

Machine Learning Tools Overview: A Survey

Dr. D. Mabuni¹, Dravidian University²

¹ Assistant Professor, Dept. of Computer Science and Technology

² Kuppam, Chittoor (District), Andhra Pradesh, India

Abstract—Machine learning systems allow to analyse data, learn relationships among the attributes from data and make effective decisions for improving the business with increased profits. The most important goal of machine learning systems is to develop an automatic tool or set of automatic tools to complete any task or sub task successfully. Machine learning is an advanced technology consisting of collections of different algorithms, techniques and tools. Machine learning tools are predominantly being used in vast varieties of real applications. Different categories of machine learning tools have been developed for the usage in the real-world different types of applications. All functionalities of many advanced machine learning tools have been automated in more than 90% of situations. Machine learning model and machine learning tool developers must have an idea for selecting the right tool for developing the desired application. The ultimate goal of any machine learning technique is to optimize hyper parameters. Machine learning models are perfectly suitable for automating the decision-making process over the large datasets. Developers should have the knowledge for selecting the right tool.

Index Terms—machine learning, machine learning tools, advanced machine learning tools, machine learning models, easy model building, automation capabilities, large quality data, model interpretability, model explainability, deep learning, patterns and trends, IOT, Auto-AI, Doctor-AI, machine learning concept, Pandas, Python, Numpy, Amazon Machine Learning, hyper parameter optimization.

I. INTRODUCTION

Machine learning is one of the most important technological advancement and it has significantly improved many companies. Important features of machine learning are: automation, scalability, generalization, predictive modelling, adaptiveness and self-adjustable. Machine learning theory draws many important elements from statistics and theory

of computation and it executes tasks such as creating mathematical models for capturing important tasks of machine learning procedure. Machine learning is a technology useful for successfully executing many tasks automatically without intervention of human beings. Automatically task execution facility causes to increase productivity, profits, fastness, handling of more tasks in less time, number of business transactions and decreases costs and time of work completion. Machine learning tools empowers and enriches data analysts, system developers, data scientists and data-engineers and provides facilities for easy model building, model deployment, model running and model management with the help of automatically task completion abilities.

Open source frameworks such as PyTorch, TensorFlow and scikit-learn are combined with machine learning for increased automation, productivity, speed of model building, effective decision making, scalability of business tasks, cost reductions etc. Machine learning systems can make better decisions with tremendous speed and in more than 95% of the cases the decisions are accurate and effective. Machine learning systems are advanced technologies that are simplifying lives of human beings so much. Many companies are successfully implementing machine learning applications and the success comes only with large quality data.

Machine learning field is an emerging paradigm based on the intelligent idea that computers can learn from data, can identify special features, desired patterns and hidden patterns, and can make decisions and predictions with minimal human intervention. Machine learning based predictive models are becoming popular for exploring and analysing large datasets. Data quality is an important feature for obtaining high accurate, reliable, scalable, high quality predictable results and validity of many survey findings. Machine learning techniques have

been successfully applied in many fields. Intelligent algorithms are used for machine learning model construction. The goal of machine learning is to create intelligent computers that can analyse, think and behave like humans.

1.1 Machine Learning Facts

Main parts of machine learning system are - data, algorithms, automation, iteration, scalability and modelling. Advantages of machine learning systems are - automation, decision making is faster, self-adjustment, adaptability, scalability, innovation, technology and workloads, extensibility, insight, business simplification, business growth, increased profits and outcomes will be good. Machine learning methods are - supervised learning, un-supervised learning, semi-supervised learning and reinforcement learning. Machine learning is a big concept and different categories of branches of machine learning are –deep learning, transfer learning, computational learning theory, statistical learning, active learning, adversarial machine learning, automation learning, quantum machine learning, decision making, predictive data analysis, robot machine learning, grammar induction, self-adjustment, self-improvement and meta-learning. Machine learning is a data driven technology, which finds various patterns in the given data. It can learn from past data and improves learning knowledge automatically.

Machine learning applications in bio-informatics are - genomics, proteomics, microarrays, system biology, text mining.

1.2 Machine Learning Models

To handle different types of datasets and to manage different types machine learning applications there is a need to develop different types of machine learning models. In order to select a specific machine learning model, one must assess abilities of existing development process, production infrastructure facilities, data quality preparation facilities, skills details of teams, ability of the model training, existing tools availability and usage details of tools and other supporting details. Complex machine learning models can produce accurate prediction results but model explainability is difficult. The main aim of machine learning is to make the computers to learn and remember the relationships in the input data and use this knowledge to solve a particular selected

problem. Machine learning means learning of a computer by its own from the past experience. Quality of the machine learning solution directly depends on the characteristics of the input data and the performance of the learning algorithms. Availability of input data is considered as the key element for constructing a machine learning model. Data science and machine learning researchers use various types of datasets for different purposes.

1.3 Machine Learning Applications

List of machine learning applications

- 1) E-mail classification
- 2) Fraud-detection
- 3) Pattern recognition
- 4) Text interpretation
- 5) Computer vision
- 6) Internet of Things (IOT)
- 7) Natural language processing
- 8) Speech processing
- 9) Online recommendation systems
- 10) Cyber security
- 11) Predictive analytics
- 12) Image recognition
- 13) Face detection
- 14) Bioinformatics
- 15) Video games

1.4 Machine Learning Advantages

Advantages of Machine Learning: Machine learning advantages are unlimited and it helps to create modernized technologies. Strictly speaking advantages of machine learning are endless.

- Machine learning algorithms and techniques can process large volumes of data in order to find desired patterns and trends.
- Powerful striking feature of machine learning is that no human intervention is needed because machine learning allows executing tasks automatically. Machine learning software systems already contain many automated tools. Machine learning can make predictions, can take good decisions and improves the algorithms on its own.
- Machine learning technology has the capability for continuous improvements in completing tasks efficiently.

- Correctness of decision making, prediction and efficiency will be increased as the experience of the machine learning algorithms execution increases.
- Machine learning has the ability to handle multidimensional, multiclass labels data as well as different varieties of data, data types and data formats dynamically.
- Now-a-days the numbers of machine learning applications are increasing rapidly.

List of machine learning advantages are:

- 1) Best algorithms, techniques and tools are available for education improvements.
 - 2) Varieties of features, good development environment facilities, abilities and capabilities are included in machine learning software for efficient handling of data.
 - 3) Machine learning algorithms and techniques have wide range of applications.
 - 4) Machine learning allows automation of every task.
- Machine learning software is a special type of artificial intelligence software. Machine learning uses quality data for effective decision making as well as to predict the outcome results of a specific situations and conditions. Machine learning tools are greatly useful for organizations in order to improve their efficiency, effectiveness, business opportunities and increased profits. Smarter business decisions, correct and better prediction outcomes can be achieved through automation of advanced machine learning tools.

1.5 Types of Machine Learning

Machine learning types are categorized into 4 types:

- 1) Supervised Learning
- 2) Unsupervised Learning
- 3) Semi-supervised Learning
- 4) Reinforcement Learning

II. LITERATURE REVIEW

L. V. Rueden et al. [1] said that machine learning systems are facing some problems with insufficient training data as well as insufficient labelled datasets. Authors have presented structured overview of multiple approaches in the machine learning field. Authors have proposed an informed machine learning concept and introduced a taxonomy of informed

machine learning approaches in the form of classification framework. Based on the taxonomy a survey was conducted successfully by describing different knowledge representations. In the field of informed machine learning many papers have been reviewed and evaluated for finding important concepts and key issues.

Machine learning has shown a great success in analysing, training, model building and model management especially in pattern recognition in multiple domains such as speech recognition, text classification, computer vision, document classification and AI based game playing. Deep learning technology, models and applications use is increasing with tremendous speed. Whenever sufficient data is not available it is not possible to create accurate machine learning model. Now-a-days many companies are developing and deploying very complex machine learning models and these models must be well documented, easily understandable, well structured, easily interpretable and clearly explainable. Labelling and feature engineering procedures are useful for integrating the knowledge into machine learning systems.

A. Karpatne et al. [2] said that data science advanced models are successfully implemented and getting good results in many diversified fields but the implementation of data science models in scientific applications is strictly limited. Authors have followed a new paradigm called theory-guided data science which is useful for increasing the applicability of machine learning models to many scientific applications. The new paradigm has helped a lot for successfully implementing many scientific applications such as medical diagnosis, quantum chemistry, robotic operations, bio-medical science, drug testing, bio-marker discovery, disease analysis, climate science, whether simulation applications, hydrology, material discovery, element property examination, turbulence modelling and research. Authors have clearly explained the necessity and importance of theory-guided data science.

M Wang et al. [3] said that machine learning analyses data closely, clearly and completely and then provides facilities to make the computers to output high-quality prediction results in solving many real time problems. Now-a-days large scale machine learning techniques are needed to find patterns from big datasets with high performance speeds.

Scalability is the main required feature of the large-scale machine learning systems. Model simplification, optimization, and parallelization are the main desired features of the largescale machine learning paradigms. Authors said that machine learning analyses data closely, clearly and completely and then provides enabling facilities to make the computers to output high-quality prediction results. X. Wu et al. [4] said that Big data deals with large volume of datasets, multiple data sources, multiple data formats, multiple input procedures, autonomous sources and complex tasks. Big data applications usages are growing fast in many fields such as research, science, engineering, medical data analysis, scientific, business, banking, biological and bio medical sciences. Authors have analysed in detail data modelling, security features, automation features, quality data preparation plans and Big data revolution plans.

Z. Akata et al. [5] have already created several benchmark objective functions in support vector machines for classifying large-scale images. All these results are one-versus many multiclass applications. The special algorithm called stochastic gradient descent allows to scale training data to very large data sizes and too many class labels. Authors experimentally proved that ranking-based algorithms do-not outperform the one-versus-many technique. Authors have suggested some good practices such as use regularization strategy for training online algorithms and cross-validation allows the optimal rebalancing of positive and negative examples.

Jarrah et al. [6] said that advanced machine learning systems are being created using intelligent software. These machine learning systems are data and performance driven, which means the main requirement is obtaining maximum classification accuracy. This paper has presented many reviews of the experimental data modelling literature with respect to large scale data relating to model capacity, model efficiency, model quality, model management, structure and design methodologies. Though machine learning applications are developing rapidly, still some people are thinking that machine learning based automated systems are not fully efficient and scalable for handling large datasets.

M. belkin et al. [7] said that deep learning techniques are aiming at the side of getting high generalization accuracy of classifiers. Deep learning techniques

require a greater number of parameters but have the capability to fit the training data exactly with correct model. To fit noisy data Gaussian kernel is required. Deep learning models usually produce overfitting results but their test results always produce high performance results.

T. Chen and C. Guestrin [8] said that tree boosting technique are reliable, efficient, highly effective and predominantly using machine learning techniques. Authors have clearly described a popular algorithm called XGBoost, which is accurate, robust, scalable and end-to-end powerful tree boosting algorithm. Authors have proposed a new sparsity aware algorithm for sparse datasets. In order to provide scalable boosting algorithm authors have given cache memory access patterns, data storage, data compression and data representation details.

Y. Roh et al. [9] said that data collection process is a toughest procedure in many machine learning systems development methodologies. Data collection is difficult because machine learning is widely used for developing many applications without having sufficient labelled data. Latest research data collection techniques combining both machine learning and data management techniques. In this paper, authors have thoroughly described data collection procedures with respect to data management point of view. Integration of machine learning techniques and data management techniques for data collection provides a very good opportunity for integrating Big data analytics and Artificial Intelligence (AI) techniques.

C. Chai et al. [10] said that the importance of machine learning is growing rapidly and the machine learning technology has revolutionized in many companies. Current machine learning technology is suffering with many challenging problems. Large amounts of labelled quality data are not available within short-time durations. Large training data sizes causes inefficiency in training and inference results. For a given machine learning task it is necessary to train lot of machine learning models. This is really a difficult task. To alleviate these difficulties one solution is applying good database techniques. There are 3 important steps in machine learning system creation process, Pre-ML, in-ML and Post-ML. Pre-ML includes high-quality data preparation, which consists of data collection, data pre-processing, data cleaning, data labelling and data storing in the

appropriate data structure. In-ML step is for creating an efficient machine learning model. The main steps in the Post-model step are – storing the model, querying, deploying and debugging the model.

R. Almutiri et al. [11] said that every action and activity is happening because data is driving the entire world. Very large amounts of data are continuously generating in many companies. This large data must be managed properly. So, there is a need to develop good and automated data processing tasks for quality data creation and management. Machine learning technologies are responsible for providing reliable procedures, methods, functionalities, rules and standards for effective processing of large data in order to gain quality data-based results. Many companies are integrating machine learning techniques and data management techniques for creating automated tools, which are used for developing and deploying many advanced machine learning systems. In this regard, authors have conducted a survey on the application of machine learning techniques in big data processing. The survey has discussed different techniques and algorithms, also discussed big data qualities, advantages, big data handling procedures. Also discussed varieties of machine learning techniques and machine learning tools that have used on a variety of fields.

Various machine learning methods that are useful for handling big data processing are: Deep learning, distributed and parallel learning, active learning, transfer learning, kernel-based learning. Deep learning uses artificial neural networks techniques. Deep learning mainly uses either supervised or unsupervised learning techniques for automatically learning hierarchical relationships. Learning from large volumes of data is not only a time consuming but also a challenging task. Active learning is a special type of learning where a sample subset of tuples is labelled and then problem is solved by learning these small labelled sets. Transfer learning comes into picture when collecting training data is costly or complex. In transfer learning the learned knowledge is collected from multiple sources and already learned knowledge is directly applied on new problems. When data sizes are very large, distributed techniques are convenient to divide and then process data in parallel.

J. Kamiri and G. Mariga [12] have explained clearly about the current machine learning research methods, emerging techniques and technologies and their implications. Authors have systematically analysed more than 100 earlier research articles published in IEEE papers. The study revealed that many researchers are using more than one algorithm for finding a solution. Feature engineering is crucial in finding the performance of the machine learning algorithms. Confusion matrix and its related functionalities are still dominating in the process of finding performances of algorithms. Some researchers are also considering processing times of algorithms. The survey has revealed that Python programming languages and its libraries are the most used tools in creating, training, testing, deployment and managing machine learning models. Research also revealed that the most used algorithms for classification and prediction are – decision trees, random forest, support vector machine, naïve Bayes and artificial neural networks.

Chiarello et al. [13] said that engineering design is the standard and systematic process of solving technical problems in order to create artefacts by satisfying constraints and following requirements. Authors also said that Data science is the advanced emerging inter-disciplinary field which uses computational systems to extract knowledge from multiple sources. The benefits will be increased a lot if these two fields are unified into a single field. Authors have identified tools, algorithms, functions and various data sources for improving these unified fields. Authors have tried to identify challenges in both the fields of research and also in their interfaces. The identified challenges are classified in the order of solving them easily. Engineering design combined with data science methodologies will create more and better opportunities for modern designers.

Iqbal said that [14] artificial intelligence is a leading technology now-a-days. Artificial intelligence techniques will create facilities for incorporating human intelligence, thinking, knowledge and other ideas into computers. Artificial intelligence-based modelling is a key element for building automated, dynamic, intelligent, robust and self-adjustable and modifiable advanced smart systems. To find solutions of many real-world issues, varieties of artificial intelligence systems such as data analytical systems, interactive systems, design systems, computational

systems, visual systems, textual and document-oriented systems, learning systems and many other automated systems. Creating an effective, dynamic and automated model is really a challenging task because application requirements, data and technologies are continuously changing. Authors have given comprehensive details of various artificial intelligence-based modelling techniques, functionalities, qualities, advantages, limitations and other comparative features. Also, automation of different real applications such as banking, health care, business, medical data, finance, trajectory management, agriculture, IOT, virtual systems management, cyber security, distributed automations and smart cities.

Companies are showing their interests in the three key points – automation, intelligent and smart computing. Automation means reducing human interactions and human efforts to the maximum extent by using data driven technologies and machine learning techniques. Intelligent means the system's ability to extract patterns, features, usable knowledge and many other insights from the data. Smart computing step consisting of self-data analysing, self-data modification, self-data cleaning, self-monitoring, self-feature engineering. Ultimately use automated, smart and state-of-the-art technologies for speed up decision making process, finding correct predictions with high quality outcomes, towards the increasing productivity and profits of companies. Artificial intelligent techniques make the computers to think in a smart and intelligent way for finding solutions of many problems. Artificial intelligence techniques enable the computers to execute many tasks such as decision making, quality data preparation, problem-solving, thinking, perception, self-adjusting, comprehension and self-improvement. Several artificial intelligence categories are – Analytical-AI, Functional-AI, Business-AI, Interactive-AI, Doctor-AI, Textual AI, automation-AI, medical-AI etc.

Kumeno and Fumihiko [15] said that in the recent times deep learning techniques and automation procedures have achieved a significant-breakthroughs. A notable point is that the actual outcomes of real software engineering researches are not effectively applying in the design, development, deployment and management of machine learning applications. The reason for this inability is that there exist remarkable differences between machine

learning procedures and traditional software engineering methodologies. Machine learning techniques are developing rapidly whereas software procedures are not developing with the same speed. Software engineering tools, procedures, plans and methodologies are very difficult to apply in developing and deploying machine learning applications. Machine learning algorithms, strategies, models, plans and other related procedures are rapidly evolving and changes and challenges are emerging continuously with high speed.

Khan and Zubair [16] said that machine learning techniques are significantly useful for developing intelligence based human level artificial intelligence, which in turn enables many real time software applications to become accurate, fast and improved prediction outcomes. Combination of machine learning tools and artificial intelligence techniques increases many quality desired features in the successful execution of many applications such as research, retail, business, traffic and urban planning, trajectory management, health care, education, quality data productions, manufacturing and accurate prediction results. Abundant and highly intelligent based machine learning tools are available for quality data preparation, for automation, and developing of intelligent applications with self-adjustment, self-monitoring, self-improving and self-management capabilities. Machine learning tools are categorized into framework, data engineering, interface, feature extraction ability, library, automated capability and platform. The strength of any machine learning model is the automation ability to learn and progress as new data are entering into the system and then able to convert this quality data into actionable knowledge. In application development latest trend is towards the usage of automated machine learning tools each and every step of project building, deployment and management.

K. E. K. Chai et al. [17] said that for software development, systematic reviews are needed for effective development and management of software systems. Majority of the time consumes for the manual reviews of research titles and abstracts. Now-a-days researchers are trying to automate the analysis, screening, scanning and the reviewing of too many abstracts. To accomplish this requirement varieties of machine learning tools are required and researches are progressing in this direction. Authors have tried

to develop a research-screeners, which are able to conduct systematic reviews of software systems.

F. Majidi [18] said that usage of automated machine learning tools is increasing rapidly. Machine learning practitioners are predominantly being used many automated machine learning tools to automate, optimize and manage the entire process of feature engineering, quality data preparation, model training, model development, model management and hyper parameter optimization. Different automated machine learning tools are evaluated in terms their performance, feature engineering facilities, predictive power, management capabilities and quality data preparation standards. Authors have successfully conducted a review of machine learning automated tools by analysing different types of features of tools and their applicability, ease of use, quality and other available features. A list of machine learning tools is prepared stating that which tools are useful in which contexts.

O. Z. Richter et al. [19] said that artificial intelligence is considered as one of the emerging fields of the higher engineering studies in educational technology area. Still clear clarity is required on artificial intelligence how it can be used as pedagogical tool on a broader scope and also how it can actually impact teaching and learning in higher in higher education. Authors have systematically reviewed many of the AI applications in higher education. Z. Lv et al. [20] have improved a machine learning algorithm called optimized Light Gradient Boosting Machine Learning Algorithm. This LGBML algorithm is used in android development for big data analytics and classification. The system contains various modules such as employee registration and login, tools usage management, attendance management and work management. The proposed algorithm is good with respect to many features such as accuracy, highly efficient, time complexity, less error rates and prediction accuracy.

V. Chen et al. [21] said that machine leaning strategy is to build trust models, facilities with increased debugging and increased decision-making accuracy. Authors have clearly identified that the fundamental requirements of any machine learning algorithms are – efficiency, increased debugging facility, usage of more and more automated tools, high predictive accuracy, reliability, self-management facilities and low complexity. S. Chancellor [22] said that the goal

of machine learning is to solve problems as quickly as possible and at the same time as simple as possible with expected quality outcomes. Now-a-days many machine learning applications are successfully implemented in many diversified domains. Computer scientists are continuously developing and deploying varieties of machine learning technology based real time applications in multiple domains. Main applications of machine learning are – medical diagnosis, research, business, retail, entertainment and many other commercial applications. Human-centered machine learning procedures are using in many applications.

III. MACHINE LEARNING TOOLS

Popular Machine Learning Tools are

1. PyTorch
2. MicroSoft Azure Machine Learning Studio
3. MLJAR AutoML
4. XGBoost
5. Vertex AI
6. Tensor Flow
7. Scikit-learn
8. Google Cloud Machine Learning Engine
9. Amazon Machine Learning
10. BigML
11. Apache Mahout
12. Rapid Miner
13. Jupyter Notebook
14. Rstudio
15. Pandas
16. Numpy
17. Apache Spark
18. DataRobot
19. Oracle Machine Learning
20. Watson Studio

1) PyTorch is an open source machine learning framework, which is useful for developing and deploying machine learning models including automation tools and deep learning models and many other variants of machine learning models such as active learning, transfer learning etc. PyTorch is specifically useful for developing deep learning architectures with very large datasets. PyTorch supports various modes such as graph mode and eager mode. Graph mode is convenient for graphical applications and eager mode is for building user-friendliness related applications.

2) MicroSoft Azure Machine Learning Studio is a framework tool, which contains no-code or low-code facilities. This tool allows users to analyse, design, develop, deploy and then manage user friendly based machine learning models and tools. In all respects this tool helps users and developers throughout all the machine learning lifecycle steps. This tool contains many features such as quality data preparation and data exploration, model analysis, model training, model building, model deployment and model management. This tool provides many built-in facilities for creating frameworks and libraries.

3) MLJAR AutoML is fully Python based package and contains convenient facilities for building various types of machine learning models that use tabular data. This tool searches many available machine learning algorithms and selects one of the best algorithms for a specific application and usage purpose. In this framework, different automated tools are integrated for quality data preparation and model building. This framework executes initiating machine learning experiments and tracking machine learning projects, data preparation, automatic feature engineering, model building and deployment, analysing model performance, hyper parameter tuning and support of all required tasks throughout the machine learning life cycle management.

4) XGBoost: The expansion of XGBoost is Extreme Gradient Boosting. It is an advanced open-source machine learning software library. This tool is particularly useful for scalable, flexible, fast, distributed applications development and management. It supports many software libraries for interacting with many languages using its in built interfaces. XGBoost supports many of the machine learning techniques such as classification, regression, deep learning, ranking, active learning and transfer learning etc.

5) Vertex AI contains many pre-trained models also it allows and provides environment for building customized projects. Vertex AI contains Python based tool-kits for developing machine learning models. Hyper parameter tuning, model training, model deployment, scalability and model management functionalities are available.

6) Tensor Flow: Tensor Flow allows developers full control in developing, training and deploying machine learning models. It contains some already pre-built in machine learning models that can be used

for finding solutions of simple problems. The main attractive feature of Tensor Flow is its dataflow graphs. Tensor Flow supports to find solutions of many problems such as natural language processing, retail, reinforcement learning, computer vision, decision making, predictive machine learning solutions and network management.

7) Scikit-learn: Scikit-learn is top open source framework for machine learning model development. It is a free software machine learning library for the Python programming language. Main algorithms executed by this tool are - classification, regression, documentation, clustering, dimensionality reduction, model selection and pre-processing. Specific algorithms supported by this tool are - support vector machines, random forest, gradient boosting, K-means and DBscan clustering. It is a specialist for implementing Python numerical and scientific libraries Numpy and SciPy.

8) Google Cloud Machine Learning Engine: Google Cloud Machine Learning Engine is particularly important and useful when huge amounts of training data are present. It is a hosted platform and allows machine learning developers, data scientists and other data management persons to build and run quality, automated, efficient and optimized machine learning models. Main characteristics of this tool are: training and prediction can be used either independently or combine. It can be used especially for complex model training. Main available facilities and features of this tool are - model training, model building, model deployment and management, deep learning and predictive learning.

9) Amazon Machine Learning (AML). Amazon Machine Learning framework contains a set of machine learning tools. AML is scalable, robust, fast, general purpose and cloud-based machine learning software application, which is predominantly used for developing machine learning applications. Important features of machine learning are: AML contains many visualization tools and wizards; AML allows users to train, build, deploy, manage and make predictions. AML supports classification, clustering and regression.

10) BigML: BigML is a powerful cloud-based machine learning framework that allows users to train, build, deploy and manage predictive models within a short period of time by using user friendly interface and with automation tools. Making better

decisions and producing accurate predictive results are the striking features of BigML. It has AutoML capability that automatically selects the best machine learning algorithm, self-selection and adjustment of hyper parameters, self-improvement facility. BigML reduces experience required and time to develop machine learning models. Next attractive features of BigML are model explainability, which allows transference in making predictions and also allows users to understand and validate prediction results. BigML allows integration of many data sources, and tools and making these combined frameworks available to complete many tasks.

11) Apache Mahout: Apache Mahout is fully open source collection of machine learning libraries designed and developed with scalable and efficient implementation features of machine learning algorithms. Set of machine learning algorithms available in this tool are - classification, clustering, regression, collaborative filtering and many other algorithms. Scalability, extensibility, interoperability, adaptability and modifiability are main features of Apache Mahout.

12) Rapid Miner: Rapid Miner is a data science framework and it is considered to be an amazing interface tool particularly useful for non-programmers. This tool works on cross-platform operating systems. Allows fast testing of data, data quality, model and prediction results. Main features are user friendly platform, user friendly interface, user friendly executions and ability to test own data and models.

13) Jupyter Notebook: Jupyter Notebook is one of the predominantly using machine learning frameworks. It is fast, scalable, transparent, efficient and multipurpose tool. Python, R and Julia are the languages supported by the Jupyter. Jupyter Notebook is a wonderful feature for Python coding. Jupyter Notebook can be accessed by using other graphical user interfaces such as anaconda navigator and win python navigator.

14) Rstudio: Rstudio is an IDE platform used with the R language. Rstudio is considered to be the best interface for the language R. Main purpose of R is to execute data analytics and data related functions. Rstudio based IDE makes R a powerful tool to use in completing many applications. Rstudio is accessible through anaconda also.

15) Pandas: Pandas is a machine learning tool consisting of collection of easiest standard libraries and it is generally used in Python programming language. Pandas contain fast, efficient, convenient and powerful data structures collections for easy data manipulation and easy execution of other functions. Its main goal is to become as the most advanced data manipulation tool in the entire world.

16) Numpy: Numpy is a powerful and very useful tool in machine learning and it is exclusively used for scientific calculations. Numpy supports for easy execution of many other advanced libraries. It is a king tool for executing almost all mathematical calculations. It contains N-dimensional data structure and many other advanced tools for data calculations. Numpy is actually a Python based tool with multi-dimensional storage facility for storing generic data. It can also handle various types of databases conveniently and comfortably.

17) Apache Spark: Apache Spark is powerful tool for distributed computing. This tool is fast, convenient, efficient and scalable. It is an open-source cluster-computing frame-work containing in-memory data processing engine. This data processing engine contains much functionality such as machine learning tools, ETL facilities, big data analytics facilities, stream processing data facilities and data analytics capabilities. Stream processing facility is one of the attractive features for distributed computing with scaling ability to large clusters.

18) DataRobot: DataRobot framework supports to develop end-to-end scalable AI applications. It provides data preparation for machine learning, for automating machine learning tools, design, deploy and manage all AI models conveniently in a scalable manner. DataRobot extends data science functionalities, prediction accuracies, decision making capabilities and interfacing capabilities. Features of DataRobot are - connectivity with multiple data sources, quality data preparation facilities, automatic data formats selection, interactive data analysis, data visualization features, data cleaning, data transformation and data enrichments.

19) Oracle Machine Learning: Oracle Machine Learning is useful for completing advanced data analytics functionalities. It combines Oracle SQL, Oracle data miner and R programming language into a single suite. Oracle Machine Learning allows to find solutions of many enterprise business problems

and accelerates design, train, deployment, management and continuously governing data science projects. Features available in Oracle Machine Learning are - security, scalability, accuracy, advanced data analytics facilities, automation functionalities, machine learning model building capabilities, quality data preparation, data transformation, data cleaning, data structuring, model building, model evaluation, model management and enrichment with latest updating functions.

20) Watson Studio: Watson Studio learning tool provides collaborative framework for building, deploying, running and easy management of many machine learning systems. Its main advantages are automated machine learning model creation and automated model management facilities. Effective model management is possible throughout the model deployment lifecycle. Watson Studio creates natural language processing models and it contains Auto-AI component which automatically completes quality data preparation, feature engineering, model building, model deployment, model running, model management including hyperparameter optimization. Automated tools, procedures, policies, frameworks, workflows enable organizations for quality data preparation, quality model construction, quality decision making and production of quality outcome results. It contains easy-to-use workflows for presenting data visual models. Main goal of machine learning is to collect right data, process right data, provide right data, operate on right data, give right data to the right user in right time.

IV. MACHINE LEARNING APPLICATIONS

- 1) Machine learning systems are used for recommendation of desired objects.
- 2) Facebook is another real-life machine learning example. It recognizes uploaded photos.
- 3) Third machine learning application is image processing.
- 4) Natural language processing is another example of machine learning.
- 5) Medical applications
- 6) Research applications
- 7) Business applications

V. CONCLUSIONS

Automated machine learning applications are rapidly developing in many companies. One must remember that before developing any machine learning application, it is very important to select appropriate machine learning tools with many desired features and containing many advanced libraries, a great and powerful multi-purpose user interface and the selected tool must support many programming languages. Required technology must be selected based on the appropriate guidance details of previous persons and experiences. Python language is becoming a standard data science language for all data science applications. A notable point is that artificial intelligence and many machine learning algorithms are combined and many integrated automated machine learning tools have been built. These built-in tools are very useful for developing and deploying real time applications. A strategic approach must be followed for selecting the right machine learning model to solve a problem. Everyone must remember an important point that there is an urgent need to analyse, design, develop and deploy intelligent, efficient, scalable, convenient, fast, automated, self-reliant, self-updatable and dynamic machine learning models to fulfil the future data processing demands. Quality data preparation, automation of all tasks, tool selection, tool training, tool usage and scalability are important features of any machine learning project.

REFERENCES

- [1] L. V. Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach and R. Heese, "Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems", Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 35, Issue: 1, 01 January 2023).
- [2] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly and Shashi, "Theory-guided data science: A new paradigm for scientific discovery from data", Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 29, Issue: 10, 01 October 2017).

- [3] Meng Wang, Weijie Fu, Xiangnan He, Shijie Hao and Xindong Wu, "A Survey on Large-Scale Machine Learning", IEEE Transactions on Knowledge and Data Engineering, Pub Date: 2020-08-11.
- [4] X. Wu, X. Zhu, GQ Wu and W. Ding, "Data Mining with Big Data", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, NO. 1, January 2014.
- [5] Z. Akata, F. Perronnin, Z. Harchaoui and C. Schmid "Good Practice in Large-Scale Learning for Image Classification", Published in: IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 36, Issue: 3, March 2014)
- [6] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha, "Efficient machine learning for big data: A review," Big Data Research, vol. 2, no. 3, pp. 87–93, 2015.
- [7] M. Belkin, S. Ma, and S. Mandal, "To understand deep learning we need to understand kernel learning," arXiv preprint arXiv:1802.01396, 2018.
- [8] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of SIGKDD, 2016, pp. 785–794.
- [9] Y. Roh, G. Heo and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective", Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 33, Issue: 4, 01 April 2021)
- [10] C. Chai, J. Wang, Y. Luo, Z. Niu and G. Li, "Data Management for Machine Learning: A Survey", Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 35, Issue: 5, 01 May 2023)
- [11] R. Almutiri, S. Alhabeeb, S. Alhumud and R. Ullah Khan, "A Survey of Machine Learning for Big Data Processing", Journal on Big Data 2022, 4(2), 97-111, Issue published 31 October 2022.
- [12] J. Kamiri and G. Mariga, "Research Methods in Machine Learning: A Content Analysis".
- [13] panelFilippo Chiarello, Paola Belingheri, Gualtiero Fantoni "Data science for engineering design: State of the art and future directions", Elsevier, Computers in Industry, Volume 129, August 2021, 103447.
- [14] Iqbal H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions", Springer, Published: 22 March 2021.
- [15] Afreen Khan and Swaleha Zubair, "Machine Learning Tools and Toolkits in the Exploration of Big Data", December 2018, International Journal of Computer Sciences and Engineering 6(12):570-575
- [16] K. E. K. Chai, R. L. J. Lines, D. F. Gucciardi and L. Ng, "Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews", Systematic Reviews volume 10, Article number: 93 published on 01 April 2021.
- [17] F. Majidi, M. Openja, F. Khomh and H. Li, "An Empirical Study on the Usage of Automated Machine Learning Tools", accepted at the International Conference on Software Maintenance and Evolution (ICSME) 2022.
- [18] O. Z. Richter, V. Marin, M. Bond and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?", International Journal of Educational Technology in Higher Education volume 16, Article number: 39 (2019)
- [19] Z. Lv, R. Lou, H. Feng, D. Chen and H. Lv, "Novel Machine Learning for Big Data Analytics in Intelligent Support Information Management Systems", ACM Transactions on Management Information SystemsVolume 13Issue 1Article No.: 7pp 1–21.
- [20] V. Chen, J. Li, J. S. Kim, G. Plumb and A. Talwalkar, "Interpretable Machine Learning", ACM Transactions on Management Information Systems, Volume 19, issue 6, January 12, 2022.
- [21] S. Chancellor, "Toward Practices for Human-Centered Machine Learning", Communications of the ACM, March 2023, Vol. 66 No. 3, Pages 78-85.