

Data Mining of Patents: A Study on Application of Machine Learning in Animal Husbandry

Alok Khode; Nishad A Deshpande; Shabib Ahmed Shaikh
CSIR Unit for Research and Development of Information Products (URDIP), Pune

Abstract – The past two decades have seen rapid and enormous growth in technology, its implementation and application, riding on the back of information technology. The availability of technology enablers such as the ease of access to internet coupled with cost effective computing devices have resulted into multi-fold adaption and utilization of IT based technology. The advent of cloud computing has enlarged the horizon of data centers and storage. This has hence resulted in huge data generation and there by leading to calls for improved, effective, and efficient data mining concepts to extract patterns or knowledge from huge set of data. The concepts of data mining are still evolving with changes in technology and application needs. Data mining concepts are utilized in almost every business needs right from healthcare to Finance, Academics to Surveillance, Research Analysis to Artificial intelligence, Farming, logistics, Energy industry and many more. One such emerging use of data mining is its application in Animal Husbandry which forms a major component of the agriculture industry. Machine learning has also evolved over a period and have been utilized and used in almost every industry now a days and the field of Animal husbandry is no exception. Many machine learning techniques are being adopted in the field of Animal Husbandry to manage and monitor various live stocks. Such applications of Machine learning in the field of Animal Husbandry have resulted into innovations for which patent protection is being sought.

Patents are form of Intellectual Property Rights (IPR) that helps in protection of innovative technologies for a fixed duration in return for public disclosure. Many a times, patents are the earliest and the only mode of technology disclosure and therefore act as important source of information. This paper showcases mining patent data which is available in the public domain to bring out insights into various use of technology, the current application trends, major players, and their market focus along with origin of inventions in the field of animal husbandry that utilize machine learning.

Keywords – Patent Mining, Patent Informatics, Patinformatics, Data Mining, Patent Analysis, Livestock, Agriculture, Animal Husbandry.

I. INTRODUCTION

The .com revolution enabled a new economy which was defined by the software and internet boom. This was coupled with the high availability of web services, high internet

bandwidth along with cost effective computing devices of various shapes and sizes having ease of access and mobility. IT and Software to be specific have thus enabled IT to become an essential part of our daily lives. Software has also become a main component of new technological advancements in almost all fields of technologies (EPO). The last decade has seen software enabled services take central stage and most of the services have become a routine utility in our daily lives. The advent of cloud computing has enlarged the horizon of datacentres and storage thereby leading to technological advancements in the field of IoT, security and surveillance, High Performance Computing, Machine Learning, Virtual Reality, and grid computing. All this has led to enormous collection of data from multiple sources in multiple formats which has led to calls for better handling of data and its mining. Even though the concept of data mining has been in existence since the 90's, it is still evolving and changing with the changing of technology, data sets and application needs from the end users.

The concept of data mining is utilized in almost all fields of technology for predictive analysis, and it has become a synonym with the term knowledge discovery in data bases. Data mining concepts are utilized right from Healthcare to Finance, Academics to Surveillance, Research Analysis to Artificial Intelligence, Farming, logistics, Energy Industry, and many more to gain insights on data and business intelligence. One such evolving usage of data mining is its application in the field of Animal Husbandry which forms a major component of the agriculture industry. Apart from data mining, machine learning has also evolved over a period and been utilized and used in almost every industry now a days for business automations and the field of animal husbandry is no exception. Many machine learning techniques are being adopted in the field of Animal Husbandry to automatically manage as well as monitor various live stocks up to some extent. Such applications of Machine learning in the field of Animal Husbandry have resulted into innovations for which patent protection is being sought.

Patents are a form of Intellectual Property Rights (IPR) which help protect an innovation for a fixed term in a particular territory in return for public disclosure of the innovation. In most cases, patents are the earliest and the only mode of technology disclosure published anywhere on the globe and hence act as an important source of technological information that is available in the public

domain. This paper showcases the mining of patent data which is available in the public domain to bring out insights into various use of technology, the current technological trend patterns, major players, and their market focus along with origin of inventions in the field of Animal Husbandry that utilise machine learning.

II. LITERATURE REVIEW

In the agriculture domain a vast amount of data from divergent areas is available. Utilization of data mining techniques helps in uncovering the hidden information and with human interpretation can lead to knowledge discovery. Majumdar (2017) have used data mining on soil and climatic data to optimize the parameters leading to maximizing crop yields. Apart from available subject specific datasets in agriculture domain, other source of information such as scientific articles, technical datasheet and patent documents can also be used to extract relevant information.

Patent document is considered as one of the reliable sources to reveal up-to-date technological information (Noh et al., 2015). It is suggested that more than 70% of technological information available in patents is never published anywhere else (Wang, 2011). Patent data is treated as a gold mine, if searched and analysed appropriately (Deshpande et al., 2014) as it creates value in R&D, mergers and acquisitions, technology assessment as well as human resource management (Ernst, 2003, Shaikh and Singhal, 2019). Analysing this enormous data also aids in ascertaining technological trends, R&D directions, competitive advantage which would not be apparent if patent documents are handled on one-on-one basis. Patent documents are lengthy multi-topic, semi-structured, rich in technical terms and available in multiple languages and very different from general news articles or web documents (Zhang et al., 2015; Khode and Jambhorkar, 2019). Hence, patent documents are considered as special documents (Khode and Jambhorkar, 2017). Due to this, analysis of patents is a non-trivial task which requires colossal human efforts. This has resulted in new area of research in data mining called patent mining which involves processing, investigating, and analysing the patent documents to help patent professionals in better patent analysis (Zhang et al., 2015, Shaikh and Singhal, 2018).

In literature, number of patent analysis approaches have been suggested to assist researchers and patent professionals to gain intelligence technological information from patent data to support R&D strategies (Yoon et al., 2013, Shaikh et al., 2020). Thus, researchers have started looking at patent data to analyse and study these technology trends (Chen et al., 2017; Deshpande et al., 2016; Li et al., 2015). Studies show that use of data mining techniques are also widely used in the patent analysis (Ju and Sohn, 2015; Vincent et al., 2017; Li et al. 2019). Apart from it, text mining has also been explored to track the path of technology development as well as to monitor technological trends (Wang et al., 2015; Wu and Leu, 2014).

Vincent et al. (2017) used patent data of the various technology domains in fisheries sector and used patent mining technique to analyse the legal status of patents, top priority countries and top assignees in this domain. Study suggested that patent data mining in the Fisheries sector helps in the early identification of innovative trends and to stay ahead in the competition.

Singh (2012) highlighted that although the domain of agriculture is vast, the patent data can be strategically utilized as data mining tool to dig out knowledge from patent information related to new designs, technology improvement, processing techniques etc. which will in turn can be improved by researchers. Study showed that despite of India being a major exporter of agricultural goods, the patenting activity is not significant.

In another study which uses patent data in the agriculture sector, Zheng et al. (2020) predict composting technology (CT) by exploiting patent-based measures. They analyse the structured and unstructured information of 11,701 composting technology patents and suggested that the focus of CT has shifted from waste disposal to organic fertiliser preparation and highlighted the potential research directions.

Grzesiak (2012) described various data mining approaches such as random forest, artificial neural network etc. applied to various issues related with animal husbandry. In the present study, patent data is used for data mining to bring out development of technologies related to application of machine learning in animal husbandry.

III. METHODOLOGY

The purpose of patenting system is exclusivity in return for information being made available to public. Thus, Patents containing scientific and technological information and being freely accessible, have been used as information resource in the present study. The present work was carried out in three broad steps: I. Data Collection II. Data Processing III. Data Summarisation and Visualisation

A. Data Collection

The patents that disclosed use of machine learning for application in animal husbandry were searched in the freely available database viz Espacenet (<https://worldwide.espacenet.com/>). The search was carried out using International Patent Classification is a hierarchical system of categorisation of patent documents depending on the technology claimed. This system is standardised and is overseen and update every year by the World Intellectual Property Organization (WIPO, 2004). The categorisation of the patent documents helps patent examiners in searching for previously filed patent documents and hence the system has been used by most of the patent granting authorities. The IPC codes listed below in table I were used in the search:

TABLE I
DESCRIPTION OF THE IPC CODES USED IN THE SEARCH

IPC Code	Description
----------	-------------

A01K	Animal husbandry in general
G06N20	Computer systems based on specific computational models-Machine learning
G06K9	Methods or arrangements for reading or recognising for recognising patterns
G06F40	Handling natural language data

The search involving various combinations of the above-mentioned IPC codes yielded 970 patent documents and the information was exported into excel spreadsheet for further use.

B. Data Processing

Before analysis, the collected data was processed for stemming, stop words removal using various python modules. The patent data available in the patent databases is available in two forms structured data and unstructured data (Vincent et al, 2017). Structured patent information includes information such patent identifier such as Granted patent number or Published application number, the date of filing of patent application or the date of grant of the patent, IPC codes assigned to the patent documents. The unstructured information is textual in nature and involves abstract, claims and the specification relating to the invention disclosed as part of the patent document. Espacenet does not allow downloading all the content of patent documents into spreadsheet (although full text of the patent document is displayed on screen and in the PDF) so other data sources were used to populate the empty fields, and these were used for data mining and further analysis.

C. Data Summarisation and Visualisation

This step makes use of datamining for summarisation of the important information for the patents that were identified from the Espacenet database. For data summarisation and visualisations, the spreadsheet was imported into python and various packages were used to process the data leading to its summarisation and Visualisation. Listed below are few python packages used for summarisation and visualisation: Pandas, collections, nltk, matplotlib, numpy, re, wordcloud, scipy, sklearn etc

IV. ANALYSIS AND DISCUSSION

In this section, the processed patent data is analysed using data mining techniques and discussed to bring out various trends with respect to patent filing, origin of invention, major assignees, application of ML in animal husbandry etc.

A. Patenting Trend

The patent data was analysed to find out the timeline of patent filings in the field of Animal Husbandry that utilise machine learning. The same is reported in figure 1. The patent application year, patent publication year and the patents priority year was analysed. Application year is the year in which a patent is applied. Patents are usually

published after a period of 18 months from filing of applications and that year is referred to as publication year. Patents are territorial in nature and hence can be filed in different nations at different times. However, the first application filing year of such patents are referred to as priority year while filed in different nations. It was observed that the first application in this field was in the year 1968. There were single patents filed in the year 1968, 1972 and 1974. Two patents were applied in 1974. 10 patents were filed in 1999 and consistent 2-digit filing was observed since 2001 however the applied patents each year were less than 20. There was a sudden increase in the application filings of patents from 2014. It was observed that only 53 patents were applied for before the year 2000, 65 patents were applied between 2005 and 2010, 109 were applied between 2010 and 2015 and a staggering 677 patents were applied for since 2015 in the field of Animal Husbandry that utilise machine learning. A dip can be observed in patent filings in the last two years; however, this may not be the actual case since an applied patent gets published only after a period of 18 months from its applications date, hence not all patents applied in the field of the study were published which are applied in the last couple of years. (Shaikh et al., 2018).

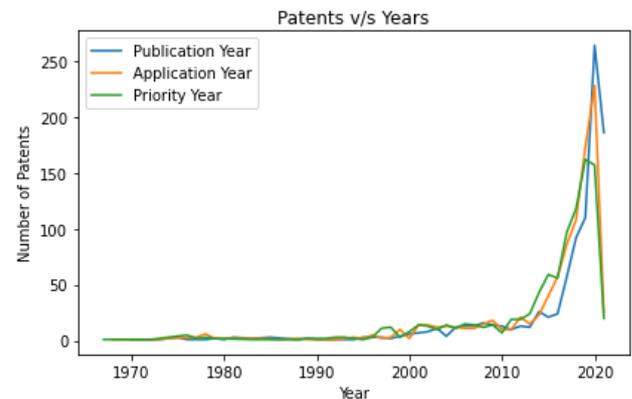


Figure 1: Patenting Trend

Analysis of publication year highlighted that the first patent was published in the public domain in 1970 and 155 patents were published till 2010. 83 patents were published between 2010 and 2015 while the bulk of patents in this filed were published in the last five years. It was observed that 733 patents were published between 2015 and 2021.

One patent had priority referring to technology patented in 1967. There were 24 patents which referred to priority year before 1990, 62 patents had priority year between 2000 and 2005 as well as similar count of 62 patents referred to priority year between 2005 and 2010. There were 164 patents refereeing to priority year between 2010 to 2015 while most of the patents, 610 referred to priority year after 2015. This showed that technology has mainly evolved in the last five years.

B. Patent Legal Status:

The legal status of the patents was also analysed to understand which all patents are still enforced. It was

observed as given in figure 2 above that majority of the patent were alive and enforced. More than 63% of the patents in the field of animal husbandry tat used ML techniques are still enforced while 24% of the patents are inactive and dead. Around 13% of the patents were in indeterminate mode highlighting that thy might not yet be granted or may be in processing stage.

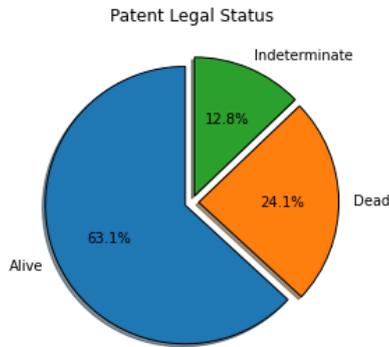


Figure 2: Legal Status of Patents

C. Assignees of Patents

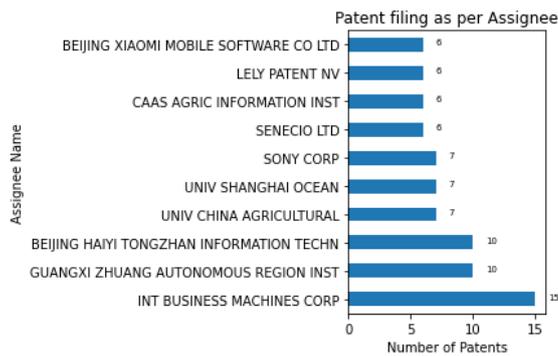


Figure 3: Patent Assignee

Patent are assigned (owned) by individuals or various institutions. It can be seen from the figure 5 that IBM is the topmost entity having patents filed in this area of study. An important point to note is that 6 institutes out of the top 10 are from China having about 58% of patents in the top 10 list.

D. Origin of invention

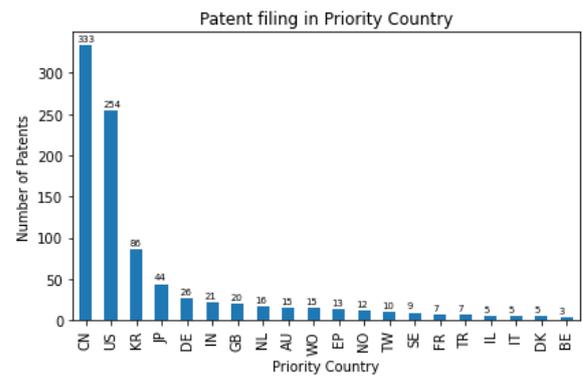


Figure 4: Patent Filing based on origin of invention

Patents are territorial in nature and a similar patent may be protected with separate applications filed in different countries. However, these patent applications will have the priority country filing mentioned in them to highlight the jurisdiction in which the first patent was filed. Usually the country of invention, where a technology is invented is the first country where protect is sought via means of technology. Hence analyzing the priority country will highlight the country from where a technology originated. Looking at the priority country filings it was observed that most of the technologies originate from the Asian subcontinent with China leading with 333 patents followed by 254 patents from USA, 86 Korea, 44 Japan while 21 patents also originated from India.

E. Technology Trend based on IPC/ CPC

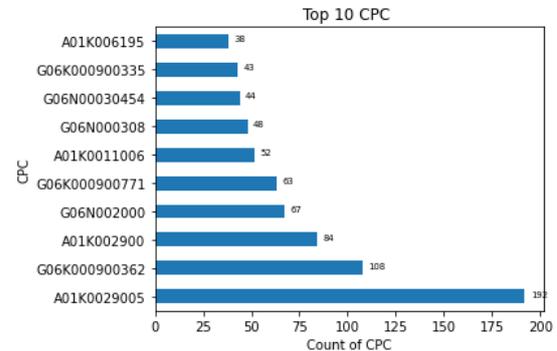


Figure 5: Patent Filing based on CPC

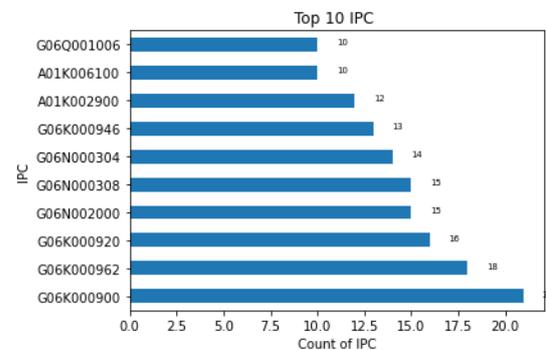


Figure 6: Patent Filing based on IPC

Technology trend can be understood by studying the filings based on CPC and IPC. CPC is Cooperative Patent Classification and IPC is international patent classification. These are hierarchical classifications scheme of technologies followed by major patent granting authorities. IPC is used to classify patent according to technology disclosed in the patent applications. CPC is an extended version having more depth level of classifications and has been developed jointly USPTO and EPO along with China. Details of the top 10 IPC / CPC code in the field related to Animal Husbandry that utilise machine learning along with their description is given in table II below:

TABLE II
DESCRIPTION OF TOP IPC/CPC

IPC/CPC	Description
A01K0011006	Automatic identification system for animals
A01K0029005	Monitoring apparatus for animal husbandry
A01K006100	Culture of aquatic animals
A01K006195	Sorting, grading of fish
G06K000900335	Recognising movements
G06K000900362	Recognising animal bodies
G06K000900771	Recognising scenes under surveillance
G06K000920	Image acquisition
G06K000946	Extraction of features or characteristics of the image
G06K000962	Methods or arrangements for recognition using electronic means
G06N000304	Computer systems based on biological models
G06N002000	Machine learning
G06Q001006	Administration or Management of Resources, workflows

F. Citation Analysis of Patent Documents

Patent Document consists of details of citations such as other cited references (backward citations), citations of non-patent literature and patents that cites the current patent (forward citations). The details citation related information is reflected in the following table III.

TABLE III
CITATIONS ANALYSIS

Number Range	Number of Patents Cited	Number of documents	Number of documents Cited Non-
--------------	-------------------------	---------------------	--------------------------------

	Patent Reference	Citing Patents	Patent Reference
1 - 5	685	849	899
5 - 10	138	31	30
10 - 20	83	33	19
20 - 30	25	21	5
30 - 50	23	18	8
50 - 75	3	3	6
75 - 100	5	3	0
100 - 150	3	4	1
150 - 200	1	5	1
200 - 260	4	3	0
Average	6.75	4.94	2.44
Min	0	0	0
Max	253	239	492

It is observed from the above table that majority of the patents have cited count between 1 – 5 for all the three types of citations. A patent on an average cites around 7 patent and 2 non patent literature for its technology. On the other hand, a patent is cited by about 5 other patents on an average.

G. Word Taxonomy

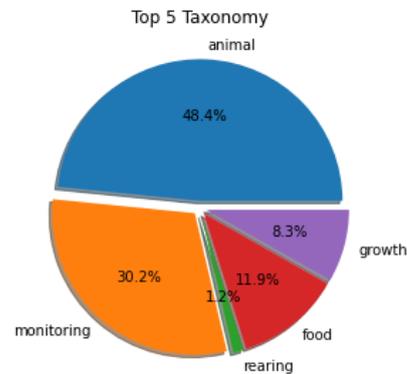


Figure 7: Word Taxonomy

Word taxonomy was analysed from the title, abstract and claims fields. It was found that more than 48% of patents had “animal” in their fields while more than 30% patents had “monitoring”. Around 12% of patents referred to “food”, 1.2% patents referred to “rearing” and 8.3% referred to “growth” in their text that was abstracted from the three fields of Title, Abstract and claims.

To understand what the technological focus is and to understand the application of ML, the NLP based keyword extraction from the title and abstract of the patent was implemented using python. To understand the changing focus over period, patents filed from year 2000 were grouped into 5 years’ time frame and the keywords and Bi-

grams and Tri- grams were created from the NLP extracted keywords.

Table no IV below summarises the technology evolution and its application areas over a period in the last 20 years

From the table IV it is observed that since year 2000, the primary focus has been on image detection and analysis and in the recent years' technologies building on deep learning, IoT have also been patented. There are various applications for which ML has been used primarily for livestock

monitoring and identification and other applications involve, carcass grading, egg measurement, milk floe profiling. In recent years, ML has been used for various purposes such as feeding appetite estimation, parasite monitoring and behaviour analysis etc.

TABLE IV
TECHNOLOGY EVOLUTION AND APPLICATION AREA OVER A PERIOD OF TIME

Time frame	Focus area	Emerging Area	Applications	Animals
2000-2004	measuring dimensions, type colour matching	biometric identification, face image	egg measurement, behaviour informatics, milk flow profiling, Livestock monitoring and identification,	Bovine
2005-2009	pattern information and verification	muzzle pattern, livestock bio identification / traceability,	monitoring and identification, diagnostics, carcass grading, aquatic counting	fish, bovine
2010-2014	image detection and analysis, bar code, body condition score	Metabolic efficiency of livestock	managing behaviour, reproductive estimation, identification, and detection	fish, bovine
2015-2019	Machine learning, machine vision	Health monitoring, Feed management,	deep learning, fish identification, neural networks, IoT, Fish Count prediction, Sow oestrous monitoring,	Fish, pig, poultry
2020-	Image analysis, ML, deep learning,	Feed management and monitoring, health monitoring, Growth analysis, Convolutional neural network	behaviour analysis, AI, Fish parasite monitoring, Feeding appetite forecasting,	Fish, pig, poultry

V. CONCLUSION

Patents act as an important source of technical information and can be used to understand the progress of technology over a period. It also helps to gauge the future directions by way of pointers received from its analysis. Data mining plays an important role in analysis of structured information encoded in patents by means of unravelling the major players, understanding the origin of inventions, technology trends etc. The present study revealed that the last five years witnessed increased patenting activity related to the use of machine learning in animal husbandry and China is the dominant player to patent such inventions. The analysis suggested that machine learning is mostly used for monitoring and identification of livestock. In recent year machine learning also find use in diagnostics such as identifying fish parasites and in appetite forecasting.

In future study, patent mining can be used to identify white spaces in domain and citation analysis can be used to determine valuable patents.

REFERENCES:

- Chen, H.S., Zhang, G.Q., Zhu, D.H., Lu, J., 2017. Topic-based technological forecasting based on patent data: a case study of Australian patents from 2000 to 2014. *Technol. Forecast. Soc. Chang.* 119 (7), 39–52.
- Deshpande, N., Ahmed, S., & Khode, A. (2016). Business intelligence through patinformatics: A study of energy efficient data centres using patent data. *Journal of Intelligence Studies in Business*, 6(3).
- Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, 25(3), 233-242.
- Grzesiak, W., & Zaborski, D. (2012). Examples of the use of data mining methods in animal breeding. In A. Karahoca (Eds.), *Data mining applications in engineering and medicine* (pp. 303-324). intechopen. DOI: 10.5772/50893

- Ju, Y., & Sohn, S. Y. (2015). Identifying patterns in rare earth element patents based on text and data mining. *Scientometrics*, 102(1), 389-410.
- Li, X., Xie, Q., Jiang, J., Zhou, Y., & Huang, L. (2019). Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology. *Technological Forecasting and Social Change*, 146, 687-705.
- Li, X., Zhou, Y., Xue, L., Huang, L.C., 2015. Integrating bibliometrics and road mapping methods: a case of dye-sensitized solar cell technology-based industry in China. *Technol. Forecast. Soc. Chang.* 97, 205–222.
- Majumdar, J., Naraseyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big data*, 4(1), 1-15.
- Noh, H., Jo, Y., Lee, S., 2015. Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Syst. Appl.* 42 (9), 4348–4360.
- Shaikh, S. A., & Singhal, T. K. (2018). An analysis of IP management strategies of ICT companies based on patent filings. *Journal of Intelligence Studies in Business*, 8(2).
- Shaikh, S. A., Deshpande, N. A., & Khode, A. (2020). Use of AI for Manuscript Writing—A Study Based on Patent Literature. *Allana Inst of Management Sciences, Pune*, 10, 1-8.
- Shaikh, S.A. and Singhal, T.K. (2019) Study on the various Intellectual Property Management Strategies used and implemented by ICT firms for Business Intelligence. *Journal of Intelligence Studies in Business*. 9 (2) 30-42.
- Singh, P. (2012). Utilizing Patent Information as a Data Mining Tool for Research in Agricultural Sector. *Available at SSRN 2070558*.
- Vincent, C. L., Singh, V., Chakraborty, K., & Gopalakrishnan, A. (2017). Patent data mining in fisheries sector: An analysis using Questel-Orbit and Espacenet. *World Patent Information*, 51, 22-30.
- Wang, S. J. (2011). The state of art patent search with an example of human vaccines. *Human vaccines*, 7(2), 265-268.
- Wang, X.F., Qiu, P.G., Zhu, D.H., Mitkova, L., Lei, M., Porter, A.L., 2015. Identification of technology development trends based on subject–action–object analysis: the case of dye-sensitized solar cells. *Technol. Forecast. Soc. Chang.* 98, 24–46.
- Wu, C.C., Leu, H.J., 2014. Examining the trends of technological development in hydrogen energy using patent co-word map analysis. *Int. J. Hydrog. Energy* 39 (33), 19262–19269.
- Yoon, J., Park, H., Kim, K., 2013. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics* 94 (6), 313–331.
- Zheng, X., Aborisade, M. A., Liu, S., Lu, S., Oba, B. T., Xu, X., ... & Ding, H. (2020). The history and prediction of composting technology: A patent mining. *Journal of Cleaner Production*, 276, 124232.
- Author 1 :** Alok Khode works as Principal Scientist at CSIR - Unit for Research and Development of Information, Pune, India. He is having more than 20 years of experience. His research interests are mainly but not limited to information retrieval, knowledge management and open access. He has master's degree in computer application from DAVV Indore and pgCert from Cranfield University, UK. He has been awarded Chevening Cyber Security Fellowship by UK government as well as SIS Fellowship by Society on Information sciences for his contribution towards open access movement in India.
- Author 2:** Nishad Deshpande is a Senior Scientist at CSIR Unit for Research and Development of Information, Pune, India. He is having more than 15 years of experience in Life sciences and as a patent analyst. He holds a PhD from Symbiosis International University and is an expert in the field of patinformatics. He has been awarded SIS young information Scientist by Society on Information sciences for his contribution towards data analytics and patinformatics.
- Author 3:** Shabib Ahmed Shaikh is working as a Senior Technical Officer at the CSIR-Unit for Research and Development of Information Products. He has more than 20 years of work experience in the field of Information Technology. He holds a PhD in Intellectual Property Management Strategies. He is a graduate in Electronic Science and a Postgraduate in Computer Management. He was awarded a merit for M.Phil. He is actively working in the domain of Patent Analytics, Data Science, Open Access, Institutional Repositories', and their Harvesting