Decision Support System for Corporate Reputation Based Social Media Listening Using a Cross-Source Sentiment Analysis Engine

R. E. Loke[a] and S. Pathak
Centre for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Keywords: Consumer Reviews, Decision Support System, Social Media Channels, Opinion Mining, Social Media Listening, Sentiment Analysis Engine.

Abstract: This paper presents a Decision Support System (DSS) that helps companies with corporate reputation (CR) estimates of their respective brands by collecting provided feedbacks on their products and services and deriving state-of-the-art key performance indicators. A Sentiment Analysis Engine (SAE) is at the core of the proposed DSS that enables to monitor, estimate, and classify clients' sentiments in terms of polarity, as expressed in public comments on social media (SM) company channels. The SAE is built on machine learning (ML) text classification models that are cross-source trained and validated with real data streams from a platform like Trustpilot that specializes in user reviews and tested on unseen comments gathered from a collection of public company pages and channels on a social networking platform like Facebook. Such cross-source opinion analysis remains a challenge and is highly relevant in the disciplines of research and engineering in which a sentiment classifier for an unlabeled destination domain is assisted by a tagged source task (Singh and Jaiswal, 2022). The best performance in terms of F1 score was obtained with a multinomial naive Bayes model: 0,87 for validation and 0,74 for testing.

1 INTRODUCTION

Decision Support Systems (DSSs) is the area of information systems (IS) that focuses on assisting and improving managerial decision making. A DSS is an information system that assists a company in making decisions that need judgment, determination, and a sequence of tasks. The information system aids an organization's mid- and high-level management by processing large amounts of unstructured data and accumulating information that can help in issue solving and decision making. A DSS can be human-powered, automated, or a hybrid of the two. A DSS supports rather than replaces decision makers. It addresses problems involving varying degrees of structured, nonstructured (unstructured or ill-structured), and semi-structured tasks, and prioritizes effectiveness over efficiency of decision processes (Eom and Kim, 2006). To be more specific, a DSS is an interactive computer-based information system that is designed to support solutions on decision problems. Distinguished from traditional management information systems, a DSS is decision focussed, user initiated and controlled, and combines the use of models and analytical techniques with traditional data access and retrieval functions (Liu et al., 2010).

We propose a DSS for social media (SM) listening in the approach of (Ducange et al., 2019) that allows businesses to use the rich insights shared over SM to generate effective, long-term business plans that comply to their digital marketing initiatives. Commercial entities can utilize the suggested system to monitor user comments posted on public SM accounts. At the system's core is a Sentiment Analysis Engine (SAE) that analyses texts retrieved from various sources of data, specifically from a set of pages and channels on a SM platform with the goal to detect polarity in opinions. A DSS helps businesses analyse their SM streams, evaluates on the success of marketing initiatives, is promptly notified of any unexpected changes, and responds properly to any unfavourable events or trends. Even though DSS has several important functionalities (Ducange et al., 2019), this paper mainly focuses on the needed SAE

[a] https://orcid.org/0000-0002-7168-090X

Loke, R. and Pathak, S.
Decision Support System for Corporate Reputation Based Social Media Listening Using a Cross-Source Sentiment Analysis Engine.
DOI: 10.5220/0012136400003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 559-567
ISBN: 978-989-758-664-4; ISSN: 2184-285X

that is commonly application domain dependent and most difficult to engineer and setup in a new domain.

An important aspect of information-gathering behaviour has always been to find out what other people think. As seen by the increased availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people can, and do, actively use information technologies to seek out and understand the opinions of others. Sentiment analysis (SA), also known as opinion mining, examines people's opinions and emotions toward entities such as products, organizations, and the attributes that are associated with them. In the modern world, SM is crucial for providing information about any product through various blogs, reviews, and comments. Different machine learning (ML) approaches are used by academics and professionals to derive useful information from people's sentiments (Liu, 2012). SA has a wide range of applications. For example, in question-answering systems, knowing the opinions of various sources can help users find better answers (Stoyanov et al., 2006). It is a valuable tool for a variety of problems in psychology, education, sociology, business, political science, and economics (Hutto and Gilbert, 2014), as well as research domains like natural language processing, data mining, and information retrieval (Zhang et al., 2014). SA may also help firms automate decision making by assisting them in better understanding the effects of specific issues on people's views about their products or services and correctly responding to these effects through marketing and communication (Sauter, 2014).

There are numerous ways to do SA and there are different methods for automatically classifying data. Lexicon-based, ML-based, and hybrid approaches are used to analyse texts based on their sentiment (Dhaoui et al., 2017). ML approaches work by first training a model to an example dataset with specific inputs and known outputs before using it later to a real dataset that has new and unknown data (Devika et al., 2016). Although lexical methods do not rely on labeled data, it is hard to create a unique lexical-based dictionary that is applicable in different contexts (Gonçalves et al., 2013) and that’s why in this paper a ML approach is adopted. ML approaches for sentiment classification are gaining interest because of (A) their ability to model many features and in doing so, capturing context, (B) their easier adaptability to changing input, and (C) their ability to measure the degree of uncertainty by which a classification is made. The most widely used models in the ML approach are supervised ones that learn from examples that have been manually categorised by people in an a priori process (Boiy and Moens, 2009).

Recent studies have focused heavily on ways for classifying sentiment across several domains (Singh and Jaiswal, 2022; Bollegala et al., 2015) or across several sources within a single domain (Ducange et al., 2019). In the latter approach a set of readily available labeled texts in a first rich source is abundantly exploited for learning required system parameters of a sentiment classifier that is generisable to another source that is much shallower (Ducange et al., 2019). In this study, the approach that has been laid out in Ducange et al. (2019), will be followed in which a ML based approach is taken for opinion mining that regards SA as a standard classification problem. The following list from Devika et al. (2016), includes some of the most well-known ML-based models that can be employed in the approach: support vector machine; n-gram SA; naïve Bayes method; maximum entropy classifier; k-nn and weighted k-nn; multilingual SA; feature driven SA.

Whereas in Ducange et al. (2019) the scenarios are those of restaurants and consumer electronics online shopping with source websites TripAdvisor and Amazon and target websites Facebook, Instagram, and Twitter, we deliberately pick a broadband scenario with source website Trustpilot and target website Facebook that is relevant for corporate reputation (CR) based SM listening in the telecommunication domain.

Application of the framework of Ducange et al. (2019) in this domain yields some challenges that need to be tackled. Firstly, as noted in Fazzolari et al. (2017) and Valdivia et al. (2018), there can be discrepancies between the sentiment assessment of the review texts and the corresponding user ratings in online reviews. One strategy to possibly deal with this issue is that only the reviews at the extremes of the evaluation scale are chosen to be used in training to achieve a high chance that the language of the reviews is coherent with the assessment label.

Secondly, datasets that have a highly imbalanced class distribution present a fundamental challenge in ML, not only for training a classifier, but also for evaluation (Raeder et al., 2012). Although it has been observed that in some domains or for some datasets, for example the Sick dataset, standard ML algorithms can induce good classifiers, even using highly imbalanced training sets (Guo et. al, 2008), imbalanced datasets typically lead to classification problems because the classes are not represented equally. Using the right tools and techniques might help in developing better classification models (Nabi, 2018). Regarding evaluation, when optimal model performance metrics are chosen for classifiers, models are most likely to yield the best performance by a considerable margin (Forman, 2003). Importantly, F1 score, precision and recall have been noted to be effective metrics for information retrieval, where the imbalance problem exists (Guo et. al, 2008).

We approach to support decision making by analysing the sentiments of SM posts that are related to broadcom products and services offered by telecom companies on Facebook in combination with those of reviews on Trustpilot. The data required for our analysis has been collected for the telecommunication company Vodafone UK as well as some SMEs found on their SM channels (Facebook) and Trustpilot. The use of a SAE for the processing of SM data in a DSS is not new, see e.g., Cresswell et al. (2020). However, it is the first time that this is done with CR measurement objectives for companies in mind.

The organization of the paper is as follows. The next section, background, will reflect on the importance of DSSs in relation to SM listening to monitor CR. The third section consists of a concise description of the research methodology and data used. The fourth section comprises the results. Finally, the last section consists of a discussion and conclusion as well as the limitations and scope of future research.

2 BACKGROUND

2.1 Social Media (SM) Listening

Internet has not only changed the way people buy music, organize vacations, and research school projects, but it has also affected how we interact socially. People can exchange photos and videos, share news headlines, post their ideas on blogs, and participate in online discussions via SM. SM allows individuals, companies, organizations, governments, and parliaments to interact with many people. In conjunction with the increase in online activity, there are certain concerns about the ways in which the personal information that is shared via the SM of users may be collected and analysed. The term SM refers to a variety of internet and mobile services that allow users to engage in online conversations, contribute to user-generated content, and join online communities (Mayfield, 2008).

SM is because of its ease of use, speed, and outreach rapidly transforming public discourse in society and defining trends and agendas in topics ranging from the environment and politics to technology and the entertainment sector. The enormity and high variability of information that spreads via large user groups on SM gives an interesting potential for harnessing that data into a form that allows for specific predictions about specific outcomes without the need for market mechanisms. Models can also be built to aggregate the collective population's opinions and get useful insights into their behaviour as well as predict future trends. Furthermore, getting information on how people talk about specific products can be useful for creating marketing and advertising campaigns (Asur and Huberman, 2010).

The amount of information and user-generated SM content is growing quickly in the digital age, and it is likely to continue to do so in the near future, driven by the current generation of web applications, the nearly infinite connectivity, and an insatiable desire for information sharing, particularly among younger generations. People using the web are constantly invited to share their opinions and preferences with the rest of the world and it has sparked an explosion of opinionated blogs, product and service evaluations, and comments on just about anything. The significance of this kind of web-based content as a source of information for various application domains is becoming more widely acknowledged (Schouten and Frasincar, 2015).

A typical metaphor for online behaviour is listening. In fact, online engagement is sometimes confused with providing a ‘voice’. Participation in online spaces such as blogs, wikis, news sites, and discussion boards has become synonymous with ‘speaking up’ (Bruns, 2008). SM creates a strong listening subject by bringing together the disparate areas of modernity in one place but also creating a gap between ideal and what is humanly manageable. For example, a Twitter user can effectively manage an online presence for friends, family, and co-workers as well as voters or customers. However, there are complex ramifications for how this capacity can be managed across one’s working life, family and social life, and political life. There are gaps between what users are technically capable of and the constraints imposed by their schedules, desires, and bodies (Crawford, 2009).

The corporate sector was quick to recognize the value of using SM to build stronger relationships with clients, obtain product information, and improve public personae. While some politicians require their people to update their Twitter accounts, many businesses delegate this responsibility to their employees. SM can be used in many ways by businesses, for example, some companies pay professional micro bloggers to help them establish an online presence. When professionals are paid to imitate a company's or a celebrity's online presence, communications are frequently degraded to the level of an impersonal, one-way marketing broadcast. The advantage of being able to listen to consumers' opinions, reply quickly to their comments and concerns, and obtain insight into how the firm is discussed, is drastically diminished. Delegated listening is not a perfect substitute for being present.

In 2008, to improve its products and services, The Land of Nod, a division of Crate and Barrel that sells children's furniture, started monitoring comments submitted on its ratings and review pages (Stribling, 2008). This is just an example, but the point is that these days, SM is a common way for businesses to gather data and information and use it to analyse their performance.

Overall, to understand public opinion better, an increasing number of Fortune 500 firms, governmental organizations, and political campaigns are using SM. As a result, a whole cottage industry of software and platforms for SM listening has emerged (Hofer-Shall, 2010).

A lot of marketers use listening platforms to compile comments left on various SM sites. They sometimes combine the data to create simple averages. In other instances, they do not integrate the insights across venues and instead report venue-specific metrics such as the number of retweets or Facebook likes (Schweidel and Moe, 2014).

User opinions posted on SM are often about services and products of specific companies and brands. This huge amount of user viewpoints might be effectively mined and exploited as a powerful source of information in business for steering marketing strategies according to what people really think about their products and services (Balazs and Velásquez, 2016).

The importance of SM listening and customer relationship management (CRM) in current society has been discussed in Stewart et al. (2017) for example.

2.2 Corporate Reputation

CR refers to social cognitions about an organization that exist in the minds of external observers, such as information, impressions, perceptions, and beliefs. It is sometimes characterized in terms of how well a firm satisfies social expectations, such as those relating to the quality of products and services, industry leadership, and social impact (Li et al., 2013). What makes CR so vital and interesting, is that it is the first thing individuals want to know more about when deciding whether they want to invest more time, energy or other resources in an organization, and reputation is the one thing that everyone understands. Everybody can tell whether a company is good or bad based on this one factor (reputation), so you don't need to be an expert in accounting, finance, engineering, innovation, or ethics. When determining whether to work for the company, buy its products, invest in its stock, or collaborate with it, this is the first thing that people want to know. Knowing an organization's reputation is different from knowing about an organization's reputation. People stop looking for additional information about an organization once they feel like they know it. In general terms, the concept of CR can be defined as the combination of all views, decisions, and ideas of people about an organization, the belief in the organization and reliability of the organization (Karabay, 2014). There are many stakeholder groups that are related to a company that benefit from a good CR (Carroll, 2016). CR has been increasingly significant in business and plays a more decisive role in sustaining businesses' expanding market presence and long-term survival. A company's reputation is shaped, built, or destroyed during its operations in its community and market. The reputation of a corporate entity has an impact on its operations and the interactions it has with numerous other organizations. Therefore, keeping a strong reputation should be seen as a crucial aspect that will contribute to an organization's ability to create value to build a lasting reputation through time and prevent reputation erosion. Contrary to popular belief, restoring a company's reputation after a setback will not be easy to compensate (Barnett et al., 2006).

Numerous marketing researchers already acknowledged in the early days the importance of corporate image and CR in customers' purchasing behaviour (Barich and Kotler, 1991). They are still often thought of as two separate constructions that are closely linked. Given the premise that image and reputation are two socially constructed things generated mostly from a customer's view of a corporation, this relationship is intuitively appealing. Most of the studies have analysed corporate image and CR separately. At a most guarded level, some authors have expressed a potential link between the two concepts (Porter, 1985). In the present competitive environment, CR and corporate image are acknowledged as having the potential to impact on customer loyalty toward the firm. The precise nature of the relationships that exist between CR and corporate image and the understanding of their effect on customer behaviour remains a key challenge for both academia and management alike. The degree of customer loyalty tends to be higher when perceptions of both CR and corporate image are strongly favourable (Nguyen and Leblanc, 2001).

Nowadays, an important factor in improving CR is listening to the feedback of consumers and constantly working on improving products and services by using that very feedback.

The importance of CR, reputation management and big data in current society has recently been discussed in Westermann and Forthmann (2021) for example.

2.3 Importance of DSS for CR-Based Social Media Listening

In the three decades of its existence, DSS has moved from a radical movement that altered how businesses perceived information systems to a widespread commercial IT movement in which all organizations participate (Arnott and Pervan, 2015). Personal DSSs, group support systems, executive information systems, online analytical processing systems, data warehousing, and corporate intelligence are all examples in terms of modern professional practice that are dedicated to assisting and enhancing managerial decision making. Organizational decisions are vital to organizational development and DSSs support organizations in decision making and business activities and combine useful information from documents, raw data, personal knowledge, and business models to find and solve business problems. Organisational digital technologies are characterised by their ability to serve the personalised needs of the customer and to go beyond their borders, by impacting products, business processes, sales channels, and supply chains (Hess et al., 2016). The digital transformation of organisations is required given the expanding global population and the use of more digital technologies to implement predictive analytics and artificial intelligence (Heavin and Power, 2018).

The quality of a decision depends on the adequacy of the information available, the quality of the information, the number of options, and the appropriateness of the modelling effort available at the time of the decision. While it is not true that more information or even more analysis is better, however, it is true that more of the appropriate type of information and analysis is better. In fact, one may argue that improving the information collection and processing processes is necessary to improve the decision-making process (Sauter, 2014).

Most decision processes rely not only on the preferences of the decision maker but also on public opinions about possible alternatives. Therefore, user preferences have been heavily considered in the multi-criteria decision-making field. A DSS can offer different trends, scenarios, and statistics within a period and it can address structured, semi-structured and unstructured decisions (Ballouki et al., 2017).

Online SM platforms are now extremely valuable resources for supporting important business intelligence applications thanks to the development of Web 2.0 apps. The knowledge gathered from SM has the potential to help create new services that are better tailored to users' demands while also achieving the goals of the companies who provide them. Online customer views, reviews and feedback are a crucial component of SM content. When consumer evaluations are properly analysed, they not only offer useful data to support customers' buying decisions but also help retailers or product manufacturers in better understanding overall consumer attitudes toward their products in order to improve marketing campaigns.

The importance of CR related constructs such as corporate control and corporate sustainability and DSSs and big data in current society has recently been discussed in Grander et al. (2021) for example.

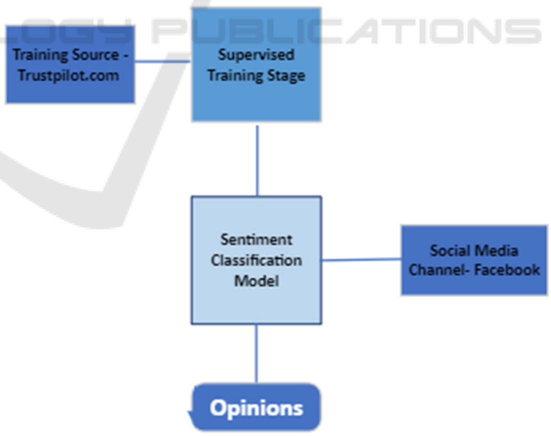

Figure 1: Scheme of the cross-source sentiment analysis; adapted from Ducange et al. (2019).

3 DATA AND METHODOLOGY

To build and apply our cross-source SAE, two data streams were used: (1) English review texts, and ratings extracted from the website Trustpilot.com and (2) English text comments from SM channels on

Facebook.com. Fig. 1 shows the data flow in the engine; it was first trained and validated on Trustpilot data and then tested on Facebook data.

To collect the data from Trustpilot a web scraper in python with scrapy was developed. In contrary, although automatized solutions are of course preferable for social listening to datastreams at scale, the Facebook data were for practical constraints manually collected from a registered user account. Importantly, to enable measurement of system performance, the Facebook comments were visually evaluated and manually rated fairly by two human, third-party experts in a three-class representation task with the following categories: (1) negative, (3) neutral, and (5) positive.

In our study, we focused on products and services in the telecommunication domain in the UK by scraping 1165 text reviews and ratings of Vodafone UK on Trustpilot and 250 text comments coming from SM channels on Facebook of five different telecom service providers and companies in the UK: Vodafone, Touch Telecom, virgin Mobile, Telecom world and Kinex Broadband. The distrubution of the rating scores in the Trustpilot dataset was as follows: 120, one star, 278, two stars, 220, three stars, 274, four stars and 273, five stars. And in the Facebook data: 156, one star, 76, five stars, 18, three stars.

In the text preprocessing and feature selection, stop words were systematically removed and lemmatization and count vectorization were algorithmatically applied (Khan et al., 2010). A total of 2504 features were extracted in the system that performed best (see Results section).

All ML classifiers were trained, validated, and tested using the vectorized training corpus by incorporating it into the classifier's code in the python sklearn library. The main task that they had focus on in the training and validation was assigning a rating score to the computed features of a review text and in the testing assigning a rating score to the computed features of a text comment.

In the training and validation phase, the selected items in the Trustpilot dataset were split into two parts with test_size in sklearn set to 0.2 and random_state to 101 to reproduce if needed. In the testing phase, all items in the Facebook dataset were used.

Since it is known from literature that there is no model that fits all, it made sense to apply several classifiers to see which one performed best in the given situation and dataset that we had under consideration. From literature we knew on the one side that random forests could possibly outperform logistic regression models in large-scale benchmarking experiments (Couronné et al., 2018) but also on the other side that there was no proof that ML algorithms performed better than logistic regression models (Christodoulou et al., 2019). We noted that the datasets in our case were rather small which suggested to us that algorithms that require a large dataset to work well, like neural networks, might not be suitable. We implemented the following classifiers in our workbench: Multinomial Naive Bayes (MNB), Random Forest, Decision Tree, Support Vector Machine (SVM), Gradient Boosting, K Neighbor (KNN), Multilayer Perceptron (MLP).

It is known that a model's behaviour can be controlled using model parameters, also known as hyperparameters (Passos and Mishra, 2022). Therefore, the parameters that could be controlled when we deployed the models using sklearn were systematically varied to determine their best settings. E.g., for random forest, the values for n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features and bootstrap were optimized using random and grid search in sklearn.

The following macro averages that measure generalized performance irrespective of a respective class were used as metrics: precision, the fraction of true positives that are predicted as positives; recall, the fraction of true positives which are actually positives; F1 score, the weighted harmonic mean of precision and recall (Chen et al., 2016); accuracy, the proportion of correct predictions i.e., the overall effectiveness of the classifier (Canbek et al., 2017).

All classifiers were run multiple times to achieve the overall unbiased estimation of performances.

4 RESULTS

Results based on macro averages have been obtained in a series of three modeling experiments. The first is when, there were equal number of reviews (120 for each of the ratings 1 and 5) in the dataset. The second is where the numbers of reviews of the extreme ratings (1 and 5) is not equal and the third one is where, reviews with three ratings (1, 3 and 5) were used to train the classifiers.

In the first experiment models performed much better than those in the second and third experiment based on the F1 score macro average that is most relevant in our case. The highest performing classifiers during validation (see Table 1) were MNB, Random Forest, Decision Tree, Gradient Boosting and MLP with the precision of 0.88, 0.80, 0.86, 0.85 and 0.86, whereas the recall was 0.87, 0.78, 0.87, 0.83 and 0.84 for each model respectively. This means that these classifiers were able to differentiate very well between text reviews with high or low rating scores.

When the top-5 performing classifiers of the validation phase were subsequently tested on the SM data, only two classifiers performed well and were able to predict the accurate ratings that were given by the expert panel (see Table 2). Specifically, the MNB and MLP performed with a precision of about 0.72 and with a recall of about 0.75. Note that we carefully manually rechecked the predicted ratings in the dataset to confirm the precise accuracy of the classifiers.

In the second and third experiment, again, MNB and MLP performed better than the other models based on F1 score; respectively, 0.82 and 0.78 for validation and 0.63 and 0.55 for testing in the second experiment and 0.72 and 0.67 for validation and 0.46 and 0.38 for testing in the third experiment.

Table 1: Results from validating the models based on macro averages, sorted on F1.

Model	Precision	Recall	F1	Accuracy
MNB	0.88	0.87	0.87	0.88
Decision Tree	0.86	0.87	0.85	0.85
MLP	0.86	0.84	0.85	0.85
Gradient Boosting	0.85	0.83	0.81	0.81
Random Forest	0.80	0.78	0.78	0.79
SVM	0.73	0.73	0.73	0.73
KNN	0.80	0.57	0.50	0.62

Table 2: Results from testing the models based on macro averages, sorted on F1.

Model	Precision	Recall	F1	Accuracy
MNB	0.73	0.75	0.74	0.76
MLP	0.71	0.74	0.71	0.72

5 DISCUSSION AND CONCLUSION

The above comparison of the performance of the several classifiers in the three modeling experiments proves that to achieve high probability that the text of a review is coherent with its evaluation label, only the extreme ratings (1 and 5) should be considered because online reviews can have inconsistencies between the sentiment evaluation of the texts and the correspondent user ratings (Valdivia et al., 2017). When the neutral sentiments (ratings=3) were used in a three-category classification task, results were worse when compared to two class classification. Although in the research conducted by Ducange et al. (2019), a three-category classification was used by using additional comments with neutral opinions from an additional independent dataset, we followed a two-class strategy because of the imbalancedness in our datasets. As mentioned by Ahmed et al. (2017) in their research of a customized SA tool for code review interactions, converting three-class dataset into two-class dataset, improved the performance in an imbalanced dataset. Importantly, test results are expected to increase when neutral center texts with respective labels will be offered to our classifiers because all neutral comments (18 out of 250) will obviously in the current situation logically result in incorrect predictions in the current situation (7.2%).

Furthermore, it is worth noting that the testing results are slightly lower than the training results. Commonly, classifiers are usually trained using data that was collected within a specific time interval. Then, yielding models are used for classifying new instances of data that are being received in online streaming such as new reviews that are being left by customers. Since the characteristics of the phenomenon under observation (in this case, reviews of telecom service providers) can change during the time elapse, the performance of the classification models may deteriorate, due to so called concept drift. Concept drift primarily refers to an online supervised learning scenario when the relation between the input data and the target variable changes over time (Gama, et al., 2014).

This study focuses on broadcom products and services provided in the telecommunication domain (by focusing on the Vodafone UK company on TrustPilot.com which is a website that contains online user reviews on products and services that generally include a text that expresses the domain and a score that may be used to label the text). Several classification models have been trained, compared, and tested to identify the most suitable ones. The best performing classification models were embedded as SAE in the respective DSS and in that role were applied to classify any text comment, extracted from related Facebook SM channels of Vodafone UK and a few other local service providers in the UK to support decision makers in understanding the implications of a particular option. In today's competitive business environment, such models are used to help clarify what and how to improve CRM. In the future, review data from the point of sale might be collected for all clients in a CRM, and data mining technologies might be used to create consumer profiles for both protected company and public online data. Such profiles might then offer managers information on trends, allowing them to change marketing campaigns or perhaps create new ones.

To test the ML models, the several models that were trained and validated with the best combination of parameters on Trustpilot data were used to predict the sentiments for SM posts in Facebook that the models were never confronted with, i.e, never had seen before. Our state-of-the-art cross company, cross source approach guarantees that external SM target data, that was never used in the setup, definition, and training of any models, can be effectively analyzed, and explored without requiring any time-consuming labeling or expensive annotation process. For building and engineering the classifiers with the source dataset, merely the reviews at the extreme values were used to maximize the correspondence of review texts with evaluation labels. Thus, a precise set of classifiers were trained and compared, each of them carrying out a 2-class classification task, tagging positive, negative opinions, and, ignoring neutral opinions.

The proposed DSS can be used by Vodafone (UK) or any other telecom company that operates within the same domain to monitor unseen, novel comments posted by their users on their public SM pages.

In the future, we aim to extend our SAE prototype for Vodafone with the use of larger datasets, implement the complete DSS around the prototyped SAE, and apply to other domains than telecom.

ACKNOWLEDGEMENTS

This paper has been inspired on the MSc project of Shubham Pathak who was involved via the master Digital Driven Business at HvA. Thanks go to D. Dey and M. Wollaert as well as several anonymous reviewers for providing some useful suggestions to an initial version of this manuscript. Rob Loke is assistant professor data science at CMIHvA.

REFERENCES

Ahmed, T., Bosu, A., Iqbal, A. & Rahimi, S. (2017). SentiCR: a customized sentiment analysis tool for code review interactions. In *32nd IEEE/ACM Int. Conf. on Automated Software Eng. (ASE)*, 106-111.

Arnott, D., & Pervan, G. (2015). A critical analysis of decision support systems research. In *Formulating research methods for information systems*, 127-168. Palgrave Macmillan, London.

Asur, S., & Huberman, B.A. (2010). Predicting the future with social media. In *IEEE/WIC/ACM int. conf. on web intelligence and intelligent agent tech.,* 1, 492-499

Balazs, J.A., & Velásquez, J.D. (2016). Opinion mining and information fusion: a survey. *Information Fusion*, 27, 95-110.

Ballouki, I., Douimi, M., & Ouzizi, L. (2017). Decision support tool for supply chain configuration considering new product re-design: An agent-based approach. *J. of Advanced Manufacturing Systems*, *16*(04), 291-315.

Barich, H., & Kotler, P. (1991). A framework for marketing image management. *MIT Sloan Management Review*, 32(2), 94.

Barnett, M.L., Jermier, J.M., & Lafferty, B.A. (2006). Corporate reputation: The definitional landscape. *Corporate reputation review*, 9(1), 26-38.

Boiy, E., & Moens, M.F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information retrieval*, 12(5), 526-558.

Bollegala, D., Mu, T., & Goulermas, J.Y. (2015). Cross-domain sentiment classification using sentiment sensitive embeddings. *IEEE Trans. on Knowledge and Data Engineering*, 28(2), 398-410.

Bruns, A. (2008). Blogs, Wikipedia, Second Life, and beyond: From production to produsage (Vol. 45). *Peter Lang*.

Canbek, G., Sagiroglu, S., Temizel, T.T., & Baykal, N. (2017). Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. In *2017 Int. Conf. on Computer Science and Engineering (UBMK)*, 821-826. IEEE.

Carroll, C.E. (Ed.). (2016). The SAGE encyclopedia of corporate reputation. *Sage Publications*.

Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: a recent review. *Artificial Intelligence Review*, *45*(1), 1-23.

Christodoulou, E., Ma, J., Collins, G.S., Steyerberg, E.W., Verbakel, J.Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. of clinical epidemiology*, 110, 12-22.

Couronné, R., Probst, P., & Boulesteix, A.L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC bioinf.*, 19(1), 1-14.

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum*, 23(4), 525-535.

Cresswell, K., Callaghan, M., Khan, S., Sheikh, Z., Mozaffar, H. & Sheikh, A. (2020). Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: a systematic review. *Health informatics j.*, *26*(3), 2138-2147.

Devika, M.D., Sunitha, C. & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, 87, 44-49.

Dhaoui, C., Webster, C.M., & Tan, L.P. (2017). Social media sentiment analysis: lexicon versus machine learning. *J. of Consumer Marketing*.

Ducange, P., Fazzolari, M., Petrocchi, M. & Vecchio, M. (2019). An effective Decision Support System for social media listening based on cross-source sentiment analysis models. *Engineering Applications of Artificial Intelligence*, 78, 71-85.

Eom, S. & Kim, E. (2006). A survey of decision support system applications (1995–2001). *J. of the Operational Research Society*, 57(11), 1264-1278.

Fazzolari, M., Cozza, V., Petrocchi, M. & Spognardi, A. (2017). A study on text-score disagreement in online reviews. *Cognitive Computation*, *9*(5), 689-701.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.*, *3*(Mar), 1289-1305.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM comp. surveys (CSUR)*, *46*(4), 1-37.

Gonçalves, P., Araújo, M., Benevenuto, F. & Cha, M. (2013). Comparing and combining sentiment analysis methods. In *Proc. of the first ACM conf. on Online social networks*, 27-38.

Grander, G., Ferreira da Silva, L. & Santibañez Gonzalez, E.D.R. (2021). Big data as a value generator in decision support systems: a literature review. *Revista de Gestão* 28(3), 205-222.

Guo, X., Yin, Y., Dong, C., Yang, G. & Zhou, G. (2008). On the class imbalance problem. In Fourth int. conf. on natural computation, 4, 192-201. *IEEE*.

Heavin, C. & Power, D.J. (2018). Challenges for digital transformation–towards a conceptual decision support guide for managers. *J. of Dec. Syst.*, *27*(sup1), 38-45.

Hess, T., Matt, C., Benlian, A. & Wiesböck, F. (2016). Options for formulating a digital transformation strategy. *MIS Quarterly Executive*, *15*(2).

Hofer-Shall, Z. (2010), The Forrester Wave: Listening Platforms, Q3, *Forrester Research*.

Hutto, C. & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proc. of the int. AAAI conf. on web and social media* 8(1), 216-225.

Karabay, M.E. (2014). Corporate reputation: a definitional landscape. In Corporate Governance, 229-240. *Springer*, Berlin, Heidelberg.

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

Li, T., Berens, G. & de Maertelaere, M. (2013). Corporate Twitter channels: The impact of engagement and informedness on corporate reputation. *Int. J. of Electronic Commerce*, 18(2), 97-126.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Liu, S., Duffy, A. H., Whitfield, R.I., & Boyle, I.M. (2010). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261-286.

Mayfield, A. (2008). What is social media? *iCrossing.*

Nabi, J. (2018). https://towardsdatascience.com/machine-learning-multiclass-classification-with-imbalanced-data-set-29f6a177c1a

Nguyen, N. & Leblanc, G. (2001). Corporate image and corporate reputation in customers' retention decisions in services. *J. of retail. & Cons. Serv.*, 8(4), 227-236.

Passos, D. & Mishra, P. (2022). A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems*, 104520.

Porter, M.E. (1985). Technology and competitive advantage. *J. of business strategy*, 5(3), 60-78.

Raeder, T., Forman, G. & Chawla, N.V. (2012). Learning from imbalanced data: Evaluation matters. In Data mining: Foundations and intelligent paradigms, 315-331. *Springer*, Berlin, Heidelberg.

Sauter, V.L. (2014). Decision support systems for business intelligence. *John Wiley & Sons*.

Schouten, K. & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Trans. on Knowledge and Data Engineering*, 28(3), 813-830.

Schweidel, D.A., & Moe, W.W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *J. of marketing res.*, 51(4), 387-402.

Singh, N. & Jaiswal, U.C. (2022) Cross Domain Sentiment Analysis Techniques and Challenges: A Survey. *4th Int. Conf. on Communication & Information Processing (ICCIP)*.

Stewart, M.C., Atilano, M. & Arnold, C.L. (2017). Improving Customer Relations with Social Listening: A Case Study of an American Academic Library. *Int. J. of Customer Relationship Marketing and Management (IJCRMM)* 8(1).

Stoyanov, V., Cardie, C., Litman, D., & Wiebe, J. (2006). Evaluating an opinion annotation scheme using a new multi-perspective question and answer corpus. In Computing attitude and affect in text: Theory and applications, 77-91. *Springer*, Dordrecht.

Stribling, W. (2008), http://www.bazaarvoice.com/blog/2008/06/20/land-of-nodturns-negatives-into-positives-for-customers/

Valdivia, A., Luzón, M.V. & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, *32*(4), 72-77.

Westermann, A. & Forthmann, J. (2021), Social listening: a potential game changer in reputation management How big data analysis can contribute to understanding stakeholders' views on organisations, *Corporate Communications: An Int. J.*, 26(1), 2-22.

Zhang, H., Gan, W. & Jiang, B. (2014), Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey, *11th Web Information System and Application Conf.*, Tianjin, China, 262-265, doi: 10.1109/WISA.2014.55.