A Bibliometric Analysis of Articles Citing the Machine Learning Algorithm

Submitted By
Md Imranul Islam
ID: 141-35-607
Department of Software Engineering
Daffodil International University

This thesis submitted in fulfillment of the requirement for the degree of Bachelor of Science in Software Engineering

Supervised By
Dr. Imran Mahmud
Assistant Professor & Associate Director (Research)
Department of Software Engineering
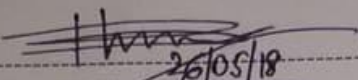Daffodil International University

Department of Software Engineering
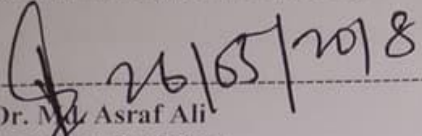
Daffodil International University

Spring – 2018

APPROVAL

This **Thesis** titled "**A Bibliometric Analysis of Articles Citing the Machine Learning Algorithm**", submitted by **Md Imranul Islam, ID:141-35-607** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Software Engineering and approved as to its style and contents.
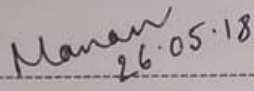
BOARD OF EXAMINERS

Dr. Touhid Bhuiyan
Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Chairman

Dr. Md. Asraf Ali
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Manan Binth Taj Noor
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Dr. Md. Nasim Akhtar
Professor
Department of Computer Science and Engineering
Faculty of Electrical and Electronic Engineering
Dhaka University of Engineering & Technology, Gazipur

External Examiner

DECLARATION

I hereby declare, that the thesis entitled "**A Bibliometric Analysis of Articles Citing The Machine Learning Algorithm"** written and submitted by us. This is my original unique work based on our individual a research. I have collected information from recent research paper on machine learning algorithm. In this case our honorable supervisor **Dr. Imran Mahmud** helped me a lot.

I also declare that neither this report nor any part of this report has been submitted elsewhere for award of any degree.

....

Md Imranul Islam

ID: 141-35-607

Batch : 13th

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

Certified By:

…………………….:

Name: Dr. Imran Mahmud

Assistant Professor & Associate Director (Research)

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

ACKNOWLEDGEMENT

I fell grateful to and wish our profound our indebtedness to **Dr. Imran Mahmud, Assistant Professor & Associate Director (Research)**, Department of SWE, Faculty of Science and Information of technology, Daffodil International University. Deep Knowledge & keen interest of my supervisor in the field of Machine Learning Algorithm, Systemic review me to carry out this project .His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan, Professor and Head,** Department of SWE, Faculty of Science and Information of technology, Daffodil International University, for his kind help to finish my thesis and also to other faculty member and the staff of SWE department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of my parent

Table of Contents

ABSTRACT

In the face of comparatively recent emergence of the machine learning algorithm, the originating article has already cited by a large number of studies, and hence it appears to have become a popular theoretical and practical choice with in the field of Machine Learning Algorithm. However as yet there have been no attempts to analyses the reason for citing the originating article . Choosing a suitable machine learning algorithm for a system is difficult because of the number of algorithms described in the literature..Such as systemic review of citations may inform researchers and guide appropriate future use of the theory .This paper therefore presents the results of a bibliometric analysis and systematic review of fifty citation of the originating article in an attempt to better understand the reason for citation, and use and adaptation of the theory . This paper presents a structured review of the literature that explore the use of machine learning algorithms in bibliometric analysis and identify new research scope. The object of this study are to define trends in the use or research of machine learning algorithms; Define clear questions in the use or research of machine learning algorithms; and helps new researchers to position new research work in this state appropriately. This paper also classifier and speculate these citation and analyses the limitation of Machine Learning Algorithm use in existing research, Speculate the use of big data technologies, Define types of machine learning algorithms. Overall, the topic summarizes the current state of knowledge of the machine learning algorithm. It creates an understanding of the machine learning algorithm for the reader by discussing the findings presented in recent research papers.

LIST OF KEYWORD

1. Machine Learning,
2. Machine Learning Algorithm ,
3. Bibliometric analysis ,
4. Big Data,
5. Systemic Review

LIST OF ACRONYMS

1. ML : Machine Learning
2. SVM : Support Vector Machine
3. DT: Decision Tree

4. SVM-RF: Support Vector Machine Random Forest

5. NB: Naïve Bayes

6. K-NN: K-Nearest Neighbor

7. ANN: Artificial Nural Network

8. LR : Logistic Regression

LIST OF FIGURE

LIST OF TABLE

CHAPTER 1: INTRODUCTION

Bibliometric methods have been used to identify relationships amongst academic journal citations. Data from citation indexes can be analyzed to determine the popularity and impact of specific author, article, and publications. Machine Learning algorithms can be initially divided as supervised, unsupervised, semi-supervised and reinforcement learning. It is worth calculating the number in each class in this systematic review and bibliometric analysis. However, the ML field does not have a clear classification design for its algorithms mainly. Because of the number of approaches and the alternative proposed in the literature (Lv & Tang, 2011). As a result, it tern complex and confusing to pick an Machine Learning algorithm that it's one's requirement when developing an Algorithm. In addition, researchers may discover it challenging to pathway the use and the trends of Machine Learning algorithms. This paper take measures a systematic review to supervise how Machine Learning algorithms used in Biliometric analysis are studied and used, and what are the object in Machine Learning algorithm research and development. It is desired that, with this systematic review, researchers and practitioners can reach more information about the Machine Learning algorithm related field, and make better execution or research decisions.

2.1 Motivation of the Research

I am working with Machine Learning Algorithm. I face some difficulties to find out with specific area based algorithm. Every researcher face those same problem. It motivates me to do the research for Bibliometric Analysis and Structured review or Systemic review on Machine Learning Algorithm.

2.2 Problem Statement

When we search in google or google scholar to find out specific Machine Learning Algorithm based paper and to know the research trends on Machine Learning Algorithm it's difficult to every researchers.

> ➤ If we search in google scholar, there would be 24,20,000 results.
> ➤ If we just search in google 10,20,00,000 results
> ➤ It is difficult to find out area based, specific algorithm based papers.

2.3 Research Objectives

This papers aims to present a citation analysis and systematic review of citations of the originating article of the ML Algorithm in order to illustrate the reasons for citation, and reveal variations in use and theoretical advancement. The overall the topic summarizes the current state of knowledge of the machine learning algorithm.

- Define the object in the use or research of machine learning algorithms.
- Identify open questions in the use or research of ML algorithms.
- Assist new researchers to position new research work in this state appropriately.
- To create a database based on year, author, country, journal, area, algorithm type, findings, limitation and future research.
- To create a software to facilitate previous objective
- It creates an understanding of the machine learning algorithm for the reader by discussing the findings presented in recent research papers.

The remaining sections of this chapter are organized as follows. The following section outlines the method utilized to undertake the analysis presented in this chapter. The Background Statement   are then presented in chapter 2, Methodology and Development tools in chapter 3 and discussed in chapter 4. Finally, the salient points of the Conclusion and Future Studies in chapter 5, which also presents the limitations and future research directions arising from this study.

CHAPTER 2: BACKGRAUND STATEMENT

I have studied lots of books existing papers related to our thesis topic. I tried to understand what are exist in present ML Algorithm. I cover the topic Bibliometric Analysis, Systemic Review or Structured Review and recent research train on Machine Learning Algorithm.

2.1 Machine Learning Algorithm

Machine Learning uses computers to assume human learning and simulate computers to identify and gain knowledge from the true world, and improve achievement of some tasks based on this new knowledge. More officially, Machine Learning is defined as follows: "A computer program is expressed to learn from experience E with important to some square of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (R. Michalski, Carbonell, & Mitchell, 1985).Although the first concepts of Machine Learning risen in the 1950s, Machine Learning was studied as a different field in the 1990s (R. S. Michalski, Carbonell, & Mitchell, 2013).

Today, Machine Learning algorithms are used in several areas besides computer science, including business, advertising and medicine. Learning is the procedure of knowledge acquisition. Humans naturally learn from experience or sense because of their power to reason.

In addition, computers do not learn by reasoning or sensing, but learn with algorithms. Today, there are a huge number of Machine Learning algorithms proposed in the literature. They can be classified based on the way used for the learning process or method. There are four main classifications: supervised, semi supervised, unsupervised, and reinforcement learning.

Supervised learning occurs when algorithms are provided with training data and right answers. The task of the Machine Learning algorithm is to learn based on the training data, and to apply the knowledge that was obtained using real data. As an example consider an Machine learning algorithm being used for food classification in a restaurant. A training set (training data + answers) can be a table relating information about each food to a correct classification. Here, information about each food may be name, test, or even every word a food contains. The Machine Learning algorithm learns with the training set. When a new food arrives at the restaurant, the algorithm can classify it based on the knowledge about book classification it has acquired.

Machine Learning algorithms can also be classified as semi-supervised. Semi-supervised learning happens when algorithms work with a training set with absence information, and still need to learn from it. An example is when an Machine Learning algorithm is provided with movie ratings. Every user not rated every movie and so, there is some absence information. Semi-supervised learning algorithms are able to learn and describe conclusions even with incomplete data.

In unsupervised learning ML algorithms do not have any training set. They are presented with some data about the true world and have to get knowledge from that data on their own. Unsupervised learning algorithms are mainly focused on finding hidden patterns in data. An example, An Machine Learning algorithm has get access to user profile information in a social network or social media. By using an unsupervised learning approach, the algorithm can several users into personality class, such as emerging and stock, allowing the social network company to target notice more straight way at specific groups of users.

Machine Learning has become completely popular recently with the increase in memory size and processor speed. As a result, the field now has a huge number of algorithms that infer data, statistical analysis to learn or use mathematical, draw conclusions. This number speared to increase as show by the number of scientific publications that propose combinations or variations of Machine Learning algorithms. As a result, MLA have been classified based on the purpose for which they are designed. few examples of classification can be found in (Shalev-Shwartz & Ben-David, 2013) and (Kulkarni, 2012), although the field still does not have any proper classifications.

2.2 Systemic Review

Systematic reviews is collect and critically analyze multiple research papers or studies, using methods that are selected before one or more research questions are formulated, and then analyzing and finding studies that relate to and answer those questions in a structured methodology. The protocol for this systematic review has main three main methods. The first method is to collect as lot of publications as possible using scientific search engines. The authors then attach initial elimination criteria and explore the studies that were recovered. The second method is attach an extra exclusion criteria and to study the summary of the remaining papers. The last method is to attach a third set of exclusion criteria or to study the entire study and gather data from it.

2.3 Bibliometric Analysis

Bibliometrics method is statistical analysis of written publications, such as article or book or journal. Bibliometric systems are often used in the field of information science and library, including scientometrics. Bibliometric methods have been used to trace relationships amongst academic Citation analysis, journal citations. Which are contract examining an item's searching documents, is analyzing their merit and used in searching for implement. Citation indices, such as Institute for Scientific Information's Web of Science, "allow users to search forward in time from a known article to more recent publications which cite the known item Data from citation indexes can be analyzed to determine the popularity and impact of specific publication, authors, and article". Using citation analysis to gauge the importance of one's work, for example, is a important part of the tenure review steps.

CHAPTER 3: RESEARCH METHODOLOGY

A summation of citation analysis and systematic review (Lee et al. 2003; Legris et al. 2003) was considered an appropriate research method to accomplish the aim of this research. A total of fifty citations of Machine Learning Algorithm originating article were identified by employing the academic journals database provided by the Thomson Scientific Web of Sciences product. The database was used to extract demographic data (including author's year, name, journal and country of publication, Keyword, Process, Limitation, feature studies) related to all cited studies. Following the approach, in order to explore the topics most frequently examined using ML Algorithm, keywords for 50 studies citing the originating article. I collect the information form fifty paper on machine learning by using Biblometric Analysis, Systemic review, Structured review method. Appropriate algorithm for appropriate type of data, Year wise algorithm development, An opportunity to visualize the research background of Machine Learning Algorithm by using those method and to create a development tools to facilitate previous objective.

3.1 Findings

The total finding of this paper show the Demographic data citation by year , Demographic data citation by area based Algorithm, Demographic data citation by algorithm, Citation by area based limitations and future studies, Demographic data citation by year wise algorithm development ,Country wise algorithm development, An opportunity to visualize the research background of Machine Learning Algorithm.

3.1.1 Demographic Data Citations by Year:

Our findings in Table 1 indicate that citations of the originating article have constantly increased since 2015. Table 1 suggests that the originating article has quickly gained acceptance and popularity amongst ML algorithm researchers. Therefore, it would be interesting and useful to explore the reasons for and nature of its citations. The analyses and findings presented here after attempt to throw light on such issues.

Table 1 : Citation by Year

Publication Year	Number of Paper	%
2015	2	4%

2016	11	22%
2017	33	66%
2018	4	8%

Figure 1 show the data of 2015~2018 year wise algorithm. This figure give the information that the 33 paper in 2017 total 50 paper .
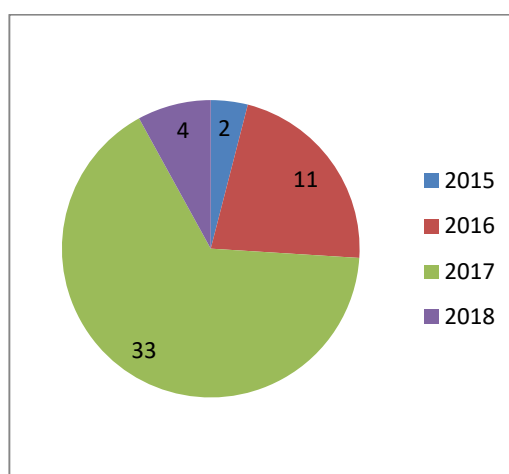
Figure 1 Citation By Year

3.1.2 Demographic Data: Citations by Area based Algorithm

 The table of 2 finding the citation show the data of area based algorithm .This table helps to find out the which area/data are most used in which algorithm . Extreme Learning, Big data, optical distance sensor area mostly used SVM, K-NN, Regression Algorithm.

Table 2 Area Based Algorithm

Area	Process Algorithm	Number of Paper	%
Motion Picture forecast in Box Office	non regression Algorithm, Forecasting Algorithm	1	2%

A computer-aided Diagnostic System	Stepwise regression algorithm	1	2%
computational methods in drug discovery	SVM-RF algorithm	1	2%
optical distance sensor	ECVFL	2	4%
Breast Cancer	SVM, Decision Tree (C4.5), N) and k-NN	1	2%
Big Data	ML-KELM Algorithm, Santander Bank Dataset, single-pass predictive approach, Multi scan machine learning Algorithm, SVM,	6	12%
Extreme learning	K-NN, EML, TL algorithms	3	6%
Materials discovery and design	machine learning in materials science	1	2%
multi-criteria decision making methodology	AMD methodology	1	2%
A computer-aided Diagnostic System	Stepwise regression algorithm	1	2%
Automatic detection	LR, SVM, DT	2	4%
medical wireless sensor and Data	DT, k-NN, Random Forests ,LR,M5P,Additive Regression (Decision Stump),RRA,,LASS	2	4%
Separation of pulsar signals from noise	ANNMLP, Adaboost, GBC, XGBoost	1	4%
Data sparsity problem	SGD, SVD	1	2%

Rainfall prediction in weather derivatives	Genetic Programming, SVM, Radial Basis Neural Networks,M5,M5 Model trees, and k-NN	1	2%
Heart rate dynamics and serum proteomics	linear and nonlinear HRV, SVM	1	2%
bus travel speeds	Artificial Neural Network , SVM , Bayes Network	1	2%
Diabetes Research	Support vector machines (SVM), data mining techniques	1	2%
Internal loop of DNA and Brain maturity	molecular machine learning, SVM	1	2%
software project effort and duration estimation	SVM, Neural Networks and Generalized Linear Models	1	2%
Other(Different area)	Tree [DT],k-Nearest Neighbor [k-NN], Naïve Bayes [NB], and Random Forests [RF]),Logistic Regression, Gaussian, Naive Bayes	20	40%

3.1.3 Demographic Data: Citations by Algorithm :

Our table 3 show the data of citation by algorithm . The table indicate the most common used algorithm of 50 citation. SVM used 23 paper of 50 and 46% , Regression Algorithm 18 paper of 50 and 36%, other mostly common used algorithm is ANN, K-NN, ELM , Random Forecast

.

Table 3 Citation by Algorithm

Algorithm	Number of paper	%
SVM	23	46%
Non Regression	2	4%
Regression Algorithm	18	36%
ML-KELM	1	2%
ANN	7	14%
K-NN	5	10%
M5	4	8%
Decision Tree	11	22%
Naive Bayes	3	6%
Random forest	9	18%
Radial Basis Neural Networks	1	2%
linear and nonlinear HRV	1	2%
ELM	6	12%
TL	4	8%
Multi scan machine learning Algorithm	1	2%
lazy learning algorithms	1	2%
TELM	1	2%

Figure 2 we get the information of MLA SVM, Regression Algorithm , Decision Tree, Random Forecast are most popular algorithm in recent research paper .
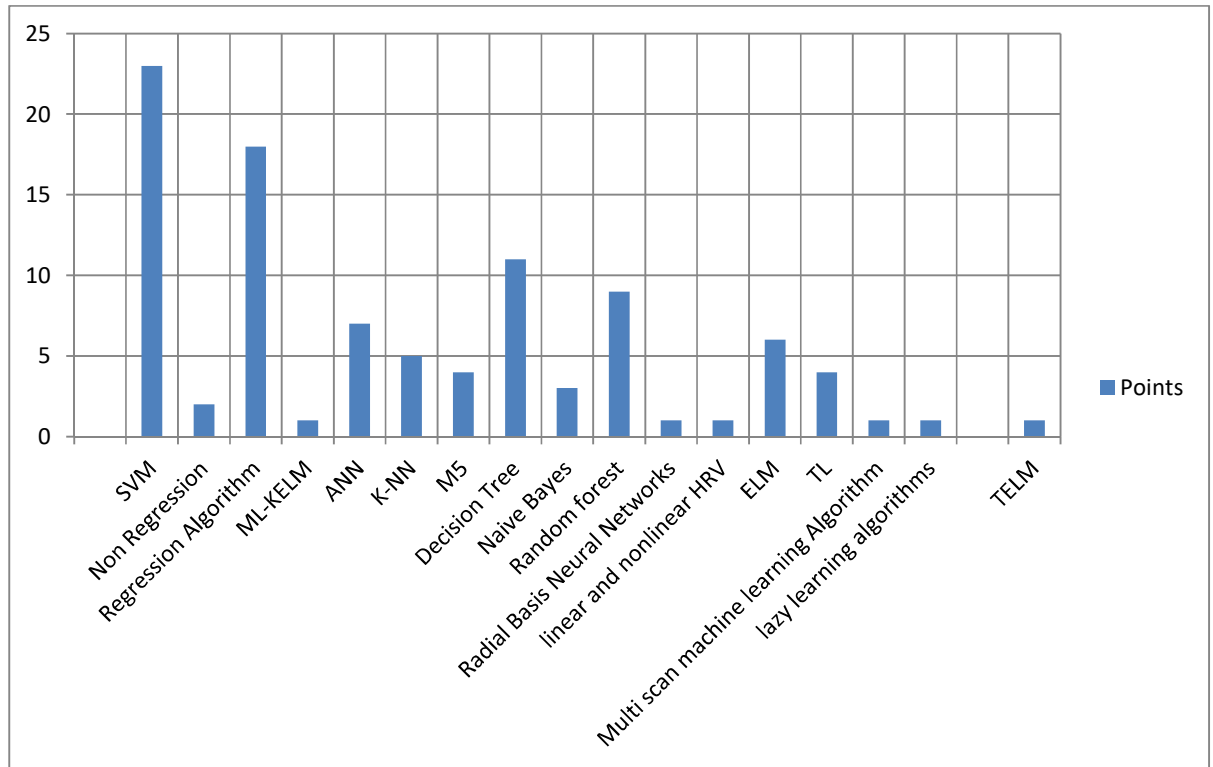

Figure 2 Citation by Algorithm

3.1.4 Demographic Data: Citations by Algorithm based Variable, Limitation and Future Studies:

Our finding of table 3.4 variable /area based limitation and future studies . Big Data are mostly used of area in future studies .This table helps to researcher to find the scope of new studies.

Table 4 Citation by limitation and future studies

Variable	Limitation	Future Studies
Big Data	Iterative learning operations avoided	Design a ML-KELM algorithm with optimal kernel parameters by sensitivity analysis of parameter settings
	biased results	more experiments will be conducted using several datasets in order to avoid biased result

	ML algorithms for RSs in different application domains	investigation of Big Data technologies, which offer a wide variety of methods to support the storage and analysis of massive data.
Focus Lens (ECVFL)	larger target range	Deploying the sensor over a larger target range
Image processing	investigating other raisin varieties such as Sun-dried and Sultana raisins	on color features need to optimize the feature extraction algorithm and increase classification accuracy of higher mixtures
Rainfall	Disconnecting the rainfall time series	time series rainfall and modifying the tuning procedure to produce a parameter setting customized to a data set climate
missing data	parameters have constant values	to change the values of parameters (learning rate, regularization) in each iteration of optimization algorithm
Frog calls	A bad anti noise ability	Modify MFCCS to improve anti noise ability
Radio frequency identification (RFID) Reading	Radio frequency identification (RFID) Reading	more data samples will be collected so as to cover as many variants of RFID readings as possible
ELM technique	ELM-TL Literature failed to report results from repeated experiments	ELM based transfer learning algorithms are credible, stable and reproducible

FOT parameters	understanding and management of asthma exacerbations by integrating ML algorithms and home monitoring using telemedicine systems based on FOT measurements	applying interpretable models to make explaining the diagnosis achieved by the classifiers easier
IF aggregation operators	proposed measure to only one dataset (clustering problem)	the IFS-IBA similarity measure with the expert given logic-based aggregation function
online dataset	No data is collected for Experiment	A larger blogger data-set covering cultural ,regional ,demographic information which remain hard to find online
radio frequency interference (RFI)	not possible to comparatively rank these algorithms on the basis of the above performance scores	apply other ML-based post-processing pipe lines such as PICS on this same data set
community wireless mesh networks (CWMNs)	online machine learning algorithms to make link quality predictions in CWMNs has not been explored yet.	the hybrid online machine learning algorithm will be evaluated using data of other CWMNs with a higher number of links and a higher percentage of variable links
ISBSG dataset, smart data preparation	performance in practice is dependent on the quality of input data	built by utilization of a cross industry ISBSG dataset which contained up to date software projects which occurred in the last decade within multiple geographically distributed organizations .

3.1.5 Demographic Data: Citations by Year based Algorithm Development

Our findings in Table 5 indicate that citations of the originating article have constantly increased since 2015. This table show the data year based ML algorithm development .SVM is most use in every year in research paper. My 50 citation 2016 SVM used 4%, 2016 SVM used 12%, 2017 SVM used 28%, 2018 used 4% , 48% of of 50 studies . Other popular algorithm is K-NN,DT, M5, Regression , Artificial neural network.

Table 5 Year Based Algorithm

Year	Algorithm	%
2015	Decision Tree ,k-NN, Random Forests ,Linear Regression ,M5P,Additive Regression (Decision Stump) RRA, least absolute shrinkage and selection operator algorithm (LASSO)	4%
2016	SVM-RF,SVM , Data mining, SVR	12%
	k-NN ,K-NN,KNN	6%
	Decision Tree (C4.5),	2%
	Naive Bayes (NB) ,	4%
	AMD methodology	2%
	Linear Discriminant analysis,	2%
	Random forest	4%
	,Artificial neural network	2%
	Multi scan machine learning Algorithm	2%
	Sparse coding ELM (ScELM) algorithm	2%
	Logistic Regression,	2%
	Gaussian ,	2%
2017	Regression Algorithm,	12%
	Forecasting Algorithm	2%

	ML-KELM Algorithm	2%
	Santander Bank Dataset	2%
	ANN	8%
	SVM	28%
	Genetic Programming	2%
	M5	6%
	K-NN	8%
	stochastic gradient descent (SGD)	2%
	SVD	2%
	DT	8%
	Random Forecast	8%
	TL Algorithm	2%
	ELM	10%
	ADAB	2%
	neural networks (NN)	4%
	Gaussian Process	2%
2018	MLFFNN	2%
	SVM	4%
	ANNMLP	2%
	K-NN	2%
	NB and RF	2%

This figure shows the data 2014 year wise algorithm. All of algorithm all most same .
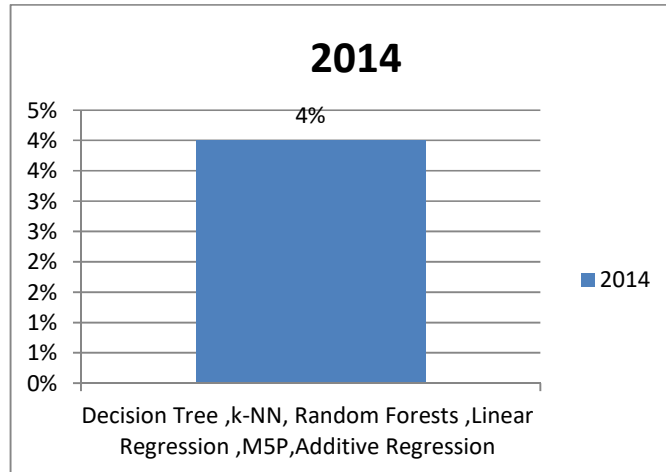

Figure 3 Citation by 2014 year wise algorithm

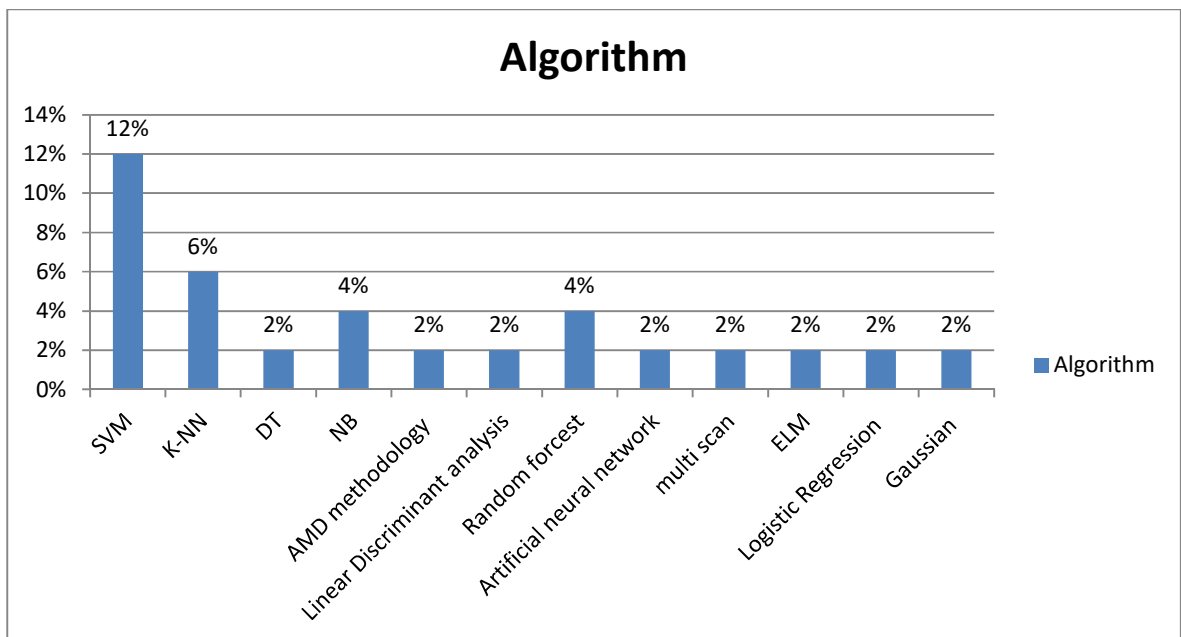This figure show the Data 2016 year wise algorithm . SVM , K-NN are mostly used.

Figure 4 Citation by 2015 year wise algorithm

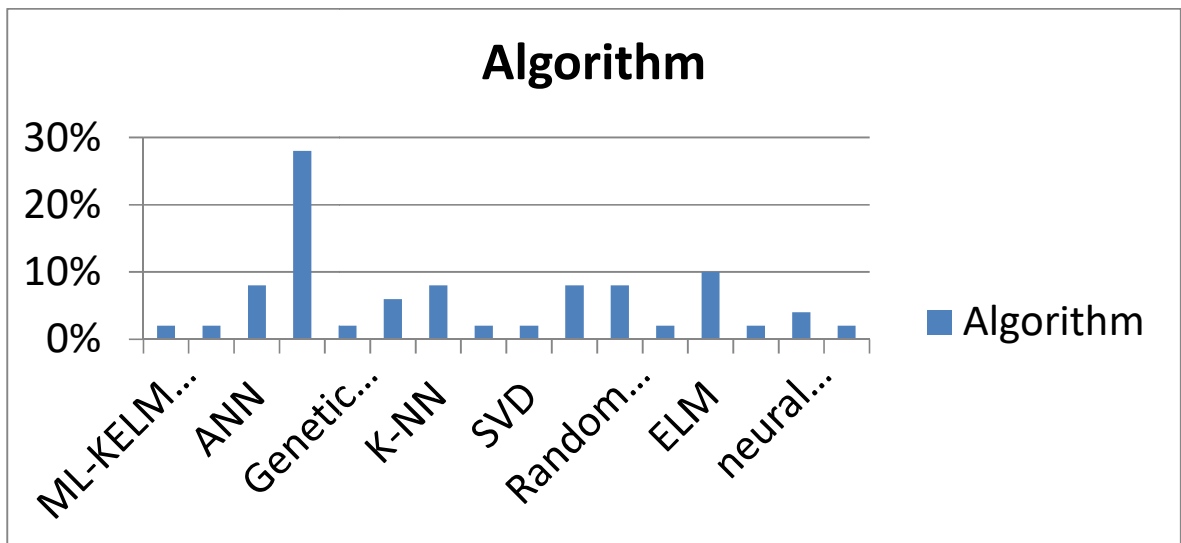This figure show the data 2017 year wise algorithm . SVM , K-NN , Regression, ELM are mostly used .

Figure 5 Citation by 2017 year wise algorithm

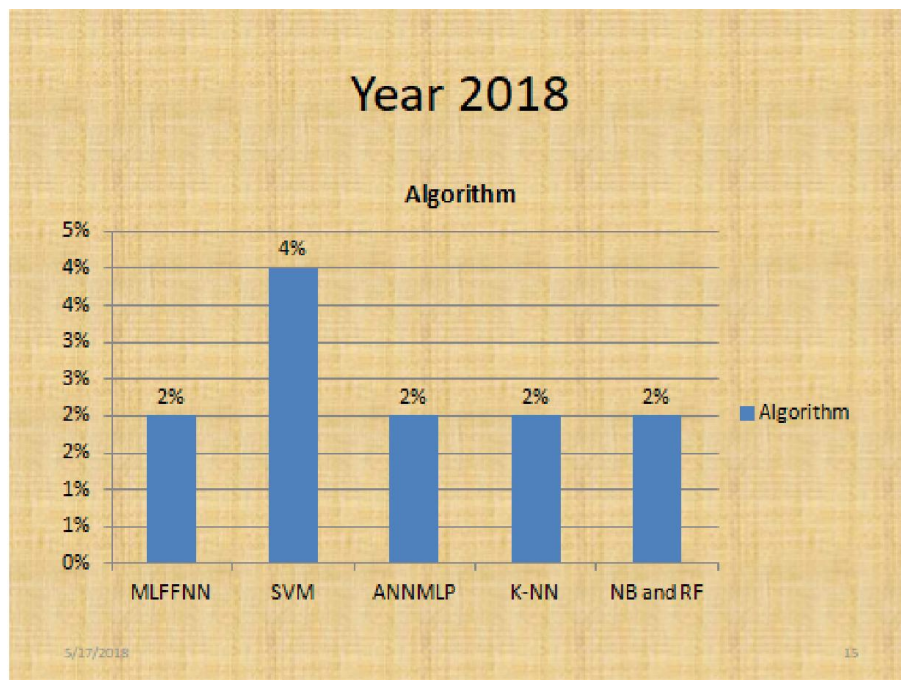This figure show the data 2016 year wise algorithm . SVM , K-NN are mostly used .

Figure 6 Citation by 2018 year wise algorithm

3.1.6 Demographic Data: Citations by Country based Algorithm Development

Our findings in Table 3.6 indicate that citations of the originating article have Country wise citation. This table show the data of 50 citation korea and china are 24% and Iran 8 % , India 10% other country 56%.

Table 6 Citation By Country wise Algorithm Development

Country	Number of Paper	%
Korea	6	12%
Greece	3	6%
Iran	5	10%
Ireland	1	2%
Morocco	1	2%
China	6	12%
Jordan	1	2%
England	1	2%
Australia	2	4%
Norway	1	2%
India	4	8%
Servia	2	2%
Brazil	2	4%
Pakistan	1	2%
UK	2	4%
Croatia	1	4%
Canada	2	4%

Spain	1	2%
Japan	1	2%
New Zealand	1	2%
Turkey	1	2%
USA	4	8%
Poland	1	2%

This figure show the data country wise algorithm development . Korea , Iran, Norway are most paper on MLA
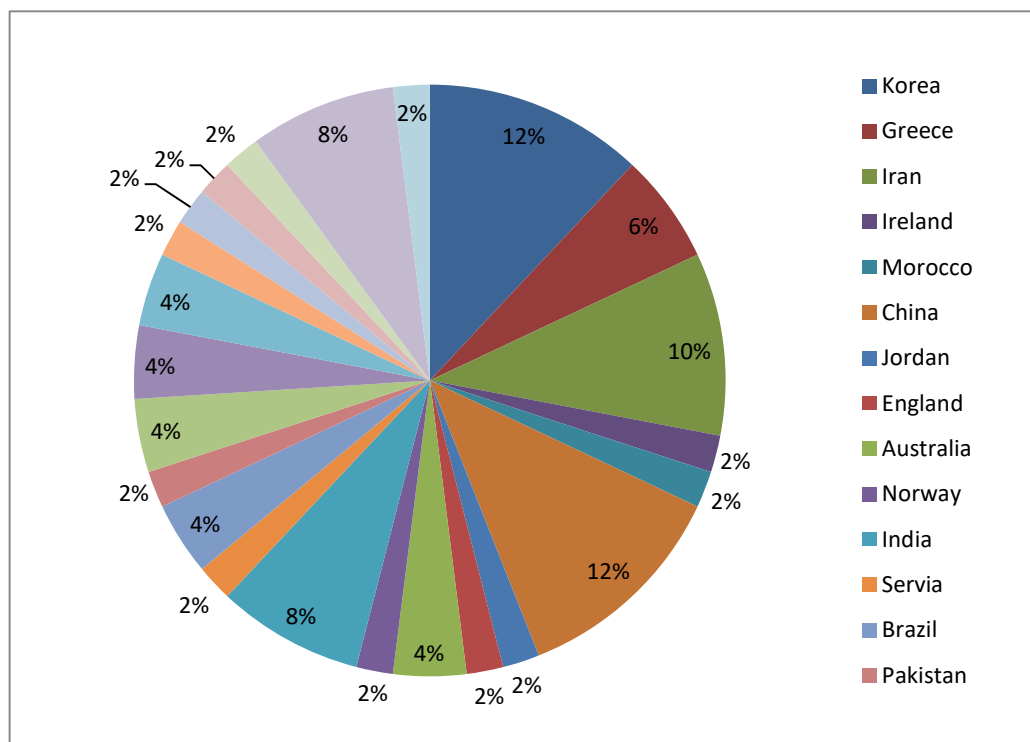

Figure 7 Citation By Country

3.1.7 Keyword Analysis :

The most frequently used keywords from all 50 studies which have cited or used all or part of Machine Learning. Our findings indicate that the keywords "Technology Machine Learing" are most used, appearing 43 times, followed by the keyword "SVM" / Machine learning Algorithm", being used in 27 studies .Here is all Key work of 50 studies Box office earning forecast, Social network service, Machine learning, Genetic algorithm, Forecast combination, Fibrosis, Shear wave elastography, Computer aided diagnosis, Ultrasonic, Classifier design, machine learning; accuracy, Drug target, Drug gable , Features, Optical distance sensing, Electronic lens, Machine learning, Polynomial regression, Regularization, Breast cancer, SVM ,NB ,C4.5 ,k-NN ,Classification, Efficiency, Effectiveness, Multilevel learning , extreme learning machine, kernel extreme learning machine, algorithm recommendation, algorithm selection; classification algorithms; classifiers recommendation, TOPSIS; Ranking classifiers, Big Data, H2O,Sparkling Water, Prediction, Spark, Santander Bank Dataset, Image processing, Golden Bleached Raisin (GBR), Bulk Textures, Textural Features, SVM & ANN, Weather derivatives Rainfall Machine learning, Data sparsity, Spectral Co-Clustering; Tensor Factorization; Collaborative filtering, Frog call classification, Sounds cape ecology , Bioacoustics, RFID Machine learning Classification False positive readings, Wireless Sensor Networks; Machine Learning Algorithms; Sensor Faults; Healthcare and patient monitoring, white rice, adulteration, discrimination, machine learning, Major depressive disorder, biomarker, proteomics, heart rate variability, machine-learning, Trajectory prediction, Real-time query response Data streams, Machine learning, Transfer learning, extreme learning machine, Clinical decision support, classification, machine learning, airway obstruction severity, forced oscillation technique, asthma, intuitionistic fuzzy sets, interpolative Boolean algebra, similarity measure, classification, clustering, Wind speed forecasting , Extreme learning machine, Outlier correction method, Twin extreme learning machine, Pattern classification, Extreme learning machine, Twin support vector machine, Sparse coding, extreme learning machine, gradient projection, Machine learning, Blogging ,Decision tree algorithm, Lazy learning algorithm, Assembling techniques ,Classification, Natural Gas, Henry Hub, Machine learning, Feature selection algorithm, Support Vector Regression Machines, Neural networks, Public transport , machine learning , Speed Prediction, Child Welfare; Analytics; Machine Learning, Machine Learning; Data Mining; Diabetes mellitus; Diabetic complications; Disease prediction models; Biomarker(s) identification, Optical, distance sensing , Electronic lens, Machine learning, Polynomial regression, Regularization, Biomarkers , Inflammation, Machine learning method, Microbial Infection, New materials discovery; Materials design; Materials properties prediction; Machine learning, Street design features, Enclosure Walk

Page 24

ability, Machine learning, Thyroid, Ridge regression, Machine learning Regularization method, Bimolecular, computation, Hyper network Machine learning Classification, developmental neuroimaging, functional MRI, infant, prematurity, multivariate pattern analysis, systematic review of the literature, recommender systems, machine learning, machine learning algorithms, application domains, performance, DRASTIC, Groundwater contamination risk, Machine learning model, Multi-mode, Machine Learning, Internet of Things, Smart Data, Smart City, Software project estimation, machine learning, effort and duration estimation, ensemble models, ISBSG.

3.2 Development Tools:

I create a web application based on year, author, country, journal, area, algorithm type, findings, limitation and future research. To create this application to facilitate previous objective. The purpose of this application is to review recent progress in a machine learning algorithm by searching. Overall, the application gives us to view summarizes the current state of knowledge of the machine learning algorithm, easily find out the recent research tends, easily view the new research scope . It creates an understanding of the machine learning algorithm for the reader by discussing the findings presented in recent research papers.

3.2.1 Software Language or Framework

I used C# language and asp.net MVC framework to build this application. Client site used Java script.

1. C# language

2. Asp.net MVC framework

3. Java Script

ASP.NET reduces the amount of code required to make big applications. With built-in Windows authentication and per-application configuration, my applications are safe and secured. It provides good performance by taking advantage of early binding, just-in-time compilation, caching services right out of the box and native optimization.

3.2.2 Database Design or Architecture

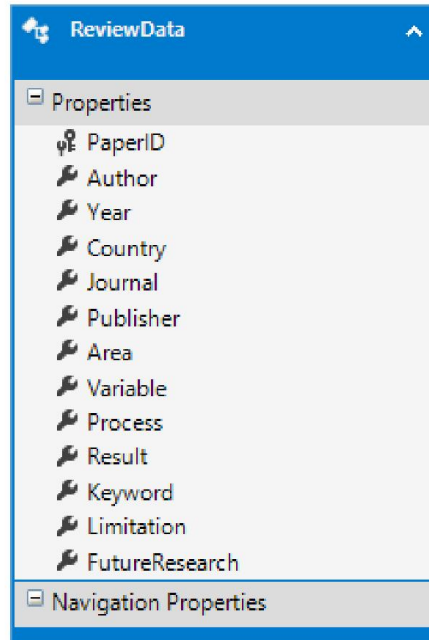This is application Database design of this application . only one database used this application.

Figure 8 Database Digram

3.2.3 User Interface Technology

User Interface I used those technology .This is web application.

- ➤ ASP.NET MVC Framework
- ➤ Html ,CSS
- ➤ Bootstrap
- ➤ J query

3.2.4 Implementation Tools & Platforms

I used those Implementation tools and this is web application

- ➢ Microsoft Visual Studio 2015
- ➢ SQL Server 2014 Management Studio
- ➢ . NET Runtime

3.2.5 Some User Interface

Any User can add data on basis of research paper . and search by using multiple keyword. User can edit, delete any data .

3.2.5.1 Add Data Page

This page show the add data interface. Any user can add data by Year , Author , Country , Journal , Process, Result, Limitation , Keyword, future Studies

3.2.5.2 View Page

User can search multiple key searching and visualization the recent research progress and total overview on machine learning Algorithm related paper

Machine Learning Algorithm Paper Review

+ Add New

Show 10 ▼ entries   Search: [ ]

PaperID	Athor	Year	Country	Journal	Publiser	Area	Variable	Process	Result	Keyword	Limitation
1	Taegu Kim	2017	Korea	Elsevier	International Journal Forecasting	Motion Picture forcast in Box Office	Earning, screening information	non regration Algorithm, Forecusting Algorithm	A new approch to forcasting the the box office earning of motion picture is proposed	Box office earning forcast , Social network service , Mchine learning , Genetic algorithm , Forecast combination	The number of seats is determined shoely by exhibitors
6	Lias Gatus	2017	Greece	Elsevier	World Federation for Ultrasound in Medicine & Biology	A computer-aided Diagnostic System	Color based Interaction	Stepwise regression algorithm	Parameters and criteria for CLD diagnosis employing SWE images	Fibrosis , Shear wave elastography , Computer aided diagnosis, Ultrasonics, Classifier design	Specific protocol include the relatively small number of individuals investigated and the fact that obese patient and

3.2.5.3 Search , Edit, Delete Page

User can search , edit , Delete any paper content . User can update database by import data.

3.2.5.4 Filtering Option

User can search by special filtering option. user can search multiple keyword.

PaperID	Athor	Year	Country	Journal	Publiser	Area	Variable	Process	Result
6017	Zingjing Jhang	2017	UK	Elsevier	International society of nephrology	Peritoneal dialysis patient with a bacterial infections	Peritoneal dialysis patients	Non linear mathematical model	Utilizing machine learning algorithm ,demonstrate that different groups of bacteria induce qualitatives distinct local immune fingerprint .

3.2.6 Summery :

Development this application to know current trend of machine learning algorithm. Any researcher can add data by review machine learning algorithm related paper. Everyone see this application to know Year wise ML Algorithm development, Country wise ML Algorithm, Limitation and future studies.

CHAPTER 4: RESULT ANALYSIS

Fifty citation of this paper show the data approximately, current trend of machine learning algorithm, year wise algorithm development , county wise algorithm development , area based algorithm , future studies and limitation of 50 studies . Those citations of the originating article have constantly increased since 2015.

Suggests that the originating article has quickly gained acceptance and popularity amongst ML algorithm researchers. Therefore, it would be interesting and useful to explore the reasons for and nature of its citations. The analyses and findings presented hereafter attempt to throw light on such issues. The citation show the data of area based algorithm and helps to find out the which area/data are most used in which algorithm .

Mostly used big data and SVM algorithm of the 50 studies Show the data of citation by algorithm and indicate the most common used algorithm of 50 citations. SVM used 23 paper of 50 and 46%, Regression Algorithm 18 paper of 50 and 36%, other mostly common used algorithm is ANN, K-NN, ELM, Random Forecast. Big Data are mostly used of area in future studies.

This paper helps to researcher to find the scope of new studies. This paper show the data year based ML algorithm development .SVM   is most use in every year in research paper. My 50 citation 2016 SVM used 4%, 2016 SVM used 12%, 2017 SVM used 28%, 2018 used 4% , 48% of 50 studies . Other popular algorithm is K-NN,DT, M5, Regression   , Artificial neural network .and indicate that citations of the originating article have Country wise citation. This table show the data of 50 citation korea and china are 24% and Iran 8 % , India 10% other country 56%.

CHAPTER 5: CONCLUSION AND FUTURE STUDIES

The following salient points emerged from the findings and discussion presented in our study .The majority of articles that cited ML Algorithm have done so as a basis for supporting an argument, or for criticizing the theory, rather than actually using the theory. Many studies reported as using ML Algorithm actually made only partial use of it. Keyword analysis revealed that the most often used keywords in papers that cited the originating article were in fact "Machine Learning Algorithm/ ML ", "SVM", "Big Data", "K-NN", "DT" and "acceptance", thus suggesting that many studies may refer to Machine learning algorithm .

It is suggested that the study presented in this paper is relevant in that ML Algorithm provides a useful tool by which to evaluate the potential for success of new technology initiation, and helps identify research tends likely to influence new technology. Identify object in the use or research of ML algorithms. Find open questions in the use or research of ML algorithms; and entire new researchers to position new research work in this domain appropriately. This paper also classifier and discuss these citation and explores the limitation of ML Algorithm use in existing research, describes the use of big data technologies, Identify types of machine learning algorithms. Overall, the topic summarizes the current state of knowledge of the machine learning algorithm. It creates an understanding of the machine learning algorithm for the reader by discussing the findings presented in recent research papers. Our systematic review of 50 citation contributes to the area of ML Algorithm/ML and research highlighting various issues, including identifying trends in terms of popular and underemployed approaches. Clearly, this is an additional work to be carried out, particularly in respect to analyzing new studies that actually make use of ML Algorithm rather than merely citing it in passing.

REFERENCES

Amin, M. J., & Riza, N. A. (2018). Machine learning enhanced optical distance sensor. *Optics Communications*, *407*, 262-270.

Ali, R., Lee, S., & Chung, T. C. (2017). Accurate multi-criteria decision making methodology for recommending machine learning algorithm. *Expert Systems with Applications*, *71*, 257-278.

Amaral, J. L., Lopes, A. J., Veiga, J., Faria, A. C., & Melo, P. L. (2017). High-accuracy detection of airway obstruction in asthma using machine learning algorithms and forced oscillation measurements. *Computer methods and programs in biomedicine*, *144*, 113-125.

Alizamir, M., & Sobhanardakani, S. (2017). A comparison of performance of artificial neural networks for prediction of heavy metals concentration in groundwater resources of toyserkan plain. *Avicenna Journal of Environmental Health Engineering*, *4*(1).

Asim, Y., Shahid, A. R., Malik, A. K., & Raza, B. (2017). Significance of machine learning algorithms in professional blogger's classification. *Computers & Electrical Engineering*.

Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science*, *89*, 117-123.

Breidenbach, L., Daian, P., Tramer, F., & Juels, A. Enter the Hydra: Towards Principled Bug Bounties and Exploit-Resistant Smart Contracts.

Bethapudi, S., & Desai, S. (2018). Separation of pulsar signals from noise using supervised machine learning algorithms. *Astronomy and Computing*, *23*, 15-26.

Bote-Lorenzo, M. L., Gómez-Sánchez, E., Mediavilla-Pastor, C., & Asensio-Pérez, J. I. (2018). Online machine learning algorithms to predict link quality in community wireless mesh networks. *Computer Networks*.

Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, *85*, 169-181.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, *10*(7), 1895-1923.

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy and Buildings*, *163*, 34-43.

Hassan, M. A., Khalil, A., Kaseb, S., & Kassem, M. A. (2017). Potential of four different machine-learning algorithms in modeling daily global solar radiation. *Renewable Energy*, *111*, 52-62.

Fernandes, D. L., Siqueira-Batista, R., Gomes, A. P., Souza, C. R., da Costa, I. T., da SL Cardoso, F., ... & Cerqueira, F. R. (2017). Investigation of the visual attention role in clinical bioethics decision-making using machine learning algorithms. *Procedia Computer Science*, *108*, 1165-1174.

Granata, F., & de Marinis, G. (2017). Machine learning methods for wastewater hydraulics. *Flow Measurement and Instrumentation*, *57*, 1-9.

Gatos, I., Tsantis, S., Spiliopoulos, S., Karnabatidis, D., Theotokas, I., Zoumpoulis, P., ... & Kagadis, G. C. (2017). A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound in Medicine and Biology*, *43*(9), 1797-1810.

Gu, G. X., Chen, C. T., & Buehler, M. J. (2018). De novo composite design based on machine learning algorithm. *Extreme Mechanics Letters*, *18*, 19-28.

Haque, S. A., Rahman, M., & Aziz, S. M. (2015). Sensor anomaly detection in wireless sensor networks for healthcare. *Sensors*, *15*(4), 8764-8786.

Jamali, A. A., Ferdousi, R., Razzaghi, S., Li, J., Safdari, R., & Ebrahimie, E. (2016). DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. *Drug discovery today*, *21*(5), 718-724.

Julio, N., Giesen, R., & Lizana, P. (2016). Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms. *Research in Transportation Economics*, *59*, 250-257.

Kim, T., Hong, J., & Kang, P. (2015). Box office forecasting using machine learning algorithms based on SNS data. *International Journal of Forecasting*, *31*(2), 364-390.

Karimi, N., Kondrood, R. R., & Alizadeh, T. (2017). An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms. *Measurement*, *107*, 68-76

Kim, E. Y., Lee, M. Y., Kim, S. H., Ha, K., Kim, K. P., & Ahn, Y. M. (2017). Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *76*, 65-71.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, *15*, 104-116.

Khosravi, A., Pabon, J. J. G., Koury, R. N. N., & Machado, L. (2018). Using machine learning algorithms to predict the pressure drop during evaporation of R407C. *Applied Thermal Engineering*, *133*, 361-370.

Luo, F., Guo, W., Yu, Y., & Chen, G. (2017). A multi-label classification algorithm based on kernel extreme learning machine. *Neurocomputing*, *260*, 313-320.

Lai, C. C. (2007). An empirical study of three machine learning methods for spam filtering. *Knowledge-Based Systems*, *20*(3), 249-254.

Lim, D. K., Long, N. P., Mo, C., Dong, Z., Cui, L., Kim, G., & Kwon, S. W. (2017). Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Research International*, *100*, 814-821.

Li, X., Mao, W., & Jiang, W. (2016). Extreme learning machine based transfer learning for data classification. *Neurocomputing*, *174*, 203-210.

Luo, M., Zheng, Q., & Liu, J. (2016). Distributed extreme learning machine with alternating direction method of multiplier. In *Proceedings of ELM-2015 Volume 2* (pp. 43-54). Springer, Cham.

Mi, X. W., Liu, H., & Li, Y. F. (2017). Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Conversion and Management*, *151*, 709-722.

Ma, H., Wang, Y., & Wang, K. (2018). Automatic detection of false positive RFID readings using machine learning algorithms. *Expert Systems with Applications*, *91*, 442-451.

Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2017). Machine learning for Internet of Things data analysis: A survey. *Digital Communications and Networks*.

Ma, H., Wang, Y., & Wang, K. (2018). Automatic detection of false positive RFID readings using machine learning algorithms. *Expert Systems with Applications*, *91*, 442-451.

Milošević, P., Petrović, B., & Jeremić, V. (2017). IFS-IBA similarity measure in machine learning algorithms. *Expert Systems with Applications*, *89*, 296-305.

Nasiri, M., Minaei, B., & Sharifi, Z. (2017). Adjusting data sparsity problem using linear algebra and machine learning algorithm. *Applied Soft Computing*, *61*, 1153-1159.

Prasad, V., Rao, T. S., & Reddy, P. P. (2016). Improvised prophecy using regularization method of machine learning algorithms on medical data. *Personalized Medicine Universe*, *5*, 32-40.

Portugal, I., Alencar, P., & Cowan, D. (2017). The use of machine learning algorithms in recommender systems: a systematic review. *Expert Systems with Applications*.

Pospieszny, P., Czarnacka-Chrobot, B., & Kobylinski, A. (2018). An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*, *137*, 184-196.

Schwartz, I. M., York, P., Nowakowski-Sims, E., & Ramos-Hernandez, A. (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience. *Children and Youth Services Review*, *81*, 309-320

Shan, L., Liu, B., & Sun, C. (2011). Comparison and improvement of feature selection method for text categorization. *Journal of Harbin institute of technology*, *43*(1), 319-324.

Salaken, S. M., Khosravi, A., Nguyen, T., & Nahavandi, S. (2017). Extreme learning machine based transfer learning algorithms: A survey. *Neurocomputing*, *267*, 516-524.

Rahman, A., & Smith, A. D. (2017). Predicting fuel consumption for commercial buildings with machine learning algorithms. *Energy and Buildings*, *152*, 341-358.

Valsamis, A., Tserpes, K., Zissis, D., Anagnostopoulos, D., & Varvarigou, T. (2017). Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction. *Journal of Systems and Software*, *127*, 249-257.

Yu, Y., & Sun, Z. (2017). Sparse coding extreme learning machine for classification. *Neurocomputing*, *261*, 50-56.

Xie, J., Towsey, M., Zhang, J., & Roe, P. (2016). Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms. *Applied Acoustics*, *113*, 193-201..

.