

# Source Recommendation System Using Context-based Classification: Empirical Study on Multi-level Ensemble Methods

Abdullah Al Kafi<sup>1</sup>, Sumit Kumar Banshal<sup>2,\*</sup>, Nishat Sultana<sup>3</sup>, Vedika Gupta<sup>4</sup>

<sup>1</sup>Department of Computer Science and Engineering, Daffodil International University, Dhaka, BANGLADESH.

<sup>2</sup>Department of Computer Science and Engineering, Alliance University, Bangalore, Karnataka, INDIA.

<sup>3</sup>Department of Computer Science, William and Mary, Williamsburg, VA 23185, USA.

<sup>4</sup>Jindal Global Business School, OP Jindal Global University, Sonapat, Haryana, INDIA.

## ABSTRACT

**Aim/Background:** This research aims to develop an automated contextual classifier for scholarly papers by utilizing established algorithms and understanding the information retention of different parts of a scholarly article, such as the Abstract, Article Title, and Keywords. It also seeks to recommend a contextual classifier-based recommender system to help academics identify credible sources. Scholarly articles from various study fields often use similar terms in their titles and keywords. However, finding a publication venue can be challenging for researchers at the beginning of a scientific inquiry. Thus, it is crucial to classify information based on its context, especially when abstracts, keywords, and titles receive equal attention. **Materials and Methods:** An ensemble model was developed and trained using 114K instances from 38 classes of the Web of Science (WoS) dataset and 40 classes of the Dimensions dataset. The ensemble approach incorporated both machine learning and deep learning algorithms to build a diverse classifier. The model was evaluated by testing it with an 80:20 train-test split to assess performance. The classifier was further integrated into a recommender system designed to suggest probable publication sources based on given article information. **Results:** The ensemble classification approach demonstrated superior performance with faster inference and efficient training time. The balanced training model, tested on 114K instances, effectively categorized scholarly articles into one of 40 categories. The recommender system was capable of recommending up to 10 probable publication sources based on the article's Title, Keywords, and Abstract. Models utilizing abstractions yielded the best results and provided a better understanding of the context in every iteration of the experiment. **Conclusion:** This study successfully developed an ensemble-based contextual classifier for academic papers, which can also function as a recommender system. The system aids researchers in choosing the most appropriate sources to publish by categorizing articles into 40 categories and suggesting credible publication venues. This approach simplifies the decision-making process for academics, enabling them to identify relevant publications and suitable sources for their work more efficiently.

**Keywords:** Scholarly Article, Recommendation System, Subject Classification, Contextual Classification, Text Analysis.

## Correspondence:

**Sumit Kumar Banshal**

Department of Computer Science and Engineering, Alliance University, Bangalore 562106, Karnataka, INDIA.  
Email: sumitbanshal06@gmail.com  
ORCID ID: 0000-0001-7795-1355

**Received:** 24-07-2023;

**Revised:** 13-10-2023;

**Accepted:** 13-03-2024.

## INTRODUCTION

The nature of academic publication has been completely transformed as a result of the progressive expansion and increased penetration of electronic formats. To be more explicit, the field of academic publication has been forced to undergo significant change due to the proliferation of new technologies. In the past few

decades, there has also been a flourishing of research that crosses disciplinary boundaries, which has outpaced transdisciplinary studies by a significant margin. When articles from a variety of Subject Categories (SCs) are mixed together in a search engine's database, it makes it significantly more difficult to locate relevant research papers as the number of academic publications increases. This is especially the case. It is conceivable that doing searches based just on keywords is no longer the method that produces the best results. Publications have the potential to unearth deeper insights across a wide variety of academic subjects if they are categorized and organized correctly. There are certain papers or publications that do not have the topic categories included in their metadata.<sup>[1]</sup> It's possible that organizing things into categories is the answer to this problem.



DOI: 10.5530/jscires.13.2.38

### Copyright Information :

Copyright Author (s) 2024 Distributed under Creative Commons CC-BY 4.0

**Publishing Partner :** EManuscript Tech. [www.emanuscript.in]

On the other hand, deciding where to publish your work may be a challenging and time-consuming endeavor. It is possible that a more insightful classification, depending on the article's subject matter, may be helpful in achieving this goal since the interdisciplinary study has developed over the last few decades. One factor that is employed more extensively is the keyword that has been adopted in several research works.<sup>[2-6]</sup> However, this is not adequate in the context of the current research trends. For a better understanding of applications that cut across disciplines, a content-based categorization is required.<sup>[1]</sup> On the other hand, the lack of multiclass categorization of academic publications using ensemble learning methods applies to both conventional and interdisciplinary articles. To satisfy in, it is recommended that a tool be developed to organize academic publications based on keywords, titles and abstracts. In addition to categorization, it also provides suggestions for publishing sources. In the first step of our experiment, we did not classify the classes of the datasets into categories before training machine learning models on them using one of two types of datasets containing 38 or 40 classes. After that, an Ensemble model is developed to combine all of the models into a single entity.

## LITERATURE REVIEW

Due to recent breakthroughs in academic research and publication, a vast number of research articles, papers and journals are now accessible. Classifying the items into their relevant categories or arranging them in the proper sequence might be difficult. Text classification commonly referred to as text tagging or text categorization, is the process of arranging text into distinct groups. It is often called text classification.<sup>[1]</sup> During this time period, new categorization and classification procedures for research publications were established.<sup>[1-4,6-8]</sup> Moreover, different studies suggest that context-based classification is more suitable than journal-based hierarchical taxonomy classification.<sup>[9-12]</sup> Other studies on scholarly articles about the literacy and importance of PID and Scopus datasets on sustainability are also present.<sup>[13-15]</sup> Gurubuz *et al.* applied traditional machine learning models such as naïve Bayes random forest and support vector machine on a scholarly dataset with 3 classes in both English and Turkish language and showed that SVM outperformed the other 2 models.<sup>[16]</sup> Several investigations, like the one by Daradkeh *et al.*, utilized Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and other common techniques.<sup>[3]</sup> CNN had the best results when compared to other search engines utilizing a dataset that includes Scopus, ProQuest and EBSCOhost. Using a bag of words on Random Forest, SVM and Decision tree, research was conducted on the dataset supplied by the library in order to classify the library books into the five provided categories. The Random Forest (RF) algorithm's 89% accuracy was the highest of all methods.<sup>[2]</sup> In separate research by Kandimalla *et al.*, scholarly publications

were classified using text mining methods, including tf-idf and unigram. In this study, 104 classes were categorized by abstract using a variety of techniques, including RF, Naive Bayes (NB), SVM, Logistic Regression (LR), Character-level Convolutional Network (CCNN) and deep averaging network. The findings demonstrated that CCNN outperformed the other approaches.<sup>[11]</sup> On three sub-domains of machine learning, bi-LSTM, Asymmetric Word Embedding and bi-LSTM were utilized to construct a binary classification strategy on 11778 articles from Arxiv using Hierarchical Attention Networks (HAN) that outperformed earlier algorithms.<sup>[17]</sup> Bi-LSTM and the knowledge graph were used on a dataset of 92,195 articles separated into 21 groups in order to estimate their degree of similarity. In addition, a description of how the combination of deep learning algorithms increases article categorization performance was presented.<sup>[4]</sup> Using BERT and ensemble learning algorithms resulted in a weighted average F1 score of 91%, according to another study on the categorization of scientific papers using abstracts.<sup>[8]</sup> In various other text classification domains, a large number of machine learning and deep learning approaches were also employed.<sup>[18-23]</sup> Various research has shown that the performance of various algorithms varies based on the kind of data. For instance, multinomial Naive Bayes outperforms Bernoulli in identifying news polarity;<sup>[24]</sup> for the BBC dataset, LR outperforms KNN, the least successful of the three was LR. Ensemble classification techniques have received a substantial amount of research on several datasets, including image and text datasets.<sup>[25-31]</sup> whilst RF outperforms RF.<sup>[32]</sup> LR fared the best for text-based classification when compared to other traditional ML methods and deep learning alternatives.<sup>[33]</sup> Aborisade and Anwar revealed that for Twitter data, LR outperforms NB,<sup>[34]</sup> however, in a different investigation, a hybrid model consisting of CNN and SVM performed better than the basic models.<sup>[22]</sup> On three independent datasets, Luo *et al.* evaluated NB, SVM and LR and found that SVM performed the best of the three algorithms.

It has been demonstrated that typical machine learning techniques, such as logistic regression, neural networks and support vector machines, perform exceptionally well in text mining applications. Moreover, techniques based on ensemble learning perform far better when used to text mining. However, the performance of ensemble learning approaches for the classification of scientific publications has not yet been evaluated. In addition, computations based on language, such as bigrams and trigrams seen in scholarly papers, might be a fascinating categorization technique for texts. On top of that, offering source recommendations based on the title, keyword and abstract might be a useful tool for aspiring scholars in any domain.

## OBJECTIVES

To bridge the gaps explored in previous studies, our work is focusing on the following objectives:

An ensemble approach of algorithms to classify scholarly articles contextually based on the content (Title, Keywords and Abstract).

A recommendation system to recommend the probable publication sources for a given article.

## DATA PREPROCESSING AND METHODOLOGY

The dataset used in this research was compiled using information from the Web of Science and Dimensions as both are very well-known databases. Dimension uses content-based sorting whereas WOS uses journal-based taxonomy and Dimension uses article-level classification. Here the two most crucial features were included in the dataset. 'Research Areas' from the WOS data and the 'Dimension Categories' both represent the label or category of the instances for respective dataset. There are 76 columns, over 9 million data instances and 178 different fields to explore in this collection. In these tests, we selected 7 columns and a sample size of 1500 for each category to ensure a well-balanced model. After these procedures, the WOS dataset and Dimension dataset were merged into one dataset depending on the DOI (Digital Object Identifier) to make a single dataset without duplicates to ensure the feasibility of the experimental design. There was a total of 1,14,000 occurrences or instances. For Dimension 40 categories and WOS 38 different categories were selected. In contrast, the 60/40 split was used for both training and testing. Figure 1 depicts the preprocessing steps that must be taken.

### Preprocessing

#### Selecting Fields

The raw Dataset have about 76 fields from which only 7 relevant fields were selected.

#### Cleaning and merging

The duplicates and null values from the initial dataset were cleaned.

#### Balancing

The initial dataset didn't have the same number of instances per category. The data was balanced in a way that each category comprises 1500 instances each.

#### Discretization

Some of the instances had overlapping categories. Which would change the problem into a multi-label classification problem.<sup>[35]</sup> Which isn't the aim of this study and would affect the prediction probability. So, the labels were separated and mutually inclusive instances were discarded.

#### Transformation

After the above steps the dataset was ready for transformation. Different Transformation strategies were used in different

sections of the experiments. Those will be discussed in detail in later sections.

The classes from each dataset are manually selected in a way that minimal variation is present each having similar data. An example of classes is given in the following Table 1.

### Proposed Ensemble Model

We created the ensemble configuration by selecting Naive Bayes, Logistic Regression and Support Vector Machine from the pool of traditional machine learning algorithms since these algorithms performed very well when it came to text classification<sup>[24,32-34]</sup> Techniques such as Convolutional Neural Networks, Artificial Neural Networks and Deep Learning were also used for the same purposes.<sup>[1,8,17]</sup> As a result, the primary objective is to design and implement an ensemble model that uses either majority voting or hard voting to choose the instance classes.<sup>[36-38]</sup> The tried-and-true Ensemble Model will provide a publication venue recommendation for each distinct collection of titles, keywords and abstracts that are provided. The workflow diagram is illustrated in Figure 2, which is a visual depiction of the diagram.

Here from Figures 2 and 3, it can be seen that our proposed model uses 2 levels of Ensemble methods. For Each level, it uses the maximum vote of the base class to predict Output 1, Output 2 and Output 3. Then another maximum voting method is used to determine the class. Maximum voting technique is a very common and widely used ensemble method.<sup>[39-41]</sup>

### Experimental Setup

The tests were conducted in two stages, with each stage consisting of six separate studies on three distinct types of data. These types of data were the "Article Title," the "Abstract," and the "Author Keyword." Throughout each step, experiments are carried out on a wide variety of input types, with testing and training taking up around 80% and 20% of each phase, respectively. During the first phase, we will be concentrating on the WOS categories and during the second phase, we will be concentrating on the dimension categories. Techniques that are considered to be more conventional for Machine Learning (ML) and Deep Learning (DL) are applied throughout the entirety of the procedure. The tf-idf vectorizer was applied to 5000 features in standard techniques for machine learning and the range of n-grams was chosen between one and three. Deep learning made use of word embedding which

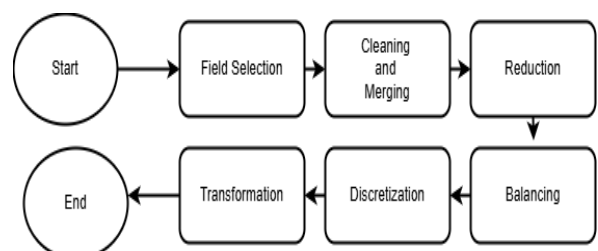


Figure 1: Preprocessing steps for Preparing the Dataset.

consisted of 64 dimensions and 15,000 often-used words. Table 1 outlines the two distinct measurements that can be used for the transformation. After collecting information on the performance of the model, the models themselves are reconstructed so that they may once again serve as the fundamental models for the Ensemble model.

Based on the designed setup mentioned above, the model was trained and tested. To understand the performance of the setup, apart from overall accuracy, other evaluation matrices were taken into consideration. Precision, Recall and F1 scores were measured over all the classes for all three features explored. The base equation of evaluation matrices has been depicted in equations no 1, 2 and 3. Further, these evaluation matrices were clubbed using the macro average concept for each of the features in the ensemble method using equation 4.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}} \quad (2)$$

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Macro Average} = \frac{\sum \text{Evaluation metrics per class}}{\text{Number of class}} \quad (4)$$

## RESULTS AND DISCUSSION

Following the previous discussion, the current investigation into a classification and recommendation system chooses to use NB, LR, SVM, CNN and ANN. Experiments make use of the data that can be found in scientific journals. This data might include titles, authors and abstracts. In order to construct the underlying models, the datasets went through a total of ten distinct experiments. Three further tests were carried out to determine whether or not the aggregated ensemble model was effective. Sets of experiments were designed with the specific intention of

making their interpretation easier to understand. The first round of studies investigated the efficiency of the fundamental model when applied to Dimension data, while the second group of trials investigated the usefulness of the model when applied to WOS data. During the last round of tests, we studied how the performance of the Ensemble model changed depending on the type of data we requested to evaluate. The prefixes exp1, exp2 and exp3 shall be used from this point on whenever there is a reference to any of the individual experiment sets. Since the experimental circumstances and settings were kept the same throughout all the trials, the only variable that was compared was the accuracy of the tests administered.

The accuracy of the exp1 method after it was applied to the Dimension dataset is shown below in terms of each of the three input categories, which are "Abstract," "Keyword," and "Article Title," respectively (exp 1). The correlation between the accuracy of their replies and their scores is seen in Figure 4, which may be found here (in percentage). Our contention that the categorization of articles needs to be carried out on the basis of the content rather than the keywords is given credence by the graphical representation that is provided here. Testing that is more accurate makes use of the abstract, which has a bigger amount of information than the keywords. In addition, the findings reveal that LR achieves the maximum degree of accuracy for the abstract as well as the other parameters among the basic model settings. This is indicated by the fact that LR earns the highest possible score. The second-place finisher is the SVM. These results lend credence to the conclusions drawn by earlier research on the prevalence of text categorization algorithms in academic writing and other comparable contexts.<sup>[25,26,42]</sup>

The same premise was put to the test in Experiment 2 (exp2), which was a continuation of Experiment 1 and employed a different dataset known as the WOS Dataset. These findings are presented in the form of a graphical depiction in Figure 5, which also includes the associated degrees of accuracy for each

**Table 1: Example of Selected Classes.**

WOS	Dimension
Computer Science	Computer Science
Business and Economy	
Optics	Optical Physics

**Table 2: Data transformation metrics.**

Parameters	Machine Learning Algorithms	Parameters	Deep Learning Algorithms
Vectorizer	TF-IDF	Tokenizer	TensorFlow Provided
Features	5000	Embedding Dimension	64
Stop words	English	Vocabulary Size	15000
n-gram Range	1-3	Max Length	Abstract: 250 words Other: 30 words

category (in percentage). This graphic illustrates how well LR and SVM perform in comparison when it comes to content-based classification (Abstract). Even while it might look like the overall accuracy is lower than it was in the last experiment, the patterns that lie beneath have not changed. The problem stems from the fact that the WOS base classification approach relies on public sources as its primary data source. As a result, the fact that the model's training was focused on this particular direction should not come as a surprise. It is important to point out that deep

learning techniques like ANN and CNN were demonstrated to be almost as accurate as LR and SVM. The accuracy patterns seen in the WOS dataset are also consistent with the hypothesis that probabilistic methods are being used in the classification of text data a growing amount more frequently.<sup>[32,33,43]</sup> In addition to this, it provides credibility to the argument in terms of the classification of scientific papers according to the subject matter that they cover.

The third and final examination examines the performance of the proposed ensemble model on both the WOS dataset as well as the Dimension dataset. The degrees of accuracy are shown on a percentage scale in the figure, which is numbered 6. When applying the ensemble method, classification based on the abstract performs much better than classification based on the other two fields for both datasets. The results of this experiment unequivocally demonstrate that content-based categorization is superior to keyword-based classification. The ensemble approach is contrasted with the baseline model in Table 2 so that it helps to distinguish the understanding of the effectiveness of the ensemble method. When the efficacy of each different kind of input is analyzed independently, it becomes abundantly evident that the ensemble model is superior to the basic model. As can be seen in Table 2, there is a considerable gap between the accuracy of the ensemble models and that of the basis models taken as a whole in their collective representation. This table takes into consideration all of the datasets. When comparing results with and without the use of abstract and author keywords in the Dimension and WOS datasets, respectively, the ensemble method achieves an accuracy improvement of over 6% and 4%, respectively (for average cases). Although there has been some improvement made to the accuracy of article titles, it has not been made to the same degree as the other two categories of input. In both instances, there was an increase that was greater than four and 4% respectively. The most accurate results obtained by using all of the basic models together are just 3% more accurate than the most accurate results obtained by using the ensemble models alone. This demonstrates the need to utilize a content-based methodology, in addition to the ensemble technique, when it comes to the categorization of scientific articles.

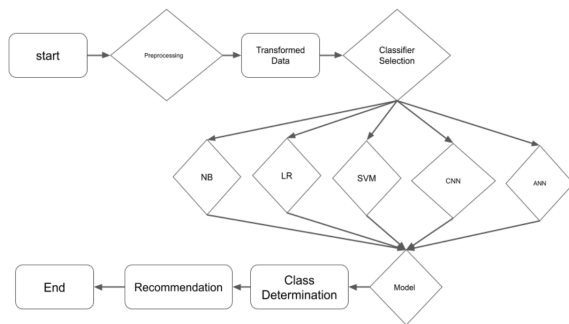


Figure 2: EWorkflow diagram for the Recommender System (2 levels).

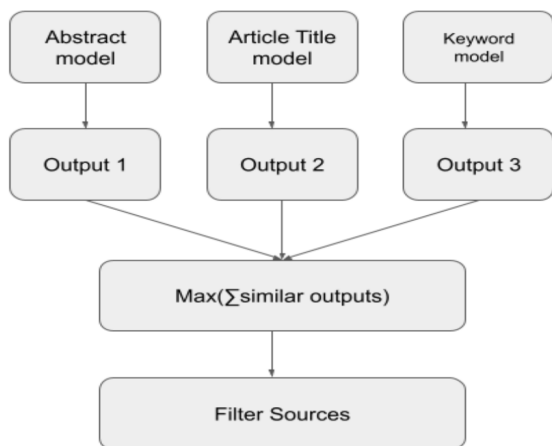


Figure 3: Ensemble Workflow Diagram.

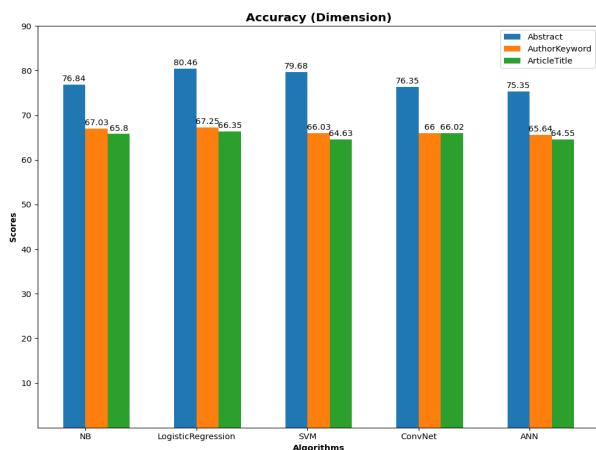


Figure 4: Base model accuracy on Dimension Dataset.

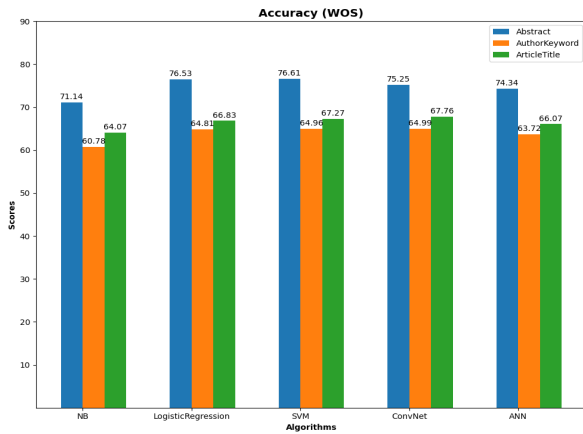


Figure 5: Accuracy of the base model on the WOS dataset.

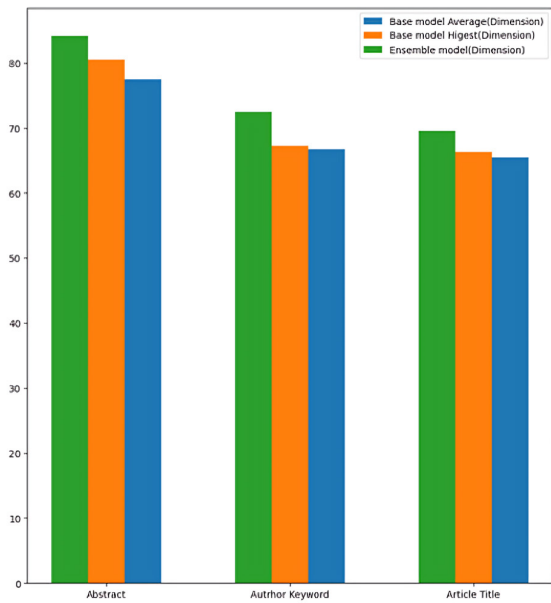


Figure 6: Accuracy of the base model vs Ensemble model (Dimension).

while the Recall value is above 67%. These matrices lend support to the case for a context-aware ensemble strategy for scholarly article categorization. Further, this macro average concept has been incorporated with all the input settings altogether. The values for these three evaluation matrices have been plotted in Figure 8. The overall performance of these matrices was also noted significantly to draw conclusions of relevance to have a contextual classification for scholarly articles (Table 4).

Here, in both datasets, the accuracy achieved is significantly higher than in base-level algorithms.

Here is a close look at the Table 5 accuracy report for the base model shows that not all models perform the same in each class. Some of the models are performing better than others. For example, while classifying Optics Class Naïve Bays perform poorly whereas Random Forest SVM, ANN and CNN perform well. A similar case happens in the Optical Physics Class.

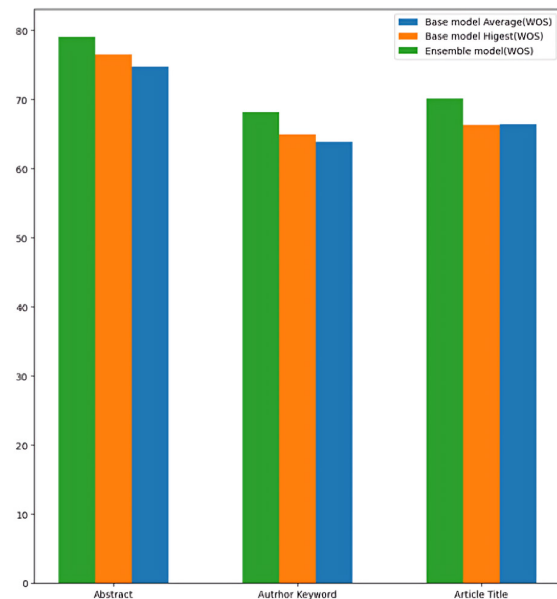


Figure 7: Accuracy of the base model vs Ensemble model (WOS).

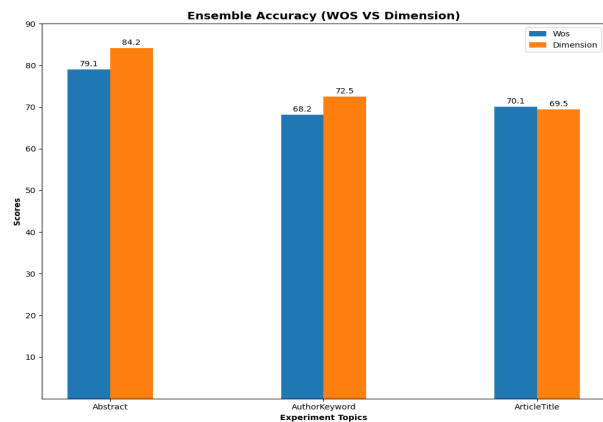


Figure 8: Accuracy of the Ensemble model on WOS and Dimension Dataset.

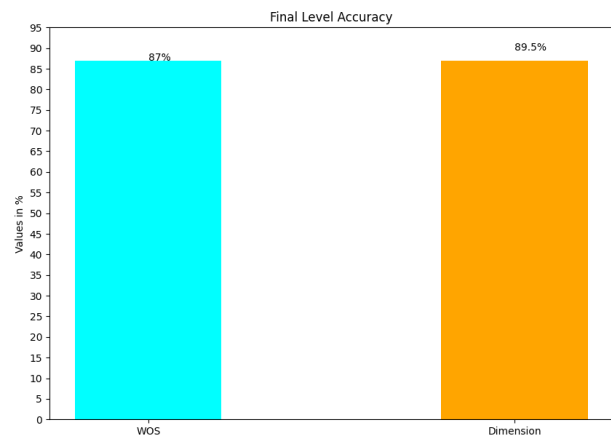


Figure 9: 2nd or Final Level Accuracy.

Interface Descriptions	Interface Functions
Enter Your Title	This option takes the title of the user’s article
Enter Keyword	This option takes the keywords selected by the users.
Enter your Abstract	This option takes the Abstract of the user’s paper.
Radio Button	This button allows the user to choose the type of model he or she wants to use. Namely WOS or Dimension.
Sidebar	The sidebar shows the available classes for the selected model type.
Compute Button	It runs the algorithm and shows the prediction table and recommendation list.
Prediction Table	The prediction table shows the predicted class and the confidence of the algorithm. Confidence values can be used to determine the correlation.
Common Publishing Sources	This list recommends the user similar publishing sources for the predicted category.

Figure 10: Interface Introduction.

**Supported Topic**

1. Agriculture
2. Astronomy & Astrophysic
3. Biotechnology Biotechnology and applied Microbiology
4. Business and Economy
5. Cardiovascular system and cardiology
6. Chemistry.
7. Computer Science.
8. Dentistry.
9. Dermatology.
10. Education and Educational Research.
11. Engineering.
12. Environmental science and ecology.
13. Fisheries.

## Welcome To The App

Enter Your Title

Text classification using capsules

Enter Your Keywords

Deep learning, text classification, capsule network, machine learning, text mining.

Enter Your Abstract

We utilized seven benchmark data sets to demonstrate that capsule network, along with the proposed routing method, provide comparable results.

	Article title	Author Keywords	Abstract
Prediction	Computer Science	Computer Science	Computer Science
Confidence	100%	100%	100%

**Recommended Publishing Source**

- Image and Vision Competing
- Acm Transaction on mathematical software
- IEEE Transactions on Effective Computing
- Graphical models
- Neural Networks
- Applied soft computing.
- IEEE transaction on effective computing.
- IEEE Transaction on Network and Service Management
- IEEE Transaction on computational intelligence and AI in games

Figure 11: The Output Interface.

Table 3: Macro Average of Ensemble Method Input Wise.

Precision		Precision Recall		F1 score	
WoS	Dimension	WoS	Dimension	WoS	Dimension
0.75	0.77	0.67	0.70	0.73	0.76
0.71	0.67	0.64	0.60	0.68	0.57
0.66	0.63	0.60	0.61	0.60	0.58

**Table 4: 2<sup>nd</sup> or Final Level Classification.**

Input type	Precision	Recall	F1 score
Dimension	0.88	0.87	0.83
WOS	0.84	0.82	0.79

**Table 5: Accuracy Report for 2 classes (Abstract).**

Model	Optics	Optical Physics
Random Forest	83%	72%
Naïve Bayes	28%	68%
CNN	75%	61%
SVM	93%	63%
ANN	66%	59%

As our Ensemble model takes the max vote count of 5 models it removes the shortcoming of classifying some of the classes. For the same classes, our Ensemble model accuracy is 92% which is better than all 5 of the base models. Because it balances the low-accuracy models.

## DISCUSSION

One of the main purposes of this research was to find the information retention of different important parts of a research paper. For this purpose, we studied the Abstract, Article title and Keywords in different experiments. As abstract has the most words among the three it retains more information and shows a better result while classifying tasks.

On the other hand, keywords and article titles are often written in different words and keywords are more precisely selected for an article but show an approximately similar level of information retention in most cases. Another interesting finding was most base-level algorithms performed better on Article titles than Keywords though keywords are selected more precisely. The same trend can be seen in Figure 8 where the ensemble model on the Article title outperformed the Author keyword on the WOS dataset.

For 2nd level ensemble model, we could have neglected any of the similar performing stacks of ensemble models such as article title or keywords models. But our findings show a different picture shown in Figure 9. In some cases, the article title fails to classify correctly but keywords show the correct class which improves the maximum likelihood of having a better accuracy and vice versa.

## The Web Interface

The model that was provided is then converted into a recommender system and made available online through a user interface. This is done to make the concept more applicable in real-world settings (Figure 10). The user interface of the website was designed to be as user-friendly and straightforward as was humanly possible so that any visitor could easily make use of all

of the website's features. The user can obtain recommendations for journals that have published articles in the same category by utilizing the web interface to compare the title, author keyword and abstract of their work with the proposed category. This comparison will result in the user receiving recommendations for journals that have published articles in the same category. This recommender system that runs on the web utilizes the established ensemble technique as its primary method of data processing. The computation of the suggestion is carried out using the inputs provided by any user and the result is presented in the format depicted in Figure 11. Figure 11 provides a visual representation of the user input panel's eight component sections, each of which has been extensively addressed thus far. Any user who has access to a scientific publication can complete the required fields, which include the Title, Keywords and Abstract of the article. Figure 11 illustrates how the results of the computation will be utilized to update the process of picking the best model and will also demonstrate how the user will be asked to complete the computation. This online recommender tool will be helpful to ambitious researchers who are faced with a plethora of publication channels from which to pick.

## CONCLUSION

This investigation is primarily focused on achieving two primary goals: first, the development of a content-based classification scheme for academic papers; and second, the development of a recommender system that can guide the way toward publishing articles. Both of these aims are being investigated as part of this particular area of investigation. Classifiers are developed by combining two distinct types of datasets with two distinct categories of algorithms. The procedure in question is referred to as "combining/merging." In addition to this, the application can provide recommendations for publishing resources that are arranged in a manner that is consistent with the categorization. This approach is not only quicker to teach than other ways, but it is also quicker to draw findings from and it is successful. These are two of the major benefits that may be gained from utilizing this method as opposed to utilizing any of the other available options. The total performance of our ensemble model was anywhere from three to six percentage points better than the performance of the base models and this improvement depended on the conditions. It suggests a total of ten different publishing sources that might be employed for each individual subject area. The preliminary ensemble technique that we have discussed has a total of 38 and 40 classes and has the potential to be expanded in either the immediate or more distant future. In the future, the performance of Machine Learning models such as NB, LR and SVM can be increased by adding more instances to each class, which will certainly require more hardware resources. On the other hand, experimenting with different n-gram sizes and feature sizes is yet to be discovered. In addition to that, transformer-based classification models are not implemented in



this study. Applying a transformers-based approach can direct to a different scenario. Moreover, the model best performed on the Dimension dataset and the model is 89.5% accurate where the precision, recall and F1 score ranges from 80%-90% which is realistic but can misclassify some documents which can create confusion among the users of the application. In the future, more deep architecture can be studied as this study aims to create base model standards using shallow models. During the course of our inquiry, we have taken into consideration a dataset that is not only incredibly well-balanced but also has a sizeable number of records that have a combined total that is greater than 100,000. In light of what has been demonstrated in this body of work, it is possible that the scope of this research might be expanded by linking the multi-class system with a probabilistic value in a multi-label classification problem.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## DECLARATION

The first draft was initially uploaded on the Orvium platform.

## REFERENCES

- Kandimalla B, Rohatgi S, Wu J, Giles CL. Large Scale Subject Category Classification of Scholarly Papers With Deep Attentive Neural Networks. *Front Res Metr Anal*. 2021;5:31.
- Caragea C, Wu J, Das Gollapalli S, Giles CL. Document Type Classification in Online Digital Libraries. 2016 [cited 2022 Dec 21]; Available from: <http://pdfbox.apache.org/>
- Daradkeh M, Abualigah L, Atalla S, Mansoor W. Scientometric analysis and classification of research using convolutional neural networks: A case study in data science and analytics. *Electronics*. 2022;11(13):2066.
- Hoppe F, Dessi D, Sack H. Deep Learning meets Knowledge Graphs for Scholarly Data Classification. In: Companion proceedings of the world wide web conference (WWW), 2021; 417-421. DOI: 10.1145/3442442.3451361
- Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaglu M, Gao J. Deep Learning-based Text Classification. *ACM Computing Surveys (CSUR)*. 2021;54(3). DOI: 10.1145/3439726
- Salatino A, Osborne F, Motta E. CSO Classifier 3.0: a scalable unsupervised method for classifying documents in terms of research topics. *International Journal on Digital Libraries*. 2022[cited 2022 Dec 21];23(1): 91-110. Available from: <https://link.springer.com/article/10.1007/s00799-021-00305-y>
- Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaglu M, Gao J. Deep Learning-based Text Classification. *ACM Computing Surveys (CSUR)*. 2021;54(3):1-40. DOI: 10.1145/3439726
- Piao G. Scholarly Text Classification with Sentence BERT and Entity Embeddings. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2021 [cited 2022 Dec 21];12705 LNAI:79-87. DOI: 10.1007/978-3-030-75015-2\_8
- Narin F, Carpenter M, Berlnt NC. Interrelationships of scientific journals. *Journal of the American Society for Information Science*. 1972[cited 2023 Oct 10];23(5): 323-31. DOI: 10.1002/asi.4630230508
- Carpenter MP, Narin F. Clustering of scientific journals. *Journal of the American Society for Information Science*. 1973[cited 2023 Oct 10];24(6): 425-36. DOI: 10.1002/asi.4630240604
- Small H, Griffith BC. The Structure of Scientific Literatures I: Identifying and Graphing Specialties. *Soc Stud Sci*. 1974[cited 2023 Oct 10];4(1): 17-40. DOI: 10.1177/030631277400400102
- Milojević S. Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative Science Studies*. 2020[cited 2023 Oct 10];1(1): 183-206. DOI: 10.1162/qss\_a\_00014
- Al-Raei M. Scopus-Based Study of Sustainability in the Syrian Higher Education Focusing on the Largest University. *Open Information Science*. 2023[cited 2023 Oct 12];7(1). DOI: 10.1515/opsis-2022-0149
- MacGregor G, Lanchon-Barrantes BS, Pennington DR. Measuring the Concept of PID Literacy: User Perceptions and Understanding of PIDs in Support of Open Scholarly Infrastructure. *Open Information Science*. 2023[cited 2023 Oct 12];7(1). DOI: 10.1515/opsis-2022-0142
- Kumar R. Students' Perspectives on the Application of Internet of Things for Redesigning Library Services at Kurukshetra University. *Open Information Science*. 2023[cited 2023 Oct 12];7(1): 201-15. DOI: 10.1515/opsis-2022-0154
- Gürbüz T, Uluyol Ç. Research article classification with text mining method. *Concurr Comput*. 2023[cited 2023 Oct 10];35(1): e7437. DOI: 10.1002/cpe.7437
- Maillette De Buy Wenniger G, Van Dongen T, Aedmaa E, Kruitbosch HT, Valentijn EA, Schomaker L. Structure-Tags Improve Text Classification for Scholarly Document Quality Prediction. [cited 2022 Dec 21]; DOI: 10.18653/v1/P17
- Dai W, Xue GR, Yang Q, Yu Y. Transferring Naive Bayes Classifiers for Text Classification. [cited 2022 Nov 6]; Available from: [www.aaii.org](http://www.aaii.org)
- Frank E, Bouckaert RR. Naive bayes for text classification with unbalanced classes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2006 [cited 2022 Nov 6];4213 LNAI:503-10. Available from: [https://link.springer.com/chapter/10.1007/11871637\\_49](https://link.springer.com/chapter/10.1007/11871637_49)
- McCallum A, Nigam K. A comparison of event models for naive bayes text classification. 2017. AAAI Conference on Artificial Intelligence. Available from: [https://https://www.researchgate.net/publication/2408220\\_A\\_Comparison\\_of\\_Event\\_Models\\_for\\_Naive\\_Bayes\\_Text\\_Classification](https://https://www.researchgate.net/publication/2408220_A_Comparison_of_Event_Models_for_Naive_Bayes_Text_Classification)
- Chen Y, Zhang Z. Research on text sentiment analysis based on CNNs and SVM. *Proceedings of the 13<sup>th</sup> IEEE Conference on Industrial Electronics and Applications, ICIEA 2018*. 2018;2731-4.
- Zhang W, Gao F. An Improvement to Naive Bayes for Text Classification. *Procedia Eng*. 2011;15:2160-4.
- Singh G, Kumar B, Gaur L, Tyagi A. Comparison between Multinomial and Bernoulli Naive Bayes for Text Classification. 2019 International Conference on Automation, Computational and Technology Management, ICACM 2019. 2019;593-6.
- Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. *Frontiers of Computer Science* 2019 14: 2. 2019[cited 2022 Nov 1];14(2): 241-58. Available from: <https://link.springer.com/article/10.1007/s11704-019-8208-z>
- Liang D, Yi B. Two-stage three-way enhanced technique for ensemble learning in inclusive policy text classification. *Inf Sci (N Y)*. 2021;547:271-88.
- Onan A. Hybrid supervised clustering-based ensemble scheme for text classification. *Kybernetes*. 2017;46(2):330-48.
- Kılınc D. The effect of ensemble learning models on Turkish text classification. *Celal Bayar University Journal of Science*. 2016;12(2):215-20.
- Shi L, Ma X, Xi L, Duan Q, Zhao J. Rough set and ensemble learning based semi-supervised algorithm for text classification. *Expert Syst Appl*. 2011;38(5):6300-6.
- Wang G, Sun J, Ma J, Xu K, systems JGD support, 2014 undefined. *Sentiment classification: The contribution of ensemble learning*. Elsevier. [cited 2022 Nov 10]; Available from: <https://www.sciencedirect.com/science/article/pii/S0167923613001978>
- Zeng K, Pan Z, Xu Y, Informatics YQJM, 2020 undefined. An ensemble learning strategy for eligibility criteria text classification for clinical trial recruitment: algorithm development and validation. *medinform.jmir.org*. [cited 2022 Nov 10]; Available from: <https://medinform.jmir.org/2020/7/e17832>
- Shah K, Patel H, Sanghvi D, Shah M. A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. *Augmented Human Research*. 2020;5(1):1-16. Available from: <https://link.springer.com/article/10.1007/s41133-020-00032-0>
- Kamath CN, Bukhari SS, Dengel A. Comparative study between traditional machine learning and deep learning approaches for text classification. In: *Proceedings of the ACM Symposium on Document Engineering*. 2018; 1-11. DOI: 10.1145/3209280.3209526
- Aborisade OM, Anwar M. Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers. *Proceedings-2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018*. 2018;269-76.
- Mustafa G, Usman M, Yu L, afzal MT, Sulaiman M, Shahid A. Multi-label classification of research articles using Word2Vec and identification of similarity threshold. *Scientific Reports*. 2021;11(1):1-20. Available from: <https://www.nature.com/articles/s41598-021-01460-7>
- Bin Habib AZS, Tasnim T. An Ensemble Hard Voting Model for Cardiovascular Disease Prediction. 2020 2<sup>nd</sup> International Conference on Sustainable Technologies for Industry 40, STI 2020.
- Delgado R. A semi-hard voting combiner scheme to ensemble multi-class probabilistic classifiers. *Applied Intelligence*. 2022 [cited 2022 Nov 6];52(4): 3653-77. Available from: <https://link.springer.com/article/10.1007/s10489-021-02447-7>
- Mishra S, Mallick PK, Tripathy HK, Jena L, Chae GS. Stacked KNN with hard voting predictive approach to assist hiring process in IT organizations. *International Journal of Electrical Engineering & Education*. 2021. DOI:10.1177/0020720921989015.
- Meel P, Vishwakarma DK. HAN, image captioning and forensics ensemble multimodal fake news detection. *Inf Sci (N Y)*. 2021;567:23-41.
- Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. *Computer Methods and Programs in Biomedicine Update*. 2021;1:100032.

40. Koushik C, Bhattacharjee R, Hemalatha CS. Symptoms based Early Clinical Diagnosis of COVID-19 Cases using Hybrid and Ensemble Machine Learning Techniques. 2021 5th International Conference on Computer, Communication and Signal Processing, ICCSP 2021. 2021;59-64.
41. Luo X. Efficient English text classification using selected Machine Learning Techniques. Alexandria Engineering Journal. 2021;60(3):3401-9.
42. Gonçalves CA, Vieira AS, Gonçalves CT, Camacho R, Iglesias EL, Diz LB. A Novel Multi-View Ensemble Learning Architecture to Improve the Structured Text Classification. Information. 2022;13(6):283. DOI: 10.3390/info13060283.

**Cite this article:** Kafi AA, Banshal SK, Sultana N, Gupta V. Source Recommendation System Using Context-based Classification: Empirical Study on Multi-level Ensemble Methods. J Scientometric Res. 2024;13(2):475-84.