Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 6 (2022) 1006 - 1013 ISSN Media Electronic: 2580-0760

The Role of Genetics in Domestic Research on Forestry Issues: A Text Mining Analysis

Titis Hutama Syah
Program Studi Kehutanan, Sekolah Tinggi Pertanian Kutai Timur
titis@stiperkutim.ac.id

Abstract

As a megadiverse country, Indonesia has plentiful genetic resources. The interest of domestic researchers in it and its relation to forestry scope is the focus of this paper. The objective is to determine the genetics aspects represented in forestry scholarly articles. Text mining analysis is carried out for the abstract articles, followed by topic modeling and trend analysis. Python libraries were used to conduct this research. Garuda website was the main source of the data collection. Natural language Toolkits (NLTK) were used to retrieve article information from Garuda. Sci-kit learn (SKLearn) of Latent Dirichlet Allocation module was used for topic modeling analysis, and pyLDAVis was used to represent it. SKLearn was also used for trending analysis. After article text retrieval, three topic clusters were found: forest diversity, products, and land use. The topics were scattered in 1966 abstract articles that were found during data retrieval. Article growth showed the quadratic pattern known after regression analysis. The trend showed the rapid growth of topics and scholars' interest, but the number of articles was low compared to the total articles on the Garuda portal.

Keywords: data science, forestry issue, genetic diversity

1. Introduction

Science has recently rapidly grown with the support of increasingly advanced information and communication technology. New information and knowledge are easy to find by accessing the internet on various gadgets. In supporting the dissemination of knowledge, the Indonesian government encourages its scientists to publish their research results in online scientific papers. Likewise, knowledge related to biodiversity based on research findings by Indonesian scholars should be disseminated online. Biodiversity itself consists of species variations, ecosystem variations, and genetic variations [1]. Moreover, Biodiversity-Based Forest Management in development planning remains the focus of Indonesia's attention in the years to come [2]. The genetic aspects of scholarly papers on forestry science are discussed further in this paper.

Scientific studies in Indonesia have grown rapidly in the last decade [3]. This fact is driven by national regulations encouraging scientists to document their studies in online publications [4]. All publications must be indexed online through SINTA as a Web-based research information system. SINTA itself is currently synchronizing its database from several sources. The main source is Google Scholar to import all online researcher documents. And then, Scopus collects information on Indonesian researchers who publish scientific articles on the international level. The publication manager has to publish their journal articles online and must be recognized by Arjuna as an online accreditation system [5]. Arjuna currently only assesses national publications that have been indexed by Garuda [6]. Garuda is a mandatory platform for scholarly publications in Indonesia. Thus, all publishers have to direct all their online articles to be indexed [7]. In other words, Garuda is an online place to collect all scholarly articles from domestic authors and publishers in Indonesia, including scientific papers related to genetics and forestry, which are the subject of this paper.

Text mining analysis can be used to extract the essence of informative dissemination in scholarly documents. Text mining can be defined as an activity to extract information and patterns from unstructured big data that will reveal its implicitness, unknownness, and its value to society. Text mining differs from data mining. Data mining deals with structured data, but text mining deals with unstructured data [8]. Thus, it is hoped that genetic information from the forestry sector disseminated

Accepted: 05-07-2022 | Received in revised: 11-11-2022 | Published: 29-12-2022

through Garuda's articles can be used to reveal its status within the scope of scientific development.

Google Scholar is a search engine often used by scholars worldwide to get information. Although Google Scholar is the main source of scientific indexes in Indonesia, using it for large-scale data analysis is impractical in the absence of APIs that supports data collection [9]. Whether Scopus has APIs support, it imposes strict restrictions on the number of articles that can be crawled [10]. It is the main obstacle in conducting text analysis on the targeted subjects in this paper. Therefore, Garuda was chosen for crawling Indonesian scholarly articles, although it was limited to articles published domestically.

Text analysis is usually conducted by text mining from a web page, then cleansing and reducing textual data noise. Thus, the text has to convert as vectors like frequent words, phrases, or grammatical structures. The built vector is commonly arranged as a Bag-of-Words (BOW), diagram of categorization, and similarity measurement of classification. Statistical analysis is commonly used to learn the models of textual data classification [8]. These steps have been used in forestry and genetics studies, but no publications reveal the role of genetics in research on forestry issues, particularly in Indonesian research environments.

Previously, a text analysis of several forestry journal articles had been carried out. Dobbertin and Nobis [11] collected research articles between 1979-2008 to determine contemporary issues' trends from the socio-economic aspect. A text mining analysis conducted by Schober et al. [12] revealed three phases of changes in sustainable forest management research issues between 1990-2016, which were the phase of exploration of general concepts of forest management, forestry-focused issues, and climate change-focused issues. Polinko and Coupland [13] revealed that there was a shift in research from issues based on sustainability to the broader term of forestry during analyzing research articles published between 1966-2019. Zhao et al. [14] have applied data mining techniques to find out trends in research topics in the forestry sector. More broadly, text analysis has also been used in various research on forestry in online media. Text analysis conducted by Wesgate et al. [15] revealed a research gap between genetics and ecosystem management issues, where genetics is a surrogated issue with other indicators. In the same article, Westgate et al. also warn to be more careful in conducting text analysis, although it can provide complementary information for researchers. Jaric et al. [16] have applied data mining techniques to find out the vernacular names of the species on online media.

A study of Indonesia's genetic biodiversity has also been carried out. Latifa [17] reveals that access to genetic research is time-consuming, more expensive, and lacks modern approaches. Rintelen et al. [18] also, reveal a growing concern about genetic biodiversity in the research community. The challenges in preserving forest genetic resources are also formidable. Geburek and Konrad [19] stated that to remove obstacles to preserving genetic diversity, it is necessary to take a political approach and provide significant information to policymakers so that it can be considered a vital policy. Raum et al. [20] stated that text analysis on online sources provides an effective tool to support researchers and other stakeholders in various fields and countries, to conduct policy analysis, especially in forest policy making. The question is, "what can domestic researchers do to face all those obstacles?". In this paper, the efforts of researchers to raise the issue of genetic diversity in the broad sense of forestry in their writings are attempted to be revealed, with the main question being: "What are forestry issues represented in domestic scholarly articles that utilize genetic knowledge?".

Based on the information previously mentioned, this paper aims to determine the genetic aspects represented in forestry scientific articles by domestic researchers. In the forestry sector, web crawling-based research has been carried out with a different focus. The genetic issue is the focus of this paper because it is one of the essential forest components of forestry science that needs to be preserved for future purposes. However, how it is currently developing from the perspective of Indonesian scholars is interesting to study more deeply.

2. Research Methods

2.1 Data collection

The text mining approach was carried out during this research. The latest data was taken from https://garuda.kemdikbud.go.id/, and the last data collection was carried out on May 25, 2022. The data was collected using the python programming language. The text is mined using the Beautifulsoup module [21]. The retrieved data was a list of authors, titles, abstracts, and journal and publisher details.

Since the articles indexed by Garuda were more than two million articles, filtering was conducted using the subject contained in Garuda itself, which were subject groups of Biochemistry, Genetics, and Molecular Biology, also subjects groups of Agriculture, Biological Sciences, and Forestry. To focus deeply on the main subjects, Genetics and Forestry, retrieved articles were selected using keywords related to those two subjects. The final data selection was to ensure that the articles retrieved were domains of forestry science. Figure 1 shows the steps taken in data collection.

2.2 Text preprocessing

The selected data to be processed was the information represented in the abstract. This process was carried out

DOI: https://doi.org/10.29207/resti.v6i6.4228

to convert words or sentences in ordinary language into readable encoded programming. Some python modules were required to perform this step. The main module was NLTK (Natural Language Toolkit), a natural language processing program. The selected data to be processed was the information represented in the abstract. This process was carried out to convert words or sentences in ordinary language into readable encoded programming formats. This module was used to classify, tokenize, stem, tag, reason, and parse the text represented [22]. An additional module to run NLTK in the Indonesian language was Sastrawi, which was used to stem Indonesian words to their base form [23]. However, retrieved abstracts were represented in both English and Indonesian. Therefore, the preprocessing included both languages.

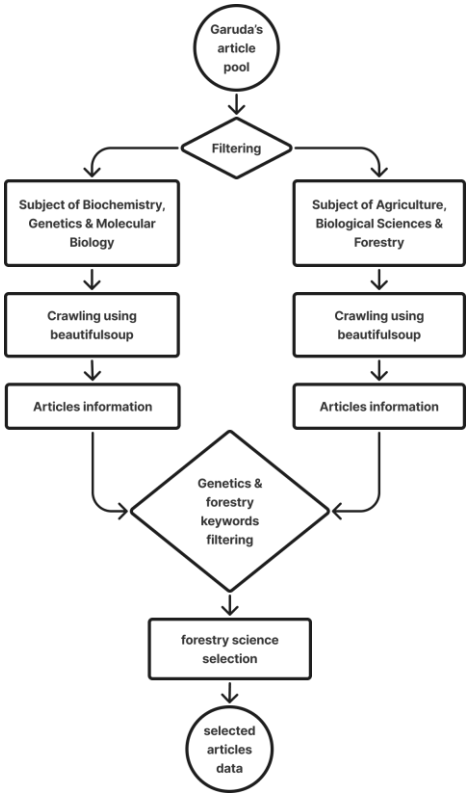

Figure 1. Data retrieval workflow from Garuda website.

Preprocessing text resulted in words in their base form. This word's format was set up for further analysis. To achieve this research's goals, the word format needed to be converted into vectors for topic modeling, which was Latent Dirichlet Allocation (LDA) modeling. This modeling aimed to determine the general topic of the represented abstract in all retrieved articles. The built topic attempted to conclude the domestic scholars' research trend.

2.3 LDA Modelling

The topic model was built according to the LDA modeling described by Blei et al. [24]. The model is built using the Sci-kit learn module of the Python library [25] and visualized using LDAVis, initially built for R statistical computing [26]. The used source code was Sklearn decomposition Latent Dirichlet Allocation script [27] and pyLDAvis for its visulizations [28]. This code was used to represent text from collected abstracts as semantic vectors. The algorithm of this module automatically discovered the text's semantic structure and statistically examined the corpus co-occurrence. The co-occurrence corpus was used to represent the topic to be concluded as trends of researchers' interest in genetics and forestry issues.

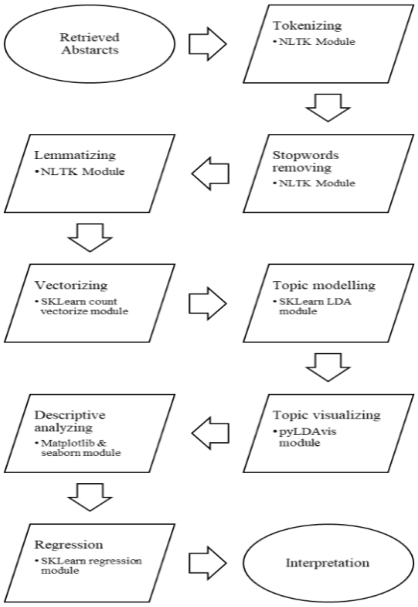

Figure 2. Workflows of research procedure and main python module used.

The scripts and code mentioned above as Python modules could not be run independently. Therefore, several other modules were also used to achieve the research objectives. Pandas, NumPy, Seaborn, and Matplotlib modules are used. Those modules created basic code, such as statistical description, data management, and data display. Those modules were important to represent supporting details, such as authors' distributions, articles, and topics. Therefore, a descriptive analysis could be done. The main procedures and modules used are shown in Figure 2.

3. Results and Discussions

Based on their abstract articles, genetics in the forestry sector remained less concern to Indonesian scholars as a megadiverse country. It was proved by data collection retrieved that domestic scholars in their publications focused only on 1966 articles. Figure 3 shows some datasets of python notebook screen capture of retrieved data.

The number of articles indexed by Garuda was more than 2 million articles. Meanwhile, data retrieved by selected subjects were 198,673 articles. The number of

DOI: https://doi.org/10.29207/resti.v6i6.4228

genetics articles in the forestry sector remained small in those two subjects, less than 1%. Table 1 represents the data in more detail.

Table 1. Distribution of retrieved articles

Pool	Number of articles	Approximate of total articles
Garuda's database	> 2 million	-
Biochemistry, Genetics and Molecular Biology subject	89,597	4.47%
Agriculture, Biological Sciences and Forestry subjects	109,086	5.45%
Selected subjects	198,673	9.93%
Genetics in forestry sector	1966	< 1%
Duplicates articles	10	< 0.001%

Although articles in the selected subjects accounted for almost 10% of the total, it was found that not all articles represented in this study had been ensured to have subjects suitable to the article scope.

The selected articles involved 4,437 scholars. The last decade has shown rapid growth. As represented in Figure 4, the authors' growth increased as the articles grew. The figure also shows that the authorship of each article might involve more than one author. The earliest article was written in 1974, and the number of articles grew flatty until the early 2000s. After that, its number surged. It may be caused by the risen of computer and informatics technologies. The previously paper-printed articles began to be digitalized. The number of articles may increase because there might be not all old articles have been converted into digital format. Moreover, there may also be any reluctance from publishers to index their journals into Garuda. Since 2012, there has been a very rapid growth in article numbers. It was spurred by government policies requiring scholars to write and publish articles online [30].

```
In [1]: import pandas as pd

In [2]: pd.read_excel('dataset.xlsx').drop(columns={'Unnamed: 0'})

Out[2]:
```

	Title	Author(s)	Abstract	Issue
0	PENDUGAAN STOCK KARBON TERSIMPAN PADA MODEL MA...	[Saputra, Hendra, Hardiansyah, Gusti, Fahrizal...	The estimation of carbon stocks is very import...	JURNAL HUTAN LESTARI Vol 6, No 1 (2018): JURNA...
1	Analisis Karakter Fenotip Beberapa Spesies Den...	[Ainiah, Nur, Rahayu, Tintrim, Hayati, Ari]	Phenotype character of orchid Dendrobium is kn...	BIOSAINTROPIS (BIOSCIENCE-TROPIC) Vol 5 No 2 (...
2	Genetic Parameter Estimates for Growth in a Pr...	[B. Hardiyanto, Eko]	Uji keturunan sengon penyerbukan terbuka (half...	0853-8670
3	ESTIMASI LAJU PERKAWINAN SILANG PADA Toona sin...	[Jayusman, Jayusman, Na'iem, Muhammad, Indriok...	ABSTRACTToona sinensis Roem or surian, belongs...	Journal Penelitian Kehutanan FALOAK Vol 2, No ...
4	JENIS TUMBUHAN DAN TIPE HABITAT DI HUTAN DATAR...	[Jitmau, Marthen, Rumbino, Amon A]	ABSTRACTMamberamo lowland areas are generally ...	VOGELKOP: Jurnal Biologi Vol 1, No 1 (2018)
...	...	...	...	...
1961	The Conservation Efforts of Wood Apple, An Ide...	[Rahayu, Enni Suwarsi, Dewi, Nur Kusuma, Herli...	Feronia limonia (.) Swingle better known as w...	Biosaintifika: Journal of Biology & Biology Ed...
1962	ANALISIS PERILAKU HARIAN HARIMAU SUMATERA (Pan...	[Tarigan, Kiandreas, Nurmawan, Wawan, Langi, M...	AbstractMedan Wildlife Park is conservation ...	COCOS Vol 5, No 5 (2021)
1963	Variasi Morfologi Empat Spesies Jati (Tectona...	[Fauzi, Mohamad Anis, Hasna, Tri Maria, Setiad...	Jati (Tectona grandis Linn.) dikembangkan sec...	Biota : Jurnal Ilmiah Ilmu-Ilmu Hayati Vol 5, ...
1964	Eksplorasi Plasma Nutfah Tanaman Pangan di Pro...	[Sri Astuti]	AbstractWest Kalimantan is are area of tropica...	Buletin Plasma Nutfah Vol 10, No 1 (2004): Juni
1965	PENAKSIRAN POTENSI KAYU DARI HUTAN RAKYAT DI K...	[La Ode Midi, La Ode Agus Salim Mando]	Penelitian ini bertujuan untuk menaksir potens...	Jurnal Ecogreen Vol 1, No 1 (2015)

1966 rows × 4 columns

Figure 3. Screen capture of python notebook of retrieved data.

belonged to those subjects. Some of the retrieved articles were completely unrelated to the selected subjects. In categorizing their journal subjects into the Garuda database, it was presumed that the journal managers did not ensure the scope of their articles correctly. It is presumed to manipulate the search engine system and can be categorized as unethical search engine optimization tricks, as it can influence journals or publishers of webometrics ranks. Although Search engine companies have tried to prevent it, web appearances are manipulated easily as 50% of weightage ranks are influenced by the visibility of web appearances, such as external links [29]. In this case, subject manipulation increases the visibility of search engine optimization. However, the selected articles

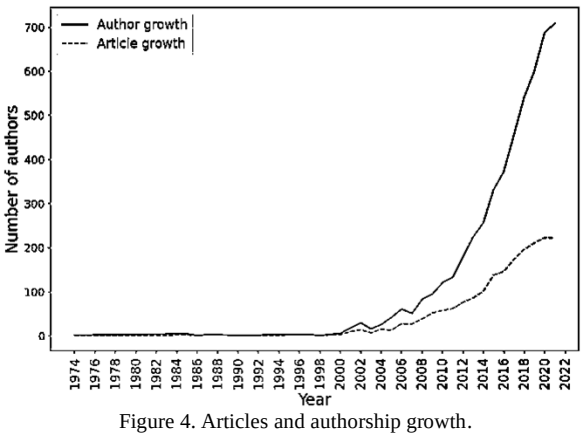

Figure 4. Articles and authorship growth.

DOI: https://doi.org/10.29207/resti.v6i6.4228

- The Topics

This research reveals that domestic researchers tend to be interested in the genetics of forestry sector issues within three topic clusters. These clusters were concluded after the LDA model was conducted. The clusters are represented based on the principal components after hyper parameters adjustment. Figure 5 shows that the principal components distributed each topic in separate groups. That was the reason for clusters being concluded. The cluster is determined by calculating the distance between topics that was projecting intertopic distance based on multidimensional scaling to determine the relative distances of each object based on its similarity [31]. Multidimensional scaling is used to construct a pairwise configuration distance between objects. It is a data reduction method used to discover the space dimension based upon a proximity matrix of distances. Principal Component Analysis (PCA), used to determine clusters in this paper, is a multidimensional scaling technique used to find a set of dissimilarities proximity matrix [32].

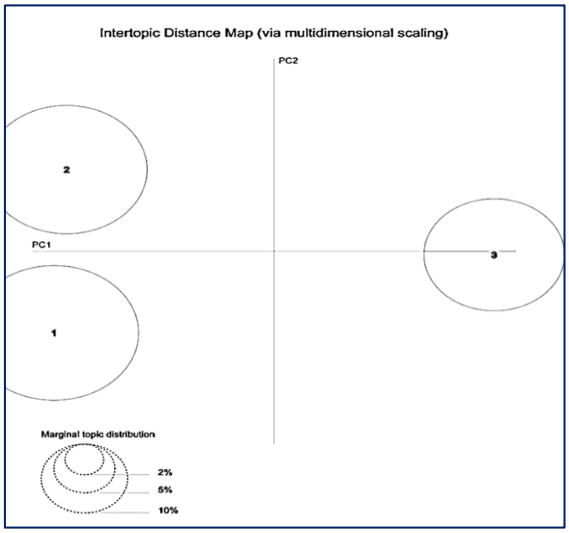

Figure 5. Visual representation of principal components of the topi

Table 2. Relevant issues of co-occurrence words

Cluster	Top 10 relevant words	Relevant issue
Topic 1	species forest area diversity research plants found plant used trees	forest diversities -
Topic 2	hutan teliti hasil kayu jenis tanam masyarakat tingkat tumbuh milik (*forest research product wood species plant community level growth owner*)	forest products
Topic 3	forest research land used species using results conservation data showed	forest land uses

Figure 6 shows the most salient term, the words that frequently occur in the retrieved scholarly abstracts. Meanwhile, the topic cluster contains a set of relevant co-occurrence words in each topic. The details are represented in Table 2. The co-occurrence words were then concluded based on the issues of forestry terms. The number of topics is represented in Figure 7. Those results indicate that the genetics issues are also used for research outside the scope of genetics itself. As part of biodiversity, genetics deals with its diversity. The scope of genetic diversity deals with taxonomy, origin and evolution, and conservation-based selection. Other elements affecting confer value of genetic diversity are population integrity, environment and ecosystem support, and its relationships within the ecosystem [33]. The built topic was only topic 1, which discussed genetic diversity. The rest discuss the elements supporting its diversity. Therefore, it concluded that most forestry scholars are interested in discussing genetics to describe its supporting elements, but not about genetics itself.

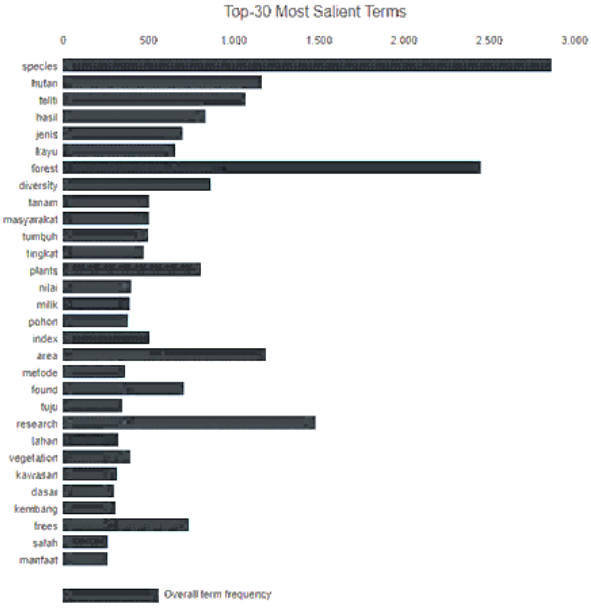

Figure 6. Top 30 relevant terms of the topics.

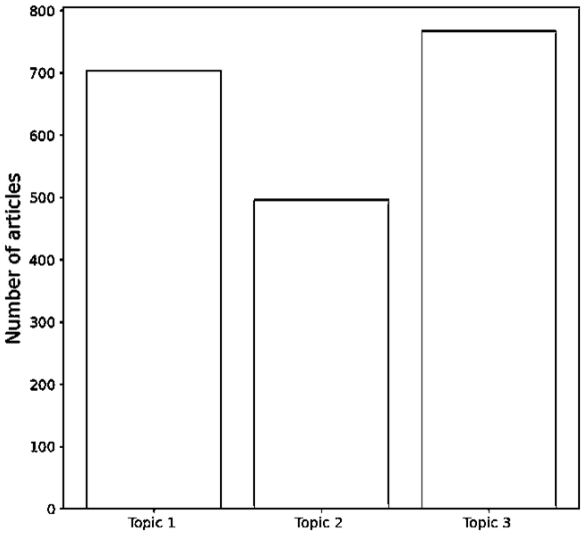

Figure 7. Number of articles for each topic

DOI: https://doi.org/10.29207/resti.v6i6.4228

Scholars' interest in each topic also rapidly increase. Figure 8 is represented the increment of scholars' interest. Topics 1 and 3 coincided with each other throughout time. Meanwhile, Topic 2 has fewer scholars interested in it and started to decline recently. On the contrary, as represented in Figure 9, the scholar's interest in topics 1 and 3 has become more focused in recent years. The darker color on the heatmap indicates that scholars' studies are more intensive.

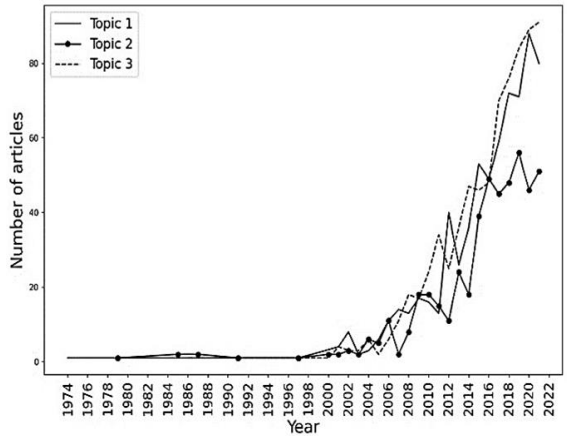

Figure 8. Topic distribution through the years

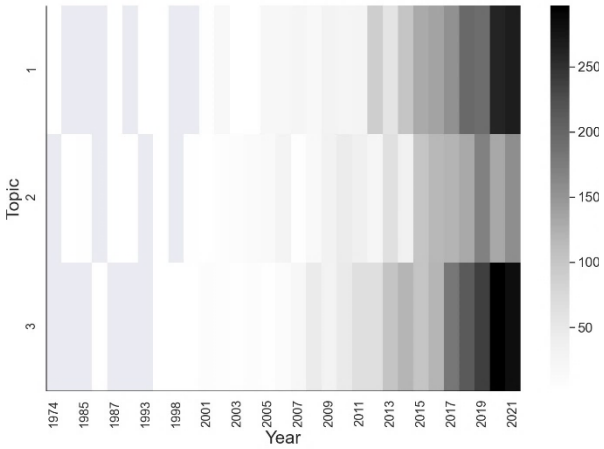

Figure 9. Heatmap of authors' interest

Scholars in forest land use research use genetics issues because land change directly impacts biodiversity decline. The declining biodiversity is mostly due to overexploitation and agriculture. Over 8000 species are threatened by overexploitation, such as logging, hunting, fishing, and plant gathering. Meanwhile, agricultural activities such as crop farming, livestock farming, timber plantation, and aquaculture have the same effect on these species. However, other activities such as urban development, pollution, and climate change also cause a decrease in biodiversity [34]. Institutional and policy factors mainly cause forest land-use changes in Indonesia. The lack of institutional capacity to control property rights and transmigration issues has led to more massive forest land conversion, such as increased oil palm plantations that have been converted from the forest area [35]. Naturally, land-use study always coincided with genetics which is affected by its diversity, as is the case with ecosystem diversity. Forest land uses itself surrounded by many massive issues.

Forests have remained a deposit of major biodiversity for human livelihoods. Forestry products are widely used in wild or domesticated forms. It produces most of the human needs, inseparable from its diversity. The products came from plants, animals, aquatic creatures, and trees. Unfortunately, it is conducted selectively and causes genetic uniformity, likely decreased livelihood survival [36]. That is why the study of forest products could be related to genetics, as obtained in Topic 2 of this study.

- The Trends

As marked as the Indonesian scholarly article revolution began, the recent decade is set to be analyzed. The surge of topics 1 and 3 have had a positive trend in the recent decade. The regression model, as represented in Figure 10 for each topic, has a very strong R^2 value of more than 0.9. it means that the model can represent the topic trends. Robustly, the Root Mean Square Error (RMSE) value indicates the best-fit prediction of the polynomial regression as quadratics trends. All topics increment on this subject also tends to grow. The model was fitted using regression through the origin, which did not involve an intercept in the model (uncentered model). It was regarded to the statement of Eisenhauer in his article [37]. Although there were a lot of controversies, his model was used to answer the curve trends that its ordinary models could not handle. It was done to reduce the standard error in the ordinary model and get a more fitted model with the observed trends, as represented in Table 3.

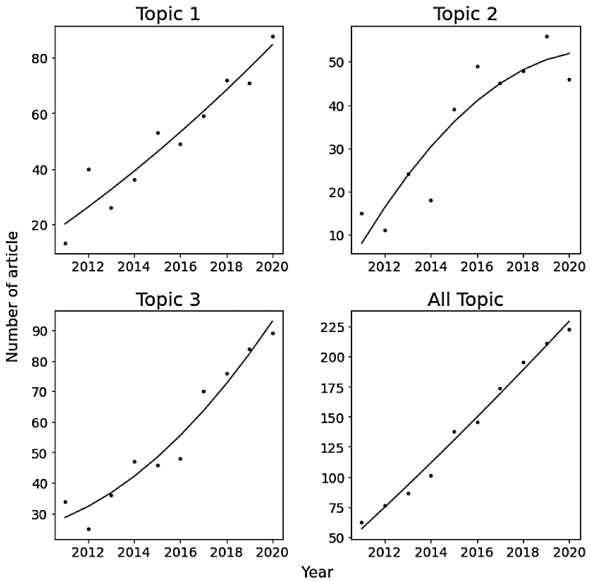

Figure 10. Quadratics trends of each topic in the recent decade

.

DOI: https://doi.org/10.29207/resti.v6i6.4228

Table 3. Comparison of the best-fit polynomial regression model (centered vs. uncentered)

		Intercept	x_1	x_2
(Centered)				
Topic 1 (R^2=0.9)	Constant	5.55E+05	-557.7568	0.1402
	s.e	1.36E+06	1352.568	0.336
	P>|t|	0.696	0.692	0.689
Topic 2 (R^2=0.8)	Constant	-1.75E+06	1730.3	-0.428
	s.e	1.28E+06	1272.735	0.316
	P>|t|	0.215	0.216	0.217
Topic 3 (R^2=0.9)	Constant	1.79E+06	-1779	0.4432
	s.e	1.04E+06	1028.37	0.255
	P>|t|	0.128	0.127	0.126
All Topic (R^2=0.9)	Constant	5.92E+05	-606.8	0.1553
	s.e.	1.30E+06	1290.717	0.32
	P>|t|	0.663	0.653	0.642
(Uncentered)				
Topic 1 (R^2=0.9)	Constant	-	-7.144	0.0036
	s.e	-	0.803	0.000
	P>|t|	-	<0.01	<0.01
Topic 2 (R^2=0.9)	Constant	-	-4.843	0.0024
	s.e	-	0.841	0.000
	P>|t|	-	<0.01	<0.01
Topic 3 (R^2=0.9)	Constant	-	-7.092	0.0035
	s.e	-	0.72	0.000
	P>|t|	-	<0.01	<0.01
All Topic (R^2=0.9)	Constant	-	-19.08	0.0095
	s.e	-	0.769	0.000
	P>|t|	-	<0.01	<0.01

R^2 is coefficient of determination; P>|t| is p-value or probability value; s.e. is the standard error of parameters; x_1 is variable x to the power of 1; x_2 is variable x to the power of 2.

Scholars' interest in the genetics of forestry issues shows an increasing trend and is predicted to be continued. However, as previously mentioned, the number of scholars interested in this issue is very low. Genetics in this paper emerged as a multidisciplinary approach to forestry issues. It is because the genetic issues differ among scholars' objectives. In addition, the difference in technological knowledge is a differentiator in the scholar's ability to analyze issues. Lack of funding also results in the issue of having very few devotees within the research community. Those have been world problems for these 50 years [38]. Nevertheless, this study shows rapid growth in these issues. It should be grateful, and some action is needed to solve the abovementioned problems.

This research seeks to uncover hidden but essential aspects of biodiversity, one of the three crucial components of diversity: genetics. This aspect can be said to be overlooked and covered by other issues that directly impact the forestry sector. Economic aspects, sustainability and climate change that have been studied previously are issues that have attracted much interest from forestry scholars [11][12][13]. Especially in Indonesia, limited funding and regulatory restrictions in sharing research resources with the global community are the main obstacles of genetics research [17]. This paper reveals that the number of genetics scholarly publications in the forestry sector is deficient, although it shows rapid development in the last decade. The issues raised are also not entirely about genetics, but its use is dominated by supporting information for written articles.

4. Conclusion

Garuda is the main pool of indexed journals in Indonesia, although domestic publisher only. This research was conducted on scholarly articles on genetic issues from the perspective of forestry, which turned out to be very limited (less than 1%). Text mining to the abstract has been carried out and clustering issues into three topics: forest diversity, forest products, and forest land use. There has been an increasing trend of author interest in the recent decade. However, it is predicted that interest in the topic of genetics in forest products issues is predicted to be declined as polynomial regression is conducted to its trends.

Several other studies have been conducted using data mining and web crawling technologies to determine the science growth. In forestry science, these technologies have been used to determine the development of sosio-economic aspect, forest sustainability, and climate change as the main issues in recent years. Meanwhile, genetics is still a less interesting study. This paper attempts to reveal the scholars' interest in the genetics issues used in forestry research and finds that genetics is used for broader multidisciplinary topics than genetics itself.

Acknowledgment

Thanks to all python and data science community for sharing their source code online and making this research possible.

Reference

[1] U. S. Rawat and N. K. Agarwal, "Biodiversity: Concept, threats and conservation," *Environ Conserv J*, vol. 16, no. 3, pp. 19–28, Dec. 2015, doi: 10.36953/ECJ.2015.16303.

[2] Kementerian Lingkungan Hidup dan Kehutanan, *Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.16/Menlhk/Setjen/Set.1/8/2020 Tentang Rencana Strategis Kementerian Lingkungan Hidup Dan Kehutanan Tahun 2020-2024*. Indonesia, 2020. Accessed: Jun. 27, 2022. [Online]. Available: http://jdih.menlhk.co.id/uploads/files/P_16_2020_RENSTRA_KLHK_menlhk_08312020090150.pdf

[3] "Paper per Subject per Lecturer (FTE) per Year (x100)," *SINTA - Science and Technology Index*. https://sinta.kemdikbud.go.id/wcu/research_output2 (accessed Jun. 27, 2022).

[4] Lukman, S. S. Ahmadi, W. Manalu, and D. S. Hidayat, *Pedoman Publikasi Ilmiah 2019*. Jakarta: Direktorat Pengelolaan Kekayaan Intelektual Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2019.

DOI: https://doi.org/10.29207/resti.v6i6.4228

[5] Kementerian Pendidikan Kebudayaan Riset dan Teknologi Republik Indonesia, "SINTA - Science and Technology Index," *Frequently Asked Questions*, 2020. https://sinta.kemdikbud.go.id/home/faq (accessed Jun. 27, 2022).

[6] *Panduan Akreditasi Jurnal Nasional (ARJUNA)*. Direktorat Pengelolaan Kekayaan Intelektual Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional, 2019. Accessed: Jun. 27, 2022. [Online]. Available: https://arjuna.kemdikbud.go.id/files/download/Manual_Book_Arjuna_(Pengusul).pdf

[7] "Garuda - Garba Rujukan Digital," *Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi*. https://garuda.kemdikbud.go.id/ (accessed Jun. 27, 2022).

[8] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text Mining in Big Data Analytics," *Big Data and Cognitive Computing 2020, Vol. 4, Page 1*, vol. 4, no. 1, p. 1, Jan. 2020, doi: 10.3390/BDCC4010001.

[9] H. Else, "How I scraped data from Google Scholar," *Nature*, Apr. 2018, doi: 10.1038/D41586-018-04190-5.

[10] "Elsevier Developer Portal." https://dev.elsevier.com/api_key_settings.html (accessed Jun. 27, 2022).

[11] M. K. Dobbertin and M. P. Nobis, "Exploring research issues in selected forest journals 1979-2008," *Ann For Sci*, vol. 67, no. 8, Dec. 2010, doi: 10.1051/FOREST/2010052.

[12] A. Schober, N. Šimunović, A. Darabant, and T. Stern, "Identifying sustainable forest management research narratives: a text mining approach," *https://doi.org/10.1080/10549811.2018.1437451*, vol. 37, no. 6, pp. 537–554, Aug. 2018, doi: 10.1080/10549811.2018.1437451.

[13] A. D. Polinko and K. Coupland, "Paradigm shifts in forestry and forest research: a bibliometric analysis1," *https://doi.org/10.1139/cjfr-2020-0311*, vol. 51, no. 2, pp. 154–162, 2020, doi: 10.1139/CJFR-2020-0311.

[14] M. Zhao, D. Li, and Y. Long, "Forestry big data platform by Knowledge Graph," *Journal of Forestry Research 2020 32:3*, vol. 32, no. 3, pp. 1305–1314, Apr. 2020, doi: 10.1007/S11676-020-01130-W.

[15] M. J. Westgate, P. S. Barton, J. C. Pierson, and D. B. Lindenmayer, "Text analysis tools for identification of emerging topics and research gaps in conservation science," *Conservation Biology*, vol. 29, no. 6, pp. 1606–1614, Dec. 2015, doi: 10.1111/COBI.12605.

[16] I. Jarić, F. Courchamp, J. Gessner, and D. L. Roberts, "Data mining in conservation research using Latin and vernacular species names," *PeerJ*, vol. 2016, no. 7, p. e2202, Jul. 2016, doi: 10.7717/PEERJ.2202/SUPP-2.

[17] E. Latifa, "Access to Genetics Resources in Indonesia: Need Further Legislation?," *Oklahoma Journal of Law and Technology*, vol. 11, no. 1, 2015, Accessed: Jun. 26, 2022. [Online]. Available: http://digitalcommons.law.ou.edu/okjolthttp://digitalcommons.law.ou.edu/okjolt/vol11/iss1/2www.okjolt.org

[18] K. von Rintelen, E. Arida, and C. Häuser, "A review of biodiversity-related issues and challenges in megadiverse Indonesia and other Southeast Asian countries," *Research Ideas and Outcomes 3: e20860*, vol. 3, pp. e20860-, Sep. 2017, doi: 10.3897/RIO.3.E20860.

[19] T. Geburek and H. Konrad, "Why the Conservation of Forest Genetic Resources Has Not Worked," *Conservation Biology*, vol. 22, no. 2, pp. 267–274, Apr. 2008, doi: 10.1111/J.1523-1739.2008.00900.X.

[20] S. Raum, F. Rawlings-Sanaei, and C. Potter, "A web content-based method of stakeholder analysis: The case of forestry in the context of natural resource management," *J Environ Manage*, vol. 300, p. 113733, Dec. 2021, doi: 10.1016/J.JENVMAN.2021.113733.

[21] L. Richardson, "beautifulsoup4 · PyPI." https://pypi.org/project/beautifulsoup4/ (accessed Jun. 28, 2022).

[22] S. Bird, E. Klein, and Edward Loper, *Natural Language Processing with Python: Analyzing Text with the Natural ... - Steven Bird, Ewan Klein, Edward Loper - Google Buku*. Accessed: Jun. 28, 2022. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=KGIbfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+Language+Processing+with+Python&ots=Y4DlA0FHL2&sig=ovuEK1a_UbZ7WZrCcJZiRuj305E&redir_esc=y#v=onepage&q=Natural%20Language%20Processing%20with%20Python&f=false

[23] H. A. Robbani, "Sastrawi · PyPI." 2021. Accessed: Oct. 26, 2021. [Online]. Available: https://pypi.org/project/Sastrawi/

[24] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003, Accessed: Jun. 29, 2022. [Online]. Available: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=https://githubhelp.com

[25] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[26] C. Sievert, K. S.-P. of the workshop on interactive, and undefined 2014, "LDAvis: A method for visualizing and interpreting topics," *aclanthology.org*, pp. 79–86, 1999, Accessed: Jul. 01, 2022. [Online]. Available: https://aclanthology.org/W14-3110.pdf

[27] M. D. Hoffman, D. M. Blei, and F. Bach, "sklearn.decomposition.LatentDirichletAllocation — scikit-learn 1.1.1 documentation," 2010. https://github.com/blei-lab/onlineldavb (accessed Jul. 01, 2022).

[28] B. Mabey, "pyLDAvis — pyLDAvis 2.1.2 documentation," 2015. https://pyldavis.readthedocs.io/en/latest/readme.html (accessed Jul. 01, 2022).

[29] I. F. Aguillo, J. L. Ortega, M. Fernández, and A. M. Utrilla, "Indicators for a webometric ranking of open access repositories," *Scientometrics*, vol. 82, no. 3, pp. 477–486, Feb. 2010, doi: 10.1007/S11192-010-0183-Y.

[30] G. Ratnasari and R. Nurislaminingsih, "Pengaruh Kebijakan Publikasi Karya Ilmiah Di E-Journal Terhadap Peningkatan Motivasi Menulis Karya Ilmiah Mahasiswa Ilmu Perpustakaan Fakultas Ilmu," *Jurnal Ilmu Perpustakaan*, vol. 6, no. 1, pp. 201–210, 2017, Accessed: Jul. 03, 2022. [Online]. Available: https://ejournal3.undip.ac.id/index.php/jip/article/viewFile/23055/21108

[31] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, pp. 74–77, 2012, doi: 10.1145/2254556.2254572.

[32] N. H. Timm, *Applied multivariate analysis*. Springer, 2002.

[33] V. Ramanatha Rao and T. Hodgkin, "Genetic diversity and conservation and utilization of plant genetic resources," *Plant Cell Tissue Organ Cult*, vol. 68, no. 1, pp. 1–19, 2002, doi: 10.1023/A:1013359015812.

[34] S. L. Maxwell, R. A. Fuller, T. M. Brooks, and J. E. M. Watson, "Biodiversity: The ravages of guns, nets and bulldozers," *Nature*, vol. 536, no. 7615, pp. 143–145, Aug. 2016, doi: 10.1038/536143A.

[35] L. Juniyanti, H. Purnomo, H. Kartodihardjo, and L. B. Prasetyo, "Understanding the driving forces and actors of land change due to forestry and agricultural practices in sumatra and kalimantan: A systematic review," *Land (Basel)*, vol. 10, no. 5, May 2021, doi: 10.3390/LAND10050463.

[36] Food Agriculture Organization of the United Nations, *Harvesting Nature's Diversity: World Food Day, 16 October*. FAO, 1993.

[37] J. G. Eisenhauer, "Regression through the Origin," *Teach Stat*, vol. 25, no. 3, pp. 76–80, Sep. 2003, doi: 10.1111/1467-9639.00136.

[38] C. van Oosterhout, "Conservation genetics: 50 Years and counting," *Conserv Lett*, vol. 14, no. 2, p. e12789, Mar. 2021, doi: 10.1111/CONL.12789.

DOI: https://doi.org/10.29207/resti.v6i6.4228