

# PROJECT REPORT

## on

# AutoVizML

NAS1001 - NASSCOM Future Skills- Associative Data Analyst  
[E21+E22+E23 slot]

---

**Submitted By:**

19MIM10024 - Raghav Agarwal  
19MIM10046 - Indrashis Paul  
19MIM10088 - Kshitij Singh Chouhan  
19MIM10038 - Anshika Mishra  
19MIM10111 - Shashwat Jha  
19MIM10078 - Raunak Ansari

---

**Guided By:**

Dr. Nilamadhab Mishra

**Index table:**

S.No.	Topic	Page No.
1	Introduction	
2	Problem Statement	
3	Existing solution and Literature Review	
4	Proposed Solution	
5	System Architecture Diagram	
6	Methodology and workflow explanation	
7	Implementation	
8	Result and discussion	
9	Limitations and Scope of Improvement	
10.	References	

# Introduction

There are many different aspects to learning. Declarative learning, the growth of physical and cognitive skills through teaching or practice, the organizing of new knowledge into universal, useful representations, and the discovery of novel facts and hypotheses through observation and experimentation are all examples of learning processes. Researchers have been working to incorporate these capabilities into computers since the dawn of the computer era. It has been and still is a challenging and interesting long-term objective in artificial intelligence to solve this issue (AI). Machine learning is the study and computer modeling of learning processes in their varied manifestations.

Due to the lack of data scientists, automated machine learning as a concept will become a popular issue in 2022. More than ever, it's critical to enhance productivity by having existing data scientists automate as much of their work as possible and get new data scientists up and running as quickly as possible. AutoML lets data scientists build models very quickly and also lets new data scientists on the ramp very quickly. They can concentrate on learning how to prepare data for AutoML algorithms rather than having to learn how to do so for each algorithm.

## Problem Statement

With about 1,000 petabytes of data being generated each day across the globe, the need of extracting something useful out of it is the need of the hour. A trained domain expert can easily look into the data, find insights out of it and even model it to predict the predictions. But, an individual with no domain knowledge can not handle this data. Therefore, there is a need for a platform upon which anyone can choose the dataset, explore it to find some insights, and even model it to predict the predictions.

## **Existing work and Literature Review**

There are few available solutions out there and one of the best one is a custom web application which is written in R programming language and uses a shiny user interface known as Radiant.

Radiant is a highly optimized web application which provides a vast variety of machine learning models and analytical tools which can be used with just some clicks by which any person with no prior knowledge of these tools working can extract some key information and can find some valuable insights which normal base level tools can't provide and the most amazing thing is that it does not require a single line of code from the user. Radiant is having a high dimensionality which makes it for everyone, which means it surely will be helpful to new comers, but a person which is having good knowledge and solid background can also leverage this application in the best possible way as it will remove some very long and repetitive process and it has a collection of some special tools all at one place. Radiant can also handle large amount of data and process it as the user wants, it can easily channel the data through the required selection which user selects.

## **Proposed Solution**

Machine Learning is a complex process that requires a fair amount of technical knowledge even to build little dummy models. Our platform "AutoVizML" is a platform built using Shiny and flex-board packages present in R programming language, that helps the user to automate the data visualization and model training part. Our platform has the following section:

1. - Data Selection
  - a. We have two options to select data:
    - i. - First is to select a dataset from the dropdown menu consisting of a list of all available datasets.
    - ii. - Second is the upload button using which the user can update his/her own dataset file in CSV format.
2. - EDA
  - a. Once a dataset is selected, the user can switch to the EDA section for getting various visualizations/
3. - Model Training

## **System Architecture Diagram**

## **Methodology and Workflow explanation**

**Implementation**

**Result and Discussion**

**Limitations and Scope of Improvement**

## References