**ABSTRACT**

The objective of this briefing is to present an overview of the machine learning techniques currently in use or in consideration at statistical agencies worldwide. Section I, outlines the main reason why statistical agencies should start exploring the use of machine learning techniques. Section II outlines what machine learning is, by comparing a well-known statistical technique (logistic regression) with a (non-statistical) machine learning counterpart (support vector machines). Sections III, IV, and V discuss current research or applications of machine learning techniques within the field of official statistics in the areas of automatic coding, editing and imputation, and record linkage, respectively. The material presented in this paper is the result of a literature review, of direct contacts with authors during conferences, and more importantly of an international call for input that was distributed on July 18, 2014 to participants from the 2014 MSIS Meeting, participants from the 2014 Work Session on Statistical Data Editing, and members of the Modernization Committee on Production and Methods. Section VI contains a list of machine learning applications in official statistics outside of the three areas mentioned above.

**INTRODUCTION**

What is Machine Learning?

In the statistical context, Machine Learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While Machine Learning involves concepts of automation, it requires human guidance. Machine Learning involves a high level of generalisation in order to get a system that performs well on yet unseen data instances.

Machine learning is a relatively new discipline within Computer Science that provides a collection of data analysis techniques. Some of these techniques are based on well established statistical methods (e.g. logistic regression and principal component analysis) while many others are not.

Most statistical techniques follow the paradigm of determining a particular probabilistic model that best describes observed data among a class of related models. Similarly, most machine learning techniques are designed to find models that best fit data (i.e. they solve certain optimization problems), except that these machine learning models are no longer restricted to probabilistic ones.

Therefore, an advantage of machine learning techniques over statistical ones is that the latter require underlying probabilistic models while the former do not. Even though some machine learning techniques use probabilistic models, the classical statistical techniques are most often too stringent for the oncoming Big Data era, because data sources are increasingly complex and multi-faceted. Prescribing probabilistic models relating variables from disparate data sources that are plausible and amenable to statistical analysis might be extremely difficult if not impossible.

Machine learning might be able to provide a broader class of more flexible alternative analysis methods better suited to modern sources of data. It is imperative for statistical agencies to explore the possible use of machine learning techniques to determine whether their future needs might be better met with such techniques than with traditional ones.

The goal **machine** **learning** program computers use example data past experience solve given problem. Many successful applications **machinelearning** exist already, including systems analyze past sales data predict customer behavior, optimize robot behavior so task can completed using minimum resources, extract knowledge from bioinformatics data. Introduction **Machine** **Learning** comprehensive textbook subject, covering broad array topics usually included introductory **machine** **learning** texts. In order present unified treatment **machine** **learning** problems solutions, discusses many methods from different fields, including statistics, pattern recognition, neural networks, artificial intelligence, signal processing, control, data mining. All **learning** algorithms explained so student can easily move from equations book computer program. The text covers topics supervised **learning,** Bayesian decision theory, parametric methods, multivariate methods, multilayer perceptrons, local models, hidden Markov models, assessing comparing classification algorithms, reinforcement **learning.** New second edition chapters kernel machines, graphical models, Bayesian estimation; expanded coverage statistical tests chapter design analysis **machine** **learning** experiments; case studies available Web (with downloadable results instructors); many additional exercises. All chapters have been revised updated. Introduction **Machine** **Learning** can used advanced undergraduates graduate students who have completed courses computer programming, probability, calculus, linear algebra. It also interest engineers field who concerned application **machine** **learning** methods. Adaptive Computation **Machine** **Learning** series

**LITERATURE REVIEW:**

There are several applications for Machine Learning (ML), the most significant of which is predictive data mining. Every instance in any dataset used by machine learning algorithms is represented using the same set of features. The features may be continuous, categoricalor binary.

If instances are given with known labels (the corresponding correct outputs) then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabeled (Jain et al. 1999).

Numerous ML applications involve tasks that can be set up as supervised. In the present paper,wehaveconcentratedonthetechniquesnecessarytodothis.Inparticular,thisworkis concernedwithclassiﬁcationproblemsinwhichtheoutputofinstancesadmitsonlydiscrete, unorderedvalues.

Wehavelimitedourreferencestorecentrefereedjournals,publishedbooks and conferences. A brief review of what ML includes can be found in (Dutton and Conroy 1996). De Mantaras and Armengol (1998) also presented a historical survey of logic and instance based learning classiﬁers. Afterabetterunderstandingofthestrengthsandlimitationsofeachmethod,thepossibility ofintegratingtwoormorealgorithmstogethertosolveaproblemshouldbeinvestigated.The objective is to utilize the strengths of one method to complement the weaknesses of another. If we are only interested in the best possible classiﬁcation accuracy, it might be difﬁcult or impossible to ﬁnd a single classiﬁer that performs as well as a good ensemble of classiﬁers.

Our next section covers wide-ranging issues of supervised machine learning such as data pre-processing and feature selection

. Logic-based learning techniques are described in Sect. 3,whereasperceptron-basedtechniquesareanalyzedinSect.4.StatisticaltechniquesforML are covered in Sect. 5. Section 6 deals with the newest supervised ML technique—Support Vector Machines (SVMs)

In Sect. 7, a representative algorithm for each learning technique is comparedtoanumberofdatasetsinorderforresearcherstohavebaselineaccuracyfornew algorithmsinthesewell-knowndatasets.

Section8presentstherecentattemptforimproving classiﬁcation accuracy—ensembles of classiﬁers. Finally, the last section concludes this work.

**CATEGORIES OF MACHINE LEARNING**

There are two main classes of machine learning techniques:

1.Supervised machine learning

2.Unsupervised machine learning.

3.Semi-supervised Learning

4.Re-inforcement Learning

1. **Supervised Learning:**

Predicted output

Predictive model

Supervised learning technique

D

A

T

A

INPUT ATTRIBUTE

DESIRED ATTR

Data set contains dependent data and labled class to analyse the data and establish the relationship between the dependent and labeled class we use supervised algorithms. Which establishes relationship between them and forms the predictive model.on the basis of predictive model it will produce the output data.

Logistic regression (statistics) vs Support vector machines (machine learning)

Logistic regression, when used for prediction purposes, is an example of supervised machine learning. In logistic regression, the values of a binary response variable (with values 0 or 1, say) as well as a number of predictor variables (covariates) are observed for a number of observation units. These are called training data in machine learning terminology. The main hypotheses are that the response variable follows a Bernoulli distribution (a class of probabilistic models), and the link between the response and predictor variables is the relation that the logarithm of the posterior odds of the response is a linear function of the predictors. The response variables of the units are assumed to be independent of each other, and the method of maximum likelihood is applied to their joint probability distribution to find the optimal values for the coefficients (these parameterise the aforementioned joint distribution) in this linear function. The particular model with these optimal coefficient values is called the “fitted model,” and can be used to “predict” the value of the response variable for a new unit (or, “classify” the new unit as 0 or 1) for which only the predictor values are known. Support Vector Machines (SVM) are an example of a non-statistical supervised machine learning technique; it has the same goal as the logistic regression classifier just described: Given training data, find the best-fitting SVM model, and then use the fitted SVM model to classify new units. The difference is that the underlying models for SVM are the collection of hyperplanes in the space of the predictor variables. The optimization problem that needs to be solved is finding the hyperplane that best separates, in the predictor space, the units with response value 0 from those with response value 1. The logistic regression optimization problem comes from probability theory whereas that of SVM comes from geometry.

Other supervised machine learning techniques mentioned later in this briefing include decision trees, neural networks, and Bayesian networks.

**classification**

Supervised algorithms are also known as classification algorithms. It is one of the machine learning task of gathering a function from the training data which are already labelled. A set of training examples includes in the training data. In this technique, each illustration is a couple comprising of an input object and a preferred output. The training data is analysed by supervised learning algorithm and produces secondary function, which are then used for mapping new examples. A best scenario will allow for the algorithm to correctly conclude the class labels for unobserved instances

**regression**

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points.

. There are many algorithms that can be used, the following three are used in this paper:

• Decision tree

• Naïve bayes

• Naïve bayes tree

**1.Decision Tree**

A flow-chart like tree is a decision tree [2], in which each internal node is a features, each branch is a values which connects features and leaf nodes is the class which terminates nodes and branches. The starting point of the tree is known as the root of the tree and continues down to the leaves. The classification of an object begins with the root of the tree, continues towards the branch till the suitable outcomes yields. And also the process continues until the leaf is met.

**2.Naïve Bayes Naïve**

Naïve bayes algorithm is based on Bayesian theorem and probabilistic knowledge. The naïve bayes classifier takes an indication from dissimilar attributes to rush up last prediction to classify the attributes. This type of classification uses bayes rule to estimate the conditional probability by inspecting the association between each attribute value and the class. Naive Bayes classifiers calculate the probabilities of a feature which has a feature value.

**3. Naïve bayes Tree**:

The combination of decision tree and naïve bayes is Naïve Bayes Tree [4]. An NBTree can be considered as a dual- level classifier, root node with the decision tree classifier and numerous leaf nodes with naïve bayes classifier. Both Naïve Bayes and decision tree don’t have a good accuracy. The NBT algorithm is more accurate than C4.5 or Naïve Bayes on certain datasets. Like the other tree based classifiers, NBT also has branches and nodes. The advantages of both decision tree and naïve bayes can be utilized by the NBTree algorithm and it overtakes these two classifiers.

**2.unsupervised learning**

Unsupervised Learning

technique

OUTPUT

Input

Data set contains only input data.data will be analysed by using unsupervised algorithms.and it will produce the output.this output can be directly used for our own understanding.

Principal component analysis (statistics) vs Cluster analysis (machine learning)

The main example of an unsupervised machine learning technique that comes from classical statistics is principal component analysis, which seeks to “summarize” a set of data points in high-dimensional space by finding orthogonal one-dimensional subspaces along which most of the variation in the data points is captured. The term “unsupervised” simply refers to the fact that there is no longer a response variable in the current setting.

**1.Dimension reduction technique**

In [statistics](https://en.wikipedia.org/wiki/Statistics), [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and [information theory](https://en.wikipedia.org/wiki/Information_theory), dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration[[1]](https://en.wikipedia.org/wiki/Dimensionality_reduction#cite_note-1) by obtaining a set of principal variables. It can be divided into [feature selection](https://en.wikipedia.org/wiki/Feature_selection) and [feature extraction](https://en.wikipedia.org/wiki/Feature_extraction).

**2.Association analysis**

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness

**3.clustering**

Cluster analysis and association analysis are examples of non-statistical unsupervised machine learning techniques. The former seeks to determine inherent grouping structure in given data, whereas the latter seeks to determine co-occurrence patterns of items.

The problem of unsupervised learning is that of trying to find hidden structure in unlabelled data. Clustering is the unsupervised learning mechanisms and it is the well- known approach used to classify the classes in the core of a group of objects. It clusters the objects based on its resemblance without any past knowledge of the true classes. The **Association analysis**

unsupervised machine learning approach depend on a classifier that has been built from clusters are found and labelled in a training set of data. The good clusters have to have intra-cluster similarity and high- inter-cluster dissimilarity. In order to classify the network traffic of unknown applications, it is a difficult problem to solve using supervised methods. It is thoroughly linked to the problem of density approximation in statistics. Still it also covers many other techniques that seek to review and explain key features of the data. Many methods in unsupervised learning are based on data mining methods used to preprocess the data. There are many algorithms that can be used, the following three are used in this paper:

• K-means

• DBSCAN

• Expectation-Maximization

**1. K-means**

K-means algorithm is a partition based clustering technique and used to classify the traffic. it is used to classify the traffic and tries to find out user-specified number of clusters i.e., K which are denoted by using centroids. To measure the similarity between flows, Euclidean distance is used. Once the natural clusters are formed, there is a step called modelling which is used to describe a rule and that allocates a new flow to cluster. the distance measured between new flow and the cluster is called Euclidean distance. if the distance is minimum, then the new flow belongs to the cluster which is spherical in shape that is produced by the K-means algorithm. A simple and standard analysis method is a K-means clustering

algorithm. The main objective is to divide n observations into K clusters, in which each observation fits to the cluster with the nearest mean. First select K initial centroids and each point is ascribed to the nearby centroids and each group of points is designated to the centroids is a cluster. Each cluster in the centroids is streamlined based on the points designated to the cluster. We repeat the update steps till the centroids keeps on same.

**2 .DBSCAN**

Density Based Spatial Clustering of Applications with Noise (DBSCAN) is a density based algorithms [6]. It regard clusters as dense areas of objects that are separated by less dense areas. It has an advantage that it is not constrained to find spherical shaped clusters but it can able to find arbitrary shaped clusters when compared with partition based algorithms. This algorithm is constructed upon the outlooks of density-reachability and density- connectivity. These outlooks are based on two input parameters: one is epsilon(eps) and other is minimum number of points(minPts). The distance around an object is epsilon and that describes its eps-neighbourhood. For a given object say q, within the eps-neighbourhood when the number of objects is atleast minPts, then that q is termed as core object. The objects within the eps- neighbourhood are called directly density-reachable from q. In addition to this, an object say p, called as density- reachable if it is within the objects’s eps-neighbourhood that is either directly density-reachable or density- reachable from the core object q. Both the objects p and q are termed as density-connected, if they are density- reachable from an object o exists. These density-reachable and density-connected notions are used to define the cluster. The set of objects in data set which are density- connected to a particular core object is termed as a cluster. Any object that is not a fragment of cluster is considered as a noise.

**3**. **Expectation-Maximization Expectation**

Maximization algorithm (EM) is a probabilistic clustering method [7]. It is used to find out the maximum likelihood for the parameters of the probability distribution in the model. It groups traffic based on the similar properties into distinct application types. Based on the feature, the flows are grouped into small number of clusters using EM algorithm and then develop classification rules from the clusters. The algorithm that generates clusters can be specified as either Hard or Soft clusters. In Hard clusters, 0assigns a given data to exactly one of several mutually exclusive groups but in soft clusters it assigns a data point to more than one group. Specify the features that don’t create any effect on the classification are detached from the input to the learning phase and the process is continued. The EM algorithm first estimates the parameters of the model in each cluster and repeatedly uses two step processes in

**3.semi-supervised learnig**

In field **machine** **learning,** semi-supervised **learning** (SSL) occupies middle ground, between supervised **learning** (in which all training examples labeled) unsupervised **learning** (in which label data given). Interest SSL has increased recent years, particularly because application domains which unlabeled data plentiful, images, text, bioinformatics. This first comprehensive overview SSL presents state-of-the-art algorithms, taxonomy field, selected applications, benchmark experiments, perspectives ongoing future research. Semi-Supervised **Learning** first presents key assumptions ideas underlying field: smoothness, cluster low-density separation, manifold structure, transduction. The core book presentation SSL methods, organized according algorithmic strategies. After examination generative models, book describes algorithms implement low-density separation assumption, graph-based methods, algorithms perform two-step **learning.** The book discusses SSL applications offers guidelines SSL practitioners analyzing results extensive benchmark experiments. Finally, book looks interesting directions SSL research. The book closes discussion relationship between semi-supervised **learning** transduction. Adaptive Computation **Machine** **Learning** series

**4**.**Reinforcement learning**

Reinforcement learning dates back to the early days of cybernetics and work in statistics, psychology, neuroscience, and computer science. In the last five to ten years, it has attracted rapidly increasing interest in the machine learning and artificial intelligence communities. Its promise is beguiling| a way of programming agents by reward and punishment without needing to specify how the task is to be achieved. But there are formidable computational obstacles to fulfilling the promise.

Reinforcement learning is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment. The work described here has a strong family resemblance to eponymous work in psychology, but differs considerably in the details and in the use of the word \reinforcement." It is appropriately thought of as a class of problems, rather than as a set of techniques.

Reinforcement learning differs from the more widely studied problem of supervised learning in several ways. The most important difference is that there is no presentation of input/output pairs. Instead, after choosing an action the agent is told the immediate reward and the subsequent state, but is not told which action would have been in its best long-term interests. It is necessary for the agent to gather useful experience about the possible system states, actions, transitions and rewards actively to act optimally. Another difference from supervised learning is that on-line performance is important: the evaluation of the system is often concurrent with learning.

**APPLICATIONS**

Artificial Intelligence (AI) is everywhere. Possibility is that you are using it in one way or the other and you don’t even know about it. One of the popular applications of AI is Machine Learning (ML), in which computers, software, and devices perform via cognition (very similar to human brain). Herein, we share few examples of machine learning that we use everyday and perhaps have no idea that they are driven by ML.

1. **Virtual Personal Assistants**

Siri, Alexa, Google Now are some of the popular examples of virtual personal assistants. As the name suggests, they assist in finding information, when asked over voice. All you need to do is activate them and ask “What is my schedule for today?”, “What are the flights from Germany to London”, or similar questions. For answering, your personal assistant looks out for the information, recalls your related queries, or send a command to other resources (like phone apps) to collect info. You can even instruct assistants for certain tasks like “Set an alarm for 6 AM next morning”, “Remind me to visit Visa Office day after tomorrow”.

Machine learning is an important part of these personal assistants as they collect and refine the information on the basis of your previous involvement with them. Later, this set of data is utilized to render results that are tailored to your preferences.

Virtual Assistants are integrated to a variety of platforms. For example:

* Smart Speakers: Amazon Echo and Google Home
* Smartphones: Samsung Bixby on Samsung S8
* Mobile Apps: Google Allo

**2. Predictions while Commuting**

*Traffic Predictions*: We all have been using GPS navigation services. While we do that, our current locations and velocities are being saved at a central server for managing traffic. This data is then used to build a map of current traffic. While this helps in preventing the traffic and does congestion analysis, the underlying problem is that there are less number of cars that are equipped with GPS. Machine learning in such scenarios helps to estimate the regions where congestion can be found on the basis of daily experiences.

*Online Transportation Networks*: When booking a cab, the app estimates the price of the ride. When sharing these services, how do they minimize the detours? The answer is machine learning. Jeff Schneider, the engineering lead at Uber ATC reveals in a an interview that they use ML to define price surge hours by predicting the rider demand. In the entire cycle of the services, ML is playing a major role.

**3. Videos Surveillance**

Imagine a single person monitoring multiple video cameras! Certainly, a difficult job to do and boring as well. This is why the idea of training computers to do this job makes sense.

The video surveillance system nowadays are powered by AI that makes it possible to detect crime before they happen. They track unusual behaviour of people like standing motionless for a long time, stumbling, or napping on benches etc. The system can thus give an alert to human attendants, which can ultimately help to avoid mishaps. And when such activities are reported and counted to be true, they help to improve the surveillance services. This happens with machine learning doing its job at the backend.

**4. Social Media Services**

From personalizing your news feed to better ads targeting, social media platforms are utilizing machine learning for their own and user benefits. Here are a few examples that you must be noticing, using, and loving in your social media accounts, without realizing that these wonderful features are nothing but the applications of ML.

* *People You May Know*: Machine learning works on a simple concept: understanding with experiences. Facebook continuously notices the friends that you connect with, the profiles that you visit very often, your interests, workplace, or a group that you share with someone etc. On the basis of continuous learning, a list of Facebook users are suggested that you can become friends with.
* *Face Recognition*: You upload a picture of you with a friend and Facebook instantly recognizes that friend. Facebook checks the poses and projections in the picture, notice the unique features, and then match them with the people in your friend list. The entire process at the backend is complicated and takes care of the precision factor but seems to be a simple application of ML at the front end.
* *Similar Pins*: Machine learning is the core element of Computer Vision, which is a technique to extract useful information from images and videos. Pinterest uses computer vision to identify the objects (or pins) in the images and recommend similar pins accordingly.

**5. Search Engine Result Refining**

Google and other search engines use machine learning to improve the search results for you. Every time you execute a search, the algorithms at the backend keep a watch at how you respond to the results. If you open the top results and stay on the web page for long, the search engine assumes that the the results it displayed were in accordance to the query. Similarly, if you reach the second or third page of the search results but do not open any of the results, the search engine estimates that the results served did not match requirement. This way, the algorithms working at the backend improve the search results.

**ADVANTAGES**

**1.Fast processing and real-time predictions**

**Machine learning algorithms tend to operate at expedited levels.** In fact, the speed at which machine learning consumes data allows it to tap into burgeoning trends and produce real-time data and predictions. For example, machine learning can optimize and create new offers for grocery and department store customers. This means that what customers might see at 1 p.m. may be different than what they see at 2 p.m.

**2. Data Input from Unlimited Resources**

**Machine learning can easily consume unlimited amounts of data with timely analysis and assessment.** This method helps review and adjusts your message based on recent customer interactions and behaviors. Once a model is forged from multiple data sources, it has the ability to pinpoint relevant variables. This prevents complicated integrations, while focusing only on precise and concise data feeds.

**3.medical**

Machine learning in medicine has recently made headlines. [Google has developed a machine learning algorithm](https://www.mercurynews.com/2017/03/03/google-computers-trained-to-detect-cancer/)to help identify cancerous tumors on mammograms [Stanford is using a deep learning algorithm](https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/) to identify skin cancer. A [recent JAMA article](https://jamanetwork.com/journals/jama/article-abstract/2588763) reported the results of a deep machine-learning algorithm that was able to diagnose diabetic retinopathy in retinal images. It’s clear that machine learning puts another arrow in the quiver of clinical decision making..

**4.Banking**

***Fraud detection*** As mentioned earlier, ML programs can identify anomalous actions for near real-time fraud detection. The value of this in cost alone is significant considering that in 2016, customers lost nearly US $16 billion in identity theft and online fraud. In banks, machine learning can establish patterns based on the historical behavior of account owners. When uncharacteristic transactions occur, an alert is generated indicating the possibility of fraud. Such predictive analytics can also be used as an anti-money laundering (AML) tool to trace the true source of money by identifying disguised illegal cash flow – a tactic commonly used when laundering money.

***Risk assessment*** – Banks are always trying to improve loan approval processes – this is another area where machine learning can play a key role. Through intelligence gained from various data sources such as credit scores, financial data, spending patterns, etc., ML algorithms can prescribe accurate risk scores and predict the possibility of a user defaulting on a loan. Armed with such information, banks are better positioned to tailor loans, craft relevant terms and conditions and customize services to suit different customer profiles.

**5.google**

Google Search and Google Maps employ Machine Learning too. When you start typing in the search box it automatically anticipates what you might be looking for and provides suggested search terms. The suggestions could be based on past searches, what is popular now, or where you are at the time.

[Self Driving Cars](https://www.google.com/selfdrivingcar/) are probably the most sophisticated example of Machine Learning in action. If you drive around Mountain View, California you will likely see a Google self-driving car. They have been on campus for several years, and have logged 700,000 miles of accident-free autonomous driving. Watch this video to see what the onboard computers “see” and how they react as they drive the car. It is truly amazing!

**6.Facebook**

Facebook uses Machine learning in every aspect. Either you are scrolling the news feed or browsing the images or videos, you have been a part of seeing Artificial Intelligence(Machine learning).

For instance, if you upload an image at Facebook with your friends, animals or simply with pretty background, then you might get a suggestion whether you wanna tag other person or not with this image. Here you don’t need to seek that person over Facebook, Facebook will already do it for you. Facebook can also tell about a person is Happy/Sad or she/he is standing or sitting or even any other type of activity can easily be distinguished with the help of AI or ML.

Theoretically this type of learning would be called as “**Supervised Learning”** in which Facebook would learn from past experience(similar images) and label it(giving suggestion of tagging) to that particular image.

**CHALLENGES FACING BY MACHINE LEARNING**

* A very common challenge of ML application in manufacturing is the acquisition of relevant data. This is also a limitation as the availability, quality, and composition (e.g. are meta-data included? are data labeled?) of the manufacturing data at hand have a strong influence on the performance of ML algorithms. Some challenges the data-set can contain are, e.g. high-dimensional data can represent for some ML algorithms, that is, it can contain a high degree of irrelevant and redundant information which may impact the performance of learning algorithms .
* After the available data are secured, the data often have to be pre-processed depending on the requirements of the algorithm of choice. *Pre*-*processing of data* has a critical impact on the results. However, there are many standardized tools available which support the most common pre-processing processes like normalizing and filtering the data. Also it has to be checked whether the training data are unbalanced. This can present a challenge for the training of certain algorithms. In manufacturing practice, it is a common problem that values of certain attributes are not available or missing in the data-setThese so-called missing values present a challenge for the application of ML algorithms.
* Human visual systems use attention in a highly robust manner to integrate a rich set of features. But at the moment, ML is all about focusing on small chunks of input stimuli, one at a time, and then integrate the results at the end.
* Object detection is still hard for algorithms to correctly identify because imagine classification and localization in computer vision and ML are still lacking. The best way to resolve this is to invest more resources and time to finally put this problem to bed.

**CONCLUSION**

Machine learning approaches applied in systematic reviews of complex research fields such as quality improvement may assist in the title and abstract inclusion screening process. machine learning approaches are particular interest considering steadily increasing search outputs and accessibility of the existing evidence is a particular challenge of the research field quality improvement. Increased reviewer agreement appeared to be associated with improved predictive performance.

**REFERENCES**

* Chris Ding, Xiaofeng He, Hongyuan Zha, and Horst Simon. Adaptive dimension reduction for clustering high dimensional data. In Proceedings of the 2nd IEEE International Conference on Data Mining, 2002
* Projected Estimators for Robust Semi-supervised Classiﬁcation Jesse H. Krijthe and Marco Loog
* Ajith K J1, Ajay Kumar C M1, Adithya S Prakash1, Ramesh G2 Machine Learning Approach for Word Extraction
* S.Dhivya1, P.Shanmugaraja2 STUDY OF INTERNET TRAFFIC CLASSIFICATION USING MACHINE LEARNING ALGORITHMS
* Gunnar R¨atsch Friedrich Miescher Laboratory of the Max Planck Society, Spemannstraße 37, 72076 T¨ubingen, Germany http://www.tuebingen.mpg.de/˜raetsch A Brief Introduction into Machine Learning
* Leslie Pack Kaelbling, Michel L.Littman,Andrew w.Moore Reinforcement Learning A survey
* Greg Hamerly, Charles Elkanfghamerly,elkang@cs.ucsd.edu Department of Computer Science and Engineering University of California, San Diego La Jolla, California 92093-0114 K-means clustering