INDO-ASIAN JOURNAL OF INFORMATION RESEARCH AND
TECHNOLOGY VOL 2 (01), JANUARY – JUNE 2025, PP. 31-49

THE SIGNIFICANCE OF PROMPT DESIGN FOR IMPROVING THE PERFORMANCE AND ACCURACY OF MACHINE LEARNING LANGUAGES

Achint Kaur Sodhi

M.Sc. Maths , Lovely Professional University

Email Id : aksodhi2828@gmail.com

ABSTRACT

A bibliometric analysis of prompting in machine learning languages' performance is conducted. The study uses bibliometric approaches to analyse scientific publications to investigate the major contributions, patterns, and research fields in this discipline. To map and illustrate the network of authors, documents, affiliations, sources, keywords, and country-specific contributions, two potent bibliometric tools—Biblioshiny and VOSviewer—were used to extract data for the study from the Scopus database. With an emphasis on the development of machine learning language performance, the study examines the geographical distribution of research, author partnerships, citation patterns, and keyword co-occurrences. The application of rapid design to improve machine learning model performance reveals notable trends, emerging fields of research, and regional variations, according to the results. The country-specific study provides a worldwide perspective on the discipline by highlighting important countries that are spearheading research in this area. In addition to offering recommendations for future study topics in this expanding subject of machine learning, this paper offers an overview of the current state of the field by highlighting significant articles, authors, and collaborative networks.

Keywords: *Machine Learning Models, Prompting, Performance Optimization, Keyword Co-occurrence, Biblioshiny, VOS Viewer*

1. INTRODUCTION

Empirical research findings can be transferred from people's brains and labs to the network and tools that assist in organizing and filtering the data, greatly accelerating machine learning and data mining research. Although there are many more applications for the massive streams of experiments being carried out to test theories, benchmark new algorithms, or model new datasets, they are frequently discarded or their specifics are lost over time. Furthermore, it is extremely difficult to demonstrate theoretically that a new algorithm is superior to others, even on a subset of problems, because learning algorithms are heuristic in nature. Although there is some theory linking certain heuristics to locating the hidden concept c, this relationship is frequently poorly understood. Furthermore, the provided data and its hidden concept are even less understood unless they are artificially generated; otherwise, there would be no need for modelling in the first place. Moreover, it is exceedingly challenging to logically establish that

a new algorithm outperforms others, even on a limited collection of situations, due to the heuristic nature of learning algorithms. Despite theoretical connections between specific heuristics and the identification of the concealed idea c, this relationship is often inadequately comprehended. Moreover, the supplied data and its underlying concept are even less comprehended until artificially manufactured; otherwise, modelling would be unnecessary from the outset.

Empirical research findings can be transferred from people's brains and labs to the network and tools that assist in organizing and filtering the data, greatly accelerating machine learning and data mining research. Although there are many more applications for the massive streams of experiments being carried out to test theories, benchmark new algorithms, or model new datasets, they are frequently discarded, or their specifics are lost over time. Thus, the objective of the study is as follows:

a. To perform a literature profile study to determine the distribution of journals, publications, authors, affiliations, and keywords.
b. To do literature profiling to understand country-specific analysis.
c. To examine patterns of keyword co-occurrence in the literature.

The study seeks to systematically delineate the existing knowledge in this swiftly evolving domain by offering an overview and formal definition of prompting strategies (Section 2). This is succeeded by a comprehensive examination of prompting techniques, ranging from fundamental aspects like prompt template engineering (Section 3) and prompt answer engineering (Section 4) to more sophisticated notions such as multiport learning methods (Section 5) and prompt-aware training methods (Section 6). Subsequently, we categorise the diverse applications of prompt-based learning methods and examine their interaction with the selection of prompting techniques (Section 7). Ultimately, we endeavour to contextualise the present status of prompting methodologies within the research landscape, establishing links to different academic disciplines (Section 8) and proposing many contemporary challenges that may be suitable for further investigation (Section 9).

2. LITERATURE REVIEW

Due to the perpetual inadequacy of datasets for developing high-quality models, early NLP models predominantly depended on feature engineering (Table 1(a); e.g., Guyon et al. , Lafferty et al. , Och et al. , Zhang and Nivre). In this process, NLP researchers or engineers utilised their domain expertise to identify and extract significant features from raw data, thereby supplying models with the necessary inductive bias to learn from this constrained data. The emergence of neural network models for natural language processing enabled the simultaneous learning of salient features during model training, thereby redirecting attention to architecture engineering, where inductive bias was introduced through the design of an appropriate network architecture that facilitates the learning of such features.

From 2017 to 2019, there was a significant transformation in the learning of NLP models, resulting in a diminishing importance for the fully supervised paradigm. The standard transitioned to the pre-training and fine-tuning paradigm. In this framework, a model with a predetermined architecture is pre-trained as a language model (LM), estimating the likelihood of observed textual data.

The abundant availability of raw textual data enables the training of language models (LMs) on extensive datasets, allowing them to acquire robust general-purpose linguistic properties. The aforementioned pre-trained language model will thereafter be tailored for various downstream tasks by incorporating supplementary parameters and refining them using task-specific goal functions.

In this framework, the emphasis shifted primarily to goal engineering, formulating the training objectives employed during both the pre-training and fine-tuning phases. Zhang et al. (2024) demonstrate that incorporating a loss function for predicting salient phrases from a document enhances the efficacy of a pre-trained language model for text summarisation. The primary component of the pre-trained language model is typically (though not universally; Peters et al. (2024)) fine-tuned to enhance its efficacy for addressing the downstream job. In 2021, we are experiencing a significant transformation, wherein the "pretrain, fine-tune" methodology is supplanted by a new approach termed "pre-train, prompt, and predict." In this framework, rather than modifying pre-trained language models for downstream tasks through goal engineering, downstream problems are restructured to resemble those addressed during the initial language model training, utilising a textual prompt. For instance, when interpreting the sentiment of a social media post stating, “I missed the bus today,” we can proceed with the cue “I felt so ” and request the language model to complete the sentence with an emotion-laden term. Alternatively, if we select the question "English: I missed the bus today." French: "), thereafter, a language model may be capable of providing a French translation to complete the sentence.

By selecting suitable prompts, we can influence the model's behaviour, enabling the pre-trained language model to predict the desired output, occasionally without any supplementary task-specific training This method's value lies in the ability to utilise a single language model, trained totally in an unsupervised manner, to address numerous tasks when provided with a suitable set of prompts. Nevertheless, like many appealing concepts, there is a drawback this approach necessitates prompt engineering to identify the most suitable prompt for enabling a language model to address the task effectively.

Supervised Learning Framework

In a conventional supervised learning framework for natural language processing, we input x, often text, and forecast an output y utilising a model P (y|x; θ). y may represent a label, text, or other forms of output. To ascertain the parameters θ of this model, we utilise a dataset including pairs of inputs and outputs and train a model to predict this conditional probability.

Prompting Basics

The primary challenge of supervised learning is the requirement for annotated data to train a model P (y|x; θ), which is often scarce for numerous jobs. Prompt-based learning methodologies for NLP seek to address this challenge by developing a language model that estimates the probability P(x; θ) of the text x itself and employing this probability to forecast y, thereby diminishing or eliminating the necessity for extensive labelled datasets. This section presents a mathematical characterisation of the most basic kind of prompting, which includes numerous studies on the subject and can be extended to incorporate further variations. Specifically, simple prompting forecasts the highest-scoring ˆy in three stages.

Further, we can also designate the initial type of prompt, which includes a gap to be filled within the text, as a cloze prompt, and the second type of prompt, where the input text precedes totally, as a prefix prompt. (2) Often, these template words are not comprised of real language tokens; they may consist of virtual words (e.g., represented by numeric identifiers) that will subsequently be embedded in a continuous space, and certain prompting techniques even produce continuous vectors directly.

Supplement to the prompt. In this phase, a prompting function fprompt(·) is utilised to transform the input text x into a prompt x' = fprompt(x).

In most prior research (ref), this function entails a two-step procedure:

Utilise a template, which is a textual string including two placeholders: an input slot [X] for input x and an answer slot [Z] for an intermediate generated response text z that will subsequently be converted into y.

Insert the input text x into slot [X].

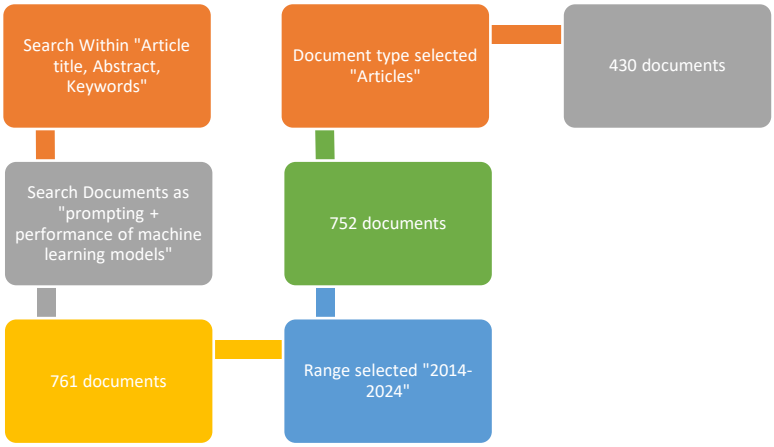
In sentiment analysis, where x = "I love this movie," the template may be structured as "[X] Overall, it was a [Z] movie." Consequently, x ′ would transform into "I adore this film." In summary, it was a [Z] film, as per the aforementioned example. In the context of machine translation, the template may be structured as "Finnish: [X] English: [Z]," wherein the input text and response are interlinked with headers denoting the language. Significantly, the aforementioned prompts will contain a vacant position for z, either centrally located within the prompt or at its conclusion. In the subsequent text, we will designate the initial type of prompt including a fillable slot inside the text as a cloze prompt, whereas the second type of question, where the input text precedes altogether, will be termed a prefix prompt. (2) Often, these template words are not comprised of natural language tokens; they may consist of virtual words (e.g., represented by numeric identifiers) that are subsequently embedded in a continuous space, and certain prompting techniques The quantity of [X] slots and [Z] slots can be adjusted according to the requirements of the work at hand.

3. RESEARCH METHODOLOGY

Widely employed as a review approach, bibliometric techniques enable academics to objectively and quantitatively examine a collection of publications, revealing important connections, patterns, and similarities within a certain topic. (Aria & Cuccurullo, 2017; Zupic & Čater, 2015).

The process of collecting data is:

Figure 1 Data Collection Process



To determine the most significant contributions, the sample was examined using frequency counts and citation patterns (Kumar et al., 2020; Martín-Martín et al., 2018). The number of yearly scientific publications and the average number of citations per publication were determined before the analysis. The journals, authors, articles, affiliations and keywords in the sample were then evaluated to ascertain their significance. Implementing Bibliometric, an R-based program for mapping literature, the profiling and citation analysis were carried out (Aria & Cuccurullo, 2017).

Using VOS viewer software, a bibliographic network is created to examine the main research topics in the performance of machine learning languages. (Perianes-Rodriguez et al., 2016). VOSviewer is used to perform a co-occurrence analysis to examine the co-occurrence of keywords. (Eck & Waltman, 2016). This study shows the frequency of occurrence of particular keywords in the gathered collection of research.

4. DATA ANALYSIS

The process of data analysis is shown in **Figure 2**:

Figure 2 Data Analysis Process

Documents are analysed on 2 grounds:

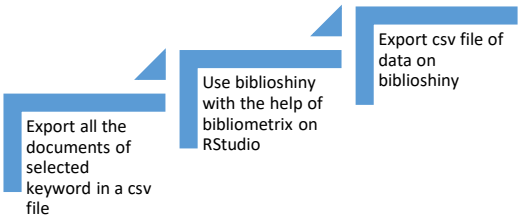
a. Search terms and sample selection:
b. Literature profiling and Co-occurrence analysis:

a. Search terms and sample selection

Sample articles are sorted based on keywords considered “performance of machine learning models” on a research database that is Scopus. The period is from 2014 to 2024. This resulted in 430 samples of documents after refining. The process is given in **Figure 1.**

Some basic information about the samples is presented in **Figure 3**.

Figure 3 Basic Information



The above figure shows the main information of the data. There are a total of 312 sources in 430 samples of documents. The annual growth rate is 63.56%. The total number of authors in all the documents is 1954. International Co-Authorship as co-authors from other countries is 28.14%. The co-Author per document is 5.11. The total number of references in the sample is 19794. The average citation per document is 16.95.

b. Literature profiling and Co-occurrence analysis

- **LITERATURE PROFILING**

The evolution of the performance of machine learning models is demonstrated through literature profiling. It has been displayed how many studies have been published in this field.

Figure 4 Annual Scientific Production

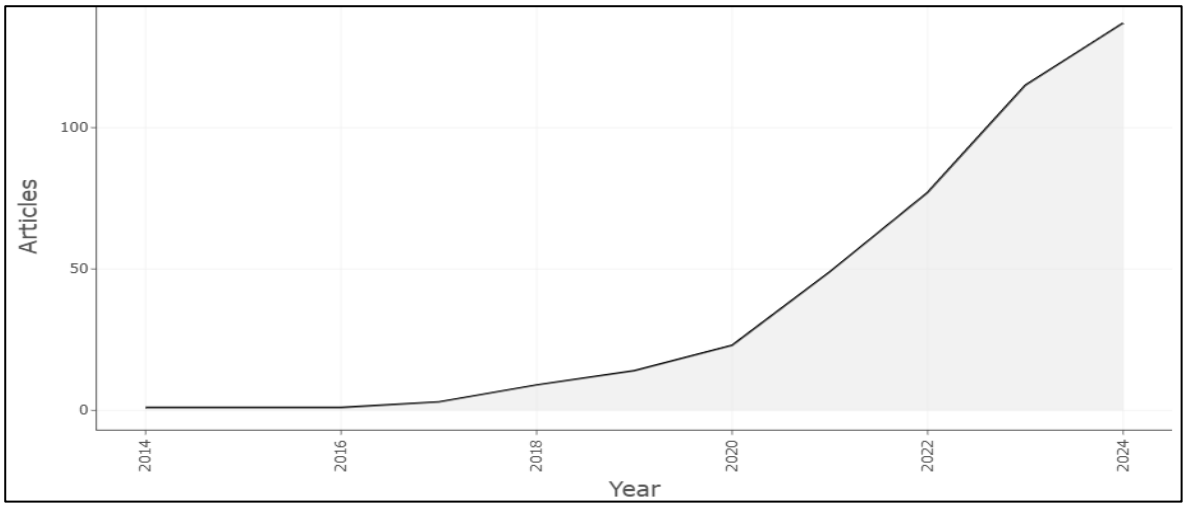


Figure 4 depicts the annual production of articles over the years. The x-axis defined the number of years, while the y-axis represented the number of articles. By analysing, it can be seen that from year 2014 to 2016, there was no such change in the publications but from 2016, the number of articles started rising and continuously rose till now. The number of articles produced is at its peak in the year 2024.

Figure 5 Average Citation per Year

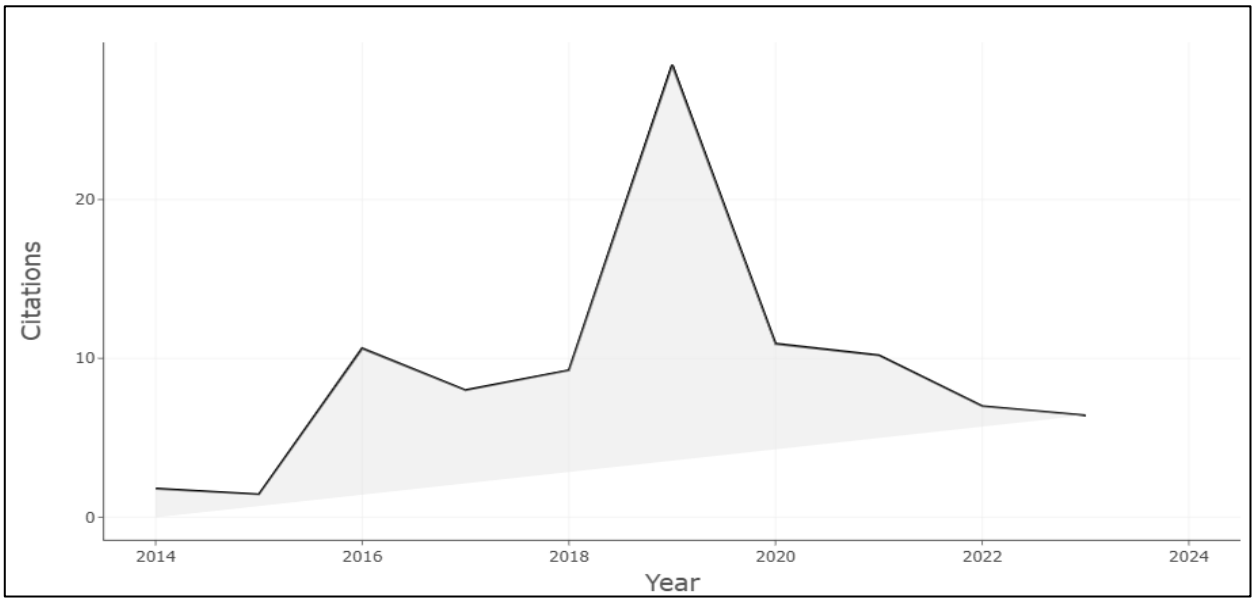


Figure 5 depicts the average citations per year. The x-axis depicts the number of years, while the y-axis represents the number of citations. It can be seen that the average citation per year started rising from the year 2015 and declined in 2017, It again increased and was at its maximum in the year 2019 but then declined in the year 2020 and it is still declining till the year 2024.

Author Analysis

The author's impact is evaluated using two criteria. The first method involves examining the most relevant author, which identifies the authors in the sample with the most publications, and the second involves the work of the top ten authors with the most documents throughout time. **Figure 6** and **Figure 7** show the top 10 relevant authors and their work over the years, respectively.

Figure 6 Top 10 relevant authors

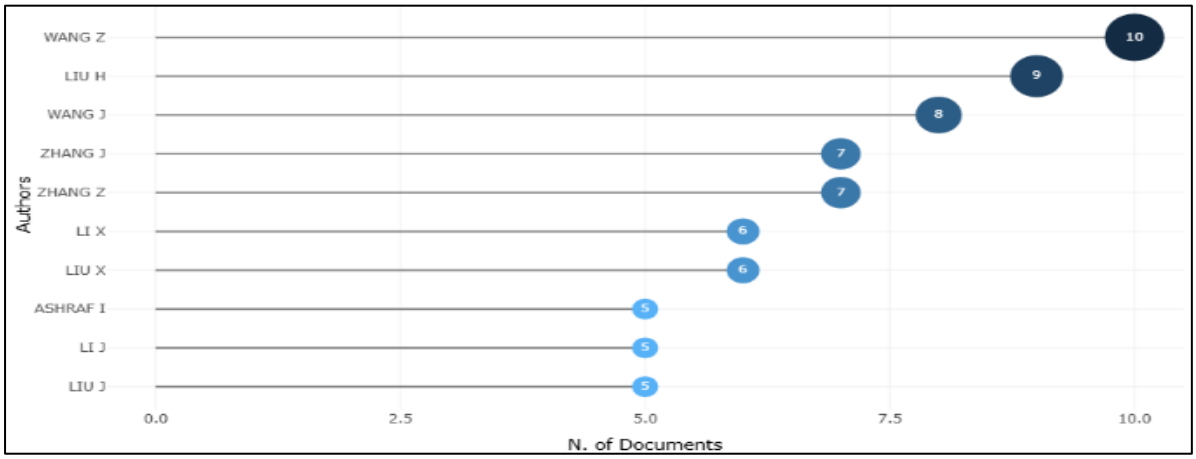


Figure 6 shows the top 10 authors with the most documents. On the x-axis, the total number of documents is shown, and on the y-axis, the authors are stated. Wang Z has 10 articles which is the maximum number. After that Liu H and Wang J have 9 and 8 articles respectively. Both Zhang J and Zhang Z have 7 articles each. Similarly, Li X and Liu X, both have 6 articles each. At last all three remaining authors i.e., Ashraf I, Li J and Liu J have 5 articles each.

Figure 7: Top 10 authors publication over the years

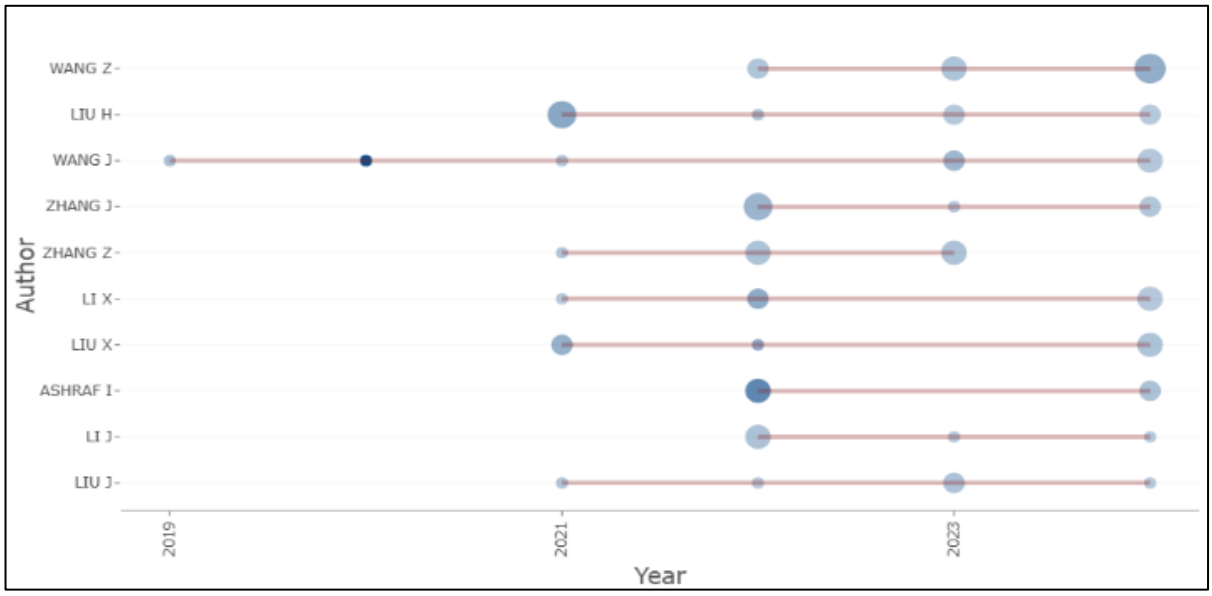


Figure 7 shows the work of the top 10 authors with the maximum number of documents over the years. On the x-axis number of years is mentioned and, on the y-axis, the top 10 authors are mentioned. No author with the maximum number of articles has publications before 2019. Wang J is the only author who has articles from 2019 to 2024. Liu H, Li X, Liu X and Liu J all have articles from 2021 to 2024 with 9, 6, 6 and 5 articles respectively. Zhang Z also has articles from 2021 to 2023. Wang Z, Zhang J, Ashraf I and Li J all have articles from 2022 with 10, 7, 5 and 5 articles respectively. Wang Z with the maximum number of articles have publications from the last three years only and is increasing year by year with 2 articles in 2022, 3 articles in 2023 and 5 articles in 2024. Wang J also the one publishing from the year 2019 is constant in publications and increasing year by year.

Analysis of Documents

Criteria which is used to evaluate the effects of documents are the most globally cited documents.

Figure 8 Top 10 globally cited documents

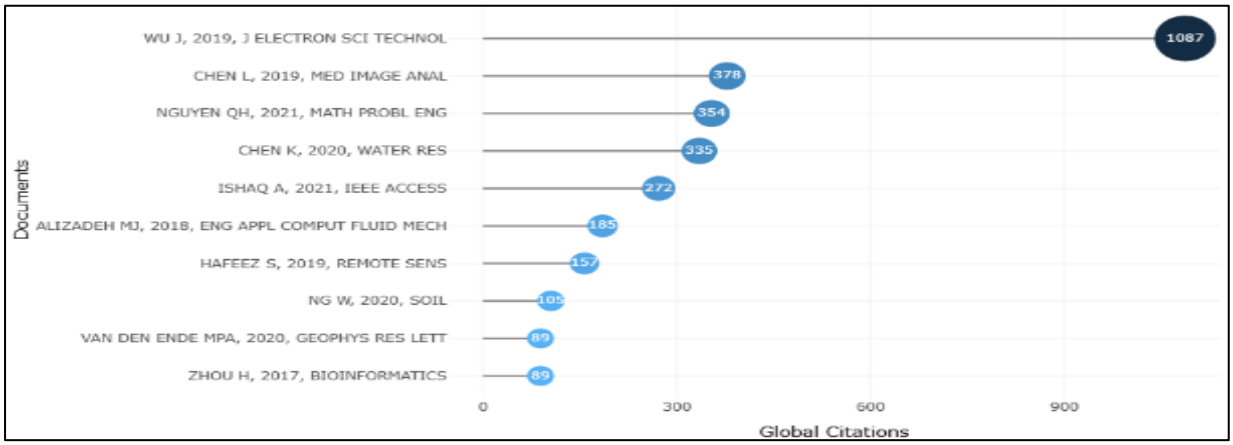


Figure 8 shows the top 10 documents with their global citations. Wu J with 1087 citations discussed the relevance of hyperparameters for machine learning algorithms since they directly influence the behaviours of training algorithms and have a substantial impact on the performance of machine learning models (Wu et al., 2019). Chen L with 378 citations investigated the application of self-supervised learning for medical picture analysis and suggested an image context restoration technique (L. Chen et al., 2019). Nguyen QH and Chen K with 354 and 335 citations; examined the effects of various data-splitting strategies on machine learning models' ability to predict soil shear strength. (Nguyen et al., 2021), and evaluated how well different machine learning models predict the quality of surface water and use big data to identify important water factors (K. Chen et al., 2020). Ishaq A. with 272 global citations in his study addressed class imbalance and improved model accuracy by utilizing SMOTE and efficient data mining approaches to improve the survival prediction of heart failure patients (Ishaq et al., 2021). (Alizadeh et al., 2018), (Hafeez et al., 2019), (Ng et al., 2020), (van den Ende & Ampuero, 2020) and (Zhou et al., 2017) have global citations of 185, 157, 105, 89 and 89 respectively.

Analysis of Affiliations

An affiliation's impact is examined by looking at the maximum number of documents by the top 10 affiliations.

Figure 9 Top 10 relevant affiliations

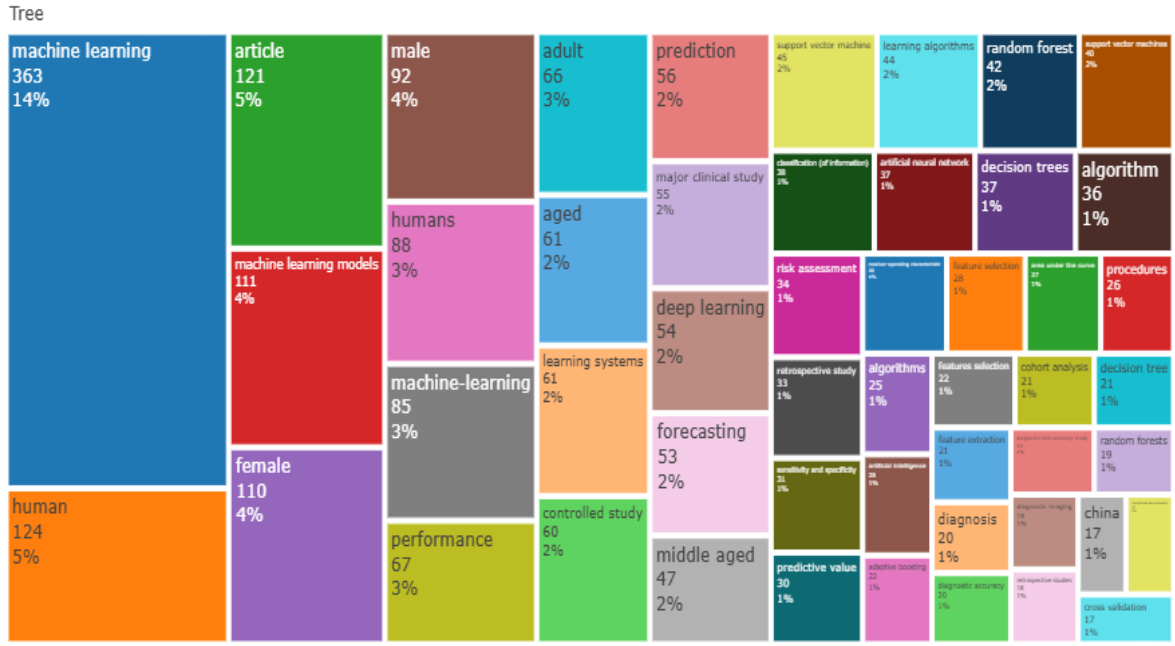


Figure 9 shows the top 10 affiliations with the most documents. On the x-axis, the total number of documents is shown, and on the y-axis, the affiliations are stated. Harvard Medical School have the highest number of documents with 26 articles. University of Electric Science and Technology of China have 20 documents. Capital Medical University and the University of California have 19 and 18 articles respectively. Centre de recherche have 17 documents. Both Sichuan University and Wuhan University have 16 documents each. The University of Toronto has 14 documents. Beijing and Hanyang University have 13 and 12 documents respectively. Most of these affiliations are of China then USA and Canada.

Analysis of Journals

A journal's impact is determined by looking at its highest number of publications.

Figure 10 Top 10 journals with maximum documents

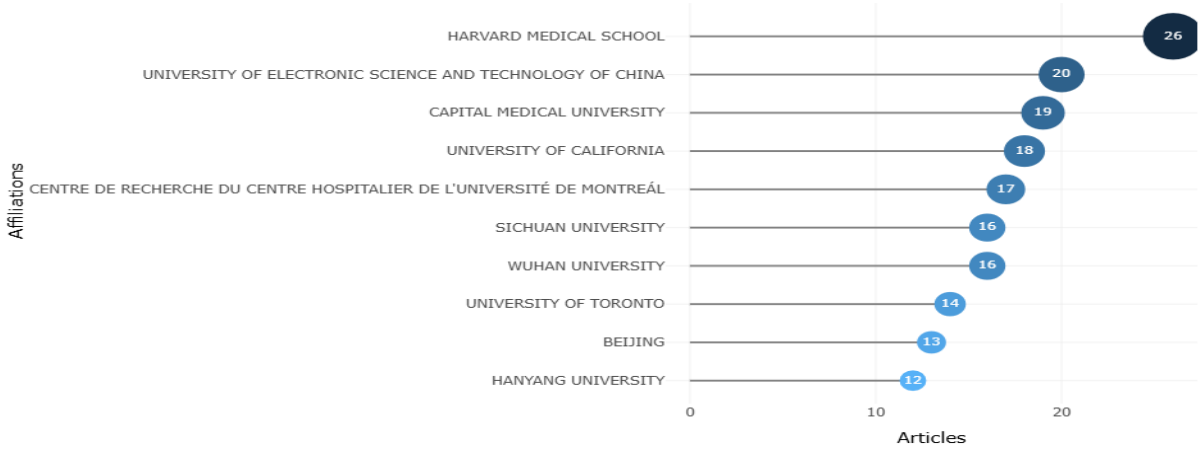


Figure 10 shows the top 10 journals with a maximum number of articles. The X-axis represents the number of documents, and the y-axis represents the top 10 journals. IEEE Access has the maximum number of documents i.e., 14. Scientific Reports and Applied Sciences (Switzerland) have 12 and 10 papers each. International Journal of Medical Informatics has 7 documents. Both Applied Soft Computing and Plos One have 6 papers each. Electronics (Switzerland) has 5 papers. Both Expert Systems with Applications and Peerj Computer Science have 4 documents each. At last, Applied Intelligence has 3 papers.

Country Specific Analysis

The Chart has been produced using biblioshiny to do the country-specific analysis of research evaluating machine learning models' performance. Figure 12 shows the production of articles from the top 10 countries over time.

Figure 11 Country-wise production of articles

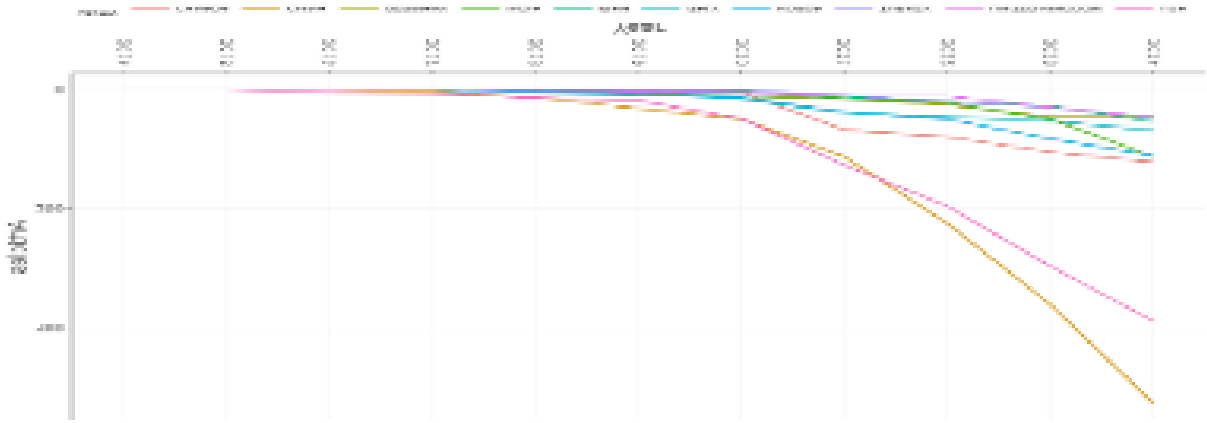


Figure 11 shows the production of articles by country over time. The number of years is shown on the x-axis, and the total number of articles generated by each country is shown on the y-axis. The figure mentions the production of articles by the top 10 countries from 2014-2024. China and the USA have a maximum number of articles of 527 and 389 respectively. Then Canada, India and Korea with 122, 113 and 111 articles respectively. Next Italy and Iran have 69 and 54 articles respectively. Both Turkey and the United Kingdom have 47 articles each. At Last, Germany has a total of 46 articles.

Co-Occurrence Analysis

Co-occurrence analysis highlights the close connections between terms or keywords and offers insights into their interactions. (Eck & Waltman, 2016). On a map, the distance between two terms indicates their degree of relationship; the closer the distance, the closer the two terms are. The full counting method was selected as the counting method, and the co-occurrence analysis was carried out using the VOS viewer tool for all keywords.

Figure 12 Co-occurrence Analysis

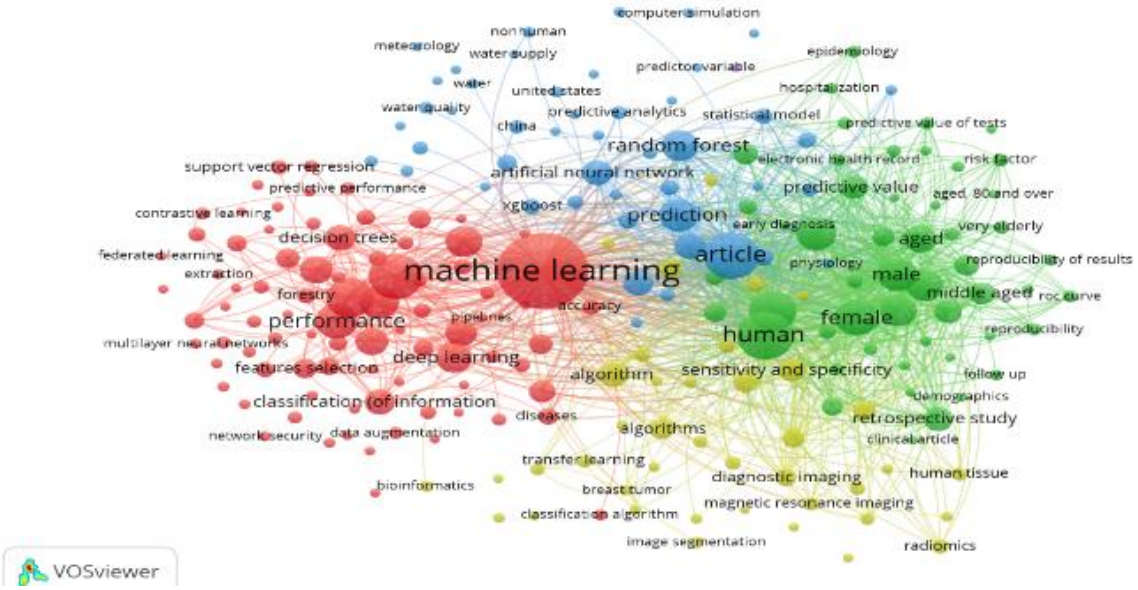


Figure 12 shows the co-occurrence of keywords in the papers on the performance of machine learning models. There were a total of 4221 keywords. The total number of keywords that showed relatedness with the threshold set at 5 was 226. It displays four clusters in various colors. The colors of a cluster are determined by the density and relatedness of the keywords that are part of that cluster. As a result, the red cluster has the maximum density, followed by the green, blue, and yellow clusters, which come in last. Cluster red includes keywords like machine learning, deep learning, performance, predictive performance, feature selection, and decision trees. Cluster green includes keywords like human, female, male, predictive value, and demographics. The blue cluster includes keywords like article, prediction, random forest, artificial neural networks, xgboost, statistical model, and k nearest neighbor. This cluster included most of the models used in the papers and their relationship. Last, the yellow cluster includes keywords like algorithms, sensitivity and specificity, diagnostic imaging, and transfer cleaning.

5. CONCLUSION

Rapid technological breakthroughs are currently at the centre of international discussion, especially in light of machine learning and artificial intelligence (AI) becoming more and more prominent. This study's examination of machine learning models' performance using co-occurrence analysis and literature profiling shows how researchers are becoming more interested in these emerging technologies. In the profile analysis section, the timespan of this research was from 2014 to 2024. The total number of documents was 430 and has an astounding yearly growth rate of 63.46%.

A thorough assessment covering a range of factors, including documents, affiliations, journals, significant authors, and the frequency of appropriate keywords, was carried out during the literature profiling phase. To determine which countries are at the forefront of research on the application of machine learning models for performance evaluation, a country-specific analysis

was also conducted. According to this analysis, India ranks 5th in terms of contributions, whereas China and the US lead the world in research production. Through Co-occurrence analysis, it was discovered that there was a total of 4221 keywords in the total documents out of which 226 keywords met the threshold of 5 set as a number of occurrences of a keyword. The machine learning keyword occurred 316 times in the articles whereas the machine learning models keyword occurred 116 times.

To fully appreciate machine learning models' potential and increase their efficacy in a range of applications, performance evaluation is crucial. Prioritizing strong assessment techniques is essential given the rapidity at which technology is developing to guarantee that these models produce accurate, dependable, and effective results. These models will be improved and their full potential for creativity and problem-solving in various industries will be unlocked with a deliberate focus on performance evaluation.

6. LIMITATIONS AND FUTURE STUDIES

A wider range of keywords might be considered in future studies, as the keywords chosen during the article search process may generate biases.

Additionally, to prevent overpowering the analysis with too many publications, it is recommended to balance the number of papers when using search engines for literature retrieval. Future research should investigate using search engines other than Scopus to improve the literature review's diversity and efficiency

REFERENCES

Alizadeh, M. J., Kavianpour, M. R., Danesh, M., Adolf, J., Shamshirband, S., & Chau, K. W. (2018). Effect of river flow on the quality of estuarine and coastal waters using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, *12*(1), 810–823. https://doi.org/10.1080/19942060.2018.1528480

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., & Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research*, *171*, 115454. https://doi.org/10.1016/J.WATRES.2019.115454

Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. (2019). Self supervised learning for medical image analysis using image context restoration. *Medical Image Analysis*, *58*, 101539. https://doi.org/10.1016/J.MEDIA.2019.101539

Eck, N. J. Van, & Waltman, L. (2016). Text Mining and Visualization. *Text Mining and Visualization*, 1–5. https://doi.org/10.1201/b19007

Hafeez, S., Wong, M. S., Ho, H. C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K. H., & Pun, L. (2019). Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. *Remote Sensing*, *11*(6), 617. https://doi.org/10.3390/RS11060617

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *IEEE Access*, *9*, 39707–39716. https://doi.org/10.1109/ACCESS.2021.3064084

Kumar, B., Sharma, A., Vatavwala, S., & Kumar, P. (2020). Digital mediation in business-to-business marketing: A bibliometric analysis. *Industrial Marketing Management*, *85*(May), 126–140. https://doi.org/10.1016/j.indmarman.2019.10.002

Martín-Martín, A., Orduna-Malea, E., & Delgado López-Cózar, E. (2018). A novel method for depicting academic disciplines through Google Scholar Citations: The case of Bibliometrics. *Scientometrics*, *114*(3), 1251–1273. https://doi.org/10.1007/s11192-017-2587-4

Ng, W., Minasny, B., de Sousa Mendes, W., & Melo Demattê, J. A. (2020). The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *SOIL*, *6*(2), 565–578. https://doi.org/10.5194/SOIL-6-565-2020

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, *2021*(1), 4832864. https://doi.org/10.1155/2021/4832864

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

van den Ende, M. P. A., & Ampuero, J. P. (2020). Automated Seismic Source Characterization Using Deep Graph Neural Networks. *Geophysical Research Letters*, *47*(17), e2020GL088690. https://doi.org/10.1029/2020GL088690

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120

Zhou, H., Yang, Y., Shen, H. Bin, & Hancock, J. (2017). Hum-mPLoc 3.0: prediction

enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features. *Bioinformatics*, *33*(6), 843–853. https://doi.org/10.1093/BIOINFORMATICS/BTW723

Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629