Supervised Machine Learning: Algorithms and Applications

Chapter · February 2022

DOI: 10.1002/9781119821908.ch1

1

Supervised Machine Learning: Algorithms and Applications

Shruthi H. Shetty*, Sumiksha Shetty†, Chandra Singh‡ and Ashwath Rao§

Department of ECE, Sahyadri College of Engineering & Management, Adyar, India

Abstract

The fundamental goal of machine learning (ML) is to inculcate computers to use data or former practice to resolve a specified problem. Artificial intelligence has given us incredible web search, self-driving vehicles, practical speech affirmation, and a massively better cognizance of human genetic data. An exact range of effective programs of ML already exist, which comprises classifiers to swot e-mail messages to study that allows distinguishing between unsolicited mail and non-spam messages. ML can be implemented as class analysis over supervised, unsupervised, and reinforcement learning. Supervised ML (SML) is the subordinate branch of ML and habitually counts on a domain skilled expert who "teaches" the learning scheme with required supervision. It also generates a task that maps inputs to chosen outputs. SML is genuinely normal in characterization issues since the aim is to get the computer, familiar with created descriptive framework. The data annotation is termed as a training set and the testing set as unannotated data. When annotations are discrete in the value, they are called class labels and continuous numerical annotations as continuous target values. The objective of SML is to form a compact prototype of the distribution of class labels in terms of predictor types. The resultant classifier is then used to designate class labels to the testing sets where the estimations of the predictor types are known, yet the values of the class labels are unidentified. Under certain assumptions, the larger the size of the training set, the better the expectations on the test set. This motivates the requirement for numerous area specialists or even different non-specialists giving names to preparing the framework. SML problems are grouped into classification and regression. In Classification the result

*Corresponding author: shruthihshetty02@gmail.com
†Corresponding author: sumikshashetty7@gmail.com
‡Corresponding author: chandrasingh146@gmail.com
§Corresponding author: ashu1280@gmail.com

Pradeep Singh (ed.) Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications, (1–16) © 2022 Scrivener Publishing LLC

has discrete value and the aim is to predict the discrete values fitting to a specific class. Regression is acquired from the Labeled Datasets and continuous-valued result are predicted for the latest data which is given to the algorithm. When choosing an SML algorithm, the heterogeneity, precision, excess, and linearity of the information ought to be examined before selecting an algorithm. SML is used in a various range of applications such as speech and object recognition, bioinformatics, and spam detection. Recently, advances in SML are being witnessed in solid-state material science for calculating material properties and predicting their structure. This review covers various algorithms and real-world applications of SML. The key advantage of SML is that, once an algorithm swots with data, it can do its task automatically.

***Keywords*:** Supervised machine learning, solid state material science, artificial intelligence, deep learning, linear regression, logistic regression, SVM, decision tree

1.1 History

The historical background of machine learning (ML), in the same way as other artificial intelligence (AI) concepts, started with apparently encouraging works during the 1950s and 1960s, trailed by a significant stretch of accumulation of information known as the "winter of AI" [9]. As of now, there has been an explosive concern essentially in the field related to deep learning. The start of the primary decade of the 21st century ended up being a defining moment throughout the entire existence of ML, and this is clarified by the three simultaneous patterns, which together gave an observable synergetic impact. The first pattern is big data and the second one is the reduction in the expense of equal processing and memory, and the third pattern is acquiring and building up the possibility of perceptron using deep learning algorithms. The investigation of ML has developed from the actions of a modest bunch of engineers investigating whether a machine could figure out how to solve the problem and impersonate the human mind, and a field of insights that generally overlooked computational reviews, to a wide control that has delivered basic measurable computational hypotheses of learning measures.

1.2 Introduction

ML is one of the quickest developing fields in software engineering. A lot of studies have been carried out to make machines smart; learning is one of the human characters which are made as necessary aspects of the machine

too. For example, we are standing at a crowded railway station waiting for a friend. As we wait, hundreds of people pass by. Each one looks different, but when our friend arrives we have no problem picking her out of the crowd. Recognizing people's faces is something we humans do effortlessly, but how would we program a computer to recognize a person? We could try to make a set of rules. For example, our friend has long black hair and brown eyes, but that could describe billions of people. What is it about her that you recognize? Is it the shape of her nose? But can we put it into words? The truth is that we can recognize people without ever really knowing how we do it. We cannot describe every detail of how we recognize someone. We just know how to do it. The trouble is that to program a computer, we need to break the task down into its little details. That makes it very difficult or even impossible to program a computer to recognize faces. Face recognition is an example of a task that people find very easy, but that is very hard for computers. These tasks are often called artificial intelligence or AI. ML is the subset of AI [1]. Earlier data was stored and handled by the companies. For example, each time we purchase a product, visit an official page, or when we walk around, we generate data. Every one of us is not just a generator yet also a buyer of information. The necessities are needed to be assumed also interests are to be anticipated. Think about a supermarket that is marketing thousands of products to millions of consumers either at stores or through the web store. What the market needs is to have the option to predict which client is probably going to purchase which item, to augment deals and benefits. Essentially every client needs to find the best suitable product. We do not know precisely which individuals are probably going to purchase which item. Client conduct changes as expected and by geological area. However, we realize that it is not arbitrary. Individuals do not go to store and purchase things irregular, they purchase frozen yogurt in summer and warm clothes in winter. Therefore, there are definite outlines in the data.

An application of AI strategies to an enormous information base is termed data mining [4, 17]. Data mining is an enormous volume of information handled to develop a basic model with significant use, for instance, having high perspective accuracy. To be insightful, a framework that is in a changing climate ought to be able to learn. If the framework can learn and receive such change, then the framework designer need not anticipate and give answers for every conceivable circumstance. An exact range of effective programs of ML already exists, which comprises classifiers to swot e-mail messages to study that allows us to distinguish between unsolicited mail and non-spam messages. For an immense size of data, the manual foreseeing gives an unpredictable task to individuals. To overthrow this

issue, the machine is trained to foresee the future, with the assistance of training and test datasets. For the machine to be trained, different types of ML algorithms are accessible. The computer program is supposed to study from the experience E regarding few classes of task T from performance P extent. The estimated performance of a task improves with experience [8].

ML can be implemented as class analysis over supervised, unsupervised, and reinforcement learning (RL). These algorithms are structured into a taxonomy constructed on the estimated outcome.

Unsupervised learning (UL) is a kind of AI that searches for previously undetected samples in an informational set without prior marks and with the least human management. Cluster analysis and making data samples digestible are the two main methods of UL. SML works under defined instructions, whereas UL works for the unknown condition of the results. The UL algorithm is used in investigating the structure of the data and to identify different patterns, extract the information, and execute the task [12, 15].

R) can be an idea of a hit and a preliminary strategy of knowledge. For each activity performed, the machine is given a reward point or a penalty point. On the off chance that the alternative is right, the machine picks up the prize point or gets a penalty point if there should be an occurrence of an off-base reaction. The RL algorithm is the communication between the atmosphere and the learning specialist [14]. The learning specialist depends on exploitation and exploration. The point at which the learning specialist follows up on experimentation is called exploration, and exploitation is the point at which it plays out an activity-dependent on the information picked up from the surrounding

Supervised learning (SML) algorithms function on unidentified dependent data which is anticipated from a given arrangement of identified predictors.

1.3 Supervised Learning

SML is genuinely normal in characterization issues since the aim is to get the computer to get familiar with a created descriptive framework. In SML, the data annotation is termed as a training set, whereas the testing set is unannotated data. When annotations are discrete in the value they are called class labels while the continuous numerical annotations are so-called continuous target values. SML problems are grouped into classification and regression. Classification is the type of SML where the result has discrete value and the aim is to predict the discrete values fitting to

a specific class. Regression is the type of SML that is acquired from the labeled datasets and continuous-valued result are predicted for the latest data which is given to the algorithm.

In SML, every model is a pair comprising of an input object and the desired output value. SML requires pre-labeled information. For masked occurrences, an ideal situation will take into consideration to accurately calculate and decide the class labels. This requires the taking in algorithms, to sum up from the training data to unobserved states in a "sensible" way. SML algorithm investigates the training data set and produces a derived capacity, which is utilized for planning new models. By this process, the informational set should have inputs and known outputs. SML can be classified into two types: regression and classification [12]. Regression is the sort of SML that studies the labeled datasets and anticipates a persistent output for the new information set to the algorithm. In this method, the required result is in the form of a number. Taking an example, a regression model that decides the cost of a pre-owned vehicle, ought to have numerous instances of used vehicles recently sold. It should essentially know the data sources and the subsequent output to assemble a model. In classification, the algorithm needs to plan the new information that is found in any of the two classes that are present in the dataset. The classes should be planned to one or 0 which is considered as "Yes" or "No", "snows" or "does not snow", etc. The result will be both of the classes and not a number as it was in regression. For instance, the classifier decides if an individual has an illness, the algorithm should consist of sources of input and it must be able to predict the outcome.

Some of the known SML algorithms are linear regression, logistic regression, decision tree, support vector machine (SVM), etc. [3].

1.4 Linear Regression (LR)

LR is the simplest method of regression; it is a linear approach to model the link between a scalar response and one or more descriptive variables. Few examples of the LR algorithm are predicting the stock price, exam scores, etc. In other words, it is a statistical regression process used for predictive evaluation, mainly used to solve the regression problem in ML. Assume a model with a linear relationship among the input (x) and the single output value (y). Precisely that y can be estimated through a linear combination of input (x). The input with a single value is referred to as simple LR and input with multiple values is often referred to as multiple LR. For example, consider a linear equation which consolidates a set of (x) input variable

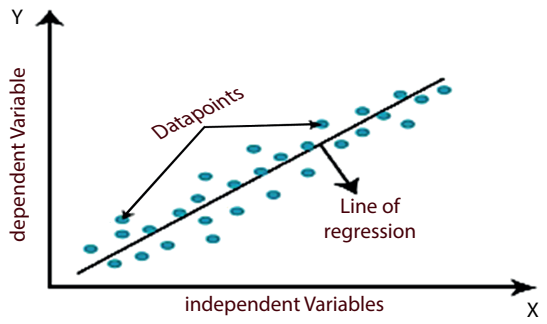


Figure 1.1 Linear regression [3].

resulting in a predicted outcome (y) for the given set of input. Hence, both the input (x) and the output value are numeric. The line equation allows one scaling factor to every input value which is called a coefficient. Another extra coefficient is added, which is often known as the intercept. To learn the LR model is to estimate the coefficient values used in the illustration of available data. Various techniques are to train the data; the most common technique used is ordinary least squares (OLS) [6]. Figure 1.1 characterizes the conspiracy between data points and LR line.

1.4.1 Learning Model

1] Simple linear regression: Single input is used to estimate the coefficients. This involves statistical calculations such as mean, standard deviations (SD), correlations, and covariance.
2] OLS: This technique is used when there is more than one input, to calculate the coefficients. This OLS method looks for minimizing the summation of the squared residuals. That is, for a given regression line through the input, the distance is calculated from every data point concerning the regression line then square it, and all together sum the squared errors. Assuming the data as a matrix, this approach uses linear algebra to calculate the coefficient values. Sufficient memory and data should be available to fit the data and to complete matrix operation [6].
3] Gradient descent: For more than one input value, the process of optimizing the coefficient values can be achieved by iteratively minimizing the errors on training data. This

procedure is termed gradient descent and works for random values for every coefficient. For every couple of input data and output, the summation of the squared errors is estimated. The coefficient values are updated in the path of diminishing the error. This procedure is repetitive up to a minimum sum-squared error is attained or no added progress is possible [6].

4] Regularization: This method looks for minimizing the sum-squared error on the training data (using OLS) and also to decrease the complexity in the model. These approaches are said to be operative when the input values are collinear and OLS overfits the training dataset [6].

1.4.2 Predictions With Linear Regression

Predicting values are more like solving an equation for the specified input. Consider an example where weight (y) is predicted from height (x). The LR equation is represented as [6]

$$y = B_0 + B_1 * x \tag{1.1}$$

or

$$weight = B_0 + B_1 * height \tag{1.2}$$

These equations can be conspired as a line in 2-dimension as shown in Figure 1.2.

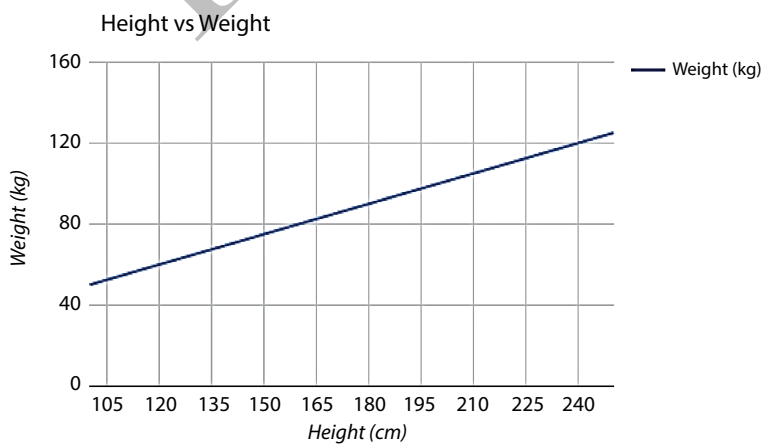


Figure 1.2 Height vs. weight graph [6].

Let B_0 be the bias coefficient and B_1 be the height column coefficient. To find the coefficients, the above learning techniques are used. Later, different height values are used to calculate the weight. For example, let $B_0 = 0.2$ and $B_1 = 0.4$, for an individual of height of 185 cm, the weight is calculated as follows [6]:

$$weight = 0.2 + 0.4 * 185 \tag{1.3}$$

$$weight = 74.2 \tag{1.4}$$

1.5 Logistic Regression

Logistic regression is well-known ML algorithms, which is under the SML technique. It is utilized for anticipating the dependent factor by making use of a given set independent factor, it is used for the classification problems, and it is dependent on the idea of probability. Logistic regression calculates the yield of a dependent variable. Thus, the outcome is a discrete value. It may be either yes or no, zero or one, and valid or invalid [3]. However, instead of giving the definite value as 0 and 1, it provides the probabilistic values which lie in the range of 0 and 1. For instance, consider that you are being given a wide scope of riddles/tests trying to comprehend which concept you are acceptable at. The result of this investigation would be considered a geometry-based issue that is 70% prone to unravel. Next is the history quiz, the chance of finding a solution is just 30%. Consider an event of detecting the spam email. LR is utilized for this event; there is a constraint of setting a limit depending on which classification is possible. Stating if the class is spam, predicted consistently is 0.4 and the limit is 0.5, the information is categorized as not a spam mail, which can prompt the outcome progressively. Logistic regression is classified as binary, multinomial, and ordinal binary can have only two possible values either yes or no or true or false where multinomial can have three or more possible values and Ordinal it manages target factors with classifications. For instance, a grade can be arranged as "very poor", "poor", "great", and "excellent".

Logistic regression is well defined as [16].

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}} \tag{1.5}$$

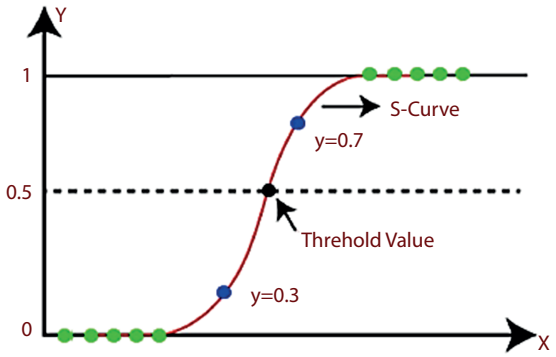


Figure 1.3 Logistic regression [3].

Figure 1.3 shows the function curve between the values 0 and 1.

1.6 Support Vector Machine (SVM)

SVMs are an influential yet adaptable type of SML which are utilized both for classification and regression. They are mainly utilized for classification problems. They use a Kernel capacity which is an essential idea for the greater part of the learning process. These algorithms make a hyperplane that is utilized to group the different classes. The hyperplane is produced iteratively, by the SVM with the target to minimize the error. The objective of SVM is to split the datasets into different classes to locate a maximum marginal hyperplane (MMH). MMH can be located using the following steps [10].

- SVM creates hyperplanes iteratively that separates the classes in a most ideal manner.
- Then, it picks the hyperplane that splits the classes accurately.

For example, let us consider two tags that are blue and black with data features p and q. The classifier is specified with a pair of coordinates (p, q) which outputs either blue or black. SVM considers the data points which yield the hyperplane that separates the labels. This line is termed as a decision boundary. Whatever tumbles aside of the line, will arrange as blue, and anything that tumbles to the next as black.

The major terms in SVM are as follows:

- Support Vectors: Datapoints that are nearby to the hyperplane are called support vectors. With the help of the data points, the separating line can be defined.
- Hyperplane: Concerning Figure 1.4, it is a decision plane that is parted among a set of entities having several classes.
- Margin: It might be categorized as the gap between two lines on data points of various classes. The distance between the line and support vector, the margin can be calculated as the perpendicular distance.

There are two types of SVMs:

- Simple SVM: Normally used in linear regression and classification issues.
- Kernel SVM: Has more elasticity for non-linear data as more features can be added to fit a hyperplane as an alternative to a 2D space.

SVMs are utilized in ML since they can discover complex connections between the information without the need to do a lot of changes. It is an incredible choice when you are working with more modest datasets that have tens to a huge number of highlights. They normally discover more precise outcomes when contrasted with different calculations in light of their capacity to deal with little, complex datasets.

Figure 1.4 shows the hyper-plane that categorizes two classes.

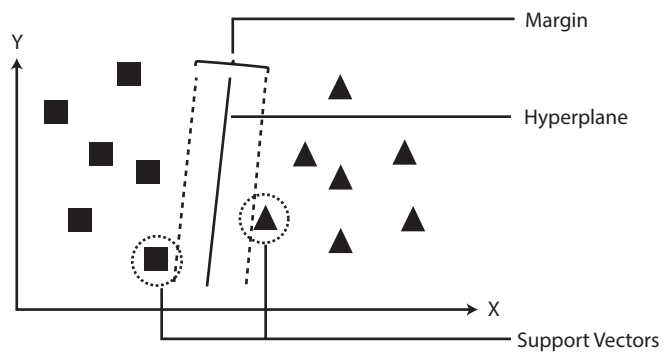
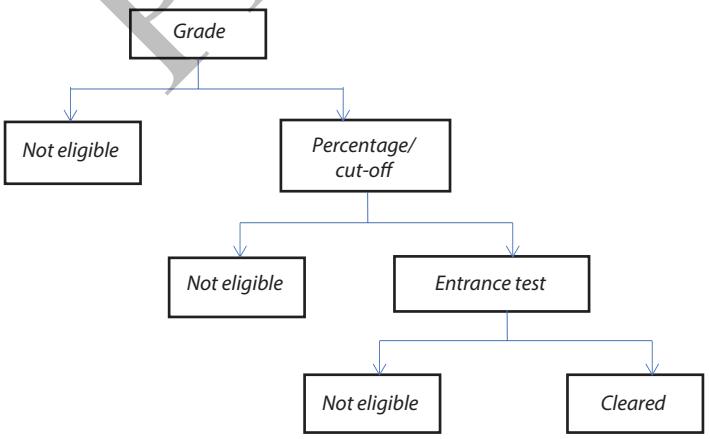


Figure 1.4 SVM [11].

1.7 Decision Tree

Decision tree groups are dependent on the element values. They utilize the strategy for Information Gain and discover which element in the dataset, give the best of data, making it a root node, etc., till they can arrange each case of the dataset. Each branch in the decision tree speaks to an element of the dataset [4]. They are one of the most generally utilized calculations for classification. An analysis of the decision tree, the decision tree is utilized to visually and signify the decision and the process of decision making. As the term suggests it utilizes a tree-like representation of choices. Tree models are the objective variable that can take a discrete arrangement of values termed as classification trees; in this tree model, leaves signify the class labels, and combinations of features of class labels are signified by the branches.

Consider an example of listing the students eligible for the placement drive. Now, the scenario is whether the student can attend the drive or not? There are "n" different deciding factors, which has to be investigated for appropriate decision. The decision factors are whether the student has qualified the grade, what is the cut-off, whether the candidate has cleared the test, and so on. Thus, the decision tree model has the following constituents:



Q1 **Figure 1.5** Decision tree.

- Root Node: The root node in this example is the "grade".
- Internal Node: The intermediate nodes with an incoming edge and more than 2 outgoing edge.
- Leaf Node: The node without an out-going edge; also known as a terminal node.

For the currently developed decision tree in this example, initially, the test condition from the root hub is tested and consigns the control to one of the active edges; thus, the condition is again tried and a hub is allocated. The tree is supposed to be ended when all the test conditions lead to a leaf hub. The leaf hub consists of class-labels, which vote against or in favor of the choice.

1.8 Machine Learning Applications in Daily Life

Some of the main areas where we use ML algorithms are in traffic alert systems in Google maps, social media sites like Facebook, in transportation and commuting services like Uber, Product recommendation systems, virtual personal assistant systems, self-driving cars, Google translators, online video streaming services, fraud detection, etc.

1.8.1 Traffic Alerts (Maps)

Nowadays, when we decide to go out and in need of assistance for directions and traffic situations on the road we have decided to travel, we usually take the help of Google maps. If in case you decided to travel to a city and decide to take the highway, and the Google traffic alert system suggested that "Even though there is heavy traffic, you are on the fastest route to your destination", how does the system know all these things? In short, it is a combined data of people actively using the service, the previous data of the route collected over the years, and also involves some own tricks which are acquired by the company to efficiently calculate the traffic. Most of the people who are currently using the Google maps service is indirectly providing their location, speed, and the routes they are going to take in which they are traveling, which helps Google collect data about the traffic, which will help the Google map algorithm predict the traffic and recommend the best routes for future users.

1.8.2 Social Media (Facebook)

Social media applications like Facebook use ML to detect and recognize faces that are used for automatic friend tagging suggestions. The algorithm compares the detected faces with the database of pictures it already has and gives users suggestions. Facebook's DeepFace algorithm which uses deep learning runs behind the Facebook application to recognize faces and identify the person in the picture. It also provides alternative tags to images already uploaded on Facebook.

1.8.3 Transportation and Commuting (Uber)

Transportation and commuting apps like Uber use ML to provide good services to their clients. It provides a personalized application that is unique to you, for example, it automatically detects your location and gives options either to go home or office or any other frequent places which will be purely based on your search history and patterns. The application uses a ML algorithm on top of historic data on trips to make accurate ETA predictions. There was an increase of 26% in the accuracy of delivery and pickup after implementing ML on their application.

1.8.4 Products Recommendations

This tells you how powerful is the ML recommendation systems are these days. Take for example, you liked an item on Amazon, but add it to your wish list because you cannot afford the item at the current price. Surprisingly, the day after, when you are watching videos on YouTube or some other application you encounter an ad for the item which you have wish-listed before. Even when you switch to another app, say, Facebook, you will still see the same ad on that website. This happens because Google tracks your search history and recommend ads depending on the activities you do. About 35% of Amazon's wealth is generated by using product recommendation systems like these.

1.8.5 Virtual Personal Assistants

Here, virtual assistant finds some useful information when the user asks some questions via text or voice. There are many applications of ML which are being in these kinds of applications. Applications involve speech verification and identification systems, speech-text conversions, NLP, and text-to-speech conversion. The only thing you have to do is ask a simple question like, "What is my schedule for tomorrow?" or maybe "Show my

upcoming booking", then assistants search for information related to questions to collect information. Recently, chatbots use a personal assistant, which is being used in many food ordering company applications, online coaching or training sites, and also many in many transport applications.

1.8.6 Self-Driving Cars

This may be one of the most breath taking the implementation of ML in the modern world. Tesla uses deep learning and other algorithms to build a self-driving car. As the computation required for this is very high, we need matching hardware to run these algorithms, NVIDIA provides the necessary hardware to run these computationally expensive models.

1.8.7 Google Translate

Before when you remember the times when you go to a new place where the language used there is completely new to you and you find it difficult to communicate with the locals or find places you wanted to go, this was mainly because you could not understand what is written on the local spots. But nowadays, Google's GNMT is a neural machine algorithm that has a dictionary of thousands of millions of words of many different languages, uses natural language processing to very efficiently and accurately translate any sentences or words. Even the tone of every sentence matters, it uses techniques like NER.

1.8.8 Online Video Streaming (Netflix)

More than a 100 million users use Netflix, and there is no doubt that it is the most-streamed web service in the whole world. Netflix application use ML algorithms which collect a massive amount of data about the users, when the user pauses, rewinds, or fast forwards. It also takes data depending on the day you watch the content, the date and time, and mainly the rating pattern and search pattern. The application collects these data from each of their users they have and use their recommendation systems and a lot of algorithms related to ML approaches.

1.8.9 Fraud Detection

Currently, online credit card fraud detection is 32 billion dollars market in 2020. That is approximately higher than the profit made by many MNC companies combined. Nowadays, the number of payment channels

like a credit card, debit card, numerous wallets, UPI, and much more has increased the number of criminals. ML approaches fraud detection as a classification problem.

1.9 Conclusion

In the anticipating years, ML will embrace a significant reason in the divulgence of data from the abundance of information that is at present open in a different zone of utilization. The supervised learning strategies are developing constantly by the information researchers, which contain an enormous arrangement of algorithms. This zone has the consideration of numerous engineers and has picked up generous advancement in the most recent decade. The learning strategies accomplish magnificent execution that would have been hard to get in the earlier many years. Given the reckless development, there is a lot of room for the engineers to work effectively and to develop the SML strategies.

Q2 ## References

1. Iqbal, M. and Yan, Z., Supervised Machine Learning Approaches: A Survey. *Int. J. Soft Comput.*, 5, 946–952, 2015, 10.21917/ijsc.2015.0133.
2. Nasteski, V., An overview of the supervised machine learning methods. *Horizons. B.*, 4, 51–62, 2017, 10.20544/HORIZONS.B.04.1.17.P05.
3. Sharma, R., Sharma, K., Khanna, A., Study of Supervised Learning and Unsupervised Learning. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET),* 8, VI, June 2020.
4. Kotsiantis, S.B., Supervised Machine Learning: A Review of Classification Techniques, in: *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth*, HCI, Information Retrieval and Pervasive Technologies, The Netherlands, 2007.
5. Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O., Akinjobi, J., Supervised Machine Learning Algorithms: Classification and Comparison. *Int. J. Comput. Trends Technol. (IJCTT)*, 48, 3, 128–138, June 2017.
6. Brownlee, J., Linear Regression for Machine Learning – Machine Learning Mastery, [online] Machine Learning Mastery https://machinelearningmastery.com/linear-regression-for-machine-learning/.

7. Brownlee, J., Logistic Regression for Machine Learning – Machine Learning Mastery, [online] Machine Learning Mastery Available at: https://machinelearningmastery.com/logistic-regression-for-machine-learning/.
8. Mitchell, T., *Machine Learning*, 432pp., McGraw Hill, Carnegie Mellon University, 2015.
9. https://www.javatpoint.com/machine-learning
10. https://www.tutorialspoint.com/machine_learning_with_python/classification_algorithms_support_vector_machine.htm
11. https://www.edureka.co/blog/supervised-learning/
12. https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning algorithms/
13. Fabris, F., Magalhães, J.P., Freitas, A.A., A review of supervised machine learning applied to ageing research. *Biogerontology*, 18, 2, 171–188, 2017.
14. Ayodele, T., Types of Machine Learning Algorithms, InTech Open, 2010, https://www.intechopen.com/chapters/10694
15. Musumeci, F. *et al.*, An Overview on Application of Machine Learning Techniques in Optical Networks. *IEEE Commun. Surv. Tutorials*, 21, 2, 1383–1408, Secondquarter 2019.
16. https://en.wikipedia.org/wiki/Logistic_regression
17. https://www.edureka.co/blog/decision-trees/
18. Mohri, M., Rostamizadeh, A., Talwalker, A., *Foundations of machine learning*, MIT Press, Cambridge, MA, 2012.
19. Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., De Felice, F., Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. *Sustainability*, 12, 492, 2020.
20. Behera, R. and Das, K., A Survey on Machine Learning: Concept, Algorithms and Applications. *Int. J. Innov. Res. Comput. Commun. Eng.*, 5, 2, 1301–1309, 2017.
21. Domingos, P., A few useful things to know about machine learning. *Commun. ACM*, 55, 10, 14–20, 2012.