
	Concept Based	Approach	Search	Information	Term	Meaning	Present	Keyword	Text	
D_1	0.01379	0.01379	0.00689	0.00689	0.06200	0.02068	0.00689	0.00689	0.01379	$2.80638798E^{-18}$
D_2	0.01680	0.01680	0.00840	0.00840	0.03361	0.00840	0.00840	0.00840	0.01680	$1.23423709E^{-18}$
D_3	0.00819	0.00819	0.00819	0.00819	0.01639	0.00819	0.00819	0.00819	0.00819	$6.14390154E^{-20}$
D_4	0.01515	0.00757	0.00757	0.01515	0.00757	0.01515	0.00757	0.00757	0.00757	$1.69641436E^{-19}$

Step 8 In our example among 4 abstract 3 are used for recommendation process and lowest rank is left because it is least relevant.

With the same procedure we have generate recommendation for other subject field research papers. The performance measures namely as precision, recall, and *F*-score are calculated for the introduced technique. Precision estimates the chastity percentage and recall measures the perfection of the summarizer and uniting of precision and recall, *F*-score is evaluated.

$$\text{Precision} = \frac{\text{relevant research paper} \cap \text{retrieved research paper}}{\text{retrieved research paper}}$$

$$\text{Recall} = \frac{\text{relevant research paper} \cap \text{retrieved research paper}}{\text{relevant research paper}}$$

$$\text{F-Measure} = \frac{\text{Recall} * 2 * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Table- XI: Knowledge Base Retrieved Research Paper

No.	Knowledge Base Area	Retrieved	Not Retrieved	Not Pertinent	Not Pertinent not Retrieved
1	Big Data	**36**	**10**	**5**	**49**
2	Artificial Intelligence	**45**	**6**	**3**	**46**
3	Document Recommendation	**50**	**13**	**5**	**32**
4	Network Security	**32**	**8**	**3**	**57**
5	Data Mining	**34**	**4**	**2**	**60**

1. **The values of precision are calculated as,**

For Big Data, Precision = 36/ (36+5) = 0.8780
For Artificial Intelligence, Precision = 45/ (45+3) = 0.9375
For Document Recommendation, Precision = 50/ (50+5) = 0.9090
For Network security, Precision = 32/ (32+3) = 0.9142
For Data Mining, Precision = 34/ (34+2) = 0.9444

2. **The values of recall are calculated as,**

For Big Data, Recall = 36/ (36+10) =0.7826
For Artificial Intelligence, Recall =45/ (45+6) = 0.8823
For Document Recommendation, Recall = 50/ (50+15) = 0.7936
For Network security, Recall = 32/ (32+8) = 0.8000
For Data Mining, Recall = 34/ (34+8) = 0.8947

3. **The values of f-measure are calculated as,**

For Big Data, f-measure = 1.3742/1.6606= 0.8275
For Artificial Intelligence, f-measure = 1.6543/1.8198= 0.9090
For Document Recommendation, f-measure = 1.4427/1.7026= 0.8473
For Network security, f-measure = 1.4627/1.7142= 0.8532
For Data Mining, f-measure = 1.6899/1.8391= 0.9188

Table- XI: The Result of Precision, Recall, F- Measure

No.	Knowledge Base Area	Precision	Recall	F-measure
1	Big Data	0.8780	0.7826	0.8275
2	Artificial Intelligence	0.9375	0.8823	0.9090
3	Document Recommendation	0.9090	0.7936	0.8473
4	Network Security	0.9142	0.8000	0.8532
5	Data Mining	0.9444	0.8947	0.9188

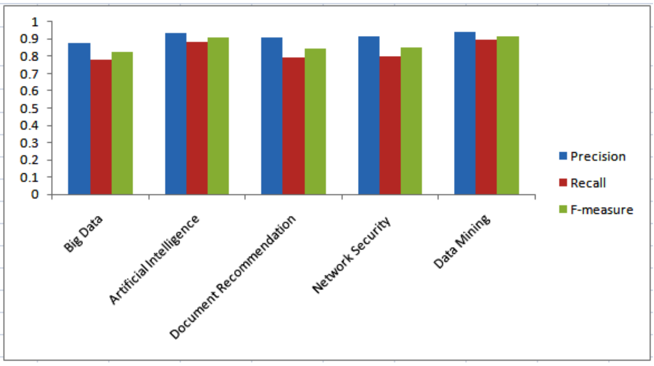

Figure 3: The Performance Evaluation of Precision, Recall, F- Measure Outcome

The X-axis represents the knowledge base of the research paper like Artificial Intelligence, Network Security, Data Mining, Big Data, and Document Recommendation and the Y-axis indicates the scale of intervals. If recall, f- measure, precision reaches near 1 it shows that the system is more effective.

V. CONCLUSION

In this paper we have introduced the Bayesian classification method for recommending research paper. This method is favorable for our purpose. As it performed a better classification method so we are easily able to identify the class of the given paper, then by using a naïve base NB10 approach we have computed the rank of selected categories' papers and highest rank papers are recommended to the user. The paper uses only abstract that are most of the time publicly available as compare to other baseline method that uses whole text that are not always publicly available. Our proposed method has presented the expressive reformation better than baseline method of evaluate the executions and potency to retrieve legitimate result with advantageous publication on the uppermost part of the recommendation list. Consequently, outcome of this recommendation has acme of recall and precision. Subsequently we can elaborate the technique for real time application also.

REFERENCES

1. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds): Recommender Systems Handbook, pp. 1–35. Springer, Berlin , 2011.
2. Nikhat Akhtar, Dr. Devendera Agarwal, "A Literature Review of Empirical Studies of Recommendation Systems" International Journal of Applied Information Systems (IJAIS), ISSN : 2249-0868, Foundation of Computer Science FCS, NewYork, USA, Volume 10 – No.2, Pages 6 – 14, Dec 2015, DOI:10.5120/ijais2015451467
3. Basu, C., H. Hirsh, and W. Cohen. Recommendation as classification: Using social and content-based information in recommendation. In Recommender Systems.Papers from 1998 Workshop. Technical Report WS-98-08.AAAI Press, 1998.
4. Yusuf Perwej, "An Experiential Study of the Big Data," International Transaction of Electrical and Computer Engineers System (ITECES), USA,ISSN (Print): 2373-1273 ISSN (Online): 2373-1281, Vol. 4, No. 1, page 14-25, March 2017, DOI:10.12691/iteces-4-1-3.
5. Nikhat Akhtar, Firoj Parwej, Yusuf Perwej, "A Perusal Of Big Data Classification And Hadoop Technology," International Transaction of Electrical and Computer Engineers System (ITECES), USA, ISSN (Print): 2373-1273 ISSN (Online): 2373-1281, Vol. 4, No. 1, page 26-38, May 2017, DOI: 10.12691/iteces-4-1-4
6. Nikhat Akhtar, "A Model- Based Research Material Recommendation System For Individual Users" for published in the Transactions on Machine Learning and Artificial Intelligence (TMLAI), Society for Science and Education, United Kingdom (UK), ISSN 2054-7390,Vol.5, Issue 2, Pages 1 - 8, March 2017, DOI : 10.14738/tmlai.52.2842
7. M. Lapata, Automatic Evaluation of Information Ordering: Kendall's Tau, Computational Linguistics, vol. 32, no. 4, pp. 471–484, 2006
8. Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles, "Citation recommendation without author supervision," in Proc. of WSDM, pp.755–764, 2011
9. Nikhat Akhtar, Dr. Devendera Agarwal, "A Survey of Imperfection of Existing Recommender Systems for Academic Fraternity", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 20, Issue 3, Ver. III, PP 08-15, May - June. 2018, DOI:10.5120/ijais2015451467
10. Beel, J., et al. A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In Proceedings of the international workshop on reproducibility and replication in recommender systems evaluation. 2013. ACM
11. D. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang. CARES: A Ranking-Oriented CADAL Recommender System. In Proc. of the 9th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2009),pages 203–211, 2009
12. Al Alshaikh, M., G. Uchyigit, and R. Evans, A research paper recommender system using a dynamic normalized tree of concepts model for user modeling, 2017
13. M. D. Ekstrand, P. Kannan, J. A. Stemper, J. T. Butler, J. A. Konstan,and J. T. Riedl, "Automatically building research reading lists," in Proc.ofRecSys, 2010, pp. 159–166.
14. O. Ku¨c¸u¨ktunc¸, E. Saule, K. Kaya, and U . V. C¸ atalyu¨rek, "Towards a personalized, scalable, and exploratory academic recommendation service," in Proc. of ASONAM, 2013
15. C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in Proceedings of the 17th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD "11. New York, NY, USA: ACM, pp. 448–456, 2011
16. Sholom M. Weiss, Nitin Indurkhya, and Tong Zhang. Fundamentals of Predictive Text Mining. Springer, 2010.
17. C. Caragea, A. Silvescu, P. Mitra, and C. L. Giles, "Can"t see the forest for the trees?: a citation recommendation system," in Proc. Of JCDL, pp. 111–114, 2013
18. Raamkumar, A.S., S. Foo, and N. Pang, Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. Information Processing & Management, 2017. 53(3): p. 577-594
19. K. El-Arini and C. Guestrin, "Beyond keyword search: discovering relevant scientific literature," in Proc of KDD, pp. 439– 447, 2011
20. Beel, J., et al. Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. In Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017)-(isi2017), Humboldt-Universität zu Berlin, 2017
21. W. Huang, S. Kataria, C. Caragea, P. Mitra, C. L. Giles, and L. Rokach,"Recommending citations: translating papers into references," in Proc.of CIKM, pp. 1910–1914, 2012
22. Yang chang Zhao. R and data mining: Examples and case studies. In Yangchang Zhao, editor, R and Data Mining, pages 1–4. Academic Press, 2013.
23. X. Liu, Y. Yu, C. Guo, and Y. Sun, "Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation,"in Proc. of CIKM, pp. 121–130, 2014
24. Tarus, J.K., Z. Niu, G. Mustafa, "Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning", Artificial Intelligence Review, Volume 50, Issue 1, Pages 21-48, 2018
25. Yusuf Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online) : 2278-1021, ISSN (Print): 2319-5940, Vol.4, Issue 2, Pages 10 - 16, February 2015, DOI: 10.17148/IJARCCE.2015.4203
26. H. Jingand K. McKeown. Cut and paste based text summarization. In Prodeedings of NAACL 2000
27. T. Fujiki, H. Nanba, and M. Okumura, „Automatic acquisition of script knowledge from a text collection", EACL 2003 in Budapest, Hungry, 91-94
28. Mylopoulos, J., "An overview of knowledge representation", 1981, ACM press New York, NY, USA
29. M. Kerber, "Knowledge Representation I", Introduction to AI 06-08775, 2004/05

30. Laurence A. F. Park, Marimuthu Palaniswami, and Ramamohanarao Kotagiri, "Internet document filtering using fourier domain scoring," in Principles of Data Mining and Knowledge Discovery, Luc de Raedt and Arno Siebes, Eds, number 2168 in Lecture Notes in Artificial Intelligence, pp. 362–373, Springer-Verlag, . September 2001
31. Lin, C.,&Hovy, E. (2001). Neats: A multi document summarizer. In Proceedings of the document understanding workshop (DUC)
32. I. Ben-Gal, Bayesian Networks, Encyclopedia of Statistics in Quality and Reliability, 2007
33. Yusuf Perwej, Firoj Parwej, M. M. Mohamed Hassan, Nikhat Akhtar, "The Internet-of-Things (IoT) Security: A Technological Perspective and Review", International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5, Issue 1, Pages 462-482, February 2019, DOI: 10.32628/CSEIT195193
34. Lincy Liptha R, Raja K., G.Tholkappia Arasu," Enhancing Text Clustering Using Concept-basedMining Model", International Journal of Electronics and Computer Science Engineering, ISSN-2277-1956, Volume1,Number 2,pp 550-556, 2016
35. Pradnya Randive, Nitin Pise," Improving Text Clustering Quality by Concept Mining", Int. Journal of Engineering Research and Applications, ISSN : 2248-9622, Vol. 3, Issue 5, pp.1701-1704, Sep-Oct 2013
36. Cem Rıfkı Aydın, Ali Erkan, Tunga Güngör and Hidayet Takçı," Dictionary-based concept mining: an application for turkish", International Conference on Foundations of Computer Science & Technology (CST 2014), Zurich, Switzerland, January 2014, DOI:10.5121/csit.2014.4103
37. K. N. S. S. V. Prasad, S. K. Saritha," A Survey Paper on Concept Mining in Text Documents", International Journal of Computer Applications (0975 – 8887) Volume 166 – No.11, May 2017
38. K.N.S.S.V. Prasad, S. K. Saritha," Concept Mining in Text Documents using Clustering", International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 48, April 2019

AUTHORS PROFILE

Nikhat Akhtar, Research Scholar Ph.D (Computer Science & Engg.) in the Department of Computer Science & Engineering, Babu Banarasi Das University, Lucknow, India. She has authored a number of papers in different journal. Her research interests include Recommender System, Bayesian Network, Text Categorization and Clustering, Cryptography, Big Data. He is a member of IEEE.