



A Novel Approach for Data Analysis using Explainable AI

Nirupama P

University of Bangalore, Alumni of Raja Reddy College, Bangalore, India. Email: nirupamachidanand2015@gmail.com

Crossref

DOI: http://doi.org/10.38177/ajast.2021.5210

Copyright: ©2021 Nirupama P et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 15 February 2021

Article Accepted: 10 May 2021

Article Published: 12 June 2021

ABSTRACT

Explainable artificial intelligence (XAI) has become popular in the last few years. The Artificial intelligence (AI) community in general and the Machine learning (ML) community in particular is coming to the realization that in many applications, for AI to be trusted, it must not only demonstrate good performance in its decision making, but it also must explain these decisions and convince us that it is making the decisions for the right reasons. This paper presents a novel way of representing statistical features from machine learning datasets to image based representation and exploring Conventional neural networks (CNN's) based networks for better data interpretation. This technique found to be extremely powerful as compared to ML and found to improve more than 10% accuracy on various statistical datasets available.

1. Introduction

Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explainability also helps an organization adopt a responsible approach to AI development. given in points, same as in the MS Word font size points. Times New Roman is the preferred font.

As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm came to a result. The whole calculation process is turned into what is commonly referred to as a "black box" that is impossible to interpret. These black box models are created directly from the data. And, not even the engineers or data scientists who create the algorithm can understand or explain what exactly is happening inside them or how the AI algorithm arrived at a specific result [1].

ML models are often thought of as black boxes that are impossible to interpret. Conventional Neural networks used in deep learning are some of the hardest for a human to understand. Bias, often based on race, gender, age or location, has been a long-standing risk in training AI models. Further, AI model performance can drift or degrade because production data differs from training data. This makes it crucial for a business to continuously monitor and manage models to promote AI explainability while measuring the business impact of using such algorithms. Explainable AI also helps promote end user trust, model auditability and productive use of AI.

This work summarizes below contributions.

- Presents a unified, hierarchical approach for preprocessing work by identifying/replacing NA values,
 Normality transformation outliers/leverage/influential data.
- o Presents detailed analysis of traditional Machine learning algorithm and 1D CNN algorithm.



- Presents a unique approach for representation statistical data and numerical data to image based representation using different types of graphs such as Pie chart, Bar chart, Box Plot, Scatter plot. So that 2D CNN can understand to assist classification algorithm.
- Comparison of unique representation of statistical data using 2D CNN with respect to transform Machine Learning algorithm found to be superior.

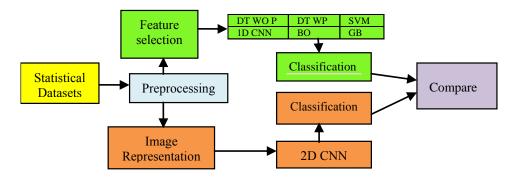


Fig.1. Machine Learning Model and Explainable AI

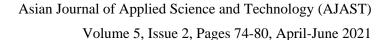
2. Methodology

2.1 Data preprocessing

Data Preprocessing is a technique that is used to convert the raw data into a clean data. Data Preprocessing is the process of making data suitable for use while training a machine learning model. The dataset initially provided for training might not be in a ready-to-use state, for e.g. it might not be formatted properly, or may contain missing or null values. Solving all these problems using various methods is called Data Pre-processing, using a properly processed dataset while training will not only make life easier for you but also increase the efficiency and accuracy of our model.

Steps in Data Preprocessing:

- (a) Splitting of the data set in Training and Validation sets: Train Test Split is one of the important steps in Machine Learning. It is very important because our model needs to be evaluated before it has been deployed. And that evaluation needs to be done on unseen data because when it is deployed, all incoming data is unseen. The main idea behind the train test split is to convert original data set into 2 parts i.e., Train and Test, where train consists of training data and training labels and test consists of testing data and testing labels.
- (b) Handling Missing Data: If your data set is full of NaNs and garbage values, then surely your model will perform garbage too. So taking care of such missing values is important. Let's take a dummy data set to see how we can tackle this problem of taking care of garbage values Here, we imputed the missing values by using Mean [9] in which, replacing the missing value for a particular attribute for the average value for that attribute.
- (c) Taking care of Categorical Features: You can guess that since machine learning models are based on mathematical equations you can intuitively understand that it would cause some problem if we keep the text here in the categorical variables in the equations because we would only want numbers in the equations[11]. So that's why we need to encode the categorical variables. That is to encode the text that we have here into numbers.





(d) Feature Scaling: We use feature scaling to convert different scales to a standard scale to make it easier for Machine Learning algorithms [2].

2.2 Decision Tree

A decision tree [3] is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

(a) **Pre-pruning:** Pre-pruning deals with noise during learning. An alternative method to prevent overfitting is to try and stop the tree-building process early, before it produces leaves with very small samples. This heuristic is known as early stopping but is also sometimes known as pre-pruning decision trees. At each stage of splitting the tree, we check the cross-validation error. If the error does not decrease significantly enough then we stop. Early stopping may underfit by stopping too early. The current split may be of little benefit, but having made it, subsequent splits more significantly reduce the error.

(b) Post-pruning: The aim of pruning is to discard parts of a classification model that describe random variation in the training sample rather than true features of the underlying domain. This makes the model more comprehensible to the user, and potentially more accurate on new data that has not been used for training the classifier. When pruning, an efficient mechanism is needed for distinguishing parts of a classifier that are due to chance effects from parts that describe relevant structure.

Statistical significance tests are theoretically well-founded methods for determining whether an observed effect is a genuine feature of a domain or just due to random fluctuations in the sampling process. Thus they can be used to make pruning decisions in classification models. Reduced-error pruning (4), a standard algorithm for post-pruning decision trees, does not take statistical significance into account, but it is known to be one of the fastest pruning algorithms, producing trees that are both accurate and small.

2.3 Support Vector Machine

Support-vector machines [5] are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis and supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text.

It is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

(a) Linear and Non-linear: It can solve linear and non-linear problems and work well for many practical problems. The algorithm creates a line or a hyper plane which separates the data into classes. Linear means something related to a line. All the linear equations are used to construct a line. A non-linear equation is such which does not form a straight line. It looks like a curve in a graph and has a variable slope value.



2.4 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) [6] is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. A *convolutional* neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain.

A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to trains limited for image processing and natural language processing. network is a system of hardware and/or software patterned after the operation of neurons in the human brain.

2.5 Bayesian Optimization

Bayesian Optimization [7] is an approach that uses Bayesian Theorem to direct the search in order to find the minimum or maximum of an objective function.

Optimization is at the heart of machine learning. When we fit a model, we rely on finding the minimum of the associated cost function. Fortunately, many optimization problems are relatively easy. For example, the cost function for a logistic regression model is convex with a single minimum and closed-form derivatives and so any sensible gradient-based method will find the minimum.

Optimization is at the heart of machine learning. When we fit a model, we rely on finding the minimum of the associated cost function. Fortunately, many optimization problems are relatively easy. For example, the cost function for a logistic regression model is convex with a single minimum and closed-form derivatives and so any sensible gradient-based method will find the minimum.

2.6 Gradient boosting

Explicit regression and classification boosting algorithms were subsequently developed by Friedman (2001, 2002), who coined the name "gradient boosting "and paid a special attention to the case where the individual components are decision trees.

Gradient boosting [8] is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient boosted trees, which usually outperforms random forest. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

3. Results

From the below table, we observe that, by using DS (Data Structures) like, IB (Indian diabetes Datasets), EA (Employee Attrition Datasets, BK (Banking datasets) for analysis of traditional Machine learning algorithm and exploring CNN's based networks.

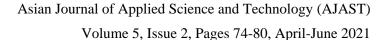


Table 1. Analysis of traditional Machine learning algorithm and 1D CNN and 2D CNN algorithm before preprocessing

DS	I	SVM						CNN		во	GB	
	DTPR	DTPO							1D CNN	2D CNN		
				Li		N. Li			WOI	WI		
			\boldsymbol{A}	P	R	\boldsymbol{A}	P	R				
ID	.70	.77	.77	.78	.80	.78	.81	.52	.84	.88	.80	.82
EA	.80	.87	.87	.88	.90	.88	.91	.62	.94	.98	.90	.92
BK	.60	.67	.67	.68	.70	.68	.71	.42	.74	.78	.70	.72

Table 2. Analysis of traditional Machine learning algorithm and 1D CNN and 2D CNN algorithm After preprocessing

DS	I	DΤ	SVM						CNN		во	GB
	DTPR	DTPO					1D CNN	2D CNN				
			Li			N. Li			woi	WI		
			A	P	R	A	P	R				
ID	.72	.78	.78	.79	.80	.79	.83	.55	.86	.90	.82	.84
EA	.84	.89	.89	.90	.92	.89	.92	.65	.96	.99	.92	.94
ВК	.63	.69	.69	.70	.72	.70	.73	.44	.76	.80	.72	.74





In Table 1, we observe that unique representation of statistical using CNN (Convolutional neural network) based network found to be more accuracy than the traditional technique machine learning algorithm like DT (Decision tree) - DTPR (Decision Tree pre-pruning), DTPO (Decision Tree post-pruning) + SVM (Support Vector Machine) + BO (Bayesian Optimization) + GB (Gradient Boosting).

In Table 2, We observe that hierarchical approach for preprocessing work by identifying/replacing NA values, Normality transformation outliers/leverage/influential data found to be improve the accuracy level.

4. Conclusion

In this paper, a unique representation of statistical features from machine learning datasets to image based representation and exploring CNN's based networks for better data interpretation. The 2D CNN based technique is found to be more accuracy than traditional technique of machine learning.

Decision Tree, Support Vector Machine, Bayesian Optimization, Gradient Boosting approaches are used for analyzing the accuracy level but Comparing unique representation of statistical data using 2D CNN with respect to transform Machine algorithm found to be superior. So the 2D CNN can better understand to assist classification algorithm.

Declarations

Source of Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The author declares no competing financial, professional and personal interests.

Consent to participate

Not Applicable.

Consent for publication

Author declares that he/she consented for the publication of this research work.

Availability of data and material

Author is willing to share data and material according to the relevant needs.

References

[1] Pantelis Linardatos, Vasilis Papastefanopoulos, Sotiris Kotsiantis, Explainable AI: A Review of Machine Learning Interpretability Methods, University of Patras, 2021.

[2] K Uma, M Hanumanthappa, Data Collection Methods and Data Preprocessing Techniques for Healthcare Data Using Data Mining, International Journal of Scientific & Engineering, June-2017.



- [3] Xiangxiang Zeng, Isi Yuan, You Li & Quan Zou, Decision Tree Classification Model for Popularity Forecast of Chinese Colleges, 2014.
- [4] Quinlan, Selecting a classification method by cross-validation, 1987.
- [5] Durgesh K, Srivasta, Lekha, Bhmbhu, Data classification using SVM, Journal of Theoretical and Applied Information Technology.
- [6] Sakshi Indoliaa, Anil Kumar Goswamib, S. P. Mishra, Pooja Asopa, Deep Learning Approach.
- [7] Stewart Greenhill, Santu Rana, Sunil Gupta, Pratibha Vellanki, Svetha Venkatesh, Bayesian Optimization for Adaptive Experimental Design: A Review, 2020.
- [8] G. Biau, B. Cadre, L. Rouvière, Accelerated gradient boosting, 2019.