Electronic Markets (2019) 29:93–106
https://doi.org/10.1007/s12525-019-00336-z

RESEARCH PAPER

Cognitive computing for customer profiling: meta classification for gender prediction

Robin Hirt[1] · Niklas Kühl[1] · Gerhard Satzger[1]

Received: 25 September 2017 / Accepted: 29 January 2019 / Published online: 21 February 2019

Abstract
Analyzing data from micro blogs is an increasingly interesting option for enterprises to learn about customer sentiments, public opinion, or unsatisfied needs. A better understanding of the underlying customer profiles (considering e.g. gender or age) can substantially enhance the economic value of the customer intimacy provided by this type of analytics. In a design science approach, we draw on information processing theory and meta machine learning to propose an extendable, cognitive classifier that, for profiling purposes, integrates and combines various isolated base classifiers. We evaluate its feasibility and the performance via a technical experiment, its suitability in a real use case, and its utility via an expert workshop. Thus, we augment the body of knowledge by a cognitive method that enables the integration of existing, as well as emerging customer profiling classifiers for an improved overall prediction performance. Specifically, we contribute a concrete classifier to predict the gender of German-speaking Twitter users. We enable enterprises to reap information from micro blog data to develop customer intimacy and to tailor individual offerings for smarter services.

Keywords Cognitive computing · Micro blog data · Gender detection · Meta machine learning · Meta classifier

JEL classification C · M · O

Introduction

Many service industries thrive on an improved understanding of customers and on the exploitation of this knowledge via tailored offerings—thus pursuing a "customer intimacy" strategy (Habryn 2012; Treacy and Wiersema 1993). As potential customers willingly put forward information via social media (Baird and Parasnis 2011), this source of "large-scale" information has been recognized and already tapped into to retrieve information valuable to marketing or innovation managers, e.g., in eliciting sentiments (Narr et al. 2012) or disclosing unfulfilled needs (Kühl et al. 2016; Wieneke and Lehrer 2016). Heimbach et al. (2015) show the importance of social media data like Twitter, and the importance of the authors' profiles. Detailed knowledge of authors' demographics allows marketers to build customer segments and to tailor individualized, segment-specific offers. However, as Twitter does not publicly provide any reliable information about its users, there is little to no automatically retrievable information on the tweet authors' personal characteristics. E.g., the "big five" factors (Allport and Odbert 1936) described from a psychological perspective, which include gender, age, location, or personality traits, are not given. Thus, in order to use social blog data to implement customer intimacy strategies on a "large scale," methods and tools are needed to elicit information about the user in an automated fashion.

Within user demographics, gender plays a particularly important role (Arnold and Bianchi 2001). As Fischer and Arnold (1994) show, the gender of a customer has considerable influence on consumer behavior. Related work already

Responsible Editor: Haluk Demirkan

This article is part of the Topical Collection on Smart Services: the move to customer- orientation

✉ Robin Hirt
robin.hirt@kit.edu

Niklas Kühl
niklas.kuehl@kit.edu

Gerhard Satzger
gerhard.satzger@kit.edu

[1] Karlsruhe Institute of Technology, Karlsruhe Service Research Institute, Kaiserstr. 89, D-76133 Karlsruhe, Germany

shows the possibilities of predicting gender based on a variety of data sources (Burger et al. 2011). However, each individual classifier is treated as a separate demographic feature, while the aggregation power of cognitive approaches has not been exploited yet. It is known, e.g., that the human brain digests separate pieces of information from visual or audio input to merge insights from different methods and data sources (Grimes 1990). Cognitive computing, as the brain's artificial intelligence equivalent, "aims to develop a coherent, unified, universal mechanism inspired by the mind's capabilities" (Modha et al. 2011, p. 62). Thus, we aim to answer the following research question (RQ):

How can we determine customer characteristics like gender based on multiple heterogeneous sets of data and different prediction techniques?

In order to answer this question applying a design science research (DSR) approach, we draw on information processing theory and meta machine learning to propose a modular, extendable, and cognitive classifier for profiling purposes that can integrate and combine various isolated base classifiers. We do an evaluation which is divided into three parts. First, we evaluate the feasibility and the statistical performance of our approach for predicting gender via a technical experiment, specifically considering the case of twitter data posted in German. By comparing the meta classifier's performance to the underlying base classifiers, we draw on the performance increase of our approach, and, further, we include a separate third-party base classifier to elaborate on the flexibility. Second, we evaluate the meta classifier's suitability via a real-world test in the field of e-mobility. In addition, we make the artifact available as a public web service to enable readers to explore and test its capabilities themselves. Third, in an expert workshop with practitioners, we draw first insights on the practical utility of the gender prediction service.

Thus, we augment the body of knowledge by proposing a dynamic cognitive method that enables easy integration of existing as well as emerging customer profiling classifiers, to provide improved overall prediction performance. As an instantiation, we contribute a concrete (meta) classifier to predict the gender of German-speaking Twitter users. We further contribute by enabling enterprises to reap information from micro blog data in order to segment customer groups, detect trends, and build customer intimacy as a base for novel, tailored service and product offerings (Neuhofer et al. 2015).

The paper proceeds as follows: We detail our research design in section 2, and then raise awareness for the problem by screening related work and practitioner views in section 3. In section 4 we conceptualize a method for combining heterogeneous classifiers based on the cognitive paradigm. Next, in section 5, by describing our data set, the available single-source classifiers, and their combination through a "cognitive" meta classifier, we develop and implement the artifact instantiation for detecting the gender of German language tweet authors. Subsequently, in Section 6, we evaluate the method along different dimensions. Section 7 concludes this paper by acknowledging limitations, highlighting theoretical contributions, and managerial implications, as well as indicating future research avenues.

Research design

To guide and structure our research, we follow a Design Science Research (DSR) approach. Different DSR methodologies have been proposed in the field of information systems, most commonly those of (Hevner et al. 2004), Peffers et al. (2007), as well as Kuechler and Vaishnavi (2012). While there are no general criteria or rules to choose particular DSR variants (Gregor and Hevner 2013), given an implementation-heavy approach like ours, we favor clearly differentiating between an abstract *suggestion* and a concrete, more technical, implementation-centric *development*. Thus, we follow the approach by Kuechler and Vaishnavi (2012) that explicitly emphasizes these aspects—as depicted in Fig. 1.

The design of our method is guided by information processing theory (Craik and Lockhart 1972) and meta machine learning (Giraud-Carrier et al. 2004) which provide justificatory knowledge and function as kernel theories (Gregor and Jones 2007). In the case of predicting the gender of German-speaking Twitter authors, information processing theory's specific instantiation is additionally based on existing techniques for customer profiling (Burger et al. 2011; Liu and Ruths 2013).

In terms of knowledge contribution, the work presented here, according to Gregor and Hevner (2013), is an *improvement* as the work applies a novel solution—the cognitive combination of different machine learning models with heterogeneous data sources—to the challenge of gender detection in micro blog data, specifically in our instantiation of gender detection of German language tweets' authors. In order to evaluate the artifact, we use a technical experiment that measures the statistical performance of the approach with self-labeled Twitter profiles. Additionally, we use the suitability of our instrument in an illustrative application scenario in the e-mobility domain to demonstrate economic value (Peffers et al. 2012). Furthermore, we aim to gain first insights about the practical utility of the developed artifact in an expert workshop.

Related work and contribution to theory

In order to identify state-of-the-art solutions from literature, we conducted a comprehensive literature review (Webster and

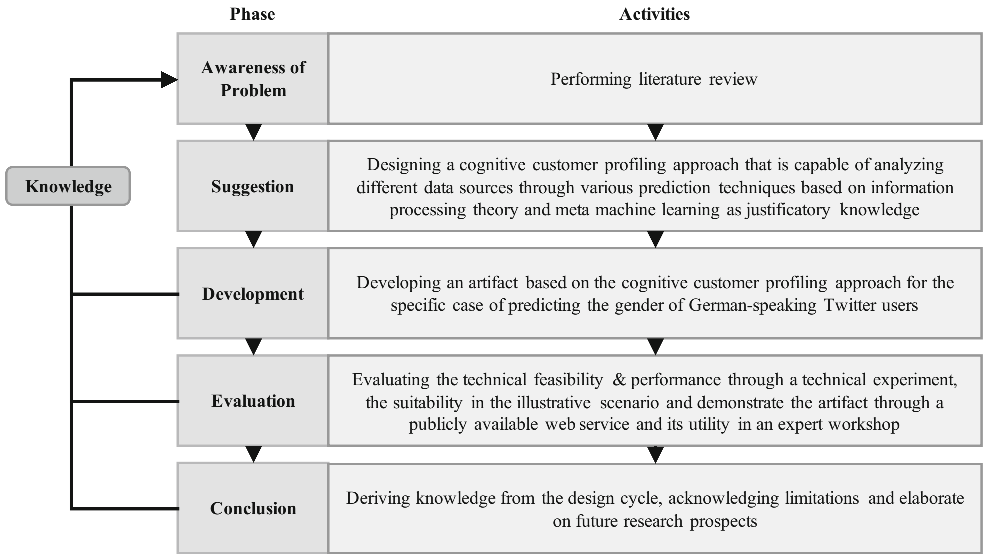


Fig. 1 DSR activities of this work based on Kuechler and Vaishnavi (2012)

Watson 2002). First, our review elaborates on common techniques for automated customer profiling. Then, we examine research on the combination of different data sources for customer profiling. Lastly, we show how our work deviates from existing literature.

For many insightful attributes of individual authors, it has been shown that their prediction is, to a certain degree, feasible, as for age (Burger et al. 2011), for personality traits (Nguyen et al. 2013), for ethnicity (Bergsma et al. 2013), or for political affiliation (Gottipati et al. 2014). Table 1 provides an overview of studies that aim to classify different attributes, while also introducing different approaches to classification.

Methods that aim to classify social media users by certain attributes focus either on the text's linguistic features (e.g., the text of a tweet), or on other publicly available profile information, such as the name of the user (Liu and Ruths 2013) or even the user-selected profile colors (Alowibdi et al. 2013). Based on supervised machine learning, Liu and Ruths (2013) use the author's name as feature for gender classification of English-speaking users. Although using names could be a promising basis for classification in determining gender, other demographic attributes such as personality traits might not be similarly derivable.

Due to user attributes being strongly correlated with the user's language (Tausczik and Pennebaker 2010), texts are often used as a basis for classification. User classification of demographics based on longer texts, such as e-mails (Estival et al. 2007) or blog posts (Argamon et al. 2009), is a known technique in many media settings. However, these approaches leverage and analyze long texts, while tweets or other short messages have different, unique characteristics such as the use of fixed phrases.

Schwartz et al. (2013) studied Facebook messages of a similar length to tweets, showing that even in short messages gender, age and language are statistically correlated. In contrast to tweets, these messages are private, users are likely to be more open and, therefore, will implicitly reveal more personal information in their messages. Furthermore, a tweet contains Twitter-specific elements, like mentions or hashtags.

Arroju et al. (2015) classify age, gender and personality traits of a Twitter user in English, Dutch, Spanish, and Italian by applying an n-gram based supervised machine learning classifier. To enhance their classifications, external dictionaries are linked to the used words. Similarly, Burger et al. (2011) focus on the prediction of gender and age of a user by applying supervised machine learning techniques to tweets, description, screen name, and display name. They do not limit their collected tweets to a certain language, but treat different languages such as English, Portuguese, Spanish, and Dutch without attention to linguistic differences. Although they consider different sources of information, they treat the user's name and description as an extension of the user-written texts. They test their classification approach with each information source separately and can therefore assess the usefulness of each source.

One key problem of all the mentioned approaches is the limitation of information gathering from one source, e.g., textual data only, or of applying one classifier to more than one source. Taking different, heterogeneous information sources into account and classifying them separately in the classification of users, could be a promising approach. Ikeda et al. (2013) classify age, gender, area, occupation, hobbies, and marital status using a hybrid approach of supervised learning, network analysis, and a statistical estimation method. They perform a combined classification by applying a hybrid estimator that determines the final prediction based on statistical decisions.

Rao et al. (2010) apply supervised machine learning techniques to classify English-speaking Twitter users by age, gender, regional and political origin. They consider the

Table 1 Overview of related work in the field of user characteristics prediction

Articles	Data source	Text-based	Image-based	Name-based	Location-based	Network-based	Color-based	Hybrid approach	Language	Target
Estival et al. (2007)	E-mails	x							English	Gender, age, geographic origin, level of education, native language
Argamon et al. (2009)	Blog posts	x							English	Gender, age, neuroticism, native language
Rao et al. (2010)	Tweets	x							English	Gender, age, regional origin, political orientation
Burger et al. (2011)	Tweets, user description, user name	x							Multilingual	Gender
Alowibdi et al. (2013)	Twitter profile colors						x		Language independent	Gender
Bergsma et al. (2013)	Twitter user name, location			x	x			x	English	Gender, age, location, origin, ethnicity, race
Ikeda et al. (2013)	Tweets, network	x				x		x	Japanese	Gender, age, area, hobby, occupation, marital status
Liu and Ruths (2013)	Twitter user name			x					Language independent	Gender
Nguyen et al. (2013)	Tweets	x							Dutch	Gender, age
Schwartz et al. (2013)	Facebook messages	x							English	Gender, age
Arroju et al. (2015)	Tweets	x							English, Spanish, Dutch, Italian	Gender
This work	*Tweets, Twitter user name, profile image*	x	x	x				x	*German*	*Gender*

sociolinguistic features of a user and n-grams of his or her tweets, testing each feature type on its own and combining those two by a stacked support vector machine model. In some cases, they increase the classifier performance slightly. However, the authors neither test different classifier algorithms, nor give detailed information about their model and the validation techniques they use. Further, both underlying base classifier types rely on the user's tweets. They do not combine different sources of information.

As related work shows, there are different approaches to analyzing user attributes. In the fast-changing environment of social media, new approaches can emerge and coexist with older ones. Additionally, social media platforms, such as Twitter, offer a wide range of data sources that could require consideration during analysis.

Thus, the question arises how different, heterogeneous information sources and classifiers can be (re-)combined and used for holistic, superior predictions. In contrast to previous work, our goal is to analyze different heterogeneous data sources that appear in a Twitter profile, and to utilize meta machine learning, a technique that is typically used to increase the prediction performance (Dietterich 1997). Meta machine learning combines the prediction of several base classifiers to create one aggregated prediction (Quinlan 2006).

For each base classifier, we build on base classifiers used in techniques examined in related work, such as a text-based classification of Twitter users similar to Burger et al. (2011). We focus specifically on German-speaking users, and a language-independent name-based classification of Twitter users similar to Liu and Ruths (2013).

In contrast to Rao et al. (2010) who also combine different prediction techniques, we do not take only one source of information into consideration, but a combination of different sources for a more flexible meta machine learning classification.

Lastly, there is no work on automatically predicting the gender of German-speaking Twitter users—an obvious research gap given the increasing number of German-speaking Twitter users recently, and their activity on the microblogging platform that had an estimated 5.76 million users in 2016 (Statista 2016). Thus, in contrast to all the mentioned studies, we fully automate the gender classification process of one or multiple Twitter users who write in German, by creating a web service architecture that can be integrated into any other smart service or analytics applications.

Summarizing, we can identify the lack of (a) a method that is capable of dynamically combining different data sources and prediction techniques for customer profiling, and (b) a German-based twitter gender predictor. Thus, our objective is to design a cognitive classifier concept that is able to combine prediction models based on heterogenous data sources, and is capable of predicting the gender of German-speaking Twitter users.

A cognitive approach to customer classification through meta machine learning

We are interested in a method that allows for flexible (re-)-combination of various classifiers that are based on heterogeneous data sources, to realize customer profiling. To illustrate such customer profiling, we investigate the gender classification of social media users. To be more specific, in the case of gender classification for German-speaking Twitter users, the question arises as to how one can identify the gender of a user in general. Looking at such a gender profile is difficult due to two limiting factors. First, social media users might want to stay anonymous and, therefore, may not actively share their gender. They might even put forward false information by using pseudonyms, and entering misleading description texts or profile pictures. Second, even if they do not have any anonymity intention, they might still not want to share their personal information publicly. Thus, reliance on only one source of information for classification would most likely result in too large a number of false classifications. For coping with this issue, the presented approach is inspired by the perception and decision making, i.e. the cognition, of a human being, and it is based on information processing theory as a core theory (Miller 1956; Modha et al. 2011; Rumelhart and McClelland 1986). Information processing theory describes the human mind as a processor of input derived from the senses, which is finally translated into a behavioral response (Miller 1956). If a human had to classify the gender of a Twitter user, the first step would be to look at and 'analyze' the available sources of information. The name, picture, description text, or the tweets of a user can be indicators of his/her gender. A human would screen every information source in the user profile and make a *holistic* decision in identifying the "bigger picture." The research field of multimodal fusion faces similar problems as it aims to analyze multimedia content such as videos, by first separately processing visual and audio content and then aggregating it (Atrey et al. 2010; Kludas et al. 2008). Additionally, the term "cognitive computing" describes processing paradigms which are inspired by the capabilities of the human mind (Modha et al. 2011). Similarly, we aim to use a cognitive paradigm to combine heterogeneous information sources by applying meta machine learning as a second core theory (Todorovski and Džeroski 2003). Meta machine learning propagates the use of more than one layer of machine learning to enable "learning about learning" to accumulate knowledge from different domains or problems (Vilalta and Drissi 2002). This work aims to mimic human cognitive abilities through a meta machine learning paradigm.

We propose an analogous cognitive approach of different, independent classifiers, each of which analyzes one information source for which it creates an independent classification including a corresponding certainty factor. These independent classifier results are then combined into one feature vector without data-specific traits rendering the meta classification data-agnostic. As a basis for creating the independent single-source classifiers, we build on a combination of existing approaches to customer profiling (Burger et al. 2011; Liu and Ruths 2013).

As already stated, we aim to instantiate the presented approach for gender prediction. It is, however, flexible and modular, so it can also be applied to other user-specific attribute classifications. To achieve this, we rely on a *text (tweet)* machine learning classifier, a *name dictionary* machine learning classifier, a third-party image classifier, and a *superordinate meta* machine learning classifier. We strive to develop two instantiations of a meta classifier: the first solely uses underlying classifiers that we develop on our own so that we can control the evaluation setup to make reliable statements regarding the performance reached. The second makes use of a third-party classifier that has not been developed for use in meta classification, to show the extensibility of our approach, and to show possible improvements due to considering additional classifiers.

The goal of the *text* classifier is to determine the gender of a Twitter user by analyzing the user's available tweets. Similar to Burger et al. (2011), we use NLP and supervised machine learning approaches to classify the gender of a user. The *name* classifier considers the "screen name" of a Twitter User (e.g., "@John_Doe1978") and the "display name" that is shown at the top of the profile of a user (e.g. "John Doe"). The suggested name classifier first analyzes, cleans, and builds variations of the display and the screen name. Then, these variations are compared to a dictionary of gender-tagged names and matched to similar ones. After that, a heuristic determines the most likely gender and the probability for the classification. Additionally, we make use of a third-party classifier. We use the IBM Visual Recognition API (IBM 2016) to classify the faces that appear on user profile pictures as a third-party image classifier to predict the gender.

Finally, the separate classifiers predict the gender of a profile independently and calculate a confidence score for the meta prediction. The result of each classifier is used to generate a meta feature, which represents the prediction result of all underlying classifiers. We employ two separate meta classifiers that predict the gender based on the results of the underlying classifiers. The first meta classifier (A) considers only the name and text classifier. The second (B) includes the results of the third-party image classifier.

The meta classification itself can be designed in various ways, as a cascading model (Gama and Brazdil 2000), a voting mechanism (Breiman 1996), or a meta machine learning model (Džeroski and Ženko 2004). We implement the third option, using machine learning to train an additional classifier to make a *meta prediction*. The results of the underlying base classifiers are used as an input (the meta features) and the meta output then is a binary classification (female or male). We follow this third option and implement a second layer of machine learning to enhance the performance of our base classifiers; therefore, we minimize the uncorrelated error of the base classifiers (Todorovski and Džeroski 2003).

To visualize our general idea of a holistic cognitive classifier, in Fig. 2 we schematically portray the amount of correctly classified profiles for each classifier, base and meta. Every base classifier S is able to predict the gender of a set of profiles correctly (true positives). For example, the name classifier could be very accurate in case a user displays his or her real name and a clear gender implication exists. However, if a user shows a fake name, the name classifier could propagate a false result. A meta classifier would ideally be able to identify, in which situation which classifier's prediction is correct and "select" the correct class for a given data point. In Fig. 2, we depict two cases: one meta classifier utilizing the two base classifiers text and name (*meta classifier A*) and one utilizing the results of all three base classifiers, text, name, as well as image (*meta classifier B*). The underlying principle here is to improve the ability of a meta classifier to detect and minimize the uncorrelated error of different sub classifiers (Džeroski and Ženko 2004; Todorovski and Džeroski 2003).

Further, with enough training, the meta classifier is able to identify cases where all underlying base classifiers yield a false class for a data point. If this is achieved, the meta classifier has learnt about situations that indicate such a situation, in which case the meta classifier would make a contrary prediction. However, Fig. 2 represents an idealistic setting where the meta classifiers are able to successfully determine the true positives of every underlying base classifier.

Additionally, to show the utility of such a meta classifier, we intend to automate the data input and output, and make the presented meta classifier A accessible. We suggest an integrated web service architecture that can be accessed through a public application programming interface (API).

Development of an artifact for gender detection in twitter data

To instantiate the proposed method in an artifact, we use a real-world data set based on Twitter data. In what follows, we describe the data set, the base classifiers and their combination through a meta classification, and lastly, the deployment of the developed artifact as a web service.

Data set

In order to implement the suggestion, we needed a data set of gender-labeled Twitter profiles for training and testing the machine learning classifiers. For this purpose, we randomly crawled German-speaking Twitter users by using the official Twitter API. We searched for German tweets that contain common words, e.g. the neuter definite article "das". We manually sorted out company accounts, as well as bots, and disregarded them in the further steps. Additionally, we then took the whole profile (including profile image, name, tweets, and other publicly available profile information) into consideration to manually determine the gender of a user. As we had slightly more male than female users, we randomly selected and disregarded 217 male users to balance our distribution. Thus, we arrived at a data set of 2916 profiles, of which 1342 profiles are labeled as male and 1125 profiles as female. For each subsequent training and testing process, we under-sampled the data set to reach a balance between female and male users.

In the following subsections, we describe the development of the single modules of our artifact, namely the text classifier, name classifier, image classifier, meta classifier, as well as their implementation into a web service.

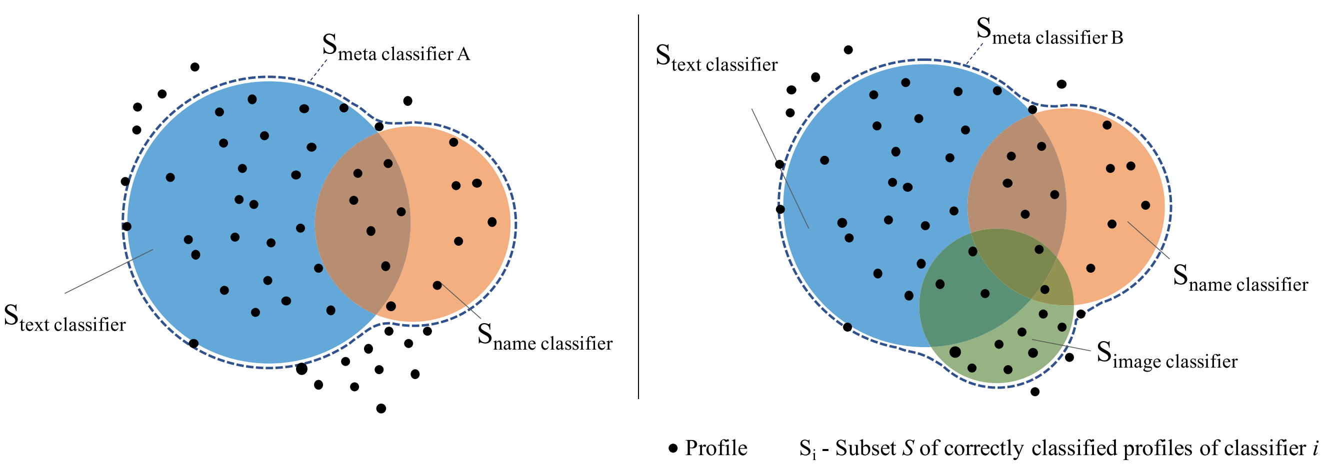


Fig. 2 Schematic, qualitative depiction of profiles and their subset for each classifier that get correctly classified (true positives) in an idealistic setting

Tweet-based classification

Similar to Burger et al. (2011) we use NLP methods and a supervised machine learning approach to classify the gender of a user. We first have to determine the machine learning-relevant features which are extracted from the tweets. Because we assume that with an increasing number of processed tweets the classification performance will increase, we run tests on different numbers of $t \in \{1, 10, 20, 30, 40, 50\}$ tweets per user. We also test different preprocessing methods to optimize our classifier. Then, the resulting text is stemmed (Lovins 1968) and vectorized. Finally, the numeric feature vector and the corresponding binary gender value is used to train a classifier. After training the classifier with the training set, the same procedure of pre-processing and feature extraction is applied to the validation set. Now the trained classifier can be used to classify the validation set and the resulting predictions are compared to the actual gender values. We use 10-fold cross-validation to produce meaningful results (Stone 1977).

For each cross-validation run, the metrics are calculated and averaged afterwards. To find the best configuration and to determine the performance of the proposed method, we evaluate different combinations of pre-processing steps and classifier algorithms by applying a basic grid search (Hsu et al. 2008; Snoek et al. 2012). We choose the classifiers out of three popular algorithm groups (Michie et al. 1994): Bayes classifiers (Gaussian Naïve Bayes, Bernoulli Naïve Bayes), Support Vector Machines (Linear SVC, SVC), and tree-based classifiers (Random Forest). Additionally, we also test ensemble classifiers (Adaboost, Bagging) that combine different algorithms and logistic classifiers (Stochastic Gradient Descent), since they have proven to be successful for other text mining tasks (Zhou et al. 2005). Another option would be to use deep learning for classifying the gender of a user. However, in comparison to other supervised machine learning approaches, deep learning needs large amounts of data for reaching acceptable performance in classification tasks (Goodfellow et al. 2016). As our data set contains only 2916 profiles and we want to narrow our experimental setup, deep learning is not considered as a method for classifying users according their gender.

Name-based classification

At first the display and screen names were cleaned and possible name candidates selected by performing a tokenization. Then, these sub names were compared to all names in a database[1] to find similar names. If, in this procedure, the Levenshtein distance (Levenshtein 1966) between a sub name and a name out of the name database is smaller than a defined threshold, a match is found. The Levenshtein distance thus considers insertions, deletions, and substitutions (Navarro 2001). Based on these matches, we calculated the probability of the gender of the user by finding the maximum occurrences per gender class. In contrast to the text classifier and the meta classifier, the name classifier did not perform a binary male or female classification, but also classified a neutral class. Due to the possibility of some names being both male and female, or of a name not being known, the neutral class and corresponding probabilities are essential. Although the comparison using the Levenshtein distance could yield in false results, we assume that our meta classifier detects faulty cases and handles them accordingly.

Image-based classification

We aim to show that the suggested approach is generic and that gender classifiers of any kind can be integrated; therefore, we used the IBM Visual Recognition API to predict the gender of a user by analyzing the profile picture. We sent a single user's profile picture to the API and received a classification result in return. The classifier itself first finds the position of a face in the picture and then runs a gender prediction based on the recognized face. Among other information, the response contains the predicted gender and a confidence score of the classification.

Combining predictions through meta classification

Figure 3 shows the workflow of a meta classification: First, the separate base classifiers predict the gender of a profile independently and calculate a confidence score of the prediction. Next, the result of each classifier is used to generate a meta feature. The meta classifier itself is machine learning based and identifies when a certain base classifier's prediction is correct.

Similar to the text classifier procedure, the implementation of the meta classifier was divided into a training and a validation phase. The predictions and corresponding confidence scores for each profile were added to a vector that built up the meta feature vector for a certain Twitter user. We performed a nested cross validation to ensure the maximization of testing data and statistical correctness of our results.

We developed two instantiations of a meta classifier: The first one is based on the predictions of the name and text classifier. The second one additionally includes the predictions of the third-party image classifier.

[1] *namedict.txt* from https://heise.de/ct, softlink 0717182 by Jörg Michael, last accessed 15-11-2016

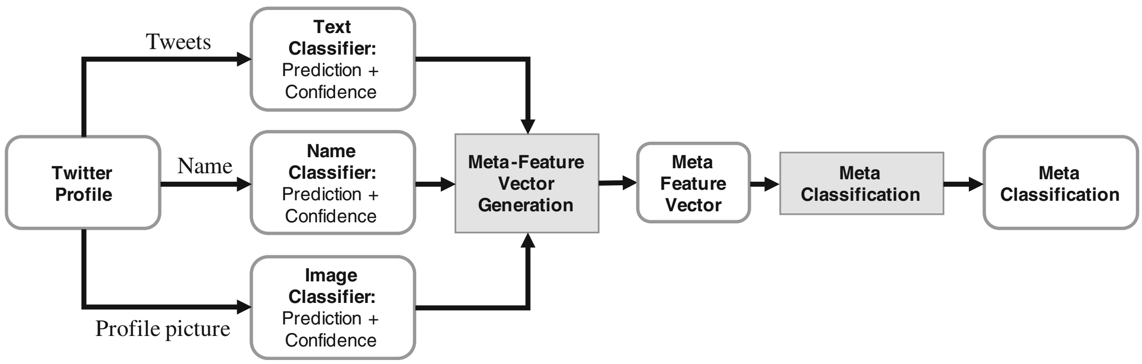


Fig. 3 Meta classifier workflow

Deployment of artifact as a web service

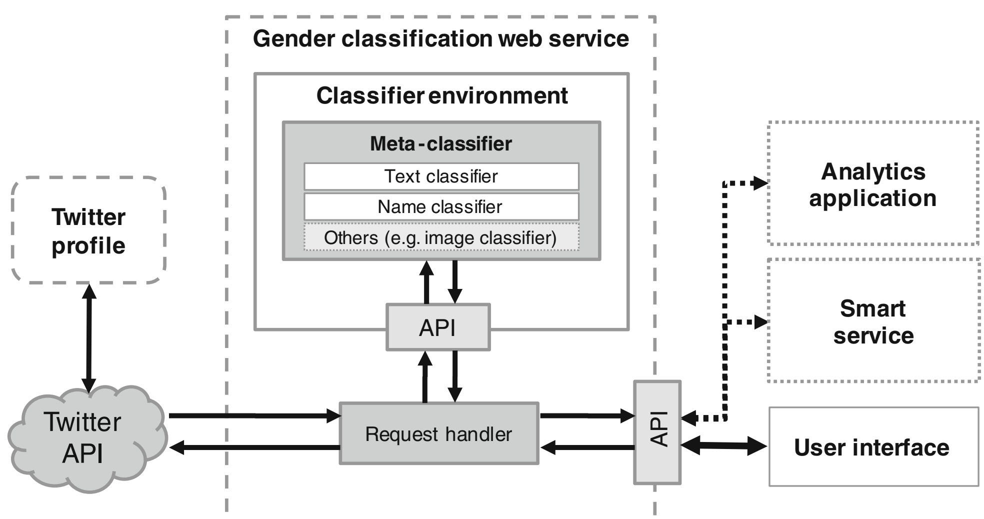
To make the developed artifacts available, we aim to deploy them in a web service. However, as our second meta classifier makes use of a commercial third-party classifier, there are legal constraints determining we can only implement a web service for the meta classifier that combines the text and name base classifier's results.

After implementing, serializing, and exporting the text, name, and meta classifier, a holistic artifact can be built which automates the gender classification and provides a web service. We implement an API for sending a classification object and responding with a corresponding prediction and probabilities. To demonstrate our classifiers, we develop a browser application with a simple user interface. When the user inserts a screen name, a request to the backend triggers the crawling of the user's profile throughout the Twitter API. After receiving the response, the backend checks the validity and prepares the profile for sending a request to the classification API. After the classification, the classifier environment responds with a classification that gets forwarded to the frontend immediately. Figure 4 illustrates the web service architecture.

Evaluation

Sonnenberg and Vom Brocke (2012) distinguish between different evaluation activities throughout the design process. The evaluation of the presented artifact has three facets: First, the feasibility and performances of the presented classifiers are independently tested and analyzed in a technical experiment (Peffers et al. 2012) to verify the correctness of the overall classification. By comparing the single classifiers (text, name, and third-party image classifier) with the two developed meta classifiers, we show the performance increase achieved by the use of the meta classifier concept. This supports our contribution as an improvement (Gregor and Hevner 2013) as, compared to single-source classifiers, ours is a more efficient, significantly improved classifier – particularly considering the multiple heterogeneous data sources and their integration in one meta model. Additionally, we show the flexibility of our cognitive approach that enables a flexible combination of different classifiers. Second, the suitability of the gender classification as a web application itself is shown in an illustrative scenario in general in the field of e-mobility to determine customer attributes on a large-scale. We apply it to a set of

Fig. 4 General architecture of the web service



over 5000 tweets to determine the gender distribution of tweet authors over a period of time. Further, we enable readers to test the classifier themselves by making it available through a web service. Third, we gain first insights on the practical utility of the gender prediction service for German-speaking Twitter users in an expert workshop with practitioners.

Technical experiment

We claim that our cognitive classifier approach contributes as an improvement to existing single-source classifier approaches. However, as we cannot directly compare our cognitive classifier to classifiers described and developed in related work, we performed a technical experiment in an isolated setting. For that, we evaluated the performance of two single-source classifiers that we developed based on related work, and one commercial solution. By comparing the performance of these classifiers to our cognitive classifiers, we can show the improvement.

As described, all results are derived with a 10-fold cross-validation. For statistical performance measurement and comparison, we calculated the metrics of *accuracy*, *precision, recall,* F_β*-score* (Blair 1979) for each class (male and female users). We then averaged the values for both classes to derive the averaged metric.

The *precision* indicates the share of true positives over all predicted positives ("How many predicted observations are relevant?"). It reaches its best value at 1 (100%) and its worst at 0 (0%). In our example, the precision of a random classifier would be 50% as the distribution between male and female users is balanced. It is used as a central metric if it is not important how many of the relevant observations are predicted, but how high the *share* of all predicted observations is.

$$precision = \frac{TP}{TP + FP}$$

The *recall* indicates the share of true positives over all positive predictions ("How many relevant observations are selected?"). It reaches its best value at 1 (100%) and its worst at 0 (0%).

$$recall = \frac{TP}{TP + FN}$$

As a more meaningful measure, which also takes imbalances into account, the F_β*-score* denotes a weighted compromise of recall and precision and reaches its best value at 1 (100%) and its worst at 0 (0%) (Blair 1979).

$$F_\beta = \left(1 + \beta^2\right) * \frac{precision * recall}{precision + recall}$$

For $\beta = 1$, the F_1*-score* delivers a harmonic mean between precision and recall. It is a popular measure since it allows integration of the two extremes of precision and recall within one meaningful metric (Powers 2011). Thus, we chose the F_1*-score* as our main metric for our technical experiment.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

As part of the grid search, we iterate through all different combinations of tweet count (number of regarded tweets per user), pre-processing steps, and classification algorithms. The best text classifier reaches an F_1*-score* of 69.60%, a *recall* of 69.90%, a *precision* of 70.18%, and an *accuracy* of 69.74%. For further development and evaluation of the text classifier, we use the Statistic Gradient Descent classifier with the above described pre-processing configuration.

In contrast to the text, image and meta classifiers, the name classifier does not use machine learning, but a dictionary-based approach to predict the gender of a user. This means the complete data set of 2916 profiles of female and male profiles is used for the evaluation. The name classifier reaches high values in terms of precision (86.75%), but a very low recall of 38.33%. The F_1-score is slightly lower than the one of the text classifier at 69.06%. This means that if the name classifier detects a certain gender, it is a fairly precise. But in contrast, it has a poor recall of gender classes. The third-party image classifier reaches an F_1-score of 25.37% but a precision of 85.75% with a low recall of 15.04%. Note, that even if two classifiers reach comparable performances, such as the text and name classifier in terms of F_1-score, the set of correctly and falsely classified data points (true and false positives) per classifier can differ.

To evaluate the suggested meta classifiers, the data set is divided into a global validation set and two training sets as we previously described. For the meta classifier A, that combines the text and name base classifiers, the Bernoulli Naïve Bayes algorithm performs the best throughout all iterations: out of eight tested algorithms it has an F_1-score of almost 80% as depicted in Table 2. For the meta classifier B, that combines the third-party image classifier with the text and name classifier, similarly, Bernoulli Naïve Bayes algorithm performs the best with an F_1-score of 81.46% (+1.62% compared to meta classifier A).

Table 2 Averaged best classification results (in %)

	F_1-score	Precision	Recall
Text classifier	69.79	70.50	69.69
Name classifier	69.06	86.75	38.33
Image classifier	25.37	83.23	15.04
Meta classifier A: Text + Name	79.63	80.10	79.59
Meta classifier B: Text + Name + Image	81.46	81.54	81.53

As we observe, a performance increase for meta classifier A (+14.01%) and meta classifier B (+16.72%) to the best performing base classifier, we show the improvement of our cognitive approach over the single-source classifiers. It is observable, that the cognitive approach is able to outperform base classifiers by identifying their strengths and weaknesses through its capability to minimize their uncorrelated error.

Although meta classifier B outperforms meta classifier A, for legal reasons we do not include it in the final web service. In addition, the IBM Watson service is non-transparent and the source code is not publicly available. Nonetheless, we can demonstrate how any another classifier can easily be included in the meta classification.

Illustrative scenario

Applying the developed prototypical front end of the web service already allows us to gain insight on the process of classification and its results. As stated in section 5.5, our web service is limited to the first meta classifier that is based on text and name classification.

With this web service we can classify any given Twitter user of which we know he/she uses German in twitter postings on their gender, as is exemplarily shown in Fig. 5. To enable the reader to test the web service, and to demonstrate the automated approach, we prototypically make single user classifications publicly available at http://gender-prediction.science.

In order to show possible applications of the artifact, we analyzed a sample set of user-generated tweets from the domain of e-mobility. We chose the domain of e-mobility—as defined in Scheurenbrand et al. (2015)—since it is rich in micro blog data, topical, and important to the ecological and commercial future (European Commission 2017). We collected tweets from May 2015 to May 2016 by utilizing the official Twitter Streaming API with a keyword-based search. The keywords[2] were determined by an expert workshop[3] and the best-selling electric vehicles in Germany (Kraftfahrt-Bundesamt 2014). The collected ~1 million tweets were rigorously reduced to ensure only user-generated and German-speaking tweets without duplicates were regarded. That left us with 5143 relevant tweets written by 2544 individual users, to which we automatically applied the presented web service of gender classification (Fig. 6).

Although the total shift of classification between the text and meta classifier may be small (the number of female users deviate by only 29), the individually different meta classification still might be more accurate. The name classifier fails to classify 1045 profiles, possibly due to the names of those users not being retrievable. However, as discussed previously, if the name classifier classifies a profile according a gender, the result is more likely to be correct. In this case we can expect the meta classification to be more stable than a single-source prediction.

Since our artifact performs without any errors in a short period of time (80 ms on average per profile), we show its' feasibility for large-scale applications. With the presented artifact, such analyses can easily be performed to deliver interesting insights—in this case the relatively low share of female Twitter users posting about e-mobility. With such knowledge about the demographics of a certain domain, service, or product, marketing efforts can be customized. As Fischer and Arnold (1994) state, the consumer behavior of men and women differs. For the case of e-mobility, this could mean one should design product placement to appeal more to male audiences, or one could actively engage in measures to address more female customers.

Besides analyzing large user groups automatically, it is also possible to get insight from one individual profile that got classified. Additionally, if service providers know the Twitter account of a service consumer, they can derive knowledge about that consumer's gender and tailor their service specifically to such users. The developed artifact can easily be integrated using the API.

Expert workshop

To gain first insights on the practical suitability of the developed gender prediction service, we conducted an expert workshop with a large German utility provider whose annual revenue is more than EUR 45 billion. The company needs to gain insight on the demographics of social media users in their field on a regular basis. The workshop participants had different roles in different divisions within the company. In all, 7 employees took part, among them a product and innovation manager, a senior project manager, the head of a competence center, a group leader, a user researcher, and a R&D project manager.

To evaluate the prototype, each participant used the artifact for up to 10 min. Then, they filled out a short questionnaire regarding their experiences with using it, and we openly discussed positive and negative feedback. In conclusion, the participants liked the high accuracy level ("true positives") of the artifact and its ability to profile gender demographics of large numbers of Twitter users with a single click. However, they would like to see additional profiling capabilities

[2] Keywords (alphabetical, case-insensitive): bmw i3; e-tankstelle; eauto; ecar; egolf; electric mobility; electric vehicle; elektroauto; elektrofahrzeug; elektromobilitaet; elektromobilität; e-mobility; emobility; eup; fortwo electric drive; ladesaeule; ladesäule; miev; nissan leaf; opel ampera; peugeot ion; renault zoe; tesla model s.

[3] The expert workshop took place in October 2014 with four experts from the domain of e-mobility, as well as two research associates. The experts were identified by recent publications and activities in publicly-funded e-mobility research projects.

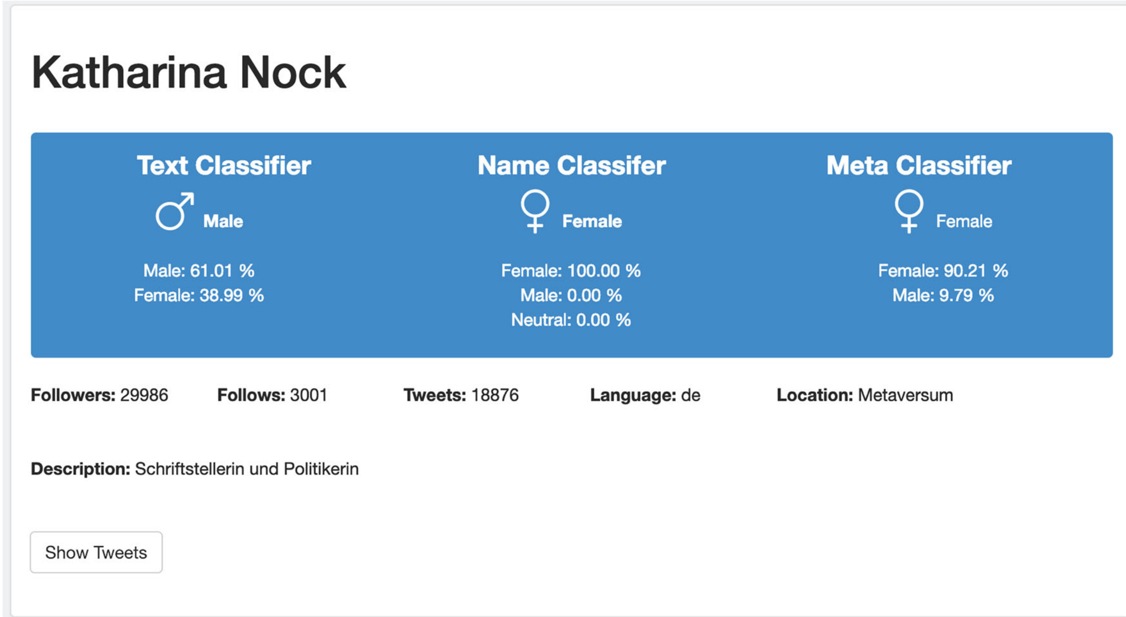


Fig. 5 Exemplary classification of a German-speaking Twitter user

implemented, especially ones that detect age, as well as the big five personality factors.

Discussion and conclusion

In this work, we have designed a dynamic cognitive method for combining heterogeneous data sources and prediction techniques for customer profiling. Based on this method, we implemented and deployed an artifact that is able to automatically predict the gender of German-speaking Twitter users on a large scale, based on their names, tweets, and profile pictures.

Having evaluated the artifact in different dimensions, we can answer our research question: *How can we determine customer characteristics like gender based on multiple heterogeneous sets of data and different prediction techniques.* By employing a cognitive approach founded on a meta learning technique, we can create a combined classification based on user-written text, profile images, and user names. We evaluated the technical feasibility and performance of the approach across almost 3000 user profiles and achieved convincing results with F_1-scores of over 80%. Additionally, we could see a steady increase of the cognitive classifier in comparison to all base classifiers. We evaluated the approach in a technical experiment across almost 3000 user profiles, gaining convincing results with F_1-scores of over 80%. Thus, we have succeeded in designing cognitive classifiers that are able to utilize multiple customized classifiers for different information sources. Further, we developed two instantiations of cognitive classifiers, one that aggregates a text and name classifier and one that additionally considers a third-party image classifier. Our findings show that by integrating different base classifiers, we can significantly increase the overall prediction performance.

Additionally, we show the utility of gender classification on German Twitter data in an illustrative scenario in the field

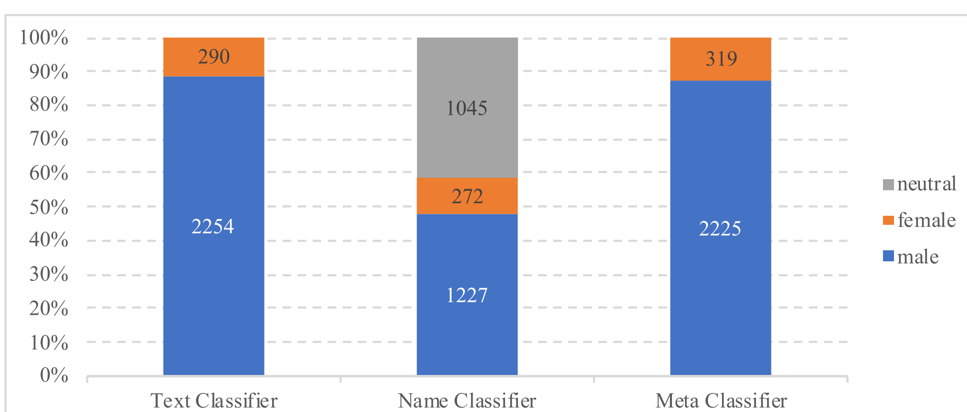


Fig. 6 Gender classification results of text, name, and meta classifier for German e-mobility tweets, from May 2015 to May 2016

of e-mobility. The approach is able to segment users into female and male groups on a large scale. A web service architecture enables us to fully automate the approach and make it freely available via an API. To draw further on first insights about the practical utility of a gender classifier for German-speaking Twitter users, we performed an expert workshop. As an outcome, users appreciated the artifact's high accuracy level, but expressed a need for more capabilities regarding user classification (e.g., age prediction).

Finally, we have to acknowledge a number of limitations. First, the data set used to train the classifiers is still small compared to the universe of tweets. Thus, we can expect the classifier performance to further improve with a larger-scale analysis. Second, our true gender values are based on the qualified labelling of a researcher, not on the objectively verified true gender of tweet authors. Thus, we still suffer a degree of bias. A sampling of tweets via identified authors may be a valuable alternative; however, that poses another bias threat in that such tweets might not be representative of the overall author's universe. Regarding the performed evaluation, the artifact needs further assessment in future research regarding its flexibility to combine other classifiers, as well as its capability for developers and data scientists to include it in larger applications. For now, we have focused solely on a narrow set of evaluations in an e-mobility context, and assessed the technical aspects of the approach. The expert workshop is a first step in the direction of extending our artifact's applicability, and its results already represent a starting point for future research endeavors.

Besides remedying the limitations above, a number of additionally promising research opportunities arise from this DSR project: First, our meta classification is built to imitate and automate the cognitive human perception and intelligence process based on information processing theory. A deeper understanding of the criteria used in human cognition would allow us to target specific additional base classifiers. Second, a broader coverage of author profiles should be reached by applying similar approaches to other author characteristics such as personality traits, education level, nationality, or profession. Although gender represents a strong consumer characteristic, analyzing personality traits (Schwartz et al. 2013) in a similar way using different sources (e.g., tweets, likes, activity) could enable even better-tailored services, as was also suggested by the expert workshop's participants. Additionally, the application to different languages is only preliminary and future research needs to address the challenge of how different languages mark personal features. Third, from a methodological point of view, we have not considered the quality of the data from the base classifiers for our meta classification yet. Fourth, integrating a gender (or broader demographic) detection model into other fields of study, such as social network analysis, topic mining, social media dynamics (Cranshaw et al. 2012), as well as user and consumer behavior, should bear fruitful results.

Lastly, the projectability of the designed solution needs to be expanded. Especially, the method of using meta machine learning in a cognitive paradigm to perform analyses based on distributed, heterogeneous data sources could greatly impact the smart service context, where data is produced in different entities. Thus, future research needs to focus on a generalization of the meta method itself in a wider context to examine the possibilities it has for a smart service system comprised of different data sources.

Still, our work in its current form should already provide valuable contributions to the body of knowledge. For example, the conceptual notion of mimicking human cognition processes by base and meta classifiers in machine learning models might have an impact in many IS areas. Our developed model not only shows the feasibility of the approach for gender detection, but it is also extendable in two ways: on the one hand, additional classifiers could solve the gender detection problem ("input-oriented"), on the other hand, additional elicitation objectives could be pursued ("output-oriented").

Cognitive classifiers allow providers as well as consumers to (re)combine existing smart services that aim to predict insights. In contrast, to develop holistic analytics services capable of analyzing heterogeneous data sources, this modular approach enables the combination of encapsulated, highly customized elementary services. This way, every analytics service can fully exploit its individual strengths using a superordinate cognitive classifier to identify relevant features and make a combined prediction.

Not only in the field of customer profiling in social media, but in every setting with multiple data sources, a cognitive architecture as shown in this work could enable better performances. However, we need more research regarding the opportunities for cognitive computing in smart service systems (Hirt and Kühl 2018).

Managerial implications and avenues to finding commercial value are evident: We enable detecting the gender of a German-speaking Twitter user in a fully automated way. Thus, any application that monitors certain topics on Twitter could benefit from enriched gender information, e.g., opinion- or need-mining, as well as sentiment analyses. In addition, other types of platforms could also be enabled to personalize their users' experience by reacting to gender, e.g. using chat bots (Jenkins et al. 2007). Looking beyond gender detection, similarly built meta classification models could be implemented to sketch a more comprehensive picture of the unknown social media user. This will enable higher levels of customer or user "intimacy" as a key strategic lever, particularly for service providers (Habryn 2012).

Overall, we are confident that our extendable meta classifier approach will assist in better eliciting author characteristics and, therefore, will enhance customer understanding. Thus, we hope to contribute a thorough method for drawing insight on customers from micro blog data and to turn it into business value.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: a psycho-lexical study. *Psychological Monographs, 47*, 171–220.
Alowibdi, J. S., Buy, U. a., & Yu, P. (2013). Language independent gender classification on Twitter. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, (May), 739–743.
Argamon, S., Koppel, M., Pennebaker, J., & Schler, J. (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM, 52*(2), 119–123.
Arnold, K. A., & Bianchi, C. (2001). Relationship marketing, gender, and culture: implications for consumer behaviour. In C. G. Mary & M. L. Joan (Eds), *Advances in consumer research* (vol. 28, pp. 100–105). Valdosta: Association for Consumer Research.
Arroju, M., Hassan, A., & Farnadi, G. (2015). Age, gender and personality recognition using tweets in a multilingual setting. In *6th Conference and Labs of the Evaluation Forum (CLEF 2015): Experimental IR meets multilinguality, multimodality, and interaction, Toulouse, France*, pp. 23–31.
Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). *Multimodal fusion for multimedia analysis: A survey. Multimedia Systems* (Vol. 16). https://doi.org/10.1007/s00530-010-0182-0.
Baird, C. H., & Parasnis, G. (2011). From social media to social customer relationship management. *Strategy & Leadership, 39*, 30–37.
Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., & Yarowsky, D. (2013). Broadly improving user classification via communication-based name and location clustering on Twitter. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013), Atlanta, USA*, pp. 1010–1019.
Blair, D. C. (1979). Information retrieval, 2nd edn. *Journal of the American Society for Information Science, 30*(6), 374–375. https://doi.org/10.1002/asi.4630300621.
Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(421), 123–140.
Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on Twitter. *Association for Computational Linguistics, 146*, 1301–1309.
Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X.
Cranshaw, J., Schwartz, R., Hong, J. I. & Sadeh, N. (2012). The livehoods project: utilizing social media to understand the dynamics of a city. In *Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM'12), Dublin, Ireland, AAAI Press*, pp. 58–65.
Dietterich, T. G. (1997). Machine-learning research. *AI Magazine, 18*(4), 97. https://doi.org/10.1609/aimag.v18i4.1324.
Džeroski, S., & Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one? *Machine Learning, 54*(3), 255–273.
Estival, D., Gaustad, T., Pham, S. B., Radford, W., & Hutchinson, B. (2007). Author profiling for English emails. *10th conference of the Pacific Association for Computational Linguistics*, 263–272.
European Commission. (2017). *Reducing CO2 emissions from passenger car*. Retrieved June 21, 2018, from http://ec.europa.eu/clima/policies/transport/vehicles/cars/index_en.htm.
Fischer, E., & Arnold, S. J. (1994). Sex, gender identity, gender role attitudes, and consumer behavior. *Psychology and Marketing, 11*, 163–182.
Gama, J., & Brazdil, P. (2000). Cascade Generalization. *Machine Learning, 41*(3), 315–343.
Giraud-Carrier, C., Giraud-Carrier, C., Vilalta, R., Vilalta, R., Brazdil, P., & Brazdil, P. (2004). Introduction to the special issue on Meta-learning. *Machine Learning, 54*, 187–193.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Retrieved June 21, 2018, from http://www.deeplearningbook.org/.
Gottipati S., Qiu M., Yang L., Zhu F., & Jiang J. (2014). An integrated model for user attribute discovery: a case study on political affiliation identification. In V. S. Tseng, T. B. Ho, Z. H. Zhou, A. L. P. Chen & H. Y. Kao (Eds.), *Advances in knowledge discovery and data mining. PAKDD 2014. Lecture Notes in Computer Science* (vol. 8443). Cham: Springer.
Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science types of knowledge in design science research. *MIS Quarterly, 37*(2), 337–355.
Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems, 8*(5), 1–25.
Grimes, T. (1990). Audio-video correspondence and its role in attention and memory. *Educational Technology Research and Development, 38*(3), 15–25.
Habryn, F. (2012). *Customer intimacy analytics: leveraging operational data to assess customer knowledge and relationships and to measure their business impact*. KIT Scientific Publishing. https://doi.org/10.5445/KSP/1000028159.
Heimbach, I., Gottschlich, J., & Hinz, O. (2015). The value of user's Facebook profile data for product recommendation generation. *Electronic Markets, 25*(2), 125–138.
Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.
Hirt, R., & Kühl, N. (2018). Cognition in the era of smart service systems: Inter-organizational analytics through meta and transfer learning. In *Proceedings of the Thirty Ninth International Conference on Information Systems (ICIS), San Francisco, CA, USA, 13th–16th December 2018.*
Hsu, C., Chang, C.-C., & Lin, C.-J. (2008). A practical guide to support vector classifcation. *Bioinformatics, 1*(1), 1–15.
IBM. (2016). Watson Visual Recognition service. Retrieved October 16, 2016, from http://www.ibm.com/watson/developercloud/doc/visual-recognition/.
Ikeda, K., Hattori, G., Ono, C., Asoh, H., & Higashino, T. (2013). Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems, 51*, 35–47.
Jenkins, M.-C., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of user interaction with service oriented Chatbot systems. *Human Computer Interaction, 4552*, 76–83.
Kludas J., Bruno E., & Marchand-Maillet S. (2008). Information fusion in multimedia information retrieval. In N. Boujemaa, M. Detyniecki & A. Nürnberger (Eds.), *Adaptive multimedia retrieval: retrieval, user, and semantics. AMR 2007. Lecture Notes in Computer Science* (vol. 4918). Berlin, Heidelberg: Springer.
Kraftfahrt-Bundesamt. (2014). *Anzahl der Neuzulassungen von Elektroautos im Zeitraum von 2011 bis 2014*.
Kuechler, W., & Vaishnavi, V. (2012). A framework for theory development in design science research: multiple perspectives. *Journal of the Association for Information Systems, 13*(6), 395–423.
Kühl, N., Scheurenbrand, J., & Satzger, G. (2016). Needmining: Identifying micro blog data containing customer needs. *Proceedings of the 24th European Conference on Information Systems*, 1–16.
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*(8), 707–710.
Liu, W., & Ruths, D. (2013). What's in a name? Using first names as features for gender inference in Twitter. *Analyzing Microtext: Papers from the 2013 AAAI Spring Symposium*, 10–16.

Lovins, J. B. (1968). Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics, 11*(June), 22–31.

Michie, E. D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification. *Technometrics, 37*(4), 459.

Miller, G. A. (1956). The magical number 7, plus or minus 2 - some limits on our capacity for processing information. *Psychological Review, 63*, 81–97. https://doi.org/10.1037/h0043158.

Modha, D. S., Ananthanarayanan, R., Esser, S. K., Ndirango, A., Sherbondy, A. J., & Singh, R. (2011). Cognitive computing. *Communications of the ACM, 54*(8), 62.

Narr, S., Hulfenhaus, M., & Albayrak, S. (2012). Language-independent twitter sentiment analysis. *Knowledge discovery and machine learning (KDML), LWA*, 12–14.

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys, 33*(1), 31–88.

Neuhofer, B., Buhalis, D., & Ladkin, A. (2015). Smart technologies for personalized experiences: a case study in the hospitality domain. *Electronic Markets, 25*(3), 243–254.

Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013). How old do you think I am ?: A study of language and age in Twitter. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 8-11 July 2013, Cambridge, Massachusetts, USA*, 439–448.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77.

Peffers K., Rothenberger M., Tuunanen T., & Vaezi R. (2012). Design science research evaluation. In K. Peffers, M. Rothenberger & B. Kuechler (Eds.), *Design science research in information systems. Advances in theory and practice. DESRIST 2012. Lecture Notes in Computer Science* (vol. 7286). Berlin, Heidelberg: Springer.

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Quinlan, J. R. (2006). Bagging, boosting, and C4.5. *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, *5*(Quinlan 1993), 725–730.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in twitter. *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents - SMUC '10*, 37.

Rumelhart, D. E., & Mcclelland, J. L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: Foundations*. MIT Press: Cambridge.

Scheurenbrand, J., Engel, C., Peters, F., & Kühl, N. (2015). Holistically defining E-mobility: a modern approach to systematic literature reviews. *Karlsruhe Service Summit*, 17–27. https://doi.org/10.5445/KSP/1000045634.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One, 8*(9), e73791.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems, 25*, 1–9.

Sonnenberg, C., & Vom Brocke, J. (2012). Evaluations in the science of the artificial - reconsidering the build-evaluate pattern in design science research. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-29863-9_28.

Statista. (2016). Anzahl der monatlich aktiven Nutzer von Twitter in Deutschland in den Jahren 2014 und 2015 sowie eine Prognose für 2016 (in Millionen). Retrieved August 4, 2016, from http://de.statista.com/statistik/daten/studie/546761/umfrage/anzahl-der-monatlich-aktiven-twitter-nutzer-in-deutschland/.

Stone, M. (1977). Asymptotics for and against cross-validation. *Biometrika, 64*(1), 29–35.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.

Todorovski, L., & Džeroski, S. (2003). Combining classifiers with meta decision trees. *Machine Learning, 50*(3), 223–249.

Treacy, M., & Wiersema, F. (1993). Customer intimacy and other value disciplines customer intimacy and other value disciplines. *Harvard Business Review, 71*(9301), 84–93.

Vilalta, R., & Drissi, Y. (2002). A perspective view and survey of meta-learning. *Artificial Intelligence Review, 18*(2), 77–95. https://doi.org/10.1023/A:1019956318069.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii https://doi.org/10.2307/4132319.

Wieneke, A., & Lehrer, C. (2016). Generating and exploiting customer insights from social media data. *Electronic Markets, 26*(3), 245–268.

Zhou, G., Shen, D., Zhang, J., Su, J., & Tan, S. (2005). Recognition of protein/gene names from text using an ensemble of classifiers. *BMC Bioinformatics, 6*(1), 1.