

Article

Automatic Literature Mapping Selection: Classification of Papers on Industry Productivity

Guilherme Dantas Bispo ^{1,*}, Guilherme Fay Vergara ^{1,*}, Gabriela Mayumi Saiki ¹,
Patrícia Helena dos Santos Martins ², Jaqueline Gutierrez Coelho ¹, Gabriel Arquelau Pimenta Rodrigues ¹,
Matheus Noschang de Oliveira ¹, Letícia Rezende Mosquera ², Vinícius Pereira Gonçalves ^{1,*},
Clovis Neumann ¹ and André Luiz Marques Serrano ¹

- ¹ Department of Electrical Engineering, University of Brasilia, Federal District, Brasilia 70910-900, Brazil; gabriela.saiki@redes.unb.br (G.M.S.); jaqueline.coelho@redes.unb.br (J.G.C.); gabriel.arquelau@redes.unb.br (G.A.P.R.); oliveira.matheus@aluno.unb.br (M.N.d.O.); clovisneumann@unb.br (C.N.); andrelms@unb.br (A.L.M.S.)
- ² Department of Economics, University of Brasilia, Federal District, Brasilia 71966-700, Brazil; patricia.martins@redes.unb.br (P.H.d.S.M.); mosquera@redes.unb.br (L.R.M.)
- * Correspondence: guilherme.bispo@redes.unb.br (G.D.B.); guilherme.vergara@redes.unb.br (G.F.V.); vpgvinicius@unb.br (V.P.G.)

Abstract: The academic community has witnessed a notable increase in paper publications, whereby the rapid pace at which modern society seeks information underscores the critical need for literature mapping. This study introduces an innovative automatic model for categorizing articles by subject matter using Machine Learning (ML) algorithms for classification and category labeling, alongside a proposed ranking method called SSS (Scientific Significance Score) and using Z-score to select the finest papers. This paper's use case concerns industry productivity. The key findings include the following: (1) The Decision Tree model demonstrated superior performance with an accuracy rate of 75% in classifying articles within the productivity and industry theme. (2) Through a ranking methodology based on citation count and publication date, it identified the finest papers. (3) Recent publications with higher citation counts achieved better scores. (4) The model's sensitivity to outliers underscores the importance of addressing database imbalances, necessitating caution during training by excluding biased categories. These findings not only advance the utilization of ML models for paper classification but also lay a foundation for further research into productivity within the industry, exploring themes such as artificial intelligence, efficiency, industry 4.0, innovation, and sustainability.

Keywords: intelligence; automatic classification; innovation; productivity in industry; sustainability



Citation: Bispo, G.D.; Vergara, G.F.; Saiki, G.M.; Martins, P.H.d.S.; Coelho, J.G.; Rodrigues, G.A.P.; Oliveira, M.N.d.; Mosquera, L.R.; Gonçalves, V.P.; Neumann, C.; et al. Automatic Literature Mapping Selection: Classification of Papers on Industry Productivity. *Appl. Sci.* **2024**, *14*, 3679. <https://doi.org/10.3390/app14093679>

Academic Editor: Andrea Prati

Received: 4 March 2024

Revised: 9 April 2024

Accepted: 10 April 2024

Published: 26 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Literature mapping is one of the activities carried out by researchers to identify relevant articles for their research during a Systematic Literature Review (SLR). This is one of the daily activities of all researchers and creators of new studies and investigations. In general, there are four stages to carrying out a consistent literature review, which are as follows: (i) Establishing a methodology with a basis and raising relevant research questions; (ii) Collecting a database with applications to support the research; (iii) Evaluating the base of articles collected, in the bibliometric search, considering title, abstract, keywords, number of citations and relevance of the article to the topic of study; and (iv) Structuring the information regarding the results found by the bibliometric search.

Identifying and analyzing the most relevant studies on a given subject is a challenge of literature mapping, as it is an exhaustive activity to find answers to specific research questions so that the information can be relevant to the study. Another problem to be considered is the possible existence of bias in the research when this stage is carried out based on the exclusive individual interpretation of a single reviewer, so it is necessary to have structured studies to

minimize bias and make the research more transparent and robust [1]. In the same article, the authors point out that time is another problem to be overcome, not only because of the time that reviewers spend selecting articles but also because of the time that reviewers will have to dedicate in the coming years if this is not automated, because every year a more significant number of studies appear, so the work of evaluating and reviewing each of these articles has become increasingly complex and time-consuming.

Faced with the challenges of identifying and analyzing relevant studies during literature reviews, machine learning (ML) models are emerging as promising solutions. ML models are revolutionizing several areas, including academia, because they have the ability to understand and interpret texts in a similar or superior way to humans [2]. In this context, it is essential to explore how these technologies can improve and automate the literature mapping process, helping reviewers to make decisions on relevant articles based on textual analysis of the articles [3]. Reinforcing this understanding, [4] states that in the current era of Industry 4.0, knowledge of artificial intelligence and especially mastery of Machine Learning algorithms is crucial for analyzing data in an intelligent and automated way.

Therefore, ML models are expected to be increasingly inserted into society's daily activities and practices, as they have been radically improved and applied to the development of automated and semi-automated systems and methodologies [5]. This is because ML models can understand and comprehend the patterns generated in a way that minimizes human action in the training phase [6]. This makes it possible to develop tools to automate or semi-automate processes, not only to speed up the analysis stage but also to ensure that information can constantly be updated, especially at the article selection stage, to ensure the quality of the review, and because it is the variable that most impacts on the time spent by the reviewer [7].

Several tools and methodologies are being developed to demonstrate the relevance of Machine Learning research aligned with article selection [1]. An example of this is [8], which for the first time pointed out the advantages of using automatic classification of citations for articles within the systematic review of documents relating to the efficacy of the class of drugs in situations where treatment of diseases is required, demonstrating that it is possible to reduce by more than 50% the topics to be reviewed during a manual review.

In reference [1] proposed a systematic literature review in the area of food safety, applying machine learning techniques to automatically classify articles relevant to the area using naive Bayes and a support vector machine, resulting in a reduction of more than 30% of the articles to be read by the reviewer. In [7] evaluated classification models within tokenization, lemmatization, empty word removal, TF-IDF (Term-Frequency/Inverse-DocumentFrequency), decision tree algorithms, and support vector machines for classification algorithms within the field of Software Engineering (SE) with an FScore result of 92%.

Finally, this research aims to contribute to these issues by demonstrating a new methodology that uses machine learning algorithms to improve quality, speed, and standardization along with a Scientific Significance Score-SSS, a scoring system for best literature mapping methods to assess productivity-related issues in the sector. According to the results, it was possible to reduce the number of articles to be evaluated by the reviewer. In some cases, the results indicated a reduction of up to 96.00% in the Innovation and Artificial Intelligence classification category, while the other category had a reduction of more than 90% of the relevant articles in each specific topic.

The research is structured as follows: Section 2 discusses the methods and methodologies used to carry out this study. The category classifier is first presented, which uses Machine Learning models to train and label the category. It then moves on to the relevance classifier, which uses a scoring system called SSS alongside the Z-score to classify the finest papers. The categories used in the research related to productivity within the industry are presented, and the work to collect the database and analyze this data using Bibliometrix is described. Section 3 contains the results of the classifiers by category, using the accuracy, precision, recall, and F-measure metrics to compare the best Machine Learning model between Multinomial Naive Bayes (MNB), SGD Classifier, Support Vector Machine (SVM), and Decision Tree models to be used by the category classifier. In Section 4, the main con-

clusions of the research are presented, in addition to pointing out the main improvements for future studies, as well as issues not addressed by the research.

2. Methodology

This section describes the main methodologies and methods used to structure the category and relevance classifiers. First, a framework is presented for the whole process, including how the database will be handled, the Machine Learning classification models tested, the validation metrics, an explanation of the SIR and how the Z-score and Standard Deviation measures were applied to classify relevance, how the database was structured, and an analysis of the database.

2.1. Integrated Framework for Literature Mapping

We propose a comprehensive framework integrating data extraction from the Web of Science, machine learning classification algorithms, and relevance calculations to conduct literature mapping efficiently.

The database is built from the Web of Science, ensuring the inclusion of high-relevance and reliable articles. Next, machine learning algorithms are employed to automatically classify articles, reducing manual workload and enhancing process efficiency.

Following the initial classification, each category undergoes a relevance calculation. This process involves the application of specific metrics, such as citation frequency, journal quality, and article timeliness. The combination of these parameters results in a relevance score for each article.

The final stage of the framework involves the selection of the most relevant articles based on the calculated relevance scores. Predefined criteria and a specific threshold are established to identify each category’s finest papers.

Figure 1 illustrates the integrated flow of the proposed framework. It begins with data extraction from the Web of Science, followed by applying machine learning classification algorithms. After classification, relevance calculation takes place, considering metrics such as citation frequency, journal quality, and article timeliness. The last step is the selection of the most relevant articles based on calculated relevance scores, establishing a specific threshold. This systematic approach aims to optimize identifying significant works in systematic literature reviews.

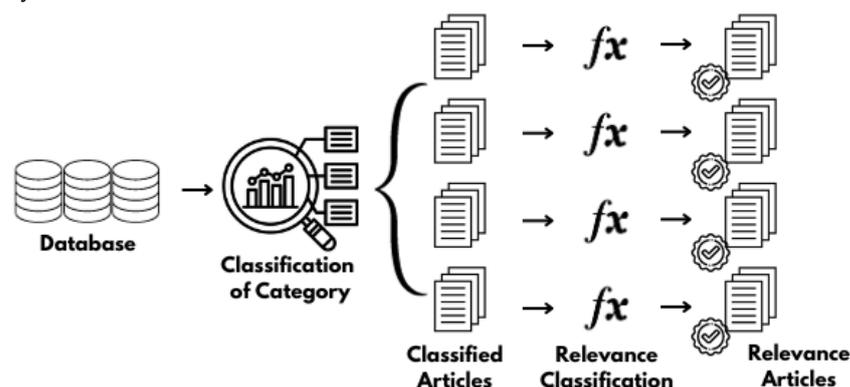


Figure 1. Integrated Framework for Systematic Literature Review.

2.2. Database

Web of Science is a platform with an essential database for the area of exact sciences; it has a series of articles from engineering, economics, and natural sciences. Therefore, it becomes essential to survey articles focused on productivity and industry. Given the increasingly massive use of data and the requirement to generate information quickly, the literature review also has to adapt computationally to analyze the articles, thus becoming faster and more automatic. WoS allows for exporting a list of information about the articles searched for external analysis. A computational tool used to analyze article data is the Bibliometrix package from RStudio [9], which makes data analysis simple, with studies

involving the prominent authors, countries, articles, affiliations, thematic evolution, clusters of themes, and publications over the years.

The steps for data extraction from the Web of Science platform, database consolidation, and Bibliometrix analysis follow the following structure, composed of five steps:

- Definition of research categories within the macro-themes industry and productivity (Artificial Intelligence, Efficiency, Industry 4.0 and Industry 5.0, Innovation, Sustainability, Business, Gross Domestic Product (GDP), Sustainable Development Goals (SDGs), Public policy).
- Search within the macro-themes (industry and productivity) by category.
- Extraction of data in Excel format for consolidation of labeled database (to be used in the training process of machine learning models) and Bibtex for consolidation of base for analysis in Bibliometrix (as it is a more recommended format for the use of the database within the environment in biblioshiny).
- Creation of a new column within the Labeled Database to label with the category's name corresponding to each article according to the search result in WoS.
- Removal of duplicates for Bibliometrix analysis.

Each article's data for machine learning training are Title, Abstract, Author, Journal, Year, Country, Keywords, DOI, and Citations, as shown in Figure 1.

2.3. Categories of Productivity in Industry

In our investigation, we undertook a Literature Mapping using the conventional methodology involving manual keyword selection and filtering. Subsequently, we employed both Category Classification of Scientific Articles and Relevance Classification of Scientific Articles as equitable indicators of workload to make a comparison and exhibit its efficacy.

We are poised to adhere to a stringent approach to conducting an extensive bibliometric examination centered on productivity in the industry. This entails applying specific filters to ensure the pertinence and contemporaneity of the encompassed studies. To encompass a substantial temporal scope, we intend to scrutinize articles published from 2014 to 2023, drawing upon various reputable sources.

In industrial productivity, six primary dimensions (Table 1) exist that play pivotal roles in driving progress and reshaping the sector. These dimensions encompass innovation, sustainability, Industry 4.0, Industry 5.0, efficiency, and artificial intelligence. Each of these will be examined below:

Innovation catalyzes augmenting industrial productivity by introducing novel concepts, technologies, and processes that streamline operations and amplify efficiency. It also fosters adaptability to market shifts and inventive solutions to industrial challenges. Sustainability is critical in the industry, entailing practices that curtail resource consumption, minimize environmental impact, and advocate for energy efficiency. Companies adopt sustainable strategies to enhance productivity while diminishing their ecological footprint.

Industry 4.0 Industry 4.0 integrates cutting-edge digital technologies, including IoT, big data, and artificial intelligence, to establish intelligent factories. This leads to automation, heightened quality control, and augmented productivity.

Industry 5.0 represents a fusion of advanced technologies, such as collaborative robotics and artificial intelligence, with human labor to bolster productivity. Humans and machines collaborate, harnessing their distinct capabilities to amalgamate efficiency and creativity.

Efficiency is a cornerstone of industrial productivity, entailing the optimization of processes, waste reduction, and quality enhancement. This is achieved through a perpetual process analysis and refinement cycle for more efficient operational performance.

Artificial Intelligence drives industrial productivity by ushering in automation, refining supply chain logistics, and steering data-powered decision-making. These advancements result in heightened efficiency and enhanced quality.

These six categories interact with each other, forming a set of approaches and technologies that contribute to boosting productivity in the industry. Adopting and integrating

these elements into industrial strategies and processes can result in significant efficiency, competitiveness, and sustainability gains.

Table 1. Categories —Industry and Productivity.

Categories	
Innovation Industry 4.0 Artificial Intelligence	Efficiency Industry 5.0 Sustainability

2.4. Analysis of the Database

This first part analyzes the data used in the data preparation process and provides an overview of the 7872 articles. This is an important step that precedes the cleaning and training of machine learning models. Since 70% of the labeled database will be used to train the algorithm, it is necessary to understand what prior information from these data can be generated and the articles’ initial structure to guide the research’s next steps.

With the requirement for increasingly faster information generation from big data or even from smaller databases, it becomes essential and evident how for the adequacy of data analysis and information generation to occur automatically and computationally, several tools for bibliometric analysis have emerged to allow this process to take place more intuitively with the support of new technologies, such as Scimago, BibExcel, and SciVal. In this way, Bibliometrix emerges as a bibliometric analysis tool for massive analysis of data related to articles, whereby platforms such as Web of Science, Scopus, Dimensions, Lens.org, PubMed, and Cochrane Library can have raw files imported into the Bibliometrix environment, as well as data frames which can also be analyzed by the tool, provided there is adequate data treatment. It is interesting to remove the replications of articles if there is more than one database to be analyzed, given that article platforms may contain the same article within the list of journals or even in cases where several bases were extracted from the same platform by using different combinations.

From the six predefined categories, a base of 9842 journals from 2014 to 2023 is obtained while filtering only the articles. In total, 1970 replications are analyzed; without the replications, the base results in 7872 articles. Bibliometrix analyzes a publication growth rate of 15.5%, 9842 sources, 17,380 authors, and 666 authors of documents with single authorship. It is analyzed that the number of single authors is insignificant compared to most articles containing more than one author per document, including 29.82% of articles with international co-authorship, which points to significant international collaboration between countries, resulting in 20,056 keywords and 284,432 references that were used, with an average of 15.28 citations per document and an average age of 4.07 years per document, which shows that most of the exported articles are recent studies within less than 5 years of publication, which indicates the productivity and industry is high within the nine categories surveyed (Figure 2).



Figure 2. Main Information.

The main authors and the main keywords dealt with in the articles. The main authors are Li Y., Wang Y., Li J., Lin B., and Chen X.

The keywords are China, productivity, efficiency, data envelopment analysis, and industry. It is analyzed that the main authors and the authors who publish the most on the

subject are Chinese. It is observed that China appeared as the most used keyword, which demonstrates the country's dominance within the macro-themes productivity and industry. The other two keywords that stand out in the research are efficiency and Data Envelopment Analysis (DEA); DEA is used for efficiency analysis; it is the only methodology that appears among the most addressed topics.

Three groupings resulting from the 8621 articles are analyzed. Two clusters are formed by the macro themes of productivity, identified by the green color, and industry, identified by the blue color. It is observed that there is a tendency for the articles to be located in some thematic pole.

The efficiency category is more related to industry, while the innovation and policy categories are more related to productivity. The third grouping labeled with red is the sustainable category.

2.5. Category Classification of Scientific Articles

Categorizing scientific papers into pertinent sections is crucial for effectively arranging and retrieving information from an extensive corpus of scholarly literature.

In this subsection, we elucidate the approach undertaken to train and assess the efficacy of four distinct machine learning models—Multinomial Naive Bayes Classification, Stochastic Gradient Descent, Support Vector Machines, and Decision Tree Classifier—applied to classify scientific articles across six distinct domains: Innovation, Sustainability, Industry 4.0, Industry 5.0, Efficiency, and Artificial Intelligence. A visual depiction of the complete process executed can be observed in Figure 3.

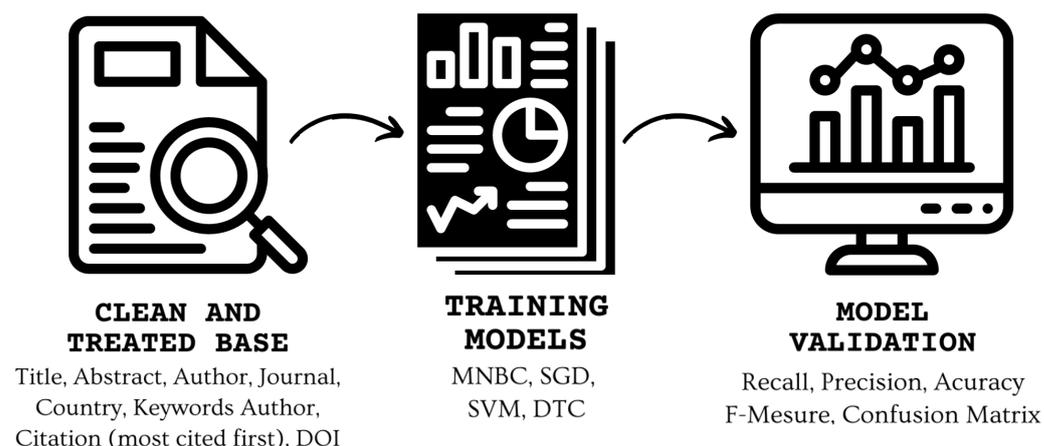


Figure 3. Category classification of scientific articles.

2.5.1. Data Preparation

The dataset employed in this research was compiled from six distinct queries on the Web of Science (WoS), guaranteeing a diverse assortment of articles for each category. Articles devoid of title, abstract, or keyword data were eliminated from the dataset due to their insufficiency for accurate classification.

Additionally, measures were taken to ensure the exclusivity of each article to a single category. Any articles duplicated in more than one category have been removed to ensure that each article is in only one category. These duplicates were discovered by comparing the DOI of the articles found in each category. This strategy was adopted to ensure the dataset was more balanced. This meticulous data cleaning and preparation procedure safeguarded the reliability of the training and validation sets. The distribution of article categories post-duplication removal is illustrated in Figure 4.

In terms of textual representation, a corpus was built by employing the TF-IDF (Term Frequency-Inverse Document Frequency) methodology, as outlined in the study conducted by Qaiser et al. (2018) [10]. This technique transformed the textual content into numerical vectors. Furthermore, stemming was employed to truncate words to their root forms, curtailing feature space dimensions and enhancing model performance.

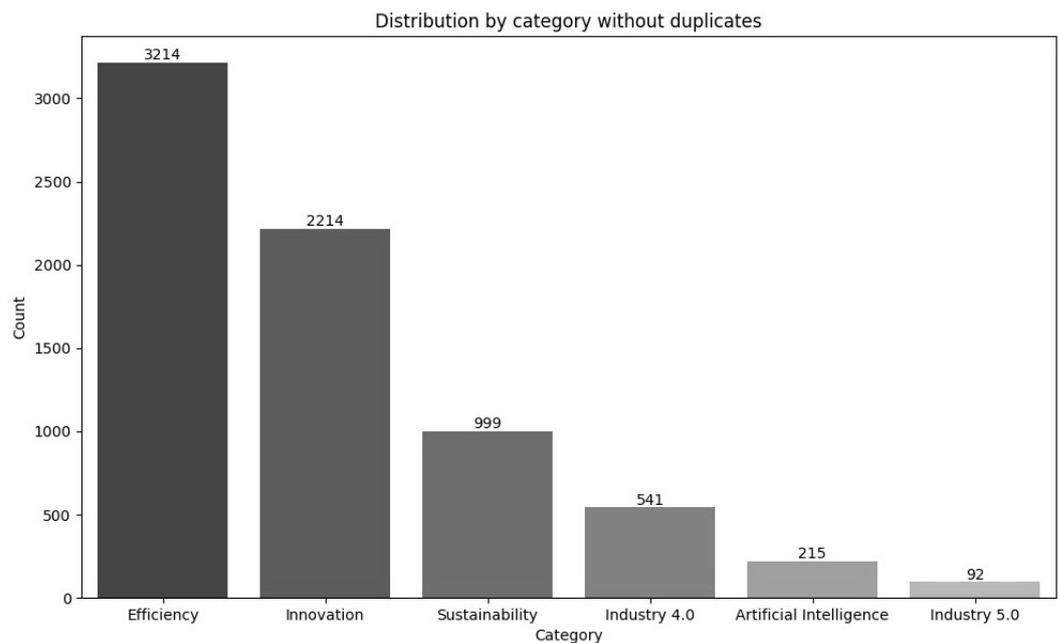


Figure 4. Distribution by category without duplicates.

Subsequently, the dataset was randomly partitioned, allocating 70% for training and 30% for validation, thereby guaranteeing both subsets were representative of all six categories.

2.5.2. Machine Learning Models

Having readied the training dataset, the four designated machine learning models underwent training using the prepared data.

- Multinomial Naive Bayes Classification (MNBC) is a prevalent technique for text classification, establishing a classifier from a network of trained data. According to [11], MNBC utilizes the Frequency Estimate parameter to accentuate data frequencies. This feature contributes to the reliability of results and renders MNBC an effective tool for data classification.
- Stochastic Gradient Descent (SGD) is a classifier capable of comprehending increasingly intricate functions and generalizing over parameterized data, as indicated by [12]. It is noteworthy that Dynamic Multiclass Strategy (DMS) frequently initiates learning with straightforward classifiers, deferring the use of complex ones. The rationale lies in preserving the information gleaned from the initial learning phase via the simpler classifiers. Consequently, the DMS initiates learning in a predominantly linear manner.
- Support Vector Machines (SVMs) operate based on the minimum structural risk principle found within the statistical theory [13]. Consequently, SVM plays a pivotal role in data classification by employing linear regression to generate outcomes. SVM transforms a nonlinear input dataset into a linear form by applying a kernel function [14].
- The Decision Tree Classifier (DTC) serves to simplify intricate issues, making them more comprehensible and aiding in decision-making [15]. As described by [16], DTC is characterized as a hierarchical classifier, furnishing multi-level classification and divulging the specific pattern to which each datum belongs. Moreover, it offers adaptability in handling both binary and multi-class classifications.

The hyperparameters in each model were optimized using GridSearchCV (Alhakeem et al., 2022) [17]. As can be seen in Figure 5, this approach enables a grid of possible values to be created for the hyperparameters and then involves examining all the available combinations so that the best configuration can be determined. The GridSearchCV implementation uses cross-validation to analyze the performance of each set of hyperparameters. By employing this method, we were able to discover the best combination of hyperparameters for each model and thus optimize its effectiveness. The metric chosen for the best

sets of hyperparameters was the average of the cross-validation results. By adopting this methodology, we could pinpoint each model’s optimal blend of hyperparameters, elevating its overall performance to its pinnacle.

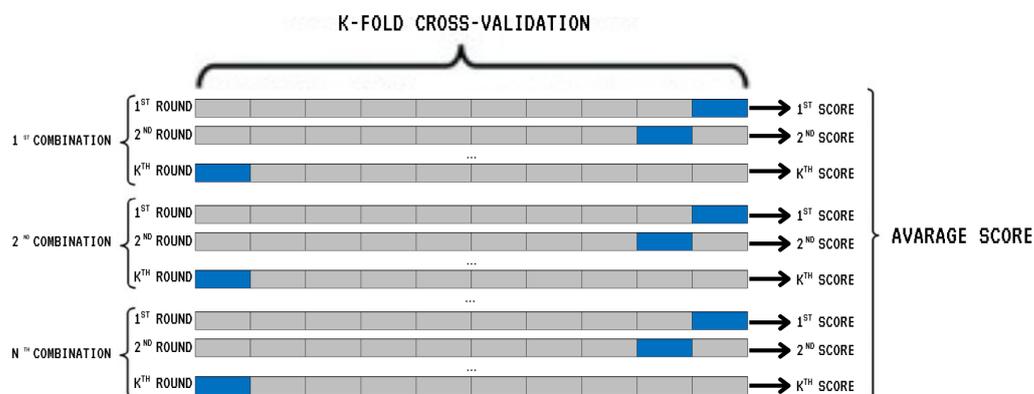


Figure 5. GridSearchCV method.

2.5.3. Method Validation Metrics

A crucial aspect of the development of machine learning systems is the validation of the model. This process allows us to assess how well the model performs and generalizes when faced with new and unseen data. Numerous key elements of model validation are underscored:

- Recall calculates the ratio of true positives to the total count of positive samples within the testing dataset. This measurement holds significance when the objective is to reduce false negatives, thus preventing the model from erroneously labeling positive samples as negative.
- Precision represents a measurement that gauges the fraction of accurate positive predictions (true positives) to the complete count of positive predictions generated by the model. This metric is valuable when the emphasis diminishes false positives, preventing the model from categorizing negative samples as positive.
- Accuracy serves as a prevalent assessment measure, quantifying the ratio of accurate forecasts produced by the model relative to the predictions rendered. This calculation entails adding the correct forecasts and dividing this sum by the overall count of instances. Accuracy proves advantageous in scenarios where classes are evenly distributed, signifying that they possess comparable instance quantities.
- The F-measure represents a measurement that amalgamates precision and recall, forging a singular gauge. This metric comes in handy when class distribution is uneven, as it considers false positives and negatives alike. The F-measure is computed as the harmonic average of precision and recall, offering a cohesive yardstick to portray the comprehensive prowess of the model’s performance.
- The confusion matrix serves as a technique to assess the outcomes of machine learning-based classification. It furnishes a thorough overview of the model’s predictions through a juxtaposition with the factual actual values. We can pinpoint instances of false positives and false negatives by conducting tests with labeled test data.

While validating a machine learning model, it becomes imperative to scrutinize these factors to cherry-pick the most suitable model coupled with finely tuned parameters. This process aims to attain peak performance, gauged by a spectrum of metrics encompassing accuracy, precision, recall, and the F-measure. The choice of which metric to prioritize hinges on the distinctive requirements of the problem and the classes under consideration.

2.6. Scientific Significance Score

Assessing the significance of scientific papers holds paramount importance in pinpointing the most noteworthy contributions within a particular research field or category. A prevalent technique employed for this assessment involves the utilization of a proposed

relevance classifier called SSS alongside two pivotal metrics: the standard deviation of the mean and the Z-score. In the following section, we will delve into these concepts and their pivotal role in categorizing scientific articles based on their relevance.

Let SSS represent the Scientific Significance Score, a metric proposed in this paper to rank journals based on their scientific significance. The formula for calculating SSS is as follows:

$$SSS = \frac{SJR \times 10^{SF-1} + NC}{1 + CY - PY} \tag{1}$$

where:

- *SF* = the number of decimal places of the rounded value of the Scimago Journal, Country Rank (*SJR*).
- *NC* = Number of Citations of the paper.
- *CY* = Current Year.
- *PY* = Publication Year, when the paper was published.

This formula represents a scoring mechanism where SSS is determined based on an international index *SJR*, the number of citations (*NC*) of the paper, and the temporal difference between the current year (*CY*) and the publication year (*PY*).

To identify the top-tier articles, we leveraged the criterion of the article’s citation count. To cherry-pick the finest articles, we employed an outlier detection method, such as the Z-score and the standard deviation of the mean [18].

Z-scores typically identify outliers that notably exceed (positively or negatively) the majority of the data. When a Z-score surpasses a predetermined threshold (e.g., ±2 or ±3), it can indicate that a specific data point is an outlier. In our study, we classify articles as Finest if their Z-score exceeds 3, while articles are deemed relevant if their Z-score falls below −3 and is greater than the upper range of the standard deviation, and the articles below the upper bound of the standard deviation were classified as Good.

3. Results

This segment unveils this study’s outcomes by presenting and examining how four distinct classification models fared when applied to the challenge of categorizing scientific articles.

3.1. Category Classification

The assessment encompasses the Multinomial Naive Bayes (MNB), SGD Classifier, Support Vector Machine (SVM), and Decision Tree models. Each of these models underwent multiple runs with different hyperparameter configurations, and their effectiveness was gauged by metrics such as the F-measure [19,20]. The primary objective is to discern the most fitting model for the given challenge of subject classification. This scrutiny involved processed data transformed into vectors. The dataset is composed of labeled documents spanning diverse subject categories. The task involves training the models to aptly classify new documents into their appropriate categories.

Table 2 shows the values of the F-measure metric for each model. This is the most suitable metric when we have an unbalanced database. We highlight the remarkable rate for this metric for the Decision Tree model, which reached a percentage of over 74%.

Table 2. Web of Science results table.

Variables	MultinomialNB (Naive Bayes)	SGDClassifier	SVM	Decision Tree
F-measure	0.6606	0.7134	0.7367	0.7441

- (1) Multinomial Naive Bayes (MNB). This classifier hinges on Bayes’ theorem and is acknowledged for its simplicity and computational efficiency. When evaluating MNB’s performance

in our subject classification endeavor (Figure 6), we observed that the F-measure rests at 66.06%.

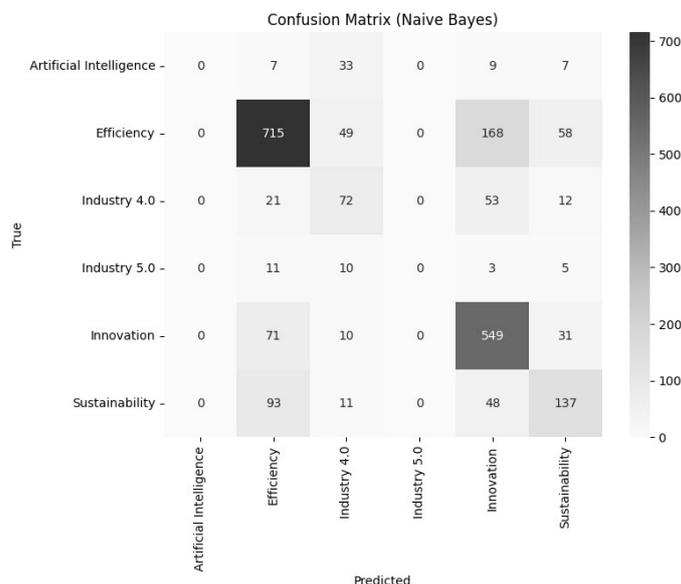


Figure 6. Confusion Matrix (Naive Bayes).

Despite the advantages of its simplicity and computational efficiency, the outcomes suggest that MNB might not be the optimal choice for our specific subject classification task. In this context, exploring alternative models that have demonstrated more promising outcomes becomes pivotal. This exploration aims to pinpoint the model that best aligns with the project’s requirements and objectives.

- (2) The Stochastic Gradient Descent (SGD). The SGD classifier, a linear classifier employing stochastic gradient descent for optimization, was subjected to performance evaluation in our subject classification undertaking. Our findings demonstrate that the SGD classifier attained the F-measure of 71.34%. This outcome underscores the model’s superiority over MNB, and this improvement can be attributed to its enhanced capability to tackle more intricate classification challenges (Figure 7).

The enhanced F-measure demonstrated by the SGD classifier, compared to MNB, can be attributed to its inherent characteristics and proficiency in handling linear relationships among problem features. This ability proves especially valuable when dealing with intricate classification challenges. Consequently, these outcomes suggest that the SGD classifier presents a more favorable prospect for the subject classification task than MNB. However, it remains crucial to thoroughly consider the other evaluated models to ascertain the most suitable option that aligns with the precise demands of this study.

- (3) Support Vector Machines (SVM). Recognized for its efficacy in high-dimensional spaces, the Support Vector Machine (SVM) stands as a frequently employed classifier in classification tasks. Our findings reveal that the SVM garnered an F-measure metric of 73.67%. These demonstrate that the SVM surpassed the performance of both MNB and SGD classifiers. This highlights many true positives, evident within the associated confusion matrix (Figure 8).

The remarkable outcomes observed in accuracy, recall, and precision, when contrasted with those of the MNB and the SGD Classifier, strongly validate the SVM’s potency for subject classification tasks. Notably, it shines in scenarios characterized by expansive data dimensionality and intricate classification intricacies.

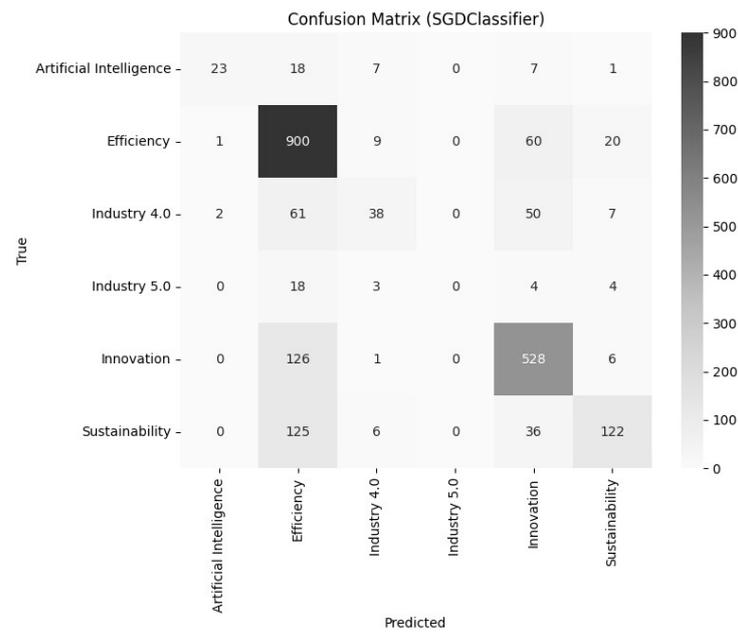


Figure 7. Confusion Matrix (SGD).

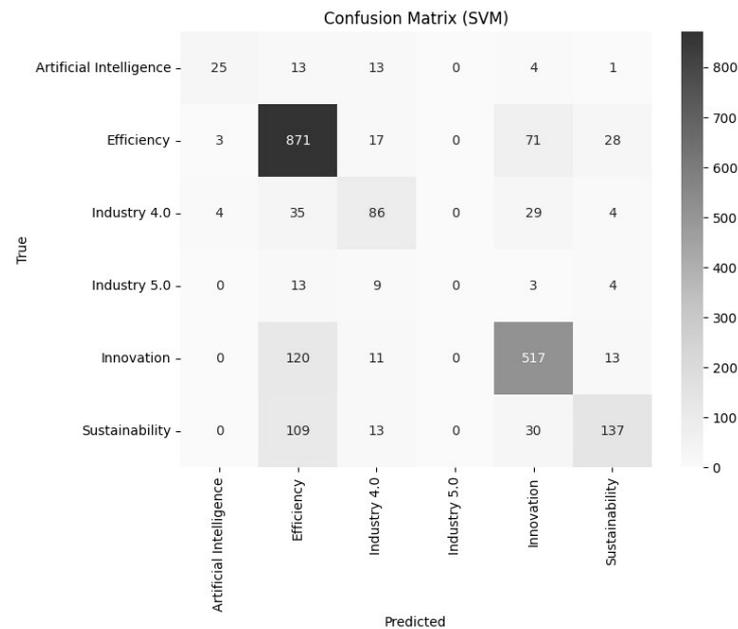


Figure 8. Confusion Matrix (SVM).

Consequently, these findings underscore the SVM’s stature as a remarkably competitive and accomplished choice for the subject classification endeavor. Nevertheless, it remains imperative to weigh additional factors like runtime and interpretability when choosing the most fitting model that aligns with our project’s unique prerequisites and objectives.

- (4) The Decision Tree, recognized for its interpretability and reliance on decision rules, stands out as a model choice. Our findings reveal that the Decision Tree model attained an F-measure of 74.41%. This outcome reigns supreme among the models under examination, solidifying how the Decision Tree excelled in the subject classification undertaking.

The commendable outcomes achieved in accuracy, recall, and precision affirm that the Decision Tree is a formidable contender for subject classification tasks. Notably, its

interpretability is a considerable asset, offering more precise insights into the model's decision-making process and enabling the recognition of classification trends. In light of the results presented, it is evident that the Decision Tree emerged as the top-performing model for the subject classification undertaking. The interpretability feature of the Decision Tree brings a substantial advantage to the table when navigating the realm of subject classification.

3.2. Relevance Classification

The 7872 articles classified by subject were further classified into three categories of relevance: Good, Relevant, and Finest (Table A1).

(1) Innovation. In Innovation, we classified as follows (Figure 9):

- Good: 2600
- Relevant: 27
- Finest: 23

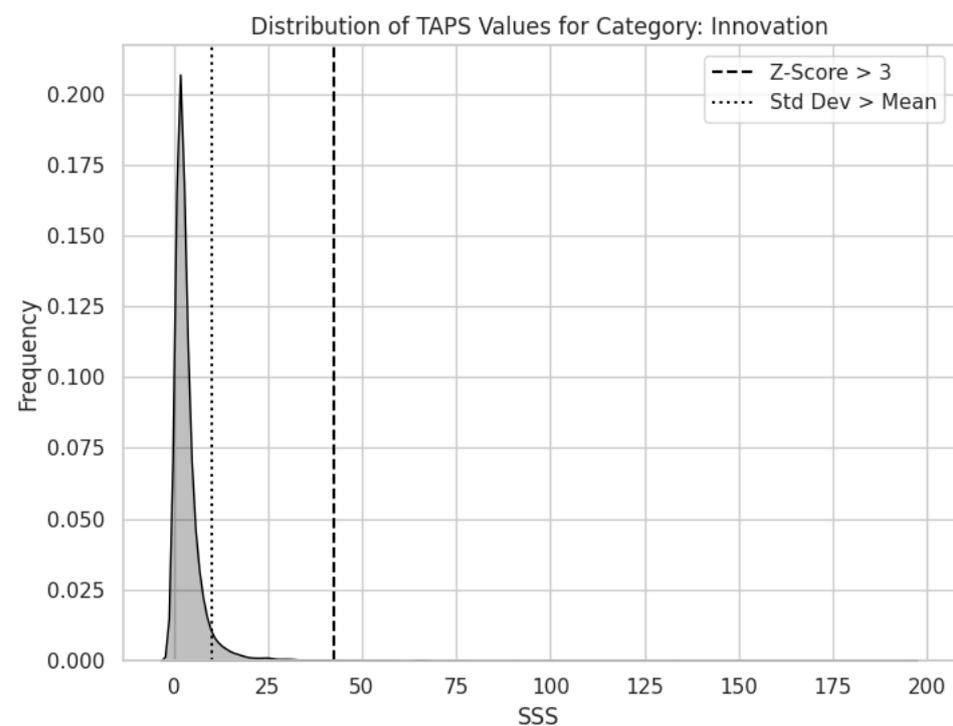


Figure 9. Relevance article classification for Innovation.

The finest article is Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs.

(2) Efficiency. In Efficiency, we classified as follows (Figure 10):

- Good: 4556
- Relevant: 24
- Finest: 24

The finest article is Impact of Energy Conservation Policies on the Green Productivity in China's Manufacturing Sector: Evidence from a Three-stage DEA Model.

(3) Sustainability. In Sustainability, we classified as follows (Figure 11):

- Good: 715
- Relevant: 311
- Finest: 20

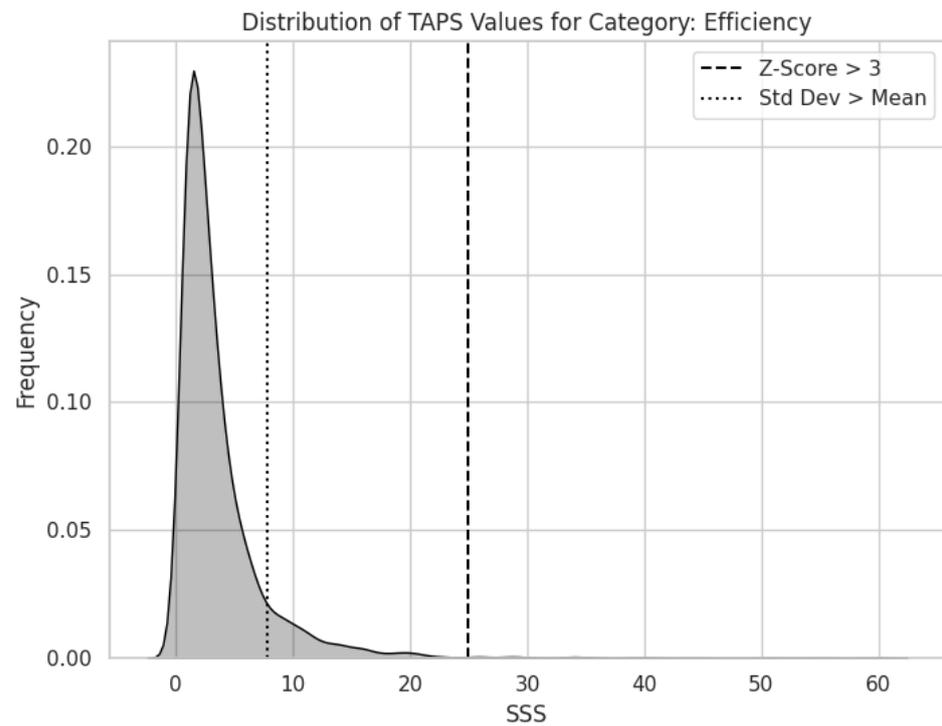


Figure 10. Relevance article classification for Efficiency.

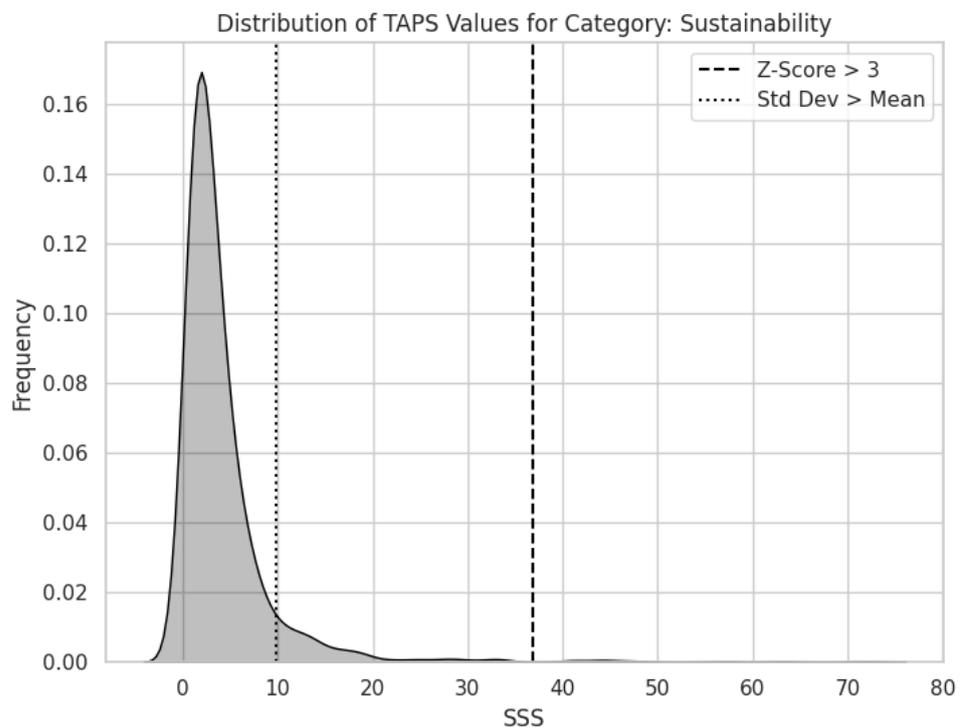


Figure 11. Relevance article classification for Sustainability.

The finest article is Natural Fiber Reinforced Polymer Composites in Industrial Applications: Feasibility of Date Palm Fibers for the Sustainable Automotive Industry.

(4) Industry 4.0. In Industry 4.0, we classified as follows (Figure 12):

- Good: 684
- Relevant: 3
- Finest: 3

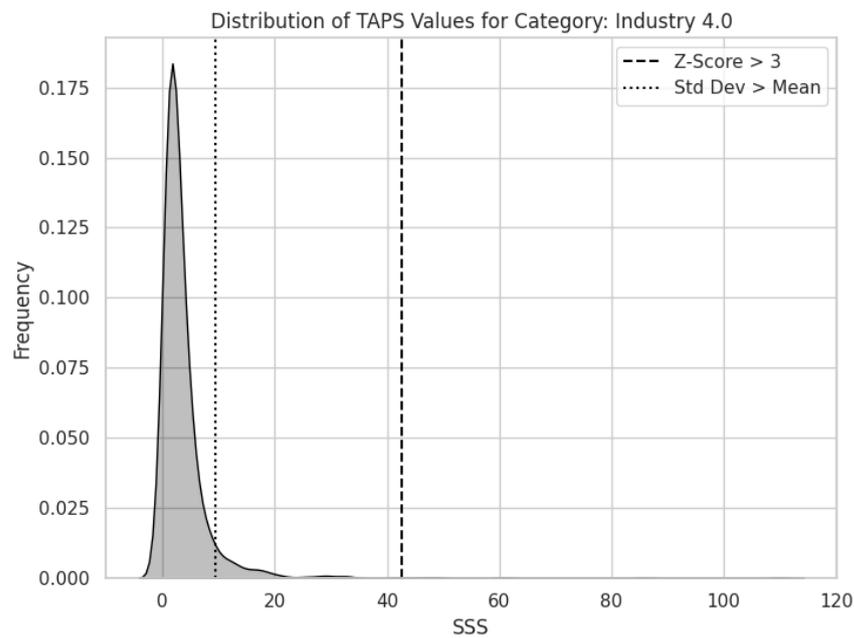


Figure 12. Relevance article classification for Industry 4.0.

The finest article is State-of-the-Art in Surface Integrity in Machining Nickel-based Superalloys.

(5) Artificial Intelligence. In Artificial Intelligence, we classified as follows (Figure 13):

- Good: 168
- Relevant: 59
- Finest: 5

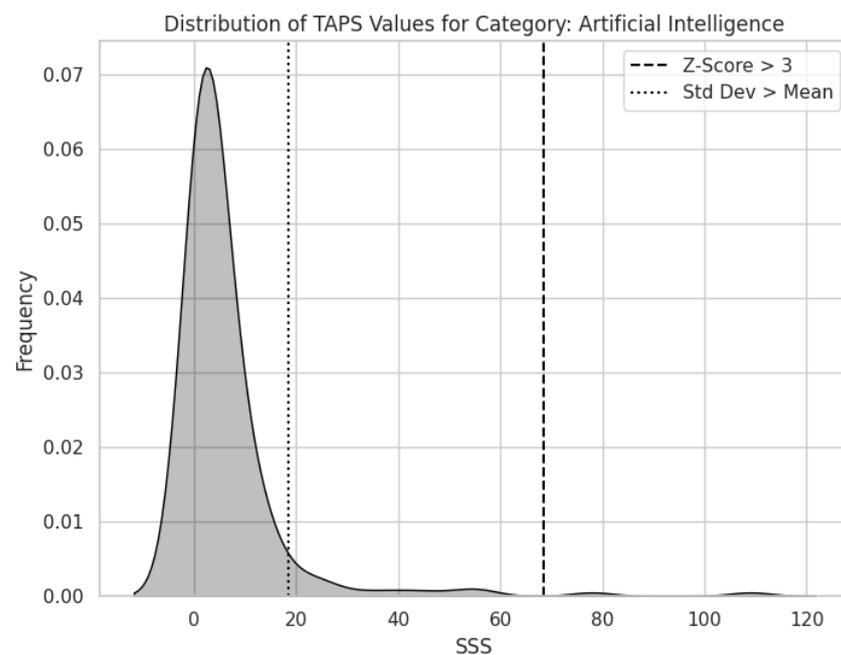


Figure 13. Relevance article classification for AI.

The finest article is Brave New World: Service Robots in the Frontline.

4. Conclusions

In summary, this study aimed to compare various approaches to systematic literature reviews, seeking the most efficient method for swift and effective review processes. Our

investigation categorized industry productivity into six core sectors: Innovation, Sustainability, Industry 4.0, Industry 5.0, Efficiency, and Artificial Intelligence. Subsequently, systematic literature reviews were carried out for each of these sectors. To streamline the classification, we leveraged four Machine Learning models: Multinomial Naive Bayes Classification (MNBC), Stochastic Gradient Descent (SGD), Support Vector Machines (SVM), and Decision Tree Classifier (DTC). Among these, the Decision Tree Classifier (DTC) exhibited remarkable performance, boasting a classification accuracy surpassing 75%.

A relevance classifier was employed to pinpoint significant contributions in scientific articles across various research domains or categories. This classifier integrated two crucial metrics: the standard deviation of the mean and the Z-score.

As a prospective recommendation for future research, we underscore the importance of incorporating supplementary data sources and considering the h-index as an alternative method for ranking article relevance. By broadening the data scope and integrating diverse evaluation metrics, upcoming studies have the potential to amplify the precision and depth of systematic literature reviews. Also, we recommend the creation of a human curation stage, in which the results of relevant articles are evaluated manually to endorse the results of the evaluation presented. Ultimately, this contributes to more comprehensive insights and well-informed decision-making within industrial productivity.

Author Contributions: In this study, each author played a distinct role. G.D.B. and G.F.V. led the methodology as well the validation and formal analysis alongside. Data curation was handled by G.M.S. Initial drafting was primarily by P.H.d.S.M. and J.G.C., with input from others. Visualization was carried out by M.N.d.O. and G.A.P.R. Review and editing were led by V.P.G. and L.R.M., A.L.M.S. and C.N. supervised the project and managed administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available in github <https://github.com/GuiFay/Automatic-Literature-Mapping-Selection>.

Acknowledgments: The authors would like to thank the LATITUDE Laboratory of the University of Brasilia for its technical and computational support, the Federal Attorney General’s Office for TED 01/2019 (AGU Grant 697. 935/2019), TED 01/2021 from the National Secretariat for Social Assistance-SNAS/DGSUAS/CGRS for the SISTER City Project-Safe and Effective Real-Time Intelligent Systems for Smart Cities (Grant 625/2022), the “Project Control and Unification System for the Federal District Government-Sispro-DF” Project (Grant 497/2023), the Dean of Research and Innovation-DPI/UnB and FAP/DF, the Brazilian National Service for Industrial Learning (SENAI) of the National Confederation of Industrial (CNI).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Relevance articles extracted from WoS classified by category.

Category	Title
Artificial Intelligence	Brave New World: Service Robots in the Frontline
	Industry 5.0—A Human-Centric Solution
	Agriculture 4.0: Broadening Responsible Innovation in an Era of Smart Farming
	Industrial Internet of Things: Recent Advances, Enabling Technologies, and Open Challenges
	From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges

Table A1. Cont.

Category	Title
Efficiency	Impact of Energy Conservation Policies on the Green Productivity in China's Manufacturing Sector: Evidence from a Three-stage DEA Model
	The Role and Impact of Industry 4.0 and the Internet of Things on the Business Strategy of the Value Chain-The Case of Hungary
	Energy and CO ₂ Emissions Performance in China's Regional Economies: Do Market-oriented Reforms Matter?
	Modeling the Role of Environmental Regulations in Regional Green Economy Efficiency of China: Empirical Evidence from Super Efficiency DEA-Tobit Model
	Enhancing Microalgal Biomass Productivity by Engineering a Microalgal-Bacterial Community
Industry 4.0	State-of-the-art in Surface Integrity in Machining of Nickel-based Super Alloys
	The Cost of Additive Manufacturing: Machine Productivity, Economies of Scale, and Technology-Push
	The Link between Industry 4.0 and Lean Manufacturing: Mapping Current Research and Establishing a Research Agenda
	Multi-response Optimization of Minimum Quantity Lubrication Parameters using Taguchi-based Grey Relational Analysis in Turning of Difficult-to-Cut Alloy Haynes 25
	A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective
Innovation	Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs
	Influence of Tribology on Global Energy Consumption, Costs, and Emissions
	Understanding the Implications of Digitization and Automation in the Context of Industry 4.0: A Triangulation Approach and Elements of a Research Agenda for the Construction Industry
	Environmental Regulation and Competitiveness: Empirical Evidence on the Porter Hypothesis from European Manufacturing Sectors
	Different Types of Environmental Regulations and Heterogeneous Influence on Green Productivity: Evidence from China
Sustainability	Natural Fiber Reinforced Polymer Composites in Industrial Applications: Feasibility of Date Palm Fibers for Sustainable Automotive Industry
	Nanotechnology in Sustainable Agriculture: Recent Developments, Challenges, and Perspectives
	'Green' Productivity Growth in China's Industrial Economy
	Sustainable Manufacturing in Industry 4.0: An Emerging Research Agenda
	Can Carbon Emission Trading Scheme Achieve Energy Conservation and Emission reduction? Evidence from the Industrial Sector in China

References

- Bulk, L.; Bouzembrak, Y.; Gavai, A.; Liu, N.; Heuvel, L.; Marvin, H. Automatic classification of literature in systematic reviews on food safety using machine learning. *Curr. Res. Food Sci.* **2022**, *5*, 84–95. [[CrossRef](#)] [[PubMed](#)]
- Shafay, M.; Ahmad, R.; Salah, K.; Yaqoob, I.; Jayaraman, R.; Omar, M. Blockchain for deep learning: Review and open challenges. *Clust. Comput.* **2023**, *26*, 197–221. [[CrossRef](#)] [[PubMed](#)]
- Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–4 June 2019; pp. 4171–4186.
- Sarker, I. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)] [[PubMed](#)]
- Islam, M.; Ahmed, M.; Barua, S.; Begum, S. A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Appl. Sci.* **2022**, *12*, 1353. [[CrossRef](#)]
- Lee, K.; Yoo, J.; Kim, S.; Lee, J.; Hong, J. Autonomic machine learning platform. *Int. J. Inf. Manag.* **2019**, *49*, 491–501. [[CrossRef](#)]
- Watanabe, W.; Felizardo, K.; Candido, A., Jr.; Souza, E.; de Souza, É.F.; de Campos Neto, J.E.; Vijaykumar, N.L. Reducing efforts of software engineering systematic literature reviews updates using text classification. *Inf. Softw. Technol.* **2020**, *128*, 106395. [[CrossRef](#)]
- Cohen, A.; Hersh, W.; Peterson, K.; Yen, P. Reducing Workload in Systematic Review Preparation Using Automated Citation Classification. *J. Am. Med. Inform. Assoc.* **2006**, *13*, 206–219. [[CrossRef](#)] [[PubMed](#)]

9. Derviş, H. Bibliometric analysis using bibliometrix an R package. *J. Scientometr. Res.* **2019**, *8*, 156–160. [[CrossRef](#)]
10. Qaiser, S.; Ali, R. Text mining: Use of TF-IDF to examine the relevance of words to documents. *Int. J. Comput. Appl.* **2018**, *181*, 25–29. [[CrossRef](#)]
11. Chebil, W.; Wedyan, M.; Alazab, M.; Alturki, R.; Elshaweesh, O. Improving Semantic Information Retrieval Using Multinomial Naive Bayes Classifier and Bayesian Networks. *Information* **2023**, *14*, 272. [[CrossRef](#)]
12. Nakkiran, P.; Kaplun, G.; Kalimeris, D.; Yang, T.; Edelman, B.; Zhang, F.; Barak, B. Sgd on neural networks learns functions of increasing complexity. *arXiv* **2019**, arXiv:1905.11604.
13. Güven, I.; Şimşir, F. Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Comput. Ind. Eng.* **2020**, *147*, 106678. [[CrossRef](#)]
14. Leong, W.; Bahadori, A.; Zhang, J.; Ahmad, Z. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *Int. J. River Basin Manag.* **2021**, *19*, 149–156. [[CrossRef](#)]
15. Priyanka; Kumar, D. Decision tree classifier: A detailed survey. *Int. J. Inf. Decis. Sci.* **2020**, *12*, 246–269. [[CrossRef](#)]
16. Wang, F.; Wang, Q.; Nie, F.; Li, Z.; Yu, W.; Ren, F. A linear multivariate binary decision tree classifier based on K-means splitting. *Pattern Recognit.* **2020**, *107*, 107521. [[CrossRef](#)]
17. Alhakeem, Z.; Jebur, Y.; Henedy, S.; Imran, H.; Bernardo, L.; Hussein, H. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. *Materials* **2022**, *15*, 7432. [[CrossRef](#)] [[PubMed](#)]
18. Howell, D.C. *Statistical Methods for Psychology*; PWS-Kent Publishing Co.: Worcester, UK, 1992.
19. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
20. Dalianis, H.; Dalianis, H. Evaluation metrics and evaluation. In *Clinical Text Mining: Secondary Use of Electronic Patient Records*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 45–53.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.