Machine Learning with Applications 7 (2022) 100249

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Initialization of profile and social network analyses robot and platform with a concise systematic review

Burak Omer Saracoglu

Orhantepe Mahallesi, Tekel Caddesi, Istanbul, Turkey

ARTICLE INFO

Keywords:
Decision Making Trial and Evaluation Laboratory
DEMATEL
Investment
Location
Profile analysis
Social network analysis

ABSTRACT

This paper presents profile and social network analyses on concise systematic review corpora. It suggests two new robots and platforms for profile and social network analyses, that will serve previously proposed data, expert, and event-driven robots and platforms for energy and power industry. The literature is collected and stored in three topic clusters "*location*", "*investment*", and "*DEMATEL*" to prepare corpora. Twenty-five publications are selected in each sample corpus. A sample dataset of each corpus is prepared for thirty-one features such as "author's full name and surname", "applied methods", and "publisher". Afterward, "authors network matrices" are prepared in spreadsheet software. Data input files (*.csv) are prepared for each dataset. Gephi 0.9.2 201709241107 (free open-source software) is used for social network analyses with built-in layout and statistics algorithms on a desktop Windows 10 Pro, Intel(R) Core(TM) i5 CPU 650 @ 3.20 GHz, 6,00 GB RAM personal computer in an offline and active cybersecurity software environment. Force Atlas, Force Atlas 2, Fruchterman–Reingold, OpenOrd, Yifan Hu, and Yifan Hu Proportional layout algorithms with Noverlap layout algorithm are run one by one. Runtimes range 2–120 s. All default statistic algorithms are run for several metrics like average degree, average weighted degree, betweenness centrality, closeness centrality, harmonic closeness centrality, eccentricity, and density. Authors in "*location*" cluster have a centralized network, but authors in "*investment*" and "*DEMATEL*" clusters have distributed networks. General profile analyses are conducted based on authors' publications in the literature without any data and information on social media sites and platforms. Two new profile analysis metrics are proposed as "researcher's past research focus index", and "researcher's future research focus prediction index". Detailed profile analysis is performed for only Burak Omer Saracoglu. All analyses and findings are compared and summarized in the end.

1. Introduction

This research presents the concepts of profile analyses robot and platform (*PARP*), and social network analyses robot and platform (*SNARP*). They may help the development activities of proposed umbrella frameworks entitled "*Global Power Prediction Systems: GP2S*", "*Global Power Plants Developers: GP2D*", "*Global Power Plants Engineers: GP2E*", and "*Global Power Plants Owners: GP2O*". *GP2S*, *GP2D*, *GP2E*, and *GP2O* are radical, but not utopic robots and platforms. They are effective solutions for global problems like consumption, demographic change, immigration, income inequality, pollution, warming, and wastage (Electronic Supplementary Material: ESM.1–3). *GP2S* presents and predicts power consumption and generation considering many probable and possible alternatives. *GP2D* focuses on finding, defining, identifying, describing, selecting, and clustering all features for power plants and their investments. Moreover, it generates location alternatives and recommends the best ones. *GP2E* performs design and engineering activities and recommends the best power plant designs. *GP2O* opens and offers all registered power plant investments on its platform to the World for crowd investment (examples of electronic trading platforms for other purposes ESM.4). The shares of a single power plant, large power plants as groups, or all power plants as one huge group in any project development, power plant construction, or power plant operation stages can be sold and bought in a long-term perspective on that transparent robot and platform. Shareholders and potential shareholders can access *GP2O* anytime and anywhere. It will allow ordinary people to buy and sell shares and get their earnings with only the net profit fair sharing principle. That means everybody and anybody in the World, and machines can buy and sell shares based on some predefined constraints. Above all, there is not any interest rate in the system. In other words, there is only earnings sharing. Thus, the interest rate is 0% in a different manner. Hence, it obeys Halal, Islamic, and Sharia investment guidelines. All these proposed umbrella frameworks will be founded on "geographic information systems (GIS)", "internet", "real-time", "near real-time", "web", and "state-controlled" principles. The "state control" idea is for organizing a transparent and accountable international technical governmental body for technical

E-mail address: burakomersaracoglu@hotmail.com.

https://doi.org/10.1016/j.mlwa.2022.100249
Received 2 February 2021; Received in revised form 4 December 2021; Accepted 7 January 2022
Available online 17 January 2022
2666-8270/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Nomenclature, Distinctions, Symbols, Abbreviations, and Acronyms

Abbreviations

AI	Artificial Intelligence
AIAA	American Institute of Aeronautics and Astronautics
ANN	Artificial Neural Network
BERT	Bidirectional Encoder Representations from Transformers
BOS	Burak Omer Saracoglu
CEO	Chief Executive Officer
CluE	Clustering-Enhanced Neural Model
CNN	Convolution Neural Network
COVID-19	Coronavirus 2019
CTO	Chief Technical Officer
CV	Curriculum Vitae
D3&D	Development, Demonstration, Deployment, and Diffusion
DEMATEL	Decision-Making Trial and Evaluation Laboratory Method
DF	Document Frequency
DLVN	Deep-Learning Vocabulary Network
DNN	Deep Neural Network
DoD	Department of Defense
DOE	Directors of Engineering
DRD	Directors of Research and Development
ESM	Electronic Supplementary Material
GIS	Geographic Information System
HITS	Hyperlink-Induced Topic Search
HTML	Hypertext Markup Language
IC	Intelligence Community
IDF	Inverse Document Frequency
IG	Information Gain
IMDb (IMDB)	Internet Movie Database
NN	k-Nearest Neighbor
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
MDPI	Multidisciplinary Digital Publishing Institute
ML	Machine Learning
MLP	Multilayer Perceptron Neural Network
MSc.	Master of Science
NER	Named Entity Recognition
NLP	Natural Language Processing
NN	Neural Network
PC (plural: PCs)	Personal Computer (plural: Personal Computers)
Ph.D.	Doctor of Philosophy
PNC	Number of Publications in Each Cluster
PNT	Total Number of Publications in All Clusters
Pos	Part Of Speech
RD3&D	Research, Development, Demonstration, Deployment, and Diffusion
RFRFPI	Researcher's Future Research Focus Prediction Index
RM	Reference Material
RNN	Recurrent Neural Networks
RPRFI	Researcher's Past Research Focus Index
RR	Ridge Regression
SLR	Systematic Literature Review
SNA	Social Network Analysis
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SSXUSVUAVC	Special Hybrid Vehicle – Unmanned Aerial Vehicle Unmanned Surface Vehicle – for surveillance, search and rescue, and similar including evacuation, firefighting, city and urban planning, reforestation, and similar operations doi:10.13140/RG.2.1.4013.8001 Saracoglu, 2013-2016
SVM	Support Vector Machine
TextCNNs	Text Convolutional Neural Networks
TF	Term Frequency
TF–IDF	Term Frequency–Inverse Document Frequency
TL	Turkish Lira (monetary currency of the Republic of Turkiye)
UAV	Unmanned Aerial Vehicle
UMLS	Unified Medical Language System
$Universe_{researchers}$	the universe of the researchers
$Universe_{researchdocuments}$	the universe of the research documents
USA	United States of America
USD	United States Dollar
USV	Unmanned Surface Vehicle
VV&A	Verification, Validation, and Accreditation
WEKA	Waikato Environment for Knowledge Analysis

Algorithms Layout Algorithms

Force Atlas	Force Atlas in the Gephi https://gephi.org
Force Atlas 2	Force Atlas 2 in the Gephi https://gephi.org
Fruchterman-Reingold	Fruchterman–Reingold in the Gephi https://gephi.org
Noverlap	Noverlap in the Gephi https://gephi.org
OpenOrd	OpenOrd in the Gephi https://gephi.org
Yifan Hu	Yifan Hu in the Gephi https://gephi.org
Yifan Hu Proportional	Yifan Hu Proportional in the Gephi https://gephi.org

issues like "fair taxation", "fair retirement", "fair health security", and "fair social security" of shareholders on the platforms. All these robots and platforms will work on any digital device like desktops, laptops, and smartphones. These robots and platforms will help to build net-zero or negative emissions power grids, reduce income inequality, and reduce poverty in the World. Accordingly, the net emissions in the World shall be zero or negative. King and Saracoglu (2018), Ohunakin and Saracoglu (2018), Saracoglu (2014, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g, 2018a, 2018b, 2018c, 2018d, 2019a, 2019b, 2019c, 2019d, 2019e, 2020a, 2021a, 2021b, 2021c, 2021d), Saracoglu et al. (2018), Saracoglu and de Simon Martin (2018), Solangi, Shah, Zameer, Ikram, and Saracoglu (2019), Tercan and Dereli (2020) and Tercan, Eymen, Urfali, and Saracoglu (2021)

Statistic Measures (Metrics) and Algorithms

Average Clustering Coefficient	Clustering Coefficient in Average Clustering Coefficient in the Gephi https://gephi.org
Average Degree	Average Degree in the Gephi https://gephi.org
Average Weighted Degree	Average Weighted Degree in the Gephi https://gephi.org
Betweenness Centrality	Betweenness Centrality in Network Diameter in the Gephi https://gephi.org Betweenness Centrality in Average Path Length in the Gephi https://gephi.org
Closeness Centrality	Closeness Centrality in Network Diameter in the Gephi https://gephi.org Closeness Centrality In Average Path Length in the Gephi https://gephi.org
Density	Density in Graph Density in the Gephi https://gephi.org
Eccentricity	Eccentricity in Network Diameter in the Gephi https://gephi.org Eccentricity in Average Path Length in the Gephi https://gephi.org
Eigenvector Centrality	Eigenvector Centrality in the Gephi https://gephi.org
Harmonic Closeness Centrality	Harmonic Closeness Centrality in Network Diameter in the Gephi https://gephi.org Harmonic Closeness Centrality in Average Path Length in the Gephi https://gephi.org
Hubs And Authority	Hubs and Authority in HITS
Pagerank	Pagerank in the Gephi https://gephi.org
Size	Size in Modularity in the Gephi https://gephi.org Size in Connected Components in the Gephi https://gephi.org

Common Math

β	beta

General

approx.	Approximately
e.g.	exempli gratia (for example)
i.e.	id est (that is)

Language

Islam	Islam as a religion to distinguish, discriminate, and differentiate from Islam as a name (first name, second name, or surname) The usage may differ and be very complex in texts and speeches. This word can come up with a few words in different combinations. Special care should be taken into account according to the context. Some usage is directly related to hateful conduct, Islamophobia, psychological war, violence, and war against Islam. Some usage is directly related to crypto in materials of newspapers, magazines, television and radio programs, cartoons, comics, and similar (see cryptology in Simmons, 2016). Some materials are related to destabilization, generation and demographic change, reconstruction, redesigning, regime change, transforming, and invasion of the Greater Middle East and North Africa.
Turkiye	Turkey as country boundaries to distinguish, discriminate, and differentiate from animal turkey as food. "A turkey is a large bird that is kept on a farm for its meat. Turkey is the flesh of this bird eaten as food". Collins (2021) For example, "HAPPY TURKEY DAY", "EAT TURKEY", "TURKEY FOR SALE", "Happy Turkey Day", "Eat Turkey", "Turkey For Sale" Some usage is directly related to crypto in materials of newspapers, magazines, television and radio programs, cartoons, comics, and similar (see cryptology in Simmons, 2016). Some materials are related to destabilization, generation and demographic change, reconstruction, redesigning, regime change, transforming, and invasion of the Greater Middle East and North Africa.

Machine Learning Datasets and Platforms ML Datasets

AG's News	AG's News Topic Classification Dataset https://www.kaggle.com/amananandrai/ag-news-classification-dataset
Amazon Product Reviews	Amazon Review Data (2018) https://nijianmo.github.io/amazon/index.html
Amazon Q/A	Amazon question/answer data http://deepx.ucsd.edu/public/jmcauley/qa/
COVID-19	COVID-19 Dashboard (Dong, Du, & Gardner, 2020) https://github.com/CSSEGISandData/COVID-19 https://gisanddata.maps.arcgis.com/apps/
Corona Virus Tagged	Coronavirus tweets NLP — Text Classification https://www.kaggle.com/datatattle/covid-19-nlp-text-classification

Similarly, the new proposed *PARP* and *SNARP* are also revolutionary, but not utopic robots and platforms. While developing, demonstrating, and presenting *GP2S*, *GP2D*, *GP2E*, and *GP2O* with their valuable publications' review processes, it has been realized that all of them have been getting important day by day for the World and they can be developed more cost-effectively, collaboratively, collectively, and inclusively in all aspects with a voluntary crowd-development strategy. For that reason, *PARP* and *SNARP* are proposed as straightforward automatic tools. They can assist social networking and profiling of researchers, finding and managing very large research groups, taking attention and interest of researchers into *GP2S*, *GP2D*, *GP2E*, and *GP2O*, and making them focus on developing, demonstrating, and improving all these robots and platforms. *PARP* and *SNARP* are ground-breaking ideas, because after a reasonable searching period it has been realized that there have not been any profile and social network analyses tools in the software product market that can be used directly for any similar aim and/or manner. Hence, *PARP* and *SNARP* development activities are required in a long RD3&D perspective. In short, the first motivation of this study is to present the *PARP* and *SNARP* concepts.

Credit Card Fraud Detection	Credit Card Fraud Detection https://www.kaggle.com/mlg-ulb/creditcardfraud
DBpedia	DBpedia http://wikidata.dbpedia.org/develop/datasets
Fake and real news	Fake and real news, classifying the news https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
Fake News	Fake News https://www.kaggle.com/c/fake-news/data
IMDB movie reviews	IMDB Dataset of 50 K Movie Reviews https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
Parkinsons Data Set	Parkinsons Data Set https://archive.ics.uci.edu/ml/datasets/parkinsons
Social circles: Facebook	Social circles: Facebook https://snap.stanford.edu/data/egonets-Facebook.html
Social circles: Google+	Social circles: Google+ https://snap.stanford.edu/data/egonets-Gplus.html
Social circles: Twitter	Social circles: Twitter https://snap.stanford.edu/data/egonets-Twitter.html

ML Platforms

CrowdAI	https://www.crowdai.org/
CrowdAnalytix	https://www.crowdanalytix.com/
DrivenData	https://www.drivendata.org/
Kaggle	https://www.kaggle.com/
KNIME Analytics Platform	https://www.knime.com/knime-analytics-platform
OpenML (beta 2 version)	https://www.openml.org/home

Proposed Robots and Platforms (1)

GP2D	Global Power Plants Developers https://www.researchgate.net/project/Global-Power-Plants-Developers https://github.com/burakomersaracoglu/PublicationScripts/tree/
GP2E	Global Power Plants Engineers https://www.researchgate.net/project/Global-Power-Plants-Engineers https://github.com/burakomersaracoglu/PublicationScripts/tree/
GP2O	Global Power Plants Owners https://www.researchgate.net/project/Global-Power-Plants-Owners https://github.com/burakomersaracoglu/PublicationScripts/tree/
GP2S	Global Power Prediction Systems https://www.researchgate.net/project/Global-Power-Prediction-Systems https://github.com/burakomersaracoglu/PublicationScripts/tree/
PARP	Profile Analyses Robot and Platform
SNARP	Social Network Analyses Robot and Platform

Synonymous

Adjacency lists (plural: Adjacency lists)	Edge list (plural: Edge lists)
Cluster (plural: Clusters)	Group, Set (plural: Groups, Sets)
Edge (plural: Edges)	Link, Tie (plural: Links, Ties)
Feature (plural: Features)	Criterion, Factor, Measure, Parameter, Variable (Criteria, Factors, Measures, Parameters, Variables)
Features Universe	Features Planet, Features Ocean, Features Lake, Features List, Features Pool, Features Set, Features Warehouse (different data and information sizes)
Gray	Gray
Method (plural: methods)	Techniques (plural: techniques)
Network (plural: Networks)	Graph (plural: Graphs)
Node (plural: Nodes)	Vertex (plural: Vertices)

Social Networking Platforms

Facebook	Facebook https://www.facebook.com/
Google+	Google+ www.plus.google.com/
Google Scholar	Google Scholar https://scholar.google.com/
LinkedIn	LinkedIn https://www.linkedin.com/
ResearchGate	ResearchGate https://www.researchgate.net/
TikTok	TikTok https://www.whatsapp.com/
Twitter	Twitter https://twitter.com/
WhatsApp	WhatsApp https://www.whatsapp.com/

Software

Apache OpenOffice	Apache OpenOffice https://www.openoffice.org/
Carrot^2	Search Results Clustering Engine https://search.carrot2.org
DeepFakes	DeepFake software and tools https://github.com/deepfakes https://github.com/perov/DeepFaceLab https://github.com/llSourcell/deepfakes https://github.com/aerophile/awesome-deepfakes

It is worth mentioning some key representative terms and success factors related to the development and operation of *GP2S*, *GP2D*, *GP2E*, *GP2O*, *PARP*, and *SNARP* as follows: "accountability", "accuracy", "bias freeness", "cognition", "collaboration", "easiness", "fact", "intellectualism", "openness", "perception", "realism", "re-runability", "repeatability", "reproducibility", "reusability", "replicability", "simplicity", "traceability", "transparency", "trustworthiness", and "truth". Accordingly, data, document, information, network, network analysis, profile, profile analysis, social network, and social network analysis are crucial for those robots and platforms. Standardization, correct data, and correct information are crucial for profile, network, and social network analyses. Standardized data and information universe and its files can play key roles in the effectiveness of *PARP*, and *SNARP*. During the research period of *PARP* and *SNARP*, the author has corresponded

Gephi 0.9.2	Gephi https://gephi.org
LibreOffice	LibreOffice https://www.libreoffice.org/
Microsoft Office Excel	Microsoft Office Excel https://products.office.com/en-in/excel
Microsoft Paint 3D	Microsoft Paint 3D https://www.microsoft.com/en-us/p/paint-3d/9nblggh5fv99
Paint.NET 4.2.12	Paint.NET 4.2.12 https://www.getpaint.net/index.html
Wikipedia	Wikipedia https://en.wikipedia.org/
Some Websites for Data and Information Gathering or Acquisition	
Google	Google www.google.com
Websites of Some Publishers	
Elsevier	Elsevier http://www.elsevier.com/
Springer	Springer https://www.springer.com/

[1] The proposed robots and platforms by Burak Omer Saracoglu in their development phases as prototype versions 2020

with many publishers like Elsevier (http://www.elsevier.com/), and Springer (https://www.springer.com/). The author has asked them for those sorts of data and information files related to publishers' publications. The author has also requested information whether they have had any preparation, development, and publishing ideas of those sorts of files. The author has learned that publishers have not had any data and information files of their publications for profile, network, and social network analyses. In addition, they have not had any plans or intentions of developing anything for those purposes. These findings are the facts and basic RD3&D gaps. In that sense, the last motivation of this study is to contribute to prepare standardized corpora, and standardized files presenting publication data and information to support and feed artificial intelligence (AI), and machine learning (ML) infrastructures, platforms, and applications.

This paper has six sections. The second section presents ML-related literature. The third section proposes a simple ML-related solution. The fourth section describes the selected metrics in the literature and newly proposed metrics in this study. The fifth section presents the systematic review, corpora, and sample dataset preparation in this study. The final section summarizes the accomplishments, limitations, and further research.

2. Related machine learning literature in a broader view

PARP and *SNARP* are related to text analysis, text classification, text categorization, text clustering, topic analysis, topic classification, topic categorization, topic clustering, social network analysis, natural language processing (NLP), sentiment analysis, clustering, regression, and anything similar in a general ML perspective. There are numerous publications on these ML topics in the literature. A few of them with their critical points are summarized in this section.

The research publications were as follows:

Agarap (2021) researched text classification and text clustering with neural networks (NN) on NLP. The author pointed out the shortcomings of n-Grams ("each word acting as a token") and Term Frequency–Inverse Document Frequency (TF–IDF) ("scoring the importance of a word with respect to a document and to the entire corpus") such as the lack of dimensionality, and context capturing, and then explained word embeddings and sentence embeddings. The author emphasized the contribution by disentangling the text representation idea with feed-forward NN and convolutional neural network (CNN). The author performed the experiments on only one dataset (AG's News) due to time and computational limitations. There were 7600 documents amongst 119 843 documents as a subset with removed stop words and removed words less than three characters in the training dataset. There were some internal and external criteria for clustering evaluation like Davies–Bouldin Index, Silhouette Score, Normalized Mutual Information, and Adjusted Rand Index. Ahmed, Iqbal, and Bakhsh (2019) explained the difficulties to extract meaningful and reliable opinions, and emotions in social media contents like tweets with ML (sentiment analysis), and how ML had failed in the past. They summarized the literature very well by giving the information of datasets, encoding techniques, and classifiers. A few examples of datasets, encoding techniques, and classifiers were respectively IMDB movie reviews, Twitter, Google, Wikipedia English (datasets); Word2vec with n-gram, CNN (encoding techniques); and multilayer perceptron neural network (MLP), support vector machine (SVM) (classifiers). They proposed a new methodology for data extraction, data pre-processing, feature encoding, and so forth. They tested their methodology in some benchmark datasets with some performance metrics (i.e. accuracy, specificity, recall, precision, F-measure). AlHarbi, AlHarbi, AlZahrani, Alsheail, and Ibrahim (2020) focused on a crucial subject, "cyberbullying". It had previously been learned from many different sources that cyber bullying had taken human lives. It was an instrument for "murder", "genocide", "homicide", "terror", and "war". Hence, AlHarbi et al. (2020) studied a crucial topic. They reviewed ML algorithms for the detection of cyberbullying automatically. They came up with a recommendation for using Ridge Regression (RR) and Logistic Regression (LR) techniques to increase accuracy in cyberbullying detection in Arabic social media. Boddu and Devi (2020) also tried to detect and present cyberbullying on social media. The authors presented some bullying 1-gram examples. Their paper also gave an idea of what teenagers and pre-teens had thought about cyberbullying attacks (sad, depressed, sorrow). They proposed an SVM algorithm and tested it. However, the feedback of victims was not expressed in the paper to understand whether their ML model had failed or not. Chai, Zhang, and Jin (2020) presented a study in neural text classification that was important for spam detection, spoken dialogue systems, and so forth. The authors proposed an ad hoc Clustering-Enhanced neural model (CluE) to overcome some drawbacks in the deep learning methods such as TextCNNs (Text Convolutional Neural Networks), and BERT (Bidirectional Encoder Representations from Transformers). The datasets in their experiments were AG's News, DBpedia, and IMDB movie reviews. The authors presented some visualization for the impact (text representation, clustering centroid, cluster-token alignment). An informative confusion matrix was also given with actual and predicted labels for error analysis. Gannarapu, Dawoud, Ali, and Alwan (2020) developed a system for bot detection and removal on social media platforms to make them free of false information. The authors' system classification with columns like research area (bot detection in Twitter, in social media platforms), signature text data (public dataset, critical terms, and hashtags, network traffic cache data), a user history (user-generated content, structural patterns, critical terms, and hashtags), correlation selection (linear programming, SVM, decision tree models, multivariate Hawkes process), univariate selection (recursive feature elimination, F-Test, Bayesian probabilistic graphical model), bot analysis (motif detection, network visualization, communication frequencies) were very informative. Gurung and Wagh (2017) studied topic identification in big documents and small documents. The authors underlined the importance of topic identification for reliable information retrieval. Their generic model consisted of items such as pre-processing of text (lower case conversion, tokenization, stop word removal, stemming, number removal, and punctuations removal), preparing term-document matrix, and representing vector space model; applying clustering algorithm, and identifying the topic. The authors applied the k-means algorithm to abstracts and full papers to compare the findings with final words

emphasizing the importance of "availability of big documents on web". Weißer, Saßmannshausen, Ohrndorf, Burggraf, and Wagner (2020) presented a topic filtering in a systematic literature review (SLR) method based on NLP, article metadata, and k-means clustering algorithm. Kim and Gil (2019) proposed a system that had extracted key terms from abstracts of papers and topics by Latent Dirichlet allocation (LDA) scheme. The authors used a k-means clustering algorithm to classify the whole paper into different subjects with TF–IDF values. The author presented the system flow diagram as five blocks (crawling, data management, topic modeling, TF–IDF, classification). Many steps and activities were presented in an informative way (e.g. abstract before and after preprocessing, keyword sets, topic sets by LDA, TF, and document frequency calculations). That journal article could be used as a template structure for journal papers too. Malik and Jain (2020) studied comparatively with 10 000 tweets in the test set and 40 000 tweets in the training set for logistics regression, naive Bayes, SVM, random forest, and k-Nearest Neighbor (k-NN). The highest accuracy, precision, recall, and F1-score for the types of word N-gram, Char N-gram, and Word-Char N-gram of models were respectively 0,7902 (logistics regression, Word-Char), 0,8678 (logistics regression, Word), 0,7944 (logistics regression, Word-Char), and 0,7935 (logistics regression, Word-Char). It seemed that the authors' logistics regression model was much better than their k-NN, naive Bayes, random forest, and SVM models. A few other research publications were Mombini, Li, Zhang, Korkin, and Tulu (2020) in social media analysis for rare diseases like Trigeminal Neuralgia, Gokarn Nitin, Gottipati, and Shankararaman (2019) in topic analysis and clustering, Pourvali, Orlando, and Omidvarborna (2018) in topic analysis and text cluster labeling, Sato, Kawashima, Okuda, and Oku (2018) in new topic detection, topic analysis and clustering, Pugachev and Burtsev (2021) in short text clustering with sentence vector representations of transformers, Xu et al. (2017) in short text clustering with flexible self-taught convolutional neural networks, Zhang, Dong, Yin, and Wang (2021) in short text clustering with adversarial training integrated unsupervised clustering, Suleymanov, Kalejahi, Amrahov, and Badirkhanli (2019) in Azerbaijani language text classification, Yohannes and Assabie (2021) in Amharic text clustering, Wang, Gao, Wei, Ma, Liu et al. (2020) in text clustering on tweets, Yi, Zhang, Zhao, and Wan (2017) in text clustering with deep-learning vocabulary network (DLVN), Rao, Sreeram, and Raju (2020) in fake accounts on social media, social networks, social online platforms, and their detection, Merlini and Rossini (2021) in text categorization ML application in Waikato Environment for Knowledge Analysis (WEKA), and Tan, Liu, and Hu (2019) in social network analysis with deep representation learning.

The review and survey publications were as follows:

Aggarwal and Zhai (2012) surveyed text-clustering algorithms. They defined clustering as "finding groups of similar objects in the data". They explained feature selection methods (e.g. document frequency-based selection, term strength); Latent Semantic Indexing based methods, non-negative matrix factorization techniques, distance-based clustering algorithms, and so forth. The authors explained everything in sentences and equations. Surprisingly, there were not any summarizing figures or tables in that book chapter. Ali et al. (2020) reviewed text-clustering algorithms too. The authors focused on big data clustering and divided the main algorithms into four groups (hierarchical algorithms: Balanced Iterative Reducing and Clustering Using Hierarchies, Clustering Using Representatives; partition-based algorithms: k-means clustering, fuzzy c-means clustering; density-based algorithms, ordering points to identify the clustering structure; grid-based algorithms, STatistical Information Grid). Surprisingly, their summary figure had only 4 main groups, not subgroups. Allahyari et al. (2017) presented a brief survey on text mining, in other words knowledge discovery from text (information retrieval, NLP, information extraction from text, text summarization, unsupervised learning methods, supervised learning methods, probabilistic methods for text mining, text streams and social media mining, opinion mining and sentiment analysis); text representation and encoding (text preprocessing: tokenization, filtering, lemmatization, stemming; vector space model); text classification (naive Bayes classifier: multi-variate Bernoulli model, multinomial model; nearest neighbor classifier; decision tree classifiers; SVM); text clustering (hierarchical clustering; k-means clustering; probabilistic clustering: topic modeling); information extraction (named entity recognition, hidden Markov models, conditional random fields, relation extraction); and applications in biomedicine and healthcare domains (e.g. Unified Medical Language System: UMLS, Systematized Nomenclature of Medicine Clinical Terms: SNOMED CT, Named Entity Recognition: NER). Their study showed very well how deep, large, and wide the ML topic had been in theory and applications. Thangaraj and Sivakami (2018) presented a literature review in text classification techniques. The authors summarized their review with an informative figure. The text classification was divided into two parts (statistical, ML) at first. Then ML part was divided into three parts as supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning had two main classifiers as a parametric classifier with two methods logistic regression, and naive Bayes; a non-parametric classifier with five methods SVM, decision tree, rule induction, k-NN, and NN. Fuzzy c-means, k means clustering, and hierarchical clustering were grouped under unsupervised learning. Co-training, self-training, transductive SVM, and graph-based methods were grouped under semi-supervised learning. The authors also tabulated advantages, disadvantages, and applications in a pretty way. Dalal and Zaveri (2011) summarized an automatic text classification generic strategy based on a technical review. That strategy included pre-processing (removing stop-words, stemming: "reducing words to their root or base form", removing Hypertext Markup Language: HTML tags, and similar), feature extraction/selection (TF–IDF, latent semantic indexing, multi-word), selecting appropriate ML model (naive Bayes, decision tree, NN, SVM, and similar), training classifier, and similar. Khan, Baharudin, Lee, and Khan (2010) reviewed text and documents classification with ML algorithms pointing out feature extraction with common steps like tokenization, removing stop words, stemming word; feature selection with feature evaluation metrics like gain ratio, information gain (IG), Chi-square, conditional mutual information, document frequency (DF), TF, IDF, expected cross-entropy, odds ratio, weight of evidence of text, term frequency and document frequency (TF/DF); semantic issues like sentence splitting, part-of-speech (pos) tagging, word sense; and automatic text classification ML algorithms like Rocchio's Algorithm, k-NN, decision tree, decision rules classification, naive Bayes algorithm, Artificial Neural Network (ANN), and SVN. Kowsari, Meimandi, Heidarysafa, Mendu, Barnes et al. (2019) presented a pretty well organized survey in text classification algorithms. The authors gave information in many parts of the subject by comparing the traditional methods and deep learning techniques for their interpretability and accuracy. According to their figure, linear regression was the most interpretable but the least accurate method, and deep neural networks were the least interpretable but the most accurate method. The color scheme did not show which methods were the traditional methods and which were deep learning techniques according to the authors' point of view. The authors also collected information about the performance measures for the evaluation of algorithms (i.e. accuracy, sensitivity, specificity, precision, recall, F1-Score). The paper was finalized with a summarizing discussion section showing the advantages and limitations of feature extraction models, text classification comparison algorithms, and metrics pitfalls. A few other review and survey publications were Minaee, Kalchbrenner, Cambria, Nikzad, Chenaghlu et al. (2021) in text classification with deep learning methods, Asgari-Chenaghlu, Feizi-Derakhshi, Farzinvash, Balafar, and Motamed (2021) in Twitter topic detection and tracking techniques, Gupta (2017) in keyword extraction, and Li (2018) in texts analysis of patent data.

The importance of datasets in ML was clearly emphasized in many publications in the literature. The datasets related issues and some datasets were also explained in the following papers Botelho and Antunes (2011), Di Capua, Di Nardo, and Petrosino (2016), Edwards, Wattam, Rayson, and Rashid (2016), García (2015), Gauen et al. (2017),

Gebru et al. (2021), He and McAuley (2016), Islam, Ashad Kabir, Ahmed, Kamal, Wang et al. (2018), Kayacik and Zincir-Heywood (2015), Li, Su, Shen, Li, Cao et al. (2017), McAuley and Leskovec (2012), McAuley, Targett, Shi, and van den Hengel (2015), McAuley and Yang (2016), Ni, Li, and McAuley (2021), Norregaard, Horne, and Adali (2019), Wan and McAuley (2016), Whytock et al. (2020), and Yavanoglu and Aydos (2017).

In this section, it was also worth mentioning to give some references of ML books that had presented many ML algorithms and models in detail such as Barber (2017), Barocas, Hardt, and Narayanan (2020), Hutter, et al. (2019), MEAP (2021), Mohri, Rostamizadeh, and Talwalkar (2018), and Moitra (2014).

According to the literature review, it was understood that ML was a very deep and wide topic with its many algorithms and models like maximum entropy, Recurrent Neural Networks (RNN), deep neural networks (DNN), k-NN, and SVM. There were also very informative sources in ML topics for developers like Google (Google, 2021a, 2021b, 2021c, 2021d).

In short, there were many ML algorithms with many ML models applied in different text, topic, social network, and language-related problems. The authors used many different datasets that had been presented by others or prepared by their efforts before their ML applications. They came up with many different findings in their ML applications. There were not any baseline or benchmark datasets in most cases in the literature. Also, there was not any verification and validation activity in the literature. In short, the authors did not get any opinions and feedback for their ML applications from text, topic, and speech owners (e.g. cyberbullying victims, authors of texts, owners of tweets). All studies depended on ML algorithms and models, but nobody investigated whether their findings were relevant to text, topic, and speech owners' thoughts or not. Those sorts of issues were very clear research gaps in the literature. It was also realized that surveying and studying ML had to be organized in more manageable ways like focusing on only text preprocessing (e.g. text cleaning, syntactic word representation), dimensionality reduction (e.g. component analysis, autoencoder), or classification techniques (e.g. boosting, bagging). In that manner, the scope of publications would be narrower, so some better findings could be gathered in a sufficient period. Of course, general overviews needed to continue with covering all general issues in a more summarizing way. In addition, the challenges of ML were presented well enough in the literature.

According to the difficulties of ML underlined in the literature, a simple D3&D solution is presented for the current development activities of *PARP* and *SNARP* in the next section.

3. Proposed simple machine learning related solution

The mind behind the design philosophy of *PARP* and *SNARP* with their future ML related development activities is only the "simplicity" based on the following mottos:

"Nature is pleased with simplicity. And nature is no dummy". Isaac Newton, "Simplicity is the ultimate sophistication". Leonardo da Vinci, "That's been one of my mantras – focus and simplicity. Simple can be harder than complex". Steve Jobs, "Everything should be made as simple as possible, but not simpler". Albert Einstein

A step-by-step and very practical D3&D development roadmap is planned in this study to overcome the challenges of ML, to improve the way how ML research has been conducted, and to follow a "simplicity" design philosophy in *PARP* and *SNARP*. The first steps of this D3&D development roadmap are related to dataset preparation, basic evaluation, and basic analysis that include social network analysis and regression analysis. After sufficient achievements and accomplishments are made in the following years, automation and automization can be the focus of D3&D activities including automatic data acquisition from text and social media. In the end, near real-time or real-time applications can be developed and demonstrated to the World. Until some great achievements and accomplishments are made during that long D3&D path, everything and anything related to the proposed robots and platforms will be studied and developed one by one in small and manageable parts. Accordingly, this study focuses on dataset preparation, social network analysis, and profile analysis. The sample datasets are for developing some baseline and benchmark datasets with sufficient and reliable data, and information.

In this section, it is also wise to give some ML platforms and datasets as examples to describe the proposed solution better. A few examples of ML platforms are CrowdAI (https://www.crowdai.org/), CrowdAnalytix (https://www.crowdanalytix.com/), DrivenData (https://www.drivendata.org/), KNIME Analytics Platform (https://www.knime.com/knime-analytics-platform), Kaggle (https://www.kaggle.com/), and OpenML (beta 2 version) (https://www.openml.org/home). A few examples of datasets are Fake News (https://www.kaggle.com/c/fake-news/data), Parkinson's Data Set (https://archive.ics.uci.edu/ml/datasets/parkinsons), and Credit Card Fraud Detection (https://www.kaggle.com/mlg-ulb/creditcardfraud).

In this section, it is also wise to give a few key terminologies related to the current context as follows:

Label: "*An answer for a prediction task - either the answer produced by a machine learning system, or the right answer supplied in training data. For example, the label for a web page might be "about cats"*" (Zinkevich in Google Developers, 2021).

Label set: "*A label set is the set of labels you want the human labelers to use to label your images. For example, if you want to classify images based on whether they contain a dog or a cat, you create a label set with two labels: "Dog" and "Cat".*" (Google Cloud AI Platform, 2021a, 2021b).

Dataset: "*A dataset is the collection of data items you want the human labelers to label. It contains representative samples that you want to classify or analyze. Well labeled dataset can be used to train a custom model.*" (Google Cloud AI Platform, 2021c). Attention to human labelers please.

Feature: "*A property of an instance used in a prediction task. For example, a web page might have a feature "contains the word "cat""*" (Zinkevich in Google Developers, 2021).

Feature Column: "*A set of related features, such as the set of all possible countries in which users might live. An example may have one or more features present in a feature column. Feature column is Google-specific terminology. A feature column is referred to as a "namespace" in the VW system (at Yahoo/Microsoft), or a field.*" (Zinkevich in Google Developers, 2021).

Metric: "*A number that you care about. May or may not be directly optimized.*" (Zinkevich in Google Developers, 2021).

"*Classification tasks, where labelers assign one or more labels to each text segment. You could specify the number of labelers to label each text segment. Data Labeling Service does a majority vote to determine the proper labels.*" (Google Cloud AI Platform, 2021a, 2021b, 2021c, 2021d)

The differences in the terminology for the same or similar items and issues are expressed with a good example by Zinkevich in Google Developers (2021) as presented above. Whether the developers prefer to use A, B, or C for the same element, item, issue, and subject like "feature column", "namespace", or "field", they do or perform the same things in AI and ML scope. The current text is organized in cooperation with the terminology above.

If the proposed D3&D approach of this study is accepted by researchers all over the World, all researchers may share their datasets on those platforms or similar in the next years. After a while, if and when some or all publishers start to think and make their business models similar to the proposed approach in this study, they can start to prepare very large datasets (e.g. publication topic by authors' decision, publication topic by editors' and reviewers' decision, topic by publishers' decision) and share them on their websites, those platforms or similar. Afterward, if and when they may come up with a common decision to share their publications' corpus and corpora in a standard manner (e.g. without any stop words, punctuations), they can prepare and share them on their websites, those platforms, or similar. Then ML researchers will be able to be succeeding in testing their ML algorithms and models appropriately. Those activities shall be true scientific achievements and accomplishments.

4. Selected and proposed metrics

There are many metrics for different aims in the literature like density, the minimum number of actors (De et al., 2012), authority centrality (Nurek & Michalski, 2020), and normalized atedge length (Venturini, Jacomy, & Jensen, 2019) for network analysis related subjects like social network analysis (SNA). Some of them can be used in *PARP* and *SNARP* for ML models and applications. Of course, there is always room to propose new metrics too. *PARP* and *SNARP* require metrics for not only their development and demonstration but also for their operation, in other words running automatically. According to their design approach and their current development stage, some metrics are related to SNA, and some others are related to regression analysis. The metrics that have been taken directly in the literature (e.g. average clustering coefficient, betweenness centrality, degree-centrality) and newly proposed ones during this study are presented shortly (Table 1).

There are forty-two metrics in the table that can be used for *SNARP*. Some of them are the same metrics in the different terms as synonymous. Some of them may also be the same or different according to the different perspectives of the authors. There are two new metrics proposed for profile analysis of *PARP* during this D3&D period. The first one is the researcher's past research focus index (RPRFI) for a historical or ex-post analysis. The second one is the researcher's future research focus prediction index (RFRFPI) for prediction or ex-ante analysis. These two new metrics also show the invention capability of any researcher. If these two new metrics are extended for some studies and applications in the researchers', engineers' and any similar's backgrounds files like curriculum vitae (CV) datasets, the future possible positions and possible earnings of any person (e.g. Chief Executive Officer: CEO, Chief Technical Officer: CTO, Directors of Engineering: DOE, Directors of Research and Development: DRD, 2 000 000 TL/year, 5 000 000 TL/year, 2 000 000 USD/year, 5 000 000 USD/year) can be predicted under the assumption of the same condition (e.g. war, health). Of course, that sort of information may be used for "criminal attacks" like "espionage", "bullying", "cyberbullying", "sabotage", "poisoning", "assassination", or intelligence monitoring, and security precautionary and preventive actions (before incidents/events). Those sort of "criminal attacks" are in the scope of "biological", "chemical", "cyber", "economic", "financial", "food", "health", "hybrid", "water", and any similar warfare types. And of course, as vice-versa that sort of information may be used for criminal investigations, after incidents/events like cyber warfare or assassination. Their proposed equations and descriptions are short as follows:

Researcher's Past Research Focus Index (RPRFR)		
Equation.1	$RPRFI = \frac{PNC}{PNT}$ where PNC is the number of publications in each cluster, that may be for topics, industries, or similar subject PNT is the total number of publications in all clusters.	(1)
Description.1	The ratio of publications in a subject (cluster) to the total number of publications can present the researcher's study area in the past very well. The RPRFI value ranges between 0 and 1. The higher the value in RPRFI shows that the researcher has focused on that subject more than the other subjects.	

Researcher's Future Research Focus Prediction Index (RFRFPI)		
Researcher's Past Research Focus Index (RPRFR)		
Equation.2	The general simple linear regression equation is $y = a + bx$ (Berk & Carey, 2010; Zafarani, Abbasi, & Liu, 2014) where y is the dependent variable, a is the intercept or constant term (coefficient), b is the slope (coefficient), and x is the independent variable (predictor). Here y represents the ratio of publications in a subject (cluster) to the total number of publications in a time series as future prediction, x is the time variable, years (annuity), a represents the RFRFPI and b represents the simple linear regression equation coefficient for calculating the ratio of publications in a subject (cluster) to the total number of publications in a time series as a future prediction. Please see the following sections for examples.	(2)
Description.2	A simple linear regression model (simple regression model) in a time series is sufficient to present the trend of publications of a researcher in a subject (cluster). The RFRFPI value may be − and +. The higher the positive value in RFRFPI shows that the researcher will focus on that subject more than the other subjects. The lower the positive value in RFRFPI shows that the researcher will focus on other subjects more than that subject. When the value is negative, the researcher will not study that subject in the future.	

In this section, it is also wise to underline that data is the key for any SNA. There are many social networking platforms available for users and potential users like Facebook (https://www.facebook.com/), Google+ (www.plus.google.com/), Google Scholar (https://scholar.google.com/), LinkedIn (https://www.linkedin.com/), ResearchGate (https://www.researchgate.net/), TikTok (https://www.whatsapp.com/), and WhatsApp (https://www.whatsapp.com/). They are great data providers and excellent information sources for SNA. Publishers can be great data providers for these kinds of SNA, *PARP*, and *SNARP*. These kinds of SNA are very difficult and time incentive without any technical support from publishers. *PARP* and *SNARP* cannot be near real-time or real-time without the contributions of publishers.

5. Concise systematic review, corpora, and sample dataset preparation

There are many power technologies in *GP2S*, *GP2D*, *GP2E*, and *GP2O*. For instance: hydro, solar, wind. Hydropower plants are generally classified into six technology sub-groups by their installed power or installed capacity in the literature according to the author's review activities until 01/06/2020 (Saracoglu, 2015b; ESM.5–6). Those six hydropower technology sub-groups are large, medium, small, mini, micro, and pico. They are divided into eight sub-groups in this D3&D activity for better investment discrimination. These eight sub-groups are giant or huge, mega or very large, large, medium, small, mini,

Table 1
A few metrics in the literature and new ones in this study.

Appropriate for SNARP				
No	Metrics	Equation	Description	References
1	Average clustering coefficient	See RM	See RM	Gephi (2021) Gephi github (2021)
2	Average degree	See RM	See RM	Gephi (2021) Gephi github (2021)
3	Average path length	See RM	See RM	Grandjean and Jacomy (2019)
4	Average weighted degree	See RM	See RM	Gephi (2021) Gephi github (2021)
5	Betweenness centrality	See RM	See RM	De et al. (2012) Gephi (2021) Gephi github (2021) Grandjean and Jacomy (2019) Nurek and Michalski (2020) Verma and Goyal (2021)
6	Centralization-centrality	See RM	See RM	De et al. (2012)
7	Closeness	See RM	See RM	De et al. (2012)
8	Closeness centrality	See RM	See RM	De et al. (2012) Gephi (2021) Gephi github (2021) Grandjean and Jacomy (2019) Nurek and Michalski (2020) Verma and Goyal (2021)
9	Clustering coefficient (average clustering coefficient)	See RM	See RM	Gephi (2021) Gephi github (2021)
10	Clustering coefficient (total triangles)	See RM	See RM	Gephi (2021) Gephi github (2021)
11	Degree	See RM	See RM	De et al. (2012) Jacomy (2020)
12	Degree-centrality	See RM	See RM	De et al. (2012) Verma and Goyal (2021)
13	Density	See RM	See RM	De et al. (2012) Gephi (2021) Gephi github (2021) Grandjean and Jacomy (2019)
14	Eccentricity	See RM	See RM	Gephi (2021) Gephi github (2021)
15	Eigenvector centrality	See RM	See RM	Gephi (2021) Gephi github (2021) Grandjean and Jacomy (2019) Nurek and Michalski (2020)
16	Eigenvector centrality (number of iterations)	See RM	See RM	Gephi (2021) Gephi github (2021)
17	Eigenvector centrality (sum change)	See RM	See RM	Gephi (2021) Gephi github (2021)
18	Flow betweenness-centrality	See RM	See RM	De et al. (2012)
19	Geodesic distance (synonymous of the shortest path)	See RM	See RM	De et al. (2012)
20	Graph distance average path length	See RM	See RM	Gephi (2021) Gephi github (2021)
21	Graph density	See RM	See RM	Gephi (2021) Gephi github (2021)
22	Graph distance diameter	See RM	See RM	Gephi (2021) Gephi github (2021)
23	Graph distance radius	See RM	See RM	Gephi (2021) Gephi github (2021)
24	Harmonic closeness centrality	See RM	See RM	Gephi (2021) Gephi github (2021)
25	Hub centrality	See RM	See RM	Nurek and Michalski (2020)
26	Indegree centrality	See RM	See RM	Nurek and Michalski (2020)
27	Katz–Bonacich Centrality	See RM	See RM	Verma and Goyal (2021)

(*continued on next page*)

Table 1 (*continued*).

Appropriate for SNARP				
No	Metrics	Equation	Description	References
28	Local clustering coefficient	See RM	See RM	Nurek and Michalski (2020) Verma and Goyal (2021)
29	Modularity	See RM	See RM	Gephi (2021) Gephi github (2021)
30	Modularity number of communities	See RM	See RM	Gephi (2021) Gephi github (2021)
31	Modularity with resolution	See RM	See RM	Gephi (2021) Gephi github (2021)
32	Network average path length	See RM	See RM	Gephi (2021) Gephi github (2021)
33	Network diameter	See RM	See RM	Gephi (2021) Gephi github (2021)
34	Network radius	See RM	See RM	Gephi (2021) Gephi github (2021)
35	Number of weakly connected components	See RM	See RM	Gephi (2021) Gephi github (2021)
36	Outdegree centrality	See RM	See RM	Nurek and Michalski (2020)
37	Pagerank (synonymous of page rank, typo differences)	See RM	See RM	Gephi (2021) Gephi github (2021) Nurek and Michalski (2020)
38	Path length	See RM	See RM	De et al. (2012)
39	Radiality	See RM	See RM	De et al. (2012)
40	Reach	See RM	See RM	De et al. (2012)
41	Shortest path (synonymous of geodesic distance)	See RM	See RM	De et al. (2012)
42	Size	See RM	See RM	Gephi (2021) Gephi github (2021)
Appropriate for PARP				
No	Metrics	Equation	Description	References
43	Researcher's past research focus index[a] (RPRFI)	See Equation.1	See Description.1	this study
44	Researcher's future research focus prediction index[a] (RFRFPI)	See Equation.2	See Description.2	this study

Note: (1) some of these metrics may be the same, but have not been discriminated during this study, because of the reference materials and their authors' usage, (2) RM: reference material, please see citations, please use them as in their original publications.

[a]Metrics proposed in this study: RPRFI: Researcher's past research focus index, RFRFPI: Researcher's future research focus prediction index.

micro, and pico. All of them have different challenges, environmental issues, economic and technical lifespans, initial investment or capital requirements, operational and maintenance issues, and risks. The collected documents in this study cover all hydropower technology families and investments from mainly small to giant installed capacities (ESM.5–6).

GP2S, *GP2D*, *GP2E*, and *GP2O* have interlinked many system components for each power technology family. *GP2D* runs for features (factors, variables, etc.) and locations. It has many modules as unique system components. All selected features will help to research and develop the optimum 100% renewable ecological power grids (Saracoglu, 2017a, 2020a; Saracoglu et al., 2018). Investment features module and location features module serve for only feature analysis and its related topics to understand bankability, environmental sound friendliness, and financial analysis. All methods in the literature will be adopted and used in these modules. In that sense, the collected documents in this study cover investment and location features of hydropower technology families and investments.

There were two separate literature review periods related to this publication (ESM.7, Fig. 1). The first one had been a systematic one in the first RD3&D steps of the proposed robots and platforms (Li & Saracoglu, 2021; Saracoglu, 2017a, 2020a; Saracoglu et al., 2018). The second one was both unsystematic or lateral one and systematic one during the other studies and this study (King & Saracoglu, 2018; Ohunakin & Saracoglu, 2018; Saracoglu, 2014, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g, 2018a, 2018b, 2018c, 2018d, 2019a, 2019b, 2019c, 2019d; Saracoglu & de Simon Martin, 2018; Solangi et al., 2019; Tercan et al., 2021; Tercan, Saracoglu, Bilgilioglu, Eymen, & Tapkin, 2020).

In both literature review periods, all search key terms were in word n-grams (N-gram) forms (ESM.7–8). For instance, word 1-gram had only one word such as DEMATEL, and hydropower, word 2-grams had two words such as facility location and hydropower location; word 5-grams had five words such as Decision Making Trial and Evaluation Laboratory, and small hydropower location selection criteria. All key terms were manually searched with an unlimited truncation and full-text option on many scientific databases, and search engines with or without some tools (ESM.7, ESM.9). The hit documents were automatically and/or manually reviewed on their titles, abstracts, keywords, full text, and references.

The irrelevant and incomplete documents were eliminated and only the relevant ones were saved in their clustered folders. These publications were directly related to the current review topics (ESM.10). The other topics, indirectly related to the topics like text analysis, fuzzy systems, gray systems, or neural networks were not presented as sample datasets in this study, because of the scope, length, and type of this paper. The publications in "*location selection*", "*location selection factors*", and "*location alternatives generation*" topics were archived in the "*location folder*". Hence, they were in "*location cluster*" and "*location corpora*". The publications in "*investment selection*", "*investment selection factors*", and "*investment alternatives generation*" topics were archived in the "*investment cluster*". Hence, they were in "*investment cluster*" and "*investment corpora*". They were related to investment, bankability, and financial analysis subjects. The publications in "*Decision Making Trial and Evaluation Laboratory*" or "*DEMATEL*" topic were in "*DEMATEL folder*". Hence, they were in "*DEMATEL cluster*" and "*DEMATEL corpora*". Those specific folders were called "*corpora*" because they consisted of many authors' publications on the same topic.

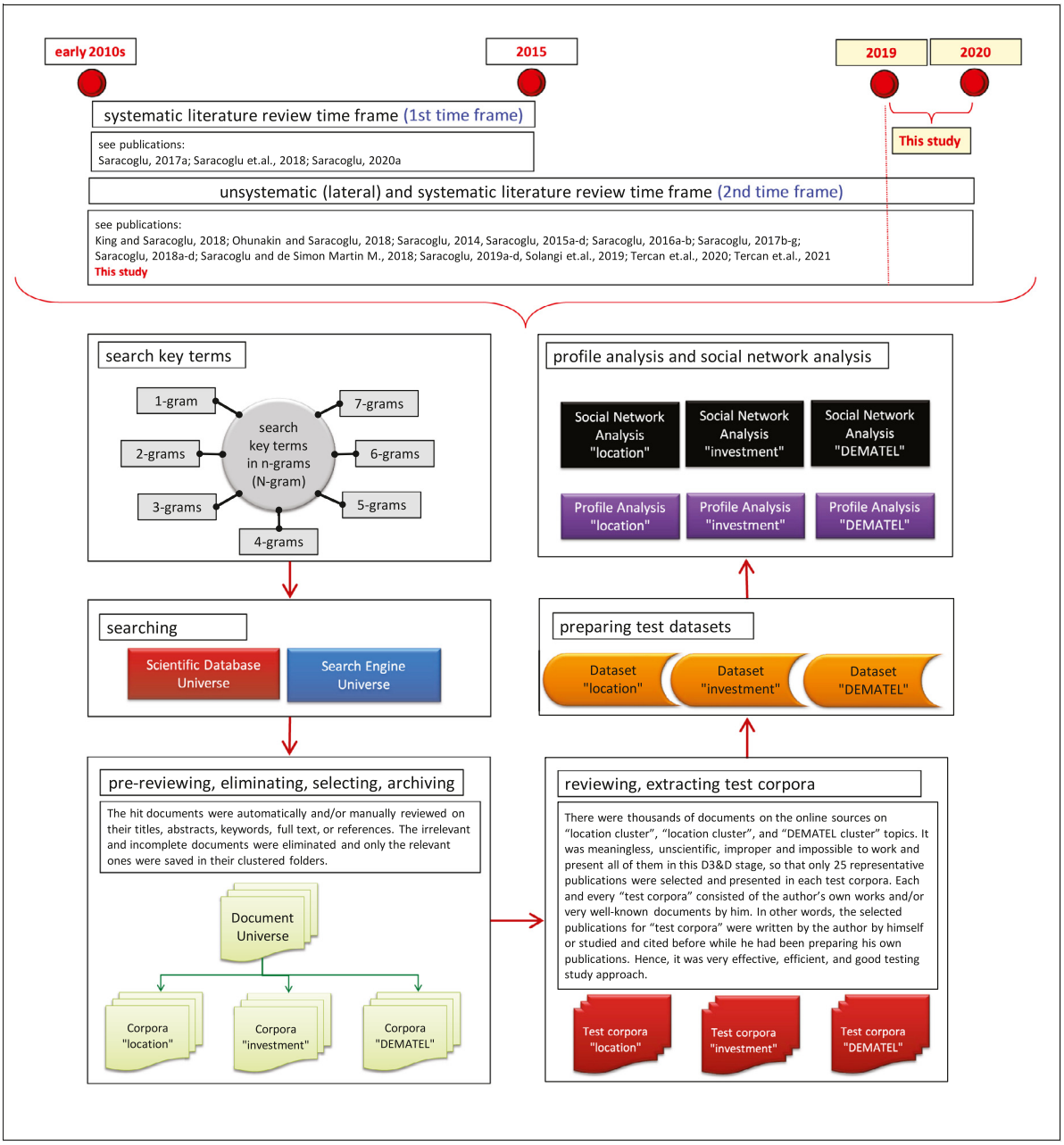


Fig. 1. The overview of this research study (visualization on spreadsheet software) (ESM.7).

The unique group of publications for each author was called a "*corpus*". This terminology was the preference in this paper. All publications in the same corpora were reviewed in detail like in the previous studies (Li & Saracoglu, 2021; Saracoglu, 2017a, 2020a; Saracoglu et al., 2018).

There were thousands of documents on the online sources for each topic. It was early, improper, impossible, meaningless, and unscientific to work and present all publications at once in this D3&D stage. Therefore, only twenty-five representative publications were selected and presented in each sample corpora (ESM.7, ESM.11–13). They were entitled as "*location sample corpora*", "*investment sample corpora*", and "*DEMATEL sample corpora*". The sample term was used as synonymous with "example". It also distinguished, discriminated, and differentiated from the terms "training" and "testing" in training and testing datasets of ML applications. Each "*sample corpora*" consisted of Burak Omer Saracoglu's (BOS) publications and very well-known documents by BOS. In other words, the selected publications for "*sample corpora*" were written by BOS or studied and cited before this study by BOS, while BOS had been preparing his publications. Thus, it was a very effective, efficient, and good sampling approach.

These three "*sample corpora*" were later used to prepare "*sample datasets*" entitled "*location sample dataset*", "*investment sample dataset*", and "*DEMATEL sample dataset*". Each "*sample dataset*" consisted of thirty-one different "*features*" (Table 1, ESM.11–13). Each "*feature*" could also be evaluated as a "*cluster*". Hence, there were thirty-one different clusters in each sample dataset. There was a maximum of seven authors in the publications of "*location sample corpora*", "*investment sample corpora*", and "*DEMATEL sample corpora*". For this reason, there were only seven authors as features in the "*sample datasets*". Similarly, there were only seven features for publication countries. The descriptive statistical analyses for the most important twenty-three features according to the preference of the author were prepared to better understand the literature (Table 2, ESM.14–16). For instance: "documents by the year", "documents by the first author", "documents by the uncertainty", and "documents by the publisher".

Table 2
Features (ESM.17-18).

No	Features	Description
1	Publication year	This feature presented the year of the publication. It was presented on the document by the publisher. It might be the first appearance of the online publication as the publication year, or it might be the first appearance of the paper-print as the publication year, or similar. It helped to understand the attentions, interests, improvements, and developments in the literature.
2	First author's full name-surname	This feature presented the full name and surname of the first author of the publication. The full name and surname were taken together to define and discriminate the information of the researchers or academics very well. It might be important to present, who had most contributed to the publication. However, this situation was not often the case in academic publications. For instance, there had been cases with the most contribution of the last author. In other words, any author in the authors' list had possibly contributed more than others. Moreover, it was possible that the real authors of the publications had not even been presented in the publications. There were also many dissertations, Master of Science (MSc.) thesis and Doctor of Philosophy (PhD.) thesis cases in some countries, where the authors had found some academics on some web pages and bargained with them to buy some dissertations and thesis for themselves. In other words, students did not prepare their dissertations and theses. They bought them and the academics accepted and approved them. It was an ordinary and daily thing for the academics in the universities and similar. A good example for those countries having that problem was Turkiye with huge criminal networks of universities, central governmental bodies, local governmental bodies, and private companies (ESM.18). The worst thing was that mentality had not been accepted or seen as a crucial problem in the universities, central governmental bodies, local governmental bodies, and private companies. It was a business as usual issue according to them. As a result, they were biased and unqualified. Turkiye was preferred in this text instead of Turkey considering the different usages like "turkey" and "Turkey". *Please search the news about that problem*. In short, the order of authors did not necessarily present the order of contributions. In other words, it was very difficult to find who contributed the most to the idea and publication without any interviewing activity and the analysis of the previous works. In any case, it helped to find the networks and perform profile analysis and social network analysis.
3	Second author full name-surname	This feature presented the full name and surname of the second author of the publication. See the feature "First author's full name-surname".
4	Third author full name-surname	This feature presented the full name and surname of the third author of the publication. See the feature "First author's full name-surname".
5	Fourth author full name-surname	This feature presented the full name and surname of the fourth author of the publication. See the feature "First author's full name-surname".
6	Fifth author full name-surname	This feature presented the full name and surname of the fifth author of the publication. See the feature "First author's full name-surname".
7	Sixth author full name-surname	This feature presented the full name and surname of the sixth author of the publication. See the feature "First author's full name-surname".
8	Seventh author full name-surname	This feature presented the full name and surname of the seventh author of the publication. See the feature "First author's full name-surname".
9	First author address as country	This feature presented the country of the first author of the publication. The country was taken to define and discriminate the work environment and living conditions of the researcher or academic. It might define how well the research environment and conditions had been in a country. However, this situation was not often the case in academic publications. For instance, there had been many cases related to being an immigrant, a visiting fellow, and similar. The origin and background of the author played an important role in these kinds of issues. In other words, the author might contribute to the publication in another country, and then moved to another country and finalized the publication. In other words, it was very difficult to find where the most contribution had been made without any interviewing activity and the analysis of the previous works. In any case, it helped to find the networks and perform profile analysis and social network analysis.

(*continued on next page*)

Table 2 (*continued*).

10	Second author address as country	This feature presented the country of the second author of the publication. See the feature "First author address as country".
11	Third author address as country	This feature presented the country of the third author of the publication. See the feature "First author address as country".
12	Fourth author address as country	This feature presented the country of the fourth author of the publication. See the feature "First author address as country".
13	Fifth author address as country	This feature presented the country of the fifth author of the publication. See the feature "First author address as country".
14	Sixth author address as country	This feature presented the country of the sixth author of the publication. See the feature "First author address as country".
15	Seventh author address as country	This feature presented the country of the seventh author of the publication. See the feature "First author address as country".
16	Research scope region	This feature presented the regions of the focus of the publication. The research scope region was found in the publication to define and discriminate the regions studied by the authors. Some studies focused on a region or some regions in a country. Some of them focused on a country or some countries. Some of them focused on the World. It would help to analyze the site expertise of the author. The author might improve his or her knowledge in a specific region or similar. This feature helped to find the networks and perform profile analysis and social network analysis.
17	Publication category (type)	This feature presented the category or type of publication. The publication category was found in the publication to define and discriminate the categories of the study. There were three publication categories or types in this study. These categories were theory, application, and review. This feature helped to find the networks and perform profile analysis and social network analysis.
18	Main topic	This feature presented whether the study was in "*location cluster*", "*investment cluster*", or "*DEMATEL cluster*". The main topic was found in the publication to define and discriminate the topics of the study. This feature helped to find the networks and perform profile analysis and social network analysis.
19	Industry	This feature presented the industries studied in the publication. The industries related to the publication were found in the publication to define and discriminate the industrial classification of the study. For instance, renewable energy, hydropower, real estate. This feature helped to find the networks and perform profile analysis and social network analysis.
20	GIS perspective	This feature presented the GIS design, usage, and perspective of the publication. The GIS information in the publication was found to define and discriminate the classification of the study. There were two categories of this feature in this study. These categories were GIS-based studies that had presented the GIS applications and none GIS-based studies that had not presented any GIS applications. This feature helped to find the networks and perform profile analysis and social network analysis.
21	Factors methods	The feature presented the methods used for factors and their analysis. There might be several methods used for those purposes such as Decision Making Trial And
		Evaluation Laboratory, Delphi method, expert opinion, Impact Matrix Cross-Reference Multiplication Applied to a Classification, impression article, in-depth interview, Interpretive Structural Modelling and questionnaire survey. All information about the methods used for factors and their analysis in the publications were found and listed to define and discriminate the study. This feature helped to find the networks and perform profile analysis and social network analysis.
22	Business type	This feature presented the relevancy of the publication according to the sectors. The sector information in the publication was found to define and discriminate the classification of the study. There were three categories in this study. These categories were private sector, public sector, and both private and public sector. This feature helped to find the networks and perform profile analysis and social network analysis.
23	Subindustry type	This feature presented the subdivision in an industry in the publication. For instance, mini-hydropower industry under hydropower industry, small hydropower industry under hydropower industry. This feature helped to find the networks and perform profile analysis and social network analysis.

(*continued on next page*)

Table 2 (*continued*).

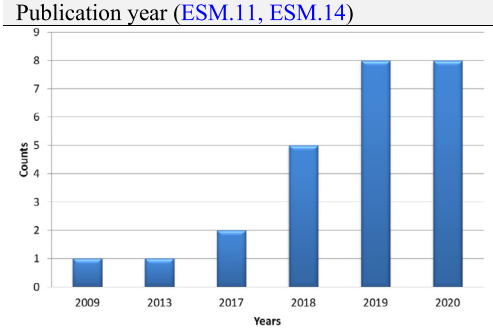
24	Applied methods	This feature presented the core methods in the publications. There might be several methods used in a study such as Multi-Objective Genetic Algorithm, Niched Sharing Genetic Algorithm/Non-dominated Sorting Genetic Algorithm, Niched Sharing Genetic Algorithm Version II, Type-1 Mamdani's Fuzzy Inference System, Analytic Network Process, Benjamin Franklin's Rule, System Advisor Model PVWatts V5, Preference Ranking Organisation Method for Enrichment Evaluations I, Preference Ranking Organisation Method for Enrichment Evaluations II, Preference Ranking Organisation Method for Enrichment Evaluations graphical analysis for interactive aid, Decision EXpert for Education. All information about the applied methods in the publications was found and listed to define and discriminate the study. This feature helped to find the networks and perform profile analysis and social network analysis.
25	Uncertainty	This feature presented the details of the uncertainty methods applied in the publications. Some studies had applied some uncertainty methods, but some had not applied any uncertainty methods. Accordingly, the applied uncertainty methods in the studies were found and presented such as crisp-deterministic without any uncertainty, possibilistic-type 1 fuzzy, possibilistic-type 2 fuzzy, and possibilistic-grey. All information about the uncertainty in the publications was found and listed to define and discriminate the study. This feature helped to find the networks and perform profile analysis and social network analysis.
26	Publisher	This feature presented the titles of the publishers. It was presented on the document by the publisher. It was observed that there had been mergers and acquisitions in the publishing industry like in other industries. For instance, it was very normal to get news that one of the hydropower plants in any project stage of a power company had been bought by another power company. There was nothing wrong, shocking, surprising, and unusual about it. This situation was very normal in the publishing industry too. Moreover, instead of closing a publisher with losing all data and information about its publications, it was wise to sell it to another publisher or investor, and would be the correct action for humankind. The title of the first and original publisher was taken into account in this feature. For instance, Elsevier, MedCrave Group, Springer. The publisher information in the publication was found to define and discriminate the classification of the study. This feature helped to find the networks and perform profile analysis and social network analysis.
27	Publication group	This feature presented the groups of the publication. There were three publication groups in this study. These groups were books, journals, and proceedings. This feature helped to find the networks and perform profile analysis and social network analysis.
28	Book category (type)	This feature presented the category or type of the book. There were two book categories or types in this study. These categories were electronic and print. This feature helped to find the networks and perform profile analysis and social network analysis.
29	Journal title	This feature presented the title of the journal if it was a journal publication. This feature helped to find the networks and perform profile analysis and social network analysis.
30	Proceedings title	This feature presented the title of the conference if it was a conference publication. This feature helped to find the networks and perform profile analysis and social network analysis.
31	Publication title	This feature presented the title of the publication. This feature helped to find the networks and perform profile analysis and social network analysis.

Note: features with descriptive statistical analyses in the orange background-color.

The features related to the author's full name and surname in these three "*sample datasets*" were then used to prepare the authors' network matrices (ESM.19, ESM.22, ESM.25). All seven authors' full name and surname features were used for the authors' network matrices. They were called "*location authors network matrix*", "*investment authors network matrix*", and "*DEMATEL authors network matrix*". The authors' network matrices presented whether there had been any relations of an author with another one. It also contained data and information about how many times an author had published studies with another. For instance, BOS had eleven publications in the "*location authors network matrix*". BOS had two publications with Olayinka S. Ohunakin in the "*location authors network matrix*". In other words, BOS and Olayinka S. Ohunakin were coauthors in two publications in the "*location sample corpora*". These authors' network matrices were then used for SNA.

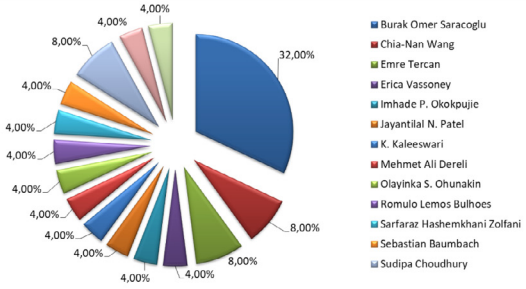
The social networks were plotted and SNA was performed on the Gephi 0.9.2 (https://gephi.org, https://github.com/gephi) (ESM.20–21, ESM.23–24, ESM.26–27). It was free open-source software. An author in a social network was defined as a node or vertex. A relation between authors in a social network was defined as an edge or link. The profile analyses and social network analyses helped to better understand the leading researchers, their research relations, their research focuses, and their research potentials or talents and capabilities. These analysis tools could also be very helpful and valuable for studies related to research publications and academia like Kiesslich, Beyreis,

Publication year (ESM.11, ESM.14)



visualization on spreadsheet software

The first authors' names-surnames (ESM.11, ESM.14)



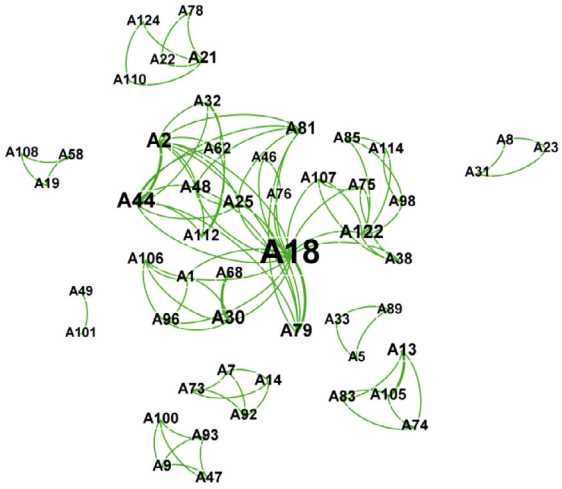
visualization on spreadsheet software

A small part of authors' adjacency matrix (ESM.19)

		A1	A2	A5	A7	A8	A9	A13	A14	A18	A19	A21	A22	A23	A25	A30	A31
		Abdurrahman Eymen	Aderemi A. Atayero	Alex Alisson Bandeira Santos	Ali Ebadi Torkayesh	Andrea Mammoliti Mochet	Andreas Dengel	Apu Kumar Saha	Arman Derakhti	Burak Omer Saracoglu	C. Vijayalakshmi	Chia-Nan Wang	Chih-Chiang Su	Claudio Comoglio	Damola S. Adelekan	Emre Tercan	Erica Vassoney
A1	Abdurrahman Eymen	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
A2	Aderemi A. Atayero	0	2	0	0	0	0	0	0	1	0	0	0	0	1	0	0
A5	Alex Alisson Bandeira Santos	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A7	Ali Ebadi Torkayesh	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
A8	Andrea Mammoliti Mochet	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1
A9	Andreas Dengel	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
A13	Apu Kumar Saha	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
A14	Arman Derakhti	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
A18	Burak Omer Saracoglu	1	1	0	0	0	0	0	0	11	0	0	0	0	1	1	0
A19	C. Vijayalakshmi	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
A21	Chia-Nan Wang	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0
A22	Chih-Chiang Su	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
A23	Claudio Comoglio	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1
A25	Damola S. Adelekan	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
A30	Emre Tercan	1	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0
A31	Erica Vassoney	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1

visualization on spreadsheet software

Authors' network visualization (ESM.20, files in LSGephi folder)



visualization with OpenOrd and Noverlap layout algorithms in the Gephi 0.9.2 (Jacomy, 2009; Bastian et al., 2009)

Fig. 2. The location sample corpora (full-scale image files in the ESM).

Zimmermann, and Traweger (2021), McMahan and McFarland (2021), and Savage and Olejniczak (2021).

The findings of this study were presented in a standardized manner for each "*sample dataset*" so that "*sample corpora*" and "*document cluster*" in this main text incorporating the ESM files. The paragraphs were structured in a standardized manner in the following sections. The publication years were presented as bar graphs. The first author's names and surnames were given in pie charts. All other features like the fourth author's names and surnames or the research scope region were given as bar plots and pie charts in the ESM (ESM.14–16). The authors' names and surnames, their alphanumeric codes, and their relations were given as the adjacency matrices and the network plots with different layout algorithms in the Gephi 0.9.2 201709241107 (Figs. 2–4). The whole adjacency matrices and the nodes and edges related to those adjacency matrices were only given in the ESM (ESM.20–21, ESM.23–24, ESM.26–27). The network statistics were calculated with different layout algorithms in the Gephi 0.9.2 (Tables 2–3, ESM.21, ESM.24, ESM.27). Some results of the clustering and SNA were presented in a comparative manner (Tables 2–3, ESM.21, ESM.24, ESM.27). The network analysis and its data files were submitted to the journal as ESM files in *.csv, *.xlsx, *.gephi, *.html, and *.png formats (see ESM files and ESM.28–29). Only two representative publications in each "*sample dataset*" were summarized according to the personal preference of BOS in this text.

5.1. Location sample corpora and profile analysis

The "*location sample corpora*" had documents related to the topics of "*location selection*", "*location selection factors*", and "*location alternatives generation*". There was a total of twenty-five publications in the sample corpora. Twenty-two of them were journal papers and three of them were conference proceedings (Ahmed, Tan, Solangi, & Ali, 2020; Baumbach, Khan, Ahmed, & Dengel, 2020; Bulhoes, de Santana, & Santos, 2020; Choudhury, Howladar, Majumder, & Saha, 2019; Choudhury & Saha, 2019; Dereli & Tercan, 2020; Kaleeswari, Johnson, & Vijayalakshmi, 2018; Ohunakin & Saracoglu, 2018; Okokpujie, Okonkwo, Akinlabi, Okokpujie, & Atayero, 2019; Patel & Rana, 2018; Saracoglu, 2013, 2017a, 2018c, 2019c, 2019d, 2020a; Saracoglu, Insel, & Helvacioglu, 2009; Saracoglu et al., 2018; Solangi et al., 2019; Tercan & Dereli, 2020; Tercan et al., 2020; Vassoney, Mochet, & Comoglio, 2017; Wang, Thanh, & Su, 2019a; Wang, Tsai, & Huang, 2019b; Zolfani, Yazdani, Torkayesh, & Derakhti, 2020) (Fig. 2, ESM.10–11, ESM.14, ESM.19–21, LSGephi.gephi, and LSGephi folder).

The summary of findings related to the "*location sample corpora*" was as follows. One of the publications in this cluster was published in 2009, two of them in 2013, two of them in 2017, five of them in 2018, eight of them in 2019, and eight of them in 2020. Hence, the "*location sample dataset*" was an up-to-date dataset. There were fifteen different first authors. The maximum and the second maximum number of publications were respectively eight by BOS, and two by Chia-Nan Wang, Emre Tercan, and Sudipa Choudhury. There were eleven different countries of the first author. The maximum and the second maximum number of publications were respectively eleven in Turkiye, and four in India. There were eight different research scope regions. The maximum and the second maximum number of publications were respectively eight on the global scale (the whole World), and four in India (the country political boundary). There were two different publication categories.

Twenty-four publications were in the application publication category and one publication was in the review publication category. Twenty-one publications were in the location selection main topic cluster, four publications were in the location selection factors main topic cluster, and two publications were in the location alternatives generation main topic cluster. There were thirteen different industries. The maximum and the second maximum number of publications were respectively eight in solar power, and three in hydropower. According to the uncertainty classification, there were four possibilistic publications, and fourteen deterministic (crisp) publications. One publication was a review paper. There was not any uncertainty method in its design and presentation. The possibilistic publications were in different methods. Type-1 fuzzy models were presented in nine publications, gray systems were applied in four publications, neural networks were used in two publications, and type-2 fuzzy models were presented in one publication. There were thirteen different publishers. Multidisciplinary Digital Publishing Institute (MDPI) had five publications. Springer also had five publications. They had the maximum number of publications. Elsevier had four publications. It was the second on the list after MDPI and Springer.

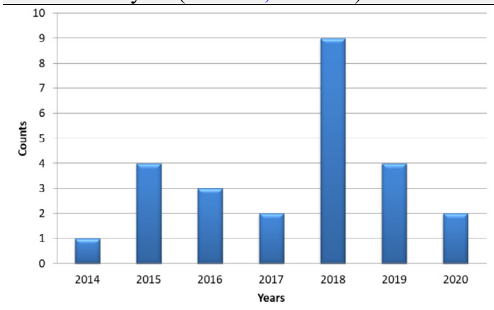
The researchers' network and profile analyses showed that BOS was the most active, dominant, and influencing researcher. BOS had eight publications as the first author, two publications as the second author, and one publication as the fifth author. As a result, BOS had a total of eleven publications. BOS had the oldest publication. It was in 2009. BOS worked with researchers in five different countries. Those counties were China, India, Nigeria, Turkiye, and USA. That researcher focused on two main industries as such shipbuilding from 2009 to 2013, and power from 2017–2020. BOS studied in several subindustries such as concentrated solar power, concentrated photovoltaic solar power, photovoltaic solar power, small hydropower, and new shipbuilding. That researcher studied a wide range of regions. Those regions were World (global), Nigeria, Pakistan, and Turkiye. BOS researched "location selection", "location selection factors", and "location alternatives generation" topics. That researcher had both GIS-based and not GIS-based studies. BOS applied many methods and techniques in location factors topic such as DEMATEL, Delphi method, expert opinion, in-depth interview, impression article, Impact Matrix Cross-Reference Multiplication Applied to a Classification, Interpretive Structural Modeling, Search Results Clustering Engine (Carrot^2) Lingo/Suffix Tree Clustering/K-means, Simple Additive Weighting, text analysis, Weighted Product Method; Political, Economic, Social and Technological, and questionnaire survey. That researcher used different methods for location selection models like fuzzy ordered weighted average, fuzzy weighted average, Analytical Hierarchy Process, Elimination, and Choice Translating Reality III-IV, Decision Expert for Education, and Consistency Driven Pairwise Comparisons. Moreover, BOS used not only deterministic (crisp), but also possibilistic approaches like gray, type 1 fuzzy, and type 2 fuzzy systems. Finally, that researcher's studies were accepted by some different publishers' editors and reviewers like DergiPark, Elsevier, Inderscience, SNAME, Springer, and Taylor & Francis. It was deducted and thought that BOS would continue and deepen in those research topics in the following years.

The first study summarized in this subsection was by Tercan et al. (2020). They presented an experimental GIS-based photovoltaic power plant location alternatives generation study in Antalya, Burdur, Isparta provinces in Turkiye. There were many sites in different classes in that study. It was a small D3&D activity of a large robot system. They studied five weighting methods. Those weighting methods were equal weight method, rank-sum weight method, inverse or reciprocal weights method, rank order centroid, point allocation. Those methods were grouped into two different theoretical classes. They were objective and subjective. The equal weight method was in the objective weighting methods group. Rank sum weight method, inverse or reciprocal weights method, rank order centroid, point allocation methods were in the subjective weighting methods group. Furthermore, those subjective weighting methods were discriminated against into two different theoretical classes again. Those classes were ranking methods and direct weighting methods. Rank sum weight method, inverse or reciprocal weights method, and rank order centroid were in the ranking methods group. The point allocation method was in the direct weighting method. The authors also researched two different group decision-making approximation methods. Those methods were geometric mean and arithmetic mean. They applied the weighted linear combination technique as a multi-criteria decision making or multiple-criteria decision analysis method or decision rule on the raster data. Those data were the values of map cells of raster map layers. There were five main criteria in their model. Those main criteria were electricity generation resource features, essential features, obligatory features, infrastructural and complementary features, and logistics features. There were fourteen sub-criteria like Global Horizontal Irradiation, aspect, slope, elevation, and land use, and seventy-nine limitations. The main factors were mental and natural language clusters of sub-factors so that they were some relative factors. In other words, they were aggregation factors. The sub-factors were essential or principal factors. The authors excluded some land cover classes before their evaluation. Their publication was one of the promising ones for web and smart tools and devices in the future.

The second study was by Zolfani et al. (2020). They aimed to find the best temporary hospital locations in Istanbul, Turkiye during the Coronavirus disease outbreak (2019-nCoV). They proposed a decision-support framework with gray-based criteria importance through the inter-criteria correlation and combined compromise solution methods. The gray term and the gray term were synonymous. They studied five candidate locations. Those locations were Beykoz, Bakirkoy, Buyukcekmece, Eyup, and Pendik. They had ten criteria in their model. Those criteria were traffic congestion, accessibility via roads and airports, health centers in the district, distance from populated residential areas, land price, transportation cost, future expansion potential, distance from industrial areas, and local regulation. There were ten experts in their application. Eight of them were male and two of them were female. One of them had a Doctorate of Medicine degree, four of them had Master of Science degrees, and five of them had Doctor of Philosophy degrees. Those degrees were in the fields of medical doctor, civil engineer, industrial engineer, data scientist, and architecture. Although their method was not so easy to understand and apply without any difficulty, their publication's core idea was one of the promising ones for future pandemics. Their approach could be transformed into a web-based platform enriching with many different methods to use in all international, national, and regional governmental organizations. An international tool would help to prevent pandemics like Zika, Ebola, Sudan virus, Reston virus, or severe acute respiratory syndrome and protect people against them. Of course, that sort of tool would be supported if and only if governments and governmental staff would not want to kill ordinary people like serial killers, and mass murderers. In addition, that sort of emergency tool would be supported if and only if governmental bodies wanted to do what they had been responsible for, and paid for. 2019-nCoV was one of the good examples of all wrongdoings in governmental bodies. It was possibly a criminal case in which many ordinary people were murdered slowly. Moreover, almost all 2019-nCoV positive or infected cases would have some health problems with their blood vessels, hearts, kidneys, livers, lungs, or minds in the following years. They would possibly live in poor or miserable health and life conditions and die slowly in pain day by day. Nothing would be the same as before 2019-nCoV.

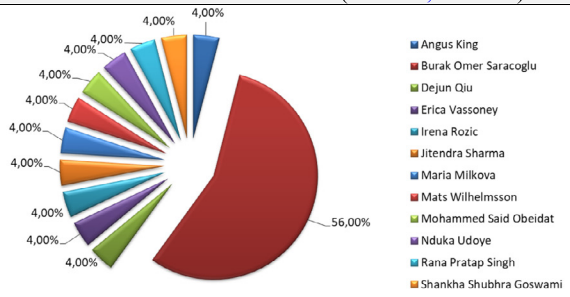
As a result, the problems in the "*location sample corpora*" were one of the basic decision-making problem groups. Many researchers tried to handle many real-world problems in many fields like finding some protected areas or some spacecraft landing sites. Moreover, some researchers also focused on these topics in the renewable power industry like solar power. The profile analyses and social network analyses would help to find experts, their expertise, and their professional relations.

Publication year (ESM.11, ESM.15)



visualization on spreadsheet software

The first authors' names-surnames (ESM.11, ESM.15)



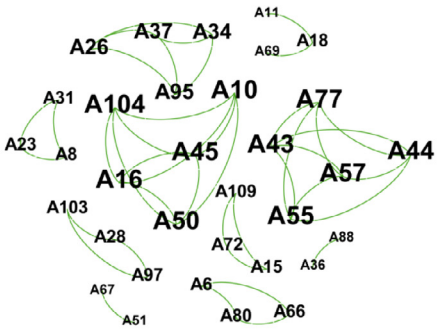
visualization on spreadsheet software

A small part of authors' adjacency matrix (ESM.22)

		A6	A8	A10	A11	A15	A16	A18	A23	A26	A28	A31	A34	A36	A37	A43	A44	A45
		Alexander Andreichikov	Andrea Mammoliti Mochet	Andrej Ferdo Gubina	Angus King	Aseel Khanfar	Besim Imamovic	Burak Omer Saracoglu	Claudio Comoglio	Dejun Qiu	Dhiren Kumar Behera	Erica Vassoney	Gozde Gulseven Ubay	Hans Peter Nachtnebel	Hasan Dincer	Igbinosa Ikpotokin	Imhade P. Okokpujie	Irena Rozic
A6	Alexander Andreichikov	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	Andrea Mammoliti Mochet	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
A10	Andrej Ferdo Gubina	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1
A11	Angus King	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
A15	Aseel Khanfar	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
A16	Besim Imamovic	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1
A18	Burak Omer Saracoglu	0	0	0	1	0	0	15	0	0	0	0	0	0	0	0	0	0
A23	Claudio Comoglio	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
A26	Dejun Qiu	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
A28	Dhiren Kumar Behera	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
A31	Erica Vassoney	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
A34	Gozde Gulseven Ubay	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
A36	Hans Peter Nachtnebel	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A37	Hasan Dincer	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
A43	Igbinosa Ikpotokin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
A44	Imhade P. Okokpujie	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
A45	Irena Rozic	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1

visualization on spreadsheet software

Authors' network visualization (ESM.22, files in ISGephi folder)



visualization with OpenOrd and Noverlap layout algorithms in the Gephi 0.9.2 (Jacomy, 2009; Bastian et al., 2009)

Fig. 3. The investment sample corpora (full-scale image files in the ESM).

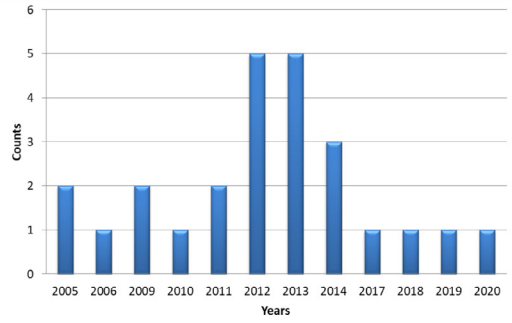
5.2. Investment sample corpora and profile analysis

The "*investment sample corpora*" had documents related to the topics of "*investment selection*", "*investment selection factors*", and "*investment alternatives generation*". There was a total of twenty-five publications in the sample corpora. Twenty of them were journal papers, three of them were conference proceedings, and two of them were digital books (Goswami, Behera, & Mitra, 2020; King & Saracoglu, 2018; Milkova, Andreichikova, & Andreichikov, 2018; Obeidat, Qasim, & Khanfar, 2018; Qiu, Dincer, Yuksel, & Ubay, 2020; Rozic, Imamovic, Pavlicevic, Gubina, & Halilcevic, 2018; Saracoglu, 2014, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b, 2017f, 2018a, 2018d, 2019a, 2019b, 2019e; Saracoglu & de Simon Martin, 2018; Sharma, 2018; Singh & Nachtnebel, 2016; Udoye, Okokpujie, Okeniyi, Dirisu, & Ikpotokin, 2019; Vassoney et al., 2017; Wilhelmsson & Zhao, 2018) (Fig. 3, ESM.10, ESM.12, ESM.15, ESM.22–24, ISGephi.gephi, and ISGephi folder).

The summary of findings related to the "*investment sample corpora*" was as follows. One of the publications in this cluster was published in 2014, four of them in 2015, three of them in 2016, two of them in 2017, nine of them in 2018, four of them in 2019, and two of them in 2020. Hence, the "*investment sample dataset*" was an up-to-date dataset. There were twelve different first authors. The maximum and the second maximum number of publications were respectively fourteen by BOS, and one by all others. There were eleven different countries of the first author. The maximum and the second maximum number of publications were respectively fourteen in Turkiye, and two in India. There were twelve different research scope regions. The maximum and the second maximum number of publications were respectively ten in Turkiye, and three on the global scale (the whole World). There were two different publication categories. Twenty-three of them were in the application publication category and two of them were review publication category. Twenty publications were in the investment selection main topic cluster, and six publications were in the investment alternatives generation main topic cluster. There were nine different industries. The maximum and the second maximum number of publications were respectively twelve in the hydropower industry and three in the solar power industry. According to the uncertainty classification, there were six possibilistic publications and nineteen deterministic (crisp) publications. Two publications were review papers. There were not any uncertainty methods in their design and presentation. The possibilistic publications were in different methods. Type-1 fuzzy models were presented in two publications, type-2 fuzzy models were applied in two publications, and stochastic models were used in two publications. There were eighteen different publishers. Creative Decisions Foundation had three publications. It had the maximum number of publications. DergiPark, Inderscience, MDPI, MedCrave Group, and Springer had two publications. They were the second on the list after Creative Decisions Foundation.

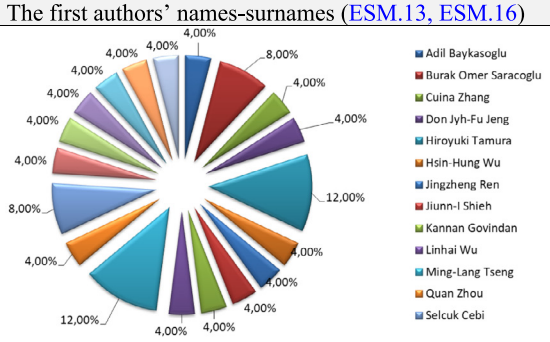
The researchers' network and profile analyses showed that BOS was the most dominant and influencing researcher. That researcher had fourteen publications as the first author and two publications as the second author. As a result, BOS had a total of ten publications. That researcher had the oldest publication. It was in 2014. BOS worked with researchers in two different countries. Those counties were Australia and Spain. That researcher focused on one main industry as the power industry from 2014–2019. BOS studied in several subindustries such as mini-hydropower, small hydropower, photovoltaic solar power, and concentrated solar power. That researcher studied a wide range of regions. Those regions were World (global), Australia, USA, and Turkiye. BOS researched "investment selection" and "investment alternatives

Publication year (ESM.13, ESM.16)



visualization on spreadsheet software

The first authors' names-surnames (ESM.13, ESM.16)



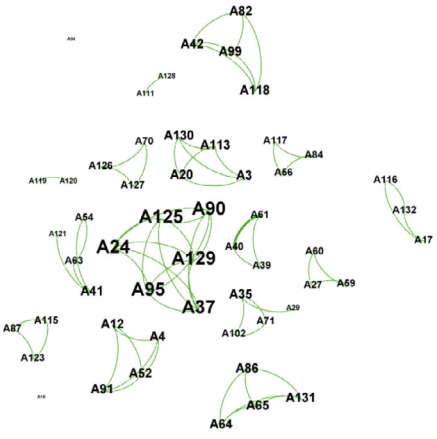
visualization on spreadsheet software

A small part of authors' adjacency matrix (ESM.25)

		A3	A4	A12	A17	A18	A20	A24	A27	A29	A35	A37	A39	A40	A41	A42	A52	A54
		Adil Baykasoglu	Alessandro Manzardo	Antonio Scipioni	Bo Feng	Burak Omer Saracoglu	Cenk Sahin	Cuina Zhang	Devika Kannan	Don Jyh-Fu Jeng	Gwo-Hshiung Tzeng	Hasan Dincer	Hiroki Okanishi	Hiroyuki Tamura	Hsin-Hung Wu	Hu-Chen Liu	Jingzheng Ren	Jiunn-I Shieh
A3	Adil Baykasoglu	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A4	Alessandro Manzardo	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
A12	Antonio Scipioni	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
A17	Bo Feng	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A18	Burak Omer Saracoglu	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
A20	Cenk Sahin	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A24	Cuina Zhang	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
A27	Devika Kannan	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
A29	Don Jyh-Fu Jeng	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
A35	Gwo-Hshiung Tzeng	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0
A37	Hasan Dincer	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
A39	Hiroki Okanishi	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
A40	Hiroyuki Tamura	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0
A41	Hsin-Hung Wu	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1
A42	Hu-Chen Liu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
A52	Jingzheng Ren	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
A54	Jiunn-I Shieh	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1

visualization on spreadsheet software

Authors' network visualization (ESM.25, files in DGephi folder)



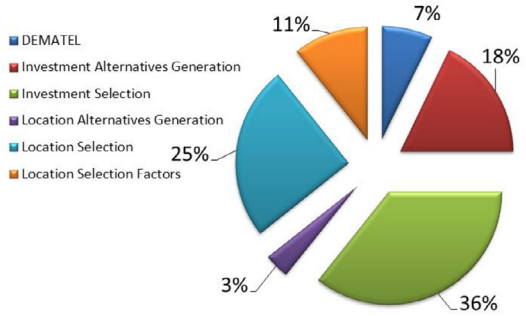
visualization with OpenOrd and Noverlap layout algorithms in the Gephi 0.9.2 (Jacomy, 2009; Bastian et al., 2009)

Fig. 4. The DEMATEL sample corpora (full-scale image files in the ESM).

generation" topics. That researcher did not have any direct study in "investment selection factors". BOS had both GIS-based and not GIS-based studies. That researcher applied many methods and techniques in investment factors of investment selection problems such as expert opinion, Search Results Clustering Engine (Carrot^2) Lingo/Suffix Tree Clustering/K-means, and literature review. However, BOS never focused and researched directly on only "investment factors". That researcher only used those methods for indirect purposes. In other words, those methods were used indirectly in "investment selection", and "investment alternatives generation" topics. BOS used many methods for investment selection models like Mamdani's fuzzy inference system, Analytical Hierarchy Process, Analytic Network Process, Benjamin Franklin's Rule, Elimination and Choice Translating Reality III-IV, Decision Expert for Education, Preference Ranking Organisation Method for Enrichment Evaluations I–II-Graphical Analysis for Interactive Aid, and Shannon's entropy weighting. Moreover, that researcher used not only deterministic (crisp), but also possibilistic approaches like type 1, and type 2 fuzzy systems in that researcher's investment selection models. BOS used many methods for "investment alternatives generation" models like Multi-Objective Genetic Algorithm, Niched Sharing Genetic Algorithm/Non-dominated Sorting Genetic Algorithm, Niched Sharing Genetic Algorithm Version II, System Advisor Model PVWatts V5, and System Advisor Model Empirical Trough Model. Moreover, that researcher used not only deterministic (crisp) but also probabilistic approaches like stochastic systems in his "investment alternatives generation" models. Finally, that researcher's studies were accepted by some different publishers' editors and reviewers like Creative Decisions Foundation, DergiPark, Grand Journals, Hindawi Publishing Corporation, MedCrave Group, OmniaScience, and Springer. It was deducted and thought that BOS would continue and deepen in these research topics in the following years.

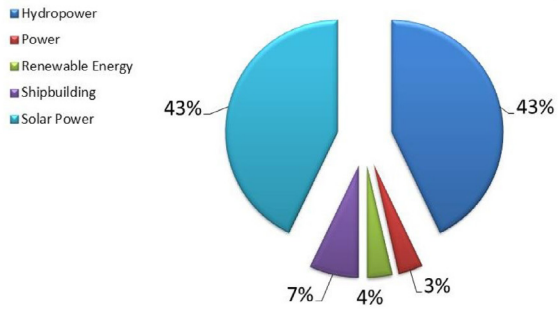
The first study summarized in this subsection was by Singh and Nachtnebel (2016). They developed a decision framework for prioritizing small hydropower plant investments in Nepal. Small hydropower plants in Nepal had been classified as hydropower plants with the installed power between 1 MW to 25 MW. They underlined that some detailed analyses such as social impact assessment, cumulative impact assessment, or environmental impact assessment had been requested in Nepal. Their Preference Ranking Organisation Method for Enrichment Evaluations Graphical Analysis for Interactive Aid model would support those small hydropower plants studies. They presented the main factors in their small hydropower plant investment selection approach as goals or objectives. Those goals were economic, social, environmental, political, and uncertainty. Those goals had their sub-goals. For instance, power generation capacity (maximization), cost of investment (minimization) in economic objective. Eighty-five experts from different backgrounds like technical professionals, sociologists, finance professionals, environmentalists attended their case study. Their response rate was 0,77. In another expression, it was 85/110. They evaluated five small hydropower schemes. They ranked Modi hp developed by the private Modi as the first, and NEA as the second. Their publication was one of the promising ones for web and smart tools in the future. However, there were some critical issues to be studied and explained in their model. For instance, "power generation capacity of the plant" aimed to be maximized in their model. This was not a well-defined objective, because selecting the largest capacity did not necessarily mean selecting the highest power generation. More specifically, as an example, there might be a 1 MW small hydropower plant with

RPRFI in research topic, ex-post analysis (ESM.30)



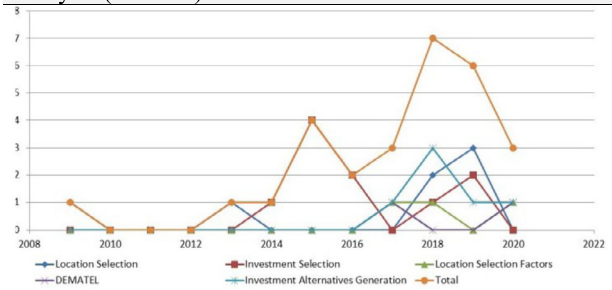
visualization on spreadsheet software

RPRFI in industry, ex-post analysis (ESM.30)



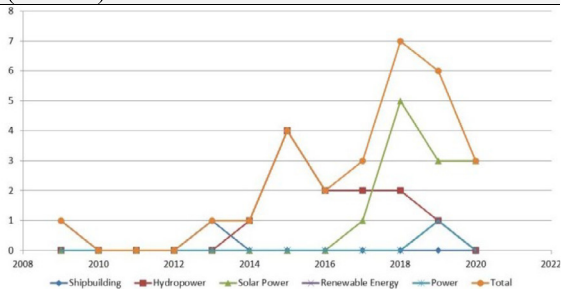
visualization on spreadsheet software

Timeline of publications in research topic, ex-post analysis (ESM.30)



visualization on spreadsheet software

Timeline of publications in industry, ex-post analysis (ESM.30)



visualization on spreadsheet software

RPRFI in research topic, ex-ante analysis (ESM.30)		RPRFI in industry, ex-ante analysis (ESM.30)	
Investment Alternatives Generation	(+) 0,0057 y = 0,0057x - 11,554	Solar Power	(+) 0,0125 y = 0,0125x - 25,12
Investment Selection	(+) 0,0042 y = 0,0042x - 8,5233	Hydropower	(+) 0,0052 y = 0,0052x - 10,53
Location Selection	(+) 0,0034 y = 0,0034x - 6,7713	Renewable Energy	(+) 0,0011 y = 0,0011x - 2,2611
Location Selection Factors	(+) 0,0029 y = 0,0029x - 5,777	Power	(+) 0,0011 y = 0,0011x - 2,2611
DEMATEL	(+)0,002 y = 0,002x - 4,019	Shipbuilding	(-) -0,0017 y = -0,0017x + 3,5278

Fig. 5. The summary of profile analysis (full-scale image files in the ESM).

a capacity factor 0,95. It would generate 1 MW × (24 × 365 + 6) total hours/year × 0,95 = 8.327,7 MW-h/year. At the same time, there might be a 3,8 MW small hydropower plant with a capacity factor 0,25. It would generate 3,8 MW × (24 × 365 + 6) total hours/year × 0,25 = 8.327,7 MW-h/year. Furthermore, when their initial costs were compared similarly, 1 MW small hydropower plant might be the least costly one. Accordingly, those factors needed some further investigation.

The second study was by Wilhelmsson and Zhao (2018). They aimed to analyze the risks of housing or real estate investments from lenders' point of view. Those risks were credit risks, and market risks. The credit risks consisted of debt, mortgage, and loan-to-value ratio risks. The market risks consisted of systematic and unsystematic equity price risks. They presented a feasible portfolio in the different market segments in Stockholm, Sweden. They developed four different models. Those models were entitled hedonic pricing, beta (β), Analytical Hierarchy Process, and Value-at-Risk. They used several features in their models. A few of them were price, living area, rooms, age, fee, top floor, and ground floor. Their core publication idea was very good, however, their publication presentation was not good enough to be understood and applied with ease. Their paper needed more explanation. Their approach could be transformed into a web-based platform enriching with many different methods.

As a result, the problems in the "*investment sample corpora*" were one of the basic decision-making problems, but "investment factors" were not studied well in the literature. Many researchers tried to handle many real-world problems in many fields like selecting the best house private investment or developing small agricultural farms. Moreover, some researchers also focused on "investment selection" and "investment alternatives generation" topics in the renewable power industry like solar power. None of them focused directly on "investment factors" and hydropower plants' investment factors.

5.3. DEMATEL sample corpora and profile analysis

The "*DEMATEL sample corpora*" had documents related to the DEMATEL methodology. There was a total of twenty-five publications in the sample corpora. They were all journal papers (Baykasoglu, Kaplanoglu, Durmusoglu, & Sahin, 2013; Cebi, 2013a, 2013b; Fan, Suo, & Feng, 2012; Fu, Zhu, & Sarkis, 2012; Geng & Chu, 2012; Govindan, Kannan, & Shankar, 2014; Hu, Lu, & Tzeng, 2014; Jeng & Tzeng, 2012; Ren, Manzardo, Toniolo, & Scipioni, 2013; Saracoglu, 2017a, 2020a; Shieh, Wu, & Huang, 2010; Si, You, Liu, & Zhang, 2018; Tamura & Akazawa, 2005a, 2005b; Tamura, Okanishi, & Akazawa, 2006; Tseng, 2009a, 2009b; Tseng, Chen, & Geng, 2012; Wu & Tsai, 2011; Wu, Zhang, Shan, & Chen, 2013; Yeh & Huang, 2014; Zhang, Li, Xia, Yuan, Dincer et al., 2020; Zhou, Huang, & Zhang, 2011) (Fig. 4, ESM.10, ESM.13, ESM.16, ESM.25–27, DGephi.gephi, and DGephi folder).

The summary of findings related to the "*DEMATEL sample corpora*" was as follows. Two of the publications in this cluster were published in

2005, one of them in 2006, two of them in 2009, one of them in 2010. two of them in 2011, five of them in 2012, five of them in 2013, three of them in 2014, one of them in 2017, one of them in 2018, one of them in 2019, and one of them in 2020. Hence, the "*DEMATEL sample dataset*" was an up-to-date dataset. There were nineteen different first authors. The maximum and the second maximum number of publications were respectively three by Hiroyuki Tamura and Ming-Lang Tseng, and two by BOS and Selcuk Cebi. There were six different countries of the first author. The maximum and the second maximum number of publications were respectively eight in China and Taiwan, and five in Turkiye. There were six different research scope regions. The maximum and the second maximum number of publications were respectively eight in China and Taiwan, and three on the global scale (the whole World), Japan, and Turkiye. There were two different publication categories. Twenty-four publications were in the application publication category and one of them was in the review publication category. The original DEMATEL theory papers were not included in the "*DEMATEL sample corpora*". Twenty-five publications were in the DEMATEL main topic cluster. There were fifteen different industries. The maximum and the second maximum number of publications were respectively four in information technology, and two in food, health, telecommunication, and tourism. According to the uncertainty classification, there were twenty-nine possibilistic publications and six deterministic (crisp) publications. One publication was a review paper. There was not any uncertainty method in its design and presentation. The possibilistic publications were in different methods. Type-1 fuzzy models were presented in thirteen publications, gray systems were applied in five publications, stochastic models were used in two publications, type-2 fuzzy models were presented in one publication, and vague sets were used in one publication. There were five different publishers. Elsevier had nineteen publications. It had the maximum number of publications. NIT Laboratories Polish Center for Accreditation had three publications. It had the second maximum number of publications.

The researchers' network and profile analyses showed that Hiroyuki Tamura and Ming-Lang Tseng were the most dominant and influencing researchers. Hiroyuki Tamura and Ming-Lang Tseng had three publications as the first author. Each one had three publications. Hiroyuki Tamura had the oldest publications. It was in 2005. That author worked with the researchers in only one country. It was Japan. That author did not have any GIS studies. That author used many methods in their studies like expectation principle, maximum–minimum principle, maximum–maximum principle, Monte Carlo method, and questionnaire survey. Moreover, the author used not only deterministic (crisp) but also possibilistic approaches. That author's possibilistic studies were only stochastic. Finally, that author's studies were accepted by only one publisher's editors and reviewers. It was NIT Laboratories Polish Center for Accreditation. It was deducted and thought that author would continue and deepen in these research topics in the following years. Ming-Lang Tseng's first publications were in 2009. That author worked with some researchers in two different countries. They were in China and Taiwan. That author focused on two main industries as real estate, and tourism from 2009–2012. That author studied in three different subindustries. They were hot spring tourism, English hospitality, and real estate agent. He studied in only two different regions. They were China and Taiwan. That author did not have any GIS studies. That author used expert opinion, questionnaire survey, and Cronbach's coefficient alpha methods. Moreover, he used only possibilistic approaches. They were type 1 fuzzy, and gray models. Finally, that author's studies were accepted by only one publisher's editors and reviewers. It was Elsevier. It was deducted and thought that he would continue and deepen in those research topics in the following years.

The first study summarized in this subsection was by Si et al. (2018). They presented an informative DEMATEL review paper with 346 international journal papers from 2006 to 2016. They grouped their collection into five categories. 30,3% of them were classical, 18,2% of them were fuzzy, 3,5% of them were gray, 44,5% of them were Analytic Network Process, and 3,5% of them were in other groups. They explained the DEMATEL method well enough. They also shared their observations in each group. These observations were important and showed some critical subjectivity and variability issues in the DEMATEL method. Accordingly, it was obvious that many researchers might get many different findings in the same problem. They all depended on their evaluations and decisions. For instance, researchers could come up with different Influential Relation Map threshold values. Their publication was one of the important ones for the foundation of other review papers on DEMATEL subject. However, a critical problematic part of their review was their grouping. The classical, fuzzy, and gray classification was related to uncertainty approaches; but Analytic Network Process was a Multi-Criteria Decision Making, Multiple Criteria Decision-Making, Multiple–Criteria Decision Analysis, or Multiple Criteria Decision Aiding method. That method could not be evaluated in the same classifier perspective with the uncertainty approaches. This grouping was a poor classifier approach in this manner.

The second study was by Yeh and Huang (2014). They aimed to determine the location selection factors of wind farms. They tried to define factors with the Goal/Question/Metric method. They tried to find the factors' relations with their words' correlations and DEMATEL. Finally, they ranked factors with Analytic Network Process. They evaluated six dimensions. They were safety and quality, economy and benefit, social impression, environment and ecology, regulation, and policy. They came up with twenty-eight evaluation criteria. For example, secure setup distance, regular wind farm testing, land costs, expenses for construction, visual coordination, noise and sharp light, and administrative compatibility. They conducted a questionnaire survey with some energy industry experts in electrical, mechanical, and energy areas and some professors in the academy. They found the most important factors as safety and quality, and environment and ecology. Their publication was one of the promising ones for web and smart tools in the future, but their study required some further explanation with their Goal/Question/Metric application.

As a result, the DEMATEL was a mature method. It was studied and used a lot in the literature. Some researchers tried to handle many real-world problems in many fields like understanding the relations in the health industry or tourism sector. Moreover, it was wise to use the DEMATEL in every problem to understand the problem and its factors very well.

5.4. Social network analysis

SNA for each sample dataset was performed on the Gephi 0.9.2 201709241107. That analysis was like the other ones in that research field (Boguna, Pastor-Satorras, Diaz-Guilera, & Arenas, 2004; Golbeck, 2013; Grandjean & Jacomy, 2019; Perer & Shneiderman, 2006, 2008a, 2008b; Tournay, Jacomy, Necula, Leibing, & Blasimme, 2019; Venturini, Jacomy, Bounegru, & J., 2018; Welser, Gleave, Fisher, & Smith, 2007) (Table 3, ESM.20–21, ESM.20–21, ESM.23–24, ESM.26–27, LS-Gephi.gephi, ISGephi.gephi, DGephi.gephi, and Gephi folder). All layout and statistics algorithms in this study were built-in algorithms on the Gephi 0.9.2. They were run by the default plugins on the Gephi 0.9.2 201709241107. The applied layout algorithms in this study were Force Atlas, ForceAtlas2, Fruchterman–Reingold, OpenOrd, Yifan Hu, Yifan Hu Proportional, and Noverlap (Fruchterman & Reingold, 1991; Gephi github, 2021; Hu, 2005; Jacomy, Venturini, Heymann, & Bastian, 2014; Martin, Brown, Klavans, & Boyac, 2011). The applied statistics algorithms and metrics in this study were for the average degree, average weighted degree, network diameter, graph density, Hyperlink-Induced Topic Search (HITS), modularity, page rank, connected components in the network overview; average clustering coefficient, eigenvector centrality in the node overview; and average path length in the edge overview. They corresponded to the algorithms and metrics of degree in average degree; degree in average weighted degree; betweenness centrality, closeness centrality, harmonic closeness centrality, eccentricity

in network diameter; density in graph density; hubs and authority in HITS; size in modularity; pagerank in pagerank; size in connected components; clustering coefficient in average clustering coefficient; eigenvector centrality in eigenvector centrality; betweenness centrality, closeness centrality, harmonic closeness centrality, eccentricity in average path length (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Brandes, 2001; Brin & Page, 1998; Gephi github, 2021; Kleinberg, 1999; Lambiotte, Delvenne, & Barahona, 2009; Latapy, 2008; Tarjan, 1972; Verma & Goyal, 2021).

The authors' adjacency matrices were manually prepared in *.xlsx format (ESM.19, ESM.22, ESM.25). The diagonal elements of authors' adjacency matrices were not zero unlike in many other studies in the literature, because the author's publication without any co-authorship could be analyzed within this approach. In other words, any researcher's publication alone could easily be found by that available data and information. The diagonal elements showed the number of publications with only one author. For instance, if the diagonal element of any researcher was one, it would show that there was only one publication of that author in the sample dataset. Moreover, it indicated that the researcher was alone. If the diagonal element of any researcher was nine, it would show that there were nine publications of that author in the sample dataset. Moreover, it indicated that the researcher was alone. Hence, self-loops were allowed in the visualizations. The nodes and edges sheets were manually prepared in *.xlsx format based on the authors' adjacency matrices (ESM.20, ESM.23, ESM.26). Then those sheets were saved as a unique file showing the nodes and edges (see NodesEdges.xlsx in the ESM). Afterwards, that unique file was used to prepare separate files in *.xlsx and *.csv for each sample dataset (LSEdges.xlsx, ISEdges.xlsx, DEdges.xlsx, LSEdges.csv, ISEdges.csv, DEdges.csv, LSNodes.xlsx, ISNodes.xlsx, DNodes.xlsx, LSNodes.csv, ISNodes.csv, DNodes.csv, ESM.28, and Gephi folders). Afterward, those *.csv files were transformed into the Gephi data tables by importing them as the Gephi input data files *.csv (ESM.21, ESM.24, ESM.27). Three Gephi files in *.gephi format were generated for each sample dataset (LSGephi.gephi, ISGephi.gephi, DGephi.gephi, and Gephi folder).

The *.gephi files were run one by one for those six layouts and eleven statistical algorithms on a desktop personal computer (PC) without any internet connection and in an active cybersecurity software condition. The PC had a Windows 10 Pro, Intel(R) Core(TM) i5 CPU 650 @ 3.20 GHz, 6,00 GB RAM configuration. All findings and outcomes were compared with each other (Tables 3–4). At first, Force Atlas, ForceAtlas2, Fruchterman–Reingold, OpenOrd, Yifan Hu, and Yifan Hu Proportional layout algorithms were run one by one with self-looping and without self-looping. After each layout algorithm, Noverlap layout algorithm was run to prevent node overlapping of the graph. The layout properties were the Gephi default values in each layout algorithm. The default values could be seen with their property descriptions on the screen views and Gephi files (ESM.21, ESM.24, ESM.27, LSGephi.gephi, ISGephi.gephi, DGephi.gephi). The statistical algorithms were run for undirected edge relations with self-looping and without self-looping because the directions were not important in this study (see directed edges, undirected edges, weighted edges, valued edges, and similar in Golbeck, 2013).

The author opened *.gephi files and run OpenOrd layout algorithm and then Noverlap layout algorithm with selecting degree partition, and without any filtering. After that, BOS selected attributes and checked the boxes for the degree, weighted degree, closeness centrality, betweenness centrality, Eigenvector centrality statistical measures, and others to display and see the values on the graph (ESM.21, LSGephi.gephi, ISGephi.gephi, DGephi.gephi). According to those values, he was able to make his distinctions as follows.

There was a relatively centralized or dense network in the "*location sample dataset*" (Tables 2–3, ESM.19–21, LSGephi.gephi). Nine authors' clusters were observed in the network. BOS with the metrics of degree, weighted degree, closeness centrality, betweenness centrality, Eigenvector centrality as approximately 18; 56; 0,77; 185; 1,0 was the most connected and important author with a high self-looping. It indicated a high publication rate alone. Aderemi A. Atayero with respectively the same statistical measures of approximately 11; 22; 0,56; 28,5; 0,74 and Imhade P. Okokpujie with respectively the same statistical measures of approximately 11; 24; 0,56; 28,5; 0,74 were the second most connected authors with also self-loops.

There was a distributed or uncentralized network in the "*investment sample dataset*" (Tables 3–4, ESM.24, ESM.27, ISGephi.gephi). Eleven authors' clusters and singletons were observed in the network. Andrej Ferdo Gubina, Besim Imamovic, Igbinosa Ikpotokin, Imhade P. Okokpujie, Jerko Pavlicevic, Joseph Dirisu, Joshua Olusegun Okeniyi, Nduka Udoye, Suad Smail Halilcevic had the same statistical measures with a low self-loop.

There was a distributed or uncentralized network in the "*DEMATEL sample dataset*" (Tables 3–4, ESM.27, DGephi.gephi). Seventeen authors' clusters and singletons were observed in the network. Cuina Zhang, Hasan Dincer, Ruobing Li, Serhat Yuksel, Yixing Yuan, and Yun Xia had the same statistical measures with a low self-loop.

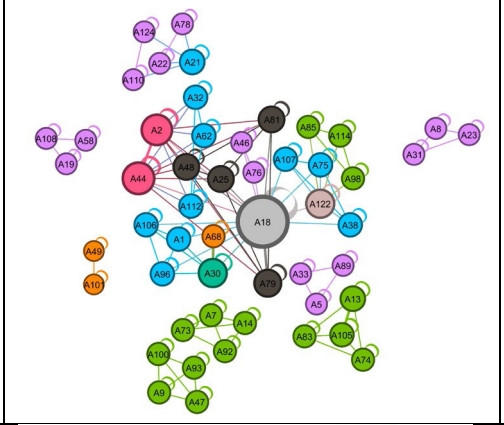
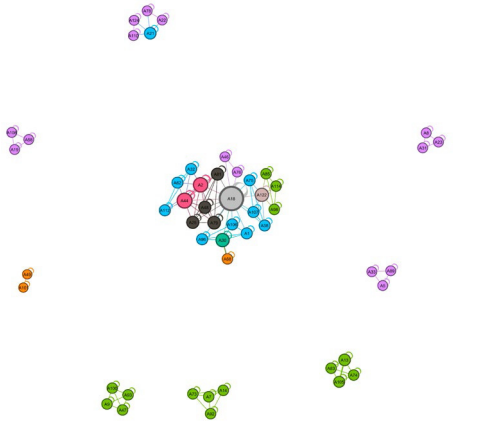
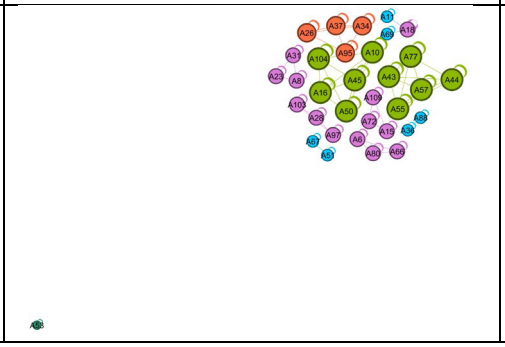
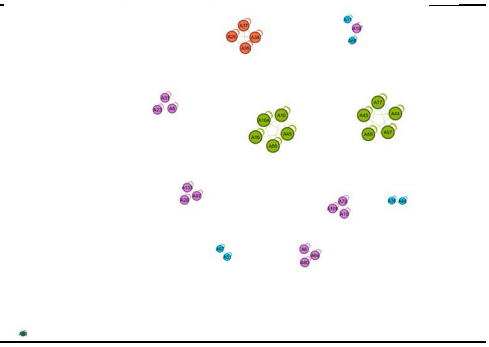
In summary, it was proved well that SNA was very helpful to understand the social relations and find the key people in the social networks. For instance, in this study, the authors in the "*location sample corpora*" were in a centralized network, but the authors in the "*investment sample corpora*", and the "*DEMATEL sample corpora*" were in distributed networks. The most important, influencing, and key researcher in the "*location sample corpora*" was BOS. It was obvious with the statistical measures. There was not only one key researcher in the "*investment sample corpora*", and the "*DEMATEL sample corpora*". Those were more diverse.

5.5. Profile analysis

In cooperation with the previous analysis and findings in the previous sections, a more specific profile analysis was conducted for the authors in the datasets ("*location sample dataset*", "*investment sample dataset*", "*DEMATEL sample dataset*") (ESM.30) based on the RPRFI for a historical, or ex-post analysis and the RFRFPI with simple linear regression for prediction or ex-ante analysis (Fig. 5). Those three separate datasets were combined and one dataset for profile analysis was prepared during this activity. A meaningful profile analysis could only be made on BOS data and information because all other authors had three (i.e. Emre Tercan, Hiroyuki Tamura, Imhade P. Okokpujie, Katsuhiro Akazawa, Ming-Lang Tseng) or less than three (e.g. Sebastian Baumbach, Miguel De Simon Martin, Chia-Nan Wang) publications in that dataset. More specifically, their data and information were very little to make some reliable analysis.

When the main research topic of BOS was analyzed from a historical, or ex-post analysis perspective, it was found out that "DEMATEL", "investment alternatives generation", "investment selection", "location alternatives generation", "location selection", and "location selection factors" had respectively the RPRFI values of 0,07; 0,18; 0,36; 0,04; 0,25; and 0,11. That showed that BOS focused on the topics in the following order "investment selection" 0,36, "location selection" 0,25, "investment alternatives generation" 0,18, "location selection factors" 0,11, "DEMATEL" 0,07, and "location alternatives generation" 0,04 in the past. That finding meant that BOS had focused on mainly "investment selection", "location selection", "investment alternatives generation", and "location selection factors". When the main industries of BOS were analyzed, it was found out that there were five general industries as "hydropower", "power", "renewable energy", "shipbuilding", and "solar power". In that evaluation, the power group could include "hydropower", "renewable energy", and "solar power", and "renewable energy" could consist of "hydropower", and "solar power". According to the first datasets, the nature of publications tended to the current classification. In short, "hydropower", "power", "renewable energy", "shipbuilding", and "solar power" had the RPRFI values of 0,43; 0,04; 0,04; 0,07; and 0,43. If those classes were reorganized,

Table 3
Comparisons of layout algorithms in each sample dataset.

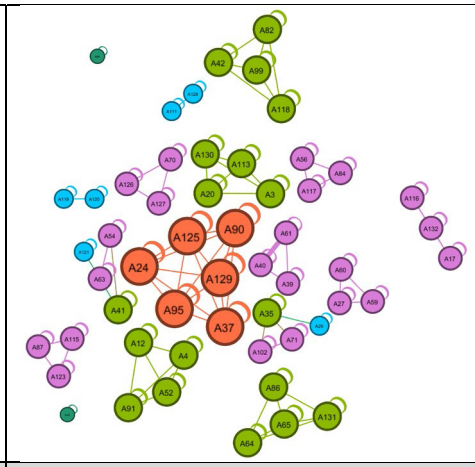
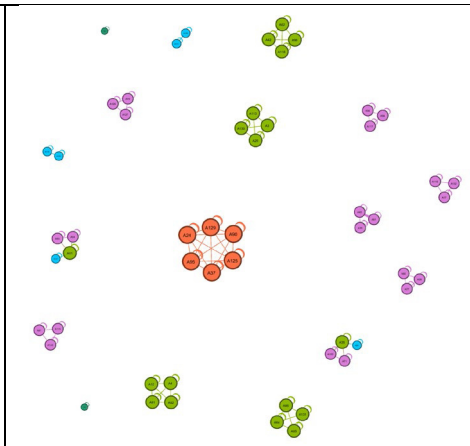
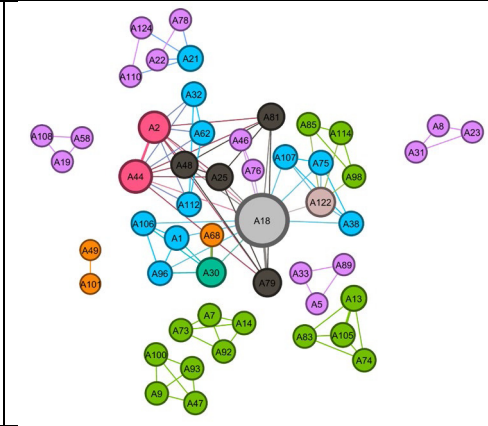
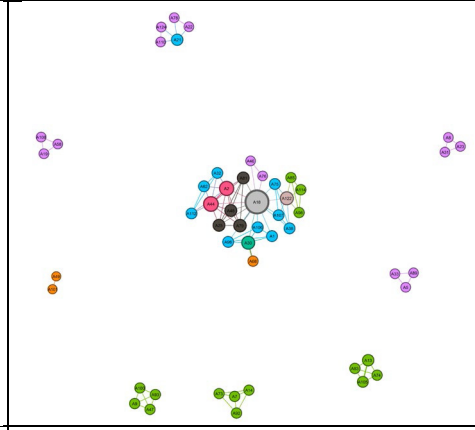
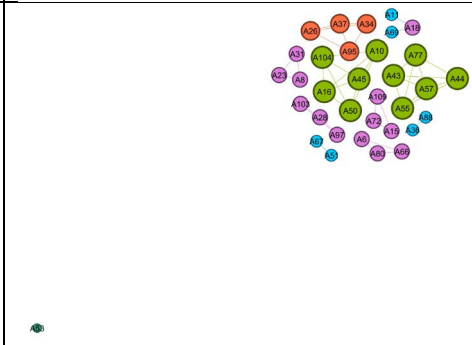
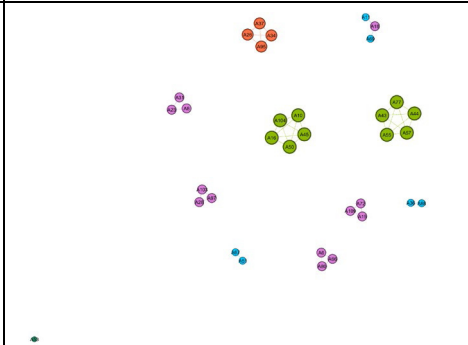
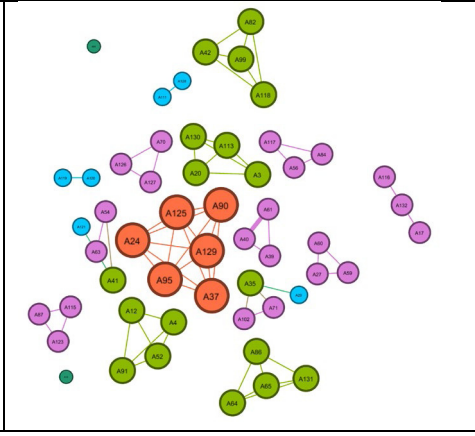
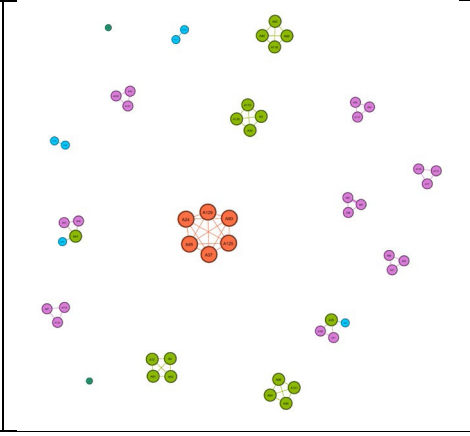
	Location	Investment	DEMATEL
File	LSGephi.gephi	ISGephi.gephi	DGephi.gephi
Layout Algorithms Runtime Details (approximate duration with manual start stop of stopwatch)			
Force Atlas	start: manual stop: manual 25 seconds	start: manual stop: manual 8 seconds	start: manual stop: manual 10 seconds
ForceAtlas2	start: manual stop: manual 25 seconds	start: manual stop: manual 50 seconds	start: manual stop: manual 2 minutes
Fruchterman-Reingold	start: manual stop: manual 25 seconds	start: manual stop: manual 4 seconds	start: manual stop: manual 3 seconds
OpenOrd	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds
Yifan Hu	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds
Yifan Hu Proportional	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds	start: manual stop: automatic 2 seconds
Network Overview Visualization			
	OpenOrd + Noverlap (Partition: degree, Filter: None)	Yifan Hu + Noverlap (Partition: degree, Filter: None)	
Location			
Investment			

(*continued on next page*)

it would be found out that “renewable energy” 0,89 (“hydropower”, “renewable energy”, “solar power”), and “power” 0,93 (“hydropower”, “renewable energy”, “solar power”, “power”). That finding meant that BOS had focused on mainly the renewable power industry, and more specifically in “solar power” and “hydropower”. When the publication time series was studied in both topic and industry-wise, it was found out that there was an increasing trend in BOS’s publications. That finding meant that BOS improved BOS’s researcher technical and technological knowledge. The productivity of BOS increased year by year very much.

At that point, when prediction or ex-ante analysis based on the RFRFPI with simple linear regression was made, the following findings were gathered (1) BOS would continue in researching “investment alternatives generation” topic, that meant that many prescreening, screening, prefeasibility and feasibility studies, and very interesting

Table 3 (*continued*).

DEMATEL		
	OpenOrd + Noverlap (Partition: degree, Filter: Edges Self-Loop)	Yifan Hu + Noverlap (Partition: degree, Filter: Edges Self-Loop)
Location		
Investment		
DEMATEL		

Notes: Nodes: 52, Edges: 146, Undirected graph.

Table 4
Comparison of statistics with filter in edges by self-loop.

Item	Location selection	Investment selection	DEMATEL
File	LSGephi.gephi	ISGephi.gephi	DGephi.gephi
Network overview			
Average degree	3.615	2.471	2.481
Average weighted degree	7.538	4.941	5.111
Network diameter	4	2	2
Network radius	1	0	0
Network average path length	2,0445859872611467	1.0232558139534884	1.056338028169014
Graph density	0.071	0.075	0.047
Modularity	0.710	0.842	0.901
Modularity with resolution	0.710	0.842	0.901
Modularity number of communities	11	11	17
Number of weakly connected components	9	11	17
Node overview			
Clustering coefficient average clustering coefficient	0.934	0.963	0.971
Clustering coefficient total triangles	87	28	44
Eigenvector centrality number of iterations	100	100	100
Eigenvector centrality sum change	0.0054497617151377745	0.006628283670101331	0.011638831177009694
Edge overview			
Graph distance diameter	4	2	2
Graph distance radius	1	0	0
Graph distance average path length	2.0445859872611467	1.0232558139534884	1.056338028169014

Notes: Nodes: 52, Edges: 146, Undirected graph.
Warning: It was experienced that the Gephi results in November 2020 and June 2021 were surprisingly not the same. In other words, the values of results were different with the same data and preferences. The reason could not be found during the publication process. It might be because of versions, updates, operating systems, or similar. Hence, the Gephi results on your device may present different values when you run them.

findings might come up, (2) BOS would continue in researching "investment selection" topic, that meant that many different methods and models for investment selection would be presented, (3) BOS would continue in researching "location selection" topic, that meant that many different methods and models for location selection would be presented, (4) BOS would continue in researching "solar power" topic, that meant that many prescreening, screening, prefeasibility and feasibility studies, and very interesting findings might come up, and many thousands of MW scale solar power plant projects could be presented, (5) BOS would continue in researching "hydropower" topic, that meant that many prescreening, screening, prefeasibility and feasibility studies, and very interesting findings might come up, and many thousands of MW scale solar power plant projects could be presented, (6) BOS quit working in shipbuilding industry, that meant that there would not be any new studies in the future related to shipbuilding industry.

In short, BOS would possibly be a guru in investment alternatives generation, investment selection, location selection, and location selection factors topics in renewable power industry like solar and hydro in the long term (Fig. 5). In other words, BOS could deal with trillions of monetary currencies like USD in the long term. It looked like BOS has a great and strong personality and research talent. Also, BOS might move or immigrate to wealthy countries and prosperous communities to work within large research communities and huge investment groups in the long term too, not stay in Turkiye.

5.6. Validation and verification efforts

Verification, validation, and accreditation (VV&A) includes techniques such as animation, comparison to other models, extreme condition tests, historical data validation, historical methods, static testing, and dynamic testing are the key for dealing with errors, and improving the confidence and credibility of models, software, and systems (Balci, 1997, 1998; Balci, Ormsby, Carr, & Saadi, 2000; Cook & Skinner, 2005; Oberkampf & Trucano, 2008; Sargent, 2010; Wang, Li, Hongbo, Li, Akhtar et al., 2019). They are shortly defined as follows:

Verification: "*process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the mode*" adopted from the 1998 AIAA Guide in (AIAA, 1998; Thacker, Doebling, Hemez, Anderson, Pepin et al., 2004).

Validation: "*process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model*" adopted from the 1998 American Institute of Aeronautics and Astronautics (AIAA) Guide in (AIAA, 1998; Thacker et al., 2004).

Accreditation: "*official certification that a model or simulation is acceptable for use for a specific purpose*" Department of Defense (DoD) Directive 5000.59 in (Balci, 1998; ESD, 2021).

A few verification and validation efforts of *GP2S*, *GP2D*, *GP2E*, and *GP2O* with their direct publications are (King & Saracoglu, 2018; Saracoglu, 2017f, 2018a). In all other publications of *GP2S*, *GP2D*, *GP2E*, and *GP2O*, there are verification and validation efforts indirectly in the focus (Saracoglu, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b, 2017c, 2017d, 2017e, 2017g; Tercan et al., 2020)

According to experiences in those studies, the following issues were noted for the current study. The datasets were checked by the author's efforts. The datasets could not be verified and validated by the original data providers, publishers, because they were not involved in this study. The datasets could not be verified and validated by any third party, because there was not any third party involved in this study. However, their data and information were reliable, because they were mostly the author's work. The author knew better than anybody what the author had prepared in the author's studies, nobody else could do. Hence, the datasets were reliable. Similarly, PARP and SNARP run online and near-real-time or real-time with the cooperation of publishers, BOS recommended that the data and information could be supplied as much as possible by the authors, and then approved by the editors, reviewers, and publishers.

The SNA findings like "BOS was the most connected and important author with a high self-looping" were also reliable, because the datasets were consisted of mainly BOS publications and BOS had lead the publications. Hence, verification and validation of the SNA were easy too. Those publications were related to BOS's inventions, ideas, and proposals. BOS knew everything about the kick-off and the finish of the publications processes.

The nature of profile analysis was personal and related to the plans of any researcher so that in-depth interviews or questionnaires were necessary to collect information. Fortunately, this activity was very simple in this study, because the datasets were very much related

to the author of this paper. Hence, verification and validation of the profile analysis were easy too. The ex-post analysis with RPRFI worked perfectly. The most interesting research topics were matched correctly "investment alternatives generation", "investment selection", "location selection", and "location selection factors". The ex-ante analysis with RFRFPI with simple linear regression predicted the future studies very well. BOS planned to continue researching in "solar power", "hydropower", and "renewable energy". The researcher would like to present online prototypes of *GP2S*, *GP2D*, *GP2E*, and *GP2O* in the future. Although BOS had a sort of invention and a new idea of unmanned hybrid vehicle (unmanned surface vehicle and unmanned aerial vehicle: SSXUSVUAVC concept) since 2013 for surveillance, search and rescue including firefighting, city and urban planning etc., the researcher would not like to continue developing and demonstrating SSXUSVUAVC concept due to time, health and some other personal security and similar issues (Saracoglu, 2013-2016). Hence, the shipbuilding finding (BOS was negative (–) to shipbuilding industry) was also correct. Turkiye was not a safe and secure country. Many researchers had died and got sick suspiciously. Strange and stupid things had been going on for many years in the governmental, state and justice systems. It might be happen all over the World too.

In short, it was observed that if that sort of data and information was collected and analyzed in an unbiased manner, the ex-post and ex-ante analysis would be reasonable and trustworthy. The predictions would be accurate, valuable, and reliable.

6. Conclusions and future research

The proposed robots and platforms, *GP2S*, *GP2D*, *GP2E*, and *GP2O* will for sure improve the quality of living on our planet. They will help people to respect the Mother Nature. They allow understanding the facts very well and guide to design net-zero and negative emission green ecological power grids. Profile analysis and social network analysis will allow finding appropriate researchers to develop appropriate researchers groups for D3&D activities. Thus, profile analysis and social network analysis tools are crucial. Therefore, this particular study focuses on profile analysis and social network analysis to prepare standardized corpora and standardized files representing literature for artificial intelligence and machine learning infrastructures, platforms, applications and proposes, recommends profile analyses robot and platform (*PARP*), and social network analyses robot and platform (*SNARP*).

This study presents a concise, systematic, and specific review in three topic clusters. These are "*location cluster*", "*investment cluster*", and "*DEMATEL cluster*". The collected and clustered documents help to prepare "*location corpora*", "*investment corpora*", and "*DEMATEL corpora*". Those three corpora are used for developing three "*sample corpora*" and "*sample dataset*", which are then used for profile analysis and social network analysis.

There are some significant core contributions during this study. For example (1) three sample corpora for three topic clusters from many electronic books, electronic journal articles, and electronic conference proceedings are prepared and presented, (2) three sample datasets related to those three sample corpora are prepared and presented, (3) two publications in each sample dataset are summarized, (4) that activity helps finding potential research directions and researchers directions in the future, (5) research trends are found and shown, (6) social network analysis is performed, (7) foundations of research documents universe $Universe_{researchdocuments}$ are built, (8) foundations of researchers universe $Universe_{researchers}$ are built, (9) a few D3&D activities are finalized for automated profile analysis and social network analysis systems, (10) prepared datasets are shared with ML community, (11) Corpora also can be gathered with easy, because all citations are made perfectly.

There are also many limitations to this study. For example (1) the number of documents in each corpus is very small as they are text corpora, (2) there is not any citation data and information, (3) there is not enough data and information for expert elicitation (see Chan, Anadon, Chan, & Lee, 2011 for expert elicitation), (4) there is not any data and information about the characteristics of authors such as health, behavior, criminality, or politics, (5) there is not any data and information about keywords and their relations, (6) there is not any network analysis of publications, (7) there is not any analysis for biases (see Wikipedia, 2020 for cognitive biases)

In addition, it is realized and shown that profile analysis and social network analysis can be used for analyzing all social networks. For instance: criminal structures. They are great tools for intelligence studies, criminal justice investigations, and community justice investigations. Hence, intelligence community (IC) and justice community have to learn, understand and use them very well. Also, they can help to deal with headaches and troubles caused by some tools and criminals using those tools like DeepFakes or Deep Fakes (see Shen, Liu, Bai, & Li, 2018; Perov et al., 2020 for DeepFakes or Deep Fakes).

The next studies, it is planned to encourage researchers, former intelligence, and justice analysts to present similar publications in their fields to develop and improve the literature. Moreover, it is planned to start investigating biases, falsifications, lies, and similar in the publications and their authors for developing detection tools. In addition, ML researchers may apply many ML algorithms and can come up with their findings in the same corpora and datasets. They can compare their findings with the findings of this study and share their observations and findings with the author of this paper to improve the current state of the science.

CRediT authorship contribution statement

Burak Omer Saracoglu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Amazon Review Data (2018). Amazon Review Dataset. Retrieved from https://nijianmo.github.io/amazon/index.html. Accessed January 21, 2022 (Ali, Kadhim, & Abid, 2020; De, Dehuri, & Wang, 2012).

Acknowledgments

The author would first and foremost like to sincerely express his deepest gratefulness to The Almighty, The Eminent, The Exalted, The Exalted in Might and Power, The Mighty, The Mighty One, The Powerful, The Powerful and Incomparable, The Self Sufficient, The Victorious, and then thankfulness to Dr. Ted Fuller, T.R. Massey, Dr. Andrzej Kobryn, Dr. Luiz Flavio Autran Monteiro Gomes, Dr. Frederico Silva Valentim Sallum, ScienceOpen (Nina), OneLook (Harvey Beeferman, Datamuse), Dr. Barbara Kosztra, Dr. Jennifer Golbeck, Dr. Charles Zaiontz and the editors, reviewers, and journal staff of this publication and the publications of "Location selection factors of concentrated solar power plant investments", "Location and investment factors of hydropower plants" in some different journals for expert elicitation comment. Any opinion, method, data, information, graph, picture, photo, clause, sentence, finding, remark, conclusion, and recommendation expressed and presented herein are the authors' or the referenced authors' and do not necessarily reflect the views of any governmental, national or international parties, bodies or organizations. Please feel free to send your ideas, feedbacks and criticisms to the author's e-mail at any time.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.mlwa.2022.100249.

References

Agarap, A. F. (2021). Text classification and clustering with annealing soft nearest neighbor loss. arXiv preprint, arXiv:2107.14597v.

Aggarwal, C. C., & Zhai, C. X. (2012). *A Survey of Text Clustering Algorithms*.

Ahmed, R. I., Iqbal, M. J., & Bakhsh, M. (2019). Sentiment analysis of social media contents using machine learning algorithms. *Technical Journal, University of Engineering and Technology (UET) Taxila*, *24*(4), 47–55.

Ahmed, W., Tan, Q., Solangi, Y. A., & Ali, S. (2020). Sustainable and special economic zone selection under fuzzy environment: A case of Pakistan. *Symmetry*, *12*, 242.

AIAA (1998). *Guide: Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998(2002))*. http://dx.doi.org/10.2514/4.472855.001.

AlHarbi, B. Y., AlHarbi, B. S., AlZahrani, N. J., Alsheail, M. M., & Ibrahim, D. M. (2020). Using machine learning algorithms for automatic cyber bullying detection in arabic social media. *Journal of Information Technology Management*, *12*(2), 123–130. http://dx.doi.org/10.22059/jitm.2020.75796.

Ali, Z. H., Kadhim, A. A., & Abid, A. M. (2020). Subject review: text clustering algorithms. *International Journal of Engineering Research and Advanced Technology (IJERAT)*, *6*, 9. http://dx.doi.org/10.31695/IJERAT.2020.3646.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint, arXiv:1707.02919v2.

Asgari-Chenaghlu, M., Feizi-Derakhshi, M. R., Farzinvash, L., Balafar, M. A., & Motamed, C. (2021). Topic detection and tracking techniques on Twitter: A systematic review. *Complexity*, Article 8833084. http://dx.doi.org/10.1155/2021/8833084.

Balci, O. (1997). Verification, validation, and accreditation of simulation models. In *Proceedings of the 1997 winter simulation conference* (pp. 135–141).

Balci, O. (1998). Verification, validation, and accreditation. In *Proceedings of the 1998 winter simulation conference* (pp. 41–48).

Balci, O., Ormsby, W. F., Carr, J. T., & Saadi, S. D. (2000). Planning For verification, validation, and accreditation of modeling and simulation applications. In *Proceedings of the 2000 winter simulation conference* (pp. 829–839).

Barber, D. (2017). *Bayesian Reasoning and Machine Learning*.

Barocas, S., Hardt, M., & Narayanan, A. (2020). *Fairness in Machine Learning Limitations and Opportunities*.

Baumbach, S., Khan, J., Ahmed, S., & Dengel, A. (2020). Quvis the question of visual site selection. arXiv preprint, arXiv:2005.12375.

Baykasoglu, A., Kaplanoglu, V., Durmusoglu, Z. D., & Sahin, C. (2013). Integrating fuzzy DEMATEL and fuzzy hierarchical TOPSIS methods for truck selection. *Expert Systems with Applications*, *40*(3), 899–907.

Berk, K. N., & Carey, P. (2010). *Data Analysis with Microsoft® Excel Updated for Office 2007®*. Boston, MA, USA: Brooks/Cole.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10*, P10008.

Boddu, R. N., & Devi, G. R. (2020). Detect and prevent the cyber bullying conversation's on social media using machine learning algorithm. *Mukt Shabd Journal*, 2025–2033, IX, VII.

Boguna, M., Pastor-Satorras, R., Diaz-Guilera, A., & Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical Review E*, *70*(5), Article 056122.

Botelho, J., & Antunes, C. (2011). Combining social network analysis with semi-supervised clustering: a case study on fraud detection.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(2), 163–177.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on the world wide web (WWW1998)* (pp. 107–117).

Bulhoes, R. L., de Santana, E. S., & Santos, A. A. B. (2020). Use of analytic hierarchy process for wind farm installation region prioritization–case study. *Energies*, *13*(9), 1–22.

Cebi, S. (2013a). Determining importance degrees of website design parameters based on interactions and types of websites. *Decision Support Systems*, *54*(2), 1030–1043.

Cebi, S. (2013b). A quality evaluation model for the design quality of online shopping websites. *Electronic Commerce Research and Applications*, *12*(2), 124–135.

Chai, Y., Zhang, H., & Jin, S. (2020). Neural text classification by jointly learning to cluster and align. arXiv preprint, arXiv:2011.12184v.

Chan, G., Anadon, L. D., Chan, M., & Lee, A. (2011). Expert elicitation of cost, performance and RD & D budgets for coal power with CCS energy. *Procedia*, *4*, 2685–2692.

Choudhury, S., Howladar, P., Majumder, M., & Saha, A. K. (2019). Application of novel MCDM for location selection of surface water treatment plant. *IEEE Transactions on Engineering Management*.

Choudhury, S., & Saha, A. K. (2019). Location selection for installation of surface water treatment plant by applying a new Sinusoidal Analytical Hierarchy Process: application of new MCDM in location detection. *International Journal of Energy Optimization and Engineering*, *8*(3), 20–42.

Collins (2021). Turkey and turkey" collins online english dictionary. Retrieved from https://www.collinsdictionary.com/dictionary/english/turkey Accessed June 10, 2021.

Cook, D. A., & Skinner, J. M. (2005). How to perform credible verification, validation, and accreditation for modeling and simulation. In *Patents & Software Technology Conference, CROSSTALK the Journal of Defense Software Engineering* (pp. 20–24).

Dalal, M. K., & Zaveri, M. A. (2011). Automatic text classification: A technical review. *International Journal of Computer Applications*, *28*(2), 37–40.

De, S. S., Dehuri, S., & Wang, G. N. (2012). Machine learning for social network analysis: A systematic literature review. *IUP Journal of Information Technology*, *8*(4), https://www.researchgate.net/profile/Sagar-S-De/publication/251236864_Machine_Learning_for_Social_Network_Analysis_A_Systematic_Literature_Review/links/59e8914e458515c363133f6a/Machine-Learning-for-Social-Network-Analysis-A-Systematic-Literature-Review.pdf.

Dereli, M. A., & Tercan, E. (2020). Comparison of GIS-based surrogate weighting methods for multi-directional landfill site selection in west Mediterranean Planning Region in Turkey. *Environment, Development and Sustainability*.

Di Capua, M., Di Nardo, E., & Petrosino, A. (2016). Unsupervised cyber bullying detection in social networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 427–432). Cancún, México: Cancún Center, 978-1-5090-4846-5/16.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, Correspondence*, *20*(5), 533–534. http://dx.doi.org/10.1016/S1473-3099(20)30120-1, COVID-19 Dashboard https://github.com/CSSEGIS{and}Data/COVID-19 and https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6.

Edwards, M., Wattam, S., Rayson, P., & Rashid, A. (2016). Sampling labelled profile data for identity resolution.

ESD (2021). Department of defense (DoD) directive 5000.59. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500059p.pdf?ver=2018-10-22-083243-490 Accessed 02 September 2021.

Fan, Z. P., Suo, W. L., & Feng, B. (2012). Identifying risk factors of IT outsourcing using interdependent information: An extended DEMATEL method. *Expert Systems with Applications*, *39*(3), 3832–3840.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, *21*(11), 1129–1164.

Fu, X., Zhu, Q., & Sarkis, J. (2012). Evaluating green supplier development programs at a telecommunications systems provider. *International Journal of Production Economics*, *140*(1), 357–367.

Gannarapu, S., Dawoud, A., Ali, R. S., & Alwan, A. (2020). Bot detection using machine learning algorithms on social media platforms in citisia. In *Proceedings of 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications*. http://dx.doi.org/10.1109/CITISIA50690.2020.9371778.

García, S. (2015). The importance of good labels in security datasets. https://www.linkedin.com/pulse/importance-good-labels-security-datasets-sebasti%C3%A1n-garc%C3%ADa Accessed 30 2021.

Gauen, K., Dailey, R., Laiman, J., Zi, Y., Asokan, N., Lu, Y. H., Thiruvathukal, G. K., Shyu, M. L., & Chen, S. C. (2017). Comparison of visual datasets for machine learning.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé Iii, H., & Crawford, K. (2021). Datasheets for datasets. arXiv:1803.09010v7.

Geng, X., & Chu, X. (2012). A new importance–performance analysis approach for customer satisfaction evaluation supporting PSS design. *Expert Systems with Applications*, *39*(1), 1492–1502.

Gephi (2021). Gephi. website Retrieved from https://gephi.org Accessed June 6, 2021.

Gephi github (2021). Gephi github website. Retrieved from https://github.com/gephi Accessed June 6, 2021.

Gokarn Nitin, M., Gottipati, S., & Shankararaman, V. (2019). Clustering models for topic analysis in graduate discussion forums. In *Proceedings of the 27th International Conference on Computers in Education* (pp. 1–10). Taiwan: Research Collection School of Computing and Information Systems.

Golbeck, J. (2013). *Analyzing the Social Web*. Morgan Kaufmann/ Elsevier.

Google (2021a). Google ML developers tutorials, google ML text classification. https://developers.google.com/machine-learning/guides/text-classification/. Accessed 30 2021.

Google (2021b). Google cloud AI platform. Data Labeling Service https://cloud.google.com/ai-platform/data-labeling/docs. Accessed 30 2021.

Google (2021c). Google developers machine learning guides. https://developers.google.com/machine-learning/guides. Accessed 30 2021.

Google (2021d). Google people AI guidebook. https://pair.withgoogle.com/guidebook. Accessed 30 2021.

Google Cloud AI Platform (2021a). Google cloud AI platform creating label sets. https://cloud.google.com/ai-platform/data-labeling/docs/label-sets Accessed 30 2021.

Google Cloud AI Platform (2021b). Google cloud AI platform creating instructions for human labelers. https://cloud.google.com/ai-platform/data-labeling/docs/instructions Accessed 30 2021.

Google Cloud AI Platform (2021c). Google cloud AI platform creating datasets. https://cloud.google.com/ai-platform/data-labeling/docs/datasets Accessed 30 2021.

Google Cloud AI Platform (2021d). Google cloud AI platform submitting text labeling requests. https://cloud.google.com/ai-platform/data-labeling/docs/text-request Accessed 30 2021.

Goswami, S. S., Behera, D. K., & Mitra, S. (2020). A comprehensive study of weighted product model for selecting the best laptop model available in the market. *Brazilian Journal of Operations & Production Management*, *17*(2), Article e2020875.

Govindan, K., Kannan, D., & Shankar, K. M. (2014). Evaluating the drivers of corporate social responsibility in the mining industry with multi-criteria approach: A multi-stakeholder perspective. *Journal of Cleaner Production*, *84*, 214–232.

Grandjean, M., & Jacomy, M. (2019). *Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics. Digital Humanities*. Utrecht, Netherlands: ffhalshs-02179024ff.

Gupta, T. (2017). Keyword extraction: A review. *International Journal of Engineering Applied Sciences and Technology*, *2*(4), 215–220.

Gurung, P., & Wagh, R. (2017). A study on topic identification using k means clustering algorithm: Big vs. Small documents. *Advances in Computational Sciences and Technology*, *10*(2), 221–233.

He, R., & McAuley, J. (2016). *Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering*. WWW.

Hu, Y. F. (2005). Efficient and high quality force-directed graph drawing. *The Mathematica Journal*, *10*, 37–71.

Hu, S. K., Lu, M. T., & Tzeng, G. H. (2014). Exploring smart phone improvements based on a hybrid MCDM model. *Expert Systems with Applications*, *41*(9), 4401–4413.

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.), (2019). *Automated Machine Learning Methods, Systems, Challenges*.

Islam, R., Ashad Kabir, M., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, *6*, 8. http://dx.doi.org/10.1007/s13755-018-0046-0.

Jacomy, M. (2020). Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. *Big Data & Society*, 1–21. http://dx.doi.org/10.1177/2053951720949577.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS One*, *9*(6), Article e98679. http://dx.doi.org/10.1371/journal.pone.0098679.

Jeng, D. J. F., & Tzeng, G. H. (2012). Social influence on the use of clinical decision support systems: revisiting the unified theory of acceptance and use of technology by the fuzzy DEMATEL technique. *Computers & Industrial Engineering*, *62*(3), 819–828.

Kaleeswari, K., Johnson, T., & Vijayalakshmi, C. (2018). Application of fuzzy AHP in water treatment plant location. *Journal of Advanced Research in Dynamical and Control Systems*, 10, 01-Special Issue.

Kayacik, H. G., & Zincir-Heywood, H. (2015). Analysis of three intrusion detection system benchmark datasets using machine learning algorithms.

Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, *1*(1), http://dx.doi.org/10.4304/jait.1.1.4-20.

Kiesslich, T., Beyreis, M., Zimmermann, G., & Traweger, A. (2021). Citation inequality and the journal impact factor: median, mean, (does it) matter? *Scientometrics*, *126*, 1249–1269. http://dx.doi.org/10.1007/s11192-020-03812-y.

Kim, S. W., & Gil, J. M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-Centric Computing and Information Sciences*, *9*, 30. http://dx.doi.org/10.1186/s13673-019-0192-7.

King, A., & Saracoglu, B. O. (2018). Greenough river solar farm case study & validation initialization. *International Journal of Energy Applications and Technologies*, *5*(2), 82–97. http://dx.doi.org/10.31593/ijeat.420701.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632.

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, *10*(150), http://dx.doi.org/10.3390/info10040150.

Lambiotte, R., Delvenne, J. C., & Barahona, M. (2009). Laplacian dynamics and multiscale modular structure in networks.

Latapy, M. (2008). Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science*, *407*, 458–473. http://dx.doi.org/10.1016/j.tcs.2008.07.017.

Li, G. (2018). A literature review on patent texts analysis techniques. *International Journal of Knowledge and Language Processing*, *9*(3), 1–15.

Li, Y., & Saracoglu, B. O. (2021). Location and investment factors of hydropower plants. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, http://dx.doi.org/10.1080/15567036.2021.1963015.

Li, Y., Su, H., Shen, X., Li, W., Cao, Z., & Niu, S. (2017). DailyDialog: A manually labelled multi-turn dialogue dataset. arXiv:1710.03957v1.

Malik, N., & Jain, S. (2020). Comparative study of machine learning algorithms for social media text analysis. In *REDSET 2019 CCIS 1230* (pp. 223–235). http://dx.doi.org/10.1007/978-981-15-5830-6_19.

Martin, S., Brown, W. M., Klavans, R., & Boyac, K. W. (2011). OpenOrd: An Open-Source Toolbox for Large Graph Layout. In *Proceedings of SPIE Conference on Visualization and Data Analysis (VDA) The International Society for Optical Engineering*.

McAuley, J., & Leskovec, J. (2012). *Learning to Discover Social Circles in Ego Networks*. NIPS.

McAuley, J., Targett, C., Shi, J., & van den Hengel, A. (2015). *Image-Based Recommendations on Styles and Substitutes*. SIGIR.

McAuley, J., & Yang, A. (2016). *Addressing Complex and Subjective Product-Related Queries with Customer Reviews*. World Wide Web (WWW).

McMahan, P., & McFarland, D. A. (2021). Creative destruction: The structural consequences of scientific curation. *American Sociological Review*, *86*(2), 341–376. http://dx.doi.org/10.1177/0003122421996323.

MEAP (2021). MEAP edition manning early access program grokking machine learning version 15. https://livebook.manning.com/book/grokking-machine-learning/welcome/v-4/ Accessed 30 2021.

Merlini, D., & Rossini, M. (2021). Text categorization with WEKA: A survey Machine Learning with Applications, 4, 100033.

Milkova, M., Andreichikova, O., & Andreichikov, A. (2018). Venture capitalists decision making: applying analytic network process to the startups evaluation. *International Journal of the Analytic Hierarchy Process*, *10*(1), 2–19.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning based text classification: A comprehensive review. arXiv:2004.03705v3.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning* (2nd ed.). Cambridge, MA: The MIT Press.

Moitra, A. (2014). Algorithmic aspects of machine learning.

Mombini, H., Li, R., Zhang, Y., Korkin, D., & Tulu, B. (2020). An exploratory study of social media analysis for rare diseases using machine learning algorithms: A case study of trigeminal neuralgia. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*. https://hdl.handle.net/10125/64211.

Ni, J., Li, J., & McAuley, J. (2021). Justifying recommendations using distantly-labeled reviews and fine-grained aspects.

Norregaard, J., Horne, B. D., & Adali, S. (2019). NELA-GT-2018: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles. In *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)* (pp. 630-638).

Nurek, M., & Michalski, R. (2020). Combining machine learning and social network analysis to reveal the organizational structures. *Applied Sciences*, *10*(1699), http://dx.doi.org/10.3390/app10051699.

Obeidat, M. S., Qasim, T., & Khanfar, A. (2018). Implementing the AHP multi-criteria decision approach in buying an apartment in Jordan. *Journal of Property Research*, *35*(1), 53–71.

Oberkampf, W. L., & Trucano, T. G. (2008). Verification and validation benchmarks. *Nuclear Engineering and Design*, *238*, 716–743, 2008.

Ohunakin, O. S., & Saracoglu, B. O. (2018). A comparative study of selected multi-criteria decision making methodologies for location selection of very large concentrated solar power plants in Nigeria. *African Journal of Science, Technology, Innovation and Development*, *10*(5), 551–567. http://dx.doi.org/10.1080/20421338.2018.1495305.

Okokpujie, I. P., Okonkwo, U. C., Akinlabi, E. T., Okokpujie, K., & Atayero, A. A. (2019). Multi-criteria decision analysis towards the selection of a perfect location for establishing crude oil refinery in Niger Delta Nigeria. *Journal of Physics: Conference Series*, *1378*(3), Article 032100.

Patel, J. N., & Rana, S. C. (2018). A selection of the best location for a small hydro power project using the AHP-weighted sum and PROMETHEE method. *Pertanika Journal of Science and Technology*, *26*(4), 1591–1603.

Perer, A., & Shneiderman, B. (2006). Balancing systematic and flexible exploration of social networks. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 693–700.

Perer, A., & Shneiderman, B. 2008a. Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In *Proceedings of the 13th international conference on intelligent user interfaces*, (pp. 109-118).

Perer, A., & Shneiderman, B. 2008b. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 265-274).

Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Ume, C., dpfks, Mr., Facenheim, C. S., L, RP., Jiang, J., Zhang, S., Wu, P., Zhou, B., & Zhang, W. (2020). DeepFaceLab: A Simple, flexible and extensible face swapping framework. arXiv preprint, arXiv:2005.05535v4.

Pourvali, M., Orlando, S., & Omidvarborna, H. (2018). Topic models and fusion methods: a union to improve text clustering and cluster labeling. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(4), 28–34. http://dx.doi.org/10.9781/ijimai.2018.12.007.

Pugachev, L., & Burtsev, M. (2021). Short text clustering with transformers. arXiv preprint, arXiv:2102.00541v1.

Qiu, D., Dincer, H., Yuksel, S., & Ubay, G. G. (2020). Multi-faceted analysis of systematic risk-based wind energy investment decisions in E7 economies using modified hybrid modeling with IT2 fuzzy sets. *Energies*, *13*(6), 1423.

Rao, KS., Sreeram, G., & Raju, BB. (2020). Detecting fake account on social media using machine learning algorithms. *International Journal of Control and Automation*, *13*(1), 95–100.

Ren, J., Manzardo, A., Toniolo, S., & Scipioni, A. (2013). Sustainability of hydrogen supply chain, part I: Identification of critical criteria and cause–effect analysis for enhancing the sustainability using DEMATEL. *International Journal of Hydrocarbon Engineering*, *38*(33), 14159–14171.

Rozic, I., Imamovic, B., Pavlicevic, J., Gubina, A. F., & Halilcevic, S. S. (2018). The energy sustainability of the small agricultural farms isolated of electric power grid. *International Review of Electrical Engineering*, *13*(3), 229–236. http://dx.doi.org/10.15866/iree.v13i3.14413.

Saracoglu, B. O. (2013). Selecting industrial investment locations in master plans of countries. *European Journal of Industrial Engineering*, *7*(4), 416–441.

Saracoglu (2013-2016). Ssxusvuavc concept. http://dx.doi.org/10.13140/RG.2.1.4013.8001, https://www.researchgate.net/publication/294721631_SSXUSVUAVC_Concept.

Saracoglu, B. O. (2014). Wsc18-SS1.2: A proposal and its application on type-2 fuzzy mamdani's method for selection classification of private small hydropower plant investments in the very early investment stages in Turkey. In *Presentation in the 18th online world conference on soft-computing in industrial applications (wsc18), special session 1: application of soft computing in energy industry – part 1*. http://www.fti.itb.ac.id/wsc18/tutorial/.

Saracoglu, B. O. (2015a). An experimental research study on the solution of a private small hydropower plant investments selection problem by ELECTRE III/IV, Shannon's Entropy & Saaty's subjective criteria weighting. *Advances in Decision Sciences*, Article 548460. http://dx.doi.org/10.1155/2015/548460.

Saracoglu, B. O. (2015b). An AHP application in the investment selection problem of small hydropower plants in Turkey. *International Journal of the Analytic Hierarchy Process*, *7*(2), 211–239. http://dx.doi.org/10.13033/ijahp.v7i2.198.

Saracoglu, B. O. (2015c). An experimental research of small hydropower plant investments selection in Turkey by Carrot2, DEXi, DEXiTree. *Journal of Investment and Management*, *4*(1), 47–60. http://dx.doi.org/10.11648/j.jim.20150401.17.

Saracoglu, B. O. (2015d). A comparative study of AHP, ELECTRE III & ELECTRE IV by equal objective & Shannon's entropy objective & Saaty's subjective criteria weighting on the private small hydropower plants investment selection problem in Turkey. *International Journal of the Analytic Hierarchy Process*, *7*(3), 470–512. http://dx.doi.org/10.13033/ijahp.v7i3.343.

Saracoglu, B. O. (2016a). A qualitative multi-attribute model for the selection of the private hydropower plant investments in Turkey: by foundation of the Search Results Clustering Engine (Carrot2), hydropower plant clustering, DEXi and DEXiTree. *Journal of Industrial Engineering and Management*, *9*(1), 152–178. http://dx.doi.org/10.3926/jiem.1142.

Saracoglu, B. O. (2016b). A PROMETHEE i, II and GAIA based approach by saaty's subjective criteria weighting for small hydropower plant investments in Turkey. *International Journal of Renewable Energy Technology*, *7*(2), 163–183. http://dx.doi.org/10.1504/IJRET.2016.076094.

Saracoglu, B. O. (2017a). Location selection factors of small hydropower plant investments powered by SAW, grey WPM & fuzzy DEMATEL based on human natural language perception. *International Journal of Renewable Energy Technology*, *8*(1), 1–23. http://dx.doi.org/10.1504/IJRET.2017.080867.

Saracoglu, B. O. (2017b). Long term electricity demand & peak power load forecasting variables identification & selection. *Science Journal of Circuits, Systems and Signal Processing*, *6*(2), 18–28. http://dx.doi.org/10.11648/j.cssp.20170602.13.

Saracoglu, B. O. (2017c). G2EDPS'S first module & its first extension modules. *American Journal of Applied Scientific Research*, *3*(4), 33–48. http://dx.doi.org/10.11648/j.ajasr.20170304.13.

Saracoglu, B. O. (2017d). Comparative study on experimental 2 to 9 triangular fuzzy membership function partitioned type 1 Mamdani's FIS for G2EDPS. *Global Journal of Researches in Engineering: J General Engineering*, *17*(2), 1–18.

Saracoglu, B. O. (2017e). Comparative study on experimental type 1 & interval & general type 2 Mamdani FIS for G2p3s. *Global Journal of Researches in Engineering: J General Engineering*, *17*(2), 27–42.

Saracoglu, B. O. (2017f). SEGS VI & TOpaz Solar Farm SAM empirical trough & PVWatts models case studies & validation. *International Journal of Research in Advanced Engineering Technologies*, *1*(1), 28–41.

Saracoglu, B. O. (2017g). An experimental fuzzy inference system for the third core module of the first console on the global grid peak power prediction system & its forecasting accuracy measures' comparisons with the first and the second core modules. *Journal of Energy Systems*, *1*(2), 75–101.

Saracoglu, B. O. (2018a). Solar star projects SAM version 2017.9.5 pvwatts version 5 model case study & validation. *International Journal of Energy Applications and Technologies*, *5*(1), 13–28.

Saracoglu, B. O. (2018b). Experimental FWA & FOWA aggregated vlcspps' LUR estimation for GIS based VEED. *International Journal of Engineering Technologies*, *4*(2), 70–79.

Saracoglu, B. O. (2018c). An experimental fuzzy expert system based application for go/no-go decisions to geospatial investigation studies of the regions on very large concentrated solar power plants in European super-grid concept. *International Journal of Multidisciplinary Studies and Innovative Technologies*, *2*(2), 01–06.

Saracoglu, B. O. (2018d). *Analytic network process vs. benjamin franklin's rule to select private small hydropower plants investments*. MedCrave Group LLC.

Saracoglu, B. O. (2019a). An experimental study on fuzzy expert system: proposal for financial suitability evaluation of commercial and participation banks in power plant projects in Turkey. In B. Ane, A. Cakravastia, & L. Diawati (Eds.), *Advances in intelligent systems and computing: Vol. 864, Proceedings of the 18th online world conference on soft computing in industrial applications (wsc18). wsc 2014* (pp. 81–91). Cham: Springer, http://dx.doi.org/10.1007/978-3-030-00612-9_8.

Saracoglu, B. O. (2019b). An experimental case study on fuzzy logic modeling for selection classification of private mini hydropower plant investments in the very early investment stages in Turkey. In B. Ane, A. Cakravastia, & L. Diawati (Eds.), *Advances in intelligent systems and computing: Vol. 864, Proceedings of the 18th online world conference on soft computing in industrial applications (wsc18). wsc 2014* (pp. 69–80). Cham: Springer, http://dx.doi.org/10.1007/978-3-030-00612-9_7.

Saracoglu, B. O. (2019c). Experimental FOWA aggregated location selection model for VLCPVPPs in MENA region in the very early engineering design. In B. Ane, A. Cakravastia, & L. Diawati (Eds.), *Advances in intelligent systems and computing: Vol. 864, Proceedings of the 18th online world conference on soft computing in industrial applications (wsc18). wsc 2014* (pp. 36–45). Cham: Springer, http://dx.doi.org/10.1007/978-3-030-00612-9_4.

Saracoglu, B. O. (2019d). Experimental FuzzyWA aggregated location selection model for very large photovoltaic power plants in global grid in the very early engineering design process stage. In B. Ane, A. Cakravastia, & L. Diawati (Eds.), *Advances in intelligent systems and computing: Vol. 864, Proceedings of the 18th online world conference on soft computing in industrial applications (wsc18). wsc 2014* (pp. 25–35). Cham: Springer, http://dx.doi.org/10.1007/978-3-030-00612-9_3.

Saracoglu, B. O. (2019e). *Multiobjective evolutionary algorithms knowledge acquisition system for renewable energy power plants*. MedCrave Group LLC.

Saracoglu, B. O. (2020a). Location selection factors of concentrated solar power plant investments. *Sustainable Energy, Grids and Networks*, *22*, Article 100319. http://dx.doi.org/10.1016/j.segan.2020.100319.

Saracoglu, B. O. (2021a). Global power prediction systems. Retrieved from https://www.researchgate.net/project/Global-Power-Prediction-Systems (Accessed June 10, 2016).

Saracoglu, B. O. (2021b). Global power plants developers. Retrieved from https://www.researchgate.net/project/Global-Power-Plants-Developers (Accessed June 10, 2016).

Saracoglu, B. O. (2021c). Global power plants engineers. Retrieved from https://www.researchgate.net/project/Global-Power-Plants-Engineers (Accessed June 10, 2016).

Saracoglu, B. O. (2021d). Global power plants owners. Retrieved from https://www.researchgate.net/project/Global-Power-Plants-Owners (Accessed June 10, 2016).

Saracoglu, B. O., Insel, M., & Helvacioglu, I. H. (2009). Identification of location selection criteria for new building shipyards. *Marine Technology*, *46*(1), 34–44.

Saracoglu, B. O., Ohunakin, O. S., Adelekan, D. S., Gill, J., Atiba, O. E., Okokpujie, I. P., & Atayero, A. A. (2018). A framework for selecting the location of very large photovoltaic solar power plants on a global/super grid. *Energy Reports*, *4*, 586–602. http://dx.doi.org/10.1016/j.egyr.2018.09.002.

Saracoglu, B. O., & de Simon Martin, M. (2018). Initialization of a multiobjective evolutionary algorithms knowledge acquisition system for renewable energy power plants. *Journal of Applied Research on Industrial Engineering*, *5*(3), 185–204. http://dx.doi.org/10.22105/jarie.2018.144919.1052.

Sargent, R. G. (2010). Verification and validation of simulation models.

Sato, Y., Kawashima, H., Okuda, H., & Oku, M. (2018). Trend-based document clustering for sensitive and stable topic detection. In *22nd pacific asia conference on language, information and computation* (pp. 331–340).

Savage, W. E., & Olejniczak, A. J. (2021). Do senior faculty members produce fewer research publications than their younger colleagues? Evidence from Ph.D. granting institutions in the United States. *Scientometrics*, *126*, 4659–4686. http://dx.doi.org/10.1007/s11192-021-03957-4.

Sharma, J. (2018). Economics of decision making: exploring analytical hierarchical process (AHP). *Theoretical Economics Letters*, *8*, 3141–3156.

Shen, T., Liu, R., Bai, J., & Li, Z. (2018). Deep fakes using generative adversarial networks (GAN).

Shieh, J. I., Wu, H. H., & Huang, K. K. (2010). A DEMATEL method in identifying key success factors of hospital service quality. *Knowledge-Based Systems*, *23*(3), 277–282.

Si, S. L., You, X. Y., Liu, H. C., & Zhang, P. (2018). DEMATEL Technique: A systematic review of the state-of-the-art literature on methodologies and applications. *Mathematical Problems in Engineering*.

Simmons, G. J. (2016). Cryptology. *Encyclopedia Britannica*, https://www.britannica.com/topic/cryptology (Accessed 30 August 2021).

Singh, R. P., & Nachtnebel, H. P. (2016). Decision aid for hydropower project prioritisation in nepal by applying visual PROMETHEE. *International Journal of Multicriteria Decision Making*, *6*(4), 316–342.

Solangi, Y. A., Shah, S. A. A., Zameer, H., Ikram, M., & Saracoglu, B. O. (2019). Assessing the solar PV power project site selection in Pakistan: based on AHP-fuzzy VIKOR approach. *Environmental Science and Pollution Research*, *26*(29), 30286–30302. http://dx.doi.org/10.1007/s11356-019-06172-0.

Suleymanov, U., Kalejahi, BK., Amrahov, E., & Badirkhanli, R. (2019). Text classification for azerbaijani language using machine learning and embeddings.

Tamura, H., & Akazawa, K. (2005a). Stochastic DEMATEL for structural modeling of a complex problematique for realizing safe, secure and reliable society. *Journal of Telecommunications and Information Technology*, 139–146.

Tamura, H., & Akazawa, K. (2005b). Structural modeling and systems analysis of uneasy factors for realizing safe, secure and reliable society. *Journal of Telecommunications and Information Technology*, 64–72.

Tamura, H., Okanishi, H., & Akazawa, K. (2006). Decision support for extracting and dissolving consumers' uneasiness over foods using stochastic DEMATEL. *Journal of Telecommunications and Information Technology*, 91–95.

Tan, Q., Liu, N., & Hu, X. (2019). Deep representation learning for social network analysis. *Frontiers in Big Data*, *2*(2), http://dx.doi.org/10.3389/fdata.2019.00002.

Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, *1*(2), 146–160.

Tercan, E., & Dereli, M. A. (2020). Development of a land suitability model for citrus cultivation using GIS and multi-criteria assessment techniques in Antalya province of Turkey. *Ecological Indicators*, *117*, Article 106549.

Tercan, E., Eymen, A., Urfali, T., & Saracoglu, B. O. (2021). A sustainable framework for spatial planning of photovoltaic solar farms using GIS and multi-criteria assessment approach in Central Anatolia, Turkey. *Land Use Policy*, *102*, Article 105272. http://dx.doi.org/10.1016/j.landusepol.2020.105272.

Tercan, E., Saracoglu, B. O., Bilgilioglu, S. S., Eymen, A., & Tapkin, S. (2020). Geographic information system based investment system for photovoltaic power plants location analysis in Turkey. *Environmental Monitoring and Assessment*, *192*(297), http://dx.doi.org/10.1007/s10661-020-08267-5.

Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). *Concepts of model verification and validation, la-14167-ms*. Los Alamos National Laboratory.

Thangaraj, M., & Sivakami, M. (2018). Text classification techniques: A literature review. *Interdisciplinary Journal of Information, Knowledge, and Management*, *13*, 117–135. http://dx.doi.org/10.28945/4066.

Tournay, V., Jacomy, M., Necula, A., Leibing, A., & Blasimme, A. (2019). A new web-based big data analytics for dynamic public opinion mapping in digital networks on contested Biotechnology fields. *OMICS A Journal of Integrative Biology*, *23*, http://dx.doi.org/10.1089/omi.2019.0130.

Tseng, M. L. (2009a). A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach. *Expert Systems with Applications*, *36*(4), 7738–7748.

Tseng, M. L. (2009b). Using the extension of DEMATEL to integrate hotel service quality perceptions into a cause–effect model in uncertainty. *Expert Systems with Applications*, *36*(5), 9015–9023.

Tseng, M. L., Chen, Y. H., & Geng, Y. (2012). Integrated model of hot spring service quality perceptions under uncertainty. *Applied Soft Computing*, *12*(8), 2352–2361.

Udoye, N. E., Okokpujie, I. P., Okeniyi, J. O., Dirisu, J. O., & Ikpotokin, I. (2019). Techno-economic evaluation of coalfired power plant in South East Nigeria. a review. *International Journal of Mechanical Engineering and Technology*, *10*(3).

Vassoney, E., Mochet, A. M., & Comoglio, C. (2017). Use of multicriteria analysis (MCA) for sustainable hydropower planning and management. *Journal of Environmental Management*, *196*, 48–55.

Venturini, T., Jacomy, M., Bounegru, L., & J., Gray. (2018). Visual network exploration for data journalists. In Scott A. Eldridge, & Bob Franklin (Eds.), *The Routledge HandBook of Developments in Digital Journalism Studies*. Routledge, 9781138283053. ffhal-01672282ff.

Venturini, T., Jacomy, M., & Jensen, P. (2019). What do we see when we look at networks an introduction to visual network analysis and force-directed layouts.

Verma, P., & Goyal, R. (2021). Influence propagation based community detection in complex networks. *Machine Learning with Applications*, *3*, Article 100019. http://dx.doi.org/10.1016/j.mlwa.2020.100019.

Wan, M., & McAuley, J. (2016). Modeling ambiguity, subjectivity, and diverging viewpoints in opinion question answering systems. In *International conference on data mining (icdm)*.

Wang, L., Gao, C., Wei, J., Ma, W., Liu, R., & Vosoughi, S. (2020). An empirical survey of unsupervised text representation methods on Twitter data. arXiv preprint, arXiv:2012.03468v1.

Wang, Y., Li, J., Hongbo, S., Li, Y., Akhtar, F., & Imran, A. (2019). A survey on VV & A of large-scale simulations. *International Journal of Crowd Science*, *3*(1), 63–86. http://dx.doi.org/10.1108/IJCS-01-2019-0004, 2398–7294.

Wang, C. N., Thanh, N. V., & Su, C. C. (2019a). The study of a multicriteria decision making model for wave power plant location selection in Vietnam. *Processes*, *7*(650).

Wang, C. N., Tsai, T. T., & Huang, Y. F. (2019b). A model for optimizing location selection for biomass energy power plants. *Processes*, *7*(353).

Weißer, T., Saßmannshausen, T., Ohrndorf, D., Burggraf, P., & Wagner, J. (2020). A clustering approach for topic filtering within systematic literature reviews. *MethodsX*, *7*, Article 100831.

Welser, H. T., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, *8*(2), 1–32.

Whytock, RC., Swiezewski, J., Zwerts, J. A., Bara-Słupski, T., Pambo, AFK., Rogala, M., Bahaa-el din, L., Boekee, K., Brittain, S., Cardoso, A. W., Henschel, P., Lehmann, D., Momboua, B., Opepa, C. K., Orbell, C., Pitman, R. T., Robinson, H. S., & Abernethy, K. A. (2020). Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods Ecology Evolution*, *12*, 1080–1092. http://dx.doi.org/10.1111/2041-210X.13576.

Wikipedia (2020). List of cognitive biases. Retrieved from https://en.wikipedia.org/wiki/List_of_cognitive_biases. (Accessed January 23, 2020).

Wilhelmsson, M., & Zhao, J. (2018). Risk assessment of housing market segments: the lender's perspective. *Journal of Risk and Financial Management*, *11*(4), 69.

Wu, H. H., & Tsai, Y. N. (2011). A DEMATEL method to evaluate the causal relations among the criteria in auto spare parts industry. *Applied Mathematics and Computation*, *218*(5), 2334–2342.

Wu, L., Zhang, Q., Shan, L., & Chen, Z. (2013). Identifying critical factors influencing the use of additives by food enterprises in China. *Food Control*, *31*(2), 425–432.

Xu, J., Xu, B., Wang, P., Zheng, S., Tian, G., Zhao, J., & Xu, B. (2017). Self-Taught convolutional neural networks for short text clustering. *Neural Networks*, *88*, 22–31.

Yavanoglu, O., & Aydos, M. (2017). A review on cyber security datasets for machine learning algorithms. http://dx.doi.org/10.1109/BigData.2017.8258167.

Yeh, T. M., & Huang, Y. L. (2014). Factors in determining wind farm location: integrating GQM, fuzzy DEMATEL, and ANP. *Renewable Energy*, *66*, 159–169.

Yi, J., Zhang, Y., Zhao, X., & Wan, J. (2017). A novel text clustering approach using deep-learning vocabulary network. *Mathematical Problems in Engineering*, Article 8310934. http://dx.doi.org/10.1155/2017/8310934.

Yohannes, D., & Assabie, Y. (2021). Amharic text clustering using encyclopedic knowledge with neural word embedding. In *The 2nd workshop on african natural language processing (africanlp)*.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining an introduction*. Cambridge University Press.

Zhang, W., Dong, C., Yin, J., & Wang, J. (2021). Attentive representation learning with adversarial training for short text clustering. arXiv:1912.03720v2.

Zhang, C., Li, R., Xia, Y., Yuan, Y., Dincer, H., & Yuksel, S. (2020). Analysis of environmental activities for developing public health investments and policies: a comparative study with structure equation and interval type 2 fuzzy hybrid models. *International Journal of Environmental Research and Public Health*, *17*(6), 1977.

Zhou, Q., Huang, W., & Zhang, Y. (2011). Identifying critical success factors in emergency management using a fuzzy DEMATEL method. *Safety Science*, *49*(2), 243–252.

Zinkevich, M. (2021). Machine learning guides rules of machine learning, best practices for ML engineering. https://developers.google.com/machine-learning/guides/rules-of-ml (Accessed 30 August 2021).

Zolfani, H. S., Yazdani, M., Torkayesh, E. A., & Derakhti, A. (2020). Application of a gray-based decision support framework for location selection of a temporary hospital during COVID-19 Pandemic. *Symmetry*, *12*, 886.