Guilherme Nuno Campos Góis

Licenciado em Ciências da Engenharia

Sentiment Analysis on Product-Service Systems

Dissertação para obtenção do Grau de Mestre em

Engenharia Eletrotécnica e de Computadores

Orientador: Rui Alexandre Nunes Neves da Silva, Associate Professor,
NOVA University of Lisbon

Júri

Presidente: Professora Doutora Maria Helena Fino
Arguente: Professor Doutor André Damas Mora

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2018

Sentiment Analysis on Product-Service Systems

Este documento foi gerado utilizando o processador (pdf)LaTeX, com base no template "novathesis" [1] desenvolvido no Dep. Informática da FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

To everyone that is a part of my life. Specially dedicated to my grandparents.

Acknowledgements

"No one can pass through life, any more than he can pass through a bit of country, without leaving tracks behind, and those tracks may often be helpful to those coming after him in finding their way.- Robert Baden-Powell.

Durante a minha vida tive o privilégio de ter vários trilhos para me ajudarem no meu caminho. Muitas pessoas contribuíram para este caminho sendo que as citadas neste documento são apenas algumas das que mais me marcaram. Cumpre-me, antes de mais, agradecer a Deus, por ter colocado todas estas pessoas no meu caminho.

Em primeiro lugar, quero agradecer à minha mãe Marina, que durante toda a vida lutou para garantir que conseguia ultrapassar todas as adversidades; ao Jorge, meu "pai afetivo", que desde muito cedo me considerou e me tratou como se seu filho fosse; aos meus avós Alcina e António Campos, que sempre me "estragaram"com mimos; ao meu pai Nuno, que soube balancear o ser confidente com o "dar na cabeça"; ao meu irmão Tomás, puto regila e sem emenda, que me ajudou a crescer com as suas birras; e, ao meu padrinho José Carlos Castanheira, pelo exemplo de trabalho e determinação que sempre foi para mim.

Num plano seguinte, mas talvez tão importante quanto o primeiro, quero agradecer a todos os meus amigos que me acompanharam neste percurso. Neste âmbito destaco o António Ferreira, por todas as aventuras que vivemos e mais recentemente pelas refeições revitalizantes que prepara; a Pipa, que me ajudou a ser capaz de viver com níveis elevados de ruído; o Diogo Vieira, que conheço desde miúdo e com quem partilhei grandes momentos da minha juventude; o Paulo Pereira e à sua rigidez militar; a Mafalda Martins, rapariga a quem agradeço especialmente o ter-me aturado na adolescência; o Flávio Jacinto, que me apresentou a realidade dos trabalhos de grupo da faculdade; o Ricardo Ferreira, por me ter apresentado desafios nos trabalhos de grupo; o Rui Baptista e à sua sempre alegre disposição; ao Gabriel grande colega e festivaleiro; a Lúcia Salvador e a Margarida Freixo, pelas casas nos dias de festa.

Agradeço também a todos os escuteiros agrupamento 1100 Parque das Nações, pelo crescimento pessoal e desenvolvimento de competências que me proporcionaram, bem como ao Rui Neves da Silva, pela orientação na escrita desta dissertação, e também conjuntamente ao Paulo Pina e à Rita Campos, que me possibilitaram a participação no Projeto *Diversity* em que desenvolvemos este projeto.

Por fim, não deixarei de agradecer à gloriosa FCT (e a todos os seus docentes e equipas

de apoio), por ser ontem, hoje e amanhã a casa daqueles que por lá passam.

Abstract

The main goal of this dissertation is to develop a tool to help each company reduce the amount of failed product-service systems that were avoidable due previous experience. By using tools and ideas already available and build them in a way they can interact with each other, this tool aims to give designers a better as faster way to view data. This was identified as a possible improvement since for the past 20 years the economy evolved into a consumer driven market, this led to the development of an extremely competitive economy. Companies need to strive for innovation and quality of products and services, faster than never. Products and services also need to match the expectations and needs of customers. Analyzing where product and service systems are lacking in terms of customer requirements is crucial. Currently it might take some time for information to travel from customer to producer, since the connection may include stores and local representatives before reaching the products' and services' designers. Although this information is readily available in social networks, the issue resides in efficiently merging and showing it in a simple and meaningful way to the designer of new products and systems. By shortening the time spent for information travel between costumer and producer, might lead to better and more innovative products.

Keywords: sentiment; opinion; product-service system; social networks; prediction;

Resumo

O principal objetivo desta tese e desenvolver ferramenta que ajude as empresas a reduzir a quantidade de Sistemas Produto-Serviço falhados, que são evitaveis devido a experiências anteriores. Ao utilizar ferramentas e ideias já existes e desenha-las de uma forma integrada esta nova plataforma trará aos designers uma forma mais rápida de visualizar dados. Isto foi identificado como possível melhoria dado que nos últimos 20 anos a economia evoluiu para um marcado guiado pelo consumo.

Esta realidade levou a que os diversos mercados se tornassem muito mais competitivos. Às empresas impõe-se uma focalização na inovação e na garantia da qualidade dos seus produtos e serviços, os quais terão de ser adequados às necessidades do público alvo. Esta realidade levou a que os diversos mercados se tornassem muito mais competitivos. Às empresas impõe-se uma focalização na inovação e na garantia da qualidade dos seus produtos e serviços, os quais terão de ser adequados às necessidades do público alvo. Apesar desta informação ficar rapidamente disponível nas redes sociais, o problema prende-se com a capacidade de analisar toda essa informação de forma organizada e facilmente percetível. Ao encurtar-se o tempo de entre a partilha de opiniões entre consumidor e produtor, potencia-se a criação de melhores sistemas produto-serviço. Se, além de se encurtar esta barreira temporal, conseguirmos também identificar palavras chave e quantificá-las, a informação partilhada com os produtores torna-se de mais fácil leitura, potenciando uma maior agilidade e rapidez na reação.

Palavras-chave: sentimento; opinião; sistema produto-serviço; redes sociais; predição;

Contents

List of Figures

List of Tables

Acronyms

CSM Cosine-Similarity Measure.

LCA Lowest Common Ancestor.

NLP Natural Language Processing.

OSM Owned Social Media.

PDM Product Data Management.
PSS Product-Service System.

R&D Research and Development.

SFE Social Feedback Extraction.
SPL Software Product Line.

Chapter 1

Introduction

This chapter contains the motivation that led to the work done, contribution to of the dissertation and an overview of the document structure.

1.1 Motivation

With every passing year, Internet is a bigger part of our lives. Almost everyone in Developing and Developed countries has a social media footprint. Everyone has an opinion, a voice, thoughts that can be shared throughout continents fast and effortless with a click of a button. For companies the perception of their users is important, but filtering throughout everyone's opinion takes time, and sometimes not all of them have the same importance.

Sentiment Analysis has been done by companies for years, trying to achieve better and more efficient products. Either by feedback forms, or in recent years by attending to costumer complaints and approvals on social media. The latest can take a lot of time given the number of costumers a company might have.

Each day more products are launched into the market, these were designed by companies trying to thrive in a consumer driven economy. At the same time costumers keep asking for newer products, services and combinations. Let's take the example of phones, in the past 20 years costumers went from wanting a phone just for calls, to a mini portable computer that also receives calls. The key to success is to develop new and better products, this is what this tool is aimed at. This tool was also developed in partnership with the project Diversity, project funded by the European Union's Horizon 2020 GA no:636692.

1.2 Contributions

Developing this tool might help companies reduce cost and improve on existing and future products. Designing an environment where multiple tools may be integrated, and companies can store knowledge for future use, helps companies design better and more successfully products for the costumer. By designing a tool capable of the previous statement, it is also expected that the market might take a better approach into Product-Service System creation.

1.3 Document Overview

The document starts with this chapter, where the motivation, as well as the contributions for the dissertation are presented. In the second chapter, the related state of the art is presented, giving an overview of network relations, sentiment extraction algorithms and PSS definition followed by the approach methodology on the third chapter. Server Architecture is explained on chapter 3 where the deployment and operations of the platform are fully detailed. The Code Structure that allows this platform to analyze big amounts of data on chapter 5 and consequent results on the following chapter, being the document finished with the conclusions and future work in chapter 7.

Chapter 2

State of the Art

In this chapter there will be provided current information about each part of the final solution, as how social networks are build and organized. What is a Product-Service System, and when did the concept emerge. What is Sentiment Analysis, and why should the companies investigate this concept. Natural Language Processing and current solution. The benefits of analyzing products similarities. Finally, what is visual analytics and what should considered when using it.

2.1 Social Networks

Social Network Theory and Analysis, these are the area that study the currently emerging networks.
Network communication can be found all around us, human bodies have them [21], physics, politics, computer science, etc. Socially, in later years, networks have been developed using many important websites like Facebook, Twitter, Instagram. This has created a new opportunity for marketing and analysis. Since this work is mainly focused on feedback this research will focus more on that topic.

Before defining different network types, it is important to connect common definitions from social networks to common language.

- Nodes: In this case can be associated with Authors or Posts, may apply to a specific person or a post on social media depending if we are evaluating user network or posts networks

- Relations: This defines the tie between authors or posts. Ties where both nodes are related in the same way (e.g. Two different authors commented on same post) are called indirect.

- Relationships: When relations are different on both ends, they are considered directed. These connections can be unilateral, when an author is replying to a post from a different author, or bilateral when multiple replies generate a conversation or debate.
- Weighted Relations: Not all ties are alike, some nodes may have stronger connections while others might have a weaker bond between each other.
- Network: The collection of authors and ties. When multiple networks are created and overlapped it becomes one Multiplexed network. This evolved state is present in many social media, creating multiple and different relations between authors.

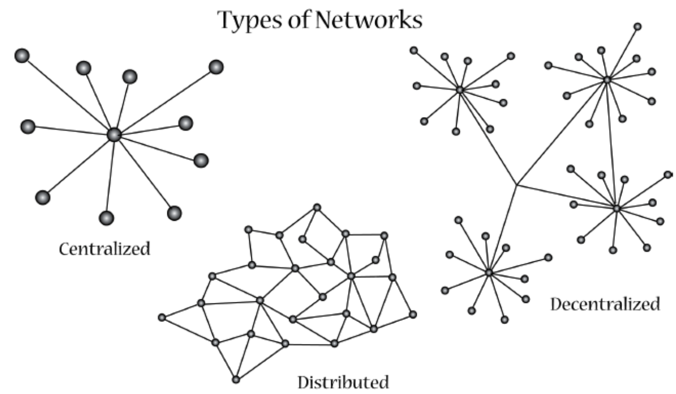


Figure 2.1: Network Types

Figure 2.1 shows some examples of network diagrams. Centralized networks are highly advantageous connection wise. Everyone can give information to each other within 2 steps but fail when the centralized node is offline for some reason, this stops the entire network.

Before internet, companies had to rely on centralized networks to gather feedback, some still do. By decentralizing, the number of nodes that fail when the upper node disappear is lower, although it still happens. The big advantage of this layout is that it can easily become a Distributed network. Outer and lower connection nodes can easily connect to each other and create redundancy [8].

Distributed networks occur when initially isolated nodes start to make connections with other nodes, something like in Facebook when a user starts adding friends-of-friends or following a page instead of waiting for a friend to share information from it figure 2.2. This means that everyone has the same, or close, importance in the network. A layout like this is extremely utopic when thinking about feedback analysis.

While everyone has their own valid opinion on a product, people like celebrities can spread their opinion faster than a kindergartner. Realistically, decentralized networks are the most common occurrence.

Social networks have changed the way of doing business [27]. Both on the competitive way, and how clients choose products based on the 4C's model: Consumer wants and

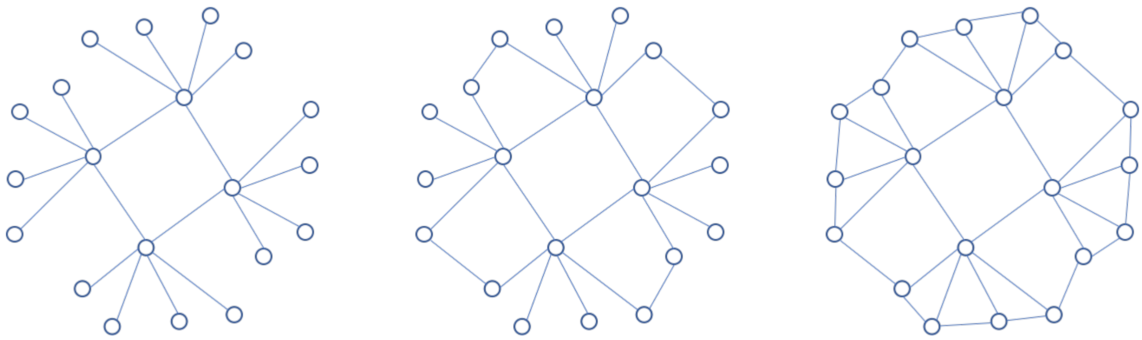


Figure 2.2: Decentralized to Distributed

needs; Cost to satisfy; Convenience to buy and Communication [18]. In recent years multiple research has been done to assess how big data and social media are impacting success of businesses. By faster dissemination of experience and opinions. According to [28] these changes have even greater impact on the fashion industry. Usage of this platforms allows better communication to maintain favorable customer ratings, by providing a broader understanding of the market. By paying attention to factors like promotions, loyalty, service quality and brand loyalty, customers perceive their feedback is being heard, with this a company has a better chance to raise their consumer database.

While sales promotion has a good short-term impact on customer satisfaction [27], this kind of approach is not as impactful on social media when compared to direct contact with costumers, through support appreciation and complaints managements.

Social Networks influences consumers on multiple stages, some of them being, brand awareness, intent to purchase and satisfaction. Brand awareness is mostly about simple cues to costumers such as ads, while purchase intent is a more direct approach, the search of information regarding the product to best assess if it meets their expectations. Lastly customer satisfaction, elaboration of feedback to share opinions with other future costumers. Social media feedback can be divided into two types, OSM (Owned Social Media) and ESM (Earned Social Media). Specific pages or accounts created by companies are categorized as OSM while feedback any other place is called ESM. OSM is more relevant to the work to be developed [7].

2.2 Product-Service System

Is the 60's a new idea surfaced, a change from product based to service-based businesses, allowing for industry improvements and for new jobs to be created. This paired with research stating that consumers see more value in products with better benefits led to a shift in mindset. Throughout the years both business operations evolved. Companies realized that merging them was the best way to satisfy customer needs. This generated a necessity for a new business concept, a Product Service System (PSS). This started with products and some services that fit together, until a carefully thought system that lets product and service necessities intertwine as if made specifically for each other [4].

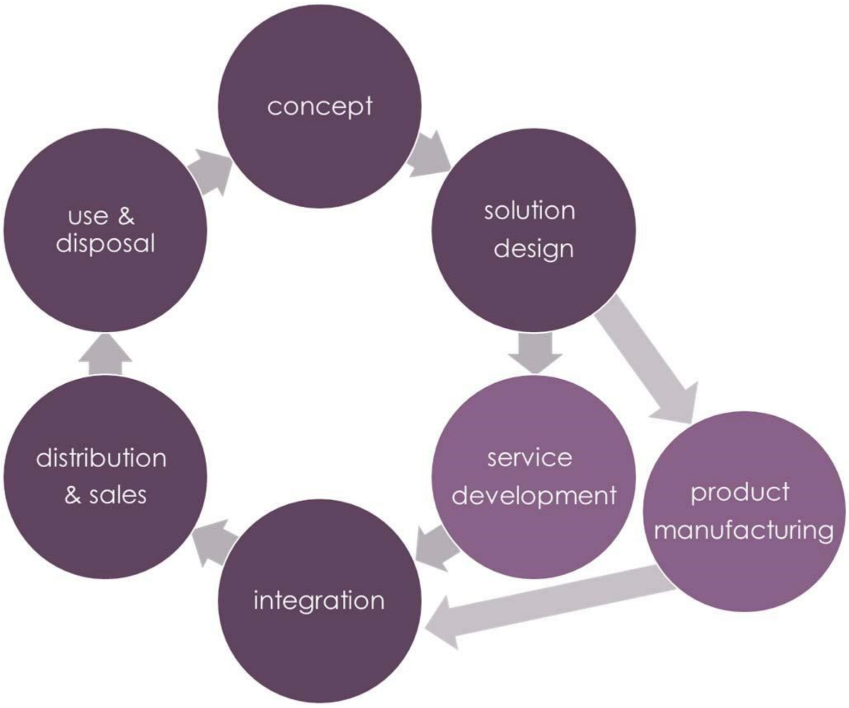


Figure 2.3: PSS Life Cycle

A PSS has multiple life cycle stages, an overall overview can be seen on figure 2.3. Starting at the concept stage, companies elaborate the PSSs objectives and improvement to be successful. In this challenging stage, company needs to consider the interests of all stakeholders and the required mindset change. This is where knowledge about how others have fared is used to achieve better results sales-wise. After this fine-tuning and the design are completed, service and product manufacturing are developed in parallel to ensure that all objectives are met. Integration then guarantees that the product and services created are compatible before launching it to the market on the next stage. The final stage on a PSS life cycle is use & disposal, all the knowledge gathered from this last stage is then fed into the creation of future PSSs, making this a semi-closed cycle [23, 26].

In earlier developments of this concept, sustainability and environmental impact were the main goals of the required change. Since PSS relies on a well-defined feedback loop it is inheritable sustainable so, in later years, companies began shifting the focus onto customer satisfaction and economic benefits. One of key activities that PSS implementation shines is control over the entire life cycle. By offering not only the product but also maintenance, installation, and dispose of product this concept shifts what were costumer responsibilities into the company [30]. This becomes particularly useful when seen from the costumer's perspective by being relieved of these responsibilities, costumers have opportunity to create new value by better management of their resource utilization.

The process of adding values to product by adding services, creating a PSS, is called servitization. Throughout the years many companies found this approach to be successful.

As early as the 60's companies started developments in this new process. IBM and Caterpillar realized that while their customers enjoyed their products, sometimes the required investment was too large for some smaller companies, so they started offering rental and leasing services. Xerox also discovered that by introducing a pay-per-user service, customers could keep track of costs and the quality of byproducts would be guaranteed, later Xerox bundled all services together into a yearly subscription, successfully transitioning from a sell focused company to a constant revenue. Michelin was also a pioneer on services by offering a pay-by-mile [17]. While most of the applications of servitization were successful there were some that failed, IBM was one of them, IBM survived by selling the product line to Lenovo and focusing on being a service-only company. Offering both systems on the same company environment, sometimes does not give the best results. So, most of these companies, in later years, split into two companies one for raw products business and another service-based.

The advantage of servitization is becoming more prominent with the rapid evolution of internet. With the digitization of information, Internet of Things (IoT) and Industry 4.0, companies can better track the movement of products and the utilization of services. Informatization becomes important because it allows the enhancement of existing products or service-level agreements (SLAs) [12].

2.3 Sentiment Analysis

Sentiment Analysis, also known as Opinion Mining, refers to the use of computational algorithms to determine the attitude of the author of the statement. On cases like *These sneakers are amazing* the attitude is simple to infer, the real challenge starts when multiple figures of speech are in place, like *I bought these amazing sneakers and my feet feel like they are comfortably sitting under a boulder*, in this case all adjectives are positive, but when the full sentence is read a negative reaction is discovered.

Sentiment Analysis can be split into different detection steps, basic evaluation by using adjective detection, objectiveness and feature.

While humans find it easy to infer double meaning in adjectives for machines the work becomes harder if only detecting adjectives is considered [25]. Natural Language Processing enhances the previous feature by evaluation the entire phrase, devising a Semantic Orientation Calculator to give scores to adjectives and then score them all together helps to better soften each adjectives weight in the final calculation [33].

Identifying the text opinion type (objective vs subjective) is also important [24], unless clearly stated, determining this might prove a difficult task for machine learning. However, this might not be successful, given that the training is done by humans, and so subjectivity also plays a part in this definition [20].

Feature evaluation is another important step, companies more and more are developing PSSs with multiple features. These features may have different impact on customer,

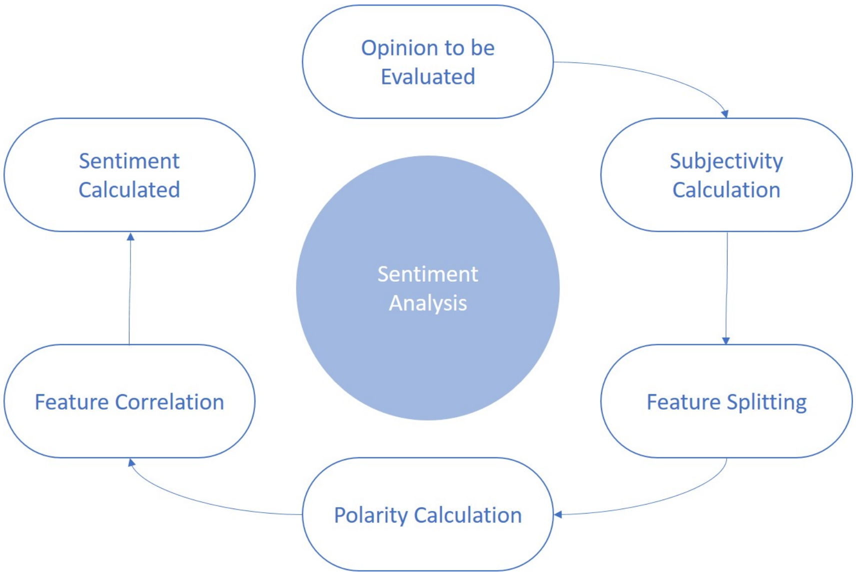


Figure 2.4: Example of a Sentiment Analysis Process

which might only comment on specific features that left a very good, or very bad impression [5, 13]. Furthermore, linking specific opinions to the correct feature or product gains importance when multiple instances of these are present on the same post, e.g. *Sneakers are very comfortable, but they won't keep the water out unless you use the WaterAway Spray, a little pricey but if you do then they are just perfect*. The previous example refers to two different products and gives two opposite opinions about each, and not in a clearly defined order, perfect is applied to both but only when used together. This concludes that for Sentiment Analysis to be successfully and relevant, relying simply on adjective polarity is not enough. A mix of all, as seen in figure 2.4, presents a more reliable method.

In practice detecting this kind of patterns becomes even harder on social networks, where more sporadic users rely on abbreviations to type faster, and typographic errors are also recurrent. However Social Networks public opinion sharing, almost always, are linked to a main post that, usually, are carefully written and that products are easily identifiable. One approach to circumvent the previous issue is assuming that all conversation under a certain post are talking about the same product. Doing this provides a simpler approach to feature detection [23]. Tags identification like Twitter #(Hash-tag) functionally, or Emoticons are other way to tackle this issue [16], while both methods have their own strength and might even complement each other, they are still far from the perfect method.

With the development and application of machine learning, Sentiment Analysis is abandoning the early stages and becoming a real future possible for companies to gather reliable information. It is expected that the traditional word-based evaluation will be replace by a more concept-centering aspect-level [31]. Some opinions rely on premises that are often not seen in word-based methods, e.g. *These sneakers are extremely slippery in wet surfaces*, slippery sole is easily identified as the feature talked about. However, its negative connotation is only applied because sneakers are supposed to be used in wet floors, if talking about how hard it is to walk with swim fins on sand a negative sentiment can be detected but swim fins are only supposed to be wore while swimming.

2.4 Natural Language Processing

Natural Language Processing (NLP) can be defined as a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts. These can be done at one or more levels of linguistic analysis for achieving human-like language processing for a range of tasks or applications [6]. The definition can the split into smaller definitions. Range of computational techniques since there are multiple methods, approaches to this analysis. Naturally occurring texts means that it is not language-bound the only requirements is that the analysis is done upon a language used by humans to communicate with each other. Levels of linguistic analysis since in most languages subtle changes in the way the phrase is constructed can give a different meaning to the words used, sarcasm for example. Human-like language processing infers that it can be considered an AI or Machine Learning derived technology, and since NLP is mostly used as a mean to a goal we reach the final definition for a range of tasks and applications.

The perfect NLP system must not only be able to process text but also infer meaning from it, translate it to another language and give its own reply about it. There is not a perfect way to successfully test a NLP system. One of the proposed test is the Turing Test, this test is considered successfully passed when a system cannot be distinguished from a human when answering from a terminal. This test has one flaw, the goal is well defined but intermediate evaluations are not possible [2]. The main barrier of successful NLP systems is the ambiguity found in all languages, these can be:

- Simple, a "mouse" can mean both the animal and the computer peripheral.
- Structural, "I saw him with glasses" it can both be interpreted as him having glasses on, or by the author only being able to see him by wearing glasses.
- Semantic, the word play is a good example of this, if asking both a DJ and a guitar player the same question "Play me some music" the outcome would be different.
- Pragmatic, "Can you give me a ride?" Might be a yes/no question or a request for a ride.
- Referential, "I saw Jack's car being towed with a flat tire, ..." while the remainder of the sentence might clarify who had the flat tire, the flat tire might refer to Jack's car or the tow truck

These ambiguities can also be considered as the linguistic levels identified earlier.

Figure 2.5 shows one if the first steps of NLP, sentence splitting and tokenization. Sentence splitting is rather simple, e.g. "Mother died today. Or, maybe, yesterday; I

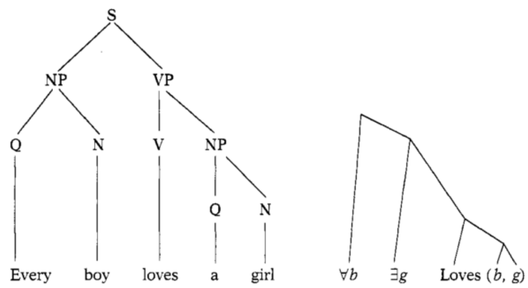


Figure 2.5: Word Splitting on NLP Analysis - S: Sentence, NP: Noun Phrase, N: Noun, V: Verb, Q: Quantifier

can't be sure" phrase becomes, "Mother", "died", "yesterday;", "I", "today.", "Or,", "can't", "be", "maybe,", "sure.". Then tokenization process treats these all these single words into meaningful elements, this process, one the other hand, is quite complicated and many libraries and tools have already been developed, one of them coreNLP works in the following way [3].

```
> getToken(annotation)$token
 [1] "Mother"    "died"      "today"     "."         "Or"        ","
 [7] "maybe"     ","         "yesterday" ";"         "I"         "ca"
[13] "n't"       "be"        "sure"      "."
```

Figure 2.6: coreNLP Tokenization

Punctuation marks are important to understand meaning so they become tokens by default, *can't* is the equivalent of can not, so it's also split. This specific library also tries to identify sentences. In this case the first 4 token were considered first sentence and the rest the second sentence. Different tokens might have the same meaning, i.e., *gone* or *going* can all be identified as the same lemma *go*, converting all word to singular form is also done in this step. In the next stage a parsing generates a tree to link token together and generate a tree like the one seen on figure 2.5

NLP can be applied on many scientific fields besides opinion mining, [9] found that applying this technology to electronic medical record could improve patient re-admission risk.

2.5 Product Similarities

On the topic of Sentiment Analysis, it is safe to assume that product similarity might prove useful when designing new PSS. As we see on smart phone topic it is expected that these have closer similarity, then comparing with an "old-school" phone. This calculation needs to consider multiple factors, for example target audience, components of product, utilization, features.

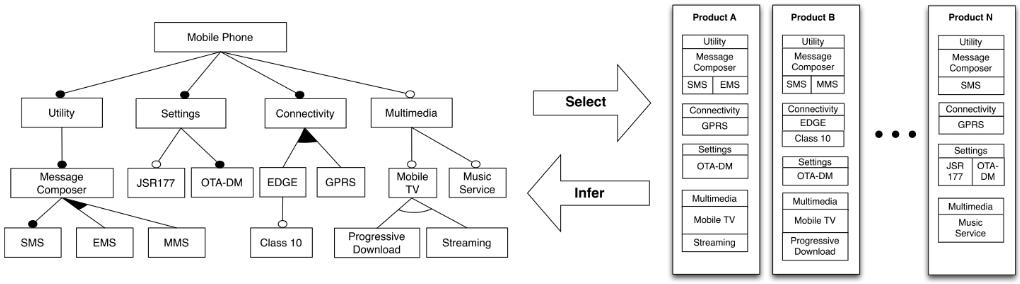


Figure 2.7: Extract from a 2005 Motorola Phone SPL [10]

On figure 2.7 it is shown a breakdown of an old phone features. The process shown displays the possible of selecting features to find similarities and inferring products from features. Meaning that a company can pick a product and improve it with better features or pick features that might be relevant at the time and get an idea for a new PSS. This gives companies the power to target PSS development to what the target audience is expecting to receive, generating better revenue expectations and credibility.

Studies found that product similarity might not be equally effective on all market segments. When comparing two opposite quality brands it was found that on low end products product similarities matter little, while on the other hand when comparing two high end products similarities are largely considered by the costumer. Sustaining that on cheap products costumers look for saving while on high end more of a feature evaluation is considered [19]. Successful product similarity calculation can also be used for company similarity calculation, by comparing product diversity and similarity in software market, [11] found that companies could be classified as direct competitor when a specific threshold of product similarity was met. They also found that product and company similarity would suffer large changes after major demand shocks [11]. This reveals the importance of evaluating current similarity and the importance of R&D.

Regarding PSS, product and service similarity are equally important. Service reuse is difficult to achieve in an enterprise environment [1], not because it is not useful but because migration is inefficient. This leads to service duplication, so while the services are not the same, they remain very similar, evaluating this similarity is important to assure that the PSS is following the desired guidelines. The effect of exploitation on International Joint Ventures new product performance shifts from negative to positive as product similarity increases [14]. IJVs with local companies are important when a company tries to enter a new market on a new culture. The local company knows how the customers usually evaluate products, and what they are expecting. When this venture occurs, companies need to assess if they are not competing against the venture itself. To find this, product similarity evaluation is done on both companies, if the new products have high similarity with the ones currently on the market, exploitation is more effective while on lower similarity exploration is the best option. Given these facts of heterogenic and fragmented markets, the benefit of creating new product variants is to meet the requirements of individual customer groups as accurately as possible. The competence to offer customized solutions at a competitive price is therefore a key success factor for companies. [32]

Cost-wise development of entire new products creates a big investment strain on the company, that sometimes does not turn out as expected. Let's give the example of Nokia phones, Nokia was once a big phone company, they kept investing on new, sturdier products. At some point in time the market changed, seeking lower lifespan products but with newer technology each time. Developing new product by recycling similar products enables to small investments throughout the development, and it allows for an easier focus change, depending on costumer reviews and market perception. Market is also volatile, what seems like a perfect deal today, might not be as good tomorrow. Developing services one by one while accessing their viability is the safest bet, especially for small companies.

2.6 Visual Analytics

Visual Analytics is the science of analytic reasoning supported by interactive visual interfaces [34]. On today's information driven society, it is not feasible for a person to analyze. Throughout the last decades, multiple automated tools were developed to help humans perceive of this data.

Due to the complexity required by some problems or systems, human interaction is always required on the early stages of the analysis process. Therefore, Visual Analytics can combine the human mind advantages (creativity, flexibility and background knowledge) with current storage and processing power of machines, leading to a faster result calculation of data. The human interaction is then limited to result analysis and interpretation, effectively reducing the impact of big data on the end user.

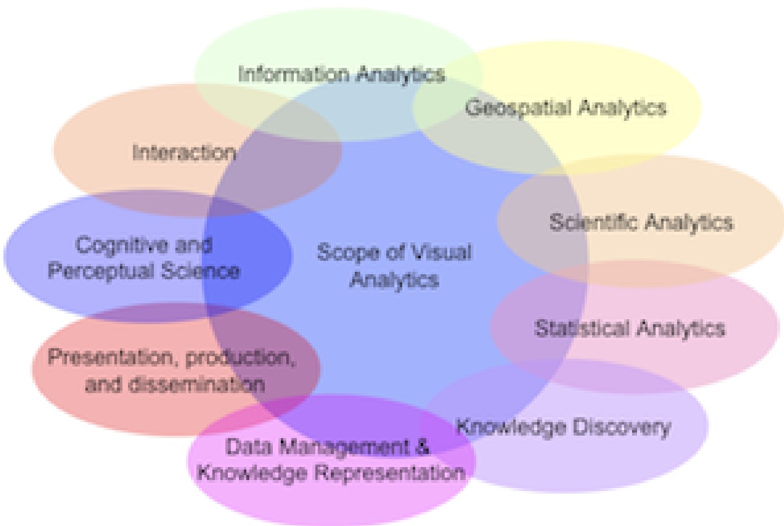


Figure 2.8: Scope of Visual Analytics

Visual Analytics overlaps with two others well know techniques (see figure 2.8), scientific and information analytics, there is no clear consensus on the boundaries of each field. While scientific visualization focuses on naturally structured data like MRI data or wind flows, information visualization if more often used to abstract data structures. Visual Analytics diverges from these two processes by focusing on coupling interactive visual representation with underlying analytic processes like data mining. This improves versus the other two field; Visual Analytics provides tool for more complex and high-level problems.

Visual Analytics improves human capabilities on six main areas [34]:

- with the help of a visual resource it effectively expands human memory
- representation of large amounts of data in compact form, reduced search effort.
- representing information depending on its relations makes for an easier pattern recognition
- due to the large data capabilities allows for multiple event analysis at the same time
- dynamic views provide 3D data analysis, opening many analysis perspectives when compared to simple graphics

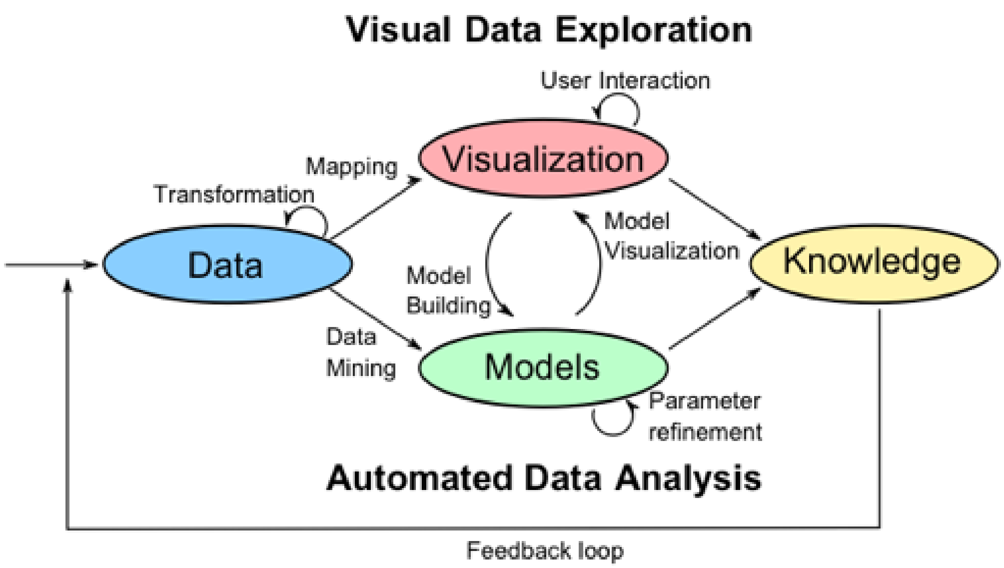


Figure 2.9: The Visual Analytics Process

The Visual Analytics process is depicted on figure 2.9. Although it is mostly cyclical, the first iteration requires a certain order. First stage is data transformation, Heterogeneous data needs to be integrated, in a way that the result becomes a unified data structure that can later be processed. After transformation the analyst may pick two paths, visual analysis or automatic analysis, automatic analysis use data mining algorithms to create models. These models almost always require tweaking by the analyst.

Alternating between automatic and visual analysis is what refines the models. Sometimes mistakes or misleading results might be found on this process. Finding these mistakes early on, helps developing sturdier and more confident models. If visual analysis is done first, then the analyst may steer model building to the information he is expecting to get from the process. The process final stage is getting knowledge from the gathered data and generating new data so that the model keeps getting more refined with each iteration. This knowledge is gathered from all processes, visualization, automatic analysis, previous models, etc.

The growing use of Visual Analysis tools in the design of complex systems, environmental informatics, and public policy, strengthens the case for exploring their use in Sustainable Lifecyle Designs [29]. These systems while helpful still have limitations, e.g. the data gathering process is mostly complex. The existing approaches for gathering and synthesizing information flows from lifecycles usually require adjustments, which leads to overhead work to implement this data.

2.7 Chapter Overview

This research led to better understanding how a social network can be mined for opinions. The importance of understanding relationships between users revealed how an approach to the current market could be improved. PSS life cycles itself, although well defined, do not have clear feedback loops which can also be improved. Sentiment Analysis proved to be a good approach to the identified issues on PSS life cycles using social networks feedback. NLP solutions allow the conversion of text inputs to sentiment values needed for analysis. Visual Analytics supports the creation feedback loops to improve model visualization. Although not directly related to the previous defined problem, product similarities allow for product comparison and possibly extrapolation of results.

Chapter 3

Concept

On this chapter it is presented an overview of the development objectives and how the tool is going to tackle issues like, author influence, product similarity, among others.

3.1 Introduction to Sentiment Analysis

Sentiment analysis is used to acquire feedback from end-users and stakeholders across the life-cycle of the PSS. This feedback allows to better monitor and understand the general sentiment from users towards a PSS. PSSs containing just one Product are also permitted, even though it is not a PSS by definition and are currently rare in the market. Three focus groups were identified to achieve the functionalities that constitute Sentiment Analysis: opinion modelling, opinion extraction and opinion prediction. Further analysis of these modules derived into four functional groups:

- **opinion modelling** where users define which PSS to monitor for sentiment and where to look for them;
- **opinion monitoring** which regularly scans for new posts and processes them;
- **opinion extraction** where user get the result from the processed opinions; and
- **opinion prediction** where user can infer the evolution of sentiment for a specific PSS by using data from similar PSSs and their opinion history

When a designer wants to get a better understanding of customer feedback and sentiment towards a specific PSS, the designer defines the model that will guide the rest of the process (**Define** step on figure 3.1). Sentiment analysis requests inputs from Social Feedback Extraction to gather relevant data that will later be presented to the designer,

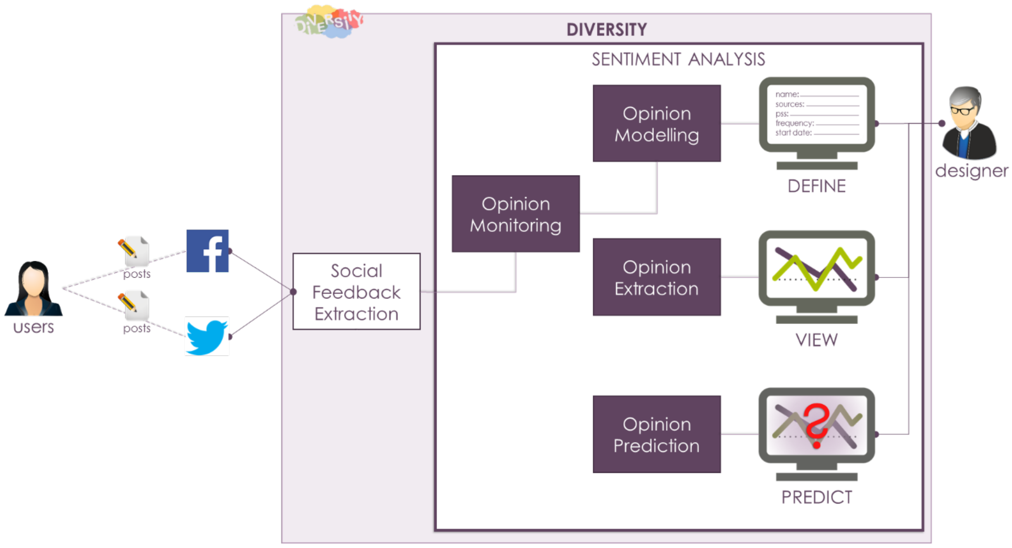


Figure 3.1: Sentiment Analysis Conceptual Structure

this information may come from Facebook, Twitter or Amazon. The information shown to the designer considers all calculation done in reach and influence, and other important information (**View** step on figure 3.1). In the end the designer might want to take advantage of lessons learned from this PSS to improve it or design a new one (**Prediction** step on figure 3.1).

3.2 Strategy

3.2.1 Sentiment Analysis

Sentiment Analysis evaluates opinions, sentiments and attitudes towards multiple topics based on textual input. While the main goal is to identify the polarity towards the main object of the opinion, this discipline has the potential of tapping into the large pool of opinions that are social networks. For companies that successfully harness this potential it might provide cutting edge advantages against their competitors.

When considering PSS, and the possibility of extending operational life-cycles, the importance of opinions gains strength. It enables the possibility of driving the industry into new requirements and improvement that may develop new services and functionalities. Seen in figure 3.2 sentiment analysis can also be classified as the vehicle for communication between users and companies.

With the emerging of social network and the ability of users to share their opinion with the entire world with a simple click, users now have a stronger voice to influence companies and products. This also means that companies also have an easier access to what the users really want and need.

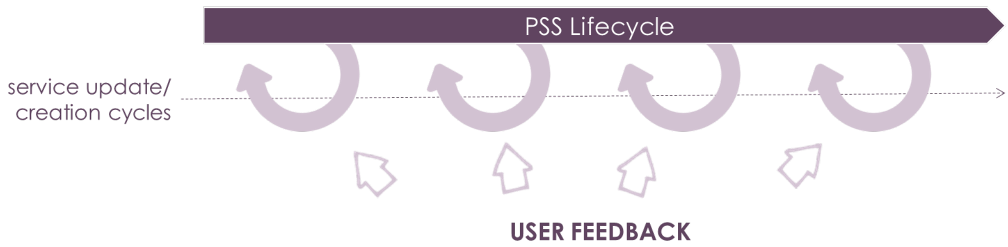


Figure 3.2: User feedback as driver to the extension of the PSS lifecycle

3.2.2 How is Sentiment Analysis obtained?

Sentiment Analysis is based on opinions and posts. Users write posts about products and services. Other users then react to these posts, either by appreciating, depreciating or simply replying to it, creating conversations and discussions. These discussions reflect individual opinions and can either be aggregated by opinions for each post, or an overall opinion.

Table 3.1: Main concepts user in Sentiment Analysis

Term	Description
Post	Content published by a user on a social network, providing a sentiment on a specific subject.
Original Post	Post that originates a thread or conversation regarding a subject.
Comment	Response to an original post.
Source	Social network or communication medium where the post was created.
Opinion	Intrinsic sentiment related to an object. This may be a conversation or a single post.
Polarity	Value representing the degree of satisfaction or positivity or negativity of sentiment regarding the subject of the opinion. This value ranges from 100, very positive to 0, very negative.
Reach	Level of propagation of an opinion measured in terms of view, likes and comments to the post.
Influence	Effect of an author in other users calculated by the history of the conversations where this author participated, measured in terms of views, likes and comments to these past conversations.
Global Sentiment	Sentiment emerging from existing opinions within a time interval and indicating a polarity towards the subject of sentiment.
Strength	Number of posts created about the subject over a time interval.
Intensity	Number of posts per user about the subject over a time interval.
Range	Number of unique authors.

The process of Sentiment Analysis is based on posts that trigger conversations or discussions. An author creates an initial post about a topic and the audience eventually reads, likes and comments on it. The influence of an author depends on the audience the author has: this is measured not only by the number of people who read the post, but by their engagement in the conversation and by their own influence over their audiences. The audience of an individual message defines its reach, while the reach of past messages sets the influence of the author. Figure 3.3 depicts these concepts graphically.

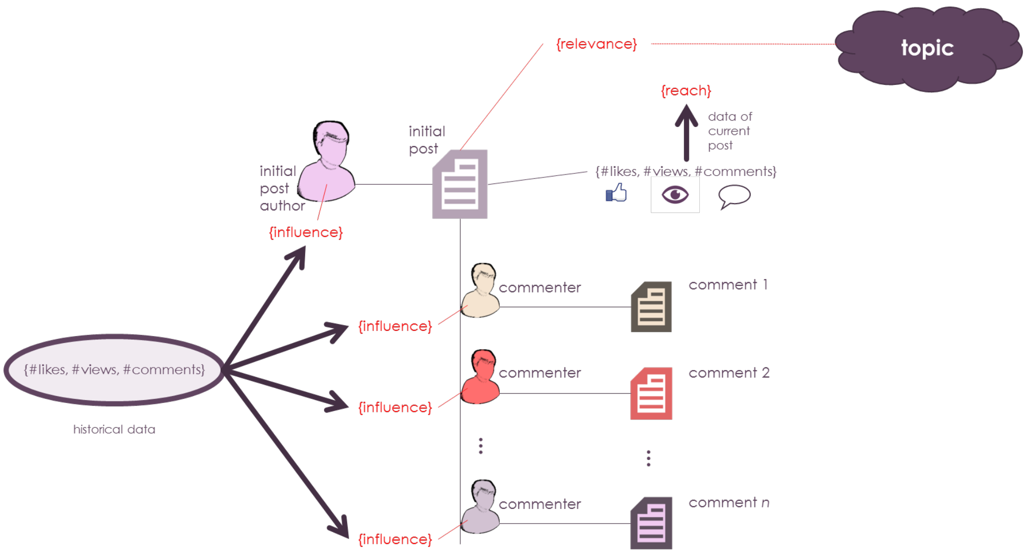


Figure 3.3: Posts, Opinions, Reach and Influence

3.2.3 Obtaining Sentiment values

Since the dimension of social networks does not allow for an open search for a subject, the user has the responsibility of identifying the relevant sources to monitor. The desired sampling frequency is also necessary, this allows for the system to manage the number of times needed to pool the sources for new posts. Finally, the search can also be limited to a specific sub-set of posts (e.g. only posts about materials, only posts provided by people over 60) to discard any information identified *a priori* as not relevant.

3.2.4 Polarity

Extracting Polarity from text is a complex field in computer science, the nuances of human speech, with the use of contrary modifiers like irony, present a challenge for computer understanding. Nevertheless, after testing and analysis of available tools the IBM Watson Natural Language Understanding was deemed the most effective with above 70% of correct assessment when compared to a human evaluation [15]. Watson NLU returns polarity values on the same range as mentioned earlier. This functionality is implemented by Social Feedback Extraction described on section 3.3.

3.2.5 Calculating Reach and Influence

Not all opinions impact networks in the same way, especially with Social Networks. The amount of connections we have (friends, pages, games) impact how much people see our opinions. Outside the Internet, sharing an opinion is restricted to our immediate surroundings, even if the people we contact directly share it after some iterations the original author will no longer be known. On the Internet a single opinion might be read by a user on the other side of the world and credited to original author within few seconds.

The electronic word-of-mouth is critical for the success or failure of a PSS and depends on the reach of the message. As seen recently some TV Shows that were canceled were rebooted after the audience conveyed to the internet to ask for a continuation. Therefore, a message that has a wider audience has a bigger impact than one that is only able to reach a small amount of people. Although a bigger audience provides a more impactful opinion, it is not linear, diminishing returns are in effect when the connections to the original author is no longer direct but through a mutual connection. People are more likely to relate to people with the same opinion, or similar interests. This results in the inner circle of connections to be more affected then outer circles.

Conflicting opinions are found throughout all topics in social networks. Relationships between these opinions are even more complicated since they are multilateral. The influence of authors becomes a key factor. When two opinions compete in the same circles, influence is crucial, both because typically authors with more influence are opinion leaders and because their opinions propagate indirectly more easily.

The influence of an author is determined by the reach of past posts. An author who usually reaches wider audience is usually one with a bigger influence.

Equation 3.1 is used to calculate the reach of a post:

$$reach(post_i) = \omega_c . \frac{\#comments(post_i)}{\phi comments_{global}} + \omega_l . \frac{\#likes(post_i)}{\phi likes_{global}} + \omega_v . \frac{\#views(post_i)}{\phi views_{global}} \tag{3.1}$$

where $\phi comments_{global}$ means the average number of comments per post, globally, within the universe of data present for that specific platform, posts on Facebook tend to have more interaction then on twitter.

Influence is calculated by the average comments, likes and views to an author's post. As with reach equation 3.2 is used:

$$influence(author_j) = \omega_c . \frac{\#comments(author_j)}{\phi comments_{global}} + \omega_l . \frac{\#likes(author_j)}{\phi likes_{global}} + \omega_v . \frac{\#views(author_j)}{\phi views_{global}} \tag{3.2}$$

On both equations the variables ω_c, ω_l and ω_v are weight variables to be set by the user on deployment.

3.2.6 Global Sentiment

Global Sentiment shows variations of the overall sentiment over time towards a specific PSS as seen on figure 3.4. Its polarity is calculated by averaging the individual opinions of each interval, weighted by their respective reach.

Global Sentiment allows for designers to understand the acceptance variations towards a PSS or the goods it produces and identify possible future improvements.

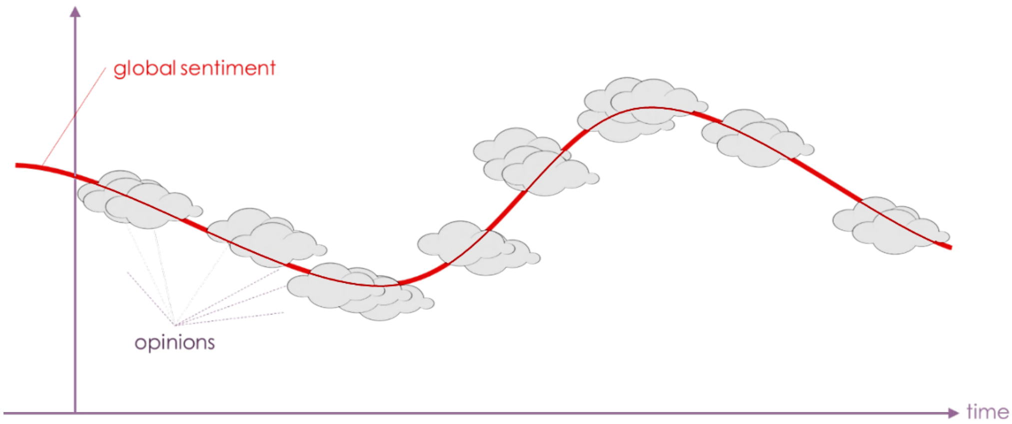


Figure 3.4: Global Sentiment

3.2.7 Prediction

The main goal of predicting the global sentiment towards a PSS is to anticipate situation that require some type of strategy change by the company, specially targeted at the design team, see figure 3.5. Prediction can be calculated from a current trend analysis or by using analysis on previous company PSS and the sentiment that was calculated for them.

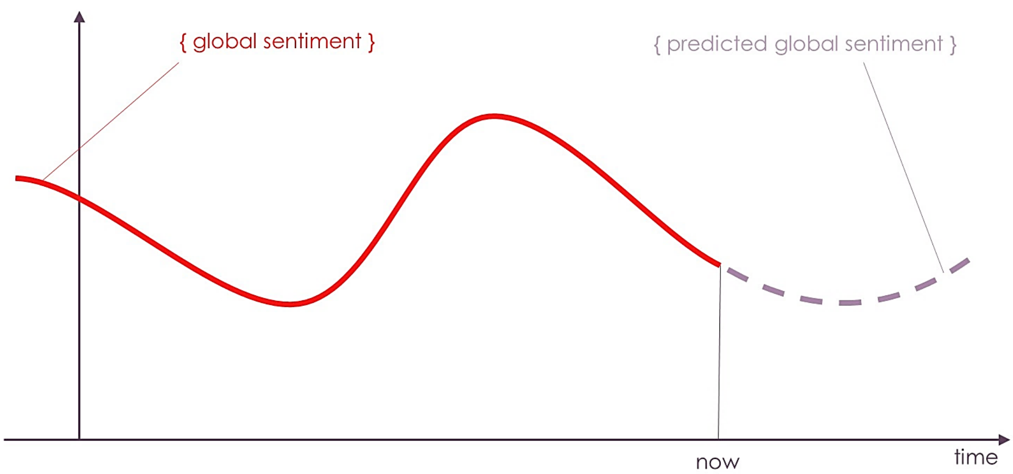


Figure 3.5: Global Sentiment Extrapolation

3.2.7.1 Extrapolation

The simplest approach to prediction is based on the extrapolation of current time series to the near future. This relies on the usual projection techniques, e.g. least squares, where the user has the possibility of selecting the size of the estimation window and prediction horizon. The estimation window is the data set used to estimate the model parameters. In the case of the existence of cyclic patterns, a comprehensive model consolidates the information from previous cycles. For instance, it is possible that positive opinions on a seasonal PSS (e.g. boat rental) are more frequent in one specific season. As the length of the dataset increases, the prediction module identifies the existence of such patterns, using standard correlation techniques, and derive the extrapolation model accordingly. The variance of the estimation allows the computation of the confidence level. The extrapolation is based on models and fitting methods. Models relate an independent variable (the one we want to estimate) with one or several dependent variables. There are various models described in detail in numerous papers in the literature. Common models include: polynomials, exponential, Fourier series, Gaussian, power, sum of sines, etc. Fitting models define the parameters of the models, i.e. they fit the data to a model. The Opinion Prediction component will implement estimation of polynomial models to support the extrapolation, given by equation 3.3.

$$y = \sum_{i=1}^{n+1} p_{ix} n - i + 1 \tag{3.3}$$

where n is the order of the model, y is the output and x is the independent variable (time).

Polynomials are often USED for this task for their simplicity, and ease of characterizing data using a global fit. One of the main advantages of a polynomial approach relies on the flexibility, for not highly complicated data, and due to their linearity, the fitting process is rather simple. The disadvantage is that higher orders tend to be unstable and with many oscillations, when the requirement for such oscillation is not necessary this effect is more present and while inside the data range the result might provide a good fit, it can diverge drastically outside it.

In this approach the need for an order above 3 is deemed necessary. This allows for linear results and seasonal concavities.

The third-degree polynomial as the structure of equation 3.4

$$y = p_1 x^3 + p_2 x^2 + p_3 x + p_4 \tag{3.4}$$

Fitting methods determine the coefficients that a parametric model uses to relate the response data to the predictor data. The 3rd order model has four coefficients, p_i. The result of the fitting process is an estimate of the model coefficients, $\hat{p}_i$. To obtain the coefficient estimates, the least-squares method minimizes the summed square of residuals. The residual for the $i^t h$ data point r_i is defined as the difference between the observed response value y_i and the fitted response value $\hat{y}_i$, and is identified as the error associated with the data as seen in equation 3.5

$$r_i = y_i - \hat{y}_i \tag{3.5}$$

The summed square of residuals is given by

$$S = \sum_{i=1}^{m} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3.6}$$

Being m the amount of data points and S the sum of square error estimate.

The extrapolation process follows the following step:

1. Opinion extraction sends the current data-set to be extrapolated
2. Using standard (3rd order) polynomial fit (equation 3.4, the data-set is modelled for the current time span
3. The predicted values are computed from the model obtained in 2, to a prediction horizon of 1/3 of the current data-set time span
4. The results are sent back to the opinion extraction processing for displaying together with the confidence levels

3.2.7.2 Sentiment Prediction

The second prediction option relies on the use of past knowledge to derive the expected acceptance of the market toward new combinations of products and service in a PSS. Taking this approach need to take into consideration critical thinking, using information from past posts about certain PSSs to estimate the feedback of a new PSS. The accuracy of this method will mostly rely on the similarity against already mapped and processed PSS, even if only an update is being done. The development of a PSS can range from a minor modification, to the creation of an entire new PSS. The tool should cover these and all situation in between.

The algorithms used for this calculation are of the utmost importance, verifying if enough data exists for an estimation should also be detected by the tool. A minimum input size should be present allowing for statistical treatment of data. Even when this minimum is met, the tool should provide an indicator accessing the level of confidence of the prediction of the PSS provided. Thus, the basic assumption for the Opinion Prediction approach is that the necessary data is available in the repositories to compute the predicted sentiment. Then, the computation of the confidence level will tell us if the prediction is of any value to the user, and in what extent.

As the repository of data grows, the sentiment towards existing PSSs is accumulated, the opinion prediction module creates a correlation between each product version and the observed sentiments, figure 3.6. The same is applied for service versions.

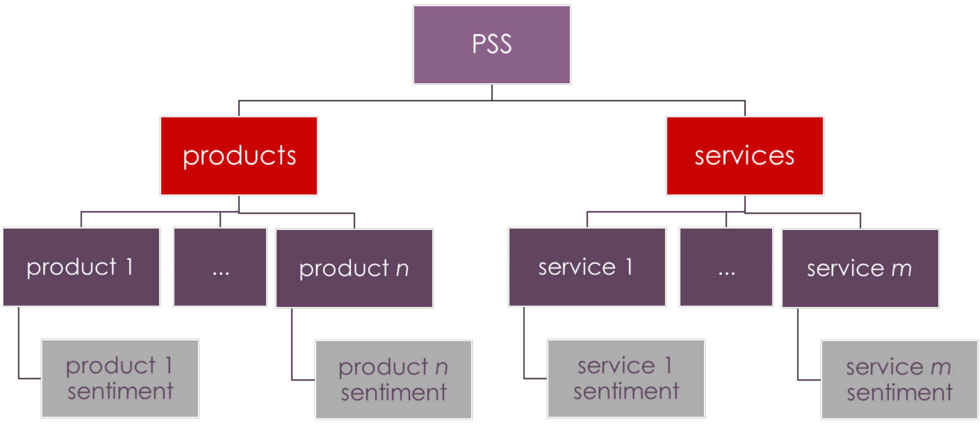


Figure 3.6: Association of a Sentiment with existing Products and Services

When a new PSS design is considered, the module needs to access if it derives or not from a combination of existing products and services, the module then estimates the sentiment prediction from a combination of the past sentiment towards each of contained elements. On the other end if the services and products are new, the system then identifies similarity between the characteristics of these new elements and previously existing ones. The module computes the confidence level of the prediction the degree of similarity with the selected existing elements.

3.2.7.3 Similarity in hierarchical models

Cosine-Similarity Measure (CSM), usually used to define vector similarity based on the angle between them, has also been proven very popular for query-document and document-document similarity in text retrieval. This notion can be generalized to compare generic objects treated as vectors in an n-dimensional space, where n is the cardinality of the element domain. The cosine of the angle between two objects is then used as a measure of their similarity.

If these object models are enhanced with the addition of hierarchic relationships, an improved measure of similarity is inferred from common characteristics between objects.

Let U be a rooted tree, with all nodes carrying a distinct label. We do not impose any restrictions on the shape of U. Each node can have arbitrary fan-out, and the leaves of U can be at different levels.

Conventionally, the depth of a node is the number of edges between the root of U and that specific node. Considered any two leaves l_1 and l_2 in U, the Lowest Common Ancestor $LCA(l_1, l_2)$ is the node with the lowest depth that is both ancestor of l_1 and l_2.

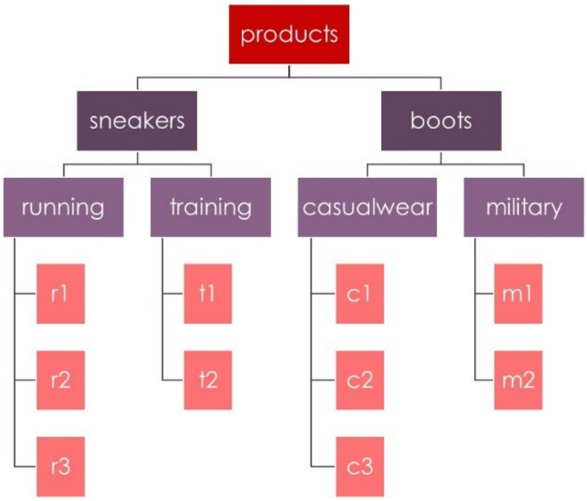


Figure 3.7: Example of rooted tree of Products

This LCA is always defined since the worst-case scenario occurs where the only common ancestor is the root node, and only one common ancestor can have the same depth. As seen on figure 3.7 $LCA(r1, r2)$ = running, and $LCA(c1, m1)$ = boots.

Identically to the standard Vector-Space Model's on the matter of weights usage. For any two elements l_1 and l_2 the definition is as follows:

$$Sim(l_1, l_2) = \vec{l_1}.\vec{l_2} = \frac{2.depth\{LCA_U(l_1, l_2)\}}{depth(l_1) + depth(l_2)} \tag{3.7}$$

This definition is consistent with the previous analysis, since the right side of this equation always lies between 1 and 0. Achieve 1 when, and only when, $l_1 = l_2$, and 0 when the root is the LCA.

The Sentiment Prediction process follows the following steps:

1. Opinion prediction user interface sends the information about the specific product(s) and service(s) being selected to compose a new PSS.

2. Opinion prediction navigates through the trees of products and services and select all products and services with a degree of similarity above a threshold specified by the designer e.g. 75%.

3. Opinion prediction requests all opinion models (which includes the sentiment) of all PSS that include either similar product(s) or similar service(s).

4. Opinion prediction computes the statistical distribution of the sentiments for each product and service, weighted by the degree of similarity.

5. Opinion prediction computes the predicted sentiment by averaging the global sentiment of all similar products and services. The predicted sentiment of the PSS is the average of these two results.

6. Predicted sentiment and the statistical distributions are sent to the opinion prediction user interface.

3.3 Social Feedback Extraction

This tool was designed by another partner in the consortium, it was adapted to serve the purpose required to test the work done in this dissertation.

Social Feedback Extraction (SFE) serves only two purposes:

- To search social media platforms (e.g. Twitter, Facebook) and retrieve at regular intervals posts from accounts being monitored by the opinion monitoring module. Posts in the current context represent not only textual data but also the relevant metadata, like account name, sex, location of author and likes and other reactions.

- Provide an interface to external services that evaluate sentiment based on the textual analysis.

These two purposes are rather simple, these functionalities were not included on the main system since the project developed defined this module as an external component. This tool is not the focus of this dissertation and for this reason is only mentioned for validation purposes.

Chapter 4

System Architecture

On this chapter is explained the overview of the system. What was built, how the different modules interact with each other, and how the tool was designed to give the user meaningful information.

4.1 Introduction

Modern enterprises, acting at the global market, need powerful engineering environments to allow for multi-directional exchange of knowledge between product design, service design and manufacturing as well as customers and suppliers. The exchange of knowledge must be assured along the entire life cycle of the product-service systems.

To support dynamic building of PSS, there is a need for collaboration among various actors across the value chain. This requires dynamic feedback loops exchanging knowledge between the design, manufacturing and product-service use phases. It is particularly important to analyze the combined feedback from the users of the product-services, both business customers and final consumers, to create or update the product-services.

The classical product engineering and PDM systems do neither meet the requirements concerning the effective support of concurrent PSS design, nor facilitate acquisition and re-use of tacit knowledge. Industrial companies need high flexibility from tools to allow capturing dynamically changing requirements and experience of various actors. On top of that, they need solutions to support knowledge sharing across the entire life-cycle and among various actors involved in dynamically changing value chains.

Cloud technologies, as emerging service-oriented knowledge-based systems, provide new possibilities for collaborative design of PSS within such distributed enterprise. These are adaptable to highly dynamic conditions under which enterprises are developing and manufacturing their product-services.

Therefore, this system relies on a combination of cloud technologies and social software solutions to meet the requirements of distributed manufacturing enterprises allowing for effective PSS engineering utilizing manufacturing intelligence and experience of all actors in the value chain, including both business customers and consumers. On top of that, the large amount of knowledge to be gathered and shared under dynamically changing conditions, and diffused to a wide spectrum of actors involved, having different expertise and working conditions/cultures, asks for effective context sensitive solutions for knowledge capturing, analysis and diffusion. The tiers that compose this software are:

- The **presentation tier** is the focal point for interaction with the user and explained on section 4.4
- A local dedicated database shared by all modules reside in the **data tier**. This tier is analyzed later, on chapter 4.2
- The **application tier** is composed of the processing layer and the data and integration layer. The first processes interactions with the user and manages the functionality needed for that. It handles communications with internal modules and with the data tier for manipulating the data that supports the processing layer. This later is analyzed on section 4.3

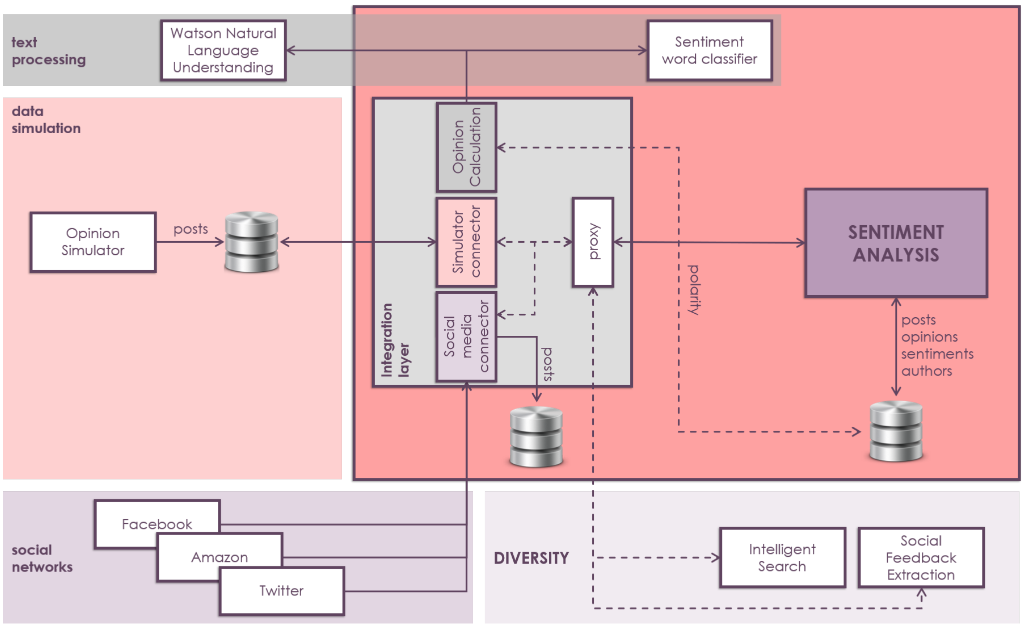


Figure 4.1: Server Processing Architecture

4.2 Data Tier

MySQL is an open-source relational database management system and was the tool chosen for the job of designing a database to hold the amount of data generated figure 4.2.

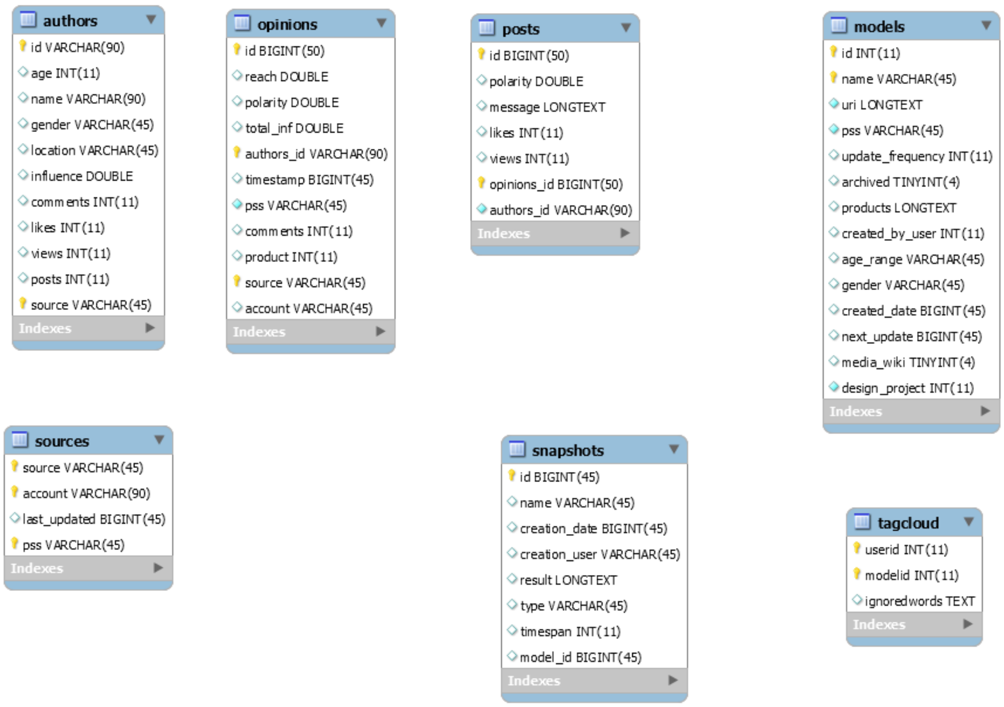


Figure 4.2: Database Architecture

Starting at the source table, this table is used to keep track of which sources and accounts are being tracked, it also gives an option to assign a PSS instead of letting the software discover it automatically. Since recalculating the entire results from a source would be a waste of resources a last updated column keeps track of this info, so only new information is requested.

A social media user is called an author, it is identified by its unique id and the name of the platform itself. This way it is possible to distinguish users with the same name in different platforms. The fields influence, comments, likes, views and posts are calculated fields populated by the Data Processing Module. These keep information about the user impact on the social media.

Since it is expected that all replies remain on the topic of the original post, the set of original post and replies is defined as an Opinion. An Opinion can be identified with its own unique id, paired with social media source and the id of the author. The account field refers to the page or tag used in the social media chosen.

The posts table represent raw data received from the social network, apart from polarity that is calculated based on the message using a Watson NLU.

Snapshots, Models and Tag Cloud are table more oriented for extraction and will be further analyzed on following chapters.

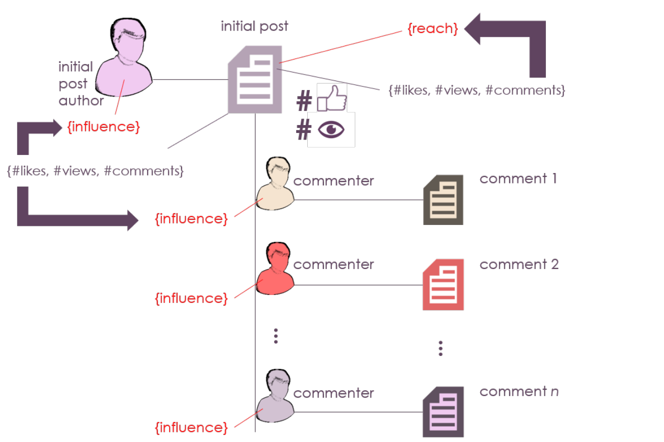


Figure 4.3: Author

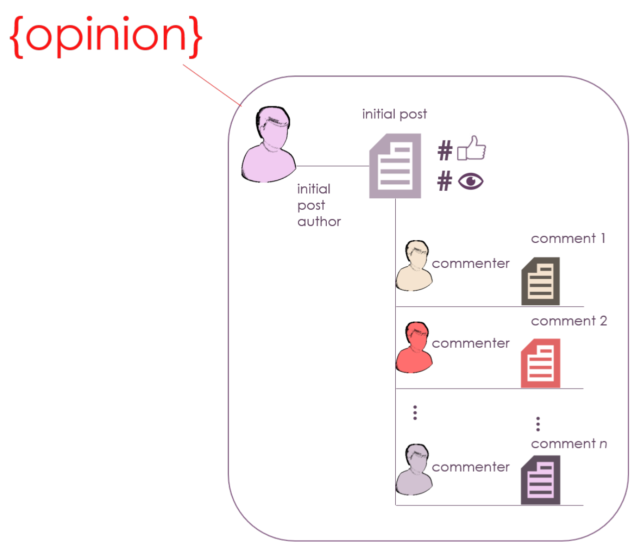


Figure 4.4: Opinion

4.3 Application Tier

Application Tier is composed of two modules available to the user and one for internal data calculations.

- **Opinion Modelling**: module where users define the targeted PSS and the sources to be used for acquiring posts and calculating opinions
- **Opinion Extraction**: module where users visualize the results of sentiment analysis for the existing data
- **Opinion Monitoring**: internal module responsible for managing opinion models, fetching information from social networks and calculating opinion values

4.3.1 Opinion Modelling

Opinion Modelling is the process of defining search and visualization parameters. A Model contains information of what product is being monitored, from what sources and accounts see figure 4.2. One model may have more than one source and account, effectively allowing for one model for an entire product, and multiple models source specific without extra strain for the opinion monitoring module. A Model important option is source and account combination , at least one combination must be present otherwise there would be no information to fetch. Other options are the PSS that the model is analyzing, if it generates a final product, and data granularity. Granularity being a visual aspect, sometimes seeing values per week might be more interesting than seeing them per day. By default, the model start monitoring from the date of creation, option for a custom date exists and changes the first data point that the model will have on extraction.

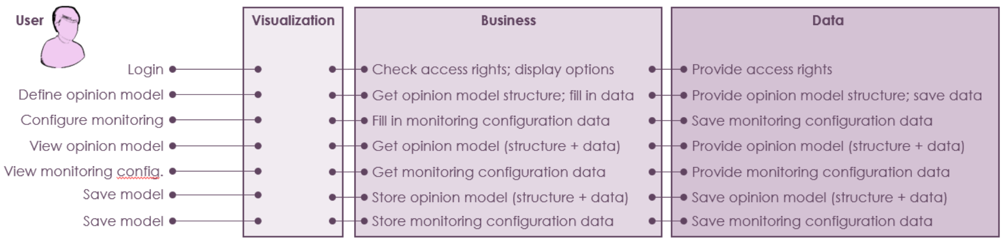


Figure 4.5: Opinion Modelling

When a designer creates a new model or adds new sources to old models, this module checks if that specific combination already exists in the database. When a new combination is entered, this module requests opinions monitoring to start fetching data from those sources. Date entered is important, this allows opinion monitoring to filter fetch data after the date defined.

4.3.2 Opinion Monitoring

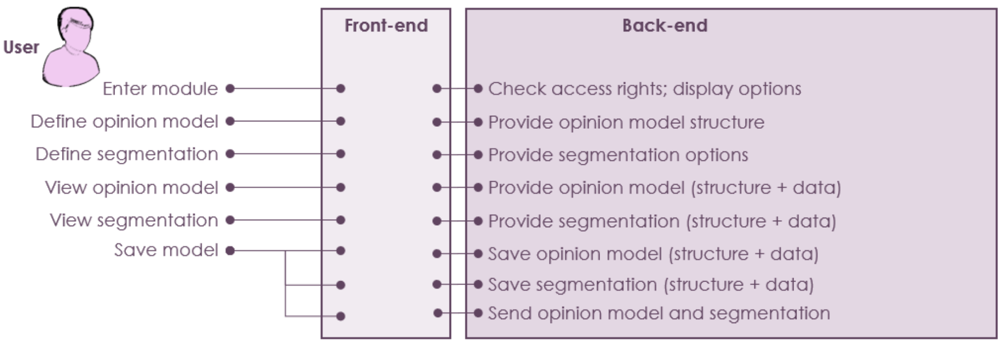


Figure 4.6: Opinion Monitoring

Once an opinion model is saved, opinion monitoring starts pooling for new posts at the defined frequency and processes opinions and sentiment. This is done at specific time of the day defined at server start, as to not overload the server at normal run-time. Monitoring starts by organizing what information it is going to fetched. Checking which source/account pair is required to be fetched, since some models might share these combinations.

When these combinations are all ordered by social network, the monitoring model starts requesting data, data requests are done one social network at a time, but a thread pool assures that multiple accounts are queried at the same time.

As soon as the first new entry appears in the database another process starts. This process requests polarity calculations from Watson, this is also the longest process of the opinion monitoring, usually keeping the calculation from finishing, while all non-sentiment dependent calculations are done at the same time. Values like global opinion sentiment wait for this process to finish before being calculated. This module is also responsible for keeping track of the last time each source was pooled to keep the duplicate information fetched to a minimum.

4.3.3 Opinion Extraction

Whenever the designer wants to view results, the designer selects the corresponding model. Opinion extraction then provides the results, including number of opinions, distribution of polarity, sentiment, reach of opinions (i.e. their potential audience) over time, and variation of sentiment over time. A Tag Cloud and the most relevant posts for that model are also provided.

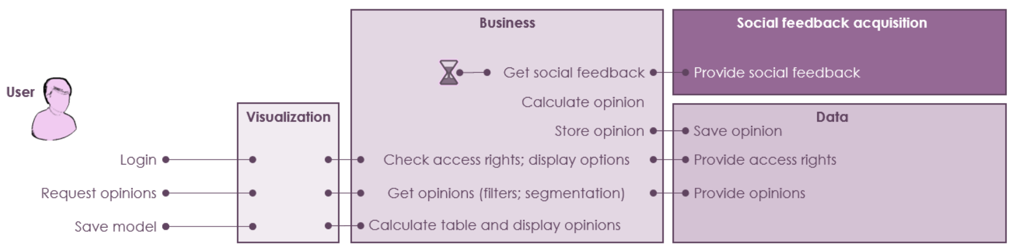


Figure 4.7: Opinion Extraction

These graphs shown by the opinion extraction, have the option to segment data, three options are available, gender, age and location. By using this option, instead of one line in the charts, there will be one for each option on the specific filter, by age (<30, 30-60, >60), location (East, West), and gender (Male, Female). Filtering out the results is also possible, by picking only one of each option, or even any combination of filter and segmentation.

A Tag Cloud containing the model most common words seen in posts is also presented, and allows filtering based on the presence of this word on opinions. If the designer sees a recurring word that is particularly negative for the PSS, the designer might choose to restrict showing only discussions where that word was used. The designer at any point might want to check results from a specific time frame, so one more filtering option exists. The user can click the graphs to filter for only the results that were considered when calculating that specific data point. Still on the view model page, a trending analysis is also possible by the extrapolation, giving the designer an expectation of how the PSS will behave based on the current trend.

While all this data is provided closest to live as possible, the designer might be interested in storing the current results in order show them later. Opinion Extraction enables the possibility to create a snapshot of what is currently being shown, this snapshot is not limited to the current filters. Snapshots also saves segmentation possible choices.

4.4 Presentation Tier

4.4.1 Tutorial

Throughout all web pages there is a tutorial available about the functionalities present on that page. When users enter a new page for the first time, a message is displayed asking if the user wants to see a field-by-field tutorial for the page (as shown in figure 4.8)

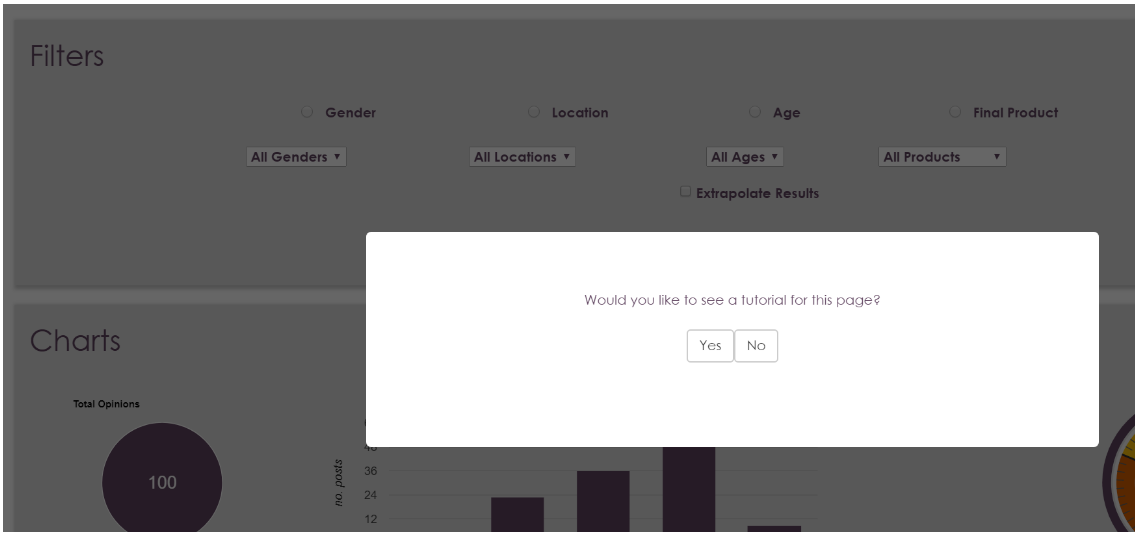


Figure 4.8: Tutorial Notification

If the user accepts this request, for every field a small description will be provided to clarify its functions (figure 4.9).

This tutorial is also available by clicking the question mark on the top right corner of all pages.

4.4.2 Home

Sentiment Analysis Home Page, shown in figure 4.10, has an overview of the tools available in the interface. Two panels constitute this view. On the Option panel, three sections exist: Define, View and Predict. The first section, DEFINE ("1" in figure 4.10), the

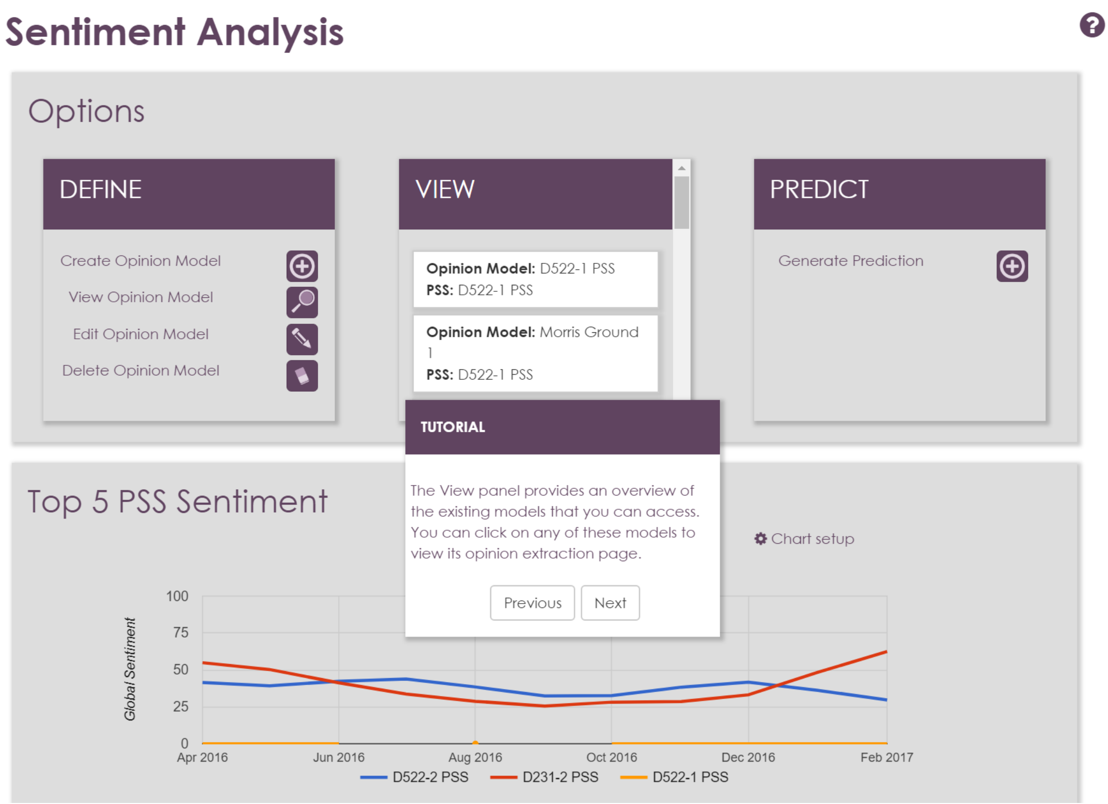


Figure 4.9: Tutorial Example

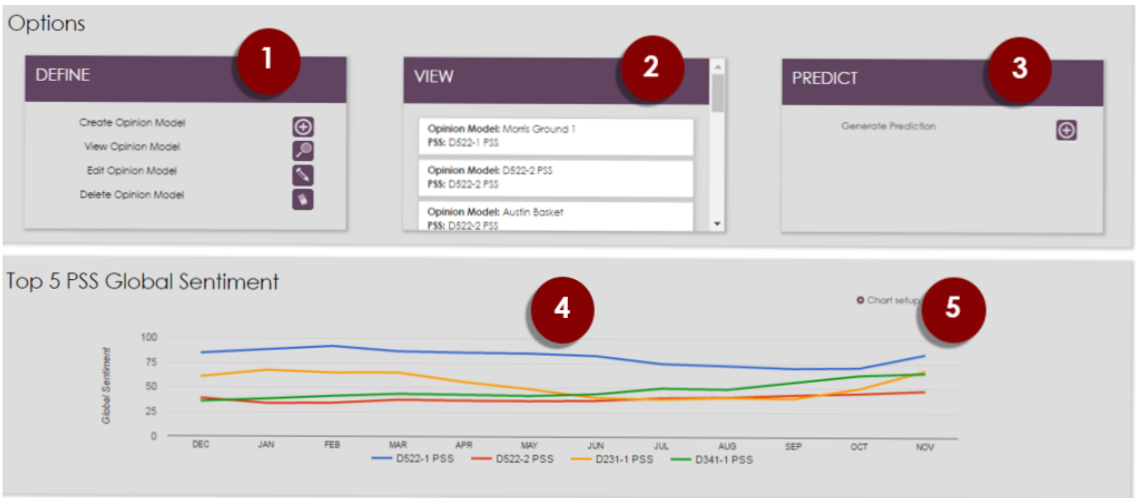


Figure 4.10: Sentiment Analysis Home Page

designer is presented with four different options that are connected to creation and edit of opinion models. The second section, VIEW ("2" in figure 4.10), displays the models currently available to the user as well as the associated PSS. The third section, PREDICT ("3" in figure 4.10) contains the feature of predicting PSS global sentiment.

The bottom panel, Top 5 PSS Global Sentiment ("4" in figure 4.10), displays a global sentiment chart of five PSSs with the highest reach. The user can hover the lines to get a more accurate value of that month, as displayed on figure 4.11. Chart Option ("5" in figure 4.10) allows the user to specify overall graph settings.

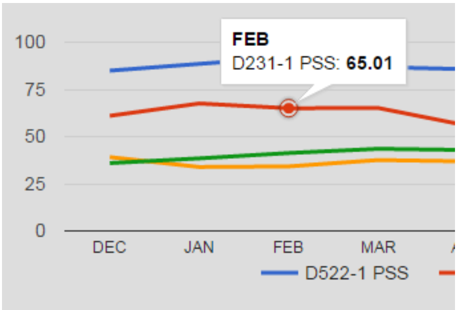


Figure 4.11: Top 5 PSS Global Sentiment Chart Tooltip

4.4.2.1 DEFINE Panel

Each of the components allows the designer to interact with the opinion models in a different way. On the Define panel the user has options to create, update and delete opinion models. In the following paragraphs these options will be better detailed.

Create Opinion Model:

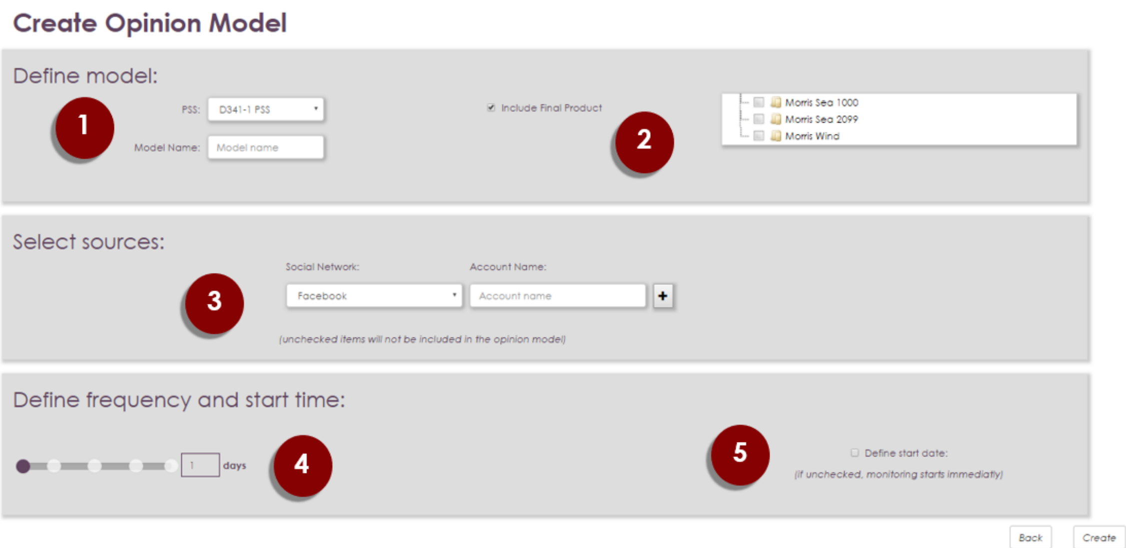


Figure 4.12: Opinion Modelling - Create Opinion Model

In the first section, the designer picks the PSS and defines a model name ("1" in figure 4.12). After selecting a PSS from the dropdown list, the user can decide whether to include that PSS final products by toggling the correspondent checkbox, which displays a tree view containing all the final products related to the selected PSS ("2" in figure 4.12). In this tree view, the user can click each product to toggle its check mark and include it in the opinion model.

In the middle section ("3" in figure 4.12) the designer selects the accounts that are to be monitored to this specific model by selecting the social network from the left dropdown list and entering the account on the right input box. After this clicking on the "+"saves the entry into the model. Checkboxes appear in case the user ends up not wanting that specific entry, see figure 4.13.

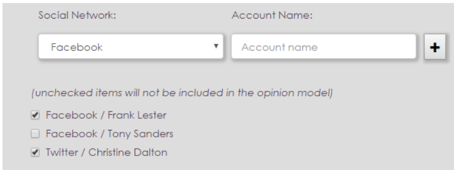


Figure 4.13: Opinion Modelling - Sources and Accounts

This process can be repeated multiple times until the designer adds all intended sources and accounts. Unchecked entries will not be saved into the model.

On the final section the designer can pick the minimum frequency the sources should be pooled for data, either by entering a value or sliding the slider ("4" in figure 4.13). On the right side the user also has the option to pick a start date for the monitoring process figure 4.14 ("5" in figure 4.12). If the checkbox if left unchecked monitoring will start as soon as model is created.

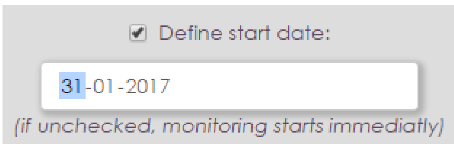


Figure 4.14: Opinion Modelling - Start Date

When all the necessary information is filled in, the user can complete the creation process by clicking the Create button or click the Back button to discard the changes and return to the Sentiment Analysis Home page.

View Opinion Model: If the user chooses the option to view model, the DEFINE panel changes to display a dropdown list of existing models, seen in figure 4.15. This allows the designer to choose which model configuration to view. By clicking the View Model button, the user is redirected to a page like the one in figure 4.12 but with information already filled in read-only mode.

Edit Opinion Model: The Edit Opinion Model option work in a similar way to view model but instead of a read-only view, the user is presented with an edit mode page, seen

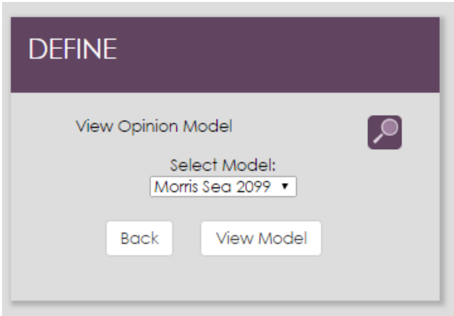


Figure 4.15: View Opinion Model

in figure 4.16. This allows the user to edit almost every input of the model, except for model name and PSS associated.

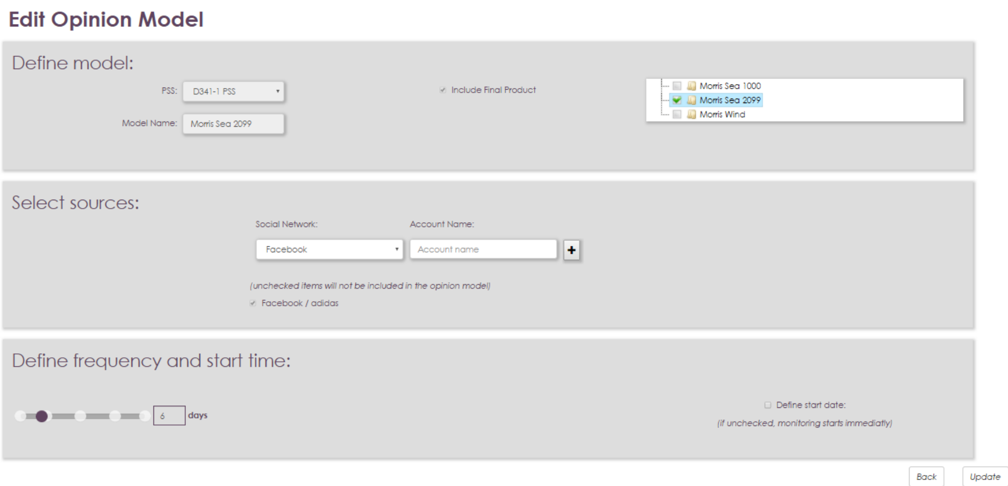


Figure 4.16: Edit Opinion Model

Delete Model: this option allows the user to delete an opinion model. After clicking, the DEFINE panel changes to display a dropdown list like the one shown in Figure 4.15, with delete button instead of view, where the user can choose the model to delete. The user can then click on Delete Model to delete the selected model, which will prompt a confirmation. If the user clicks "OK" in the confirmation dialog, the model is deleted. Otherwise, if the user clicks "Cancel" or closes the window, no models will be deleted.

4.4.2.2 View Panel

As mentioned earlier in this section, designers can click the models in the middle section, named VIEW, to get an opinion extraction based on the selected model. This is achieved through the Opinion Extraction page, represented in figure 4.17.

In this page, the user can get a detailed view of several properties related to the selected model. These properties are presented in the form of charts, which will be detailed below.

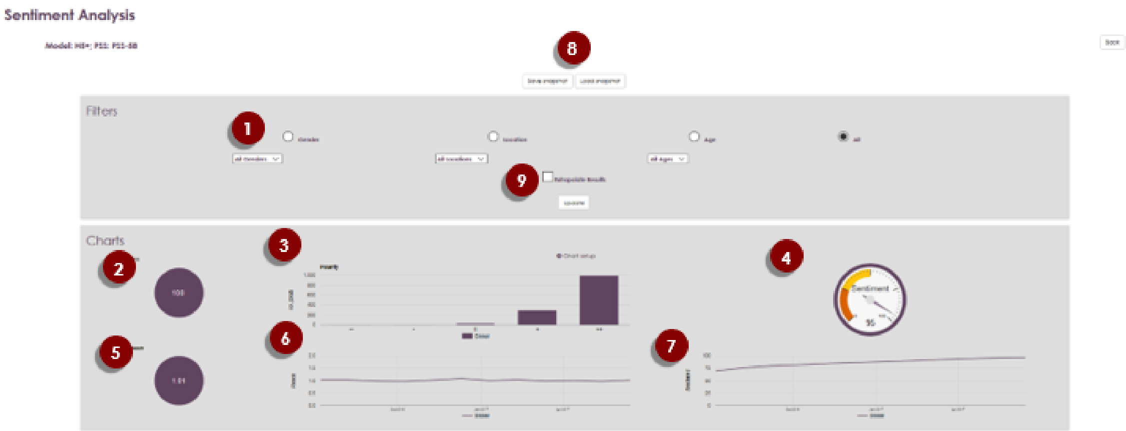


Figure 4.17: Opinion Extraction Page

When the user first enters this page, the data is displayed without any filters, this means the user is getting the values of all genders, locations and ages all averaged into the same point. Using the radio buttons and dropdowns in the top section ("1" in figure 4.17), the user can then apply filters or segment data to each of these settings, which will cause the charts to be redrawn with the new data filters.

The top left chart ("2" in figure 4.17) displays the total number of opinions monitored. In Figure 4.17, this number equals 100. This means that the displayed data is calculated based on 100 different opinions from the accounts defined in the opinion model creation.

The top middle chart ("3" in figure 4.17) is a histogram that shows the polarity of the selected PSS. The polarity is measured by assigning a value depending on polarity, "--" <21, -" 21-40, "0" 41-60, "+" 61-80 and "++" >81, to each post, and then count all the posts of each value to draw the chart. In other words, this chart shows the number of posts with a "--" (very bad) polarity, the number of posts with a "++" (very good) polarity, and all in between.

Figure 4.18, displays this chart with a filter by age, which splits each polarity value in three different age ranges.

The top right chart ("4" in figure 4.17) is a gauge chart that displays the global sentiment for the selected PSS now. This value ranges from 0 to 100 and represents how satisfied a user is with a given PSS.

The bottom left and bottom middle ("5" and "6" in figure 4.17) charts show the average opinion reach values. Opinion reach is a value that considers the number of views,

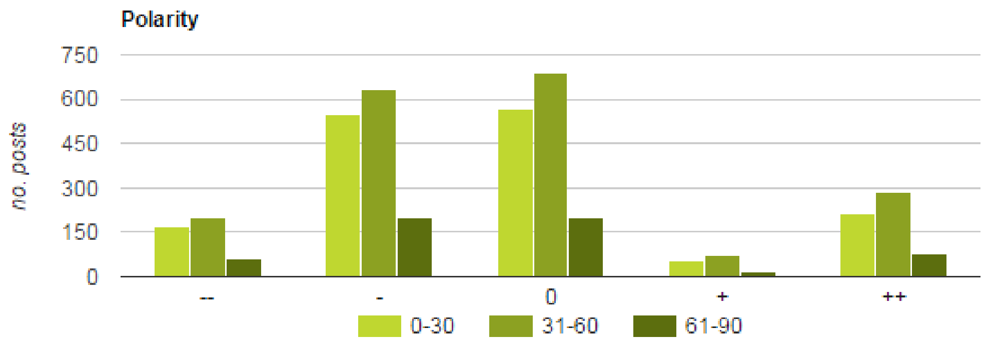


Figure 4.18: Opinion Extraction Page - Polarity Chart

comments and likes to represent an opinion visibility. Figure 4.19 displays the reach chart with a gender filter, which shows the reach values for Male and Female individuals separately.

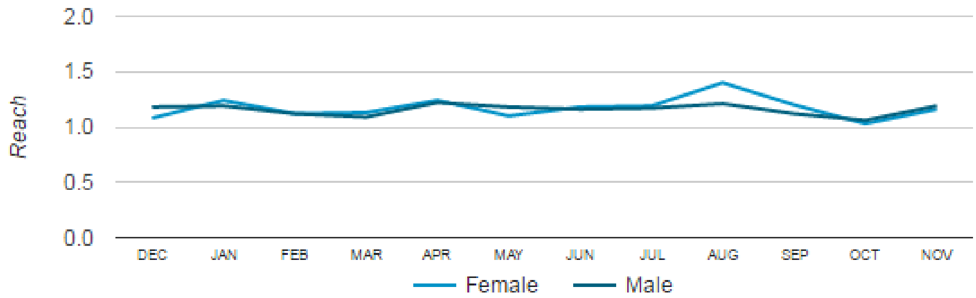


Figure 4.19: Opinion Extraction Page - Reach Chart

The bottom right chart ("7" in figure 4.17) displays the sentiment value throughout time. This represents how satisfied the users are with the selected PSS. Figure 4.20 shows this chart with a location segmentation, which shows that in this example users from Asia are generally more satisfied with the PSS than users from Europe.

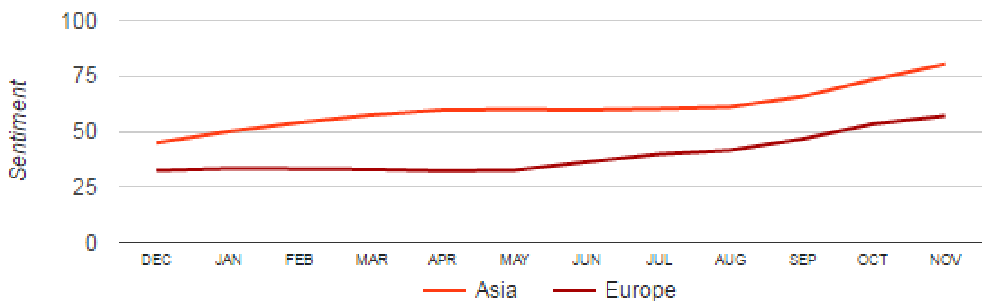


Figure 4.20: Opinion Extraction Page - Sentiment Chart

Two more sections are present on the page that are not seen on the previous figure, one of them is the Tag Cloud (figure 4.21) and the other one is the Top 5 Opinions (figure

4.22. The Top 5 Opinions show the five posts with biggest reach. This table provides detailed information about the top 5 opinions, such as the author, the post content, date, polarity, reach, influence, location, gender and age.

If the user clicks on one of these posts, an internal pop-up window is displayed, showing the original post and its comments. If the user selects a point in the Sentiment Chart (see figure 4.20), the top 5 list presents the posts with biggest reach for the corresponding time interval (and segmentation, if selected). When no point is selected, the considered time interval is equivalent to the whole chart time span.

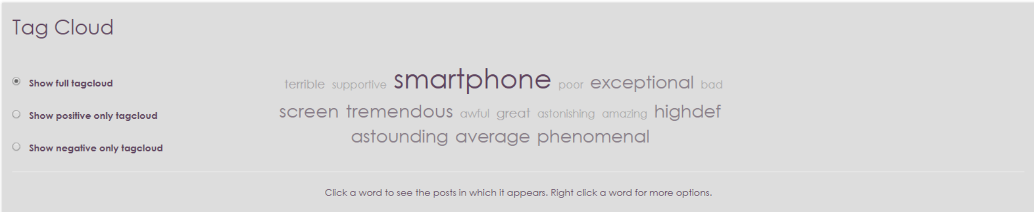


Figure 4.21: Opinion Extraction Page - Tag Cloud

Top 5

Clear selection

Original Author	Post	# Comments	Date	Polarity	Reach	Influence	Location	Gender	Age
Sandra Goodrich	Tell me what you think of the new my style casual! I say outstanding! No?	15	2016-12-10	71.0	2.63	0.0	Europe	Female	15
Eunice Waymoon	Check the new my style casual! What a sensational shoe!	15	2017-01-11	70.0	2.63	0.0	Asia	Female	45
Jimmy Plant	Have you tested the my style casual? This shoe is fair!	15	2016-12-28	55.0	2.63	0.0	Asia	Male	75
Bobby Gundrum	They launched the new my style casual! ...fair shoe!	15	2017-01-07	55.0	2.63	0.0	Europe	Male	35
Christa Paffgen	Check the new my style casual! What a splendid shoe!	15	2016-12-02	74.0	2.63	0.0	Europe	Female	65

Figure 4.22: Opinion Extraction Page - Top 5 Opinions

The tag cloud is computed by processing the contents of all posts that contributed to the extraction results. The most frequent words are displayed in this cloud.

The tag cloud is fully responsive, i.e., whenever a filter is applied to the result or a specific month selected, the contents of the cloud change accordingly. It is also possible to filter results on the tag cloud by selecting only positive or negative sentiments. In this case, only contents of posts with opinions in these intervals are displayed. Finally, the designer is also able to remove specific words from the cloud by right-clicking on them and removing them. The system maintains the list of removed words for each user, assuring that they no longer appear on later requests.

Snapshots ("8" in figure 4.17) allows for storing of an image of the feedback about a PSS on that moment. Users can then load these snapshots to evaluate changes to PSS acceptance, keep track of previous PSS sentiment values or for the design of new PSS. When users save snapshots, they must only define a name for that snapshot and save it.

When users want to load previously saved snapshots, a modal window is displayed with a list of existing snapshots (seen in figure 4.23)

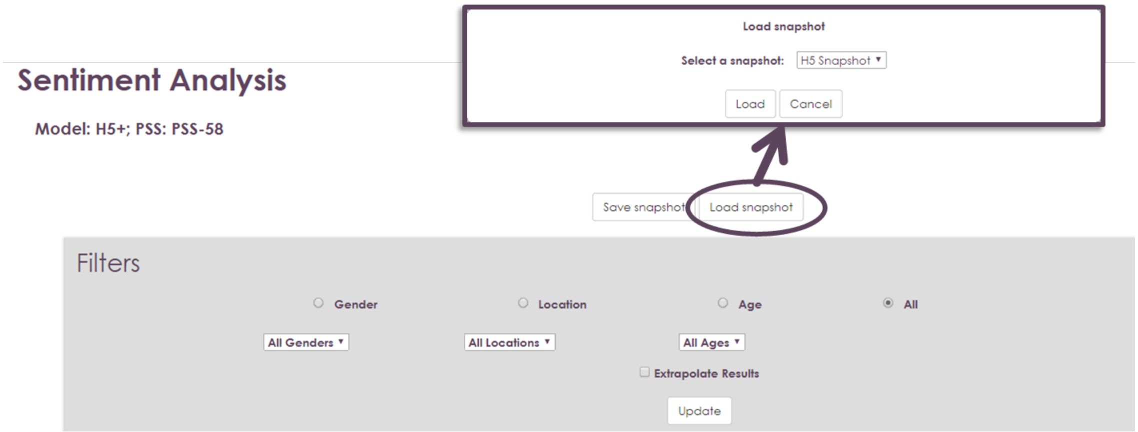


Figure 4.23: Opinion Extraction Page - Load Snapshot

When the selected snapshot is loaded, the resulting page is very similar to a regular extraction page, apart from the fact that the filtering area is no longer available (highlighted on figure 4.24). This is due to data management constraints. This page also indicates the name of the snapshot, along with author and creation date.

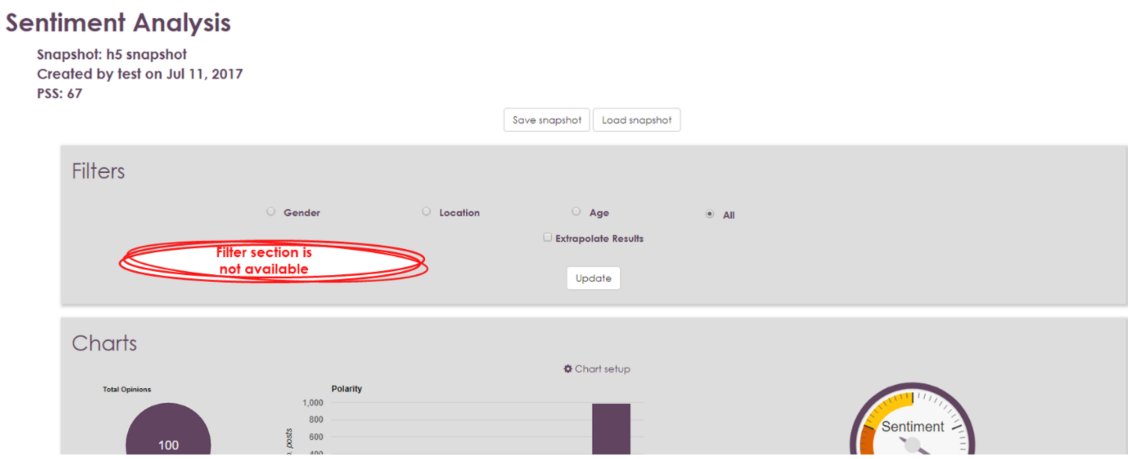


Figure 4.24: Opinion Extraction Page - Snapshot Example

If the extrapolation feature is selected ("9" in figure 4.17), the chart for sentiment over time displays an additional line depicting the evolution trend for that sentiment. Figure 4.25 shows the look and feel of this feature.

Finally, the user can click the Home button to return to the Sentiment Analysis Home page.

4.4.3 Prediction

This feature was designed to work as a sandbox for users (especially designers) to try different sets of products and services (see figure 4.26) and see the predicted sentiment

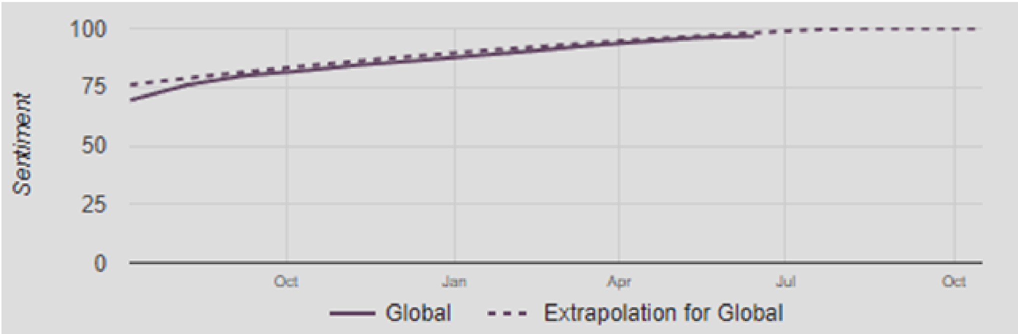


Figure 4.25: Opinion Extraction Page - Extrapolation Example

outcomes for each set (see figure 4.27).

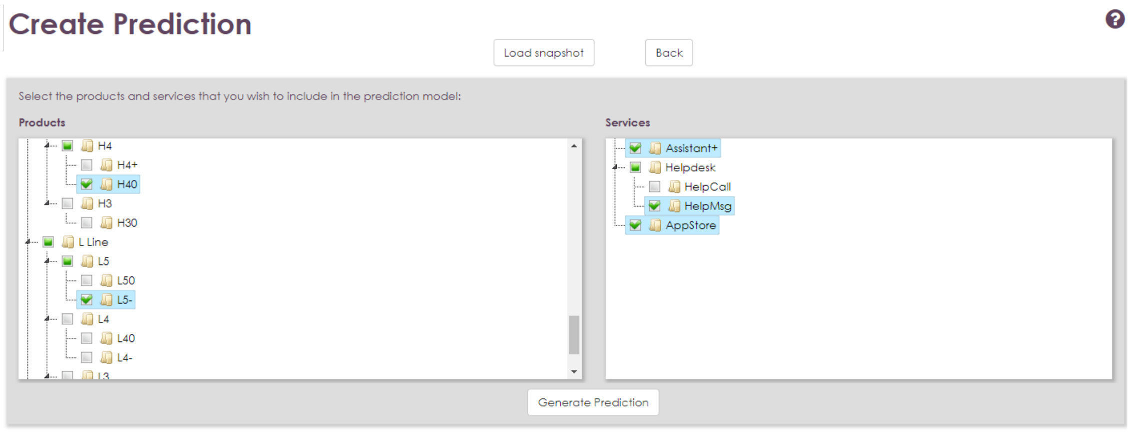


Figure 4.26: Prediction Creation

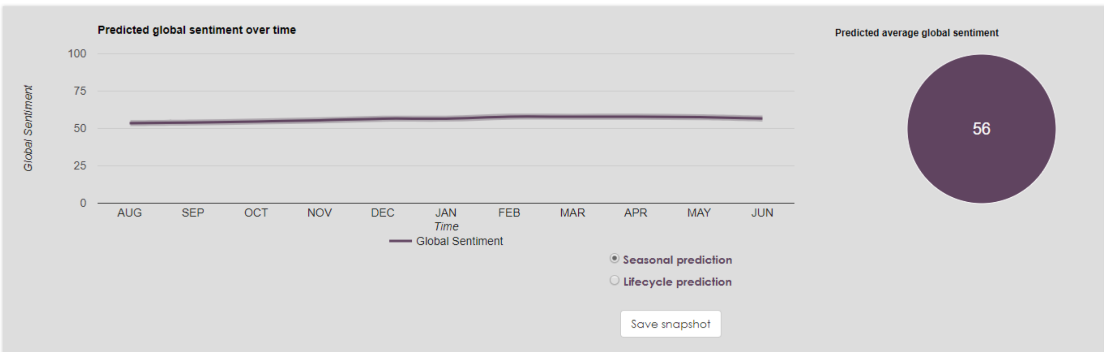


Figure 4.27: Prediction Results

4.5 Social Feedback Extraction

The Social Feedback Extraction Component retrieves Posts from Social Media Sources and stores them so that Opinion Monitoring may pool these results from time to time . The Social Feedback Extraction component retrieves posts from the following sources:

- Facebook accounts
- Twitter Users
- Amazon Product Pages

Although Amazon product pages are not posts, the user review and textual feedback that user leave contain relevant data for processing. Information is retrieved at regular intervals.

The Social Feedback Extraction (SFE) can be organized in four modules:

- Facebook module - In charge of all requests towards Facebook
- Twitter module - In charge of all requests towards Twitter
- Amazon module - In charge of all requests towards Amazon
- Management module - Responsible for organizing and queuing requests for each Social Network

The SFE component supports connection to three online platforms, two Social Media Networks, namely Facebook and Twitter and one online store Amazon Review pages. Since Amazon is not a Social Networks like the other two platform the data retrieved is more like ratings and reviews of registered users instead of the usual conversation seen in Facebook and Twitter. Twitter and Amazon posts are retrieved on one simple request; Facebook on the other hand, and due to the big amount of data, demands making multiple connections to retrieve Post metadata (also called scraping). Facebook API uses a pagination mechanism, so sometimes the modules requires multiple requests until one opinion is completely fetched.

SFE has no visualization component and works only as a gathered and endpoint creator. These endpoints as mostly used by opinion monitoring to let SFE know which sources and account should start or stop being requested.

Chapter 5

Code Structure

5.1 Sentiment Analysis

Most of Sentiment Analysis tool was developed from scratch. The exception was the calculation of the polarity, in this case Watson Natural Language Understanding from IBM was used. This tool also was built on top of existing open source technologies (Table 5.1)

Table 5.1: APIs and Libraries used by Sentiment Analysis

Technology	Description
MySQL Connector Java 5.1.39	Library to handle communication between server end-point and the MySQL server.
Org JSON	Library to define and access the object that transfer information to and from the front-end page.
Apache Tomcat v9.0	Library for all the integration and creation of a server end-point.
JUNIT 4	API for testing purposes on the Backend(server).
Selenium	API for testing purposes on the Frontend(webpage).
Java EE	Common Java EE libraries for server usage.

The development was structured according to the architecture seen of figure 5.1.

Source Code is divided into packages. The following tables describe the packages and how they were implemented.

General - Package where all the program settings, and definition of classes that are used throughout the program exist. Examples in 5.2.

Extraction - Package where all the program settings, and definition of classes that are used throughout the program exist. Examples in 5.3.

Modelling (table 5.4) - Package that handles model data modification

Table 5.2: General Classes

Class	Represents	Example Information
Author	A Single Author	Name, Age, Location
Backend	Endpoint for all front-end requests	N/A
Company	A Single Company	Name, Id
Data	Data Class	Definition of all stored information variables
Model	A Single Model	Name, PSS, Products
Opinion	A Single Opinion	Reach, Author, Replies
Post	A Single Post	Author, Message, Polarity
Product	A Single Product	Name, PSS, Supplied by
PSS	A Single PSS	Id, Company, Name
Settings	Static Variables for common usage	Local Database Replication
Startup	Startup sequence	N/A
DBHandler	Logger Connection to the Database	N/A
Design Project	A Design Project	Id, Name, Time Created, Team
Loader	Loader Class manages all new and old information calculations.	N/A
LoadThreads	Definition of some Thread Classes to be used by the Loader.	N/A
Logging	Modified Logger Class to implement custom settings.	N/A
Operations	Front-end Operation Request Decoder	getConfig -> Code 12
Role	A Single Role from User	Permissions, Role name
User	Employee Information	Username, First and Last name

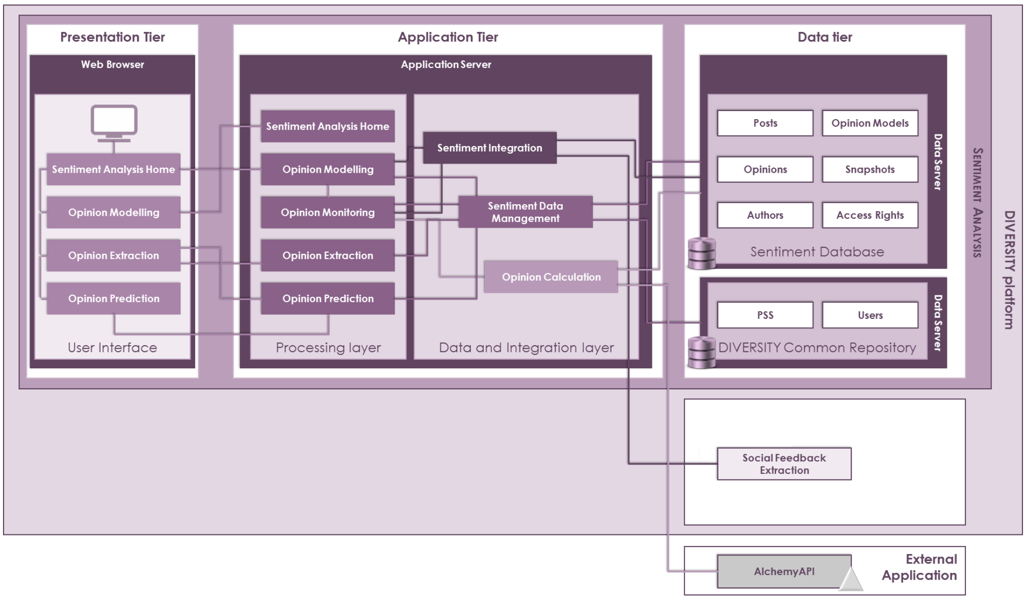


Figure 5.1: Sentiment Analysis Conceptual Architecture

Table 5.3: Extraction Classes

Class	Brief Description
GetComments	Extracts all replies information from a specific parent post
GetPosts	Extracts all parent posts from a specific PSS, or products
GetProducts	Extracts all products from PSS or Company
GetReach	Extracts Reach calculated by 'Data' depending on the user inputs
GlobalSentiment	Extracts Global Sentiment calculated by 'Data' depending on the user inputs
Extrapolation	Manages all requests and calculations required to correctly show extrapolation results based on user inputs
Prediction	Extracts Prediction results based on user, service and products combinations, selections on front-end
Snapshots	Saves and Loads user snapshots on request.
TagCloud	Generates tag cloud based on model requested.
Collaboration	Extracts Team Information to be show on user request.

Class	Brief Description
GetModels	Handles all edit, create and delete actions of models.

Table 5.4: Modelling Class

Class	Brief Description
Monitor	Handles requests, from the user, to start and stop monitoring specific sources and accounts, and reports these changes to SFE
Oversight	Time recurring task that handles data update, and data consistency. Run once every day fetching all data from last update until that specific time.

Table 5.5: Monitoring Classes

Class	Brief Description
Roles	Check database for Roles so that when entering the Homepage user gets the correct permissions.

Table 5.6: Security Class

Class	Brief Description
GetNumPosts	Returns number of posts present on the database about a specific PSS
GetNumSources	Returns number of sources present on the database about a specific PSS
GetPrediction	Generates Prediction Based on GET input
GetSnapshots	Returns List of Snapshots from a specific PSS

Table 5.7: Endpoints Classes

Monitoring (Table 5.5) - Package that contains the class Oversight, that handles daily updates of the information on the software and manages source update times.

Security (Table 5.6) - – Package that handles data security, checks database for user permissions when data is requested. Allowing said information to be displayed only if the user can see it.

Endpoints (Table 5.7) - Package that contains all endpoint's code present in the software.

Test (Table 5.8) – Testing package all JUNIT testing exists only inside this package.

5.2 Social Feedback Extraction

SFE is mainly composed of APIs that were then designed to work together and gather data the most efficient way possible.

Class	Brief Description
BackendTest	Testing class to verify if everything implemented keeps working as intended, after new implementations.

Table 5.8: Test Class

Table 5.9: APIs and Libraries used by SFE

Technology	Description
MySQL Connector Java 5.1.39	Library to handle communication between server end-point and the MySQL server.
facebook4j	Library for retrieving Posts and Metadata from Facebook.
Apache Tomcat v9.0	Library for all the integration and creation of a server end-point.
twitter4j	Library for retrieving Posts and Metadata from Twitter.

While Facebook and Twitter APIs exist and are easily found. Amazon does not have an actual API.

Code is split into packages that are organized as follows.

Table 5.10: Amazon Classes

Class	Brief Description
AmazonAgent	Organizing Page Requests and Crawls Page for useful information.
Review	A Single found review.

Table 5.11: Twitter Classes

Class	Brief Description
TwitterAccount	One Single Account
TwitterAgent	The connection and parameters required to successfully connect to Twitter
TwitterHandler	Organizing Page Requests and parse management
TwitterPost	A Single Post.

Table 5.12: Facebook Classes

Class	Brief Description
FacebookAgent	Organizing Page Requests and parse management
FacebookUtil	The connection and parameters required to successfully connect to Twitter
MyAuthor	A Single Author
MyPost	A Single Post.

The AmazonAgent works by parsing the entire page found and looks for identified keyword that point to reviews. It then parses those inputs along to Review that parses one review in a way that can be stored in the database.

Table 5.13: Endpoints Classes

Class	Brief Description
GetSimulatedData	Endpoints that return the information present on the database that matched the parameters requested
LoadPosts	Endpoint to force Post Request from Sources
RegisterAccount	Endpoint for Opinion Monitoring to register new account to be pooled

Chapter 6

Results

The tool requires large amount of inputs (posts) to be able to generate results. Because of this, and since real data proved not be well processed by available NLP tools, generating posts, conversations and actors is required. To do this a different tool was used, OpinionSim [22]. This tool allowed for the generation of conversations based on requirements like PSS/Product talked about, amount of posts, authors with different amounts of created posts, allowing for different influences and sentiment for different users and segments. This tool generated inputs with keyword like awesome, mediocre and awful which were then identified by this work's tool and given a previously defined sentiment. This array of options allowed for the testing of all the features with the creation of 5 PSSs with different trend. The Product Trees are defined as follows (see figures 6.1, 6.2, 6.3) .

Three branches were defined on product level. Notice that they all share the same root node (Company A). So, when using calculating similarities between product from different branches, the LCA is the Company.

On the service level and to reduce complexity only one level of services was introduced on this testing environment (figure 6.4).

With these Product and Services created it is now possible to define PSS. Three PSS

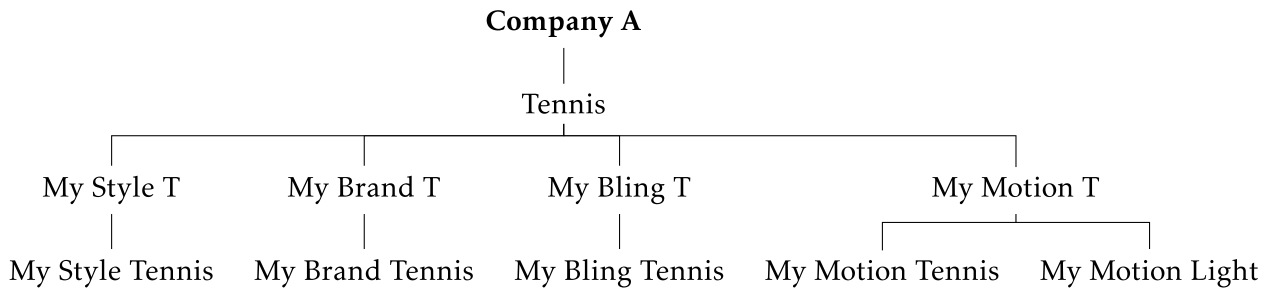


Figure 6.1: Product Tree Part 1

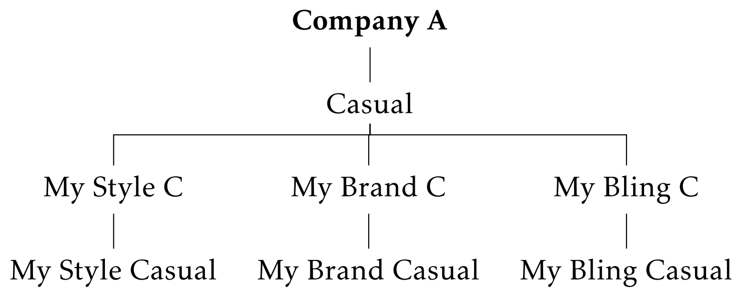


Figure 6.2: Product Tree Part 2

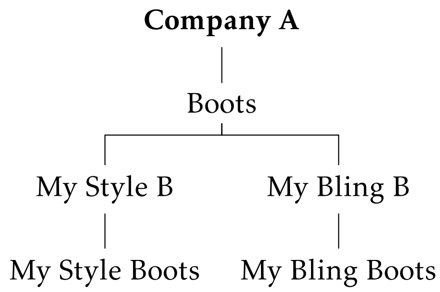


Figure 6.3: Product Tree Part 3

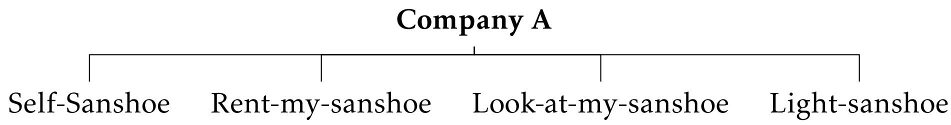


Figure 6.4: Services Tree

were then created, *My Motion Tennis PSS* (1), *My Style Boots PSS*(2) and *My Brand Casual PSS*(3).

PSS (1) consist of the product My Motion Tennis and the service Self-Sanshoe.

PSS (2) consist of the product My Style Boots and the service Self-Sanshoe.

PSS (3) consist of the product My Brand Casual and the service Light-Sanshoe.

Self-Sanshoe is a service that allows customer to personalize their shoes, picking the color they want, materials. While Light-Sanshoe is more of a cheaper version, with not so much room for customization.

The following images show the sentiment trend for each PSS.

We can see in figure 6.5, PSS (1) sentiment can be described as mild appreciation in the beginning of the PSS life-cycle with a rise on mid-life and stable end-life.

On figure 6.6, PSS (2) sentiment can be described as the opposite of PSS (1), high sentiment in the beginning and strong decline until an average appreciation toward the end of the PSS life-cycle.

Figure 6.7 gives hints to deficient product, start with an already low appreciation and never recovers. Seeing results like this the designer will investigate PSS (3) to try to understand what the issue really is. So, checking the Opinion Extraction page the user

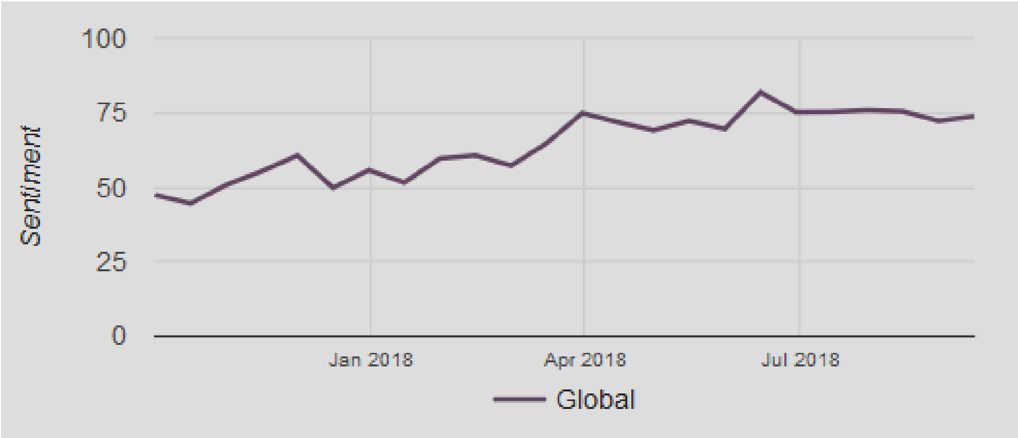


Figure 6.5: My Motion Tennis PSS Sentiment

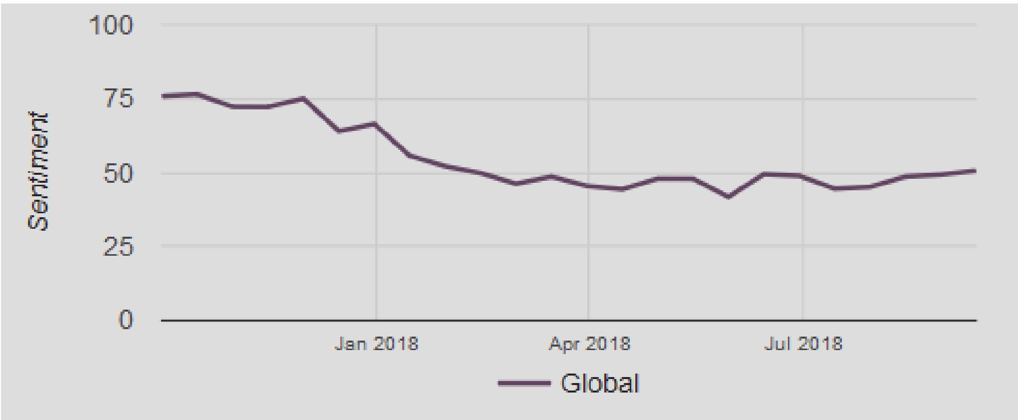


Figure 6.6: My Style Boots PSS Sentiment

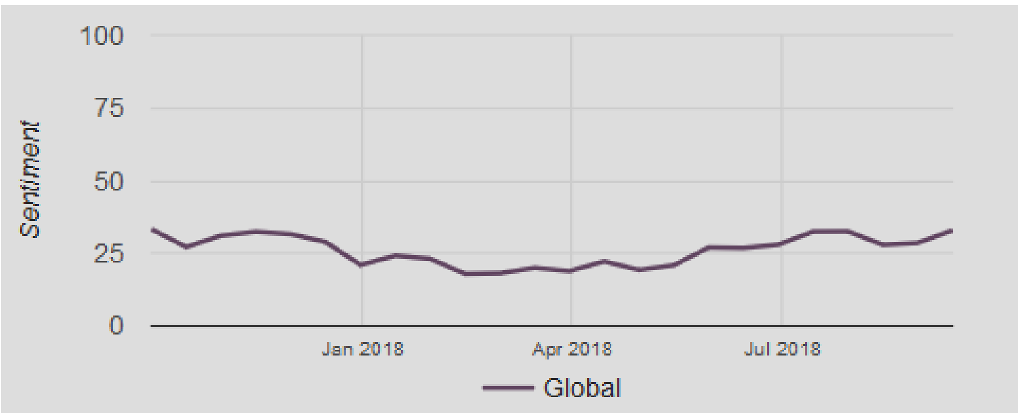


Figure 6.7: My Brand Casual PSS Sentiment

sees the interface on figure 6.8.

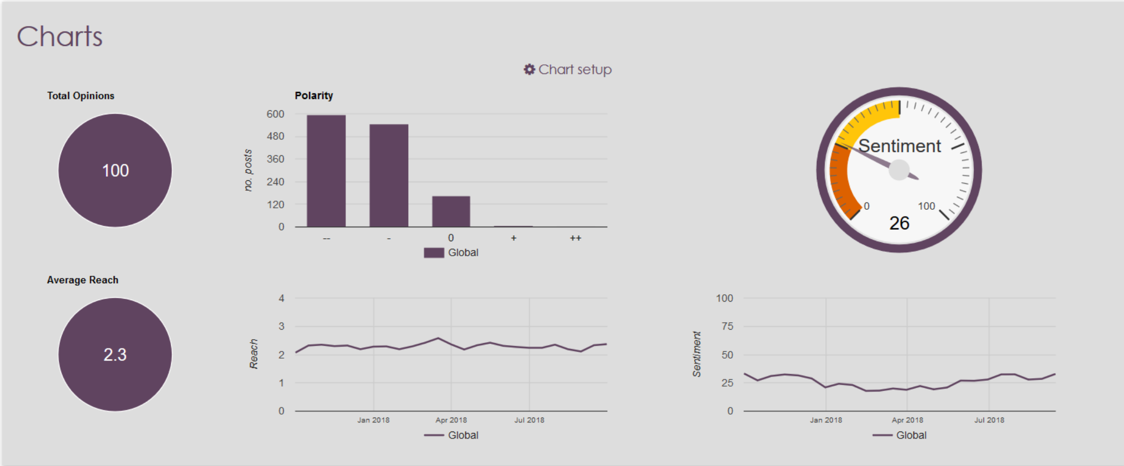


Figure 6.8: My Brand Casual Extraction Page

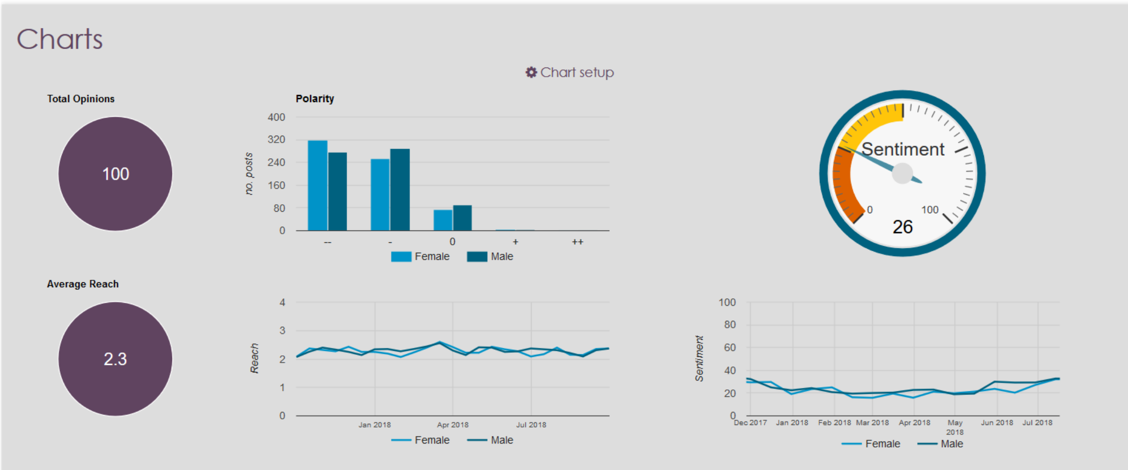
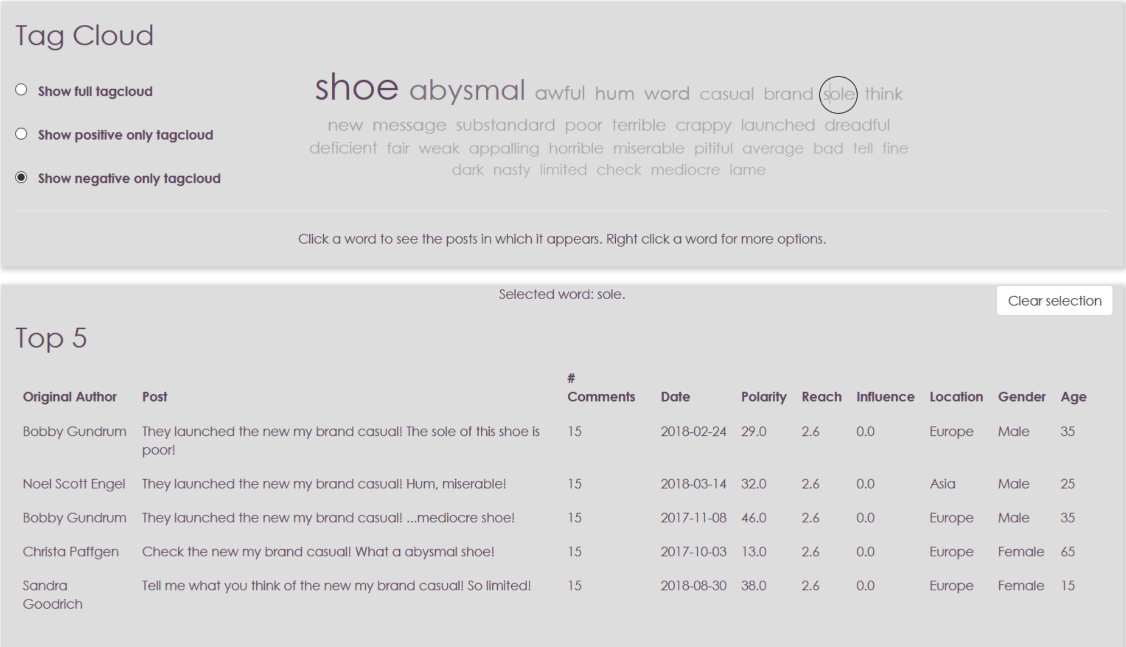


Figure 6.9: My Brand Casual Extraction Page by Gender

Since casual shoes can be worn by everyone, the designer tried to understand if the issue is a defective product or if the target audience was not the correct one, by looking at gender values figure 6.9. After analysis the designer reaches the conclusion that it must be a problem with the product itself since both genders perceive the product in the same way.

The designer then checks the Tag Cloud for common word that might give any clues about the issue (see figure 6.10). In the middle of many negative word the designer sees **sole**. By clicking in the word sole Top 5 shows only opinions where the sole is mentioned, and without even drilling down to the comments, the designer, can see that one of the original authors classifies the sole as poor. One user even goes as far as classifying as awful (figure 6.11).

The designer successfully identified the issue. People are not happy with the current sole on this kind of product. A new PSS is required or an update to the current one.



Original Author	Post	# Comments	Date	Polarity	Reach	Influence	Location	Gender	Age
Bobby Gundrum	They launched the new my brand casual! The sole of this shoe is poor!	15	2018-02-24	29.0	2.6	0.0	Europe	Male	35
Noel Scott Engel	They launched the new my brand casual! Hum, miserable!	15	2018-03-14	32.0	2.6	0.0	Asia	Male	25
Bobby Gundrum	They launched the new my brand casual! ...mediocre shoe!	15	2017-11-08	46.0	2.6	0.0	Europe	Male	35
Christa Paffgen	Check the new my brand casual! What a abysmal shoe!	15	2017-10-03	13.0	2.6	0.0	Europe	Female	65
Sandra Goodrich	Tell me what you think of the new my brand casual! So limited!	15	2018-08-30	38.0	2.6	0.0	Europe	Female	15

Figure 6.10: My Brand Casual Tag Cloud and Top 5

Figure 6.11: User Reply

The problem remains, how is the company supposed to recover this PSS, what should be changed or added.

The designer then decides to simulate what would happen if a different combination of services were used. Products are much less versatile to change mid-life so the designer

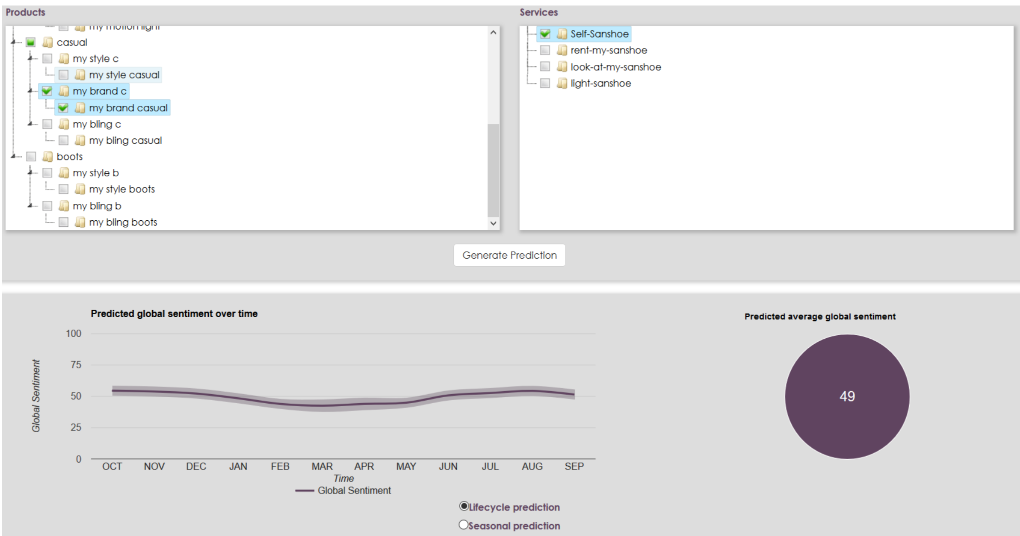


Figure 6.12: Prediction with a different Service

combines the existing product with another existing services figure 6.12. The designer found one possible solution, it is expected that if this new service is added, Self-Sanshoe described earlier, the PSS should recover not to high sentiment levels but at least to more average values.

Another option would be designing a new PSS. This process would be costlier when compared to service reallocation and according to the predicted result the difference would not be significant (figure 6.13).

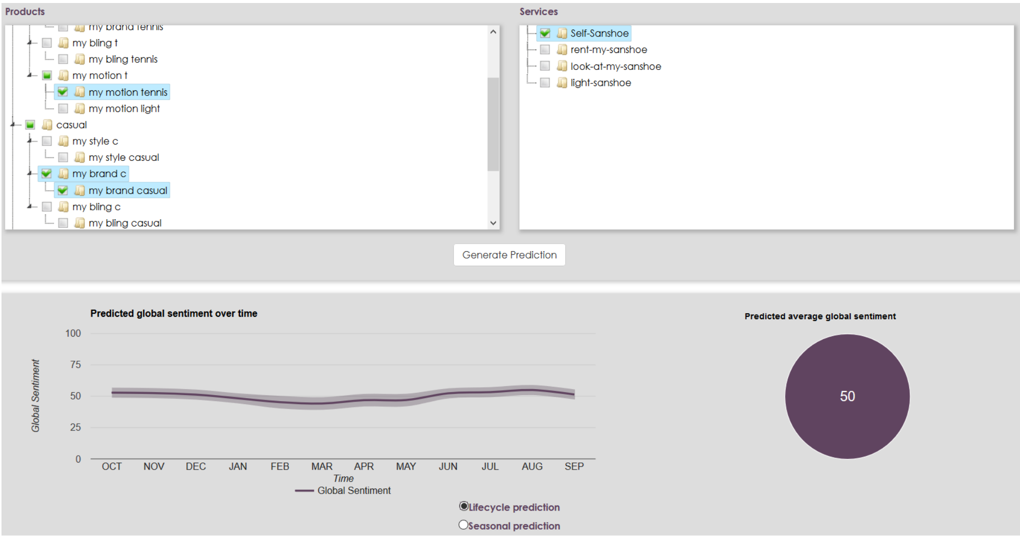


Figure 6.13: Prediction of a New PSS

So, the designer would probably decide to redesign the current PSS with the existing service, to mitigate issues, and for future PSS designs this product would have a low appreciation and therefore designers would avoid repeating the same mistake.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Social Networks capabilities are barely tapped by companies to gather feedback. The amount of data generated each day in a Social Network like Facebook turn a simple feedback analysis a full-time job. The upside of having almost instant feedback on these works leads to the downside of too much time to evaluate this feedback.

By using simulated data, the benefits of a platform like this were proven real, the way the designer easily identified the problem and started to plan PSS redesign to recover customer appreciation is something that can make the difference between total product failure, and a little hiccup on the road to success.

This tool was validated by all 3 companies that were part of the consortium. These companies tested the tool using generated data, as it was shown in this dissertation, and with real data fetched from social networks. Data from social networks was also pooled into the platform but the volatility of the comments and the language limitation of the external tool, only the most common languages are supported, lead to unreliable textual sentiment evaluation. These issues were already discussed on section 2.4, and further improvements to sentiment extraction tool are required to better analyze real data. However even with all the limitation, using a small range of data, the platform's Tag Cloud identified issues in some shoes, e.g. not water-proof or too high cost.

Current NLP tools are not developed enough to sustain a platform like this but with the development of more advance machine learning algorithms, a robust Sentiment Analysis platform can be deployed and successfully used by companies to better understand customer needs.

The tool that was developed to support this dissertation could be applied for other markets, e.g. by changing PSS for something like school departments and teacher, it could give good feedback about how students are evaluating the teachers. One company in currently attempting to gather investors to develop this tool into a marketable tool.

7.2 Future Work

While Prediction is mathematically correct, it still requires further validation by deploying on a real company situation. With the ever-emerging concept of machine learning, it is expected that applying machine learning to algorithms to work developed in this dissertation could further improve it. By taking into consideration reach and influence equations (3.1 and 3.2) and using machine learning to fine tune the weight variables could provide a more objective weight evaluation instead of what a user might think are the correct weights.

Bibliography

[1] V. Alkkiomäki and K. Smolander. "Anatomy of one service-oriented architecture implementation and reasons behind low service reuse." In: *Service Oriented Computing and Applications* 10.2 (2016), pp. 207–220. ISSN: 18632394. DOI: 10.1007/s11761-015-0181-2. URL: http://link.springer.com/10.1007/s11761-015-0181-2.

[2] J. F. Allen. "Natural language processing." In: *Encyclopedia of Computer Science* (2003), pp. 1218–1222.

[3] T. Arnold and L. Tilton. *Humanities data in R: Exploring networks, geospatial data, images, and text.* 2015, pp. 1–211. ISBN: 9783319207025. DOI: 10.1007/978-3-319-20702-5.

[4] M. Boehm and O. Thomas. "Looking beyond the rim of one's teacup: A multidisciplinary literature review of Product-Service Systems in Information Systems, Business Management, and Engineering & Design." In: *Journal of Cleaner Production* 51 (2013), pp. 245–250. ISSN: 09596526. DOI: 10.1016/j.jclepro.2013.01.019. URL: http://dx.doi.org/10.1016/j.jclepro.2013.01.019.

[5] M. Cataldi, A. Ballatore, I. Tiddi, and M. A. Aufaure. "Good location, terrible food: detecting feature sentiment in user-generated reviews." In: *Social Network Analysis and Mining* 3.4 (2013), pp. 1149–1163. ISSN: 18695469. DOI: 10.1007/s13278-013-0119-7.

[6] G. G. Chowdhury. "Natural Language Processing." In: *Annual Review of Applied Linguistics* 37.1 (2003), pp. 51–89. ISSN: 0267-1905. DOI: 10.1017/S0267190500001446. arXiv: 0812.0143v2.

[7] A. Colicev, A. Malshe, K. Pauwels, and P. O'Connor. "Improving Consumer Mindset Metrics and Shareholder Value Through Social Media: The Different Roles of Owned and Earned Media." In: *Journal of Marketing* (2018). ISSN: 0022-2429. DOI: 10.1509/jm.16.0055.

[8] M. Denny. "Social Network Analysis." In: *Institute for Social Science Research* 10.September (2014), pp. 1–20.

[9] J. L. Greenwald, P. R. Cronin, V Carballo, G Danaei, and G Choy. "A Novel Model for Predicting Rehospitalization Risk Incorporating Physical Function, Cognitive Status, and Psychosocial Support Using Natural Language Processing." In: *Medical Care* 55.3 (2016), pp. 261–266. ISSN: 1537-1948. DOI: 10.1097/MLR.0000000000000651. URL: http://www.embase.com/search/results?subaction=viewrecord{\&}from=export{\&}id=L612246047{\%}0Ahttp://dx.doi.org/10.1097/MLR.0000000000000651{\%}0Ahttp://zp9vv3zm2k.search.serialssolutions.com/?sid=EMBASE{\&}issn=15371948{\&}id=doi:10.1097{\%}2FMLR.0000000000000651{\&}atitle=A+N.

[10] M Harman, Y Jia, J Krinke, W. B. Langdon, J Petke, and Y Zhang. "Search Based Software Engineering for software product line engineering: A survey and directions for future work." In: *18th International Software Product Line Conference, SPLC 2014* 1 (2014), pp. 5–18. ISSN: 9781450327404 (ISBN). DOI: 10.1145/2648511.2648513. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84907846595{\&}partnerID=40{\&}md5=ec7935a979d0e3defebcf6fa236fe79b.

[11] G. Hoberg and G. Phillips. "Text-Based Network Industries and Endogenous Product Differentiation." In: *Journal of Political Economy* 124.5 (2016), pp. 1423–1465. ISSN: 0022-3808. DOI: 10.1086/688176. arXiv: arXiv:1011.1669v3. URL: http://www.journals.uchicago.edu/doi/10.1086/688176.

[12] K. Hribernik, M. Franke, P. Klein, and E. Coscia. "Towards a Platform for Integrating Product Usage Information into Innovative Product-Service Design." In: *23rd ICE/IEEE International Technology Management Conference*. 2017, pp. 1449–1455. ISBN: 9781538607749.

[13] M. Hu and B. Liu. "Mining and summarizing customer reviews." In: *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04* (2004), p. 168. ISSN: 1581138889. DOI: 10.1145/1014052.1014073. arXiv: 0112017 [cs]. URL: http://portal.acm.org/citation.cfm?doid=1014052.1014073.

[14] J. L. Jin, K. Z. Zhou, and Y. Wang. "Exploitation and Exploration in International Joint Ventures: Moderating Effects of Partner Control Imbalance and Product Similarity." In: *Journal of International Marketing* 24.4 (2016), pp. 20–38. ISSN: 1069-031X. DOI: 10.1509/jim.15.0164. URL: http://journals.ama.org/doi/10.1509/jim.15.0164.

[15] Joe Kozhaya. *Improve your machine learning solution with established performance metrics - IBM Code*. 2018. URL: https://developer.ibm.com/code/2018/04/10/ai-machine-learning-solution-performance-accuracy-precision-recall/ (visited on 01/09/2019).

[16] E. Kouloumpis, T. Wilson, and J. D. Moore. "Twitter sentiment analysis: The good the bad and the omg!" In: *Icwsm* 11 (2011), pp. 538–541.

[17] C. Kowalkowski, H. Gebauer, B. Kamp, and G. Parry. "Servitization and deservitization: Overview, concepts, and definitions." In: *Industrial Marketing Management* 60 (2017), pp. 4–10. ISSN: 00198501. DOI: 10.1016/j.indmarman.2016.12.007.

[18] R. Lauterborn. "New Marketing Litany: Four Ps Passé: C-Words Take Over." In: *Advertising Age* 61.24 (1990), p. 26.

[19] R. A. Litz and J. M. Pollack. "Interfirm Rivalry between Small Hardware Stores and "Big Box"Retailers: Market Commonality and Product Mix Similarity as Antecedents to Competitive Response." In: *Journal of Small Business Management* 53.2 (2015), pp. 436–449. ISSN: 1540627X. DOI: 10.1111/jsbm.12071.

[20] R. Mihalcea, C. Banea, and J. Wiebe. "Learning multilingual subjective language via cross-lingual projections." In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* June (2007), pp. 976–983. ISSN: 0736587X. DOI: citeulike-article-id:3270776. URL: http://acl.ldc.upenn.edu/p/p07/p07-1132.pdf.

[21] G. Neves, S. F. Cooke, and T. V. P. Bliss. "Synaptic plasticity, memory and the hippocampus: a neural network approach to causality." In: *Nature Reviews Neuroscience* 9.1 (2008), pp. 65–75. ISSN: 1471-003X. DOI: 10.1038/nrn2303. URL: http://www.nature.com/doifinder/10.1038/nrn2303.

[22] R. Neves-silva and P. Pina. "OpinionSim : modelling and simulating stakeholders ' feedback to product-service systems." In: (2017), pp. 1473–1478.

[23] R. Neves-Silva, P. Pina, P. Spindler, G. Pezzotta, D. Mourtzis, M. Lazoi, D. Ntalaperas, and A. R. Campos. "Supporting Context Sensitive Lean Product Service Engineering." In: *Procedia CIRP*. Vol. 47. 2016, pp. 138–143. ISBN: 2212-8271. DOI: 10.1016/j.procir.2016.03.103.

[24] B. Pang and L. Lee. "Opinion Mining and Sentiment Analysis." In: *Foundations and Trends® in Information Retrieval* 2.1-2 (2008), pp. 1–135. ISSN: 16130073. DOI: 10.1561/1500000001. arXiv: 1703.04009.

[25] B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2002), pp. 79–86. ISSN: 1554-0669. DOI: 10.3115/1118693.1118704. arXiv: 0205070 [cs]. URL: http://portal.acm.org/citation.cfm?id=1118693.1118704.

[26] P. Pina, G. Góis, and F. Mano. "A sentiment analysis solution for the re-design of product-service systems." In: *23rd ICE/IEEE International Technology Management Conference* (2017), pp. 1342–1347.

[27] U. Ramanathan, N. Subramanian, and G. Parrott. "Role of social media in retail network operations and marketing to enhance customer satisfaction." In: *International Journal of Operations & Production Management* (2017). ISSN: 0144-3577. DOI: 10.1108/IJOPM-03-2015-0153. arXiv: /dx.doi.org/10.1108/BIJ-10-2012-0068 [http:].

[28] U. Ramanathan, N. Subramanian, and G. Parrott. "Role of social media in retail network operations and marketing to enhance customer satisfaction." In: *International Journal of Operations and Production Management* 37.1 (2017), pp. 105–123. ISSN: 17586593. DOI: 10.1108/IJOPM-03-2015-0153. arXiv: /dx.doi.org/10.1108/BIJ-10-2012-0068 [http:]. URL: http://www.emeraldinsight.com/doi/10.1108/IJOPM-03-2015-0153.

[29] D. Ramanujan, W. Z. Bernstein, S. K. Chandrasegaran, and K. Ramani. "Visual Analytics Tools for Sustainable Lifecycle Design: Current Status, Challenges, and Future Opportunities." In: *Journal of Mechanical Design* 139.11 (2017), p. 111415. ISSN: 1050-0472. DOI: 10.1115/1.4037479. URL: http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4037479.

[30] W. Reim, S. Lenka, J. Frishammar, and V. Parida. "Implementing Sustainable Product-Service Systems Utilizing Business Model Activities." In: *Procedia CIRP* 64 (2017), pp. 61–66. ISSN: 22128271. DOI: 10.1016/j.procir.2017.03.130. URL: http://dx.doi.org/10.1016/j.procir.2017.03.130.

[31] K Schouten, F. F.I. T. on Knowledge, and U. 2016. "Survey on aspect-level sentiment analysis." In: *Ieeexplore.Ieee.Org* 28.3 (2016), pp. 813–830. URL: http://ieeexplore.ieee.org/abstract/document/7286808/.

[32] G. Schuh, M. Riesener, and M. Ebi. "Cost-orientated product variant evaluation using similarity analysis." In: *PICMET 2017 - Portland International Conference on Management of Engineering and Technology: Technology Management for the Interconnected World, Proceedings* 2017-Janua (2017), pp. 1–6. DOI: 10.23919/PICMET.2017.8125399.

[33] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. "Lexicon-Based Methods for Sentiment Analysis." In: *Computational Linguistics* 37.2 (2011), pp. 267–307. ISSN: 0891-2017. DOI: 10.1162/COLI_a_00049. URL: http://www.mitpressjournals.org/doi/10.1162/COLI{_}a{_}00049.

[34] J. J. Thomas and K. A. Cook. "Illuminating the Path: The R&D Agenda for Visual Analytics." In: *National Visualization and Analytics Center* January 2005 (2005).