Study of Machine Learning No-Code Platform

Kartikey Ahlawat

University Institute of Engineering, Chandigarh University, Mohali, Punjab, India.

Email: kartikeyprit@gmail.com

ABSTRACT: *Study on a web application based on no-code platform ML Modeller, which targets to provide predictions using various machine learning algorithms under classification and regression. Users have to upload data on which predictions are needed, the no-code platform will automatically train on the data and even show you the accuracy is achieved, now users can get required predictions. No-Code platforms will reduce Manual Data Analyzing efforts and overcome the problem of false insights and guessing. Website's performance is also evaluated if people are using it or not.*

KEYWORDS: *Machine learning, Regression, Classification, No-code platform, Automation, Website, Prediction, and Analysis.*

INTRODUCTION

Machine learning no-code platform has a huge use case in today's scenario. Machine learning-based no-code platform, to help general people train machine learning models on their datasets, just by uploading them and making predictions on the input dataset. Less than 1% of working people are software developers and the main objective of no-code platforms is to disperse the power of software to the rest 99%. Moreover, this industry is expected to cross $21 billion by the end of 2022 [1]. It will be useful for people doing business or caring for some research work or any task requiring data analysis and making predictions. As most of these people do not know how to analyze or make predictions so that they can take decisions accordingly. Currently, they might be going through data manually, which is time-consuming and might even give false insights in most cases, which could eventually result in making bad decisions. Almost anyone with the relevant data can use it and get the required predictions with just a few clicks. Gartner predicted that by 2024 at least 65% of all new applications will be using or created with low-code or no-code platforms. No-code platforms will decrease the time and cost to the company during performing tasks, eventually benefiting the company and shareholders.

Moreover, it will help in developing and deploying applications fast, resulting in a decrease in development to production time [2]. During the pandemic, in online mode of working it was observed that its workload increased substantially as compared to offline mode, majority organizations stated that IT was not able to complete all tasks in that year and the picture could have been different if some automated tool was there to ease the load [3]. ML Modeller is based on a no-code platform, which means even a person with no coding skills can avail of benefits. Users do not require any prior machine learning knowledge. ML Modeller works on images and comma-separated files (.csv) as well, containing numeric and character values. These datasets are more common, easily available, and in more use than audio datasets. (According to the current scenario).

It will help in decision-making for businessmen, recruiters, analysts, students, teachers, and the list continues. Less time-consuming and more accurate predictions than manual analysis. It will be useful for people doing business or caring for some research work or any task which requires data analysis and making predictions. Users have to upload two excel sheets, one would be on which model will get trained (Training set) and the other which contains input values. And then click upload to get the corresponding predictions to input values and accuracy and metrics for evaluation (MSE, RMSE, and MAE). And the best part is no need to log in to do so, moreover, websites would not store any user data. Moreover, a Blog platform will also be integrated where users can sign-up and start creating content for the community [4].

OBJECTIVE

The main objective is to make the power of machine learning easily available for normal people having no prior knowledge of coding or machine learning. Under no-code platform technology, users have to just upload a dataset and choose the appropriate machine learning model/algorithm and then can have insights as well as predictions. Moreover, users can check the accuracy of the model as a classifier or metrics for evaluation like root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE). Less time-consuming and more accurate predictions than manual analysis [5].

REVIEW OF LITERATURE

At the time of the creation of ML Modeller, there was no such application in the market. And the existing state-of-the-art is writing machine learning code manually and then getting predictions or making observations according to one's insightfulness, this requires the knowledge of machine learning and coding. But now there has been more development made in the field of ml-based no-code platforms such as Google had launched a similar app- 'Teachable Machine', which only works on image and audio datasets. AWS launched 'SageMaker' accepts all kinds of data like text, numeric, image, etc, moreover automating the ml pipeline i.e. create, train, and deploy. Similarly, in 2019 Microsoft released the AutoML feature in ml.net, for the purpose of automating the ml pipeline.

ML Modeller works on images and moreover on comma-separated files (.csv) as well, containing numeric and character values. These datasets are more common, easily available, and in more use than audio datasets (according to the current scenario) [1]. ML Modeller is based on a no-code platform, which means even a person with no coding skills can avail of benefits. Users do not require any prior machine learning knowledge. It will help in decision-making for businessmen, recruiters, analysts, students, teachers, and the list continues. Less time-consuming and more accurate predictions than manual analysis. The only benefit of writing ML code manually is that one can build data specific ml model which might perform better than the website (Higher accuracy) but only if the person knows ML enough to do so [5].

PROBLEM STATEMENT AND SOLUTION

In this section, we discuss the data prediction problems that people around the globe face irrespective of their knowledge in technical terms.

Problem Statement

Machine learning is considered a branch of artificial intelligence, the main idea is that computers not only use pre-written algorithms but also learn how to solve problems. Starting from the analysis of traffic jams and ending with self-driving cars, more and more tasks are transferred to self-learning machines. But the problem arises when people with no prior machine learning knowledge or coding skills require machine learning models as decision-making tools in their daily life. Apart from this, Data Analyzing and prediction making models are costly and due to which even most of the technology companies are unable to afford this technique but here an introduction to self-learning machine models and their better accuracy plays a crucial role to avail them the same benefit as a Machine Learning knowledgeable person or tech giant companies.

Solution

As stated in the above problem statement self-learning machine models play a crucial role to overcome this situation and also it makes the machine models more accurate, reliable, effective, less effort, and time-consuming model at the cheapest affordability. ML Modeller is designed to take care of research-related work and business alliances [5].

Table 1: Drawbacks of existing state of the art and how ML Modeller overcome them.

Sl.	The existing state-of-the-art	Drawbacks in existing state of art	Overcome
1.	Coding the ml model from scratch.	→ Need to have programming knowledge. → Need to know how to code like python, R, etc. → Need to have knowledge about different ml models and when to apply which. → Time-consuming	→ No need to have programming knowledge as the process is automated, you only have to check predictions and metrics for evaluation (MSE, RMSE and MAE) and accuracy. → No need to know how to code as it is done at the backend for you. → No need to have prior knowledge about ML models →Saves time, as users do not have to write code.
2.	Going through data manually	→ Time-consuming → False insights → This results in making bad decisions.	* (Above points apply)

ML Modeller is invented to solve:
(1) Saving time: Saves time, as users do not have to write code.
(2) Data: The website is not storing any user data.
(3) No need to have programming knowledge as the process is automated, you only have to check predictions and metrics for evaluation (MSE, RMSE, and MAE) and accuracy.
(4) No need to know how to code as it is done at the backend for you.
(5) No need to have prior knowledge about ML models.

(6) No need to go through data manually, as it might give false insight and eventually result in making bad decisions [5].

PERFORMANCE EVALUATION

The application has been deployed on the web in August 2021. And ever since application is given consistent growth. Moreover, let's analyze the service-share per country with per month growth rate from the stats below.
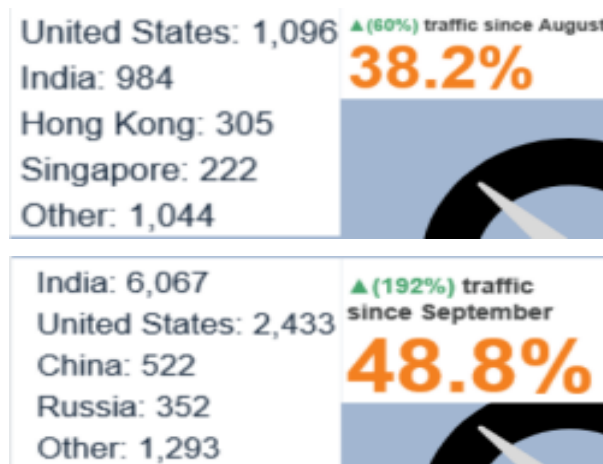

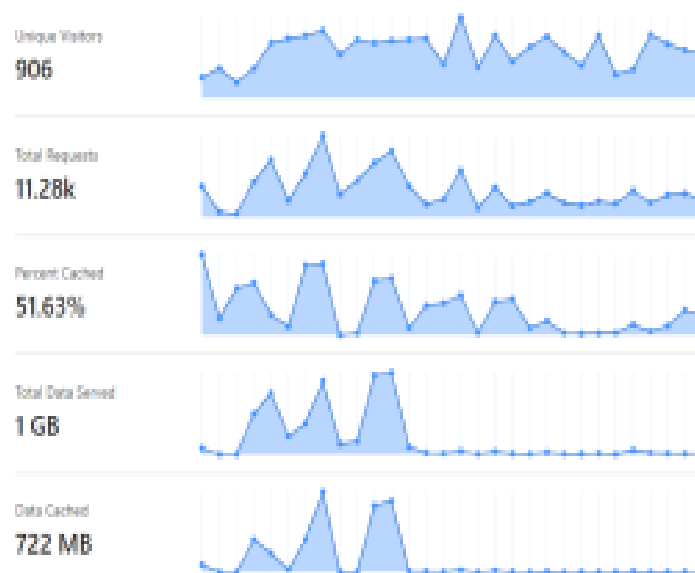**Figure 1: Web traffic country wise and increase from the previous month**

[Source: newsletter@cloudflare.com] [3]

Figure 2: Basic website statistics [3]

So by observing the above data (Figure 1.) there is significant growth in traffic. Moreover by observing figure 2 (data is from 9th October to 8th November); unique visitors/users are also

4

considerably balanced with existing visitors/users. It is as important as existing visitors/users to revisit as it is to gain or attract new visitors/users. Total request is 11.28K, which is considered decent for a new service-based web application. Percent cached is 51.63%, which is poor. The reason being is that lately more users are from India and the server on which the app is running is in the USA, this causes CDN misses (CDN→ Content Delivery Network), as the local servers do not always cache the user data. But if you are data is cached then the next time you visit the app URL it will take less time to load and increase the throughput and decrease the overhead processing [3].

Figure 3: HTTP versions used by visitors [3]

On observing figure 3, we observed that most of the users access the website via HTTP/1.1 closely following HTTP/3. HTTP/3 is the latest and faster as it follows UDP and its previous versions follow TCP. HTTP/1.0 is the oldest among these and least used to access 'ML Modeller' [3].
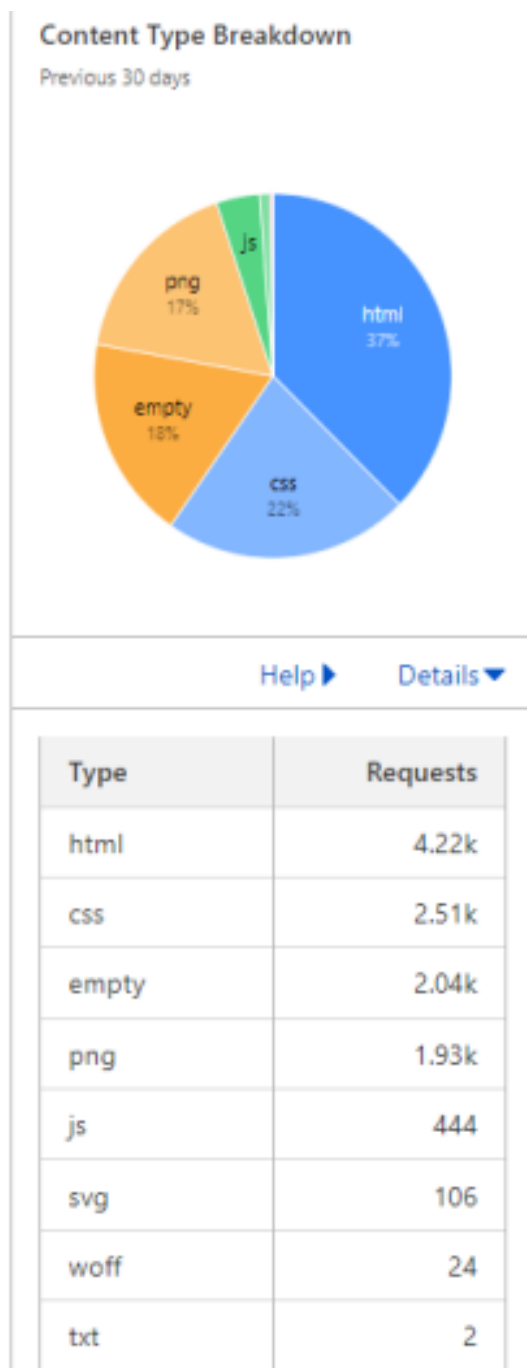
Content Type Breakdown
Previous 30 days

Type	Requests
html	4.22k
css	2.51k
empty	2.04k
png	1.93k
js	444
svg	106
woff	24
txt	2

Figure 4: Content-Type Breakdown [3]

Data in figure 4 represents the breakdown by content type of all traffic flowing through Cloudflare to the website (including both cached and uncached responses). The classifications include: "jpeg",

"HTML", "png", "gif", "CSS", "javascript", "JSON", "octet-stream", "plain", "ocsp-response", "x-shockwave-flash", "XML", "mixed", "jpg", "SVG", "webp" Note: For "empty" there was either no content-type header or the content header was empty [3].

WORKING OF APPLICATION

Users have to upload two excel sheets, one would be on which model will get trained (Training set/dataset file) and the other which contains input values (Input file). And then click upload to get the corresponding predictions to input values and accuracy and metrics for evaluation (MSE, RMSE, and MAE). And the best part is no need to log in to do so, moreover, websites would not store any user data [5].

Note: 'Input file' will only contain independent feature values. 'Dataset file' will contain training data, containing both dependents as well as independent features.

Figure 5: Instruction to be followed while uploading dataset on the website [5]

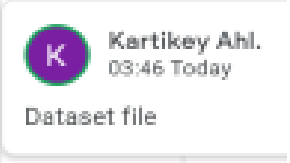Below are the snapshots (Figure 6 and Figure 7), which might help in understanding file structure:

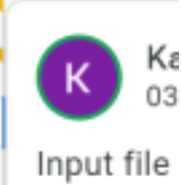**Figure 6: Dataset file on which model is trained [5]**

Figure 7: Input file on which user need predictions [5]

METHODOLOGY

Collection of Datasets

Have to gather data according to requirement/specifications. In our case Data training and input data are given by the user [6].

Installation of Libraries

Before getting our hands dirty in code, we have to install some required libraries. These libraries have a diverse set of ml functions and also increase the execution speed of programs [6].

Data Pre-processing

In this step, data is pre-processed and converted into a well-structured table of information. There are five different steps involved under it, as stated under:

Importing the Libraries

In this, we import desired libraries needed in developing models. Like some of the usual libraries are NumPy → for solving mathematical expressions, matplotlib.pyplot → for visualization of the result, and Pandas → for importing dataset [6].

Importing the Dataset

At this step with the help of the pandas library, we import the dataset, on which our model will be built. After this, we have to split the data set into independent (X) and dependent (y) variables [6].

Taking Care of Missing Data

Sometimes data contains missing values in columns. These missing values can reduce a model's accuracy. We replace these missing values (Nan) with the mean of the whole column [6].

Splitting the Dataset into the Training Set and Test Set

Via using the train_test_split method we split the dataset into training (on which model is trained) and test set (on which model is tested). Splitting results in 4 variables- X_train, X_test, y_train and y_test [6].

Feature Scaling

Sometimes values in our dataset differ/deviate very much. Like one column containing values from 100 to 1000 and another column containing values from 0 to 30. So in this case, by default first column will have more effect on the dependent variable (output) than the other. So to scale all features in one scale so there is no partiality on basis of value, on the dependent variable [6].

Choosing a Model

'ML Modeller' provides the service of classification and regression models.

Classification model users can choose from:
- Decision Tree
- Kernel SVM
- K-nearest neighbour
- Linear SVM
- Naive Bayes
- Random Forest

Regression model users can choose from:
- Simple Linear Reg.
- Multiple Linear Reg.
- Polynomial Reg.
- Support Vector Reg.
- Decision Tree Reg.
- Random Forest Reg. [6]

Training Model

Now website will train the chosen model on the user dataset. The model will then be able to find trends present in the dataset [6].

Evaluating

In this step, via scikit-learn will import the accuracy_score, mean_squared_error, and mean_absolute_error, from which in the case of the classifier model website will be able to get the model's accuracy by calculating the count of false negative and false positive. And in the case of the regressor model website will get Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as metrics for evaluation [6].

Hyperparameter Tuning

It is the problem of choosing a set of optimal hyperparameters for a learning algorithm (the algorithm used for the training model, e.g. SVM, KNN, etc). A hyperparameter is a parameter whose value is used to control the learning process. After performing it, we get optimal parameters that will give us the best accuracy. We put these parameters in the training step, i.e step 7.5. [2].

Note: In ML Modeller we have kept ideal hyperparameters and we are not implementing it explicitly [6].
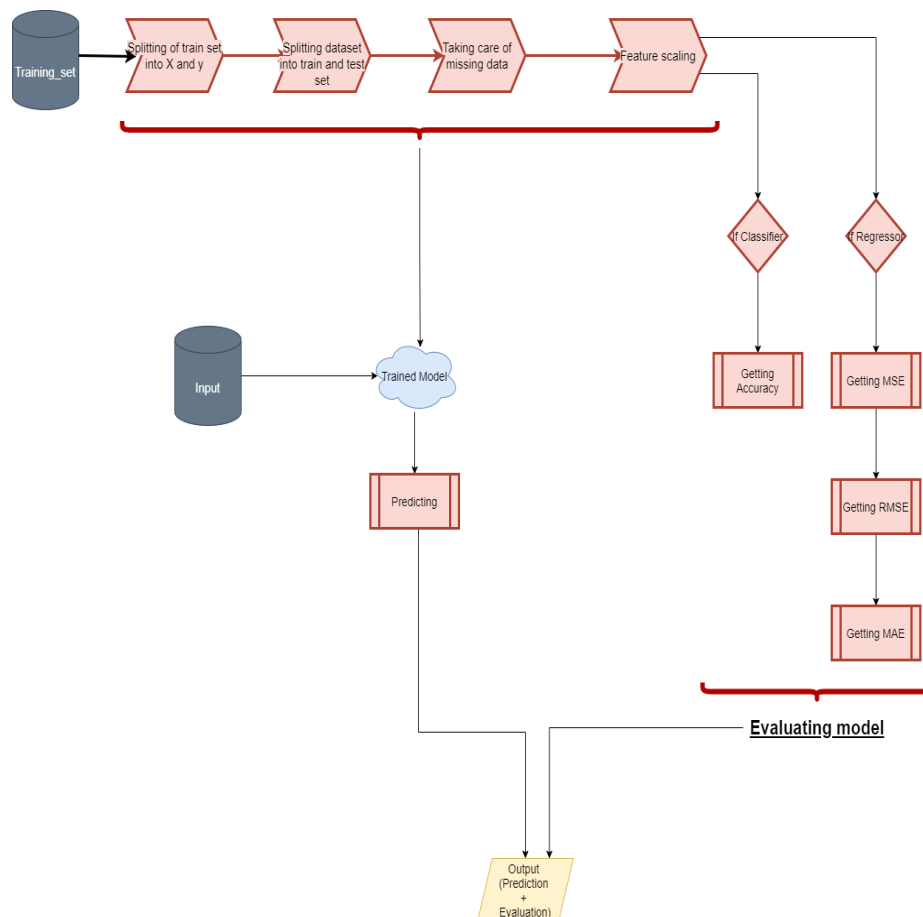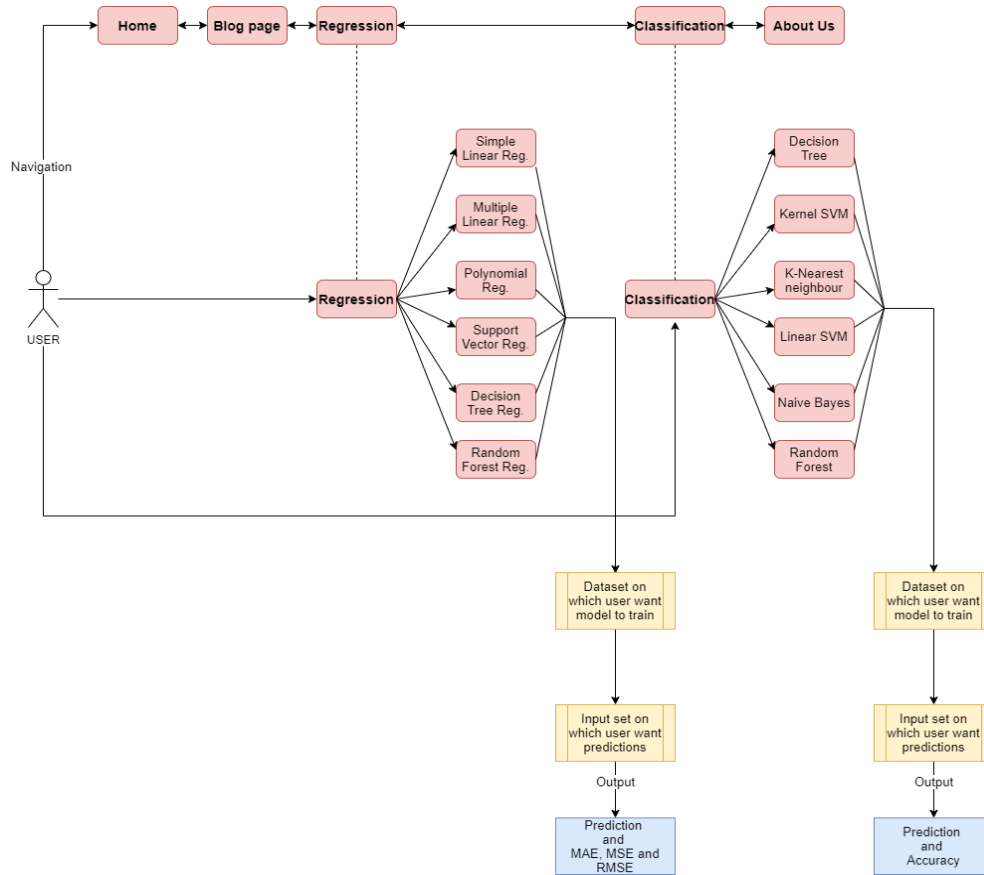

Figure 8: Data Flow Diagram

Figure 9: ER diagram of no-code platform

PROGRAMMING LANGUAGES USED

ML Modeller consists of 4 programming languages, namely: HTML, CSS, javascript, and python.
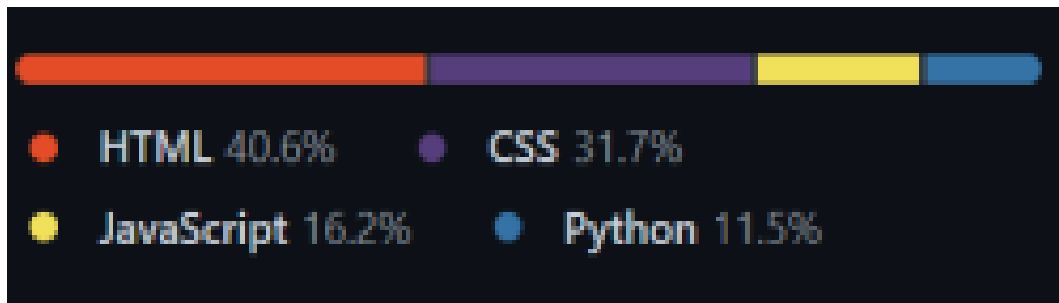

Figure 10: Line graph representing % of languages used in website

HTML used for creating the skeleton for the website, to which user can interact, an UI. CSS is used for styling of the HTML page, also used bootstrapping as well. JavaScript is used to create dynamic alerts and add DOM functionalities such as on-click or mouse hover effect. Finally, python is used for backend like, in creating machine learning codes for classification and regression models, and developing API which is the backbone of the application [6].

FUTURE PROSPECTS

In the near future we are planning to add one more service which will tell users which machine learning model suits best for their dataset by checking evaluation metrics. This service would be useful for both type of visitors, one being who does not have prior machine learning knowledge and are dependent on the website, and secondly for those who know machine learning but wants to save time on deciding a model for training [7]. Users with no prior machine learning knowledge would then not have to go to each model for raining and evaluating the accuracy and then deciding the best one. This would also attract visitors with Machine learning knowledge and ML Modeller's market hold will expand with all types of different users [8].

THE EXISTING STATE OF THE ART

Currently, there is no such application in the market. And the existing state-of-the-art is writing machine learning code manually and then getting predictions or making observations according to one's insightfulness, this requires the knowledge of machine learning and coding. Google had launched a similar app- 'Teachable Machine', which only works on image and audio datasets [9]. Our app works on images and moreover on comma-separated files (.csv) as well, containing numeric and character values. These datasets are more common, easily available, and in more use than audio datasets (according to the current scenario) [1]. ML Modeller will be based on a no-code platform, which means even a person with no coding skills can avail of benefits. Users do not require any prior machine learning knowledge. Help in decision-making for businessmen, recruiters, analysts, students, teachers, and the list continues. Less time-consuming and more accurate predictions than manual analysis. The only benefit of writing ML code manually is that one can build data specific ml model which might perform better than the website (Higher accuracy) but only if the person knows ML enough to do so [5].

PRACTICALITY IN DIFFERENT PROFESSIONS

Teachers and People in the Education System

ML Modeller has the potential to make educators more efficient by completing tasks such as classroom management, scheduling, etc. In turn, educators are free to focus on tasks that cannot be achieved by AI, and that require a human touch. ML Modeller in the form of learning analytics can help teachers gain insight into data that cannot be gleaned by using the human brain and give conclusions that positively impact the teaching and learning process. With it, educators can make conclusions about things that may happen in the future. For instance, using a data set of middle school students' cumulative records, predictive analytics can tell us which ones are more likely to drop out because of academic failure or even their predicted score on a standardized exam, such as the ACT or SAT. It can grade student assignments and exams more accurately than a human can. It may require some input from a human being, but the results will have higher validity and reliability [5].

Doctors and in Healthcare

Doctors can use ML Modellers to reduce the cost of healthcare, for example uploading EMR datasets on our website and getting results for specific inputs. Similarly, doctors can upload a dataset for diabetes, heart disease, pneumonia, and different types of cancer and diagnose if a person is having any of these ailments or not. The ultimate goal is to improve and provide

healthcare facilities at a lower cost. Moreover, doctors can analyze oncology data, providing insights that allow oncologists, pharmaceutical companies, payers, and providers to practice precision medicine and health. Similarly, providing a dataset can predict illness and treatment to help physicians and payers intervene earlier, predict population health risk by identifying patterns and surfacing high-risk markers and model disease progression, and more [5].

Recruiters and Hiring Team

Recruiting a perfect candidate or fixing a payroll is often very tireless work. But if you have relevant past data, then an ML Modeller will help you in finding the best suit for a job and with how much you should pay him or her according to your company policies. It would become easier to identify top candidates from large candidate pools [5].

Real Estate

Since ML Modeller has the ability to analyze patterns in vast amounts of data, it can be used to make reasonable predictions of the future value of a property. For example, it can combine current market data from the marketplace and CRM as well as consider public information such as transportation network characteristics, crime rates, schools, and buying trends. The number of property attributes or market data points can exceed tens of thousands, which is definitely a kind of analysis no human analyst or market research is capable of conducting. Similarly, ML Modeller can help students of school and colleges with their projects, debate, home tasks, research, and much more [5].

So the list of people who can benefit from it is countless.

CHALLENGES FACED

Initially we also tried to add a module for deep learning in ML Modeller and was successful in doing so but soon we had to discontinue it as deep learning requires high processing speed and if you are making a budget-friendly project then it is not recommended to use deep learning in it as at the time of deploying on the cloud you have to purchase service with higher CPUs. And if you try to run it on a low CPU, as we did initially, then the time required to train the model will increase exponentially, which is again not considerable. That is why deep learning still is not much used in industries and is only limited to research work due to its cost inefficiency. The other challenge we faced was applying hyper-parameter tuning. It can be almost described as a brute force technique, as it checks all possible parameters and tells the best possible match in which models give the best performance with high accuracy. But hyper-parameter tuning is time-consuming and again requires more CPUs to process fast. Thus impractical to apply. To encounter this issue we chose to keep parameters ideal while training.

CONCLUSION

As per the present scenario, we have researched and observed there is a high demand for no-code platforms as per the performance evaluation done on ML Modeller, which is not currently available in the market. Still, stepping out as the first community to introduce this technology offering people high standard prediction experience with our automated self-learning models with the least error rate, maximum accuracy, and easy-to-function capabilities. Most people with no prior knowledge of programming and machine learning can take benefit of such applications train their dataset and

get predictions. Otherwise, the existing approach was very naive, time-consuming, and less accurate as the dataset has to be analyzed manually and then taken the decision as per the individual's perspective and insight. Even after taking a decision person would have no clue that what is going to be its probability of getting correct. But in the case of no-code platforms such as ML Modeller, we not only get predictions with their corresponding accuracy rate which makes it easier to rely on. On a performance basis, the website is giving consistent growth with new visitors/users every day and engaging existing visitors/users as well as the re-visit also. Most of the users are from USA or India.

REFERENCES

[1]   Teachablemachine.withgoogle.com, 2021. [Online]. Available: https://teachablemachine.withgoogle.com/.

[2]   K. Eremenko, 2021. [Online]. Available: https://www.udemy.com/course/machinelearning/learn/lecture/ 20203260?start=0#overview.

[3]   [Online]. Available: https://dash.cloudflare.com/.

[4]   [Online]. Available: https://www.w3schools.com/.

[5]   "ML MODELLER ! No-code platform", Mlmodeller.in, 2021. [Online]. Available: https://www.mlmodeller.in/.

[6]   "Ml Modeller", 2021. [Online]. Available: https://github.com/kartikeyahl/mlmodeller_blog.

[7]   "At +26.1% CAGR, Low-code Development Platform Market size is Expected to reach 65.15 Billion by 2027, says Brandessence Market Research", Brandessence Market Research And Consulting Private Limited, 2021.

[8]   M. Tommasi, "Democratizing software with no-code", Makerpad Ltd., 2021.

[9]   V. Afshar, "92% of IT leaders comfortable with business users using low-code tools", ZDNet, 2021.