A Review of Machine Learning Methods Applicable to Quality Issues

Cyrus Hoseini*, M. Affan Badar, and A. Mehran Shahhosseini
*Corresponding Author
Department of Applied Engineering and Technology Management
Indiana State University
Terre Haute, Indiana 47809, USA
choseini@sycamores.indstate.edu, m.affan.badar@indstate.edu,
Mehran.Shahhosseini@indstate.edu

Christopher J. Kluse
Department of Engineering Technology
Bowling Green State University
Bowling Green, Ohio 43403, USA
ckluse@bgsu.edu

Abstract

Disruptive technology, especially machine learning (ML), is changing the paradigm in many fields, including quality. Advancements in data science, increasing processing powers of computers, and the availability of massive datasets, have made machine learning a useful tool to solve the problem at scale. In this work, a systematic review of literature has been conducted to analyze the type of industry and quality problems that can be detected with ML. ML applications in industries such as service, manufacturing, food, software/IT, and healthcare to detect quality issues and detect fraud in healthcare and health insurance have been presented. The paper has also summarized the common themes in applying ML in detecting quality problems and discussed the advantages and disadvantages of various ML algorithms in detecting quality issues and anomalies, including fraud, in various industries.

Keywords
Machine Learning, Supervised Learning, Quality Assurance, Quality Management, Quality 4.0.

1. Introduction

Machine learning (ML) is a collective name for a group of computer algorithms and statistical methods that use computational learning and artificial intelligence principles to enable computers to learn a pattern or make a prediction without being explicitly programmed (Simon, 2013; Kohavi & Provost, 1998). ML methods have been around for decades, but recent increases in the computers' processing power, made them available to a broader user community. Advancements in data processing, like matrix, parallel, and distributed computation, also made it possible to leverage massive datasets and implemented scalable deep-learning solutions. As a result, the performance of the algorithms improves as data grows. All these made ML a popular and economical method to solve many operational and quality problems. ML algorithms use probabilistic models. A probabilistic model is a way to prove the existence of a structure by creating a probability space (choosing random elements) and "prove any random element from the space has both a positive probability and the properties sought after" (Glen, n.d.). ML algorithms are commonly clustered in four primary categories (Fumo, 2017; Ayodele, 2010) as follows:

1. *Supervised learning*: the principal function of supervised learning is to infer a function from labeled data. In this group of algorithms, the researcher uses historical data as a baseline (training data) to produce an inferred function that predicts a pattern in the future data (test data) (Mohri et al., 2014). A regression model is an example of the supervised model in which a probabilistic model predicts the future behavior of a dependent variable (test data) based on existing data for independent variables (training data). The most used supervised learning algorithms are

Nearest Neighbor, Naive Bayes, Decision Trees, Linear Regression, Support Vector Machines (SVM), and Neural Networks (NN). The biggest challenge with supervised learning is the labeling process, usually a manual, time consuming, and expensive process.
2. *Unsupervised learning*: These types of machine learning algorithms do not require pre-labeled training data and analyze a set of unlabeled data to find and model hidden structures in the data set. Cluster analysis is an example of unsupervised methods in which an algorithm analyzes and divides a dataset into different clusters. K-means clustering and Association Rules (AR) algorithms are examples of unsupervised learning methods.
3. *Semi-supervised (or Hybrid) learning*: semi-supervised algorithms use a combination of labeled and unlabeled data for training. In these algorithms, a small set of labeled data is used to identify specific groups of data elements present in the labeled data and can be used to label unlabeled data. The algorithm is then trained on the unlabeled data to define the boundaries of those data elements and even found new characteristics for labeling (Castle, 2018).
4. *Reinforcement learning:* These types of algorithms use an iterative approach to learning. In these methods, the algorithm (a.k.a. agent) gathers observations from the interaction with the environment and tries to optimize the outcome (minimize risk or maximize profit) using a reward feedback (a.k.a. reinforcement signal) and continues with the iterations until it explores the full range of possible states (Fumo, 2017). Q-learning, Temporal Difference (TD), and Deep Adversarial Networks (DAN) are examples of reinforcement learning methods.

In most cases, detecting quality issues is a form of classification problem and could be formulated as a supervised learning problem if labeled data is available. The input to a classification ML model could be structured or unstructured data. Structured data refers "to any data that resides in a fixed field within a record or file" (Beal, 2018). Examples of structured data are data that reside in a relational database or a spreadsheet. Unstructured data is defined as "information that either does not have a pre-defined data model or is not organized in a pre-defined manner" (Wikipedia, n.d.). Examples of unstructured data are emails, PDFs, images, videos, audios, social media posts, text files, websites, MP3 files, and most sensor data. About 80% of the data in the world is unstructured (Cano, 2014).

In this work, a systematic review of literature has been carried out to find the application of ML in different industries such as service, manufacturing, food, software/IT, and healthcare to detect quality issues and to detect fraud in healthcare and health insurance have been presented. The paper has been organized as follows. Section 2 presents the literature review. Section 3 summarizes the same in a tabular form. Section 4 discusses the criteria to select ML algorithms in quality. Section 5 includes the conclusion and discussion.

2. Application of machine learning to solve quality problems

Machine learning techniques have been used to solve production control problems since the late 90s. As an example, Bowden and Bullington successfully applied an unsupervised learning method called the Genetic algorithm rule discovery system (GARDS) to solve a flexible cellular manufacturing system problem (Bowden & Bullington, 1996). However, leveraging machine learning to solve quality problems is a much newer trend. For the present work, authors have reviewed the literature extensively to find examples of machine learning applications to solve quality problems. In most cases, machine learning methods are used to detect quality issues in a product, process, or service. More than one ML model has been used in some studies, and their performance in solving the quality problem has been compared. More details can be found in Hoseini (2020)

2.1. Using ML to detect quality issues in the service industry

Yussupova and colleagues used several machine learning techniques, including Text Mining, Aspect Sentiment Analysis, and Decision Trees, to emulate the service quality and customer satisfaction for hotel customers (Yussupova et al., 2016). Peddamuthu & Srivastava used statistical and rule-based natural language processing (NLP) techniques to categorize sentiment and emotions, detect quality issues (e.g., anger and delays) to augment the quality assurance (QA) process for calls center conversations (Peddamuthu & Srivastava, 2014). Ucar et al. (2018) leveraged Extreme Learning Machine (ELM) to detect smart-grid Power Quality Events (PQE) in a dataset. They combined a histogram-based method with a Discrete Wavelet Transform (DWT) to improve their algorithm's performance and then tested the result on a real-world like PQE database and validated the classification accuracy method. To augment the QA process and detect inaccurate coding in financial contribution, Blomquist & Möller (2015) applied supervised classification methods (e.g., SVM) and tested ten classification models. They concluded the Adaboost procedure performed better and more steadily on most models, primarily when an ensemble classifier

supports it. Honda leveraged Text Analytics and Machine Learning to extract and classify useful information from the unstructured feedback it receives from a pull of over 20 million customers resulting in an 80% reduction in the time required by Quality Assurance staff (Kumar, 2018). To estimate the mobile phone provider service quality using social media (Twitter), Calvin & Setiawan (2014) developed a supervised learning model and Naïve Bayes classifier. This study has several limitations, including limited use of training data, trying just one supervised learning algorithm, and a lack of proper validation method of test results. Paprzycki et al. (2004) compared the performance of Six ML algorithms to predict the quality of call center services. They compared Multi-layer perceptron (MLP), Linear neural networks (LNN), Probabilistic neural networks (PNN), Classification and regression trees (CART), SVM (with a third-degree polynomial kernel), and a hybrid decision tree-ANN. They concluded the CART algorithm performed the best in overall prediction accuracy and resulting in lowest false positive and false negative prediction (over 80% customer service satisfaction prediction accuracy and close to 90% in predicting business need satisfaction).

2.2. Using ML to detect quality issues in a product or manufacturing process

As an alternative to expensive chemical QC test for olive oils, Ordukaya & Karlik (2017) used classification ML algorithms. First, they used Principal Component Analysis (PCA) to reduce the number of predictors from 32 to 8. Then they compared the performance of classifiers algorithms, including SVM, ANN, Naïve Bayesian, -Nearest Neighbors (-NN), Linear Discriminate Analysis (LDA), and Decision Tree. Then performances of these classifiers were compared according to their accuracies finding Naïve Bayes performs the best as a classifier algorithm in their case. Escobar & Morales-Menendez (2018) applied supervised learning classification methods, including the least absolute shrinkage and selection operator (LASSO) and Logistic Regression, to detect rare quality defects in a high-conformance manufacturing environment with a near-perfect result. Ribeiro (2005) proposed a classification model based on SVMs (C-SVM and ν-SVM algorithms) to monitor plastic injection molding quality issues. The study used process data (cycle time, metering time, injection time, barrel temp, cushion, and injection velocity) as predictors (independent variables). The accuracy results ranged from 70.83% up to 99.16%, which shows an acceptable performance. The study also validated the model using a radial basis function (RBF) NNs as an alternative classification method and concluded SVMs have a performance advantage over NNs (Ribeiro, 2005). ML and Predictive modeling in junction with IoT is used for predictive maintenance and predicting the equipment failure, which resulted in a significant increase in the overall equipment efficiency (OEE) and reliability (Tracy, 2018). This work used a massive dataset containing three years of process and sensor data from 350 lines across 33 manufacturing plants and 15 countries. The author built a machine learning model using N-dimensional Euclidean distance-based scoring algorithms normalized that predicted potential quality failures of a process and then applied the model to 1,400 manufacturing lines and over 20 million sensor events resulting in a potential $300 million annual saving (RTInsights Team, 2017). Lieber et al. (2013) proposed a hybrid ML model consisting of unsupervised (k-Means) and supervised algorithms (nearest neighbor and SVM) to predict intermediate products' quality in interlinked manufacturing processes. The k-NN algorithm (k=11) achieved a 97% accuracy level in prediction, while SVM produced about 90% prediction accuracy. Irgens et al. (2012) proposed a hybrid model consisting of unsupervised (Cluster Analysis) and Supervised ML algorithms (SVM) to improve the product and process quality in the manufacturing process. In a detailed study, Bai et al. used the manufacturing quality control data. They proved deep learning algorithms (specifically deep restricted Boltzmann machine and the stacked autoencoder) have a performance advantage over shallow learning methods (in this case, a feed-forward neural network with one hidden layer and the least squares support vector machine with no hidden layers). Also, they found the performance of deep learning algorithms is proportional to the sample size (Bai et al., 2017). Chou et al. (2018) collected ten years' worth of data from 20 reservoirs in Taiwan and applied four ML algorithms (ANN, SVM, classification and regression trees, and linear regression) and concluded the ANN model performs best in predicting the water quality. To correlate arc sound with the weld quality, Sumesh et al. (2015) used ML classification algorithms (namely J48 and Random Forest). They concluded that J48 is outperforming the random forest to detect weld quality issues (88.69% accuracy for J48 vs. 70.78% for random forest).

2.3. Using ML detect quality issues in the food industry

To identify quality issues in Salmon using a computer vision, Sture (2015) applied ML and classification methods using Support Vector Machines (SVM) with a nonlinear kernel function (Radial basis function), with success and then validated the geometric classification results using nearest-neighbor classification, with slightly less accuracy. Lotfi and the team successfully applied vision processing techniques augmented by an ML method based on neural networks to detect French fries' quality issues (Lotfi et al., 2008). To expand machine vision and measure the quality of mixed raisins, Karimi et al. (2017) built an ML model with 146 predictors. They then applied a Principal

Components Analysis (PCA) method to reduce the predictors to an optimal level. They used SVM and ANN to classify the raisin mixtures and found SVM surpasses ANN in this classification problem and produces a high-level accuracy in prediction. Gonzalez Viejo et al. (2017) leveraged ML and used physical measurements of color and foam (as independent values) to predict beer quality (defined by the intensity levels of sensory descriptors like perceived foam–related parameters and beer color). Their research used principal component analysis to identify the most relevant predictors and an artificial neural network (ANN) regression algorithm, which resulted in an R = 0.91correlation to predict the beer's quality. To check Pistachio's quality automatically and through machine vision, Çitak & Genç (2017) formulated a classification problem and used support vector regression (SVR), deep convolutional networks, and a classification method based on deep neural networks with over 98% accuracy.

2.4. Using ML to detect quality issues in Software and Information Technology

ML and predictive modeling have long been used to augment the quality assurance and testing practices in software engineering and information technology. Morales (2017) used a generic title for all such algorithms as "Predictive quality analytics" and defined it as "the process of extracting useful insights from test data from various sources by applying statistical algorithms and machine learning to determine patterns and predict future outcomes and trends." The core statistical algorithms used in such methods include various classification, clustering, regression, time series, and association techniques (Morales, 2017). Parra et al. (2015) used a supervised learning model and text analytics approach to classify the software requirements (into two classes of good and bad quality requirements), emulating quality experts' assessments using two different methods. They compared the accuracy, stability, and efficiency for six different machine learning algorithms: PART, C4.5, bagging PART, bagging C4.5, boosting PART, and boosting C4.5. They concluded C4.5 has the highest efficiency and accuracy rate (87.72%) but has the highest standard deviation (7.12%). They found bagging-PART has the lowest standard deviation (2.44%) and is the most stable algorithm with the second-highest accuracy of 87.02% (Parra et al., 2015). Khoshgoftaar & Seliya (2003) used tree-based software quality classification models, software metrics, and SPRINT decision tree algorithm to build a predictive model to forecast if a software module is fault-prone or not. They also collected software metrics from extensive telecommunications systems and compared the SPRINT decision tree algorithm's result with a CART decision tree algorithm (a more generic form of SPRINT). They found advantages in how the SPRINT decision tree algorithm uses a unique tree pruning technique based on the Minimum Description Length (MDL) principle resulting in improved accuracy and stability of the model (Khoshgoftaar & Seliya, 2003). Many researchers have leveraged machine learning to predict which module in software has a higher chance of failure. Elish & Elish (2008) compared the performance of eight different ML algorithms, including SVM, Logistic regression (LR), K-nearest neighbor (KNN), Multi-layer perceptrons (MLP), Radial basis function (RBF), Bayesian belief network (BBN), Naïve Bayes (NB), Random forest (RF), and Decision tree (DT). They found SVM outperforms other algorithms in predicting defect-prone modules on the four NASA datasets (Elish & Elish, 2008). In a similar attempt, Gondra (2008) used the same NASA open-source dataset and 21 software product metrics as predictors and compared SVM and ANN's performance as supervised learning methods. He concluded SVM outperforms ANN in this binary classification problem (87.4%, to 72.61% accuracy rate). He also noted the advantage of SVM and ANN as nonlinear approaches to solving this problem compared to linear ones and even the benefit of sensitivity analysis to PCA in selecting software metrics that are more robust indicators of a defect in a module (Gondra, 2008). To overcome the challenge of skewness in defect-prediction datasets, Pelayo & Dick (2007) applied the Synthetic Minority Over-sampling Technique (SMOT). This improved the geometric mean classification accuracy by over 20%. Bouguila et al. (2008) applied a Bayesian model based on finite Dirichlet mixture models to predict software quality of fault-prone and non-fault-prone program modules. They used Gibbs sampler to implement their Bayesian algorithm. ML algorithms like SVM and Bayesian methods are widely applied to predict the quality of Web Services (Kai et al., 2016).

2.5. Using ML to detect quality issues in healthcare

A group of researchers used over 7,500 human-rated samples and applied specific supervised ML techniques (kernelized Support Vector Machine and Gradient Boosted Decision Trees classifiers) to detect meshes of failing quality. They improved the accuracy of quality assessment of MRI-derived data conducted by medical professionals resulting in a human workload reduction by 30-70% (Petrov et al., 2017). Gupta (2018) applied NN and SVM algorithms to determine wine quality dependency on different physicochemical characteristics. He first used linear regression to measure the dependency of the target variable (wine quality) on predictors and then leveraged the SVM and NN and concluded that SVM performs better than NN to predict the wine quality.

2.6. Using ML to detect fraud in healthcare and health insurance

ML in healthcare services is frequently used to detect Fraud, Waste, and Abuse (FWA). Anomaly detection, which mainly uses "unsupervised learning" methods, has been used often for this purpose (Anbarasi & Dhivya, 2017). Some researchers advanced anomaly detection methods through a combination of several algorithms. For example, to detect Korean outpatient clinics with abusive utilization patterns, Shin et al. (2012) proposed an algorithm consisting of a scoring model to measure the degree of abusiveness and segmentation method to cluster clinics with similar utilization patterns. Their algorithm leveraged decision tree for clustering and conditional probability distributions of the composite degree of anomaly (CDA) score to categorize clinics in intervention or non-intervention subgroups (Shin et al., 2012). Kose et al. (2015) proposed an interactive machine learning approach combined with the pairwise comparison method of analytic hierarchical processing (AHP) for weighting the actors and attributes and expectation maximization (EM) to detect electronic fraud and abuse in the healthcare system. To overcome the limitation of normality and outlier free assumptions of parametric methods, they invented a non-parametric unsupervised method (Kose et al., 2015). Gallardo (2017) used convolutional deep neural networks (CNN), and recurrent neural networks (RNN) have also been patented as an algorithm to create an analytics engine for detecting medical frauds. van Capelleveen et al. (2016) used multivariate clustering and outlier detection, along with an expert system to detect fraud cases in Medicaid dental claims. A group of experts manually reviewed the flagged cases and validated 71% of cases proved to be fraudulent, a considerable improvement over the conventional method, which is, on average, detect 10% of fraud cases (van Capelleveen et al., 2016).

Medicaid and Medicare are government-funded health programs and subject to fraud, waste, and abuse. As a result, many researchers studied machine learning algorithms' application in detecting fraud cases in Medicare and, especially, Medicaid. Most ML methods used for fraud detection in Medicaid are based on unsupervised methods. In supervised Medicaid fraud detection techniques, subject matter experts use prior information on class membership to select a set of training data (Travaille et al., 2011). This training data set is used by the algorithm to label each reviewed value as a suspicious or legal item. For example, subject matter experts review a set of paid and denied Medicaid claims and label each one legitimate or fraudulent. There are many supervised techniques listed as potential algorithms for Medicaid Fraud detection, including (Copeland et al., 2012; Phua et al., 2005; Getchius, 2014; Li et al., 2008):

- Multi-layer perceptron network
- Artificial Neural Network including Backpropagation (backward propagation of errors)
- Support Vector Machine
- Decision Trees
- Fuzzy Logic
- Bayesian Network including Bayesian Belief Network (BBN)
- Probable Graph Model (PGM)
- Linear Gaussian Model

Researchers also presented some innovative methods to evaluate supervised methods' performance by calculating their return on investment and costs versus benefits. At least two metrics are introduced in research to measure the performance of anomaly detection methods tradeoff between detection probability and false alarm ratio and the tradeoff between false alarm ratio and detection delay (Siris & Papagalou, 2004). Phua et al. (2005) described how statistical methods should be classified, utilized, interpreted, and validated to detect healthcare fraud. They compared statistical methods applied to health care fraud detection by focusing, reviewing, and analyzing the investigations conducted in this field (Phua, et al., 2005). While supervised methods are more accurate in detecting previous fraud types, they also suffer from some shortcomings, primarily when used to detect Medicaid fraud. The shortcomings include:

- Bolton and Hand indicated that supervised methods' accuracy relies on the accurate identification of fraudulent and non-fraudulent transactions in the historical dataset where information is missed or limited (Bolton et al., 2001).
- Creating training data is a tedious task and requires expensive subject matter experts to develop training data sets (especially non-fraudulent or legal data sets). One way to overcome this burden is to create just fraudulent datasets, which are much more limited.

- The constant change in the healthcare landscape changes the claim patterns. This includes demographical changes (e.g., aging population, shrinkage of middle-class, baby boomer retirement) and policy changes (e.g., the Affordable Care Act). As a result, there is a constant need to update training data sets.
- Periodic changes in healthcare systems force the users of supervised methods to update their training data sets regularly. Examples of such changes include implementing the hospital's performance-based payment, HIPAA 5010, and ICD-10 code migrations.

Unsupervised Medicaid fraud detection techniques do not assume prior class labels of legitimate or fraudulent behavior (Travaille et al., 2011). These algorithms mine the entire dataset and try to find a Medicaid data pattern and identify outliers. For example, an unsupervised algorithm could analyze the utilization of a specific prescription drug among a group of Medicaid beneficiaries and identify the anomalies. All anomalies then need to be researched and validated. The important fact about all probabilistic fraud detection models is that they will find likely fraud cases, and each case requires future research and validation. These algorithms' strengths are their sensitivity (detecting most true positives) and specificity (reducing false positives). Examples of unsupervised methods that could be used for fraud detection (including Medicaid fraud) are Anomaly Detection, K-mean, Benford's Law, Linear Regression, Modified Batch Library Method (MBLM), Adaptive threshold algorithm, and CUSUM (Cumulative SUM) algorithms (Heino & Toivonen, 2003; Siris & Papagalou, 2004; Issa & Vasarhelyi, 2011; Tsung et al., 2007). Unsupervised Medicaid Fraud algorithms also have strengths and weaknesses. A few of these are listed below:

- The neural network is an excellent method to handle complex data sets and manage noisy data, but its black-box approach limits the researchers' understanding of how a system works (Li et al., 2008).
- Anomaly Detection is not robust toward the number of metrics used to flag fraudulent cases. This means too many metrics negatively impact the anomaly detection method's efficiency. When metrics features are increasing, more cases will be flagged as potential fraud cases, increasing the number of false-positive and making the process ineffective (Copeland et al., 2012).
- A decision tree is useful for handling missing data and generates rule from a tree (white-box approach) but is not suitable for complex datasets (Li et al., 2008).
- Fuzzy Logic allows approximate reasoning but is difficult to tune and lack sufficient learning capability (Li et al., 2008).
- Genetic Algorithm could be used very well for systematic random search but is difficult to tune.

3. Summary of examples ML application to detect quality issues in different industries

Table 1 presents a summary of the main elements of the studies reviewed in the literature and discussed in Section 2 to draw some conclusions.

Table 1- Summary of literature review of the application of ML to detect quality issues.

Industry	Quality Problem	Data Type[1]	Inputs	BDA Techniques ML Algorithms Applied	Notes about the methods and results	Reference
Food	Detect the quality issues in Salmon	U	Images	- Support Vector Machines (SVM) with a nonlinear kernel function (Radial basis function) [B][2] - Nearest-neighbor classification		(Sture, 2015)
Food	QC test for olive oils	U&S	Chemical and visual measures	- Naïve Bayesian [B] - SVM - ANN - -Nearest Neighbors (-NN) - Linear Discriminate Analysis (LDA) - Decision Tree	Researchers first used Principal Component Analysis to reduce the number of predictors from 32 to 8	(Ordukaya & Karlik, 2017)

[1] S: Structured, U: Unstructured
[2] [B] Best performing algorithm in the study (if any).

Food	Detect quality issues in French fries	U	Images	- Vision processing techniques augmented by an ML method based on neural networks		(Lotfi et al., 2008)
Food	Measure the quality of mixed raisins	U	Images	- SVM [B] - ANN	First used Principal Components Analysis (PCA) method to reduce 146 predictors to an optimal level	Prediction (Karimi et al., 2017)
Food	Predict wine quality	S	Physicochemical characteristics	- SVM [B] - NN	First used linear regression to measure the dependency of the target variable (wine quality) on predictors	(Gupta, 2018)
Food	Predict the quality of the beer (defined by the intensity levels of sensory descriptors like perceived foam–related parameters and beer color)	U&S		- ANN - Regression	Researchers first used principal component analysis to identify the most relevant predictors. The final resulted in an R = 0.91correlation to predict the quality of the beer	(Gonzalez Viejo et al., 2017)
Food	Check the quality of Pistachio automatically and through machine vision	U	Physicochemical characteristics	- Support vector regression (SVR) - Deep convolutional networks - Deep neural networks		Çitak & Genç (2017)
Manufacturing	Detect rare quality defects in a high-conformance manufacturing environment	S	Sensor data	- least absolute shrinkage and selection operator (LASSO) - Logistic Regression (LR)	Supervised Learning	(Escobar & Morales-Menendez, 2018)
Manufacturing	Predict potential quality failures of a process	S	Process and sensor data	- N-dimensional Euclidean distance-based scoring algorithms	Resulting in a potential $300 million annual saving	(RTInsights Team, 2017)
Manufacturing	Detect manufacturing quality issues	S	QC data	- Deep Learning (deep restricted Boltzmann machine and the stack autoencoder) [B] - Shallow learning (feed-forward neural network with one hidden layer and the least-squares SVM with no hidden layers)	Researchers proved deep learning methods outperforming shallow learning methods. They also found the performance of deep learning algorithms is proportional to the sample size	(Bai et al., 2017)

Manufacturing	Predict the quality of intermediate products in interlinked manufacturing processes	S	Process and sensor data	- A hybrid model of unsupervised (k-Means) and supervised (k-nearest neighbor [B] and SVM) algorithms	With the k-NN algorithm (k=11), researchers achieved a 97% accuracy level in prediction	(Lieber et al., 2013).
Manufacturing	Improve the product and process quality in the manufacturing process	S	Process data	- The hybrid model consists of unsupervised (Cluster Analysis) and Supervised ML algorithms (SVM)		(Irgens et al., 2012)
Manufacturing	Predict the quality of welding	U	Arc sound and images	- J48 [B] - Random Forest	88.69% accuracy rate for J48 vs. 70.78% for random forest	(Sumesh et al., 2015)
Manufacturing	Monitor quality issues in plastic injection molding	S	Process data (cycle time, metering time, injection time, barrel temp., cushion, and injection velocity)	- SVMs (C-SVM and v-SVM) radial basis function (RBF) NNs	- Accuracy results ranged from 70.83% up to 99.16% - SVMs have a performance advantage over NNs	(Ribeiro, 2005).
Call Center	Detect quality issues (e.g., anger or delay) and categorize sentiment and emotions	U	Calls, Text	- Rule-based natural language processing (NLP)		(Peddamuthu & Srivastava, 2014)
Call Center	Predict the quality of service in a call center	S	Call Performance Evaluation data	- CART [B] - Multi-layer perceptron (MLP) - Linear NN - Probabilistic NN - SVM (with a third-degree polynomial kernel) - Hybrid decision tree-ANN	Over 80% of customer service satisfaction prediction accuracy and close to 90% in predicting business need satisfaction	(Paprzycki et al., 2004)
Hospitality	Measure service quality and customer satisfaction	U	Text	- Aspect Sentiment Analysis - Decision Trees		(Yussupova et al., 2016)
Utilities	Detect smart-grid Power Quality Events (PQE)	S	Network data	- Discrete Wavelet Transform (DWT)		Ucar, et al., 2018)
Financial services	Detect inaccurate coding in	S	Financial data	- Adaboost supported by ensemble classifier [B] - SVM	Supervised Learning	(Blomquist & Möller, 2015)

	financial contribution					
Medical	Improve the accuracy of quality assessment of MRI-derived data conducted by medical professionals	U	Medical images	- Kernelized SVM - Gradient Boosted Decision Trees classifiers	Supervised learning using over 7,500 human-rated samples resulting in a human workload reduction by 30-70%	(Petrov et al., 2017)
Natural Resource	Measure water quality	S	Lab data	- ANN [B] - SVM - Classification and regression trees (CART) - LR	Huge sample size (10 years' worth of data from 20 reservoirs in Taiwan)	(Chou et al., 2018)
Telecom	estimate the mobile phone provider service quality using social media	U	Twits from Twitter	- Naïve Bayes	Lack of proper validation method of test results	(Calvin & Setiawan, 2014).
Telecom	Predict if a software module is fault-prone or not	S	Software metrics	- The SPRINT decision tree [B] - CART decision tree	They found the way SPRINT decision tree leverages a unique tree pruning technique based on the Minimum Description Length (MDL) principle improves the accuracy and stability of the model	(Khoshgoftaar & Seliya, 2003)
Software	Predict which module in software has a higher chance of failure.	S	Software metrics	- SVM [B] - LR - KNN - Multi-layer perceptrons (MLP) - Radial basis function (RBF) - Bayesian belief network (BBN) - Naïve Bayes (NB) - Random forest (RF) - Decision tree (DT)		(Elish & Elish, 2008)
Health Insurance	Detect fraud in outpatient clinics claims	U	Hospital visit information	- Unsupervised clustering methods: - Decision tree	Used conditional probability distributions of the composite degree of anomaly (CDA) to adjust the risk score	(Shin et al., 2012)
Health Insurance	Detect fraud in electronic claim data	U	Hospital visit information	- Non-parametric unsupervised method	AHP was used for weighting.	(Kose et al., 2015)

Health Insurance	Detect fraud in electronic claim data	U	Hospital visit information	- Convolutional deep neural networks (CNN) - recurrent neural networks (RNN)	Patented proprietary algorithm.	(Gallardo, 2017)
Health Insurance	Detect fraud in healthcare insurance claims	S	Insurance claims	- Analytic hierarchical processing (AHP)		(Anbarasi & Dhivya, 2017)
Health Insurance	Detect fraud in healthcare insurance claims	S	Insurance claims	- N/A	A literature review of various statistical, computerized, and ML method	(Travaille et al., 2011)
Health Insurance	Detect fraud in healthcare insurance claims	S	Insurance claims	- Neural Network - Decision Tree	A survey of various ML methods	(Li et al., 2008)
Health Insurance	Detect fraud in healthcare insurance claims	S	Insurance claims	- Multi-layer perceptron network - Artificial Neural Network including Backpropagation (backward propagation of errors) - Support Vector Machine - Decision Trees - Fuzzy Logic - Bayesian Network including Bayesian Belief Network (BBN) - Probable Graph Model (PGM) - Linear Gaussian Model	A survey of various ML methods	(Phua, C. et al., 2005)
Health Insurance	Detect fraudulent refunds	S	Insurance claims	- Neural network - Bayesian network - SVM - K-means		(Issa & Vasarhelyi, 2011)
Health Insurance	Detect dental insurance fraud	S	Medicaid dental insurance claims	- Multivariate clustering - Outlier detection - Time series		(van Capellevee n et al., 2016)
Service	Detect fraud service	S	Customer information	- Multi-way Principal Component Analysis (MPCA) - SPC	Applying manufacturing batch techniques to fraud detection with incomplete customer information	

4. Criteria to select ML algorithms to detect a quality problem

Many factors should be considered before one of several ML algorithms are selected to solve a quality problem. The following criteria are among the most critical factors that a quality practitioner should consider before choosing an ML algorithm to solve a quality problem:

- *Model's performance*: various metrics could be used to measure the performance of an ML model, including Accuracy, Recall, Precision, F1 score (the weighted average of precision (positive predicted value) and recall

(sensitivity). It is also known as the F-score or F-measure), Average squared error (ASE), Area under the curve (C-statistic), Area under the ROC curve, Captured response, Cumulative captured response, Kolmogorov- Smirnov statistic (KS), False discovery rate, Gini, KS (Youden), Lift, Misclassification (Event), Misclassification (MCE), ROC separation, The root-mean-square deviation (RMSD), and root-mean-square error (RMSE)

There are also many other metrics used to precisely evaluate the performance of supervised learning models. Each has its advantages and disadvantages based on the problem statement and the nature of the data. Researchers have offered various metrics for model validation and selecting the most appropriate model for each particular problem (Shah et al., 2016; Marc-Oliver Arsenault, 2017).

Perhaps the most popular metric to select an appropriate machine learning algorithm is the model accuracy, which is defined as total true positives and true negatives of all results produced by an algorithm. Numerous articles compared the accuracy of different supervised classifiers in solving different types of problems. As an example, Singh & Kaur (2016) compared the accuracy of J48 and REP Tree to predict computer science students' performance and found J48 outperforms REP Tree in model accuracy 67.37% to 56.78%. Accuracy is easy to calculate and understand. However, when dealing with rare events as the target group (dependent variable), accuracy could be misleading. This is because of many true negatives in the denominator; accuracy would be a near-perfect measure even for poor performing models. Recall, precision, and F-measure (a balanced measure of precision and recall) focus on the target category (a rare event) and are more relevant to the problem. When the cost of false negatives is high (e.g., in detecting cancerous tumors), recall becomes a critical metrics to consider. If our concern is more sensitive toward lower false positives, then precision becomes more important. F-measure (F1 score) is another essential metric to consider, especially recall, and accuracy should be controlled. ROC separation is also widespread; however, if it may not work and F-1score for quality issues. Because quality issues are rare events, detecting a quality issue in a large dataset with ML usually deals with imbalanced data. Research shows F1-score superiority in selecting the best performing model when dealing with imbalanced data (Yahya, 2018).

- *Model Stability*: an algorithm is defined as β-stable (or stable, in general) when its losses incurred by the corresponding hypotheses on two similar but different datasets are equal or less than β (Mohri et al., 2014). There are specific limitations to how big β could grow and still a training algorithm to be convergence. Convergence to a particular measure is required for the learning algorithm to have an endpoint. Convergence to the optimal point is necessary to ensure the algorithm effectively solves the problem (it stops when it finds the global optimum or a reasonable local optimum).

- *Model interpretability vs. prediction accuracy:* there is an inherited trade-off between the statistical models' interpretability and prediction accuracy (James et al., 2015). Many flexible and accurate ML algorithms (e.g., SVM and NN) lack interpretability. Such algorithms are black-box approaches and are not good choices if there is a need to look inside the box. When there is a lack of trust in the ML model, or stakeholders demand the model logic to be clearly explained, the choice is to compromise the flexibility and choose a more interpretable model (e.g., LASSO) (Beeravalli et al., 2018; Rácz et al., 2019).

- *Bias-variance trade-off:* In the context of ML, bias is the overall fitness of the model to the training data, and variance is the overall fitness of the model to the test and unseen data. As a rule of thumb, as model flexibility and complexity increase, the bias will decrease, and variance will increase, so the complex models (e.g., NN and SVM) have an inherited tendency to over-fit and should be used with cautious (James et al., 2015). Brain & Webb (1999) found the dataset's size may impact variance and bias in an ML model. Specifically, the hypothesized variance can be expected to decrease as training set size increases, but no apparent effect of training set size on bias was observed (Brain & Webb, 1999). These results have profound implications for data mining from large data sets, indicating that developing practical learning algorithms for large data sets is not merely a matter of finding computationally efficient variants of existing learning algorithms.

- *Computation and tuning time:* when dealing with massive datasets and complex models, the computation time will increase (sometimes exponentially). It is essential to understand the computational power required for each ML algorithm. Some ML algorithms (e.g., SVM) are known to require substantial computational power, which may hinder the researcher's ability to run their models properly.

- *The impact of data reduction*: one way to increase the efficiency of a machine learning model is to appropriately reduce the version of data or a lower number of attributes. El-hasnony et al. (2015) compared the accuracy of classification nine algorithms using data reduction techniques like Correlation Feature Selection (CFS), Rough Set Attribute Reduction (RSAR), Fuzzy Rough Feature Selection (FRFS), PCA, and gain ratio. These are the list of classification algorithms they tested in their research: C4.5, fuzzy rough nearest neighbor, Multi-layer perceptron (MLP), Nearest-neighbor-like algorithm using non-nested generalized exemplars (NNGE), Fuzzy nearest neighbor, sequential minimum optimization (SMO), classification via clustering, NB-tree and Naïve Bayes. They concluded that fuzzy rough feature selection outperforms rough set attribute selection, gain ratio, correlation feature selection, and principal components analysis (El-hasnony et al., 2015). The importance of this study is not only in the selection of the nine popular classification techniques (that could be used as a data input to select the most commonly used classifier algorithms for this study) but also in understanding the performance of different data reduction techniques should we need to use them in data validation and testing phases.

- *Model fitness*: Akaike's Information Criterion (AIC) is a statistical measure of the goodness of fit for a particular model. It maximizes the expression $-2(LL+k)$ where k is the number of features, and LL is the maximized value of the log-likelihood function for the given model. The smaller the AIC, the better the model fits the data. Because of the k term, the smaller number of model parameters is favored (Dean, 2014).

5. Conclusion and discussion

In this work, a systematic review of the literature has been conducted. Machine learning applications in industries such as service, manufacturing, food, software/IT, and healthcare to detect quality issues and detect fraud in healthcare and health insurance have been presented. The literature review shows the most common application of machine learning in the examined cases is to detect quality issues through supervised learning. This problem could be formulated as a classification problem. The second most common application is to predict quality using supervised or semi-supervised learning. Depending on the quality measure (variable vs. attribute), the problem could be formulated as a regression or a classification problem. Supervised learning has more applications in detecting quality issues. Fraud detection methods are mainly based on unsupervised methods. This is due to the changing nature of the fraud patterns and the cost of data labeling. When it comes to machine learning algorithms, it is found that there exists no magic bullet for the best algorithm. Each algorithm may perform better in solving a specific problem. The most popular choices among algorithms are Decision Tree (various configurations for both supervised and unsupervised learning), SVM (for supervised learning for binary classification problems), Logistic Regression (supervised), NN (for unsupervised and semi-supervised learnings), and Naïve Bayes (supervised). It is also observed that complex (aka black box) algorithms like SVM and NN may perform better in many cases, but they are hard to explain. Computation time for complex models also tends to be higher. For example, SVM requires massive computational power for mixed and massive datasets, a common scenario in real-world quality problems. Simpler models like decision trees also have disadvantages, including low stability. Research shows the biggest challenge in leveraging supervised learning to solve quality problems is the labeling process, usually a tedious and expensive process. Dimensionality reduction techniques like Principal Components Analysis (PCA) are frequently used to reduce the number of features without compromising the model performance to reduce the model complexity without losing the prediction power. Another important finding is that most quality problems are a specific type of supervised learning dealing with rare events and unbalanced data. Various techniques (including event-based sampling and oversampling) are used to overcome this challenge.

References

Anbarasi, M. S., & Dhivya, S. (2017). Fraud detection using outlier predictor in health insurance data. *2017 International Conference on Information Communication and Embedded Systems (ICICES)*, 1–6. https://doi.org/10.1109/ICICES.2017.8070750

Ayodele, T. O. (2010). Types of Machine Learning Algorithms. In Y. Zhang (Ed.), *New Advances in Machine Learning*. InTech. https://doi.org/http://dx.doi.org/10.5772/46845

Bai, Y., Sun, Z., Deng, J., Li, L., Long, J., & Li, C. (2017). Manufacturing Quality Prediction Using Intelligent Learning Approaches: A Comparative Study. *Sustainability*, *10*, 85.

Beal, V. (2018). *What is Structured Data?* Webopedia Definition. https://www.webopedia.com/TERM/S/structured_data.html

Beeravalli, V., Krishna, N., & Pandian, R. (2018, September 30). *Comparison of Machine Learning Classification Models for Credit Card Default Data*. Medium.Com. https://medium.com/@vijaya.beeravalli/comparison-of-machine-learning-classification-models-for-credit-card-default-data-c3cf805c9a5a

Blomquist, H., & Möller, J. (2015). *Anomaly detection with Machine learning – Quality assurance of statistical data in the Aid community*. http://www.utn.uu.se/sts/cms/node/908

Bolton, R. J., Hand, D. J., & H, D. J. (2001). Unsupervised Profiling Methods for Fraud Detection. *Proc. Credit Scoring and Credit Control VII*, 5–7. https://doi.org/10.1.1.24.5743

Bouguila, N., Wang, J. H., & Hamza, A. Ben. (2008). A Bayesian approach for software quality prediction. *2008 4th International IEEE Conference Intelligent Systems*, *2*, 11–54. https://doi.org/10.1109/IS.2008.4670508

Bowden, R., & Bullington, S. F. (1996). Development of manufacturing control strategies using unsupervised machine learning. *IIE Transactions*, *28*(4), 319–331.

Brain, D., & Webb, G. I. (1999). On The Effect of Data Set Size on Bias And Variance in Classification Learning. *Proceedings of the Fourth Australian Knowledge Acquisition Workshop (AKAW '99)*, 117–128.

Calvin, & Setiawan, J. (2014). Using Text Mining to Analyze Mobile Phone Provider Service Quality (Case Study: Social Media Twitter). *International Journal of Machine Learning and Computing*, *4*(1), 106–109. https://doi.org/10.7763/IJMLC.2014.V4.395

Cano, J. (2014). *The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity*. https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity

Castle, N. (2018). *What is Semi-Supervised Learning?* https://www.datascience.com/blog/what-is-semi-supervised-learning

Chou, J.-S., Ho, C.-C., & Hoang, H.-S. (2018). Determining quality of water in reservoir using machine learning. *Ecological Informatics*, *44*, 57–75. https://doi.org/10.1016/J.ECOINF.2018.01.005

Çitak, E., & Genç, Y. (2017). Machine Learning For Product Quality Inspection. *Signal Processing and Communications Applications Conference (SIU)*, 1–4.

Copeland, L., Edberg, D., Panorska, A. K., & Wendel, J. (2012). Applying Business Intelligence Concepts to Medicaid Claim Fraud Detection. *Journal of Information Systems Applied Research (JISAR)*, *5*(1), 51–61. www.aitp-edsig.org

Dean, J. (2014). Big data, data mining, and machine learning: value creation for business leaders and practitioners. In *Wiley & SAS business series*.

El-hasnony, I. M., Bakry, H. M. El, & Saleh, A. A. (2015). Comparative Study among Data Reduction Techniques over Classification Accuracy. *International Journal of Computer Applications*, *122*(2), 8–15. https://doi.org/10.5120/21671-4752

Elish, K. O., & Elish, M. O. (2008). Predicting defect-prone software modules using support vector machines. *Journal of Systems and Software*. https://doi.org/10.1016/j.jss.2007.07.040

Escobar, C. A., & Morales-Menendez, R. (2018). Machine learning techniques for quality control in high conformance manufacturing environment. *Advances in Mechanical Engineering*, *10*(2), 1–16. https://doi.org/10.1177/1687814018755519

Fumo, D. (2017). *Types of Machine Learning Algorithms You Should Know*. https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

Gallardo, K. (2017). *Analytics Engine for Detecting Medical Fraud, Waste, and Abuse* (Patent No. US20170270435A1). https://patents.google.com/patent/US20170270435A1/en

Getchius, J. M. (2014). *Healthcare fraud detection with machine learning* (Patent No. US20140149128 A1).

Glen, S. (n.d.). *Probabilistic: Definition, Models and Theory Explained*. StatisticsHowTo.Com: Elementary Statistics for the Rest of Us! Retrieved October 30, 2020, from https://www.statisticshowto.com/probabilistic/

Gondra, I. (2008). Applying machine learning to software fault-proneness prediction. *Journal of Systems and Software*, *81*(2), 186–195. https://doi.org/10.1016/j.jss.2007.05.035

Gonzalez Viejo, C., Fuentes, S., Torrico, D., Howell, K., & Dunshea, F. R. (2017). Assessment of beer quality based on foamability and chemical composition using computer vision algorithms, near infrared spectroscopy and machine learning algorithms. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/jsfa.8506

Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques. *Procedia Computer Science*, *125*, 305–312. https://doi.org/10.1016/J.PROCS.2017.12.041

Heino, J., & Toivonen, H. (2003). Automated Detection of Epidemics from the Usage Logs of a Physicians' Reference Database. *7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 180–191. https://doi.org/10.1007/978-3-540-39804-2_18

Hoseini, C. (2020). *Leveraging Machine Learning to Identify Quality Issues in the Medicaid Claim Adjudication*

Process. PhD Dissertation, Indiana State University.

Irgens, C., Wuest, T., & Thoben, K.-D. (2012). Product state based view and machine learning: A suitable approach to increase quality? *IFAC Proceedings Volumes*, *45*(6), 1733–1738. https://doi.org/10.3182/20120523-3-RO-2023.00083

Issa, H., & Vasarhelyi, M. A. (2011). Application of Anomaly Detection Techniques to Identify Fraudulent Refunds. In *Available at SSRN 1910468* (pp. 1–19). https://doi.org/10.2139/ssrn.1910468

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). An Introduction to Statistical Learning. In *Springer Texts in Statistics*. https://doi.org/10.1016/j.peva.2007.06.006

Kai, J., Miao, H., & Gao, H. (2016). *A Survey of Quality Prediction Methods of Service-oriented Systems*. *9*(4), 183–198.

Karimi, N., Ranjbarzadeh Kondrood, R., & Alizadeh, T. (2017). An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms. *Measurement*, *107*, 68–76. https://doi.org/10.1016/J.MEASUREMENT.2017.05.009

Khoshgoftaar, T. M., & Seliya, N. (2003). Software Quality Classification Modeling Using the SPRINT Decision Tree Algorithm. *International Journal on Artificial Intelligence Tools*, *12*(03), 207–225. https://doi.org/10.1142/S0218213003001204

Kohavi, R., & Provost, F. (1998). *Glossary of Terms*. Machine Learning. http://ai.stanford.edu/~ronnyk/glossary.html

Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing Journal*. https://doi.org/10.1016/j.asoc.2015.07.018

Kumar, A. (2018). *How Honda is using cognitive search to drive real changes*. IBM Big Data & Analytics Hub. http://www.ibmbigdatahub.com/blog/how-honda-using-cognitive-search-drive-real-changes-quality-assurance

Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health Care Management Science*, *11*(3), 275–287. https://doi.org/10.1007/s10729-007-9045-4

Lieber, D., Stolpe, M., Konrad, B., Deuse, J., & Morik, K. (2013). Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning. *Procedia CIRP*. https://doi.org/10.1016/j.procir.2013.05.033

Lotfi, E., Yaghoobi, M., & Pourreza, H. R. (2008). A new approach for automatic quality control of fried potatoes using machine learning. *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, 1–4. https://doi.org/10.1109/UKRICIS.2008.4798934

Marc-Oliver Arsenault. (2017). *KOLMOGOROV–SMIRNOV TEST. A needed tool in your data science… | by Marc-Olivier Arsenault | Towards Data Science*. https://towardsdatascience.com/kolmogorov-smirnov-test-84c92fb4158d

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2014). Foundations of Machine Learning. In *Adaptive Computation and Machine Learning Serries*. MIT Press.

Morales, K. (2017). *Predictive Quality Analytics: delivering better quality, faster*. Atlassian Blog. https://www.atlassian.com/blog/add-ons/predictive-quality-analytics-delivering-better-quality-faster

Ordukaya, E., & Karlik, B. (2017). Quality Control of Olive Oils Using Machine Learning and Electronic Nose. *Journal of Food Quality*. https://doi.org/10.1155/2017/9272404

Paprzycki, M., Abraham, A., Guo, R., & Abraham, A. (2004). Analyzing Call Center Performance: A Data Mining Approach. *2.Softcomputing.Net*, *Section 3*. http://2.softcomputing.net/icfai-km.pdf%5Cnhttp://link.springer.com/chapter/10.1007/978-3-540-24677-0_112

Parra, E., Dimou, C., Llorens, J., Moreno, V., & Fraga, A. (2015). A methodology for the classification of quality of requirements using machine learning techniques. *Information and Software Technology*, *67*, 180–195. https://doi.org/10.1016/J.INFSOF.2015.07.006

Peddamuthu, B., & Srivastava, S. (2014). *QualityAssurance in Real-time (QART)*. Xerox Research Centre India. http://www.xrci.xerox.com/quality-assurance-in-real-time-qart

Pelayo, L., & Dick, S. (2007). Applying Novel Resampling Strategies To Software Defect Prediction. *NAFIPS 2007 - 2007 Annual Meeting of the North American Fuzzy Information Processing Society*, 69–72. https://doi.org/10.1109/NAFIPS.2007.383813

Petrov, D., Gutman, B. A., Shih-Hua, Yu, van Erp, T. G. M., Turner, J. A., Schmaal, L., Veltman, D., Wang, L., Alpert, K., Isaev, D., Zavaliangos-Petropulu, A., Ching, C. R. K., Calhoun, V., Glahn, D., Satterthwaite, T. D., Andreasen, O. A., Borgwardt, S., Howells, F., … Thompson, P. M. (2017). *Machine Learning for Large-Scale Quality Control of 3D Shape Models in Neuroimaging*. http://arxiv.org/abs/1707.06353

Phua, C., Lee, V., Smith-Miles, K. and Gayler, R. (2005). A Comprehensive Survey of Data Mining-based Fraud

Detection Research. *School of Business Systems, Faculty of Information Technology, Monash University*.
Rácz, A., Bajusz, D., & Héberger, K. (2019). Multi-Level Comparison of Machine Learning Classifiers and Their Performance Metrics. *Molecules (Basel, Switzerland)*, *24*(15), 2811. https://doi.org/10.3390/molecules24152811
Ribeiro, B. (2005). Support Vector Machines for Quality Monitoring in a Plastic Injection Molding Process. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, *35*(3), 401–410. https://doi.org/10.1109/TSMCC.2004.843228
RTInsights Team. (2017). *Predicting Quality Assurance Outcomes for Process Manufacturing*. https://www.rtinsights.com/predicting-quality-assurance-outcomes-for-process-manufacturing/
Shah, N., Chatterjee, A., Bhargava, A., & Joshi, N. (2016). *Model Validation, KS Test and Lorenz Curve*. Kharagpur Data Analytics Group-KDAG. https://kgpdag.wordpress.com/2016/03/14/model-validation-ks-test-and-lorenz-curve/
Shin, H., Park, H., Lee, J., & Jhee, W. C. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2012.01.105
Simon, P. (2013). *Too Big to Ignore: The Business Case for Big Data*. Wiley.
Singh, W., & Kaur, P. (2016). Comparative Analysis of Classification Techniques for Predicting Computer Engineering Students ' Academic Performance. *International Journal of Advanced Research in Computer Science RESEARCH*, *7*(6).
Siris, V. a., & Papagalou, F. (2004). Application of anomaly detection algorithms for detecting SYN flooding attacks. *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, *4*, 2050–2054. https://doi.org/10.1109/GLOCOM.2004.1378372
Sture, Ø. (2015). Automatic Quality Control of Salmon - Using Machine Learning Algorithms based on Input from a 3D Machine Vision System. *138*. https://brage.bibsys.no/xmlui/handle/11250/2352578#.WwSR2hm3SjU.mendeley
Sumesh, A., Rameshkumar, K., Mohandas, K., & Babu, R. S. (2015). Use of machine learning algorithms for weld quality monitoring using acoustic signature. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2015.04.042
Tracy, R. B. (2018). *How Machine Learning Can Transform Quality Management*. EtQ Blog. https://blog.etq.com/how-machine-learning-can-transform-quality-management
Travaille, P., Thornton, D., Müller, R. M., & Hillegersberg, J. van. (2011). Electronic Fraud Detection in the U . S . Medicaid Healthcare Program : Lessons Learned from other Industries. *Seventeenth Americas Conference on Information Systems*, 1–10. http://doc.utwente.nl/78000/1/Travaille11electronic.pdf
Tsung, F., Zhou, Z., & Jiang, W. (2007). Applying manufacturing batch techniques to fraud detection with incomplete customer information. *IIE Transactions*, *39*(6), 671–680. https://doi.org/10.1080/07408170600897510
Ucar, F., Alcin, O., Dandil, B., & Ata, F. (2018). Power Quality Event Detection Using a Fast Extreme Learning Machine. *Energies*. https://doi.org/10.3390/en11010145
van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., & van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *International Journal of Accounting Information Systems*. https://doi.org/10.1016/j.accinf.2016.04.001
Wikipedia. (n.d.). *Unstructured data*. Retrieved June 11, 2018, from https://en.wikipedia.org/wiki/Unstructured_data
Yahya. (2018). *F1 Score vs ROC AUC*. Stack Overflow. https://stackoverflow.com/questions/44172162/f1-score-vs-roc-auc
Yussupova, N., Kovács, G., Boyko, M., & Bogdanova, D. (2016). Models and methods for quality management based on artificial intelligence applications. *Acta Polytechnica Hungarica*.

Biographies

Cyrus Hoseini defended his Ph.D. in Technology Management dissertation at Indiana State University, USA. This work is from his dissertation. He has more than 20 years of work experience as a director of quality and analytics. He is now the Vice President of Quality and Analytics Advisory Services with QualMine, Inc. in California. He holds an M.B.A, a Bachelor of Science in Applied Mathematics, and has several ASQ certificates, including Six Sigma Black Belt, Certified Manager of Quality & Organizational Excellence (CMQ/OE), Certified Software Quality Engineer (CSQE), and Certified Quality Auditor (CQA). He is a voting member of US TAG to ISO/TC 176.

M. Affan Badar, Ph.D., CPEM is a Professor and former Chair of the Department of Applied Engineering and Technology Management at Indiana State University, USA. From 2016 to 2018, he was Professor and Chair of the Department of Industrial Engineering and Engineering Management at the University of Sharjah, UAE. He received a Ph.D. Degree in Industrial Engineering from the University of Oklahoma in 2002, MS in Mechanical Engineering from King Fahd University of Petroleum and Minerals in 1993, and MSc in Industrial Engineering in 1990 and BSc (Hons) in Mechanical Engineering in 1988 from Aligarh Muslim University. Dr. Badar has published more than 70 articles in refereed journals and proceedings in the areas of quality, reliability, coordinate metrology, lean manufacturing, health care, and supply chain. He is a member of IISE, IEOM, ASEM, and ASME.

A. Mehran Shahhosseini, D. Eng., is a Professor and Director of Ph.D. in Technology Management Program in the Department of Applied Engineering and Technology Management at Indiana State University, USA. He received his doctorate degree in Mechanical Engineering from Lamar University, Beaumont, Texas, USA in 1999, M.Sc. in Materials Engineering, Isfahan University of Technology, Isfahan, Iran in 1991, and B.Sc. Metallurgical Engineering, Tehran University, Tehran, Iran, in 1988. He has published over 50 articles in different journals and conference proceedings. He has served as principal and co-principal investigator for research projects sponsored by National Science Foundation, Ford Motor Company, and the US Army. Before working at Indiana State University, he was a faculty at the University of Louisville for ten years. He also has over four years of industrial experience working in different steel manufacturing plants and research laboratory of Ford Motor Company as a Senior Researcher. He is a member of ASME, ASEE, and ATMAE.

Christopher J. Kluse, Ph.D., is the program coordinator and Assistant Professor of Quality Systems at Bowling Green State University, USA. Prior to his full-time appointment in academia, he was a lecturer of Quality Management at Eastern Michigan University and an adjunct instructor for Southern New Hampshire University. Dr. Kluse was employed full-time for 25 years in the automotive industry, holding various quality management-related positions. Kluse holds a Ph.D. in Technology Management and is an ASQ Certified Manager of Quality & Organizational Excellence (CMQ/OE).