Understanding the Evolution of Machine Learning Algorithms

Anagha Negi, Hetal Bhavsar

Department of Computer Science and Engineering, Faculty of Technology and Engineering, The Maharaja Sayajirao University of Baroda, Vadodara, India

Corresponding author: Anagha Negi, Email: anagha.negi.india@gmail.com

Initially the computers used to perform complex calculations and stored the data as well the instructions of how the data was handled. The evolution of computers was mainly focused on managing, manipulating and storing the data while following the instructions. Furthermore, algorithms were created that read large amounts of data and could apply classification and prediction rules that understood the data and extract insights from that data. This was the beginning of Machine Learning which is a subpart of Artificial Intelligence. The revolution of Artificial Intelligence and Machine Learning is the talking point of almost every conversation that takes place today. In this paper, we take you through the history of Machine Learning and AI through its algorithms. The development and advancement of algorithms and how it is changing the scenario today. We also provide a comprehensive and comparative analysis of different Machine Learning algorithms so that it is understood in its entirety.

Keywords: Supervised Learning, Unsupervised Learning, Reinforcement Learning, Neural Networks, Instance-based Learning.

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 255–265. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-22

1 Introduction

Humans have used a variety of instruments and tools to make certain jobs easier for ages. This ingenuity of the human mind resulted in the development of all the machines that exist today. These devices made life easier for individuals by allowing them to address a variety of demands, such as travel [1], agriculture[2]-[3], industry [4]-[5], sports [6] and computing [7]-[8]. One among them is machine learning.

In 1950, Alan Turing conceived the idea of using computers to replicate intelligent behaviour and analytical thinking. Alan Turing gave the idea of "learning machines" with the concept of computers being able to replicate intelligent behavior and analytical thinking [9]. John McCarthy in the year 1956 coined the term "Artificial Intelligence" and he defined it as 'the science and engineering of making intelligent machines.' This was the conceptualization of computers acting like humans, which we know today as intelligent machines. The term Machine learning was popularized by Arthur Samuel in 1959, he defined it as the branch of study that enables computers to learn without being explicitly programmed [10]. Arthur Samuel was well-known for his computerized checkers game, and he popularized the term Machine Learning [10].

Machine learning is used when humans are unable to interpret or understand a humungous amount of data. It teaches computers how to operate with data efficiently. It is a tool that helps transform the data into analytical information and helps extract information from the data to predict future happenings.
Every data-related problem is unique, and the same type of approach cannot be employed for each issue. Therefore, Machine learning employs a variety of algorithms for data analysis and prediction. It can read data in numerous ways. The type of method used is determined by what exactly you are searching for from the data, the number of variables involved, the best model to use, and other parameters to consider.

Let us look into different categories of Machine Learning Algorithms to understand their working and the solutions that they provide for a wide range of problems.

2 Supervised Learning

Supervised learning is defined by its use of labeled datasets to train algorithms that precisely classify data or predict the outcomes (By: IBM Cloud Education, 2020)[11]. Thus, supervised learning algorithms could be used to solve classification problems or for regression.

3 Classification

It is used when the dataset is to be categorized into discrete sets of classes. It determines the class label for an unlabeled test case. There are varied applications of classification like email filtering, speech recognition, handwriting recognition, biometric identification, document identification et. al. Decision Tree, Naïve Bayes, Linear Discriminant Analysis, k-nearest neighbor, neural networks, and Support Vector Machines are among the most used algorithms for classification problems.

4 Decision Tree Algorithm

Among all the algorithms mentioned above, Decision Tree is one of the most frequently used algorithms for classification. It is also sparsely used for regression analysis. Let us discuss a brief history of Decision Tree algorithm.

In 1959, William Belson in his paper “Matching and Prediction on the Principle of Biological Classification”, suggested an empirical way of prediction where he used punch cards for binary classification [12]. Here, the first-order predictor is divided into two groups A and non-A, and this is done with as near as possible a 50/50 split.

For the second-order predictor, the search would be different in both the A and non-A groups, but the division remains precisely the same. The split continues till either the size of the sub-group is minimized to a point that is unintelligible for analysis or the predictive power of the next predictor is too low for the concerned sub-group. This “biological classification” could be called the basis for the decision trees that we use today.

The mid 90’s to early 2000s was a time when a lot of statisticians conceived a lot of algorithms relating to decision trees and ensemble learning. Breiman’s work in machine learning and statistics is notable as he proposed the Classification and Regression Trees in 1984. CART used Gini index as the splitting criterion for the nodes to create a binary decision tree [13].

Table 1. Comparison of Decision Tree Algorithms

Methods	CART	C4.5	C5.0	CHAID	QUEST
Proposed by author in the year	Breiman, 1984	Ross Quinlan, 1993	Ross Quinlan, 1994	Kass, 1980	Loh and Shih, 1997
Measures for selecting Input variables	GINI Index	Entropy, Information Gain	Entropy, Information Gain	Chi-square Analysis	Chi-square for Categorical variables, ANOVA for continuous variables
Type of split at each node	Binary Split	Multiple Splits depending on the features	Multiple Splits depending on the features	Multiple Splits depending on the features [14]	Binary Split
Type of dependent variable	Categorical/ Continuous	Categorical/ Continuous	Categorical/ Continuous	Categorical	Continuous
Type of Independent Variable	Categorical/Continuous				
Pruning	Pre-pruning using single pass algorithm[15]	Pre-pruning using single pass algorithm [16]	Post-Pruning [17]	Pre-pruning using the Chi-Square test [14]	Post-Pruning[18]
Advantages	Simple to understand and takes less effort for data preparation	Oversees missing values very well.	C5.0 is very dependable when it comes to missing	Output is highly visual and easy to interpret.	It is unbiased in split-variable selection

			data. The training time is also low.		[19]
Disadvantages	May create over complex trees and do not generalize data very well i.e. does not handle overfitting optimally	Poor attribute splitting technique leading to overfitting. High learning cost	Even though better than its antecedents it still suffers from high variance.	It requires larger datasets to work withas it creates a wide tree with multiway splits.	It is suitable for multiple category variables but can process only binary data.

Naïve Bayes One of the other often used classifiers is Naïve Bayes. It is a probabilistic classifier based on Bayes theorem which was proposed in 1760s by Reverend Thomas Bayes. Bayes theorem focuses on getting hypothesis(H) from given evidence(E) [20]. It can be written as:

$$P(H|E) = (P(E|H) * P(H)) / P(E) \quad (1)$$

Probability of Hypothesis after getting evidence

$$= \frac{\text{Probability of E given H * Probability of H}}{\text{Probability of E.}}$$

Conceptually bayes theorem alone will not work in a real-world scenario where we have multiple input variables and multiclass labels as output. We consider multiple input variables (x1, x2, x3,…xn) to predict the output (Y). Also for Naïve Bayes assumes that all the input variables must be independent of each other [21].
Bayesian classification model is been used in wide range of applications. It is used in wireless sensor network for identifying the correctness of data scheme that combine the data into cluster heads [22]. Naïve bayes classifier is also used for Web page classification using URL in the ever growing world wide web [23].

5 Support Vector Machine

One of the other commonly used classification and regression algorithm is Support Vector Machine. The algorithm was originally proposed in 1963 by Vladimir N. Vapnik and Alexey Ya. Chervonenkis [24]. Boser et al. in their, the paper “A Training Algorithm for Optimal Margin Classifier”, suggested a training algorithm to maximize margins between training patterns in n-dimensional spaces with two classes [25]. In the consecutive paper [26], “Support Vector Networks” the authors proposed to map the input vectors in a non-linearly manner with a high dimension feature space. Whereas the previous paper [25], presented the solution as a linear blend of supporting patterns. Over the years the SVM classifier has been used widely for different classification and regression purposes. For example, 1)we have decomposition-based SVMs that breakdown the larger problem into smaller optimal sets that tackle only a couple of variables chosen carefully for an optimized output 2) the others are variant-based algorithms that help reduce the training time, as opposed to the decomposition algorithms that require a lot of time to process, here the trade-off is accuracy 3) Multiclass SVMs, these are the set of algorithms developed to handle multiclass classification problems [27]. Initially, the SVM's were designed for binary classification, but with time and evolution of data binary classification seems rather rudimentary, therefore multiple SVM algorithms like Multistage SVM[28], Hierarchical SVM [29], Binary Tree of SVM [30] came into existence for multiclass classification. Each of these algorithms is unique on their own with varied applications.

6 Regression

The advent of regression dates to the years 1805-1809, with the method of least square [31]-[32]. Whereas the word regression originally was coined by Sir Francis Galton for a biological phenomenon where the taller descendants would lean down or "regress" toward average height people [33]. Nonetheless, linear regression was the first ever algorithm to be used in statistics. Here, the weights of the next-generation seeds were compared against the parent seed where comparison of the median weights of the offspring to that of the original seed displayed a straight line with a positive slope of less than 1. However, there is a lot of research done in the field of regression analysis and diverse types of regression exist with various applications.

7 Ensemble Learning

Ensemble learning is the process of systematically generating and combining many models, such as classifiers or regressors, to tackle a specific computational intelligence issue. Ensemble learning is generally used to improve a model's performance or lessen the risk of an unintentional poor model selection. Ensemble learning is also used for assigning a confidence level to the model's decision, selecting optimal features, data fusion, incremental learning, nonstationary learning, and error correction.

8 Bagging

Let us look into a few primitive ensemble learning methods that have been the basis for modern works in the ensemble space. Papers like "Bagging Predictors", "Out-of-bag estimation", "half-and half" by Breiman, have been one of the initial works where multiple algorithms were combined together to create an ensemble [34]-[35]. His work is even more relevant today with the upsurge of Artificial Intelligence and Machine Learning. Creation of multiple variants of the predictors and aggregated together to get better accuracy [36]. In bagging predictors for regression problems, the average value was considered whereas, for classification majority voting was done. He also introduced the idea of bringing randomness to function estimation procedures to improve performance. For this New training sets were generated using the original dataset by randomising the output. In 1999, he proposed Adaptive bagging which was a hybrid of bagging and boosting [37]. The base learners were bagged together instead of a regular base learner that is generally used in boosting procedures. Also, for every boosting step it replaced the out-of-bag residuals instead of ordinary residuals.

9 Boosting

Diertterich proposed a modified version of the C4.5 algorithm is used to create an ensemble by randomizing the internal decision of splitting the tree at each node [38]. 33 different domains were tested for randomization, bagging and boosting over C4.5. An ensemble of 200 classifiers was used for bagging and randomization, for AdaBoost 100 or less classifiers were used based upon the weighted errors

[38]. The ensembles created were evaluated based on pruned as well as unpruned decision trees. The results showed that out of 33 domains tested all the ensembles had better performance than C4.5 alone [38]. The comparison was done between bagged, randomized and boosted algorithms with no noise added, and 20% added noise to the dataset. Also, the size of the dataset was increased to see if there were any changes in the results. It was found that best method in the different scenario was bagged C4.5, randomized C4.5 also was similarly good, while Adaboost did not prove to be a better choice. To sum up, the efficiency of bagging reduced with the increasing training set, unless the size of the corresponding decision tree grew along with it.

Freund & Schapire in a precursor to AdaBoost algorithm, presented an algorithm that improves the accuracy of binary classification. After which the authors have given an online allocation model which has applications in a variety of learning problems [39]. Here's when AdaBoost algorithm was initially developed and introduced. Considering Δ, Ω, λ as decision space, outcomes space and bounded loss function different examples were used to show the applications of the model [39]. For example, The k-array prediction problem was named after the model that might be used to predict the series of letters over an alphabet with size k. Another example was given that showed λ as a game of "rock, papers, scissors" [39]. AdaBoost was used for drastically reducing the error of any learning process by creating a classifier that somewhat better than arbitrary guessing.

In the following paper, the authors carried out two sets of experiments where first they compared AdaBoost with Breiman's Bagging and in the second series of experiments they made a detailed study on the performance of boosting on an OCR problem over a nearest neighbor classifier [40].

Now that we have seen the history of the ensemble learning algorithms, let us dive into the recent advances and applications of the different ensemble learning methods.

10 Semi-Supervised Learning

Semi-supervised machine learning combines the advantages of both supervised and unsupervised machine learning methods. When unlabelled data is readily available but, collecting the labelled data is a time-consuming procedure, semi-supervised learning is what you need [41]. It trains the model with an exceedingly small amount of labelled data and a big amount of labelled data. The expense of labelling a large dataset entirely is infeasible and expensive. Thus, Semi-supervised learning can be extremely useful in these instances [41]. When combined with a modest amount of labelled data, unlabelled data can increase learning accuracy significantly.

11 Reinforcement Learning

Reinforcement learning is a branch of machine learning that studies how software agents should behave in a given environment to maximize cumulative rewards[42]. It is used by a variety of software and computers to determine the best feasible action or path in a given situation. Reinforcement learning differs from supervised learning in that supervised learning includes the answer key, allowing the model to be trained with the correct answer, whereas reinforcement learning does not include an answer and instead relies on the reinforcement agent to decide what to do to complete the task [43]. As there is no training dataset it must learn through its experiences. Reinforcement learning, along with supervised and unsupervised learning, is one of the three main machine learning paradigms. Reinforcement Learning is applied for creating games emulating chess. Games like DeepBlue use parallel processing of tree-based search and customized hardware that uses 'brute force' to calculate the moves to win the game [44]. While AlphaGo uses Bayesian optimization and Monte Carlo tree search methodology along with physically watching the world champions to optimize the model and eliminate the need to use ‘brute force’ which increases the overhead when there are multiple moves and possible combinations that would make computation extremely complex [44].

12 Neural Network

A neural network consists of a series of algorithms that attempt to recognize underlying relationships in a batch of data using a method that mimics the way in which the neural connections in a human brain are built. As new neural connections are created in the human brain with every added information coming in through the senses, similarly the neural networks keep adapting to the new input data fed to the model [45]. This way, they can produce the best possible outcome without requiring the output criteria to be redesigned. The artificial intelligence-based notion of neural networks is quickly gaining traction due to its highly adaptable nature.

Consider every node to be a separate linear regression model, with input data, weights, a bias (or threshold), and an output [46]. Weights are assigned after a determination of the input layer. These weights aid in determining the significance of each variable, with larger weights having a greater impact on the final result than smaller ones. Next, each input is multiplied by its corresponding weight before being added together. The output is then determined by an activation function once the output has been passed through it. This activates the node, sending data to the network's next layer, if the output increases beyond a predetermined threshold [47]. As a result, the input of one node becomes the output of the following node. This neural network is a feedforward network since data is passed from one layer to the subsequent layer throughout this procedure [47].

13 Instance Based Learning

Instance-based learning refers to a group of learning algorithms that compare the new data with the instances present in the training samples rather than explicit generalization. These algorithms are commonly referred to as "lazy" since computation is postponed until a fresh instance is encountered[48]. Instance-based learning is called so as it creates hypotheses directly from the training examples. Instance-based learning has an advantage over other machine learning methods in that it can instantaneously modify its model to previously unknown data. Learners who use instances can simply store a new instance or discard an old one [49].

Apart from the categories of algorithms that Machine Learning offers, we have dimensionality reduction techniques that are essential for feature selection and feature extraction.

Table-II: Comparison of Different Machine Learning Algorithms

	Regression Models		Classification Models				Ensemble Models			Neural Network
Algorithm	Linear Regression	Ridge Regression	Logistic Regression	Naïve Bayes	SVM	Decision Tree	Random Forest	AdaBoost	Gradient Boosting	
Description	This algorithm shows a linear relationship between the inputs and the continuous output variable. It works	It shrinks the feature coefficient near to zero by penalizing features with lower predictive outputs	It shows a linear relationship between input and categorical output. Inspite of the name Logistic Regression is a classification algorithm	An algorithm that classifies every pair of features independent of each other. It requires less time to train	SVM creates a decision boundary that segregates n-dimensional space into different classes to classify the	It is a tree-based model where all the branches are the decision rules, the parent nodes represent the features	It combines multiple decision trees into a single model and uses majority vote for classification whereas for	It creates an ensemble of multiple weak learners by exploiting the exponential error function at each	It uses a differentiable loss function to compute the pseudo-residual values which is further used to optimize	It recognizes the relationships between features of the data set using the biological reference of the oper-
Regression/ Classification	Regression	Regression	Classification	Classification	Both	Both	Both	Both	Both	Both
Accuracy	Low	Low	Low	Low	Moderate	Moderate	High	High	High	High
Training Speed	Fast	Fast	Fast	Fast	Fast	Fast	Slow	Slow	Slow	Slow
Prediction Speed	Fast	Fast	Fast	Fast	Fast	Fast	Moderate	Fast	Moderate	Fast
Parameter Tuning	Regularization required	Regularization required	Regularization required	Required for feature extrac traction	Required	Required	Required	Required	Required	Lot of parameter tuning is required
Performance with smaller observation	Good	Good	Good	Good	Average	Average	Below Average	Below Average	Below Average	Below Average
Handling irrele-	Does not handle	Better than linear	Does not handle	Handles irrele-	Bad when target	Better with less	Does a decent job	Worsen with	Works very well	May overfit the

vant features	very well	regres- gression	very well	vant features well	classes are overlapping	number of features	when the noise ratio is low	the introduction of irrelevant features		model with irrelevant features
Automatically learns feature interactions	No	No	No	No	Not with linear SVM	Yes	Yes	Yes	Yes	Yes
Parametric	Yes	Yes	Yes	Yes	Linear SVM is parametric	No	No	No	No	No
Feature Scaling	Not required if regularized	Not required if regularized	Not required if regularized	No	Required	No	No	No	No	Required
Overfitting	Prone to overfitting						Tackles the overfitting issue of single-model algorithms			

The table above shows different types of classification, regression and ensemble learning algorithms used for supervised learning. It is observed that in general the ensemble learning algorithms have better predictive performance, accuracy and robustness as compared to single model algorithms. But the execution time is better for single model algorithms. Thus there is a trade-off between the two types based on user preference. Also, ensemble learning methods are widely popular among data scientist as they are less prone to overfitting and seem to have less errors. Also with large amounts of data deep learning has also caught the attention of the Artificial Intelligence (AI) community.

14 Conclusion

With the upsurge of AI and Machine Learning, a strong understanding of the basics of all the algorithms that constitute these domains will be very crucial. This paper provides the history of where it all began, and its applications. This understanding will help one better understand the applications of these algorithms in different fields. Also the detailed differentiation between different algorithms in this paper will surely help the researcher to deeply understand the algorithms in an easy and better manner.

References

[1] Yadav, V.: Conquer data science in travel industry: DataTrained, https://www.datatrained.com/post/usage-of-data-science-in-travel-and-tourism-industry/.
[2] Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: A Review. Sensors. 18, 2674 (2018).
[3] Burdett, H., Wellen, C.: Statistical and machine learning methods for crop yield prediction in the context of Precision Agriculture. Precision Agriculture. 23, 1553–1574 (2022).
[4] May, M.C., Frenzer, M., Lanza, G.: Teaching machine learning in learning factories with industry 4.0 use-cases. SSRN Electronic Journal. (2022).
[5] Gourisaria, M.K., Agrawal, R., Harshvardhan, G.M., Pandey, M., Rautaray, S.S.: Application of machine learning in industry 4.0. Studies in Big Data. 57–87 (2021).
[6] Richter, C., O'Reilly, M., Delahunt, E.: Machine learning in sports science: Challenges and opportunities. Sports Biomechanics. 1–7 (2021).
[7] Ma, L., Sun, B.: Machine learning and AI in marketing – connecting computing power to human insights. International Journal of Research in Marketing. 37, 481–504 (2020).
[8] Auslander, N., Gussow, A.B., Koonin, E.V.: Incorporating machine learning into established bioinformatics frameworks. International Journal of Molecular Sciences. 22, 2903 (2021).
[9] Luchini, C., Pea, A., Scarpa, A.: Artificial Intelligence in oncology: Current applications and future perspectives. British Journal of Cancer. 126, 4–9 (2021).
[10] Lee, J.A.N.: Arthur Lee Samuel, https://history.computer.org/pioneers/samuel.html.
[11]"What is supervised learning?," IBM. [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning.
[12] Belson, W.A.: Matching and prediction on the principle of biological classification. Applied Statistics. 8, 2, 65 (1959).
[13] Breiman, L. et al.: Classification and regression trees. Biometrics. 40, 3, 874 (1984).
[14] Ye, J. et al.: A chi-mic based adaptive multi-branch decision tree. IEEE Access. 9, 78962–78972 (2021).
[15] Takamitsu, T. et al.: Pre-pruning decision trees by local association rules. Lecture Notes in Computer Science. 148–151 (2004).
[16] Du, J. et al.: Cost-sensitive decision trees with pre-pruning. Advances in Artificial Intelligence. 171–179 (2007).
[17] Shamrat, F.M.J. et al.: A comprehensive study on pre-pruning and post-pruning methods of decision tree classification algorithm. 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI). (2021).
[18] Lin, C.-L., Fan, C.-L.: Evaluation of CART, CHAID, and Quest Algorithms: A case study of construction defects in Taiwan. Journal of Asian Architecture and Building Engineering. 18, 6, 539–553 (2019).
[19] Zwartjes, A. et al.: Quest: Eliminating online supervised learning for efficient classification algorithms. Sensors. 16, 10, 1629 (2016).
[20] Kurama, V.: Introduction to naive bayes, https://blog.paperspace.com/introduction-to-naive-bayes/.
[21] Hand, D.J., Yu, K.: Idiot's bayes?not so stupid after all? International Statistical Review. 69, 3, 385–398 (2001).
[22] Chu, S.-C. et al.: Identifying correctness data scheme for aggregating data in cluster heads of wireless sensor network based on Naive Bayes classification. EURASIP Journal on Wireless Communications and Networking. 2020, 1, (2020).
[23] Rajalakshmi, R., Aravindan, C.: A naive Bayes approach for URL classification with supervised feature selection and rejection framework. Computational Intelligence. 34, 1, 363–396 (2018).
[24] Schölkopf, Bernhard, Zhiyuan Luo, and Vladimir Vovk, "Empirical inference. Festschrift in honor of Vladimir N. Vapnik.", Springer Science & Business Media, 2013.

[25] Boser, B.E. et al.: A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory - COLT '92. (1992).
[26] Cortes, C., Vapnik, V.: Support-Vector Networks. Machine Learning. 20, 3, 273–297 (1995).
[27] Franc, V., Hlavac, V.: Multi-class support vector machine. Object recognition supported by user interaction for service robots. (2002).
[28] Xiao-Peng Liu et al.: A multistage support Vector Machine. Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693). (2003).
[29] Yangchi Chen et al.: Integrating support vector machines in a hierarchical output space decomposition framework. IEEE International IEEE International IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004. (2004).
[30] Fei, B., Jinbai Liu: Binary tree of SVM: A new fast multiclass training and classification algorithm. IEEE Transactions on Neural Networks. 17, 3, 696–704 (2006).
[31] Gauss, C.F., Gauss, C.F.: Theoria Motus Corporum Coelestium in sectionibus conicis Solem Ambientium. Sumtibus Frid. Perthes et I.H. Besser, Hamburgi (1809).
[32] Norman, J.: Carl Friedrich Gauss & adrien-Marie Legendre discover the method of least squaresj, https://www.historyofinformation.com/detail.php?entryid=2707.
[33] Jang, S.-H.: An analytical and numerical study of Galton-Watson branching processes relevant to population dynamics. In: ProQuest Dissertations And Theses. pp. 3129–465NASA Astrophysics Datasystem (2007).
[34] Breiman, L.: Out-of-Bag Estimation. (1996).
[35] Breiman, L.: Half and Half Bagging and Hard Boundary Points. (1998).
[36] Breiman, L.: Bagging predictors. Machine Learning. 24, 2, 123–140 (1996).
[37] Breiman, L.: USING ADAPTIVE BAGGING TO DEBIAS REGRESSIONS. (1999).
[38] Diertterich, T.G.: An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. (1999).
[39] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences. 55, 1, 119–139 (1995).
[40] Freund, Y., Schapire, R.: Experiments with a New Boosting Algorithm. (1996).
[41] Chapelle, O. et al.: Semi-supervised learning. MIT Press, Cambridge, MA. (2010).
[42] [38] Lee, M.: History of Reinforcement Learning, http://www.incompleteideas.net/book/ebook/node12.html#:~:text=Farley%20and%20Clark%20described%20another,Mendel%20and%20McClaren%2C%201970).
[43] Mahoney, C.: Reinforcement Learning: A Review of the Historic, Modern, and Future Applications of this Special Form of Machine Learning, https://towardsdatascience.com/reinforcement-learning-fda8ff535bb6, (2021).
[44] Li, T. et al.: https://blog.rebellionresearch.com/blog/ibm-s-deep-blue-vs-google-s-alphago-gary-kasparov, (2021).
[45] Neural Networks, https://www.ibm.com/cloud/learn/neural-networks, (2020).
[46] Hardesty, L.: Explained: Neural networks, (2017).
[47] T. Mummert, D. Subramanian, L. Vu, and N. Pham, “What is Reinforcement Learning?,” IBM developer, 14-Sep-2022. [Online]. Available: https://developer.ibm.com/learningpaths/get-started-automated-ai-for-decision-making-api/what-is-automated-ai-for-decision-making/.
[48] Mitchell, T.: Instance Based Learning. McGraw Hill (1997).
[49] Russell, S.J. et al.: Artificial Intelligence: A modern approach. Pearson, Harlow (2022).