A Naive Bayes-Based Model for Predicting Users' Information Needs in Digital Library Systems

Bekkamov Fayzi Absoatovich

[*] *International Islamic Academy of Uzbekistan, Senior Lecturer, PhD, Tashkent, Uzbekistan*

bfayzi93@gmail.com

Article type	Received	Accepted	Published
Original Research	15 November 2025	26 February 2026	12 March 2026

ABSTRACT

The contemporary digital library systems have been functioning in the environment that is characterized by the rapidly increasing volume of information resources, so it is becoming more challenging to be able to offer the users the relevant materials in the most efficient way. Classical methods of search only react to explicit user requests and do not always detect the hidden or new information requirements. This paper represents a smart model to predict the information needs of the users in library systems based on the Naive Bayes classification algorithm. The proposed model examines demographic, academic and behavioral user characteristics such as age, profession, education, search history, download and even preferred subject areas to predict the likelihood of user interest in a particular type of information. The system will create individualized recommendations of resources based on the projected information requirements. The experimental analysis of the approximate accuracy of the proposed method based on user data in the Libsmart digital library system indicates that the given method has an accuracy of 86.4, which proves its efficiency in providing adaptive and user-based information services. The intelligent recommendation modules in digital library setting based on the proposed model can act as a viable basis.

Keywords: *Library systems, information need prediction, Naive Bayes, intelligent recommendation, digital libraries, user behavior analysis.*

DOI: https://zenodo.org/records/20536026 **How to cite:** *Bekkamov F.A. (2026). A Naive Bayes-Based Model for Predicting Users' Information Needs in Digital Library Systems. Nexus International Journal of Science and Technology, 15(1), 125–131.*

1. INTRODUCTION

The increased development of digital information resources has greatly changed the role of the modern library systems. Libraries are no longer seen as places of storing and accessing printed materials only and has grown into smart information spaces which have to ensure quick, precise and individualized access to various knowledge support. In this regard, information needs of users have emerged as one of the most crucial issues in the library and information science to comprehend and anticipate. The users of library system are varied in terms of age, profession, educational level, field of research and information seeking behavior. Consequently the same set of resources can be related to quite different requirements of different users. The conventional systems of searching by keywords work only in case the user can clearly define queries. But in practice, in most cases, users do not actually state their needs, or their interests can also evolve with time. This means that the library systems today need smart systems that are capable of analyzing user characteristics and behavior to enable forecasting their information demand in future and give appropriate advice.

The opportunities of solving such problems have been opened with the help of artificial intelligence techniques [1], [2]. Widely used studies on recommender systems and machine learning methods have been in information

DOI: 10.5281/zenodo.20536026

filtering, user modelling and personalised retrieval environments [3], [4], [5]. Collaborative filtering methods have also become a significant part of the current research in recommendations and serve as a valuable basis of individual information access [6]. Machine learning methods are particularly helpful with analyzing user data patterns and determining probable areas of interest among them. The Naive Bayes classifier is one of the most appropriate techniques towards this end [7], [8]. The technique is premised on probabilistic arguments and enables prediction of a label of a class given a combination of observed features. However, despite its simplicity, Naive Bayes has been shown to be quite good in several classification and recommendation problems. A Naive Bayes-based model can be applied in library systems to approximate whether a user is interested in a specific subject area or in a specific type of resource according to demographic and behavioral traits. These kinds of indicators can be age and profession, the level of education, the past search query, the documents downloaded, ratings, history of borrowing, and the favorite topics. With processing such features, the system is able to forecast the most probable information requirement of the user and assist more adapting and user-based services. This paper is aimed at coming up with an intelligent model based on the Naive Bayes that will predict the information requirement of users in library systems. The paper provides the conceptual framework of the model, explains the key characteristics to be used in prediction and how the mechanism of probabilistic classification can be implemented in a library setting. The suggested solution can be used to develop smart recommendation systems in case of digital libraries and automatic information systems [2]. General recommender-system principles and implementations have also been synthesized in past survey and overview research [9], [10].

2. LITERATURE REVIEW

2.1 Information Needs and User Modeling in Library Systems

The academic literature on information needs in library systems has evolved from passive query-response models toward predictive and adaptive approaches. Manning, Raghavan, and Schütze [11] established that an information need arises when a user recognizes a knowledge gap required to solve a problem or accomplish a professional goal. Bekkamov, Babajanov, and Berdimurodov [9] further explored Bayesian methods for predicting users' information needs in digital libraries, demonstrating that probabilistic models effectively capture the relationship between user attributes and subject-area interests. Bekkamov and Rustamov [10] additionally provided a broader overview of recommender system architectures, situating probabilistic user modeling within the wider landscape of personalization techniques.

2.2 Recommender Systems and Naive Bayes Classification

Recommender systems have been widely studied across e-commerce, media, and educational platforms, with collaborative filtering emerging as one of the most influential paradigms [3], [6]. Probabilistic classifiers, and Naive Bayes in particular, have demonstrated strong performance in recommendation tasks due to their computational efficiency and interpretability. Valdiviezo-Diaz et al. [1] showed that Naïve Bayes-augmented collaborative filtering improves prediction accuracy, while Özcan, Göz, and Temel [7] confirmed its effectiveness in e-learning recommendation using demographic and behavioral features. Rrmoku, Selimi, and Ahmedi [8] further validated that Naive Bayes produces consistent and interpretable results even in the presence of noisy user data, making it particularly suitable for institutional library environments.

3. RECOMMENDER SYSTEMS AND INFORMATION NEED PREDICTION

3.1 Information Needs in Library Systems

The information need is typically viewed as a condition when a person understands he/she does not know something necessary to address a problem, solve a task, or fulfil a personal/professional goal [11]. Information requirements in library systems are manifested in different ways such as catalog, subject, borrowing, downloading, and resource reviews. Such interactions are very useful as they offer important signals that can be utilized to respond to user interests and their future resource needs.

Recommender systems are also widely researched in the literature, both in survey and handbook papers and foundational machine learning literature. Collaborative filtering continues to be among the most influential personalized recommendation paradigm [6], whereas Bayesian and probabilistic methods have demonstrated good

DOI: 10.5281/zenodo.20536026

potential of scalable and explainable user modeling [12]. Moreover, Naive Bayes-based recommendation algorithms have been used in the semantic-user recommendation setting, e-learning, and movie recommendation setting [13], [14], [15]. Domain-specific recommender models have also been suggested in more general situations of personalization like travel recommendation [16]. Overview studies also talk of general recommender-system concepts and architectures [8].

In the conventional library practice, the information need of the user is typically responded to once it is clearly stated. Nevertheless, this reactive kind of method is not always an effective way of delivering service particularly in the online world where customers demand quick and personalized service. Because of this, more interest is being shown in predictive models that are able to anticipate information requirements and help users more proactively.

3.2 Intelligent Recommendation in Digital Libraries

The recommendation systems are common in e-commerce, media platforms, and online learning systems [8]. They are primarily meant to propose items which will tend to be interesting to a user based on past interactions, interests and similarities between users or items. Recommendation systems in digital library setting can be applied to propose books, articles, theses, and reports among other resources that are of academic or professional interests to the user. There are a number of recommendation strategies that are used. Content-based filtering is concerned with the properties of resources that the user has already liked [12]. One of the most popular paradigms of recommendation that has been widely used in the literature and studied is collaborative filtering. Naive Bayes-based recommendation techniques have also been used in recent studies in systems that are related to education and the media. Complex application Hybrid recommendation models have also been suggested in complex application settings like travel recommendation and large scale probabilistic personalization [16]. Collaborative filtering relies on the likes of the similar users to suggest new items. Demographic filtering is based on social and demographic characteristics like age, occupation, and the level of education. Each approach has its merits in library systems, however, they are also limited to cases where data about the users are sparsely distributed, or where preferences are not explicitly represented. Thus, probabilistic classification models are a good substitute [9], [17], [18].

3.3 Naive Bayes as a Prediction Procedure

Naive Bayes is a probabilistic classifier that is built on the Bayes theorem. It presumes that the features that are being used to make predictions are conditionally independent with reference to the name of the class. This assumption is simplified but the approach works quite well in most real-life applications. Naive Bayes works based on the idea that one can compute on what can be understood as the posterior probability of given a set of observed features. In this research, the class can be a subject category or a form of information resource to the user. These characteristics can be demographic and behavioral attributes based on the activity of the users within the library system. The concept of similar Bayesian and Naive Bayes-based methods has been effectively applied to the task of recommending competent users, learning strategies, and personalized content in various fields of application.

The Bayes theorem would be expressed as follows:

$$P(C|X) = \frac{P(X|C)\cdot P(C)}{P(X)} \tag{1}$$

where:

C - is the target class,

X - represents a set of observable features of users,

$P(C|X)$ - is the posterior probability of class C given features X,

$P(X|C)$ - is the probability of X to be observed in the class C,

$P(C)$ - is the prior probability of class C,

$P(X)$ - is the evidence.

Under the Naive Bayes assumption, the likelihood term is simplified as:

$$P(X|C) = \prod_{i=1}^{n} P(x_i|C) \tag{2}$$

where, x_i represents the individual features of the user profile.

DOI: 10.5281/zenodo.20536026

Therefore, the prediction rule will be:

$$\hat{C} = \arg max\, P(C) \prod_{i=1}^{n} P(x_i|C) \tag{3}$$

This implies that the user is assigned to the class that has the highest probability of occurrence in the system. Naive bayes possesses a number of strengths, which render it appropriate to library systems. To begin with, it is computationally efficient and can be applied to quite small datasets. Second, it can be interpreted readily, which is significant in the academic and institutional setting where open-minded decision-making is appreciated. Third, it works well with categorical and organized user data, a typical library environment. Lastly, the algorithm can be incorporated into the current digital library system without the need to use very complicated computer infrastructure.

3.4 Model Design for User Information Need Prediction

The proposed smart model will utilize the information requirements of the users in a library system by processing demographic, academic and behavioral characteristics. The model incorporates personal user characteristics, e.g., age, occupation, education level, search history, downloaded content, ratings, and the subject favors, so as to determine the most likely information-interest category. Based on this classification, the system would come up with the relevant resources recommendations which would be based on the predicted area of interest of the user. In such approach, it will be possible to leave the old query-based retrieval paradigm behind and deliver more dynamic, more personalized, and user-oriented library services. The input features of the model include: age, gender, profession, education level, department or specialization, previous search queries, downloaded resources, borrowing history, preferred document type, ratings or feedback, frequently selected subject areas.

These characteristics have been chosen due to the relative stability of the user and active nature of the information seeking behavior. The age, profession and education level are examples of stable characteristics whereas the search history, downloaded resources, ratings and areas of preferences are examples of dynamic characteristics. All these characteristics make the model more complete in its description of the user and enhance the accuracy of information-need prediction.

The target classes are the key thematic groups of information resources that are accessible in the library system. These groups reflect the subject-areas of the majority of users where a user is supposed to show interest. The target classes can be the following, depending on the organization of the library collection: Computer Science, Information technology, Education, Economics, Medicine, Literature, Law.

The system suggests the likelihood of a user being in each category and classifies the user to the category with the most probability. The proposed model works on the basis of gathering user profile and behavioral information, converting them to a feature vector and the use of the Naive Bayes classifier to approximate the most likely information-need category. The system suggests the pertinent library resources based on the category that was predicted. Fig. 1 shows the overall design of the proposed model.

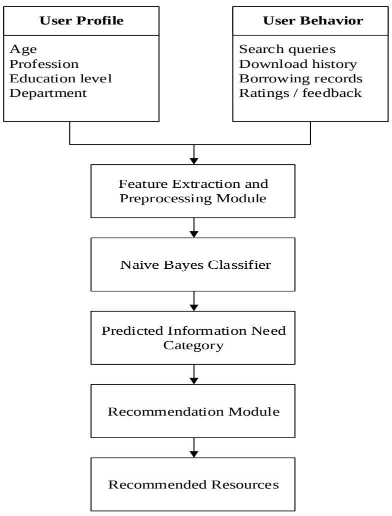

Fig. 1. General structure of the proposed Naive Bayes-based model for predicting users' information needs in library systems.

DOI: 10.5281/zenodo.20536026

4. METHODOLOGY

4.1 Data Collection and Dataset Description

In order to test the suggested model, a dataset comprising of 1,000 clients of the Libsmart library system was employed. The data used comprised demographic, academic and behavioral data including age, profession, education level, search history, download history, rating, and favorite subject areas. Individual user records were transformed into feature vectors, whereas the target category was the primary category of information needs.

To use as experimental subjects, the information-need categories were clustered into some large subject areas, such as Computer Science, Education, Economics, Literature, and Medicine. Training stage The Naive Bayes classifier computed prior probabilities of each category and conditional probabilities of the observed values of the features. In the prediction phase, the model used the posterior probabilities to place each user into a category with the most probability. Fig. 1 presents the overall design of the proposed model.

According to the experiment, it is possible to apply the suggested approach to user data of the Libsmart library system to assist in personalized recommendation and adaptive library services.

5. RESULTS AND DISCUSSION

The suggested model of Naive Bayes was tested on the basis of 1,000 users of the Libsmart library system. The data was split into 80/20 into training and testing sets. The label of the target of every user was determined based on the most frequent subject in the history of user interaction. The purpose of the experiment was to foretell the prevailing information-need category of each of the users based on the demographic, academic and behavioral characteristics. The results that have been obtained show that the proposed approach is applicable to personalized recommendation jobs of library systems. The distribution of predicted information-need categories is shown in Tab. 1.

Table 1. Distribution of predicted information-need categories

Category	Number of users	Percentage
Computer Science	320	32.0%
Education	210	21.0%
Medicine	180	18.0%
Economics	160	16.0%
Literature	130	13.0%
Total	1000	100%

Tab. 1. shows the distribution of users across the predicted information-need categories. The highest user proportion was allocated to the Computer Science category as illustrated then Education and Medicine. Such distribution shows that the chosen data set is mostly represented by users whose information interests are concentrated in the fields of technical, academic, and knowledge-based subjects. The prevalence of these categories can be attributed to the educational and research-based character of the Libsmart library environment since in this case the users have higher chances of searching the materials related to study, scientific activity and professional growth. Consequently, the findings in Table I offer the suggestion that the proposed model can define the key thematic trends of user interest and can be used to recommend the resources more generally and specifically. The distribution of predicted information-need categories among users is illustrated in Fig. 2.

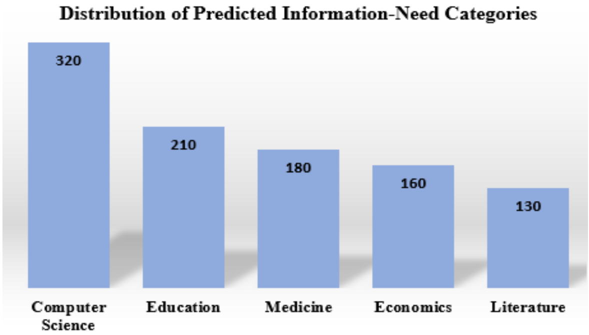

Fig. 2. Distribution of predicted information-need categories among users of the Libsmart library system.

DOI: 10.5281/zenodo.20536026

Fig. 2 illustrates the variation in users' interests across key subject categories, highlighting the uneven distribution of information needs. The number indicates that the estimated categories are not evenly spread, but focused in some prevailing subject areas. This tendency suggests that the given model is capable of determining the differences in the information preferences of the users according to the demographic and behavioral data that is available. Therefore, this classification is useful in enhancing the better guidance of the resources and it will help in generation of adaptive and user friendly library services. The performance of the proposed Naive Bayes model is presented in Tab. 2.

Table 2. Performance of the proposed naive Bayes model

Metric	Value
Accuracy	86.4%
Precision	84.9%
Recall	85.7%
F1-score	85.3%

In order to evaluate the efficiency of the proposed model even further, the standard classification performance measures were used. Tab. II presents the results of the experiment. The suggested Naive Bayes-based model demonstrated the accuracy of 86.4, which means that the general percentage of user dominant information-need category classification is high. Meanwhile, the precision (84.9%), recall (85.7%), and F1-score (85.3%) values were more or less similar to each other, which indicates a balanced and stable predictive performance. Such uniformity in the metrics used in the evaluation indicates that the classifier can make consistent predictions without having a significant drop in both sensitivity and correctness measures. Thus, the received findings prove that the suggested model can be successfully applied to the user-oriented recommendation and adaptive information support in the digital library system.

The model is simple and interpretable with low cost of computation making it one of the primary benefits of the proposed model. These qualities render it applicable in the digital library systems. Meanwhile, the model facilitates active information services through anticipating the interests of the users prior to them being explicitly defined as search queries. The model, however, is also relying on the quality and the completeness of the user data, and the assumption of independence of Naive Bayes could make the real relationships between features simpler. However, the results that have been received prove that the suggested approach can be an efficient foundation of individualized recommendation and adjustive library services in the Libsmart environment.

6. CONCLUSION

This paper introduced an intelligent model that uses Naive Bayes to identify the information requirements of the users in digital library systems. The given method applies demographic, academic, and behavioral traits of users to categorize them into the most likely information-interest groups and create personal suggestions. Results of experiment using the Libsmart digital library dataset indicate that the proposed model can be used with high accuracy in terms of classification and consistent performance based on various evaluation measures. The Naive Bayes classifier is very simple, interpretable, and computationally efficient, thus making it very applicable in a real life digital library setting. Future studies can aim at combining more behavioral predictors, testing the model with bigger data, and comparing its results with other machine learning methods, including decision trees and neural networks.

ACKNOWLEDGMENT

The author expresses sincere gratitude to the National Library of Uzbekistan for data access and institutional support. The author declares no competing financial interests. This research received no specific external funding.

REFERENCES

[1] P. Valdiviezo-Diaz, F. Ortega, E. Cobos, and R. Lara-Cabrera, "A collaborative filtering approach based on naïve Bayes classifier," IEEE Access, vol. 7, pp. 108581–108592, 2019, doi: 10.1109/ACCESS.2019.2931235.

[2] H. Li, H. Li, S. Zhang, Z. Zhong, and J. Cheng, "Intelligent learning system based on personalized recommendation technology," Neural Computing and Applications, vol. 31, pp. 4455–4462, 2019, doi: 10.1007/s00521-017-2993-2.

[3] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, 2013, doi: 10.1016/j.knosys.2013.03.012.

DOI: 10.5281/zenodo.20536026

[4] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, Springer, 2015, ISBN: 978-1489976376.

[5] T. M. Mitchell, Machine Learning, New York, NY, USA: McGraw-Hill, 1997, ISBN: 978-0070428072.

[6] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Advances in Artificial Intelligence, vol. 2009, pp. 1–19, 2009, doi: 10.1155/2009/421425.

[7] M. Özcan, F. Göz, and T. Temel, "New recommender system using Naive Bayes for e-learning," in Research Highlights in Education and Science, 2016, pp. 62–68.

[8] R. Rrmoku, B. Selimi, and L. Ahmedi, "Application of trust in recommender systems utilizing Naïve Bayes classifier," Computation, vol. 10, no. 1, p. 6, 2022, doi: 10.3390/computation10010006.

[9] F. Bekkamov, M. Babajanov, and M. Berdimurodov, "Using Bayesian methods to predict users' information needs in a digital library," in Proc. Int. Scientific and Practical Conf. Environment. Technology. Resources, vol. II, Latvia, 2025, pp. 37–41.

[10] F. Bekkamov and A. Rustamov, "Recommender systems: An overview," Bukhara State University Scientific Reports, no. 3(85), pp. 2–12, 2021.

[11] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008, ISBN: 978-0521865715.

[12] G. Chen, F. Zhu, and P. A. Heng, "Large-scale Bayesian probabilistic matrix factorization with memo-free distributed variational inference," ACM Trans. Knowl. Discov. Data, vol. 12, no. 3, pp. 1–24, 2018, doi: 10.1145/3183713.

[13] N. Wang, M.-H. Abel, J.-P. Barthes, and E. Negre, "Recommending competent users from semantic traces using a Bayes classifier," in Proc. IEEE Int. Conf. Systems, Man, and Cybernetics, 2015, pp. 1351–1356, doi: 10.1109/SMC.2015.239.

[14] A. Saleh, N. Dharshinni, D. Perangin-angin, F. Azmi, and M. I. Sarif, "Implementation of recommendation systems using the naïve Bayes classifier algorithm," Sinkron, vol. 8, no. 1, 2023, doi: 10.33395/sinkron.v8i1.

[15] S. Liu and F. Sen, "Design and implementation of movie recommendation system based on Naive Bayes," J. Phys.: Conf. Ser., vol. 1345, 2019, doi: 10.1088/1742-6596/1345/5/052023.

[16] S. P. R. Asaithambi, R. Venkatraman, and S. Venkatraman, "A thematic travel recommendation system using an augmented big data analytical model," Technologies, vol. 11, no. 2, 2023, doi: 10.3390/technologies11020042.

[17] M. A. Rakhmatullayev, Sh. B. Normatov, and F. A. Bekkamov, "Fuzzy relations based intelligent information retrieval for digital library users," in Proc. Int. Scientific and Practical Conf. Environment. Technology. Resources, vol. II, Latvia, 2023, pp. 80–83.

[18] R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, and M. Willemsen, "Bayesian statistics and modelling," Nature Reviews Methods Primers, vol. 1, no. 1, pp. 1–26, 2021, doi: 10.1038/s43586-020-00001-2.

DOI: 10.5281/zenodo.20536026