

# Empowering Innovation: The Future Landscape of Machine Learning with AutoML

<sup>1</sup>Mohammed Tajuddin,<sup>2</sup>Mohammad Sajid Khan,<sup>3</sup>Mairaj Hussain

<sup>1</sup>Associate Professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

<sup>2,3</sup>B.E Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

[mdtajuddin@lords.ac.in](mailto:mdtajuddin@lords.ac.in), [rafaykhan1710@gmail.com](mailto:rafaykhan1710@gmail.com), [hmairaj94@gmail.com](mailto:hmairaj94@gmail.com)

**Abstract:** Machine learning techniques are easier to apply and reduce the need for human records. Automated device learning (Auto ML) turn out to be superior as a development environment that attempts to determine the selection, series and evaluation of system mastering fashions for powerful class of duties and data. In this newsletter, we first offer information of Auto ML. Second, we have a look at man or woman segments of the Auto ML pipeline and cover them briefly. We additionally provide case studies on the enterprise use and effect of Auto ML, with a focus on sensible company programs. Finally, we recall open research issues and destiny regions of studies. This article presents the dominion of the artwork of Auto ML, focusing on commercial enterprise packages. It offers the maximum brand new outcomes from the maximum effective Auto ML algorithms.

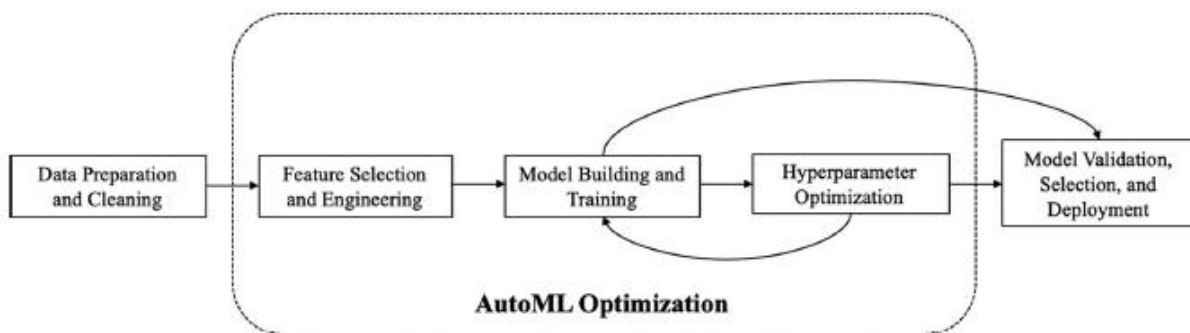
**Keywords:** AutoML algorithm, machine learning, data analysis, Hyperparameter Optimization.

## I. INTRODUCTION

Many synthetic intelligence structures are constructed with facts from multiple information supply using non-deterministic mathematical tools consisting of neural networks. An exact way to consider the usage of facts to advantage knowledge in that version. The reality is that maximum facts are often loose from mistakes. There will also be corporations which have a mixture of lacking records, incorrect records, terrible information, unlawful characters, and so forth. [1]. this is frequently the supply of risky remedy. Many studying techniques use a couple of records collection techniques, allowing topics or averages to fill in lacking data [2]. This approach, even though standard, is that it tends to rely upon the shape of the facts.

Validation involves extracting significant responses and procedures in addition to regulations and approaches. These fashions may be used to make predictions, diagnoses, scores and extra.

Although the mastering system has many benefits, the effective use of the device by way of human specialists calls for a whole lot of attempt, because there is no rule that could offer results collectively all in all questions. Although fitness researchers are acquainted with scientific facts, they often do no longer have the gear that look at the important records to aid those ideas in big datasets. Health researchers can create panels with statistical professionals [3], however the dialogue generally calls for less time and effort for every time. This isn't constantly necessary; however statistics and human information are frequently tough to obtain, especially in scientific settings [4]. Therefore, it's far tough to sing and release the tool to capture the reaction. The exercising starts with thorough research, figuring out qualified applicants, and consists of a consistent rotation of gadget learning specialists and professionals. Automating most of the additional duties that require human know-how can permit a healthcare company to quick design, put in force and configure reaction control equipment, achieving the effects of improvement excellent patient care. Driven with the aid of the application of this motive throughout industries, Auto ML has emerged as the new state of the art for optimizing legacy property for understanding pipelines, in line with as proven in Figure 1.



**Fig.1**An addition to the flute for troubles. The first step is to prepare the information. This consists of loading and cleansing the logs used in the device, as well as the use of the effects of interception, interception or get entry to. The next step is to perceive gadgets with the aim of the use of them to increase the version.

In the present day 12 months, many Auto ML solutions have emerged to enhance one or greater extra products, lots of which may be merchandise from the 2015 and 2018 Auto ML Challenge competition. Cha Kawm Auto ML's expertise of the need of occasions to method the supervisor to understand issues. With human intervention, provide a few boundaries to the calculator. However, the contemporary calculation problems aren't specific to each

opposition and generally relate to the time restriction (20 minutes for teaching and checking out) and image usage restricts (24 GB of RAM for the first 3 towers and fifty six GB after that). ). Some games cowl GPU audio, however references to these strains are uncommon. This complicated process is designed to create black content material that eliminates people's most common know-how of the usage of devices to clear up diverse troubles and additionally contributes to decreasing information and supporting the capacity of an insufficient variety of researchers with know-how at the subject. There are already three hours of hard events, every with simple formulas and statistical devices. There is now an ordinary office of statistics regarding the frame and the cause of the computer's thoughts and feelings hobby. A particular evaluation of Auto ML demanding situations from 2015 to 2018 is discussed in [5]

This article is meant to provide an accurate evaluate of traits inside the Auto ML discipline in current years. We well known the man or woman content of Auto ML and summarize the improvements made in latest years. The motivation for this newsletter comes from the shortage of mixed studies within the United States currently on Auto ML. While we understand the desirability of diverse studies, their motive is to offer an in-intensity observe of a specific stage of Auto ML, to offer a beta assessment of a couple of structures in use, or to approximate the intensity of insights into the structure.

The main findings of this newsletter are indexed below.

We destroy down the Auto ML pipeline into content and monitor engagement in each phase.

We discovered numerous new tools types now available for Auto ML and evaluated them.

We've also seen enhancements in system mastery that seem to have come via deep gaining knowledge of in latest years.

## II. REVIEW OF LITERATURE

AutoML aspires to improve the contemporary way of building ML packages through automation. ML experts can take advantage of AutoML by automating tedious responsibilities such as hyperparametric optimization (HPO), resulting in increased efficiency.

**Gatish Mehta et al. [2020]** Automated system studying is the system of using and automating the look at of facts for the problems of the arena. Automated system mastering (Auto ML) covers the whole procedure, from raw facts sets to model attention to deployable gadgets.

Auto ML originated because the sub domain device becomes used to determine the exceptional robotically suitable approach to the tool's mastering trouble. In the context of synthetic intelligence, artificial intelligence in computer systems is used in lots of approaches, together with pre-processing, elimination, and Bayesian optimization. Tools and algorithms are used in all software (enterprise, advertising and advertising, commercial enterprise, clinical and car) and labelled and defined.

**RadwaElshawi et al. [2019]**with the steady boom of the amount of information in our digital global, it is stated that the quantity of statistics scientists cannot boom to resolve tough issues. There is consequently a sturdy tendency to automate the production gadget to meet standards. In latest years, several standards and frameworks have been delivered to automate the choice of integrated and (effective) hyper parameter tuning methods inside the system gaining knowledge of space. The most important goal of the machine is to reduce the human position in the cycle and fill the gap left for clients concerned about machines without intelligence by way of gambling at the professional domain area. In this paper, we provide a complete analysis of efforts to clear up the complexities of coins.

Additionally, we highlight research on automating the back steps of the complicated device getting to know (Auto ML) manner, from facts intelligence to implementation launch. In addition, we offer suitable coverage on numerous varieties of system and frames brought to the sphere. Finally, we speak a certain variety of factors of view at the orientations and open conditions that need to be introduced together to jot down the innovative and prophetic dreams of the Auto ML generation.

**Fabio Fabris et al. [2019]**with the increasing quantity of strategies and pre-defined instructions, manually setting up an critical set of rules and hyper parameter settings (as an instance, the performance of the optimization function) is a tedious project. Automatic machine studying (Auto-ML) techniques have now been proposed to solve this problem. The Auto-ML device is intended to pick out the first-class classifiers for statistical methods. This paper analyzes the accuracy and effectiveness of the car getting to know era from the instance region, which reproduces the excessive-stage process for non-public facts. This phrase has three parts. First, we classify three sorts of automobile over fitting, regarding modifications in the estimation of accuracy measured in particular information: faculty elements (for intelligence and validity of the inner version) and store the validation procedure used in the

final evaluation. Second, we take a look at the distribution of elegance styling fashion alternatives for SK-magnificence motors throughout the seventeen digits. Third, we compare the convergence of predictive accuracy beneath facts uncertainty and unique strategies of generalization. Today, hyper-allocation degrees are considered in lots of statistical offerings, and the business enterprise's preference of trees is a ordinary choice.

**Susto et al. [2015]** this article is able to doing many things which might be understood about the procedure technique (ML) of upkeep (PdM). PdM is a vital manner to deal with protection issues as improvement desires to lessen the time and cost involved. A tough problem with PdM is the creation of so-referred to as "health devices", or quantitative indicators, famous gear for unique care troubles and their information of debt history and operational failures. The PdM generation idea lets in the usage of dynamic choice approaches for manipulate and can be used with large statistics and manipulate issues. This is accomplished with the aid of a number of observe modules with great bets to provide a great change of average overall performance in the context of frequent and unusable explosions and which then use this information inside the safety choice. The effectiveness of the technique is examined the use of a simulated example and using electronic equipment to develop troubles.

**Duan et al. [2018]** Introduce effective tools that assist you to higher understand the malicious detection technology in the safety industry. Our strategies encompass engineering work or growing new capabilities using human resources for non-purposeful obligations. We create DNF models from the previous functions, extract the Boolean functions, take benefit of Fourier evaluation to generate new ability equations, and cause them to in step with the largest variety of Fourier coefficients. We present real-global commercial enterprise case studies that enhance the engineering characteristics and overall performance of various type and clustering algorithms. Compared to the uncooked splendour of the capacity, the design has reached 50.6% hypocrisy, no extra than zero.47% incorrect. We are nonetheless searching for a higher separation from the horrific businesses while doing the meeting of the designs. A small quantity of engineering paintings is often higher than a massive amount of information, in step with our satisfaction measures. Our virtual signature generation additionally preserves interpretation that is a first-rate difficulty in digital safety packages.

**Katz et al. [2016]**one of the hardest factors to grasp is the temporal feature. We are leaving behind the research tools, the foundation of IT work. Product studies makes use of unique equipment that learn how to apprehend the whole choice capability that calls for the software program of the competitor's recent work to reap a chain of details about the nearby features. Product evaluation creates a huge range of filtering abilities by combining the information of specific buildings to attain a usual overall performance estimate primarily based on the choice standards with the assist of a character. This manner leads to properly use of new sources and better consequences compared to present day work alternatives. We reveal the effectiveness and robustness of our approach using the analysis of 25 datasets and 3 precise algorithms. We discovered that engines like Google can gather a similar provide of 20% because of bad first-rate.

**Kaul et al. [2017]**In recent years, the importance of enterprise engineering has been validated through the entire success of a deep information of ideas, which leads to those projects in numerous ways paintings For others, function engineering calls for heuristic attempt in layout and selection and is regularly tedious and non-scalable. Introducing AutoLearn, a totally regression-based commercial enterprise procedure. In preserving with fact, it does not want to be registered in the location and is therefore widely known. This explanation is performed by extracting meaningful associations the usage of pairs, analyzing linear or non-linear interactions among each pair, the use of regression, studying relationships exact, and improve prediction performance. Our scientific trials in 18 UC Irvine and seven gene expression databases, across forte areas, offer evidence that the effects obtained from our model can enhance the accuracy of the regular prediction by means of thirteen.28% compared to the location of the original characteristics and five.87% of the model. The maximum crucial distinction, across the 8 unique ratings without using any nearby understanding.

**Efimova et al. [2017]**to keep away from the log identification trouble, we need to describe legitimate techniques and alter their hyper parameters to enhance the perfect response. We don't forget the wonderful work and endorse a typically development-primarily based approach, which incorporates fashion facts integration time and hyper parameter optimization. The consequences display that the selection of extra studies methods has restricted has an effect on at the consequences showed through our techniques, which can be superior to the experimental results of Auto-WEKA and TPOT.

**IMPACT OF PRACTICAL AUTOMATED ML**

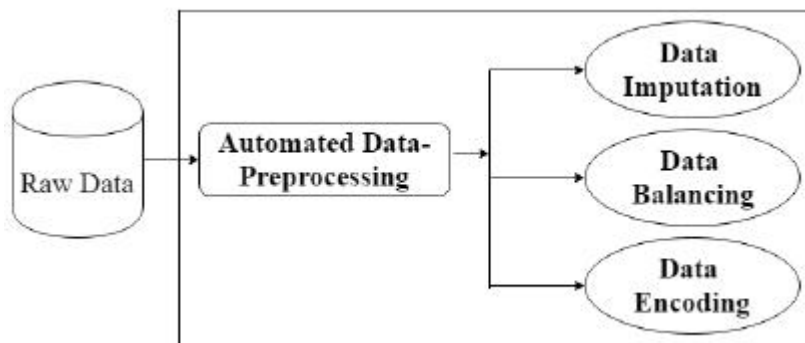
In latest years, wise era has been used in an increasing number of fields. Business programs, as an instance, consist of predictive protection and vulnerability scanning to make organizations more efficient and enhance operations. In health, the records furnished by the affected individual allows within the remedy of complex illnesses, which includes a few sclerosis, and helps medical doctors decide on the maximum suitable treatment. In the insurance and banking enterprise, the hazard related to the closing of money owed [14] and claims can be anticipated, making it viable to confirm the accuracy of the fraud. Finally, the growth in revenue and revenue forecast facilitates to improve the delivery chain [15]. However, this effective tool to decide which fashions can upload price to the company is time-consuming and mistakes-inclined, if performed manually; The average performance of different fashions have to be in comparison, which include desirable algorithms, brilliant parameter tuning and function selection. This model is repeatable and, therefore, an amazing candidate for automation. With Auto ML, the information world frees itself from this difficult challenge and can comprehend more changes, ensuing in better price for the commercial enterprise. New business fashions may be diagnosed, evaluated and verified thru speedy prototyping. In practice, Auto ML can provide correct facts. At the initial degree, using this information in facts can offer remarks approximately the suitability of the information for forecasting functions. When exceptional styles of distinct algorithms do not carry out higher than the basic ones, this could imply that there isn't always sufficient electricity in the data. At exceptional, but, precise models can be created, and the surveyor has the option of making a prototype or developing a complicated and rapid model from diverse fashions. Finally, there's a by-product in Auto ML of some of the extra optimizations during the manner: estimates of values may be received by way of determining the preliminary evaluation of the version, depending on the capacity to apply as enter. For samples. Therefore, introducing the Auto ML device directly into the business can greatly boom the price of the information scientist's art work. As an example of one of our predictive maintenance projects, wherein a crew of our full-time researchers worked for numerous weeks, the machine gained know-how of a model with a community out version below the curve (AUC) of 0. Eighty-one have been created. A few months after this fulfilment, the model of the Auto ML device after confirming the DSM turned into used, the usage of the equal set of facts (and without special assist or statistics textual content) as a measure for evaluating the high-quality of this device.

The consequences are converted into robotic turbines that perform higher than guide fashions with AUC from scratch.

**III. PROPOSED METHODOLOGY**

Prior statistics guarantees that the provision of great statistics is ensured via precise content. This is crucial due to the shortage of high-quality statistics, as most records produced and saved is regularly partial or maybe unstructured. However, although it is a critical a part of a digital device, it is the least complicated. The authors write that among 60% and 80% of researchers discover its task the toughest, the most mundane and the dullest.

In Auto ML, several statistical pre-processing, which are then implemented to a specific dataset, are encoded in cosy companies, thereby increasing the overall clarity and price of the information. We have frequently labelled these operations into the subsequent organization, specially based totally on our new research, as shown in Figure 2.



**Fig.2**Data pre-processing pipeline

**1) Data Imputation:**Data frequently additionally have missing values for serious reasons (human error or loss of information). There is easy lacking facts way: missing records at random (MAR) and lacking completely at random (MCAR). The randomness of the MCAR data is high sufficient that there's no sample bias of a particular type, unlike the MAR information, which is liable for bias. In the calculation of accuracy, we include parameters such as NaNs, fields, integer values, varieties of false information, and so forth. This problem is solved by means of changing these values with several strategies, which consist of a cheap choice wherein all the distinctive prices are removed and - the option costs take their area. Another method is to apply the suggest or median of the column throughout the data set to adjust for missing values. It uses features that include lower back-calculation bias and



regression evaluation to calculate the exchange value for a statistical column. Some strategies of calculating the facts with the time extensive use the Auto-WEKA technique. XG improvement coverage settings are also extensively utilized in TPOT and Auto-WEKA tools to combine the reality.

**2) Data Balancing:** The records evaluation is the case, even supposing one or greater lessons of the content supplied have better writing than the others. Accepting this unequal content material ends in injustice for maximum college students inside the class. It is viable to alter the version we've with statistical evaluation to make a nice complete to lessen the difference in style and this problem. In a terrific feel, the strategies used to discover the truth are inconsistencies. Files are saved or deleted in pairs to preserve dynamic beauty at some stage in the recording. Another way to enhance the worker's standard overall performance in the occasion of a computer virus is hardware from Bergman's artwork. The statistics is presently being compiled to create greater statistical information to enhance instructor facts. One way to handle statistics inconsistencies is to use price-sensitive information rather than editing statistical information. Good understanding of how to take benefit of misclassification trades to vicinity it in an unbalanced protecting style. This is important to gain fantastically skewed information in this real minority organization. In Auto ML, operators with TPOT show a way to use their API to alter elegance sensitivity to govern bias.

**3) Data Encoding:** To make the facts human-readable, college facts are frequently checked with words. Data encoding refers to changing the written text of present works into digital form to allow a computer to interpret them. Some kinds of information encoding are ordinal, warm, binary; hash encoding, goal encoding and lots of others. Goal coding is a surrogate device for a particular value of the aim/implicit mediator/model instance. Although first-rate tag encoding assigns incremental values or binary traces to every tag, the values they will be assigned do no longer represent any value inside the given data. Objective coding provides a number of crucial files that are made out of particular records, including the values of the statistics in unique objectives. H2O.AI is Auto ML that uses target coding in its API. Auto-Sklearn uses a unique and user-friendly method to coding information.

#### **4) Feature Selection**

Feature selection creates a subset of features based on the usual set of special characters by reducing dotted features or redundant capabilities. This tends to simplify the build and

therefore prevents you from over fitting and improves model performance. Specific characteristics are often heterogeneous and especially associated with assets.

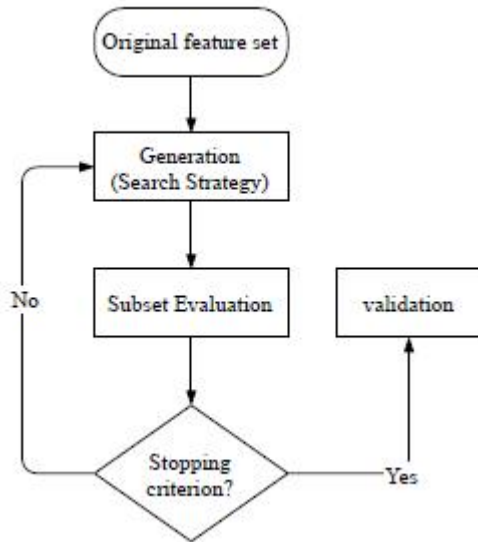


Fig.3 The iterative process of feature selection

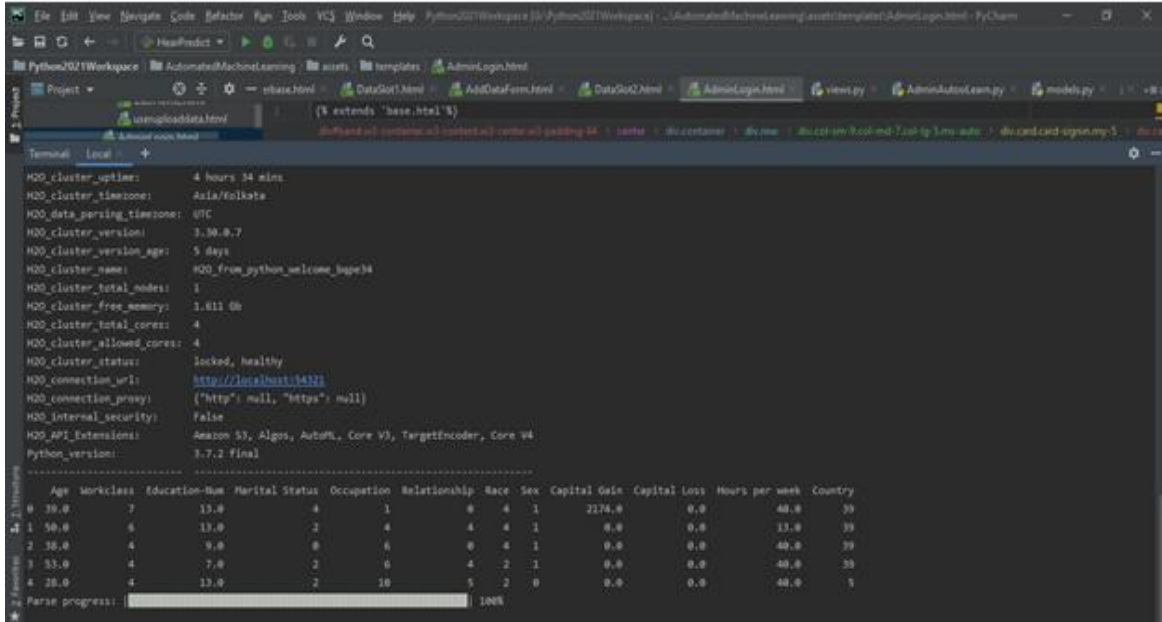
The look for strategy for characteristic assortment involves three types of algorithms: complete search, heuristic search, and chance search.

#### IV. ANALYSIS

In this examination stage, we are by means of Python 3.7 version to carry out and appraise the planned algorithm. AutoML Server used to analyse the data.

```
[22/Jul/2020 16:18:09] "GET /userlogin/ HTTP/1.1" 200 6427
Login ID = Alex Password = Alex@141
Status is = activated
User id at 2 activated
[22/Jul/2020 16:18:40] "POST /UserloginCheck/ HTTP/1.1" 200 5991
Checking whether there is an H2O instance running at http://localhost:54321 . connected.
-----
H2O_cluster_uptime: 4 hours 34 mins
H2O_cluster_timezone: Asia/Kolkata
H2O_data_parsing_timezone: UTC
H2O_cluster_version: 3.39.0-7
H2O_cluster_version_age: 5 days
H2O_cluster_name: H2O_from_python_welcome_page34
H2O_cluster_total_nodes: 1
H2O_cluster_free_memory: 1.011 Gb
H2O_cluster_total_cores: 4
H2O_cluster_allowed_cores: 4
H2O_cluster_status: locked, healthy
H2O_connection_url: http://localhost:54321
H2O_connection_proxy: {"http": null, "https": null}
H2O_internal_security: False
H2O_API_extensions: Amazon S3, Algos, AutoML, Core V3, TargetEncoder, Core V4
python_version: 3.7.2 final
-----
```

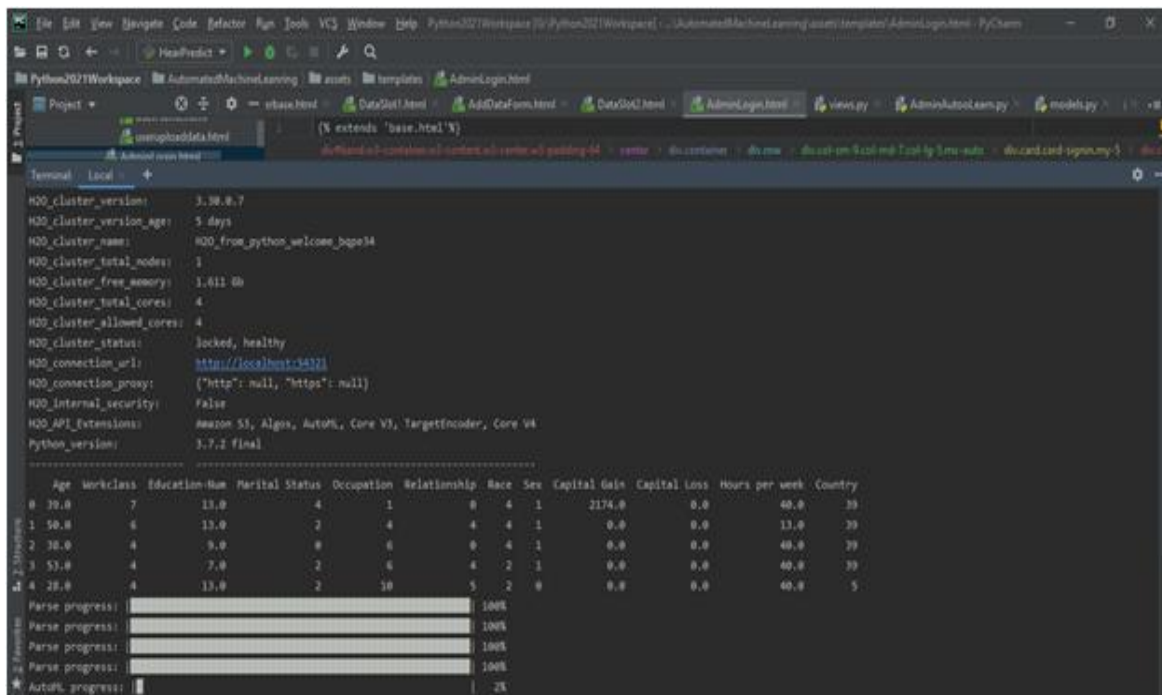
**Fig.4**AutoML Server has started to fetch data



**Fig.5**Fetching Data from shape server

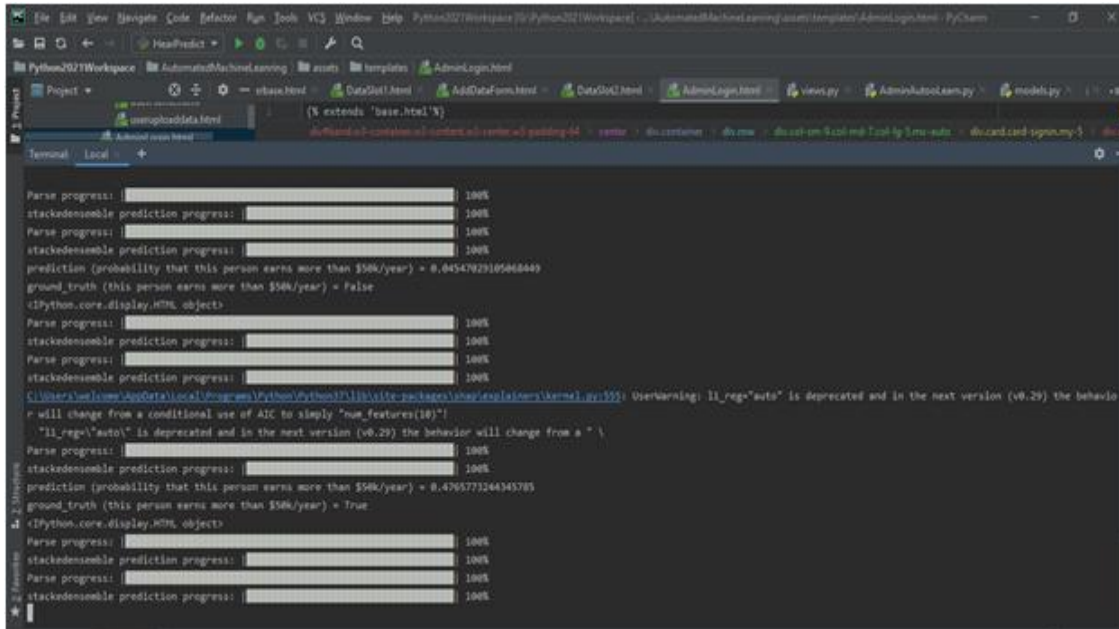
As shown in the figure 4, after server started

**Training AutoML**



**Fig.6 Training AutoML**

**Testing AutoML**



**Fig.7 Testing AutoML**

**AutoML results**

**Table.1**AutoML result table

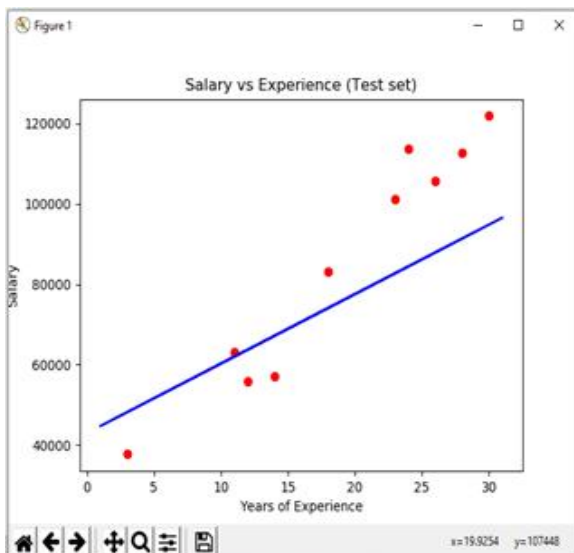
	model_id	auc	logloss	aucpr	mean_per_class_error	rmse	mse
0	StackedEnsemble_AllModels_AutoML_20200727_162032	0.918305	0.308159	0.804799	0.177644	0.309214	0.095614
1	StackedEnsemble_BestOfFamily_AutoML_20200727_1_	0.916268	0.312327	0.801839	0.176929	0.311859	0.097256
2	GBM_3_AutoML_20200727_162032	0.912290	0.413188	0.791602	0.186498	0.357587	0.127869
3	GBM_4_AutoML_20200727_162032	0.911895	0.426079	0.786651	0.178021	0.364889	0.133144
4	GBM_grid_1_AutoML_20200727_162032_model_1	0.910124	0.346165	0.788076	0.177121	0.323605	0.104720
5	GBM_1_AutoML_20200727_162032	0.909217	0.388501	0.781646	0.191528	0.345138	0.119121
6	GBM_2_AutoML_20200727_162032	0.907870	0.405938	0.767621	0.183145	0.355160	0.126139
7	DeepLearning_grid_1_AutoML_20200727_162032_mo_	0.907017	0.318388	0.770385	0.185953	0.319424	0.102031
8	GLM_1_AutoML_20200727_162032	0.904723	0.324298	0.762277	0.189650	0.321746	0.103521
9	DeepLearning_1_AutoML_20200727_162032	0.899646	0.334766	0.753148	0.206121	0.329101	0.108307
10	GBM_5_AutoML_20200727_162032	0.896027	0.463310	0.726895	0.185710	0.385155	0.148344
11	DRF_1_AutoML_20200727_162032	0.872614	1.202999	0.718824	0.207485	0.337948	0.114209
12	XRT_1_AutoML_20200727_162032	0.870409	0.994401	0.719207	0.197025	0.338989	0.114914

**Table .1** shows the result for the autoML

**Perform prediction analysis**

**Table.2** Prediction analysis

S.No	Years of Experience	Salary
1	1.1	39343.0
2	1.3	46205.0
3	1.5	37731.0
4	2.0	43525.0
5	2.2	39891.0
6	2.9	56642.0
7	3.0	60150.0
8	3.2	54445.0
9	3.2	64445.0
10	3.7	57189.0
11	3.9	63218.0
12	4.0	55794.0
13	4.0	56957.0
14	4.1	57081.0
15	4.5	61111.0
16	4.9	67938.0



**Fig.8** Salary Vs experience (Test set)

The figure.8 shows the final comparison for the salary Vs experiences for test set data

**V. CONCLUSION**

Automated system gaining knowledge of (Auto ML) is a brand new research location in pc technological know-how that has the ability to help people who do not know the way to use device getting to know out of the box. As the cost of virtual data increases rapidly, facts-

pushed techniques along with shallow and deep system gaining knowledge of are increasingly more used, increasing the want for green and comprehensive Auto ML. In this newsletter, we provide readers with statistics on several elements of Auto ML from a angle. Each of those areas has one of a kind thoughts which might be in brief described to offer a short evaluate. We additionally mentioned much stuff that has been found in current years, including signs and symptoms of unstudied areas that want unique attention. In addition, we endorse a few destiny instructions that may be explored for in addition studies within the area of auto ML.

## REFERENCES

1. Demuth, H. B., Beale M. H, De Jess, O. and Hagan, M. T. (2014). *Neural Network Design*. 2nd ed., Martin Hagan, Oklahoma, USA: Oklahoma State University.
2. Graupe, D. (2007). *Principles of Artificial Neural Networks*. 2nd Ed. New Jersey, USA: World Scientific
3. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380(14):1347–58
4. Auffray C, et al. Making sense of big data in health research: towards an EU action plan. *Genome Med* 2016;8(1):71.
5. Guyon I, et al. Analysis of the AutoML Challenge series. 2017. p. 2015–8.
6. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.
7. Elshawi, R., Maher, M., & Sakr, S. (2019). *Automated Machine Learning: State-of-The-Art and Open Challenges*. arXiv preprint arXiv:1906.02287.
8. Fabris, F., & Freitas, A. A. (2019). Analysing the Over fit of the auto-sklearn Automated Machine Learning Tool. In *Machine Learning, Optimization, and Data Science*, Vol. 11943, pp. 508- 520.
9. G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, 2015.
10. J. Duan, Z. Zeng, A. Oprea, and S. Vasudevan, "Automated generation and selection of interpretable features for enterprise security," in *2018 IEEE International Conference on Big Data (Big Data)*, pp. 1258– 1265, IEEE, 2018

11. G. Katz, E. C. R. Shin, and D. Song, "Explore kit: Automatic feature generation and selection," in Data Mining (ICDM), 2016 IEEE 16<sup>th</sup> International Conference on, pp. 979–984, IEEE, 2016.
12. A.Kaul, S. Maheshwary, and V. Pudi, "Auto learn- automated feature generation and selection," in Data Mining (ICDM), 2017 IEEE International Conference on, pp. 217–226, IEEE, 2017.
13. Efimova, A., & Shalamov, V. (2017). Fast Automated Selection of Learning Algorithm And its Hyperparameters by Reinforcement Learning. In International Conference on Machine Learning AutoML Workshop.
14. M. Handzic, F. Tjandrawibawa, and J. Yeo, "How neural networks can help loan officers to make better informed application decisions," *Informing Science*, vol. 6, pp. 97–109, 2003.
15. G. Tsoumakas, "A survey of machine learning techniques for foodsales prediction," *Artificial Intelligence Review*, pp. 1–7, 2018.
16. Prasadu Peddi and Dr. Akash Saxena (2014), "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", *International Journal of Emerging Technologies and Innovative Research* (www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available: <http://www.jetir.org/papers/JETIR1701B47.pdf>
17. Prasadu Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", *International Journal of All Research Education and Scientific Methods (IJARESM)*, ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.