University of Twente

Master Thesis Business Administration

Department of Behavioural, Management and Social Sciences

# Assessing Radicality Using Patent Content Text-Mining: A Tesla Case Study

*Author:*
J.B. Sigger

*Student Number:*
1461133

*Supervisors*:
dr. I. Skute
dr. M. De Visser

March 9, 2023

UNIVERSITEIT TWENTE.

# Contents

# 1 Introduction

Today's world is one that is often referred to as VUCA, an acronym that stands for Volatile, Uncertain, Complex and Ambiguous. It is a term used to describe the current state of the world, characterized by rapidly changing and unpredictable conditions, high levels of ambiguity, and complex systems and relationships (Millar, Groth, & Mahon, 2018). In this VUCA world, traditional approaches to decision making and problem solving may no longer be effective, and organizations and individuals must be able to adapt to new challenges. Lifecycles of businesses are becoming shorter and organizations are faced with high levels of technological and market uncertainty. Managers have to be quick in their decision making, while struggling with a lack of information and possible information asymmetry (Setyoko & Kurniasih, 2022). A recent and extreme example of VUCA is the emergence of the COVID-19 virus, and all the restrictions the pandemic brought with it (Peñarroya-Farell & Miralles, 2021). In addition, companies have to deal with increasingly more present sustainability issues. Reducing emission and the use of fossil fuels are topics that almost every company is confronted with nowadays. To handle all these uncertainties and challenges, it is important for a business to effectively manage their research and development (R&D) portfolios and to allocate the available resources towards innovations (Wheelwright & Clark, 1992).

One of the driving factors for creating competitive advantage in high-tech industries is a firm's innovativeness. Innovativeness refers to a company's capability to engage in innovation, or in other words, to introduce new processes, products, or ideas in the organization (Burns & Stalker, 1961). According to Hurley, Hult & Thomas (1998), a strong capacity to innovate leads to greater competitive advantage and firm performance. Distinguishing different types of innovation can be done in multiple ways. One way is by dividing innovative projects into incremental innovations and radical innovations. In Joseph Schumpeter's renowned book "The Theory of Economic Development", he discusses his theory of "creative destruction" (Schumpeter, 1934). According to this theory, the driving force behind economic growth and development is not incremental improvements, but rather the process of creative destruction, where new and innovative businesses displace old and outdated ones. This theory stresses the importance of the radicality of innovation, as it drives progress and creates new opportunities for growth and development.

One way of analysing a firm's innovativeness is by looking at its patents. Patents are often considered a good representation of a firm's innovative projects because they serve as public disclosures of the firm's R&D activities. By filing for a patent, a firm publicly announces its invention and provides details on how it works and how it is novel and useful. This information can then be used by researchers and competitors to gauge a firm's technological capabilities and future R&D direction (Lerner, 1994). Additionally, the granting of a patent provides the firm with legal protection for the invention, giving it a competitive advantage in the market (Hall & Ziedonis, 2001).

Patents can be analysed in multiple ways, and with multiple purposes. One purpose of analysing patents is to determine the radicality of a firm's innovative projects. Using patent data to analyse radicality is not uncommon. For example, in the research by Riita Katila (2000), a method is developed to determine the radicality of innovation, and in extension, firm performance. However, the research by Katila approaches patents from a citation analysis perspective, and measures innovation radicality based on citation-weighted patents. This is the case for the majority of studies that focus on assessing radicality using patent analysis, while there is a lack of studies that analyse radicality from a patent content perspective. One reason that most studies focus on citation analysis rather than content analysis is because companies can have portfolios of thousands of patents, and reading and assessing these patents in terms of radicality can take a tremendous amount of time. One way to reduce this time is by using text mining techniques to form clusters of patents, and to assess the radicality of these clusters instead of the individual patents.

The goal of this research is to construct a reproducible approach to assess the radicality of a company by analysing the content of a firm's patent portfolio. The ability to assess radicality can be important for a company to be able to compare their radicality with the radicality of competing firms and the industry as a whole. Furthermore, it can be important for investors to be able to assess radicality in order to determine which firm has the competitive advantage, which may be an indicator for future financial performances. The research question of this study reads: *"How can the radicality of a company be assessed using the content of its patent portfolio?"*

In order to answer the research question, one company is selected to perform a case study on, and to test the designed approach for assessing radicality. Because this study is focused on innovation, it is preferred that the selected company is active in a highly innovative environment, which is open to disruptive changes and receives a large amount of attention from the public. For this research, a company from the automotive industry, and more specifically the electric vehicle industry, is chosen to be subject of analysis. Due to the increasing level of interest in sustainable solutions to reduce the use of fossil fuels, electric vehicles have become more and more popular over the last decade. Where the count of electric vehicles sold worldwide did not exceed 120 000 in 2012, it has risen up to 6.6 million of electric cars per year in 2021 (International Energy Agency, 2022). While the electric vehicle industry is home to many of interesting companies, the largest and most well-known manufacturer of EVs is definitely Tesla. With a market capitalization of over 500 billion USD it is currently one of the largest companies in the world, and it dominates the EV market. In 2021, 70% of all electric vehicles sold worldwide were Tesla cars. For these reasons, Tesla will be the company to analyse in this research. Answering the first research question will be done by answering a second research question: *"How did Tesla's radicality develop over the last ten years?"*

To find an answer to this second research question, unsupervised text mining techniques will be used on Tesla's patent portfolio in order to structure it in such a way that it can be scored in terms of radicality. The techniques and frameworks used in this research on themselves are not novel, but the combination of these established methods is what makes this research unique. In the remainder of this document, firstly, an introduction to Tesla as a company will be given. Secondly, some insights from academic literature on the field of patents, innovation and text mining will be reported. In addition, a framework of portfolio management, which is the foundation of this research, will be introduced. Thereafter, the methods used in this study to achieve the desired results will be explained. Then, the results and their interpretations will be discussed comprehensively. To conclude, the acquired results will be discussed, and finally, recommendations for future research will be made.

## 2 An introduction to Tesla

Tesla is an American multinational corporation that specializes in electric vehicles, energy storage and solar panel manufacturing based in Austin, Texas. The company was founded in July 2003, and specializes in electric cars, lithium-ion battery energy storage, and residential photovoltaic panels. Tesla Motors is a public company that trades on the NASDAQ stock exchange under the symbol TSLA.

The company's first product, the Tesla Roadster, a sports car that could travel almost 400 km on a single electric charge, was released in 2008. Tesla's second vehicle, the Model S, an electric luxury sedan, debuted in 2012 and was followed by the Model X, a crossover SUV, in 2015. In 2016, Tesla released the Model 3, a lower-priced sedan that has become the best-selling electric car in the world, and in 2019 the Model Y, a mid-size SUV, was released. In addition to its electric vehicles, Tesla has also developed the Powerwall, Powerpack and Megapack, rechargeable lithium-ion battery stationary energy storage products for home, commercial, and utility-scale projects. The company has also installed solar panels and solar roof tiles on residential and commercial buildings.

Tesla has faced a number of controversies, including a 2015 "Consumer Reports" report stating that the Model S had the worst reliability of any car the magazine had ever tested, and a 2016 National Highway Traffic Safety Administration (NHTSA) investigation into the high number of fires that occurred in the Model S. In 2016, CEO Elon Musk was criticized for announcing a merger with SolarCity, a solar panel manufacturer in which he was the largest shareholder, and for his handling of a Model S driver's death in a collision while using the car's Autopilot feature. Despite these controversies, Tesla has achieved significant milestones, including becoming the first publicly traded American car company to have a market capitalization greater than $100 billion in 2019, and the first company ever to have a market capitalization of over $100 billion and no profits, according to Forbes.

Tesla's leadership is largely centered around CEO Elon Musk, who first invested in the company in 2004 and has served as CEO since 2008. Musk is known for his ambitious plans and bold personality, and has made headlines for a number of high-profile projects, including SpaceX's efforts to reduce space transportation costs and establish a human settlement on Mars, Tesla's work on electric and autonomous vehicles, and the Boring Company's plans to build a network of underground tunnels to reduce traffic congestion. However, Musk has also faced criticism and controversy. Some have questioned his inconsistent behavior and controversial tweets, and he has faced legal issues, including a defamation lawsuit and a Securities and Exchange Commission (SEC) investigation. In addition, some have criticized Musk's management style, claiming that he has a history of mistreating employees and overpromising on the company's projects.

In addition, Elon Musk has had a hectic relationship with cryptocurrency. In 2017, he tweeted about his support for bitcoin and revealed that he owned some of the cryptocurrency. However, he later distanced himself from the asset and said that he had sold all of his bitcoin holdings. In late 2020, Musk's company Tesla announced that it had purchased $1.5 billion worth of bitcoin and would begin accepting the cryptocurrency as a form of payment for its products. This announcement caused the price of bitcoin to surge. However, in May 2021, Tesla announced that it would no longer accept bitcoin as payment due to concerns about the environmental impact of bitcoin mining, which is energy-intensive. In addition to his involvement with bitcoin, Musk has also praised dogecoin, a cryptocurrency that was originally created as a joke. Musk has tweeted about dogecoin and has even referred to himself as the "CEO of Dogecoin." His tweets have had a significant impact on the price of dogecoin, and some have criticized him for manipulating the market.

Tesla has a somewhat unconventional approach to patents. In 2014, the company announced that it would not initiate patent lawsuits against anyone who used its technology "in good faith." This decision was part of Tesla's efforts to promote the development of electric vehicles and accelerate the shift away from fossil fuels. However, Tesla has also stated that it will take action to defend itself if it believes its patents are being used in bad faith, such as if a company is using its technology without permission or attempting to steal its trade secrets. In addition, the company has used patents as a way to negotiate partnerships and licensing agreements with other companies.

There are a number of reasons why Tesla is worth investigating from a research perspective. For one, the company is at the forefront of the electric vehicle market, which is rapidly expanding and is expected to play a major role in the future of transportation. Additionally, Tesla's innovative approach to energy storage and solar panel manufacturing make it a leader in the renewable energy industry. Finally, the company's unique corporate culture and leadership, particularly CEO Elon Musk, make it an interesting case study in business and leadership.

# 3 Patents & innovation

Patents are legal protections granted to inventors and creators for their unique and novel ideas or inventions. Patents give the holder the exclusive right to make, use, and sell their invention for a certain period of time, usually 20 years from the date of filing. Patents are important because they encourage innovation by giving inventors the opportunity to profit from their creations and invest in further research and development (R&D) (Alberts et al., 2017).

Patent analysis is the study of patents and their impact on innovation and industry. Researchers can use patent analysis to study a variety of topics, including the history of technology, market trends, and the competitive landscape of an industry. For example, an academic research study might use patent analysis to examine the development of renewable energy technologies over time or to identify the key players in a particular market.

There are several methods that can be used to analyze patents, including bibliometric analysis, citation analysis, and content analysis. Bibliometric analysis involves analyzing the characteristics of a patent, such as the number of citations it has received or the technology area it belongs to (Daim, Rueda, Martin, & Gerdsri, 2005). Citation analysis involves examining the relationships between patents, such as how often one patent cites another (Karki, 1997). Content analysis involves analyzing the text of a patent to identify trends or patterns (Tseng, Lin, & Lin, 2007).

Many different studies have been performed using patent analysis. For example, Ardito et al. (2018) use patent analysis to provide a comprehensive overview of the current trends regarding Industry 4.0 and examine which technologies play a role in enabling supply chain management-marketing (SCM-M) integration. Suominen, Toivanen and Seppänen (2017) have performed a research based on approximately 160 000 full text patents on the leading telecommunication providers between 2001 and 2014, to demonstrate the benefits and constraints of the use of machine learning in industry level patent analysis. Finally, Chen and Chang (2010) use patent analysis to examine the relationship between a firm's patent quality and its market value in the pharmaceutical industry in the United States.

There is a strong link between patents and innovation. Patents provide inventors with the incentive to invest in R&D and to bring new and innovative products to market. In turn, this drives technological progress and economic growth (Crosby, 2000). Patent analysis can be used to identify trends and patterns in innovation, such as which industries or companies are the most active in R&D.

One element to consider when analysing innovation is its radicality. Radical innovation refers to the development and introduction of new ideas, products, or processes that significantly change the way things are done or challenge existing norms, standards, or practices in a particular industry or field (McDermott & O'Connor, 2002). Radical innovations are typically disruptive, meaning they can fundamentally alter the way a market or industry operates, and often lead to the creation of new markets or industries. Garcia and Calantone (2002) define radical innovation as innovations that have the potential to cause both marketing and technological discontinuity on both a macro (industry- or market-wide) and micro (firm) level. Examples of radical innovation might include the introduction of the personal computer, which revolutionized the way people work and communicate, or the development of the internet, which transformed how people access and share information. More historical examples of radical innovation include the invention of the wheel, paper, or the steam engine.

Radical innovation can have a significant impact on the product architecture of certain products. In order for products to allow for radical changes, different modules of a product need to be connected using de-coupled interfaces. This way, individual modules can change as long as the interface remains the same, allowing for radical innovation at the component level (Ulrich, 1995; Pil & Cohen, 2006). However, when considering truly radical innovation, the impact on a product may exceed single modules and cause a greater impact on the product architecture. For example, when replacing the combustion engine of a car with an electric motor, the impact of the innovation is not limited to this single module. The gear box will no longer be needed and there will be no use for oil circuits and fuel tanks. Additionally, the battery storage and charging systems need to be accommodated. In other words, the entire architecture of the vehicle needs to be rethought.

Radical innovation is often contrasted with incremental innovation, which involves making small improvements or modifications to existing products, processes, or systems. While incremental innovation is important for improving efficiency and competitiveness, radical innovation can have a much larger impact and is often more difficult to achieve. To be truly radical, an innovation must be both novel and impactful (Chandy & Tellis, 2000). It must offer a solution to a problem or meet a need in a way that is significantly different from existing approaches. It must also have the potential to significantly change the market or industry in which it is introduced. From a Jordanian study performed by Al-Khatib and Al-Ghanem (2021), it showed that a high level of radical innovation has a positive relation with a firm's competitive advantage, especially when it operates in a high-tech industry.

There are several key factors that can contribute to the radicality of an innovation. These include the degree of change it represents, the magnitude of its impact, and the level of uncertainty or risk involved in its development and implementation. The research by Hurmelinna-Laukkanen et al. relate radicality, or radicalness, with a high level of market and technological uncertainty, new market creation and current product cannibalization (Hurmelinna-Laukkanen, Sainio, & Jauhiainen, 2008).

# 4 Portfolio management

In order to effectively allocate resources like time and money to their different innovative projects, it is crucial for organizations to focus on portfolio management. In their research "Crafting R&D Project Portfolios", Ian MacMillan and Rita McGrath (2002) discuss the process of creating a portfolio of research and development (R&D) projects, including how to identify and prioritize potential projects and how to manage the portfolio once it has been established. The authors argue that a carefully crafted R&D portfolio can help organizations achieve a balance between short-term and long-term goals, as well as between high-risk and low-risk projects. They also discuss the importance of aligning the R&D portfolio with the overall strategy of the organization.

In their research, MacMillan & McGrath (2002) discuss two types of uncertainties for R&D projects: Technical uncertainty and market uncertainty. Technical uncertainty can refer to a lack of knowledge surrounding the technical feasibility and viability of a project. In projects with a high technical uncertainty, it may not be clear whether the technology will be feasible to produce or whether it will perform as expected. This may be related to staffing problems, availability of equipment, availability and cost of raw materials, technological barriers and the ability to overcome them and production capacity.

Market uncertainty on the other hand refers to the uncertainties in the market, questioning whether the demand is sufficient and whether new products will be commercially viable. When addressing market uncertainty, one has to consider the potential revenues that can be generated from the project, whether distributors and suppliers can meet demands, the rate of acceptance in the market, how the competitors will respond to the new service or product, and whether the project will face regulatory problems.

To assess the technical uncertainty and market uncertainty, MacMillan and McGrath proposed a questionnaire with questions that can be scored on a scale from 1 (certain) to 7 (highly uncertain). These questions can be seen in tables 3 and 4 in appendix A. With the results of answering these questions a market uncertainty score and a technical uncertainty score can then be calculated by simply averaging the answers of the questions. A score between 1 and 3 resembles a low uncertainty, a score between 3 and 5 resembles medium uncertainty, and a score between 5 and 7 resembles high uncertainty. These scores can then be used to assess a project using figure 1, which basically places a project in one of 5 different categories: Positioning options, scouting options, stepping-stone options, platform launches and enhancement launches.

Positioning options have a high level of technical uncertainty because of a lack of insights towards feasibility or a lack of dominant design, while maintaining confident on the matter which markets or segments to address (MacMillan & McGrath, 2002).

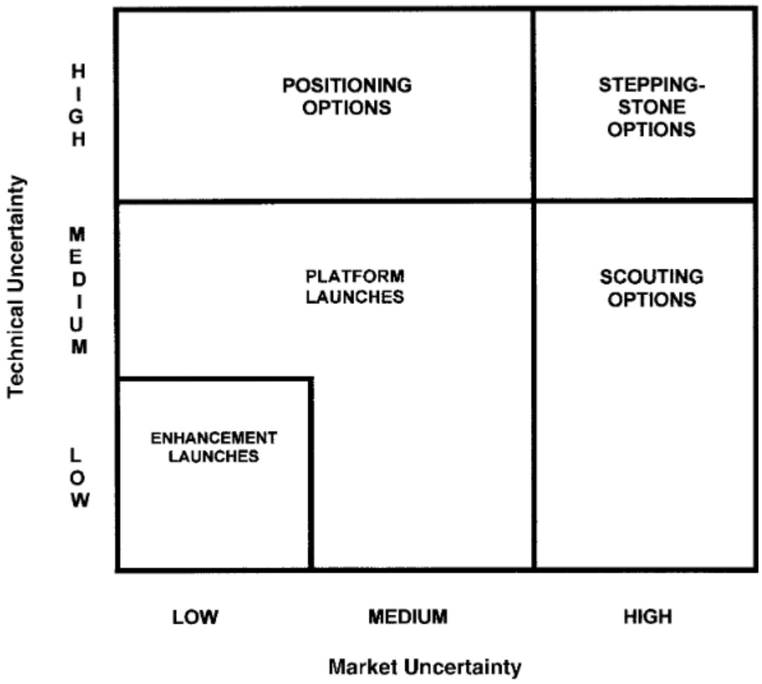



Figure 1: Technical uncertainty vs Market uncertainy. Obtained from MacMillan & McGrath (2002).

Scouting options on the other hand are options in which the owners are confident that the technology can be developed, but there is a lack of knowledge of the demands and wishes of the market. Finally, stepping-stone options consist of options that have both a high level of technical uncertainty and a high level of market uncertainty. These options usually focus on making small improvements in less challenging markets to gain insights and eventually build new technology in increasingly challenging markets.

Besides three types of options, the authors also mention two types of launches. Launches are innovative projects which such a low level of uncertainty that it is best to make an outright launch without further delay (MacMillan & McGrath, 2002). Launches can be categorized into two categories: Platform launches and enhancement launches. Enhancement launches consist of incremental improvements into existing products and services with a low level of market and technological uncertainty. Platform launches on the other hand are launches with a slightly higher level of uncertainty that can be used to create a basis for future business or competitiveness using next generation platforms (Wheelwright & Clark, 1992).

By successfully identifying different categories and allocating projects to these categories, companies may effectively manage their R&D portfolio, which can help to align the organization's resources and goals. A well crafted project portfolio ensures that the organization's R&D efforts align with its overall business strategy, and allows organizations to allocate their resources, such as funding, personnel, and time, efficiently. This may lead to an improved return on investment and helps organizations to gain a competitive advantage over competitors. Organizations can use several strategies for their portfolio management. One strategy can be to focus mainly on potentially disruptive projects, that have the potential to create new opportunities for growth. In contrary, a second strategy could be to focus on low-risk projects with a high potential for success, to avoid spending resources on a failing project. While this may be the safest option, it does not provide new opportunities for radical growth. A third strategy could be to establish a more balanced portfolio, with a mix of projects that have a high potential for success and projects that are more experimental or uncertain, but have a higher disruptive potential.

# 5 Text mining

In order to analyse large quantities of patent texts or other forms of textual data without spending enormous amounts of time, it is possible to use data mining, or more specifically, text data mining. Data mining is the process of automatically discovering patterns and knowledge from large data sets. It involves the use of algorithms and statistical models to extract useful information from data and transform it into an understandable structure for further use (Lew & Mauch, 2006). Text mining, also known as text data mining, is the process of extracting useful, interesting, and non-trivial information from text. It involves the use of natural language processing (NLP) techniques, machine learning algorithms, and other tools to extract insights from text data. The goal of text mining is to extract actionable information from text and use it to support decision making and other tasks. (Aggarwal & Zhai, 2012). The most important difference between data mining and text mining is that data mining is usually performed on structured data, like spreadsheets or databases, while text mining is usually performed on large, unstructured sets of textual data, like news articles, tweets, patents or customer reviews. Where data mining on structured data can usually be performed directly, text mining of unstructured data requires more data preparation before the actual analysis can be performed.

According to Žižka et al. (2020), text mining can be defined as a knowledge-intensive process in which a user uses a set of analytic tools on a collection of documents to identify and explore patterns that are potentially interesting. In their book "*Text Mining with Machine Learning: Principles and Techniques*", they list an extensive set of tasks that can be performed using text mining. These tasks include: The categorization of documents, clustering documents, summarising, information retrieval, extracting the meaning of documents or their parts, information extraction, association mining, trend analysis and machine translation. To analyse documents using text mining, multiple machine learning techniques can be used, both supervised and unsupervised. These techniques include, but are not limited to, Bayes classifiers, nearest neighbors, decision trees, random forests, support vector machines (SVM), deep learning using artificial neural networks (ANN), clustering, and topic modeling (Žižka, František, & Svoboda, 2020).

Many studies have already been performed using text mining to analyze large sets of textual data. The research by Kayser and Shala (2020) uses topic modeling on tweet hashtags to develop future scenarios for the topic of "quantified self". Suominen et al. (2017) also use topic modelling, more specifically, Latent Dirichlet Allocation (LDA) on full text patents to find technological diversification within companies. Li & Wu (2010) use sentiment analysis, k-means clustering and support vector machine to find a distribution of interest in a number of sports forums.

The process of text mining usually starts with tokenizing the corpus of documents that is to be analysed. Tokenization is the process of dividing a text into pieces, or tokens. These tokens can either be words (unigrams) or multiple words (n-grams) (Silge & Robinson, 2017). To analyse these tokens, the tokenized documents can then be transformed to either a tidy format or a document-term matrix. The tidy text format can be described as a table with one token per row, which results in a very long table. In a document-term matrix (DTM), each row represents one document, and each column represents one term. For a DTM of size $d \times n$ with $n$ documents and $d$ terms, every $(i, j)^{\text{th}}$ cell holds the (normalized) frequency of term $j$ in document $i$ (Aggarwal & Zhai, 2012).

## 6 Research design

In this research, the goal is to asssess Tesla's innovation radicality from patent data, and the development thereof over the last decade. To accomplish this, the first step is to acquire the patents to analyse. This is done by using Orbis, a database containing comprehensive information on over 400 million companies around the world. These patents are then imported into RStudio using R, a programming language for statistical computing and graphics. Unfortunately, it was not possible to retrieve full-text patents from Orbis, but merely patent titles. For the remainder of this research, these patent titles will be referred to as the patent texts, or documents.

After importing the patent texts using R, the texts can be pre-processed and tokenized into unigrams, and thereafter converted to a document-term matrix. To account for the rarity of a term in the corpus of documents, the cells in the DTM are converted from term frequency to term frequency-inverted documented frequency (TFIDF). From these TFIDF values, a distance matrix is composed using the cosine similarity measure.

The acquired distance matrix can then be used to perform the clustering method of choice. However, first the suitable number of clusters must be determined. This can be done in various ways, but in this research, the elbow method is used. Having found the appropriate number of clusters, a hierarchical agglomerative clustering method using Ward's minimum variance is performed to divide the corpus into clusters. These clusters can be visualised using a cluster plot, silhouette plot, or dendogram.

In order to give meaningful explanations to the obtained clusters, the clusters can be visualised using wordclouds. After assigning a topic to every cluster, the clusters are classified on a scale of their market and technical uncertainty, following the model by MacMillan & McGrath that was first introduced in section 4. Next, a radicality dimension is introduced in order to assign a radicality value to each cluster. Finally, the normalised yearly cluster frequencies are multiplied by this radicality value to find the evolution of Tesla's radicality over the past decade.

In the remainder of this section, the methods used will be discussed into more detail. For this research, it was chosen to divide the patent portfolio into clusters using hierarchical agglomerative clustering, which will be discussed in section 6.7. An alternative approach could be to classify each patent individually, and assess each class in terms of market and technical uncertainty. This would however require predetermined classes, making the method untransferable to patent portfolios from other industries or markets. For this reason, the unsupervised approach is preferred over a supervised approach. Yet another approach could be to use topic modelling to identify topics, an unsupervised learning method in which every document has a probability that it belongs to a topic. In turn, clusters could then be formed by grouping documents with a high probability for the same topic. In topic modelling, each document is considered to be a combination of topics, and every topic is considered to be a combination of words. This can work well for large documents, but for short texts, like patent titles, the amount of information for the topic model is too limited, which makes it difficult for the algorithm to identify meaningful topics and determine to which topic a document belongs. Through experimentation and trial and error it became apparent that in this particular case traditional clustering proved to be the preferred approach.

## 6.1 Orbis

Orbis is a comprehensive and widely-used database developed and maintained by Bureau Van Dijk. It contains a wealth of information on millions of companies from around the globe, and is used by businesses, financial institutions, and government organizations for a variety of purposes.

One of the key features of the database is its focus on intellectual property (IP), specifically on providing insights on a company's patents and trademarks, which can be incredibly valuable for companies looking to identify new business opportunities or mitigate risk. Orbis offers rich data on patent and trademark information, including filing dates, expiry dates, and classification codes, making it an invaluable tool for IP research and analysis. This information can be used to identify potential licensees or partners, to assess the strength of a company's IP portfolio, and to identify potential IP infringement risks.

In addition to data on intellectual property, Orbis also includes data on a company's financials, ownership structure, key executives, subsidiaries and affiliates, and industry classification. The database is updated regularly to ensure that the information it contains is accurate and up-to-date, making it an essential resource for professionals in various industries such as finance, legal, marketing and strategy. The wealth of data and insights available in Orbis make it an invaluable tool for businesses looking to stay competitive in today's global economy.

Through a license that is owned by the University of Twente, Tesla's patent portfolio was retrieved from Orbis' database. However, unfortunately this license only gives access to the patent titles, and not the full text patents. The patent titles, along with patent information like status, publication data and current owner, are downloaded as a *.xlsx* file.

## 6.2 R

R is a programming language and software environment for statistical computing and graphics. It is widely used for data analysis, machine learning, and other statistical applications. R was created in the early 1990s by Ross Ihaka and Robert Gentleman at the University of Auckland in New Zealand, and is now maintained by the R Project for Statistical Computing (Venables, Smith, & The R Development Core Team, 2007).

One of the main strengths of R is its wide range of built-in and user-contributed libraries, known as packages, which provide a variety of specialized functionality. For example, there are packages for data visualization, machine learning, natural language processing, bioinformatics, and much more. R also has a very active user community, which has contributed to the development of many packages and also offers a lot of resources such as tutorials and forums, providing support for learners and experienced users.

R is a free and open-source software, which means that anyone can use, modify, and distribute it. This has led to the development of many other tools and software that integrate with R, such as RStudio, a popular integrated development environment (IDE) for R. It is also easy to interface R with other software and systems, such as databases, web applications, and various data analysis tools, which makes it an attractive choice for data scientists, statisticians, and researchers in many fields. In this research, R is used to import, pre-process, process and analyse the patents that were acquired from Orbis.

## 6.3 Pre-processing

After importing the patent data into RStudio, the first few steps regard pre-processing the data. Pre-processing refers to the set of techniques used to prepare and clean raw data for analysis. Firstly, all patents that are not submitted between 2012 and 2021 are removed from the dataset, as the focus of this research is on this time period. Second, all patents that contain strange characters (for example, patent titles that are written in Arabic) are removed from the data. Patents with a title of 3 words or shorter are also removed, because these are difficult to assign to a cluster due to the limited availability of information. Somehow, some patents that appear in the dataset have a different company than Tesla as current owner. These titles are removed as well. Next, all patent titles are converted to lowercase and stemmed to their root form using the SnowballC package (Bouchet-Valat, 2020).

Finally, stop words are removed and the text is tokenized into unigrams using the tidytext package (Silge & Robinson, 2016). Stop words are words that are commonly used but are considered to be of little value in text analysis. These words are typically removed from text data before further processing, as they are deemed to add little to the overall meaning of a text. Examples of common stop words in English include "a," "an," "the," "and," "but," "or," etc.

Tokenizing is the process of dividing text into tokens. A token is a meaningful unit of text that can be used for analysis, and can either be a sub-word, a word, an n-gram, a sentence, a line or a paragraph. (Silge & Robinson, 2017). The "n" in n-gram represents the number of words that form a token. For text mining, it is common that a document is split in both single words, or unigrams (1-gram) as well as higher order n-grams. In this research however, the amount of text per document is limited due to the fact that only the patent title is accessible in contrast to the full text pattent. Therefore, when using higher order n-grams, the number of n-grams per document becomes very limited. Thus, for this study, the tokenization is restricted to unigrams only. After tokenization, which results in a tidy "one token per row" dataframe, the data can be casted into a document-term matrix, in which every row represents a document and every column represents a token.

## 6.4 TFIDF

In the DTM that is casted from the tokenized tidy format, every $(i, j)^{\text{th}}$ cell contains the term frequency of the $j^{\text{th}}$ term in the $i^{\text{th}}$ document. Naturally, this matrix is a sparse matrix. To account for the rarity of terms in the entire corpus of documents, the term frequency is converted to the term frequency-inverse document frequency. TFIDF is a statistical measure that is commonly used in text mining to evaluate the importance of a word in a document or a collection of documents (Jones, 1972; Salton & Buckley, 1988). It is calculated by multiplying the term frequency, which is the number of times a word appears in a document, by the inverse document frequency, which is a measure of how rare the word is across a collection of documents (Jones, 1972).

The term frequency-inverse document frequency measure is used in many natural language processing tasks, such as information retrieval (Jones, 1972), text classification (Manning, Raghavan, & Schütze, 2008), and topic modeling, as it can help to identify the most important and relevant words in a document or a collection of documents. It is particularly useful for text mining tasks where the goal is to extract important or relevant information from large amounts of text data.

One advantage of using TFIDF over the term frequency measure is that it takes into account the rarity of a word in a collection of documents (Jones, 1972). This is important because common words may appear frequently in a document, but they may not be as meaningful or relevant as rarer words that are specific to the topic of the document. By considering the rarity of a word, the TFIDF measure can more accurately identify the words that are most important or relevant to the topic of the document.

As stated above, the TFIDF is calculated by multiplying the (normalised) term frequency with the inverse document frequency. The normalised term frequency $TF_{i,j}$ is defined as the number of occurrences of term $t_i$ in document $d_j$ divided by the total number of terms in $d_j$. The document frequency $DF_i$ is defined as the number of documents that contain term $t_i$. The inverse document frequency, $IDF_i$, is the inverse of the document frequency $DF_i$ and is calculated by dividing the total number of documents $N$ by $DF_i$. However, when a very large corpus is considered, the value for $IDF_i$ will explode. Therefore, the *log* of $IDF_i$ is taken. The formula for TFIDF can be seen in equation 1.

$$TFIDF_{i,j} = TF_{i,j} \cdot log\left(\frac{N}{DF_i}\right) \tag{1}$$

## 6.5 Cosine distance

To use a clustering technique on a set of documents, it is important to find the distances between documents. The most straightforward method to determine the distance between documents is by use of the Euclidean distance. The Euclidean is defined as the square root of the sum of the squares of the differences between the coordinates of two points (D'Agostino & Dardanoni, 2009). A simple example of calculating Euclidean distance in a two dimensional space is the Pythagoras theorem, which calculates the straight-line distance between two points by taking the square root of the sum of the squares of the differences of its two dimensions. In the case of text mining, a $n$-dimensional space is considered in which $n$ is the total number of terms, or in other words the number of columns of the document-term matrix. The "points" in space, like in Pythagoras theorem, are the documents, i.e. the number of rows in the DTM. Summarizing, the formula for the distance between document $p$ and $q$ can be seen in equation 2, in which $p_i$ and $q_i$ are the TFIDF values on every dimension and $n$ is the total number of dimensions, or terms. The distance between every two documents can be calculated and stored in a distance matrix.

$$dist(p, q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{2}$$

Even though the Euclidean distance is the most common way of computing distance, it has its disadvantages when used in the context of text mining. The main disadvantage of the Euclidean distance is the fact that it is highly dependent on the length of a document. The Euclidean distance between a short title and a much longer title on the same topic can be large because of the length of the documents, which will cause them to be clustered into different categories. Therefore, in this research, the cosine distance will be used. The research by Janidis et al. shows that the cosine distance measure outperforms other text distance measures (Jannidis, Pielström, Schöch, & Vitt, 2015). Using the cosine distance instead of the Euclidean distance is faster, better suitable for dealing with sparse matrices, and independent of the length of documents (Godrey, Johns, Sadek, Meyer, & Race, 2014).

The cosine distance is defined as 1 minus the cosine similarity. This makes sense, because when the similarity is large, the distance must be small. In calculating the cosine similarity, every document $p$ is seen as a vector $\vec{p} = (p_1, p_2, ..., p_n)$, with $p_i$ being the TFIDF values for every term, as explained in section 6.4. The cosine similarity represents the cosine of the angle $\theta$ between two vectors $\vec{p}$ and $\vec{q}$. If $\vec{p}$ and $\vec{q}$ are similar, and thus move in the same direction, the angle between the two vectors is zero, and thus the cosine similarity is one. On the contrary, when $\vec{p}$ and $\vec{q}$ are perpendicular to each other, the angle between the two vectors is 90° and the cosine similarity is zero. The cosine similarity of two vectors is calculated by dividing the dot-product of the two vectors by the product of the magnitudes of the vectors (J. Han, Kamper, & Pei, 2012). The formula for calculating the cosine similarity can be seen in equation 3. Since the definition of the cosine distance equals one minus the cosine similarity, equation 3 translates to equation 4.

$$sim\left(\vec{p}, \vec{q}\right) = cos\left(\theta\right) = \frac{\vec{p} \cdot \vec{q}}{\|\vec{p}\| \cdot \|\vec{q}\|} = \frac{\sum\limits_{i=1}^{n} p_i q_i}{\sqrt{\sum\limits_{i=1}^{n} p_i^2}\sqrt{\sum\limits_{i=1}^{n} q_i^2}} \tag{3}$$

$$dist\left(\vec{p}, \vec{q}\right) = 1 - sim\left(\vec{p}, \vec{q}\right) = 1 - \frac{\sum\limits_{i=1}^{n} p_i q_i}{\sqrt{\sum\limits_{i=1}^{n} p_i^2}\sqrt{\sum\limits_{i=1}^{n} q_i^2}} \tag{4}$$

From equation 3 it follows that for the similarity to be anything other than zero, two texts have to have at least one similar word. If not, the sum of $p_i q_i$ will be zero and the distance will be one. For documents that do have at least one similar word, the distance can be calculated. As an example, the cosine distance between the following two patent titles will be calculated: "*Flux shield for electric motor*" and "*Electric motor waste heat mode to heat the battery*". For these two titles, the document-term matrix can be seen in table 1. It can be noted that all stopwords have been removed from the titles. Also, for this example, the term frequency instead of the TFIDF will be considered, as the TFIDF is dependent on the entire document corpus.

| | Flux | Shield | Electric | Motor | Waste | Heat | Mode | Battery |
|---|---|---|---|---|---|---|---|---|
| Title 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Title 2 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 1 |

Table 1: Document-term matrix of the example patent titles.

As the two patent titles have two similar words, it is already known that the cosine similarity will be unequal to 0. The numerator from equation 3 will be:

$$\sum_{i=1}^{n} p_i q_i = 1 \cdot 0 + 1 \cdot 0 + 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 + 0 \cdot 2 + 0 \cdot 1 + 0 \cdot 1 = 2$$

The magnitude of title 1 will be:

$$\sqrt{\sum_{i=1}^{n} p_i^2} = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2} = \sqrt{4} = 2$$

and the magnitude of title 2 will be

$$\sqrt{\sum_{i=1}^{n} q_i^2} = \sqrt{0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 1^2} = \sqrt{9} = 3$$

This means that the denominator in equation 3 becomes $2 \cdot 3 = 6$, and the cosine similarity will be $2/6 = \frac{1}{3}$. As the cosine distance equals one minus the cosine similarity, the cosine distance becomes $1 - \frac{1}{3} = \frac{2}{3}$.

The results of performing the calculations for the cosine distance of every combination of two documents can be stored in a distance matrix. This distance matrix is a square, symmetric matrix of size $d \times d$, in which $d$ is the total number of documents. The values on the diagonal of the distance matrix must be equal to zero, as the distance between a vector and itself is of course zero.

### 6.6 Elbow method

Once the distance matrix is acquired, the preferred clustering method can be applied. However, first the appropriate number of clusters must be determined. There are various methods to determine the number of clusters. The research by Mirkin (2011) proposes over 20 different methods to choosing the number of clusters. One method to determine the number of clusters that is fairly simple and does not take hours to compute is the elbow method. This may however not be the most exact of methods, but is more of a heuristic approach. Nevertheless, for this research, the result is sufficient to continue the analysis.

The elbow method is performed by calculating the total within-cluster sum of squares (WSS) for a range of number of clusters and plotting it as a function of the number of clusters. The WSS is a measure of the variability within a cluster. Clusters with a low value for WSS are generally more compact than clusters with a higher WSS. The total WSS is defined as the sum of the WSS values of the separate clusters and can be calculated using equation 5 (Marutho, Hendra Handaka, Wijaya, & Muljono, 2018).

$$WSS = \sum_{i=1}^{N_c} \sum_{x \in C_i} dist\,(x, \bar{x}_{C_i})^2 \tag{5}$$

In this equation, $N_c$ represents the total number of clusters, $x$ is a data point - or in this case - document, $C_i$ is the $i^{th}$ cluster and $\bar{x}_{C_i}$ is the centroid of the $i^{th}$ cluster. Once the WSS is calculated for a range of $N_c$, it can be plotted as a function of $N_c$. The resulting line will be shaped like an "arm", in which the ideal number of cluster will be at the point of the "elbow" (Humaira & Rasyidah, 2020). In other words, it is at the point where the steepness of the decline of the plot decreases drastically.

## 6.7 Hierarchical agglomerative clustering

Once the number of clusters is determined, a clustering method of choice can be applied. Numerous techniques to divide textual data into clusters exist. These can either be hard clustering methods (each document belongs to one cluster), or soft clustering (a document can belong to multiple clusters). An example of soft clustering is the Gaussian mixture model, in which the dataset is divided into normally distributed sub-populations. A different approach, which is not exactly soft clustering but does resemble it, is topic modeling. In topic modeling, each document is regarded as a combination of topics, and each topic is considered to be a combination of words (Silge & Robinson, 2017). Instead of dividing the corpus into clusters, topics are extracted by determining the probability of each word to belong to a topic.

Two examples of hard clustering methods are K-means clustering and hierarchical clustering. In K-means clustering, the dataset is randomly divided into a pre-set number of $k$ clusters, in which every cluster has its own centroid. The datapoints are assigned to the cluster with the least (squared) distance to its centroid. Then, an iterative process is initiated that optimizes the positions of the centroids, until it converges to a situation where the within cluster sum of squares does not decrease anymore (Wijnhoven, 2021). K-means clustering is a rather simple technique that requires a relatively low computational capacity. However, since it starts with a random choice of clusters, the algorithm may give a different result on each run. Results are therefore non-reproducible.

Hierarchical clustering refers to the process of grouping each document into clusters in a hierarchical fashion. There are two types of hierachical clustering: Agglomerative and divisive. In agglomerative hierarchical clustering, each data point (document) is considered to be its own cluster. Then, the algorithm finds the two clusters that are most similar, and merges them. This process is repeated until all clusters are merged, or the number of cluster reaches a pre-determined cut-off point. Divisive clustering is the opposite of agglomerative clustering, with all data points starting in one big cluster which gets smaller with every iteration by splitting off the data that is least similar to the rest, until every document ends up in its own cluster. In this research, the agglomerative method is preferred.

As stated above, in every iteration of the hierarchical agglomerative clustering method the two clusters that are most similar are merged. Multiple methods exist for determining this similarity between clusters. These methods include single linkage, complete linkage, centroid linkage, average linkage and Ward linkage. The research by Ulian and colleagues states that Ward's minimum variance method is the most effective metric to determine within cluster similarity (Ulian, Becker, Marcolin, & Scornavacca, 2021). This method attempts to minimize the sum of the squared distances between the points in seperate clusters, and is based on Joe Ward's research from 1963 (Ward, 1963).

In R, these computations can be performed using the "hclust" function from the "stats" package or the "hcut" function from the "factoextra" package (R Core Team, 2021; Kassambara & Mundt, 2020). The "factoextra" package also provides functions to visualize the results, for example in the form of a dendogram, silhouette plot or cluster plot. A dendogram is a tree diagram that shows the performed steps in the clustering process. A silhouette plot shows the silhouette width of each cluster, from which the quality of the clusters can be determined. In a cluster plot, all clusters and their data points are plotted as a function of the first two principal components. Furthermore, each cluster can be visualised using a word cloud, which is a visual representation of the word frequency in each cluster. Words that appear more frequently are give greater prominence than words that appear less frequently.

## 6.8 Cluster interpretation

After every document is assigned to a cluster, meaningful explanation can be given to each cluster. These interpretation are based on the corresponding word clouds and by manually checking the patent titles in a cluster. Some full text patents that correspond to these titles can be further evaluated using Google Patents, a search engine by Google that indexes patents and patent applications. Using this more in-depth information and the word clouds, a clear interpretation for each cluster can be established. This is done by looking for keywords, finding patterns and identifying the common themes or topics that are present within the cluster. This can help to understand what the cluster is about, what the documents in the cluster have in common, and what sets them apart from other clusters. Using the interpretation, the technical uncertainty and market uncertainty for each cluster can be determined, implementing the framework by MacMillan and McGrath that was introduced in section 4. Using the technical and market uncertainty, each cluster can then be plotted into figure 1.

As is already mentioned in section 3, Hurmelinna-Laukkanen and her colleagues state that radicalness can be regarded as a function of technical uncertainty and market uncertainty (Hurmelinna-Laukkanen et al., 2008). In this research, the studies by Hurmelinna-Laukkanen et al. and MacMillan and McGrath are combined to propose a categorization of radicality. MacMillan and McGrath propose 5 categories of innovation: Enhancement launches, platform launches and three types of options: Positioning options, scouting options and stepping-stone options. Enhancement launches have a low level of both market uncertainty and technical uncertainty. Platform launches have a medium level of uncertainty of at least one of the two variables, and options have a high level of uncertainty of at least one of the two variables.

In this research, a scoring mechanism for radicality is introduced. In this mechanism, the radicality for enhancement launches is scored "low". In other words, enhancement launches are of the incremental type of innovation. Projects that fall into the platform launches category get a radicality score of "medium", and projects in either one of the three options categories are scored "high". To further conceptualize this proposition, the radicality scores are ranked on a scale of 1 to 3, with low radicality receiving a score of 1 and high radicality receiving a score of 3.

Using the scoring mechanism for radicality that is explained above, each document in the dataset can receive a score of 1 to 3. In combination with the relative cluster appearance per year, a yearly radicality value can be determined. Furthermore, the evolution of this value over time can be analysed.

# 7 Results

In this section, the results that follow from applying the method from section 6 to Tesla's patent portfolio, which was downloaded from Orbis, are shown. The entire patent portfolio consists of 3797 patents. After applying the pre-processing steps that are described in section 6.3, 2017 patents remain. Following the process of tokenizing, stemming and removing stopwords, the patent titles are casted into a document-term matrix. This results in a sparse matrix with 2017 rows and 1255 columns. After converting the term frequencies in the cells of the DTM to TFIDF values, the cosine distances between each pair of documents can be calculated and stored in a distance matrix of size $2017 \times 2017$. Next, the within-cluster sum of squares can be calculated as a function of the number of clusters. The result of these calculations are plotted and can be seen in figure 2.

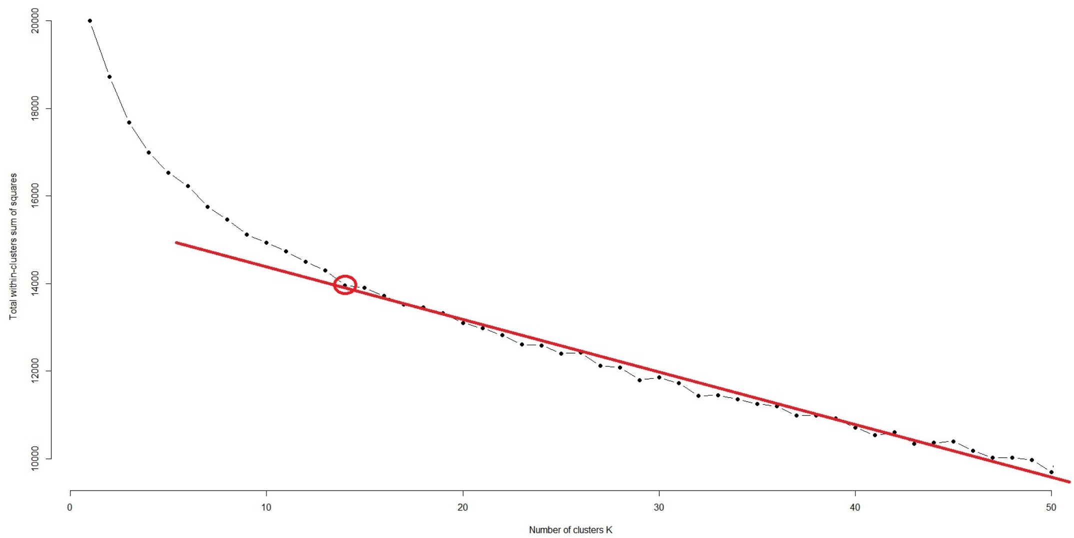



Figure 2: Within-cluster sum of squares as a function of the number of clusters.

As stated in section 6.6, the appropriate number of clusters is at the point of the elbow, when considering the line in the plot has the shape of an arm. From figure 2, the optimal number of clusters is chosen to be 14. Next, the hierarchical agglomerative clustering approach is implemented using the "hcut" function from the "factoextra" package. This results in the corpus being divided into 14 clusters of different sizes. These clusters can be visualised using tools from the "factoextra" package. In figure 3, the resulting dendogram can be seen. The 14 clusters are framed by the grey rectangles.

A second way to visualize the clusters is with the help of a silhouette plot, which shows the silhouette width of each document within a cluster, and the average silhouette of both the separate clusters and the entire corpus. The silhouette width value represents a measure of how similar a document is with respect to the cluster it is in. It ranges from -1 to 1, with a high value meaning that a document is correctly placed in its cluster, and a low value meaning that a document may be wrongly appointed to a cluster. The silhouette plot for this research can be seen in figure 4.

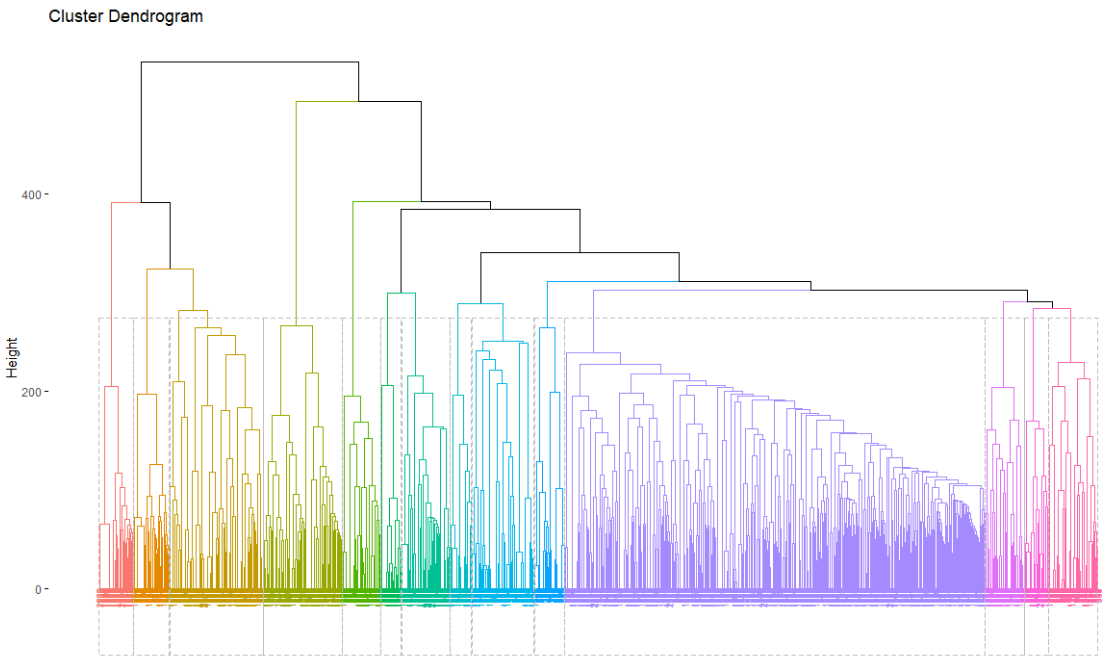



Figure 3: Dendogram of the resulting clusters.

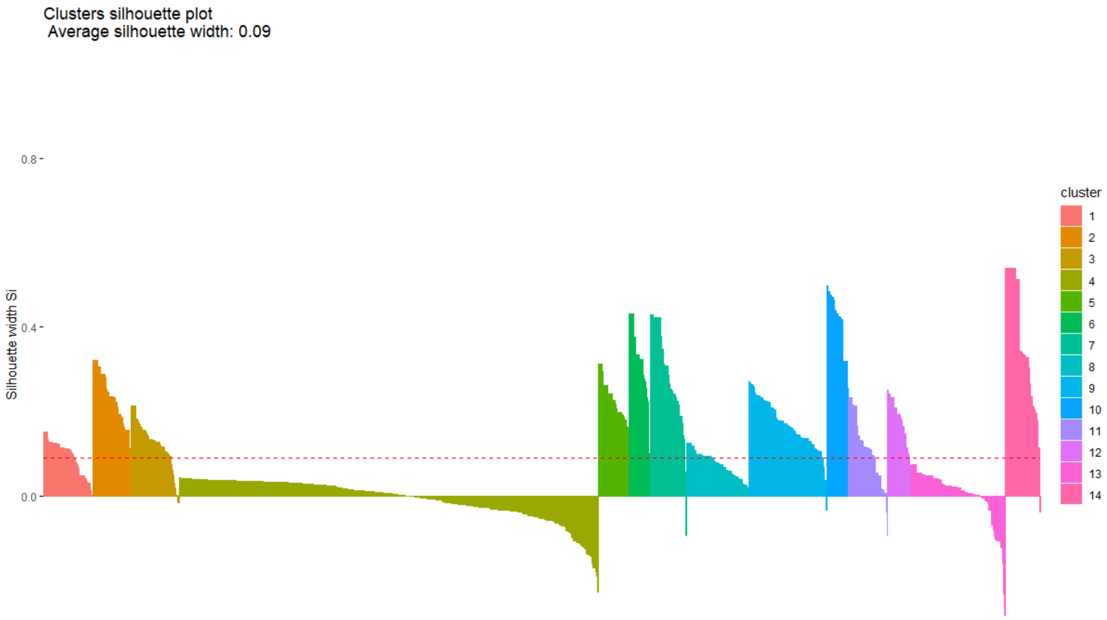



Figure 4: Silhouette plot.

From the silhouette plot, it can be seen that most clusters have a positive average silhouette value, meaning that the documents within the clusters are fairly similar. For cluster 4 however, despite it being very large, the silhouette values are awfully low.

Finally, the clusters can be visualised using a cluster plot. In a cluster plot, each cluster is plotted as a function of the first two principal components. As it does not add value towards the goal of this research, the meaning of the principal components are not further analysed. However, from figure 5, other interesting details can be seen. Despite the plot being quite messy, it can be seen that the large cluster 4 has a large level of overlap with almost all other clusters. This may explain the low silhouette width seen in figure 4.

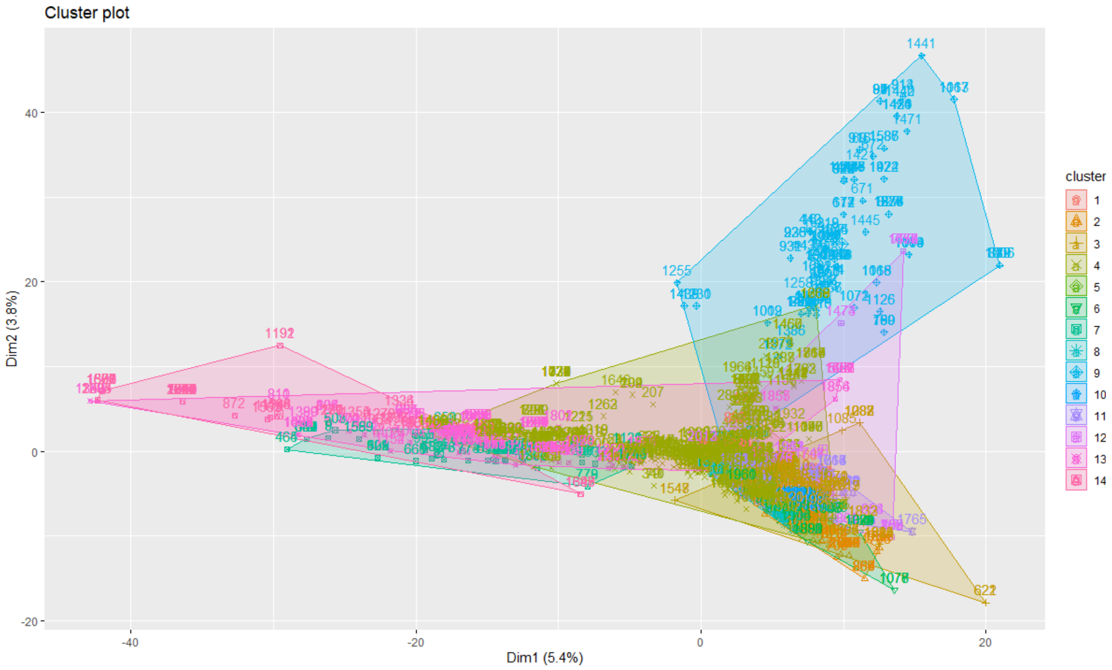



Figure 5: Cluster plot.

In the continuation of this section, the meaning of the individual clusters will be analysed. In this process, more in-depth patent information will be used, as well as word clouds and the information from figures 3, 4 and 5.

Upon analysis, it became apparent that cluster 4 is quite messy and chaotic. There is no coherent structure or overarching topic to be found from the word cloud or the patent titles itself. Many patents that were placed in cluster 4 would fit better in other clusters. This is confirmed by the silhouette plot and cluster plot in figure 4 and 5. Therefore, for the remainder of this research, cluster 4 and all the patents it contains will be disregarded from the dataset.

### 7.1 Cluster explanations

#### 7.1.1 Cluster 1: Chemical battery design

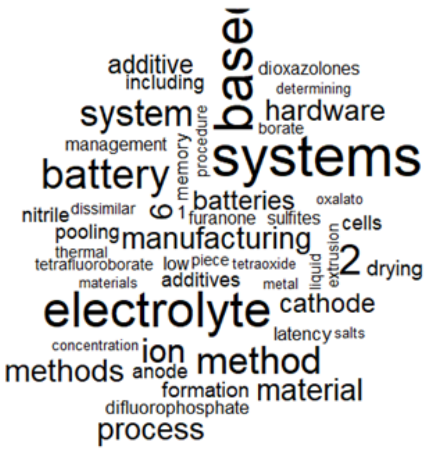



Figure 6: Word cloud of cluster 1.

**Example patents:**

*PROCESS FOR RECOVERING MATERIALS FROM SPENT RECHARGEABLE LITHIUM BATTERIES*

*NOVEL BATTERY SYSTEMS BASED ON TWO ADDITIVE ELECTROLYTE SYSTEMS INCLUDING 1 2 6 OXODITHIANE 2 2 6 6 TETRAOXEDE*

*Dioxazolone and nitrile sulfite as electrolyte additives for lithium ion batteries*

*Battery manufacturing using a liquid based thermal system*

Cluster 1 mostly holds patents that involve the manufacturing of the battery system, focusing on the chemical aspect. For example, it discusses the use of different anodes and cathodes, additives for lithium ion batteries, and different electrolytes. Moreover, most patents mention lithium ion batteries and ways to improve these. Therefore, cluster 1 will be given the title: *Chemical battery design.*

Lithium-ion batteries are a type of rechargeable battery that are commonly used in portable electronic devices, such as laptops, smartphones, and tablets. They are also used in electric vehicles and grid-level energy storage systems. Lithium-ion batteries have a high energy density, which means that they can store a relatively large amount of energy in a small space. They are also relatively lightweight and have a low self-discharge rate, which means they can hold a charge for a long time. Lithium-ion batteries work by moving lithium ions between the cathode and anode during charging and discharging. The cathode is typically made of lithium cobalt oxide or lithium manganese oxide, while the anode is made of graphite. When the battery is being charged, the lithium ions move from the cathode to the anode, and when the battery is being discharged, the ions move back to the cathode. The cathode and anode are separated by an electrolyte, usually a lithium salt in an organic solvent.

The majority of the patents in cluster 1 offer improvements for these lithium ion batteries. For example, additives for the electrolyte system, methods for producing cathode material, or methods for extracting lithium from clay minerals.

There are a few different types of lithium-ion batteries that are used in EVs, including lithium cobalt oxide, lithium manganese oxide, and lithium iron phosphate. Each of these types of lithium-ion batteries has its own set of characteristics, such as energy density, power density, and lifespan, which make them more or less suitable for different applications. For example, lithium cobalt oxide batteries have a high energy density but are relatively expensive, while lithium iron phosphate batteries have a lower energy density but are more stable and have a longer lifespan.

One of the challenges of using lithium-ion batteries in EVs is managing their temperature. Lithium-ion batteries can be sensitive to temperature and their performance can be affected by extreme temperatures. As a result, EV manufacturers often use various cooling and heating systems to help keep the batteries at an optimal temperature.

The demand for lithium-ion batteries is expected to continue to grow in the coming years as the market for electric vehicles and other applications that rely on lithium-ion batteries, such as grid-level energy storage systems, expands. As a result, there is a lot of research and development being conducted in an effort to improve the performance, cost, and sustainability of lithium-ion batteries.

One area of focus is increasing the energy density of lithium-ion batteries. Higher energy density batteries can store more energy in a smaller space, which is important for applications such as EVs where weight and size are critical factors. Researchers are experimenting with different cathode and anode materials, as well as new electrolyte formulations, in an effort to increase the energy density of lithium-ion batteries.

Another area of focus is reducing the cost of lithium-ion batteries. Currently, the high cost of lithium-ion batteries is one of the main barriers to wider adoption of EVs and other applications that rely on these batteries. Researchers are looking for ways to reduce the cost of lithium-ion batteries by using cheaper raw materials, improving manufacturing processes, and developing new designs that are more efficient.

Finally, there is also a lot of interest in making lithium-ion batteries more sustainable. This includes finding ways to reduce the environmental impact of mining and processing the raw materials used in these batteries, as well as developing recycling processes to minimize waste. Lithium-ion batteries are made from a variety of raw materials, including lithium, cobalt, nickel, manganese, and graphite. These materials are mined from the earth and then processed into the cathode and anode materials that are used in lithium-ion batteries. According to Weil, Ziemann and Peters (2018), the demand for these materials until 2050 is higher than the known resources. Therefore, improvements to these battery systems that may reduce the use of any of the less abundant raw materials are of major importance. In 2021, Tesla switched to the use of lithium-iron-phosphate (LPF) batteries instead of the previously used nickel-cobalt-aluminum oxide cells for all their standard ranged cars, because they are cheaper to produce and do not require any nickel or cobalt.[1]

Because of the increasing demand for lithium-ion batteries and the rapid rate of improvements, there is a low level of uncertainty with respect to future demand and revenue. Acceptance of improvements to these technologies will not be a problem and development should not be dependent on parallel technologies. With respect to the technical aspects, uncertainty is considerably higher. Because of the high development rate of EV batteries, time becomes essential and the costs of development rises. The availability of raw materials can become a problem, which may force the industry to use different anode and cathode materials, which in turn may raise additional technological barriers. In a 2022 study by Roland Berger, a management consulting company from Munich, four future risks for the lithium-ion battery are identified: Availability of raw materials, prices of raw materials, environmental impact and geopolitical factors, as 10%

[1] https://www.cnbc.com/2021/10/20/tesla-switching-to-lfp-batteries-in-all-standard-range-cars.html

of the world's nickel comes from Russia.[2] Therefore, the technical uncertainty of cluster 1 is considered to be high, and the market uncertainty is considered to be medium.

#### 7.1.2 Cluster 2: Electric motor and accessories

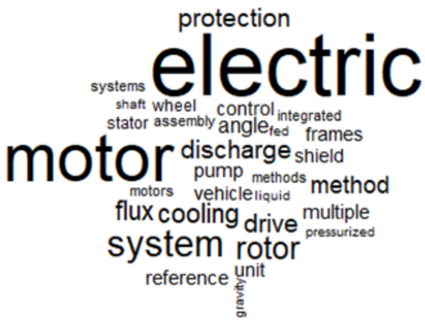



Figure 7: Word cloud of cluster 2.

**Example patents:**

*PRESSURIZED AND GRAVITY FED LIQUID COOLING OF ELECTRIC MOTOR*

*FLUX SHIELD FOR ELECTRIC MOTOR*

*Method of operating a dual motor drive and control system for an electric vehicle*

*Electric drive unit with gear shaft and rotor shaft*

Cluster 2 mostly holds patents that concern the electric motor within an electric vehicle, with systems and methods for cooling, discharge protection, flux shields and motor operation. Therefore, the topic of cluster 2 will simply be: *Electric motor and accessories.*

Electric motors are an essential component in EVs and are used to convert electrical energy into mechanical energy, which is used to power the vehicle. There are several different types of electric motors that can be used in EVs, including alternating current (AC) motors, permanent magnet synchronous motors (PMSMs), and switched reluctance motors (SRMs).

AC motors are the most common type of electric motor used in EVs, and they have the advantage of being able to be powered by a wide range of voltages. AC motors work by using alternating current to create a rotating magnetic field, which causes the rotor to follow the field and rotate. The Tesla Model S is an example of a vehicle that uses an AC induction motor.

PMSMs are similar to AC motors, but they use permanent magnets instead of electromagnets. They are often used in EVs because they are efficient and have a high power-to-weight ratio. PMSMs work by using the interaction between the magnetic field of the permanent magnets and the current in the wire coils to create motion. An example of a car with a PMSM is the Chevrolet Bolt, and the Tesla model 3. In an article from 2018, Tesla's principal motor designer Konstantinos Laskaris explains why Tesla switched from an AC indution motor for the model S to a permanent magnet motor for the model 3.[3]

A SRM also uses electromagnetic induction to rotate. It consists of a rotor made of ferromagnetic material and a stator made of wire coils. The rotation of the rotor is controlled by switching the current in the stator's wire coils on and off at specific times.

[2] https://www.mynewsdesk.com/rolandberger/pressreleases/rising-demand-for-lithium-ion-batteries-may-lead-to-shortages-in-raw-material-supply-3173780

[3] https://electrek.co/2018/02/27/tesla-model-3-motor-designer-permanent-magnet-motor/

Because of its potentially low cost, good power-speed characteristics and high reliability, the SRM is currently being researched by many research teams, including for example Toyota and Land Rover (S. Han, Diao, & Sun, 2021). However, because of problems with ripple and vibrations, no commercial vehicles have been launched yet.

The patents in cluster 2 regard multiple aspects of the electric motor. Examples of these are patents about the stator and rotor in an electric motor, cooling systems, discharge protection, flux shields and the use of dual motors. In terms of uncertainty, cluster 2 has a rather low market uncertainty, as most improvements to the electric motor can be simply implemented without major changes in demand and revenue or implications for suppliers and distributors. There are however some known (and unknown) technological barriers that must be overcome, like SRM problems with ripple and vibrations (S. Han et al., 2021). Therefore, patents from cluster 2 can be considered to be platform launches, as it has a fairly low level of market uncertainty and a medium level of technical uncertainty.

### 7.1.3 Cluster 3: Battery pack optimization

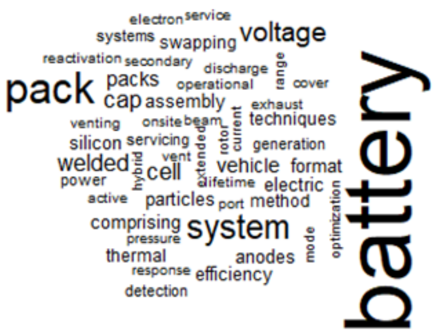



Figure 8: Word cloud of cluster 3.

**Example patents:**

*LARGE FORMAT BATTERY ANODES COMPRISING SILICON PARTICLES*

*HIGH EFFICIENCY HIGH VOLTAGE BATTERY PACK FOR ONSITE POWER GENERATION SYSTEMS*

*Electric Vehicle Extended Range Hybrid Battery Pack System*

*Electric Vehicle Battery Lifetime Optimization Operational Mode*

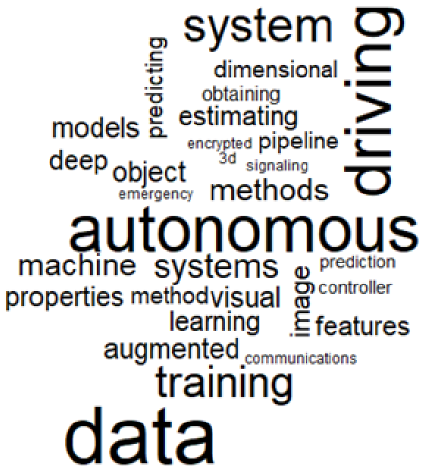
Like cluster 1, the main focus of cluster 3 is towards batteries. However, where cluster 1 mainly tends towards the chemical components within a battery, cluster 3 focuses more on battery performance, like efficiency, vehicle range and lifetime optimization, in both vehicle batteries and onsite energy storage systems. Therefore, for the remainder of this document, cluster 3 will be named: *Battery pack optimization*.

The patents in cluster 3 focus on the accessories and optimization of battery packs, both for electric vehicles and Tesla's other power storage systems. Tesla's large-scale energy storage systems, such as the Powerpack and Megapack, work by storing excess energy generated from renewable sources in a network of lithium-ion battery cells. When demand for energy is high, the stored energy can be released back into the grid to help meet that demand. The battery cells in these systems are arranged in a configuration that can be customized to meet the specific needs of a project. For example, a system designed to smooth out fluctuations in renewable energy generation may have a different configuration than one designed to provide backup power to a grid.

The systems are connected to the grid and can be controlled remotely, allowing utilities to manage the flow of energy and respond to changes in demand in real-time. They are also equipped with advanced monitoring and control systems that allow them to operate safely and efficiently. Tesla's large-scale energy storage systems are a key part of the company's efforts to accelerate the transition to renewable energy and help build a more sustainable future.

Cluster 3 regards aspects of both the utility scale storage systems and the battery packs in the electric vehicles. This includes many varieties, like welded battery caps, servicing methods for high voltage packs, battery pack dehumidifying systems, pressure monitoring in battery packs and exhaust nozzles. These battery system innovations are mainly designed to improve the efficiency, performance and energy density of battery packs. For example, in 2016 Tesla launched a new series of Powerpacks that had double the capacity of the Powerpacks that came before, without being increased in size.[4] Improvements like these help to make the use of lithium ion batteries more compelling, causing an increase in demand and revenue, but are not of a disruptive nature. The level of market uncertainty is considered low, while the level of technical uncertainty moves towards higher regions as the complexity of battery systems evolves. The patents in cluster 3 are therefore categorized as platform launches.

#### 7.1.4 Cluster 5: Artificial intelligence




**Example patents:**

*SYSTEMS AND METHODS FOR TRAINING MACHINE MODELS WITH AUGMENTED DATA*

*PREDICTING THREE DIMENSIONAL FEATURES FOR AUTONOMOUS DRIVING*

*Data pipelines and deep learning systems for autonomous driving*

*ESTIMATING OBJECT PROPERTIES USING VISUAL IMAGE DATA*

Figure 9: Word cloud of cluster 5.

Cluster 5 mostly holds patents that are linked to autonomous driving, deep learning, and data. It is made up of patents for technology that use machine learning, deep learning and other forms of artificial intelligence to make a digital twin of the outside world. Using such virtual models, the right decisions for autonomous driving and other tasks can be predicted. For the remainder of this research, cluster 5 will simply be referred to as: *Artificial Intelligence.*

Artificial intelligence (AI) is being used in the electric vehicle industry in a number of ways, including data analysis, deep learning, and autonomous driving. One example of the use of AI in the EV industry is in the analysis of data from vehicle sensors. This data can be used to improve vehicle performance, optimize the use of energy, and predict

[4]https://electrek.co/2016/10/07/tesla-powerpack-2-0-doubling-energy-capacity-new-battery-cell-2170-gigafactory/

maintenance needs. Deep learning algorithms can be used to analyze this data in order to identify patterns and make predictions.

Another area where AI is being used in the EV industry is in the development of autonomous driving systems. These systems use a combination of sensors, such as cameras, lidar, and radar, to gather data about the environment around the vehicle. Deep learning algorithms can be used to analyze this data and make decisions about how the vehicle should navigate the environment.

There are a number of companies that are using AI in the EV industry, including Tesla, Google, and Toyota. In Tesla's patent portfolio, cluster 5 holds many patents on this particular topic. These patents include matters like features for autonomous driving, estimating object properties using visual image data, training machine models with augmented data, data pipelines and deep learning. On their company website, Tesla has included an elaborate page about the different ways they use AI in their products.[5]

From a technical perspective, the integration of AI into EV technology may have a high level of technical uncertainty. There are a number of challenges that need to be overcome to make AI-powered EVs a reality. For example, the development of advanced sensors and data processing capabilities to enable the vehicle to "see" and "understand" its environment, as well as the integration of these systems into the overall vehicle design. There could also be challenges in terms of power consumption and data storage.[6] Additionally, the reliability and safety of AI-powered EVs would also be a concern and need to be carefully evaluated.

From a market perspective, the use of AI in EVs could be considered to have a medium to high level of market uncertainty. On one hand, the market for EVs is growing rapidly, and there is a growing interest in developing advanced technologies such as AI to improve the performance and functionality of these vehicles. On the other hand, the consumer acceptance of AI-powered EVs is still uncertain, and it is not clear whether consumers will be willing to pay a premium for these advanced features.[7] Additionally, it is not clear yet how much of the market will be ready to adopt the AI feature in their EV purchase. Furthermore, there are also regulatory and legal issues that need to be considered, such as ensuring the safety and security of AI-powered vehicles on the road.

Because of this high level of both technical and market uncertainty, the patents that concern the use of AI in the EV industry are classified as stepping-stone options. Despite the high potential, these patents need to be detailed thoroughly and carefully in order to avoid future liabilities.

[5] https://www.tesla.com/AI

[6] https://www.upgrad.com/blog/top-challenges-in-artificial-intelligence/

[7] https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

#### 7.1.5 Cluster 6: Vehicle door and accessories


**Example patents:**

*Power release system for front door of vehicle*

*Angle detection door handle assembly*

*Control System for Use with a Dual Hinged Vehicle Door*

*VEHICLE FRONT DOOR POWER OPENING SYSTEM*

Figure 10: Word cloud of cluster 6.

Cluster 6 is mainly about the doors of an electric vehicle, and its accessories such as door handles and hinges. The patents in this cluster describe techniques for automatic opening of the front doors, or the technologies for the sensors in the door handle. Cluster 6 will be described by the following title: *Vehicle door and accessories*.

Over the last 10 years, there have been several significant innovations in car doors. Automatic doors, for example, allow drivers to open and close their doors with the push of a button, making it easier for those with mobility issues or who are carrying heavy items. Sliding doors, found on minivans and some SUVs, are convenient to use in tight parking spaces and make it easier to load and unload passengers and cargo. Suicide doors, also known as "coach doors," are hinged at the rear and are often found on luxury vehicles for their aesthetic appeal. Butterfly doors open upwards like wings, and are commonly found on sports cars for their dramatic appearance. Gull-wing doors open upwards and outwards, giving them a shape similar to a seagull's wings and can be found on high-end sports cars. Finally, there are also hidden doors, which are concealed within the body of the car and give the appearance of a seamless exterior.

Besides these different types of door hinges and opening systems, there have also emerged more and more options for car door handles. One example is the touchless handle, which allows drivers to open their car doors by simply waving their hand in front of the handle. Electronic handles, which can be opened with a button or keyless entry system, are also becoming more common. Retractable handles, found on some sports cars, are designed to be flush with the body of the car when not in use, giving the car a sleek appearance. The use of these retractable handles is not without risk. A Washington Post article from 2019 tells the story about a man that burned alive in his car, because the retractable door handles malfunctioned and prevented bystanders from opening the doors and saving the driver. This resulted in a wrongful-death lawsuit for Tesla.[8]

For Tesla, most patents about vehicle doors filed in the last 10 years are about opening systems, different types of door handles, different types of hinges and the

[8]https://www.washingtonpost.com/business/2019/10/23/man-died-burning-tesla-because-its-futuristic-doors-wouldnt-open-lawsuit-alleges/

controlling of the door and the handle. Even though there has been plenty of innovation on this field, the impact and uncertainty on both a market and a technical level are relatively low. While different hinging systems may have an effect on the appeal of a vehicle, particularly the door handle innovations will neither be a dealbreaker nor a unique selling point for a car. Vehicle door patents will therefore be categorized in the enhancement launches region.

### 7.1.6 Cluster 7: Solar roof tiles

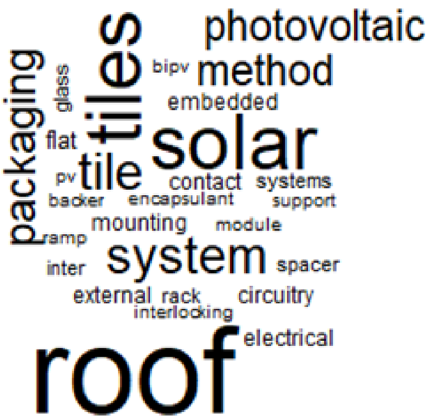



Figure 11: Word cloud of cluster 7.

**Example patents:**

*SOLAR ROOF TILE SPACER WITH EMBEDDED CIRCUITRY*

*Systems and methods for packaging photovoltaic roof tiles*

*PARALLEL CONNECTED SOLAR ROOF TILE MODULES*

*METHOD FOR IMPROVING ADHESION BETWEEN GLASS COVER AND ENCAPSULANT FOR SOLAR ROOF TILES*

Cluster 7 comprises of patents that hold methods and systems for solar roofs. While Tesla is most famous for building and selling electrical vehicles, they also specialize in selling solar roof tiles and solar panels. Thus, the patents that are included in cluster 7 are not necessarily used for the electrical vehicle industry. However, when consumers own both an electric car and a solar roof, the electricity that is produced with the solar roof can be used to charge the electric vehicle, reducing cost of transportation. Because cluster 7 focuses on solar roof tiles rather than panels or cells, it will hereby be named: *Solar roof tiles.*

Solar roof tiles, also known as solar shingles, are a type of photovoltaic (PV) panel that is designed to integrate with the roof of a building. They are installed in place of traditional roof tiles and generate electricity from sunlight in the same way as traditional PV panels. One of the main advantages of solar roof tiles is that they are less visible than traditional PV panels, which can be an important consideration for homeowners who are concerned about the appearance of their home. Solar roof tiles are also relatively easy to install, as they can be integrated into the roof during the construction process or added to an existing roof as a substitution.

Solar roof tiles can be just as efficient at generating electricity as traditional PV panels, but it's important to note that their performance can vary depending on a number of factors. The efficiency of solar panels is typically measured in terms of the percentage of sunlight that is converted into electricity, and most solar panels have an efficiency rate of around 15-20%. However, the efficiency of solar roof tiles may be slightly lower due to their smaller size and the fact that they are integrated into the roof, which can cause them to be partially shaded by other parts of the building.

Tesla launched their solar roof tiles in 2021, after purchasing SolarCity back in 2016. In an interview with Scientific America in 2017, Elon Musk predicted that Tesla's new solar shingles would be a succes.[9] However, the drawback of using solar roof tiles instead of traditional photovoltaic panels are the costs. Integrating solar roof tiles on an existing roof can cost up to 6 times the amount you would pay for regular solar panels, according to an article from CNET in 2022.[10] This is mainly because the entire roof has to be scrapped in order to replace the existing tiles with Tesla's solar roof shingles, while panels can be installed on top of existing roof tiles.

Because the technology for solar roof tiles is the same as it is for traditional solar panels, there is a low level uncertainty regarding the technical aspects of solar shingles. However, despite the growing interest in renewable energy, and solar power in particular, it is uncertain how solar roof tiles will compete with regular PV panels, and whether consumers are willing to pay the premium prices for a more aesthetic roof.[11] Therefore, the market uncertainty of cluster 7 is considered to be high. In consequence, patents regarding solar roof tiles will be considered to be scouting options.

### 7.1.7 Cluster 8: Vehicle body and interior

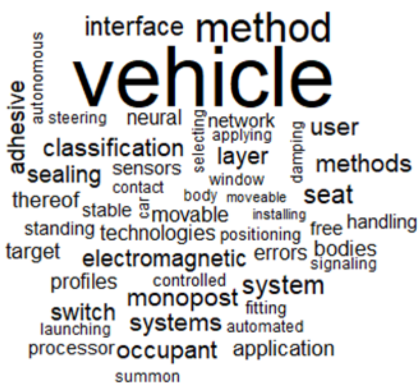



Figure 12: Word cloud of cluster 8.

**Example patents:**

*VEHICLE TECHNOLOGIES FOR AUTOMATED TURN SIGNALING*

*MONOPOST FOR FREE STANDING VEHICLE SEAT*

*Method of Selecting an Application Target Window in a User Interface*

*Method for installing sealing profile on car body or car body part using adhesive layer*

*AUTONOMOUS AND USER CONTROLLED VEHICLE SUMMON TO A TARGET*

Cluster 8 mainly holds patents that regard a vehicles interior or exterior, like the vehicle seats, turn signals, a sealing profile for the vehicle body, or an app on the user interface. However, besides these vehicle hardware related patents, some patents also hold technologies for summoning vehicles towards a location using autonomous driving. It can be argued that these patents would be better placed in cluster 5, which is composed of patents concerning artificial intelligence. However, the better part of cluster 8 contains patents regarding vehicle hardware, and thus cluster 8 will be approached as such: *Vehicle body and interior.*

For the majority, the patents in cluster 8 are focused on the interior and exterior of the vehicle. This includes topics like the vehicle user interface, steering wheel, processor,

[9] https://www.scientificamerican.com/article/will-tesla-rsquo-s-tiles-finally-give-solar-shingles-their-day-in-the-sun/

[10] https://www.cnet.com/home/energy-and-utilities/tesla-solar-roof-is-the-sleekest-solar-option-your-best-one/

[11] https://www.forbes.com/home-improvement/solar/solar-shingles-buying-guide/

vehicle seat, sensors, turn signals and sealing profiles for body parts. Over the last 10 years, many innovations have been made to the interior and exterior of a vehicle. For starters, many new vehicles now have touch screen displays that serve as the main interface for the car's infotainment system. These displays can range in size from small to very large, and they often include features such as navigation, music playback, and connectivity with smartphones. Tesla's efforts to improve their infotainment systems were not without success: In 2019, Consumer Reports announced that Tesla's infotainment system is better than any other auto brand's.[12] There has also been an increase in the use of natural materials such as wood, leather, and cloth in car interiors. These materials can give the interior a more premium feel and add to the overall comfort of the vehicle.

One trend in exterior design has been the use of more aerodynamic shapes to improve fuel efficiency. This can involve the use of features such as smooth underbody panels, active grille shutters, and rear spoilers. There has also been a trend towards the use of lighter weight materials such as aluminum and carbon fiber in vehicle construction. This can help to improve fuel efficiency and performance. Another trend has been the use of LED lighting for headlights, taillights, and other exterior lights. LED lights are more energy efficient and have a longer lifespan compared to traditional halogen bulbs.

Regarding technicality, improvements to the body and interior of an EV may have a low level of technical uncertainty. While the basic technology of EV body and interior design is well understood, there could be challenges in terms of integrating new materials or features, such as advanced aerodynamics or lightweight composites, that could improve the performance and efficiency of the vehicle (Henriksson, 2017). Additionally, there could be challenges in terms of ensuring that the new design is compatible with the existing EV platform, and in terms of cost and scalability of production. However, these challenges do not result in major uncertainties.

When looking at improvements to the body and interior of an EV from a market perspective, it may have a medium level of market uncertainty. It is not clear whether consumers will be willing to pay a premium for advanced body and interior design features, such as advanced aerodynamics, or if they will find these features important enough to influence their purchase decision. Additionally, the market for EV body and interior design improvements may also be influenced by government regulations, such as safety and emissions standards, which can vary by region and country.

[12] https://www.businessinsider.nl/tesla-infotainment-system-best-in-world-beats-bmw-consumer-reports-2019-5

#### 7.1.8 Cluster 9: Energy storage

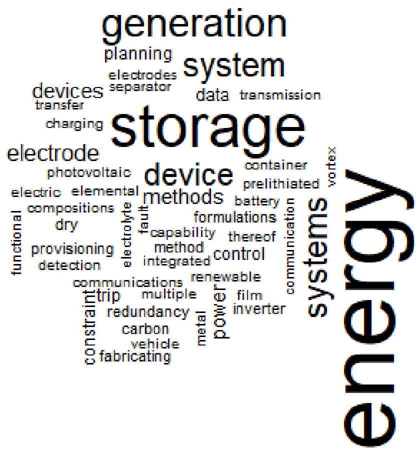



Figure 13: Word cloud of cluster 9.

**Example patents:**

*Arc fault detection for battery packs in energy generation systems*

*Multiple energy storage devices for inverter power control systems in an energy generation system*

*DRY ENERGY STORAGE DEVICE ELECTRODE AND METHODS OF MAKING THE SAME*

*ELECTRODE FOR ENERGY STORAGE DEVICE WITH MICROPOROUS AND MESOPOROUS ACTIVATED CARBON PARTICLES*

Cluster 9 clearly holds patents that regard energy storage, both for electric vehicles and for domestic and utility size storage systems. These systems can be used to store the energy that was generated by the solar roofs in domestic situations. For large scale commercial projects, Tesla also offer large energy storage units with a capacity of up to 3 MWh per unit. The patents included in cluster 9 hold technologies that are designed to improve these domestic and large scale storage systems. Moving on, cluster 9 will be known as: *Energy storage.*

Tesla has several energy storage options available for both residential and commercial use. For residential use, Tesla offers the Powerwall. The Powerwall is a battery that stores electricity for use when the grid is down or when renewable energy sources like solar panels are not producing electricity (e.g., at night). It can be used to power a home during an outage or to save excess solar energy for later use. The Powerwall has a capacity of 13.5 kWh, and can be stacked to increase capacity. In an article by an Ivy league business school from 2015, it was predicted that Tesla's powerwall could have a great impact on the way consumers use electricity.[13]

For commercial use, Tesla offers the Megapack. The Megapack is a large-scale battery system designed for use by businesses, utilities, and other organizations. It can be used to store excess energy generated by renewable energy sources, provide backup power during outages, and help smooth out fluctuations in energy demand. The Megapack has a capacity of up to 3 MWh, and can be scaled up to provide even more storage capacity as needed. Both the Powerwall and the Megapack are designed to be easily installed and require minimal maintenance. They are also equipped with advanced safety features to protect against fires and other hazards. Tesla's energy storage devices, are certainly innovative and have helped to popularize the use of battery-based energy storage systems. However, they are not the only options available in the market. There are many other companies that offer energy storage systems for both residential and commercial use.

[13] https://knowledge.wharton.upenn.edu/podcast/knowledge-at-wharton-podcast/how-teslas-powerwall-will-shift-control-to-the-consumer/

In general, energy storage systems work by charging batteries using electricity from the grid or from renewable energy sources such as solar panels, and then discharging the stored energy when it is needed. There are many different types of batteries that can be used in energy storage systems, including lead-acid, lithium-ion, and flow batteries, each with its own advantages and disadvantages. In addition to batteries, there are also other types of energy storage technologies that are available or under development. These include pumped hydro energy storage, thermal storage, and compressed air energy storage systems.

The adoption of domestic energy storage systems is still in the early stages, but it is becoming more common as the cost of renewable energy technologies continues to decline and more people look for ways to reduce their carbon footprint. In some countries, government incentives and subsidies are also helping to increase the adoption of home energy storage systems.

Despite constant improvements being made in the field of battery design, the technology used in energy storage systems is relatively mature. However, there are still uncertainties regarding the scalability of battery systems and the required infrastructure to integrate energy storage systems into the existing power grids, especially when demand continues to grow. The technical uncertainty of cluster 9 is therefore considered to be low to medium.

Concerning market uncertainty, the demand for battery energy storage systems is increasing as the cost of battery technology decreases and the need for energy storage solutions to support renewable energy integration into power grids increases. Additionally, the growth of electric vehicles and other applications that require energy storage are driving the growth of the battery energy storage market. A Bloomberg article from 2022 projected the energy storage market to grow 15-fold in just 10 years.[14] The market uncertainty for energy storage systems is thus considered to be moderate.

#### 7.1.9 Cluster 10: Overcharge detection in battery elements

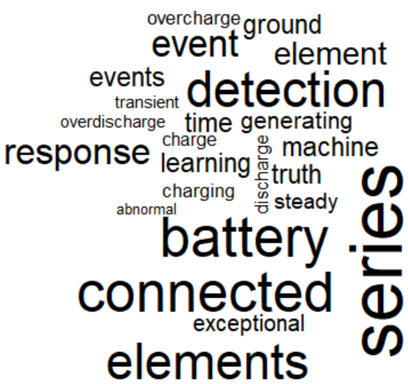



Figure 14: Word cloud of cluster 10.

**Example patents:**

*Response to over discharge event detection in series connected battery elements*

*TRANSIENT DETECTION OF AN EXCEPTIONAL CHARGE EVENT IN A SERIES CONNECTED BATTERY ELEMENT*

*Steady state detection of an exceptional charge event in a series connected battery element*

*Generating ground truth for machine learning from time series elements*

[14] https://about.bnef.com/blog/global-energy-storage-market-to-grow-15-fold-by-2030/

The patents in cluster 10 are mainly focused on detecting overcharging in a series connected battery element, both in transient and steady state. Furthermore, some patents include technology for generating ground truth for machine learning. These patents are probably included because of the use of the word “series”, and may be regarded as wrongly placed. However, because the majority of cluster 10 is focused on detecting exceptional charge events in batteries, cluster 10 will be described as following: *Overcharge detection in battery elements.*

Overcharging and over discharging can both be harmful to batteries that are connected in series. Overcharging occurs when a battery is charged beyond its maximum capacity, which can cause the battery to become damaged or even explode. Over discharging, on the other hand, occurs when a battery is drained of its charge to a level that is below its minimum recommended level, which can also cause damage to the battery.

To detect overcharging and over discharging in batteries, it is common to use a battery management system (BMS). A BMS is an electronic system that is designed to monitor and manage the charging and discharging of batteries. It can be used to prevent overcharging and over discharging by automatically shutting off the charging or discharging process when the battery reaches a certain threshold.

BMS systems can also have additional features, such as the ability to monitor the temperature of the batteries, the current being drawn from the battery, and the voltage of the individual cells within the battery. By monitoring these parameters, the BMS can provide an early warning if the battery is in danger of being overcharged or over discharged, allowing corrective action to be taken before any damage occurs. By preventing a vehicle from discharging or overcharging, the usable capacity of a vehicle battery decreases. This resulted in some discussion after the actual capacity numbers leaked in 2016.[15]

Steady state detection refers to the continuous monitoring of a battery or system to identify any abnormal charging events that may occur. In steady state detection, the system is continuously monitored, and any deviations from normal operating conditions are immediately detected and addressed. Transient detection, on the other hand, refers to the detection of short-term changes or disturbances in the battery or system. These disturbances, also known as transients, can be caused by a variety of factors, such as a sudden change in load or a sudden surge in current. Transient detection is often used to identify and address these types of events in order to protect the battery or system from damage.

Although discharge and overcharge detection are important topics for both the electric vehicle industry and the design of domestic and utility scale energy storage systems, both the technical and market uncertainty can be considered low. The technology is well developed and comprises multiple methods, and even though the market for battery systems is growing, there is little uncertainty regarding demand and revenue. Patents on the topic of discharge and overcharge detection are thus classified as enhancement launches.

[15] `https://electrek.co/2016/12/14/tesla-battery-capacity/`

#### 7.1.10 Cluster 11: Vehicle battery charging

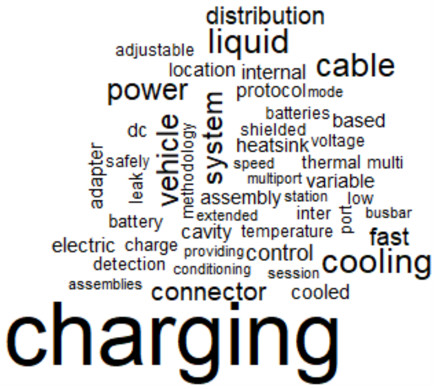



Figure 15: Word cloud of cluster 11.

**Example patents:**

*High power shielded busbars for electric vehicle charging and power distribution*

*LIQUID COOLED CHARGING CABLE AND CONNECTOR*

*MULTIPORT VEHICLE DC CHARGING SYSTEM WITH VARIABLE POWER DISTRIBUTION*

*Fast charging mode for extended travel*

Cluster 11 consists of patents that focus on the charging of vehicle batteries, including the charging cable and charging connector. This patent cluster will therefore be called: *Vehicle battery charging*. EV charging refers to the process of replenishing the battery of an electric vehicle with electricity. Charging an EV can be done at home, at work, or at public charging stations.

In the past 10 years, the charging infrastructure for EVs has significantly improved. In the early 2010s, public charging stations were scarce and often only available at specialized locations such as dealerships. However, as the popularity of EVs has increased, so has the number of public charging stations. Today, there are many options for EV owners to charge their vehicles while on the go, including at malls, supermarkets, and even parking garages.

There have also been improvements in the speed of EV charging. At the beginning of the previous decade, it could take upwards of 8 hours to fully charge an EV using a Level 2 charging station (which delivers a higher charging rate than Level 1 charging, which is available at most household outlets). Today, it is possible to charge an EV to 80% capacity in as little as 30 minutes using a DC fast charger, which is the highest charging rate currently available. Another development in EV charging is the emergence of wireless charging. Although this technology is not yet launched and applied to existing models, it could be the future of EV charging, as it allows EV owners to charge their vehicles simply by parking over a pad. This eliminates the need to plug in and unplug the charging cable, making the charging process more convenient.[16]

In addition to the physical infrastructure of EV charging, there have also been advancements in the payment process. 10 years ago, EV owners often had to pay for charging using a credit card or a specialized charging card. Today, many EV charging networks offer apps that allow users to pay for charging directly from their smartphones. This makes the charging process more streamlined and convenient.

A significant number of patents in Tesla's portfolio are on the topic of EV charging and collected in cluster 11. The patents in this cluster are about battery charging

[16] https://autovista24.autovistagroup.com/news/wireless-electric-vehicle-charging/

for electric vehicles, specifically the charging cable and connector for electric vehicles. Furthermore, it focuses on cooling systems for the charging cable, fast charge systems, leak detection and location based charging control systems. In 2015, Tesla released a liquid-cooling system for Supercharger cables that led to faster charging times while decreasing the thickness of the cable.[17]

While vehicle charging might take a disruptive turn in the future, the current technology is relatively mature and well known. Patents regarding battery charging are mainly focused on small improvements to this technology, with low levels of uncertainty. Regarding market uncertainty, the demand for battery charging is in line with the growing market of electric vehicles. Both governments and private companies are investing in the development of charging infrastructure and the standardization of charging protocols. The market uncertainty for cluster 11 is therefore also considered to be low.

#### 7.1.11 Cluster 12: Vehicle heat management

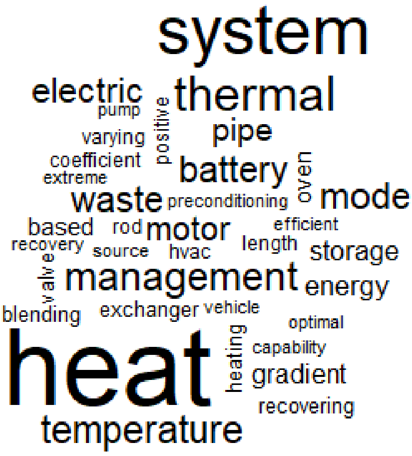



Figure 16: Word cloud of cluster 12.

**Example patents:**

*ELECTRIC MOTOR WASTE HEAT MODE TO HEAT BATTERY*

*Optimal source electric vehicle heat pump with extreme temperature heating capability and efficient thermal preconditioning*

*HEAT RECOVERING TEMPERATURE GRADIENT BASED OVEN SYSTEM*

*Thermal management system with heat exchanger blending valve*

The patents in cluster 12 focus on the heat regulation of the electric vehicle, regarding motor waste heat, efficient thermal preconditioning and heat exchangers. This cluster will therefore simple be known as: *Vehicle heat management.* Heat management is an important aspect of electric vehicle design, as it can affect the performance and range of the vehicle. There are several ways that heat is generated in an EV, including through the battery, inverter, and motors.

One key aspect of heat management in EVs is the cooling system, which is responsible for dissipating excess heat from the various components. This is typically done using a combination of air cooling and liquid cooling. Air cooling uses fans to blow air over the hot components, while liquid cooling circulates a coolant through channels in the components to remove the heat. A 2015 article by InsideEVs claims that there is room for improvement for Tesla regarding their liquid cooling systems, as the systems used by

[17] https://chargedevs.com/newswire/teslas-liquid-cooled-supercharger-cable-could-enable-faster-charge-times/

GM showed to be superior.[18]

Another important aspect of heat management in EVs, as previously mentioned in section 7.1.9, is the battery management system, which is responsible for monitoring and controlling the temperature of the battery cells. The BMS can use a variety of methods to regulate the temperature of the battery, such as heating or cooling the cells as needed, or by adjusting the charging and discharging rates.

Proper heat management is critical for the performance and life expectancy of an EV, as high temperatures can lead to reduced battery life and decreased performance. By carefully designing the cooling system and BMS, it is possible to maintain the temperature of the EV's components within an optimal range and ensure that the vehicle performs at its best.

Heat management systems for electric vehicles could be classified as having a medium level of technical uncertainty. On one hand, the technology for cooling systems and battery management systems in electric vehicles is well-established and has been used in other industries. However, electric vehicles present challenges regarding heat management due to the large amount of energy stored in the battery packs and the high power output of the electric motor, which can both generate significant heat.[19]

In terms of market uncertainty, improvements to the heat management system of electric vehicles do not create much additional market value or increase in demand, and offers little uncertainties. Combined with the medium technical uncertainty, cluster 12 is placed in the platform launches segment.

### 7.1.12 Cluster 13: Photovoltaic systems

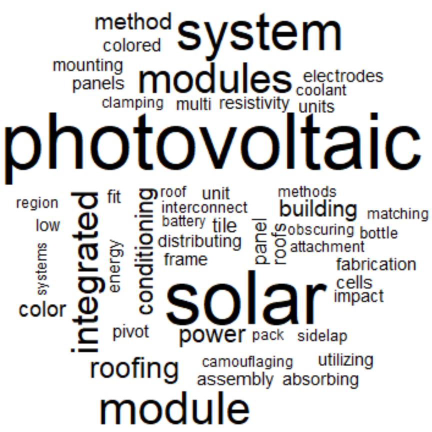



Figure 17: Word cloud of cluster 13.

**Example patents:**

*SYSTEM AND METHOD FOR IMPROVING COLOR APPEARANCE OF SOLAR ROOFS*

*Building integrated photovoltaic system with glass photovoltaic tiles*

*Building integrated photovoltaic roofing assemblies and associated systems and methods*

*Bifacial photovoltaic module using heterojunction solar cells*

Like cluster 7, the patents in cluster 13 mainly discuss improvements for solar roofs. However, where cluster 7 focuses purely on solar roof tiles, cluster 13 includes patents on all types of photovoltaic systems. This varies from the color of the panels

[18] https://insideevs.com/news/328909/tesla-or-gm-who-has-the-best-battery-thermal-management/

[19] https://www.e-motec.net/thermal-management-challenges-for-electric-vehicles

to technical specifications of the photovoltaic cells. Cluster 13 will therefore be called: *Photovoltaic systems*. A Harvard Business Review article from 2015 emphasizes the powerful combination Tesla possesses by uniting solar energy, energy storage and electric vehicles.[20]

In the past decade, there has been significant growth and development in the field of photovoltaics, or the conversion of sunlight into electricity. This technology has become increasingly cost-effective, making it an attractive option for electricity generation in both residential and commercial settings. One major development in photovoltaics over the past decade has been the improvement of solar cell efficiency. Solar cell efficiency refers to the percentage of solar energy that is converted into electricity. In the past, solar cells had low efficiency ratings, but advances in technology have led to the development of cells with much higher efficiency ratings. This means that a smaller number of cells can be used to generate the same amount of electricity, which can reduce the overall cost of a photovoltaic system.

Another development has been the decline in the cost of photovoltaic systems. The cost of solar panels has decreased significantly in recent years due to advances in manufacturing processes and an increase in demand for solar energy. As a result, solar energy has become more accessible to a wider range of people.

There have also been improvements in the durability and lifespan of photovoltaic systems. Modern photovoltaic panels are designed to withstand harsh weather conditions and have longer lifespans than older models. This means that they require less maintenance and can generate electricity for a longer period of time.

When considering technical uncertainty, the technology behind photovoltaic systems is well developed; it has been around for decades and it has been widely used in many applications such as residential, commercial and industrial settings. Photovoltaic systems are becoming more normalized and understood every day, and the engineering and manufacturing processes are becoming more streamlined.

In terms of market uncertainty, the market for photovoltaic systems is growing rapidly, driven by factors such as decreasing costs of solar panels, increasing concerns about climate change, and government incentives for renewable energy. However, the market for photovoltaic systems also depends on a variety of factors such as government policies, regulations, and the cost of competing energy sources, which can make it uncertain. Also, recycling solar panels is currently difficult, but this may improve in the future.[21] Patents from cluster 13 can be considered to have a low technical uncertainty, and a medium market uncertainty.

[20] `https://hbr.org/2015/07/tesla-is-betting-on-solar-not-just-batteries`

[21] `https://www.euronews.com/green/2023/03/06/recycling-dead-solar-panels-isnt-easy-these-australian-scientists-might-have-found-a-solut`

#### 7.1.13 Cluster 14: Mounting systems for solar roofs

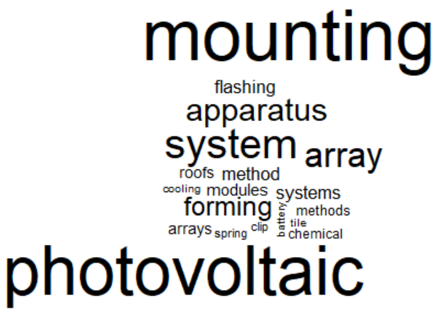



Figure 18: Word cloud of cluster 14.

**Example patents:**

*METHOD AND APPARATUS FOR FORMING AND MOUNTING A PHOTOVOLTAIC ARRAY*

*PHOTOVOLTAIC MOUNTING SYSTEM FOR TILE ROOFS*

*SYSTEMS FOR ATTACHING MOUNTING RAILS ONTO PHOTOVOLTAIC MOUNTING POSTS*

*Flashing System For Mounting Photovoltaic Arrays Onto Tile Roofs*

Cluster 14 has a high level of overlap with both cluster 7 and cluster 13, consisting of patents that mention solar roofs. However, where cluster 7 and 13 focus on the techniques behind solar and photovoltaic cells, the patents in cluster 14 focus on mounting methods for both solar tiles and panels. Therefore, this cluster will hold the following name: *Mounting systems for solar roofs*. From 2016 onward (which is the year Tesla purchased SolarCity) , the business section of Tesla's annual report form 10-K states that their business includes the design and engineering for mounting systems for photovoltaic panels (Tesla Inc., 2016).

Different methods and systems exist for the mounting of photovoltaic systems. First of all, a difference exists between systems that are mounted on sloped roofs and systems that are mounted on level roofs or on the ground. Furthermore, some systems are attached to roofs by penetrating the existing roof structure, distinguishing between different types of roofing like tiles or steel, while others are mounted with the use of concrete ballast and do not penetrate the existing roof. Mounting systems can either be as flush as possible with their mounting surface, or protrude. In the case of the latter, some systems offer the possibility to rotate and follow the sun, so a higher efficiency can be achieved.

In the case of Tesla, most patents regarding the topic of mounting photovoltaic systems and arrays consider small innovations, like changes to railing or clamps that are used in mounting the systems. These patents are relatively straightforward in terms of technology and do not offer high levels of uncertainty. Future demand and revenue from PV mounting systems are well-known too, causing patents from cluster 14 to be categorized as enhancement launches.

### 7.2 Radicality

Once all clusters are explained and scored in terms of market and technical uncertainty, the clusters can be placed in MacMillan and McGrath's uncertainty plot as seen in figure 1 in section 3. The filled uncertainty plot can be seen in figure 19. Tables 5 and 6 in appendix B show how the uncertainty scores used in figure 19 were established, using the questionnaires in tables 3 and 4 from appendix A.

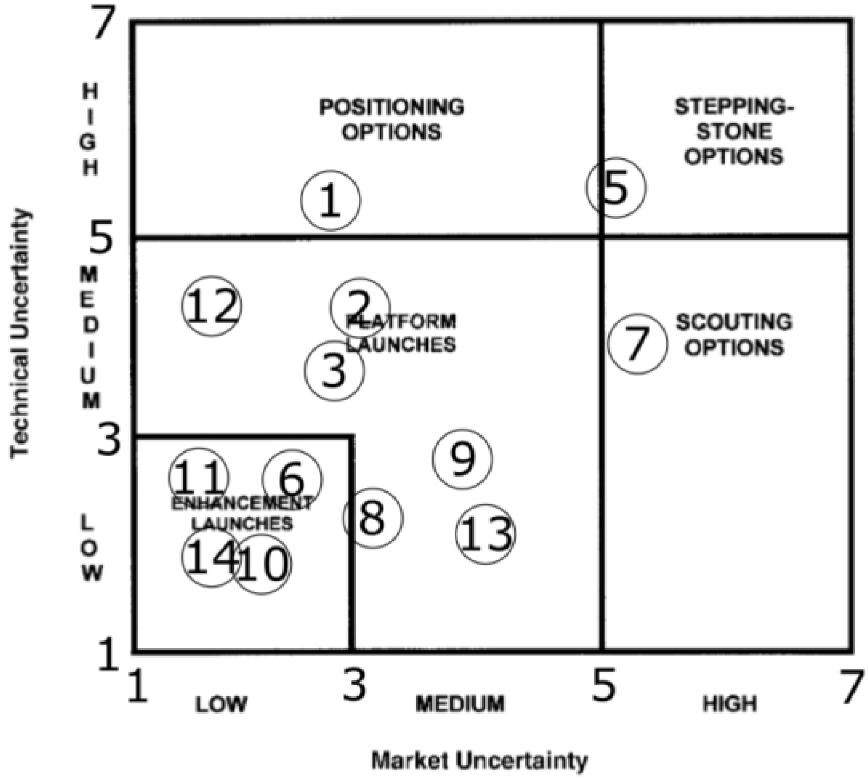



Figure 19: Uncertainty scores of the clusters.

Section 6.8 explains how every cluster in figure 19 can be scored in terms of radicality, and how these scores can be combined with the yearly relative cluster appearances to determine the evolution of a company's radicality over time. The development of the relative cluster appearance per year can be seen in figure 20 and the radicality score for each cluster can be seen in table 2.

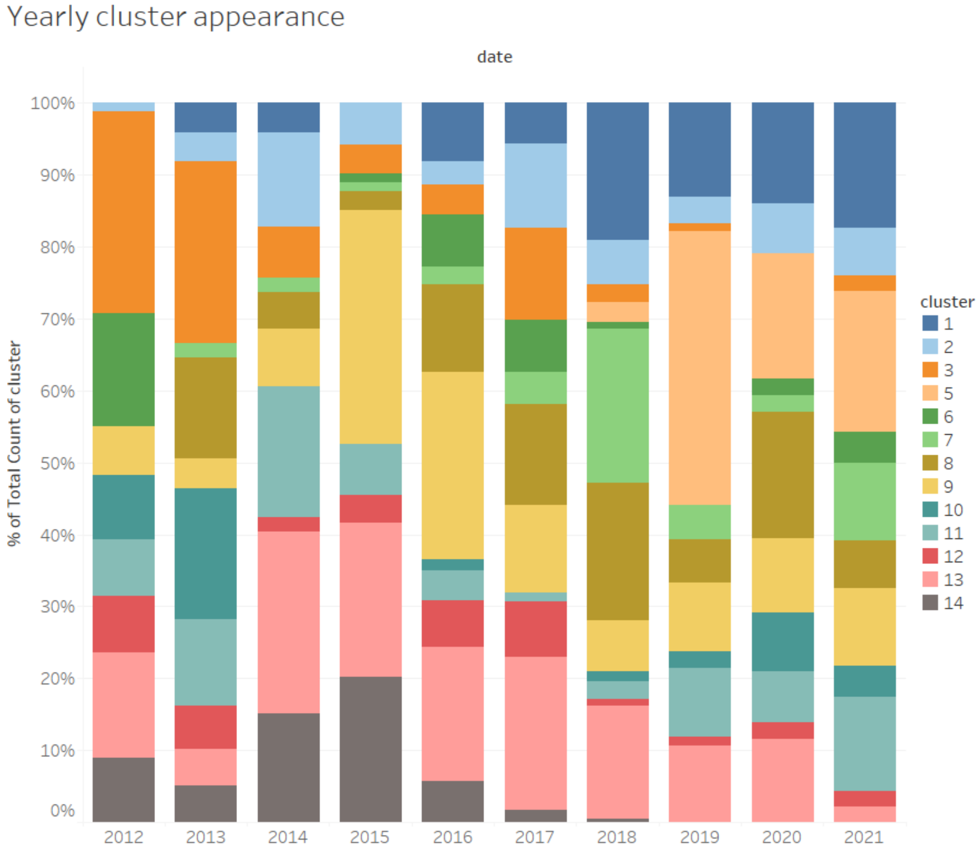



Figure 20: Relative cluster appearance per year.

| Cluster | Topic | Radicality |
|---|---|---|
| 1 | Chemical battery design | 3 |
| 2 | Electric motor and accessories | 2 |
| 3 | Battery pack optimization | 2 |
| 5 | Artificial intelligence | 3 |
| 6 | Vehicle door and accessories | 1 |
| 7 | Solar roof tiles | 3 |
| 8 | Vehicle body and interior | 2 |
| 9 | Energy storage | 2 |
| 10 | Overcharge detection in battery elements | 1 |
| 11 | Vehicle battery charging | 1 |
| 12 | Vehicle heat management | 2 |
| 13 | Photovoltaic systems | 2 |
| 14 | Mounting systems for solar roofs | 1 |

Table 2: Radicality scores of each cluster.

By combining the results from figure 20 and table 2, a radicality score can be calculated. This is done using equation 6, in which $Y_j$ is the radicality score of year $j$, $N$ is the number of clusters, $A_{ij}$ represents the relative appearance of cluster $i$ in year $j$ and $R_i$ is the radicality score of cluster $i$. Each yearly radicality score has a minimum of 1 and a maximum of 3. The results are plotted and can be seen in figure 21.

$$Y_j = \sum_{i=1}^{N} A_{ij} \cdot R_i \tag{6}$$

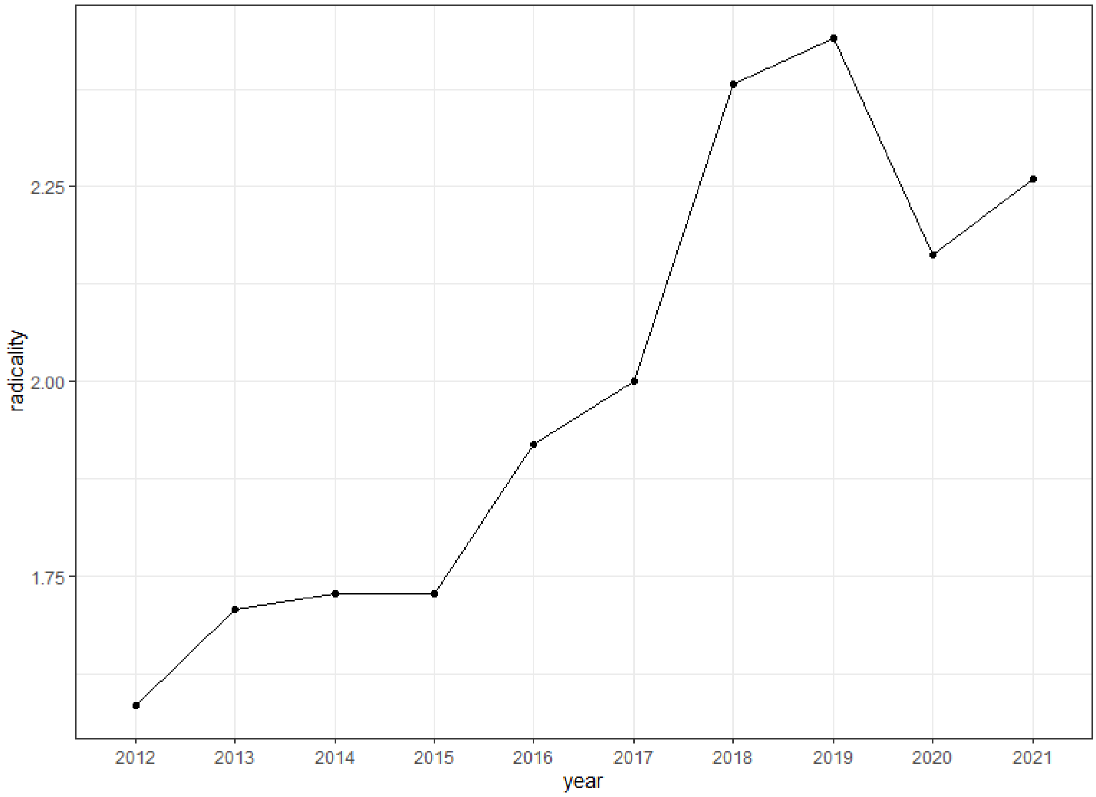



Figure 21: Yearly radicality score.

## 8 Discussion

The results shown in the previous section show some interesting outcomes. Figure 19 shows the distribution of Tesla's patent clusters as a function of their technical and market uncertainty. It can be seen that the majority of Tesla's projects is scaled in the "launches" sections, while only three clusters were scored as an "option". This is somewhat surprising, as Tesla is known for their radical innovation and disruptive projects. By introducing the Roadster in 2008, Tesla revolutionized the way people thought about electric cars, but after launching this first extremely radical product, the focus shifted to improving it in a more incremental fashion, resulting in a balanced project portfolio. The research by Cooper, Edgett and Kleinschmidt (1998) on strategic technology management concludes that a well-balanced R&D portfolio is more often present at companies that outperform their competition. Thus, among other things, a well-balanced R&D portfolio may lead to better financial performance.

Figure 20 shows the yearly relative cluster appearance. Striking from this figure is that the clusters that focus on solar energy, cluster 7, 13 and 14, where present is the portfolio even before Tesla purchased SolarCity in 2016. The reason for this is that the corresponding patents were first filed by SolarCity, and transferred to Tesla after the acquisition. It can be seen that the share of patents that focus on solar panels and the mounting of panels was prominent in the first part of last decade, but has grown to be less important in more recent years. This may be explained by the fact that the technology for photovoltaic systems has evolved to be well understood and mature. The interest in solar tiles on the other hand has increased in more recent years. By acquiring SolarCity, Tesla vertically integrated its energy generation and storage businesses and provided customers with a more extensive set of clean energy products and services. This acquisition emphasizes Tesla's absorptive capacity, a term first introduced by Cohen and Levinthal in 1989 used to describe a firm's ability to recognize, assimilate and commercialize the value of external knowledge (Lim & Falk, 2016). Cohen and Levinthal state that absorptive capacity is critical to a firm's innovative capabilities (Cohen & Levinthal, 1990).

From figure 21, it can be seen that the radicality of Tesla's portfolio has sharply risen after 2017. This is mainly caused by the increased presence of patents on chemical battery design and solar roof tiles (cluster 1 and 3), and the emergence of patents on the topic of artificial intelligence (cluster 5). Artificial intelligence is a technology that has gained in popularity over the more recent years, and comprised 38% of the total number of analyzed patents in 2019.

Besides the actual results in terms of radicality, one interesting result from this research is the method that was constructed using a combination of existing techniques. It was shown that these techniques can be used to assess radicality from patent content, but a variation on this method could also be used in different areas. The text processing techniques and clustering approach could be used on the content of other documents, like tweets or customer reviews, to be able to extract relevant information. For example, customer reviews can be clustered into different categories, to see which complaints occur most common and how this develops over time.

The goal of this research was to construct a method for quantifying the radicality of firms based on their patent portfolio. It was aimed to fill the gap in the academic literature on determining radicality with use of patent content analysis. To achieve this goal, a case study around the company Tesla was performed and the following research question was constructed: *"How did Tesla's radicality develop over the last 10 years?"* To answer this question, the study started by conducting a literature review of academic theory on relevant topics. Next, a research was designed in section 6 based on information from the literature and previously performed related studies. The results of this design, which shows the efficacy of the design, can be seen in section 7, which eventually leads to the development of a yearly radicality score in figure 21. In this figure, it can be clearly seen that the radicality of Tesla, which is measured on a scale of 1 to 3, has followed a positive trend over the last 10 years, with a minimum of 1.58 in 2012 and a maximum of 2.44 in 2019. This answers the case study research question on Tesla's radicality, and by extension, also answers the first research question: *"How can the radicality of a company be assessed using the content of its patent portfolio?"*

By constructing a method for quantifying radicality based on content analysis of patent portfolios, this study contributed to the existing literature on radicality and patent analysis by delivering a framework for future research. The method that was designed can be further optimized by scholars and researchers, and can be used to compare companies based on their radicality, both to one another and to the entire industry. It can be used as an indicator for competitive advantage, and may be of use to investors to make decisions on which companies to invest in.

# 9 Limitations and future recommendations

While the overall outcome of this research is satisfactory, there are a few flaws in the process that can be pointed out. First of all, the input for the text mining process are the patent titles as obtained from Orbis. To acquire a more accurate cluster allocation, it would be beneficial to use full-text patents instead of only the titles. By using full-text patents, the number of terms per document increases, resulting in a more accurate cosine distance between two documents, which leads to a finer classification of patents to clusters. In result, the silhouette width of the clusters would be larger, and less patents would be falsely allocated to a cluster. It may be possible that by using full-text patents, the patents that are now fitted to cluster 4, which is disregarded in this analysis due to the absence of a coherent topic, will be assigned to a more suitable cluster. The drawback of using full-text patents instead of only patent titles is that the size of the document-term matrix will increase significantly, which may have a detrimental effect on the computation time.

A second short-fall in the process, which is already mentioned above, is the disregarding of all patents that were classified in cluster 4. This portion of patents makes up for almost 40% of the total number of patents that is started with after pre-processing the documents. The author has attempted many methods to find a better classification for these patents, including latent dirichlet allocation (LDA), K-means clustering and a hybrid between K-means and hierarchical clustering, but without result. Eventually, the choice was made to remove these patents from further analysis, but for future research it would be better to try and allocate these patents more accurately. As explained above, this can for instance be done by using full-text patents.

Another dubious point in the method is the determining of the uncertainty scores using the questionnaires in appendix A. These questionnaires were filled out by the author only, which causes the output to be very reliant on the opinion and judgment of the author. It would be better to use multiple coders, and check for inter-coder reliability. Besides checking for inter-coder reliability, the validity of the uncertainty scores must also be checked. This can best be done by using a panel of experts to evaluate the results.Furthermore, the scoring of the uncertainties is time-dependent. Technologies that were very new and promising ten years ago may have been considered highly uncertain at the time, while it may be considered straightforward and well understood at present. It is difficult to adjust for this time-dependence when scoring uncertainty.

For future research, it is recommended to perform the analysis on multiple companies within the same industry to acquire reference. The radicality scores do not present much significance on their own, but can certainly be used to compare companies to each other or to the industry. Furthermore, it may be interesting to analyse the development of a firm's radicality over time, and to check if a correlation with a firm's financial performance exists.

## References

Aggarwal, C., & Zhai, C. (2012). An Introduction to Text Mining. In *Mining Text Data* (p. 1-10). Berlin: Springer.

Alberts, D., Yang, C., Fobare-DePonio, D., Koubek, K., Robins, S., Rodgers, M., ... DeMarco, D. (2017). Introduction to Patent Searching. In *Current Challenges in Patent Information Retrieval* (p. 3-45). Germany: Springer-Verlag GmbH.

Al-Khatib, A., & Al-Ghanem, E. (2021). Radical Innovation, Incremental Innovation, and Competitive Advantage, the Moderating Role of Technological Intensity: Evidence from the Manufacturing Sector in Jordan. *European Business Review*, *34*(3), 344-369.

Ardito, L., Petruzzelli, A., Panniello, U., & Garavelli, A. (2018). Towards Industry 4.0: Mapping Digital Technologies for Supply Chain Management-Marketing Integration. *Business Process Management Journal*, *25*(2).

Bouchet-Valat, M. (2020). Snowballc: Snowball stemmers based on the c 'libstemmer' utf-8 library [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=SnowballC` (R package version 0.7.0)

Burns, T., & Stalker, G. (1961). *The Management of Innovation*. London: Tavistock Publications.

Chandy, R., & Tellis, G. (2000). The Incumbent's Curse? Incumbency, Size, and Radical Product Innovation. *Journal of Marketing*, *64*(3), 1-17.

Chen, Y., & Chang, K. (2010). The Relationship Between a Firm's Patent Quality and its Market Value - The Case of US Pharmaceutical Industry. *Technological Forecasting & Social Change*, *77*(2010), 20-33.

Cohen, W., & Levinthal, D. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1990), 128-152.

Cooper, R., Edgett, S., & Kleinschmidt, J. (1998). Best Practices for Managing R&D Portfolios. *Research-Technology Management*, *41*(4), 20-33.

Crosby, M. (2000). Patents, Innovation and Growth. *The Economic Record*, *76*(234), 255-262.

D'Agostino, M., & Dardanoni, V. (2009). What's so Special About Euclidean Distance? A Characterization with Application to Mobility and Spatial Voting. *Social Choice and Welfare*, *33*(2009), 211-233.

Daim, T., Rueda, G., Martin, H., & Gerdsri, P. (2005). Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis. *Technological Forecasting and Social Change*, *73*(2006), 981-1012.

Garcia, R., & Calantone, R. (2002). A Critical Look at Technological Innovation Typology and Innovativeness Terminology: A Literature Review. *The Journal of Product Innovation Management*, *19*(2002), 110-132.

Godrey, D., Johns, C., Sadek, C., Meyer, C., & Race, S. (2014). A Case Study in Text Mining: Interpreting Twitter Data From World Cup Tweets.

Hall, B., & Ziedonis, R. (2001). The Patent Paradox Revisited: An Emprical Study of Patenting in the U.S. Semiconductior Industry, 1979-1995. *The RAND Journal of Economics*, *32*(1), 101-128.

Han, J., Kamper, M., & Pei, J. (2012). Getting to Know Your Data. In *Data mining (third edition)* (p. 39-82). Morgan Kaufmann.

Han, S., Diao, K., & Sun, X. (2021). Overview of Multi-phase Switched Reluctance Motor Drives for Electric Vehicles. *Advances in Mechanical Engineering*, *13*(9).

Henriksson, F. (2017). Introducing New Materials in the Automotive Industry: Managing the Complexity of Introducting New Materials in Existing Production Systems. *Linköping Studies in Science and Technology*(1973).

Humaira, H., & Rasyidah, R. (2020, 01). Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm..

Hurley, R., Tomas, G., & Hult, M. (1998). Innovation, Market Orientation, and Organization Learning: An Integration and Empirical Examination. *Journal of Marketing*, *62*(July 1998), 42-54.

Hurmelinna-Laukkanen, P., Sainio, L., & Jauhiainen, T. (2008). Appropriability Regime for Radical and Incremental Innovations. *R&D Management*, *38*(3), 278-289.

International Energy Agency. (2022). Global EV Outlook 2022. `www.iea.org`.

Jannidis, F., Pielström, S., Schöch, C., & Vitt, T. (2015). Improving Burrows' Delta - An Empirical Evaluation of Text Distance Measures. In *Digital humanities conference* (Vol. 11).

Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, *28*(1), 11-21.

Karki, M. (1997). Patent Citation Analysis: A Policy Analysis Tool. *World Patent Information*, *19*(4), 269-272.

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and visualize the results of multivariate data analyses [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=factoextra` (R package version 1.0.7)

Katila, R. (2000). Using Patent Data to Measure Innovation Performance. *International Journal of Business Performance Management*.

Kayser, V., & Shala, E. (2020). Scenario Development Using Web Mining for Outlining Technology Futures. *Technological Forecasting & Social Change*, *156*(2020).

Lerner, J. (1994). The Importance of Patent Scope: An Empirical Analysis. *The RAND Journal of Economics*, *25*(2), 319-333.

Lew, A., & Mauch, H. (2006). Introduction to Data Mining Principles. *Studies in Computational Intelligence (SCI)*, *38*, 1-20.

Li, N., & Wu, D. (2010). Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast. *Decision Support Systems*, *48*(2010), 354-368.

Lim, K., & Falk, M. (2016). Absorptive Capacity. In *The palgrave encyclopedia of strategic management.* United Kingdom: Palgrave MacMillan.

MacMillan, I., & McGrath, R. (2002). Crafting R&D Project Portfolios. *Research Technology Management*, *45*(5), 48-59.

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Marutho, D., Hendra Handaka, S., Wijaya, E., & Muljono. (2018). The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. In *2018 international seminar on application for technology of information and communication* (p. 533-538).

McDermott, C., & O'Connor, G. (2002). Managing Radical Innovation: An Overview of Emergent Strategy Issues. *The Journal of Product Innovation Management*, *19*(2002), 424-438.

Millar, C., Groth, O., & Mahon, J. (2018). Management Innovation in a VUCA World: Challenges and Recommendations. *California Management Review*, *61*(1), 5-14.

Mirkin, B. (2011). Choosing the Number of Clusters. *WIREs Data Mining and Knowledge Discovery*, *1*(3), 252-260.

Peñarroya-Farell, M., & Miralles, F. (2021). Business Model Dynamics from Interaction with Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(81).

Pil, F., & Cohen, S. (2006). Modularity: Implications for Imitation, Innovation, and Sustained Advantage. *Academy of Management Review*, *31*(4), 995-1011.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Salton, G., & Buckley, C. (1988). Term-weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, *24*(5), 513-523.

Schumpeter, J. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credits, Interest, and the Business Cycle.* Piscataway: Transaction Publishers.

Setyoko, P., & Kurniasih, D. (2022). SMEs Performance During Covid-19 Pandemic and VUCA Era: How the Role of Organizational Citizenship Behavior, Budgetary, Participation and Information Asymmetry? *International Journal of Social and Management Studies*, *3*(4).

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in r. *JOSS*, *1*(3). Retrieved from `http://dx.doi.org/10.21105/joss.00037` doi: 10.21105/joss.00037

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach.* The United States of America: O'Reilly Media Inc.

Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' Knowledge Profiles: Mapping Patent Data With Unsupervised Learning. *Technological Forecasting & Social Change*, *115*(2017), 131-142.

Tesla Inc. (2016). *Form 10-K 2016.* `https://www.sec.gov/Archives/edgar/data/1318605/000156459017003118/tsla-10k_20161231.htm`. U.S. Securities and Exchange Commission.

Tseng, Y., Lin, C., & Lin, Y. (2007). Text Mining Techniques for Patent Analysis. *Information Processing & Management*, *43*(5), 1216-1247.

Ulian, D. Z., Becker, J. L., Marcolin, C. B., & Scornavacca, E. (2021). Exploring the Effects of Different Clustering Methods on a News Recommender System. *Expert Systems With Applications*, *183*(2021).

Ulrich, K. (1995). The Role of Product Architecture in the Manufacturing Firm. *Research Policy*(24), 419-440.

Venables, W., Smith, D., & The R Development Core Team. (2007, October). An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics (2.6.0 ed.) [Computer software manual]. Vienna, Austria.

Žižka, J., František, D., & Svoboda, A. (2020). *Text Mining with Machine Learning: Principles and Techniques.* Boca Raton: CRC Press.

Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236-244.

Weil, M., Ziemann, S., & Peters, J. (2018). The Issue of Metal Resources in Li-Ion

Batteries for Electric Vehicles. In *Behaviour of lithium-ion batteries in electric vehicles* (p. 59-74). Switzerland: Springer.

Wheelwright, S., & Clark, K. (1992). Creating Project Plans to Focus Product Development. *Harvard Business Review*, *70*(2), 70-82.

Wijnhoven, F. (2021). *Knowing the Future.* University of Twente.

# A Questionnaire to determine technical & market uncertainty

| 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 the market uncertainty of a project. Derived 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 technologybarriers 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 the technical uncertainty of a project. Derived from MacMillan & McGrath (2002).

# B Uncertainty scores

**MARKET UNCERTAINTY**

| | C1 | C2 | C3 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **M1** | 2 | 3 | 4 | 6 | 3 | 5 | 2 | 4 | 1 | 1 | 1 | 2 | 1 |
| **M2** | 3 | 3 | 5 | 6 | 2 | 7 | 3 | 6 | 3 | 1 | 1 | 2 | 1 |
| **M3** | 3 | 3 | 5 | 6 | 2 | 7 | 3 | 6 | 3 | 2 | 2 | 5 | 1 |
| **M4** | 3 | 4 | 5 | 5 | 2 | 5 | 4 | 3 | 2 | 1 | 1 | 4 | 2 |
| **M5** | 3 | 3 | 3 | 4 | 2 | 2 | 4 | 3 | 4 | 2 | 2 | 3 | 1 |
| **M6** | 4 | 3 | 5 | 5 | 2 | 6 | 4 | 5 | 4 | 2 | 2 | 5 | 2 |
| **M7** | 2 | 4 | 4 | 5 | 4 | 6 | 2 | 5 | 2 | 2 | 2 | 3 | 2 |
| **M8** | 4 | 4 | 4 | 6 | 4 | 5 | 2 | 2 | 2 | 2 | 2 | 5 | 2 |
| **M9** | 3 | 2 | 5 | 4 | 3 | 4 | 2 | 3 | 1 | 1 | 1 | 3 | 1 |
| **M10** | 2 | 3 | 4 | 5 | 3 | 6 | 3 | 4 | 1 | 2 | 2 | 4 | 1 |
| **M11** | 4 | 4 | 5 | 4 | 5 | 5 | 3 | 4 | 2 | 2 | 2 | 5 | 2 |
| **M12** | 2 | 4 | 5 | 5 | 3 | 5 | 3 | 5 | 2 | 3 | 3 | 4 | 2 |
| **M13** | 4 | 3 | 6 | 5 | 3 | 5 | 4 | 4 | 2 | 3 | 3 | 4 | 2 |
| **M14** | 3 | 3 | 4 | 5 | 2 | 4 | 3 | 4 | 3 | 2 | 2 | 4 | 1 |
| **M15** | 4 | 4 | 5 | 6 | 2 | 5 | 3 | 3 | 2 | 1 | 1 | 4 | 1 |
| **M16** | 5 | 2 | 4 | 6 | 2 | 6 | 3 | 2 | 2 | 2 | 2 | 4 | 2 |
| **M17** | 4 | 3 | 4 | 4 | 2 | 6 | 3 | 3 | 3 | 2 | 2 | 4 | 1 |
| **M18** | 3 | 3 | 4 | 6 | 3 | 6 | 3 | 3 | 2 | 2 | 2 | 5 | 2 |
| **M19** | 2 | 4 | 4 | 6 | 3 | 5 | 4 | 5 | 2 | 2 | 2 | 5 | 1 |
| **M** | **3,2** | **3,3** | **4,5** | **5,2** | **2,4** | **5,3** | **3,1** | **3,9** | **2,3** | **1,8** | **1,4** | **4,0** | **1,5** |

Table 5: Market uncertainty scores

**TECHNICAL UNCERTAINTY**

| | **C1** | **C2** | **C3** | **C5** | **C6** | **C7** | **C8** | **C9** | **C10** | **C11** | **C12** | **C13** | **C14** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T1** | 6 | 4 | 4 | 6 | 2 | 4 | 3 | 3 | 2 | 2 | 5 | 2 | 1 |
| **T2** | 5 | 3 | 3 | 5 | 2 | 4 | 3 | 2 | 1 | 1 | 4 | 1 | 2 |
| **T3** | 5 | 2 | 3 | 6 | 1 | 3 | 2 | 2 | 1 | 1 | 4 | 2 | 2 |
| **T4** | 6 | 3 | 3 | 6 | 2 | 4 | 2 | 2 | 2 | 2 | 5 | 2 | 1 |
| **T5** | 5 | 3 | 5 | 4 | 3 | 4 | 2 | 3 | 2 | 2 | 5 | 2 | 1 |
| **T6** | 6 | 3 | 2 | 5 | 2 | 5 | 3 | 3 | 2 | 2 | 4 | 2 | 1 |
| **T7** | 7 | 4 | 2 | 4 | 2 | 5 | 3 | 3 | 2 | 2 | 5 | 1 | 1 |
| **T8** | 5 | 4 | 3 | 6 | 3 | 4 | 4 | 3 | 3 | 2 | 5 | 3 | 2 |
| **T9** | 5 | 4 | 3 | 6 | 3 | 5 | 4 | 3 | 3 | 3 | 5 | 3 | 1 |
| **T10** | 5 | 4 | 3 | 6 | 3 | 4 | 3 | 3 | 3 | 2 | 5 | 2 | 2 |
| **T11** | 3 | 3 | 3 | 5 | 3 | 3 | 4 | 4 | 2 | 2 | 5 | 2 | 1 |
| **T12** | 6 | 4 | 4 | 6 | 3 | 5 | 4 | 4 | 2 | 2 | 5 | 2 | 2 |
| **T13** | 6 | 4 | 4 | 6 | 3 | 6 | 4 | 4 | 2 | 3 | 6 | 1 | 2 |
| **T14** | 5 | 5 | 4 | 5 | 3 | 4 | 3 | 3 | 2 | 3 | 5 | 2 | 2 |
| **T15** | 6 | 4 | 3 | 6 | 2 | 5 | 2 | 3 | 2 | 2 | 5 | 2 | 1 |
| **T16** | 2 | 3 | 3 | 5 | 2 | 2 | 2 | 2 | 2 | 1 | 4 | 2 | 1 |
| **T17** | 3 | 3 | 2 | 5 | 2 | 3 | 2 | 3 | 3 | 2 | 5 | 2 | 1 |
| **T18** | 6 | 4 | 3 | 6 | 3 | 3 | 3 | 2 | 2 | 1 | 4 | 2 | 1 |
| **T19** | 6 | 5 | 2 | 5 | 2 | 3 | 2 | 2 | 2 | 1 | 4 | 1 | 2 |
| **T20** | 6 | 6 | 3 | 5 | 3 | 5 | 2 | 3 | 2 | 1 | 4 | 2 | 2 |
| **T21** | 3 | 3 | 3 | 4 | 3 | 5 | 2 | 2 | 2 | 2 | 5 | 3 | 1 |
| **T22** | 4 | 5 | 4 | 5 | 4 | 5 | 2 | 2 | 2 | 2 | 4 | 2 | 1 |
| **T23** | 4 | 4 | 5 | 6 | 3 | 5 | 2 | 3 | 2 | 2 | 4 | 2 | 2 |
| **T24** | 6 | 5 | 5 | 4 | 3 | 4 | 3 | 2 | 2 | 2 | 4 | 2 | 2 |
| **T** | **5,0** | **3,8** | **3,3** | **5,3** | **2,6** | **4,2** | **2,8** | **2,8** | **2,1** | **1,9** | **4,6** | **2,0** | **1,5** |

Table 6: Technical uncertainty scores