symmetry

MDPI

Review

Analysis of Human Behavior by Mining Textual Data: Current Research Topics and Analytical Techniques

Edgar Gutierrez [1,2], Waldemar Karwowski [1], Krzysztof Fiok [1], Mohammad Reza Davahli [1,*], Tameika Liciaga [1] and Tareq Ahram [1]

[1] Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, FL 32816, USA; edgar.gutierrezfranco@ucf.edu (E.G.); wkar@ucf.edu (W.K.); fiok@ucf.edu (K.F.); tameikaliciaga22@knights.ucf.edu (T.L.); tahram@ucf.edu (T.A.)

[2] Center for Latin-American Logistics Innovation, Massachusetts Institute of Technology, Global SCALE, LOGyCA, Bogota 110111, Colombia

* Correspondence: mohammadreza.davahli@ucf.edu

Abstract: The goal of this study was to conduct a literature review of current approaches and techniques for identifying, understanding, and predicting human behaviors through mining a variety of sources of textual data with a focus on enabling classification of psychological behaviors regarding emotion, cognition, and social empathy. This review was performed using keyword searches in ISI Web of Science, Engineering Village Compendex, ProQuest Dissertations, and Google Scholar. Our findings show that, despite recent advancements in predicting human behaviors based on unstructured textual data, significant developments in data analytics systems for identification, determination of interrelationships, and prediction of human cognitive, emotional and social behaviors remain lacking.

Keywords: text mining; human behavior; sentiment analysis; physiological profiling

Citation: Gutierrez, E.; Karwowski, W.; Fiok, K.; Davahli, M.R.; Liciaga, T.; Ahram, T. Analysis of Human Behavior by Mining Textual Data: Current Research Topics and Analytical Techniques. *Symmetry* **2021**, *13*, 1276. https://doi.org/10.3390/sym13071276

Academic Editor: Leyi Wei

Received: 16 June 2021
Accepted: 12 July 2021
Published: 16 July 2021

1. Introduction

At present, the vast amount of textual data being generated from myriad sources (e.g., formal or informal reports, interviews, call logs, emails, performance documents, blogs, tweets, comments, or social media entries) is rapidly increasing [1]. Although this increase in textual data allows for large repositories to be analyzed, summarized, and deciphered, using these data to make insightful decisions has become much more challenging. Thus, in this study, we sought to explore the current approaches through which the unstructured textual data can be analyzed by extracting valuable information to support decision-making for various purposes. Consequently, we conducted a systematic literature review of the techniques and methods used to identify, understand, and predict human behaviors by mining various textual data sources.

The problem of mining textual data has received substantial attention, owing to the proliferation of social networks that allow the distribution of opinions and sharing sentiment on diverse subject matters. This literature review is focused on the methods for understanding human psychological behavior through the use of textual data. Mining textual data can provide deep insights into an individual's views, attitudes, sentiments, and emotions toward other individuals and help predict future social behaviors [2]. Such human behaviors can be identified and understood by extracting textual data with meaningful semantic properties, including metadata such as concepts, events, keywords, categories, including symmetric and asymmetric relationships. Such knowledge can facilitate improved decision-making (e.g., personnel selection and training) or intelligence analyses [3]. According to Bornstein et al. [4], human behavior is described as "the potential and expressed capacity for physical, mental, and social activity during the phases of human life." Regarding the identification of behaviors by text mining, Tausczik et al. [5] stated that "by drawing on massive amounts of

text, researchers can begin to link everyday language use with behavioral and self-reported measures of personality, social behavior, and cognitive styles." Furthermore, Pennebaker and Stone [6] classified the use of language in the following categories: emotional experience, social relationships, time orientation, and cognitive abilities.

This present study makes two main contributions. First, it focuses on identifying the main methodological approaches to understanding human psychological behavior by analyzing people's expressions through textual communication. Second, it identifies gaps in research and the characteristics necessary for the development of analytical tools and methods to predict behaviors through textual analyses. The remainder of the paper is ordered as follows: Section 2 describes the methods and criteria used for the selection of the included literature; Section 3 presents the main results of the study; Section 4 discusses the results and answers the research questions.

2. Method

This review is based on the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [7].

The following research questions forming the basis of this systematic literature review are based on the objectives of the present study, as outlined in the abstract:

- RQ1. What has been the most relevant research reported in the scientific literature for the identification of human behaviors through text mining?
- RQ2: How can current analytical techniques for the prediction of human behavior from unstructured textual data be classified?

The inclusion criteria were as follows: (a) papers written in English; (b) peer-reviewed papers; and (c) papers depicting graphs, charts, equations, and/or tables presenting text mining techniques that initially identified research focusing only on methods of psychological analysis of behavior. The exclusion criteria were as follows: (a) papers not written in English; (b) papers determined upon evaluation to be unrelated to the research questions; and (c) opinions, letters, and editorials.

A search strategy for the review was used to identify papers applicable to answering the research questions. The strategy involved defining the search space and the vetting process to be used in identifying pertinent literature. The recent and influential literature in the field of text mining, including journal articles, textbooks, proceedings, and grey literature, were important sources in this research.

With the knowledge of the subject matter and based on widely cited articles such as [2,5], we developed a list of set of keywords, which after testing in search engines, was reduced to the 15 keywords that are presented in Table 1. Subsequently, this set was used to query databases, such as EBSCOhost, Compendex, IEEE Xplore, Google Scholar, and ProQuest. This process resulted in a reduction of the core search parameters used to identify the key components affecting the prediction and understanding of human behavior via text mining. After retrieving the articles, we then carefully chose pertinent papers. The terms used in the EBSCOhost database were as follows: ("data mining"[MeSH terms] OR "text mining"[all fields]) AND (("humans"[MeSH terms] OR "humans"[all fields] AND ("behavior"[all fields] OR "behavior"[MeSH terms] OR "behavior"[all fields])) AND (("1998/01/01"[PDAT]: "2019/12/31"[PDAT]) AND "humans"[MeSH terms] AND English[lang]).

Table 1. Keywords used for searching selected databases.

Text Mining	Behavioral Markers	Cognitive Behavior
Human behavior	Linguistic markers	Social behavior
Information extraction	Lone wolf behavior	Emotional behavior
Opinion mining	Behavioral profiling	Sentiment analysis
Linguistic inquiry and word count (LIWC)	Computational linguistics	Natural language processing

To assess the risk of bias in the present study, we used the Cochrane Risk of Bias Tool [8] as a support instrument. The relevant papers were classified among different bias domains, such as sequence generation (the methods through which the data were collected), allocation concealment (whether data allocations could have been foreseen before or during collection), blinding of participants (the people who generated the text), blinding outcomes (the people who generated the text data not having knowledge of the results), incomplete outcome data (whether the papers showed completeness of the outcome in their results) and finally selective outcome reporting (whether the authors showed outcome reporting and what was found). Figure 1 depicts the number of papers in each of these categories. Most of the papers had a low risk of bias. The use of the Cochrane Tool allowed us to reduce all possible biases that could have affected the quality of the review and thus the reliability of conclusions.

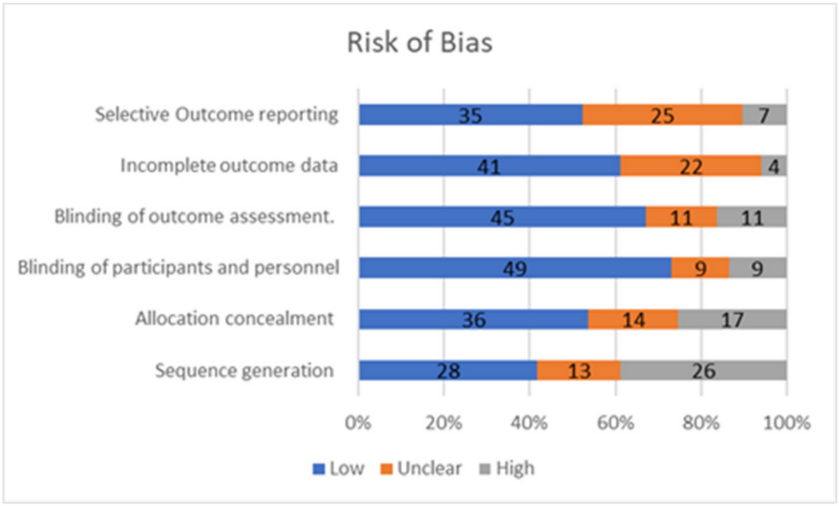

Figure 1. Assessing the risk of bias with the Cochrane collaboration's tool.

The risk of bias was evaluated using a subjective judgment (high, low, or unclear) regarding the individual elements of the domains represented in Figure 1. Once the classification was made, a percentage estimate of these judgments was obtained. On average, the different domains were approximately 58% low, 23% unclear, and 18% high risk of bias.

3. Results

To understand the evolution of the research on the prediction of human behavior on the basis of unstructured textual data, the selection procedure and the numbers of papers selected in the various stages of selection are shown in Figure 2.

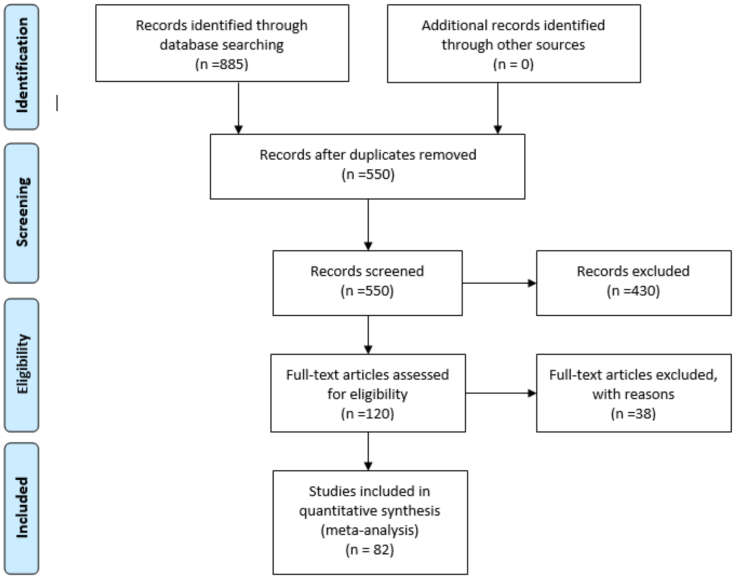

Figure 2. Flow diagram and selection process for including literature in the meta-analysis.

Then the literature review was classified into categories. After a category was identified, we proceeded to identify the main text mining approach and the main insights of each work analyzed. The characteristics of the included papers are shown in Table 2.

Table 2. Characteristics of the included papers.

Reference	Title	Research Method	Category
[2]	Sentiment Analysis and Opinion Mining	Natural language processing	Emotional
[3]	Clarifying the Linguistic Signature: Measuring Personality From Natural Speech	Document clusterization	Emotional and cognition
[5]	The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods	Natural language processing	Emotional
[6]	Words of wisdom: Language use over the life span	Natural language processing	Cognition and theory
[9]	Mobile phones as medical devices in mental disorder treatment: an overview	Natural language processing	Emotional
[10]	Opinion Mining for text classification	Document classification	Emotional
[11]	A new significant area: Emotion detection in E-learning using opinion mining techniques	Natural language processing	Emotional
[12]	Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena	Natural language processing	Emotional
[12]	Twitter mood predicts the stock market	Information extraction	Social
[13]	Sentiment classification based on supervised latent n-gram analysis	Document classification	Emotional
[14]	Supporting disease insight through data analysis: Refinements of the monarca self-assessment system	Natural language processing	Emotional
[15]	Smartphone-Based Recognition of States and State Changes in Bipolar Disorder Patients	Natural language processing	Emotional
[16]	Mining and summarizing customer reviews	Natural language processing	Emotional
[17]	Sentiment analysis with long short-term memory networks	Natural language processing	Emotional
[18]	Thumbs up? Sentiment Classification using Machine Learning Techniques	Natural language processing	Emotional
[19]	A faceted characterization of the opinion mining landscape	Natural language processing	Emotional
[20]	A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives	Web mining	Emotional
[21]	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews	Document clusterization	Emotional
[22]	Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization	Web mining	Emotional

Table 2. *Cont.*

Reference	Title	Research Method	Category
[23]	Product aspect ranking using sentiment analysis: A survey	Web mining	Emotional
[24]	Opinion mining and sentimental analysis approaches: A survey	Natural language processing	Emotional
[25]	Opinion mining and sentiment analysis	Natural language processing	Emotional
[26]	Sentiment analysis and opinion mining: A survey	Natural language processing	Emotional
[27]	#MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter	Web mining	Emotional
[28]	Understanding customers using Facebook Pages: Data mining users feedback using text analysis	Natural language processing	Emotional and social
[29]	Using causal models in heterogeneous information fusion to detect terrorists	Natural language processing	Emotional and social
[30]	INSiGHT: A system for detecting radicalization trajectories in large heterogeneous graphs	Natural language processing	Emotional and social
[31]	Harvesting and analysis of weak signals for detecting lone wolf terrorists	Natural language processing	Emotional and social
[32]	Detecting Linguistic Markers for Radical Violence in Social Media	Natural language processing	Emotional and social
[33]	Personality and language: The projection and perception of personality in computer-mediated communication	Natural language processing	Emotional and social
[34]	Hierarchical Sentiment Analysis Model for Automatic Review Classification for E-commerce Users	Natural language processing	Emotional and social
[35]	Assessing Bipolar Episodes Using Speech Cues Derived from Phone Calls	Information retrieval	Emotional and social
[36]	Using behavioral indicators to help detect potential violent acts	Natural language processing	Emotional and social
[37]	Sentiment analysis: capturing favorability using natural language processing	Natural language processing	Emotional and social
[38]	Identifying topical influencers on twitter based on user behavior and network topology	Natural language processing	Emotional and social
[39]	Language-based personality: a new approach to personality in a digital world	Natural language processing	Emotional and cognition
[40]	The efficacy of SMS text messages to compensate for the effects of cognitive impairments in schizophrenia	Natural language processing	Emotional and cognition
[41]	Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis	Natural language processing	Emotional and theory
[42]	The Development and Psychometric Properties of LIWC2015	Natural language processing	Emotional and theory
[43]	The Role of Text Pre-processing in Sentiment Analysis	Document clusterization	Emotional and theory
[44]	Text sentiment analysis based on long short-term memory	Information extraction	Emotional and theory
[45]	Analysing the presence of school-shooting related communities at social media sites	Web mining	Social
[46]	The State of the Art 2015: A literature review of social media intelligence capabilities for counter-terrorism	Natural language processing	Social
[47]	Opinion Mining platform for Intelligence in business	Natural language processing	Social
[48]	Mining the peanut gallery: opinion extraction and semantic classification of product reviews	Information extraction	Social
[49]	Analysis of Online Social Networks Posts to Investigate Suspects Using SEMCON	Information retrieval	Social
[50]	A mutually beneficial integration of data mining and information extraction	Information extraction	Social
[51]	Product aspect ranking and its applications	Information extraction	Social
[52]	Opinion zoom: a modular tool to explore tourism opinions on the web	Natural language processing	Social
[53]	Language and interaction: applying sociolinguistics to social network analysis	Information retrieval	Social
[54]	A Generic Architecture for a Social Network Monitoring and Analysis System	Document classification	Social

Table 2. *Cont.*

Reference	Title	Research Method	Category
[55]	Mind mapping: Using everyday language to explore social & psychological processes	Information extraction	Social and cognition
[56]	Information extraction: Distilling structured data from unstructured text	Information extraction	Social and theory
[57]	Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques	Information extraction	Social and theory
[58]	Affinity: A System for Latent User Similarity Comparison on Texting Data	Natural language processing	Cognition
[59]	Automated computer-based feedback in expressive writing	Natural language processing	Cognition
[60]	Intelligence analysis: Behavioral and social scientific foundations	Document clusterization	Cognition
[61]	Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language	Natural language processing	Cognition and theory
[62]	Text Mining—Knowledge extraction from unstructured textual data	Information extraction	Cognition and theory
[63]	Opinion Mining and Information Fusion: A survey	Natural language processing	Others/theory
[64]	Text mining and analysis: Practical methods, examples, and case studies using SAS	Information extraction	Others/theory
[64]	Analysis of unstructured data: Applications of text analytics and sentiment mining	Information extraction	Others/theory
[65]	Web mining research: a survey	Web mining	Others/theory
[66]	Foundations of statistical natural language processing	Natural language processing	Others/theory
[67]	Using maximum entropy for text classification	Document classification	Others/theory
[68]	Sentiment miner: a prototype for sentiment analysis of unstructured data and text	Natural language processing	Others/theory
[69]	Text mining: predictive methods for analyzing unstructured information	Natural language processing	Others/theory
[70]	Development of a Scale to Measure Problem Use of Short Message Service: The SMS Problem Use Diagnostic Questionnaire	Natural language processing	Others/theory
[71]	An introduction to text mining	NA	Others/theory
[72]	Text mining. Applications and Theory	NA	Others/theory
[73]	Text mining: challenges and future directions	Unstructured text mining objective	Others/theory
[74]	Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes	Information retrieval	Others/theory
[75]	Document-level sentiment classification: an empirical comparison between SVM and ANN	Document classification	Others/theory
[76]	How to conduct behavioral research over the Internet: a beginner's guide to HTML and CGI/Perl	Web mining	Others/theory
[77]	Emotional Text Mining: Customer profiling in brand management	Natural language processing	Emotional
[78]	Text-based emotion detection: Advances, challenges, and opportunities	Information extraction	Emotional
[79]	Opinion mining and emotion recognition applied to learning environments	Information extraction	Emotional
[80]	Emotion correlation mining through deep learning models on natural language text	Natural language processing	Emotional
[81]	Using Opinion Mining as an educational analytic: An integrated strategy for the analysis of students' feedback	Information extraction	Emotional
[82]	Text Analytics of Customers on Twitter: Brand Sentiments in Customer Support	Natural language processing	Emotional
[83]	A Suicide Prediction System Based on Twitter Tweets Using Sentiment Analysis and Machine Learning	Natural language processing	Emotional

Table 2. *Cont.*

Reference	Title	Research Method	Category
[84]	Study of coronavirus impact on parisian population from april to june using twitter and text mining approach	Information extraction	Emotional
[85]	Characterizing Twitter Interaction during COVID-19 pandemic using Complex Networks and Text Mining	Natural language processing	Emotional
[86]	Estimating Industry 4.0 impact on job profiles and skills using text mining	Information extraction	Emotional
[87]	When Emotions Rule Knowledge: A Text-Mining Study of Emotions in Knowledge Management Research	Natural language processing	Emotional
[88]	Determining a Person's Suicide Risk by Voting on the Short-Term History of Tweets for the CLPsych 2021 Shared Task	Natural language processing	Emotional
[89]	Identification and Prediction of Human Behavior through Mining of Unstructured Textual Data.	Natural language processing	Others/theory
[90]	An Exploration about the Last Mile Logistic Efficiency in Indian E-Commerce Sector- A Text Mining Approach	Natural language processing	Cognition
[91]	Saving Human Lives: What Complexity Science and Information Systems can Contribute	Natural language processing	Others/theory

The articles, which were included on the basis of the publication date, relevance, and content, were classified into three main behavioral categories: emotional, social, and cognitive. For each paper, we identified objectives, algorithms/techniques, models of computational aims, and main applications. Each of the included papers was subclassified according to the categories shown in Figure 3. The present review retrieved a combination of 82 relevant papers, which are identified by subcategory.

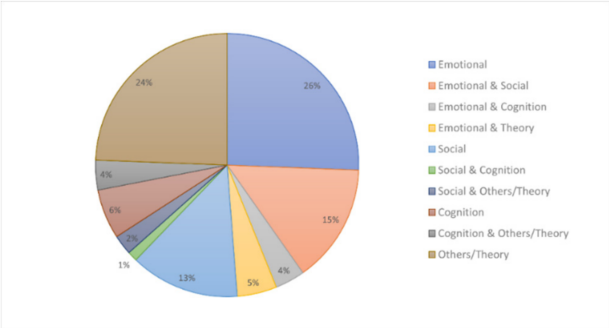

Figure 3. Sub classification of publications.

Figure 4 represents the percentages for each of the analyzed text mining approaches. The results indicated that more than 50% of the reviewed literature was completed by using natural language processing (NLP), which was one of the strongest approaches. This method was followed by information extraction (15%); document classification and clusterization (13%); and web mining, information retrieval, and summarization (20% combined).

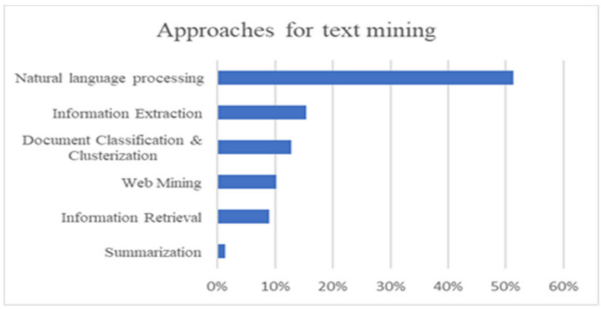

Figure 4. The literature according to the text mining behavior analysis approach.

We also provide a map of the co-occurrence of the "text mining" term in the title and abstract in Figure 5. We used VOSviewer software (https://www.vosviewer.com/, accessed on 10 June 2021) to map the bibliometric data as a network and develop keyword co-occurrence maps.

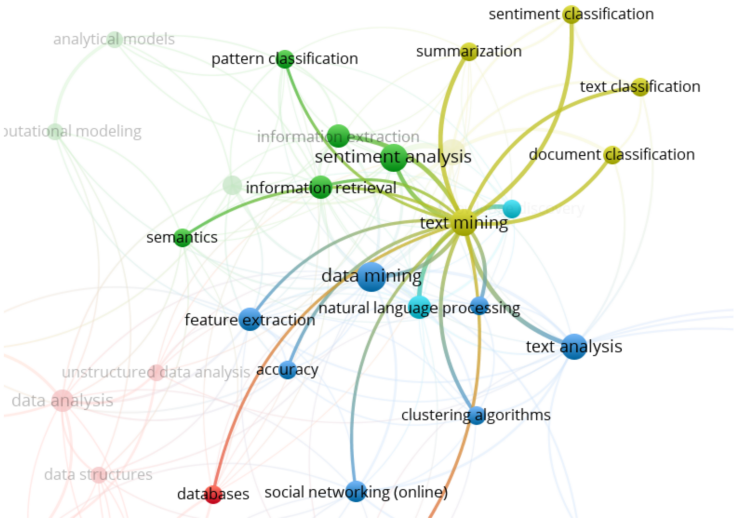

Figure 5. The map of the co-occurrence of the "text mining" term.

In this figure, links between the "text mining" node and other nodes show the co-occurrence of the terms, and their sizes indicate the frequency of occurrence.

Efficient analyses of unstructured information about people make continuous monitoring of a given individual's performance or learning effectiveness very difficult. This aspect explains the increased need for automated techniques to analyze and apply tags of human behavior signatures and human performance. Such a task, performed by a human expert, might require weeks or months when performed manually, particularly when the analyzed results are biased because of emotional, relational, and other environmental factors. The present study discusses the state of the art in the applications of behavior analysis from the mining of unstructured texts to assess attitudes, emotions, or performance at the individual level.

4. Discussion

In this section, the included papers are discussed according to their research methods and categories of human behavior.

4.1. Research Methods and Classification of Approaches and Techniques for Text Mining

Recently, studies highlighted multiple applications of text mining in a variety of forms [20–89]. Select examples of such applications, which used various data mining algorithms and methods, and are relevant to the goals of this research, are briefly reviewed below. Huang et al. [92] described a technique focusing on processing large quantities of unstructured intelligence information and used mathematical analyses for unstructured data in emergency systems. This approach focuses on visual computing, cognitive modeling, and NLP to build an intelligence information service platform. Bakshi [93] reported an unstructured data-mining method based on the data processing paradigm MapReduce. MapReduce is a programming model for processing large quantities of unstructured data and generating datasets by using a parallel, distributed algorithm on a cluster to extract sentiment information and other meaningful social and relational data. Weerdt et al. [74] proposed a method based on a combination of text mining and trace-clustering for incident reporting and predicting possible modes of action.

Wu et al. [28] described an analytical process for interpreting dialogue on social network platforms, such as Facebook or Twitter group pages. This technique focuses on critical elements of posted internet content and uses textual analyses to apply knowledge and information processing to extract key phrases from conversations. Shahbaz et al. [68] proposed a method based on the software Sentiment Miner. Sentiment Miner filters text files, such as interviews, for "opinion mining" at the sentence level by using NLP techniques and opinion mining (OM) algorithms. This approach filters users or groups of users that are relevant to the custom search query in question via an analysis of unstructured data. Chakraborty et al. [64] described an approach to analyzing unstructured textual data to extract user insights from an extensive collection of documents by using "Text-Miner" and "SAS Sentiment Analysis", which are based on artificial neural networks and use a regression model to predict target variables such as descriptive classifications of behavioral models.

Every day, humans generate vast amounts of textual information, which is stored electronically. This information is of great relevance because it includes information on moods, opinions, behavioral trends, and preferences. An example of this value is in text mining for commercial uses, such as consumer identification and purchase preferences for products and services [94]. Text mining includes the application of different methodological approaches and algorithms [71–95].

Information extraction: Information extraction enables the automatic extraction of structured information from structured or semi-structured documents or databases, which can later be used to perform calculations. It has various applications, such as consumer care and personal information management [96].

Information retrieval: Information retrieval is a process enabling organizing and retrieving information at different levels of storage, whether from metadata, images, documents, or information within a specific document or query. The process is performed according to the user's needs via a query, after which the information is indexed and filtered, and the relevant data are subsequently extracted and returned to the user for a corresponding purpose.

Document classification and clusterization: In many practical applications, data must be organized by groups according to content to facilitate information handling. The process of organizing the information often uses supervised document classification models or methods created for unsupervised clusterization, such as NMF [97] or LDA [98]. As discussed by Kowsari et al. [99], text classification is a significant challenge in many domains and fields of application. Clustering, in contrast, allows for the same organization to be performed according to groups, but in an unsupervised manner, by using clusters of the same types of data information that are not initially labeled, thereby decreasing the possibility that the data that are incorrectly assigned [100].

Summarization: Summarization creates a short representation of the original text, thus allowing readers to grasp the entire information on a general level. It can take the form of identification of key sentences in a document, which are subsequently presented as a summary with considerable relevance for the user. In contrast, in the abstract form, the summarization tool attempts to understand the information described in the text to later extract the relevant idea and present it to the user [101].

Natural language processing: NLP is the technique through which data and texts written in natural language are processed. The importance of this technique lies in the difficulties in computing systems' understanding of how humans execute communication, because communication requires not only the use of words and symbols but also accentuation, expression, and their respective meanings within a given context, which in many cases can become abstract [102]. The field of applying NLP ranges from applications such as Grammarly, Siri, or Alexa to sectors such as the economic and health sectors (e.g., the IBM supercomputer Watson).

Web mining: In web mining, techniques and algorithms are used to obtain relevant information found on the internet. It can focus on user activities, such as visited websites,

the links that users click during browsing, or the documents consulted. However, the data gathered from the internet are most often used to extract relevant information to find patterns that enable, for example, diagnosis and prediction of trends of consumption. These data also include analyses of a given user's reaction to receiving a product or service and are used to create personalized marketing strategies, thereby enabling business and closing processes to be conducted effectively [103].

OM or sentiment analysis: Sentiment analysis refers to a wide range of machine learning methods, including computational classification techniques that focus on linguistics, NLP, and textual analytics. These techniques allow for the identification of the attitudes and opinions of individuals toward a specific issue on the basis of metrics representing the characteristics of the problems of interest. Such attitudes and opinions can be evaluated via human judgments, expert evaluations of the emotional states of individuals, or the intended meaning of communication in a specific context.

One method for unstructured text mining is sentiment analysis [24,26,69], also known as OM in the context of NLP. Opinions are subjective descriptions, appraisals, and feelings expressed by individuals, whereas sentiment analysis focuses on the algorithmic extraction of various attributes of expressed opinions, such as polarity, subject matter, and ownership. Sentiment mining focuses on the computational analysis of the "subjective" information typically contained in textual sources, such as reports, reviews, blogs, posts, and comments [34]. As noted by Shahbaz et al. [68], the main aim of sentiment analysis is to categorize text at the document or sentence level and to provide information about whether the analyzed text expresses a positive, negative, or neutral sentiment toward a given topic.

Turney [21] and Pang et al. [18] discussed various approaches for categorizing the polarity of individual opinions and introduced the feature-based analysis model, similarly to the sentiment analysis used by Hu and Liu [16]. This model can be used to determine views expressed by individuals about specific features (e.g., product features) as well as user attitudes (e.g., positive, neutral, or negative) regarding specific features or aspects of an issue through the use of certain words and sentences. Sentiment analysis can be applied at the text, sentence, or sub-sentence level and can be optimized to offer a fine-grained analysis on a five-point Likert scale or a five-star rating, which is used to detect the emotion of the entity expressing the opinion in aspect-based, intent, and multilingual analyses. With the exponential increase in unstructured data, the algorithmic extraction of sentiment from expressed personal views significantly improves behavioral insights and makes behavioral analysis more efficient.

4.2. Human Behavior

This literature review proposes three main categories to classify human behavior in the context of text mining: cognitive, emotional, and social behaviors. Most of the literature conveys how textual data are analyzed to understand the activities, mental skills, and social interactions among people, with the goal of identifying emotional, social, and cognitive behaviors, whose characteristics are depicted in Figure 6 [55].

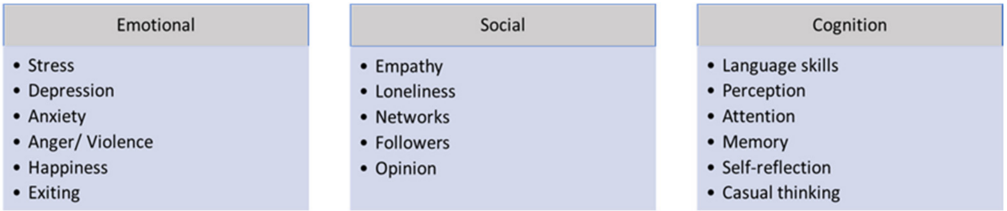

Figure 6. Classification of psychological behaviors.

Emotional behavior is correlated with mental health issues (e.g., stress, depression, anger, or violence), and the monitoring and treatment of mental disorders can be achieved by extracting textual data from communication devices. Several studies have explicitly

shown that assumptions can be made about a given person's current mood by analyzing variations in mobile usage patterns, texting, and calling [9]. Moreover, studies by Tausczik et al. [5] and Rutland et al. [70] described machine-learning methods to analyze the content of messages sent by short message service (SMS) to scan words from texts and link them to psychologically meaningful terms that can be used to asses emotions and changes in mood. In addition, audio data from mobile calls can be analyzed and translated to text, and the person's mood can be extracted to detect emotional signatures [35].

Social behavior is associated with issues of social interaction, such as empathy or loneliness. Social networks, such as Twitter, LinkedIn, or Facebook, were used to study human social behavior. Textual comments were analyzed to identify sentiments to extract information about attitudes, social activity, and interactions with other users [27]. For example, the number of ongoing or outgoing comments or conversations can reflect the current mood of a given individual. In addition, people with mental health conditions tend to increase the number of texts sent during maniac episodes, whereas low levels of texts sent can be correlated with depression [14]. In terms of societies, governments, and leader actions, interesting advances using hybrid approaches such as complexity science, symmetry, and information systems (text) presented by Helbing et al. [91] demonstrate that they can contribute to areas such as understanding geopolitical tensions by analyzing an extensive data set from newspaper articles. The published news was used to search the text of the article for mentions of a given country along with a set of keywords typically associated with tensions (for example, crisis, conflict, antagonism, clash, contention, discord, fight, attack, combat) and have predictions about the subsequent actions.

Cognitive behavior is associated with the performance of mental processes such as thinking and casual reasoning [55]. The information expressed as textual data can be used to monitor an individual's skills in different activities. Some authors divide cognitive functions into categories (e.g., perception, attention, memory, language skills, and executive functioning), which can be monitored by assessing performance in specific tasks within these categories. For example, textual analytics was used in SMS text messages from people with schizophrenia to help identify cognitive impairment [104].

4.2.1. Emotional Category

In this category, the included papers focused on using smartphones, written information, opinion mining, and customer feedback. Gravenhorst et al. [9] recognize smartphones as a promising technology for use in the treatment of mental disorders through the implementation of sensor devices to monitor illnesses. By using human–computer interfaces to support therapy and by collecting data from patients' daily lives, smartphones can be beneficial for treating people with mental disorders. Grünerbl et al. [15] explored the use of mobile phones to recognize depressive and manic states in people with bipolar disorder. This sensor-based smartphone system can support the treatment of patients with bipolar disorder as a supplementary tool for health care professionals [15]. Muaremi et al. [35] demonstrated the applicability of phone calls to assess episodes of bipolar disorder in patients. Statistics were extracted from various phone call conversations by using speech cues, and different features, social signals, and emotional properties were identified [35]. Li and Qian [44] identified how long-term memory helps classify information by analyzing different emotions in texts. This method helps classify different sentences with a corresponding emotion and may be used to project possible trends in preferences [44].

Wang et al. [80] used emotion evolution law for emotion analysis. This method evaluates natural language text from web news by using one-step and limited-step shifts as well as path transfer; it was validated on a data set of titles, bodies, and comments from news articles. This method can identify feelings such as love-anger, sadness-anger, and joy, thus providing insight into applications regarding affective interaction in network public sentiment, social media communication, and human–computer interaction. Swain et al. [83] proposed a method for detecting suicide ideation by using sentiment analysis from tweets via supervised learning. By using Python language modules and machine learning models

for opinion mining, the research using this method suggests that machine sentiment analysis can aid in timely detection and act as an alert system for suicidal tendencies. Similar work was recently presented by Bayram and Benhiba in [88], where with machine learning techniques, it was possible to identify a person's suicide risk based on the short-term history of their tweets. Fareri et al. [86], in 2020, focused on the development of a data-driven approach using text mining techniques to analyze job profiles and quantify the readiness of employees of a large firm to adopt the Industry 4.0 paradigm. This approach provides a framework for estimating the Industry 4.0 readiness of enterprises.

Mahendran et al. [10] proposed classifying written information as positive, negative, or neutral to efficiently study raw data by using traditional approaches such as Bag of Words, Naïve Bayes classifier, and frequency distribution. Tausczik et al. [5] used the computerized text analysis program Linguistic Inquiry and Word Count (LIWC) to determine the physiological meaning of textual information. In this program, words are categorized into different psychological classes to assess peoples' thought processes, emotional states, intentions, and motivations [5]. Turney [21] categorized data according to an analysis of the meanings of different words by using algorithms. For example, positive reviews (thumbs up) are determined if the review contains positive words, whereas negative reviews are determined by negative words (thumbs down). Nasukawa and Yi [37] applied a semantic analysis and achieved 70–95% precision in relating sentiments to positive or negative words in text documents from web pages and articles. Extracting information by using NLP can help determine sentiments expressed online [37]. Thakur and Han [105] presented an attractive approach for analyzing the acceptance of interaction with virtual assistants throughout different interactive devices with sentiment analysis using Natural Language Processing to explore the views, expressions, and beliefs expressed by older adults.

Pennebaker et al. [42] studied the software LIWC, which processes textual information to capture the beliefs, preferences, and sentiments of people expressed in words. This study provides evidence that the words that people use have psychological value [42]. Emotions play a critical role in the studies of human knowledge and behavior. These emotions can be determined by the environmental events of the individual or by their cognitive abilities and social skills. Knowledge management (KM) research considers them from specific angles, and, to date, a comprehensive understanding of the emotions that dominate KM and their prediction has been lacking. To offer a holistic view, this study investigated the presence of emotions in knowledge management publications by applying sentiment analysis [87].

Liu [2] introduced different aspects of sentiment analysis and opinion mining because these two fields have become the most critical approaches in analyzing people's opinions, sentiments, emotions, and attitudes through the collection of textual language. Miedema [17] explored how sentiment classification can be used to arrange documents according to sentiments. This method was used to organize feelings gathered through movie reviews for the long short-term memory. Bo Pang et al. [18] indicated that some machine learning techniques have not performed correctly in classifying texts by sentiment. This aspect has become a concern because it makes sentiment analyses more challenging. Othman et al. [24] explored approaches for opinion mining and sentiment analysis to gather and analyze information about the opinions of the public. Machine learning can help collect the responses posted on different social media platforms so that data can be used for various purposes in the industry. Acheampong et al. [78] focused on sentiment analysis through emotional detection via text mining. With the ease of sourcing for data, the analysis of text mining has led to different approaches in the design of text-based emotional detection systems as well as different proposals regarding the concepts of contributions, approaches used, datasets used, results obtained, and strengths and weaknesses.

Vinodhini and Chandrasekaran [26] established that sentiment analysis and opinion mining can help predict future behavioral trends by elucidating the preferences of customers according to what they write. This capability is valuable for economic and marketing studies. Usability in logistics and supply chain management was used recently to examine customer perceptions of companies' services. For example, Siby et al., in [90],

presented an interesting application in last-mile logistics. The research used customer reviews about their delivery experience regarding quality, service quality, product return, refund policy, information sharing issues, etc. This work recommended suggestions for redesigning processes related to last-mile logistics by introducing artificial intelligence technology. Pang and Lee [25] compared traditional analyses and sentiment-aware applications that process information about the sentiments and opinions of people. Different techniques, benchmarking, future work, and resources were also studied. Salloum et al. [20] proposed a different classification system for the different aspects of opinion mining because the challenges of correctly detecting the meanings and interpretations of different opinions can complicate opinion mining (i.e., an understanding of the domain-specific opinion is required) [20].

Greco and Polli [77] focused on the abundance and use of textual data as a source of valuable information regarding opinions and feelings and discussed the use of emotional text mining in brand management. This method is used to profile social media users' representations and sentiments about a topic by extracting information from a collection of texts such as Twitter. Raeesi Vanan [82] performed a study in which 3 million inbound tweets and outbound brand responses (tweets) were collected for brand sentiment analysis. Steps of CRIP-DM were used as a reference guide for business and data understanding, preparation, text mining, validation, and discussion of its contributions. The analytical conclusions regarding the sentiment trends were that the sentiments of customers toward a brand are significantly correlated with the brand's proper response to a brand community over social media as well as providing customers with a deep feeling of reciprocal understanding of needs.

Pang and Lee [25] presented the importance of opinion mining and sentiment analysis, which has led to the development of several techniques and machines to gather and process information about the opinions and moods of people. The challenge is to seek better approaches to sentiment-aware applications. Haddi et al. [43] used support-vector machines (SVM) to explore the importance of text pre-processing in sentiment analysis because understanding the relevance of product opinion can be very challenging, owing to the diversity and quantity of unstructured data in existence. [43].

Binali et al. [11] indicated that determining the emotional experiences of e-learning students can be difficult; however, through mining techniques, analyses can detect emotion in online students. In addition, identifying the different emotions of e-learning students can help model more suitable educational programs. Another study [16] proposed summarizing customer reviews by choosing product features on which they commented, classifying whether the opinion was positive or negative, and summarizing the results. This analysis is important because extremely high numbers of reviews prevent potential customers from reviewing every single opinion. Mate [23] proposed a ranking of essential product features from the online reviews of consumers. These aspects were identified by the number of times the product features appeared in reports and how these aspects influenced the overall opinions of consumers.

Estrada et al. [79] performed a comparison of sentiment analysis classifying techniques, machine learning, deep learning, and EvoMSA to classify education opinions in an Intelligent Learning Environment called ILE-Java. The development of two corpora expressions, sentiTEXT, which has polarity positive and negative labels, and eduSERE, which has positive and negative learning-centered emotion labels, reflected students' emotional states regarding teachers, exams, homework, and projects. EvoMSA produced the best results among the classifying techniques, with a 93% accuracy rating for the sentiTEXT corpus and an 84% accuracy rating for the eduSERE corpus. Two expressions in the programming language domain reflect the emotional states of students and their feelings regarding teachers' exams, homework, and academic projects: sentiTEXT (positive and negative labels) and eduSERE (positive and negative learning-centered emotions labels. Misuraca et al., 2021 [81] discussed OM as a combination of statistics, linguistics, and computer science that evaluates sentiments of individual opinions and highlights semantic

orientation. The discussion includes the induction of OM as a statistical text analysis tool in a learning environment to process student feedback from natural language producing useful analytics, and to explore text collections from a quantitative viewpoint.

Wu et al. [28] studied how information shared on Facebook pages can be beneficial in determining whether a company is correctly reaching its customers or the desired requirements are met. By analyzing the interactions of Facebook users and the reactions to their posts, companies can gather information, apply statistical analyses, and model behavioral trends [28]. Kaur and Bansal [34] introduced opinion mining as a powerful tool for e-commerce because it gathers information about how customers feel about different products. This collection of opinions can help companies make better decisions and align their efforts with what customers really want. The classification of e-commerce users represents an appealing area of study for marketers seeking to align their efforts to capture more consumers. [34]. Gamon [41] used large feature vectors and feature reduction to demonstrate that large, noisy data regarding customer feedback can be analyzed and classified. Feedback received from customers can present many challenges, and classifying these data is necessary to retrieve only the important information [41].

Bollen et al. [12] highlighted that many Twitter users express their emotions through this social media platform. With the use of a psychometric instrument, different social events were found to profoundly affect changes in public mood. The identification of these sentiments reflects personality trends, as well as the atmosphere and emotions of Twitter users. Basari et al. [22] examined how tweets can contain information about users' preferences regarding movies. SVM can analyze natural language to determine patterns via opinion mining. Online reviews can help predict the possible preferences of the movie audience [22]. Zengin Alp and Gündüz Öğüdücü [38] introduced a method called Personalized PageRank, which integrates the information retrieved from network topology and the information of Twitter users regarding their actions and activities. This capability has become appealing for marketers because Twitter is an online platform where users share their preferences.

Saire and Cruz [84] focused on the use of text mining of data collected from social media and search trends to analyze the effects of COVID-19 on the population of Paris, France, from 23 April 2020 to 18 June 2020. The primary findings revealed a decreasing pattern of publication/interest in the health crisis and the health and economic effects on the population resulting from the effects of COVID-19. Chire-Saire [85] used analysis of social media through complex network representation and text mining to compare the effects of COVID-19 in other countries. Focusing on South American countries, the analysis of texts via Twitter indicated the existence of patterns similar to those in complex systems and confirmed the idea of system and visualization of adjacency matrices, which may potentially identify posts made by robots as opposed to humans.

Frost et al. [14] studied the system MONARCA 2.0 to collect relevant information from bipolar patients, with an aim to provide insight into the disease for both patients and clinicians by processing subjective and objective data about patient mood. This system helps identify patterns in behaviors and factors affecting the disease [14]. Lachmar et al. [27] gathered information shared by individuals with sentiments of depression on Twitter through the hashtag #MyDepressionLooksLike. These tweets presented dysfunctional thoughts, hopeless feelings, and unlovability characteristics, thus revealing how people with depression talk about their symptoms via social networking. Pijnenborg et al. [40] discussed the benefits of using SMS to decrease the effects of cognitive impairments in patients who have schizophrenia. Because schizophrenia also involves delusions and hallucinations, improvements in the status of patients using SMS can be very modest.

Bespalov et al. [13] proposed an approach to modeling higher-order sentences to a lower order to make the classification of data viable. Supervised latent n-gram analyses can help classify sentiments that are extracted from textual information. Davis et al. [29] determined how analytical models can enhance public safety with the help of probabilistic and parametric methods, as well as different nonlinear algebraic models, by analyzing

uncertain data and identifying threats and false alarms, and detecting possible terrorist profiles [29].

Gill [33] illustrated the relationship between the language used and the personality projected by word choice. The personality traits of extraversion, neuroticism, and psychoticism can be determined by analyzing text from emails [33]. Boyd and Pennebaker [39] studied the language used by people to identify personality patterns. Rather than focusing on responses to self-reported questionnaires, language-based measures represent a new approach to model personality trends. A.S. Cohen et al. [3] applied computerized lexical analyses to determine positive or negative affectivity dimensions through natural speech. Measuring personality was possible because people with positive affectivity demonstrate high levels of positive emotions, whereas those with negative affectivity show high levels of negative emotions.

Brynielsson et al. [31] used different techniques for analyzing data to detect "lone wolf" terrorists with the goal of preventing possible attacks. Analytical models were created by using a platform to harvest and capture online information and trace possible lone wolves [31]. K. Cohen et al. [32] established the challenges of detecting lone wolves by using traditional police methods and introduced new tools and technologies that can detect weak signals in the form of linguistic markers that facilitate the identification of lone wolves' profiles [32].

Hung et al. [30] introduced a new framework and technology called INSiGHT (Investigative Search for Graph-Trajectories) that helps detect groups or individuals whose behavior suggests a potential for violence by identifying radicalization trajectories over time [30]. Paul K. Davis et al. [36] studied behavioral patterns and their usage to predict possible acts of violence.

4.2.2. Social Category

In this category, Alexander Semenov et al. [45] studied the identification of possible school shooters by analyzing the content shared by users on different social media platforms. Future shooters can be identified by analyzing the emails, chats, texts, and social media feeds of prior school shooters sharing similar behaviors [45]. Bartlett and Reynolds [46] presented how social media faces legal and ethical responsibilities, yet also can be useful to prevent terrorism and preserve public safety. Privacy can protect the public and prevent the use of social media for terrorism and propagandistic purposes [46]. Marrese-Taylor et al. [52] tested the software Opinion Zoom to gather online information about tourism opinions to propose solutions to problems in the industry. A modular tool was used because tourism opinions on the web can help predict possible traveling patterns as well as preferences of travelers.

Kastrati et al. [49] investigated the activities of users on online social networks to identify crimes by applying the objective metric SEMCON. By retrieving online posts, feeds, or users' comments, this method can determine whether a user is a suspect [49].

Bollen et al. [12] analyzed how OpinionFinder and Google-Profile of Mood States (GPMOS) can help determine the mood patterns presented on social media regarding worldwide events. This analysis can also help companies predict the behavior of customers regarding the stock market and minimize the effects of fluctuations in the stock market. Bucur [47] established that opinion mining had become a key technique for extracting and collecting relevant information needed for companies to make better decisions and that the opinions of customers are fundamental input. Opinion mining has become an appealing area of study for many businesses [47].

Dave et al. [48] extracted textual information and classified online reviews as positive or negative according to different product attributes. Opinions can be classified through semantic analysis of online reviews [48]. Zha et al. [51] introduced a ranking system for product aspects by identifying that a) the most important aspects are described by more consumers, and b) these aspects directly affect the overall opinion of consumers. Product

aspect ranking has many applications in various industries, and the main use is to gather relevant information to make better decisions.

Nahm and Mooney [50] examined how DiscoTEX can help extract data by combining data mining and information extraction. This method can locate data within documents and transform unstructured text into a structured database, as well as predict additional information for extraction from other documents. The integration of data mining and information extraction can help combine data in a more readable structure [50]. McCallum [56] investigated how unstructured data present a challenge in interpreting information. Therefore, the aim of information extraction is to create a database by gathering loosely formatted texts in which patterns can be identified by data mining [56].

Diehl [53] examined not only the structural but also the cultural aspects of social networks. Relational sociology studies have tended to examine and retrieve information from text data, whereas the importance of the implications of face-to-face interactions when analyzing network information has largely been ignored. A. Semenov et al. [54] proposed three modules for long-term monitoring of different social networks: the crawler, the repository, and the analyzer. By crawling, storing, and analyzing different sites, longitudinal data from social media sites can be examined.

Pennebaker [55] analyzed the words that people use in emails, Twitter feeds, and Facebook posts to determine their emotions, thoughts, social relationships, and personalities. The focus was on word use rather than on how people were speaking. Mind mapping can help explore social and psychological trends. Ibrahim and Ahmad [57] researched how Requirements Analysis and Class Diagram Extraction (RACE) can expedite textual extractions and improve the analysis of the data requirements that are currently performed manually. Many NLP techniques were developed to extract relevant information from textual data.

4.2.3. Cognition Category

Eichinger et al. [58] introduced Affinity, a system that can assess similarities among the text message histories of users while preserving private information. A latent format is used, which does not allow for the reconstruction of the comparison words. Chung and Pennebaker [61] distinguished the adjectives most commonly used by college students by applying computerized text analytic tools. This study has established the strengths of analyzing open-ended texts to extract information from the natural language used by different participants. This method enables the examination of cultural patterns as well as personality characteristics.

Bond and Pennebaker [59] experimented with changing pronouns to moderate the health benefits of expressive writing by alternating the focus of participants. Expressive writing can therefore affect people's physical and psychological health. Pennebaker and Stone [6] developed two projects showing the relationship between language use and aging: as people age, they tend to use more positive affect words than negative affect words and to use fewer self-references and fewer past-tense verbs.

Rajman and Besançon [62] established that text mining is a powerful technique to extract important information from a dataset by applying probabilistic associations of keywords because unstructured data can be challenging to interpret.

Fishhoff and Chauvin [106] investigated how intelligence analysis helps clear difficult situations and enhance valuable information for better decision-making by evaluating and integrating pertinent information. Intelligence analysis can help determine behavioral profiles and social conduct.

4.2.4. Other Studies

Kosala and Blockeel [65] explored the use of web mining by dividing it into three different categories—web content mining, web structure mining, and web usage mining—and studying representation issues, recess, and learning algorithms. Balazs and Velásquez [63] studied how information fusion seeks to correctly transform and compress data to trans-

form them into a more understandable representation. Fusion processes and the development of surveys to extract relevant data can be helpful as the use of opinion mining steadily increases. Nigam et al. [67] evaluated maximum entropy techniques to establish how a uniform distribution can benefit the classification of data. More studies must be performed, but this technique appears promising.

Continuous efforts have been undertaken worldwide to propose new classification algorithms such as Tsetlin Machine [107] or Dendritic Neuron Models [108]. Rutland et al. [70] evaluated how the use of SMS can be measured with the SMS Problem Use Diagnostic Questionnaire (SMS-PUDQ) to determine behavioral addiction to SMS use. The time spent using SMS and other measures of mobile phone use were detected during the study. Aggarwal and Zhai [71] explored the importance of mining text data, an appealing research topic, given that the amount of web-enabled data has increased and facilitates the exploration of vast quantities of textual data. A comparison of the classical and modern aspects of text mining was also described. Berry and Kogan [72] studied the contributions of text mining, as well as major topics associated with text mining, by categorizing text into three different components to explore keyword extraction, classification, and the clustering of information presented in textual data. Akilan [73] investigated the field of text mining to extract unstructured data and identify interesting and non-trivial patterns from text documents. An exploration of the current challenges and projected directions of this field was described [73]. Chakraborty et al. [64] prepared various case studies and performed text mining and analysis to extract important information from textual data. Different scenarios were created wherein SAS was used to perform comprehensive text analytics to help industries leverage the textual data [64].

Shahbaz et al. [68] proposed a solution to the analysis of textual information by developing a system, Sentiment Miner, to process and classify text files according to opinions stated in various sentences by using NLP techniques and opinion mining algorithms. Weiss et al. [69] introduced methods to predict and analyze unstructured information presented on textual data. Methods used for data mining could be adapted to be applied to text.

Chakraborty et al. [64,109] collected insightful information from customers by analyzing textual data from various documents to improve business operations and performance. Analyses of unstructured data are possible by extracting important information when performing text analysis and sentiment mining. Weerdt et al. [74] described the importance of retrieving data to benefit business process management by applying process mining, which uses techniques to analyze and extract knowledge and information from system event logs.

Manning and Schutze [66] established the value of using statistical NLP to extract and interpret textual data, not only for businesses but also for government agencies and individuals who could benefit from extracting information from a large amount of data. The theory and practice of these techniques are also explored. [66]

Moraes et al. [75] compared SVM and artificial neural networks to determine the differences between these two approaches in performing sentiment analysis and determined that artificial neural networks perform better than SVMs. Fraley [76] presented guidelines on how to construct web-based surveys to conduct behavioral research. Strengths and limitations of online surveys are highlighted, as well as the factors affecting the design of internet-based research.

5. Conclusions

In this paper, we analyzed published articles on different topics related to text mining and human behavior. We divided the analysis into psychological behaviors regarding emotion, cognition, and social empathy. The current research in identifying behaviors has focused primarily on detecting emotional and social behaviors, whereas studies on cognitive behavior are rarer. We found that NLP is the most common approach, which is followed by information extraction and document classification. Another main finding

in this review was that few studies have focused on detecting cognitive behavior. To our knowledge, no decision support system has used a holistic approach to analyze cognitive, emotional, and social behaviors simultaneously. The literature reviews analyzed and the articles in Table 2 focus primarily on detecting emotions or empathy. The psychological studies, for example [4] and [76], identified relationships between cognitive aspects, emotions, and empathy. For this reason, it would be helpful to develop analytical and computational systems that make it possible to identify the connections between different aspects of human behavior through text analysis. In this way, predictions of future human behaviors and explanations of past actions could be made. Furthermore, behaviors and their effects on the outcomes of human action should be distinguished in greater detail. For instance, the detection of negative words in comments can be associated with certain social behaviors (e.g., being socially aggressive), as well as with cognitive behaviors (e.g., having dementia or depression).

Through the literature review, we identified a trend in the detection of mood states that may affect a person's life. For example, technological tools can support the detection of behaviors over time (e.g., hours, days, weeks, or months). Consequently, detecting short-term emotional behaviors in users (e.g., being "socially inactive" over a long period) could, in turn, predict mood states and disorders (e.g., loneliness or depression), which could also affect long-term social and cognitive behaviors. Thus, the traceability of behaviors was studied. In the same way, many authors have demonstrated how text messages or the translation of audio or video to textual data could contain delicate information that might otherwise be missed. In addition, privacy and security are issues that must be managed through the use of anonymous analyses.

Most of the literature discussed how the formulation and understanding of human behavior were challenging and remained an evolving area of research that considerably affects analytics. On the basis of the present review, a method or platform allowing all classification methods to be combined has not been thoroughly explored. In contrast, we found that each element of behavior has generally been examined individually. Hence, future work should address this lack of information by using more systematic approaches, in which multiple behavioral aspects can be analyzed simultaneously.

We hope that this review will support the design of a system that combines sentiment mining and NLP techniques to develop an unstructured data opinion miner and index engine for polarity extraction and classification at the sentence level through the use of a variety of documents from repositories that represent or describe a given group of individuals. Such a system should also facilitate the use of progressive tracking to capture various changes in individual behavior, including the detection of behavioral changes, the detection of anomalies, risk evaluation, and monitoring. This research may serve as a reference for practitioners and researchers interested in detecting human behavior through text analysis.

Author Contributions: Conceptualization, E.G. and K.F.; methodology, E.G.; software, E.G.; validation, M.R.D., W.K. and T.L.; formal analysis, E.G.; investigation, T.A.; resources, E.G.; data curation, M.R.D.; writing—original draft preparation, E.G.; writing—review and editing, M.R.D.; visualization, E.G.; supervision, T.A.; project administration, W.K.; funding acquisition, W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported in part by a research grant from the Office of Naval Research N000141812559 and was performed at the University of Central Florida, Orlando, Florida.

Institutional Review Board Statement: Not applicable to this study.

Informed Consent Statement: Not applicable to this study.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: No potential conflict of interest was reported by the authors.

References

1. Ahram, T.Z.; McCauley-Bush, P.; Karwowski, W. Estimating Intrinsic Dimensionality Using the Multi-Criteria Decision Weighted Model and the Average Standard Estimator. *Inf. Sci.* **2010**, *180*, 2845–2855. [CrossRef]
2. Liu, B. Sentiment Analysis and Opinion Mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [CrossRef]
3. Cohen, A.S.; Minor, K.S.; Baillie, L.E.; Dahir, A.M. Clarifying the Linguistic Signature: Measuring Personality From Natural Speech. *J. Pers. Assess.* **2008**, *90*, 559–563. [CrossRef] [PubMed]
4. Bornstein, M.H. Human Behavior | Definition, Theories, Characteristics, Examples, Types, & Facts. Available online: https://www.britannica.com/topic/human-behavior (accessed on 21 March 2021).
5. Tausczik, Y.R.; Pennebaker, J. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. Lang. Soc. Psychol.* **2009**, *29*, 24–54. [CrossRef]
6. Pennebaker, J.W.; Stone, L.D. Words of wisdom: Language use over the life span. *J. Pers. Soc. Psychol.* **2003**, *85*, 291–301. [CrossRef] [PubMed]
7. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef] [PubMed]
8. Higgins, J.P.T.; Altman, D.G.; Gøtzsche, P.C.; Jüni, P.; Moher, D.; Oxman, A.D.; Savović, J.; Schulz, K.F.; Weeks, L.; Sterne, J.A.C.; et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* **2011**, *343*, d5928. [CrossRef] [PubMed]
9. Gravenhorst, F.; Muaremi, A.; Bardram, J.; Grünerbl, A.; Mayora, O.; Wurzer, G.; Frost, M.; Osmani, V.; Arnrich, B.; Lukowicz, P.; et al. Mobile phones as medical devices in mental disorder treatment: An overview. *Pers. Ubiquitous Comput.* **2015**, *19*, 335–353. [CrossRef]
10. Mahendran, A.; Duraiswamy, A.; Reddy, A.; Gonsalves, C. Opinion Mining for Text Classification. *Int. J. Sci. Eng. Technol.* **2013**, *2*, 589–594.
11. Binali, H.H.; Wu, C.; Potdar, V. A new significant area: Emotion detection in E-learning using opinion mining techniques. In Proceedings of the 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, Lake Ohrid, Macedonia, 16–19 June 2009; pp. 259–264.
12. Bollen, J.; Mao, H.; Pepe, A. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. *Proc. Int. AAAI Conf. Web Soc. Media* **2011**, *5*, 1.
13. Bespalov, D.; Bai, B.; Qi, Y.; Shokoufandeh, A. Sentiment Classification Based on Supervised Latent N-Gram Analysis. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow Scotland, UK, 24–28 October 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 375–382.
14. Frost, M.; Doryab, A.; Faurholt-Jepsen, M.; Kessing, L.V.; Bardram, J.E. Supporting Disease Insight through Data Analysis: Refinements of the Monarca Self-Assessment System. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 133–142.
15. Grunerbl, A.; Muaremi, A.; Osmani, V.; Bahle, G.; Ohler, S.; Troster, G.; Mayora, O.; Haring, C.; Lukowicz, P. Smartphone-Based Recognition of States and State Changes in Bipolar Disorder Patients. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 140–148. [CrossRef]
16. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, DC, USA, 22–25 August 2004; pp. 168–177.
17. Miedema, F. *Sentiment Analysis with Long Short-Term Memory Networks*; Vrije Universiteit Amsterdam: Amsterdam, The Netherlands, 2018.
18. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. *arXiv* **2002**, arXiv:cs/0205070.
19. Arora, R.; Srinivasa, S. A Faceted Characterization of the Opinion Mining Landscape. In Proceedings of the 2014 Sixth International Conference on Communication Systems and Networks; IEEE Computer Society: Washington, DC, USA, 2014; pp. 1–6.
20. Salloum, S.A.; Al-Emran, M.; Monem, A.A.; Shaalan, K. A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. *Adv. Sci. Technol. Eng. Syst. J.* **2017**, *2*, 127–133. [CrossRef]
21. Turney, P.D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *arXiv* **2002**, arXiv:cs/0212032.
22. Basari, A.S.H.; Hussin, B.; Ananta, I.G.P.; Zeniarja, J. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Eng.* **2013**, *53*, 453–462. [CrossRef]
23. Mate, C. Product Aspect Ranking Using Sentiment Analysis: A Survey. *Int. Res. J. Eng. Technol.* **2015**, *3*, 126–127.
24. Othman, M.; Hassan, H.; Moawad, R.; El-Korany, A. Opinion Mining and Sentimental Analysis Approaches: A Survey. *Life Sci. J.* **2014**, *11*, 321–326.
25. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends® Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
26. Vinodhini, G.; Chandrasekaran, R.M. Sentiment Analysis and Opinion Mining: A Survey. *Int. J.* **2012**, *2*, 282–292.
27. Lachmar, E.M.; Wittenborn, A.K.; Bogen, K.W.; McCauley, H.L.; Cravens, J.; Berry, N.; Radovic-Stakic, A. #MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter. *JMIR Ment. Health.* **2017**, *4*, e43. [CrossRef] [PubMed]
28. Wu, H.; Liu, K.; Trappey, C. Understanding Customers Using Facebook Pages: Data Mining Users Feedback Using Text Analysis. In Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD); IEEE: Piscataway, NJ, USA, 2014; pp. 346–350.

29. Davis, P.K.; Manheim, D.; Perry, W.L.; Hollywood, J. Using causal models in heterogeneous information fusion to detect terrorists. In Proceedings of the 2015 Winter Simulation Conference (WSC); IEEE: Piscataway, NJ, USA, 2015; pp. 2586–2597.
30. Hung, B.W.K.; Jayasumana, A.P.; Bandara, V.W. INSiGHT: A System for Detecting Radicalization Trajectories in Large Heterogeneous Graphs. In Proceedings of the 2017 IEEE International Symposium on Technologies for Homeland Security (HST), Waltham, MA, USA, 25–26 April 2017; pp. 1–7.
31. Brynielsson, J.; Horndahl, A.; Johansson, F.; Kaati, L.; Mårtenson, C.; Svenson, P. Harvesting and analysis of weak signals for detecting lone wolf terrorists. *Secur. Inform.* **2013**, *2*, 1–15. [CrossRef]
32. Cohen, K.; Johansson, F.; Kaati, L.; Mork, J.C. Detecting Linguistic Markers for Radical Violence in Social Media. *Terror. Polit. Violence* **2013**, *26*, 246–256. [CrossRef]
33. Gill, A.J. Personality and Language: The Projection and Perception of Personality in Computer-Mediated Communication. Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, 2003.
34. Kaur, J.; Bansal, M. Hierarchical Sentiment Analysis Model for Automatic Review Classification for E-commerce Users. In *Hybrid Intelligence for Social Networks*; Banati, H., Bhattacharyya, S., Mani, A., Köppen, M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 249–267. ISBN 978-3-319-65139-2.
35. Muaremi, A.; Gravenhorst, F.; Grünerbl, A.; Arnrich, B.; Tröster, G. Assessing Bipolar Episodes Using Speech Cues Derived from Phone Calls. In Proceedings of the Pervasive Computing Paradigms for Mental Health; Cipresso, P., Matic, A., Grünerbl, A., Lopez, G., Tröster, G., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 103–114.
36. Davis, P.K.; Perry, W.L.; Brown, R.A.; Yeung, D.; Roshan, P.; Voorhies, P. *Using Behavioral Indicators to Help Detect Potential Violent Acts*; RAND Corporation: Santa Monica, CA, USA, 2013.
37. Nasukawa, T.; Yi, J. Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In Proceedings of the Proceedings of the 2nd International Conference on Knowledge Capture; Association for Computing Machinery: New York, NY, USA, 2003; pp. 70–77.
38. Alp, Z.Z.; Öğüdücü, Ş.G. Identifying topical influencers on twitter based on user behavior and network topology. *Knowl. Based Syst.* **2018**, *141*, 211–221. [CrossRef]
39. Boyd, R.; Pennebaker, J. Language-based personality: A new approach to personality in a digital world. *Curr. Opin. Behav. Sci.* **2017**, *18*, 63–68. [CrossRef]
40. Pijnenborg, G.H.M.; Withaar, F.K.; Brouwer, W.H.; Timmerman, M.E.; Bosch, R.J.V.D.; Evans, J.J. The efficacy of SMS text messages to compensate for the effects of cognitive impairments in schizophrenia. *Br. J. Clin. Psychol.* **2010**, *49*, 259–274. [CrossRef] [PubMed]
41. Gamon, M. Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis. In Proceedings of the COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004; pp. 841–847.
42. Pennebaker, J.W.; Boyd, R.L.; Jordan, K.; Blackburn, K. *The Development and Psychometric Properties of LIWC2015*; University of Texas at Austin: Austin, TX, USA, 2015.
43. Haddi, E.; Liu, X.; Shi, Y. The Role of Text Pre-processing in Sentiment Analysis. *Procedia Comput. Sci.* **2013**, *17*, 26–32. [CrossRef]
44. Li, D.; Qian, J. Text Sentiment Analysis Based on Long Short-Term Memory. In Proceedings of the 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), Wuhan, China, 13–15 October 2016; pp. 471–475.
45. Semenov, A.; Veijalainen, J.; Kyppo, J. Analysing the presence of school-shooting related communities at social media sites. *Int. J. Multimed. Intell. Secur.* **2010**, *1*, 232–268. [CrossRef]
46. Bartlett, J.; Reynolds, L. *The State of the Art 2015: A Literature Review of Social Media Intelligence Capabilities for Counter-Terrorism*; Demos London; Demos: London, UK, 2015.
47. Bucur, C. Opinion Mining Platform for Intelligence in Business. *Econ. Insights Trends Chall.* **2014**, *3*, 99–108.
48. Dave, K.; Lawrence, S.; Pennock, D.M. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceedings of the 12th International Conference on World Wide Web; Association for Computing Machinery: New York, NY, USA, 2003; pp. 519–528.
49. Kastrati, Z.; Imran, A.S.; Yildirim-Yayilgan, S.; Dalipi, F. Analysis of Online Social Networks Posts to Investigate Suspects Using SEMCON. In Proceedings of the Social Computing and Social Media; Meiselwitz, G., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp. 148–157.
50. Nahm, U.Y.; Mooney, R.J. A Mutually Beneficial Integration of Data Mining and Information Extraction. In Proceedings of the AAAI/IAAI, Austin, TX, USA, 1–3 August 2000; pp. 627–632.
51. Zha, Z.-J.; Yu, J.; Tang, J.; Wang, M.; Chua, T.-S. Product Aspect Ranking and Its Applications. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1211–1224.
52. Marrese-Taylor, E.; Velásquez, J.D.; Bravo-Marquez, F. Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT); IEEE: Piscataway, NJ, USA, 2013; Volume 3, pp. 261–264.
53. Diehl, D.K. Language and Interaction: Applying Sociolinguistics to Social Network Analysis. *Qual. Quant.* **2019**, *53*, 757–774. [CrossRef]

54. Semenov, A.; Veijalainen, J.; Boukhanovsky, A. A Generic Architecture for a Social Network Monitoring and Analysis System. In Proceedings of the 2011 14th International Conference on Network-Based Information Systems, Tirana, Albania, 7–9 September 2011; pp. 178–185.
55. Pennebaker, J.W. Mind Mapping: Using Everyday Language to Explore Social & Psychological Processes. *Procedia Comput. Sci.* **2017**, *118*, 100–107. [CrossRef]
56. McCallum, A. Information Extraction: Distilling Structured Data from Unstructured Text. *Queue* **2005**, *3*, 48–57. [CrossRef]
57. Ibrahim, M.; Ahmad, R. Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques. In Proceedings of the 2010 Second International Conference on Computer Research and Development, Kuala Lumpur, Malaysia, 7–10 May 2010; pp. 200–204.
58. Eichinger, T.; Beierle, F.; Khan, S.U.; Middelanis, R. Affinity: A System for Latent User Similarity Comparison on Texting Data. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7.
59. Bond, M.; Pennebaker, J.W. Automated Computer-Based Feedback in Expressive Writing. Comput. *Hum. Behav.* **2012**, *28*, 1014–1018. [CrossRef]
60. National Research Council. *Intelligence Analysis: Behavioral and Social Scientific Foundations*; National Academies Press: Washington, DC, USA, 2011.
61. Chung, C.K.; Pennebaker, J.W. Revealing Dimensions of Thinking in Open-Ended Self-Descriptions: An Automated Meaning Extraction Method for Natural Language. *J. Res. Personal.* **2008**, *42*, 96–132. [CrossRef]
62. Rajman, M.; Besançon, R. Text Mining-Knowledge Extraction from Unstructured Textual Data. In Proceedings of the Advances in Data Science and Classification; Rizzi, A., Vichi, M., Bock, H.-H., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 473–480.
63. Balazs, J.A.; Velásquez, J.D. Opinion Mining and Information Fusion: A Survey. *Inf. Fusion* **2016**, *27*, 95–110. [CrossRef]
64. Chakraborty, G.; Pagolu, M.; Garla, S. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*; SAS Institute: Cary, NC, USA, 2014.
65. Kosala, R.; Blockeel, H. Web Mining Research: A Survey. *ACM SIGKDD Explor. Newsl.* **2000**, *2*, 1–15. [CrossRef]
66. Manning, C.; Schutze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
67. Nigam, K.; Lafferty, J.; McCallum, A. Using Maximum Entropy for Text Classification. In Proceedings of the IJCAI-99 Workshop on Machine Learning for Information Filtering, Stockholom, Sweden, 1 August 1999; Volume 1, pp. 61–67.
68. Shahbaz, M.; Guergachi, A.; Rehman, R.T. ur Sentiment Miner: A Prototype for Sentiment Analysis of Unstructured Data and Text. In Proceedings of the 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), Toronto, ON, Canada, 4–7th May 2014; pp. 1–7.
69. Weiss, S.M.; Indurkhya, N.; Zhang, T.; Damerau, F. *Text Mining: Predictive Methods for Analyzing Unstructured Information*; Springer Science & Business Media: Berlin, Germany, 2010.
70. Rutland, J.B.; Sheets, T.; Young, T. Development of a Scale to Measure Problem Use of Short Message Service: The SMS Problem Use Diagnostic Questionnaire. *Cyberpsychol. Behav.* **2007**, *10*, 841–844. [CrossRef] [PubMed]
71. Aggarwal, C.C.; Zhai, C. An introduction to text mining. In *Mining Text Data*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–10.
72. Berry, M.W.; Kogan, J. *Text Mining: Applications and Theory*; John Wiley & Sons: West Sussex, UK, 2010.
73. Akilan, A. Text Mining: Challenges and Future Directions. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 1679–1684.
74. Weerdt, J.D.; vanden Broucke, S.K.; Vanthienen, J.; Baesens, B. Leveraging Process Discovery with Trace Clustering and Text Mining for Intelligent Analysis of Incident Management Processes. In Proceedings of the 2012 IEEE Congress on Evolutionary Computation, Brisbane, Australia, 10–15 June 2012; pp. 1–8.
75. Moraes, R.; Valiati, J.F.; Gavião Neto, W.P. Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN. *Expert Syst. Appl.* **2013**, *40*, 621–633. [CrossRef]
76. Fraley, R.C. *How to Conduct Behavioral Research over the Internet: A Beginner's Guide to HTML and CGI/Perl*; Guilford Press: New York, NY, USA, 2004.
77. Greco, F.; Polli, A. Emotional Text Mining: Customer Profiling in Brand Management. *Int. J. Inf. Manag.* **2020**, *51*, 101934. [CrossRef]
78. Acheampong, F.A.; Wenyu, C.; Nunoo-Mensah, H. Text-Based Emotion Detection: Advances, Challenges, and Opportunities. *Eng. Rep.* **2020**, *2*, e12189. [CrossRef]
79. Estrada, M.L.B.; Cabada, R.Z.; Bustillos, R.O.; Graff, M. Opinion Mining and Emotion Recognition Applied to Learning Environments. *Expert Syst. Appl.* **2020**, *150*, 113265. [CrossRef]
80. Wang, X.; Kou, L.; Sugumaran, V.; Luo, X.; Zhang, H. Emotion Correlation Mining through Deep Learning Models on Natural Language Text. *IEEE Trans. Cybern.* **2020**. [CrossRef] [PubMed]
81. Misuraca, M.; Scepi, G.; Spano, M. Using Opinion Mining as an Educational Analytic: An Integrated Strategy for the Analysis of Students' Feedback. *Stud. Educ. Eval.* **2021**, *68*, 100979. [CrossRef]
82. Raeesi Vanani, I. Text Analytics of Customers on Twitter: Brand Sentiments in Customer Support. *J. Inf. Technol. Manag.* **2019**, *11*, 43–58.

83. Swain, D.; Khandelwal, A.; Joshi, C.; Gawas, A.; Roy, P.; Zad, V. A Suicide Prediction System Based on Twitter Tweets Using Sentiment Analysis and Machine Learning. In *Machine Learning and Information Processing: Proceedings of ICMLIP 2020*; Springer: Berlin/Heidelberg, Germany, 2021.
84. Saire, J.E.C.; Cruz, J.F.O. Study of Coronavirus Impact on Parisian Population from April to June Using Twitter and Text Mining Approach. In *2020 International Computer Symposium*; IEEE: Piscataway, NJ, USA, 2020.
85. Chire-Saire, J.E. Characterizing Twitter Interaction during COVID-19 Pandemic Using Complex Networks and Text Mining. *arXiv Prepr.* **2020**, arXiv:2009.05619.
86. Fareri, S.; Fantoni, G.; Chiarello, F.; Coli, E.; Binda, A. Estimating Industry 4.0 Impact on Job Profiles and Skills Using Text Mining. *Comput. Ind.* **2020**, *118*, 103222. [CrossRef]
87. Fteimi, N.; Hornung, O.; Smolnik, S. When Emotions Rule Knowledge: A Text-Mining Study of Emotions in Knowledge Management Research. *Int. J. Knowl. Manag. IJKM* **2021**, *17*, 1–16.
88. Bayram, U.; Benhiba, L. Determining a Person's Suicide Risk by Voting on the Short-Term History of Tweets for the CLPsych 2021 Shared Task. In Proceedings of the Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access, Mexico City, Mexico, 11 June 2021; pp. 81–86.
89. Davahli, M.R.; Karwowski, W.; Gutierrez, E.; Fiok, K.; Wróbel, G.; Taiar, R.; Ahram, T. Identification and Prediction of Human Behavior through Mining of Unstructured Textual Data. *Symmetry* **2020**, *12*, 1902. [CrossRef]
90. Siby, S. An Exploration about the Last Mile Logistic Efficiency in Indian E-Commerce Sector—A Text Mining Approach. In Proceedings of the International Conference on Innovative Computing & Communications (ICICC), New Delhi, India, 21–23 February 2020. Available online: https://ssrn.com/abstract=3563089 (accessed on 21 March 2021). [CrossRef]
91. Helbing, D.; Brockmann, D.; Chadefaux, T.; Donnay, K.; Blanke, U.; Woolley-Meza, O.; Moussaid, M.; Johansson, A.; Krause, J.; Schutte, S. Saving Human Lives: What Complexity Science and Information Systems Can Contribute. *J. Stat. Phys.* **2015**, *158*, 735–781. [CrossRef]
92. Huang, H.H.; Yang, Y.C.; Hsiao, C.T.; Liang, H.C.; Liu, C.S. The National Health Insurance: Decoding the Health Bill. In Proceedings of the 2010 IEEE International Conference on Management of Innovation Technology, Singapore, 2–5 June 2010; pp. 414–419.
93. Bakshi, K. Considerations for Big Data: Architecture and Approach. In Proceedings of the 2012 IEEE Aerospace Conference, Big Sky, MT, USA, 3–10 March 2012; pp. 1–7.
94. Talib, R.; Hanif, M.K.; Ayesha, S.; Fatima, F. Text Mining: Techniques, Applications and Issues. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 414–418. [CrossRef]
95. Gutiérrez, E.; Bhide, S.; Mendizabal, L.C.R. *Artificial Intelligence: Advances in Research and Applications*; Nova Science Publishers: Huappauge, NY, USA, 2018.
96. Sarawagi, S. *Information Extraction*; Now Publishers Inc.: Delft, The Netherland, 2008.
97. Wang, Y.-X.; Zhang, Y.-J. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Trans. Knowl. Data Eng.* **2012**, *25*, 1336–1353. [CrossRef]
98. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
99. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text Classification Algorithms: A Survey. *Information* **2019**, *10*, 150. [CrossRef]
100. Sisodia, D.; Singh, L.; Sisodia, S.; Saxena, K. Clustering Techniques: A Brief Survey of Different Clustering Algorithms. *Int. J. Latest Trends Eng. Technol. IJLTET* **2012**, *1*, 82–87.
101. Yeasmin, S.; Tumpa, P.B.; Nitu, A.M.; Uddin, M.P.; Ali, E.; Afjal, M.I. Study of Abstractive Text Summarization Techniques. *Am. J. Eng. Res.* **2017**, *6*, 253–260.
102. Joseph, S.R.; Hlomani, H.; Letsholo, K.; Kaniwa, F.; Sedimo, K. Natural Language Processing: A Review. *Nat. Lang. Process. Rev.* **2016**, *6*, 207–210.
103. Kumar, A.; Singh, R.K. Web Mining Overview, Techniques, Tools and Applications: A Survey. *Int. Res. J. Eng. Technol. IRJET* **2016**, *3*, 1543–1547.
104. Schmidt, C.; Collette, F.; Cajochen, C.; Peigneux, P. A Time to Think: Circadian Rhythms in Human Cognition. *Cogn. Neuropsychol.* **2007**, *24*, 755–789. [CrossRef] [PubMed]
105. Thakur, N.; Han, C.Y. An Approach to Analyze the Social Acceptance of Virtual Assistants by Elderly People. In Proceedings of the 8th International Conference on the Internet of Things, Santa Barbara, CA, USA, 15–18 October 2018; pp. 1–6.
106. Fischhoff, B.; Chauvin, C. Intelligence Analysis. *Behav. Soc.* 2011. Available online: https://www.nap.edu/read/13062/chapter/1#ii (accessed on 21 March 2021).
107. Granmo, O.-C. The Tsetlin Machine–A Game Theoretic Bandit Driven Approach to Optimal Pattern Recognition with Propositional Logic. *arXiv Prepr.* **2018**, arXiv:1804.01508.
108. Gao, S.; Zhou, M.; Wang, Y.; Cheng, J.; Yachi, H.; Wang, J. Dendritic Neuron Model with Effective Learning Algorithms for Classification, Approximation, and Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *30*, 601–614. [CrossRef]
109. Chakraborty, G.; Krishna, M. Analysis of Unstructured Data: Applications of Text Analytics and Sentiment Mining. In Proceedings of the SAS Global Forum, Washington, DC, USA, 23–26 March 2014; pp. 1288–2014.