An Effective Decision Support System for Social Media Listening Based on Cross-source Sentiment Analysis Models

Pietro Ducange[a,*], Michela Fazzolari[b], Marinella Petrocchi[b], Massimo Vecchio[c,a]

[a]SMARTEST Research Centre, eCampus University, Via Isimbardi 10, 22060, Novedrate (COMO), Italy
[b]Istituto di Informatica e Telematica - Consiglio Nazionale delle Ricerche (IIT-CNR), Via G. Moruzzi 1, 56124, Pisa, Italy
[c]OpenIoT research unit, FBK CREATE-NET, Via Cascata, 56/D, 38123 Povo (TN), Italy

Abstract

Nowadays, companies and enterprises are more and more incline to exploit the pervasive action of on-line social media, such as Facebook, Twitter and Instagram. Indeed, several promotional and marketing campaigns are carried out by concurrently adopting several social medial channels. These campaigns reach very quickly a wide range of different categories of users, since many people spend most of their time on on-line social media during the day.

In this work, a Decision Support System (DSS) is presented, which is able to efficiently support companies and enterprises in managing promotional and marketing campaigns on multiple social media channels. The proposed DSS continuously monitors multiple social channels, by collecting social media users' comments on promotions, products, and services. Then, through the analysis of these data, the DSS estimates the reputation of brands related to specific companies and provides feedbacks about a digital marketing campaign.

The core of the proposed DSS is a Sentiment Analysis Engine (SAE), which is able to estimate the users' sentiment in terms of positive, negative or neutral polarity, expressed in a comment. The SAE is based on a machine learning text classification model, which is initially trained by using real data streams coming from different social media platforms specialized in user reviews (e.g., TripAdvisor). Then, the monitoring and the sentiment classification are carried out on the comments continuously extracted from a set of public pages and channels of publicly available social networking platforms (e.g., Facebook, Twitter, and Instagram). This approach is labeled as cross-source sentiment analysis.

After some discussions on the design and the implementation of the proposed DSS, some results are shown about the experimentation of the proposed DSS on two scenarios, namely restaurants and consumer electronics online shops. Specifically, first the application of effective sentiment analysis models, created relying on user reviews is discussed: the models achieve accuracies higher than 90%. Then, such models are embedded into the proposed DSS. Finally, the results of a social listening campaign are presented. The campaign was carried out by fusing data streams coming from real social channels of popular companies belonging to the selected scenarios.

Keywords: Decision Support Systems, Digital Marketing, Multiple Social Media Channels, Social Media Listening

1. Introduction

Thanks to the widespread diffusion of Internet connected devices, such as smart phones and tablets, the number of people sharing content online is constantly increasing [1]. The information shared ranges from visual data, such as videos, photos, and presentations, to textual one, like comments, posts, and articles, which are frequently uploaded and shared on social media, represented nowadays by a considerable number of platforms [2]. Indeed, people use social media not only for leisure purposes, but also for working, studying, discussing, marketing, and so on. For example, an increasing number of companies promote their products and services through social media marketing campaigns, using more than one public page for each social media considered. Indeed, several restaurants daily promote their menus and offers on different social media such as Facebook public pages and Instagram channels. In the specialized literature, we can find a number of contributions regarding how to address and evaluate marketing strategies on social media [3, 4, 5].

User opinions posted on social media are often about services and products of specific companies and brands. Thus, this huge amount of user viewpoints might be effectively mined and exploited as a powerful source of information for the enterprises business, in particular for steering marketing strategies according to what people really think about their products and services. Nevertheless, the main challenge of relying on contents available on social media platforms is to judiciously fuse huge amounts of opinions [6] often coming from different sources published on different platforms. Essentially, the problem is to summarize such opinions, neglecting noisy comments and focusing only on what really matters. Thus, more and more often, companies rely on *social media monitoring tools* [7]. These tools allow companies to keep track of the direct mentions of their brands on social media channels (including blogs, wikis, e-news, and social networking sites), as well as user-generated content in general. In this way, it is possible to determine, for instance, the volume and the sentiment of online conversations about specific brands or topics.

There exist several tools[1] for Social Media Monitoring, which allow users to track the content coming from social networks, as well as to perform sophisticated analysis on such content. Nevertheless, free-plans usually provide only limited functionalities, while pay-as-you-go-plans are normally quite expensive. Sometimes these tools also include a sentiment analysis module, but generally its technical specifications are not available. For this reason, in the majority of cases, the design of the sentiment model remains inaccessible and the accuracy of the sentiment analysis is unknown. In general, there is a lack of free-of-charge

*Corresponding author

Email addresses: `pietro.ducange@uniecampus.it` (Pietro Ducange), `m.fazzolari@iit.cnr.it` (Michela Fazzolari), `m.petrocchi@iit.cnr.it` (Marinella Petrocchi), `mvecchio@fbk.eu` (Massimo Vecchio)

[1]https://www.g2crowd.com/categories/social-media-monitoring
note: all URLs were last accessed on November 4, 2018.

social media monitoring tools, except for the ones offered by specific social media platforms themselves (i.e., Twitter Analytics and Facebook Analytics). In this case, however, no integration across different social media platforms is provided.

In addition, in recent years, a novel paradigm is emerging in the marketing area, called *social media listening* [8, 9, 10], that is a strategic activity allowing for the extraction of useful insights and valuable inputs from social media. In detail, social media listening exploits an ensemble of techniques to mine social data and perform quantitative, qualitative, and contextual analysis of topics and keywords. It tracks conversations about a specific topic (e.g., a hashtag or a brand) and combines them to discover business opportunities, marketing campaigns, and improve the overall customer experience. Besides the analysis of numerical data, social media listening should involve finer and more complex analysis which may include the elaboration of texts by means of Natural Language Processing (NLP) techniques for sentiment analysis [11, 12, 13], for assessing the polarity of the sentiment of a given text.

According to the literature [14, 15], three kinds of techniques can be used to automatically classify texts based on their sentiment, namely lexicon-based, machine learning-based and hybrid approaches. Lexicon-based approaches consist in calculating the polarity of a text by accounting the semantic polarity of words or phrases included in the document [16]. They require a dictionary of words annotated with sentiment-polarity, such as SentiWord-Net [17]. Lexicons are easily available[2], and lexicon-based techniques behave quite well in case of general boundless contexts (i.e., without topic), with well-formed and grammatically correct texts. Unfortunately, these algorithms loose their effectiveness when executed in bounded contexts or when an informal language is used. This is mainly due to the absence of context-related words when the text is related to a certain topic or when it is extracted from posts or comments of a social media platform. In these cases, machine learning-based approaches hold a significant role for opinion mining [18, 19, 20, 21], since the sentiment analysis can be formulated as a standard classification problem. These approaches rely on the construction of some text classification models, by means of a supervised learning process. Thus, a set of labeled training texts are necessary for learning the parameters of a sentiment classifier. However, to collect labeled text for the supervised learning stage of the sentiment classification models may be not a trivial task. Indeed, such models are domain-dependent: for each context, a set of domain-specific training examples must be retrieved, read, analyzed and finally labeled with the correct associated sentiment. It is clear that manual labeling can become very time consuming. Moreover, a sentiment classification model trained on a specific context (e.g., movies) cannot be directly applied to another context (e.g., electronic products) [22]. This is mainly due to the fact that features indicating positive (resp., negative) sentiments within a certain domain (usually called *source domain)* are usually different from those other features indicating positive (resp., negative) sentiments within another domain (usually called *target domain*).

In order to handle the problem of training accurate sentiment classification models under the assumption of scarce amount of labeled texts within a specific domain, the literature considers the option of fusing together information coming from different sources and differ-

---

[2]http://mpqa.cs.pitt.edu/lexicons/

ent domains. Recently, *sentiment domain adaptation* approaches across different domains have been widely studied [23, 24, 25]. These techniques allow to train text classification models within one (or more) source domains characterized by enough amounts of labeled data, and then to apply and test them within a target domain. Indeed, for some specific target domains, it is often too difficult to retrieve enough labeled texts to train domain-specific sentiment classification models.

Since the specific domain may be also defined by the language used in the text, in the last years, several contributions have been proposed in the framework of *cross-lingual sentiment analysis* [26, 27, 28]. In this framework, labeled texts available in one or more languages are used to train sentiment classification models for texts written in another language with scarce labeled text examples. Nevertheless, the majority of the aforementioned research contributions shows that best results are obtained when a specific model is built considering training texts derived from the same domain.

In this paper, a practical application of sentiment classification algorithms is presented, based on machine learning models, in the field of digital marketing and social media communication. In particular, the design and the development of a Decision Support System (DSS) for Social Media Listening are carried out, expressly conceived for companies wanting to exploit the valuable knowledge spread across social media to build effective, long-term strategies for their business. The proposed system can be used by commercial activities (e.g., restaurants, hotels, shops) to monitor user comments posted on their public social media pages. The core of the system is represented by a Sentiment Analysis Engine (SAE) applied over texts fetched from different sources of information, specifically, from a set of pages and channels of different social media platforms. The aim is to detect positive, negative, and neutral opinions about specific aspects of the monitored activities, which are then elaborated to send appropriate real-time notifications to the commercial activity's managers. Thus, the proposed DSS is useful for companies to evaluate their share of voice on social media, to assess the outcome of marketing campaigns, to be promptly alerted about unexpected turnaround and to appropriately react to negative events or trends.

For handling the scarcity of training text patterns and mitigating the problem of manually labeling, information coming from different social media sources are extracted and exploited. We do not consider domain adaptation nor multi-lingual sentiment analysis techniques. Indeed, the proposed sentiment classification models are still based on the construction of domain-specific classifiers, but already tagged opinions are exploited, namely user reviews and the associated scores, coming from online social media specialized on user reviews, all related with the same domain. In this way, a domain-specific model for each specific social media marketing scenario can be easily embedded in the system. Then, these models can be adopted for the real-time social media listening and used to classify unlabeled texts and comments retrieved from a set of social media channels different from the ones used in the training phase. This strategy has been labeled as *cross-source sentiment analysis*.

An extensive experimental campaign has been developed, in which two specific scenarios are considered, namely restaurants and consumer electronics online shops. The sentiment classification models are built over training sets, obtained by merging tagged reviews coming

from TripAdvisor[3] (for the restaurant scenario) and from Amazon[4] (for the consumer electronics online shop scenario). The trained sentiment classification models achieve accuracies higher than 90% in cross-validation, by using Support Vector Machine as the classification model.

Upon setting up the system with the trained classifiers, the results of a simulated social media listening campaign are also shown, considering three different social networks, namely Facebook[5], Twitter[6] and Instagram[7]. In particular, different social channels of real popular restaurants and consumer electronics online shops have been considered. More than 100,000 comments have been analyzed and more than 1,000 comments have been manually verified, finding out that the proposed system is able to correctly recognize the sentiment of more than 90% of them.

In conclusion, the main novelties and contributions introduced in this work are:

- The design and the implementation *from scratch* of a novel DSS for social media listening of promotional and marketing campaigns. The listening may be carried out even concurrently on different social media channels.

- The integration into the DSS of a *very accurate* sentiment analysis engine for identifying positive, negative or neutral opinions from users' comments.

- A novel method, that has been labeled as *cross-source sentiment analysis*, for generating context-based training datasets for the Sentiment Analysis Engine.

- An intensive experimental analysis, in which the proposed DSS is tested, considering *real-world* promotional or marketing campaigns, carried out by restaurants and consumer electronics online shops.

The paper is organized as follows. Section 2 describes the most recent related works. Section 3 gives details on the design and implementation of the proposed DSS. Section 4 introduces the cross-source sentiment analysis approach. Section 5 briefly describes the adopted methods for text representation and classification. Section 6 discusses the validation of the sentiment analysis models by showing the achieved classification results. Section 7 presents the on-the-field experimentation for the proposed DSS. Section 8 draws some conclusions.

2. Related Work

The specialized literature in the field of sentiment analysis discusses several approaches for both text representation and classification. Generally speaking, to classify a text it is necessary to transform it into a numerical vector, which is then handled by a classification

---

[3]www.tripadvisor.it

[4]www.amazon.it

[5]www.facebook.com

[6]www.twitter.com

[7]www.instagram.com

model. The *Bag-Of-Word* (BOW) representation is one of simplest and most used technique for text representation [29]: the dimension of the feature space is equal to the number of relevant terms (words, stems, n-grams) that can be extracted by the corpus of training documents or that characterize a specific context. A text may be represented by a binary vector whose elements specify either the presence or the absence of a relevant term within the text itself. Alternatively, real-numbered vectors can be used and, in this case, each component specifies the frequency of the term within the text [30, 31].

Moreover, *word embedding* methods are also deeply used for text representation: words in a vocabulary are transformed into vectors of continuous real numbers. Word2Vec [32], based on Neural Networks, and GloVe [33], based on matrix factorization, are among the most used approaches for learning the word embeddings. However, two main drawbacks can be highlighted when dealing with word embedding methods: first, a huge amount of training documents are needed for building accurate models for feature representation; second, even though some pre-trained model may be adopted (some of them are trained using the entire Wikipedia records), they may not be suitable for specific sentiment classification contexts, due to their generality. For these reasons, in this work a BOW representation has been adopted, with Term Frequency-Inverse Document Frequency (TF-IDF) index [30, 31] for extracting the value of each feature.

As regards classification algorithms, classical machine learning models are still deeply used for sentiment classification. In particular, *Support Vector Machines* (SVMs) have been recently adopted for several sentiment analysis applications, such as evaluating the quality of information in product reviews [34], assessing the human papillomavirus vaccination sentiment trend from tweets [35] and identifying sentiment polarity (positive, negative, and neutral) of tweets in several languages [36]. Also *Naive Bayes* (NB) classifiers are still successfully used, such as in [21], where the authors present an emotion recognition framework based on an ensemble of classifiers. Indeed, in [21], an NB classifier, a Maximum Entropy learner and a knowledge-based tool performing deep analysis of the natural language sentences are used together with a voting strategy for emotions classification. The authors show the application of their ensemble on sentences extracted from different sources, such as news headlines, news articles and Twitter posts. Results show that the NB classifiers, which adopt a BOW text representation, achieve the highest accuracy among the three elements of the ensemble. Quite obviously, the ensemble allows to achieve the highest accuracy.

It is worth mentioning that *deep learning-based* methodologies represent a fashion trend also within the sentiment analysis research community [37]. The majority of these approaches simply adopt a word embedding text representation followed by some convolutional or recurrent neural network layers [38, 39]. However, the necessity of prohibitive amounts of training data for the average deep learning-based algorithm to effectively learn accurate models represents the main drawback also in this context. For this reason, browsing the recent scientific literature it is not difficult to read about contributions still presenting text classification models based on BOW text representation and followed by classical machine learning classification algorithms [13, 40, 41].

Our contribution follows a similar approach. In fact, it is worth highlighting that the main goal of this paper is not to propose a novel sentiment analysis algorithm, therefore, a

comprehensive comparison campaign among the various techniques available in the literature is out of the scopes of this work. Indeed, the main objective of this paper is to present a real application of sentiment analysis algorithms based on state-of-the-art methodologies for text classification. Indeed, the main features and achievements of a DSS for social media listening are shown. The DSS embeds accurate models for sentiment classification. Moreover, a cross-source sentiment training procedure has been adopted for handling the scarcity of labeled training set. Notice that, the idea of cross-source training procedure may be adapted to specific context-based sentiment analysis procedures that need a supervised learning procedure. However, and quite remarkably, the achieved accuracies are in line with those summarized in [11, 21, 42], which are obtained using recent state-of-the-art sentiment analysis approaches.

As regards DSS, we recall that such systems have been widely studied since early 1970s [43]. As stated in the work in [44], DSS is the area of the information systems discipline that is focused on supporting and improving managerial decision-making. Thus, the contribution that we propose in this paper almost perfectly fits with the definition of DSS. The basic architecture of a DSS includes i) a data base, which contains information and knowledge extracted from a specific domain, ii) an intelligent engine, which exploits the knowledge contained in the data base to give advices on the decisions to be taken, and iii) a user interface, which should be as simple as possible in order to allow users to interact with the system in a easy way. In the recent literature, we can find DSSs designed and adopted for a wide range of contexts such as medical applications [45, 46], risk analysis for project management [47], financial frameworks [48, 49], and manufacturing maintenance [50]. Methods and algorithms of the computational intelligence [47, 48, 51] and data mining [45, 50, 52] fields are deeply adopted in the implementation of the intelligent engine of the most recent DSS.

3. System Design and Implementation

In this section, some details are given regarding the development cycle of the proposed Decision Support System (DSS) for Social Media Listening. First, the requirements analysis stage for the software platform that implements the DDS is briefly described. Second, the actors of the platform and the main functionalities associated to them are presented. Third, the UML diagrams of the classes that describe the main data structures adopted in the software are shown. Forth, the functional modules of the software platform are described. Finally, the architecture of the DSS prototype implementation is discussed.

3.1. Requirements Analysis

The initial stage of the DSS design consists in the identification of the main requirements that the software platform must satisfy.

The main purpose of the system is to monitor comments of users published on the public social media pages of commercial activities, such as restaurants, hotels, shops and malls, maintained either by the owners or by the marketing managers.

The system has to continuously performs the following operations:

(i) Real-time fetching of comments on posts published by a commercial activity on its public pages on social media.

(ii) Analysis of such comments, in order to extract the expressed sentiment, i.e., positive polarity, negative polarity or neutral comments.

(iii) Notification to the social activity owner (or manager) when negative comments is received and/or negative trends are recognized, in order to adequately reply.

(iv) Generation of reports and statistics regarding both i) the sentiment extracted from a specific share of voice on social media and ii) the overall reputation of a product and/or the brand of the company.

Moreover, the system has to allow the users to carry out basic functions such us i) to upload posts, comments, pictures and videos, ii) to download and save post and comments (setting appropriate search parameters), and iii) to analyze the trend of the number of likes or followers for a specific page or channel.

The DSS software platform has to manage several accounts on social networks, even belonging to the same social network. The pair (*account, social_network*) is referred to as *social channel*.

3.2. Actors Identification

During this stage, three kinds of actors have been identified, i.e. users that can interact with the DSS software. These three actors are: the *Customer*, the *Social Account Manager* and the *System Administrator*, each of them having different privileges. In the following the description of each actor and of the functionalities associated to them are given:

- The Customer is a person - or a delegate of a company- which can own several accounts on different social media, in particular on social networking platforms (in the following, these platforms are called social networks). The most relevant functionalities associated to a Customer are: to visualize the list of social channels, to add/remove a social channel, to visualize relevant data coming from the social network page (statistics, *likes*, comments, sentiment on comments, etc), to add/remove/save contents on a social channel, to analyze the content of a social channel. Each Customer is associated to a set of privileges, established by the Social Account Manager or by the System Administrator, that specify the operations permitted among the ones described above.

- The Social Account Manager (SAM) manages the activities of a set of Customers, within their respective set of social channels. Indeed, many companies outsource their marketing campaigns to external agencies and the SAM usually is a member of these agencies, interacting with the social networks on behalf of their Customers. SAMs perform all the operations available for Customers on their social channels. In addition, SAMs can manage the Customer user profiles and modify the privileges associated to each Customer.

- The System Administrator mainly acts as account manager and can access all the functionalies enabled for Customers and SAMs that have been previously described. This user can manage the profiles and interactions of all Customers and SAMs; moreover, the System Administrator can add/modify/remove users to/on/from the platform.

3.3. Class Diagrams

In Figure 1, the main class diagrams drawn after the requirements analysis and actors identification stages are shown. For the sake of simplicity and readability, the details of the attributes and the procedures for each specific class have been omitted. Rather, the main objective is to describe the relationships among classes by using associations. Thus, the UML 2 formalism is adopted, using bi-directional associations and inheritance lines [53].
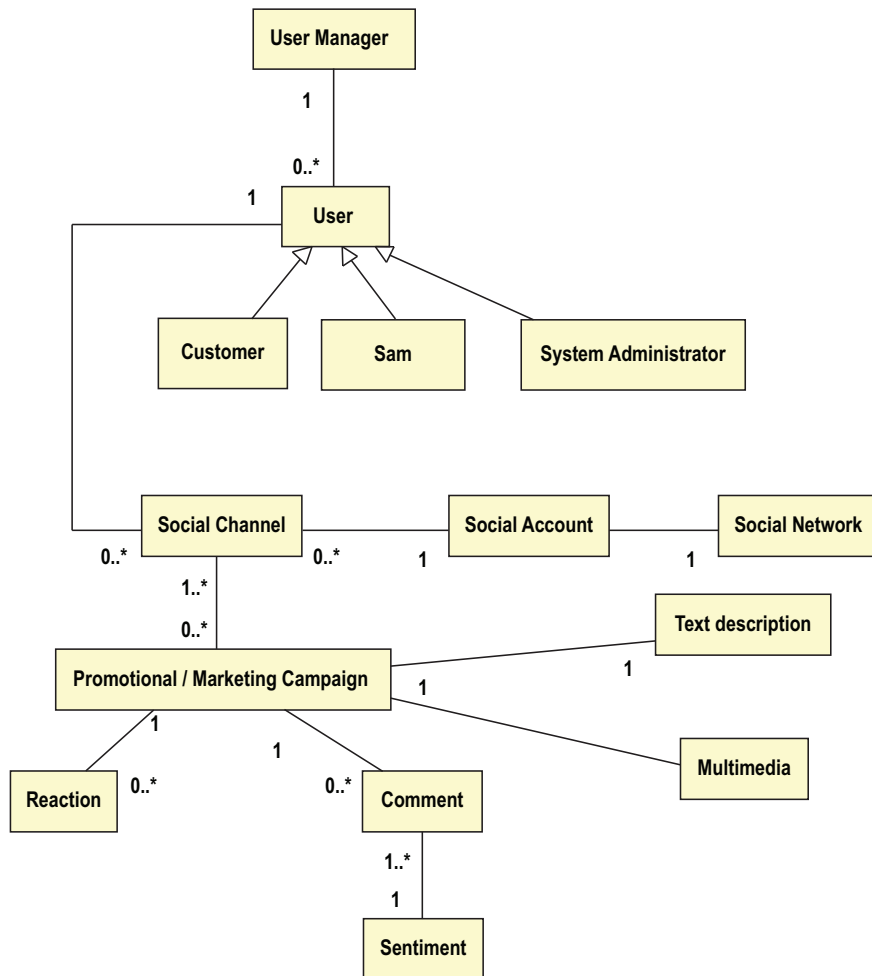
Figure 1: UML Diagrams of the Main Classes.

A *Users Manager* class has been identified, which is in charge of managing a collection of *Users*. The diagram highlights that the User class is an abstract class that can be specialized into the three identified actors, namely *Customer*, *SAM* and *System Administrator*.

The User class is associated with a collection of *Social Channels*. The Social Channel class is associated with a *Social Account* on a *Social Network*. Moreover, the Social Channel class is associated with a collection of *Promotional/Marketing Campaign*. Finally, a Promotional/Marketing Campaign class is associated with a *Text Description*, a collection of *Multimedia* (photos, videos, etc.), a collection of *Comments* and a collection of *Reactions* (likes, hate, angry, etc.). Each Comment is associated with a *Sentiment*.

The classes described above have been used as the basis for the main data structures adopted in the software implementation, especially for designing the data base schemes and tables.

3.4. Functional Modules of the DSS Software Platform

The last step of the DSS software platform design was to define the main functional modules. As shown in Figure 2, the platform is composed by three main functional modules, namely (i) the User Dashboard; (ii) the Social Media Manager; (iii) the Sentiment Analysis Engine. In the figure, the arrows connecting the functional modules represent the information flow inside the system, whereas the puzzle pieces identify the social plug-ins, introduced and discussed in the following, that allow the system to handle multiple social channels.

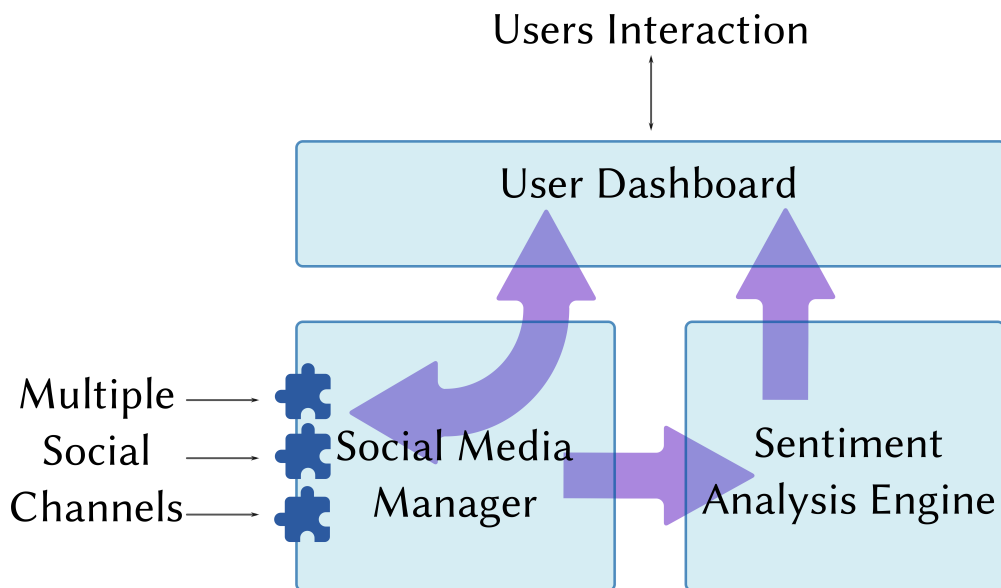In the remaining part of this section, the system modules are briefly described.

Figure 2: Functional Modules of the Social Media Listening system.

3.4.1. User Dashboard

The software design of the proposed DSS is tailored to deal with different kinds of users and to manage streams of data coming from different social media. The User Dashboard module allows users to interact with the system and to perform all the operations required

to manage the social media information and to produce a set of outputs such as simple checks of the comments and sentiment analysis report.

3.4.2. Social Media Manager

The Social Media Manager (SMM) module manages the interaction between the system and the social media channels. The different social networks are handled as plug-ins of the system. Basically, each *Social Plug-in* permits to perform the basic operations available on each social network, such as to add/modify/remove an account, to log in/log out, to publish any kind of content on a social network (post, pictures, links, *likes*, etc.), to retrieve content posted by other users on the managed social channels. Moreover, through this module, the user is able to describe the content to be monitored, according to one or more parameters (such as time interval, number of comments to retrieve and, when available, the geo-localization of the users). The SMM module also pre-processes the texts collected from different social networks, in order to unify their representation.

3.4.3. Sentiment Analysis Engine

Even though several functionalities are available in our system, the most important module, on which we mainly focus our attention in this paper, is the Sentiment Analysis Engine (SAE). This module is in charge of classifying texts and extracting their polarity. As input, it takes a string (or a stream of strings), which represents a text (or a set of texts). The strings are fetched from social channels. The module outputs an evaluation of the sentiment, associated to each string, in term of positive, negative or neutral polarity.

3.5. Implementation Details

The proposed DSS software platform has been actually implemented as a Web Application Framework. In particular, as shown in Figure 3, the platform prototype was developed according to the multi-tier architecture [54]. From a logical point of view, we tried to keep the Client-side and Server-side layers as separate as possible. The server side layer has been further subdivided into sub-layers: Front-End (which contains the jsp pages), Middleware (which contains all the modules for the business logic and for interactions with the database) and Back-End (the database). This kind of architecture has been chosen in order to make the web application as scalable as possible, with the aim of simplifying any future changes. The Client Side was developed using HTML and Javascript. The Middleware was mainly developed adopting Java Enterprise Edition as programming framework [55]. For the database, MySQL[8] has been adopted as management system.

It is worth mentioning that the current version of the DSS software is able to interact with three widespread social networks, namely Facebook, Twitter and Instagram. Thus, three different Social Plug-ins have been implemented. The integration with Twitter is performed by exploiting Twitter4J[9], an unofficial Java library, which is used to interact with Twitter API. Similarly, the interfacing with Facebook is realized with an unofficial Java library called
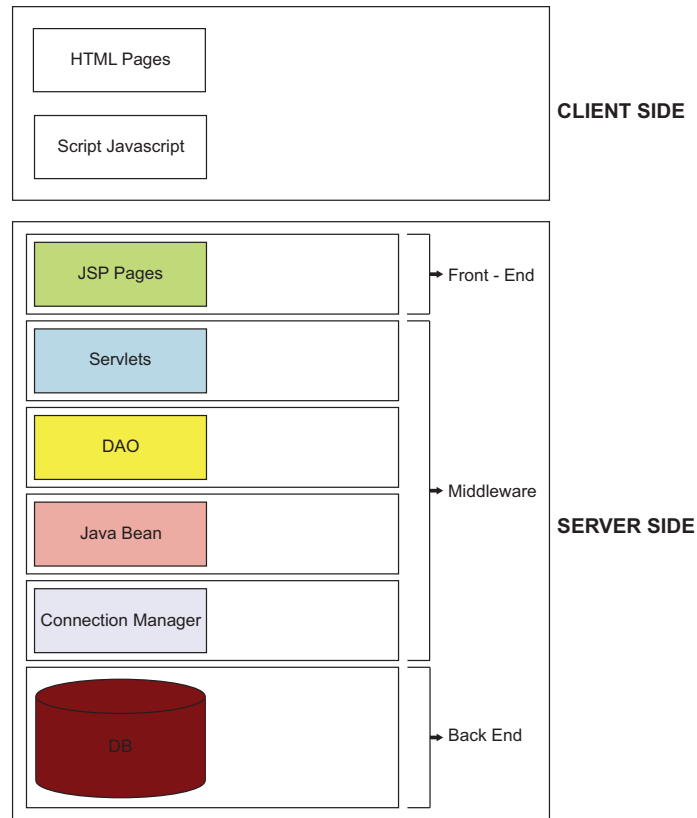
---

[8]https://www.mysql.com
[9]http://twitter4j.org/en/index.html

Figure 3: Multi-tier Architecture Implementation of the DSS Software Prototype.

Facebook4J[10], which wraps Facebook API in Java. Finally, the interaction with Instagram is carried out by means of Instagram Realtime API[11].

As regards the texts extracted from the different social networks, all the meta-information associated to a text (such as user IDs, timestamps, geographic coordinates and special characters) are identified in the text by applying a Regular Expression filter [56]. The most relevant meta-information are extracted and organized in an appropriate data structure giving a uniform representation of posts and comments in our system. This data structure includes a text represented as a simple string, where a case-folding operation has been applied, and set of features which mainly includes user ID, timestamps, number of hashtags/mentions/followers/likes, links, list of emoticons. Since the comments extracted from different social networks may contain or not some meta-information, some fields of the aforementioned data structure may be unspecified.

Finally, the text classification modules have been implemented using the Java APIs for Weka (Waikato Environment for Knowledge Analysis) [57], which include both the codes for the text elaboration and for machine learning-based classification models. Obviously, these codes have been appropriately integrated into the Middleware layer of our software.

---

[10]http://facebook4j.github.io/en/index.html
[11]https://github.com/sachin-handiekar/jInstagram

4. Cross-Source Sentiment Analysis

The SAE module of the DSS represents the core of the proposed system. In order to identify the sentiment expressed in the text, two main steps have been carried out: i) to represent the text in a numerical vector space and ii) to classify the text using a specific model. In the proposed DSS, a state-of-the-art approach is adopted, based on Bag-Of-Words (BOW) text representation and machine learning classification models, similar to the ones described in [21, 58].

When dealing with machine learning classifiers, a set of labeled patterns is needed for training the models to be used in the classification of unseen patterns. Specifically, we need a set of labeled text with the sentiment that we want to estimate. Since the sentiment classification model is domain-dependent, a specific model for each context needs to be constructed, which is a meta-environment that unifies commercial activities with similar characteristics. For example, different kinds of restaurants can be gathered into a single scenario, since the related comments usually consider the quality of food, the attitude of the staff, the appeal of the location, etc. On the contrary, if focusing on another kind of business, such as, e.g., an online shop of electronic devices, the related comments usually deal with the quality of different characteristics, such as the packaging, the delivery, the customer support, etc. Therefore, a specific classification model is suitable only for a certain scenario and, when the scenario changes, a new context-aware model needs to be learned.

To overcome the issue of labelling a specific set of texts for each model to be trained, a simple and effective method is adopted, to automatically obtain labeled set of texts to be used in the learning stage. We move from the intuition that a scenario is characterized by a set of keywords used in that domain, therefore we look for already labeled texts about a certain domain. To this aim, the huge amount of available online user reviews has been exploited. In fact, online user reviews usually come with a text, which is the expression of the domain, and a score, which can be used for the class labelling. For each considered scenario, we search online reviews suitable for the specific scenario, i.e., reviews with a domain similar to the one of the scenario considered, and we use them as a source to obtain labeled data. For example, for the restaurant scenario, it is possible to use online reviews coming from Yelp, TripAdvisor and Google. Similarly, for the consumer electronics online shop scenario suitable reviews can be collected from Amazon (worldwide), BestBuy (USA), MediaMarkt and Euronics (Europe), and so on.

In Figure 4, the flow of the adopted strategy in our application is shown, namely the *Cross-source Sentiment Analysis*.

Our analyses focus on Italian texts, thus the labeled dataset is constructed starting from reviews collected from Amazon and TripAdvisor, widely used in Italy and offering a huge amount of reviews. The details of the method used to build each labeled dataset are described in Section 6.1. In the next section, the approach adopted for the text representation and classification is described in detail.
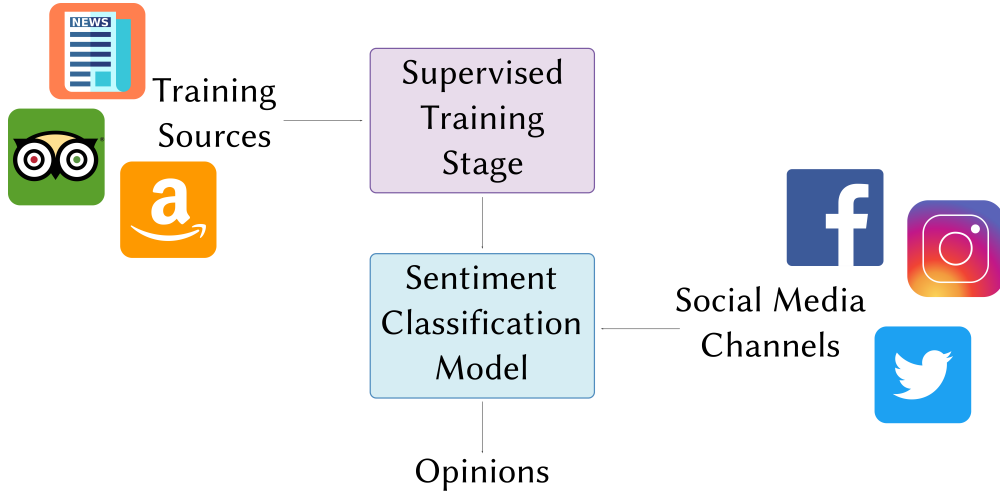
Figure 4: Scheme of the Cross-source Sentiment Analysis.

5. Text Representation and Classification

In the following, the main steps that characterize the extraction of a specific sentiment from a text are described.

5.1. Text Representation

The string representation of a text needs to be transformed into a numeric vector, before being elaborated by a classification model. Thus, the following set of text elaboration procedures are applied to the j-th text $text_j$ handled by the SAE:

1) *tokenization*: the first step of a process of text mining usually consists in transforming a string into single processing units called *tokens*, where each token can be represented by a syllable, a word or a phrase. During this step, other operations are also performed: the punctuation and other non-text characters are discarded [59], and symbols are normalized. The adopted tokenizer removes all the punctuation marks and uses a BOW representation, i.e., each text is split into words. Thus, at the end of this step, each j-th tokenized text is represented as the set of words therein contained: $text_j^T = \{t_{j1}^T, \ldots, t_{jh}^T, \ldots, t_{jH_j}^T\}$, where t_{jh}^T is the h-th token, and H_j is the number of tokens in $text_j^T$.

2) *stop word filtering*: during this step, stop-words are removed, i.e., words that provide little or no information to the text analysis, such as prepositions, conjunctions, articles, etc. Additional stop-words are words that appear very often in the sentences of the considered language (language-specific stop-words) or in the set of the texts analyzed (domain-specific stop-words). These words have no statistical significance and can be considered as noise [60]. In the proposed system, the list of Italian stop-words was downloaded from the freely available Snowball Tartarus website[12]. At the end of this

---

[12]http://snowball.tartarus.org/algorithms/italian/stop.txt

step, each text is represented as a set of relevant tokens, cleaned from stop-words: $text_j^{SW} = \{t_{j1}^{SW}, \ldots, t_{jk}^{SW}, \ldots, t_{jK_j}^{SW}\}$, where t_{jk}^{SW} is the k-th token, and $K_j, K_j \leq H_j$, is the number of relevant tokens in $text_j^{SW}$.

3) *stemming*: this step consists in reducing each token to its stem - or root - form, by removing the suffix. The aim is to gather words that have closely related semantics. The stemmer included in our tool uses the Snowball Tartarus Stemmer[13] for the Italian language, based on the Porter's algorithm [61]. At the end of this step, each text is represented as a set of stems: $text_j^S = \{t_{j1}^S, \ldots, t_{jl}^S, \ldots, t_{jL_j}^S\}$, where t_{jl}^S is the l-th stem, and $L_j, L_j \leq K_j$, is the number of stems in $text_j^S$.

4) *stem filtering*: during this step, the number of stems representing a text is reduced. In fact, each text is filtered and some stems are removed, according to the set RS of relevant stems identified during the supervised learning stage that will be described in Section 5.3. At the end of this step, the j-th text is represented as a subset of the relevant stems: $text_j^{SF} = \{t_{j1}^{SF}, \ldots, t_{jp}^{SF}, \ldots, t_{jP_j}^{SF}\}$, where t_{jp}^{SF} is the p-th relevant stem, and $P_j, P_j \leq L_j$, is the number of relevant stems in $text_j^{SF}$. The total number F of relevant stems is identified during the supervised learning stage by considering the whole set of labeled training texts.

5) *feature representation*: to apply a classification algorithm to texts, they have to be represented in a common feature space. Thus, this step consists in building a vector of numeric features for each text. The set of F features corresponds to the set $RS = \{\hat{s}_1, \ldots, \hat{s}_f, \ldots, \hat{s}_F\}$ of relevant stems. In particular, a feature for each relevant stem is considered, therefore, the j-th text is described by a vector $text_j^{FR} = \mathbf{x}_j = \{x_{j1}, \ldots, x_{jf}, \ldots, x_{jF}\}$, where each element x_{jf} is defined as follows:

$$x_{jf} = \begin{cases} TF_{jf} \cdot w_f & \text{if } text_j^{SF} \text{ contains the corresponding relevant stem } \hat{s}_f \\ 0 & otherwise. \end{cases} \tag{1}$$

In the above equation, TF_{jf} is the term frequency of $\hat{s}_f$ in the j-th text, whereas the weight w_f identifies the importance in the training dataset of the f-th feature, that is the f-th relevant stem, and it is computed during the supervised learning stage, which will be described in Section 5.3.

5.2. Text Classification

Once a text has been elaborated, i.e., transformed into a vector of numbers, it is tagged with one of the three class labels related to the expressed sentiment: positive or negative if the sentiment itself is positive or negative, respectively, neutral if the text contains something completely unrelated with the topic. Thus, if the SAE receives as an input a stream of comments, the generated output is a collection of classified texts. The classified texts may

---

[13]http://snowball.tartarus.org/algorithms/italian/stemmer.html

be also used in the reports that the system is able to generate and that may also include statistics, such as the number and the trend of the likes and/or followers for a specific post or marketing campaign.

As stated before, in order to obtain accurate and robust results, a machine learning approach is applied, by using a sentiment classification model, whose parameters are identified during a supervised learning stage described in the following section. This work focuses on Italian texts and, for each considered scenario, a set of classification models has been trained and compared to identify the most suitable one. The classifier with the best performance is finally embedded in the SAE and adopted to classify any text which can be extracted from the different social channels, handled by the SMM module.

5.3. Sentiment Analysis Engine Learning Stage

This section gives the details of the supervised learning process needed to perform the last two steps of the text elaboration described in Section 5.1 (namely the *stem filtering* and *feature representation*), and the text classification. We recall that for each domain, i.e., for each specific scenario, this stage must be carried out.

The supervised learning stage implies the following steps:

(i) The first step consists in identifying the relevant stems to be used in the features representation;

(ii) The second step consists in identifying the structural parameters of the adopted classification model.

5.3.1. Relevant Stems Identification

A set of N_{tr} labeled texts is used as training set, which have been obtained from different social media that contains labeled reviews for a specific domain. These texts are pre-processed by applying the step 1-3 described in Section 5.1. Then, the complete set of stems $CS_{tr} = \{s_1, \ldots, s_q, \ldots, s_Q\} = \bigcup_{t=1}^{N_{tr}} text_t^S$ is obtained by the union of all the Q stems extracted from the N_{tr} training texts, after the stemming step.

The importance of each stem s_q in CS_{tr} is represented by a weight w_q computed as the Inverse Document Frequency (IDF) index [62]:

$$IDF_q = ln\frac{N_{tr}}{N_q} \tag{2}$$

where N_q is the number of texts that contains the stem s_q. Therefore, each training stemmed text $text_t^S$ is represented as a vector of features, namely $\mathbf{x}_t = \{x_{t1}, \ldots, x_{tq}, \ldots, x_{tQ}\}$, where x_{tq} is defined as follows:

$$x_{tq} = \begin{cases} TF_{tq} \cdot w_q & \text{if } text_t^S \text{ contains the stem } s_q \\ 0 & otherwise. \end{cases} \tag{3}$$

where TF_{tq} is the Term Frequency of q-th stem in the t-th text of the training set. Since the weight has been computed as the IDF index, the final value of each element in the feature

vector is represented as the well-known Term Frequency-Inverse Document Frequency (TF-IDF) index [30, 31].

Then, a feature selection algorithm is applied, in order to select only a reduced set of F, $F \leq Q$, relevant stems. To evaluate the quality of each stem s_q, the Information Gain (IG) [30] value between a feature S_q (corresponding to stem s_q) and the possible output class labels in C is computed as follows:

$$IG(C, S_q) = H(C) - H(C|S_q) \qquad (4)$$

$H(C)$ is the entropy of C and $H(C|S_q)$ is the entropy of C after the observation of the feature S_q. The final set RS of F relevant stems is identified by selecting those stems presenting a positive value for IG.

5.3.2. Identification of the Sentiment Classifier Parameters

The final part of the supervised learning stage involves the application of a training algorithm for identifying the parameters of a classification model. As stated before, in order to identify the most suitable classifier, for each considered scenario, several classification algorithms commonly used in the literature have been experimented. In particular, the results achieved by four algorithms have been compared, i.e. i) Support Vector Machine (SVM) [63], ii) Naive Bayes (NB) [64], iii) C4.5 decision tree [65], and iv) k-Nearest Neighbor (k-NN) [66]. In Section 6, the aforementioned classifiers will be briefly described and the achieved results will be shown and discussed.

6. Sentiment Analysis Engine Setup

This section describes how the Sentiment Analysis Engine has been actually built for each of the two chosen scenarios. First, the experimental setup is described: in particular, the details of the adopted labeled training sets are given. Then, the classification results achieved by a set of different models are discussed and compared. Finally, for each scenario, a model is selected to be embedded into our Social Media Listening System for the on-the-field experimentation discussed in Section 7.

6.1. Experimental Setup: Details on the Labeled Training Sets

In order to train the classifiers, a distinct training set is needed for each considered scenario. As already introduced in Section 3, the information exploited is collected from sources which are different than the ones used for on-the-field experiments for social media listening. In particular, the labeled texts are extracted from social media and websites which contain user reviews on products and services. For the sake of brevity, we select only two scenarios, but the technique can be easily extended to additional scenarios.

The first scenario deals with restaurants. The training set referred to this scenario is built by considering both positive and negative online reviews coming from TripAdvisor. To this aim, we consider several restaurants in Italy and collect 1,000 positive comments, i.e., reviews associated with a score value equal to 5, and 1,000 negative comments, i.e., reviews associated with a score value equal to 1.

The second scenario is about consumer electronics online shops. In this case, we consider online reviews from Amazon, related to electronic devices. To populate the training set, we extract 1,000 positive comments, i.e., reviews with an evaluation equal to 5 stars, and 1,000 negative comments, i.e., reviews associated with an evaluation equal to 1 star.

It is worth noting that online reviews can present inconsistencies between the sentiment evaluation of the review texts and the correspondent user ratings, as pointed out in [67] and [68]. Thus, in order to achieve a high probability that the text of the reviews is coherent with the evaluation label, for each scenario, only the reviews at the extremes of the evaluation scale are selected.

For each scenario, a set of classifiers are trained and compared, each of them carrying out a 3-class classification task, tagging positive, negative and neutral opinions. Regarding neutral opinions, in the training set of both scenarios we rely on the same set of 1,000 texts, randomly extracted from different web sources, including, for example, political discussions, newspaper news and reviews on topics far from the selected scenarios. To ensure the absence of biases, the selected texts have been manually inspected for the neutral class and we assessed that they were associated to neither positive nor negative opinions. All the labeled texts, included into the training sets for both scenarios, have been extracted from reviews written along November 2017.

6.2. Comparisons of the Different Classifiers

For each scenario, four different classification models have been built and compared, namely SVM, NB, C4.5 and k-NN (with k equal to 1). In this section, the classification results achieved by applying these classifiers to the datasets described in Section 6.1 are discussed. For each classifier, the experiments are performed by using a n-fold cross-validation methodology, which consists of randomly partitioning the dataset into n folds, where the classes in each fold are represented with the same proportion as in the original data. The classification model is then trained on $n-1$ folds, while the remaining fold is used for testing purposes. This process is repeated n times, using each of the n folds only once for testing. The n results obtained by each execution are finally averaged to obtain an overall estimation of performances. In our experiments, we set $n = 5$ for both the scenarios.

To assess the performance of each classification model, well known metrics are used, namely precision, accuracy, recall, and F-score [69]. Table 1 describes the adopted metrics. For the sake of simplicity, in the table, we refer to the case of binary classification, in which there are only two classes, the positive and the negative ones. Nevertheless, the methods for assessing the performances can be easily extended to the case of multi-classification: the metrics are computed per class, considering the patterns of the specific class as positive and the remaining patterns as negative. Indeed, the aforementioned metrics are computed in terms of True Positives (TP), that is the number of actual positive samples correctly classified as positive; True Negatives (TN), i.e., the number of actual negative samples correctly classified as negative; False Positives (FP), i.e., the number of actual negative samples that are incorrectly classified as positive; False Negatives (FN), that is the number of actual positive samples incorrectly classified as negative.

Specifically, *Accuracy* represents the overall effectiveness of the classifier and corresponds to the number of correctly classified samples over the total number of samples. *Precision* is the number of correctly classified samples of a class, i.e., positive class, over the number of samples classified as belonging to that class. Precision represents how many texts, classified with a specific sentiment, are relevant. *Recall* is the number of correctly classified samples of a class, i.e., positive class, over the number of samples of that class; it represents the effectiveness of the classifier to identify positive samples (*Sensitivity*). Precision and Recall allow us to understand and measure the relevance of a specific sentiment classification model towards positive, negative and neutral sentiments. The *F-score* is the weighted harmonic mean of precision and recall.

Table 1: Metrics definition

Metric	Equation
Accuracy	$Acc = \frac{TP+TN}{TP+FP+FN+TN}$
Precision	$Prec = \frac{TP}{TP+FP}$
Recall	$Rec = \frac{TP}{TP+FN}$
F-Score	$Fscore = 2 \cdot \frac{Prec \cdot Rec}{Prec+Rec}$

Tables 2 and 3 show the results achieved by the different classification models on the restaurant and the consumer electronics online shop scenarios, respectively. More in detail, each table shows, for each classifier, the accuracy and the per-class value of recall, precision and F-score. All the values are averaged over the 5 values obtained by applying the cross validation methodology. Figures 5 and 6 summarize the results shown in Tables 2 and 3, respectively, as bar charts.

For the restaurant scenario, the best performances are achieved by the SVM, with an average accuracy of 91.95%. Also for the second scenario, the SVM is the classifier which shows the best results and in particular its average accuracy is equal to 93.01%. It is worth noticing that, for both scenarios, the precision and the recall associated with the neutral class is higher than 95%. This means the SVM is potentially able to discriminate very well between comments which actually express an interesting opinion (positive or negative) and neutral comments.

As stated in the Section 2, the achieved results are in line with those summarized in recent contributions, such as the works discussed in [11, 21, 42] and obtained using state-of-the-art sentiment analysis approaches.

According to the results shown in Tables 2 and 3, the SVM models are chosen to be embedded in the final implementation of the Social Media Listening system, specifically in the SAE. The following section reports the results of a simulated monitoring campaign of different social channels, using the proposed system.
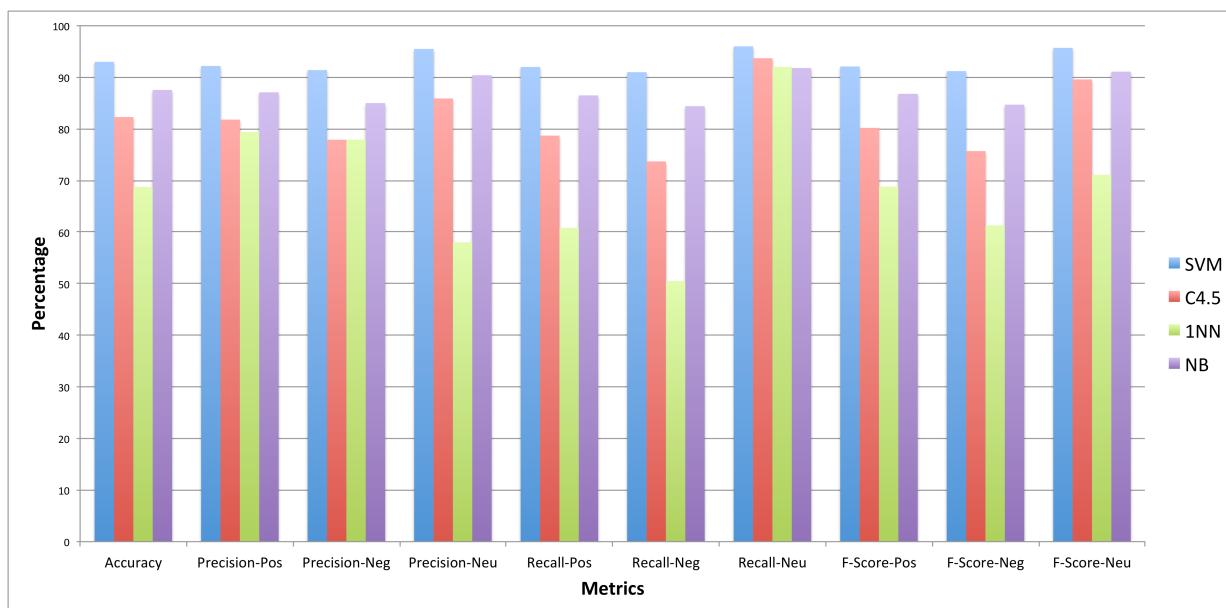
Figure 5: Restaurant scenario: bar chart of the average classification results.
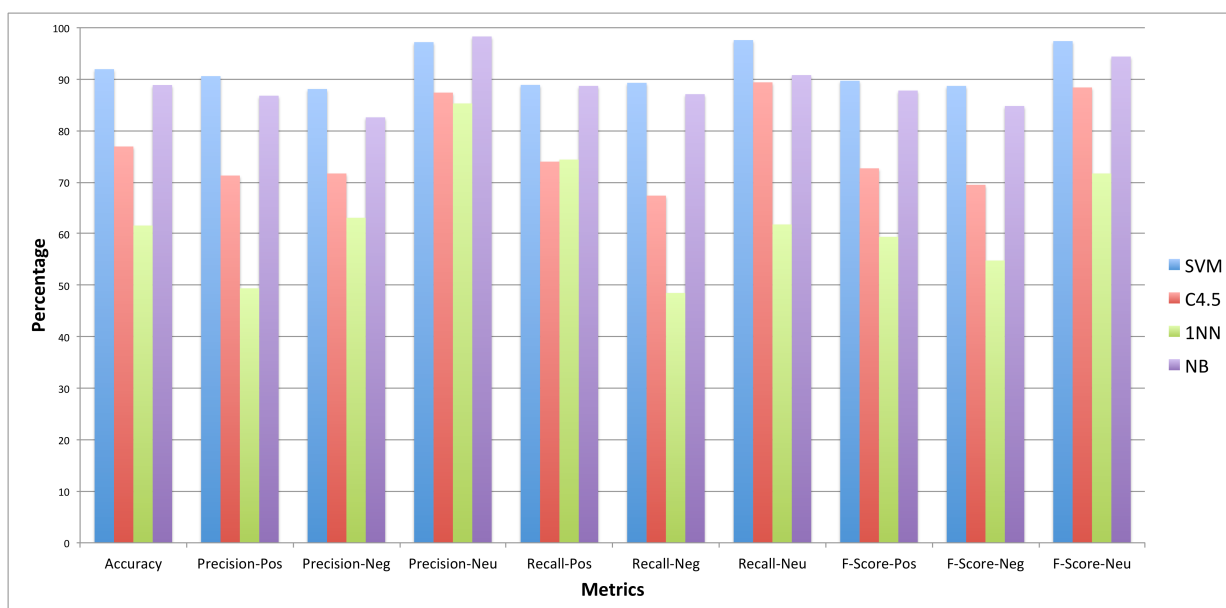

Figure 6: Electronics online shop scenario: bar chart of the average classification results.

7. On-the-field Social Media Listening

In order to test the proposed system, a simulated social media listening campaign has been performed. The term *simulated campaign* is used, since it is not possible to directly interact with the social channels of a specific commercial activity. Indeed, to this aim, we should have been the administrators of the public pages or channels of the commercial

Table 2: Restaurant scenario: average classification results.

Classifier	Accuracy (%)	Precision(%) by class			Recall(%) by class			F-Score(%) by class		
		Pos	Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu
SVM	**91.95**	**90.6**	**88.1**	97.2	**88.9**	**89.3**	**97.6**	**89.7**	**88.7**	**97.4**
C4.5	76.94	71.3	71.7	87.4	74	67.4	89.4	72.7	69.5	88.4
1NN	61.59	49.4	63.1	85.3	74.4	48.5	61.8	59.4	54.8	71.7
NB	88.87	86.8	82.6	**98.3**	88.7	87.1	90.8	87.8	84.8	94.4

Table 3: Consumer electronics online shop scenario: average classification results.

Classifier	Accuracy (%)	Precision(%) by class			Recall(%) by class			F-Score(%) by class		
		Pos	Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu
SVM	**93.01**	**92.2**	**91.4**	**95.5**	**92**	**91**	96	**92.1**	**91.2**	**95.7**
C4.5	82.3	81.8	77.9	85.9	78.7	73.7	93.7	80.2	75.7	89.6
1NN	68.74	79.4	77.9	58	60.8	50.5	92	68.8	61.3	71.1
NB	87.55	87.1	85	90.4	86.5	84.4	91.8	86.8	84.7	91.1

activity. Thus, as a first attempt, we created some fake pages on Facebook, Twitter and Instagram, arranging marketing campaigns and/or promotions. Even though we asked students and colleagues to collaborate with us and write some comments, the final results were not interesting at all. However, these fake pages were useful to test the main functionalities of the DSS, such as to write promotional posts and to define marketing strategies, even concurrently on different social channels, and to answer to users' comments.

As a second attempt, another approach has been followed, by considering public pages and channels of popular commercial activities, falling into the framework of the two chosen scenarios. We identified a set of marketing and promotional posts, in a time span between December 2017 and January 2018. The experiments were conducted as follows:

- Our system connected to the public and open social channels of the chosen commercial activities;

- A number of selected posts, regarding promotion or marketing campaigns of real commercial activities, and the associated comments were dowloaded;

- The SAE classified the streams of downloaded comments and a report for each campaign was generated.

In order to classify the downloaded comments, the SVM classification models are used for both scenarios. For the restaurant scenario, our tool must be able to classify comments related to the quality of different aspects: the food, the service, the location, and so on. As well, when considering the second scenario, the system must be able to classify comments related to the quality of the shipping, the characteristics of a product, the integrity of the packaging, the functions of a specific device, and so on.

(a) Restaurant scenario	(b) Online shop scenario

Figure 7: The stem-clouds of the first 40 relevant stems.

For the system setup, as training sets the whole datasets described in Section 6.1 are used. During the learning stage, for the restaurant scenario $Q = 5545$ features are identified, which are reduced to $F = 970$ features after the feature selection step; for the online shop scenario, the initial number of $Q = 5953$ features was reduced to $F = 1178$. In Figure 7, two stem-clouds are shown, related to the first and second scenario, respectively, which give a flavour of the stems selected by the feature selection process (in the stem-clouds, only the first 40 selected stems are depicted).

In Table 4 and Table 5, some examples are shown of the sentiment classification made by our system during the simulated social media listening campaign. Social channels on Twitter, Facebook and Instagram of both restaurants and consumer electronics online shops are considered. For privacy issues, the actual names of the chosen commercial activities are not mentioned. The tables, first report the marketing or promotional post, which is usually concurrently shared on different social channels and, then, the list of the associated comments is shown. For each comment, we specify the social channel from which it was downloaded and both the actual and estimated sentiment polarities. For the sake of simplicity and readability, both posts and comments are translated in English, avoiding to show the original Italian texts.

As shown in Table 4 and Table 5, where the misclassified texts are highlighted in bold, the system was able to correctly classify most of the selected comments. We notice that, sometimes, the SAE cannot correctly recognize the sentiment due to sarcasm (such as in the comments "*Really? You got the nerve to mention Health? Double burger with cheese paves the way for colon cancer!*" and "*Customer service is a joke*") or to too long and articulated sentences with ironic statements (such as "*Camera stocked eh? Funny I ordered a 1400 camera five days ago. I still can't pick it up. I was supposed to be able to get it today, now they say hopefully Wednesday. When I called the customer service to complain and they told me that my options were to cancel the order or to wait…great customer service and satisfaction!!!*" and "*I bought some extra chicken nuggets yesterday and I decided to eat them for lunch today. I am so glad I did not pop the entire nugget into my mouth because I found a BONE! I could have chipped my tooth or chocked on it!!*").

22

Table 4: Restaurant scenario: some examples of classified comments.

Post
Regular Margherita or regular Capricciosa? Or half and half if you can't make up your mind!

Comments	Source	Est.Pol.	Actual Pol.
If in doubt, have both. Damn, we need pizza like this in Milan!	Facebook	**Positive**	**Neutral**
Ahhhh delicious	Twitter	Positive	Positive
Hey! Are you open Mondays? Rob	Instagram	Neutral	Neutral
Bleaaah!Avoid such a Pizza and that awful pizzeria	Facebook	Negative	Negative
Do you have any gluten free options?	Facebook	Neutral	Neutral
I should venture out to try this... Usually your Pizza is magnificent!	Facebook	Positive	Positive
Hi, do you need a boy?	Facebook	Neutral	Neutral

Post
BIG NEWS First sighting of our new menu. Will feature some new proposals and some classics like our delicious dumplings

Comments	Source	Est.Pol.	Actual Pol.
When some tapas?	Facebook	Neutral	Neutral
Another one plate for me! I went there yesterday and I felt in love with the four cheeses dumplings	Facebook	Positive	Positive
Dumplings!!! #icantwait	Facebook	Positive	Positive
So excited!!! Your dumplings are delicious!	Facebook	Positive	Positive
I can have some problems in reaching you! Hey guy, are you delivering near the stadium??	Facebook	**Negative**	**Neutral**

Post
New entries in our Organic Menu: Double Veggy Burger and Spicy Nuggets. An healthy option for your meal!

Comments	Source	Est.Pol.	Actual Pol.
Really? You got the nerve to mention Health? Double burger with cheese paves the way for colon cancer!	Facebook	**Neutral**	**Negative**
You should close down! You have the worst burger I have ever seen. Every time my family& and I go there, they make us sick.	Twitter	Negative	Negative
I love your burgers. I never really had problems.	Facebook	Positive	Positive
I bought some extra chicken nuggets yesterday and I decided to eat them for lunch today. I am so glad I did not pop the entire nugget into my mouth because I found a BONE! I could have chipped my tooth or chocked on it!!!	Facebook	**Neutral**	**Negative**
Actually really good!	Facebook	Positive	Positive
Put some bacon on and it will be great! The best burger in the town:)	Twitter	Positive	Positive

Table 5: Online shop scenario: some examples of classified comments.

Post

The new s80 is released today on our site! €399.99, 32 GB Dual Camera 60 Mp, 6 in Display, Pearl White

Comments	Source	Est.Pol.	Actual Pol.
I am looking forward to check your promotions! I always follow your amazing promotional posts!	f	Positive	Positive
Are you planning any promotions also for IPHONES?	twitter	Neutral	Neutral
I bought an S7 and they send me it after 3 weeks! Very bad service...avoid this shop	instagram	Negative	Negative
I hope to be on time to get sales also from your past great promotions	f	Positive	Positive
I was wondering if I can buy your items in a shop in Rome	f	Neutral	Neutral
High prices, avoid if you can...	f	Negative	Negative
Even though I don't like Samsung phones, I found great promotions and good shipping service	f	**Negative**	**Positive**
I got a great deal on a S6 phone	f	Positive	Positive

Post

When you're camera-ready, we're camera stocked. Make the most of your memories with our cameras selection.

Comments	Source	Est.Pol.	Actual Pol.
I'm interested in the a6300 camera. I would want a bundle with 2 lenses and a bag, if possible.	f	Neutral	Neutral
They have the worst customer service home delivery ever. I will never purchase anything from them again.	f	Negative	Negative
Camera stocked eh? Funny I ordered a 1400 camera five days ago. I still can't pick it up. I was supposed to be able to get it today, now they say hopefully Wednesday. When I called the customer service to complain and they told me that my options were to cancel the order or to wait...great customer service and satisfaction!!!	f	**Positive**	**Negative**
Very good service! Thanks!	f	Positive	Positive
Customer service is a joke	f	**Neutral**	**Negative**

Post

Make your gaming experience and movie times look fantastic with curved screens and HDR Ready technology on selected monitors!

Comments	Source	Est.Pol.	Actual Pol.
Why did you guys ended the Black Friday deals so early?	f	Neutral	Neutral
What model is that in the picture?	twitter	Neutral	Neutral
Proud owner of XXX! It's amazing!	f	Positive	Positive
This is my favorite electronic store!!	f	Positive	Positive
Why don't just design/create a widescreen 50" or 60" monitors with built-in computer+tv capabilities that got a hotswap able hdd/ssd?	instagram	Neutral	Neutral
These are great for gaming and movie streaming!	twitter	Positive	Positive

Figure 8 reports an example of report that can be generated using the proposed DSS. It is a weekly trend of the sentiment generated by a specific promotion campaign regarding a 2x1 offer for pizzas in a restaurant. In particular, the graph shows the total number of comments, the positive, the negative and the neutral comments, from the first day till the seventh day after the publication of the campaign on the social media, considering all the available channels. In the example, after two days, the campaign seems to have a positive response. On the other hand, after the third day, the negative comments increase while the positive comments decrease. After the fourth day, the number of total comments decreases but the negative comments are always higher than the positive comments. By reading the comments, it is possible to realize that people, attracted by the promotion started on Wednesday, went to the restaurant during the weekend to try the 2x1 pizza offer: users that commented on social media stated that they found very a low food quality and a very bad service due to the too crowded restaurant.
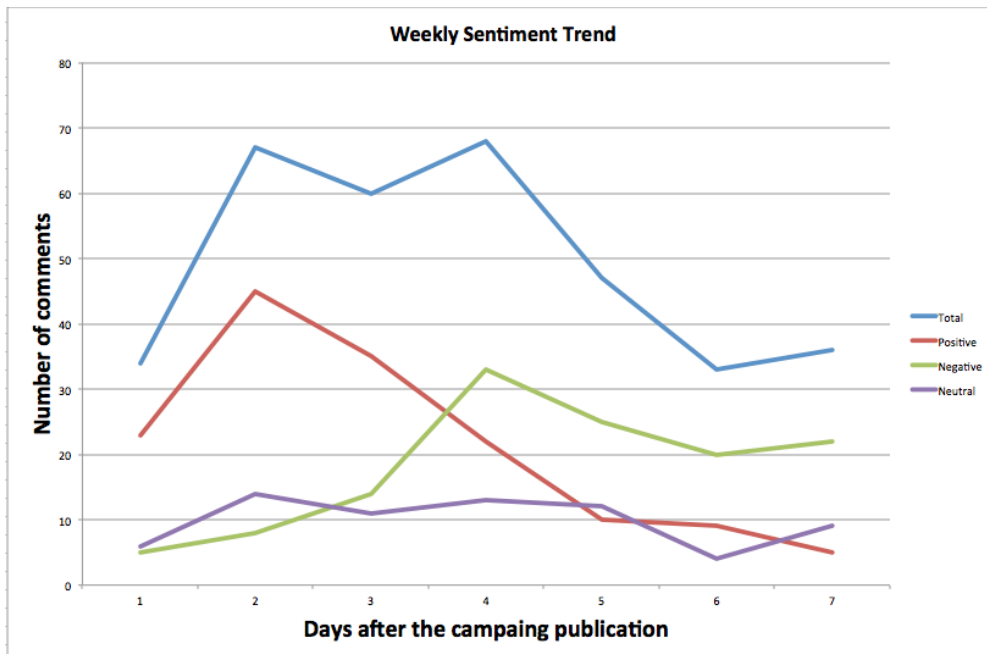

Figure 8: An Example of Weekly Sentiment Trend of a Promotional Campaign.

Overall, we downloaded more than 100,000 comments and we randomly verified, for each scenario, the correctness of the classification of more than 500 comments. The verified comments were posted along the last two weeks of January 2018. In particular, for each scenario, 3 groups have been randomly extracted containing 170 comments among the ones estimated as positive, negative and neutral, respectively. Table 6 resumes the achieved results, for each scenario. Figure 9 summarizes the results shown in Table 6 as a bar chart.

It is worth noting that the achieved results are slightly lower (more or less 1%) than the ones shown in Tables 2 and Table 3. Since those tables show the average results of a cross-validation, considering data extracted in November 2017, it is possible to affirm that the SAE of the proposed DSS does not suffer from any concept drift [70]. Indeed, when dealing
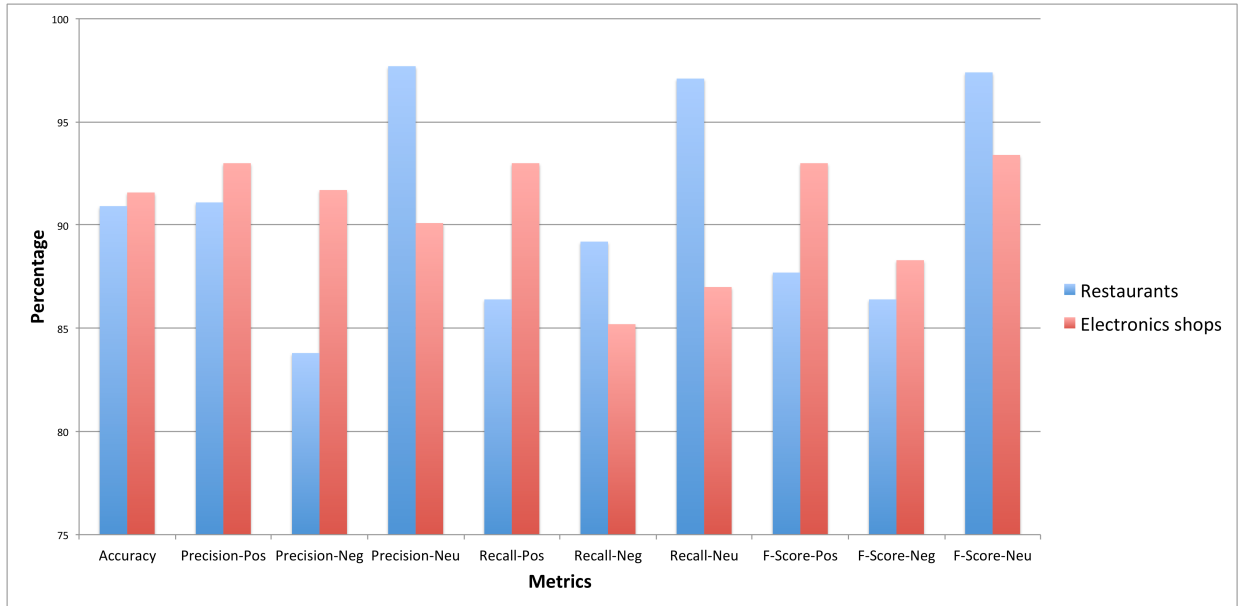
25

Figure 9: Bar chart of the results achieved during the social media listening campaign.

with continuous classification of data streams along the time, this issue should be analyzed. The classification models are usually trained using data extracted in a specific time interval. Then, the models are used for classifying the new instances received in streaming. Since the characteristics of the phenomenon under observation can change during the time, the performance of the classification models may deteriorate, due to this concept drift.

Table 6: Results achieved in the social media listening campaign.

Scenario	Accuracy (%)	Precision(%) by class			Recall(%) by class			F-Score(%) by class		
		Pos	Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu
Restaurants	90.92	91.10	83.80	97.70	86.40	89.20	97.10	87.70	86.40	97.40
Electronics shops	91.58	93.00	91.70	90.10	93.00	85.20	87.00	93.00	88.30	93.40

8. Conclusions and Future Work

In this work, a Decision Support System (DSS) for Social Media Listening has been designed and developed, based on a Sentiment Analysis Engine, which applies sentiment analysis models over comments published on business social media channels. In order to train the classification models of the Sentiment Analysis Engine, labeled texts available on user reviews websites have been exploited, such as TripAdvisor and Amazon.

Specifically, the proposed DSS is able to extract the stream of users' comments, regarding promotions and/or marketing campaigns, from a set of social channels, available as public pages or channels of social networking platforms such as Facebook, Twitter and Instagram.

The stream of comments is elaborated and classified by the Sentiment Analysis Engine in terms of positive, negative or neutral opinions.

The DSS is also able to create reports and statistics, which summarize the result of the social media listening. In particular, the trends of the likes, of the followers and of the positive and negative comments can be provided to the marketing manager of a specific enterprise. These reports can be used for evaluating the brands reputation or as a feedback of specific marketing campaigns.

We experimented the proposed system considering two application scenarios, namely restaurants and consumer electronics online shops. For each scenario, several classification models based on machine learning have been tested and compared. These classification models have been trained by using streams of user reviews extracted from TripAdvisor and Amazon. At the end of this experimental stage, the most suitable classifier for each scenario has been selected and embedded into the Sentiment Analysis Engine. Finally, a simulated social media listening campaign has been carried out, by analyzing the public social channels of popular restaurants and consumer electronics online shops. More than 100,000 comments have been downloaded and analyzed and the sentiment of around 1,000 texts has been manually verified. More than 90% of the verified texts were correctly classified.

As future work, we plan to improve the DSS with the integration of a module for profiling the activities of users that interact with the monitored promotions and marketing campaigns. By exploiting the users social profile, the system may give indications regarding how to drive promotions and marketing campaigns targeted on a specific user category. To this aim, clustering algorithms may be adopted, considering as inputs the social profiles of users. Each user profile may be described by using static information (such as her gender, age, origin, language, interests, etc.) as well as dynamic information (such as the number of comments, images, videos, etc. posted in a time interval). Users with similar profiles presumably fall in the same cluster and may be attracted by similar promotions, marketing campaigns or advertising. Moreover, we envision that the DSS may be easily adapted to monitoring the opinions and the stance of social media users on different hot topics and discussions, such as elections, immigration and terrorism.

Finally, we are currently trying to arrange some agreements with enterprises that may be interested in using the software platform that integrates the proposed DSS for handling and monitoring their marketing campaigns.

[1] F. Hussain, Internet of everything, in: Internet of Things, Springer, 2017, pp. 1–11.
[2] C. Fuchs, Social media: A critical introduction, Sage, 2017.
[3] R. Felix, P. A. Rauschnabel, C. Hinsch, Elements of strategic social media marketing: A holistic framework, Journal of Business Research 70 (2017) 118–126.

[4] B. J. Keegan, J. Rowley, Evaluation and decision making in social media marketing, Management Decision 55 (1) (2017) 15–31.

[5] S. Dahl, Social media marketing: Theories and applications, Sage, 2018.

[6] J. A. Balazs, J. D. Velsquez, Opinion mining and information fusion: A survey, Information Fusion 27 (Supplement C) (2016) 95–110.

[7] M. O. J. Laine, C. Frühwirth, Monitoring social media: Tools, characteristics and implications, in: P. Tyrväinen, S. Jansen, M. A. Cusumano (Eds.), Software Business: First International Conference, ICSOB 2010, Jyväskylä, Finland, June 21-23, 2010. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 193–198.

[8] M. Balduini, E. Della Valle, D. Dell'Aglio, M. Tsytsarau, T. Palpanas, C. Confalonieri, Social listening of city scale events using the streaming linked data framework, in: H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, K. Janowicz (Eds.), The Semantic Web - ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 1–16.

[9] D. A. Schweidel, W. W. Moe, Listening in on social media: A joint model of sentiment and venue format choice, Journal of Marketing Research 51 (4) (2014) 387–402.

[10] N. R. Rao, N. R. Rao, Social Media Listening and Monitoring for Business Applications, 1st Edition, IGI Global, Hershey, PA, USA, 2016.

[11] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowledge-Based Systems 89 (Supplement C) (2015) 14–46.

[12] A. Yadollahi, A. G. Shahraki, O. R. Zaiane, Current state of text sentiment analysis from opinion to emotion mining, ACM Computing Surveys (CSUR) 50 (2) (2017) 25.

[13] C. Catal, M. Nangir, A sentiment classification model based on multiple classifiers, Applied Soft Computing 50 (2017) 135–141.

[14] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal 5 (4) (2014) 1093–1113.

[15] C. Dhaoui, C. Dhaoui, C. M. Webster, C. M. Webster, L. P. Tan, L. P. Tan, Social media sentiment analysis: lexicon versus machine learning, Journal of Consumer Marketing 34 (6) (2017) 480–488.

[16] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational Linguistics 37 (2) (2011) 267–307.

[17] A. Esuli, F. Sebastiani, SENTIWORDNET: A publicly available lexical resource for opinion mining, in: 5th Conference on Language Resources and Evaluation, 2006, pp. 417–422.

[18] A. Tripathy, A. Agrawal, S. K. Rath, Classification of sentiment reviews using n-gram machine learning approach, Expert Systems with Applications 57 (2016) 117–126.

[19] N. P. Cruz, M. Taboada, R. Mitkov, A machine-learning approach to negation and speculation detection for sentiment analysis, Journal of the Association for Information Science and Technology 67 (9) (2016) 2118–2136.

[20] B. Agarwal, N. Mittal, Machine learning approach for sentiment analysis, in: Prominent feature extraction for sentiment analysis, Springer, 2016, pp. 21–45.

[21] I. Perikos, I. Hatzilygeroudis, Recognizing emotions in text using ensemble of classifiers, Engineering Applications of Artificial Intelligence 51 (2016) 191–201.

[22] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification, in: Association for Computational Linguistics, 2007, pp. 187–205.

[23] D. Bollegala, T. Mu, J. Y. Goulermas, Cross-domain sentiment classification using sentiment sensitive embeddings, IEEE Transactions on Knowledge and Data Engineering 28 (2) (2016) 398–410.

[24] F. Wu, S. Wu, Y. Huang, S. Huang, Y. Qin, Sentiment domain adaptation with multi-level contextual sentiment knowledge, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, ACM, New York, NY, USA, 2016, pp. 949–958.

[25] F. Wu, Y. Huang, Z. Yuan, Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources, Information Fusion 35 (Supplement C) (2017) 26–37.

[26] M. S. Hajmohammadi, R. Ibrahim, A. Selamat, Cross-lingual sentiment classification using multiple

source languages in multi-view semi-supervised learning, Engineering Applications of Artificial Intelligence 36 (2014) 195–203.

[27] A. M. Fernández, A. Esuli, F. Sebastiani, Distributional correspondence indexing for cross-lingual and cross-domain sentiment classification, Journal of Artificial Intelligence Research 55 (2016) 131–163.

[28] P. Zhang, S. Wang, D. Li, Cross-lingual sentiment classification: Similarity discovery plus training data adjustment, Knowledge-Based Systems 107 (2016) 129–141.

[29] Y. Zhang, R. Jin, Z.-H. Zhou, Understanding bag-of-words model: a statistical framework, International Journal of Machine Learning and Cybernetics 1 (1-4) (2010) 43–52.

[30] L. H. Patil, M. Atique, A novel feature selection based on information gain using wordnet, in: 2013 Science and Information Conference, 2013, pp. 625–629.

[31] C. Shang, M. Li, S. Feng, Q. Jiang, J. Fan, Feature selection via maximizing global information gain for text classification, Knowledge-Based Systems 54 (Supplement C) (2013) 298–309.

[32] Y. Goldberg, O. Levy, word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method, arXiv preprint arXiv:1402.3722.

[33] J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[34] Y.-M. Li, T.-Y. Li, Deriving market intelligence from microblogs, Decision Support Systems 55 (1) (2013) 206–217.

[35] J. Du, J. Xu, H.-Y. Song, C. Tao, Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with twitter data, BMC medical informatics and decision making 17 (2) (2017) 69.

[36] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, R. R. Suárez, O. S. Siordia, A simple approach to multilingual polarity classification in twitter, Pattern Recognition Letters 94 (2017) 68–74.

[37] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (2018) e1253.

[38] M. A. Paredes-Valverde, R. Colomo-Palacios, M. d. P. Salas-Zárate, R. Valencia-García, Sentiment analysis in spanish for improvement of products and services: a deep learning approach, Scientific Programming 2017.

[39] A. K. Uysal, Y. L. Murphey, Sentiment classification: Feature selection based approaches versus deep learning, in: 2017 IEEE International Conference on Computer and Information Technology (CIT), IEEE, 2017, pp. 23–30.

[40] L. Arras, F. Horn, G. Montavon, K.-R. Müller, W. Samek, " what is relevant in a text document?": An interpretable machine learning approach, PloS one 12 (8) (2017) e0181142.

[41] P. Hájek, Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns, Neural Computing and Applications 29 (7) (2018) 343–358.

[42] M. Ghiassi, D. Zimbra, S. Lee, Targeted twitter sentiment analysis for brands using supervised feature engineering and the dynamic architecture for artificial neural networks, Journal of Management Information Systems 33 (4) (2017) 1034–1058.

[43] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, C. Carlsson, Past, present, and future of decision support technology, Decision support systems 33 (2) (2002) 111–126.

[44] D. Arnott, G. Pervan, A critical analysis of decision support systems research revisited: the rise of design science, in: Enacting Research Methods in Information Systems, Springer, 2016, pp. 43–103.

[45] T. A. Patel, M. Puppala, R. O. Ogunti, J. E. Ensor, T. He, J. B. Shewale, D. P. Ankerst, V. G. Kaklamani, A. A. Rodriguez, S. T. Wong, et al., Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods, Cancer 123 (1) (2017) 114–121.

[46] P. Lambin, J. Zindler, B. G. Vanneste, L. Van De Voorde, D. Eekers, I. Compter, K. M. Panth, J. Peerlings, R. T. Larue, T. M. Deist, et al., Decision support systems for personalized and participative radiation oncology, Advanced drug delivery reviews 109 (2017) 131–153.

[47] M. Lachhab, C. Béler, T. Coudert, A risk-based approach applied to system engineering projects: A

new learning based multi-criteria decision support tool based on an ant colony algorithm, Engineering Applications of Artificial Intelligence 72 (2018) 310–326.

[48] M. Kraus, S. Feuerriegel, Decision support from financial disclosures with deep neural networks and transfer learning, Decision Support Systems 104 (2017) 38–48.

[49] Y. O. Serrano-Silva, Y. Villuendas-Rey, C. Yáñez-Márquez, Automatic feature weighting for improving financial decision support systems, Decision Support Systems 107 (2018) 78–87.

[50] K. Gandhi, B. Schmidt, A. H. Ng, Towards data mining based decision support in manufacturing maintenance, Procedia CIRP 72 (2018) 261–265.

[51] T. R. Rakes, J. K. Deane, L. P. Rees, D. M. Goldberg, J. Chacko, A fuzzy decision support system for pre-disaster budgeting, International Journal of Information Systems and Management 1 (4) (2018) 312–327.

[52] E. Sugiyarti, K. A. Jasmi, B. Basiron, M. Huda, K. Shankar, A. Maseleno, Decision support system of scholarship grantee selection using data mining, International Journal of Pure and Applied Mathematics 119 (15) (2018) 2239–2249.

[53] J. Rumbaugh, I. Jacobson, G. Booch, Unified modeling language reference manual, the, Pearson Higher Education, 2004.

[54] F. Buschmann, K. Henney, D. C. Schmidt, Pattern-Oriented Software Architecture, On Patterns and Pattern Languages, Vol. 5, John Wiley & Sons, 2007.

[55] M. Yener, A. Theedom, Professional Java EE design patterns, John Wiley & Sons, 2014.

[56] M. Habibi, Real World Regular Expressions with Java 1.4, APress, 2004.

[57] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2016.

[58] E. D'Andrea, P. Ducange, B. Lazzerini, F. Marcelloni, Real-time detection of traffic from Twitter stream analysis, IEEE Transactions on Intelligent Transportation Systems 16 (4) (2015) 2269–2283.

[59] A. Hotho, A. Nrnberger, G. Paa, A brief survey of text mining, LDV Forum - GLDV Journal for Computational Linguistics and Language Technology 20 (1) (2005) 19–62.

[60] Z. Yao, C. Ze-wen, Research on the construction and filter method of stop-word list in text preprocessing, in: 2011 Fourth International Conference on Intelligent Computation Technology and Automation, Vol. 1, 2011, pp. 217–221.

[61] M. Porter, An algorithm for suffix stripping, Program: Electronic Library and Information Systems 40 (3) (2006) 211–218.

[62] G. Salton, C. Buckley, Readings in information retrieval, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997, Ch. Term-weighting Approaches in Automatic Text Retrieval, pp. 323–328.

[63] J. C. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Schölkopf, C. J. C. Burges, A. J. Smola (Eds.), Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, MA, USA, 1999, Ch. Fast Training of Support Vector Machines Using Sequential Minimal Optimization, pp. 185–208.

[64] G. H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, in: Uncertainty in Artificial Intelligence, UAI'95, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[65] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[66] D. W. Aha, D. Kibler, M. K. Albert, Instance-based learning algorithms, Machine Learning 6 (1) (1991) 37–66.

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

[68] A. Valdivia, M. V. Luzn, F. Herrera, Sentiment analysis in tripadvisor, IEEE Intelligent Systems 32 (4) (2017) 72–77.

[69] G. Forman, An extensive empirical study of feature selection metrics for text classification, Journal of machine learning research 3 (Mar) (2003) 1289–1305.

[70] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM computing surveys (CSUR) 46 (4) (2014) 44.