Detecting emerging technological trends by patent text mining for the automotive industry.

Author: Stijn Lagrand
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,
In a VUCA world, firms are forced to become more dynamic. Quickly responding to new innovations is required to maintain a firm's competitive advantage. By being able to detect emerging patterns, firms can more effectively manage their R&D portfolio. This research addresses the problem of improving firms' ability to detect emerging technological trends, thereby improving their competitiveness and innovative capabilities. The findings contribute to the understanding of R&D portfolio management by detecting persistent, emerging, and fading patterns in the automotive industry. These patterns are identified by using patent text mining on the R&D portfolios of the top-5 firms over time. The study is relevant, since it assists firms in navigating through the dynamic and quickly evolving environment of the automotive industry by providing guidelines for firms' R&D strategies.

Graduation Committee members:

Dr. Igors Skute
Dr. Maximilian Goethner

Keywords
Patents, text mining, automotive, technological trends, emerging patterns, technology foresight, R&D portfolio, innovation.

1. INTRODUCTION

The world is increasingly becoming more VUCA (Volatile, Uncertain, Complex, Ambiguous) (Millar et al., 2018). Within this VUCA world, innovations are commonplace and required to remain competitive as a firm. Companies are forced to become more dynamic and agile to keep up with these innovations and new technologies. One specific industry that is impacted by this is the automotive industry. This is a high-tech industry (see 2.3), providing 16% of worldwide R&D expenditure (Center for Automotive Research, 2014). The industry consists of all the companies involved in the manufacture of motor vehicles, including the components (excluding tires, batteries and fuel) (Rae et al., 2023). This highly competitive industry is facing increasing pressures to develop itself (because of, for example, environmental concerns). Innovations are thus highly relevant for companies within this industry.

For companies, innovations are generally done by an R&D department that selects and pursues R&D projects. Companies have limited resources when it comes to pursuing R&D projects, they thus have to make strategic decisions to optimally make use of their (scarce) resources. In order to exploit its resources best, firms are at an advantage when they discover technological trends in an early phase (Day & Schoemaker, 2016; MacMillan & McGrath, 2002). Detecting emerging trends allows companies to make better-informed decisions (Irvine & Martin, 1984), and thus improve their R&D portfolio management. Companies should aim to create an R&D portfolio that is competitive (innovative projects that avoid the company from missing big opportunities) but also feasible (within their competencies) (Solak et al., 2010). Using the company's resources to their full potential, or at least better than rival firms, can be a source of competitive advantage (Mikkola, 2001). A balanced approach within the R&D portfolio and successful management of company innovations are co-determining factors for the profitability of a firm (Audretsch, 1995). With a growing number of businesses being labeled as high-tech, who per definition rely more strongly on innovations, the topic of managing R&D portfolios is becoming increasingly relevant. With cars developing towards more sustainable and autonomously functioning products, technologies are increasingly important within the automotive industry. The industry has shifted towards a high-tech industry due to this increased importance and use of technology (Center for Automotive Research, 2014).

The R&D portfolio of a company should strike a balance between the degree of competitiveness of R&D projects and their feasibility. These elements are harder to predict given the VUCA elements that currently exist, thus making it harder to optimally exploit the company's resources and manage its R&D portfolio (Petit, 2012; Solak et al., 2010). This while the R&D portfolio management of firms is playing a bigger role in firms' long-term success in an increasingly uncertain and dynamic world (Petit, 2012). Some of the reasons that R&D portfolio management is important are: avoiding sunk costs, by investing in irrelevant projects, preventing technological lock-ins, by diversifying investments, and fostering innovation, by encouraging employees to think creatively. These are all elements that are important for gaining and maintaining a firm's competitive advantage. External pressures on the automotive industry to focus on new sustainable solutions add to the importance of R&D portfolio management (Szász et al., 2021). These new solutions can open up opportunities for creating new sources of a competitive advantage. Gaining a competitive advantage based on R&D portfolio management and successfully detecting new technology trends have both become harder given the environmental uncertainties. It follows that also for investors it has become more difficult to select firms that are likely to perform well.

Research has been done on determining financial performance based on e.g. financial statements (Piotroski, 2000). This however gives an indication of the past of a company, and less of future performance (Balcaen and Ooghe, 2006). A better predictor for that would be the R&D portfolio (Petit, 2012). The challenge for firms lies in making an optimal selection of R&D projects to pursue, to effectively exploit their resources. To do this, firms must be able to predict in which direction innovations will go and have the ability to identify emerging patterns in an early stage. Emerging patterns and the direction of innovations are determined on an industry-wide scale, firms should thus investigate and map competitor's innovations to understand the industry dynamics. Day & Schoemaker (2016) refer to the ability of identifying emerging trends as *sensing*, and underline its importance for successful R&D portfolio management. *Sensing* is related to technology foresight (see section 2.3) and is an increasingly important capability for companies (Millar et al., 2018). The model from MacMillan & McGrath (2002), who identified five types of R&D projects, can help in this *sensing* element to find the projects that are most worthy to pursue. Research has also been done to determine a predictive relationship between semantic features within patents and firms' financial distress (Jiang & Zhou, 2023).

Since for companies it is best to detect trends when they are in a very early phase of development, any analysis tool that is developed for the detection of trends should be quick and dynamic. Firms should continuously monitor trends. If an automated process to analyse trends is used, patterns can be detected before the trend is formed. One such automated process is text mining, in the context of R&D portfolio management, this would be on patent information. Patents are commonly used for research on R&D portfolios, innovations and technological trends (Noh et al., 2015; Jiang & Zhou, 2023; Jung et al., 2016; Yoon et al., 2010; Yun et al., 2022). Patents provide structured information and are widely available in databases. Additionally, the nature of patents is to describe a novel method or (part of) a product, thus including an innovative element. Consequently, patents are a better source to detect emerging trends compared to other information sources (e.g. annual reports). The need for a dynamic and frequently updating analysis tool and the fact that patents have ample data extraction possibilities make it logical to use text mining.

In the automotive industry specifically, innovations are not researched through text mining yet. In this paper specific insights into the automotive industry will be found, which is relevant since innovations are industry-specific. This paper aims to detect technological trends by analyzing firms' R&D portfolios, comparing R&D portfolios and providing guidelines for how firms can improve their R&D portfolio management given the technological trends. The analysis of the technological trends and the R&D portfolios is done by text mining. Using text-mining analysis, the R&D portfolio of firms within the automotive industry will be clustered (using their patents) and analyzed. This would combine and build on the research done by Jiang & Zhou (2003), MacMillan & McGrath (2002) and Yun et al. (2022) to provide new insights into how R&D portfolios can be clustered and how these clusters can be used to identify trends. The findings should aid automotive firms in identifying technological trends in an early stage, deciding which type of R&D projects to invest in, and for external investors to analyze how a firm is likely to perform based on their innovations.

To investigate this gap in the current literature, the following question is researched: *To what extent can technology foresight and a comparison of R&D portfolios improve (high-tech) automotive firms' ability to detect new business opportunities?*

It is expected that specific technological trends for the automotive industry can be identified. This would have implications for the management of R&D portfolios within companies, as it can guide them in their decisions which R&D projects to pursue. The automated method created in this research means that managers can frequently replicate the industry-analysis, and find new trends. The method and current findings can aid in firms' degree of technology foresight. Academic implications are the extensions on the existing literature on R&D portfolio management, which are combined with new and industry-specific insights. The framework of MacMillan & McGrath (2002) is applied to a real-life scenario. Policymakers can use the insights from the research to make and/or adjust legislation appropriately. This allows for ample innovation within a country, giving it an edge over other countries, but can still protect itself and its citizens from unwanted consequences of innovations.

2. THEORETICAL FRAMEWORK

2.1 Importance of Emerging Technological Developments and R&D Management

Innovations are a driving factor of societal, technological and economic developments (Rao et al., 2001). They are the key drivers of progress and change in those areas, innovations are able to: transform norms and values within a society by challenging current beliefs, develop new technologies that improve our lives and create new markets and jobs, stimulating economic growth (Audretsch, 1995; Tewksbury et al., 1980).

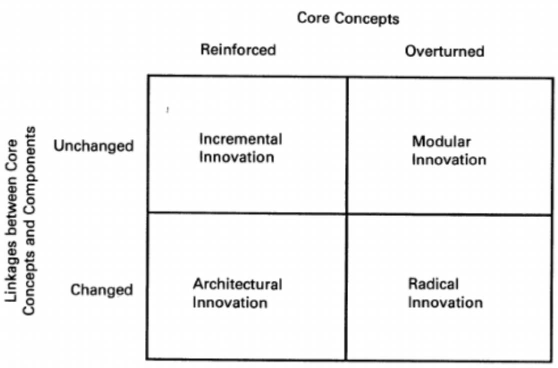

Figure 1: A framework for defining innovation. Retrieved from Henderson & Clark (1990).

Innovations come in multiple different types, a common method to distinguish innovations is based on the degree of novelty: incremental and radical innovation. Incremental innovation builds on current products/ideas/technologies and improves them, radical innovation is new and incorporates fundamental and revolutionary change (Song & Di Benedetto, 2008). Henderson and Clark (1990) have built on this model and added architectural and modular innovation types, see *figure 1*. An innovation can be classified as modular when core concepts of a technology are overturned, but the linkages between different core concepts remain unchanged. Architectural innovations develop existing technologies but change the way concepts within technologies are linked. As all of these types of innovation are relevant and important for (automotive) firms' long-term performance, they are all included in this research.

Which innovations to research is part of the management of the R&D portfolio. R&D portfolio management under uncertainty is relevant, given the importance of having a competitive R&D portfolio and the high levels of uncertainty that currently exist (Millar et al., 2018; Troise et al., 2022). Five types of R&D projects have been identified by MacMillan & McGrath (2002), using market uncertainty and technological uncertainty as dimensions to cluster the projects on. Market uncertainty here refers to lack of guarantee of sufficient demand, if the market will accept the new product, if there will be any problems regarding regulations and how competitors will respond. In the automotive industry, an example could be the (re)introduction of electrical vehicles, with Toyota's Prius being the first mass-produced electrical car (Matulka, 2014). Given its shorter range, it was uncertain what the demand for the car would be. Ettlie et al. (2021) found that R&D dynamic capabilities within the automotive sector are related to R&D project success, considering regulatory changes (a type of market uncertainty) for the automotive industry. Technological uncertainty is defined as the lack of knowledge to how the product will perform and how feasible the technology is. Considering electrical vehicles again: before the successful launch of the Toyota Prius, companies failed in developing a commercially viable electrical vehicle due to the high production costs (the EV1 by General Motors for example) (Matulka, 2014).

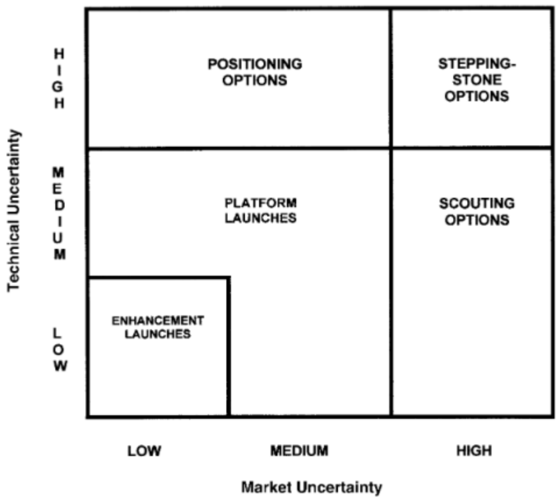

Figure 2: Technical uncertainty vs Market uncertainty. Retrieved from MacMillan & McGrath (2002).

Besides clustering R&D projects based on the aforementioned uncertainties, a relevant aspect lies in identifying technological trends. In their research, Yun et al. (2022), found a method to identify technological evolution by analyzing patents. They were successful in showing that technologies within the bio-healthcare have "evolved towards enhancing data quality or energy efficiency after ensuring functional diversity" (Yun et al., 2022, p. 58).

2.2 Text Mining in R&D Management

Patents can, among others, be used to evaluate R&D performance (Li et al., 2009) and to predict emerging technologies (Basberg, 1987). More recently, patent analysis has been done through text mining. Fattori et al. (2003) found that text mining is an effective method of classifying patents, and could even overcome limitations that exist in the conventional way of classifying patents. Text mining allows for the extraction of large amounts of data, which can provide useful insights for industry-wide research. In addition to being able to analyze larger volumes of data, compared to manual analysis of patents, the process can be automated. This is a crucial advantage of text mining in this context. The VUCA world requires making decisions under high degrees of uncertainty, and with an ever-increasing need to make decisions quickly. Manual analysis of patents for R&D portfolio management limits the speed with which analyses can be carried out, this limitation is overcome by text mining. Managers and researchers can increase the amount of data they analyze and the speed of the analysis by applying this technique. This allows for an earlier detection of technological trends.

Noh et al. (2015) provide an overview of keyword selection and processing strategies to analyze patents through text mining. They focus on four relevant factors: (a) which part of the patent documents to select keywords from, (b) what methods to use to select keywords, (c) how many keywords to select, and (d) how to turn the selected keywords into an analyzable format. They found that for (a), (b) and (c) the differences were statistically insignificant, thus that any choice for these elements are not critical in affecting the keyword-based patent analysis results. For (d), the TF-IDF method was found to be the most effective.

Analyzing patents with text mining to infer patent typologies, technological trends within industries, financial performance of firms or other business elements is already researched extensively, substantiating the predictive power of patents for R&D (Noh et al., 2015; Jiang & Zhou, 2023; Jung et al., 2016; Yoon et al., 2010; Yun et al., 2022).

In their research, Jiang & Zhou (2023), found a model to predict financial distress using patents. Financial distress is used as a dichotomous variable (yes/no) and is predicted by combining accounting and patent features. The patent features are split into statistical and semantic features. They preprocessed the data for semantic patent features by word segmentation, using the Jieba package within Python, and stop word removal. The pre-training model BERT was used to capture the semantic features within the patent text. The model was developed and tested by having a training dataset (70%) and test dataset (30%). The machine learning models used to predict financial stress are eXtreme gradient boosting (XGB), logistic regression (LR), random forest (RF) and gradient boosting decision tree (GBDT). With Shapley Additive exPlanations (SHAP) the individual feature importance is calculated. They found that 'beneficial effects', 'valid authorization ratio' and 'technological novelty' have the highest importance, amongst the patent information, for predicting financial distress.

Jung et al. (2016) used text mining to "identify the notable trends and technologies being developed applied to enable IoT in the field of logistics" (p. 624). They used KIPRIS, a Korean patent search engine, to collect the information on patents containing the keywords 'internet of things' and 'logistics'. Technology is divided into eight sections according to the International Patent Classification (IPC) system. The patents were matched with these sections, from which the authors inferred that diverse technologies are relevant for IoT within logistics. To analyse the patent abstracts, Jung et al. (2016) pre-processed the data by making corpora of the datasets, filtering out stop words, stemming, and converting the refined text data to a matrix format. They then calculated the sparsity of the word matrix to determine the extent to which the datasets must be refined. To track the trends in patent abstracts they analysed term frequency (removing the two most common terms: 'system' and 'information'). Lastly, an analysis of associated terms was used to cover an additional angle in the technological trend prediction. They showed that "technologies in the IoT for the logistics industry are evolving by combining with the cloud and big data technologies" (Jung et al., 2016, p. 631).

2.3 The Role of Technology Foresight on R&D Management

Managing their R&D portfolio and tracking emerging technologies is more relevant for high-tech companies, since innovations are occurring at a faster rate and are of bigger importance for the competitiveness of those firms. High-tech companies rely on innovations and their R&D departments more than non-high-tech companies. There is no clear consensus on a single definition of high-tech, and what industries are included or not. Combining multiple definitions, an industry can be classified as high-tech when it meets the following criteria (Center for Automotive Research, 2014):

1) R&D costs are 3% or more of the total output.
2) A minimum of 10% of the employees are technical (e.g. engineers, scientists).
3) Uses scientific and technical knowledge in the design and/or production.
4) Actively designs, develops and introduces new products.
5) Actively designs, develops and introduces innovative manufacturing processes.
6) Has a geographic cluster of innovation and allows these to develop further.

A method for high-tech companies to determine in which domains to innovate is technology foresight. Technology foresight can be defined as a systematic exercise aimed at looking into the longer-term future of science technology and innovation in order to make better-informed policy decisions (Irvine & Martin, 1984). The four most distinctive features of technology foresight are (Pietrobelli & Puppato, 2016):

1) Influencing the technology direction when attempting to predict the future.
2) Inclusion of new actors by having a participatory approach. This can increase the range of possible strategies.

3) Being pursued at multiple levels (organisational, local, governmental, international).
4) Increasing the connection between industries, universities, governments and the society.

Technology foresight includes attempting to predict the future to determine what to do in the present. This ‘look into’ the future might be more difficult than ever, as the VUCA-world implies there to be many uncertainties and unexpected events. An example from a different industry would be the difficulty in attempting to predict trends in the cryptocurrency market, which can be significantly impacted by the tweets of a single individual, Elon Musk (Ante, 2023).

Technology foresight is relevant for R&D management because firms need to start R&D projects in the present, to have the finished product in the future. High-tech R&D projects take years to develop, foresight is required to help determine if the projects are still relevant by that time. Jung et al. (2016) and Yun et al. (2022) were able to detect technological trends, which is a part of technology foresight. The latter state that “the direction of future development is proposed by referencing the evolution patterns in other sectors as an approach for discovering technology opportunities for companies” (Yun et al., 2022, p. 58). Firms can thus use the evolutionary patterns to determine what projects they need to select for their R&D portfolio, to develop projects that are in a relevant in the future. In doing so, the future is also shaped by the present, as Pietrobelli and Puppato (2016) note.

Investing in R&D projects that are relevant in the future is important for companies to remain competitive. Low technology foresight capabilities can lead to selecting irrelevant R&D projects, which will lead to unnecessary losses of resources (Pietrobelli & Puppato, 2016). Additionally, detecting technological trends and developing products that follow that trend is important since future developments often depend on developments done before. It can be more difficult for a company to start working on far-developed products, as specific knowledge and/or technologies might lack behind. A higher capability in technology foresight should thus lead to better financial performance.

For firms in the automotive industry technology foresight is, of course, also relevant. Major evolutionary patterns would be for example increasingly sustainable and fuel-efficient cars. In the Automotive Roadmap 2020-2030 there is a strong emphasis on these sustainability elements, smart technologies and safety are also found to be future development paths (Konings et al., 2020). The roadmap includes multiple directions for R&D projects (e.g. sustainable mobility, powertrain concepts for renewable energy carriers, smart mobility and AI) to remain competitive and drive changes. These directions are determined by combining societal challenges, governmental regulations and research agenda’s, one of the features of technology foresight (Konings et al., 2020; Pietrobelli & Puppato, 2016).

3. RESEARCH DESIGN

3.1 Data

Data is collected from ORBIS, a database developed by Bureau van Dijk. It contains data from millions of companies on a global scale. Financial information, ownership structure, industry classification and intellectual property are some of the key data ORBIS provides (Bureau van Dijk, n.d.).

The automotive industry as defined in the beginning of chapter 1 relates to ISIC code 29 (UN Department of Economic and Social Affairs, 2008, p. 149). To have a more specific analysis, the research focuses on the manufacture of motor vehicles (thus excluding the manufacture of components). Filtering the ORBIS database on companies with this industry classification (ISIC code 291), selecting only active and publicly listed firms (to have ample data available) that have a minimum of 1 patent, and removing firms without known financial & employee information gives a sample size of 193 firms. The analysis in this research is conducted on the top 5 biggest firms (based on the last-known revenue) from this sample, see table 1. Revenue is chosen as a metric for selection because high-revenue firms have sufficient funds to continuously innovate and discover new types of technologies. These companies are the industry leaders, they thus determine (to a large extent) the developments within the industry. Research on their patents is thus useful for the analysis to infer trends for the industry as a whole.

The data that is available for download from the ORBIS database is structured in a way that is practical for text mining. Data is exported in an .xlsx format. Patent titles, dates of the patents and company names are three of the main columns provided. The dataset will be large, since the number of patents for the top automotive firms is high. The information is separated for each company on different tabs.

Table 1. Top-5 Companies in the Automotive Industry Retrieved from ORBIS

Company name	Revenue (last-known, thousands)	Patents (granted)
Volkswagen AG	$310,607,637	65,436
Toyota Motor Corporation	$256,368,535	188,656
Stellantis N.V.	$191,552,735	42,868
Mercedes-Benz Group AG	$162,730,017	23,014
Ford Motor CO	$158,057,000	56,979

3.2 Text Pre-Processing

The number of granted patents, the titles of the granted patents, and the dates of publishing for each patent are exported as a .xlsx file. In Microsoft Excel, the data is restructured: deleting the upper section of the sheets that do not contain patent titles. Patent titles with non-Latin characters (e.g. Chinese or Arabic) are removed from the dataset. The .xlsx files are then exported to R for further pre-processing and data analysis.

The companies’ patent information are first merged into a single large dataset. They are then separated, based on the publication date of the patents, into three categories: 2010-2014, 2015-2017 and 2018-present. These intervals are chosen so that the number of patents in each category is evenly spread. To pre-process the raw data for analysis, patent titles are converted to lowercase, symbols, numbers and stop words are removed and the words will be stemmed. These are common steps in text pre-processing (Munková et al., 2013) and are the same approach as Jung et al.

(2016), Jiang et al. (2023) and Yun et al. (2022) used. The *tm* and *textclean* packages are used for all preprocessing steps. Text mining is done on n-grams, where n refers to the number of words that are analyzed together. Since the patent titles are short, using higher-order n-grams than unigrams (n=1) will result in a limited amount of n-grams per patent title. The information increase for an n-gram with n=2 is limited, while the data frame size increases substantially, which significantly slows down the analysis process. Since this analysis method is aiming to be reproduced frequently, a fast and efficient method is preferred. The patent titles are thus separated into unigrams. Additionally, patent titles shorter than three words are also removed from the dataset, since they provide too limited information to accurately perform clustering. Term frequencies can be analyzed after the preprocessing, the most common and uninformative terms are then deleted as well (e.g. system, device, vehicle et cetera).

In text mining, it is possible to use supervised or unsupervised approaches. A supervised approach allows for classifying documents into pre-specified classes. An unsupervised approach can be used to cluster documents, without the need for any ex-ante decisions. The aim of this research is to detect emerging patterns, which are not known in advance. It is thus not possible to create classes prior to the text analysis. An unsupervised approach is thus used for the text analysis.

A corpus is created from the preprocessed patent titles, in which terms that occur in less than 125 of the patent titles are removed. As described in section 2.2, Noh et al. (2015) identified that TF-IDF (equation 1) is the most effective/accurate unsupervised method to analyze a text-mining data sample. TF-IDF measures the importance of a word in the dataset. From the corpus, a term-document-matrix is created with TF-IDF weighting. Any clustering algorithm requires a similarity measure to be able to determine how distant terms are from each other (and thus whether terms belong to the same cluster). Numerous similarity measures exist, Singh & Singh (2021) found from a comparative analysis that the cosine distance (equation 2) has the greatest accuracy. The cosine distance measures how similar terms are by calculating the angle between the terms. Using the TF-IDF values, the cosine distances are calculated for all terms. From the distance values a similarity matrix is constructed.

Equation 1: $\textit{TF-IDF}\text{weight} = \sum_{i \in d} tf_{i,d} * \log(\frac{N}{df_i})$

Equation 2: $\text{similarity}(A, B) = \frac{A.B}{||A||X\,||B||} = \frac{\sum_{i=1}^{n} A_i\, X\, B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$

3.3 Hierarchical Clustering

The textual data should be clustered to be able to compare the company's R&D portfolios, the distance matrix made with the cosine distance is used for this. There are numerous methods to divide the preprocessed data into clusters. Hierarchical agglomerative text clustering, using Ward's minimum variance, will be used for this research. This is a hard clustering technique, where each patent title belongs to one cluster: patent titles are merged when they are similar until the pre-determined number of clusters is made (Burghardt & Cavanaugh, 2022). Multiple methods exist to determine this number of clusters, a commonly preferred option is the elbow-method because of its simplicity and lack of computational power required (Mirkin, 2011). By plotting the within cluster sum of squares (WSS), the appropriate number of clusters can be identified. This method will be used in the analysis process.

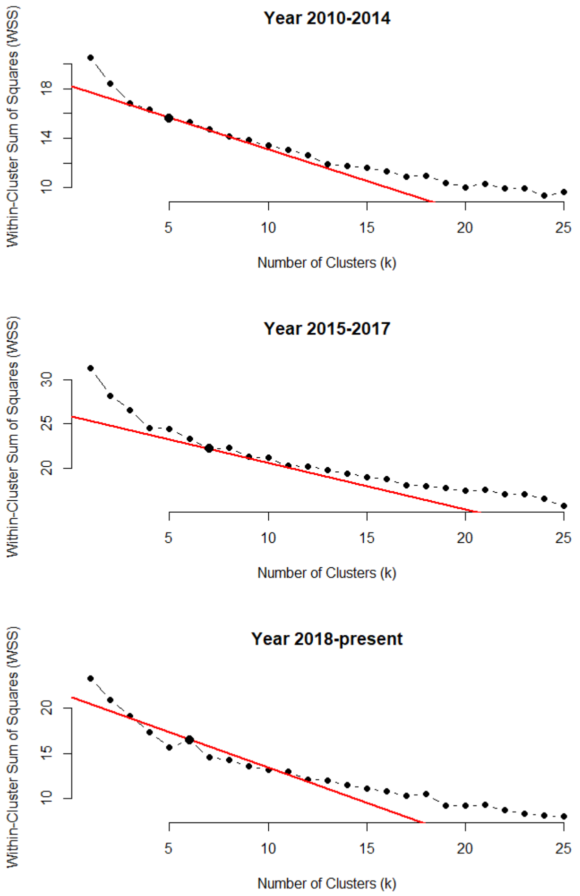

Figure 3: WSS as Function of the Number of Clusters

The plots indicate where the within-cluster sum of squares no longer decreases strongly with an increase in the number of clusters. For each of the time categories there will be 5, 7 and 6 clusters, respectively. These numbers of clusters are confirmed by manually checking the distribution of the terms when making as little as 4 clusters up to a maximum of 9 clusters for each time category. Any fewer clusters than determined by the elbow point results in a loss of information, increasing the number of clusters did not provide additional insights (e.g. a cluster with only the terms 'least' and 'one', from which not a meaningful (technological) theme can be identified).

3.4 Data Analysis & Cluster Labeling

For each company, their patents are clustered by time-category using the steps outlined in section 3.3. After the patents are clustered, the main topic for each technology cluster is identified using word clouds, term frequencies and by analyzing the structure of the dendrograms. A secondary researcher is asked to review each cluster's theme to ensure validity. Each cluster is classified on its market- and technical uncertainty using MacMillan & McGrath's (2002) questionnaire (appendix 7.2), retrieving information from the word clouds and manually looking for more detailed patent information from e.g. Google Patents. Using these classifications, plots based on MacMillan & McGrath's framework (2002) are made to depict the types of technologies in the automotive industry (see section 2.1).

The R&D portfolio plots are compared for each time category. By comparing the R&D portfolio plots for the automotive industry over time, technological trends and potential shifts in the type of R&D projects pursued can be identified.

4. RESULTS

In this section, the results of the analysis and the final themes found are depicted. The number of clusters (see 3.3) determines the number of key technological themes that are found. There are 18 clusters in total. The time category 2010-2014 is composed of 225 terms, which resulted in 5 technological themes. The years 2015-2017 and 2018-present consist of 315 and 212 terms with 7 and 6 technological themes, respectively. The distribution and clustering of the terms for each time category is visualized using dendrograms, which are attached in appendix 7.1. The terms for each cut tree in the dendrogram can then be used to determine the common themes.

A shortened version of the results is presented here to explain the analytical process, cluster 2 of 2010-2014 is analyzed as illustration for the entire process. The cluster is made up by the following terms, depicted in *figure 4* as a word cloud. A combination between the word cloud, the term frequencies and the structure of the dendrogram (appendix 7.1) are used to determine the common theme for cluster 2 of 2010-2014.

Figure 4: Cluster 2 Word Cloud in 2010-2014

Multiple terms indicate (raw) materials used in the production of automotive vehicles, e.g. alloy, steel, iron, fiber, carbon, and aluminium. Other terms also are related to the production process, e.g. tool, inspect, aftertreat, workpiece, prepare and test. The common theme identified for this cluster is thus: (raw) materials and production. The other clusters have their themes identified in similar fashion. The themes of all clusters are summarized in *table 2*.

Table 2. Technological Themes in the Automotive Industry

Cluster number	2010-2014	2015-2017	2018-present
1.	Battery-technologies	Battery-technologies	Battery-technologies
2.	Combustion engine and exhaust system	Combustion engine	Combustion engine
3.	Components and features	Exhaust system and emission control	Computer systems
4.	Variety of technologies	Variety of technologies	Variety of technologies
5.	(Raw) materials and production	(Raw) materials and production	(Raw) materials and production
6.	N/A	Energy systems	Exterior parts and components
7.	N/A	Powertrain systems	N/A

Battery technologies refers to basic elements for batteries and various types of batteries, which are used to store electric energy. Batteries are required in fully electric and hybrid vehicles. The combustion engine is the traditional manner in which automotive vehicles work: gasoline or diesel is ignited, which moves the cylinders that in turn power the axles that drive the car. The gas after ignition is emitted through the exhaust system. Emission control systems aim to reduce the harmful components that are emitted through the exhaust system. Besides a power source (like electricity, gasoline or diesel), cars include many digital systems to optimize processes and provide additional features (e.g. navigation or radio). These digital elements require the computer systems that are identified in 2018-present. The themes: 'components and features' and 'exterior parts and components' refer to specific elements inside and outside of automotive vehicles. These topics are generic, and include terms like 'airbag', 'mirror' and 'bumper'. Powertrain systems and energy systems are both related to the thrust or movement of the car, where powertrain systems relate to the components and control mechanisms required to deliver power to the wheels of a vehicle and energy systems relate to those elements that provide the means to generate that power. Although it appears as if the themes 'variety of technologies' and '(raw) materials and production' are overlapping, there are differences in the terms from which these themes are determined. The 'variety of technologies' clusters consist of many different terms, from 'air' to 'brake' and 'filter'. The size of this cluster (n=200, n=200, n=174) is a barrier to properly identifying a single theme, or clear differences between the time categories. The '(raw) materials for production' theme includes subtle differences over the time categories: in 2010-2014 the focus within the cluster is on broader-range manufacturing terms. In 2015-2017 there are more terms regarding materials used in production, with a focus on energy storage. The terms belonging to the years from 2018 to the present specify the manufacturing steps, with terms referring to heat-related processes. *Table 3* provides a selection of terms for these overlapping clusters to illustrate their differences. Three categories for the aforementioned themes are formed based on *table 2*: persistent technologies, emerging technologies and fading technologies. The category to which each theme belongs is discussed in chapter 5.

Table 3. Difference in Terms for '(Raw) Materials for Production'

Year	Term 1	Term 2	Term 3	Term 4
2010-2014	Weld	Metal	Treatment	Material
2015-2017	Lithium	Coat	Layer	Composite
2018-present	Heat	Catalyst	Pressure	Gas

The technical- and market uncertainties are identified for each of the clusters, resulting in the plot at *figure 3*. Numerous themes score very similarly on both dimensions: clusters 2,3 & 5 (2010-2014), 2, 5 & 7 (2015-2017) and 2, 5 & 6 (2018-present) form a group on the transition from enhancement launches and platform launches, clusters 1 (2010-2014), 1 & 3 (2015-2017) and 1 & 3 (2018-present) are plotted in the top-right corner of the platform launches. The substantiation for technical- and market uncertainty scores for each cluster is in appendix 7.3.

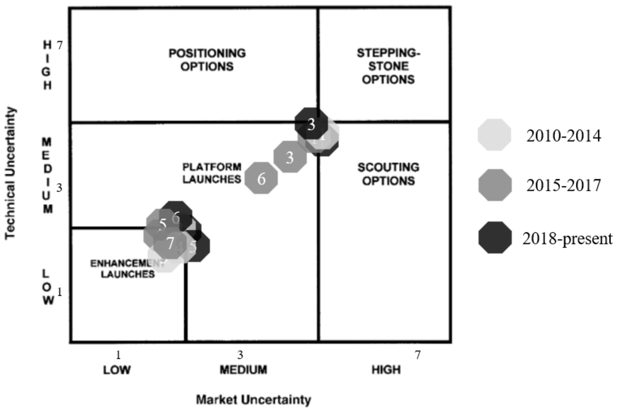

Figure 5: Automotive Industry R&D Project Typologies. (Numbers Correspond with Table 2)

Definitions for each type of uncertainty are given in section 2.1. In practice, the projects with low degrees of technical/market uncertainty are relatively sure to be adopted by consumers and have known and proven production/development methods. These projects are the enhancement launches, and overlap with the definition that Hendersen & Clark (1990) give for incremental innovation. The projects with higher degrees of both kinds of uncertainty, stepping stone options have more risks related to technical feasibility and whether the end-user will have a demand for the product. These types of projects are usually required to invest in to ensure a competitive R&D portfolio in the future, and are related to radical innovations. There are roughly two types of R&D project types that can be identified for the automotive industry: enhancement launches and platform launches to stepping-stone options. The plurality of the enhancement launches shows that automotive companies are focusing on improving existing technologies, and developing those into newer products or methods. The R&D projects that are pursued with higher scores for market- and technical uncertainty align with the definition earlier for high-tech companies, as these are the type of projects that include new products and production methods (Center for Automotive Research, 2014).

Based on the theoretical framework formed in chapter 2, two propositions are formed. These will be discussed in chapter 5, together with the main research question.

4.1.1 Theoretical Proposition 1

(High-tech) firms that actively engage in technology foresight practices are expected to have a higher capacity to detect emerging technologies, leading to a more balanced distribution of resources across positioning, stepping-stone, enhancement, platform and scouting options.

4.1.2 Theoretical Proposition 2

Firms that have a high degree technology foresight are more likely to detect both persistent and fading technologies, enabling them to allocate resources strategically for developing persistent technologies and managing the transition away from fading technologies.

5. DISCUSSION

Given the increasing VUCA-circumstances in which (automotive) companies operate currently (Miller et al., 2018), foreseeing in which technologies to invest is becoming increasingly important to gain and maintain a competitive advantage (Day & Schoemaker, 2016; MacMillan & McGrath, 2002). The aim of this study is to find the degree to which technology foresight and a comparison of R&D portfolios can help firms in detecting new business opportunities, which is further discussed in section 5.1.4. In this research, technological trends in the automotive industry are discovered by means of a method that can be easily reproduced, which is vital given the high-speed technological developments in the industry and its environment (Szász et al., 2021). The technological trends are analyzed by text mining the patents of the top 5 automotive companies (see *table 1*). The patents are divided over three time categories, to see how the types of projects pursued in the R&D departments change over time. For each time category, the patents are then clustered (based on unigrams), from which the common themes in each cluster are identified (see *table 2*). The themes are then plotted on the framework developed by MacMillan & McGrath (2002).

The results indicate numerous findings, and will be discussed in the following order: persistent technologies. technologies emerging and technologies fading away. The R&D project types indicate roughly two types of projects that are pursued by companies in the automotive industry: enhancement launches and platform launches to stepping-stone options. This is a balanced distribution of the types of R&D projects. Car companies are clearly focusing on improving existing technologies, investing in incremental innovations to exploit their knowledge and resources. The increasing uncertainty that the industry is facing is also reflected, as companies have been investing more in technologies with higher degrees of uncertainties to be prepared for the future. There is a shift over time, where in the earliest time category only one theme is classified as uncertain on both dimensions and in the third time category there are two, with the middle time category illustrating the transition.

5.1.1 Persistent technologies

The first type of technologies, persistent technologies, are: ‘combustion engine’, ‘battery technologies’ and ‘(raw) materials and production’. These technologies are present in all three year categories.

The combustion engine remained almost entirely the same, with the exception being the separation of exhaust systems from the combustion engine in 2015-2017. This is further discussed under emerging technologies. This shows that the (traditional) combustion engine is a fundamental technology within the automotive industry and that, despite increasing pressures on improving sustainability, this remains one of the main topics automotive firms are researching. Numerous patent titles about internal combustion engines are focusing on reducing emissions and improving efficiency. This indicates that the automotive industry is improving combustion engines into a more energy-efficient and less-emitting technology, incrementally innovating one of the core automotive technologies to ensure it maintains relevant and competitive.

Battery technologies have remained unchanged over the three time categories, besides the battery-cluster for 2018-present also including some broader terms. Automotive companies thus have had battery technologies on their agenda for some time, but have not made any major changes in their research direction. In other clusters, e.g. the 'variety of technologies', more battery technology related terms can be found in the later time categories. This might indicate that a larger variety of battery-related technologies are being developed, or different applications of the technology.

Automotive companies continue to invest in (raw) materials and production. Intuitively, this is logical: new production methods and new ways to utilize (raw) materials are found and to exploit them fully the companies should invest in them. There is a shift within this theme over time, where initially the focus was on basic production methods and raw materials and in the last time category more specific (new) production methods are mentioned. There is a small decrease in the patents referring to basic raw materials, such as metal. These shifts are discussed in sections 5.1.2 and 5.1.3.

5.1.2 Emerging technologies

Some technologies are emerging over time, sometimes appearing as completely new and sometimes having its terms first presented in the cluster 'variety of technologies' and the following time category as a cluster on its own. The emerging themes are 'computer systems', 'exhaust systems', 'production techniques' and 'battery technologies'.

Computer systems is a technology-type that intuitively has gotten more important over time. More and more computer systems are included in automotive vehicles, from navigation systems to systems that automatically detect slowing-down traffic in front of the vehicle. The terms in the cluster also relate to data storage, as all the information that is collected from its users is used to learn from and improve future developments. This technology deals with higher degrees of uncertainty than others, as digital developments are occurring at a high speed and with increased chances of the technologies turning obsolete (Ates & Acur, 2022).

Exhaust- and emission control systems have clearly emerged from being combined with another technology, the internal combustion engine, to an independently researched theme. These systems are responsible for managing and controlling the exhaust gases produced by the engine. The separation from the internal combustion engine implies that the importance of the technology has increased for automotive firms. External influences from environmental groups and governments might have played a role in this, as the industry is pressured to decrease its emissions. Conforming to regulations upheld by institutions is required to be able to compete, but firms are also having their own (more ambitions) goals in terms of environmental footprint. Having low-emitting cars is a competitive advantage, as customers are finding sustainability an increasingly important topic (Van Doorn et al., 2021).

As mentioned in section 5.1.1, production techniques are changing and turning from broad and general to more specific. In the most recent time category, heat-treatment and pressurized forming production methods are mentioned. These methods can improve the quality of the final product or reduce production costs, by being quicker or more efficient. New types of composite materials also require different types of production methods. The increasing number of terms related to production methods, rather than raw materials can indicate that firms shifting in where they are attempting to create a competitive advantage.

Within the battery technologies cluster, but especially within the 'variety of technologies' cluster there is a clear increase in the number of terms related to batteries. The importance of batteries appears to be increasing, but also new types of batteries are being developed. Lithium-ion batteries, a rechargeable battery, is an already known battery-type that receives more attention in the last time category. This emergence of batteries is related to the shift in cars being powered by fossil fuels to electricity, which is in turn (partly) caused by the aforementioned environmental pressures.

5.1.3 Fading technologies

The themes about 'raw materials and productions' show gradual changes over time: going from focusing on the (simple) materials used in production, to specific and advanced production techniques. This can indicate that raw materials do not represent the core value anymore and no longer provide firms with a strategic competitive advantage. The materials are considered to be the standard in the industry. Instead, production techniques (such as heat-treatment) are emerging as a more important topic, as discussed in 5.1.2.

5.2 Conclusion

This research aims to find an answer to the question:

To what extent can technology foresight and a comparison of R&D portfolios improve (high-tech) automotive firms' ability to detect new business opportunities?

The R&D portfolios of five automotive companies are compared, from which the aforementioned technological themes are found. The identified technological themes and how they have been developing since 2010 provide a guideline for automotive firms' investments. By specifically researching and investing in technologies that are emerging, new business opportunities can be found. The comparison of R&D portfolios in this paper provide an overview of these emerging technologies, thus improving automotive firm's ability to detect new business opportunities.

Overall, the study highlights the significance of technology foresight and the comparison of R&D portfolios as powerful tools for improving firms' ability to detect new business opportunities. In section 5.3 and 5.4 theoretical and practical implications are given, aiding firms in navigating through the dynamic landscape of the automotive industry and providing a guideline for the decisions regarding their R&D investments.

5.3 Theoretical Implications

This paper provides a real-life application of the framework developed by MacMillan & McGrath (2002). Practical applications of theoretical frameworks test the frameworks, increases the awareness and expands the knowledge of the framework. The findings from the automotive roadmap on macro-level trends are consistent with the findings in this research: the automotive industry is increasing the emphasis on sustainable solutions and smart technologies (Konings et al., 2020).

Other research has been done on detecting emerging trends by text mining patents, with a different methodology (Jung et al., 2016; Yun et al., 2022). Yun et al. (2022) determine the purpose of patents before clustering, thus finding trends of technological purposes rather than trends in technology types. Jung et al. (2016) researched trends for categories of technologies, classifying patents using IPC codes. In this analysis, patents were clustered and the theme was determined based on the cluster content. Jung et al. (2016) determined the theme for each patent before grouping them, and then analyzed how the composition of each group changed over time. This research provides an additional method to a plurality of researches in detecting trends. A side-by-side comparison of different methods in detecting trends can be made, to determine the most effective approach.

Technology foresight is vital for firms to have a consistent and well-balanced R&D portfolio, as discussed in section 2.3. The methodology employed in this research can enhance a firm's technology foresight capabilities by providing a structured and systematic approach to anticipate future developments. This aligns with Irvine and Martin's (1984) definition of technology foresight. Technology foresight helps shaping a firm's vision of the future, which in turn (partly) determines the types of technologies a firm will attempt to pursue. Different types of technologies require different strategies regarding resource allocation, which is elaborated upon in section 5.4.

Both theoretical propositions (see 4.1.1 and 4.1.2) can be confirmed on the basis of the results in this paper. The first theoretical proposition relates to the detection of emerging technologies and a balanced R&D portfolio. By having multiple project-themes with high uncertainty, the automotive industry meets the measured requirements for a high-tech industry (as defined in 2.3). These firms require high degrees of technology foresight to know in what specific projects to invest for the themes with higher degrees of uncertainty. The automotive industry has a significant number of emerging technologies that are detected, substantiating the high capacity of trend detection. The R&D project types are balanced (see *figure 5*), thus firms in the automotive industry distribute their resources well. The balanced distribution of resources means that investments are aligned with the potential of emerging technologies.

The second theoretical proposition concerns persistent and fading technologies, and how proper detection of these types can aid in strategic allocation of resources to develop the persisting technologies while transitioning away from the fading technologies. From the results and previous discussion, it becomes apparent that the automotive industry has indeed slowly been transitioning towards relying less on raw materials. Its focus has been on improving some traditional technologies, such as the internal combustion engine and novel production methods, instead. Technology foresight helps firms to classify technologies as being persistent or fading (or emerging), thus aiding in the strategic allocation of resources for their R&D portfolio.

5.4 Practical Implications

The findings and method of this report are useful for R&D managers, entrepreneurs and professionals. The text mining approach used in this paper can be used to detect trends in an early phase, which will help firms gain and maintain a competitive advantage. Dynamic patterns on an industry-wide scale can be detected due to the large amount of data used with the text mining approach. Because the analysis is build on text mining, the process can also be automated and repeated frequently. This means managers can continually use the method used in the paper to detect trends and patterns. Although the focus of the paper is the automotive industry, with minor adjustments the analysis can be applied to other industries as well.

The results show a balanced R&D portfolio, between improving current technologies and researching new ones. In more recent years, there have been more projects that are uncertain. This means that automotive companies are realizing the many different ways in which the industry can move, and want to prepare for what is ahead. Current R&D managers should keep focusing on improving the core technologies, but increasingly pay attention to more radical projects as well. New production methods, computer systems and an increased importance of batteries are examples of those more uncertain projects that will be vital for the future competitiveness and success of automotive companies.

Automotive firms can improve their strategic allocation of limited resources by complying with the expected actions belonging to each R&D project type. Technological themes which are identified as enhancement launches require consistent input of resources, so that technologies can be improved at a steady rate. These themes are low in uncertainty and focus on improving current technologies, and are thus a form of incremental innovation. The more uncertain themes, platform launches to stepping stone options, need less consistent input of resources but require a substantial commitment for each theme to develop. A large amount of resources might be contributed to a specific theme without there being a clear immediate return. These themes carry more risk, but are required for long-term success of firms. A firm's own risk tolerance and overall strategy should be considered for the strategic resource allocation, those elements determine the proportion of resources provided to each R&D project type: a firm aiming to provide unique automotive vehicles at a near-zero emission rate should invest a larger amount of its resources to stepping stone R&D projects than a firm aiming to provide a large range of vehicles, competing with the majority of the industry.

5.5 Limitations & Future Research

For this research, a quantitative text mining approach is used, enabling to conduct a comprehensive analysis of the entire automotive industry and detect current and emerging technological patterns. Yet, this method results in a loss of information and a reduction of insights compared to a qualitative method. Technologies that are mentioned infrequently will be discovered better by a qualitative research.

In this research, the patent titles have been analyzed. Although these provide plenty of information, they are not as informative as the patent abstracts. Nuances in different types of technologies are lost by only analyzing the titles. However, the smaller data size of titles compared to abstracts allows for the analysis of a much larger sample. Future research could imitate the method from this research, expanding on the text mining to include abstracts.

Caution must be used in interpreting the results for the automotive industry as a whole since only the top-5 companies were used for the analysis. These top-5 companies shape the industry to a large extent, thus supporting the validity of analyzing only these organizations. The analysis could be repeated on a larger scale (e.g. top-10 companies), to more strongly confirm current findings, or on a different segment (e.g. 20 companies with between 500 and 1000 patents each), to try to find trends on different industry-levels.

The trend analysis is conducted on three fixed time categories, rather than in a linear fashion. This might introduce a bias for patents introduced closely before or after the cutoff point from each of the time categories. The time categories simplify the analysis, increasing the degree to which the method can be repeated in the future the detect the emerging trends of that time.

The determination of market- and technical uncertainty is done through a range of questions. To answer the questions, the specific patent titles were researched and additional background research was done to the themes. However, an expert view on the themes is lacking in confirming the uncertainty scores given. When (automotive) companies use this method to detect emerging trends in the future, they will have the expertise to more accurately answer the questions from MacMillan & McGraths (2002) survey.

For the most recent time category (2018-present), there might be patent titles missing. The patent application process is time-costly (USPTO, 2022) For this research, only granted patents have been considered. Any pending patents are thus not included in the clustering analysis, which might results in an incomplete picture of new technologies that are forming. A similar analysis can be done on those pending patents specifically. This method should be able to detect emerging trends at an even earlier phase.

Future research can adapt the method from this research and apply it to another industry, this can confirm the text mining approach and find technological trends in other domains. The method can also be specified, discovering technological trends for smaller time periods. An approach that compares firms, instead of an industry-wide approach, can provide insights into specific companies, and make recommendations that are useful for those specific companies. A relationship can be investigated between the types of projects and the financial performance of the industry, to research whether the projects with lower uncertainty outperform projects with higher uncertainty (or vice versa).

If the automotive industry is researched further, the ideal composition of the R&D portfolio can be developed. This should be done by considering the rate of change and the nature of R&D within the industry. The resulting graph should illustrate percentages for each project type (e.g. 55% for platform launches). MacMillan and MacGrath (2002) suggest this approach in their research. This distribution of project types can then be compared with the actual R&D portfolio, found in this research. From this suggestions can be made for the automotive industry to further improve the balancing of R&D project types.

6. REFERENCES

Ante, L. (2023). How Elon Musk's Twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change*, *186*, 122112. (https://doi.org/10.1016/j.techfore.2022.122112)

Ates, A., & Acur, N. (2022). Making obsolescence obsolete: Execution of digital transformation in a high-tech manufacturing SME. *Journal of Business Research*, *152*, 336–348. (https://doi.org/10.1016/j.jbusres.2022.07.052)

Audretsch, D. B. (1995). Firm profitability, growth, and innovation. *Review of Industrial Organization*, *10*(5), 579–588. (https://doi.org/10.1007/bf01026883)

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review, 38(1),* 63-93. (https://doi.org/10.1016/j.bar.2005.09.001.)

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of the literature. *Research Policy*, *16*(2–4), 131–141. (https://doi.org/10.1016/0048-7333(87)90027-8)

Bureau van Dijk | Private company information – Orbis. (n.d.). Bvd. (https://www.bvdinfo.com/en-gb/)

Burghardt, E., Sewell, D. K., & Cavanaugh, J. E. (2022). Agglomerative and divisive hierarchical Bayesian clustering. *Computational Statistics & Data Analysis*, *176*, 107566. (https://doi.org/10.1016/j.csda.2022.107566)

Center for Automotive Research. (2014). Just How High-Tech is the Automotive Industry? In *Autos Innovate*. (https://www.autosinnovate.org/innovation/Just-How-High-Tech-is-the-Automotive-Industry.pdf)

Day, G. S., & Schoemaker, P. J. (2016). Adapting to fast-changing markets and technologies. *California Management Review*, *58*(4), 59-77. (https://journals.sagepub.com/doi/pdf/10.1525/cmr.2016.58.4.59)

Ettlie, J. E., Muammer, O., & Murthy, R. S. (2021). R&D Dynamic Capabilities in a Changing Regulatory Context. *IEEE Transactions on Engineering Management*, *1*, 98–111. (https://doi.org/10.1109/tem.2020.3045650)

Fattori, M., Pedrazzi, G., & Turra, R. (2003). Text mining applied to patent mapping: a practical business case. *World Patent Information*, *25*(4), 335–342. (https://doi.org/10.1016/s0172-2190(03)00113-3)

Henderson, R., & Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, *35*(1), 9. (https://doi.org/10.2307/2393549)

Irvine, J., & Martin, B. R. (1984). *Foresight in Science: Picking the Winners*. Pinter Pub Limited.

Jiang, C., & Zhou, Y. (2023). Mining semantic features in patent text for financial stress prediction. *Technological Forecasting and Social Change, 190(1),* (https://doi.org/10.1016/j.techfore.2023.122450)

Jung, J. U., Kim, H. J., & Choi, H. J. (2016). Patent Trend Mining for Internet of Things in Logistics. *Lecture Notes in Computer Science*, 624–634. (https://doi.org/10.1007/978-3-319-46257-8_67)

Koning, G., Schijndel-de Nooij, M., & Willems, F. (2020). HTSM Automotive Roadmap 2020 - 2030. In *Holland High Tech* (RAI_AINL_20-001). (https://hollandhightech.nl/_asset/_public/Innovatie/Technologieen/z_pdf_roadmaps/Roadmap-Automotive-v1-1_Signed.pdf)

Li, Y., Wang, L., & Hong, C. (2009). Extracting the significant-rare keywords for patent analysis. *Expert Systems With Applications*, *36*(3), 5200–5204. (https://doi.org/10.1016/j.eswa.2008.06.131)

Lin, B., Lee, Y., & Hung, S. (2006). R&D intensity and commercialization orientation effects on financial performance. *Journal of Business Research*, *59*(6), 679–685. (https://doi.org/10.1016/j.jbusres.2006.01.002)

MacMillan, I. C., & McGrath, R. G. (2002). Crafting R&D project portfolios. *Research-Technology Management*, *45*(5), 48-59. (https://doi.org/10.1080/08956308.2002.11671522.)

Matulka, R. (2015). *The History of the Electric Car*. Energy.gov. (https://www.energy.gov/articles/history-electric-car#:~:text=The%20first%20turning%20point%20many,mass%2Dproduced%20hybrid%20electric%20vehicle)

Millar, C. C. J. M., Groth, O. J., & Mahon, J. F. (2018). Management Innovation in a VUCA World: Challenges and Recommendations. *California Management Review*, *61*(1), 5–(14. https://doi.org/10.1177/0008125618805111)

Mikkola, J. H. (2001). Portfolio management of R&D projects: implications for innovation management. *Technovation*, *21*(7), 423–435. (https://doi.org/10.1016/s0166-4972(00)00062-6)

Mirkin, B. (2011). Choosing the number of clusters. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, *1*(3), 252–260. (https://doi.org/10.1002/widm.15)

Munková, D., Munk, M., & Vozár, M. M. (2013). Data Pre-processing Evaluation for Text Mining: Transaction/Sequence Model. *Procedia Computer Science*, *18*, 1198–1207. (https://doi.org/10.1016/j.procs.2013.05.286)

Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems With Applications*, *42*(9), 4348–4360. (https://doi.org/10.1016/j.eswa.2015.01.050)

Petit, Y. (2012). Project portfolios in dynamic environments: Organizing for uncertainty. *International Journal of Project Management*, *30*(5), 539–553. (https://doi.org/10.1016/j.ijproman.2011.11.007)

Pietrobelli, C., & Puppato, F. (2016). Technology foresight and industrial strategy. *Technological Forecasting and Social Change*, *110*, 117–125. (https://doi.org/10.1016/j.techfore.2015.10.021)

Piotroski, J. D. (2000). Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, *38*, 1. (https://doi.org/10.2307/2672906)

Rae, J. Bell and Binder, . Alan K. (2023, March 10). automotive industry. Encyclopedia Britannica. (https://www.britannica.com/technology/automotive-industry)

Rao, S., Ahmad, A., Horsman, W., & Kaptein-Russell, P. (2001). The Importance of Innovation for Productivity. *International Productivity Monitor*, *2*, 11–18. (https://www.researchgate.net/profile/Someshwar_Rao/publication/24051620_The_Importance_of_Innovation_for_Productivity/links/00b4951cb5a2933b28000000.pdf)

Singh, R. S., & Singh, S. (2021). Text Similarity Measures in News Articles by Vector Space Model Using NLP. *Journal of Institution of Engineers (India) Series B*, *102*(2), 329–338. (https://doi.org/10.1007/s40031-020-00501-5)

Solak, S., Clarke, J., Johnson, E. L., & Barnes, E. R. (2010). Optimization of R&D project portfolios under endogenous uncertainty. *European Journal of Operational Research*, *207*(1), 420–433. (https://doi.org/10.1016/j.ejor.2010.04.032)

Song, M., & Di Benedetto, C. A. (2008). Supplier's involvement and success of radical new product development in new ventures. *Journal of Operations Management*, *26*(1), 1–22. (https://doi.org/10.1016/j.jom.2007.06.001)

Szász, L., Csíki, O., & Rácz, B. (2021). Sustainability management in the global automotive industry: A theoretical model and survey study. *International Journal of Production Economics*, *235*, 108085. (https://doi.org/10.1016/j.ijpe.2021.108085)

Tewksbury, J. G., Crandall, M. S., & Crane, W. E. (1980). Measuring the Societal Benefits of Innovation. *Science*, *209*(4457), 658–662. (https://doi.org/10.1126/science.209.4457.658)

Troise, C., Corvello, V., Ghobadian, A., & O'Regan, N. (2022). How can SMEs successfully navigate VUCA environment: The role of agility in the digital transformation era. *Technological Forecasting and Social Change*, *174*, 121227. (https://doi.org/10.1016/j.techfore.2021.121227)

UN Department of Economic and Social Affairs. (2008). International Standard Industrial Classification of All Economic Activities. In *United Nations*. (https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf)

USPTO. (2022, July). *1120-Eighteen-Month Publication of Patent Applications*. (https://www.uspto.gov/web/offices/pac/mpep/s1120.html)

Van Doorn, J., Risselada, H., & Verhoef, P. C. (2021). Does sustainability sell? The impact of sustainability claims on the success of national brands' new product introductions. *Journal of Business Research*, *137*, 182–193. (https://doi.org/10.1016/j.jbusres.2021.08.032)

Yoon, B., Lee, S., & Lee, G. (2010). Development and application of a keyword-based knowledge map for effective R&D planning. *Scientometrics*, *85*(3), 803–820. (https://doi.org/10.1007/s11192-010-0294-5)

Yun, S., Cho, W., Kim, C., & Lee, S. (2022). Technological trend mining: identifying new technology opportunities using patent semantic analysis. *Information Processing and Management*, *59*(4), 102993. (https://doi.org/10.1016/j.ipm.2022.102993)

7. APPENDIX

7.1 Dendrograms

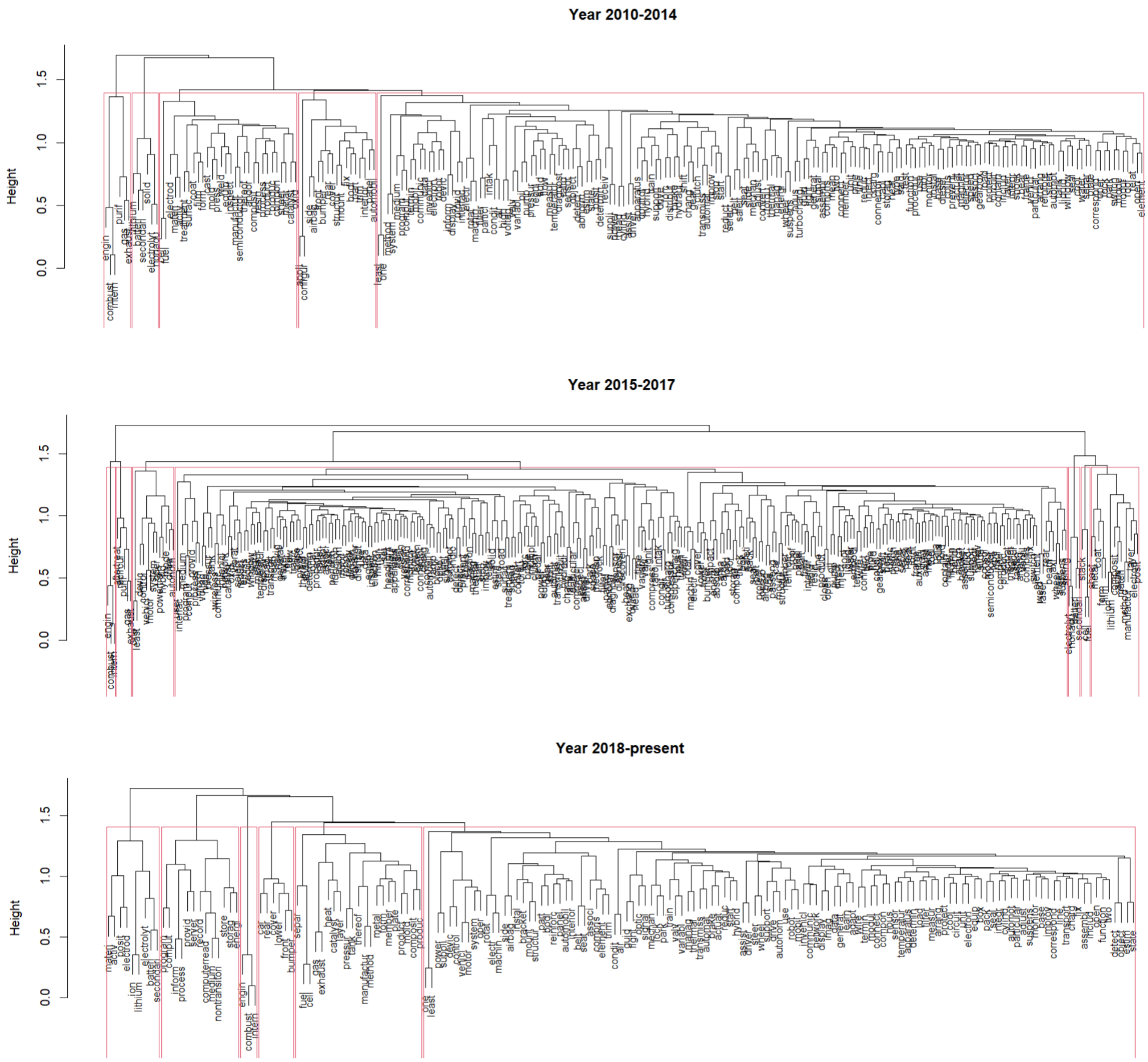

Figure 4: Dendrograms

7.2 MacMillan & McGrath's survey (2002)

MARKET UNCERTAINTY

How certain are you of the following? Score on scale of 1 (certain) to 7 (highly uncertain)	
M1	Market demand for future products using the fruits of the project
M2	Total future revenues from these products
M3	The stability of the revenue stream generated
M4	Extent to which you will be able to obtain needed support from distributors and suppliers
M5	Extent to which premium pricing can be expected
M6	Extent to which premium pricing can be sustained
M7	The speed with which products will be accepted in the market
M8	The speed with which products will be approved by necessary regulatory bodies
M9	Who the major competitors will be
M10	The probability that competitors will rapidly imitate us
M11	The probability of other technologies matching our offerings
M12	The probability of having our technology blocked by others
M13	Whether the technology has the potential to be licensed
M14	Degree to which we will have to constantly change designs
M15	The degree to which parallel technologies will be needed
M16	Whether such parallel technologies will be available in time
M17	Degree to which technical specifications will be required in the industry
M18	Degree to which technical specifications will be standardized in the industry
M19	The probability of profits being disrupted by third-party intervention (governments, distribution channels, labor unions, etc.)

Table 3: Questions to Determine Market Uncertainty. Retrieved from MacMillan & McGrath (2002).

TECHNICAL UNCERTAINTY

How certain are you of the following? Score on scale of 1 (certain) to 7 (highly uncertain)	
T1	The time it will take to complete development
T2	The type of skills needed for development
T3	The availability of necessary skills
T4	The cost of staffing those skills
T5	The type of equipment needed for development
T6	The availability of equipment needed
T7	The cost of equipment that is needed
T8	The systems needed for development
T9	The availability of systems needed
T10	The cost of systems needed
T11	The raw materials that will be needed
T12	The availability of needed raw materials
T13	The cost of raw materials
T14	Total costs of development
T15	The infrastructure that needs to be created
T16	Our ability to access necessary complementary technologies
T17	The cost of access to needed complementary technologies
T18	The technology barriers we will face
T19	Our ability to overcome technology barriers we will face
T20	The cost to overcome technology barriers
T21	The required level of product quality
T22	Required levels of support and service
T23	How much production capacity will be needed
T24	The commitment level of senior management

Table 4: Questions to Determine Technical Uncertainty. Retrieved from MacMillan & McGrath (2002).

7.3 Survey results

MARKET UNCERTAINTY																		
	C1-1	C1-2	C1-3	C1-4	C1-5	C2-1	C2-2	C2-3	C2-4	C2-5	C2-6	C2-7	C3-1	C3-2	C3-3	C3-4	C3-5	C3-6
M1	4	3	1	N/A	3	4	2	5	N/A	3	3	2	4	2	4	N/A	3	1
M2	4	3	2	N/A	3	4	2	5	N/A	3	3	2	4	2	4	N/A	3	2
M3	5	3	2	N/A	3	5	3	6	N/A	3	3	3	5	3	5	N/A	3	2
M4	5	3	2	N/A	2	5	3	4	N/A	2	4	3	5	3	4	N/A	2	2
M5	4	4	3	N/A	3	4	4	4	N/A	3	4	4	4	4	3	N/A	3	3
M6	4	4	3	N/A	4	4	4	5	N/A	4	4	4	4	4	6	N/A	4	3
M7	4	2	2	N/A	2	4	1	3	N/A	2	3	1	4	1	4	N/A	3	2
M8	5	3	4	N/A	3	5	1	4	N/A	3	4	1	5	1	4	N/A	3	4
M9	4	1	2	N/A	2	4	1	3	N/A	2	3	1	4	1	4	N/A	3	2
M10	4	4	1	N/A	1	4	3	3	N/A	1	3	3	4	2	4	N/A	2	1
M11	5	3	2	N/A	2	5	2	3	N/A	2	4	2	5	2	5	N/A	2	2
M12	6	3	2	N/A	2	6	3	3	N/A	2	4	3	6	3	5	N/A	2	2
M13	3	2	3	N/A	3	3	2	3	N/A	3	2	2	3	2	3	N/A	3	3
M14	4	2	2	N/A	2	4	2	4	N/A	2	5	2	4	2	5	N/A	3	2
M15	4	2	1	N/A	1	4	2	4	N/A	1	4	2	4	2	4	N/A	2	1
M16	5	3	2	N/A	2	5	3	5	N/A	2	4	3	5	3	5	N/A	3	2
M17	4	2	2	N/A	2	4	1	4	N/A	2	4	1	4	1	4	N/A	2	2
M18	5	2	2	N/A	2	5	2	5	N/A	2	3	2	5	2	4	N/A	2	2
M19	5	3	3	N/A	2	5	3	5	N/A	2	3	3	5	3	5	N/A	3	3
Avg.	**4,4**	**2,7**	**2,2**	**N/A**	**2,3**	**4,4**	**2,3**	**4,1**	**N/A**	**2,3**	**3,5**	**2,3**	**4,4**	**2,3**	**4,3**	**N/A**	**2,7**	**2,2**

Note that due to the high variety of technologies in the cluster labeled: ‘variety of technologies’, no accurate scores could be assigned to determine market uncertainty.

C1 refers to time category 2010-2014, C2 to 2015-2017 and C3 to 2018-present. The second number of the code relates to the cluster number, as specified in *table 2*. To illustrate: C1-1 is the first cluster of 2010-2014: battery-technologies.

TECHNICAL UNCERTAINTY																		
	C1-1	C1-2	C1-3	C1-4	C1-5	C2-1	C2-2	C2-3	C2-4	C2-5	C2-6	C2-7	C3-1	C3-2	C3-3	C3-4	C3-5	C3-6
T1	5	1	2	N/A	2	5	1	4	N/A	2	3	1	5	1	5	N/A	3	2
T2	4	2	1	N/A	1	4	1	4	N/A	1	3	1	4	1	4	N/A	2	1
T3	5	2	2	N/A	2	5	2	3	N/A	2	4	2	5	2	5	N/A	3	2
T4	4	2	2	N/A	2	4	2	3	N/A	2	3	2	4	2	5	N/A	2	2
T5	4	2	2	N/A	2	4	2	4	N/A	2	4	2	4	2	3	N/A	2	2
T6	4	3	2	N/A	2	4	2	5	N/A	2	3	2	4	2	4	N/A	2	2
T7	4	3	2	N/A	2	4	3	5	N/A	2	3	3	4	3	4	N/A	3	2
T8	4	2	3	N/A	3	4	2	4	N/A	3	3	2	4	2	4	N/A	3	3
T9	3	2	2	N/A	2	3	2	5	N/A	2	3	2	3	2	4	N/A	2	2
T10	4	3	2	N/A	2	4	3	4	N/A	2	3	3	4	3	4	N/A	2	2
T11	3	2	2	N/A	2	3	1	3	N/A	2	4	1	3	1	3	N/A	2	2
T12	6	2	2	N/A	4	6	2	3	N/A	4	4	2	6	2	5	N/A	4	2
T13	5	3	3	N/A	3	5	2	3	N/A	3	3	2	5	2	4	N/A	3	3
T14	5	3	3	N/A	3	5	3	4	N/A	3	4	3	5	3	5	N/A	3	3
T15	4	3	3	N/A	3	4	2	3	N/A	3	4	2	4	2	4	N/A	3	3
T16	4	3	2	N/A	2	4	2	4	N/A	2	4	2	4	2	4	N/A	2	2
T17	3	2	2	N/A	2	3	2	4	N/A	2	3	2	3	2	4	N/A	2	2
T18	5	3	3	N/A	3	5	3	4	N/A	3	5	3	5	3	5	N/A	3	3
T19	4	3	3	N/A	3	4	3	4	N/A	3	5	3	4	3	5	N/A	3	3
T20	5	3	2	N/A	3	5	3	4	N/A	3	4	3	5	3	5	N/A	3	2
T21	3	2	2	N/A	2	3	2	5	N/A	2	4	2	3	2	4	N/A	2	2
T22	3	2	1	N/A	2	3	2	4	N/A	2	3	2	3	2	4	N/A	3	1
T23	4	2	2	N/A	2	4	2	4	N/A	2	3	2	4	2	5	N/A	3	2
T24	2	3	2	N/A	2	2	3	3	N/A	2	3	3	2	3	4	N/A	2	2
Avg.	**4,0**	**2,4**	**2,2**	**N/A**	**2,3**	**4,0**	**2,2**	**3,9**	**N/A**	**2,3**	**3,5**	**2,2**	**4,0**	**2,2**	**4,3**	**N/A**	**2,6**	**2,2**

Note that due to the high variety of technologies in the cluster labeled: 'variety of technologies', no accurate scores could be assigned to determine technical uncertainty.

C1 refers to time category 2010-2014, C2 to 2015-2017 and C3 to 2018-present. The second number of the code relates to the cluster number, as specified in *table 2*. To illustrate: C1-1 is the first cluster of 2010-2014: battery-technologies.