MASTER THESIS

Exploring the integration of automated text classification solutions in roadmapping

Tom Benerink

Behavioural, Management and Social Sciences

Dr. Erwin Hofman

EXAMINATION COMMITTEE

Dr. Igors Skute

Dr. Ingo Nee

DOCUMENT NUMBER

ROSEN empowered by technology UNIVERSITY OF TWENTE.

Table of Contents

Abstract

Technology is of key strategic importance for delivering competitive advantage and value to companies and the industrial networks in which they operate. This importance increases when faced with high costs, complexity, globalization and fast technology change rates (Phaal et al., 2004). Correctly managing technology is thus of key strategic importance. A powerful tool to enable and support successful management and planning of technology is a Technology Roadmap (Phaal et al., 2004). Over time the roadmapping process has been described in detail and effective blueprints on how to roadmap are available, such as the T-Plan (Phaal et al., 2001) or the Scenario Driven Roadmap (Siebelink et al., 2016). Large firms with more diverse portfolios and capabilities will require more extensive roadmaps, increasing the overall scope of the roadmapping process and outputs from the workshop phase. The output from the workshop phase exists out of strategic focus areas and preconditions, which require processing to make a selection of outputs to put on the eventual roadmap. The existing blueprints on how to develop a roadmap do not address the scalability challenge of processing and selection of the workshop outputs. Currently, a number of automated solutions have successful applications to analyse, cluster or classify text documents, such as machine learning classifiers or computer aided text analysis (Short et al., 2010). Can these automated solutions be used to make processing of workshop outputs within the roadmapping process more effective and efficient? This research experimented with word frequency based machine learning classifiers & clustering and computer aided text analysis. These different solutions were used to classify the workshop outputs within the roadmapping process on novelty and value criteria or categorize the output into logical categories. The performance on these classification and categorization tasks was then compared to manual classification and categorization.

The quality and sample size of the dataset used provided challenges for the used automated solutions, as the assumptions that distinct features can be identified and that the vocabulary between groups is distinct were not met in this specific case. Automated exploratory data analysis and the use of computer aided text analysis did enable a convenient overview and a priori creation of base categories. The developments within natural language processing/text mining are taking great strides. Thus, it could be possible that with more advanced models, which can understand meaning of text or with a higher quality dataset, on the long term simple or more advanced automated solutions could be used reliably and prove useful to separate good and bad inputs.

Additionally, the lessons learned on why automated solutions to classify and cluster experienced difficulties to perform, resulted in a suggestion to improve the data collection design in the preparation phase of the roadmapping process. These lessons can already be used for near future iterations when the amount of participants is high and the workload to process the generated outputs is expected to be large. This study suggests to put more effort in the design of data processing at the preparation phase and transfer the scoring on criteria and the categorization partly to the participants in the workshop phase. Effectively linking the format of data collection in the workshop phase to the intended processing and selection.

1. Introduction

In organizations a big part of the work consists out of the acquisition, sharing and application of information and knowledge (Purser & Montuori, 1995). This is especially critical when generating new ideas for a business to pursue. Assuming that the ideas are discrete and requires hard choices to be made between them, there is often a multi round funnel or tournament approach to select the best ideas. This idea selection task you can consider as a prediction task. Organizations face uncertainty and within it they try to select the ideas they expect to be the best choice, which is even with perfect criteria a difficult prediction to make. This is even further complicated by the number of ideas that are available for selection, as less time will be available to devote to each unique idea. Resulting in a situation in which the more ideas there are, the more likely it becomes that the person selecting is not able to analyse thoroughly on essential criteria such as novelty or value.

Technology is of key strategic importance for delivering competitive advantage and value to companies and the industrial networks in which they operate. This importance increases when faced with high costs, complexity, globalization and fast technology change rates (Phaal, Farruk & Probert, 2004). In order to manage technology correctly under these challenging circumstances an effective system which facilitates idea generation, and the funnelling of these ideas needs to be in place. ROSEN Technology and Research Center GmbH (ROSEN), as a firm depending on advanced technological products and processes recognized this and collaborated with the University of Twente to work on a roadmapping process and model based on Scenario Driven Roadmapping (Siebelink, Halman &, Hofman, 2016). Roadmaps have great potential to support development and implementation of technology and product plans. Functioning as a radar by extending planning horizons and identifying threats and opportunities (Phaal et al., 2004).

The first appearance of roadmapping was identified in the U.S. automotive industry (Probert and Radnor, 2003). Motorola and Corning developed systematic approaches in the late 70's/early 80's. This more visible approach of Motorola caused the adoption of roadmapping techniques by others in the consumer electronic industry, such as Philips (Groenveld, 1997), Lucent Technologies (Albright et al., 2003), and the SIA (Kostoff & Schaller, 2001). The chain reaction continued from this point onward resulting in adoption by government and consortia who supported sector wide research collaboration. The motivation to adopt this technique has been concretely defined by Phaal et al. (2004, p 9.): *'Technology roadmapping represents a powerful technique for supporting technology management and planning, especially for exploring and communicating the dynamic linkages between technological resources, organizational objectives and the changing environment.'* An example from practice of a benefit is the statement of Motorola and Philips that roadmapping enables them to match the pace of a fast changing business environment (Simonse, Hultink, & Buijs, 2015).

The developments of this process however did not tackle the scalability problem of idea selection. Scalability refers in this case to the amount of ideas that can be reasonably processed by the average manpower a firm devotes to the data processing. The larger a firm and the more diverse its portfolio is, the more employees and potential participants there are in the roadmapping process, thus the more potential options can be generated that could be selected to be put on the firm's roadmap. Resulting in a large sample of good and bad ideas. This is more demanding on the processing. For example, 100 options would be perfectly possible to assess manually, but 3000 options would be weeks of work. So, when the scale of roadmapping increases, an effective form of automated classification is required to retrieve, analyse, curate and annotate documents (in this case the strategic options) (Kowsari et al., 2017). To effectively separate the ideas into good and bad ideas and find the very best ones.

A variety of solutions has been developed by researchers to solve this document classification problem. These solutions aim to relieve a person from reading/scanning every document and decide on a classification. Instead this is done automatically and faster by a machine. The information retrieval field first focussed on search engine basics such as indexing and dictionaries (Manning et al., 2008). Upon these basics additional work was performed providing improvements by introducing feedback and query reformulation (French, Brown & Kim, 1997) (Kowsari et al., 2015). More recent work focussed on the employment of data mining and machine learning techniques. One of the most accurate being the support vector machine (SVM) (Joachims, 1999; Tong & Koller, 2001; Fernandez-Delgado et al., 2014). This method uses kernel functions to discover separating hyperplanes in a high-dimensional space (Kowsari et al., 2017). Although accurate, SVM are difficult to interpret, therefore many information retrieval systems use Naïve Bayes (McCallum & Nigam, 1998; Kim et al., 2006) or decision trees (French et al, 1997). These methods are easier to interpret and therefore enable easier query reformulation, at the cost of some accuracy. Newer methods can be found in the deep learning field. Deep learning is an efficient form of neural networks (Hinton & Salakhutdinov, 2006), which can perform unsupervised, supervised and semi-supervised tasks (Johnson & Zhang, 2014). Image processing already saw extensive use of deep learning, but recently these methods have been leveraged to other domains such as data and text mining. This field is under heavy development, stimulated by big tech companies such as Google, Microsoft, Amazon and Facebook. Developments are at an unprecedented rate in the last three years with the introduction of Transformer models (Roberts & Raffel, 2020). Resulting in computers being able to actually 'understand' more or less natural language, for the first time outperforming manual classification on major NLP benchmarks (Devlin et al., 2018). These methods are able to leverage the content of large datasets to specific tasks, which makes them distinct from preceding neural networks, which needed thousands or millions of task specific training examples. This is essential as data for a specific task are often stuck in the middle, with too little data to effectively use traditional methods based on word frequencies and far too little for neural network based methods, but still providing a huge task to process manually.

The goal of this research is to identify the challenges and possibilities there are to make data processing more efficient and effective within the workshop phase of the roadmapping process. To reach this goal first the roadmapping inputs, process and outputs will be explored, after which the challenges and opportunities there are for more effective and efficient data processing are identified. When this is established various ways of processing text data automatically are introduced, which will be applied to find out if the current manual processing of text data in roadmapping can benefit from (computer aided) automated or semi-automated text processing.

2. ROSEN

The company for which this research was performed is the ROSEN Group, more specifically the ROSEN Technology and Research Center GmbH Lingen. ROSEN has been founded by Hermann Rosen in 1981 and is a privately owned family business which is currently predominantly active within oil & gas, mining, transportation, process and manufacturing industries. They are active and have facilities worldwide. Their portfolio exists out of services and products. Examples of services are inspection and integrity as well as research and development solutions. Their product portfolio is characterized by deep vertical integration, resulting in 85% of the products made in house. Their product portfolio is too large to fully describe here, but among other there are products such as: sensor and data acquisition technologies, pipeline cleaning and inspection tools and pipeline interior coatings. Intelligent products that combine elastomer properties with sensors are also part of their offerings. Apart from hardware solutions ROSEN also is a leading supplier in customized software. ROSEN lives up to its credo of Empowered by Technology as it is an extremely high tech and R&D focussed company with a large and complex portfolio developed almost completely in house. This requires effective processes to guide technology development and portfolios over time.

2.1 Confidentiality

Due to the collaboration with ROSEN, sensitive/specific information on ROSEN has been blurred out of the public version of this thesis. This does not influence the results or readability.

3. Research Questions

To reach the goal of this study a set of research questions was developed to guide the research process. The main research question of this study is:

How can data processing in roadmapping become more effective and efficient using automated solutions?

To aid in answering this central research question a set of sub questions have been developed:

1. How is the process of business roadmapping structured?
    a. What data are typically collected?
    b. How are the data processed?
    c. What conclusions are typically drawn that together allow us to sketch the roadmap?

2. Which of the steps in the roadmapping process provide an opportunity to automize using text mining tools? To what extent can existing automated solutions optimise the roadmapping process?
    a. What are applications of machine learning?
    b. Can the manual classification of options be replicated using natural language processing techniques?
    c. How do different solutions perform?
    d. Which data processing strategy is recommended for future roadmapping iterations?

4. Roadmapping: the inputs, processes, outputs and roadmap explained

Firstly, it is important to understand what a roadmap is, what inputs the roadmap requires to be drawn and which processes create these inputs. By fully understanding the process and inputs the current and future data processing methods can be evaluated based on the needs and characteristics of the roadmapping case. In this section the roadmapping process is explained and the opportunities to improve the current manual processing standard are identified.

4.1 Technology Roadmapping

Simonse et al. provide three basic characteristics of the roadmap object*: '(1) a visual portrait, which provides an (2) outline of market, product, and technology plans, with elements that (3) are plotted on a timeline'* (Simons et al., 2015, p. 910). Other scholars developed insights on the process of road mapping, such as using workshops in the development (Phaal, Farrukh and Probert, 2007) and the roadmap architecture (Phaal and Muller, 2009). The form and purpose of a roadmap is flexible, making it suitable for different innovation and strategic contexts, functioning as a common language for exploring, mapping and communicating the evolution and development of business systems (Phaal & Muller, 2009). This function of common language is valuable, as technology can be considered as a specific type of knowledge due to being applied, resulting in a focus on the 'know-how' of an organisation, combining 'hard' technology (science & engineering) with 'soft technology' (the enablers of successful technology implementation, new product development, innovation and organisational structures e.a.) (Phaal et al., 2001).

The business roadmap is a useful tool to implement and formulate strategies (Vishnevskiy, Karasev, & Meisner, 2015). Due to providing a comprehensible visual representation of the evolution over time of markets, products, capabilities and technologies. Resulting in high communicative and directive power. The two critical components of constructing a business roadmap are: formulation of strategy and developing it into a roadmap (Goffin and Mitchell, 2005). The T-Plan approach developed by Phaal et al. covers these two components. It exists out of three stages: planning, workshop and rollout (Phaal, 2001). Additionally Albright and Kappel (2003) introduced the concept of focus areas, which are areas defined during environmental analysis in the workshop stage, in which the firm can identify opportunities that must be expressed in the form of necessary capabilities and concrete products. In this process scholars used different systematic and formalized analyses (Groenveld, 2007; Albright and Kappel, 2003; Phaal et al., 2001), based on traditional strategic planning, such as PESTEL, SWOT and Porter's Five Forces (Porter, 1980). These are suitable for identifying threats and opportunities in the external environment and strength and weaknesses in the internal environment (Siebelink et al., 2016). These formalized systematic analyses however assume that the future will be more or less like the present, making them unsuitable for dealing with uncertainty and discontinuity. Strategic literature contains numerous examples of claims that firms need to continuously adapt and deal with uncertainty (Siebelink et al., 2016). The literature on business roadmaps however did not assess yet the obvious strategic need to deal with uncertainty. Saritas and Aylen (2010), Strauss and Radnor (2004) and Siebelink et al. (2016) contributed to this research gap by proposing to integrate scenario planning into the roadmap process.

Scenario planning incorporates multiple futures, 'probing the future' (Brown and Eisenhardt, 1998). It increasingly has been viewed as a tool to assess discontinuity, thus emphasis automatically shifts to the aspects expected to change in the future (Derbyshire and Wright, 2016). However, it is distinct from forecasting, because forecasting focusses on continuing trends, assessing change along the same trajectory as in the recent past (Derbyshire and Giovannetti, 2017). The focus in scenario planning is not on probability, but on plausibility, allowing the consideration of extreme outcomes, such as complete market (non) acceptance. Facilitating the consideration of actions to avoid or facilitate these extreme outcomes (Derbyshire and Giovannetti, 2017). It has been identified by multiple scholars a field that could provide the solution for coping with the existence of uncertainty and multiple possible futures and incorporate it into roadmapping to construct a robust roadmap (Siebelink et al., 2016; Geschka and Hahnenwald, 2013; Petrick and Martinelli, 2012; Saritas and

Aylen, 2010; Strauss and Radnor, 2004). Important here was to maintain the clear process and communicative and directive strengths attributed to the roadmapping process. The latest contribution being the Scenario-Driven Roadmapping of Siebelink et al. (2016) does so, but retains some weaknesses: the time-consuming nature of the process and the required additional analysis. To reduce the time needed and improve accuracy this research proposes computer aided text analysis as a possible tool to do so.

The aforementioned flexibility in form and purpose is highlighted by Phaal et al. (2007) defining eight different purposes and eight different roadmap formats, although hybrid forms exist. One of the identified formats is text and the other graphical formats often have text-based reports associated with them (Phaal et al, 2007). In the scenario-driven roadmapping developed by Siebelink et al. (2016) there is after the workshop phase output in text in the form of strategic options created which requires processing in order to use it for roadmap development. To use the workshop output for roadmap development it needs to be processed to select the focus areas and preconditions that are used to construct the roadmap.

This research builds upon this roadmapping literature, specifically it complements the T plan of Phaal et al. (2001) and the scenario-based roadmapping of Siebelink et al. (2016). Firstly, the T plan approach is described and then the scenario driven roadmap approach.

4.2 T-Plan approach

The structure of this section (4.1.2) and its examples, figures and descriptions are adapted from Phaal et al. (2001). To understand the T-Plan approach it is important to first understand the variety in purposes and formats of roadmaps that have been identified by Phaal et al. (2001).

4.2.1 Purposes

1. *Product planning*
   The most common type of a technology roadmap, focusses on the combination of technologies and products, often contains more than one generation of a product.
2. *Service/capability planning*
   Focussing on how technology supports organisational capabilities, rather similar to type 1.
3. *Strategic planning*
   Adds a strategic dimension to the roadmap, enabling the assessment of opportunities and threats, often at a business level.
4. *Long-Range planning*
   Unique to this roadmap is the extension of the planning horizon, resulting in execution on a national level.
5. *Knowledge asset planning*
   Business objectives alignment with knowledge assets and initiatives.
6. *Programme planning*
   Focussing on the implementation of strategy, directly relates to project planning.
7. *Process planning*
   Usage for the management of knowledge, specifically when the focus is on one specific area.
8. *Integration planning*
   Used for the evolution and/or integration of technology. Focussing on the combination of technologies within systems or products or the forming of new technologies. The time dimension is often not explicitly shown.

4.2.2 Formats

a) *Multiple layers*
   This is the most common format of a technology roadmap. It exists out of a number of layers, such as technology, product and market. Opening up the possibility to explore the evolution

within each layer and the inter-layer dynamics. This results in the facilitation of integrating technology into business systems, products and services.
Example: A Philips roadmap that illustrates the integration of product and process technologies, supporting the development of functionalities in future products.

b) *Bars*
Illustration in the form of a set of bars for each layer or sub-layer. It simplifies and unifies the required outputs. This is advantageous because it facilitates communication, integration and the development of software to support roadmapping.
Example: The Motorola roadmap (Willyard and McClees, 1987). It depicts the evolution of car radio product features and technologies

c) *Tables*
Sometimes a roadmap is put into a table format, for example time vs. performance. It is especially suited if the performance is quantifiable and activities are clustered in time periods.
Example: a table roadmap (EIRMA, 1997). Incorporating the performance dimension for products and technology against time.

d) *Graphs*
If performance of a technology is quantifiable the roadmap can take the form of a graph or plot. Mostly each sublayer has its own plot. Also known as an ‘experience curve’, this format is closely related to technology ‘S-curves’.
Example: A set of products and technologies that co-evolve shown by a roadmap in graph form (EIRMA, 1997).

e) *Pictorial representations*
A more creative approach in the form of a pictorial representation in order to communicate technology and integration plans. Occasionally metaphors are used as support for the objective.
Example: A Sharp Roadmap, using the metaphor of a tree, it relates to the development of products and product families.

f) *Flow charts*
A distinct form of pictorial representation, used to relate objectives, actions and outcomes.
Example: A NASA roadmap, it shows the relation between the vision of the organization with its mission, primary business areas, contribution to US national priorities, fundamental scientific questions, and goals.

g) *Single layer*
A subset of format ’a’, now only focussing on one layer. Less complex at the costs of not showing the linkages between layers.
Example: the example of ‘b’ is a single layer roadmap; it focusses only on the layer of technological evolution.

h) *Text*
Sometimes roadmaps are mostly or entirely text based. Instead of graphically displaying issues as other formats do, they are described.
Example: The ‘white papers’ of the Agfa, these papers support understanding of market and technological trends that will influence a sector.

This variety of purposes and formats is graphically summarized in figure 1. There are 8 purposes and 8 formats, however the data processing for each purpose or format should be more or less similar, depending on the choice of approach taken to tackle the process of constructing a roadmap. It could be influenced by the need to adapt the approach to every specific situation. Roadmaps can contain elements of more than one of the purpose/format categories defined above, therefore not always fitting in nicely in a category. As a result, custom, situation specific hybrid forms are developed.

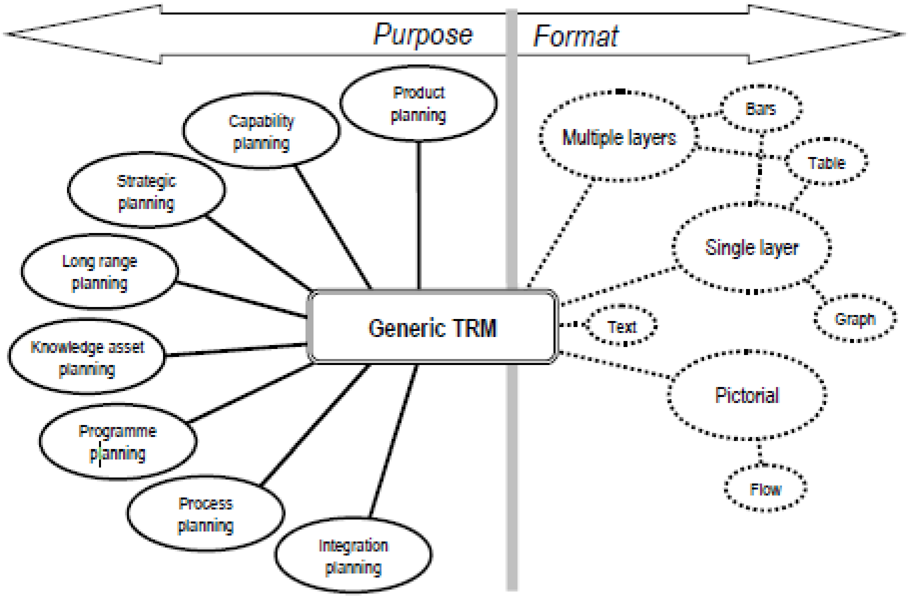


Figure 1. Characterisation of roadmaps: purpose and format. Adapted from Phaal, R., Farrukh, C., & Probert, D. (2001). *T-Plan: the fast-start to technology roadmapping: planning your route to success*. University of Cambridge, Institute for Manufacturing.

4.2.3 Process

The T Plan approach is grounded in practice as it is developed during a three-year applied research programme. In this research more than 20 roadmaps in several industry sectors have been developed together with different types of companies (Table 1.). The application of T-Plan approach aims to:

‘1. Support the start-up of company specific TRM processes.
2. Establish key linkages between technology resources and business drivers.
3. Identify important gaps in market, product and technology intelligence.
4. Develop a ‘first-cut’ technology roadmap.
5. Support technology strategy and planning initiatives in the firm.
6. Support communication between technical and commercial functions.’ (Phaal et al., 2001)

Furthermore, the T-Plan approach comes in two ‘flavours’:

1. The standard approach, suitable for supporting product planning (Phaal et al., 2000).
2. Customised approach, providing guidance on a broader application of the T-Plan.

Table 1

Applications of T-Plan fast-start TRM process

#	Company ID	Case type	Product / sector	Business aims	TRM type - purpose*	TRM type - format*
1	A	Exploratory	Industrial coding systems	Product planning for inkjet printing	1	a
2	A	Development	Industrial coding systems	Product planning for laser printing	1	a
3	B	Exploratory, development & test (10 applications)	Postal services	Integration of technology and research into business	2, 3, 4, 6, 8	a
4	C	Development	Security / access systems	Product family planning	1	a
5	D	Test (2 applications)	Software (labelling)	Exploration of product opportunity	1	a
6	E	Test & development	Surface coatings	New product development plan	1, 7	a
7	F	Test & development	Power transmission & distribution	Exploration of business opportunity for new technology	1, 3	a
8	G	Test & development (2 applications)	Railway infrastructure	Capital investment and technology insertion planning	3, 8	a
9	H	Test & development	Automotive sub-systems	Reliability services planning	2	a
10	I	Test & development (2 applications)	Medical packaging	Exploration of new business model	3	a
11	J	Test	Building controls	Exploration of new business model	3	a

** See sections 3.2.1 and 3.2.2.*

Adapted from Phaal, R., Farrukh, C., & Probert, D. (2001). *T-Plan: the fast-start to technology roadmapping: planning your route to success*. University of Cambridge, Institute for Manufacturing.

The standard process uses four facilitated workshops. The three key layers of the roadmap are focussed on in the first three workshops: market/business, product/service and technology. The final workshop is reserved to bring the layers together using a time basis to construct the graphical roadmap. As seen in figure 2.

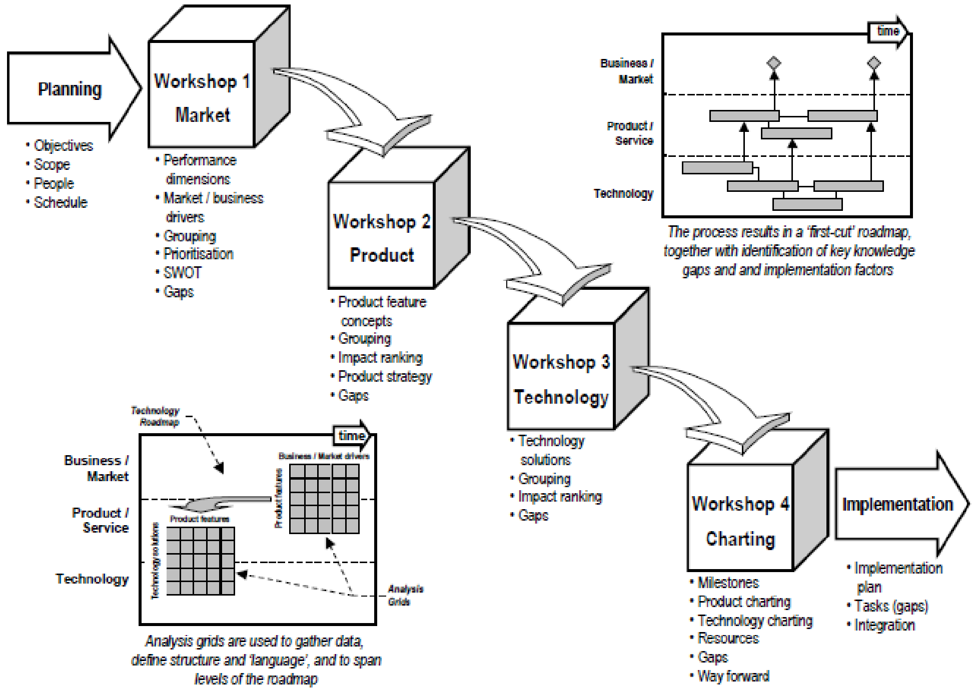


Figure 2. T-Plan: standard process steps, showing linked analysis grids. Adapted from Phaal, R., Farrukh, C., & Probert, D. (2001). *T-Plan: the fast-start to technology roadmapping: planning your route to success*. University of Cambridge, Institute for Manufacturing.

Although not specifically mentioned yet it is also important to keep the parallel management activities in one's mind. This entails process coordination, planning/facilitation of workshops and follow-actions.

Not a single case of roadmapping is identical due to different environments, structures, processes etc. Thus, to reap the full benefits of roadmapping it is safe to assume that the T-Plan approach needs customising. When a customized approach is chosen the multi-layer roadmap is often chosen as a format, due to being most flexible in its application. The following dimensions can be adapted to suit specific needs (Phaal et al., 2001):

- Time: flexible in the sense that the time horizon can be adapted from short to long term, the scale can be altered to a logarithmic format to create more space for the short term and intervals can be continuous or in periods of for example six months. Additionally, the roadmap can reserve space for an extremely long range vision or considerations while also showing the current state to identify the gaps between them.

- Layers: the vertical axis of a roadmap is important because it needs to fit the organisation and problem that is being assessed. Typically, a large initial part of the roadmapping process is dedicated to identifying the layers and sublayers on the vertical axis. Often the layers are constructed such that the top layer reflects the organizational purpose ('know-why'), the bottom layer represents the resources that can be used to meet demands of the top layers ('know- how') and the middle layer functions as a bridge or delivery mechanism between the purpose and resources ('know-what'). Most of the time this middle layer represents product development, which functions as a deployment method to meet customer and market needs. This results in a roadmap that often is in the format presented in figure 3. However, if other applications are aimed for the middle layer can represent capabilities, services, risk, systems or opportunities if more fitting to understand the delivery of technology to create benefits in the case at hand.

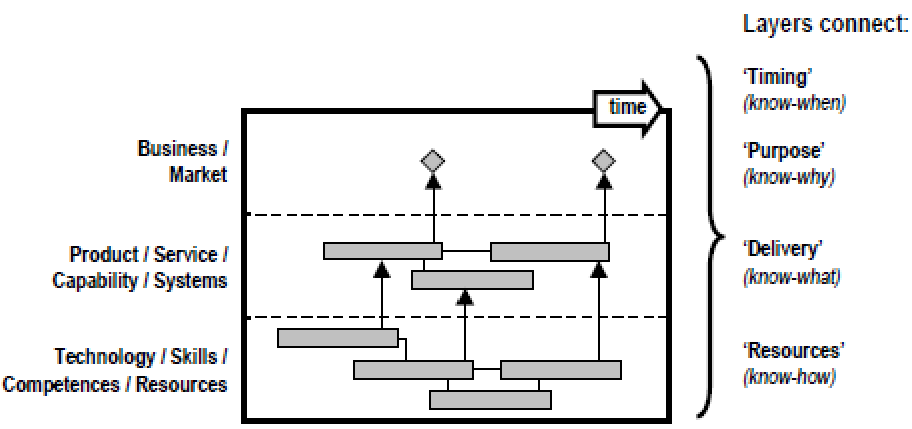


Figure 3. Generic technology roadmap. Adapted from Phaal, R., Farrukh, C., & Probert, D. (2001). *T-Plan: the fast-start to technology roadmapping: planning your route to success*. University of Cambridge, Institute for Manufacturing.

- Annotation: There is a possibility to store extra information in the roadmap that is not incapsulated within a layer, such as:
  - Linkages
  - Supplementary information
  - Other graphic devices

- Process: The process of roadmapping is different for every organization. As the process is contingent on many factors: resources (people, funding, time) are available to support the roadmapping process, characteristics of the issue at hand, available information and other ongoing processes and management structures within an organisation.

It is critical to assess *planning* when customizing a roadmap and the complementary process (Phaal et al., 2001). It involves clearly stating the process and business objectives. Then considering carefully how the generic roadmapping process can help to achieve these objectives. Roadmap ownership distributes itself over time in the organisation, starting with a single designated person or group, to the people participating in creation and eventually to a wide range of people within an organization as a communication tool. Aligning the business goals and context with the capabilities of roadmapping is important to achieve a proper roadmap process and structure. It could be helpful to appoint a designated person to manage the process and workshops, most preferably a person familiar with technology roadmapping (Phaal et al., 2001).

4.3 The Scenario Driven Roadmap approach

The scenario-driven roadmap process consists out of six phases divided over three layers, based on the T-Plan approach from Phaal (2001). The preparation, workshop setting and implementation layer. This specific roadmapping approach was developed to bring Scenario Planning into roadmapping, introducing plausible scenarios which should stimulate the ability of a roadmap to deal with uncertainty and more extreme outcomes. It is important to understand this variation of the roadmapping approach as it is used by ROSEN for which this research is conducted, but more importantly because it facilitates more extreme outcomes with higher variation, which makes classification more challenging. Below the process and resulting roadmap format of the scenario-driven roadmap approach is highlighted, explaining the stages of roadmap development and the resulting graphical roadmap.

<u>Preparation</u>

1. Preparing the workshops

This phase requires the forming of a project team that guides the development of the roadmap and the preparatory actions for the workshops. This team should (at least) exist out of an employee who possesses knowledge on the organisation, members with diverse backgrounds and analytical skills and an external or internal expert on (strategic) innovation and scenario planning that acts as a facilitator. In dialogue with senior management this team defines the scope of the business roadmap, it designs the layout, agrees on the workshop schedule and determines the various analyses required in the process. Complementary it selects, informs and prepares the workshop attendees. These attendees should represent strategic and technical levels to ensure broad knowledge, commitment and diverse views that lower bias.

This first phase results in workshops that are prepared properly and are able to provide useful results.

<u>Workshop setting</u>

2. Analysing the current situation

Currently the offerings of a firm and the market demands are supposed to be matched, however it is questionable that these offerings are still marketable in the future, as the market demands are uncertain and likely to change. This boils down to the question which markets are going to be important and what the market demands are going to be. The key thing to understand are the factors that shape this market demand, the driving forces. This includes environmental elements, such as economic climate and social developments, and their interrelationships which are subject to change. As the world is highly likely to change differently than expected it would be foolish to only assume one direction in which these driving forces would change, as business is then just based on one view of the future. The driving forces are thus subject to state uncertainty. To tackle the problem and arrive at an overview existing out of a comprehensive set of driving forces and the state uncertainty the company is facing the driving forces need to be assessed on different environmental levels: macro, meso and micro.

Eventually at the end of phase two the company will have a set of its strengths and weaknesses, a set of driving forces and an overview of current activities and served markets. If the goal is to formulate a new corporate strategy, then the driving forces and opportunities and threats should relate to this strategy.

3. Exploring future business environments

The driving forces determined in the previous phase form the foundation for developing scenarios. This scenario planning enables exploration of various possible future states, enabling the ability to cope with the environmental uncertainty. Each driving force can have multiple alternative projections, economic growth vs economic crisis for example, which are used to develop various scenarios with basic scenario planning methodologies.

At the end of phase 3 this results in multiple scenarios that represent a plausible environmental future state.

4. Determining robust areas

Using the scenarios developed in phase 3 robust areas can be identified. As each scenario provides implications for the firm, it indicates possible responses. Although each scenario is based on different unique projections of the driving forces there will be implications that are more or less similar for each scenario developed. These shared implications derive from driving forces of which the future is certain or from a unique combination of projections in each scenario. These share implications form the basis for the business roadmap, decreasing the uncertainty surrounding the driving forces.

The shared implications are either an opportunity or a threat. To condense them into high level areas that can be further elaborated in the business roadmap a SWOT (Strengths, Weaknesses, Opportunities and Threats) Analysis is used. To avoid to complex roadmaps only a few of the areas will be included in the roadmap. Phaal and Muller (2009) recommended using a maximum of eight sub-layers per main layer, which Siebelink et al. (2016) followed. The areas identified are then separated into focus areas and preconditions. This is done so to aid a firm in covering all relevant future areas, but preventing it from focussing on the eye catchers only. Focus areas are those that enable the firm to differentiate itself and make money. While preconditions are required to be met in order to excel in focus areas, compete in the market and meet minimum customer requirements. Thus, preconditions should be met in order to survive and focus areas in order to flourish.

At the end of phase 4 there will be a list of robust, high-level focus areas and preconditions that are options to include in the roadmap. These need to be evaluated in order to select the most strategically relevant and promising will be included, taking into account a healthy ratio between focus areas and preconditions. This selection process can be aided by various criteria such as: *'consistency with strategy and scope for the roadmap, (financial) feasibility, uniqueness and inspiration, risks versus potential margins, consequences for the organization, clarity, and a robustness verification (indeed visible in all scenarios?'* (Siebelink et al., 2016, p. 231-232).

5. Designing the business roadmap

For each focus area and precondition, it then has to be decided which segments are going to be prioritized for the coming years. Aims are then set per segment and the key requirements of the segment in the future year are hypothesised. Finalizing this process the firm then decides on which products and processes it wants to develop or acquire in these segments, determining the chain of markets, products, capabilities and processes that are required to move from the current portfolio in year x to the desired future of year y. Doing so will decrease response uncertainty through discussing multiple options and consequences of each decision.

At the end of this phase 5 the business roadmap able to deal with uncertainty and based on robust high-level focus areas and preconditions is complete.

Implementation

6. Implementing the roadmap

The resulting roadmap needs to be implemented in the firm. To do so the firm needs to communicate the roadmap. Additionally, the roadmap need to be kept up to date to reflect events in the current situation. The development process of a roadmap is continuous, it needs updating, in order to provide flexibility and prevent inertia which could lead to the death of a company. The roadmap needs thus evaluation and if required improvement. The rate of this iterations should depend on the rate of change in the industry in which the firm is active.

This process is graphically depicted in figure 4. In figure 5. the chain of markets, products or processes for one focus area is illustrated, including the knowledge layers of why, what and how. Although this Scenario-driven roadmap is an advanced concept the infrastructure to execute it with is still low-tech basic processing in programs such as Microsoft Word and separate drawing tools. If the goals is to integrate the scenario-driven roadmap principle in strategic planning it could benefit immensely from being fully integrated in the processes within the company, unlocking easier altering of the roadmap, continuous development, increase accessibility and visibility throughout the whole firm. Strengthening its directive and communicative power.

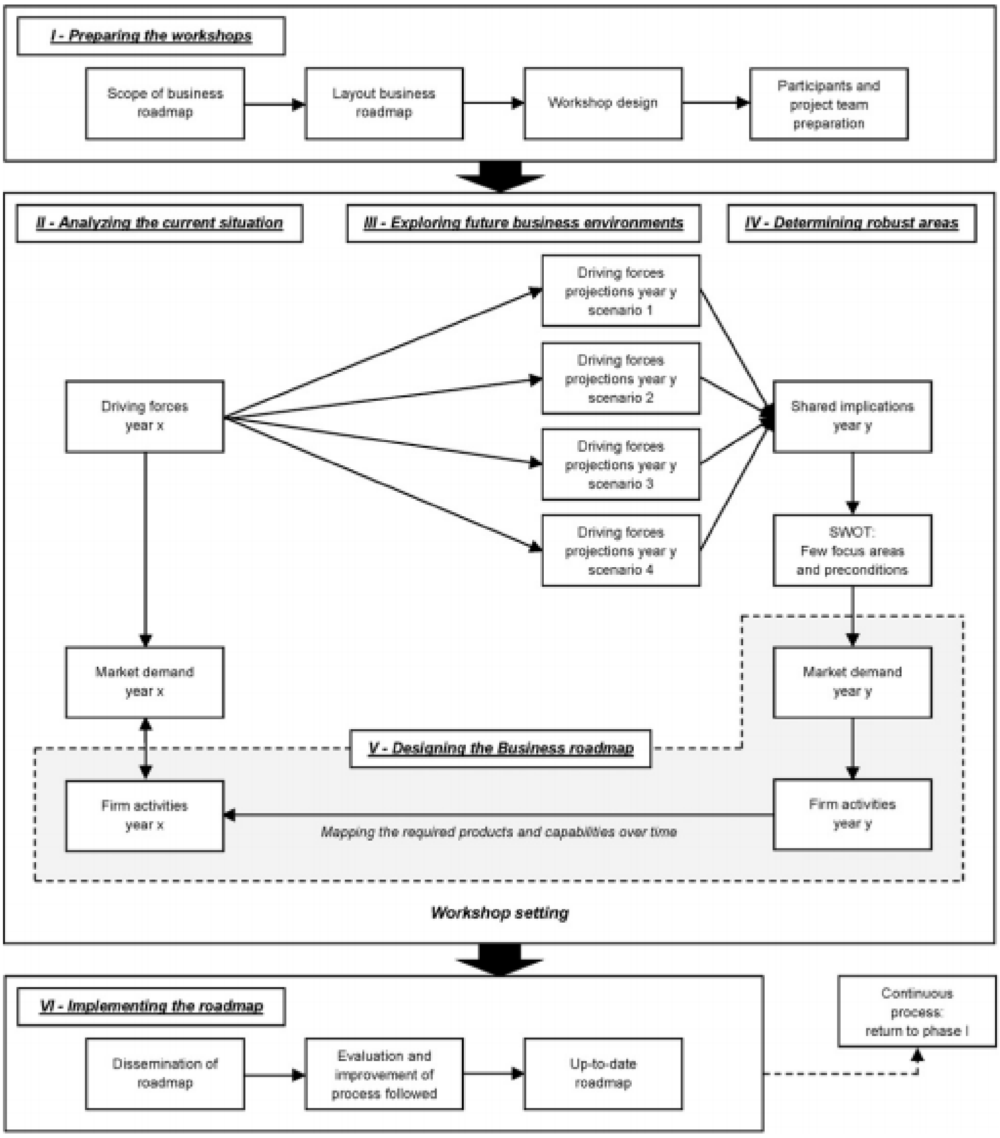


Figure 4. The rationale behind the Scenario-Driven Roadmapping approach. Adapted from Siebelink, R., Halman, J. I., & Hofman, E. (2016). Scenario-Driven Roadmapping to cope with uncertainty: Its application in the construction industry. *Technological forecasting and social change*, *110*, 226-238.

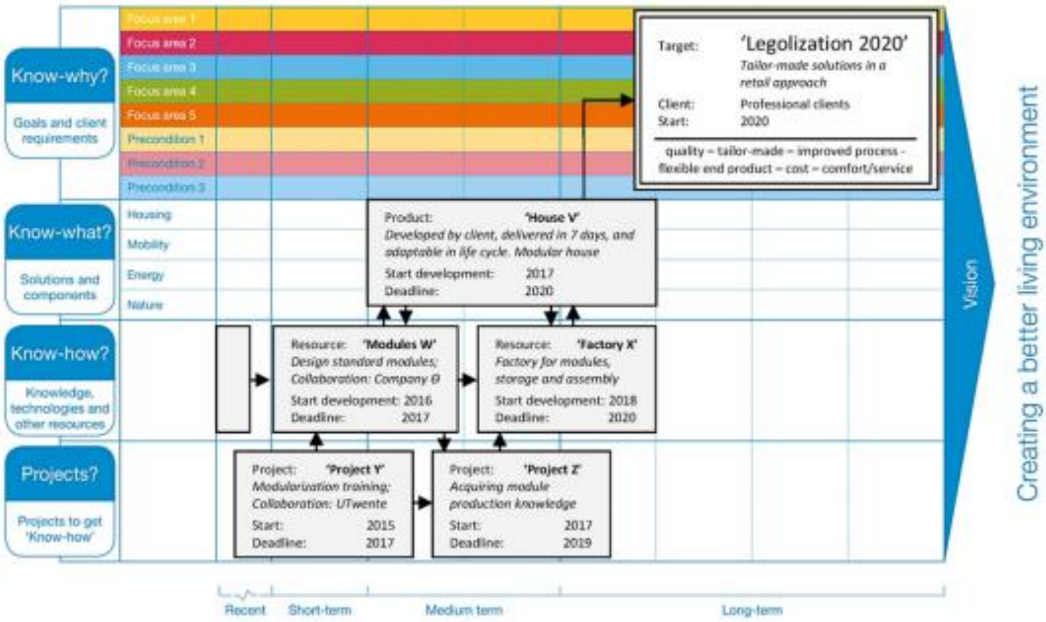


Figure 5. Illustration of a chain of markets, products or processes for one focus area on the business roadmap of Ballast Nedam. A complete roadmap will show various chains and their interrelations. Adapted from Siebelink, R., Halman, J. I., & Hofman, E. (2016). Scenario-Driven Roadmapping to cope with uncertainty: Its application in the construction industry. *Technological forecasting and social change*, *110*, 226-238.

4.4 Opportunities for machine learning

At the end of phase 4 in the Scenario-Driven roadmap approach a list of high-level focus areas and preconditions is developed. The selection of these on the criteria proposed by Siebelink et al. (2016) is a task that requires an expert or even better multiple experts that have extensive understanding of a firm. This part of data analysis has been selected to explore automated solutions for. If the workshop is performed with few people and the total length of the list would be 20 or so, then human coding works fine. However, if you would scale-up and perform workshops firm wide with over for example 1000 employees, that each provide 3 entries it results in 3000 entries to be evaluated. Resulting in a large time investment by high level manager(s), which is expensive and causes him/her to not be able to work on other tasks. Additionally, reading that many entries will probably fatigue a human, resulting in diminishing evaluation performance. Additionally, a person has his own beliefs and thoughts on what a business should pursue, so ideally you would need at least 2 independent raters to avoid biases in the selection process. So, it would be greatly beneficial if a machine could be used to relieve some or in an optimal world all human effort without deteriorating performance.

The criteria on which the outputs are judged also complicate automated processing of the roadmapping outputs, as they are firstly multiple. Secondly the criteria are not binary, using the example of a novelty criterium something can be extremely novel (no competitor or other firm has a certain technology yet), novel for the roadmapping firm or not novel at all. A simpler evaluation task would be to evaluate options in a binary good/bad fashion. This would increase the classification performance, however the usefulness of the classification would decrease. So there needs to be a balance identified between classification performance and the usefulness of the classification for further analysis.

Thirdly when designing a strategy to process all outputs from the roadmapping process it is important to realize that the roadmap is a communication tool and that its ownerships disperses through a firm, therefore having a data processing strategy that is transparent and supported by the stakeholders is critical.

As Phaal mentions planning is the most important considerations within the customization of a roadmap (Phaal et al., 2001). This can be extended to the planning of data processing. Up front it has to be decided what workable data formats are, which are easy to analyse but also work in a workshop setting (for example plain text files). And the structure in which they are organised. Additionally, categories or criteria developed are not easily altered when analysing with a computer, knowledge you gain on for example frequent keywords that indicate a certain category or score on a criteria are more or less locked-in. To change criteria or categories somewhere in the future makes a lot of the knowledge gained on set categories obsolete. Therefore, the decision on how to actually evaluate the options is critical as it should provide useful insights over an extended period of time.

Lastly, building upon this argument another characteristic of the scenario driven roadmap is the focus on aspects that are plausible to happen, scenarios are used to probe the future. Therefore, the generation of new options is likely. Truly new options are difficult to classify or evaluate, as assessment based on a comparison to previous options or firm activities is not or to a small degree possible. Therefore, when using the scenario driven roadmap it is especially important to have a broad and in depth understanding of the internal and external context of the organization for which the roadmap is being developed.

In addition to the part of data processing that is selected here, other areas exist within the Scenario Driven Roadmapping process that could benefit from automated solutions. Such as the scanning for trends to assist in creating a picture of the future business environment or the analysis of the current situation a company is in. These are however not focussed on within the scope of this research.

5. Automated methods to classify documents

Instead of manually processing all data and labelling them with categories or scores the goal is to use an automated process that predicts these scores or categories. Similar predictive models are used in a variety of domains, from sentiment analysis, medical diagnostics to news classification. These models are constructed from experience (Dreiseitl, & Ohno-Machado, 2002). The data can be expressed in a set of rules as used in knowledge-based expert systems or be used as a training set for machine learning models. This section will describe the different approaches that exist for classifying text data and their respective benefits and drawbacks. Furthermore, the way text is understood by a computer will be explained.

5.1 NLP tasks

As stated in the previous chapter the data collected in the roadmapping process is almost completely in a textual format, therefore the focus in this research is on predictive models that are able to deal with text/natural language.

Assessing work that has been done in the field of text analysis various fields of application can be considered: filtering of spam email, sentiment analysis (for example online reviews), patent analysis, social media mining, biomedical text mining among others. Different techniques are used to extract knowledge out of text, such as: Information Extraction, Text Summarization, Text Clustering, Dimensionality Reduction & Topic Modelling, Text Classification, Sentiment Analysis (Aggarwal & Zhai, 2013).

Text Classification seems to be the appropriate technique to use to categorize the ideas based on the criteria specified within the roadmapping approach. Additionally, clustering techniques could enable clustering into categories, without the need for labelled historic data. After which categories could be prioritized, resulting in the most promising categories being assessed first.

5.2 Machine learning based on word frequencies

Machine learning is considered an application of artificial intelligence. Enabling systems to automatically learn and improve from experience, instead of being programmed. Thus, it differs from traditional programming in terms of input required and the resulting output. See figure 6. Three basic steps of machine learning are: observe instances, infer on the process that generated the instances and this enables then the prediction of unseen instances (MIT, 2016).

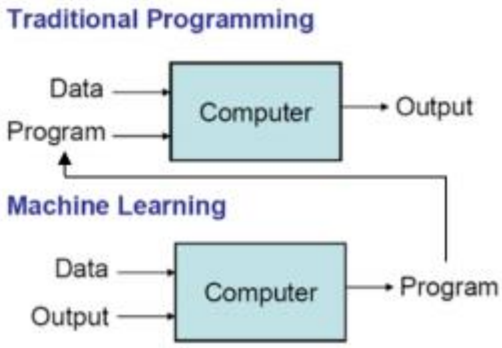


Figure 6. Machine learning vs. traditional programming. Adapted from MIT, 2016. Lecture 11: Introduction to Machine Learning. [Online] Available at: *URL https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/lecturevideos/lecture-11-introduction-to-machine-learning/*

5.2.1 Supervised Machine Learning

The two variations of machine learning are supervised and unsupervised learning. Supervised learning implying that the data is trained on existing data which has already a label in order to build a model to assess new data. It thus requires the acquisition of historic data, which is then cleaned and randomly split into two sets: the training set (70-80% of the data) and the testing set (20-30% of the data). A classifier, which is an algorithm used to identify the label an instance belongs to, is trained on the training set. After which the resulting model performance is assessed by letting the classifier (model) classify the testing set, after which the results of the classifier can be compared to the known historical label.

The classification is performed by a model that relies on the training data to learn and an algorithm that decides based on its training what the predicted class of a new instance will be. Different algorithms exist for classifying, each of them having respective benefits or drawbacks. Below the most common algorithms will be introduced with their respective benefits and drawbacks.

Decision Trees

Decision trees share some similarity with rule-based classification. It constructs true/false queries in a tree like structure, in which end nodes represent the categories and branches the connection of features leading from the root node to the end node. So, a document would start in the root node and travels along the branches of the tree to end up in a category. A decision tree is simple to understand and interpret, avoiding the black box that some algorithms cause. The tree however aims to classify on as few tests as possible, therefore performance degrades when the number of relevant features is relatively high. Additionally this could lead to overfitting, if you would for example classify political news to be about the United States and a decision tree uses as a first node the occurrence of the word Trump, this tree would perform poorly when the presidency of the USA has changed.

Random Forest

Random Forest is an ensemble learning method, using multiple randomized uncorrelated decision trees (Breiman, 2001). Each of those trees cast a vote on the class to which the test document belongs to, the most voted class will then be assigned to the document. This is called bagging (Breiman, 1996). The larger the number of predicting features is, the more trees need to be 'grown' in order to achieve good performance. Individual trees are highly flexible and thus prone to overfitting (Domingos, 2012; Sebastiani, 2002). To solve this the random forest thus combines the results of uncorrelated trees. Randomness and decorrelation are ensured by either randomly selecting training data subsets or random feature selection. The hierarchical structure of decision trees enables the learning of more complex feature interactions, modelling non-linear data and the automatic selection of features. Making it more suitable for situations in which context is important (Hartmann et al., 2019).

Naïve Bayes

The Naïve Bayes classifier is a simple probabilistic classifier (Yang, 1999). The classifiers first estimates $P(d|c)$ from the training documents, which is the class-conditional document distribution. Then it applies Bayes theorem to estimate $P(c|d)$ for the test documents. To compute the conditional probabilities efficiently the NB classifier uses a naïve assumption, assuming that every feature is independent. This assumption is seen as a reasonable trade-off between performance and computational costs (Hartmann et al., 2019). Research showed that NB even performs well in a situation with interdependent features (Domingos and Pazzani, 1997). The generative model is furthermore easy to explain and interpret (Netzer et al., 2019). NB being a generative classifier with inherent regularization is also recommended to use for smaller sample sizes, as it is less prone to overfitting if compared to discriminative classifiers (Domingos, 2012). A limitation to the NB classifier is the inability to model interaction effects that occur between features. Thus, it is more suitable for situations with strong signal words and simple relationships between text features and the classes in which they need to be classified in.

Support Vector Machines

Support Vector Machines are discriminative classifiers, using hyperplanes that aim to separate the training data by a maximal margin. Initially they were being developed as binary linear classifiers

(Cortes & Vapnik, 1995). However, by using kernel functions they can be used for nonlinear higher dimensional problems (Scholkopf & Smola, 2001). Their capacity to fit the training data is high, but compared to other classifiers with the same capacity SVM are less prone to overfitting and generalize better (Bennett & Campbell, 2000). The margin maximizing hyperplane is solely determined by the support vectors, other than providing the position of the hyperplane these support vectors carry little information (Bennett & Campbell, 2000). If the numbers of features and the sample size are large the computation of the hyperplanes can be costly due to being a convex optimization problem (Moraes et al., 2013). Effective examples of the application of SVM are available for certain text problems such as news categorization and sentiment prediction (Joachims, 1998; Pang et al., 2002). Which is not surprising due to the ability of SVM to deal with high dimensionality of data (Bermingham & Smeaton, 2010; Wu et al., 2008). However, by the limited information carried by the support vectors the SVM might be less able to model more nuanced patterns of the training data (Domingos, 2012). Which at the same time is beneficial as it results in less overfitting compared to more flexible methods such as neural networks or Random Forests (Hartmann et al., 2019).

The classifiers considered above are mostly used for what is known as 'traditional' machine learning. Which in the case of text mining/natural language processing means that they are applied in situations where word frequency-based Vector Space Models are used. Recently a trend towards classifiers based upon neural networks that outperform traditional machine learning has been developing, which will be introduced in section 4.4.

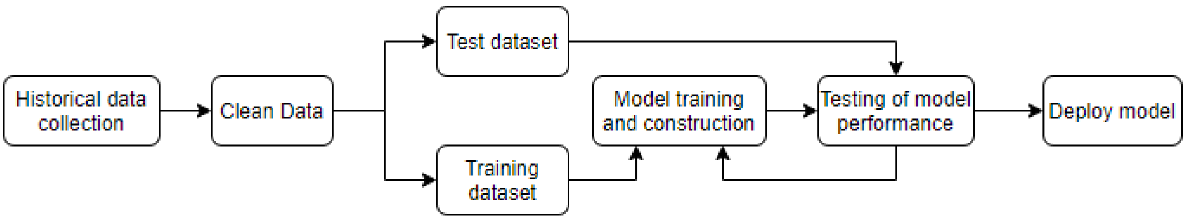


Figure 7. Supervised machine learning process illustrated.

5.2.2 Unsupervised Machine Learning

Unsupervised learning does not require training and assesses data without being trained on already known data. It aims to infer on latent features by clustering training instances into nearby groups (MIT, 2016). Clustering aims to minimize the dissimilarity of all clusters (C), thus being an optimization problem. The formulas below represent this problem, in which c represents a single cluster and e represents a single instance. Without incorporating the constraints of minimum distance between clusters or minimum number of clusters, the formula depicted in figure 8 (MIT, 2016) would provide a quite simple solution, as each instance would be a cluster, resulting in variability and dissimilarity of zero. The researcher thus has to specify the number of clusters he wants to extract. An unsupervised method is able to uncover latent relationships or categories overlooked in manual classification, downside being that there is no performance assessment from the environment possible (Suominen, Toivanen,& Seppänen, 2017).

$$Variability(c) = \sum_{e \in c} Distance(mean(c), e)^2$$

$$Dissimilarity(C) = \sum_{c \in C} Variability(c)$$

Figure 8. Clustering optimization problem. Adapted from MIT, 2016. Lecture 11: Introduction to Machine Learning. [Online] Available at: *URL https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/lecturevideos/lecture-11-introduction-to-machine-learning/*

K-Nearest Neighbour

A common algorithm used for clustering is the K-Nearest Neighbour algorithm. This algorithm works by following the following steps:

1. For each cluster a random centroid is selected
2. Distance of all datapoints to the centroid is measured
3. Datapoints are assigned to the closest cluster
4. New centroids for each cluster are found by finding the mean of all datapoints per cluster
5. Steps 2-3-4 are repeated until all points converge and the centroids stop moving.

Downside of the KNN algorithm that it uses all features in computing the distance, making it computationally expensive with large datasets (Aggarwal & Zhai, 2012; Sebastiani, 2002), additionally not relevant or noisy features degrade its performance considerably, requiring exponentially more examples to generalize when there are many features (Hartmann et al., 2019). Furthermore, a method using KNN requires as previously mentioned the number of categories to be specified, to determine this amount of categories a technique such as the elbow method can be used.

5.3 The Vector Space Model

Unlike humans computers cannot 'read'. Essentially computers are calculators and to let them work with text, the text needs to be transformed into numbers. In order for machine learning techniques (both supervised and unsupervised) to be used on text the corpus of text needs to be transformed into a Vector Space Model (VSM). The most basic approach to do so is by using the Bag of Words (BoW) model. The BOW model consists out of two components:

1. Vocabulary
2. Measure for the presence of words from the vocabulary

To illustrate how the BOW model works we take three example sentences about fruits and their colour:

1. The apple is red and a fruit
2. Bananas are a fruit and yellow
3. Peaches can have different colours

Using the BOW model, the example sentences will be converted into the following matrix:

Table 2

Example of a BOW Vector Space Model

	The	apple	is	red	and	a	fruit	Bananas	are	yellow	Peaches	can	have	different	colours
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

As word frequency is used as a scoring measure it means that some words could dominate a document, while not containing discriminative/informational content to the classification model as rarer class specific words. To compensate this, you can rescale the frequency scoring by the total occurrence of the word in all documents. This variation on the BoW model is known as Term Frequency – Inverse Document Frequency (TF-IDF).

- Term Frequency: frequency of a word in a document
- Inverse Document Frequency: a logarithmically scaled inverse fraction of the documents that contain the word.

These relatively simple featurization techniques are often used with success, but do impose limitations. First there is the sparsity, as the vocabulary is built from all occurring words in the sample, although the needed frequency in the sample for the word to be included into the vocabulary can be specified. Still the fraction of terms that one document will have in common with the complete vocabulary is very small, resulting in a sparse vector. Additionally, the semantic meaning of words is lost, thus a document with alternating word usage but the same semantic meaning will be mapped to a completely different vector (Zhao & Mao, 2017). As seen in our example with sentence 3, which is obviously about a fruit and considers the colour of a fruit, but does not have any similarity according to the BoW model to the other two sentences. This could also result in two completely opposite statements being seen as very similar. Additionally, out of vocabulary words in the set you use your trained model on will not be considered when classifying new documents. So, representing human language, which all its subtle differences and the huge potential vocabulary that can be used by people is difficult.

5.4 Neural networks

Overtime different approaches have been taken to overcome the weaknesses of the BoW model, currently the state-of-the-art models within natural language processing are the Transformer models that are based on neural networks. Natural data in its raw form was always difficult to process for conventional machine learning techniques. To construct a machine learning system required considerable domain knowledge and careful engineering to develop a feature extractor that transformed raw data into a feature vector from which a learning subsystem could identify or classify patterns in the input (LeCun et al., 2015).

Deep learning is inspired by how the human brain works. Neurons, which connect to the input layer learn patterns inductively from the training data to make predictions on test data (Efron & Hastie, 2016). The most basic form exists out of one input and output layer. With computational progression the ability to include more layers in between, so called hidden layers, was acquired (LeCun et al., 2015). The number of nodes in the hidden layer is dependent on the complexity of the task (Detienne, Detienne & Joshi, 2003).

Text classification has benefited from deep learning architectures due to their potential to reach high accuracy with less need of engineered features. Deep learning enables the learning of more subtle differences in text (Hartmann et al., 2019). But deep learning algorithms require much more training data than traditional machine learning algorithms, the exact number of tagged examples varies greatly per task. Would a deep learning model be used to detect if squares are white or black, then only a few examples would suffice, recognizing if the picture is of a dog or a cat is already more difficult and requires more data. In general the more high dimensional and sparse the classification problem is, the more training data is required. In most applications the required training examples would rapidly increase to multiple thousands. The problem however is that most downstream tasks do not have thousands or more of tagged examples.

To bridge this gap researchers focussed on general purpose language representation models using the surplus availability of unannotated text on the web, which is known as pre-training. This pretrained model can then be fine-tuned for a task specific application with a small dataset. One of the latest state-of the-art Transformer models based on this principle is the BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) model from Google. It delivered state-of-the-art results on different NLP benchmarks and is open source available. So, what makes BERT perform so well? Its basis is in recent work on unidirectional contextual representations: Semi-supervised Sequence Learning, Generative Pre-Training, ELMo and ULMFit. What makes BERT different is that it is the first deeply, unsupervised, bidirectional language representation, using only a plain text corpus for pretraining (which in the case of initial BERT release was Wikipedia) (Devlin et al., 2018).

The foundation of these models can be found when context was first introduced to NLP tasks. Harris distributional hypothesis (Harris, 1954) states that words that appear in a similar context have similar meaning. On basis of this hypothesis more advanced approaches than the BoW model were developed to form word representations. Methods involve grouping words into clusters based on the context they are in (Brown et al., 1992; Uszkoreit & Brants, 2008) and representing words as high dimensional sparse vectors in which each entry is an association between a word and a context (Turney & Pantel, 2010; Baroni & Lenci, 2010). The focus shifted towards representing words as dense vectors by leveraging knowledge from the field of neural networks, a concept known as word embeddings. With the introduction of the Word2Vec model in 2013 by Mikolov (Mikolov, 2013) new opportunities were created for introducing semantic relationships between words. With the Word2Vec model the development of so-called Word Embeddings/Representation models was popularized. With the Word2Vec model pretrained vectors were created and open source available to use.

The way BERT differs from those initial context-free representations is that context-free models such as Word2Vec or GloVe create a single embedding for each vocabulary word. Thus, the word ‘bank’ will have the same representation in ‘I sit on a bank’ and ‘I deposit money on my bank account’. Contextual models do consider context and generate a representation for each word accordingly (Devlin et al., 2018). A unidirectional contextual representation uses “I deposit money on” for representation, but ignores ‘account’. BERT does take into account the whole sentence when forming a representation.

This seems to be a simple concept, considering the context before and after a word, but BERT is the first to use this bidirectional concept. Because unidirectional models are trained efficiently by predicting each word conditioned on the previous words within in a sentence. This is however not possible for bidirectional models, as conditioning it on the preceding and next words would result in the word that is being predicted being allowed to “see itself” in a multi-layer model. Google’s researchers avoided this problem by masking out some of the words in the input and condition each word bidirectionally to predict the words that are masked (Devlin et al., 2018).

BERT is part of the family of models known as Transformer models, which revolutionized natural language processing when they were introduced in 2017. Development and progress of these Transformer models is so fast that it is difficult to assess which iteration/improvement is most meaningful and how effectively models can be combined (Roberts & Raffel, 2020). In 2019 alone there was a huge development of new transformer-based methods such as Reformer, RoBERTa, ALBERT, XLNet, and MT-DNN. All these models provide state-of-the-art results on several major NLP benchmarks which can be found on https://gluebenchmark.com/.

These models provide great opportunity for Transfer Learning, leveraging knowledge on language from huge data sources, which makes a model truly understand language more or less, into task specific applications with little training data available. However, understanding, fine tuning and successful application requires excellent knowledge on these models and computational power. Beneficially most models are made open source by the big tech companies such as Google that develop them.

5.5 Computer Aided Text Analysis

Machine learning models are not the only techniques used to extract information out of text. Computer Aided Text Analysis, also called content analysis, uses dictionaries with keywords to extract information. It is a simpler approach focussed on word occurrences and leveraging scientific theories/constructs. Making it a rule-based system that classifies based on manually selected word occurrences.

Short et al. propose two approaches when using computer aided text analysis: deductive (similar to supervised learning, with known concepts from theory as categories) and inductive approach (deriving labels from data, similar to unsupervised learning). They argue in favour of a deductive approach, which however needs theory in order to design the coding scheme. Making it difficult to use this method on fields that have no existing dictionaries (Short, Broberg, & Cogliser, 2010). As Krippendorf stated: *‘Most content analyses would benefit from the construction of special purpose dictionaries, but developing a dictionary from scratch can be a formidable task. It is not surprising, therefore, that content analysts usually try to build on available dictionaries before they* *attempt to develop their own'* (Krippendorf, 2018, p. 251). To create such a custom dictionary, it is recommended to first create a word frequency list and examine the key words and phrases (Neuendorf, 2016). This suggestion by Neuendorf however implies an inductive approach which focuses on a particular narrative instead of an independent developed deductively generated dictionary (Short et al., 2010). A recommended procedure for an inductive approach is given by Short et al.:

Inductive content analysis

'1. Identify commonly used words from narrative text of interest using DICTION or other CATA software

2. Identify or create a working definition of the construct of interest to guide word selection

3. Identify words that match the construct of interest

4. Establish initial interrater reliability

5. Refine and finalize word lists' (Short et al., 2010, p. 327)

They also provide an approach for deductive content analysis:

Deductive content analysis

'1. Create working definition of construct of interest (use a priori theory when possible)

2. Initial assessment of construct dimensionality based on existing literature

3. Develop an exhaustive list of key words from the formal definition to capture the construct of interest. (If the construct is hypothesized to be multidimensional, multiple discrete word lists should be created for each subdimension)

4. Validate word lists using content experts and assess rater reliability' (Short et al., 2010, p. 327)

5.6 Software

Computer aided textual analysis and machine learning requires the selection of software to execute it with. To have control over the process and to be accountable for each step the use of a programming language is preferred over a tool. The most obvious choices are R or Python, due to their wide application, user friendliness, supporting communities and being open source. As the main goal of this research is to cluster/categorize findings R is suitable, as it can pre-process, associate, cluster, summarize, categorize and has an API (Application Programming Interface, enabling extensions via plug-ins) (Meyer, Hornik,& Feinerer, 2008). This also true for Python as it can do the aforementioned things and has been successfully implemented in research using text mining techniques. The preference in this instance is on Python, due to being known to the organization with whom the Scenario-Driven Roadmap was developed, simplifying the communication of the process and results. Would neural networks be used then it is also advised to use Python, as most of the natural language processing methods based on them are in Python.

<u>NLP Python Libraries</u>

Within the Python community and environment, the most common used term is Natural Language Processing and the various libraries that provide the tools needed to execute text mining are referred to as NLP libraries. This section will discuss the NLP packages available and compare their features.

TextBlob

Textblob provides a simple API that enables common NLP tasks such as part of speech tagging, sentiment analysis, noun phrase extraction, classification etc. Downside is it speed and the relative basic features it offers.

Spacy
Together with NLTK one of the most popular libraries for NLP. It differs from NLTK by only providing on algorithm: the-state-of-the-art. It is specifically designed for production use. It can build natural language understanding or information extraction tools and pre-process data for deep learning. Another benefit of Spacy is its speed compared to NLTK and CoreNLP.

Gensim
Python library focussed on similarity retrieval, topic modelling and document retrieval. It is able to work with large corpora (collection of documents). The most specialized library listed here. There are a lot of examples using Gensim and it looks to be a promising library to use in this research.

Polyglot
Similar to Spacy, focussed specifically on multilingual applications. Especially useful if Spacy does not support a language one is working with.

CoreNLP
Stanford CoreNLP is able to identify the base form of words, recognize entities, normalize dates, times, and numeric quantities, parts of speech, mark up the structure of sentences regarding syntactic dependencies and phrases, noun phrases that point to the same entities, sentiment indications, extract open-class or particular dependencies between entity mentions, quotes people said, e.a. CoreNLP is in Java, it could be used in Python using a wrapper, this however reduces the speed.

Natural Language Toolkit
Probably the most popular platform used for NLP related problems. It can handle over 50 corpora and integrates lexical resources such as WordNet and text processing libraries suitable for classification, stemming, parsing, tokenization, semantic reasoning. It also provides wrappers to achieve industrial strength NLP libraries. Drawbacks are its speed and complexity, resulting in a steeper learning curve.

5.7 Summarization of chapter 5
This chapter highlighted the principles of machine learning and how they can be used to automatically assign labels or cluster documents, additionally Computer Aided Text Analysis (CATA) has been introduced as an alternative in which the computer assists but a human rater develops the rules for classification instead of the computer.

Furthermore, in this chapter the transformation of text into a Vector Space Model to make it suitable for a computer to process and the basic machine learning algorithms to automatically classify or cluster have been analysed. The drawbacks and benefits of each VSM and algorithm are summarized in table 3 and 4.

Table 3

Comparison of word representation methods

Model	Advantages	Limitations
Bag of Words	• Easy to implement • Similarity is easily computed between two documents • Understandable and transparent	• Order of words is lost • Unable to capture semantic meaning • Unable to coop with synonyms • Common words effect the results
TFIDF	• Easy to implement	• Order of words is lost

	• Similarity is easily computed between two documents • Understandable and transparent • Common words do not affect the results as strongly due to IDF	• Unable to capture semantic meaning • Unable to coop with synonyms
Word embeddings	• Able to deal with synonyms • Leverage general embeddings generated from for example Wikipedia to specific tasks • Able to deal with out of vocabulary words	• Words with double meanings such as bank are assigned the same vector • Large dataset required to generate embeddings for a specific task • Higher user barrier, more advanced knowledge required
Contextualized Transformer	• Able to incorporate context • Differentiates the same words based on their context • First text representations that equals human classification performance on benchmarks	• Development rates are high, difficult to assess what the dominant algorithm will be • New, so less examples and knowledge available on how to apply it properly • Difficult to interpret and understand • Large datasets required to generate embeddings for specific tasks

Table 4

Comparison of Machine Learning Algorithms

Model	Advantages	Limitations
Naïve Bayes	• Easy implementation • Computational inexpensive • Easy to understand and implement • Less prone to overfitting • Suitable for smaller sample sizes	• Unable to coop with interaction effects • Limited ability to classify more complex documents • Strong assumptions on the shape of data distribution
Support Vector Machine	• Can model non-linear decision boundaries • Less prone to overfitting problems	• Kernel choice requires manual expertise • Lack of transparency in result due to low

		informative value of support vectors • Limited ability to incorporate nuances compared to more flexible methods
Decision Tree	• Easy to understand and visually inspect • Transparent	• Struggles with out of sample predictions (if there is no similarity to a decision node) • Prone to overfitting
Random Forest	• Variance reduced compared to decision trees • Able to model more complex interactions • Automatic feature selection	• Not easy to visually interpret • Prone to overfitting • Requires manual selection of number of trees
Deep Learning	• Less need for engineered features • Enables leveraging larger datasets to specific tasks • Possibility to make context relevant and identify semantic meaning	• Large amounts of data required • Computationally expensive • Difficult to design a fitting architecture, experts are needed • Black box, not transparent how classifications are made • Large development effort/cost
Computer Aided Text Analysis	• No black box • Computational inexpensive • Does not require a VSM • Great control and learning from constructing dictionaries	• Need for distinct words for a category • Requires manual updating • Synonyms need to be identified
K-Nearest Neighbour for clustering	• Suitable for short texts • Easy to understand the algorithm	• Need for manual selection of clusters • Curse of dimensionality: requiring large training sets to generalize for many features • Computationally expensive for large high dimensional data sets

6. Methodology

6.1 Dataset

The output from the workshop phase of the Scenario Driven Roadmap process consists out of strategic options that describe three different knowledge layers: what, why and how. The what layer describes what service/product/process a company should develop, for example: *develop security for smart products*. The why layer describes why it should be done, continuing the example the why statement could be: *to protect the integrity of the data generated by smart products.* The how layer then describes which capabilities are required for development: *work together with the existing supplier of IT security.* For classification a dataset is available that covers one of these three knowledge type layers generated within the roadmapping process: the what layer. This layer describes the business activity that ROSEN should pursue, a typical example of a what option would be: *'Change the strategy from almost only the oil and gas market. They need to invest in more upcoming markets like telecommunication and manufacturing.'*. This data has been collected using workshops in which students of the University of Twente participated during the research on developing a Scenario-driven roadmap. The workshops were held over two days, some students were provided a Scenario Analysis and others a SWOT analysis, based on these and a short introduction of ROSEN they generated the strategic options on the what, why & how layer. The options are thus not generated by ROSEN employees or industry experts. For the what layer this resulted in 384 strategic options.

The options are written down in one or two sentences with an average length of 24,8 words. Not all options limit themselves to a description which is fitting for a 'what' layer. The following options: *'Introduce measurement systems to measure flow and usage of gas and oil, as they are not being developed by other companies.'* & *'ROSEN Group has to focus on other business models because fossil fuels are at the end of life.'* illustrate this, as it also answers *why* ROSEN should do this.

This mixing of layers should be taken into account in classification as it causes ambiguity, for example the second sentence describes 'other business models' which indicate a move away from fossil fuel activities, but at the same time contains the word fossil fuels which signals that the option is about doing something with fossil fuels.

Table 5

Descriptive statistics of the data sample

Number of options	Mean	Std	Min	25%	50%	75%	Max
384	24,8	13,5	4	14	22	33	105

The what layer already has been manually classified on novelty and value on a scale ranging from -4(low) till 4(high). These scores were defined manually during the classification process, this was done so by three experienced ROSEN managers. Based upon their knowledge of ROSEN's current activities, technologies, knowledge and business environment an assessment was made on the degree of novelty and value.

For this research the data was split twice for each construct, once on the 0 point, treating everything with a value higher than 0 as either novel or valuable. The data were also split at the -2 and +2 score, creating four categories that are low/mediocre/moderate/high value or novelty. For each of these splits a word frequency list was generated after removing stopwords. For each split there has been checked for duplicates twice using two frequency thresholds (minimum frequency a word has to have to be taken into account, adapted to fit the sample size and provide a comprehensible overview), to check if the most frequent occurring words are occurring between the splitted samples and if the words in a more broader frequency range are duplicate between samples (appendix 1-4).

Table 6

Observations of the data when splitted for different criteria values

Criterium	Split	Observation	Frequency Threshold	Appendix
Novelty	0	Off all words occurring at least ten times in both samples there is large overlap. The most frequent terms occur in both samples apart from the words [redacted]	10	1
Novelty	0	Now even more overlap occurs and the first terms that are unique appear around the count of 14.	5	1
Novelty	-2 +2	The words [redacted] are the only unique words.	5	2
Novelty	-2 +2	Only [redacted] seem to remain as a signal word for the not novel group.	3	2
Value	0	From the most frequent words [redacted] seem to differ	10	3
Value	0	Similar to the minimum frequency of 10, [redacted] however is now a common word	5	3
Value	-2 +2	[redacted] appear only in high value options, while [redacted] are only in low value options.	5	4
Value	-2 +2	Similar but now [redacted] are shared between samples.	3	4

6.2 Method

Current manual classification in Roadmapping relies on two different aspects, firstly there is the classification in categories that form an umbrella for strategic options, example of those categories could be *pipeline inspection* or *digital transformation*. Doing so enables an overview of trends and structures the data, making it easier to retrieve, analyse, curate and annotate documents. Secondly all options are evaluated based on (or a selection of) the criteria as proposed by Siebelink et al. (2016): *'consistency with strategy and scope for the roadmap, (financial) feasibility, uniqueness and inspiration, risks versus potential margins, consequences for the organization, clarity, and a robustness verification'* This classification is an evaluating step that by scoring the options assesses the quality of each option, to filter out those options that are considered of high quality and enable an assessment on how effective the roadmapping process was in generating high quality strategic options. Both the categorization and the evaluation of options will be experimented with to discover if they can be (partially) replaced by automated processes. For the evaluation of options, the focus will be on the value and novelty criteria. For each technique it will be indicated if it is used for classification based on criteria or categories.

First exploratory data analysis is executed. Exploring the words occurrences and descriptive statistics of the sample. The most frequent uni, bi and trigrams are extracted and reviewed. Considering the data exists out of 1 or 2 sentences per document these N-grams should be representative for the main topics of all documents. Additionally, the sample is pre-processed to remove noise and make it suitable for a computer to process:

- Removal of interpunction
- Lower case conversion
- Removal of special characters (&, ; etc.)

After these steps the data input for machine learning clustering and classification algorithms requires some extra pre-processing. The first three steps listed below are used to reduce the number of features.

- Removal of stopwords using the NLTK stopwords list. Additionally, the words ROSEN & group were removed.
- Stemming (using the PorterStemmer)
- Lemmatization
- Transforming the options into simple Bag of Words and TFIDF vectors

More advanced models and concepts such as deep learning methods do perform better in theory and have the potential for use on downstream tasks using transfer learning. However, they are complex and require excellent knowledge and a large time investment to make them work. Therefore, the choice is made in this research to focus on dictionary and word frequency based- machine learning techniques and word representations. Because the majority of firms that decide to create a roadmap do not have advanced knowledge and experience with natural language processing. Additionally the selection of ideas to put on the roadmap requires transparency. A deep learning approach would make the process a black box, removing some of the learning experience that comes with processing data and feature extraction. Additionally, the complexity and knowledge required to downstream state of the art methods based on deep learning principles is a challenging task and requires a large investment of time, money and manpower. Thirdly within the field of deep learning surrounding textual applications are so fast at the moment (Roberts & Raffel, 2020) that if you would research one, it would probably be outdated once you are done. Once developments slow down a bit and a dominant model is surfacing more guides and examples on how to effectively downstream a deep learning model will be available. Therefore, if a deep learning approach was taken now, it would be difficult to apply it in other roadmap cases.

Thus, the choice for the Bag of Words and TFIDF approach is made due to the low complexity and understandability of the word representation generation, making them suitable to easily be adopted and finetuned for a specific case. Additionally, a dictionary approach is used, in this approach there is no black box and the choice for keywords need to be made by a person. Either deductive, leveraging knowledge on categories from for example ambidexterity literature (March, 1991). Or inductive by deriving keywords for classes from the data itself. A machine learning algorithm more or less works the same as this inductive approach in the sense that it learns from training data what words or word combinations appear predominantly in a different classes, based on this knowledge an algorithm then predicts unseen classes. In situations with distinct categories and little data a human is likely to match the ability of a computer to recognise patterns/keywords in text.

After pre-processing different techniques are experimented with to see which is most promising to replicate the performance of manual classification or aid manual classification. The method section is divided in to two parts, the first describes Computer Aided Text Analysis, the other the machine learning classification.

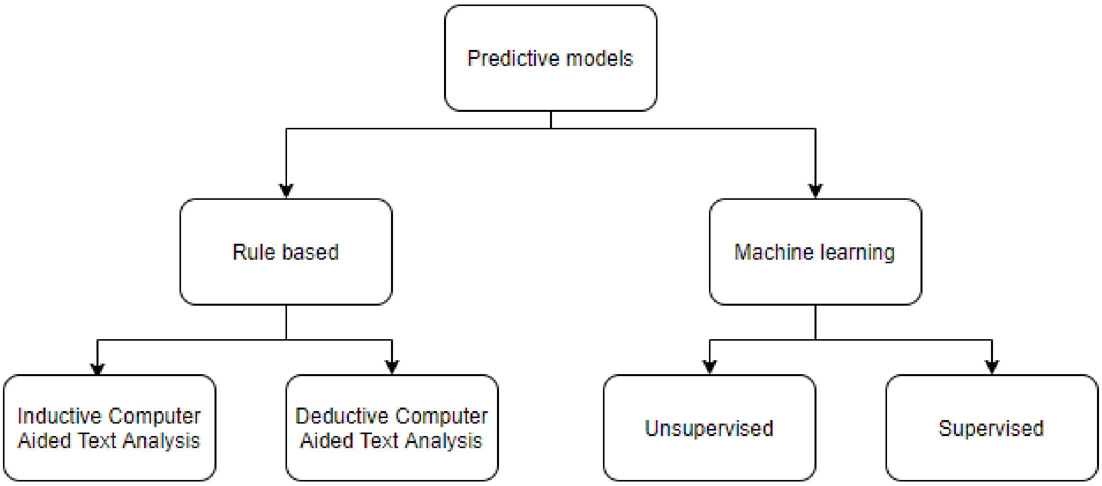


Figure 9. Overview of the predictive models used.

6.2.1 Computer Aided Text Analysis Experiments

Inductive - Categories

The approach here is an inductive approach in which frequent occurring N-grams within the sample are identified and those are used to form pre-defined logical categories with corresponding keywords in which the strategic options can be categorized. These categories will then be manually validated to see if they indeed form a logical group and the novelty and value scores between categories will be evaluated, to see if the categories formed differ on these values, providing an indication that the categories are indeed distinct from one another.

Deductive - Criteria

For Computer Aided Text Analysis a deductive approach is taken in this case. To do so the knowledge and research on ambidexterity is used. Within the research on ambidexterity researchers have identified words that are associated with exploration and words that are associated with exploitation. It is expected that words that are linked to exploration correlate with high novel options and exploitation words vice versa with low novel options. The benefit of using such generic concepts is that it potentially can be applicable (1) to a wide range of industries, (2) over extended periods of time and (3) to a broad scope of corporate actions (Uotila et al., 2009).

Defining Exploration vs. Exploitation

The work of Uotila et al. (2009) describes their efforts to identify relative exploration orientation of companies from news articles published about those companies. For their operationalization of exploration/exploitation they used the conceptual definition of March (1991). Defining exploration as: ‘things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation’ (March, 1991, p. 71) and exploitation as: ‘such things as refinement, choice, production, efficiency, selection, implementation, execution’ (March, 1991, p.71). Based on this conceptual definition a dictionary was derived. This was then used for simple word count analysis, several studies have shown that a simple word count analysis, counting words, is able to produce similar results to more labour intensive context-dependent manual or computer assisted coding (Laver, Benoit & Garry, 2003; Porac, Wade & Pollock, 1999). They found in explorative articles 3.0/1000 exploitation words and 1.5/1000 explorative words & in explorative articles 4.5/1000 explorative words and 1,6 exploitative words. As the documents in the sample are short for this research the relatively small the dictionary of Uotila et al. is not the only dictionary used, also the more elaborate dictionaries on ambidexterity of Moss et al. (2014), McKenny et al. (2018) and de Visser et al. (2017) are tested for their potential usefulness (see appendix 5 for all dictionaries).

Using the previously developed dictionaries on ambidexterity a simple word count analysis is executed to explore if a distinct difference in exploration/exploitation terms can be discovered between the novel and not novel labelled options. These will be placed into buckets with an interval of 2 score points, resulting in four buckets of low novel, not novel, novel and high novel.

Additionally an analysis is run using CATA with a dictionary received form a ROSEN manager, using the words: oil, gas, pigging & pipelines, which he uses for scanning routines in order to discover what is going on in those fields. These are thus terms that are expected to correlate with not novel options as ROSEN is already active in those fields.

6.2.2 Machine Learning Experiments

<u>Clustering - Categories</u>

Within this part first a method to detect categories into which the data may be categorized is developed. Explicitly this means for machine learning that a clustering algorithm will be used. This is an unsupervised method that detects clusters without the need for labelled data input. The researcher however has to set the number of clusters to be formed. When clustering with the K-Nearest Neighbour algorithm a common approach to determine the number of clusters is the elbow method. The number of clusters were the distortion flattens (an elbow shape is formed) is most likely a good representation of the true number of clusters.

<u>Supervised Classification - Criteria</u>

Classification based on value and novelty uses the same dataset and each options is assigned a novelty and value score between -4 and 4. A hard split is made to classify everything above 0 as novel/high value and everything with a score of 0 or less as not novel/low value. The choice to focus on word frequency techniques require word similarities between options with the same label. As observed in the data the words that occur frequent show high overlap between different novelty/value samples and to preserve those words that are unique for certain parts of the sample it is necessary to make as large a split as possible. Thus, this hard split is in this case necessary in order to preserve enough data to train the model with. If the criteria would not be binary the available amount options and their length would provide a problem, as text is high dimensional, and the features needed for classification would be more difficult to extract. The lesser the amount of training options provided the model would perform worse with out of sample data and generalizability would be difficult.

For supervised machine learning different classifiers will be used:

- Naïve Bayes
- SVM
- Decision tree
- Random Forest

The BoW and TFIDF features will also be transformed into their principal components and plotted in a scatterplot to observe if the categories provide an indication to be distinct. To evaluate the results the resulting confusion matrix will be evaluated on accuracy, precision and recall. Accuracy is a quite straightforward measure, but it could also be misleading. If you would have an unbalanced sample and classify everything as the dominating instance you could still achieve 80% accuracy for example. To tackle this precision and recall are also measured. Precision and recall require a designation of the positive and negative class, in this case high value/novelty is considered the positive class as those are the options that you are interested in to discover. The graph below accurately presents what recall and precision entail. Recall being the number of relevant items selected out of the true amount of relevant items. Precision measures how much of the items selected are actually relevant. High precision would mean that you have little false positives, while high recall means that you have accurately identified most true positives and have little false negatives.

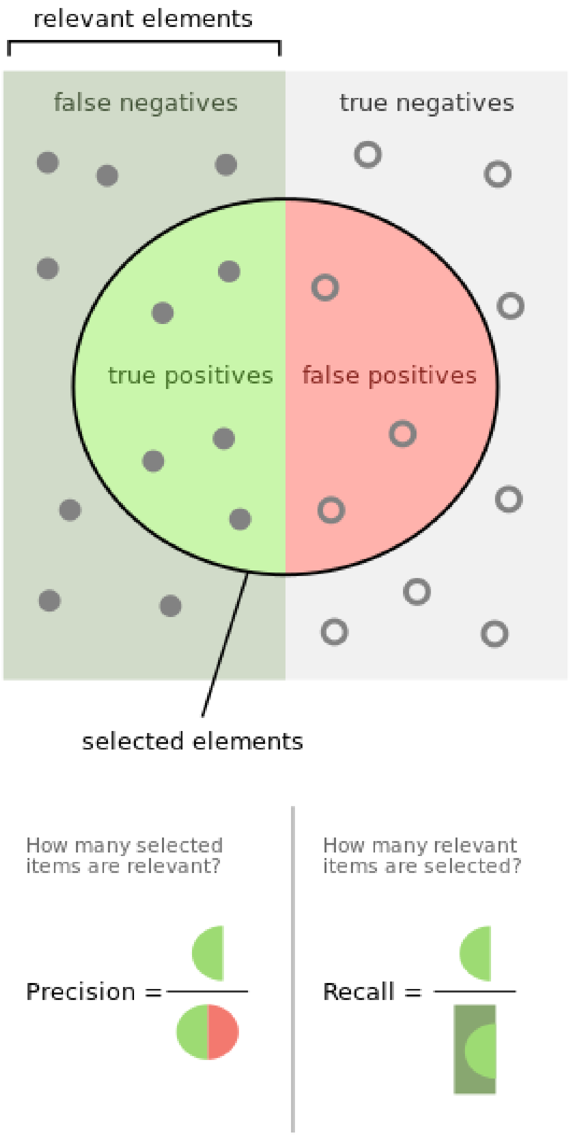


Figure 10. Precision and Recall explained. Adapted from *Wikipedia*. Retrieved May 15, 2020, from https://en.wikipedia.org/wiki/Precision_and_recall

7. Results

7.1 Explorative Data analysis

As a first step to explore the data the descriptive statistics of each group are extracted and a word cloud is made to visualize the most frequent words appearing in the total sample. The word cloud exists out of stemmed and lemmatized words.

Figure 11. Word Cloud of the complete dataset.

Table 7

Descriptive statistics for different sample splits

Group	Number of options	Mean word count per option	STD
Total	384	24,79	13,49
Novel	208	24,84	12,47
Not Novel	176	24,66	14,63
High Value	213	25,31	13,40
Low Value	171	24,08	13,60

The simple descriptive statistics do not show any significant or obvious deviations in word counts or standard deviations. Also in the case of non binary data splits the mean word count is approximately 24-25 and the standard deviation is around 13,5 for each sub sample. It is thus not the case that the length of options possess a certain indication.

Observing the results when the sample is binary split on novelty and value in table 8 (for all subsamples and both frequency thresholds see appendix 1-4) it shows that a lot of the vocabulary is shared when looking at the words that occur often or to a medium degree, which is also the case for other subsamples. This is problematic as the groups would ideally consist out of distinct vocabularies, as you need these differences to classify on. Especially as no unique bi or trigrams seem to be present, classification will be challenging, due to the lack of obvious discriminative features available. Unique words using a binary split at 0 are mostly found at frequencies lower than 10, when splitting the sample then in a training and testing set only part of that 10 will remain as a feature. If a word occurs for example only 3-5 times in the whole training sample it provides a weak signal for classification. Additionally if you dive manually into assessing differences on those low frequent words, than it would make more sense to make the whole classification manually. This means that the generalizability and use for out of sample classifications will be difficult.

Table 8

Unique Ngrams between sub samples

Novel Words		Not Novel Words		Value Words		Non-Value Words	
Word	Freq	Word	Freq	Word	Freq	Word	Freq
energi	110	oil	47	energi	71	energi	58
renew	77	ga	47	ga	63	invest	52
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]			[redacted]	[redacted]		
[redacted]	[redacted]						
[redacted]	[redacted]						
[redacted]	[redacted]						
[redacted]	[redacted]						

Note. All shared words are marked red. The minimum threshold in this table was 10, for a threshold of 5 see appendix 1,3

In addition to these scores, the options were also manually categorized in different categories (for example renewable energy), which are unknown to me, to not bias automated classification into categories. Such categories are better fit to generate without historic labels then the novelty/value scores as it is less abstract. Although less informative then value and novelty, being able to generate a number of overarching categories is still valuable to quickly identify options surrounding a topic of interest or to make sure that options are quickly send to business units that are busy working on related topics. As explained it would be possible to identify categories, as distinct categories are likely to share specific vocabulary, while value/novelty are abstract constructs that do not have a specific vocabulary.

As stated earlier the data shows that frequent words are commonly found across the samples splitted on novelty and value scores. What does stand out is that novel options have high frequencies of *renewable energy* related terms and the low novel options show *country* related terms. Diving a bit deeper into the manual exploration, the bigrams for each group drop below 10 almost immediately in the whole sample, resulting in little frequent characteristic word combinations for feature engineering, as an example the bigrams found in the Novel, Not Novel, High Value and Low Value sub sample are shown below, the remaining uni- & trigrams of the sub samples can be found in appendix 6. Considering trigrams there are almost no trigrams to be found. And there was no difference to observe in terms of option length. The exploratory analysis does provide a clear overview of the main topics that dominate the options generated. Characteristic are terms as Renewable Energy, Data, Sustainability, Pipelines, Oil & Gas or Markets. Already providing an indication of what most participants consider fields/industries that ROSEN should pursue or fits the company.

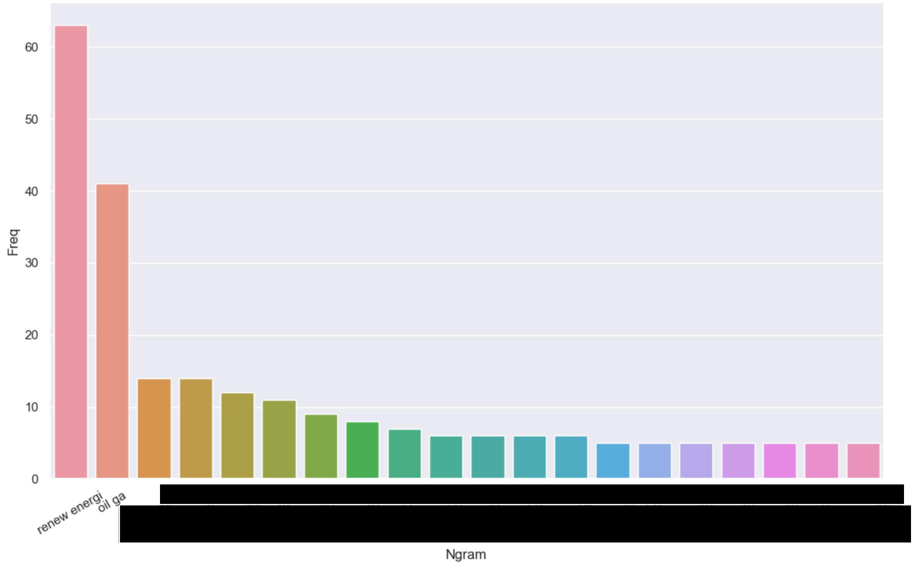


Figure 12. Bigrams for Novel options (all options with a score >0)

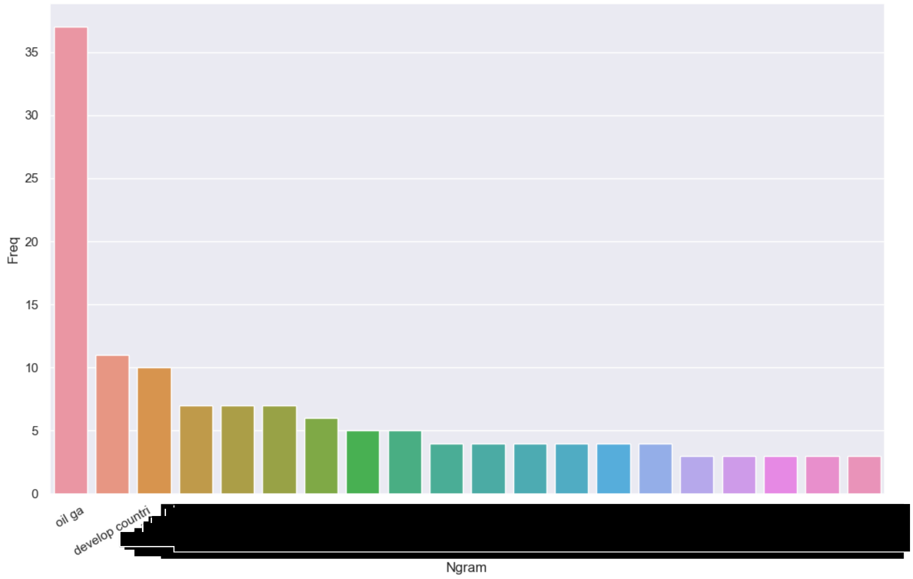


Figure 13. Bigrams for Not Novel options (all options with a score <=0)

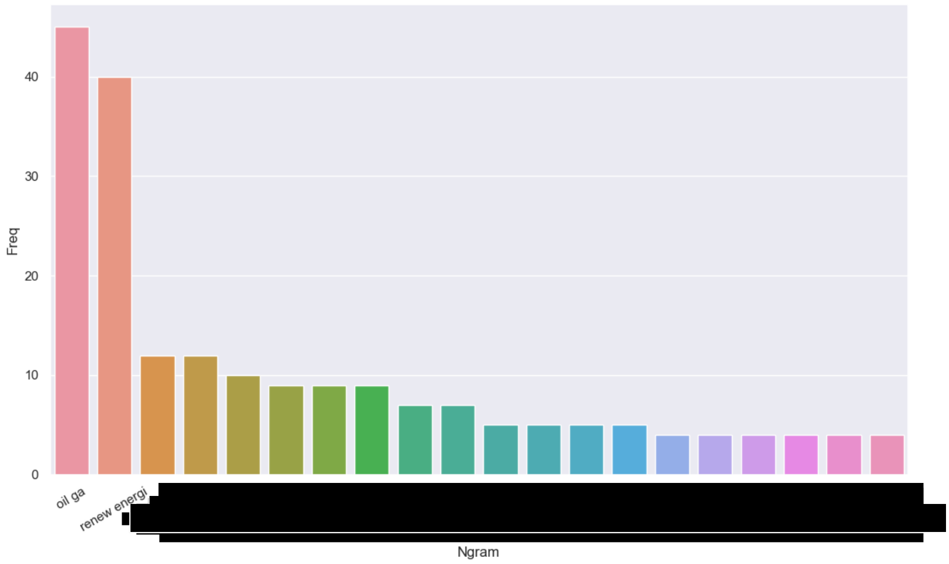


Figure 14. Bigrams for High Value options (all options with a score >0)

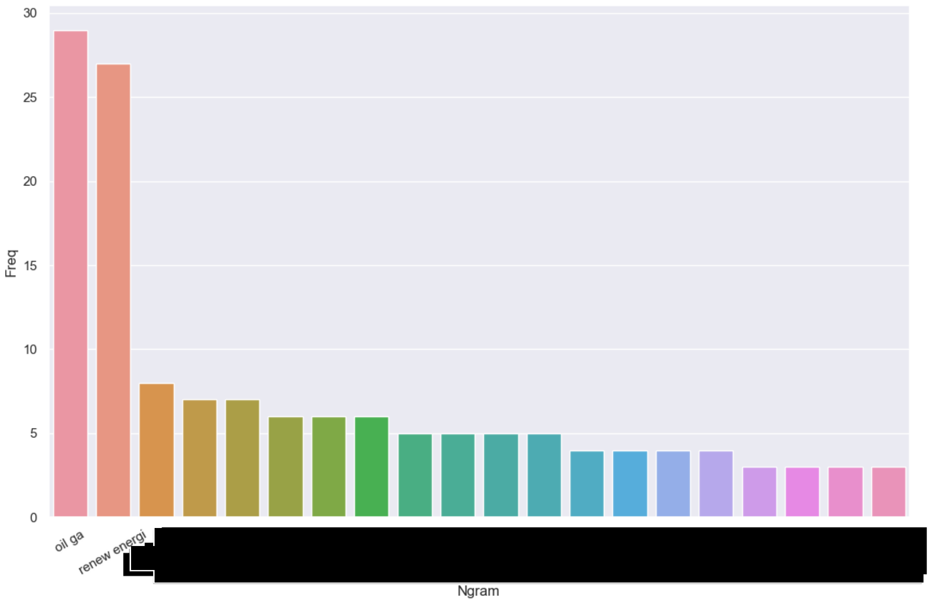


Figure 15. Bigrams for Low Value options (all options with a score <=0)

7.2 Deductive Computer Aided Text Analysis

First an analysis is run grounded in CATA, using the knowledge on words typical for exploitation or exploration from scientific literature on ambidexterity and expecting the words linked to exploration to be found more often in high novel options and the words linked to exploitation to be found in low novel options. All dictionaries used can be found in appendix 5.

7.2.1 Ambidexterity Dictionary

Extracting all keywords on ambidexterity and inspecting them learns that there seems to be no indication for a difference between the word usage associated with exploration and exploitation in the different buckets of novelty. Between each of the groups there is no clear indication that one contains significantly more terms linked to exploration or exploitation.

Table 9

Average amount of exploitation words found in strategic options for different sample splits

Exploitation		Uotila	De Visser	McKenny	Moss	Combined
Total	**Mean**	0,09	0,30	0,23	0,15	0,52
	Std	0,30	0,57	0,51	0,39	0,76
High Novel	**Mean**	0,10	0,35	0,19	0,17	0,59
	Std	0,30	0,65	0,39	0,37	0,79
Novel	**Mean**	0,07	0,25	0,25	0,12	0,47
	Std	0,25	0,48	0,54	0,35	0,72
Not Novel	**Mean**	0,12	0,34	0,26	0,18	0,55
	Std	0,33	0,64	0,55	0,45	0,86
Low Novel	**Mean**	0,10	0,34	0,19	0,11	0,50
	Std	0,37	0,60	0,44	0,37	0,65

Table 10

Average amount of exploration words found in strategic options for different sample splits

Exploration		Uotila	De Visser	McKenny	Moss	Combined
Total	**Mean**	0,06	0,60	0,23	0,15	0,86
	Std	0,27	0,88	0,51	0,39	1,08
High Novel	**Mean**	0,09	0,67	0,19	0,17	0,91
	Std	0,30	0,85	0,39	0,37	0,94
Novel	**Mean**	0,07	0,63	0,25	0,12	0,92
	Std	0,25	0,88	0,54	0,35	1,13
Not Novel	**Mean**	0,12	0,51	0,26	0,18	0,80
	Std	0,33	0,82	0,55	0,45	1,02
Low Novel	**Mean**	0,10	0,56	0,19	0,11	0,79
	Std	0,37	1,00	0,44	0,37	1,19

7.2.2 Dictionary prespecified by ROSEN manager

Leveraging the ambidexterity literature does not provide an indication that it is useful to classify options on the novelty criteria. Additionally a ROSEN manager provided words that he believed would associate with not novel activities of ROSEN. The words provided were [redacted]. These and their plurals were made into a dictionary, in this section referred to as the prespecified dictionary, and searched for in all options. At first these datapoints seem not to indicate a difference between them and the overall sample. Being evenly spread out over the sample and the mean score and standard deviation being similar to the overall sample.

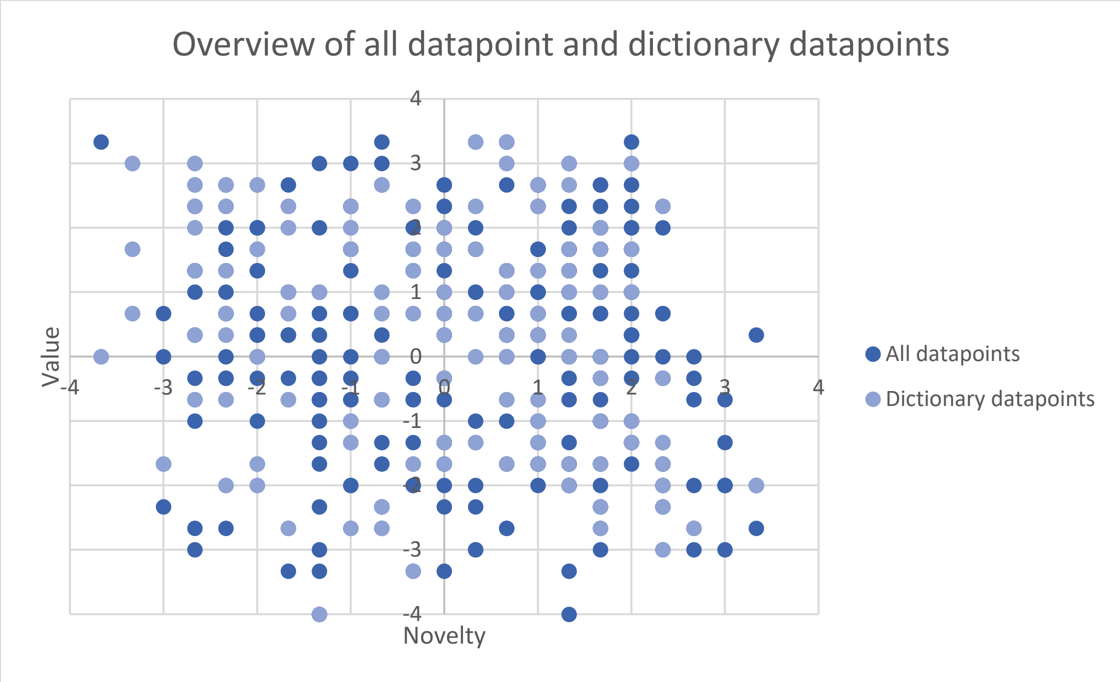


Figure 16. The novelty/value grid and the novelty/value spread of the with the prespecified dictionary filtered strategic options through the whole dataset.

Table 11

Mean scores on novelty and value of the complete sample

Sample n = 384	Novelty	Value
Mean	0,15	0,21
std	1,62	1,68

Table 12

Mean scores on novelty and value of the terms filtered with the dictionary

Sample n = 140	Novelty	Value
Mean	-0,07	0,32
std	1,61	1,71

These results were not as expected, as the options identified more or less resemble the overall sample. By overthinking these results, a possible explanation could be that although those words are used, it does not mean that it entails that the firm should pursue activities linked to the keywords, instead the option could advise to move away from activities that those keywords indicate.

Thus, the initial keywords were updated by also including frequently occurring words indicating either a leave or stay sentiment. These were identified by manually inspecting the word frequency table of the whole sample up to a frequency of 4. The words identified were used to filter out those options containing the specified terms and leave or stay terms.

The following terms were identified as either indicating a leave or a stay sentiment in a business field:

Remain word dictionary = ['expand', 'increase', "keep", "existing"]
Leave word dictionary = ['transition', "shift", "stop", "other", "leave", 'different', "away", "towards", "find", "switch", "transformation", "find", "new"]

As a result, the distribution changed. The trend within the remain word group is clearly not novel and higher value, and the leave word group is novel but relative low value. With this information you could automatically split data on oil/gas/pipeline continuation and options that want to move away from what is considered ROSEN's core business.

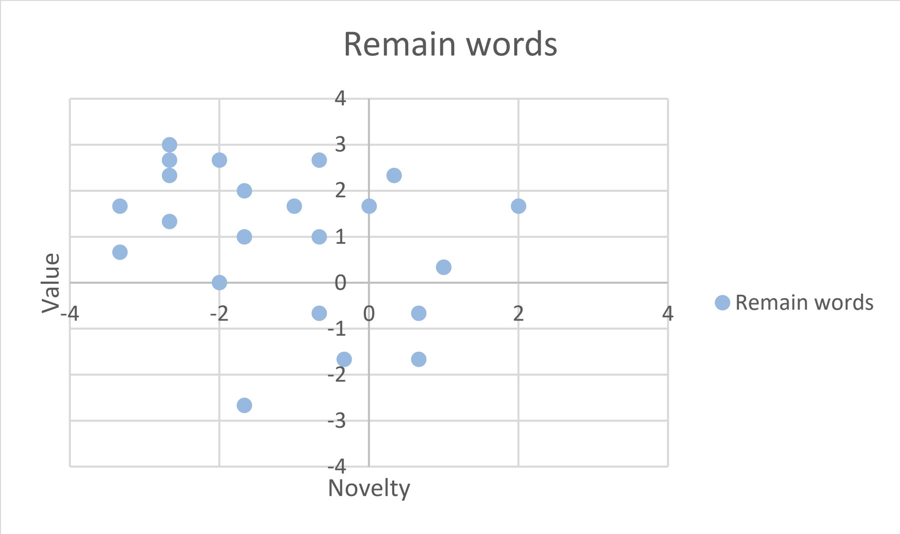


Figure 17. The novelty/value grid and the novelty/value spread of the filtered strategic options using the prespecified dictionary and the remain dictionary.

Table 13

Mean scores on novelty and value using the prespecified and remain word dictionary

Sample n = 23	Novelty	Value
Mean	-1,04	0,94
std	1,53	1,57

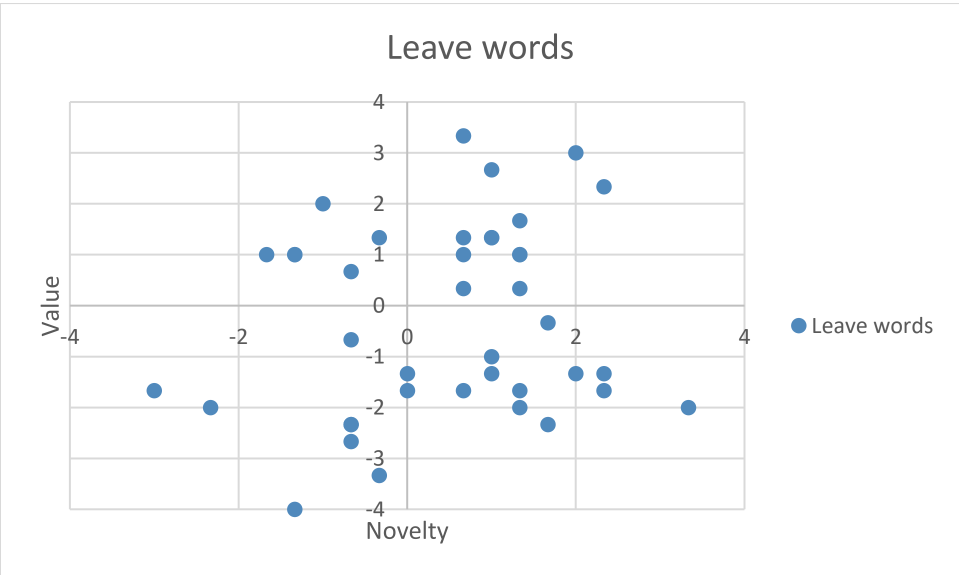


Figure 18. The novelty/value grid and the novelty/value spread of the filtered strategic options using the prespecified dictionary and the leave dictionary

Table 14

Mean scores on novelty and value using the prespecified and leave word dictionary

Sample n = 41	Novelty	Value
Mean	0,36	-0,20
std	1,51	1,91

Logically these outcomes result from this analysis, what is most interesting is that if you are able to leverage this to more complicated matters or spot similar key differences within a text dataset, it would greatly help in separating the high value or high novel options from the rest of the dataset.

7.3 Inductive Computer Aided Text Analysis

Inspecting the word frequency lists generated results in the following categories to be formed, based on common sense and logic:

Table 15

Overview of results when searching for inductive keywords

Category	1	2	3	4	5	6
Keywords	Country	Data Digital	Renewable Sustainable Green Wind Solar	Markets	Pipelines Inspection Oil Gas Fossil	Service
Number of options	25	29	101	64	163	25

Note. Each keyword that has a plural is also scanned for its plural

Although the selection of keywords is arbitrary and there is overlap between each category, the result is that in total 294 options are identified containing at least one of the specified keywords. This means that 15 words describe the topics of 294 or 76,6% of the options in the sample, assuming that due to the shortness of the options the keywords reflect the main message correctly. The benefit of this is that there are still 90 options not containing one of these keywords that could be truly unique and novel, additionally immediately a raw split can be made to focus on a certain topic with higher priority.

With this sample the results are still rather polluted due to options such as the following: '*Respond to future strict policies for oil & gas by becoming emission free. Become one of the 100% sustainable companies.'* Inspecting this option points out that it actually answers also the why question, which should be in the why layer and not in the what layer.

7.4 Clustering with K-Nearest Neighbour

Clustering with the K-Nearest Neighbour algorithm requires the manual assignment on the number of clusters. To effectively determine the number of clusters the Elbow Method is used, which plots the explained variation in the dataset as a function of the number of clusters. The elbow of the curve is the point at which the curve forms an elbow, after this point the variation explained will diminish and it is likely that actual groups are further subdivided. So, it is the point where adding an extra cluster would not better model the data.

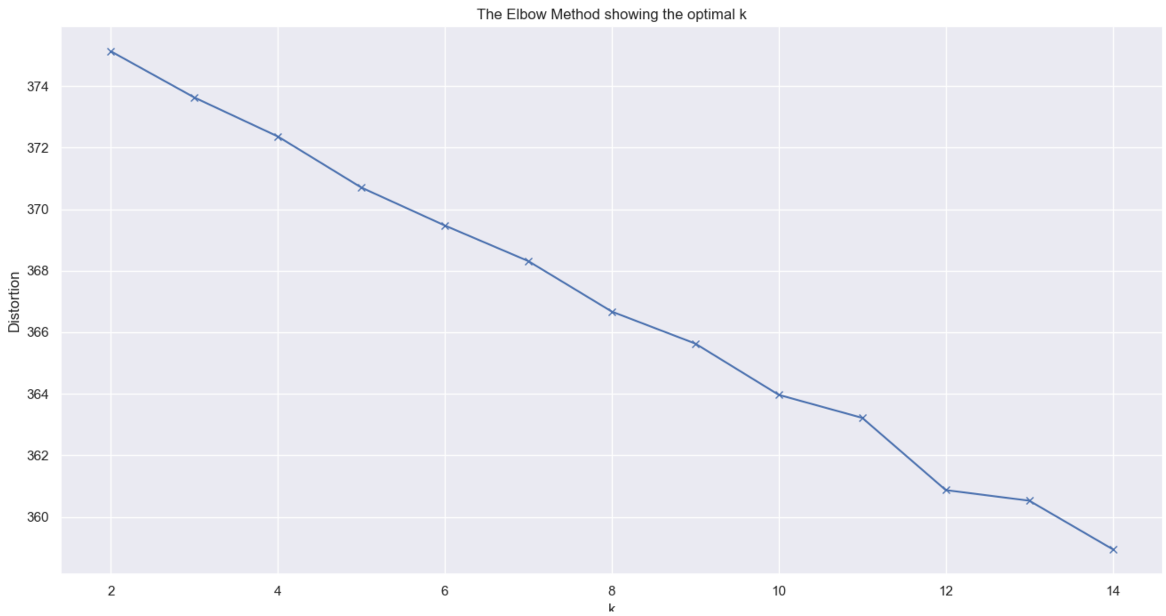


Figure 19. The graph of the elbow method to determine the optimum number of clusters.

Inspecting figure 17 it becomes clear that there is no clear elbow to be identified in the data. This provides an indication that the model is not able to identify the actual groups, or that the amount of true clusters is very large. It was thus not possible to identify the true amount of clusters so further analysis is not executed

7.5 Supervised classification

7.5.1 Supervised Classification on Novelty Criterium

First a principal component analysis is made to inspect if we can find clusters of novel and not novel options based on their principal components.

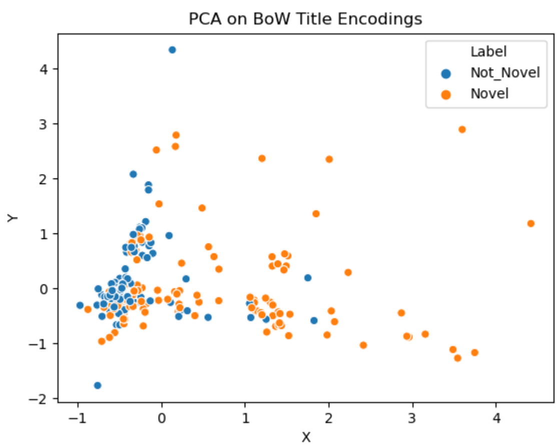


Figure 20. PCA using BoW encodings

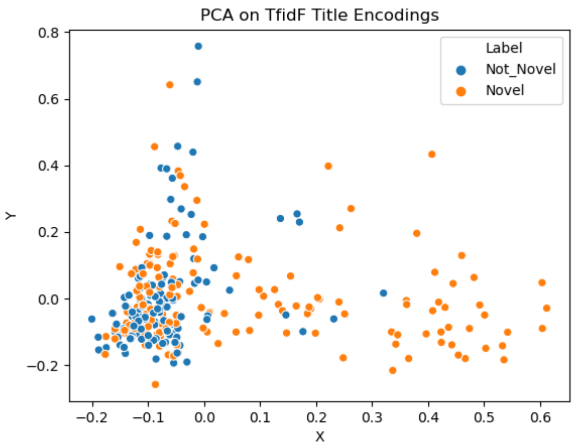


Figure 21. PCA using TFIDF encodings

Indeed, there seem to be some distinct clusters of especially novel options. Providing an indication that a classifier will provide a decent classification.

During data exploration we learned that the words around renewable energy and countries seem to differ across novel and not novel options. Apart from these words no clear features for each group could be identified. Removing the words renewable, sustainable, solar, wind, countries and energy (plus their plurals) indeed results in a loss of clear groupings of novel and not novel options within the

Principal Component Analysis. That removing few terms from the dataset immediately has such an impact on the PCA and changes its outcomes so significantly indicates that the automated classification relies too much on these features and is likely to overfit on out of sample data.

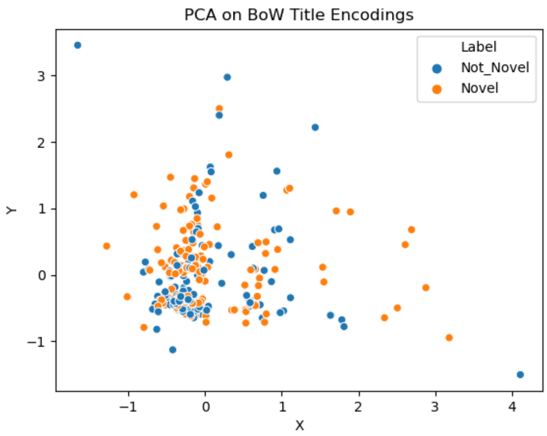


Figure 22. PCA using BoW encodings

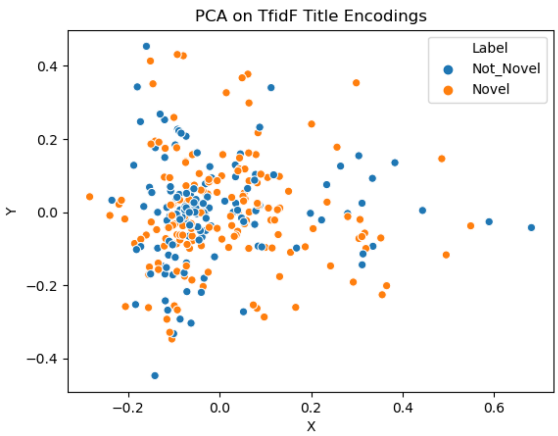


Figure 23. PCA using TFIDF encodings

The required assumption that groups are different on word usage detectable for a supervised machine learning model, using word frequency based word representations, can thus not be met with the available dataset. So supervised classification is not further pursued for the novelty criterium.

7.5.2 Supervised Classification on Value Criterium

To classify for Value again first a principal component analysis is made. Immediately it becomes clear that based on principal components there are no distinct groups of documents as in the case of novelty. Which was expected from manually assessing word frequencies of each group, in which manually there was no significant difference in vocabulary detectable. Thus also for this criterium no further supervised classification will be executed.

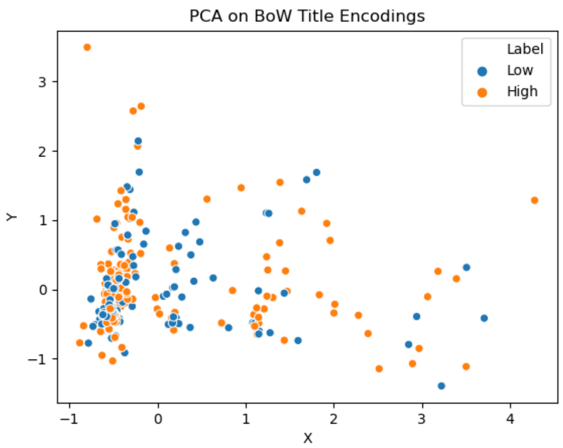


Figure 24. PCA using BoW encodings

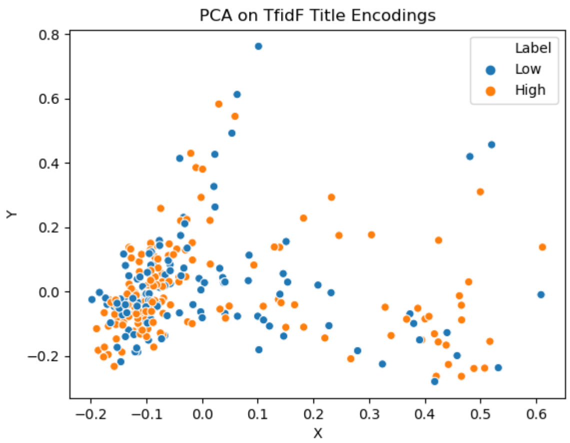


Figure 25. PCA using TFIDF encodings

8. Discussion

Roadmapping is a useful tool applicated successfully by numerous firms to assist in managing technology. There are three phases in the Scenario Driven Roadmap process, the preparation, workshop setting and implementation phase (Siebelink et al., 2016) . Especially in the workshop setting phase the roadmapping process faces some limitations related to scalability in its data processing, which is particularly challenging for large firms that have a varied portfolio and multiple business units. If a roadmapping process is started and different business units are all participating in the workshop phase, which is in essence a good thing, as it maximizes different input angles and enables for example the discovery of potential recombination of technologies to serve expected new market requirements, it will result in large data output which requires processing to select the most promising focus areas and preconditions. Processing manually will eventually become unfeasible as the manager(s) responsible would have a day job going through the strategic options while remaining to have an overview of the bigger picture. To avoid bias and create credibility you would ideally want 2 or more high level managers that compare their results, resulting in a heavy workload on key employees. To aid in data processing a potential solution was searched for in computer assisted text processing methods. Aiding in the data processing automatic categorization of data would be useful and ideally you want to filter out automatically the good and bad strategic options, reducing the strain .

8.1 Dataset

The dataset used in this research was not an internal ROSEN dataset, but instead created by students from the University of Twente participating in an experiment for scenario-driven roadmapping. This means that the dataset is unlikely to possess specific terms that are used within ROSEN. Additionally the ideas were formulated in sentences and elaborated on in depth, this resulted into signal words occurring only once or twice, while in a longer piece of text the signal words are likely to be repeated, providing a stronger signal. The options were also classified on value and novelty by three ROSEN managers; thus it may be that the bias of raters has an influence on their ratings. Additionally of the why, what and how layers only the what layer was available for this research, so the complete classification of all process output was not possible, combining those layers would maybe open up new possibilities to classify them automated.

It is not surprising that with the data at hand machine learning tools struggle. As the data is rather homogenous in word usage. There is a need for more distinct use of words which could be enabled by making design choices before you start collecting data.

A priori you could also decide to make the dataset easier to analyse, by asking participants to summarize their contribution in 3-5 words you would get indicators for the category an option belongs to without noisy context and comparisons. This is important because comparisons cause word occurrences that may not be representative for the subject of an option. Another design choice could be to provide an x amount of categories in which the idea could fit and the participant needs to choose one. The number of categories and the possible overlap between them is dependent on the context in which the roadmapping process is taking place.

The data and its context also appear to be very much determining the usefulness and the quality of an automated assessment. If you would have two categories, let’s assume sport and political news, it would be fairly easy to distinguish between them. Would you however zoom in on politics and separate local and national political news, the task would be more difficult, as the overlap between those is rather large and word usage will likely be similar, you then need more training data in order for a machine learning model to discover the discriminating patterns or words. In this roadmapping sample however the categories to fit the options in are not yet known, the criteria have high granularity, and the options are short providing little features to use for classification.

8.2 Unsupervised machine learning within roadmapping

Sparsity is a large issue clustering for K-Nearest Neighbour, resulting in clusters that do not always make sense. Signalling those words that are actually indicating a category is hard for K-Nearest

Neighbour if data is on a sentence level. Options in a cluster sometimes share not more than one word and noisy words deteriorate performance considerably, making this approach not much better than identifying words by hand and put all options containing those words in one category. Would the data consist out of an extensive proposal, then it would make more sense to use a K-Nearest Neighbour algorithm, as the sparsity goes down and it becomes less likely that a single word is representative for an idea.

8.3 Supervised machine learning within roadmapping

Within this research we saw that we could classify novelty rather well using a Naïve Bayes classifier, removing however three words: Renewable, Energy, Country and their plurals let to a drastic change in classification. Thus as the signal for classification was so heavily influenced by the absence or presence of the aforementioned words it provides a clear indication that you would need more data, as now there are 384 options and the removal of 3 terms is already highly influential. I believe you would need at least a 10.000 of these kind of quality options to have a stable robust classifier.

The difficulty using supervised machine learning approaches is that the computer actually should learn something and then apply this to unseen data. In many natural language applications this new data is rather similar in terms of word usage or categories and generated rapidly by multiple users. For example, in sentiment analysis the word usage to describe positive experiences or negative experiences are not likely to change over time. The same applies for other popular classification tasks that have stable categories with specific word usage such as news or spam email classification. The size of the training data sample would be determined by the distinctiveness of vocabularies and the granularity required for classification. The more distinct and the lower the granularity, the less data is required to classify. Roadmapping however is a process that wants to identify new or high value opportunities for a company to pursue, this is not a stable construct and the velocity by which options are generated is low compared to for example big data generated on the Internet or by the Internet of Things. Traditional machine learning is good at extracting patterns and knowledge from large datasets, but remains a 'dumb' machine that seeks for word occurrences. They are not suitable for interpreting unseen data that reflect concepts on which they have had no training. Thus, what is now considered as novel may be outdated in 5 years and what is now a potential high value option may be worthless in 3 years. Additionally, once you managed to label and collect enough data to train a robust classifier with, the labels are probably outdated or do not reflect the categories anymore in present day. As the model is only able to actually learn from previous unseen instances, incorporating environmental dynamics and changes such as for example the COVID-19 pandemic is extremely difficult using traditional word frequency-based techniques. Additionally, some high value options could be lost, sometimes a minor change to an existing low value idea can change the perception and possible success completely. For example Thomas Edison did not invent the lightbulb, but did come up with the idea to connect and model the electricity infrastructure to the existing gas infrastructure. A machine learning model would in this case maybe know lightbulb and electricity as not feasible, low value and not novel, but a slight change in the idea actually makes it feasible and high value. Additionally, the criteria on which the roadmapping process evaluates options are measuring different dimensions, to evaluate on them all would require even more data.

8.4 Inductive Computer Aided Text Analysis

Returning to the example of the polluted option in the results section: *'Respond to future strict policies for oil & gas by becoming emission free. Become one of the 100% sustainable companies.'*. As stated previously it also answers the why question, because why should ROSEN do this, because there will be future strict oil & gas policies. Such fuzzy formulated options results in options containing words of multiple categories when using inductive CATA. The following would be a better formulation: *Become a completely emission free company and thus one of the 100% sustainable companies.* This is not only a problem for inductive CATA, but for all classification methods used in this research. By separating the layers within each option more clearly and reserving some effort to inform the participants about the importance of this separation, classification and analyses will be easier.

Now inductive Computer Aided Text Analysis still suffers from the comparisons and mixing of layers within the sample, however when the ‘rules’ of formulating each layer will be made clear up front explicitly I believe this technique would be useful to quickly generate categories.

Additionally, the categories and their keywords are made by a researcher who does not have extensive experience working for ROSEN or knowledge on all their activities.

8.5 Deductive Computer Aided Text Analysis

An approach leveraging the knowledge on ambidexterity does not provide results for this specific dataset. The average signal words found are for both exploration as exploitation similar. Would there be options that are more similar to the data on which most similar research (Uotila et al., 2009; Short et al., 2010; Moss et al., 2014; McKenny et al., 2018) which used large pieces of text that told a story, then it would perhaps be able to extract a explorative or exploitative orientation, as it is indeed successful in previous research.

Using the knowledge from a ROSEN manager does indeed yield results, by using keywords that are representative for activities ROSEN is active in and combining this with words that appeared often symbolizing a leave or remain sentiment resulted in two distinct groups that are significantly different on novelty and value. Using the computer to quickly identify those different orientations could be useful, but does not help the overall processing as only a small subset of the data is targeted. If such an observation can be leveraged to a larger, or the complete part of the data then it becomes interesting, this could be achieved by asking participants to choose out of predefined categories for their input that are associated with a novelty and value score.

8.6 Further research

So traditional machine learning is difficult to implement in an idea generation/roadmapping process. Humans are in general just very good at interpreting: environments, subtle differences in meaning, capabilities of a firm and they adapt well. While machines are better at crunching huge amounts of data and spot patterns that a human might miss. A human that crunches the same data would be time consuming and challenging, but on most language related tasks provide better results (transformer models are matching human performs at some tasks currently). Building a classifier on sentence or paragraph level documents results in sparsity, resulting in classification or clustering based on the occurrence of words or word pairs. Which could also be manually analysed by creating frequency tables and interpreting the keywords that occur often, as in a sentence or paragraph they often describe the subject. You can also check as showed in this research on word such as expand or move away, to get more details on the intent of an option on for example renewable energy.

In this research the focus was on a specific part of the roadmapping process: the processing of the workshop output to select strategic focus areas and preconditions to put on the roadmap. The roadmapping process entails however more steps in which automated solutions could possibly increase the performance of the process. To name a few there could be an automated tool to scan the competitive environment and increase the understanding of current and future trends within the environment. Or analyse the current situation and trends of the firm itself by automatically scanning for the main topics that are found in internal communication.

Transformer based models are changing the automated processing of text, enabling models to actually understand language and enabling the use of machine learning on small datasets that were not suitable for machine learning, but too large for humans to efficiently classify. Still there would be the need for a ground truth set of high-quality labelled examples. The amount is very much depending on the granularity of classification and the distinctness of categories, but more is mostly better. A model such as BERT would then be able to process the data based on their *meaning*. Returning to the process of roadmapping and its purpose, which is all about identifying and gaining new strategic insights. Which is problematic, as machine learning classification always uses historic data you fed it with. You could indeed classify or cluster, but assessment is always on what you already know. But the interest of roadmapping is on what you did not know and the need for new insights, using historic data is thus counter intuitive for the purpose of a roadmapping process. Classification into categories could work, but options in a category still would need inspection, to find those ideas that are the best, unknown, provide new opportunities or are not yet identified as essential. Therefore, I believe the most promising approach to pursue in future research is similar to one of the glue benchmarks, involving Quora questions that are identified as being duplicate. This task based on similar semantic meaning is exactly what you would want for roadmapping, as all generated options that are already proposed can be accurately identified, enabling first the assessment of ideas with a truly new meaning, which are most likely to provide new strategic insights. Finetuning a pre trained transformer model for this task is complicated and requires expert knowledge, additionally as development rates are so high it is hard to predict if this is the right moment to select a model and downstream it.

9. Conclusion

Within this research the used automated solutions did not match the performance of a human rater. Important to consider is the quality of the dataset in this case, as the categories were not very distinct and the short amount of texts per document made it more challenging to work with when using automated techniques. So, based on the quality of the dataset I cannot state with certainty that automated classification is feasible or not feasible for analysing the workshop output within the roadmapping process. Roadmapping is a complicated and dynamic task subject to internal and external changes that a firm experiences. This shift is not optimal when using automated/machine learning solutions to assign scores or labels, as it requires historic training data, which is more easy to collect if the concept you are interested in is stable, as it decreases the rate at which training data becomes obsolete compared to the rate of new data collection. Collecting enough data (what enough data is will be dependent on the diversity of the inputs and the granularity of the scores/labels) before the data collected becomes outdated is therefore difficult. To evaluate the output of the workshop phase the categories and criteria scores that are typically used are rather fine grained, together with the high dimensional nature of text data and the diversity of inputs this results in large data samples to be collected in order to be able to construct a generalizable model. Thus, it is an immensely complicated task for word frequency-based document classification methods to match the detailed classification or scoring of a human rater. As humans are learning from all kind of historic and every day experiences instead of just historic strategic options, making them more widely informed and flexible.

Using a computer is definitely useful when exploring the data, extracting n-gram frequencies provides a clear indication of the main topics that are in your output. Using these insights make manually rating the options easier as a priori some categories can already be constructed and with the click of a button a fairly accurate summary of the general trends within the datasets can be created.

When focussing on the planning aspect of roadmapping, it became clear that the importance of planning in the preparation phase has not been fully optimized for the data processing. When roadmapping I suggest considering beforehand more carefully in which format the options should be delivered and on what criteria they will be assessed. By doing so you could already leverage some of the work to the front end. My suggestion would be to ask participants, *after* they have formulated their strategic options, to choose from supplemented keywords that are known to represent activities/categories of interest. If their option is not under the umbrella of supplemented terms, they should write down a maximum of three keywords that summarize their option. Doing so decreases noise and mixing of layers, decreasing disturbance of a potential automated analyses.

A dictionary containing these keywords can be constructed using the knowledge gained over multiple iterations of the roadmap. Inspecting each iteration and forming categories with distinct keywords enables the development of a dictionary in which keywords are linked to a specific category, or multiple categories in cases of overlap. Using this dictionary to scan future options and providing keywords from the dictionary to participants allows for more organized workshop output to start analysing. A dictionary can then be used to develop a classifier that estimates how likely it is that an option belongs to a certain category, based on the matching keywords belonging to that option.

Thus, when looking at classic machine learning tools based on word frequencies, I would argue that it will cost more time and resources to develop a functional automated solution then it would take to manually assess the generated options. The velocity of the data generated, high dimensionality and the fine-grained scoring required for further analyses make it challenging to implement automated solutions successfully. On the short term I would therefore recommend to start creating dictionaries that are carefully constructed around topics/categories of interest and provide them to participants in the workshop phase to assign the keywords to their options. This already structures the data and provides easy access to options centred around a topic of interest. If an organization has the means, I would on a long-term basis work on a deep learning model that detects if an option already has been suggested. On the dataset of Quora questions within the glue benchmark there are already working examples with high accuracy. Using the latest developments by using a deep learning model that is able to understand words in context enables a computer to recognise semantic similarity and thus to also recognize options with similar meaning, but different vocabulary. Such a model only relies on what options already have been formulated historically and therefore does not need to be enriched with internal and external changes or knowledge. A drawback is that these models are complicated and acquiring the right people and knowledge to design such a system are currently substantial. Using such a model as a duplicate filter reduces the workload.

All in all the task to develop an automated solution to classify the textual workshop outputs within the roadmapping is not to be underestimated. Especially on short sentences expert human raters are difficult to outperform and the data should meet the assumptions made by an automated solution.

However, for example in classifying research by analysing abstracts, which provide more data and potential features to work with, automated solutions are currently successfully and widely applied. The current advances in natural language processing techniques, including the development of solutions that can be enriched with external knowledge and the capability of semantic understanding, are possibly unlocking new applications for automated text classification which were challenging before. So performance increases on short text classification tasks are likely. Potentially enabling non-experts to reliably automate challenging classification tasks, if these more advanced methods mature and become more accessible for application by others than data-scientists.

9.1 Blueprint for the data collection of strategic options on the short term

1. Start with choice of a digital format to collect the data. A good choice would be plain .txt files.
2. Consider the classification and evaluation of the options, will categories (i.e. 'Digital Transformation', criteria scores (i.e. value, novelty) or both be used?
3. Explain explicitly that each layer of the strategic options should only consider that layer. To avoid comparisons and the mixing of layers.
4. Start the workshop and generate the options
5. After the generation of options ask the participants to make a choice out of prespecified categories. If their option does not fit in, then they can describe the category in maximum three words.
6. Provide the selection criteria and a maximum of X points to be divided over the criteria. In case of 5 criteria you could provide 25 points to be divided on a 10-point scale. This forces the participant to critically assess the criteria for his/her option. Assuming the participant has some affinity with the organisation, their assessment on categories and criteria scores should be indicative. This however needs to be validated in future iterations.
7. Execute similar explorative data analysis as used in this research, making uni- bi- and trigram frequency lists.
8. Validate the performance of participants on assessing criteria.

Using this blueprint transfers part of the workload to the start of the process and makes it easier to use machine learning tools as the options will be less polluted.

10. Acknowledgements
I would like to thank ROSEN Technology and Research Center GmbH and especially Dr. Ingo Nee for their support, time and the opportunity to conduct my master thesis at ROSEN. The corona pandemic broke out at the start of my thesis, thus it was fantastic that I still had the ability to complete it. Furthermore, I would like to thank Dr. Erwin Hofman and Dr. Igors Skute for their support and time during the writing of this thesis.

11. References

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, *37*, 3-13.

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. Springer Science & Business Media.

Albright, R. E., & Kappel, T. A. (2003). Roadmapping in the corporation. *Research-Technology Management*, *46*(2), 31-40.

Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. *Acm Sigkdd Explorations Newsletter*, *2*(2), 1-13.

Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage?. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1833-1836).

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

de Visser, M., Faems, D., Visscher, K., & de Weerd-Nederhof, P. C. (2017). Sequential ambidexterity in practice: a longitudinal study on shifts from exploration to exploitation and vice versa. In *PDMA Annual Research Forum 2017*.

Derbyshire, J., & Giovannetti, E. (2017). Understanding the failure to understand New Product Development failures: Mitigating the uncertainty associated with innovating new products by combining scenario planning and forecasting. *Technological Forecasting and Social Change*, *125*, 334-344.

Derbyshire, J., & Wright, G. (2017). Augmenting the intuitive logics scenario planning method for a more comprehensive analysis of causation. *International Journal of Forecasting*, *33*(1), 254-266.

Detienne, K. B., Detienne, D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, *6*(2), 236-265.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78-87.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, *29*(2-3), 103-130.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5-6), 352-359.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference* (Vol. 5). Cambridge University Press.

EIRMA (1997), 'Technology roadmapping - delivering business vision', Working group report, Eisenhardt, K. M., & Brown, S. L. (1998). Time pacing: Competing in markets that won't stand still. *Harvard business review*, *76*(2), 59-70. *European Industrial Research Management Association,* Paris, No. 52.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *The journal of machine learning research*, *15*(1), 3133-3181.

French, J. C., Brown, D. E., & Kim, N. H. (1997). A classification approach to Boolean query reformulation. *Journal of the American Society for Information Science*, *48*(8), 694-706.

Geschka, H., & Hahnenwald, H. (2013). Scenario-Based Exploratory Technology Roadmaps-A Method for the Exploration of Technical Trends. In *Technology Roadmapping for Strategy and Innovation* (pp. 123-136). Springer, Berlin, Heidelberg.

Goffin, K., & Mitchell, R. (2010). Innovation management. *NY: Palgrave Macmillan*, *416*.

Groenveld, P. (1997). Roadmapping integrates business and technology. *Research-Technology Management*, *40*(5), 48-55.

Groenveld, P. (2007). Roadmapping integrates business and technology. *Research-Technology Management*, *50*(6), 49-58.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*(1), 20-38.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504-507.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine Learning, ECML-98,* 137–142.

Joachims, T. (1999, June). Transductive inference for text classification using support vector machines. In *Icml* (Vol. 99, pp. 200-209).

Johnson, R., & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks. *arXiv preprint arXiv:1412.1058*.

Kang, I. S., Na, S. H., Kim, J., & Lee, J. H. (2007). Cluster-based patent retrieval. *Information processing & management*, *43*(5), 1173-1182.

Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of advances in information technology*, *1*(1), 4-20.

Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H. (2006). Some effective techniques for naive bayes text classification. *IEEE transactions on knowledge and data engineering*, *18*(11), 1457-1466.

King, G., & Lowe, W. (2003). An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, *57*(3), 617-642.

Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *IEEE Transactions on engineering management*, *48*(2), 132-143.

Kowsari, K., Brown, D. E., Heidarysafa, M., Meimandi, K. J., Gerber, M. S., & Barnes, L. E. (2017, December). Hdltex: Hierarchical deep learning for text classification. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)* (pp. 364-371). IEEE.

Kowsari, K., Yammahi, M., Bari, N., Vichr, R., Alsaby, F., & Berkovich, S. Y. (2015). Construction of fuzzyfind dictionary using golay coding transformation for searching applications. *arXiv preprint arXiv:1503.06483*.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Macintosh, A., Filby, I., & Tate, A. (1998). *Knowledge asset road maps*. University of Edinburgh, Artificial Intelligence Applications Institute. *Management,* Sept.-Oct., pp. 13-19.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Evaluation in information retrieval. *Introduction to information retrieval*, *1*, 188-210.

McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, *44*(7), 2909-2933.

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of statistical software*, *25*(5), 1-54.

Mikolov, T., Chen, K., Corrado, G., Dean, J., Sutskever, L., & Zweig, G. (2013). word2vec. *URL https://code. google. com/p/word2vec*, *22*.

MIT, 2016. Lecture 11: Introduction to Machine Learning. [Online] Available at: *URL https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/lecturevideos/lecture-11-introduction-to-machine-learning/*

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621-633.

Moss, T. W., Payne, G. T., & Moore, C. B. (2014). Strategic consistency of exploration and exploitation in family businesses. *Family Business Review*, *27*(1), 51-71.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, *56*(6), 960-980.

Neuendorf, K. A. (2016). *The content analysis guidebook*. sage.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Petrick, I. J., & Echols, A. E. (2004). Technology roadmapping in review: A tool for making sustainable new product development decisions. *Technological Forecasting and Social Change*, *71*(1-2), 81-100.

Phaal, R., & Muller, G. (2009). An architectural framework for roadmapping: Towards visual strategy. *Technological forecasting and social change*, *76*(1), 39-49.

Phaal, R., Farrukh, C. J., & Probert, D. R. (2007). Strategic roadmapping: A workshop-based approach for identifying and exploring strategic issues and opportunities. *Engineering Management Journal*, *19*(1), 3-12.

Phaal, R., Farrukh, C., & Probert, D. (2001). *T-Plan: the fast-start to technology roadmapping: planning your route to success*. University of Cambridge, Institute for Manufacturing.

Porter, M.E. (1980). Competitive Strategy: Techniques for Analyzing Industries and Competitors (Reprint 1980 Ed.). New York, NY: The Free Press.

Roberts, A., & Raffel, C. (2020). Exploring Transfer Learning with T5: the Text-To-Text Transfer Transformer. Accessed on: 23-07-2020.

Saritas, O., & Aylen, J. (2010). Using scenarios for roadmapping: The case of clean production. *Technological forecasting and social change*, *77*(7), 1061-1075.

Scholkopf, B., & Smola, A. J. (2001). *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* Cambridge, MA: MIT Press.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, *34*(1), 1-47.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320-347.

Siebelink, R., Halman, J. I., & Hofman, E. (2016). Scenario-Driven Roadmapping to cope with uncertainty: Its application in the construction industry. *Technological forecasting and social change*, *110*, 226-238.

Simonse, L. W., Hultink, E. J., & Buijs, J. A. (2015). Innovation roadmapping: Building concepts from practitioners' insights. *Journal of product innovation management*, *32*(6), 904-924.

Strauss, J. D., & Radnor, M. (2004). Roadmapping for dynamic and uncertain environments. *Research-Technology Management*, *47*(2), 51-58.

Strauss, J., Radnor, M., & Peterson, J. (1998, March). Plotting and navigating a non-linear roadmap: knowledge-based roadmapping for emerging and dynamic environments. In *Proceedings of the East Asian Conference on Knowledge Creation Management* (pp. 6-7).

Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technological Forecasting and Social Change*, *115*, 131-142.

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of machine learning research*, *2*(Nov), 45-66.

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: analysis of S&P 500 corporations. *Strategic Management Journal*, *30*(2), 221-231.

Vishnevskiy, K., Karasev, O., & Meissner, D. (2015). Integrated roadmaps and corporate foresight as tools of innovation management: the case of Russian companies. *Technological Forecasting and Social Change*, *90*, 433-443.

Wang, X., & Dass, M. (2017). Building innovation capability: The role of top management innovativeness and relative-exploration orientation. *Journal of Business Research*, *76*, 127-135.

Willyard, C.H. and McClees, C.W. (1987), 'Motorola's technology roadmap process', *Research*
Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1-37.

Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval*, *1*(1–2), 69–90.

Zhao, R., & Mao, K. (2017). Fuzzy bag-of-words model for document representation. *IEEE transactions on fuzzy systems*, *26*(2), 794-804.

12. Appendices

Appendix 1 Unique Words Between Novelty Samples

Novel Words		Not Novel Words	
Word	Freq	Word	Freq
energi	110	oil	47
renew	77	ga	47
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]			
[redacted]	[redacted]		
[redacted]	[redacted]		

Novel Words		**Not Novel Words**	
Word	Freq	Word	Freq
energi	110	oil	47
renew	77	ga	47
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

Appendix 2 Unique Words between High Novel and Low Novel Sample

High Novel Words		Low Novel Words	
Word	Freq	Word	Freq
energi	21	countri	22
invest	17	oil	21
[redacted]	[redacted]	[redacted]	[redacted]

High Novel Words		**Low Novel Words**	
Word	Freq	Word	Freq
energi	21	countri	22
invest	17	oil	21
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]
		[redacted]	[redacted]

Appendix 3 Unique Words between Value and Non-Value Sample

Value Words		Non-Value Words	
Word	Freq	Word	Freq
energi	71	energi	58
ga	63	invest	52
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]			
[redacted]	[redacted]		
[redacted]	[redacted]		

[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

Value Words		**Non-Value Words**	
Word	Freq	Word	Freq
energi	71	energi	58
ga	63	invest	52
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	
[redacted]	[redacted]	

Appendix 4 Unique Words between High Value and Low Value Sample

High Value Words		Low Value Words	
Word	Freq	Word	Freq
technolog	23	more	14
pipelin	21	energi	13
[redacted]	[redacted]	[redacted]	[redacted]

High Value Words		Low Value Words	
Word	Freq	Word	Freq

technolog	23	more	14
pipelin	21	energi	13
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

Appendix 5 Ambidexterity dictionaries

keywordsdeVisserexplor = ['research', 'researching', 'unveiled', 'search', 'search for', 'look for', 'seek', 'forage for', 'fish around for', 'fish about for', 'look through', 'hunt through', 'explore', 'go through', 'examine', 'inspect', 'check', 'variation', 'difference', 'dissimilarity', 'disparity', 'contrast', 'discrepancy', 'imbalance', 'change', 'alteration', 'diversification', 'deviation', 'variance', 'divergence', 'departure', 'fluctuation', 'development', 'adaptation', 'alteration', 'mutation', 'transformation', 'modification', 'risk', 'chance', 'uncertainty', 'unpredictability', 'precariousness', 'instability', 'insecurity', 'probability', 'likelihood', 'danger', 'threat', 'menace', 'fear', 'prospect', 'experimentation', 'investigation', 'trial', 'examination', 'observation', 'assessment', 'evaluation', 'appraisal', 'analysis', 'study', 'research', 'testing', 'tests', 'try out', 'play', 'amuse oneself', 'entertain oneself', 'enjoy one self', 'have fun', 'relax', 'mess around', 'amusement', 'entertainment', 'relaxation', 'recreation', 'diversion', 'flexibility', 'adaptability', 'adjustability', 'variability', 'versatility', 'open-endedness', 'freedom', 'latitude', 'tolerance', 'discovery', 'uncovering', 'realization', 'recognition', 'revelation', 'disclosure', 'invention', 'origination', 'pioneering', 'find', 'invention', 'breakthrough', 'innovation', 'innovation', 'change', 'alteration', 'revolution', 'transformation', 'metamorphosis', 'breakthrough', 'modernization', 'novelty', 'newness', 'creativity', 'originality', 'inspiration', 'inventiveness', 'new', 'up to date', 'latest', 'current', 'state-of-the-art', 'contemporary', 'advanced', 'recent', 'modernization', 'cutting-edge', 'novelty', 'original', 'fresh', 'creative', 'experimental', 'different', 'unfamiliar', 'unknown', 'tacit', 'implicit', 'learning', 'studying', 'education', 'research', 'knowledge', 'understanding', 'research', 'investigation', 'experimentation', 'testing', 'analysis', 'fact-finding', 'fieldwork', 'examination', 'study', 'inquire', 'probe', 'explore', 'analyze', 'review', 'look through', 'look into']

keywordsdeVisserexploit = ['refinement', 'improvement', 'fine-tuning', 'finishing off', 'revision', 'editing', 'reworking', 'choice', 'selection', 'election', 'choosing', 'picking', 'decision', 'alternative', 'range', 'variety', 'production', 'manufacture', 'making', 'construction', 'building', 'fabrication', 'assembly', 'creation', 'mass-production', 'composition', 'yield', 'productivity', 'efficiency', 'order', 'regulation', 'coherence', 'effectiveness', 'selection', 'choice', 'pick', 'option', 'preference', 'implementation', 'apply', 'put into effect', 'put into practice', 'carry out', 'perform', 'enact', 'fulfill', 'accomplish', 'achieve', 'realize', 'effectuate', 'execution', 'implementation', 'carry out', 'accomplishment', 'engineering', 'attainment', 'realization', 'performance', 'existing', 'prevailing', 'occuring', 'explicit', 'clear', 'plain', 'straightforward', 'understandable', 'precise', 'exact', 'specific', 'detailed', 'comprehensive', 'standardization', 'systematize', 'consistent', 'uniform', 'comparable', 'regulate', 'normalize', 'equalize', 'homogenize', 'regiment', 'scale up', 'increase', 'expand', 'augment', 'build up', 'add to', 'step up', 'boost', 'escalate']

keywordsUotilaexplor = ["exploration", "explore", "explores", "exploring", "search", "searching", "searchs", "variation", "variations", "risk", "risks", "experiment", "experiments", "experimenting", "play", "plays", "playing", "flexible", "flexibility", "discovery", "discover", "discovers", "discovering", "innovate", "innovates", "innovation"]

keywordsUotilaexploit = ["exploit", "exploitation", "exploits", "refine", "refines", "refinement", "choice", "choices", "production", "efficient", "efficiency", "select", "selects", "selecting", "implement", "implements", "implementation", "execute", "executes", "execution"]

keywordsMossexplor = ['discoverable', 'discoverably', 'discovered', 'discoverer', 'discoverers', 'discoveries', 'discovering', 'discoverist', 'discoverists', 'discoverment', 'discoverments', 'discovers', 'discovery', 'experiment', 'experimental', 'experimentalism', 'experimentalist', 'experimentalists', 'experimentalize', 'experimentally', 'experimentarian', 'experimentarians', 'experimentation', 'experimentations', 'experimentative', 'experimentator', 'experimented', 'experimenter', 'experimenters', 'experimenting', 'experimentist', 'experimentists', 'experimentor', 'experimentors', 'experiments', 'explorability', 'explorable', 'explorable', 'explorate', 'explorates', 'exploration', 'explorationist', 'explorationists', 'explorations', 'explorative', 'exploratively', 'explorator', 'explorators', 'exploratory', 'explore', 'explored', 'explorement', 'explorer', 'explorers', 'explores', 'exploring', 'exploringly', 'flexibility', 'flexible', 'flexibleness', 'flexibly', 'innovate', 'innovated', 'innovates', 'innovating', 'innovation', 'innovational', 'innovationist', 'innovationists', 'innovations', 'innovative', 'innovatively', 'innovativeness', 'innovator', 'innovators', 'innovatory', 'play', 'played', 'player', 'players', 'playful', 'playing', 'playingly', 'playlike', 'plays', 'research', 'risk', 'risked', 'risker', 'riskers', 'riskful', 'riskier', 'riskiest', 'riskily', 'riskiness', 'risks', 'risky', 'search', 'searchable', 'searchableness', 'searched', 'searcher', 'searchers', 'searches', 'searching', 'searchingly', 'variation', 'variational', 'variationally', 'variations', 'variative', 'discoverable', 'discoverably', 'discovered', 'discoverer', 'discoverers', 'discoveries', 'discovering', 'discoverist', 'discoverists', 'discoverment', 'discoverments', 'discovers', 'discovery', 'experiment', 'experimental', 'experimentalism', 'experimentalist', 'experimentalists', 'experimentalize', 'experimentally', 'experimentarian', 'experimentarians', 'experimentation', 'experimentations', 'experimentative', 'experimentator', 'experimented', 'experimenter', 'experimenters', 'experimenting', 'experimentist', 'experimentists', 'experimentor', 'experimentors', 'experiments', 'explorability', 'explorable', 'explorable', 'explorate', 'explorates', 'exploration', 'explorationist', 'explorationists', 'explorations', 'explorative', 'exploratively', 'explorator', 'explorators', 'exploratory', 'explore', 'explored', 'explorement', 'explorer', 'explorers', 'explores', 'exploring', 'exploringly', 'flexibility', 'flexible', 'flexibleness', 'flexibly', 'innovate', 'innovated', 'innovates', 'innovating', 'innovation', 'innovational', 'innovationist', 'innovationists', 'innovations', 'innovative', 'innovatively', 'innovativeness', 'innovator', 'innovators', 'innovatory', 'play', 'played', 'player', 'players', 'playful', 'playing', 'playingly', 'playlike', 'plays', 'research', 'risk', 'risked', 'risker', 'riskers', 'riskful', 'riskier', 'riskiest', 'riskily', 'riskiness', 'risks', 'risky', 'search', 'searchable', 'searchableness', 'searched', 'searcher', 'searchers', 'searches', 'searching', 'searchingly', 'variation', 'variational', 'variationally', 'variations', 'variative', 'variatively', 'adapt', 'adapting', 'adaptive', 'adaptors', 'create', 'created', 'creates', 'creating', 'creation', 'creative', 'creator', 'develop', 'developed', 'developer', 'developers', 'developing', 'development', 'developmental', 'develops', 'inventions', 'laboratories', 'laboratory', 'labs', 'patent', 'patented', 'patents', 'pioneer', 'pioneered', 'prospect', 'prospecting', 'prospective', 'prospectively', 'prospects', 'research', 'researcher', 'researchers', 'researching', 'scientist', 'scientists']

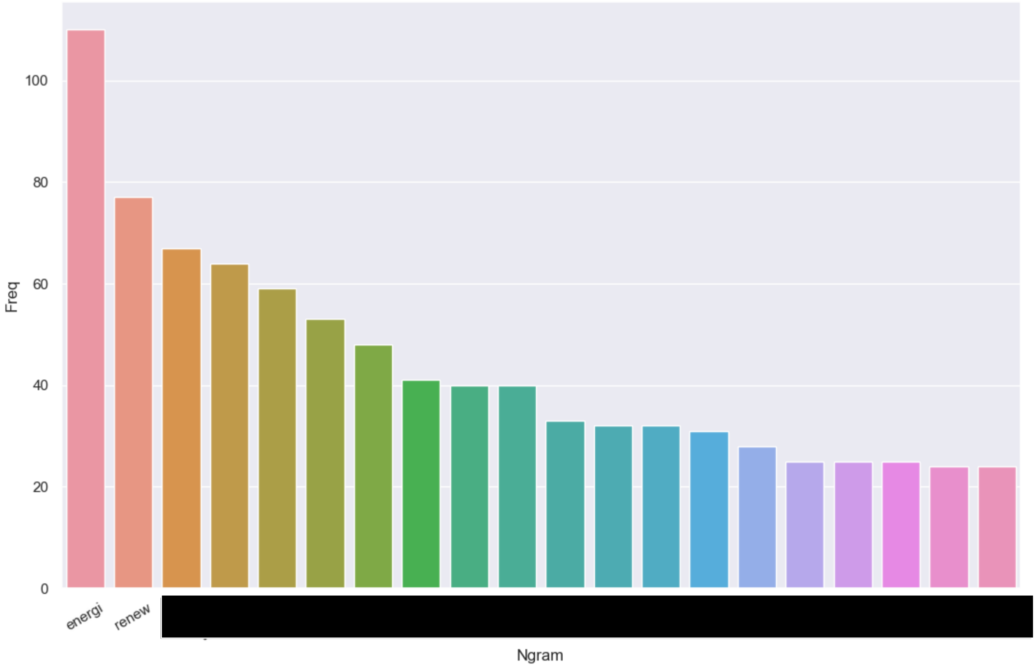
keywordsMossexploit = ['choice', 'choicer', 'choices', 'choicest', 'efficience', 'efficiencies', 'efficiency', 'efficient', 'efficiently', 'executable', 'executant', 'executant', 'executants', 'execute', 'executed', 'executer', 'executers', 'executes', 'executing', 'execution', 'execution', 'executional', 'executioner', 'executioners', 'executions', 'executions', 'executively', 'executiveness', 'executor', 'executorial', 'executors', 'executorship', 'executory', 'exploit', 'exploitability', 'exploitable', 'exploitation', 'exploitational', 'exploitationally', 'exploitations', 'exploitative', 'exploitatively', 'exploitatory', 'exploited', 'exploiter', 'exploiters', 'exploiting', 'exploitive', 'exploitively', 'exploits', 'exploiture', 'implement', 'implementable', 'implemental', 'implementation', 'implemented', 'implementer', 'implementers', 'implementing', 'implementor', 'implementors', 'implements', 'production', 'productional', 'productions', 'productivity', 'refine', 'refined', 'refinedly', 'refinedness', 'refinement', 'refiner', 'refineries', 'refiners', 'refinery', 'refines', 'refining', 'select', 'selectability', 'selectable', 'selected', 'selectedly', 'selecting', 'selection', 'selectional', 'selectionalism', 'selectionist', 'selectionists', 'selections', 'selective', 'selectively', 'selectiveness', 'selectivities', 'selectivity', 'selectly', 'selectness', 'selector', 'selector', 'selectors', 'selectors', 'selects', 'accountant', 'accountants', 'administering', 'administration', 'administrative', 'advertise', 'advertised', 'advertisement', 'advertisements', 'advertiser', 'advertisers', 'advertising', 'assemble', 'assembled', 'assembler', 'assemblers', 'assemblies', 'assembly', 'audited', 'auditing', 'auditors', 'audits', 'automate', 'automated', 'automatic', 'automatically', 'automating', 'automation', 'commercialization', 'commercialize', 'commercialized', 'commercializing', 'commercials', 'commoditized', 'commoditizing', 'commodity', 'conventional', 'deploy', 'deployable', 'deployed', 'deploying', 'deployment', 'deployments', 'distributor', 'distributors', 'increment', 'incremental', 'incrementally', 'increments', 'launch', 'launched', 'launches', 'maintain', 'maintained', 'maintaining', 'maintains', 'manufacture', 'manufactured', 'manufacturer', 'manufacturers', 'manufacturing', 'marketed', 'marketer', 'marketers', 'marketing', 'optimization', 'optimize', 'optimizer', 'optimizing', 'optimum', 'procured', 'procurement', 'promotion', 'promotional', 'promotions', 'replicated', 'replication', 'replicators', 'routine', 'routinely', 'salesforce', 'salespeople', 'salespersons', 'standardized', 'throughput']

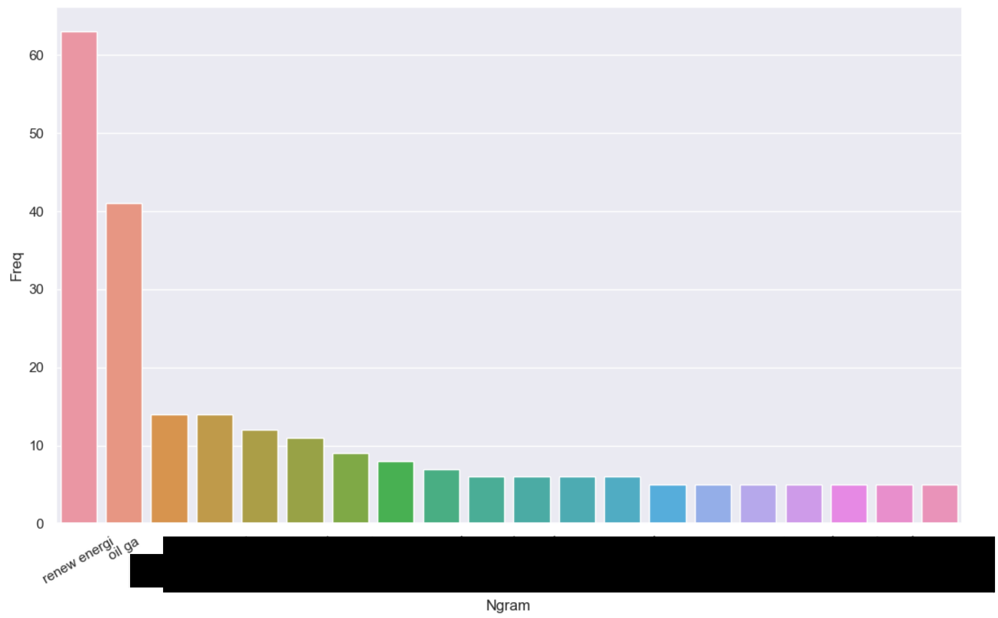
keywordsMcKennyexplor = ['beta-phase', 'beta-testing', 'breakthrough', 'breakthroughs', 'clinical studies', 'clinical study', 'clinical test', 'clinical testing', 'clinical tests', 'clinical trial', 'clinical trials', 'creative', 'develop', 'developed', 'developing', 'development', 'developmental', 'developments', 'develops', 'experiment', 'experimental', 'experimentalism', 'experimentalist', 'experimentalists', 'experimentalize', 'experimentally', 'experimentarian', 'experimentarians', 'experimentation', 'experimentations', 'experimentative', 'experimentator', 'experimented', 'experimenter', 'experimenters', 'experimenting', 'experimentist', 'experimentists', 'experimentor', 'experimentors', 'experiments', 'innovate', 'innovated', 'innovates', 'innovating', 'innovation', 'innovations', 'innovative', 'innovativeness', 'innovator', 'innovators', 'innovatory', 'inventions', 'ipr&d', 'iprd', 'laboratories', 'laboratory', 'labs', 'launch', 'launched', 'launches', 'launching', 'new drug', 'new drugs', 'new generic product', 'new generic products', 'new mobile product', 'new mobile products', 'new offering', 'new offerings', 'new product', 'new products', 'new program', 'new programming', 'new programs', 'new system', 'new systems', 'new technologies', 'new technology', 'novel', 'patent application', 'patent applications', 'patent development', 'patent developments', 'phase 1', 'phase 1a', 'phase 1b', 'phase 2', 'phase 2a', 'phase 2b', 'phase 3', 'phase 4', 'phase i', 'phase i/ii', 'phase ia', 'phase ib', 'phase ii', 'phase iia', 'phase iib', 'phase iii', 'phase iv', 'pioneer', 'pioneered', 'preclinical', 'pre-clinical', 'proof of concept', 'prototype', 'prototypes', 'prototyping', 'r&d', 'research', 'researching', 'unveiled']

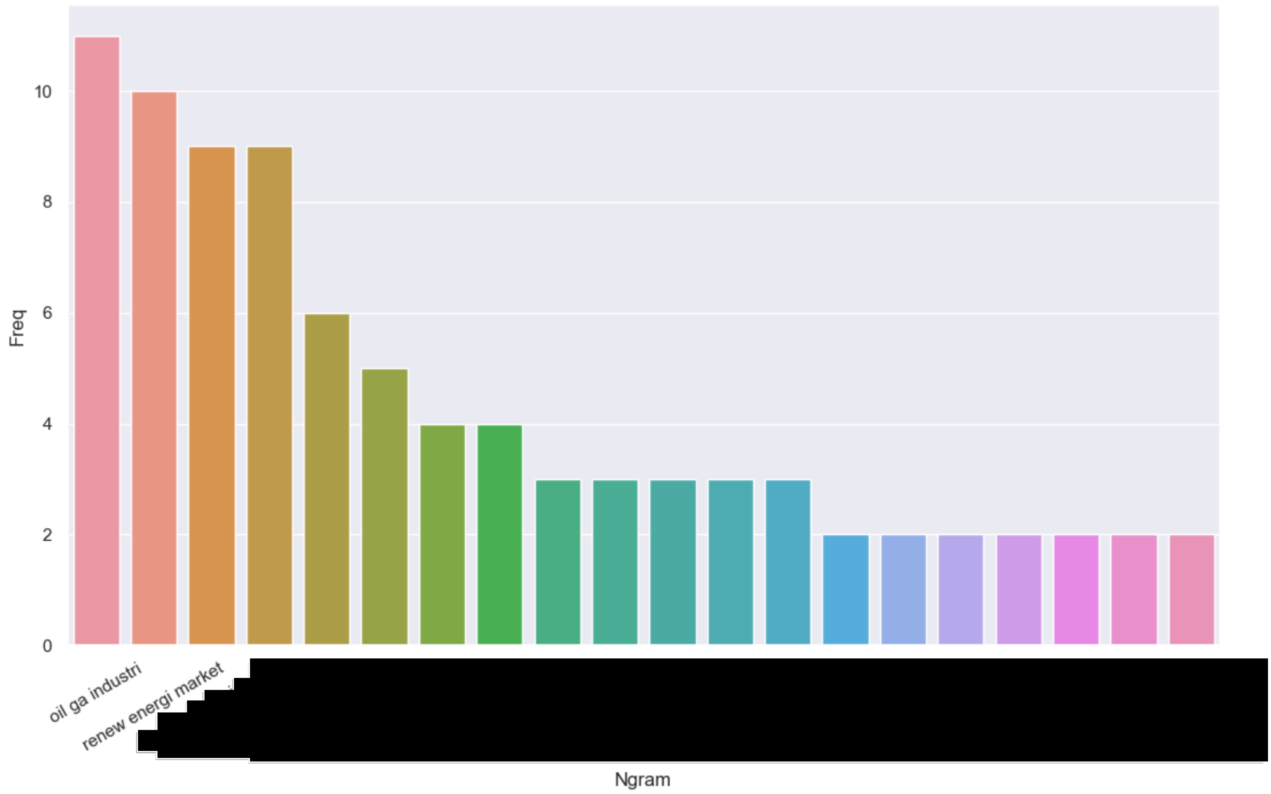
keywordsMcKennyexploit = ['adaptations', 'advertising', 'commercialization', 'commercialize', 'commercialized', 'commercializes', 'commercializing', 'commoditized', 'commoditizing', 'current offering', 'current offerings', 'current product', 'current products', 'efficience', 'efficiencies', 'efficiency', 'efficient', 'efficiently', 'existing offering', 'existing offerings', 'existing product', 'existing products', 'existing technology', 'exploit', 'exploitability', 'exploitable', 'exploitation', 'exploitational', 'exploitationally', 'exploitations', 'exploitative', 'exploitatively', 'exploitatory', 'exploited', 'exploiting', 'exploitive', 'exploitively', 'exploits', 'exploiture', 'extension', 'extensions', 'implement', 'implementable', 'implemental', 'implementation', 'implementations', 'implemented', 'implementer', 'implementers', 'implementing', 'implementor', 'implementors', 'implements', 'integrate', 'integration', 'maintenance', 'manufacture', 'manufactured', 'manufacturing', 'marketed', 'marketing', 'new features', 'new formulation', 'new formulations', 'new indication', 'new indications', 'optimization', 'optimize', 'optimized', 'optimizes', 'optimizing', 'optimum', 'produce', 'produced', 'produces', 'producing', 'production', 'productions', 'productivity', 'promotion', 'promotional', 'promotions', 'redesign', 'reengineering', 're-engineering', 'refine', 'refined', 'refinedly', 'refinedness', 'refinement', 'refinements', 'refines', 'refining', 'reformulated', 'reformulating', 'reformulation', 'refreshed', 're-launch', 'replicated', 'replication', 'replicators', 'retooled', 'salesforce', 'salespeople', 'salespersons', 'standardized', 'streamline', 'throughput', 'upgrade', 'upgraded', 'upgrades', 'upgrading', 'version', 'versions']

Appendix 6 Overview of Ngrams

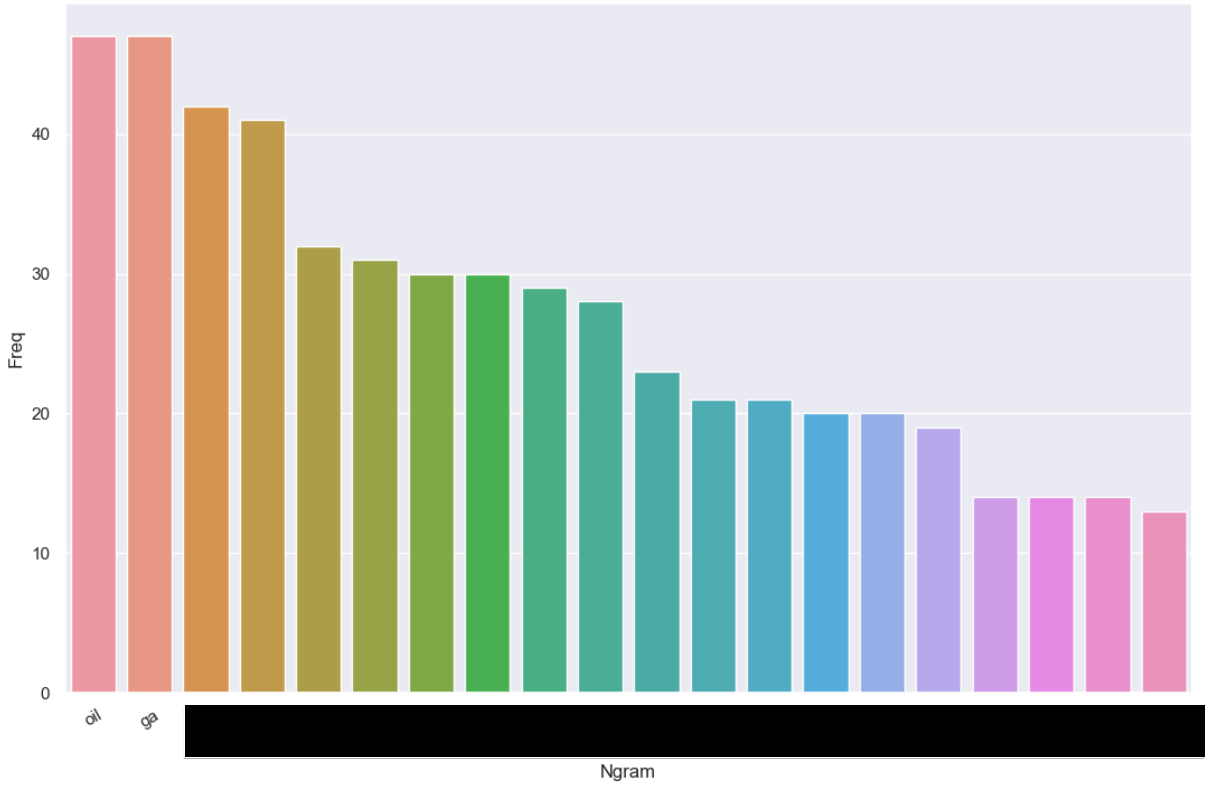
Novel sample

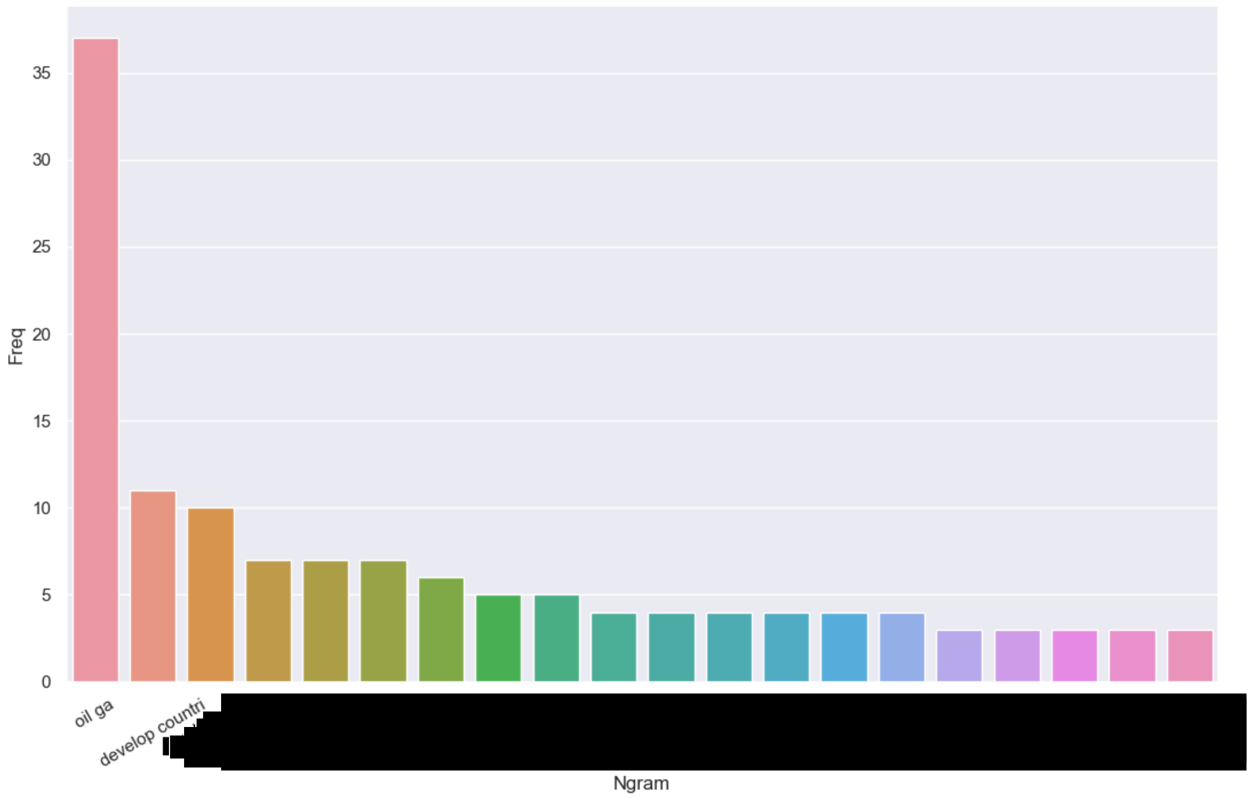


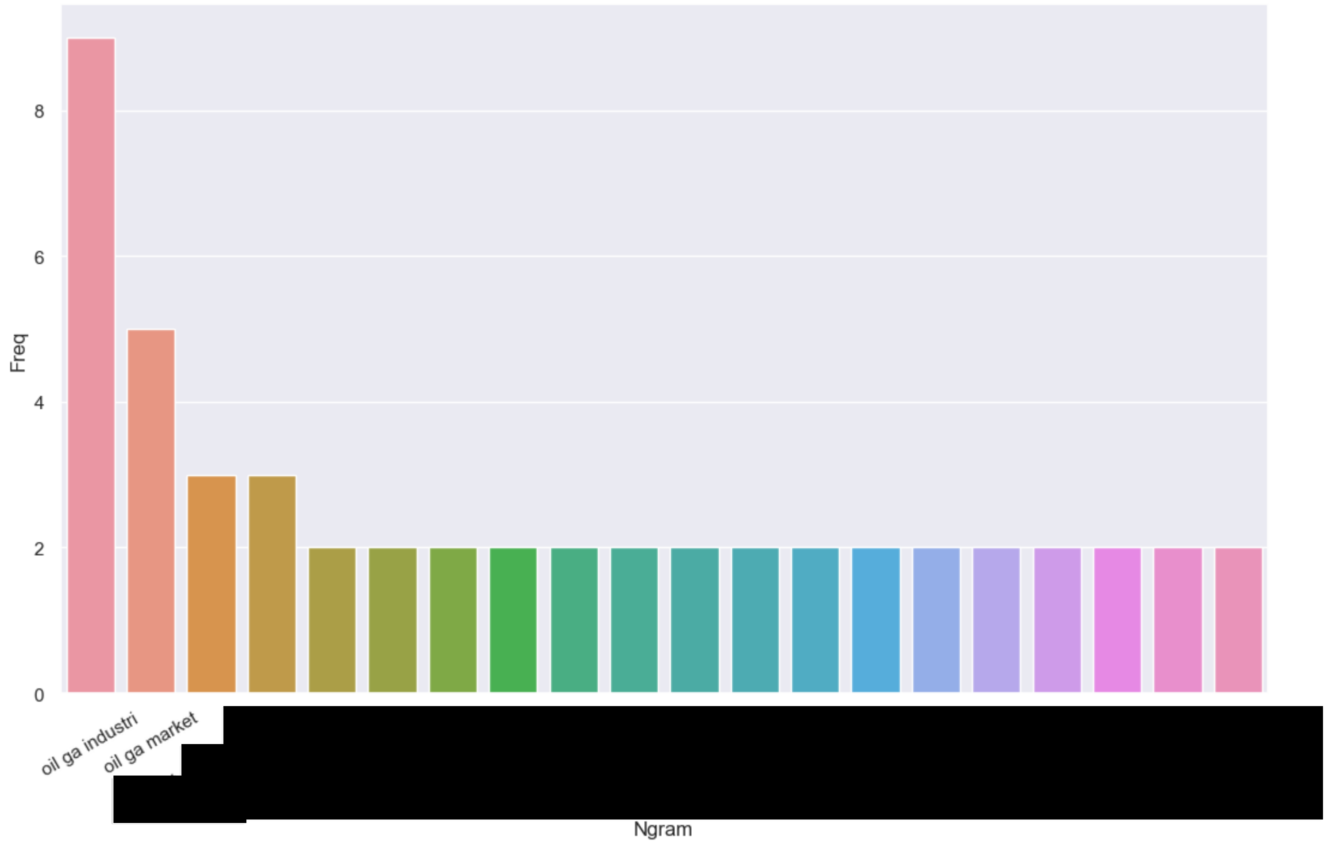



Freq
10
8
6
4
2
0
oil ga industri
renew energi market
[redacted]
Ngram

Not novel sample

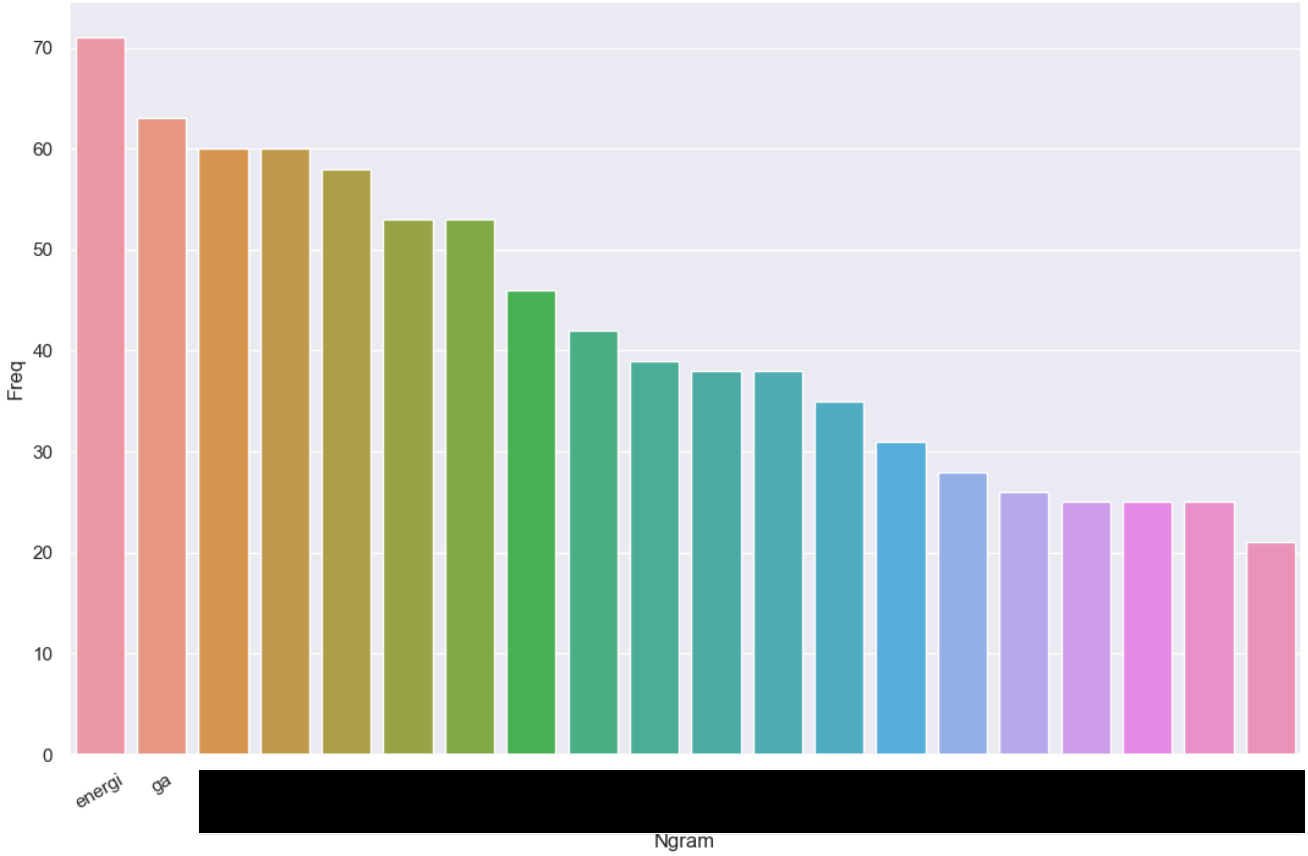


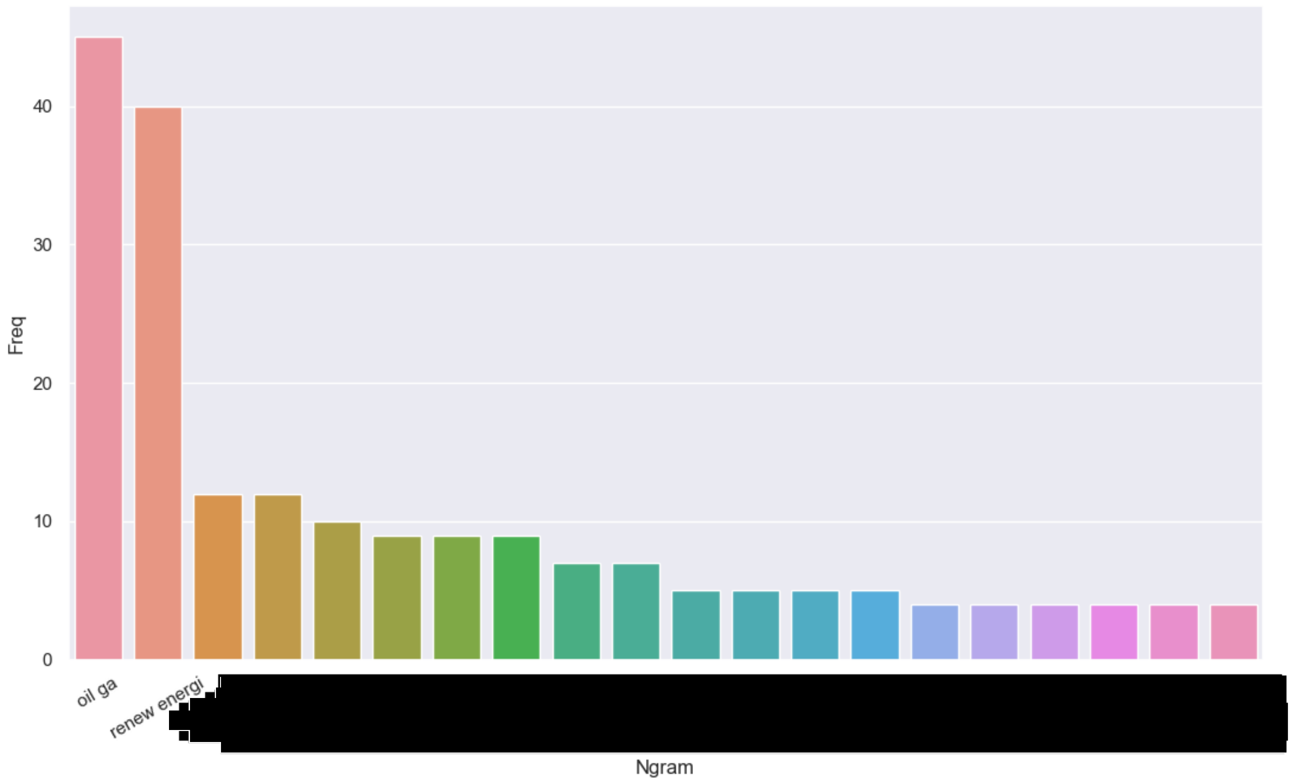


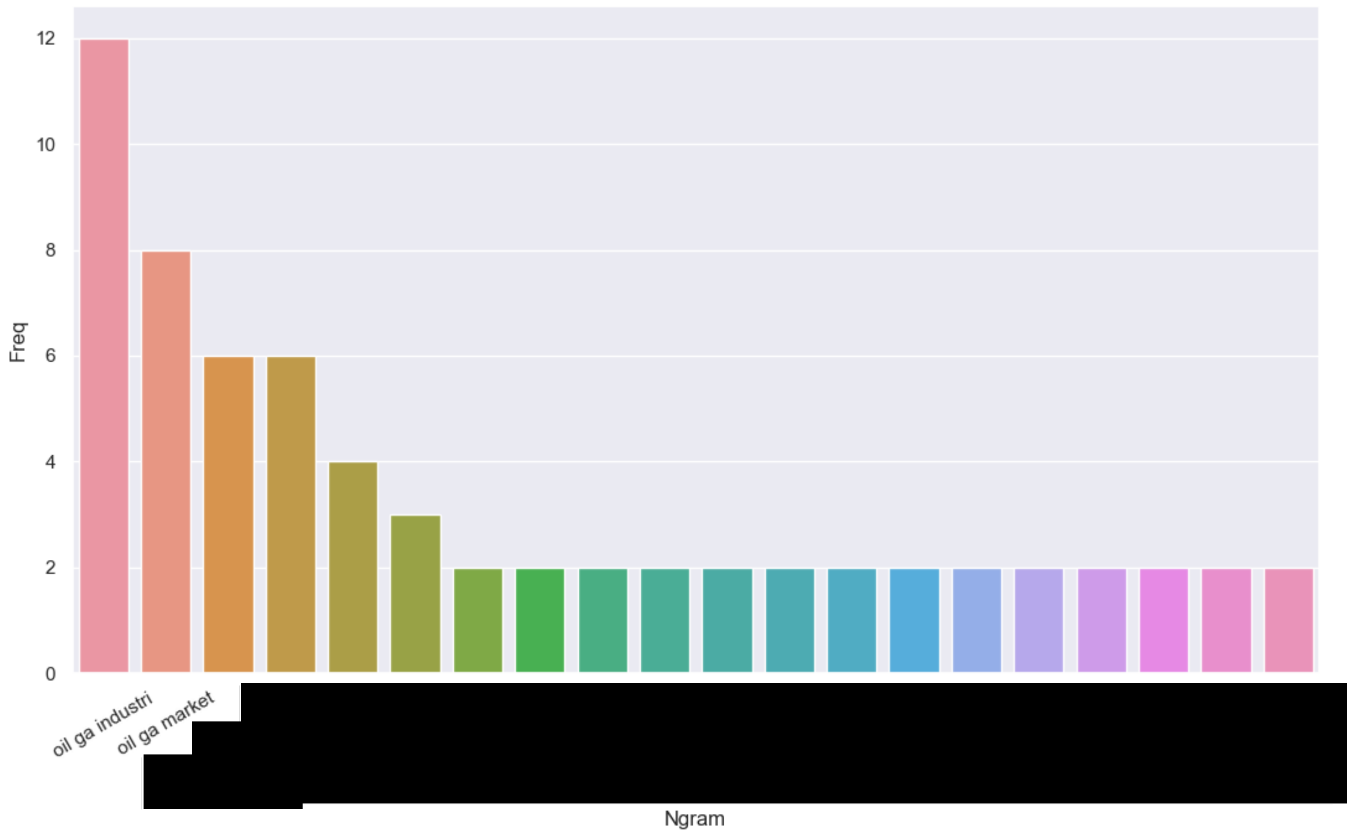


8
6
4
2
0
Freq
oil ga industri
oil ga market
[redacted]
Ngram

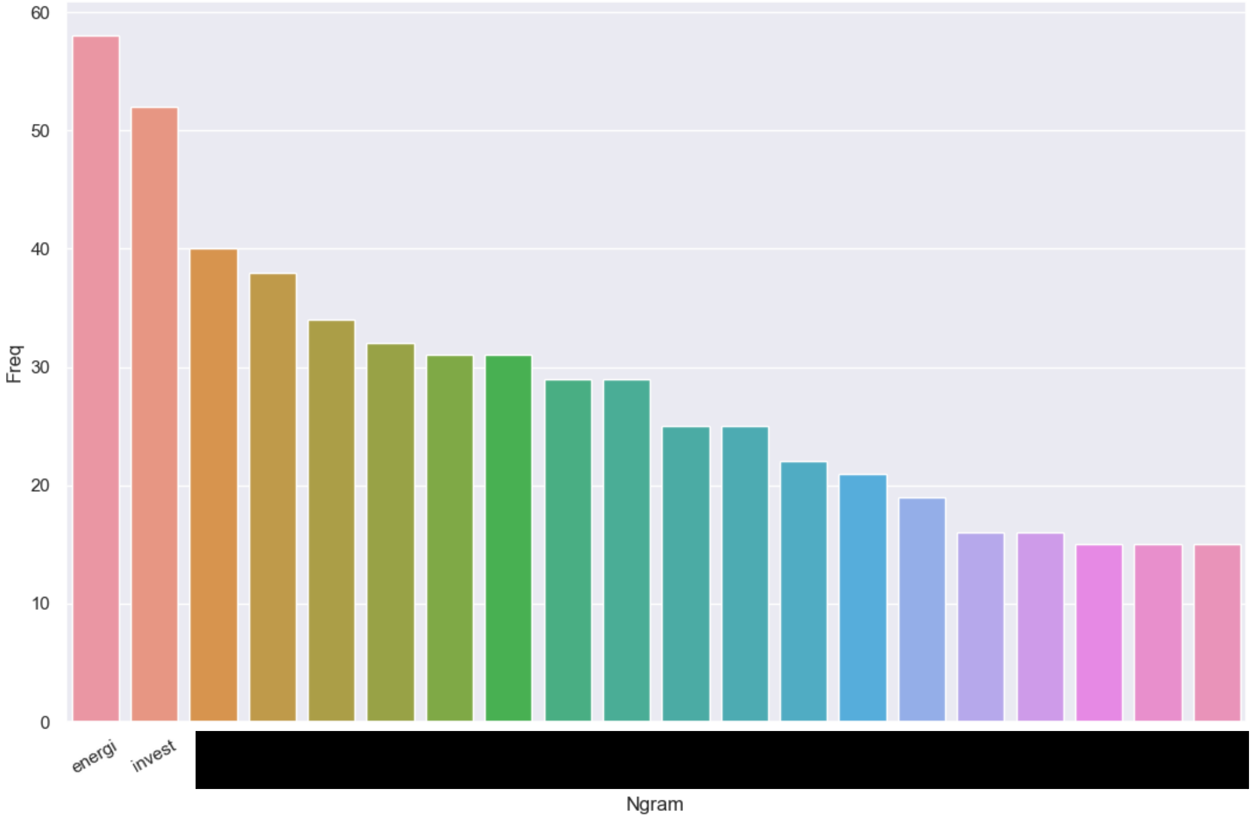
High value sample

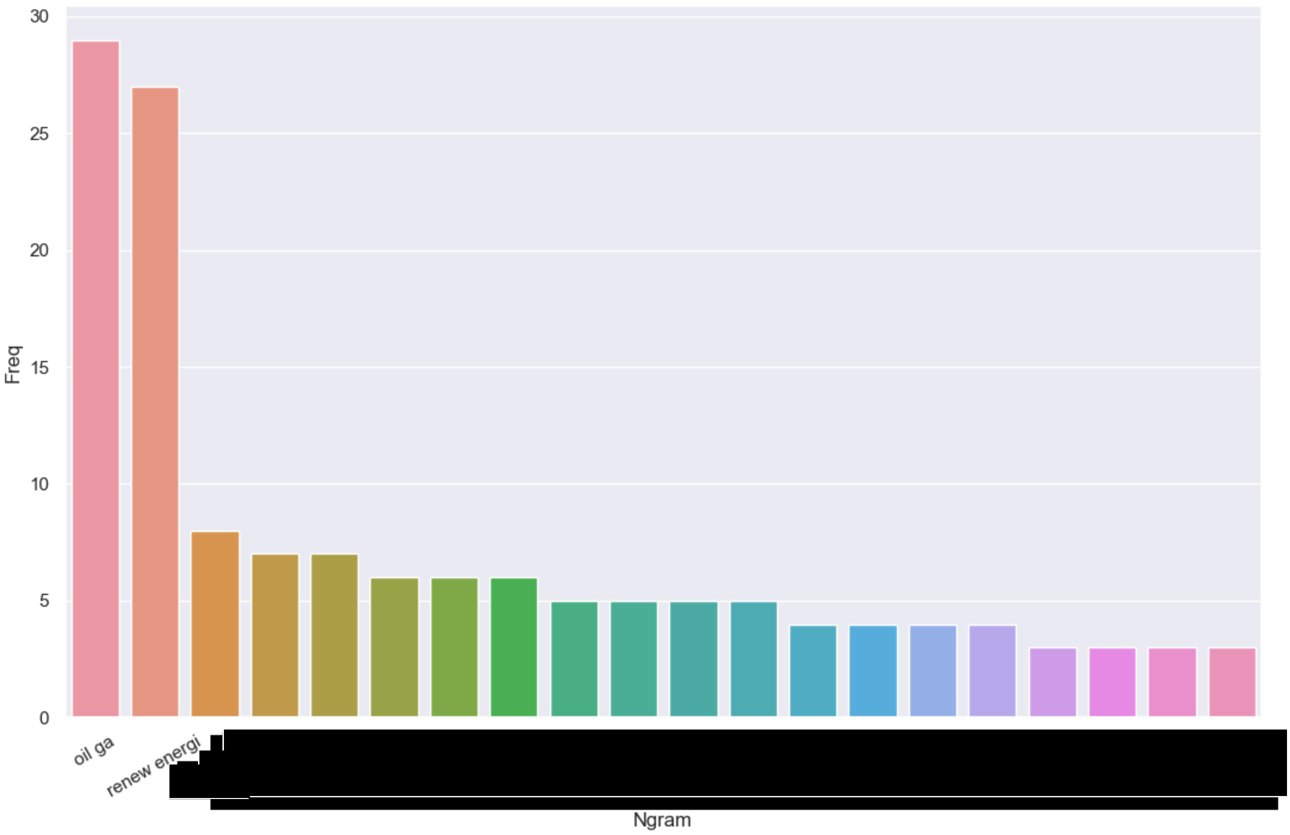


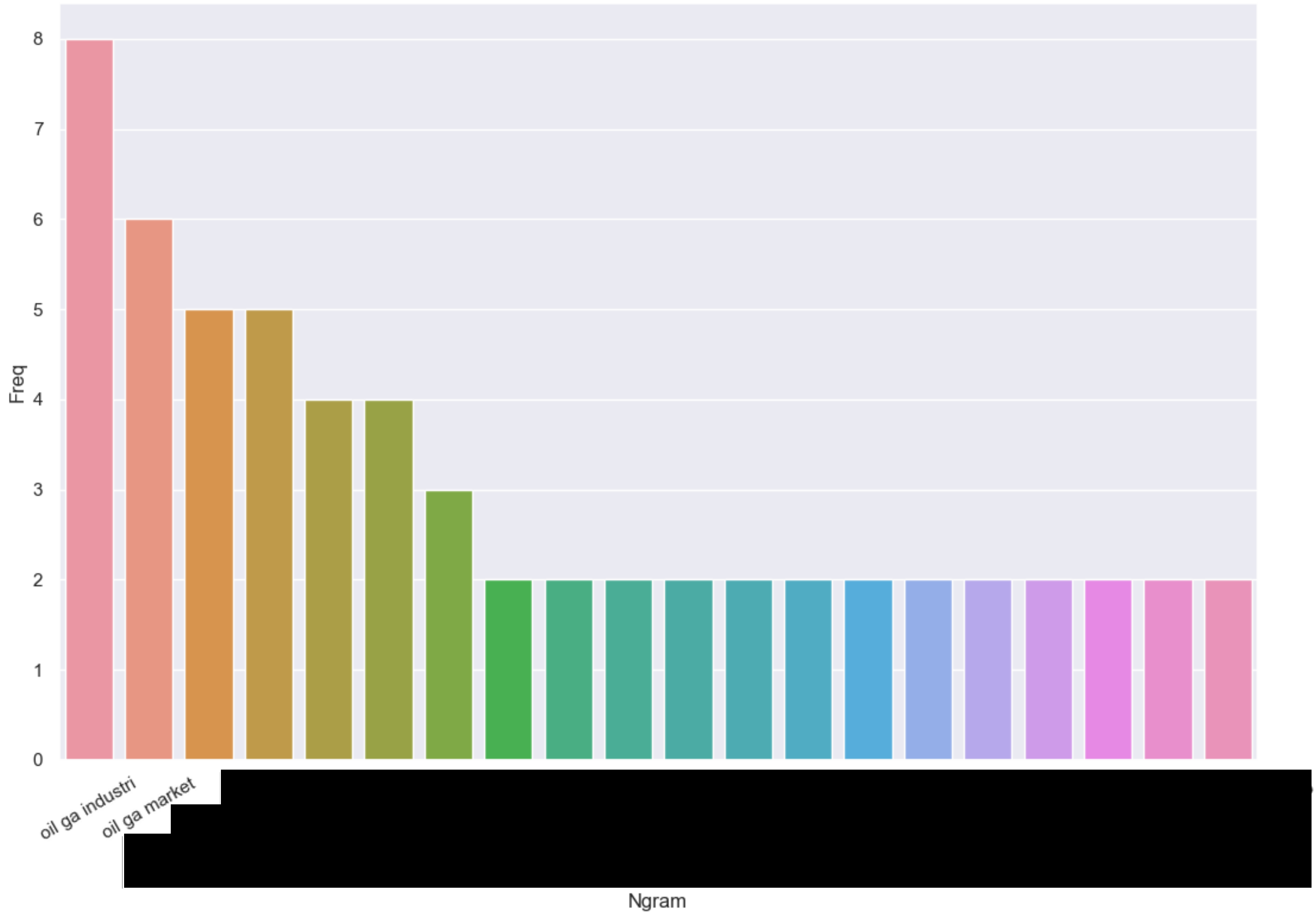



12
10
8
6
4
2
0
Freq
oil ga industri
oil ga market
[redacted]
Ngram

Low value sample



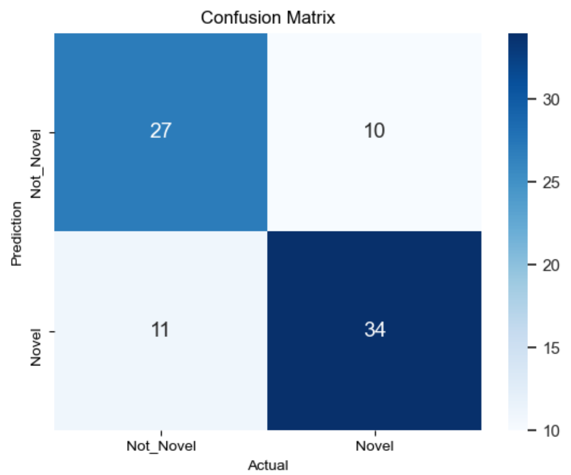



Freq
8
7
6
5
4
3
2
1
0
oil ga industri
oil ga market
[redacted]
Ngram

Appendix 7 Confusion Matrices of Supervised Classification on Novelty

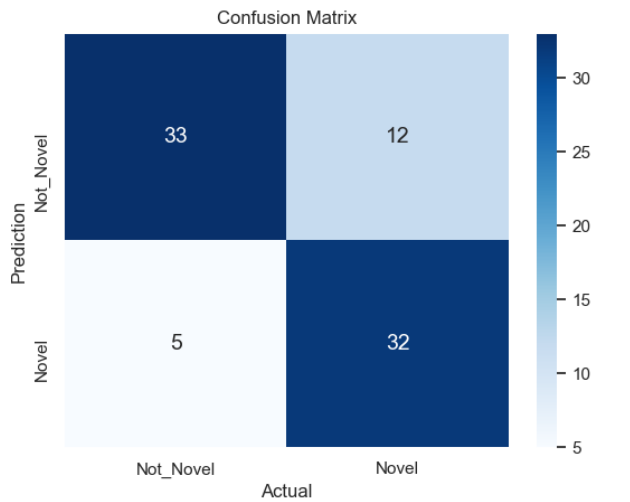
Naïve Bayes

BoW




```
F1: 0.764 | Pr: 0.756 | Re: 0.773 | AUC: 0.742 | Accuracy: 0.744
```

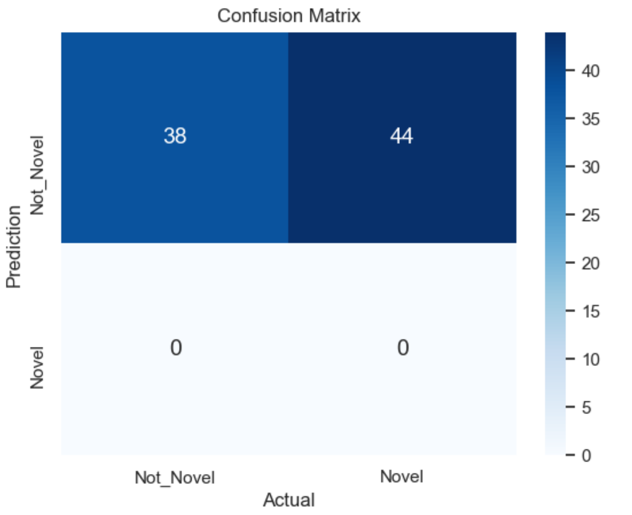
TFIDF



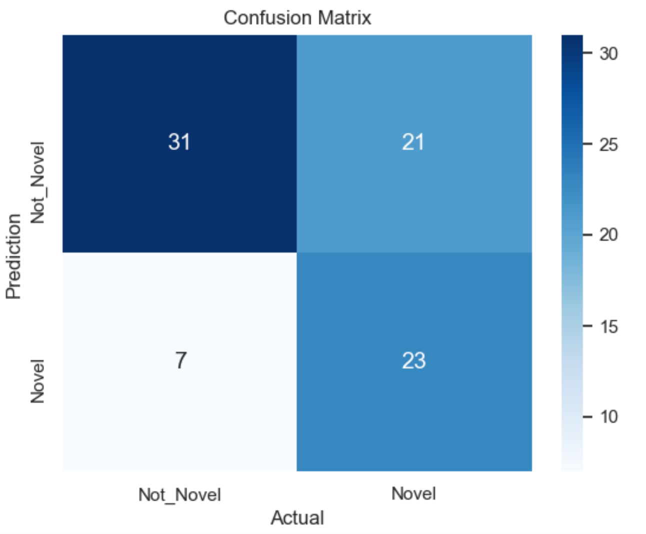

```
F1: 0.790 | Pr: 0.865 | Re: 0.727 | AUC: 0.798 | Accuracy: 0.793
```

Tree

For BoW encodings the decision tree predicts everything as not novel.



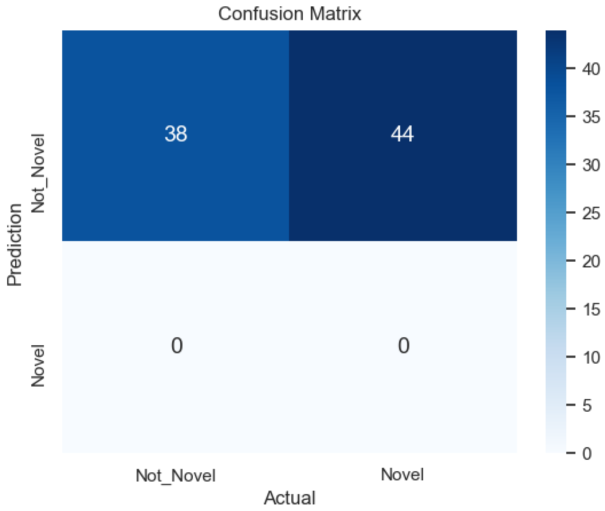
TFIDF




```
F1: 0.622 | Pr: 0.767 | Re: 0.523 | AUC: 0.669 | Accuracy: 0.659
```

SVM/Random Forest

Both the SVM and the Random Forest classify everything as not novel, for both the BoW and TFIDF word representations.

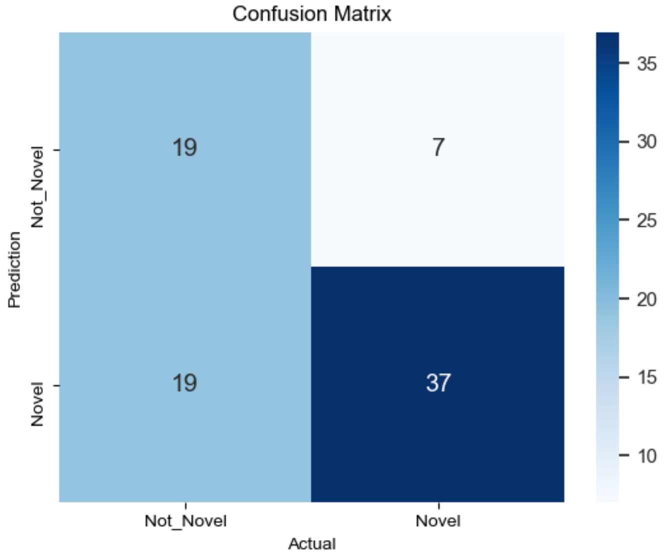


Confusion Matrix
38
44
0
0
Not_Novel
Novel
Prediction
Not_Novel
Novel
Actual
40
35
30
25
20
15
10
5
0

Results after removal of distinct terms

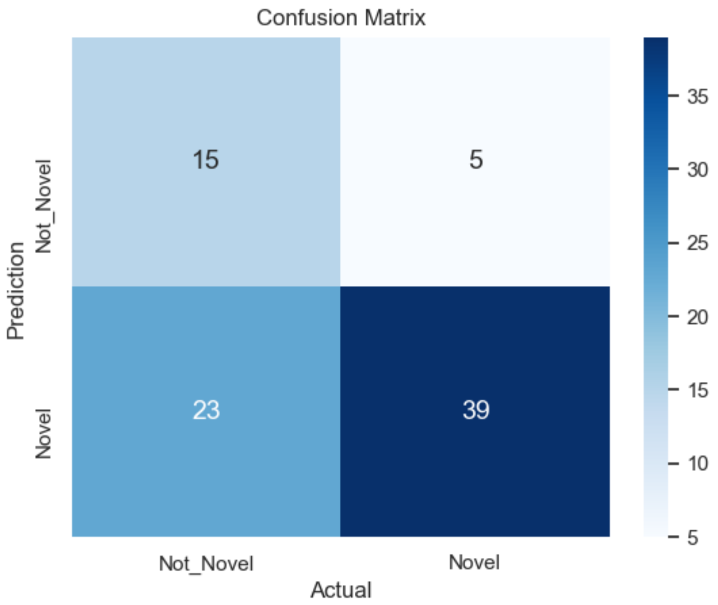
Naïve Bayes

BoW




```
F1: 0.740 | Pr: 0.661 | Re: 0.841 | AUC: 0.670 | Accuracy: 0.683
```

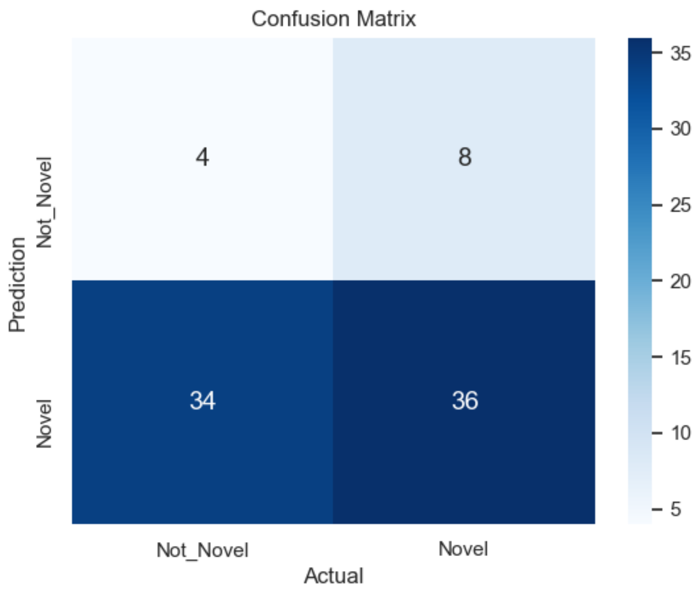
TFIDF




```
F1: 0.736 | Pr: 0.629 | Re: 0.886 | AUC: 0.641 | Accuracy: 0.659
```

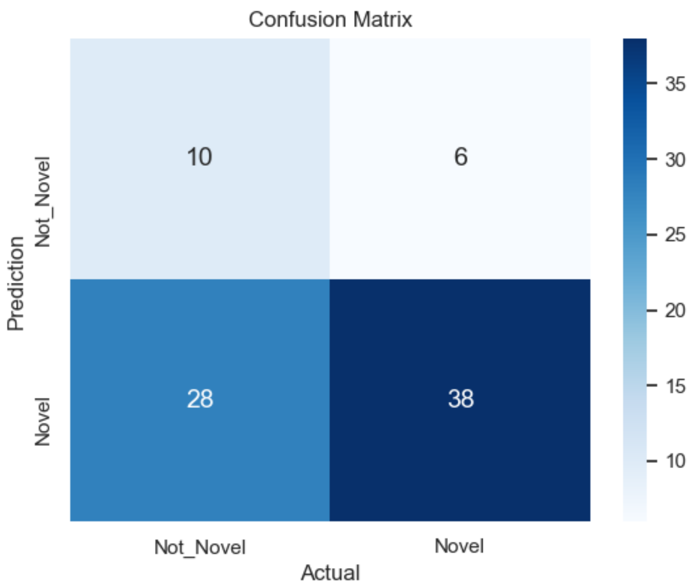
Tree

BoW



F1: 0.632 | Pr: 0.514 | Re: 0.818 | AUC: 0.462 | Accuracy: 0.488

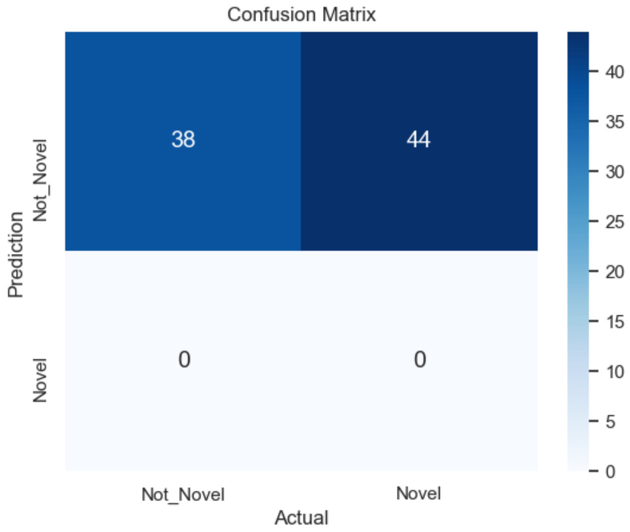
TFIDF



F1: 0.691 | Pr: 0.576 | Re: 0.864 | AUC: 0.563 | Accuracy: 0.585

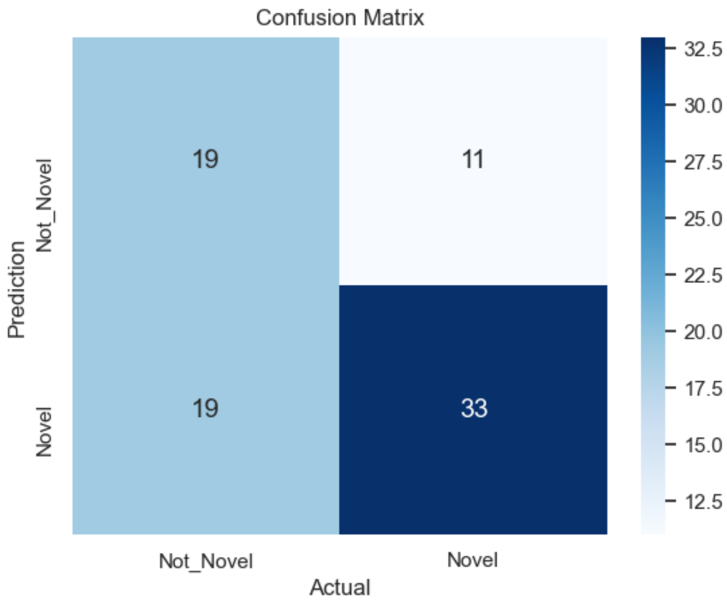
SVM

BoW



F1: 0.000 | Pr: 0.000 | Re: 0.000 | AUC: 0.500 | Accuracy: 0.463

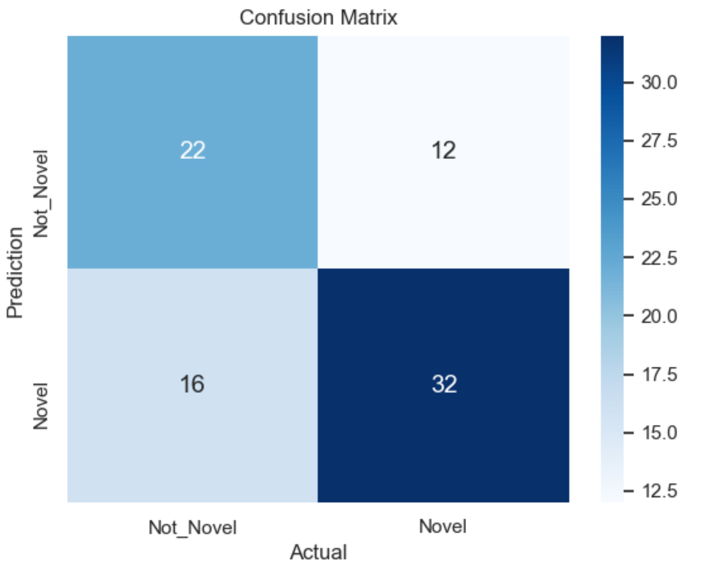
TFIDF



F1: 0.688 | Pr: 0.635 | Re: 0.750 | AUC: 0.625 | Accuracy: 0.634

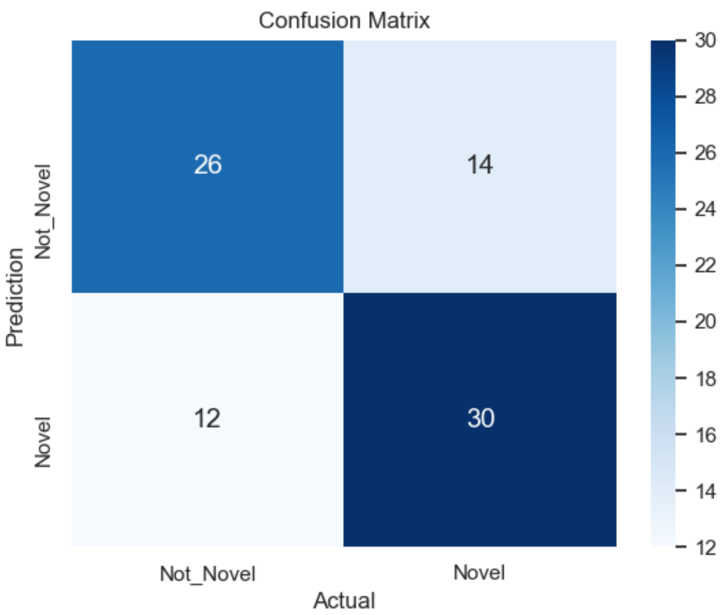
Random Forest

BoW




```
F1: 0.696 | Pr: 0.667 | Re: 0.727 | AUC: 0.653 | Accuracy: 0.659
```

TFIDF

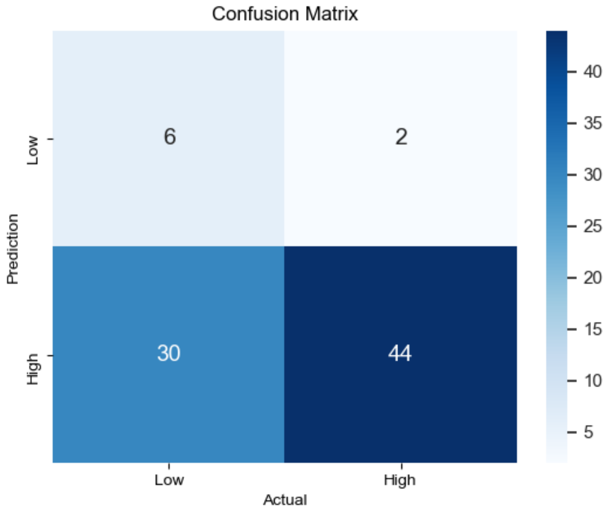



```
F1: 0.698 | Pr: 0.714 | Re: 0.682 | AUC: 0.683 | Accuracy: 0.683
```

Appendix 8 Confusion Matrices of Supervised Classification on Value

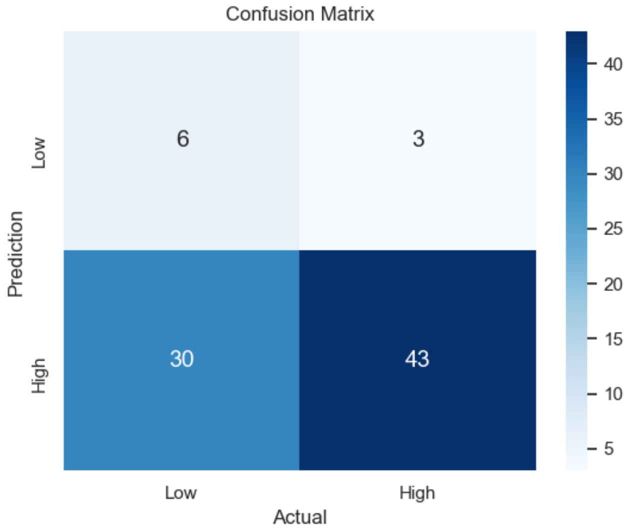
Naïve Bayes

BoW



F1: 0.733 | Pr: 0.595 | Re: 0.957 | AUC: 0.562 | Accuracy: 0.610

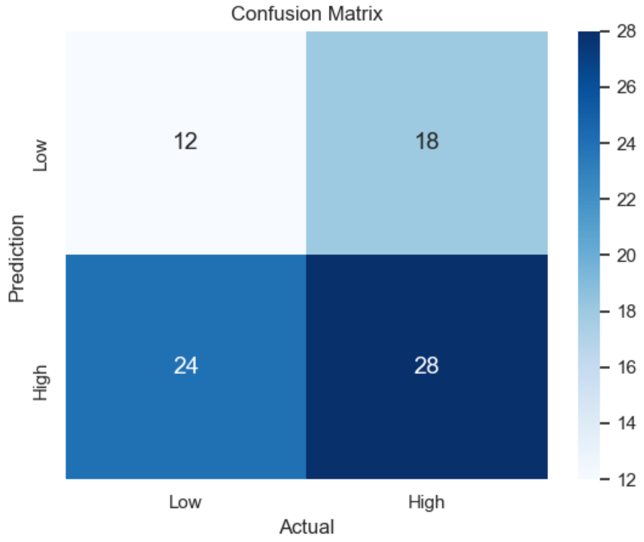
TFIDF



F1: 0.723 | Pr: 0.589 | Re: 0.935 | AUC: 0.551 | Accuracy: 0.598

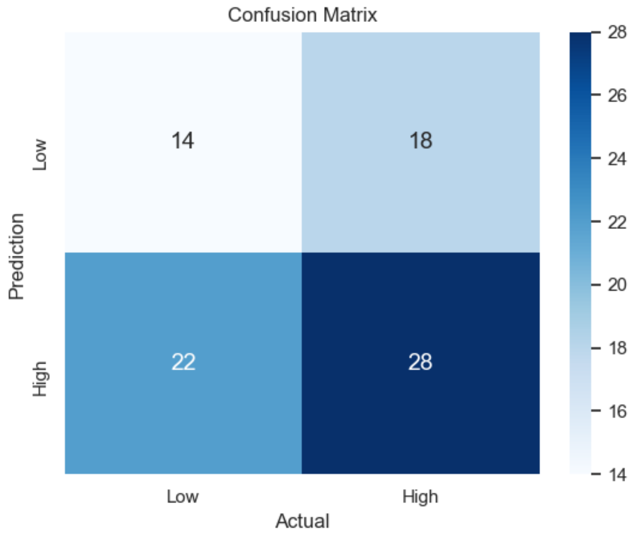
Tree

BoW




```
F1: 0.571 | Pr: 0.538 | Re: 0.609 | AUC: 0.471 | Accuracy: 0.488
```

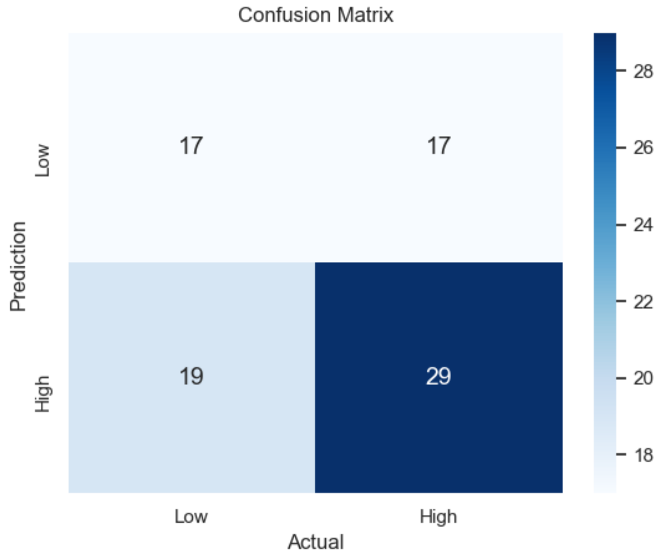
TFIDF




```
F1: 0.583 | Pr: 0.560 | Re: 0.609 | AUC: 0.499 | Accuracy: 0.512
```

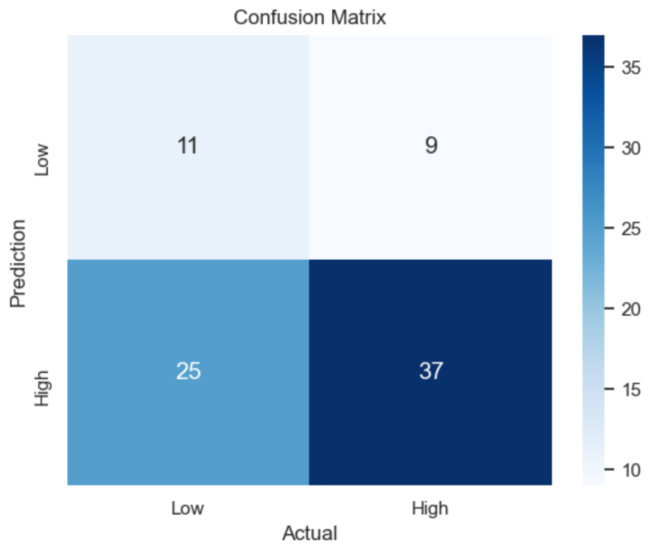
SVM

BoW



`F1: 0.617 | Pr: 0.604 | Re: 0.630 | AUC: 0.551 | Accuracy: 0.561`

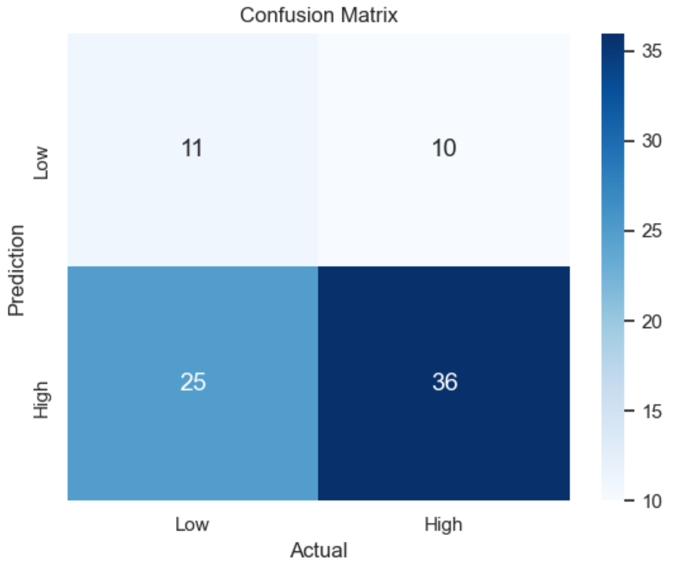
TFIDF



`F1: 0.685 | Pr: 0.597 | Re: 0.804 | AUC: 0.555 | Accuracy: 0.585`

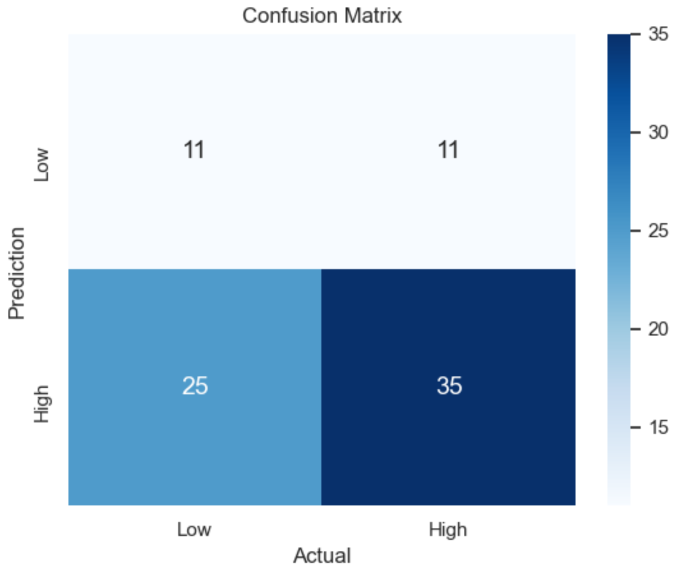
Random Forest

BoW




```
F1: 0.673 | Pr: 0.590 | Re: 0.783 | AUC: 0.544 | Accuracy: 0.573
```

TFIDF




```
F1: 0.660 | Pr: 0.583 | Re: 0.761 | AUC: 0.533 | Accuracy: 0.561
```

Appendix 10 Supervised Classifier

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

```
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]test)
print("End classifiers")
```

Appendix 11 Ngram finder and Clustering

```
[redacted]
```

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Appendix 12 Countfinder

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]