Overview of Text Analysis Tasks, Applications and Approaches

Journal:	*IEEE Access*
Manuscript ID	Access-2020-58605
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	04-Dec-2020
Complete List of Authors:	Saeed, Muhammad; Government College University Faisalabad, Department of Software Engineering; Government College University Faisalabad, Computer Sciences Awais, Muhammad; Government College University Faisalabad, Department of Software Engineering; Government College University Faisalabad, Department of Software Engineering Younas, Muhammad ; Universiti Teknologi Malaysia, Software Engineering; Government College University Faisalabad, computer Science Shah, Muhammad Arif; Universiti Teknologi Malaysia, Software Engineering Zareei, Mahdi; Tecnologico de Monterrey, KHAN, ATIF ; Islamia College Peshawar, computer science Goudarzi, Shidrokh ; Universiti Kebangsaan Malaysia,
Keywords: Please choose keywords carefully as they help us find the most suitable Editor to review:	Computer applications, Data mining, Information analysis
Subject Category Please select at least two subject categories that best reflect the scope of your manuscript:	Computers and information processing, Information theory
Additional Manuscript Keywords:	Game theory,, Artificial Intelligence, Context Understanding, Pattern of Information

LP-FTSM-2020-007

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.DOI

Overview of Text Analysis Tasks, Applications and Approaches

MUHAMMAD YAHYA SAEED[1,2], MUHAMMAD AWAIS[2] MUHAMMAD YOUNAS[1], MUHAMMAD ARIF SHAH[3], MAHDI ZAREEI[4], ATIF KHAN[5] and SHIDROKH GOUDARZI[6]

[1]Department of Computer Science, Government College University Faisalabad, Allama Iqbal Road Faisalabad, 38000, Pakistan
[2]Department of Software Engineering, Government College University Faisalabad, Allama Iqbal Road Faisalabad, 38000, Pakistan
[3]Department of Software Engineering, Pak-Austria Fachhochschule Institute of Applied Sciences and Technology, Haripur, 22620, Pakistan
[4]Tecnologico de Monterrey, School of Engineering and Sciences, Zapopan 45201, Mexico
[5]Department of Computer Science, Islamia College, Peshawar, Peshawar, 25120, KP, Pakistan
[6]Centre for Artificial Intelligent (CAIT), Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia

Corresponding author: Shidrokh Goudarzi (e-mail: shidrokh@ukm.edu.my)

ABSTRACT The rising use of internet & social media has many real life applications. The conveniences of online resources causes the public to express their thoughts and opinions about everyday activities, or local/global issues. Social media evolution is becoming vital as a progressive contributing platform for understanding social activities and beliefs. Online text over social media is a firsthand survey for businesses, governments, and services providers. Big data gives certain up to date social point of views in real-time. The digitization of every aspect in life is increasing this trend. Global software development has enlarged the scope of this field to furnish new sciences in this direction. Many firms express the rising need for analyzing social reviews relating to users and consumer opinions for business purposes. Business firms extensively rely on online customer reviews to improve the firm's standards or services. This paper presents applications of social media text extractions and presents nine areas of text analysis. This paper highlights categories of text mining in the form of context understanding, information patterns understandings, knowledge discovery, and general semantics. The study bears classification and insight of the existing researches in described areas.

INDEX TERMS : Text Mining, Game theory, Artificial Intelligence, Information Processing, Ontologies, Context Understanding, Pattern of Information Understanding, Knowledge Discovery, Semantic Analysis

I. INTRODUCTION

THE large growth of digital text creates the need for innovative data processing approaches for extracting desired text. In this paper, we have presented text and data mining studies. Unstructured textual data is the combination of worldwide digital data, web resources and textual archives, processed by artificial intelligence, machine learning, etc. Online data largely exists in the forms of blogs, chats, marketing data, social networks, etc. This data helps in mitigating social, commercial, and even political issues. This paper gives a survey of text mining approaches by presenting related studies. Online media is swiftly replacing offline media, resulting in audience involvement in real-time discussions to give independent thoughts on all issues. Online media platforms are the base for sharing daring ideas. They have quick responses with active feedback on certain issues like textual posts e.g., Twitter owns more than two hundred million monthly users with about more than five hundred million daily tweets. This is a good resource for public opinion gathering from all societies. Its analysis requires principles of detecting suitable text for extracting patterns of opinion. [1], [2].

Social Media Mining(SMM) is an emerging field of text mining. Social media generates such a large amount of data like millions of messages per day, and it is almost impossible to work with such volumes without special programs and algorithms for sorting and grouping data into a form convenient for analytical work. SMM reflects the buyer's decision-making process. Nowadays, the relationship between the brand and consumer has changed dramatically by becoming a dialogue in which the consumer's influence is greater than ever before. Digital marketing allows firms to establish direct links between the brand and the user. The final goal has become by no means just the purchase of a product by the consumer. Every brand has a responsibility to leverage

SMM to maintain customer loyalty and, most importantly, to encourage the spread of brand positivity. In short, studying the behaviour of social groups and individuals, segmentation of users based on their interests and the nature of their online interactions, are the possibilities that SMM may provide. This study summarizes the approaches to the analysis of SMM based on text analysis and analysis of social networks, which have developed to date.

Techniques for Text Mining(TM) are diverse, by nature of applying various tools of natural language processing. Text analysis tools analyze any text-based data-set, including social media, surveys, forum posts, support tickets, call transcripts, and more. Text analysis helps to process social media data to answer a wide variety of questions about consumers, brands, products, or any other topic. Text analysis attempts to understand sentiment and emotions expressed about a brand, product, or topic. It measures public voice in order to understand what percentage of a conversation exists about a brand, product, or topic. It identifies key topics, words, and phrases. Text analysis drills down within any conversation to understand what drives it and how the content of the conversation has changed over time. Text analysis quantifies a writer's desire and identifies the intent to behave.

Application of SMM has vital aspects for data analysis firms. Operational work with social media data consists of a quick response to a complaint and solving problems. For example, consider the case when a certain information feed becomes public. In such a situation, SMM serves as a tool to prevent negative responses by necessitating text analysis of the wave of negative information. Firms mostly use long-term analysis of data obtained over a long period of time to solve several marketing problems. The goal of SMM is to create a positive company image on the Internet and taking notice to consumer/user input. Quantitative analysis of reviews makes the identification of driver factors, barriers, loyalty, the brand image on the Internet, etc. Qualitative analysis of SMM targets general qualitative analysis which may be the unstructured or non-targeted search for insights, determining the values and needs of users, as well as a model of their relationship with the brand. Discourse analysis focuses on consumer behaviour through basic assumptions like taken for granted assumptions and rhetorical ways of argumentation and justification of their own positions. SMM Analysis is used by brands to study the activity of the brand and competitors, as well as the reactions of users depending on their psychographic characteristics. This analysis determines the optimal brand strategy on the Internet for both the general vector of activity and the content strategy. Analysis of SMM work and adjustment of the content strategy apply online survey of community members.

In this study, we have shown the current work in Figure 1, and it depicts the basic areas over which more than one hundred research papers have been selected: Text Mining (TM), Game theory (GT), Artificial Intelligence (AI), Information Processing (IP), Ontologies, AI & GT. These are the crucial research areas in Social Media (SM) based Text and Data Mining (TDM). The text extraction happens mainly in four main areas i.e., Context Understanding (CU), Pattern of Information Understanding (PIU), various types of Knowledge Understandings (KU) and Semantics Analysis (SA) [3], [4]. Based on these possible outcomes the papers divided into nine sections as shown in Figure 2. This figure has two parts. The right-hand side shows the grouping of the papers as per the basic research outcome. The left-hand side shows the four major forms of the result outcomes [5], [6]. We have presented these as a summarized table in each section.

All selected papers either fall in one or more of these four categories. Their general description is below:

1) Context Understanding: Context represents the topics or main issues of written text. Before context analysis, the text quality assessment comes first i.e. assessing the bias, randomness of topics, text-noise, and polarity. We show this outcome as a flag "A" in the section-wise summary tables.
2) Trivial Information Patterns Understanding: There exist various types of named entities, instance relations, time stamps, etc., We show this outcome as a flag "B" in the section-wise summary tables.
3) Knowledge Discovery: Knowledge is the context processed for the hidden related issues by matching text relevance. We show this outcome as a flag "C" in the section-wise summary tables.
4) Semantic Analysis: Semantics help to understand the basic reasoning or meaning of the text. It targets context to find the actual meaning behind the written text. We show this outcome as a flag "D" in the section-wise summary tables.

There are eleven sections in this paper. The first section is the introduction of this paper. The sections from Section-II to Section-X, are listed on the right-hand side of Figure 2. The last section is the conclusion of this paper.

II. DECISION MAKING

This section presents the text engineering for decision support and decision-making. This section covers various topics like weather forecasts, natural language and image processing, online social media, cyber-criminal, the probabilistic generative model of the semantic web, generating multiple-choice questions with a certain difficulty level from ontologies, intelligent content generation, equity criteria applied to assess serial events identification, crime pattern similarities indexing, identifying series and regularity of domains differences, measuring community diseases risks, medical decision-making, Kavli health project, evolutionary decision-making methods, Twitter future events prediction for decision-making, automation of decision-making, machine learning integration, the performance of predictive modelling etc., [5]–[15]

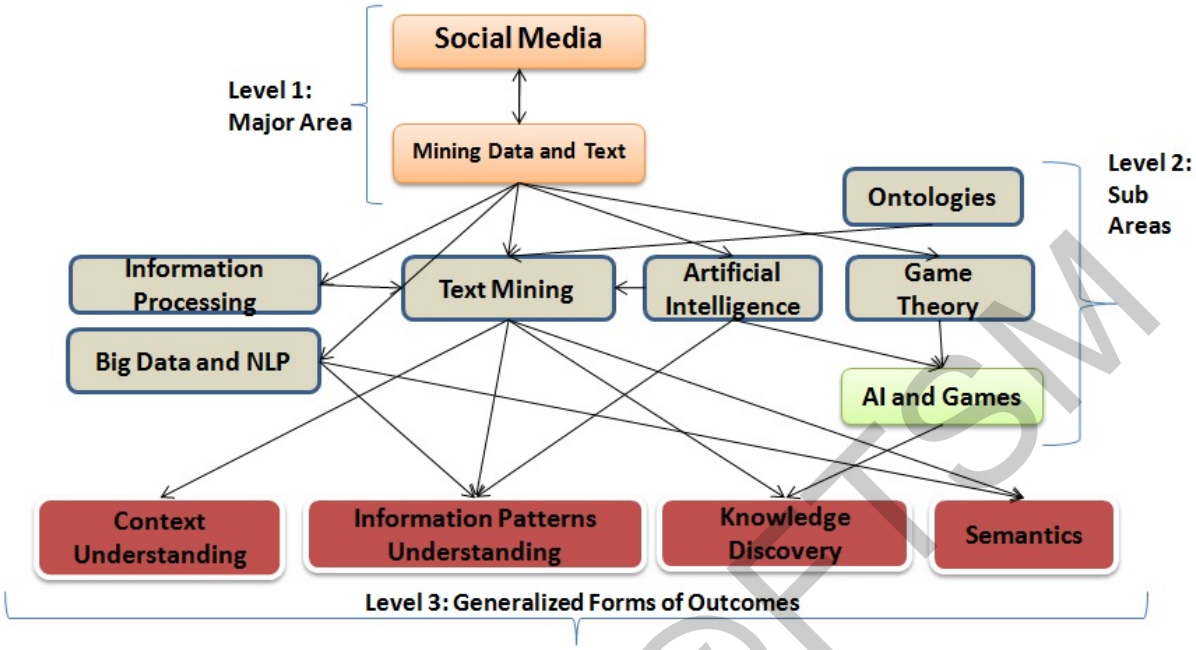

FIGURE 1. Level-wise Distribution of Study Areas and Outcomes,

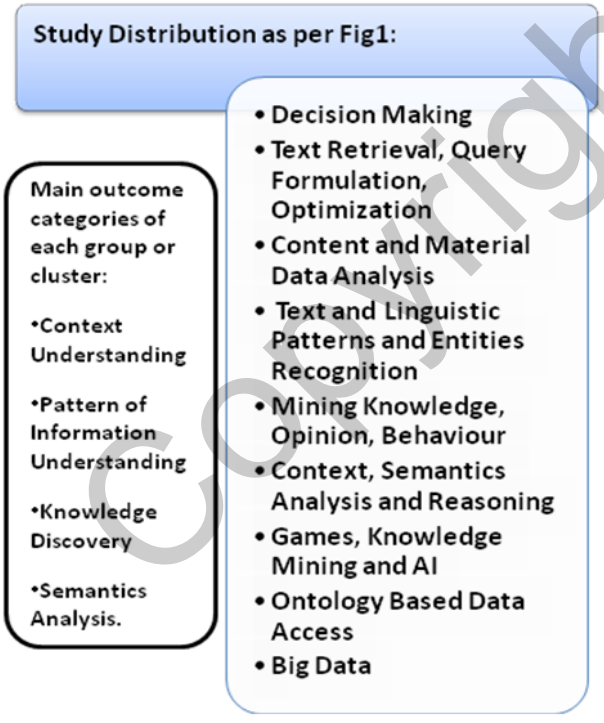

FIGURE 2. Domain Distribution of Research Papers.

A. WEATHER FORECASTS

The decision-making process depends on context assessment by examining uncertain text-data, reliability estimates and relevance probabilities as per the nature of the online text. This study compares different types of informative assessments of uncertain data. The researchers have measured the influences of uncertain data on the decision-making in the field of weather forecasts. They proved Natural Language Processing (NLP) is a tool to enhance decision-making under conditions of uncertainty. They also enhanced NLP application as compared to the previous practice of just using it as a state-of-the-art graphical presentation method [5].

B. FORENSIC EXAMINATION OF CYBER-CRIMES

Researchers focused on online social cyber-criminal "dark markets". Their proposed method detects online cyber-crime networks and facilitates forensic examination of online cyber-crimes. The Probabilistic generative model is designed, developed, and reinforced by the context-sensitive algorithm using Gibbs sampling [6].

C. SEMANTIC WEB & DECISION SUPPORT SYSTEM

This article presents a structured review of the literature on semantic web technologies in 'Decision Support System (DSS).' The researchers based their analysis on the interview results of DSS staff and developers belonging both to industry & research organizations. For study validation, the literature survey is conducted in a structured fashion and data is taken from databases, reputable conferences, and journals. All this data relates to two main topics i.e. semantic Web and DSS. The competitive results are generated in a dual assessment form, based on a set of relevant keywords, representing the intersection of semantic Web and DSS [7].

D. GENERATING QUESTIONS FROM ONTOLOGIES

In this study, an automated method is proposed for generating multiple-choice questions by existing ontologies under given domain ontology. The aspects investigated are: (1) Finding the complexity-levels of generated tests (2) The heuristic rules identification to select a small relevant set of questions, related to any subject domain (3) Designing a test with 'high', 'medium' or 'low' difficulty levels. The study proved capable of resolving similar content generations issues. This work helps not only to decide the difficulty level but also helps to verify it by the scores mentioned in the ontologies [8].

E. SERIAL CRIMES CORE

NLP based crime prediction provides an efficient text assessment way to decision-makers regarding the likelihood of future possible crime possibilities. These tools are gaining popularity for online crime mitigation. Authors examined several equity criteria recently applied to assess the types of serial crimes identification. Their suggested criterion simultaneously satisfies all prediction constraints, for example, if the occurrences of specific crimes do not decrease in crime groups over time. The time-line study combined with the principles of NLP is their major claimed contribution. They developed the hidden core of crime-series, presented as the driving factor in their research [9].

F. CRIME PATTERN SIMILARITY

The aggregate of crimes data is an important thing to determine common crime pattern similarities. The issue is assessed from past criminal histories, based over the pattern similarity history indexing. The proposed method used pattern detection models for identifying a series of crimes. This process carried out by attention on existing pattern differences relating regularities of crime investigation domains [10].

G. KAVLI HEALTH PROJECT POTENTIAL

Kavli Health Project (KHP) is intended for the medical decision-maker in assessment and measuring community disease risks. The study uncovered the poor behavioural measurements about diet, smoking, and lack of exercise etc., which limits understanding of preventive health measures. The Kavli Project consists of a wide range of areas deeply explored to measure behavioural phenotypes related to medical decision making [11].

H. KAVLI HEALTH PROJECT

KHP explored from the synoptic and granular point of views about human health. The life cycle proposed in a way that how health issues evolve differently for different people. The authors argue to develop new scientific approaches in the analysis of big data for decision making relating to medical practices [12].

I. QUANTIFIED SELF (QS)

Companies offer a variety of new platforms both in the form of hardware and software allowing tracking of increasingly quantified aspects of everyday life referred to as QS. With the rise in the number of qualified participants in the collection of certain types of data, now firms own even more data than ever before. Ultimately, firms need evolutionary decision-making methods to know what to do best. This article discusses the possibilities, potentials and problems relating to QS [13].

J. DECISION-MAKING ON FUTURE EVENTS

Online text processing firms process explicit links of Twitter for future events prediction and try to provide an automatic monitoring system for real-time business-related events. This study describes a system that extracts chat events from Twitter and defines future possibilities about entities in tweets, relating to clusters of overlapped tweets in a certain threshold. The study focuses on summarized clusters of events for users and assesses the conditions for describing the events occurring in an uncontrolled manner [14].

K. DECISION MAKING AUTOMATION

The Online Global Society increasingly relies on predictive data models in the automation of decision-making. By the existence of noise in the online text observations, these models can systematically infringe on decision-making. Various studies depend on static observations and these studies have limited flexibility in decision making. Rather than relying on specific values, this research is primarily intended for data mining with principles of machine learning for the development of integrated performance criterion for better predictive modelling [15].

L. SENSOR-BASED INTELLIGENT HEALTHCARE MONITORING FRAMEWORK

There exists a dire need to expand the availability and improve the quality of medical services for an ever-increasing global population. Especially the elderly population, who need smart IT-solutions to decrease unit costs for medical care. Medical costs increase with an increase in the volume of financing of the health care system. Modern techniques like sensors embedded SMM tools help to search for solutions for the diagnosis and treatment of diseases like diabetes, that can cover broad segments of the population. An effective solution to this problem is the use of ML-based social media tools & technology for monitoring the state of human health for the prevention of diseases and predicting the risk of chronic diseases [109].

M. HEART DISEASE PREDICTION BASED ON ENSEMBLE DEEP LEARNING AND FEATURE FUSION

This study uses a multi-level structure of the ML system for heart disease, in which each level of the system will provide an optimal solution to problems and achieve the target function of the level. It uses intelligent algorithms

TABLE 1. Studies relating to the Domain of Decision Making

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Gkatzia, Dimitra et al. (2017)	Compares information presentations.	Weather forecasts. Natural language + Image processing.	A, B	5
2	Lazarus, Suleman et al (2019)	Online social cybercriminal "dark markets".	The Probabilistic generative model designed.	A, B, C	6
3	Blomqvist, Eva. (2014)	A structured review of the literature.	Semantic Web and DSS.	A, B, C, D	7
4	EV, Vinu et al. (2017)	Deciding to generate multiple-choice questions with a certain difficulty level from ontologies.	Heuristic rules-based Intelligent content generations.	A, B, D	8
5	Choledochal, Alexandra. (2017)	Prediction of serial crimes.	Equity criteria applied to assess serial events identification.	A, B, D	9
6	Wang, Tong, et al. (2015)	Crime pattern similarities indexing.	Identifying series and regularity of domains differences.	A, B, C	10
7	Ausiello, Dennis et al (2015)	Measuring community diseases risks, and medical decision-making.	Kavli Health Project (KHP) strength evaluated.	A, B, C, D	11
8	Azmak, Okan, et al. (2015)	KHP explored from the synoptic and granular point of views.	Kavli Health Project strength evaluated.	A, B, D	12
9	Fawcett, Tom. (2015)	Quantified aspects of everyday life (QS).	Evolutionary decision-making methods proposed.	A, B, D	13
10	Kunneman, Florian et al. (2016)	Twitter for future events prediction for decision-making.	Focused summarized clusters in descriptions of events.	A, B, D	14
11	Žliobaitė, Indrė. (2017)	Automation of decision-making.	Machine learning integrated, performance predictive modelling.	A, B, C, D	15
12	Ali, Farman, et al. (2020)	Medical decision making	Machine learning integrated with sensors and Social Media Communication.	A, C, D	109
13	Ali, Farman, et al. (2020)	Heart disease real time monitoring.	Heart devices linked with social media & health centers.	A, C, D	110
14	Raza, Mohsin, et al. (2020)	Real-time disaster monitoring through social media.	Real-time NLP processing of social communication applications such as Facebook to act in case of a public disaster.	A, B, C, D	111

for the operation of the levels of the system consisting of the inclusion of the number of sensors for the registration of biomedical signals. It targets the sampling rate of the recorded digital data and helps in the decision-making criteria. The suggested process cares about the boundaries of the change in the individual norm of medico-biological indicators, taking into account the changing state of the patient's heart data. It provides informational support of the patient when the patient's medical and biological indicators go beyond the boundaries of the change in the individual norm [110].

N. EFFECTIVE COMMUNICATIONS IN DISASTER BY SMM

Currently, the use of artificial intelligence in solving social problems is limited by the availability of data and the lack of specialists in the field of AI. In solving infrastructure problems, AI applications include smart traffic lights to maximize road capacity, planning preventive maintenance of public transport systems, and identifying potentially malfunctioning public infrastructure components in case of any disaster. Using AI, the system processes a huge amount of incoming data in real-time and analyzes it to determine the likelihood of additional problems and the need for different services to work together. For example, if a power supply cut occurs due to a fire, you need to call not only firefighters but also the electric utility. This study suggests a smart AI advisor. Using AI, advanced statistics, and machine learning, this smart advisor applies SMM over real-time data to fill blind spots and alert agencies to the potential onset of complex emergencies, from large events to linked incidents. By detecting patterns and anomalies sooner, agencies can act faster and coordinate more effectively to reduce the effects on communities, resources, and staff. When a potential event needs detection, the system detects it in real-time from social media and it notifies users who then decide whether to act on, share or dismiss the report. By detecting more crucial connections sooner, public safety agencies become better equipped to create safer, more resilient communities for citizens through everywhere available social media [111].

III. TEXT RETRIEVAL, QUERY FORMULATION, OPTIMIZATION

This section is based on better query processing for improved and enhanced text retrieval capabilities. This section covers contents such as the engineering perspective of compression related to data mining, text-retrieving techniques for information contained in program artefacts, methods of text searches and NLP in software engineering, attribute family and fuzzy relationships, implementation of a 'Black Hole Algorithm', geo-social k nearest query, semantic analysis of documents and performed ranking of Wikipedia pages, unstructured information market analysis, blogs contribution in the democratic public sphere, social networks relating young multilingual people on Facebook, interpretation of news values by comparison of written language and image modules, etc., [16]–[26]

TABLE 2. Studies relating to the Domain of Text Retrieval, Query Formulation, Optimization

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Oswald, C., Anirban I. Ghosh et al. (2015)	Engineering perspective of compression related to data mining.	Huffman coding refined by including frequent set of data mining elements.	A, B	16
2	Haiduc, Sonia. (2014)	Text-retrieving technique of information contained in program artefacts.	Developer's hard time solved while understanding code with Good or Bad queries.	B, C	17
3	Bhardwaj et al (2020)	Methods of text searches and NLP in software engineering.	Discussed their applications in various fields.	A, B, C, D	18
4	Boixader, D., J. Recasens. (2017)	Attribute family and Fuzzy relationship.	A method proposed for identifying families of new attributes with fewer elements generates the same similarity indexes.	B, C	19
5	Hatamlou, Abdolreza. (2013)	Implementation of 'Black Hole Algorithm (BHA)'.	A heuristic population-based algorithm based on the BHA phenomenon presented.	A, C, D	20
6	Shim, Changbeom, et al. (2018)	A new type of geo-social query called K nearest l-close query to friends.	Three approaches used: neighbour cell search, friend-cell search, and personal-mobile search.	B, C	21
7	Dornescu, Iustin et al. (2014)	Semantic analysis of documents and performed ranking of Wikipedia pages.	Used already existing available links regarding ranking articles relevant.	B, C, D	22
8	Ferrucci, David et al. (2014)	Unstructured information market analysis.	The proposed architecture of middleware for processing unstructured information.	C, D	23
9	Mummery, Jane et al. (2013)	Focused media specifically blogs contribution in the democratic public sphere.	Examined three major Australian politics-oriented blogs.	A, C, D	24
10	Androutsopoulos, Jannis. (2014)	Social networks relating young multilingual people on Facebook focused.	Proposed improved negotiation and learning strategies of language selection and its effects.	B, C	25
11	McKeown et al (2020)	Interpretation of news values by comparison of written language and image modules.	Model for joint image and text processing proposed for better news understandings.	B, C	26

A. DATA COMPRESSION

The article targets the perspective of compression in data mining. There are various techniques of data mining which require huge storage for Big Data. These algorithms are improved by applying compression tools. In the current study, efficient data compression by Huffman coding is applied and further refined, by including frequent sets of data mining elements. This combination improved data mining skills as an additive step in the query optimizing association of rules [16].

B. PROGRAM ARTEFACTS AND NLP

In this study, they review a text-retrieving technique used for textual information, which contained program artefacts, and is used to help developers solve their everyday tasks. Developers face a difficult time while understanding code and carrying out the search results. "Bad" queries result in wastage of time and effort. Querying at previously needed inessential information, result in a bottleneck in query processing. The current approach discarded redundant data which makes for an improved query. The researchers reformatted the query without certain pointers as to the improvement in this technique [17].

C. NLP SURVEY FOR SOFTWARE ENGINEERING

In this paper, a technical briefing is presented over modern methods for text searches. NLP is used in software engineering and discussed with its applications in various broader areas. This research pointed out the need for NLP in the requirement engineering phase and relating its patterns in the next phases of the software development life cycle. Authors suggested the NLP based software testing over the software configuration reports. This work provided the framework of optimal NLP techniques useful in everyday software practices [18].

D. ATTRIBUTE FAMILY AND FUZZY RELATIONSHIP

The study considers the average similarities in the family of attributes of a software model. The attribute similarities help to classify the inputs and outputs of a software system. In this research a method proposed for identifying families of new attributes that have fewer elements and could generate the same similarity indexes. More generally, this article considers the structure of the class of fuzzy relationships in developing more accurate results [19].

E. BLACK HOLE (BH) ALGORITHM

Researchers presented a heuristic algorithm based on the BH phenomenon. It is a population-based algorithm. It starts with the initial or basic elements about a population. It searches for solutions regarding optimization of social problems. It focuses objective or intent functions, estimates and initials. Researchers worked over the self-balancing property of this model and justified its usefulness in real-life scenarios [20].

F. K NEAREST CLOSE QUERY

In this paper, the author defined a new type of geo-social query called K-nearest-close-query. It returns K nearest data objects from the number of user hops in the user's friend requests. Authors listed three approaches of processing: neigh-

bour cell search, friend-cell search, and personal-mobile search. In this work, they proposed the optimal path selection principle in the miscellaneous entities based on the nature of their relationship [21].

G. SEMANTIC INDEXING

Researchers proposed semantic analysis of documents and performed ranking of Wikipedia pages. They presented similarities established by tasks like wikification, binding of named identified objects etc., These techniques lack tight rules in making of standard uniformity regarding the named entities identification. This study used already existing available links in ranking more relevant articles to the specific document. It describes an explicit form of semantic indexing which allows a semantic search useful in NLP applications [22].

H. UNSTRUCTURED INFORMATION MARKET ANALYSIS (UIMA)

UIMA represents the growing need for processing unstructured information. This study led to the development of the architecture of middleware for processing unstructured information. It is based on powerful search capabilities, data-driven development, compilation, and distributed deployment of analytical systems etc., In this article, researchers explored UIMA by focusing on the design points of architectural analysis [23].

I. DEMOCRATIC PUBLIC ONLINE SPHERE

Researchers studied and explained how mass media can serve consumers' interests in multiple dimensions. They justified the public sphere of the media, capable to achieve more than only satisfying common market demands. The study focused specifically on blogs that how these can contribute to the democratic public sphere of responsibility by the deliberative exchange of ideas. Authors examined three major Australian political blogs. They worked on classification, text similarity in opinions and ways to express ideas over these blogs. They have presented various matrices of their study as well as various interview evaluation criteria [24].

J. MULTILINGUAL LANGUAGE SELECTION AND NEGOTIATION

This document explored empirical data from social networks relating young multilingual people on Facebook for improved negotiation and learning strategies of language selection. Relying on the sociolinguistic basis of design, they made research over sociolinguistics multilingualism analysis. Researchers presented that the choice of language is significant in social initiating and responding. They also focused on metapragmatic language style negotiations and the role of the English language as a dominant resource among the network of online blog writers [25].

K. INTERPRETATION NEWS BY WRITTEN LANGUAGE

In this article, the authors investigated the interpretation of news values by comparison of written language and image modules. They emphasized attitude assessment and positioning of text and images. From their point of view, news values are not only the beliefs that a journalist adheres to, but these are also the social values created by the choice of language and images as well. Authors proved that it is necessary to pay attention to both i.e. text and images to gain a better understanding of events in becoming "news" [26].

IV. CONTENT AND MATERIAL DATA ANALYSIS

This section contains the studies over the data analysis topics for context understanding and related improvements. This section covers topics like common sense and linguistics calculations for data flow, an effective approach for detection of emotion polarity in natural language text, under-represented data with its imbalanced distribution, scientist's online conversation ethical quality check, NLP used for 'listening devices', conflicts in academic blog discussions, qualitative analysis of interactive interaction objects, the concept of "neutralism" of interviewing strategies in the UK and USA, four methodological issues relating to socio-linguists online data, appropriation and evidence in the news discourse, etc., [27]–[36].

A. COMMON SENSE AND LINGUISTICS CALCULATIONS FOR DATA FLOW

Researchers worked on methods of computational intelligence combined with calculations of common sense and linguistics for the analysis of data flows. Authors proposed to automatically decipher on people's emotions, sentiment and opinions expressed by natural language. This study takes the opinion bias as the pivot for the entire study. Text selection becomes difficult to justify, in the absence of assessing the language writing styles [27].

B. RECOGNITION OF EMOTIONS, DETECTING POLARITY IN NL TEXT FOR ELM

This research presents an effective approach in the recognition of emotions and detecting polarity in natural language text. The study gives a direction to how to use NLP tools for achieving a breakthrough in the field of statistical learning theory. It guides on how to effectively build the NLP educational machine. The NLP educational machine is used in estimating the productivity of the designed models in the analysis of a large pool of social data [28].

C. IMBALANCES, UNDER-REPRESENTED DATA AND LEARNING ALGORITHMS

The presence of under-represented data and its imbalanced distribution effects the performance of machine learning algorithms. It results in inherent complex characteristics and algorithm pieces of training. It requires an effective transformation of a huge amount of data into the correct presentation

TABLE 3. Studies relating to the Domain of Content and Material Data Analysis

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Poria, Soujanya, et al. (2015)	Common sense and linguistics calculations for data flow.	Automatic decipher people's emotions, sentiment and opinions expressed.	A, B	27
2	Oneto, Luca, et al. (2016)	An effective approach for detection of emotions polarity in natural language text.	statistical learning theory (SLT) for effectively building an extreme educational machine (ELM).	B, C	28
3	He, Haibo et al. (2009)	Focused on under-represented data with its imbalanced distribution.	Effective transformation methodologies of huge data into the correct presentation compared.	A, B	29
4	Neff, Gina, et al. (2017)	Scientist's online conversation ethical quality check.	Socio-material mechanisms for conversation data generation presented.	B, C	30
5	Dale, Robert. (2017)	NLP used for 'Listening Devices'.	Proposed algorithm to significantly improve speech recognition.	A, B, C	31
6	Luzón, María José. (2013)	Focuses on the conflicts in academic blog discussions.	Analyzed: frequency of conflicts, strategies, purpose of the conflict.	C, D	32
7	Giles, David, et al. (2015)	Qualitative analysis of interactive interaction objects.	Analyzed the archived materials.	A, C, D	33
8	Tolson, Andrew. (2012)	Concept of "neutralism" of interviewing strategies in the UK and USA.	Political interviewing strategies proved partial regarding failing in "neutralism".	A, B	34
9	Bolander, Brook et al. (2014)	Four methodological issues relating to socio-linguists online data studied.	Moral values, multi-modality, methodologies concerning online and offline settings, annotations of web data.	B, C	35
10	Schafer, Svenja (2020)	Appropriation and evidence in the news discourse.	A functional linguistic paradigm focused.	A, B, C	36

of information and knowledge. This paper presents a comprehensive review of such developments relating to unbalanced data [29].

D. PRODUCING ETHICAL DATA SCIENCE

This article focuses on scientist's online conversation, criticizing etc., relating to scientific data. This study identified the social and organizational mechanisms for ethical data in online scientific debates. Authors summarized the four usual criticisms: the data is inherently prone to needing interpretations, the data is inseparable from the context, the data is always based on the socio-material mechanisms that produce them, and the data serves as a vehicle for negotiation and transfer of values. Authors presented qualitative research with scientific data of similar projects, data of specialized and interdisciplinary engineering groups [30].

E. NLP AND LISTENING DEVICES

The researchers proposed an algorithm, which can significantly improve speech recognition, reinforced by improvements in the ability to comprehend recognized speech. It added NLP as a tool for smart speakers and other devices etc., This work is good for small phrases and helps in expressing basic emotions. Authors proved that these devices benefited from their work [31].

F. INTERPRET CONFLICTS IN ACADEMIC BLOG DISCUSSIONS

The article focuses on the conflicts in academic blog discussions in the comment sections. It analyzes strategies used to interpret conflicts and its impact on the average features, and determine how the conflict could have a new identity on the Internet. Studies consisted of discussions of nine academic web journals. Authors analyzed data in the following aspects: (i) the frequency of conflicts in each academic web-blog (ii) strategies used to interpret the conflict (iii) the purpose of the conflict. The analysis showed a high prevalence of conflict in the discussions of academic blogs, ranging from mild criticism to disagreement and more severe expressions of conflict, such as bold criticism, difficult questions or even insults in some cases [32].

G. QUALITATIVE ANALYSIS OF INTERACTIVE INTERACTION OBJECTS

This article related to micro-analysis of online data. Researchers explored the possibility of conducting a qualitative analysis of interactive objects. 'Text engineering methodologies' is used in analyzing the archived materials, which had previously been made visible to millions on internet users. This analysis serves as the framework for estimating the larger or smaller elements of online text. This technique tries to relate the loosely related information elements with previously believed major attributes in online information [33].

H. ANALYSIS OF INTERVIEW STRATEGIES

In particular, the concept of "neutralism" is central to the analysis of interviewing strategies in the UK and USA for all types of public interviews but not specifically for political interviews. The study classifies political interviewing strategies. Its research finding turned quite different from general interviews. The results of this study is based on text analysis techniques. As a future recommendation, researchers presented a framework of NLP based interviews, particularly with a built-in self-verification mechanism [34].

I. SOCIOLINGUISTS ON ONLINE DATA

This article focuses on four methodological issues relating to sociolinguists online data values and their associated use: (1) moral values (2) multi-modality (3) mix of the relationship of methodologies between online and offline settings (4) annotations of web data. The study revealed that although there are many publication ethics for communicating sociological research, fewer publishers focus on empirical linguistic research recommendations in this direction [35].

J. PIECES OF EVIDENCE IN THE NEWS DISCOURSE

Authors considered previous studies regarding online texts based on appropriations and evidence in the news discourse. They suggested improving the understanding of the text by referring to some key concepts, which they referred to as the evaluation framework. It is an approach of evaluation of language developed within the framework of the functional linguistic paradigm [36].

V. TEXT AND LINGUISTIC PATTERNS AND ENTITIES RECOGNITION

This section presents researches which focused on text patterns and considerations about entities in written text. This section covers issues like recognizing named entities by text structure of Wikipedia, multiple instances of learning monotone predicate problems, finding primary schools using minimum quotas, calculation of sets of logical program responses, robust unsupervised spectral feature selection, heterogeneous analysis of the information network, eliminating erroneous interpretations and including negation, machine translation methodologies, unfolding life events by small stories, etc., [37]–[46].

A. RECOGNIZING NAMED ENTITIES BY TEXT STRUCTURE OF WIKIPEDIA

The study suggests automatically creating multilingual good standard pre-trained annotations. These annotations are used for recognizing named entities using the structure of Wikipedia text. By taking additional links and heuristics, researchers proved that good automatic annotation can be developed with better standards. This work proved efficient in exploring relations within named entities [37].

B. MULTIPLE INSTANCES OF LEARNING (MIL)

Researchers have shown that MIL is important in pattern recognition but effective solutions are lacking in this field. Comparative sorts of studies are not available to highlight the behavioural characteristic methods in this regard, so this article focuses on the problems relating to such classification. People inter-mix the concept of name entities and multiple instances which is not an effective way of understanding the entities behaviour. This paper has uncovered various similar issues relating to multiple instances [38].

C. MONOTONE PREDICATE PROBLEM (MPP)

This work provides new ways of solving known functional problems, including simple implicates, literals, independent variables etc., in text engineering. Moreover, this work targets the development of efficient algorithms. The paper outlines several areas for future research relating to the expansion of the problem of MPP. The current study has related this issue to the use of pre-defined variables in text extraction. They carried out text assessment comparison by previously applied functional procedures [39].

D. FINDING PRIMARY SCHOOLS USING THE MINIMUM QUOTAS

Researchers proposed an online algorithm capable of providing primary schools selection information, meeting all the minimum requirements of students. Authors developed two strategy-proof mechanisms i.e. first is the best trending cycles and the second based on the leading trending cycles. This work suggested ideas to consider multiple school places with less observing and processing online available quotas. This technique served as the NLP based recommender system and its framework proved applicable in various similar situations of places recommendations [40].

E. CALCULATION OF SETS OF LOGICAL PROGRAM RESPONSES

This approach performed calculations over the sets of logical program responses and targeted the concept's verification of answer sets. Authors developed the algorithm based on the answer set programming for the first time. Their approach proved that logic responses are more efficient and they evaluated various methods for dynamic programming in the described directions. Their technique applied cause and effect approaches in user response presentation [41].

F. ROBUST UNSUPERVISED SPECTRAL FEATURE SELECTION

Researchers proposed a reliable approach for the self-expression of functions, which eliminates the effect of nonessential text elements. A low-rank constraint with optimized parameters list designed on the weight matrix. This list preserved the global structure among the samples based on the correlation between the text traits. Both local and global correlations between objects used to study the dynamic matrix of the internal space of the original data to preserve the local structure among the samples [42].

G. HETEROGENEOUS ANALYSIS OF THE INFORMATION NETWORK

In this article, the authors presented the basic concepts of heterogeneous analysis of information networks and considered solutions to various problems in data mining. They classified heterogeneous networks as per text mining techniques. They listed the causes and parameters of certain variations which help to identify similar and dissimilar information networks. They also pointed out future possible directions [43].

TABLE 4. Studies relating to the Domain of Text & Linguistic Patterns and Entities Recognition

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Nothman, Joel, et al. (2013)	Recognizing named entities by text structure of Wikipedia.	Suggested automatically creating multilingual good standard annotations.	A, B	37
2	Amores, Jaume. (2013)	Multiple Instances of Learning.	Behavioural characteristic classification methods discussed.	B, C	38
3	Marques-Silva, Joao et al. (2017)	Monotone predicate Problem.	Suggested new ways of solving known functional problems.	A, B, C	39
4	Hamada, Naoto, et al. (2017)	Finding primary schools using the minimum quotas.	An algorithm capable of providing primary schools selection information, meeting all the minimum requirements of students.	A, B	40
5	Gebser, Martin et al. (2012)	Calculation of sets of logical program responses.	Logical program responses presented based on concepts of verification of answer sets.	A, B	41
6	Zhu, Xiaofeng, et al. (2017)	Robust unsupervised spectral feature selection.	Eliminated the effect of non-essential functions.	B, C	42
7	Park, Sangwon (2020)	Heterogeneous analysis of the information network.	Related to data mining problems.	A, B	43
8	Blanco, Eduardo et al. (2014)	Eliminating erroneous interpretations and including negation.	Negative-expression fixation model by detecting and identifying implicit concepts of positive values.	B, C	44
9	Rapp, Reinhard et al. (2016)	Machine translation methodologies explored. The study is in the form of a literature survey.	The semantics focused dimensions presented with the concept of word attachments.	A, B, C	45
10	West, Laura E. (2013)	Unfolding life events by small Stories. Blogging and Facebook capered.	Proved that blog can perform differently from Facebook.	A, B	46

H. ELIMINATING ERRONEOUS INTERPRETATIONS AND INCLUDING NEGATION

This paper presents a negative-expression fixation model by detecting and identifying implicit concepts of positive values. Negative proposition is logically presented depending on the scope and orientation of the negation. The new approach determines the direction of negation and thus eliminates misinterpretations. Researchers included it as part of the construction of semantic relations. It gave a rich semantic representation of text, including hidden pins. [44]

I. MACHINE TRANSLATIONS HIGHLIGHTS

The article contains developments in the field of automatic translation. Authors highlighted the updated definitions of previous machine translation methodologies and examined the related bilingual models of language translations. They developed neural networks and explored them to get the hidden representation of the context with a smaller number of dimensions. Their proposed semantics uses dimensions concerning the concept of word attachments. [45]

J. UNFOLDING LIFE EVENTS BY SMALL STORIES

This article targets unfolding live events e.g., buying, selling, painting etc., on Facebook through small story messages. These messages have a huge impact on opinion-making. Online users prefer short texts with one issue in two to three lines. With small messages, someone might choose to simultaneously share the same event in a blog with a similar small description. The author compared the real-time social goals and proved that a blog can also perform differently for Facebook users in multiple regards [46].

VI. MINING KNOWLEDGE, OPINION, BEHAVIOUR

This section covers opinion and behaviour understanding researches. This section highlights topics such as the determination of hybrid knowledge base, the model of social contact networks in Wikipedia, detecting personal online risks in one-time classification, social agents subgroups, modularity technique based theories, non-parametric hierarchical Bayesian model, formatting collective classifier on data transformations, the significance of public health data over the internet, the existence of heterogeneous behaviour less probably related to network density, the algorithm of structured learning, extracting automatic probabilistic logical positions of the opponent's strategies, including human interpreted formations, the fuzzy system to extract and analyze statements, shrinking procedures to save time and cost by the dependence map, etc., [47]–[59]

A. SEMANTICS FOR HYBRID KNOWLEDGE BASES

Authors characterized three-valued knowledge base models in stable sub-processes over sub-sections of long text. They evaluated partially stable models by logic programming. The intuitively understandable concept of stable sections is facilitated to discover the theoretic-proof method, for the existence determination of hybrid knowledge base. Authors performed partial partitions and tested their stability by calculating the variable points. They also performed various consistency checks over partial stable models of logic programming [47].

B. RELIABLE INTERPRETATION OF SOCIAL CONTACT NETWORK

The main contribution of the article is a model of a social contact network in Wikipedia. It used to derive behavioural data and reliable interpretation for the social contact network. Other materials also discussed as improved versions

TABLE 5. Studies relating to the Domain of Mining Knowledge, Opinion, Behaviour

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Liu, Fangfang, et al. (2017)	Authors characterized three-valued MKNF models.	Determination of hybrid knowledge base.	A, B	47
2	Jankowski-Lorek, Michał, et al. (2016)	Reliable interpretation of social contact network.	Model of social contact network in Wikipedia.	B, C	48
3	Barrera-Animas, Ari Yair, et al. (2017)	Risk detection one-time classification problem.	Detecting personal online risks in one-time classification.	A, B	49
4	Zhao, Yiyi, et al. (2017)	Social agents study in the social network.	Divides social agents into two subgroups: 'leaders of public opinion', 'followers of public opinion'.	B, C	50
5	De León, Hernán Ponce, et al. (2018)	Unwanted process behaviours detection.	The modularity technique based theories.	A, B, C	51
6	Liu, Yezheng, et al. (2018)	Non-parametric hierarchical Bayesian model.	Suggests an improved non-parametric hierarchical Bayesian model.	B, C	52
7	Perez-Chacón, R., et al. (2020)	Forecasting and data transformations.	Combined two principles to test the hypothesis for formatting collective classifier on data transformations.	A, B	53
8	Akbari, Mohammad, et al. (2017)	Health analysis Study over the Internet.	significance of public health data over the internet.	B, C	54
9	Buskens, Vincent, et al. (2016)	Characteristics concentrated data parameters relating heterogeneity of social behaviour.	Existence of heterogeneous behaviour less probably related to network density.	A, B, C	55
10	Lippi, Marco. (2016)	Probabilistic logical positions describing an opponent's strategies.	Algorithm of structured learning, extracting automatic probabilistic logical positions of the opponent's strategies, including human-interpreted formations.	A, B	56
11	Krestel, Ralf, et al. (2014)	Specific single topic data exaction.	The fuzzy system presented to extract and analyze statements.	B, C	57
12	Church, Kenneth. (2017)	Measuring having 'Good and Bad'.	Students pressure to publish more articles.	A, B	58
13	Lee, Kichun, et al. (2013)	Dependence map of distance values.	Shrinking procedure to save time and cost by the dependence map.	B, C	59

of behavioural networks, based on behaviour and history of discussions in Wikipedia [48].

C. RISK DETECTION ONE-TIME CLASSIFICATION PROBLEM

Researchers have shown how to develop a mechanism for detecting personal online risks in one-time text classification. Authors trained several classifiers based on the daily occurrence of certain online subjects. As the focus of this research, authors checked the accuracy of the classifiers to identify anomalies that previously not included in the classifier training process [49].

D. SOCIAL AGENTS STUDY IN SOCIAL NETWORK

In light of the influence of opinion study, this article divides social agents in a social network into two subgroups: 'leaders of public opinion' and 'followers of public opinion.' Then Authors created a new 'dynamic model of trust' as a roadmap for opinion leaders and followers to calculate the evolution of the opinion impact for the group of various types of communicating agents [50].

E. UNWANTED PROCESS BEHAVIOURS DETECTION

In this paper, researchers presented the behaviour detection process. This function can be critical for those processes, which are less complex. These functions are appropriate in many situations but these are not completely accurate. However, these are good for generalizing correct behaviour in the underlying log event. This technique is based on modularity theories combined with other process detection approaches. This is carried out as a post-processing step to simplify complex models [51].

F. NON-PARAMETRIC HIERARCHICAL BAYESIAN MODEL

The article suggests an improved non-parametric hierarchical Bayesian model. It investigated some generative relationships with internal factors preferences. The proposed model used a three-level generation system based on the influence of internal factors on user preferences. The current research on Facebook data showed that the said model can mark valuable hidden aspects from social contents. It identified the internal motivations of user behavioural choices [52].

G. FORECASTING AND DATA TRANSFORMATIONS

Authors combined two principles to test the hypothesis for formatting collective classifiers on data transformations to increases the accuracy of classification in time series. These classifiers are built over two main domains i.e., a timeline of information generation and frequency of information change over this timeline. For the time series analysis, the authors used a set of measures as elastic distances. The textual data grouped and marked for communicating as per the timeline of information frequency [53].

H. WELLNESS AND HEALTH STUDY OVER THE INTERNET

Authors studied the significance of public health data over the internet and suggested a method to consider two types of wellness knowledge: (1) the temporal sequence of health attributes (2) the heterogeneity of wellness characteristics in the patient data in a large population. Authors conducted extensive experiments to evaluate their proposed structure in addressing three main objectives: predicting health attributes, defining community health & wellness and the community health-related medicine issues. Authors showed experimental results for two real data sets and practically demonstrated the ability of their proposed approach in making effective representations of health-related issues [54].

I. HETEROGENEOUS SOCIAL BEHAVIOR STUDY

Researchers studied the characteristics of concentrated data parameters by relating them to the heterogeneity of social behaviour in online comments. The authors showed that heterogeneity behaviour is retained if the knowledge network is more segmented and less centralized. As a major research finding, the authors proved that the existence of heterogeneous behaviour relatively & less probably, relate to network density [55].

J. PROBABILISTIC LOGICAL POSITIONS DESCRIBING OPPONENT'S STRATEGIES

Algorithms for logical inferences used to solve the problem for opponents to find Nash and Pareto-optimal solutions. Algorithms of structured learning used to extract automatic probabilistic logical positions describing the opponent's strategies with the help of human-interpreted formations. Experiments conducted using Markov logical networks [56].

K. SINGLE TOPIC DATA EXACTION

The growing number of publicly available sources of information makes it impossible for individuals to keep track of all the different opinions of a single topic or single very specific entity. The fuzzy system is used to extract and analyze statements of opinions from newspaper articles. These systems are modelled by using the theory of fuzzy sets, applied after extraction of information by natural language processing. Fuzzy models are used to accept or reject a hypothesis based on a set of customizable strategies. These strategies are meant to focus on the very basic or main issues in a given text [57].

L. MEASURING HAVING THE GOOD AND BAD

Authors stated that students are currently under pressure to publish research articles quickly and ultimately have more publications than ever before. Researchers publish too much these days, over which the authors of this paper raised the questions i.e., have people time to originate great scientific thoughts and spend time to learn more about things, at such a large scale. They pointed out some cases in which previous researches by the same authors may not be directly relevant to the next publications etc., They suggested that there are long-term macro trends in the pace of publication, which are beyond the control to classify it as good or bad. These trends last for years and will continue [58].

M. DEPENDENCE MAP

Authors introduced the concept of internal distance between the points in the point-wise expansion of the statistical dependence between variables. These variables are text quality, names entities relations, current trends like political issues, etc., Authors focused on shrinking the procedure to save time and cost by the dependence map. Its theoretical base is linked to other similar methods and linked with the empirical behaviour of named entities in real data sets. [59]

VII. CONTEXT, SEMANTICS ANALYSIS AND REASONING

This section contains the reasoning based studies of context and semantics like product related user's reviews polarity, online-contents quantitative reasoning, mining hidden definition of aspects and mood, calculating the semantic similarity of different sentences, aggregate sentiments of certain social subjects, knowledge collection from web corpus, social semantic web and data paradigm, contextual and semantic information extraction, representing preferences as part of user-profiles, commercial use of NLP survey, Twitter joke trial, analysis of the news contents, etc., [60]–[71].

A. PRODUCT REVIEWS POLARITY

User reviews about product distributions have polarity in their ratings. Therefore, these estimates based on different products reviews are mostly numerically skewed. The study showed that interim reviews have proven potentially useful by performing user training over product overview. Researchers suggested using a consistency model to introduce temporary relationships into user and product views to improve the effectiveness of document-level user's mood analysis [60].

B. ONLINE CONTENTS QUANTITATIVE REASONING

In this study, preliminary knowledge-oriented reasoning of qualitative nature-focused on online contents, expression, objects, archived histories, and social actions. Authors demonstrated the concept of the sufficiency of knowledge. They used it to justify different conclusions previously based on the incomplete binding for knowledge mining [61].

C. MINING HIDDEN DEFINITION OF ASPECTS AND MOOD

Researchers proposed a new probabilistic supervised collaborative aspect model for the one-go-semantics understanding. The study reviewed the document process based on pairs of opinions. It simultaneously stimulated the terms of the aspect to corresponding reviews relating hidden aspect of human mood. Authors used general sentimental ratings relying on

TABLE 6. Studies relating to the Domain of Context, Semantics Analysis and Reasoning

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Chen, Tao, et al. (2016)	Product users reviews polarity.	Interim reviews have proven potentially useful to handle polarity.	A, B	60
2	Davis, Ernest, et al. (2017)	Online-contents quantitative reasoning.	Sufficiency of knowledge for justifying many conclusions.	B, C	61
3	Hai, Zhen, et al. (2017)	Mining hidden definition of aspects and mood.	General sentimental ratings relying upon surveys like surveillance data.	B, C	62
4	Li, Yuhua, et al. (2016)	Calculating semantic similarity of different sentences.	The structured lexical database used.	A,B	63
5	Schouten, Kim, et al. (2016)	Aggregate sentiments of certain social subjects.	Sentiments analysis at the aspect level.	B, C	64
6	Hua, Wen, et al. (2017)	Knowledge collection from Web Corpus (WC).	Knowledge-based WC approach proven better than traditional methods in some specific tasks.	A, B, C	65
7	Benitez-Andrades, Jose Alberto, et al (2020)	Social semantic web and data paradigm.	The integration policy of social and semantic Web technologies.	A, B	66
8	Saif, Hassan, et al. (2017)	Contextual and semantic information extraction.	Lexicon enriching with contextual and semantic information and improving the calculation of meanings.	B, C	67
9	Polo, Luis, et al. (2014)	Representing preferences as part of user profiles.	New ontological name base assessments and preferences proposed.	A, B, C	68
10	Dale, Robert. (2017)	Commercial use of NLP Survey.	State of the is the use of commercial NLP presented.	A, B	69
11	Kelsey, Darren, et al. (2014)	Twitter Joke Trial.	Target environmental factors and social influence on social text production.	B, C	70
12	De Smet, Wim, et al. (2013)	Analysis of the news contents.	Contents classified into terms or named entities based on probabilistic models of content comparison.	A, B	71

online surveys like surveillance data. This technique proved capable to infer semantic aspects, aspects of less or more meaningful mood levels and predicting the general temperament of the reviewers [62].

D. CALCULATING SEMANTIC SIMILARITY OF DIFFERENT SENTENCES

Authors worked in the two directions, one for the comparison and significance of short sentences and another on sentence semantic similarities. They considered word order to reflect the ordering of words in the proposed model. To analyze this ordering, the structured lexical database is used for calculating the semantic similarity of the two sentences. They also developed sentence related statistics [63].

E. AGGREGATE SENTIMENTS OF CERTAIN SOCIAL SUBJECTS

Authors worked on semantic analysis of various levels of social aspects. This review focused on the analysis of human thoughts regarding certain levels to find aggregate sentiments over concerned subjects. An in-depth review of the current state of the art progress is presented over user sentiments. The sentiment analysis at the aspect level provided accurate information about the sentiment in various applications areas [64].

F. KNOWLEDGE COLLECTION FROM WEB CORPUS (WC)

Authors created a prototype system for the understanding of online text. It used semantic knowledge provided by a well-known knowledge base automatically collected from a web corpus. This knowledge-based WC approach proved better than traditional methods in tasks like segmentation of text, marking part of speech and marking a concept. Authors performed a comprehensive performance evaluation based on real-time data [65].

G. SOCIAL SEMANTIC WEB AND DATA PARADIGM

In this paper, the authors presented an integration policy of social and semantic Web technologies with a related students learning data paradigm. This joint new methodology improved entities inter-activities. Educational environments are focused on simultaneously by putting students under the control of their learning processes encompassing various tools and services. Results of this study supported the significance of the study [66].

H. CONTEXTUAL AND SEMANTIC INFORMATION EXTRACTION

In this paper, researchers presented an approach for the adaptation of lexicons using contextual and semantic information extraction. They extracted textual data from DBPedia to assess the significance of the weighted feelings of words and to add new words to the lexicon dictionary. The research evaluated three different sets of Twitter data and showed that enriching the lexicons with contextual and semantic information improve the calculation of meanings [67].

I. REPRESENTING PREFERENCES AS PART OF USER PROFILES

This article presents a 'domain-independent application' for natural languages representing preferences as part of user

profiles. It also describes the translation issues of these languages using the new ontological name base assessments and preferences [68].

J. COMMERCIAL NLP SURVEY

Authors conducted a survey for NLP business-oriented uses and highlighted commercial uses of natural language processing which began more than 35 years ago. However, from the last few years, it became more noticeable because of the intense interest of researchers towards artificial intelligence in social media. The research presented the state of art commercial NLP today. Authors showed the main industry work and current progress in this area [69].

K. TWITTER JOKE TRIAL

The authors referred to this study as a theoretical discussion about 'Twitter jokes'. The authors focused on the growing need to understand the individual's communicative degree and power of interpretation in online data. However, this is a difficult task for online communication. The study targeted social factors and influence on production, interpretation, and consumption of social media data [70].

L. NEWS CONTENTS ANALYSIS

Authors studied several methods for presenting, merging, and comparing the presentations of news contents. A vector spatial model is used with both latent semantic analysis and probabilistic models. Contents classified into terms or named entities based on several models of content comparison. Authors proved that simple methods can surpass the current state of the art technology [71].

VIII. GAMES, KNOWLEDGE MINING AND AI

This section presents studies over extracting knowledge based on games with the help of AI, behaviour modelling of buyers and suppliers, polymorphic stationary states of the multi-user social games, the social dilemma of games, equilibrium game strategies, information feedback for optimization in games, stochastic models for the large population, mixed Nash equilibrium, weak or strong rarity value evolution of cooperation in rewarding compensation and punishment, HCI's three key points of views on social media, student performance assessment by solution point qualification rating, the survey on search optimization, logical programming paradigm for a set of answers, mitigation of adult material in children games, player's complaint classifications, game intelligence IEEE operations survey, artificial intelligence algorithms survey, fitness assessment methodology, etc., [72]–[89].

A. MARKET ANALYSIS

Researchers modelled the behaviour of buyers and suppliers as a two-step complete game process. They explored the issues related to existence, efficiency, and the computational complexity of the equilibrium in two-party games. To overcome situations where equilibrium does not exist or exists but is very inefficient, authors considered those scenarios where supplier lower prices algorithmically to satisfy the potential buyers [72].

B. MULTI-USER GAME DILEMMAS

Authors studied polymorphic stationary states of the players for analysis of various types of multi-user social game dilemmas. They presented the concept implementation theory of coalition games models. They treated it as the game based social dilemma study, through the evolutionary game theory. They proved their work can help to promote cooperation in participating entities. [73]

C. SOCIAL DILEMMA OF GAMES

Axiomatic type's derivation and categorization of the different type of social games dilemmas focused. Researchers classified social dilemma games having payouts of arbitrary nature. They also classified the types of game payouts. They based their work on stable equilibrium of actions and game rewards. They compared the corresponding game dynamics regarding the aspect of optimal rewards and payouts [74].

D. EQUILIBRIUM STRATEGIES

A new model for maximizing accumulative payments proposed for giving new lives during games. Equilibrium strategies evaluated so that the model may well approximate the equilibrium in related games having a large finite number of players. This model proposed the relation of players' sentiment and game difficulty level. This model evaluated over various types of online social games [75].

E. STATE STRUCTURE INFORMATION FEEDBACK

Authors studied a scalar linear-quadratic form of the differential in the games using the information feedback. An algorithm of numerical nature is presented, capable of determining whether a specific game can have an equilibrium. In the case of having a unique equilibrium, these sorts of algorithms provide chances for auxiliary equilibrium [76].

F. STOCHASTIC MODELS FOR LARGE POPULATION

Games and software having complex structures mostly have risks of fewer audiences. To overcome the complex nature, these games require stochastic models. Modelling the large populations with the help of game theory is intensifying in many crucial ways not only in games but also in real-time software solutions. Authors presented various dynamic games applications as a constructive resource and roadmap for evolutionary population-based solutions development [77].

G. MIXED NASH EQUILIBRIUM

Authors discussed the combination of the consequences of time delays in games with two mixed Nash equilibrium. It showed if there is stable internal equilibrium then an arbitrary minor delay makes for stochastic stability [78].

TABLE 7. Studies relating to the Domain of Games, Knowledge Mining and AI

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Caragiannis, Ioannis, et al. (2017)	Modelled the behaviour of buyers and suppliers.	Price rated fluctuations focused to see consumer behaviour.	A, B, C, D	72
2	Płatkowski, Tadeusz. (2016)	Polymorphic stationary states of the multi-user social game.	Worked on solving the theory of coalition games models.	B, C, D	73
3	Płatkowski, Tadeusz. (2017)	The social dilemma of games.	General nature having payouts of arbitrary nature were classified based on stable equilibrium.	B, C, D	74
4	Więcek, Piotr. (2017)	Equilibrium game strategies.	Approximated the equilibrium in games having a large finite number of players.	A, B, C, D	75
5	Engwerda, Jacob. (2017)	Information feedback for optimization in games.	An algorithm of numerical nature presented for equilibrium and auxiliary equilibrium.	B, C, D	76
6	Broom, Mark, et al. (2015)	Stochastic models for the large population.	Games applications proven constructive resource for evolutionary population-based solution approaches.	A, B, C, D	77
7	Carmona, , et al (2020)	Mixed Nash Equilibrium.	Stable internal equilibrium focused.	B, C, D	78
8	Joosten, Reinoud. (2016)	Weak or strong rarity value	Reward fixation issue focused.	A, B, D	79
9	Sasaki, Tatsuya. (2014)	Evolution of cooperation in rewarding compensation and punishment.	Evolutionary game dynamics focused.	A, B, C	80
10	Yannakakis, Georgios N., et al. (2015)	HCI has taken from three key points of view on social media.	AI methods, the ratio of each user region, Placing area with HCI perspective.	A, B	81
11	Kawatsu, Christopher, et al. (2017)	Student performance assessment by solution point qualification rating.	Concept of solution point qualification rating presented.	A, B, D	82
12	Togelius, Julian, et al. (2011)	Survey on search optimization.	Overview of published articles having game contents generated by searches over-optimization.	B, C	83
13	Stanescu, Marius, et al. (2016)	Logical programming paradigm for a set of answers.	Probabilistic based approaches and methods are taken in-game knowledge.	A, B, D	84
14	Cheong, Yun-Gyung, et al. (2015)	Mitigation of adult material in children games.	Mitigated predatory behaviour available in chats by the methods or algorithms of machine learning.	A, B, D	85
15	Balcı, Koray, et al. (2017)	Player's complaint classifications.	Complaints are taken from dual aspects algorithmically.	B, C, D	86
16	Lucas, Simon M. (2009)	Game intelligence IEEE operations survey.	Survey of good quality archives published material	A, B, C, D	87
17	Frutos-Pascual, Maite, et al. (2017)	Artificial intelligence algorithms survey.	Decision making AI training based games focused to analyze the trends of AI.	A, B, D	88
18	Luo, Linbo, et al. (2017)	Fitness assessment methodology.	Assessment of fitness of methodology integrating simulation processes of AI nature-based learning models.	B, C, D	89
19	Minaee, Shervin, et al. (2020)	Deep leaning and text classification.	AI based text classifications models and applications.	A, B, C, D	112

H. WEAK OR STRONG RARITY VALUE

Authors presented a study to distinguish between weak and the strong rarity values. For a strong/weak option, the total symmetric Pareto-effective rewards are higher/lower than those results obtained through keeping the reward fixed at the high level of the resource-stock. This sort of dynamic calculation improved the game satisfaction level [79].

I. EVOLUTION OF COOPERATION IN REWARDING COMPENSATION AND PUNISHMENT

Rewards are a practical way to socially promote association in social games. Researchers studied the cooperation evaluation in rewarding game compensations. They formalized and investigated game participants cooperation in public games. A complete classification is presented in the study of these games' evolutionary dynamics. In cases where the fines are found to be large, it caused instability of game cooperation. These fines eventually lead to better cooperation where compensation in a public game is kept optimal. [80]

J. CAUSING MEANINGFUL INTERACTION

Authors examined and analyzed Human-Computer Interaction (HCI) from three key points of views in social media: 1) the leading AI methods 2) the ratio of each user region 3) placing the area with HCI perspective. Besides, for each of these areas, researchers considered the informal possibility of interaction in each area. For meaningful interaction grouping, researchers classified the nature of various forms of players' interactions [81].

K. STUDENT PERFORMANCE ASSESSMENT BY SOLUTION POINT QUALIFICATION RATING

Authors simulated the performance of students, giving each student several qualification scores and exactly one learning goal. Unlike most common market software applications, students may play against each other directly. Every time the students decided a strategy for any game level, his behaviour is evaluated as per given metrics. Authors declared this action of playing like a game of 'solution point qualification rating.' [82].

L. SURVEY ON SEARCH OPTIMIZATION

This article contains an overview of all published articles having game contents developed for search optimization. Authors provided an overview of important issues of search optimization. This work provides a comprehensive briefing over search techniques used in various games. There exist a lot of games where certain hidden elements finding relates to awarding scores [83].

M. LOGICAL PROGRAMMING PARADIGM FOR SET OF ANSWERS

This article proposed predicting probable grouping of produced adversary units in each period. The logical programming-based paradigm for semantics suggested suitable in reasoning in case of uncertainty and improper knowledge tags. Probabilistic approaches and methods are considered in-game knowledge. Consistent combinations considered having the game observations [84].

N. MITIGATION ADULT MATERIAL IN CHILDREN GAMES

Games on social media are a real risk for children, as children exposed to adults' chats have possible sexual abuses. Data previously available does not highlight the age parity between adults and children and do not show the real chats in online multiplayer games. Researchers extracted real chat data from 'Movie Star Planet' multiplayer game for children. The researches mitigated predatory behaviour prevailing in chats by machine learning algorithms [85].

O. CLASSIFICATION OF PLAYER COMPLAINTS

A new structure of automatic classification presented for player complaints in the social gaming platform. Authors used functions that describe both sides of the complaints namely, the prosecutor and the suspect, as well as the interaction principles of the games [86].

P. GAME INTELLIGENCE IEEE OPERATIONS SURVEY

This work surveys computational intelligence research focused on the AI perspective, published over the IEEE platform. A survey of good quality archives published with material over various aspects. These aspects relate to games and their methodologies relating to real life. The results of this study is also suggested for the social games [87].

Q. ARTIFICIAL INTELLIGENCE ALGORITHMS SURVEY

Decision making AI training-based games are becoming popular. Researchers focused on and analyzed the trends of AI and gaming in this direction. They proposed salient features of these games useful for certain types of real-life management and social issues [88].

R. FITNESS ASSESSMENT METHODOLOGY

In this paper, the authors developed a process for the assessment of the fitness methodology, by integrating AI-based learning models simulation processes. Authors calculated fitness by a scenario-based function, which could automatically build based over proposed methodology. To improve its effectiveness, the timing impact of events added in the scenario [89].

S. CAPSULE NETWORKS

A capsule is a collection of nested neural layers. When designing a conventional neural network, new layers are added as needed, and in capsule networks, a new, additional layer is added inside another layer. In other words, inside one layer there is a nest from other layers. Signal processing in capsules is as follows: neurons inside the capsule capture the properties of one object within the image. The capsule outputs a vector to represent the existence of an object, with the orientation of the vector representing its properties. The vector is sent to all possible parents in the neural network, and then the prediction vector is calculated based on the multiplication of its own weight and the weight matrix. The bond of the capsule whose parent has the largest scalar prediction vector increases. The efficiency and productivity of text classification solutions achieved using artificial neural networks that trained by examples for solving an NLP problem, but they are not ideal either, the lack of spatial perception of objects makes neural networks useless in some cases. The solution to this problem is capsule networks, the layers of which consist of groups of neurons, allowing them to consider the location of objects relative to each other, as a human would do. [112].

T. HYBRID APPROACHES AND DEEP NEURAL NETWORKS

Neural networks are still struggling to cope with the hierarchies of concepts. Attempts to represent hierarchical structures in neural network models lead either to the retraining of neural networks or to the fact that they cannot fully differentiate recognition objects when going down the hierarchy. Hybrid models have a combination of CNN and LSTM facilitate in deep learning text analysis problems for certain reasons. For deep learning systems to work, a lot of high-quality data is required, which must be pre-cleaned and labelled by a specialist. The deeper the degree of the system, the more data is required. Deep learning systems work only with the types of data on which the training took place, and they still cannot generalize and transfer the found patterns to data of other types, even very close ones. It is very difficult for deep learning systems to work with hierarchical structures, so language processing is very difficult for them since natural language is a very deep hierarchical structure. Because of the previous deep learning systems, it is very difficult to perceive inaccurate and fuzzy data, they often do not see the difference where for a person the difference is obvious. Deep learning systems inherited from artificial neural networks exacerbated the problem of high complexity (to the point of practical impossibility) of explaining the results obtained and the inference produced. Deep learning systems do not consider the existing body of knowledge but

instead are retrained on the input data, simply interpreting them in their own way. Identifying causal relationships and separating them from simple correlations is an important task, but deep learning systems are still struggling to do this. Deep learning systems easily becomes misleading, especially if it is on the verge of over fitting. This vulnerability opens a wide scope for a variety of attacks, the consequences of which are hardly understood. They did not even begin to solve this problem. In such a situation hybrid AI approaches give optimal text processing results [112].

IX. ONTOLOGY-BASED DATA ACCESS

Ontology-based studies focused on this area in the field of text and data mining like ontological reasoning to find an appropriate application, ontologies and scientific web, ontologies for a structured vocabulary of document elements, ontological modelling of analytical applications, web ontology interfaces, etc., [90]–[94].

A. ONTOLOGY-BASED ENHANCED DATA ACCESS PERFORMANCE

Researchers used the ontology reasoning to find an appropriate application for the extended database enhanced by ontologies. They presented the outcome of their analysis in the form of logical statements that generate new intentional knowledge to facilitate auto-answering for user's queries. Various queries auto-responded by logical propositions compiled through the related ontology. These results are equivalent to the real-life answers to the queries [90].

B. ONTOLOGIES AND SCIENTIFIC WEB

This paper suggested a solution to several problems that arise when trying to accept ontology in an existing corporate environment. These problems contrasted with the excessively large advantages of semantic web literature. This research focused on several scientific parameters for applying ontologies to the semantic web community [91].

C. ONTOLOGIES FOR STRUCTURED VOCABULARY OF DOCUMENT ELEMENTS

Authors focused the ontologies providing a common structured vocabulary of document elements to describe parts of the document. Besides, for the formal description of the ontology, its utility demonstrated through several of its solutions and other works of the semantics that rely on annotated and extracted components [92].

D. ONTOLOGY MODELLING ANALYTICAL APPLICATIONS

Authors introduced an extension of the W3C consortium into the ontology modelling analytics. They used applications and applied data processing in a variable type manner. Authors introduced smoothness measures for solutions applying such ontology [93].

E. WEB ONTOLOGY INTERFACES

Authors discussed the web ontology interfaces as the minimal model based on several known ontologies. The data set inter-linked various related data principles. Authors described possible scenarios for using the data set to show its potential. The study helped in web mining from multiple heterogeneous sources [94].

X. BIG DATA

This section has the papers focused on Big Data analysis like identifying a particular user on several social networking sites, picture and tags text coherence, review of large data applications, the confidentiality of data and communication management, web data on forensic topics, the impact of tasks on tracking nodes, big data and the fourth paradigm, game data social modelling, scientific community data mining, diversity and equity in data analysis, traditional behavioural data and large data, empirical design for studying open information, coordination game derivation, price information mining, etc., [95]–[108]

A. CHAT FRIEND IDENTIFICATION

Authors used a method based on friendly relations to identify one user on several social networking sites. It is based on the concept that one cannot fake relationships with friends. The unique online entity, with multiple instance names, is identified by the proposed algorithm using 'AI' approaches. [95]

B. PICTURE AND TAGS TEXT COHERENCE

In this paper, the method of modelling the inherent visual concept of data structures based on hierarchical-multilayered label is focused on. They implemented the random forest model capable of correlating structured visual pieces of information. This model detected a more accurate semantic correlation between text tags and visual features. It provided a favourable visual semantics framework of interpretation of the text to image, useful even with few or incomplete tags. [96]

C. REVIEW OF LARGE DATA APPLICATIONS

This paper gives an understanding of the big data studies, including various applications. This work presented the possibilities and problems of handling the larger data. Moreover, this paper covers the methods and technologies that currently adopted to solve the problems of large data. Authors compared and recommended several basic methodologies used in the big data processing [97].

D. CONFIDENTIALITY OF DATA AND COMMUNICATION MANAGEMENT

Confidentiality of data over online media has become a vital research area. The increase in textual information creates many problems related to confidentiality. Confidentiality of data is susceptible to personal behaviour, due to the overall structure of the information on social media platforms.

TABLE 8. Studies relating to the Domain of Ontology-Based Data Access

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Cali, Andrea, et al. (2012)	Ontological reasoning to find an appropriate application.	Query base response to a logical proposition compiled by ontology.	A, B	90
2	Oberle, Daniel. (2014)	Ontologies and scientific web	Focused several scientific parameters for applying ontologies to the semantic web community.	C, D	91
3	Constantin, Alexandru, et al. (2016)	Ontologies for a structured vocabulary of document elements.	Semantics made to rely on annotated and extracted components.	B, D	92
4	Lebo, Timothy, et al. (2017)	Ontological modelling of analytical applications.	Smoothness measure introduced for environment applying such ontology.	A, D	93
5	Dojchinovski, Milan, et al. (2016)	Web ontology interfaces.	Described several possible scenarios for using the data set for WOI.	B, C	94

TABLE 9. Studies relating to the Domain of Big Data

Sr#	Authors	Focused IssueWorking Principle	Proposed/Findings/Processing	Form of Output	Ref#
1	Viswam, Anju, et al. (2017)	Identifying a particular user on several social networking sites.	The algorithm proposed based on AI approaches.	A, B	95
2	Wang, Jingya, et al. (2017)	Picture and tags text coherence.	More accurate semantic correlation finder between text tags and visual features.	B, C, D	96
3	Chen, CL Philip, et al. (2014)	Review of large data applications.	Discussed several basic methodologies in the processing of the data.	C, D	97
4	Al-Rabeeah, Nahi A.A, et al. (2017)	Confidentiality of data and communication management.	A model-based theory of 'communication management of confidentiality' and the 'theory of planned behaviour' proposed.	A, D	98
5	Pandi, et al (2020)	Web data relating forensic topics.	Application of statistical and computational methods on web data with new forensic methods of predictive text analysis.	B, D	99
6	Song, Guojie, et al. (2017)	Impact of task on tracking nodes.	The effect of maximization aimed at determining the maximum joint impact under one static network and focuses on tracking the number of influential nodes.	C, D	100
7	Hitzler, Pascal, et al. (2013)	Big data and the fourth paradigm.	Related data of WWW global presented for identifiers and references.	A, B	101
8	Khan, Abhimanyu, et al. (2016)	Game data social modelling.	Game coordination proposed for social solutions.	A, B, C	102
9	d'Alessandro, Brian, et al. (2017)	Scientific community data mining.	How discrimination is widely measured focused.	B, C	103
10	Drosou, Marina, et al. (2017)	Diversity and equity in data analysis.	Overview of recent technical work.	C, D	104
11	Shmueli, Galit. (2017)	Traditional behavioural data and large data.	The relationship among data, subjects and research questions differs in the context of traditional behavioural in large data.	A, B, C, D	105
12	Chen, Yanping, et al. (2017)	Empirical design for studying open information.	Multiphase empirical design for studying open information presented.	A, D	106
13	Iwata, Tomoharu, et al. (2017)	Coordination game derivation.	Proposed an effective Bayesian procedure.	C, D	107
14	Sun, Yefei, et al (2020)	Price information mining.	Proportions represented for the tendency of purchase under influence.	B, D	108

This study presented a model based on the theory of 'communication management of confidentiality' and the 'theory of planned behaviour' with the help of new confidentiality factors that influence the online user's attitudes. [98]

E. WEB DATA ON NEW FORENSIC TOPICS

This study analyzed several discussions on the research and application of statistical and computational methods on web data with new forensic methods of predictive text analysis. It provides the information useful for creating a statistical index for describing quality and efficiency from a law and judiciary point of view. [99]

F. IMPACT OF TASK ON TRACKING NODES

Authors worked on the influential routing of nodes and calculated impact of the task on tracking nodes as an extension of the traditional problem of maximizing the impact in dynamic social networks. The effect of maximization aimed at determining the maximum joint impact under one static network and focused on tracking the number of influential nodes, which retains the maximum influence as the network throughput. Using the smoothness of the evolution of the network structure, an effective algorithm is proposed [100].

G. BIG DATA ANALYSIS PARADIGM

Authors related data of the web identifiers and references to various internet resources. They applied them to raw data

for classification purposes. They developed an online web information filtration mechanism. Authors presented this as the paradigm of online data identification and classification. The web identifiers are grouped and elaborated for multiple forms of web mining [101].

H. GAME DATA SOCIAL MODELLING

This research highlighted the issue of the better understating of agent's coordination in large social games. They discussed social process modelling, which proved that the coordination gadgets of games result in better player interactions. The study focused on those features which relate to observational network options to facilitate convergence to effective real-time coordination. [102]

I. SCIENTIFIC COMMUNITY DATA MINING

Authors dealt with discrimination issues for the scientific community data. They studied the usual processes of data mining by providing a taxonomy of common practices which can lead to unintentional discrimination. Authors studied how we could widely measure discrimination and suggested how to develop processes to mitigate the discriminatory capacity of information processing systems. [103]

J. DIVERSITY AND EQUITY IN DATA ANALYSIS

Researchers provided an overview of recent technical work on diversity, especially in the field of selection of tasks and discussed the relationship between diversity and equity to identify promising future directions. They pointed out important issues regarding the professional responsibilities for data mining. [104]

K. TRADITIONAL BEHAVIOURAL DATA AND LARGE DATA

Behavioural Data and Large Data refers to huge and rich multidimensional sets of social media text. It contains social actions, online interactions between companies, governments, and researchers. Many researchers of the social sciences and management are analyzing social media text to make valuable knowledge discoveries. This study showed the relationships and differences among data, subjects, and research questions in the context of traditional behavioural big data [105].

L. EMPIRICAL DESIGN FOR STUDYING OPEN INFORMATION

It is a difficult task to obtain information from a large amount of data in the public domain. This work proposed to create a network structure to measure and solve this problem. This study presented an empirical design for studying open information, consisting of three phases: document events detection, constructing a network of events and the analysis of the network of events. For the implementation and comprehension, examples presented for studying open information through a network of events [106].

M. BAYESIAN PROCEDURE

Researchers proposed an effective Bayesian procedure for analyzing the coordination in-game derivation, over the larger social groups. They evaluated their model over the Gibbs samples. By performing various experiments, the authors showed the efficiency of the proposed technique using artificial and real sets of large data [107].

N. PRICE INFORMATION MINING

In this paper a new thematic model is proposed for analyzing procurement data with price information. Price is evaluated as an important factor in consumer behaviour. The proposed model assumes that each commodity has its price distribution for each categorization. Not only the prices, but the commodities also have a statistical distribution of their marketing strategies. A rationale proportion by NLP analysis is represented for the tendency of consumer purchase under the influence of the user intention concerning the commodity prices & marketing. [108]

XI. CONCLUSION

This paper gives an overview of research in various areas of social media analysis. The paper focused nine major dimensions i.e. 'Decision Making', 'Text Retrieval, 'Query Formulation Optimization', 'Content and Material Data Analysis', 'Text and Linguistic Patterns and Entities Recognition', 'Mining Knowledge, Opinion, Behaviour', 'Context Semantics Analysis and Reasoning', 'Games, Knowledge Mining and AI', 'Ontology-Based Data Access' and 'Big Data'. The selected papers in each category are presented in the form of summary tables to show the basic inputs and outputs of each study. The outputs have four aspects i.e. Context Understanding, Pattern of Information Understanding, Knowledge Discovery, and Semantic Analysis. The approaches used in the selected papers are lexicon approaches, support vector machine, intelligent techniques, evolutionary computation, random forest model, Fuzzy-Rule, mining by the association-rule, analysis based on the formal concept, algorithms of game theory, etc., The main contribution of this research is the breakdown structure of papers for readers, so that we can help provide insight into certain researches carried out recently. This survey presents various precise ideas to help in looking forward to various solutions. In this paper, we focused on important applications like marketing, politics, uncertainty determination predictions etc., We hope that this survey can help new researchers in this field to get a head start in contributing to new techniques for social media mining.

XII. ACKNOWLEDGEMENT

We are thankful to Government College University Faisalabad (GCUF), Pakistan to provide resources for this research. Also, we are thankful to the Ministry of Higher Education Malaysia and Universiti Kebangsaan Malaysia (grant ID: GGPM-2020-029 and grant ID: PP-FTSM-2020) for supporting and funding this work.

REFERENCES

[1] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis." Knowledge-Based Systems 89 (2015): 14-46.

[2] Li, Jie-Jun, Han Yang, and Hao Tang. "Feature Mining and Sentiment Orientation Analysis on Product Review." Management Information and Optoelectronic Engineering: Proceedings of the 2016 International Conference on Management, Information and Communication (ICMIC2016) and the 2016 International Conference on Optics and Electronics Engineering (ICOEE2016). 2017.

[3] Pallavicini, F., P. Cipresso, and F. Mantovani. "Beyond Sentiment: How Social Network Analytics Can Enhance Opinion Mining and Sentiment Analysis." Sentiment Analysis in Social Networks. 2017. 13-29.

[4] Yadollahi, Ali, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. "Current state of text sentiment analysis from opinion to emotion mining." ACM Computing Surveys (CSUR) 50.2 (2017): 25.

[5] Gkatzia, Dimitra et al. "Data-to-Text Generation Improves Decision-Making Under Uncertainty." IEEE Computational Intelligence Magazine 12.3 (2017): 10-17.

[6] Lazarus, Suleman et al. "The bifurcation of the Nigerian cybercriminals: Narratives of the Economic and Financial Crimes Commission (EFCC) agents." Telematics and Informatics 40 (2019): 14-26.

[7] Blomqvist, Eva. "The use of Semantic Web technologies for decision support–a survey." Semantic Web 5.3 (2014): 177-201.

[8] EV, Vinu et al. "Automated generation of assessment tests from domain ontologies." Semantic Web 8.6 (2017): 1023-1047.

[9] Chouldechova, Alexandra. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments." arXiv preprint arXiv:1703.00056 (2017).

[10] Wang, Tong, et al. "Finding patterns with a rotten core: Data mining for crime series with cores." Big data 3.1 (2015): 3-21.

[11] Ausiello, Dennis et al. "Real-time assessment of wellness and disease in daily life." Big data 3.3 (2015): 203-208.

[12] Azmak, Okan, et al. "Using Big data to understand the human condition: the kavli HUMAN project." Big data 3.3 (2015): 173-188.

[13] Fawcett, Tom. "Mining the quantified self: personal knowledge discovery as a challenge for data science." Big Data 3.4 (2015): 249-266.

[14] Kunneman, Florian et al. "Open-domain extraction of future events from Twitter." Natural Language Engineering 22.5 (2016): 655-686.

[15] Žliobaitė, Indrė. "Measuring discrimination in algorithmic decision making." Data Mining and Knowledge Discovery(2017): 1-30.

[16] Oswald, C., Anirban I. Ghosh et al. "Knowledge engineering perspective of text compression." India Conference (INDICON), 2015 Annual IEEE. IEEE, 2015

[17] Haiduc, Sonia. "Supporting Query Formulation for Text Retrieval Applications in Software Engineering." Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on. IEEE, 2014

[18] Bhardwaj et al. "Generative Model for NLP Applications based on Component Extraction." Procedia Computer Science 167 (2020): 918-931.

[19] Boixader, D., J. Recasens. "Reduction of Attributes in Averaged Similarities." Information Sciences (2017).

[20] Hatamlou, Abdolreza. "Black hole: A new heuristic optimization approach for data clustering." Information sciences 222 (2013): 175-184.

[21] Shim, Changbeom, et al. "Nearest close friend search in geo-social networks." Information Sciences 423 (2018): 235-256.

[22] Dornescu, Iustin et al. "Densification: Semantic document analysis using Wikipedia." Natural Language Engineering 20.4 (2014): 469-500.

[23] Ferrucci, David et al. "UIMA: an architectural approach to unstructured information processing in the corporate research environment." Natural Language Engineering 10.3-4 (2014): 327-348.

[24] Mummery, Jane et al. "The role of blogging in public deliberation and democracy." Discourse, Context & Media2.1 (2013): 22-39.

[25] Androutsopoulos, Jannis. "Languaging when contexts collapse: Audience design in social networking." Discourse, Context & Media 4 (2014): 62-73.

[26] McKeown et al. "Exploring the metadiscursive realisation of incivility in TV news discourse." Discourse, Context & Media 33 (2020): 100367.

[27] Poria, Soujanya, et al. "Sentiment data flow analysis by means of dynamic linguistic patterns." IEEE Computational Intelligence Magazine 10.4 (2015): 26-36.

[28] Oneto, Luca, et al. "Statistical learning theory and ELM for big social data analysis." ieee CompUTATionAl inTelliGenCe mAGAzine 11.3 (2016): 45-55

[29] He, Haibo et al. "Learning from imbalanced data." IEEE Transactions on knowledge and data engineering21.9 (2009): 1263-1284.

[30] Neff, Gina, et al. "Critique and contribute: A practice-based framework for improving critical data studies and Data Science." Big Data 5.2 (2017): 85-97.

[31] Dale, Robert. "The pros and cons of listening devices." Natural Language Engineering 23.6 (2017): 969-973.

[32] Luzón, María José. "“This is an erroneous argument”: Conflict in academic blog discussions." Discourse, Context & Media 2.2 (2013): 111-119.

[33] Giles, David, et al. "Microanalysis of online data: The methodological development of “digital CA”." Discourse, Context & Media 7 (2015): 45-51.

[34] Tolson, Andrew. "“You'll need a miracle to win this election”(J. Paxman 2005): Interviewer assertiveness in UK general elections 1983–2010." Discourse, Context & Media 1.1 (2012): 45-53.

[35] Bolander, Brook et al. "Doing sociolinguistic research on computer-mediated data: A review of four methodological issues." Discourse, Context & Media 3 (2014): 14-26.

[36] Schafer, Svenja. "Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions." Computers in Human Behavior 103 (2020): 1-12.

[37] Nothman, Joel, et al. "Learning multilingual named entity recognition from Wikipedia." Artificial Intelligence 194 (2013): 151-175. B1

[38] Amores, Jaume. "Multiple instance classification: Review, taxonomy and comparative study." Artificial Intelligence 201 (2013): 81-105.

[39] Marques-Silva, Joao et al. "Minimal sets on propositional formulae. Problems and reductions." Artificial Intelligence 252 (2017): 22-50.

[40] Hamada, Naoto, et al. "Strategy-proof school choice mechanisms with minimum quotas and initial endowments." Artificial Intelligence 249 (2017): 47-71.

[41] Gebser, Martin et al. "Conflict-driven answer set solving: From theory to practice." Artificial Intelligence 187 (2012): 52-89.

[42] Zhu, Xiaofeng, et al. "Local and Global Structure Preservation for Robust Unsupervised Spectral Feature Selection." IEEE Transactions on Knowledge and Data Engineering (2017).

[43] Park, Sangwon, et al. "Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data." Annals of Tourism Research 84 (2020): 102973.

[44] Blanco, Eduardo et al. "Retrieving implicit positive meaning from negated statements." Natural Language Engineering 20.4 (2014): 501-535.

[45] Rapp, Reinhard, et al. "Recent advances in machine translation using comparable corpora." Natural Language Engineering 22.4 (2016): 501-516.

[46] West, Laura E. "Facebook sharing: A sociolinguistic analysis of computer-mediated storytelling." Discourse, Context & Media 2.1 (2013): 1-13.

[47] Liu, Fangfang, et al. "Three-valued semantics for hybrid MKNF knowledge bases revisited." Artificial Intelligence 252 (2017): 123-138.

[48] Jankowski-Lorek, Michał, et al. "Verifying social network models of Wikipedia knowledge community." Information Sciences 339 (2016): 158-174.

[49] Barrera-Animas, Ari Yair, et al. "Online personal risk detection based on behavioural and physiological patterns." Information Sciences 384 (2017): 281-297.

[50] Zhao, Yiyi, et al. "Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective." Information Sciences (2017).

[51] De León, Hernán Ponce, et al. "Incorporating negative information to process discovery of complex systems." Information Sciences422 (2018): 480-496.

[52] Liu, Yezheng, et al. "Identifying impact of intrinsic factors on topic preferences in online social media: A nonparametric hierarchical Bayesian approach." Information Sciences 423 (2018): 219-234.

[53] Perez-Chacón, R., et al. "Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand." Information Sciences 540 (2020): 160-174.

[54] Akbari, Mohammad, et al. "Wellness Representation of Users in Social Media: Towards Joint Modelling of Heterogeneity and Temporality." IEEE Transactions on Knowledge and Data Engineering 29.10 (2017): 2360-2373.

[55] Buskens, Vincent, et al. "Effects of network characteristics on reaching the payoff-dominant equilibrium in coordination games: a simulation study." Dynamic games and applications 6.4 (2016): 477-494.

[56] Lippi, Marco. "Statistical relational learning for game theory." IEEE Transactions on Computational Intelligence and AI in Games 8.4 (2016): 412-425.

[57] Krestel, Ralf, et al. "Modeling human newspaper readers: The Fuzzy Believer approach." Natural Language Engineering 20.2 (2014): 261-288.
[58] Church, Kenneth. "Emerging trends: Inflation." Natural Language Engineering 23.5 (2017): 807-812.
[59] Lee, Kichun, et al. "Dependence maps, a dimensionality reduction with dependence distance for high-dimensional data." Data Mining and Knowledge Discovery 26.3 (2013): 512-532.
[60] Chen, Tao, et al. "Learning user and product distributed representations using a sequence model for sentiment analysis." IEEE Computational Intelligence Magazine 11.3 (2016): 34-44.
[61] Davis, Ernest, et al. "Commonsense reasoning about containers using radically incomplete information." Artificial Intelligence 248 (2017): 46-84.
[62] Hai, Zhen, et al. "Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach." IEEE Transactions on Knowledge and Data Engineering 29.6 (2017): 1172-1185.
[63] Li, Yuhua, et al. "Sentence similarity based on semantic nets and corpus statistics." IEEE transactions on knowledge and data engineering 18.8 (2016): 1138-1150.
[64] Schouten, Kim, et al. "Survey on aspect-level sentiment analysis." IEEE Transactions on Knowledge and Data Engineering 28.3 (2016): 813-830.
[65] Hua, Wen, et al. "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge." IEEE transactions on Knowledge and data Engineering 29.3 (2017): 499-512.
[66] Benitez-Andrades, Jose Alberto, et al . "Social network analysis for personalized characterization and risk assessment of alcohol use disorders in adolescents using semantic technologies." Future Generation Computer Systems 106 (2020): 154-170.
[67] Saif, Hassan, et al. "Sentiment lexicon adaptation with context and semantics for the social web." Semantic Web 8.5 (2017): 643-665.
[68] Polo, Luis, et al. "User preferences in the web of data." Semantic Web 5,1 (2014): 67-75.
[69] Dale, Robert. "The commercial NLP landscape in 2017." Natural Language Engineering 23.4 (2017): 641-647.
[70] Kelsey, Darren, et al. "Discipline and resistance on social media: Discourse, power and context in the Paul Chambers 'Twitter Joke Trial'." Discourse, Context & Media 3 (2014): 37-45.
[71] De Smet, Wim, et al. "Representations for multi-document event clustering." Data Mining and Knowledge Discovery (2013): 1-26.
[72] Caragiannis, Ioannis, et al. "Efficiency and complexity of price competition among single-product vendors." Artificial Intelligence 248 (2017): 9-25.
[73] Płatkowski, Tadeusz. "Evolutionary coalitional games." Dynamic Games and Applications 6.3 (2016): 396-408.
[74] Płatkowski, Tadeusz. "On Derivation and Evolutionary Classification of Social Dilemma Games." Dynamic Games and Applications 7.1 (2017): 67-75.
[75] Więcek, Piotr. "Total reward semi-Markov mean-field games with complementarity properties." Dynamic Games and Applications7.3 (2017): 507-529.
[76] Engwerda, Jacob. "A Numerical Algorithm to Calculate the Unique Feedback Nash Equilibrium in a Large Scalar LQ Differential Game." Dynamic Games and Applications 7.4 (2017): 635-656.
[77] Broom, Mark, et al. "Dynamic Games and Applications: Second Special Issue on Population Games: Introduction." Dynamic Games and Applications 5.2 (2015): 155-156.
[78] Carmona, , et al (2020). "Pure strategy Nash equilibria of large finite-player games and their relationship to non-atomic games." Journal of Economic Theory 187 (2020): 105015.
[79] Joosten, Reinoud. "Strong and Weak Rarity Value: Resource Games with Complex Price–Scarcity Relationships." Dynamic games and applications 6.1 (2016): 97-111.
[80] Sasaki, Tatsuya. "The evolution of cooperation through institutional incentives and optional participation." Dynamic games and applications 4.3 (2014): 345-362.
[81] Yannakakis, Georgios N., et al. "A panorama of artificial and computational intelligence in games." IEEE Transactions on Computational Intelligence and AI in Games 7.4 (2015): 317-335.
[82] Kawatsu, Christopher, et al. "Predicting Students' Decisions in a Training Simulation: A Novel Application of TrueSkill™." IEEE Transactions on Computational Intelligence and AI in Games (2017).
[83] Togelius, Julian, et al. "Search-based procedural content generation: A taxonomy and survey." IEEE Transactions on Computational Intelligence and AI in Games 3.3 (2011): 172-186.
[84] Stanescu, Marius, et al. "Predicting opponent's production in real-time strategy games with answer set programming." IEEE Transactions on Computational Intelligence and AI in Games 8.1 (2016): 89-94.
[85] Cheong, Yun-Gyung, et al. "Detecting Predatory Behavior in Game Chats." IEEE Transactions on Computational Intelligence and AI in Games 7.3 (2015): 220-232.
[86] Balcı, Koray, et al. "Automatic Classification of Player Complaints in Social Games." IEEE Transactions on Computational Intelligence and AI in Games 9.1 (2017): 103-108.
[87] Lucas, Simon M. "Computational intelligence and AI in games: a new IEEE transactions." IEEE Transactions on Computational Intelligence and AI in Games 1.1 (2009): 1-3.
[88] Frutos-Pascual, Maite, et al. "Review of the use of AI techniques in serious games: Decision making and machine learning." IEEE Transactions on Computational Intelligence and AI in Games 9.2 (2017): 133-152.
[89] Luo, Linbo, et al. "Design and evaluation of a data-driven scenario generation framework for game-based training," IEEE Transactions on Computational Intelligence and AI in Games(2017).
[90] Cali, Andrea, et al. "Towards more expressive ontology languages: The query answering problem." Artificial Intelligence 193 (2012): 87-128.
[91] Oberle, Daniel. "How ontologies benefit enterprise applications." Semantic Web 5.6 (2014): 473-491.
[92] Constantin, Alexandru, et al. "The document components ontology (DoCO)." Semantic Web 7.2 (2016): 167-181.
[93] Lebo, Timothy, et al. "A five-star rating scheme to assess application seamlessness." Semantic Web 8.1 (2017): 43-63.
[94] Dojchinovski, Milan, et al. "Linked Web APIs Dataset." Semantic Web Preprint: (2016) 1-11.
[95] Viswam, Anju, et al. "An efficient bitcoin fraud detection in social media networks." Circuit, Power and Computing Technologies (ICCPCT), 2017 International Conference on. IEEE, 2017.
[96] Wang, Jingya, et al. "Discovering visual concept structure with sparse and incomplete tags." Artificial Intelligence (2017).
[97] Chen, CL Philip, et al. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." Information Sciences 275 (2014): 314-347.
[98] Al-Rabeeah, Nahi A.A, et al. "Data privacy model for social media platforms." Student Project Conference (ICT-ISPC), 2017 6th ICT International. IEEE, 2017.
[99] Pandi, et al. "Exploration of Vulnerabilities, Threats and Forensic Issues and its impact on the Distributed Environment of Cloud and its mitigation." Procedia Computer Science 167 (2020): 163-173.
[100] Song, Guojie, et al. "Influential node tracking on dynamic social network: an interchange greedy approach." IEEE Transactions on Knowledge and Data Engineering 29.2 (2017): 359-372.
[101] Hitzler, Pascal, et al. "Linked Data, Big Data, and the 4th Paradigm." Semantic Web 4.3 (2013): 233-235.
[102] Khan, Abhimanyu, et al. "Network characteristics enabling efficient coordination: A simulation study." Dynamic Games and Applications 6.4 (2016): 495-519.
[103] d'Alessandro, Brian, et al. "Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification." Big Data 5.2 (2017): 120-134.
[104] Drosou, Marina, et al. "Diversity in Big Data: A Review." Big Data5.2 (2017): 73-84. hundred
[105] Shmueli, Galit. "Research dilemmas with behavioral big data." Big Data 5.2 (2017): 98-119.
[106] Chen, Yanping, et al. "Exploring open information via event network." Natural Language Engineering (2017): 1-22.
[107] Iwata, Tomoharu, et al. "Robust unsupervised cluster matching for network data." Data Mining and Knowledge Discovery (2017): 1-23.
[108] Sun, Yefei, et al. "Residents' sentiments towards electricity price policy: Evidence from text mining in social media." Resources, Conservation and Recycling 160 (2020): 104903.
[109] Ali, Farman, et al. "An intelligent healthcare monitoring framework using wearable sensors and social networking data." Future Generation Computer Systems 114 (2020): 23-43.
[110] Ali, Farman, et al. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." Information Fusion 63 (2020): 208-222.
[111] Raza, Mohsin, et al. "Establishing effective communications in disaster affected areas and artificial intelligence based detection using social media platform." Future Generation Computer Systems 112 (2020): 1057-1069.

[112] Minaee, Shervin, et al. "Deep learning based text classification: A comprehensive review." arXiv preprint arXiv:2004.03705 (2020).

MUHAMMAD YAHYA SAEED is working as Lecturer in Department of Software Engineering, Government College University Faisalabad (GCUF). He has done MSc(Computer Science) MSc(Statistics), MS(CS). He is doing PhD (CS)
E-mail: m_yahya_saeed@gcuf.edu.pk
Website: https://profiles.gcuf.edu.pk/profile/myahyasaeed

DR. MUHAMMAD AWAIS received the Master degree in Computer Science from the University of Agriculture, Faisalabad, Pakistan, in 2001, MPhil degree in Computer Science in 2004 from the University of Agriculture, MPhil degree in Applied Information Science in 2008 from the Albert-Ludwigs University, Freiburg, Germany and Ph.D. degree in Applied Information Sciences from University of Bayreuth, Germany in 2013. In 2005, he joined the Government College University Faisalabad as a Lecturer and became an Assistant Professor in 2016. His current research interests include HRI, Intention Estimation, Machine and Deep Learning, Text Engineering, Computer Vision and Software Engineering.
E-mail: muhammadawais@gcuf.edu.pk
Website: https://profiles.gcuf.edu.pk/profile/drmuhammadawais

DR. MUHAMMAD YOUNAS has completed his PhD. degree from School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM). He is working as a Assistant Professor, in Computer Science Department, Government College University Faisalabad, Pakistan. His research interests are in software engineering, agile software development, cloud computing and code clone detection
E-mail: younas.76@gmail.com
Website: https://profiles.gcuf.edu.pk/profile/muhammadyounaslatif

DR. MUHAMMAD ARIF SHAH graduated from the Department of Software Engineering, Faculty of Engineering, School of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia. He is currently an Assistant Professor of Software Engineering with the Pak-Austria Fachhochschule Institute of Applied Sciences and Technology, Haripur, Pakistan. He is also a member of the Software Engineering Research Group (SERG).
E-mail: arif.websol@gmail.com

MAHDI ZAREEI (M'17 – SM'20) received the M.Sc. degree in computer network from the University of Science, Malaysia, in 2011, and the Ph.D. degree from the Communication Systems and Networks Research Group, Malaysia-Japan International Institute of Technology, University of Technology, Malaysia, in 2016. In 2017, he joined the School of Engineering and Sciences, Tecnologico de Monterrey, as a Postdoctoral Fellow, where from 2019 he started working as a Research Professor. His research mainly focuses on wireless sensor and ad hoc networks, energy harvesting sensors, information security and machine learning. He is a member of the Mexican National Researchers System (level I). He is also serving as an Associate Editor for the IEEE ACCESS and Ad Hoc & Sensor Wireless Networks Journals
E-mail: m.zareei@ieee.org

ATIF KHAN received his M.Sc. degree in Computer Science from University of Peshawar, Pakistan, in 2004, and Ph.D. degree in Computer Science (Text Mining) from Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2016. From 2016 onwards, he is working as Assistant Professor at Islamia College Peshawar, KP, Pakistan. He is a technical committee member in many international conferences and a reviewer in many international conferences, journals. His current areas of research interest include data mining, text mining, sentiment analysis and opinion mining, recommender systems, and machine learning. He is the recipient of Best Student Award and Pro- Chancellor Award at UTM during his Ph.D., for his excellent contribution in the field of text mining. He is also serving as an Associate Editor for ACM Transactions on Asian and Low-Resource Language Information Processing.
E-mail: atikhan@icp.edu.pk

SHIDROKH GOUDARZI received her Ph.D. degree in communication system and wireless network from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM). In 2014, She received three year full scholarship to study Ph.D. at (UTM).Then, She joined the Department of Advanced Informatics School at Universiti Teknologi Malaysia as a Postdoctoral Fellow from 2018 to 2019. Currently, she is a senior lecturer at Universiti Kebangsaan Malaysia (UKM). She also serves as reviewer for IEEE Transactions on Industrial Informatics, IEEE Systems Journal, Canadian Journal of Electrical and Computer Engineering, KSII Transactions on Internet and Information Systems Journal,Journal of Engineering and Technological Sciences, Mathematical Problems in Engineering and IEEE Access. Her research interests are in wireless networks, artificial intelligence, machine learning, next generation networks, Internet of Things (Iot) and Mobile/distributed/Cloud Computing.
E-mail: shidrokh@ukm.edu.my

[112] Minaee, Shervin, et al. "Deep learning based text classification: A comprehensive review." arXiv preprint arXiv:2004.03705 (2020).

MUHAMMAD YAHYA SAEED is working as Lecturer in Department of Software Engineering, Government College University Faisalabad (GCUF). He has done MSc(Computer Science) MSc(Statistics), MS(CS). He is doing PhD (CS)
E-mail: m_yahya_saeed@gcuf.edu.pk
Website: https://profiles.gcuf.edu.pk/profile/myahyasaeed

DR. MUHAMMAD AWAIS received the Master degree in Computer Science from the University of Agriculture, Faisalabad, Pakistan, in 2001, MPhil degree in Computer Science in 2004 from the University of Agriculture, MPhil degree in Applied Information Science in 2008 from the Albert-Ludwigs University, Freiburg, Germany and Ph.D. degree in Applied Information Sciences from University of Bayreuth, Germany in 2013. In 2005, he joined the Government College University Faisalabad as a Lecturer and became an Assistant Professor in 2016. His current research interests include HRI, Intention Estimation, Machine and Deep Learning, Text Engineering, Computer Vision and Software Engineering.
E-mail: muhammadawais@gcuf.edu.pk
Website: https://profiles.gcuf.edu.pk/profile/drmuhammadawais

DR. MUHAMMAD YOUNAS has completed his PhD. degree from School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM). He is working as a Assistant Professor, in Computer Science Department, Government College University Faisalabad, Pakistan. His research interests are in software engineering, agile software development, cloud computing and code clone detection
E-mail: younas.76@gmail.com
Website: https://profiles.gcuf.edu.pk/profile/muhammadyounaslatif

DR. MUHAMMAD ARIF SHAH graduated from the Department of Software Engineering, Faculty of Engineering, School of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia. He is currently an Assistant Professor of Software Engineering with the Pak-Austria Fachhochschule Institute of Applied Sciences and Technology, Haripur, Pakistan. He is also a member of the Software Engineering Research Group (SERG).
E-mail: arif.websol@gmail.com

MAHDI ZAREEI (M'17 – SM'20) received the M.Sc. degree in computer network from the University of Science, Malaysia, in 2011, and the Ph.D. degree from the Communication Systems and Networks Research Group, Malaysia-Japan International Institute of Technology, University of Technology, Malaysia, in 2016. In 2017, he joined the School of Engineering and Sciences, Tecnologico de Monterrey, as a Postdoctoral Fellow, where from 2019 he started working as a Research Professor. His research mainly focuses on wireless sensor and ad hoc networks, energy harvesting sensors, information security and machine learning. He is a member of the Mexican National Researchers System (level I). He is also serving as an Associate Editor for the IEEE ACCESS and Ad Hoc & Sensor Wireless Networks Journals
E-mail: m.zareei@ieee.org

ATIF KHAN received his M.Sc. degree in Computer Science from University of Peshawar, Pakistan, in 2004, and Ph.D. degree in Computer Science (Text Mining) from Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2016. From 2016 onwards, he is working as Assistant Professor at Islamia College Peshawar, KP, Pakistan. He is a technical committee member in many international conferences and a reviewer in many international conferences, journals. His current areas of research interest include data mining, text mining, sentiment analysis and opinion mining, recommender systems, and machine learning. He is the recipient of Best Student Award and Pro- Chancellor Award at UTM during his Ph.D., for his excellent contribution in the field of text mining. He is also serving as an Associate Editor for ACM Transactions on Asian and Low-Resource Language Information Processing.
E-mail: atikhan@icp.edu.pk

SHIDROKH GOUDARZI received her Ph.D. degree in communication system and wireless network from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM). In 2014, She received three year full scholarship to study Ph.D. at (UTM).Then, She joined the Department of Advanced Informatics School at Universiti Teknologi Malaysia as a Postdoctoral Fellow from 2018 to 2019. Currently, she is a senior lecturer at Universiti Kebangsaan Malaysia (UKM). She also serves as reviewer for IEEE Transactions on Industrial Informatics, IEEE Systems Journal, Canadian Journal of Electrical and Computer Engineering, KSII Transactions on Internet and Information Systems Journal,Journal of Engineering and Technological Sciences, Mathematical Problems in Engineering and IEEE Access. Her research interests are in wireless networks, artificial intelligence, machine learning, next generation networks, Internet of Things (Iot) and Mobile/distributed/Cloud Computing.
E-mail: shidrokh@ukm.edu.my

Original Manuscript ID: Access-2020-44893

Original Article Title: "Overview of Text Analysis Tasks, Applications and Approaches"

To: IEEE Access Editor

Re: Response to reviewers

Dear Editor,

Thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers' comments.

We are uploading (a) our point-by-point response to the comments (below) (response to reviewers), (b) an updated manuscript with yellow highlighting indicating changes, and (c) a clean updated manuscript without highlights (PDF main document).

Best regards,

Muhammad Yahya Saeed et al.

Reviewer#1, Concern # 1: several bad English constructions, grammar mistakes, and misuse of articles

Author response:

Author action: We updated the manuscript by fixing the grammatical errors

---

Reviewer#1, Concern # 2: is not well organized and does not link well with important literature on NLP research, e.g., see Minaee et al.'s recent review on deep learning based text classification. Also, check latest trends in using capsule networks for challenging NLP applications. Finally, check recent hybrid AI approaches that inject semantic information into deep neural networks, e.g., Sentic LSTM.

Author response:

Author action: We updated the manuscript by adding the said contents.

Reviewer#2, Concern # 1: The abstract is not attractive, some sentences should be changed.

Author response:

Author action: We updated the manuscript by updating the abstract

---

Reviewer#2, Concern # 2: In introduction, the main contribution is not clear. The first, second, and third paragraph should be about social media mining, techniques for texting mining, and application of social media mining, respectively.

Author response:

Author action: We updated the manuscript in introduction section as per direction given by reviser.

Reviewer#2, Concern # 3: ML

'An intelligent healthcare monitoring framework using wearable sensors and social networking data',

'Establishing effective communications in disaster affected areas and artificial intelligence based detection using social media platform',

'A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion'.

Author response:

Author action: We updated the manuscript by adding the contents of said paper and highlighted as well.

Reviewer#2, Concern # 4: The Authors should thoroughly check the manuscript and remove typos and grammatical errors.

Author response:

Author action: We updated the manuscript by fixing the typos.

Reviewer#2, Concern # 5: The Spaces are required in text (See column Form of Output in tables)

Author response:

Author action: We updated the manuscript by Spaces added in the Column of Form of Output

---

Note: *References suggested by reviewers should only be added if it is relevant to the article and makes it more complete. Excessive cases of recommending non-relevant articles should be reported to ieeeaccesseic@ieee.org*

Original Manuscript ID: Access-2020-44893

Previous Manuscript ID: Access- 2020-51106

Original Article Title: "Overview of Text Analysis Tasks, Applications and Approaches"

To: IEEE Access Editor

Re: Response to reviewers

Dear Editor,

Thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers' comments.

We are uploading (a) our point-by-point response to the comments (below) (response to reviewers), (b) an updated manuscript with yellow highlighting indicating changes, and (c) a clean updated manuscript without highlights (PDF main document).

Best regards,

Muhammad Yahya Saeed et al.

Reviewer#1, Concern # 1: Your paper was read with interest; however, proper grammar is a requirement for publication in IEEE Access and therefore we must reject. If needed, IEEE offers a 3rd party service for language polishing,

Author response: we agreed

Author action: The manuscript has been proof read by native English speaker