IOSR Journal of Mobile Computing & Application (IOSR-JMCA)
e- ISSN: 2394-0050, P-ISSN: 2394-0042.Volume 12, Issue 1 (Jan. - Feb. 2025), PP 61-72
www.iosrjournals.org

Development, Underdevelopment: The Synergetic State Of Research Output in Sub-Saharan Africa

Maxwell Scale UwadiaOsagie and Derek Ahamioje
Department of Physical Sciences, Faculty of Science, Benson Idahosa University, Benin City

Abstract:
Research is a determinantfor global impact, and societal development revolves around quality research. The hub of institutional development is based on the fundamental principle of synergetic state of collaboration of sound research tied to industryand other area of human endeavours which has a clear cut of supervise regulations. Sub-Saharan Africais considered to be a developing region and its role in the development of infrastructures and other critical infrastructures has been engendered in the past due to none-synergetic state of research collaboration among researchers, regulating agencies and industry. However, recent studies have asserted that researchers in Sub-Saharan Africa have increased its state of collaboration and this is gear towards actualization of the state of the art infrastructures on the region called Sub-Saharan Africa. Furthermore, the idea of these research collaborationshas been found to be useful owning to the need for quality research collaboration amongst researchers. For validation purpose, this paper create a new framework for analysing regionally-based scientific research output by combining Density-Based Clustering and Thread-Based Aggregation Techniques.To achieve this, the study datasets were classified to be"Title, Citation and Location. Using the LDA model the system was able to mined scientific research output with two or more authors in three fields of studies; Computer Science/Information Technology, Medicine and Engineering for a period of 10 years (2010 – 2020) using goggle scholar as the primary source. For analysis, the Density-Based Spatial Clustering Application with Noise (DBSCAN) was used. The result gotten shown from Figure 4.1 and 4.2 respectively show clear indication of collaborative model that is gear towards developmental state of modern infrastructure in Sub-Saharan Africa.

Keywords: *Clustering, google scholar, Sub-Saharan, Africa, Density-Based, Analysis, Research*

I. INTRODUCTION

In the developed countries, the proliferation of infrastructure and its underneath attributes have been a product of institutional research which on a regular basis experience some policies that constantly make it viable in the scheme of things. Research is key to the development of economy transformation and as classified, research is seen as change agent. One problem the developing region such as Sub-Saharan Africa had faced over the years is the lack of understand behind research collaboration, hence little or no attention has been given in the past.

[1] took time to explain in their work the significant meaning of research and while it must be carefully done. They opined that research is the methodical and rational call for a drastically change in method of thinking in arriving at a new knowledge destination which bothers on specific issue. It involves all aspects of investigation as it relates to the objective and methodical analysis of scientific and social problems aim at producing results. This investigation unveil a hidden part of something which indicate further exposition /clarity on the part of the end users. Beyond the unveiling of hidden part, research create several stages of knowledge driving fact and this could be negating the formal ideal and replacing it with the new fact as substantiated by the new study or adding to the already existing knowledge. At such, research is the only way to make advancement in the body of knowledge in a particular field of study. Societal development and Research has a tin line and it can be said that both represent civilization. Civilization is a social process that validate societal development and advancement. It has capacity to turn around all aspects of economy such as commerce, intellectual and geopolitical prosperity. It is evidential that scientific and other research areas does not only have capacity to change society for the better but it also weaponized to create the needed blueprint for long term planning and sustainability

Research goes beyond science and technology other areas of research such as Languages, literature, history, sociology, accounting, religious study etc. have thus far contributed meaningfully to rapid development among developed and developing nations. The end product of research is to make it proactive, methodical, and systematic process of discovering, interpreting, evaluating of facts, events, behaviours, ideas practical, concept and theories for the purpose of implementing research findings. In research, various methods such as study,

experiments, observation, analysis, comparison, and reasoning are employed. In reality, research can be found all around us.

Its primary goals are to:

I. Learn new information
II. Confirm and test critical information
III. Find out the cause and effect relationship of an event, process, or phenomena
IV. Develop new scientific tools, concepts, and theories to solve and understand scientific and nonscientific problems,
V. Find solutions to scientific, nonscientific, and social problems, and
VI. Develop new scientific tools, concepts, and theories to solve and understand scientific and nonscientific problems to overcome or solve challenges that arise in our daily lives.

Scientific research has been defined by [2] as research conducted in a planned manner, to contribute towards science by systematically collecting, interpreting, and evaluating data

For validation purpose, this paper create a new framework for analysing regionally-based scientific research output by combining Density-Based Clustering and Thread-Based Aggregation Techniques. To achieve this, the study datasets were classified to be Title, Citation and Location. Using the LDA model, the system was able to mined scientific research output with two or more authors in three fields of studies; Computer Science/Information Technology, Medicine and Engineering for a period of 10 years (2010 – 2020) using goggle scholar as the primary source. For analysis sake, the Density-Based Spatial Clustering Application with Noise (DBSCAN) was used.

II. LITERATURE REVIEW

[3] Adams *et al*. (2014) opined that though, there is an incremental state of growth as it relates to Africa's research output, but raised concern on the speed of growth when measured with that of established economy. However, while there is a strong correlation between investment and GDP in Africa,Adam *et al.*, 2014 in their work also believed that some of the continent's wealthier countries are foot-dragging significant investments in knowledge-driven economy?

As shown in the Figure 2.1, 2000 show a progression of 13,271 papers indexed in Thomson Reuters Web of Science for Africa; by 2012, it double more than half which was over 35,000 publications. .

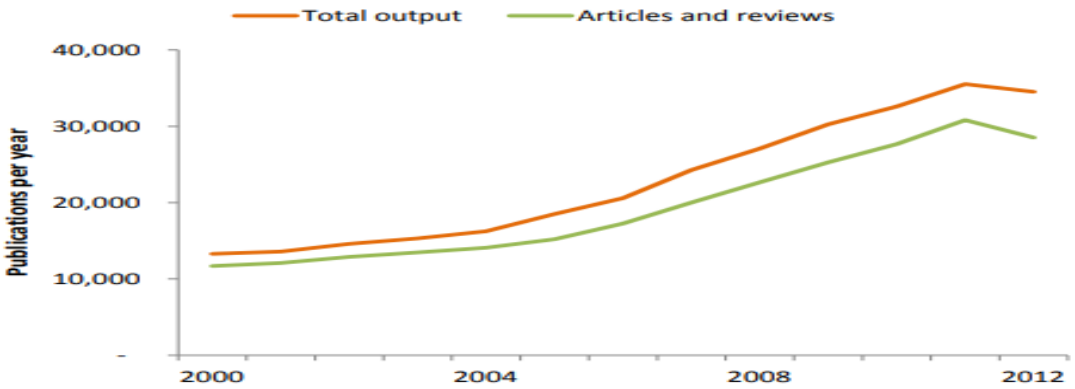

Figure 2.1: Africa's output of publications indexed [3]

[4]; [5]; [6]; [7] in their separates work made clarifications on the distinctive act of research collaboration. Their assertions were that international partnership are favoured over local collaboration because it has been established over the years to have more impact.In a separate findings, [8] stated that while high degrees of regional technical specialization and diversity were negatively related with domestic extra-regional partnerships, technological variety was positively connected with the readiness to collaborate with foreign partners. The fact indicate a cut in the significant role of academics to collaborate internationally rather than domestic synergy.The improvement of technology in recent times has improved quality of learning across the globe and this has brought about far reach distance communication with a commensurate cost [9]. From the point asserted, there is a corresponding result of progress on sustainable development through partnership. One evidential proof is that co-authored articles have been found to be more people oriented and high impact factor on citation perspective [10].

Synergistically, international collaboration has become attractive destination for Africa's researchers and this was evidence in a research work carried out by [11; 12]. To this end, the awareness and curiosity of Africa researcher have risen significantly. In recent times, the community of research and corporate entities are

becoming very much interested in the percentage rate of research collaboration amongst African stateand this is to bring about cohesive state of inter-development. Notwithstanding, [3] previously reported that intraregional and interregional collaboration rates were seen to be low by previous studies. [13] gave a classical example through the study of the Southern African Development Community (SADC), where it was reported that only 5% of the entire SADC articles published between 2005 and 2008 were co-authored by a SADC researcher and another African researcher. This indeed supported the claim of [3] which categorically reported the low rate of within collaboration. In another study by [14], it was unveiled that collaborative research has a fundamental role in the internal politics of Africans as well as the contributing factors that aid understanding of African politics. The world is challenged at the moment and many of the issues cannot be resolved by divided research output rather a more synergetic state of affair in the production of quality research good enough to address these issues is of great concern to policies maker. According to [15], Ministers from African governments met and agreed in Kigali on a Joint Call for Action in March 2014 to implement a pragmatic mechanism with strategic investments in science, technology etc. to exponentially increase Africa's state by developing knowledge-based society within one generation. The representation were those from World Bank's Partnership for Applied Science, Engineering, and Technology (PASET), which helps African governments and their partners in strengthening the role of applied science, engineering, and technology in the development agenda.

III.METHODLOGY FOR DATA CRYSTALLIZATION

In order to create roadmap for the actualization of this work, the author choseGoogle Scholar to assess trends in research growth in Sub-Saharan Africa. While this study acknowledged that metrics based on peer-reviewed research outputs do not fully represent all research activities in Sub-Saharan Africa, it is the most comprehensive and objective analysis base available. Although, prior studies have looked at research output trends in Sub-Saharan Africa, this is the first attempt to use this proposed technique to analyse these datasets. Populating the Sub-Saharan Africa data collection was carried out by collecting data on the total number of articles published in eight (8) nations in the region over 8 years (2010-2020).The information gathered was divided into categories based on the amount of output produced in each country per year. Furthermore, thirty (30) articles were sampled for each year, and for the three Fields of study. These data comprises of;

i. Title of the article
ii. Number of Citation
iii. Number of authors

Figure 3.1 depicts the number of publications published throughout the research period and the sampling locations (countries).

Figure 3.1: Research sampling locations

The method integrates the examination of two types of data (text and citation count) acquired from the same source (Google Scholar). The citation information is sent to the Crowd Dynamics Analyser module, whereas the Threads Discovery module takes text as input. Both modules function separately to discover unusual research activity in the area, believing that this research implies the presence of something uncommon. The Threads Exploration Module handles the content (text) of the article in order to deduce groups of output related to the same subject. Finally, both modules feed the Threads Ranking whose main aim is to check if the detected threads effectively correspond to unusual crowd behaviour. Figure 3.2 below shows the outlook of the two modules as described above.

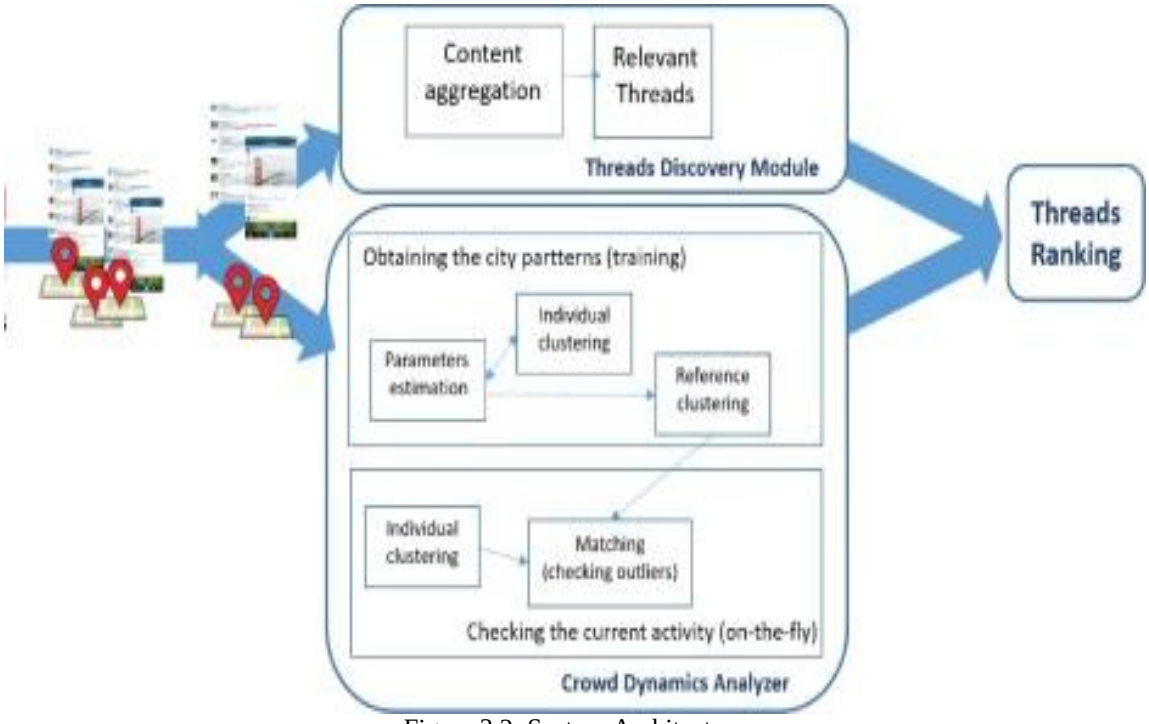

Figure 3.2: System Architecture

Combining clustering algorithms with the number of citations each publication has received over eight years can successfully bring to light discoveries, relevant activities that are tied to specific areas and, by extension, regions. Both modules (density-based clustering and aggregation) are more powerful when used together than when used separately**.** The Crowd Dynamics Analyser is a module that analyses the researchers' activity in terms of publications and citations to detect unusual behaviours in the region under investigation. The method begins by inferring the region's activity pattern, i.e. the typical behaviour (number of citations). This is, in fact, the primary goal of the training phase of our technique, and one of the features that distinguished the approach from others are literatures that do not offer the number of clusters.The thread finding module examines common trends by examining dominating subjects as evidenced by keywords used throughout the topic examined.It's tough to get relevant and desire information with the growing number of data in recent years, most of which is unstructured.

3.1 Cluster Identification

Cluster are exhibition of traits and has attributes of same characteristic. To solve this, resemblance and difference in data obtained from Google scholar through clustering with the help of density-based clustering technique. The following steps were followed for the actualization of the techniques in Python; (1) Import libraries (2) Training and Classification of the points (3) Discard noise (4) Matplot Lib tool for virtualization of clusters identified by the trained model

The imported libraries for the analysis includes; Scikit learn, Pandas, it provides tools like DataFrame for analysingdataset, and Matplot Lib, is tool for plotting graphs. For data points classification, noise points, and border points epsilon value was set to 1. The value is the maximum distance between two samples of the grouped cluster. The min_sample was set to 3 and indicate the minimum number of samples in a neighbourhood data point qualification. The training was completed through the setting of parameters into the datasets. Figure 3.1 shows the training model with an espsilon value of 1 and min_sample of 3.Noise points were discarded and clusters were assigned to the core points. Figure 3.2 shows the core points that were identified by the training

model. These core points are total of seven hundred and seventy seven (777) out of the nine hundred (900) sample data used.

```
In [34]:  1 model = DBSCAN(eps=1, min_samples=3).fit(x[:,(1,2)])
          2 labels = model.labels_
          3 print (model)

DBSCAN(eps=1, min_samples=3)
```

Figure 3.1: Training the dataset to identify and classify points as the core

```
array([  0,   1,   2,   3,   4,   6,   7,   9,  10,  11,  12,  13,  14,
        15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,  26,  27,
        28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,  39,  40,
        41,  42,  43,  44,  45,  46,  47,  48,  50,  51,  52,  53,  54,
        55,  56,  57,  58,  59,  60,  61,  62,  63,  64,  65,  66,  67,
        68,  69,  70,  71,  72,  73,  74,  75,  76,  77,  78,  79,  81,
        82,  83,  84,  85,  86,  87,  88,  89,  90,  91,  93,  94,  95,
        96,  97,  98,  99, 100, 101, 102, 103, 104, 105, 106, 107, 108,
       109, 110, 111, 112, 113, 114, 115, 116, 117, 119, 120, 121, 122,
       123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135,
       136, 137, 138, 139, 141, 142, 143, 144, 145, 147, 149, 150, 151,
       152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164,
       165, 166, 167, 168, 169, 170, 171, 172, 173, 176, 177, 178, 179,
       180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192,
       194, 195, 196, 197, 199, 200, 201, 202, 203, 204, 205, 206, 207,
       208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222,
       224, 225, 226, 227, 228, 229, 230, 231, 233, 234, 235, 236, 237,
       238, 239, 240, 241, 243, 244, 246, 247, 248, 249, 250, 251, 252,
       253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 265, 267,
       268, 270, 271, 272, 273, 274, 275, 276, 278, 279, 280, 281, 282,
       283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295,
       296, 297, 298, 299, 301, 302, 304, 306, 307, 308, 309, 310, 311,
       312, 313, 314, 315, 316, 317, 318, 319, 320, 322, 324, 325, 326,
       327, 328, 329, 330, 331, 333, 334, 335, 336, 338, 339, 340, 341,
       342, 344, 345, 346, 347, 348, 349, 350, 352, 353, 354, 356, 357,
       358, 361, 362, 363, 364, 365, 366, 367, 368, 370, 372, 373, 375,
       376, 377, 378, 380, 381, 382, 383, 384, 385, 386, 387, 388, 391,
       392, 394, 395, 396, 399, 400, 401, 402, 403, 404, 405, 406, 407,
       408, 410, 413, 414, 416, 417, 418, 420, 421, 422, 424, 425, 426,
       427, 429, 432, 433, 434, 435, 436, 438, 439, 440, 441, 442, 443,
       444, 445, 446, 448, 450, 451, 452, 453, 454, 456, 458, 459, 461,
       464, 465, 466, 467, 469, 470, 472, 473, 475, 476, 477, 479, 481,
       483, 485, 486, 488, 490, 491, 492, 493, 494, 495, 497, 498, 499,
       500, 501, 503, 504, 505, 507, 508, 509, 510, 511, 512, 513, 514,
       516, 517, 519, 523, 524, 525, 527, 528, 530, 533, 534, 535, 536,
       537, 538, 539, 545, 546, 547, 548, 549, 550, 552, 554, 555, 557,
       561, 562, 564, 565, 567, 568, 569, 571, 577, 578, 579, 583, 584,
       585, 586, 587, 588, 591, 592, 593, 594, 595, 598, 599, 600, 601,
       602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614,
```

Figure 3. 2: isolated core samples in the dataset

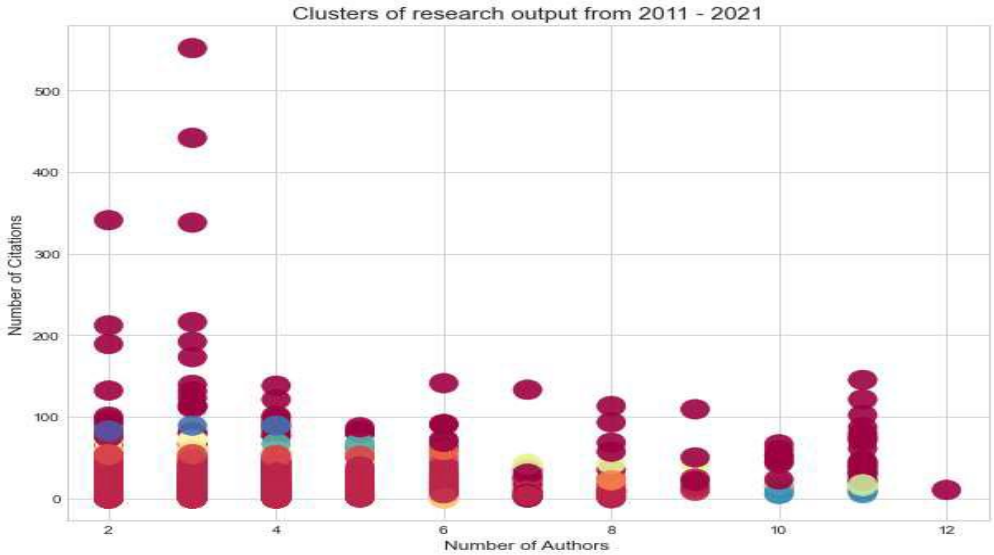

Figure 3.3: Clusters Pictorial View

3.2 Topic Modelling for Identification of Dominant Topics

The enlisting and validation of dominant topics were analysedby the use of the dataset to isolate topics that dominated research space in the reviewed period. For actualisation, the steps below were followed;

(a) Loading data

The dataset of papers published and obtained from Google scholar which is one of the most prestigious research citation databases were used. The CSV data file contains information on three hundred (300) article papers each from three fields of study (Computer Science/ICT, Medicine, and Engineering), that were published from 2011 until 2020 (10 years!). These papers discuss a wide variety of topics in their respective fields. Figure 3.4 displays the input data used for the analysis.

```
df = pd.read_csv('top.csv',encoding="ISO-8859-1")
df.head()
```

Out[2]:

	Field	title	Citation	Authors
0	Engineering	engineering education factors and academic per...	0.0	3.0
1	Engineering	locating static var compensators for loss mini...	0.0	4.0
2	Engineering	drought spatiotemporal characterization using ...	15.0	5.0
3	Engineering	application of response surface methodology (r...	5.0	5.0
4	Engineering	resistivity tomography for geo engineering inv...	1.0	3.0

Figure 3.4: Display of five sample data from the dataset to be used for DBSCAN analysis

(b) Data Cleaning and Pre-processing

It is understood that the goal of this analysis is to perform topic modelling, so, the text data from each paper was the focus. Other metadata columns such as the number of references and the number of co-authors were removed. A total number of 900 papers were used.Pre-processing was performed on the content of the paper_text column to make them more amenable for analysis, and reliable resultsusing regular expression to remove any punctuation, and then lowercase the text as shown in Figure 3.5

```
# Remove punctuation
df['title_processed'] = \
df['title'].map(lambda x: re.sub('[,\.!?]', '', x))
# Convert the titles to lowercase
df['title_processed'] = \
df['title_processed'].map(lambda x: x.lower())
# Print out the first rows of papers
df['title_processed']
```

Out[4]:

```
0      engineering education factors and academic per...
1      locating static var compensators for loss mini...
2      drought spatiotemporal characterization using ...
3      application of response surface methodology (r...
4      resistivity tomography for geo engineering inv...
                           ...
744    fuzzy expert system for the management of hype...
745     review of application of genetic algorithms i...
746    an intelligent mediating model for collaborati...
747     a collaborative algorithm for ontological mat...
748     issues and challenges in clustering technique...
Name: title_processed, Length: 749, dtype: object
```

Figure 3.5: The pre-processed text from each document title

(c) Exploratory Analysis

To verify whether the pre-processing was done, a word cloud was made using the wordcloud package to get a visual representation of the most common words in each topic. This is important for understanding the data and ensuring we are on the right track, and if any more pre-processing is necessary before training the model. The size of the keywords is dependent on the score value assigned to such word in the particular topic. Figures 3.6, 3.7 and 3.8 shows a pictorial view of topic keywords in each topic.

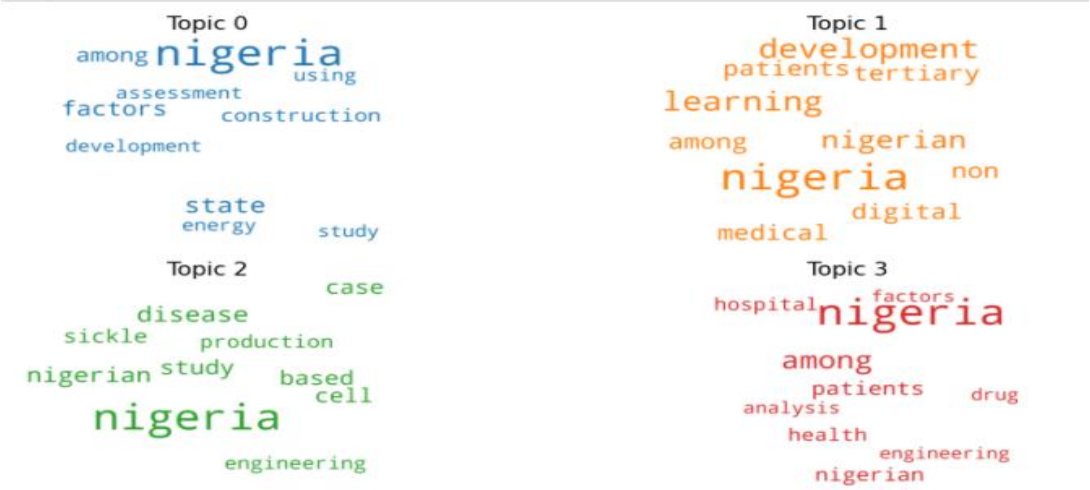

Figure 3.6: Visual representation of most common words for topic 0, 1, 2, and 3

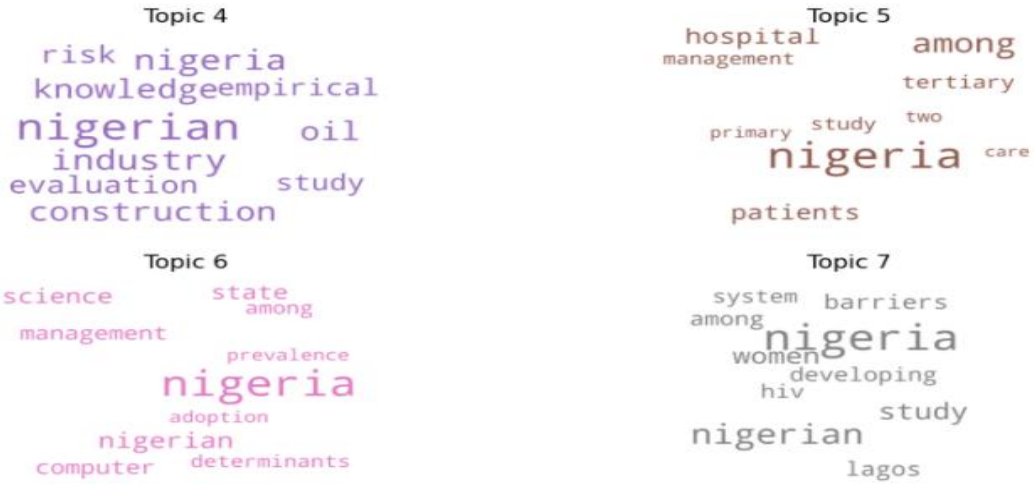

Figure 3.7: Visual representation of most common words for topic 4, 5, 6, and 7

Figure 3.8: Visual representation of most common words for topic 8 and 9

(d) Prepare data for LDA Analysis

This aspect involves the transformation of the textual data into a format that serves as input for training the LDA model. First step was tokenizing the text and removing stopwords using the gensim libraries. Stop words are words which does not add much meaning to a sentence and that can be ignored. Next, the tokenized object is converted into a corpus and dictionary. Gensim creates a unique id for each word in the document. Figure 3.9 showing samples of the tokenized objects such as 'engineering', 'education', and others. Figure 3.10 shows the corpus which was produced by the model, which is a mapping of word_id, with word_frequency. For instance topic 1 is coded as 0 occurring once.

(e) LDA model training

The study carefully trained the model for few guidelines such 10 topics model built, and each topic has a keywords, and the keyword has a percentage of weightage as topic. The gensimas used is demonstrated in Figure 3.11. Example is Keywords such as Nigeria, state and factors in topic 0 had highest score value of 0.061, 0.013, and 0.011 respectively. For energy, development and study is seen having lowest and equal values 0.007.

```
import gensim
from gensim.utils import simple_preprocess
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
stop_words.extend(['from', 'subject', 're', 'edu', 'use'])
def sent_to_words(sentences):
    for sentence in sentences:
        # deacc=True removes punctuations
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))
def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc))
             if word not in stop_words] for doc in texts]
data = df.title_processed.values.tolist()
data_words = list(sent_to_words(data))
# remove stop words
data_words = remove_stopwords(data_words)
print(data_words[:1][0][:30])
```

```
['engineering', 'education', 'factors', 'academic', 'performance', 'south', 'western',
```

Figure 3.9: Sample display of tokenized objects in the dataset

```
In [7]:
```

```
import gensim.corpora as corpora
# Create Dictionary
id2word = corpora.Dictionary(data_words)
# Create Corpus
texts = data_words
# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]
# View
print(corpus[:1][0][:30])
```

```
[(0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1)]
```

Figure 3.10: Mapping of word_id with word_frequency

```
[(0,
  '0.061*"nigeria" + 0.013*"state" + 0.011*"factors" + 0.010*"among" + '
  '0.009*"construction" + 0.008*"assessment" + 0.008*"using" + 0.007*"energy" '
  '+ 0.007*"development" + 0.007*"study"'),
 (1,
  '0.018*"nigeria" + 0.009*"development" + 0.009*"learning" + 0.007*"nigerian" '
  '+ 0.006*"tertiary" + 0.006*"among" + 0.006*"digital" + 0.006*"medical" + '
  '0.006*"non" + 0.006*"patients"'),
 (2,
  '0.053*"nigeria" + 0.010*"disease" + 0.009*"nigerian" + 0.009*"study" + '
  '0.009*"based" + 0.008*"case" + 0.008*"cell" + 0.008*"sickle" + '
  '0.007*"production" + 0.006*"engineering"'),
 (3,
  '0.048*"nigeria" + 0.018*"among" + 0.009*"patients" + 0.009*"nigerian" + '
  '0.008*"health" + 0.008*"hospital" + 0.007*"analysis" + 0.007*"drug" + '
  '0.007*"factors" + 0.007*"engineering"'),
 (4,
  '0.017*"nigerian" + 0.010*"industry" + 0.010*"nigeria" + '
  '0.008*"construction" + 0.008*"knowledge" + 0.008*"oil" + 0.007*"risk" + '
  '0.007*"evaluation" + 0.007*"empirical" + 0.006*"study"'),
 (5,
  '0.045*"nigeria" + 0.022*"among" + 0.013*"hospital" + 0.012*"patients" + '
  '0.009*"tertiary" + 0.008*"study" + 0.008*"management" + 0.007*"two" + '
  '0.006*"primary" + 0.006*"care"'),
 (6,
  '0.055*"nigeria" + 0.011*"nigerian" + 0.010*"science" + 0.010*"state" + '
  '0.009*"computer" + 0.009*"management" + 0.008*"among" + '
  '0.007*"determinants" + 0.006*"adoption" + 0.006*"prevalence"'),
 (7,
```

Figure 3.11: score value for each keyword in each topic

(f) Extracted Percentage Contribution

In LDA models, each document is composed of multiple topics. But, typically only one of the topics is dominant. The code applied here, extracts the dominant topic for each sentence and determines the weight of the topic and the keywords in nicely formatted output. This can be seen in Table 3.1

This is to help identify which document associates predominantly with each topic.

Table 3.1: Dominant topics, their Keywords, text and topic percentage contribution

	Docyment_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
1	1	2.0	0.9400	nigeria, disease, nigerian, study, based, case, cell, sickle, production, en...	[locating, static, var, compensators, loss, minimisation, using, linear, mix...
2	2	0.0	0.9307	nigeria, state, factors, among, construction, assessment, using, energy, dev...	[drought, spatiotemporal, using, self, calibrating, palmer, drought, severit...
3	3	4.0	0.7499	nigerian, industry, nigeria, construction, knowledge, oil, risk, evaluation,...	[application, response, surface, methodology, rsm, optimization, energy, gen...
4	4	9.0	0.9470	nigeria, medical, information, study, among, science, south, communication, ...	[resistivity, tomography, geo, engineering, investigation, subsurface, defec...
5	5	9.0	0.9250	nigeria, medical, information, study, among, science, south, communication, ...	[comparative, analysis, radiotherapy, linear, accelerator, downtime, failure...
6	6	0.0	0.8012	nigeria, state, factors, among, construction, assessment, using, energy, dev...	[relationship, project, performance, measures, project, management, practice...
7	7	6.0	0.9307	nigeria, nigerian, science, state, computer, management, among, determinants...	[stochastic, modelling, spatial, variability, petrophysical, properties, par...
8	8	2.0	0.9250	nigeria, disease, nigerian, study, based, case, cell, sickle, production, en...	[conceptual, model, technical, sustainability, integration, electrical, elec...
9	9	4.0	0.9250	nigerian, industry, nigeria, construction, knowledge, oil, risk, evaluation,...	[investigating, barriers, building, information, modeling, bim, implementati...

In this case, samples of texts that most represent a given topic in the documents were obtained by implementing a code that gets the most exemplar sentence for each topic. Then the resulting output was visualized as represented in Table 4.2. It shows, for instance, topic one have texts like ‘attitudes’ , ‘beliefs ’, ‘mental’, ‘illness’, ‘among’, ‘church’, ‘based’, ‘lay’, ‘health’, ‘workers’, etc.

Table 3.2: Representation text from each dominant topic

	Topic_Num	Topic_Perc_Contrib	Keywords	Representative Text
0	0.0	0.9526	nigeria, state, factors, among, construction, assessment, using, energy, development, ...	[attitudes, beliefs, mental, illness, among, church, based, lay, health, workers, expe...
1	1.0	0.9437	nigeria, development, learning, nigerian, tertiary, among, digital, medical, non, pati...	[assessment, contribution, online, information, resources, open, distance, learning, m...
2	2.0	0.9526	nigeria, disease, nigerian, study, based, case, cell, sickle, production, engineering	[modelling, impact, intervention, measures, total, accident, cases, nigeria, using, bo...
3	3.0	0.9609	nigeria, among, patients, nigerian, health, hospital, analysis, drug, factors, enginee...	[time, series, predcomputer, scienceion, five, mortality, rates, nigeria, comparative,...
4	4.0	0.9526	nigerian, industry, nigeria, construction, knowledge, oil, risk, evaluation, empirical...	[microstructure, mechanical, behaviour, stir, cast, al, mg, si, alloy, matrix, hybrid,...
5	5.0	0.9571	nigeria, among, hospital, patients, tertiary, study, management, two, primary, care	[multiple, linear, quadratic, models, estimating, road, crashes, semi, urban, two, lan...
6	6.0	0.9470	nigeria, nigerian, science, state, computer, management, among, determinants, adoption...	[exclusive, breastfeeding, hiv, aids, crossectional, survey, mothers, attending, preve...
7	7.0	0.9526	nigeria, nigerian, study, women, barriers, lagos, among,	[high, prevalence, hiv, chlamydia, gonorrhoea, among, men,

IV. EXPOSITORY ANALYSIS OF RESULT FROM LDA MODEL

The successful models as trained creates steps for visualization of the topics for understand. To enable the carry-out of the visualization, a common visualization software pyLDAvis was used. pyLDAvis is meant to assist in a hands-on manner with:

a. Understanding the topics better and interpreting individual topics.

Each topic was manually selected to view its topmost frequent AND/OR "relevant" terms, using different values of the λ parameter. This can be useful when attempting to give each topic a human-interpretable name or "meaning."

b. Give a better understanding of the connections between the subjects. The Intertopic Distance Plot can be used to understand more about how subjects are related to one another, as well as possible higher-level structure between groups of topics.

3.2 Results Evaluation

Table 4.1: List of dominance topics obtained from the LDA topic modelling on the sample data for topics 1 - 3

Doc Number	Field / Discipline	Topic	No. of Citations	No. of Authors
1	Medicine	Attitudes and beliefs about mental illness among church-based lay health workers: experience from a prevention of mother to child HIV transmission trial in Nigeria	27	11
2	Computer Science	Assessment of the contribution of online information resources in open distance learning mode to the development of lifelong learning in South West, Nigeria	4	2
3	Engineering	modelling the impact of intervention measures on total accident cases in Nigeria using box Jenkins methodology: a case study of Federal road safety commission	2	4
4	Computer Science	Time series predComputerScienceion of under 5 mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt Winters exponential smoothing and autoregressive integrated moving average models.	4	2
5	Engineering	Microstructure and mechanical behaviour of stir-cast Al-Mg-Sl alloy matrix hybrid composite reinforced with corn cob ash and silicon carbide	51	3
6	Engineering	multiple linear and quadratic models in estimating road crashes in semi-urban two-lane roads: case study of Nsukka municipal council, southeastern Nigeria	0	2
7	Medicine	Exclusive breastfeeding and HIV/AIDS: a crossectional survey of mothers attending prevention of mother to child transmission of HIV clinics in southwestern Nigeria	30	3
8	Medicine	High prevalence of HIV, chlamydia and gonorrhoea among men who have sex with men and transgender women attending trusted community centres in Abuja and Lagos, Nigeria.	76	11

Table 4.2: List of dominance topics obtained from the LDA topic modelling on our sample data

Doc Number	**Field / Discipline**	**Topic**	**No. of Citations**	**No. of Authors**
9	Engineering	Analysis of the effect of carbon/nitrogen (C/N) ratio on the performance of biogas yields for non-uniform multiple feedstock availability and composition in Nigeria.	22	3
10	Computer Science	Using android and open data kit Technology in Data Management for research in resource-limited settings in the Niger Delta region of Nigeria: cross-sectional household survey.	27	3

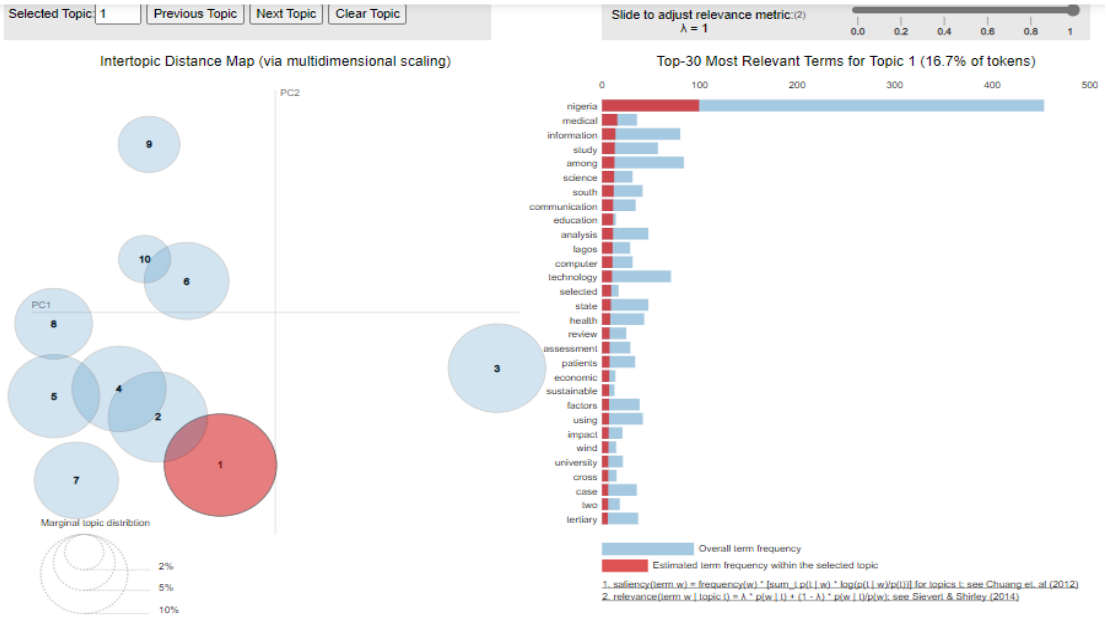

Figure 4.1: The layout of LDAvis, with the global topic view on the left, and the term bar charts (with Topic 1 selected) on the right. Linked selections allow users to reveal aspects of the topic-term relationships compactly

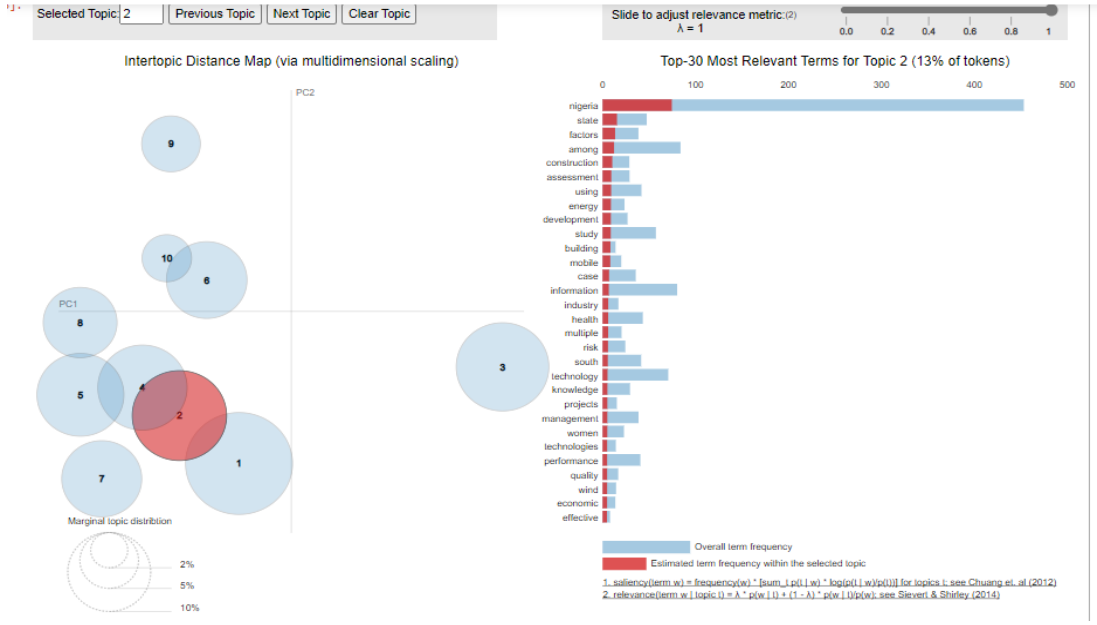

Figure 4.2: The layout of LDAvis, with the global topic view on the left, and the term bar charts (with Topic 2 selected) on the right. Linked selections allow users to reveal aspects of the topic-term relationships compactly

The work has revealed the state of research work among researchers in Sub-Saharan Africa and with the number of citations and co-authored articles in Table 4.2 is a clear indcation that more work are still be carried out by scietific scholars, though not on the pace as expected but there is a sign of progress on the part of the researchers and with the relatioship index as capture in Figure 4.1 and 4.2 respectively it shows that there are rooms for more state of collaboration among researcher and effort should be geared towards the actualization of the neeeded state that would further aid the speady growth of development in Sub-Saharan Africa nations.

REFERENCE

[1]. Rajasekar, S., Philominathan, P., and Chinnathambi, V. (2013). Research methodology. eprint. arXiv preprint physics/0601009, 1-53.

[2]. Çaparlar, C. Ö., and Dönmez, A. (2016). What is scientific research and how can it be done?.*Turkish journal of anaesthesiology and reanimation*, 44(4), 212.

[3]. Adams, J., Gurney, K., Hook, D., and Leydesdorff, L. (2014). International collaboration clusters in Africa. *Scientometrics*, 98(1), 547-556

[4]. Jeong, S., Choi, J. Y., and Kim, J. Y. (2014). On the drivers of international collaboration: The impact of informal communication, motivation, and research resources. *Science and Public Policy*, 41(4), 520-531.
[5]. Leydesdorff, L., Park, H. W., and Wagner, C. (2014). International coauthorship relations in the Social Sciences Citation Index: Is internationalization leading the Network?.*Journal of the Association for Information Science and Technology*, *65*(10), 2111-2126.
[6]. Tang, L., and Shapira, P. (2012). Effects of international collaboration and knowledge moderation on China's nanotechnology research impacts. *Journal of Technology Management in China*, 7(1), 94– 110
[7]. Wagner, C. S., Whetsell, T., Baas, J., and Jonkers, K. (2018). Openness and impact of leading scientific countries. *Frontiers in research metrics and analytics*, 3, 10.
[8]. Ebersberger, B., Herstad, S. J., and Koller, C. (2014). Does the composition of regional knowledge bases influence extra-regional collaboration for innovation?.*Applied Economics Letters*, 21(3), 201-204
[9]. Leydesdorff, L., Wagner, C., Park, H. W., and Adams, J. (2013). International collaboration in science: The global map and the network. arXiv preprint arXiv:1301.0801.
[10]. Wilsdon, J. (2011). Knowledge, networks and nations: Global scientific collaboration in the 21st century. London: *The Royal Society*.
[11]. Mêgnigbêto, E. (2013). Scientific publishing in Benin as seen from Scopus. *Scientometrics*, 94(3), 911-928.
[12]. Mêgnigbêto, E. (2013). Scientific publishing in West Africa: Comparing Benin with Ghana and Senegal. *Scientometrics*, *95*(3), 1113-1139.
[13]. Boshoff, N. (2010). South-South research collaboration of countries in the Southern African Development Community (SADC). *Scientometrics*, 84(2), 481-503.
[14]. Dodsworth, S. (2019). The challenges of making research collaboration in Africa more equitable. *Oxford Research Encyclopedia of Politics.* Pp 641
[15]. Blom, A., Lan, G., and Adil, M. (2016). *Sub-Saharan African science, technology, engineering, and mathematics research: a decade of development*. World Bank Publications. Washington, DC: World Bank. doi:10.1596/978-1-4648-0700-8. Retrieved March 23rd 2021.