University of Pisa

Data Science and Business Informatics

Strategic and Competitive Intelligence

Prof. Antonella Martini

Prof. Filippo Chiarello

Andriotto Alessandro

Ciompi Marco

Di Tommaso Antonio

Gorrasi Pasquale

Academic year 2019/2020

INDEX

FIGURE INDEX

TABLE INDEX:

RESEARCH QUESTION 1: What are the technologies and the emerging applications in precision agriculture

1.1 : EXECUTIVE SUMMARY

The main topic of our project work is the *precision agriculture*, a new way of doing farming. But what is exactly this precision agriculture?

'Precision agriculture is a management strategy that gathers, processes and analyses temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production'[1].

Precision agriculture support a sustainable way to do agriculture, which is nowadays a trend topic since are raising movements who claims a less heavy human footprint for our ecosystems.

Those new strategies developments involve not only the environmental respect, but also a concrete help for farmers, that make possible a productivity improvement, minimizing costs (e.g. by using pesticides and fertilizer in a more efficient way and only where it is needed). To do this, new technologies are a key factor as well as the development of industrial sectors that are mixing different competencies, in order to face new problems. This is specifically the purpose of our project work, to find out what kind of new technologies are being developed by industrial sectors and how different industrial sectors are being combined, in order to reach the goals mentioned above.

The project is composed by different research questions: the first one aims to have a foresight on which are the affirmed technologies and the emerging applications. In the second research question we try to help Vitibot, a French start-up working in this field, providing it a patent landscape report of its product to have an overview of possible partners and competitors. With the third research question we want to analyse three companies, Vitibot, VinBot and Agrobot, in order to grasp their positioning in the market and understand if they are potential competitors, or they could collaborate in the future.

1.2 : CONTEXT OF THE PROBLEM

Agriculture is changing thanks to the shift from traditional agriculture, based on traditional machines, to a new concept of farming, defined as precision agriculture. This evolution is due to social changes as well as technological development. Social changes have been influencing this evolution focusing on environmental respect: people start to pay more attention to the use of pesticides and fertilizer in farming process, taking care of our Earth and minimizing soil pollution.

Precision agriculture, also defined as "Agriculture 4.0", is a mix of agriculture and technological development (e.g. IoT, IT and Big Data) with the aim of creating new farming management strategies to improve business efficiency. Since this is a rich emerging field, some firms are interested in it to get an economic revenue. To do this, they invest in their R&D branches by developing new products in order to reach a good positioning in the market and to get competitive advantages. For this reason, they are creating new patents to protect their new ideas.

We must define what is a patent. 'A Patent is the right provided to an inventor not for the use or practice of the invention but for preventing others from practicing or using the invention. In the recent years the role of patents has changed dramatically. Patents used to be the concern of only a bunch of legal practitioners or specialists and no one else was really bothered. But with changing times and ever-increasing competition in business, patents have become a key factor for any business' [2].

For the purpose of this research we have used Patent Analysis, a functional tool to address the strategic management of the firm's technology as well as product or service development process. So, we got some patents on precision agriculture and through patent analysis we found out what are the affirmed technologies.

1.3 : METHODOLOGY AND WORKFLOW

For the purpose of our analysis, we started reading some scientific papers in order to obtain knowledge about patents in general and the methodologies that are involved to analyse them. Some of the papers will be cited during our discussion. After gaining the necessary information, we started projecting our workflows for the assigned research question. We followed two different approaches in order to find affirmed technologies and the emerging ones. To clarify, we are going to call RQ1 A for affirmed technologies and RQ1 B for emerging applications on precision agriculture.

1.3.1 : METHODOLOGY RQ1 A

To better grasp the methodology, first we need to explain the workflow shown in Figure 1.

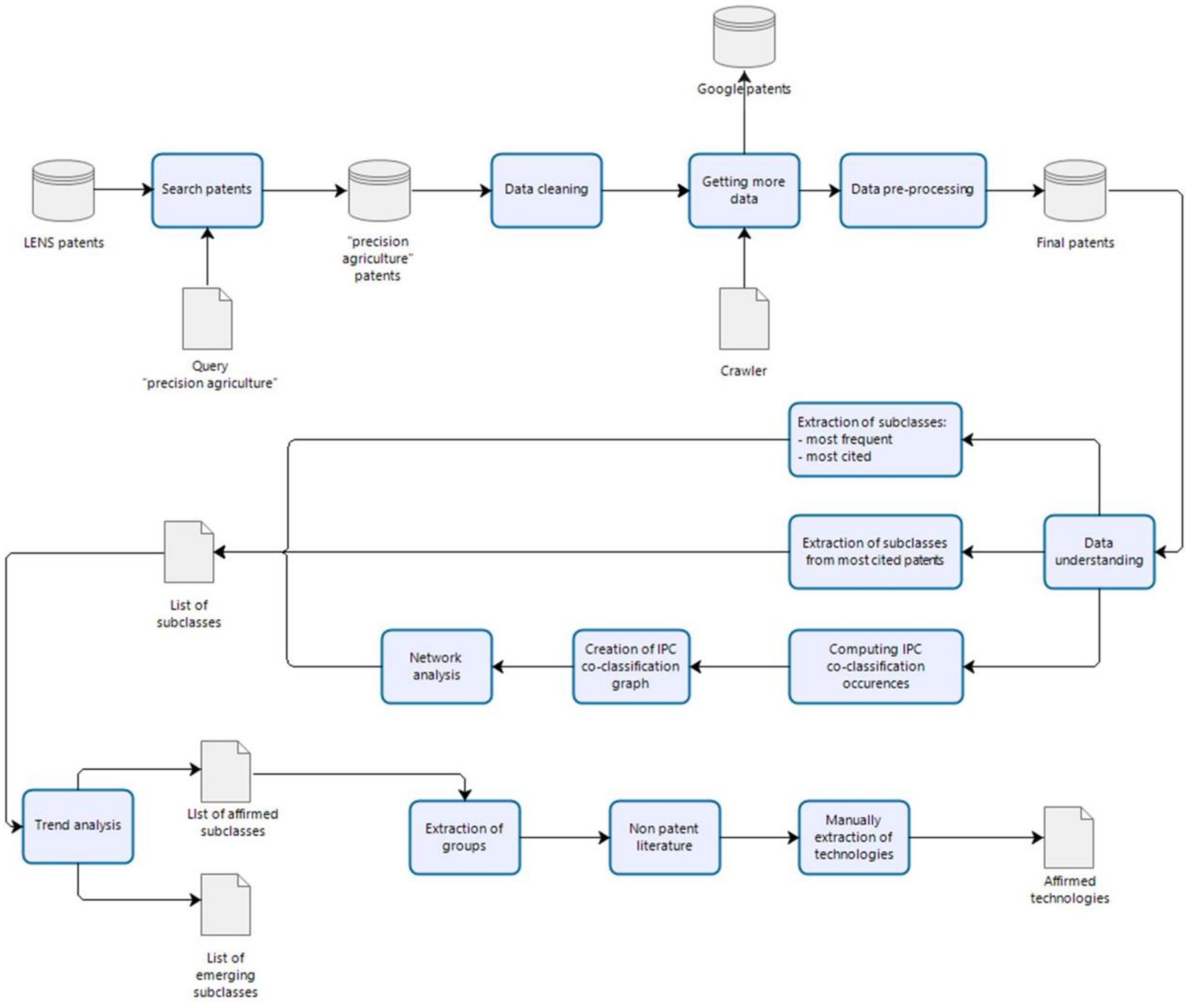


Figure 1: Workflow RQ1A

Firstly, we started defining the main problem and from there we collected patents about precision agriculture through Lens databases. We used the query ‘precision agriculture’ getting only patent applications. We obtained a dataset of 84039 records with 30 features, we carried out a data cleaning and pre-processing on it.

In data cleaning phase we removed records with missing values and duplicates as well as some irrelevant features until we obtained a dataset composed by 75243 rows and 5 columns, that can be summarized in:

- **Publication_Number**: it represents the unique publication number assigned to a patent when it is released
- **Publication_Year**: the publication year of the patent

- **Title**: the title assigned to a patent
- **Applicants**: the entity or person which or who presents an application for the grant of an industrial property right
- **IPCR_Classifications**: a list of all the classification ids that are assigned to a single patent.

In addition to the data cleaning phase, we corrected applicants' names removing orthographic errors thanks to FuzzyWuzzy library applying it on the first 300 applicants that released more patents.

In order to proceed with our analysis, we implemented a data crawler to catch more data. The data crawler is implemented in Python and, using BeautifulSoup library, it scans data from Google Patents by searching for the publication number of each patent. Thanks to this tool, we reached informations about:

- **Abstract**: represents the written content of the patent
- **Claims_count**: number of patent's claims
- **Citation_count**: number of patent's received citations
- **Citation_ids**: list of the publication numbers that cited the patent
- **Citation_dates**: is the list of dates when patents in Citation_ids cited the patent
- **Cited_count**: number of patents cited by the patent
- **Cited_ids**: list of publication numbers cited by the patent
- **Cited_dates**: list of dates when the patent cited the patents in Cited_ids.

The core of our analysis is based on IPC classification codes, because "patents are oriented towards the legal protection of technologies and therefore the classification of patents is based on technologies or products which use specific technologies"[3]

In every patent, multiple IPC classification codes are assigned. The IPC classification is a hierarchical alphanumeric code representing the characteristics of a patent. More precisely, the first three-digits of IPC code represent the class, the first four-digits code represent the subclass and the code until the slash represents the group.

Starting from those definitions, during data pre-processing phase we assigned the class and the subclass to each patent, generating two new features:

- Main_Class
- Main_Subclass

To do this, for each patent we took the list of IPC codes, and we extracted class and subclass from each patent, assigning at the two features cited above the most frequent IPC value.

At the end of these processes we obtained a final dataset composed by 72224 records and 15 columns. We got a different number of records with respect to the initial dataset, because the data crawler cannot catch data from every patent, so again we removed missing values.

After this, we began the data understanding phase to have an overview on our dataset (i.e. we searched for actors, countries, and so on).

Then, we decided to extract a list of subclasses representing only the technological sectors. Going deeper on the analysis, we also mined groups coming from subclasses to identify technologies. To find subclasses, we followed three different approaches:

- *extraction of most frequent and most cited subclasses*
- *extraction of subclasses from the most cited patents*
- *Network analysis based on the occurrences of co-classification IPC codes.* To better explaining our analysis, we need to specify what is co-classification: it is the set of all IPC codes which is assigned to a patent. Based on this we counted the occurrences of each co-classification IPC codes' couple for each patent. After this we created an undirect network, where nodes represent the technological sectors

and arcs represent the links between them. Each arc is weighted by occurrences' count. We considered an indirect graph because, for instance, the link between C12N and A61K is the same of A61K and C12N. We decided to proceed in this way because co-classification is a potential indicator of linkages among technological sectors and technologies. More precisely, we analysed the network thanks to Gephi software, considering two measures:

1. *Betweenness Centrality* that measures the node's centrality in the network; in other words, it indicates how many paths are going through each node. Nodes with high betweenness centrality have considerable influence in the network.
2. *Degree Centrality* that measures the counts of how many neighbours a node has. A higher degree implies that a node is significant.

At the end of these three approaches, we obtained a list of subclasses representing the most meaningful technological sectors. Based on this list, we executed a trend analysis, showing how many patents are published every year for every technological sector. We performed this trend analysis because "a rapid increase in patent applications filed related to a specific technology can represent its dissemination into new and often unrelated technological areas, which signals technology emergence and industry acceptance". [4]

Of course, if the increase of patents' publication it did not happen in the last few year, we consider that technological sector as an affirmed one, otherwise, we classify it as an emerging one. Based on this assumption, we generated two different lists of subclasses, one containing affirmed technological sectors and the other emerging technological sectors. The list of emerging technological sector will be used for RQ1 B, to mine emerging applications in precision agriculture.

For the selected subclasses, we extracted each group code from IPC classification since it can represent the specific technology related to a technological sector.

The basic idea is to manually analyse the meaning of these groups in order to find out technologies through Lens Classification Explorer. However, we first had to get some information about what kind of technologies are nowadays involved in precision agriculture. We did it analysing the non-patent literature, surfing different websites and watching videos. After this analysis, we obtained affirmed technologies, that are in the "OutputFiles" folder, as also the lists of extracted subclasses.

1.3.2 : METHODOLOGY RQ1 B:

The methodology to extract emerging applications follows the workflow shown is Figure 2. Our main idea is to find which are emerging technologies. We began from the data understanding phase of RQ1 A, and we implemented an approach based on "The Emerging Cluster Model".

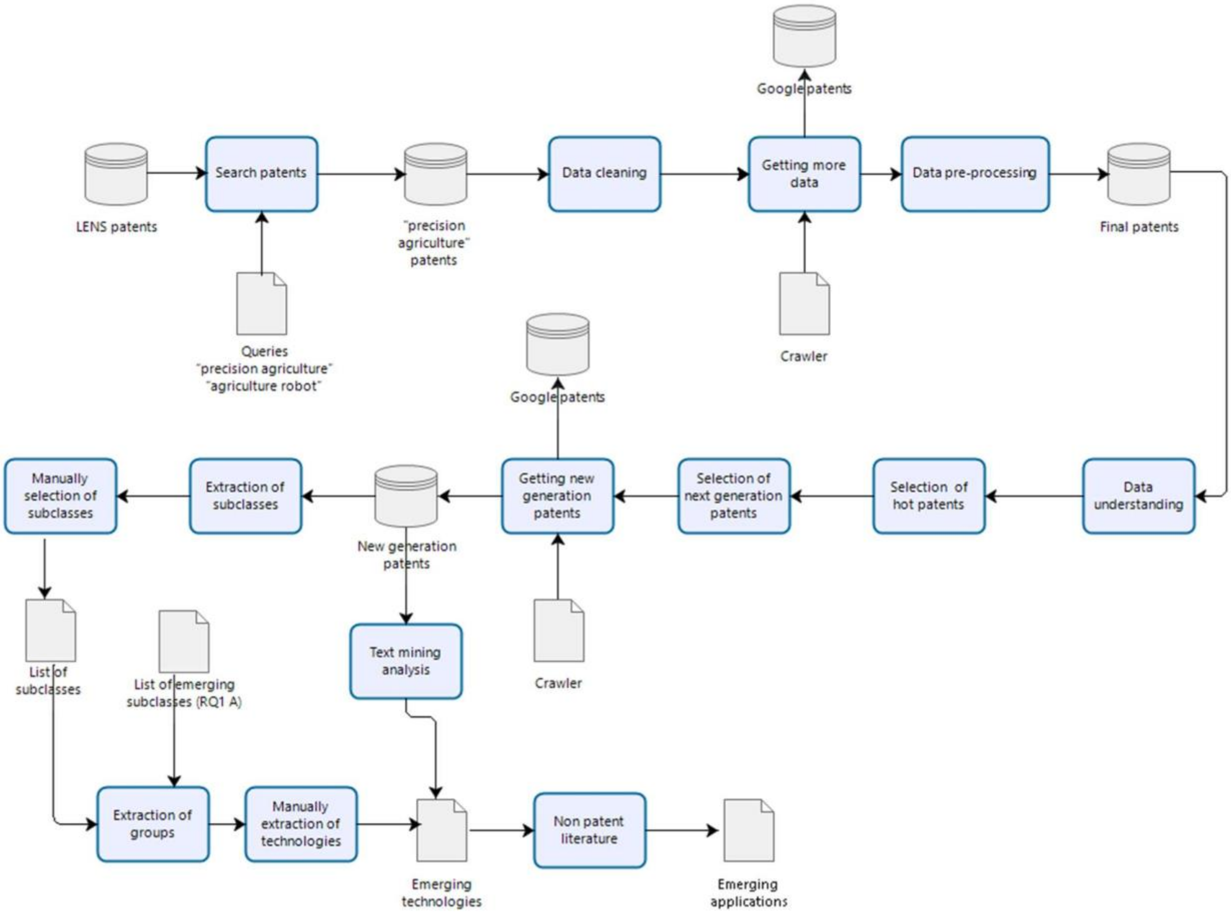


Figure 2: Workflow RQ1B

The *Emerging Clusters Model* is based on patent citations techniques, designed to identify what is coming out lately. The basic idea behind patents citation analysis is that highly cited patents tend to contain relevant technological information because they form the basis for many new innovations, so they are frequently cited by later patents.

So, the purpose of our model is to identify hot patents as well as next generation patents in order to extract the emerging technologies, and from these emerging applications checking the non-patent literature.

- Hot patents are being cited by large number of patents, mainly in the most recent period.
- Next generation patents cite hot patents in the most recent time.

Let's analyse in detail the definition of hot patent. Firstly, we have to say that a patent can be hot in a certain year rather than others (i.e. it can be hot in 2012, 2014, 2015 but not in 2013).

A patent P is classified as hot patent if it satisfies two criteria in the year T:

P is highly cited by patent released in the most recent period. To determine if P is highly cited in the year T, we used a simple threshold based on the citation count. Specifically, P is a hot patent in the year T, if it belongs to the top 10% of patents being referenced by patents released in years T and T-1.

Citations from recent patents must represent a high percentage of the total citations it has received. To determine it, we check if the percentage of citation received in the years T and T-1 is greater than a certain threshold that will be explained below. The percentage is simply the ratio between the number of citations P received in the years T and T-1 and the total number of citations received.

We are not limiting to check if it is just a high value. The reason is that a candidate hot patent can be of any age. Indeed, a recently released patent is more likely to have a high percentage of its citations from recent patents, with respect to a patent that has been accumulating citations for many years.

The threshold used for checking is the following linear function that consider the age effect of cited patents:

$$threshold = (-\frac{1}{80}) * age\_of\_patent \; + \; \frac{5}{8}$$

where *age_of_patent* represents the age of the candidate patent, computed as the difference between the current year and the age of publication of the patent. For a detailed explanation of the coefficients concerned, we refer to the lecture to the scientific paper "The Emerging Clusters Model: A tool for identifying merging technologies across multiple patent systems". [5]

Once a hot patent is defined for a certain year T, we identified its next generation patents, that is a list of patents which cite the hot patent in year T or T-1. After having obtained next generation patents using the same data crawler of RQ1, we obtained data about them from Google Patents.

We built up a new dataset for next generation patents with a shape of 1219 record and 6 columns composed by the following features: Publication_Number, Title, Abstract, IPCR_Classification, Main_Class, Main_Subclass. More precisely, we took the Abstract of the next generation patents with the data crawler; the last two features are instead assigned with the same methodology of RQ1 A.

At this point we made two different analyses:

- **text mining analysis**: it consists in analysing the most frequent trigrams extracted from title and abstract, because usually these two fields cite the protected technology. Once titles and abstracts are extracted, we removed the stop words (i.e. is, am, this, the, an etc.) considered as noises in the text, using NLTK library. After that we did lexical normalization, that consider another type of noise in the text; similar words are reduced to a common word (es. connection, connect, connectivity as 'connection"). We applied *stemming*, a linguistic normalization process, which reduces words to their word root or chops off the derivational affixes. After that, we applied *lemmatization* process, to reduce words to their base word, which is linguistically correct lemmas. It transforms root word using vocabulary and morphological analysis. At the end of this pre-processing phase we extracted our trigrams.
- **IPC classification analysis**: to understand which technological sectors are involved on emerging technologies. This analysis is implemented facing as first step the 'extraction of subclasses' as we made for RQ1 A. After this, we selected the subclasses checking their meaning on Lens Classification Explorer. Then we did 'extraction of groups' and 'manually extraction of technologies' always following the same methodology implemented for RQ1 A. As we can see from the workflow in Figure 2, we can notice that after the extraction of subclasses, we integrated the result with the emerging subclasses obtained from the trend analysis of the RQ1 A.

From the results obtained by those two approaches, we generated a file containing emerging technologies. This file is available on "OutputFiles" folder. Then we checked the non-patent literature comparing what it emerges from our analysis to find out new applications.

1.4: RQ1 A ANALYSIS: AFFIRMED TECHNOLOGIES' MINING

We started plotting a bar chart representing the number of released patents for each year, till obtaining the plot shown in Figure 3.

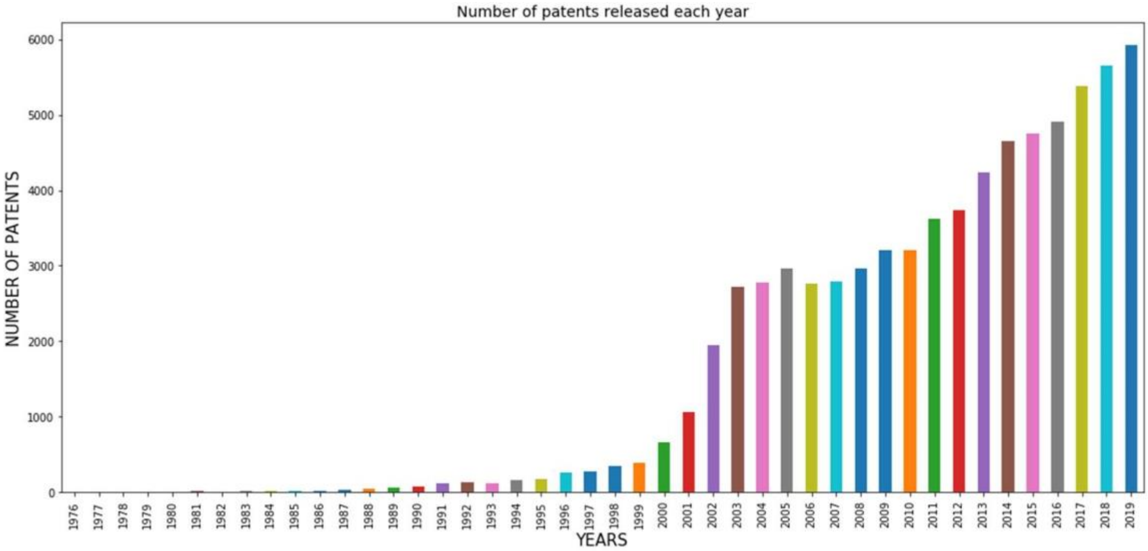


Figure 3: Patents' publication per year

We noticed that the number of released patents increases in the last years. For this reason and for the purpose of the first research question we decided to remove patents published before 2000.

After removing these patents, we performed an actor analysis based on the most important applicants. We have chosen Applicants' field because an applicant is the entity or person who presents an application for the grant of an industrial property right (normally a company or an organization).

To select an appropriate number of applicants, we used the *long tail distribution theory* that suggests selecting a threshold in correspondence of the value where distribution's trend change, which in this case is indexed by the red line, as we can notice from Figure 4. From this point, when we will speak about "long tail distribution theory" we will talk about the same methodology using different threshold values based on the distribution of the case.

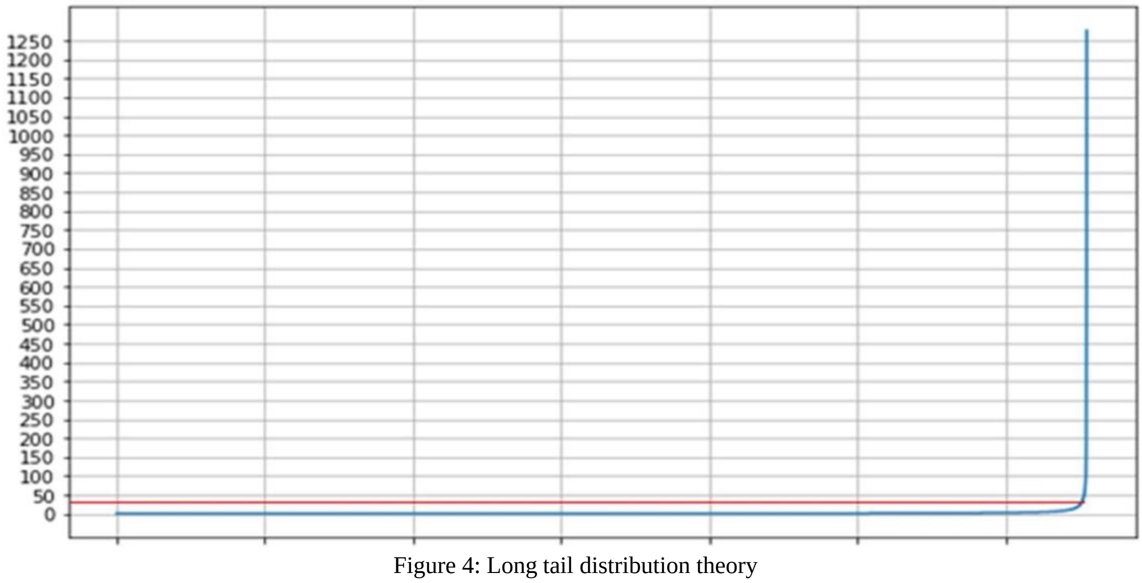


Figure 4: Long tail distribution theory

We did it because we saw that there are a lot of applicants who released few patents that are irrelevant for our analysis. So, we obtained the top applicants that represent the 10,7% of the total number of patents in our dataset, as shown in Figure 5.

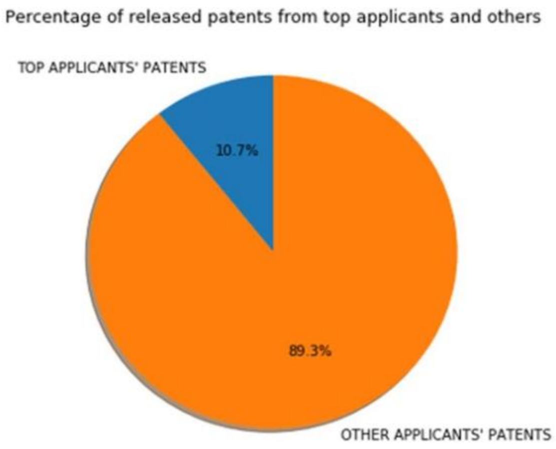


Figure 5: Top applicants/other applicants' pie chart

We analysed firms composing these top applicants as well as their nationality to make actor and country analysis. To perform actor analysis, we plotted the distribution of the number of patents filed by the top applicants, as shown in Figure 6.

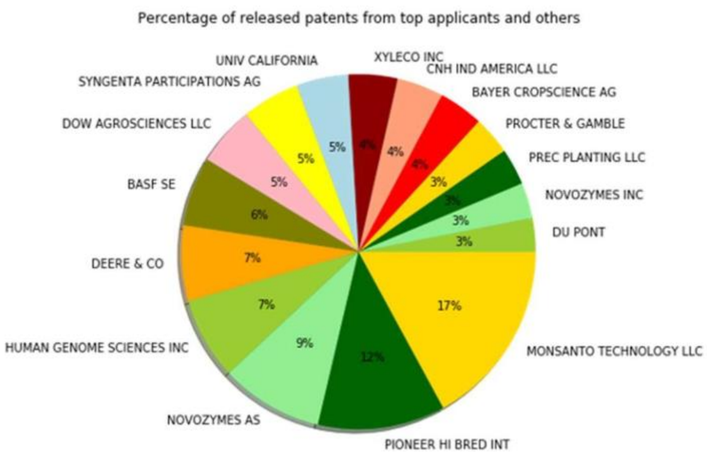


Figure 6: Pie chart actor analysis

This plot is derived from Table 1, representing the number of patents released by each top applicant in a more accurate way.

MONSANTO TECHNOLOGY LLC	1276
PIONEER HI BRED INT	869
NOVOZYMES AS	701
HUMAN GENOME SCIENCES INC	557
DEERE & CO	514
BASF SE	471
DOW AGROSCIENCES LLC	401
SYNGENTA PARTICIPATIONS AG	393
UNIV CALIFORNIA	355
XYLECO INC	366
CNH IND AMERICA LLC	317
BAYER CROPSCIENCE AG	305
PROCTER & GAMBLE	261
PREC PLANTING LLC	250
NOVOZYMES INC	242
DU PONT	229

Table 1: list of applicants from which figure 6's pie chart is derived

After this, we looked for every Applicant on Web, in order to understand their nationality and to perform country analysis, till obtaining the plot in Figure 7.

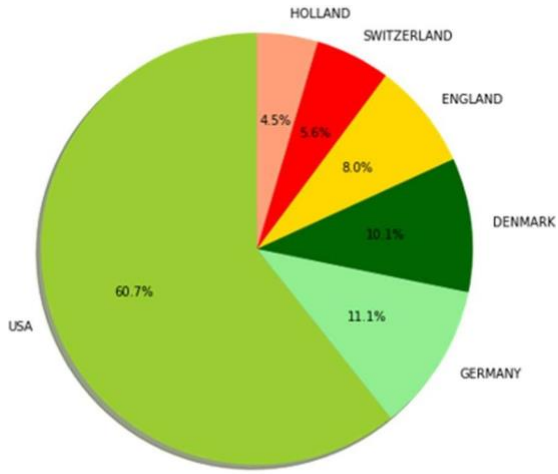


Figure 7: Top applicant's countries

From this first analysis we can see that firms who invest in precision agriculture are operating in the sector of agrochemical (focusing on the production of herbicides, fungicides, pesticides, hybrid seeds etc.), agriculture biotechnologies, biopharmaceutical and agriculture machinery.

Regarding countries, we can see that the most important countries involved in this field are USA, Germany, Denmark, England, Switzerland, Holland.

Then, as we told on the methodology, we extracted a list of subclasses representing the technological sector, following three different approaches:

1. Most frequent and most cited subclasses
2. Subclasses belonging to the most cited patents
3. Network analysis based on the occurrences of co-classifications

1.4.1 : Most frequent and most cited subclasses

Regarding the first approach, we divided our work in most frequent subclasses' analysis and most cited subclasses analysis.

In order to perform the *first step*, we counted how many patents were released for each subclass, so as to have an idea of what are the most meaningful technological sectors. After counting, we sorted the subclasses based on the number of occurrences and then we applied the long tail distribution theory to select the most relevant subclasses. The result of this process generated the following subclasses:

'C12N', 'A61K', 'A01H', 'G01N', 'A01N', 'C07K', 'C12Q', 'C12P', 'A01G', 'A01C', 'G06F', 'C07D', 'G06Q', 'A61P', 'A01B', 'C02F', 'A23L', 'A01D', 'C07C', 'B01D', 'B01J', 'G06K', 'G01S', 'A01K', 'C08L', 'A61B', 'G05D', 'C08F', 'B32B', 'A23K', 'B05B', 'C12M', 'A01M', 'C40B', 'G05B', 'C07H', 'C10L', 'A61L', 'G06T', 'C08K', 'B65D', 'B29C', 'C08G', 'H04L', 'B62D', 'G01J', 'C05F', 'A01F', 'G01C', 'C11D', 'B01L', 'C05G', 'H01L', 'H04W', 'F16H', 'A01P'

In the *second step* we considered the most cited subclasses, since we thought that from a highly cited subclass, we could have extracted a technology that is used in a wide range of technological fields. For instance, the technology of GPS is used for smartphones, cars, marine instruments etc. so we can consider GPS as an affirmed and largely used technology.

Going deeper in our analysis, we counted how many citations each subclass received. Then we sorted them, and we applied the long tail distribution theory in order to highlight the most relevant subclasses.

What is emerged is the following list of subclasses:

'C07H', 'A61L', 'C08F', 'B01J', 'C08L', 'H04L', 'C07C', 'H05B', 'A01B', 'H01L', 'A23L', 'G05D', 'G01S', 'E21B', 'C40B', 'C02F', 'H01F', 'G06K', 'B01D', 'A01C', 'A61B', 'A61P', 'A01G', 'C07D', 'H02J', 'C12P', 'A01H', 'C12Q', 'A01N', 'G06Q', 'C07K', 'G06F', 'G01N', 'A61K', 'C12N'.

1.4.2 : Subclasses belonging to the most cited patents

Based on the same idea exposed in the second step of the first approach, we considered the most cited patents and from them we extracted their subclasses applying long tail distribution theory. The result is the following list of subclasses:

'C12N', 'H02J', 'G06F', 'G06Q', 'C07K', 'H01F', 'A61K', 'A61B', 'H05B', 'H01L', 'A01H', 'C07D', 'H04L', 'C12Q', 'A01N', 'G01N', 'B01D', 'H01P', 'G01C', 'G06K', 'B60Q', 'G08B', 'A63F', 'A61L', 'G05D', 'B01F', 'B60L', 'G10L', 'C09D', 'B24D', 'B24B', 'G05B', 'G09G', 'H03B', 'A61P', 'G01P', 'A23K', 'H04W', 'H01B', 'A61F', 'C08F', 'G09B', 'G06G', 'C08B', 'H04N', 'C40B', 'F24S', 'H04Q', 'G01S', 'C12P'

1.4.3 : Network analysis based on the occurrences of co-classification

As we explained in the methodology section, we built a network of subclasses based on the co-classification occurrences of our dataset.

We computed the count of each couple of co-classification IPC codes and we took only the most meaningful links using the long tail distribution theory. From these data, we built up the network by means of Gephi software and we analysed it using two different measures: Betweenness Centrality and Centrality Degree.

- *Betweenness Centrality*: the output of this measure is the network in Figure 8. We selected subclasses with betweenness centrality different from zero, obtaining the following list: 'A01C', 'A01B', 'G01N', 'C12Q', 'C12N', 'C07K', 'A23L', 'A01N', 'A01H', 'C08L', 'C08J', 'C12P', 'C40B', 'C07H', 'A61P', 'A01G', 'A01D', 'G06F', 'G06K', 'H04L', 'C07D', 'C07C', 'A01P', 'C05G', 'A61K', 'C10L', 'C08G', 'G05B', 'B01J'

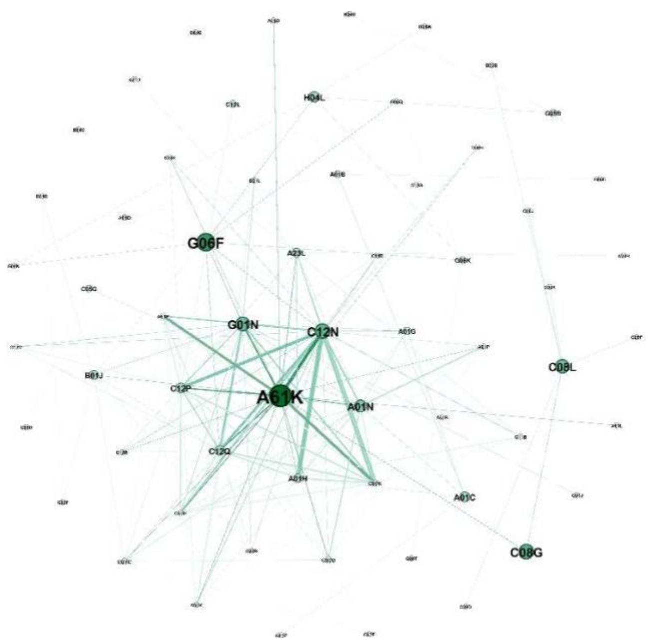


Figure 8: Network analysis with betweenness centrality

- *Centrality Degree*: the output of this measure is the network in Figure 9. Since centrality degree can assume lot of values, we selected subclasses using long tail distribution theory, obtaining the following subclasses list: 'C12N', 'C07K', 'A61K', 'C12Q', 'C12P', 'G01N', 'A01H', 'A61P', 'C07H', 'A01N', 'A23L', 'C07D', 'A01K', 'A01P', 'C11B'

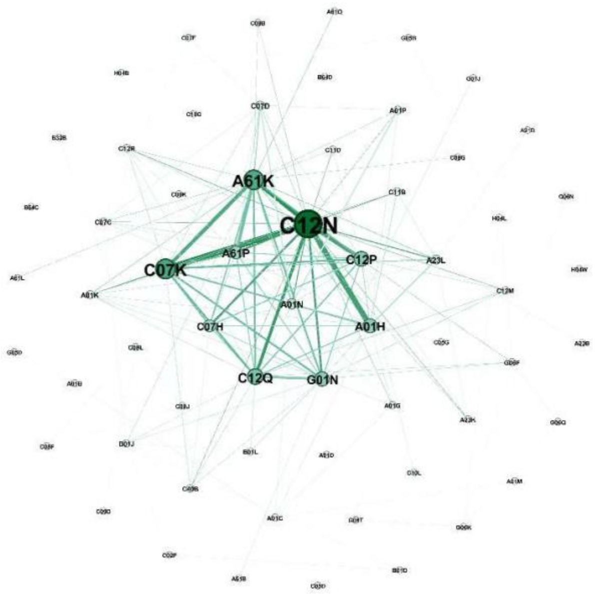


Figure 9: Network analysis with Centrality Degree

At the end of these three approaches, we merged the lists obtaining this final list of subclasses: 'G09G', 'C10L', 'C02F', 'H01P', 'C07H', 'C12P', 'H04Q', 'B62D', 'B01L', 'H04N', 'A01G', 'G09B', 'A01D', 'F16H', 'A61L', 'B65D', 'C12M', 'C08K', 'A61B', 'C07K', 'C08L', 'G06K', 'A01N', 'G01S', 'B32B', 'C11B', 'A01C', 'B01F', 'G06G', 'A01M', 'G01P', 'G06T', 'A01F', 'H01F', 'C08G', 'G06F', 'A61P', 'B05B', 'E21B', 'B01J', 'A23K', 'C08B', 'C12N', 'A23L', 'H01B', 'A63F', 'G01C', 'C09D', 'H05B', 'F24S', 'H04L', 'H02J', 'G01J', 'H01L', 'C05G', 'C11D', 'G05B', 'B24D', 'B24B', 'A01H', 'B01D', 'H03B', 'C12Q', 'A01B', 'C08F', 'G06Q', 'G01N', 'C40B', 'A01P', 'G08B', 'A61F', 'B60Q', 'B29C', 'B60L', 'C05F', 'G05D', 'H04W', 'G10L', 'C07C', 'C08J', 'A01K', 'C07D', 'A61K'.

We performed the trend analysis based on this list in order to identify what are the affirmed technological sector and what are the emerging ones.

1.4.4 : TREND ANALYSIS

By plotting the number of patents released for each subclass, we have classified as emerging technology the ones that has registered growth in the last few years; otherwise they have been classified as affirmed.

We deleted subclasses having a decreasing trend and subclasses having few patents' publications. To make this concept clearer, we show some examples:

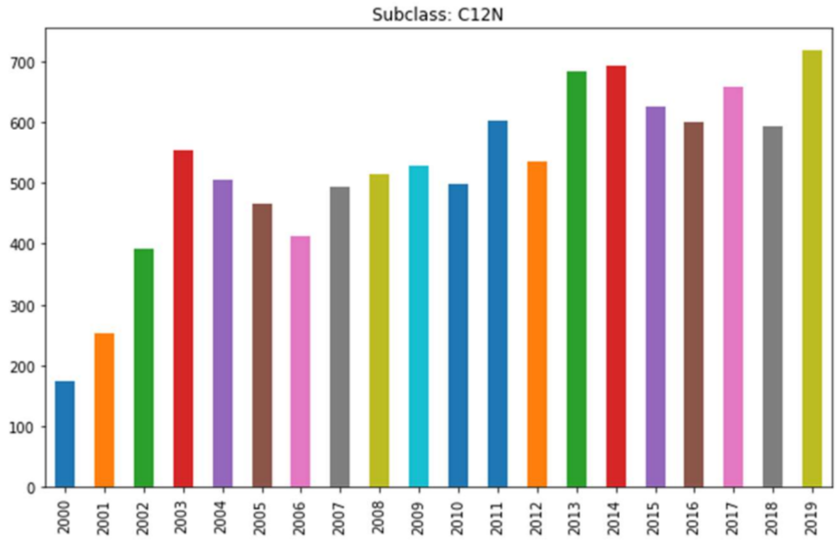


Figure 10: Affirmed technological sector's trend example

Figure 10 shows the trend of subclass C12N representing "microorganisms or enzymes compositions thereof biocides, pest repellents or attractants, or plant growth regulators". We consider this subclass as an affirmed technological sector because it increases from 2001 to 2003, but then its trend becomes quite constant in terms of number of released patents.

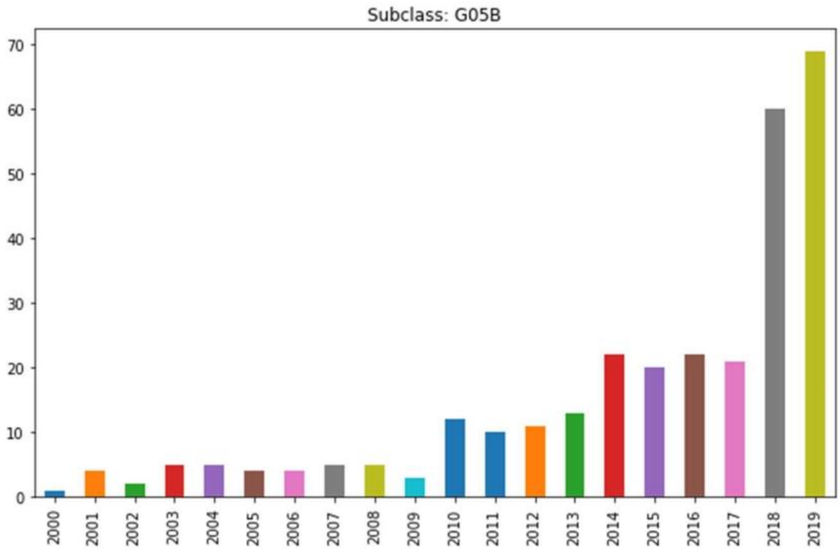


Figure 11: Emerging technological sector trend's example

Figure 11 shows the trend of subclass G05B, that is "control or regulating system in general". We consider this subclass as emerging because of its growth during the last two years.

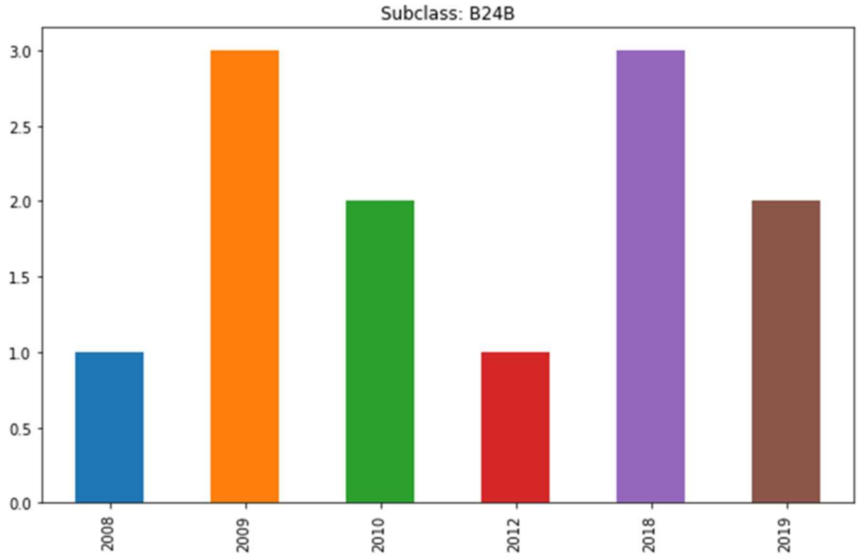


Figure 12: Example of discarder technological sector

Figure 12 shows the trend of subclass B24B representing "machines, devices, or processes for grinding or polishing". We did not consider this class because there are few publications.

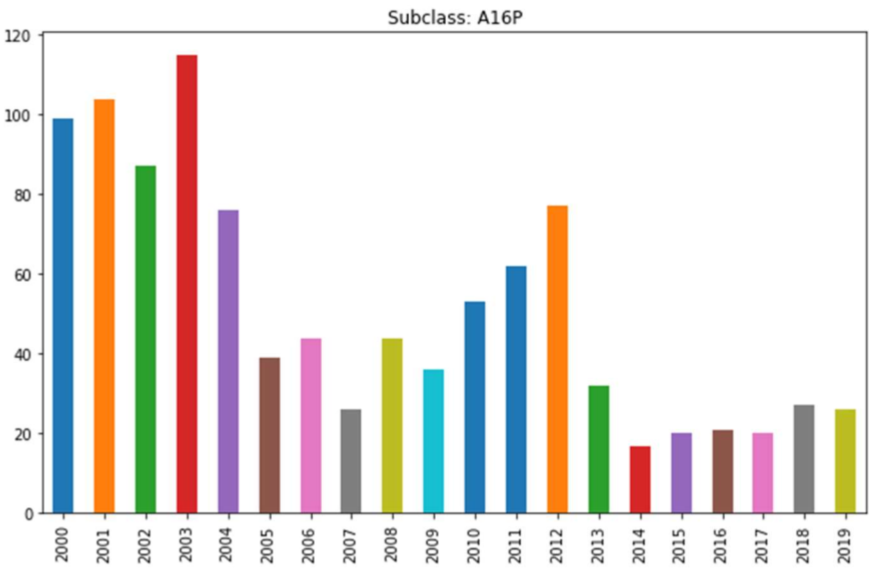


Figure 13: Decreasing technological sector's trend example

Figure 13 shows the trend of subclass A61P, that is "specific therapeutic activity of chemical compounds or medicinal preparations" containing the use of antibiotics and similar compounds. We can notice that there is a decreasing trend due to a more restrictive use of these compounds in agriculture.

At the end of the trend analysis we obtained the following list of subclasses:

'C12N','A61K','G01N','A01H','A01N','C07K','C12P','G06Q','C12Q','A01C','G06F','A01G','A01B','C02F','A01D','C07D','A61B','B05B','A23K','B62D','A01F','C12N','G01N','A01N','C12Q','A01C','G06F','A01G','A01B','C02F','A01D','A61B','B05B','H02J','H01L','G06F','A61B'.

1.4.5 : EXTRACTION OF GROUPS

From the list of selected subclasses, we extracted groups to identify the technologies. As we told in the methodology section, from this list of groups we extracted the affirmed technologies checking the non-patent literature and the meaning of each group on Lens Classification Explorer. We reported some examples of what emerged below. We are not reporting all the results, because all of them can be found in “Affirmed Technologies” file.

From the subclass A01C (representing the technological sector of soil-working) we extracted the following groups:

- A01C7, that is “Sowing with centrifugal wheels A01C17/00 arrangements for driving working parts A01C19/00 representing sowing machines.
- A01C23, that is “Distributing devices specially adapted for liquid manure or other fertilising liquid”

From the subclass A01G (that represent the technological sectors of horticulture cultivation) we extracted the following groups:

- A01G25, that is “watering gardens, fields, sports grounds or the like special apparatus”, (i.e. irrigators)
- A01G9, that is “cultivation in receptacles, forcing-frames or greenhouses” representing greenhouses.

From the subclass B05B (that represents spray apparatus) we extracted the group:

- B05B1, that is “nozzles, spray heads or other outlets, with or without auxiliary devices”.

1.5 : RQ1 B ANALYSIS: EMERGING TECHNOLOGIES MINING AND THEIR APPLICATIONS

As we told before, RQ1 B is dedicated to emerging applications, to obtain next generation patents. From their analysis we want to mine which are technologies involved in emerging applications.

From these patents we conducted two different types of analysis: text mining analysis and subclasses analysis.

1.5.1 : TEXT MINING ANALYSIS

For this analysis we extracted the most frequent trigrams from title and abstract of next generation patents, in order to understand what are the technologies involved in these patents.

In order to cut irrelevant trigrams, we applied the long tail distribution theory until we obtained 143 trigrams that are stored in the ‘List of trigrams.txt’ file. The number beside trigram is the frequency of that trigram on all title and abstracts. We report some examples below:

- Aerosol delivery device
- Wireless power transmission
- Data processing systems
- Wireless power manager
- Microbiocidal oxadiazole derivatives
- Vaporize components aerosol
- Wireless power receiver

So, we can say that emerging technologies deals with wireless power transmission. We can also suppose that there are others aerosol delivery systems as well as new data processing systems.

1.5.2 : SUBCLASSES ANALYSIS

With the same methodology used to extract subclasses used in RQ1 A, we obtained the subclasses for the next generation patents. We performed the long tail distribution theory to these subclasses, but this time we took values under the threshold, since an emerging technology represents a weak signal of changing[6], for this reason, we supposed that an emerging technology should have few patents publication. To these subclasses we added the emerging ones found with trend analysis in RQ1 A. From the merged list, we deleted the subclasses classified as affirmed in the previous analysis, until we obtained the following list of subclasses: 'B64C', 'A01K', 'G08G', 'A24B', 'B29C', 'A63F', 'H01G', 'A24F', 'B65D', 'B60L', 'G01C', 'A61F', 'G06N', 'B65H', 'D04H', 'B32B', 'G06T', 'G09B', 'A47J', 'H02M', 'A23B', 'B65G', 'G07F', 'G07C', 'G02B', 'A63B', 'G05B', 'G05D', 'H04B', 'C07F', 'H04N', 'B60W', 'B65B', 'B05C', 'H04M', 'Y02T', 'A23L', 'H03K', 'C12Y', 'H01M', 'C05G', 'G09G', 'A01M', 'B60K', 'G08B', 'A61M', 'G01R', 'H01F', 'C07C', 'G08C', 'G01F', 'B23P', 'A47K',

'Y02A', 'G16H', 'G16B', 'G06K', 'Y02P', 'B67D', 'A23P', 'A61P', 'G01J', 'H05B', 'B64D', 'H01Q', 'B26D', 'G01G', 'H04L', 'C40B', 'A61C', 'H04W', 'G01S'.

We proceeded to analyse them on Lens Classification Explorer in order to identify what are the subclasses that are most representative for our problem. Indeed, if we selected subclasses under the threshold of the long tail distribution theory, it would be possible to have subclasses not strongly connected to precision agriculture (i.e., the class A63F represents "card, board, or roulette games indoor games using small moving playing bodies video games not otherwise provided for", which is not correlated with our goal).

After selecting the most coherent subclasses with our topic, we obtained these subclasses: 'B64C','H04N','G08G','Y02A','Y02P','G16H','A01B','A01D','G01S','G05B','G06T','H04L','G01C','H04W','A01K','G06K','A01M','G05D','H05B','G06N','C05G','G16B'.

From this list we extracted the groups with the same methodology implemented in RQ1 A. After this, we checked every group on Lens Classification Explorer in order to find out emerging technologies. We report some examples of technologies:

From subclass G05B, which is "control or regulating systems in general […] systems or elements fluid-pressure actuators or systems acting by means of fluids" we extracted the group:

- G05B13: "Adaptive control systems, i.e. systems automatically adjusting themselves to have a performance which is optimum according to some preassigned criterion". We think it could be a useful technology to manage the spray area in which pesticides are distributed.

From subclass A01K, that is "animal husbandry care of birds, fishes, insects fishing rearing or breeding animals" we extracted:

- A01K63: "receptacles for live fish"
- A01K61: "culture of aquatic animals' receptacles for live fish" i.e. aquaria.

The reason why we selected these trivial groups of A01K is because there are new management systems growing up and integrating with precision agriculture scenario, as "Aquaponic", that will be discussed in "Discussion and insight" section.

From subclass Y02A, that is "technologies for adaptation to climate change" we extracted the following groups:

- Y02A40, that is "adaptation technologies in agriculture, forestry, livestock or agricultural production"
- Y02A90, that is "technologies having an indirect contribution to adaptation to climate change"

We have chosen these groups for the related subclass (Y02A) because greenhouses are becoming more technological. Indeed, new sensors and new ITC systems can improve climate condition inside the greenhouses looking at the weather outside, to increase production's quality.

At the end of text mining analysis and subclasses analysis we merged the results in the file "Emerging technologies". We use these emerging technologies to find emerging applications, that will be discussed in "Discussion and insight" sections and can be found in "Emerging applications" files in "Output files" folder.

An interesting observation came out during our analysis: two subclasses, and their related technologies, are both affirmed and emerging. Going deeper, observing the trend of released patents we noticed that the number of patents during the years is constant, while in the last years it suddenly raises. The concerned subclasses are 'A01B' and 'A01D' from which we identified the following technologies: harrows, combine harvester, mower, shakers for plants. We need to clarify that all the previous mentioned technologies have been always used in traditional agriculture. The point is that companies are working to make these tools, in order to have a better vision of the work done. Figure 14 shows how an old and affirmed technology is integrated.

Figure 14: Cooperation of an affirmed technology with modern systems

1.6 : DISCUSSION AND INSIGHT

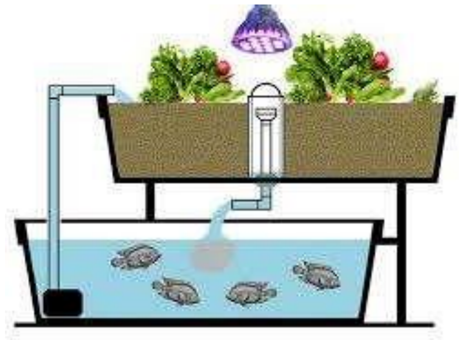
Our analysis started mining which are affirmed technologies applied in “precision agriculture” field.
Firstly, we need to say that the term “precision” related to “agriculture” has been changing over time. Until some years ago, precision agriculture involved different technological sectors, mainly chemistry, mechanical engineering and physics. From our analysis it emerges that affirmed technologies are related to what today we consider traditional agriculture: biocides, pesticides, fertilizers, sowers, combined harvesters, mowers, shakers and so on. With the years passing, technological developments brought improvement in agriculture too: from our technological mining analysis, we also extracted some technologies used nowadays. These technologies are pieces of new management strategies, that transform agriculture in “agriculture 4.0”.

These new management strategies are leading to a more sustainable agriculture thanks to the integration of advanced technologies to bring efficiency and effectiveness in this field. But what kind of new management systems are coming out?
Some new ways in doing agriculture can be summarized with those two new management systems, that are supported by high-tech instrumentation and knowledge:

- Hydroponic
- Aquaponic

Hydroponic (Figure 15) is an emerging field based on soilless cultivation: plants grow in a little portion of nutritive composition and inserted in a system giving them the nutrients needed, only when it is necessary. This new agriculture methods can save until the 90% of space and can optimize the use of chemical compounds. Often hydroponic is implemented in greenhouses, that are becoming more technological too and self-adaptive thanks to sensors implementations.

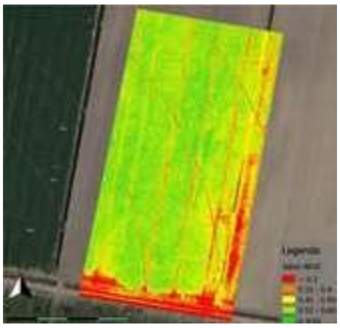
Figure 15: Hydroponic application’s example



Aquaponic (Figure 16) is another emerging field, derived from hydroponic: it is the combination of agriculture and sustainable breeding. Plants take nutrients from fish excrements. Once water is given to plants, it is filtered and reinserted on fishes' aquaria.

Figure 16: Aquaponic explanation

There is also an improvement in traditional agriculture, which is becoming more technologic too.



We can cite the Vigor Map (Figure 17), which is taken by means of satellites showing the production of a field with great precision. Thanks to this tool, farmers can have an idea of where it is better to use products such as fertilizer and pesticides in a more efficient way, in order to cut costs and to obtain an improvement on production.

But how this Vigor Map is implemented? This map is a set of data that are processed by computers. These computers guide the activity of new agricultural robots, that is another emerging tech too.

Figure 17: Vigor map's example

The emerging robot can perform different tasks, such as spray products only where the cultivation needs or pick up only mature fruits with a great precision, treat soil (Figure 18) and so on. These robots are often electrical and autonomous.

Figure 18: Agricultural robot's example

Another example of emerging applications in precision agriculture is the ever-increasing use of Drones, to monitor crop yields, climate conditions and they represent a new way to spray products in a precise area (Figure 19).

We can say that agriculture is becoming a high-tech field too, thanks to technological developments and merging of discovered knowledge from other complementary fields.

In conclusion, these are our ideas coming from the analysis performed. Anyway, we want to point out that we are not certain about precise results of our analysis because it was conducted without experts.

Figure 19: Agricultural drone

RESEARCH QUESTION 2: What is the Patent Landscape for Vitibot's product? Who are their competitors and potential partners?

2.1 : EXECUTIVE SUMMARY

As well as in every other sector, also in agriculture modern technologies are developing due to new knowledge and discoveries, transforming it in an industrial sector 4.0. The development of machine learning tools, thanks to the new It and IOT, systems is generating a new branch of agriculture: Precision Agriculture.

It consists in new methods for agriculture that allows farmers to reduce operative costs, increasing the profit per unit produced. One of the most incredible technologies branch in this field is the one developing agriculture robot, that merges different IT methods and data processing systems in order to build an efficient and sustainable system, exploring new horizon in doing agriculture, according to social movements for the environmental respect, that are fighting against human pollution.

These robots have multiple functions, they are designed to cut grass, give plants pesticides and fertilize, robots that pick up fruits only when is mature, robots that process data taken from the cultivation, and they do all those tasks it optimizing the waste, thanks to the support of technologies as Vigor Map, cited in the previous Research Question.

The agriculture sector is one of the most ancient of our society, it's been one of the first employment of the primitive man and it has followed the evolution of the human being throughout different ages. For this reason, the agriculture has evolved until now, with a new concept of the sector where everything is managed behind a monitor. What an amazing progress!

2.2 : CONTEXT OF THE PROBLEM

In this Research Question we aim to help Vitibot, a French start-up, in understanding the Patent Landscape Report for its product, Bakus. A Patent Landscape report is a "a snapshot of the patent situation of a specific technology, either within a given country or region, or globally. It can be useful to learn policy discussions, strategic research planning or technology transfer"[7]

We used the Patent Landscape Report to define which are the most interesting technological sector that compose this robot, based on his components and functionalities. We use this analysis to find which are the firms that are investing in the same technological fields in order to define Vitibot's potential partners and competitors. We consider potential partner a firm investing in some technologies that can be useful for Bakus products (i.e. some technologies that are not strongly developed for this product), or firms that are investing in complementary technology sectors (i.e., new pesticides, fertilizer etc). A competitor instead is a firm investing mainly in the same Vitibot's technological sectors and that are developing quite the same product, together with their dimension and role in the market.

2.3 : METHODOLOGY AND WORKFLOW

The company's objective is to place its technological breakthrough, the robot Bakus, within the framework of sustainable viticulture and conquer the French market within 2 years, before trying to land in the international market.
For this reason, it is interested in understanding the patent landscape for their product and to identify competitors and potential partners.
Our first idea was to get Bakus' patent and to extract from them the technological sector, using the same approach in RQ1. Based on the obtained subclasses, we would like to download patents of companies operating in the same sector and compare the Vitibot's portfolio with portfolios of the other companies to understand potential partner and competitors.

Since the Vitibot's patents are not available we implemented another approach, which basic idea is to get a dataset containing patents about the sector where Vitibot operates to extract the characteristics of Bakus and build the patent landscape based on these patents.

To better understand the methodology, we illustrate the workflow related to this research question, shown in Figure 20.

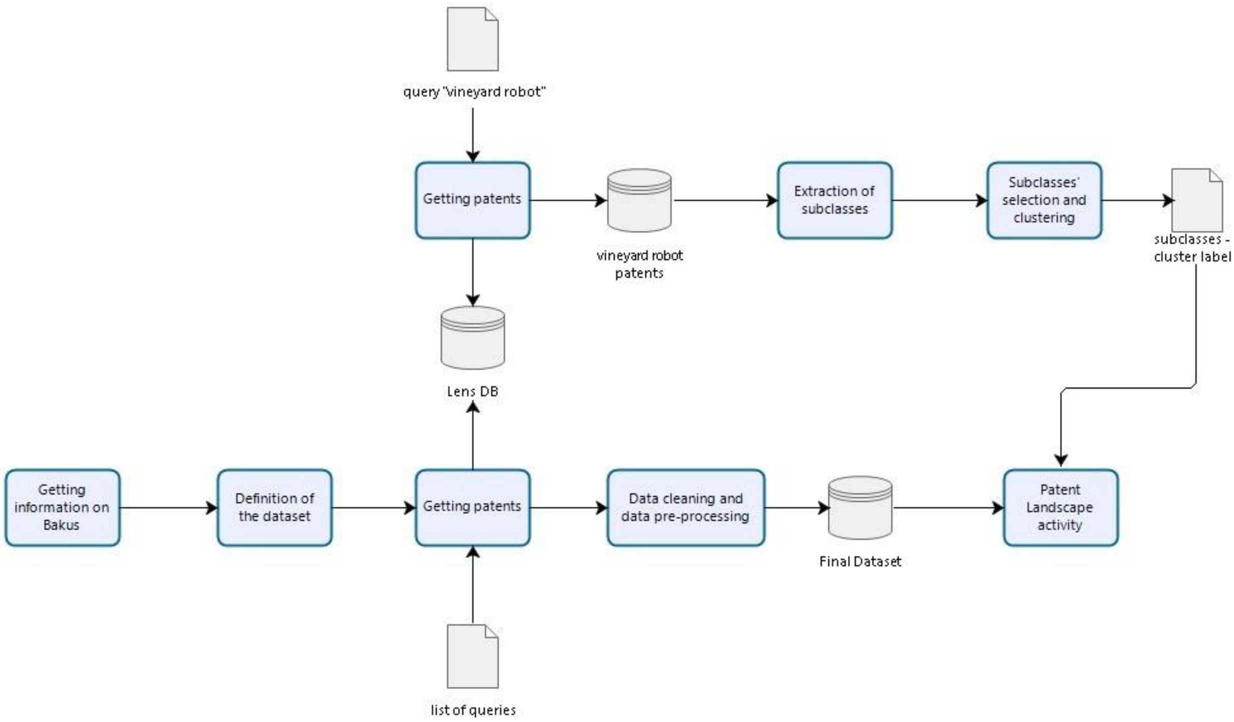


Figure 20: Workflow for RQ 2

We started reading articles about Bakus, surfing Vitibot's website and social networks to understand which can be its technological composition and its functionalities.

To define the dataset, after a brainstorming, we drew up a list of queries that can be used to search patents on Lens. As we said before, we select keywords based on the sector and functionalities of Bakus. Thinking about the sector we have chosen the query *"viticulture robot"*; answering to the question "what is Bakus?" we selected the queries *"precision vineyard robot"* and *"vineyard robot"*. Finally, according to the main functionality of Bakus (treat vine spraying products with a great precision) we extract patents about "*vineyard treatment robot".*

After catching patents, we have done data cleaning, removing duplicates, missing values and irrelevant columns in the same way of RQ1A. In the data pre-processing phase, we applied FuzzyWuzzy again to remove orthographical errors on all the applicants of the dataset, stored in the feature "Applicants"; finally, we created the feature "Cat_Applicants" that is a categorization of the feature "Applicants", obtaining in the end a dataset (that will be called "dataset 1") composed by 383 records and the following 5 features:

- Publication_Number: the unique publication number that is assigned to a patent when it is released
- Publication_Year: the publication year of the patent
- Applicants: it represents the entity or person which or who presents an application for the grant of an industrial property right
- IPCR_Classifications: a list of all the classification id that are assigned to a single patent.

- Cat_Applicants: a categorization of the feature Applicants, and can assume 3 values that are “FIRMS” if the assignee is a company, “University” if the assignee is a University and “Inventors as Assignee” if Inventor’s row coincides with the Applicant.

This categorization is made in order to have a clear idea about who is investing in these fields, be it firms or research institute such as universities.

On this dataset we performed the Patent Landscape activity. We have done an analysis about the general trend of the sector and the applicants, followed by a bubble chart analysis considering the top companies and the main subclasses involved in the sector in which Vitibot operate.

Our idea to identify these technological sectors is that we can extract the subclasses from patents that best fits with Bakus product.

To find these patents, we extracted from Lens another dataset (that is extracted with one of the queries that compose our patent landscape’s dataset) querying for *“vineyard robot”* because we think that this query is the most representative for Bakus.

We obtained 320 patents (that will be called “dataset 2”) and, with the same method used for RQ1 A, we assigned to each of them the “Main_Subclass” value, that represents the most frequent four-digits codes from the list of IPCs describing the patent.

From the values assigned, we analysed all the subclasses using the Lens Classification Explorer to select which ones are best representative for Bakus product and, at the same time, we assembled them in different clusters.

We performed this kind of subclasses’ classification considering their meaning and their first 3 digits of the IPC code, that represents the class. To explain it better, having the subclasses 'A01C', 'A01B', 'A01D there is a high probability that they belong to the same cluster because they have the same first 3 digits of IPC code.

The clusters obtained are the following:

CLUSTERS	**TASK**
- *Data computing:* contains the technological sectors about data processing (i.e. G06k is “recognition of data presentation of data record carriers handling record carriers”, G06F is “electric digital data processing”). The subclasses involved are the two in the example - *Transmission data:* contains the technological sectors operating in data transmission (i.e. H04W is “wireless communication networks”, H04L is “transmission of digital information”, H04B is “transmission systems for measured values, control or similar signals”). The subclasses of the clusters are the three explained in the examples.	The data processing through computing and transmission allows the robot to collect and elaborate information about the vineyard and the soil.
- *Soil treatment*: contains the technological sectors involved in the soil treatment (i.e. reporting the IPC classification meaning we have that A01B is “soil working in agriculture”, A01D is “harvesting mowing parts, details or accessories of agricultural machines” etc.). The complete list of subclasses belonging to the cluster are: 'A01C', 'A01B', 'B23C', 'A01D', 'B23Q', 'B26D'.	The activity carried out by the robot to take care of vineyard, the soil and vineyard treatment and the use of pesticides and fertilizer through its spraying apparatus

- *Final product treatment:* contains the technological sectors involved in the treatment of the final products (i.e. A01G is “horticulture cultivation of vegetables, flowers, rice, fruit, vines, hops or seaweed forestry watering”, A23N is “machines or apparatus for treating harvested fruit, vegetables or flower bulbs in bulk”). The subclasses that belong to this cluster are 'A01D', 'A23N'. - *Spraying apparatus:* is the cluster containing the technological sectors which cooperate to the process of spraying substances on plants as pesticides, nutritious substances etc. (i.e. G05B is “control or regulating systems […] for such systems or elements fluid-pressure actuators or systems acting by means of fluids in general”, B05D is “processes and apparatus for applying liquids or other fluent materials to surfaces”). The subclasses belonging to the cluster are: 'B05B', 'G05B', 'B05D'.	
- *Analysis and optical instruments:* contains the technological sectors which take care about optical instruments and analysis methods to recognize obstacles or persons allowing the movements of a robot, or also instruments as radars, sensors to monitor the agricultural products (i.e. G01N is “investigating or analysing materials by determining their chemical or physical properties”, G06T is “image data processing or generation”). The list of subclasses is: 'G01N', 'G06T', 'H04N', 'G06G', 'G01J', 'G08G', 'G01S', 'A61B'.	The optical instruments are fundamental to the robot to permits its mobility, avoiding obstacles.
- *Electrical devices and methods:* contain all the technological sectors involved in the electrical field (i.e. H01L is “semiconductor devices electric solid-state devices”, b60l is “propulsion of electrically-propelled vehicles”). The subclasses belonging to the cluster are: 'H01L', 'B60L', 'H05K', 'G01R'.	The electrical devices allows the robot to do its functions reducing the environmental impact.

Table 2: Technological sectors clusters created

During the selection and clustering of the technological sectors, we decided to not consider the chemical subclasses because, according to our research, Vitibot does not operate in this field.

Now, we are interested in understanding the investment portfolio of each firm, and for this purpose we perform an analysis with Bubble Chart, using the selected subclasses and clusters.

So, we decided to not take in consideration just the main technological sector (with this term we indicate the most frequent 4-digits value of the IPC codes list of a patent), but all his IPC codes: in other words, we considered all the IPC codes of a single patent, not only the main one. This was made considering that every patent is a set of technological fields, equivalent to the set of IPC codes, that converges on the same product. For this reason, whoever is investing in a product is investing in different technological fields at the same time.

We try to clarify the concept with an example.

Given the following dataset sample:

Publication_Number	Applicants	IPCR_Classifications
WO 2019/077138 A1	TOFWERK AG	H01J49/14;; G01N33/46;; G06N5/04;; H01J49/04
US 2015/0302305 A1	CLIMATE CORP	G06N5/04;; A01B79/00;; A01B21/00

Table 3: Example of original dataset

we do not consider the main IPC (extracting Main_Subclass feature) creating a dataset as the one below:

Publication_Number	Applicants	Main_Subclass
WO 2019/077138 A1	TOFWERK AG	H01J
US 2015/0302305 A1	CLIMATE CORP	A01B

Table 4: Example of dataset with "Main_Subclass" extraction

Instead, we do a kind of IPC code unpacking, obtaining a dataset as the following:

Publication_Number	Applicants	Subclass
WO 2019/077138 A1	TOFWERK AG	H01J
WO 2019/077138 A1	TOFWERK AG	G01N
WO 2019/077138 A1	TOFWERK AG	G06N
WO 2019/077138 A1	TOFWERK AG	H01J
US 2015/0302305 A1	CLIMATE CORP	G06N
US 2015/0302305 A1	CLIMATE CORP	A01B
US 2015/0302305 A1	CLIMATE CORP	A01B

Table 5: Example of "unpacked" dataset

In this way, we can observe that TOFWERK AG is operating in the technological sectors H01J, G01N and G06N, instead, considering only the main subclass, we would have included in the analysis only the sector H01J, wrongly.

During the IPC unpacking, we have discarded records with the value of Subclass different to the selected subclasses and to the chosen ones we assigned the cluster label, obtaining a dataset of 2507 record and 4 features that are: *Subclass*, *Applicants, Publication_Year* and *Subclass_Cluster* ,clearly understandable.

Based on this work, to have an intuitive oversight on the problem, we plot 4 different bubble charts considering:

- *technological sectors and years:* to analyse deeper in the cluster, considering the precise technological sectors.
- *top 20 companies and clusters:* to have an idea on which are the cluster where the top companies are investing.
- *top 20 companies and technological sectors:* the idea is the same of the previous chart, but we have a more detailed vision of the sector.

Once Patent Landscape has been done, we defined Potential Partner and Competitor basing on the definitions said before.

2.4 : ANALYSIS: Designing a Patent Landscape Report

As we explained in the methodology section, we worked on two datasets: one obtained extracting patent on Lens using multiple queries ("dataset 1"), and a dataset extracted with the query "*vineyard robot"* ("dataset 2"). From these datasets we have obtained another dataset that will be used for patent landscape report and to find partners and competitors ("dataset 3").

For our analysis we started collecting patents composing dataset 1 extracting from Lens patents obtained with multiple queries (as explained in methodology section) and doing the pre-processing phase. After we extracted from Lens the dataset 2 to find subclasses of patents that best fit with Bakus product. To do this, we have checked the meaning of every subclass composing our sample dataset thanks to Lens Classification Explorer.

From this process, we extracted this list of subclasses: 'A01C', 'A01B', 'B23C', 'A01D', 'B23Q', 'B26D', 'A01G', 'A23N', 'G06K', 'G06F', 'G01N', 'G06T', 'H04N', 'G06G', 'G01J', 'G08G', 'G01S', 'A61B', 'B05B', 'A01M', 'G05B', 'B05D', 'H04W', 'H04L', 'H04B', 'H01L', 'B60L', 'H05K', 'G01R'.

At this point we started working on dataset 3, plotting patents' publication per year, obtaining the plot shown in figure 21.

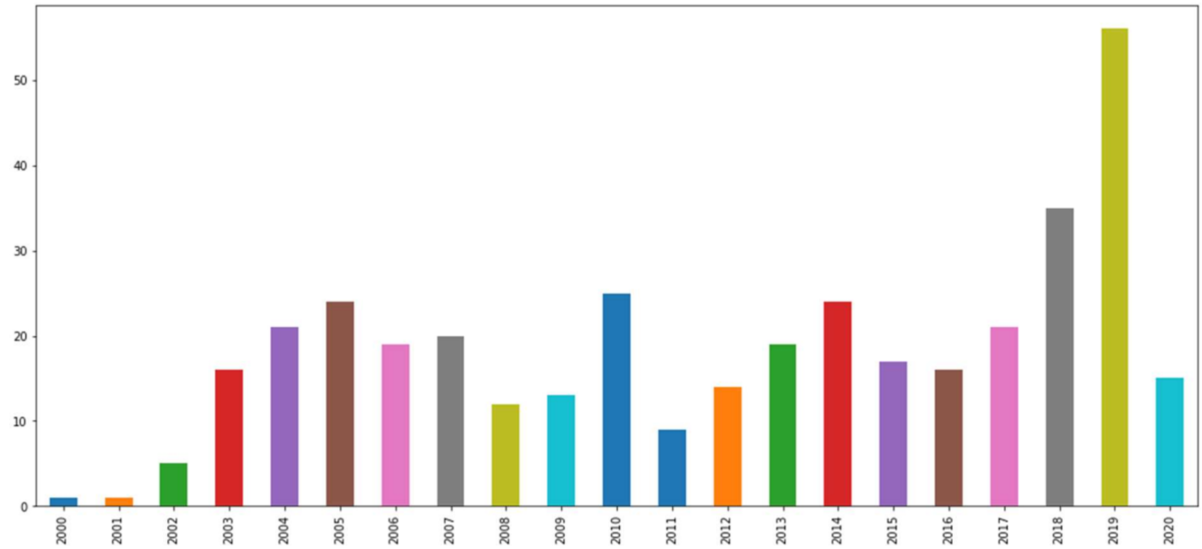


Figure 21: Patents' publications per year of dataset 3

Is possible to notice that the publication number per year about this topic has been increasing year by year. Even the year just started shows a high number of patents in these first months.

To proceed our analysis, we want to reply a simple question: who is investing more in those tools? The public sector or to private one? To answer we plotted a pie chart (shown in figure 22**)** of the "Cat_Applicants" features of dataset 3, extracted with the methodology explained in the previous section.

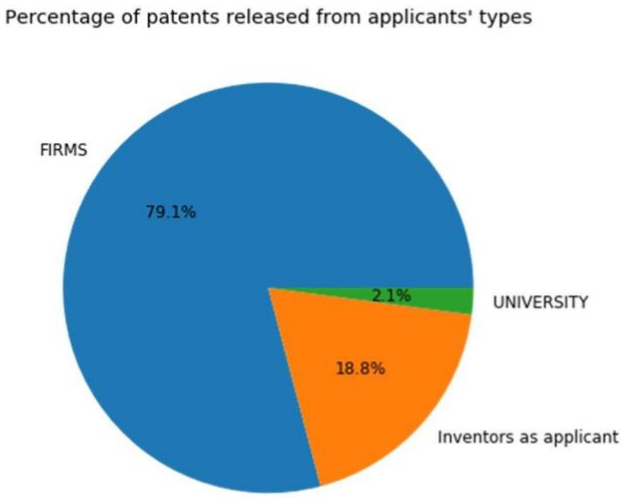


Figure 22: Firms, inventors and research institute pie chart

From this plot we can intuitively observe that who is investing in this industrial sector are mainly firms and inventors, so the majority class of our dataset belongs to the private sector. This because these new agriculture developments can bring enormous innovative and competitive advantages among other agricultural firms.

To better understand our dataset composition, we extracted a list of the top 10 applicants for patents publications, and we checked on Google their function, obtaining the table shown in Table 6:

FIRM	**PRODUCT**
BAYER CROPSCIENCE	Pesticides, fertilizers, insecticides
ADVANCED ELEMENTAL TECHNOLOGIES	Computing-based patents/products
APPLIED MATERIALS INC	Equipment, services and software to produce semiconductor chips for electronics, flat screen displays for computers, smartphones, televisions and solar products
INDIGO AG INC	Plant microbes, to improve the yields of cotton, wheat corn etc
ANDERSON NOEL WAYNE	Is an inventor for deere & co firm
LAW OFFICE OF J. GROSS	Law office
GOOGLE INC	Simply Google
WEEDOUT LTD	Bioherbicides
CRINKLAW FARM SERVICES INC	Custom vineyard services
DEERE & CO	Agricultural machines

Table 6: Top 10 applicants: firms and products/services

Is possible to see how our dataset is composed by different products: some are interesting for our purpose, other not so. We performed the unpacking procedure as we explained in the methodology section, considering patents as technologies involving the merging of different industrial sectors. This allows us to understand which is the investment portfolio of every firm: investing in one technology may involve different technological sectors too.

Then, we took from the unpacked dataset only records belonging to the selected subclasses extracted from dataset 2, we cluster subclasses together, based on their meaning and thinking about linkages between technologies to try to represent Bakus' features. Based on this we computed different bubble charts, an intuitive way to know in which technological fields firms are investing every year and other information, such as the top firms and which technologies firms are developing.

The first analysis shows the trend of cluster through years, an overview in which functionalities firms are investing in the most recent times. From this, we generate the bubble chart shown in Figure 23:

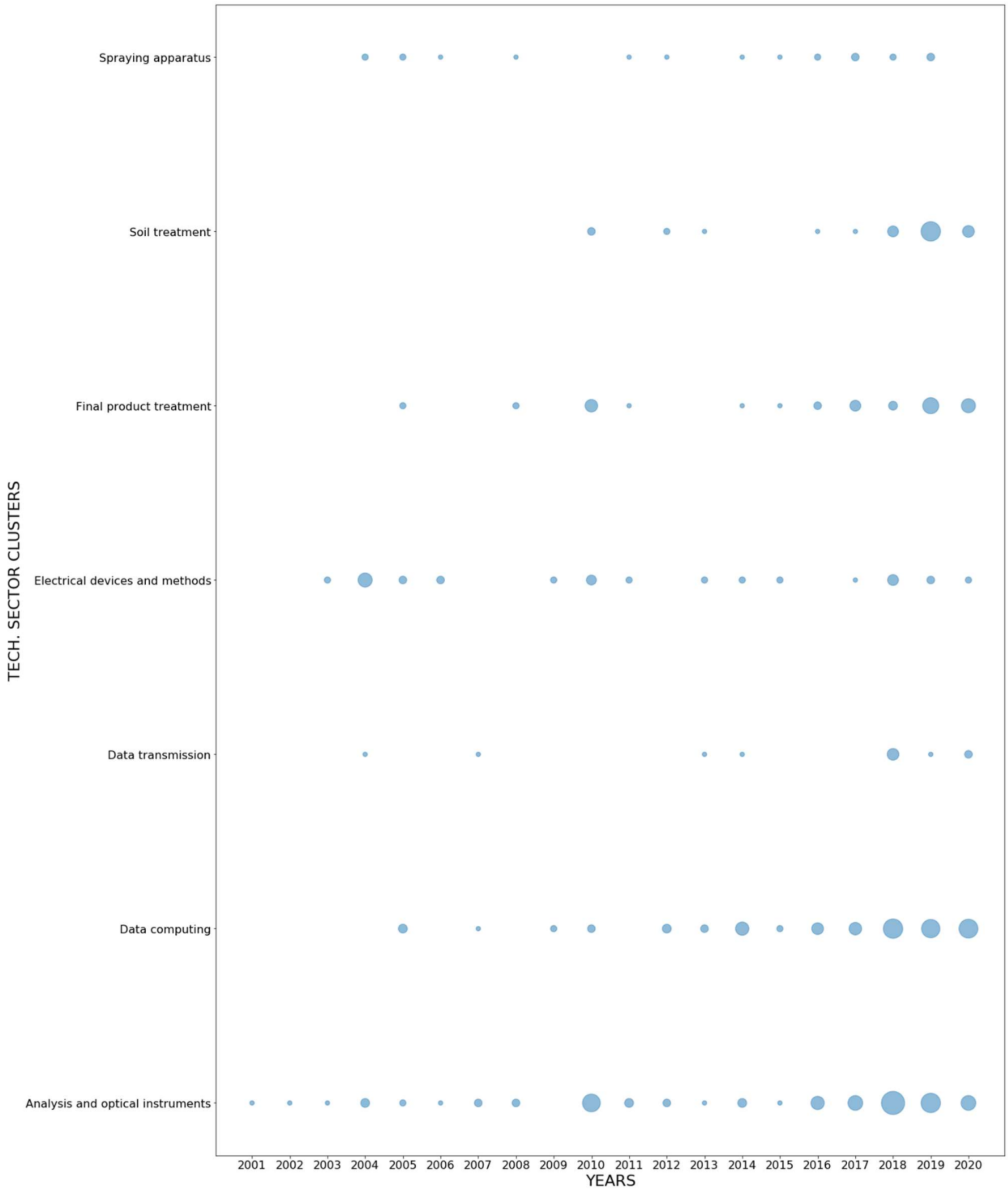


Figure 23: Clusters' trend per year

From this bubble chart we can see the growth in investment for the different technological cluster through the year. Is possible to see how the "Analysis and optical instruments", "Electrical devices and methods", "Final product treatment" and "Spraying apparatus" has been growing constantly in the last 20 years, while "Soil Treatment", "Data Computing" and "Drone" registered a sudden rise in the last years since they include emerging technologies.

To go deeper in our analysis, we computed the trend of each technological sector composing those clusters in years. To have an intuitive overview, we use the same bubble chart, extracting what it comes in figure 24.

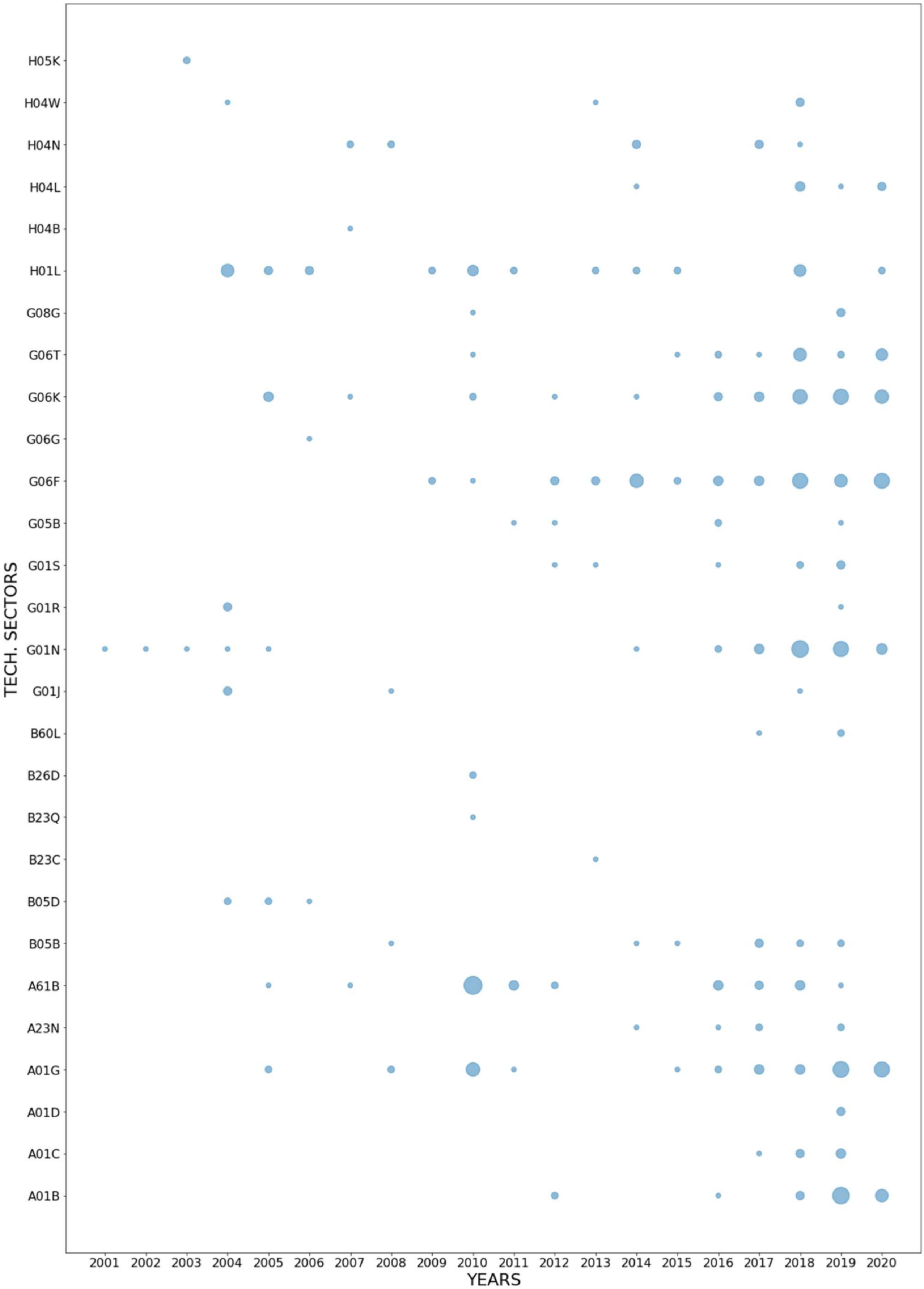


Figure 24: Technological sector's trend per year

From this bubble chart, we can see that the most meaningful technological sector (four digits IPC codes) that increased in the last few years are:

- A01G: “horticulture cultivation of vegetables, flowers, rice, fruit, vines, hops or seaweed forestry watering picking of fruits, vegetables, hops”
- A01B: “soil working in agriculture or forestry parts, details, or accessories of agricultural machines or implements, in general”
- G01N: “investigating or analysing materials by determining their chemical or physical properties”
- G06F: “electric digital data processing computer systems based on specific computational models”
- G06K: “recognition of data presentation of data record carriers handling record carriers”
- G06T: “image data processing or generation, in general”

How we can see from the chart, the technological sectors more developed in the last few years are about soil working and using of processing computers to manage data of various type. Agricultural firms are investing in those fields to innovate themselves and to increase the edge between their competitors, together with the emerging of new technologies in those sectors.

After this analysis on technology clusters and sectors trends, we proceed analysing which firms and universities (almost the bigger of our dataset) are investing in the same clusters and sectors.

We computed a bubble chart (in figure 22) representing firms and technological clusters. For graphical purpose, we give acronym to clusters.

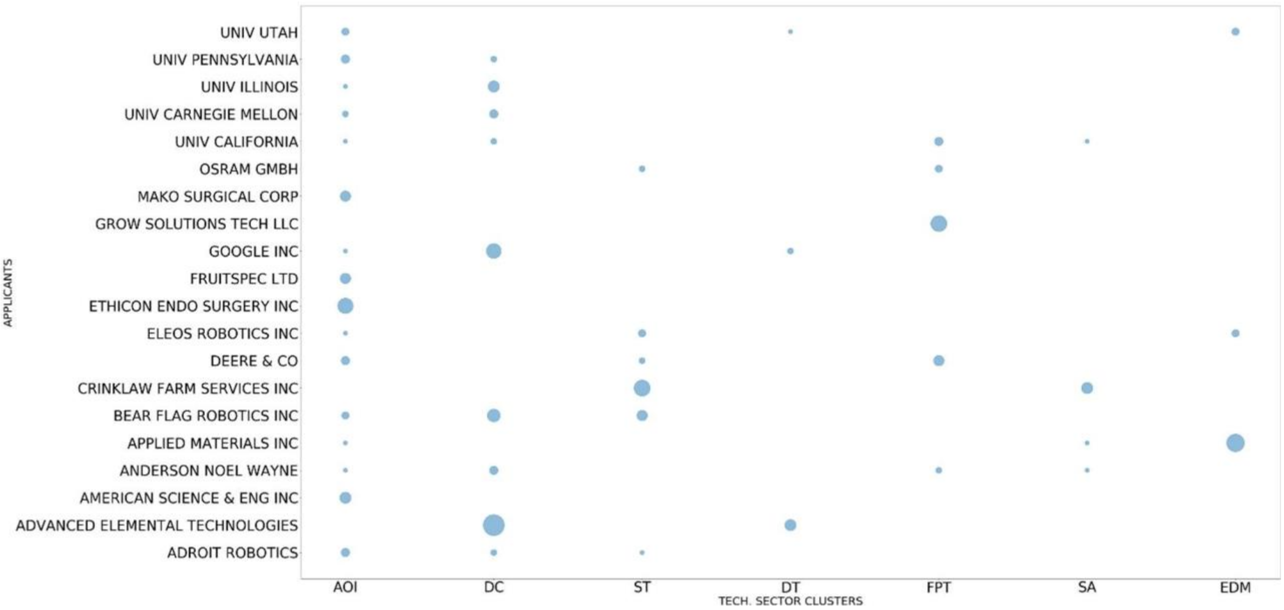


Figure 25: Top applicants/clusters bubble chart

This bubble chart gives us different points: reading the graphic vertically we can see which firms are developing quite the same product, finding which are the technological sector clusters that have more products in it and so individuate potential competitors. Analysing the bubble chart’s horizontal sense, we can see the investment portfolio of each firm.

So, based on what we have just said, we can see that “AOI” cluster, acronym for Analysis and optical instrument is a well-covered field, as the Data Computing (DC) one. We can also see that firms invests in many fields that are complementary one to another.

Going deeper, we increased the granularity of our analysis decomposing our clusters and doing the same analysis previously made, but considering every technological sector for every firm, obtaining the bubble chart shown below in figure 26.

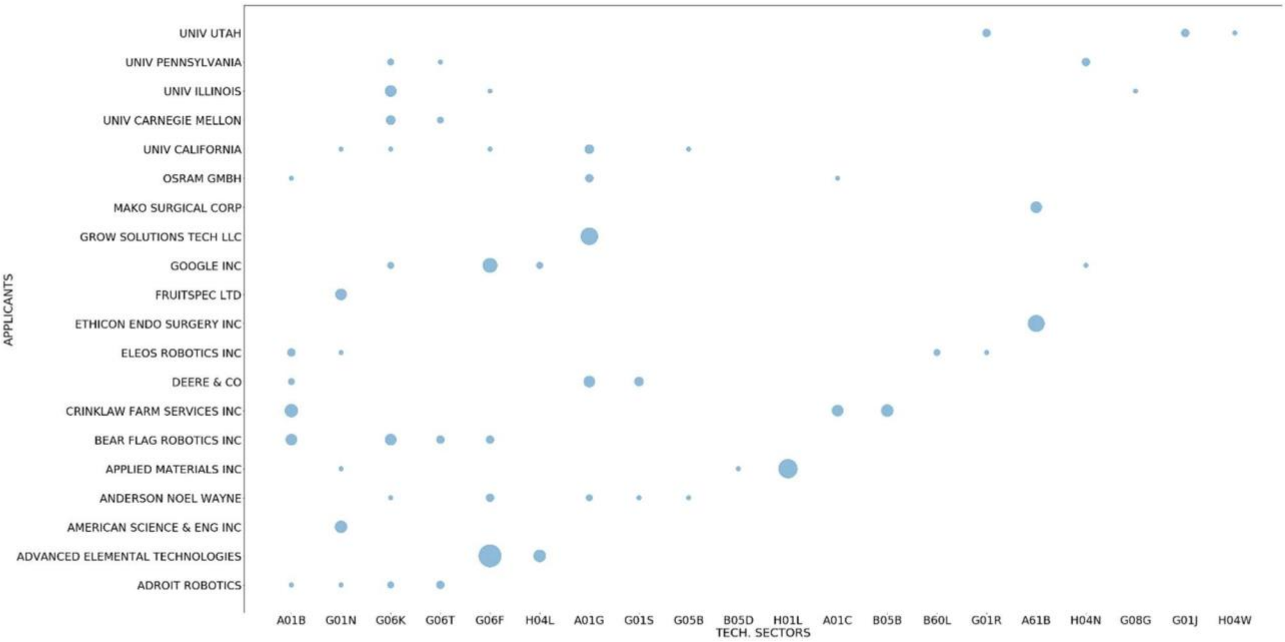


Figure 26: Applicants/technological sectors bubble chart

As we can see the technological sectors in which firms developed their products are A01B, G01N, G06K, and G06F. With respect to the previous analysis, a new class emerges: A01B, that represents “soil working in agriculture or forestry parts, details, or accessories of agricultural machines or implements, in general making or covering furrows or holes for sowing, planting”.

From this plot we can easily define which are the leader in every technological sector. We will not report every single technological sector, but only the four reported above.

For the technological sector A01B we can state that the leader is “CRINKLAW FARM SERVICE INC” that, as we explain above, is a firm offering services for vineyard management.

The leader in G01N is “AMERICAN SCIENCE & ENG INC” that mainly develop x-ray systems in order to investigate what a camion or a train contains etc. Its products can be useful in security checks on airports, at borders and so on.

The leader in G06K is “BEAR FLAG ROBOTICS INC” that is a firm developing self-driving technologies for tractors and implements; the leader of G06F in “ADVANCED ELEMENTAL TECHNOLOGIES”.

2.5 : DISCUSSION AND INSIGHT

Agriculture industrial sector is developing thanks to the use of various technologies, to create competitive advantages between firms. One of the most meaningful tools applied to improve this industrial sector are agricultural robots. Bakus, the one developed by Vitibot, is one of them but not the only, there are a lot of other firms and start-ups developing different robots for precision agriculture using more or less the same technologies, to reach different goals. We saw that Bakus mainly use a self-driving technology and a spray apparatus driven by data processing systems and computers.

Thanks to our analysis we try to analyse firms that are working in the same industrial sector, robotics applied to agriculture, in order to have an idea of which are the technologies mainly involved in this field and to search for some potential partner and competitor for this firm.

We noticed that the firms mined in our work are not always related to agricultural field: we have also firms working in surgery products, as ‘Mako surgical corp’, that is investing in surgical robotics, or the cited American Science & Eng Inc (AS&E for his website), that works in logistics and security robotics.

Among the 20 firms in our bubble charts there are also universities, but it is impossible to state which products they are developing. At least we can check for his patents’ application, but we have not be able to discover anything interesting.

We can now say the potential competitor for Vitibot that can be identified after this analysis:

- ❖ **Deere & Co and Case New Holland**: big multinational firms that develops traditional agriculture machines such as tractors, combined harvesters, sowers etc. Those firm has the possibility to buy Vitibot out of the market, using and developing its industrial property to develop innovative products.
- ❖ **Bear Flag Robotics Inc**, because it develops technologies to realize self-driving machines, that is one of the main technologies used by Bakus.
- ❖ **Fruitspec LTD**: this firm is developing a technology that use vision algorithms to count the number of fruit and to estimate fruit sizes for providing accurate early season fruit yield estimation. This is like the technology used by Bakus for the analysis of the vineyard and, even if it is a start-up, works in the same sector, the precision agriculture.
- ❖ **Eleos Robotics INC**: develops weed-killing robots that are self-driving and fully autonomous. It works in the precision agriculture sector using robots, this means it may be a possible menace for Vitibot.

On the potential partner side, we have:

- ❖ **Adroit Robotics**: that develop products that, thanks to sensors scout through the orchard, analyse the trees. This could be helpful for Vitibot to scan vineyards, to detect fruits ready to prune and to better understand which vine needs treatments.
- ❖ **Advanced Elemental Technologies**: a firm working on the data processing equipment, an important component of Bakus, and not present in the precision agriculture sector.
- ❖ **Applied Materials Inc**: a company working on materials engineering solutions used to produce virtually every new chip and advanced display in the world, this can be an important feature for Bakus. This company too works outside the agricultural sector.
- ❖ **GrowTech Solutions**: a company that provides technology products to people who grow or sell plants. Typically, that technology is related to product identification or processing and most often involves Tags, Labels, Barcodes or Merchandising Materials. It has patented different products concerning cutting implements specially adapted for horticultural purposes, an important component of Bakus.

RESEARCH QUESTION 3: Analysis about market positioning of Vitibot, Vinbot and Agribot

3.1: EXECUTIVE SUMMARY

In this research question we are going to analyze the positioning in the market of three startups working in the precision agriculture field: Vitibot, Vinbot and Agribot. Before doing that, we want to give an overview of the precision agriculture sector, showing its evolution over time and its dimension in financial terms. Then, after doing research to collect insights on their structure and their operating plan, we operated a swot analysis to compare strengths and weaknesses of the three startups; this allows us to get a clear idea of their potential. In the end we have done a comparison between the startups based on the information gained in the previous analysis and using the patents published to speculate on their market position and the possible relationship among them.

3.2 : CONTEXT OF THE PROBLEM

Vitibot, Vinbot[A] and Agrobot are three startups operating in the precision agriculture sector. Since this market segment, resulted from the merge of computer science and agriculture, is quite recent, the reality operating in this field are mainly start-up or small businesses. This landscape has created quite a few difficulties in the financial analysis of the companies operating, since most of them are in their first phase of life and have not launched their product on the market yet. This is the case of our start-ups, for this reason we have concentrated the analysis on the technologies developed by each one of them, united with the information gained on the funding received.

3.3 : METHODOLOGY AND WORKFLOW

We are going to analyze the agriculture market using bar charts to represent the growth of the sector; then we will give an overview on the different technologies used in this sector and concentrate the analysis on the precision viticulture.

After that we will consider each startup analyzing the information collected and using a SWOT analysis to represent their potential in the market and drew up a bar chart based on the technologies patented by them to compare their objectives. Finally, we gave a discussion about the possible positioning on the market and the business relations between those companies.

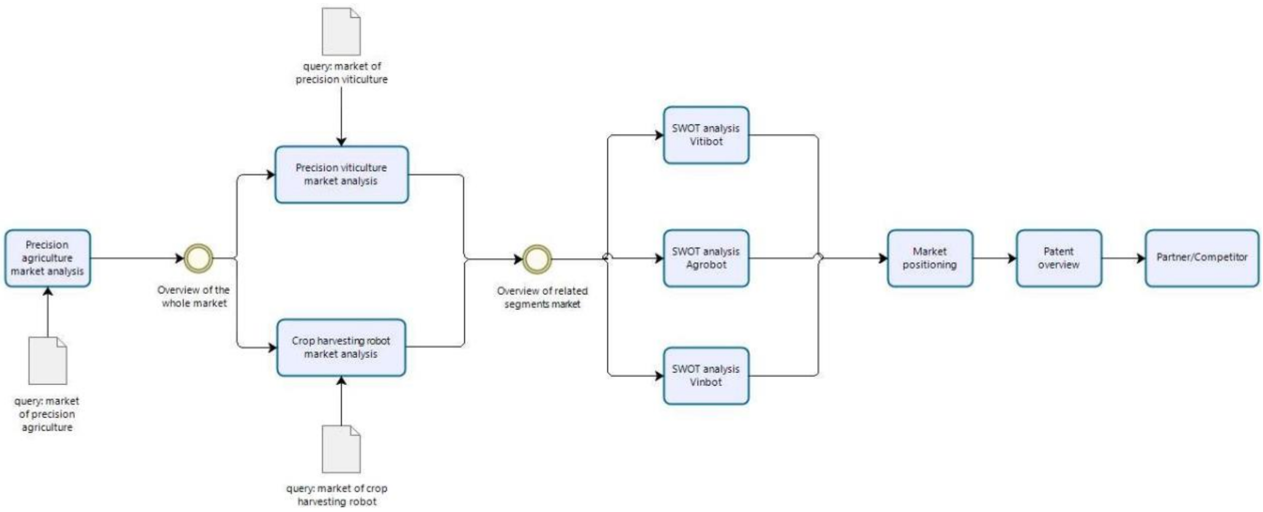


Figure 27: Workflow of RQ3

[A] Vinbot is just a project, the robot launched by the french startup Vitirover. Despite this, we will continue to name the project as Vinbot to avoid confusion.

3.4 : ANALYSIS: Brief overview on precision agriculture market

Nowadays, it's emerging a new way of doing farming: precision agriculture. This new management strategy gathers, processes and analyzes temporal and spatial data using modern instruments, in order to carry out agronomic operations while reaching data about the actual crop needs and the biochemical characteristics of the soil. From the economic point of view, the precision agriculture market has been raising in the last few years. Indeed, it is expected to grow from USD 5.1 billion in 2018 to USD 9.5 billion by 2023, at a CAGR of 13.5% (compound annual growth rate). As we can see from the following graph, American region holds the largest market share, preceded by European Region and APAC (Asia Pacific). We can see that the rest of the world, such as Africa, holds a negligible market share.

The precision farming practices are expected to keep growing as they enable farmers to accurately control changing in soil composition, increasing productivity and reducing operative costs. Farm managers and producers are leveraging the capabilities of the IoT devices such as sensors, GPS systems, automated steering systems for soil sampling, temperature monitoring, irrigation management, filed mapping and several other applications. Those new technologies provide real-time insights on how to improve farming practices for greater efficiency.[8]

Figure 28: Growth of the precision farming market and forecasts for the next years

In the precision agriculture component market, the *hardware* segment is expected to dominate in 2023 with a share above 70% due to installation of devices such as sensors or cameras onto harvesters and tractors. The remained share is occupied for 25% by *software* and for 5% by *services*.

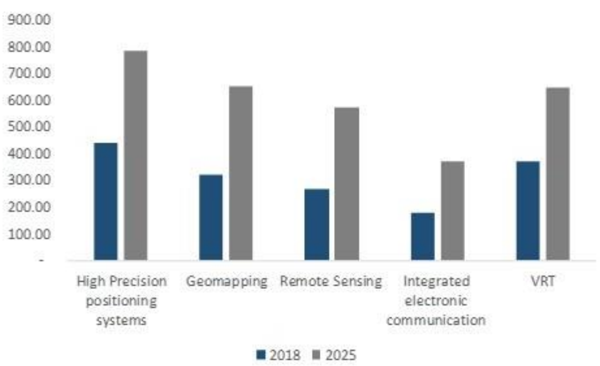
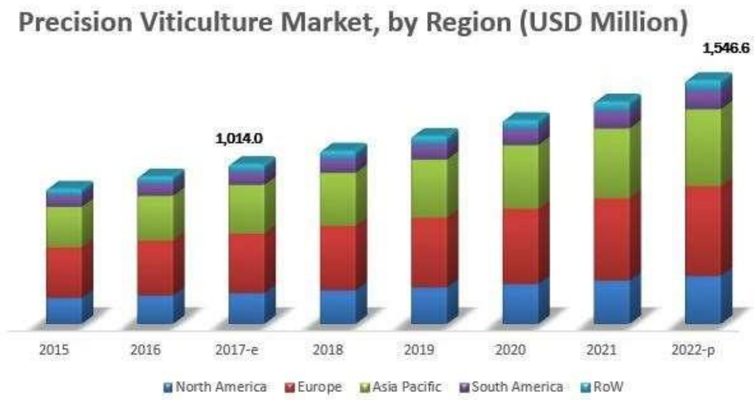


Figure 29: Composition of the precision agriculture market

The market size by technology will be dominated by **high-precision positioning systems** (27%) due to extensive use of GPS and GNSS systems (Global Navigation Satellite System). These technologies increase the efficiency of processes and reduce the unnecessary expenditure on agrochemicals, seeds and fuel. Other technologies such as **geomapping** (support system that show the growers the exact locations in the farm and give specific information regarding that location), **remote sensing** (process of detecting and monitoring the physical characteristics of an area through special cameras) and **variable-rate tech** (system that allow the farmer to vary the rate of crop inputs and detect accurately individual requirement of each crop) will cover great importance in the future of agriculture[9].

3.4.1.1 LOOKING CLOSER TO THE PRECISION VITICULTURE MARKET

An important segment of precision agriculture is the precision viticulture. It is a precision farming process applied to optimize vineyard performance, maximizing grape yield and quality while reducing risk and environmental impacts. Its market was valued over USD 936 million in 2016 and is expected to reach 1,546.6 million by 2022, with a growth rate of 8,81%. (RoW (the Middle East and Africa)).



The following graph (Figure 30) shows the size of the market by region. It turns out that Europe has a consistent share, due to the presence of many vineyard located in countries such as France, Italy and Spain. Of concern is the constant increase of the market share covered by Asian area.

Figure 30: Growth of the precision viticulture market and forecasts for the next years

Based on application, the market is widely fragmented. The main applications are:

- yield monitoring (it is still the principal application)
- field mapping
- crop scouting
- Weather tracking & forecasting
- Irrigation management
- Inventory management
- Farm labor management

Related to application we can identify three leading technologies:

- guidance systems (Global Positioning Systems and Geographic Information systems)
- remote sensing
- variable rate technologies

Based on product/services, the market is dominated by *hardware* segment (automation & control system, sensing & monitoring devices), followed by *software* (local-web based, cloud-based) and *services* (system integration and consulting, managed services, connectivity services, assisted professional services, maintenance & support).

Entry barriers in the viticulture market are noteworthy and mostly are composed by highly specific skills and huge amount of capital to invest in Research and Development.

The key players in this market include several us companies such as *Trimble* and *Teeject* Technologies while *Groupe ICV* (FR) and *Terranis*(FR) stand out in the European landscape, both of them engaged in issuing innovative services for winegrowers. The players analyzed in this report are three startups, Vitibot and Vinbot, both French, and Agribot from Spain.

3.4.1.1 : VITIBOT [10]

Vitibot is a French industrial start-up, composed by 50 employees, engaged in the market of autonomous and electronic wine robots. Indeed, it supports the winegrowers in improving their vineyards and guarantee greater hygiene and safety for wine workers. The company identify its mission in meeting the challenges of sustainable viticulture through an innovative product designed from scratch by its engineering team: **Bakus**. Bakus is an autonomous electric robot for viticulture, equipped with a sensory platform able to identify the areas where to intervene with pruning and administration of fertilizers and pesticides, both on the vineyards and on the surrounding land. Now there are 6 specimens, four operating in the Champagne fields and two used for demonstrations.

They are planning to enter the market for French and European vineyards over the next 2 years. In order to understand the actual position in the market, we conducted a SWOT analysis which showed the following points:

Table 7: Swot analysis Vitibot

STRENGHTS	**WEAKNESSES**
- Highly qualified team (50 employees, mainly electronics, computers and mechanical engineers) - Bakus designed with internal technology (100% Autonomous, 100% Electric, Multipurpose, Efficient, Flexible, 10H autonomy) - Sustainability is what the product conveys and what the market wants -Bakus can be adapted to all kinds of vineyards - several patents filed for the robot and its tools -the cost of purchasing a Bakus robot and its equipment are equivalent to that of conventional high thermal clearance tractor	- the company is quite small and it can't be self-financing - Bakus is still at the experimental stage and until now only 6 robots have been manufactured since 2017 - lack of historical data make impossible to do predictions - limited product portfolio, not yet in commerce
OPPORTUNITIES	**THREATS**
- it is operating in a fast-growing market - few competitors in France - growing demand for sustainable tool in viticulture sector - company's name is spreading all over Europe due to recent awards - possibility to get EU's contribution for financing the activity - chance to form partnership with big companies	- several skilled-startup are emerging in Europe, leading to an high level of competitiveness - small businesses can't afford this type of investment - lack of technical infrastructure and local experts in many rural areas to get through precision viticulture - technological backwardness can slow down the use of these tools -obsolescence due to rapid changes in the market

The company has been able to attract great interest from the wine and champagne producers who brought the first 3 million euros of private financing following the demonstrations; it is now reaching additional 10 million in order to fund series production.

In conclusion we can say that it’s not easy to get the real market position of an emerging startup, but from our analysis it turns out that Vitibot is expected to be a leading company in France where some Champagne winegrowers have already recognized Bakus as a tool they are willing to invest in.

3.4.1.2 : VINBOT[11]

Vinbot is a project developed by a consortium of organizations including winegrowers' associations, technology and IT companies, agro service companies and research institutes, and has been co-founded by the European Union's 7th Framework Programme. The coordinator of the consortium a Spanish company: *Robotnik Automation SLL*. The startup launching this project is Vitirover, a French company composed by 6 persons.

Vinbot is an all-terrain autonomous mobile robot with a set of sensors able to capture and analyze vineyard images and 3D data by means of cloud computing applications, in order to optimize yield management and wine quality. The purpose of this project is to help winegrowers and wine producers manage the yield and wine quality by means of a new system based on robotics and cloud-computing technologies. Vinbot responds to a need to boost the quality of European wines by implementing precision viticulture to estimate the yield. The product is not on the market yet, the company is planning to commercialize the robot over the next 2 years. In order to understand the actual position in the market, we conducted a SWOT analysis which showed the following points:

STRENGHTS	**WEAKNESSES**
- the project has both financial resources and technical knowledge by means of consortium - the robot is totally autonomous - several patents owned for the technology and the robot itself - it is suitable for any type of soil - the project is supported by a leading robotic company - it does not exist a precision viticulture tool like Vinbot	-Vinbot is still a project, indeed the robot is not on the market yet - it has not its own headquarter and it depends on the consortium - lack of historical data make impossible to do predictions - only one product at a prototype stage - the current complex technology behind Vinbot could led to exorbitant sale prices -Financing are needed to keep the project going
OPPORTUNITIES	**THREATS**
- it is operating in a fast-growing market - possibility to get EU's contribution for financing the activity - growing demand for sustainable tool in viticulture sector - Vinbot is planning to make the robot suitable for other applications besides viticulture - it is making deals with US companies that could allow to enter in the US market	- several skilled-startup are emerging in Europe, leading to a high level of competitiveness - small businesses can't afford this type of investment - lack of technical infrastructure and local experts in many rural areas to get through precision viticulture - technological backwardness can slow down the use of these tools - obsolescence due to rapid changes in the market

Table 8: Swot analysis Vinbot

In conclusion, the main emerging point is that the product could be a breakthrough innovation with no rival in the market. However, Vinbot is not in commerce yet and thus it is complicated to make prediction due to the complex technology behind the robot and the future sale price that could led to a limited target.[12]

3.4.2 : LOOKING CLOSER TO THE CROP HARVESTING ROBOT MARKET[13]

Another emerging segment in precision agriculture market is crop harvesting robot market, target market for Agribot. These types of robots move through a farm to determine plant locations and approximate position, number and size of fruits and plants. They have been gaining popularity as an effective solution for harvesting across farms in order to cut labor during the harvesting season and reduce operating costs. Farmers already using harvesting robots vary by farm size and crops. They are likely running large operations of 500 acres or more. These farmers both have more capital available and face high labor costs, creating a stronger incentive for alternatives to human labor.

Nonetheless there are still few products that have reached the commercial market, this market is expected to rise from its initial estimated value of USD 29.43 million in 2018 at a CAGR of 20.57% in the forecast period of 2019-2026, up to USD 130.56 million by 2026. An important factor that boost market growth is the global increasing in food demand. An important market driver is the increasing focus in farm mechanization, although high maintenance costs slow down the growth. Furthermore, an important market trend is identified in solar powered crop harvesting robots.

The largest market share per region is held by America (43%), while Europe and Asia have roughly 30% ad 20%.

Few of the major competitors currently working in the harvesting robot market are *Cerescon* (NE) which is busy in developing a selective harvesting solution for white asparagous, *Energid Technologies* (UK) engaged in Citrus harvesting, *Green Robot Machinary* (INDIA) which is busy in cotton harvesting robots, and *FFRobotics* (ISRAEL) that patented a robotic fruit harvester.

Consistent entry barriers decrease the number of competitors due to huge amount of capital to invest in Research and Development and technological complexity behind the product.

3.4.2.1 : AGRIBOT

Agribot is a Spanish startup engaged in the business of agricultural robot. It is working with US leading farmers and it has developed the first pre-commercial agricultural robot for gently harvest strawberries. Designed to carry out its tasks autonomously, the Agribot robot is called **E-Series** and uses a precision technology built to collect strawberries on all types of soil and implemented on small tractors and large machinery inside greenhouses. It is equipped with cameras that analyze the degree of maturity of the fruit.

With this robot, the company is trying to provide a solution to one of the key inputs and cost center in the farming system: labor shortage. Since fruit growers deal with the most delicate product and most of which must remain attractive and blemish free all the way to the grocery, fruit specific harvesting should focus on picking the best fruit without harming it. The company has a partnership with the Californian company Driscoll's that has conquered a third of the American berry market, a partnership born following the victory of the Open Mind Award by the Government of Andalusia, winning a trip to Silicon Valley. So, the company brought e-series there, working with firms that reach 2 billion dollars in revenue a year. The product is not on the market yet, the Driscoll's company is conducting tests on the land of Watsonville, a city in California that has 40% of the strawberry crops in the state, in view of a market launch.

In order to understand the actual position in the market, we conducted a SWOT analysis which showed the following points[14]:

STRENGHTS	WEAKNESSES
- first fully automated strawberry harvester - a camera system makes sure that only ripe berries are harvested - Agrobot is the first in getting in US strawberry market with this type of robot - one person is able to operate the machine and harvest an entire field of berries - several patents owned for the technology and the robot itself - the arms are fully independent and in case 1 or more arms break down it keeps working - it is relatively cheap compared with labor cost - collaboration with big companies	- The application is limited to only harvest strawberries - the robot is still at a prototype stage - it is not in commerce yet - lack of historical data make impossible to do predictions
OPPORTUNITIES	**THREATS**
- increasing adoption of automation technologies in indoor farming - declining availability of farm workers - increasing government support to adopt to new agricultural technologies - growing demand for food and agricultural supply - they are planning to extend the use to more products	- lack of technical expertise and slow adoption to newer technologies - the market is expected to have lot of competitors

Table 9: Table 9: Swot analysis Agribot

The three startups are in an embryonic state where they are looking for capital in order to start production on a large scale. Among the three Vitibot is one step higher, having already obtained private funding of 3 million.

Vinbot and Agribot also got less funding for their projects. In particular, Agribot collaborates with an American company, Driscoll's, power in the American berry harvesting sector, which is testing the E-Series, in view of a market launch on Californian lands, those of the city of Watsonville, which can boast the 40% of strawberry crops in the state.

Vitirover, Vinbot's parent company, has instead received funding from the European Union's Seventh Framework Program for Research. Both have not yet put their product on the market.

To deepen the research on the technologies used by start-ups we will analyse the patents published by the three companies and the reference sector. To do this we use the code ICP as a reference, The International patent classification, established by the Strasbourg Agreement of 1971. It creates a hierarchical system of language-independent symbols for the classification of patents and utility models according to the different technical fields to which they belong.

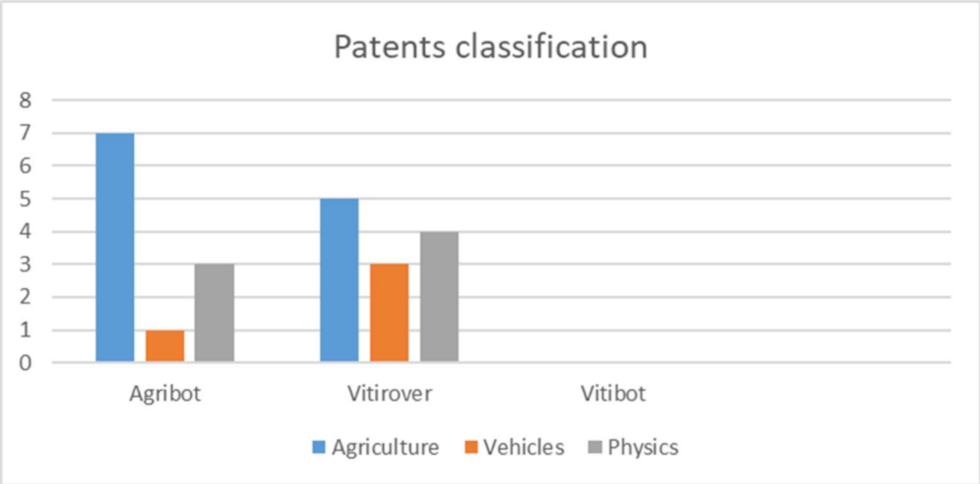


Figure 31: Patents of Agribot, Vitirover, Vitibot classified for IPC

The patents published by Vitibot are not to be found online, we have carried on a research on the possible technologies developed by them in the creation of Bakus (RQ2), so we take for granted the results obtained.

All the patents registered by the three start-ups relate to three technological fields: agriculture, vehicles and physics. Agriculture is the field with the highest number of registered patents, we assume that Vitibot follows this lead, since Bakus has several features belonging to the agricultural field (es. Pesticides and fertilizer, cutting system). Vitirover has also several patents regarding vehicles, since its main product is a robot, so as Vitibot, so we can assume a similar distribution, even without knowing the exact number of patents.

3.5 : DISCUSSION AND INSIGHT

The three start-ups work on complementary technologies, although within the same sector, in particular Vinbot and Vitibot both works in the vineyards using robots. Although they are still in an embryonic phase, the SWOT analysis was useful to better grasp their characteristics and their potential. Indeed, we can say that Vitibot has a great potential, thanks to its remarkable product (Bakus), as well as very few competitors in French. Its positioning in the market is strengthened by a highly qualified team.

Vinbot, instead, has solid financial basis, but its product is not in commerce yet. Moreover, the complex technology behind the robot could make it unaffordable.

Lastly, Agribot is entering in the US market, which is the reaches one in that field. There are several opportunities for this start-up, such as the declining availability of farm workers that could lead to increasing in crop harvesting robot demand. We think that both Agribot and Vitibot could become leading companies in their sector, while Vinbot may have some difficulties in placing its product.

Going deeper in our analysis, we found out that the technologies developed by Vinbot for the analysis of the vineyards through 3D images can be a potential competitor for Bakus, that is developing a similar technology to make the robot able to analyse the soil through a sensorial platform.

On the other side, the precision technology for harvesting developed by Agribot can provide Vitibot with a system capable of coming up with better yield from pruning the vines, especially the finest ones. The two companies may collaborate in future.

These are the only assumptions about the possible collaborations between these start-ups, currently unable to conquer an important slice of the market independently.

4.0 REFERENCES

1. Precision Agriculture - An International Journal on Advances in Precision Agriculture
2. Role of Patent Analysis in Business Strategy
3. Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organisation (WIPO)
4. Identification and Analysis of Technology Emergence Using Patent Classification
5. The Emerging Clusters Model: A tool for identifying merging technologies across multiple patent systems
6. H. Igor Ansoff, Strategic Management (1982)
7. https://www.wipo.int/patentscope/en/programs/patent_landscapes/
8. https://www.marketsandmarkets.com/Market-Reports/precision-farming-market-1243.html
9. https://www.marketsandmarkets.com/Market-Reports/precision-viticulture-market-116965282.html?gclid=Cj0KCQiAmZDxBRDIARIsABnkbYS2xWJ0_Id7Bo_3P9LuK1W5l8V4ojIlI3L3fatRMwHJKgcjS50VX-kaAvdqEALw_wcB
10. https://innovationorigins.com/start-up-of-the-day-using-robots-to/
11. https://cordis.europa.eu/project/id/605630/reporting
12. https://cordis.europa.eu/article/id/122476-an-autonomous-robot-to-predict-future-yields/it

    https://cordis.europa.eu/project/id/605630/reporting

13. https://www.technavio.com/report/global-crop-harvesting-robots-market-analysis-share-2018

    https://www.databridgemarketresearch.com/reports/global-harvesting-robot-market

14. https://agfundernews.com/harvesting-robotics-market-early-adopters-report.htmlhttps://www.pepperl-fuchs.com/global/en/27566.htm