BITS Pilani
Pilani | Dubai | Goa | Hyderabad

A REPORT
ON
IMPLEMENTING MACHINE LEARNING ALGORITHMS AND UNDERSTANDING THEIR INFERENCES USING EXPLAINABLE AI

- NIMISH BANSAL (2021B5A71179P)

AT

DYSL - CT DRDO, CHENNAI

IN FULFILLMENT OF

PRACTICE SCHOOL – I

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

JULY 2023

ACKNOWLEDGEMENT

I would like to express my gratitude to the PS Division, BITS Pilani for providing me with this opportunity to complete my Practice School-I at DYSL - CT DRDO in Chennai. I am thankful to Mr. Glynn John for his constant guidance and support in ensuring the successful completion of the project. I appreciate my parents' constant encouragement to work diligently and complete the endeavor. I am thankful to Mr. Manish Pratap Singh (Director of DYSL - CT DRDO) for allowing me to work under his supervision and for his invaluable insights on various machine learning, deep learning algorithms and explainable AI techniques. The encouragement I received from the Director and all the instructors kept me motivated throughout the course, allowing me to finish my project on time and acquire valuable insights.

- Nimish Bansal (2021B5A71179P)

INDEX

PS DETAILS

Name : Nimish Bansal

ID Number : 2021B5A71179P

Station : DYSL - CT DRDO, Chennai

Duration : 30 May 2023 to 21 July 2023

PS Mentor : Mr. Glynn John

Project Mentor : Mr. Manish Pratap Singh (Director of DYSL - CT DRDO)

ABOUT DYSL - CT DRDO

The DRDO Young Scientist Laboratories (DYSLs) are five specialized research laboratories located in Bangalore, Mumbai, Chennai, Kolkata, and Hyderabad. It was inaugurated by the Prime Minister of India on January 2, 2020. Each laboratory focuses on a specific scientific discipline, including artificial intelligence, quantum technologies, cognitive technologies, asymmetric technologies, and smart materials.

The DYSL-CT focuses on cognitive technologies. Principal areas of research include cognitive radar, cognitive radio, and cognitive surveillance systems. The lab is responsible for the design and preliminary testing of such systems, which is then transferred to higher laboratories for additional testing and mass production.

Cognitive radar is an advanced radar technology that employs cognitive principles to enhance the performance and capabilities of conventional radar systems. It integrates cognitive abilities such as adaptability, learning, and decision-making into radar systems in order to improve their detection, surveillance, and determination of targets in complex and dynamic environments.

Cognitive radio is a cutting-edge wireless communication technology that intelligently utilizes radio frequency (RF) spectrum to improve spectrum utilization and efficacy. It aims to address the growing problem of spectrum scarcity, in which the demand for wireless communication services exceeds the available frequency bands.

CHAPTER - 1

IMPLEMENTATION OF MACHINE LEARNING ALGORITHMS AND EXPLAINING THEM USING EXPLAINABLE AI ALGORITHMS

CONTENTS

ABOUT THE PROJECT

Project Title - Implementation of various machine learning algorithms and explaining them with the help of explainable AI algorithms.

The project is based on understanding the importance of trust in machine learning algorithms and models in order to deploy them into use in the real world. In order to use a new model or new technology, the user must be able to trust the technology and its predictions.

Explainable AI is a new and emerging field of AI that helps in developing trust in the model and explains the predictions made by the model. In the project, various machine learning algorithms such as **Decision Tree, Random Forest, Support Vector Machine, K - Nearest Neighbors, Naive Bayes** were implemented on a dataset which was obtained from **kaggle** and then the predictions made by them were explained using various explainable AI models such as **LIME and SHAP.**

The test and training accuracy of various models were calculated and were optimized to achieve best possible results. In the end, the models were compared in their functioning and their accuracy in making predictions.

ABOUT MACHINE LEARNING

Machine learning is a subfield of artificial intelligence (AI) that concentrates on the development of algorithms and statistical models and makes it possible for computer systems to learn and make predictions using various techniques. The primary objective of machine learning is to enable computers to learn and analyze data, recognise patterns and make accurate predictions based on this understanding.

Some of the types of machine learning approaches are:

Supervised learning:
In Supervised learning, the machine learning model is trained using labeled data, in which the input data is paired with output labels. The model learns to predict unlabeled data by generalizing from labeled training examples.

Unsupervised learning:
Unsupervised learning involves training models on unlabeled data. The model's goal is to identify patterns, structures or relationships within the data without any predefined labels or outputs.

Semi-supervised learning:
This methodology combines supervised and unsupervised learning. The model is trained using a small quantity of labeled data and a large amount of unlabeled data. The labeled data guides the learning process, whereas the unlabeled data facilitates the discovery of new patterns.

Reinforcement learning:
In reinforcement learning, an agent learns to interact with an environment and make decisions based on trial and error. The agent receives a reward based on its actions, which allows it to learn optimal behavior on its own in an unknown environment.

MACHINE LEARNING ALGORITHMS (CLASSIFIERS)

The following machine learning algorithms were used to make the predictions of the project:

1. Decision Tree
2. Random Forest
3. Support Vector Machine (SVM)
4. K-Nearest Neighbors (KNN)
5. Naive Bayes

DECISION TREE

A decision tree is a flowchart-like tree structure in which each decision node represents a classification feature, branches represent decisions, and leaf nodes represent the classifier's output.

INFORMATION GAIN: It is the measure of decrease in randomness or entropy after the data set is split into different nodes based on the threshold.

The entropy for a subset which is having K number of classes and for the i^{th} node is given by:

$$H_i = - \mathbf{\Sigma}^{n}_{k \in K} \, p(i,k) \log_2 p(i,k)$$

GINI IMPURITY: Gini Impurity is a score that evaluates the accuracy of a classification division. The Gini Impurity score ranges from 0 to 1, where 0 indicates that all observations belong to the same class and 1 represents an arbitrary distribution of the elements.

The decision tree algorithm can be explained through the following steps:

Step 1: The complete dataset is available at the root node which is put into the decision tree for classification.

Step 2: The best attribute feature from the dataset is found using Attribute Selection Measure (ASM) technique.
Step 3: The root node is divided into subsets that contain the best possible values for the best features.
Step 4: A decision node is generated based on the classification, which contains the best attribute.
Step 5: Repeatedly new decision trees are made using the subsets of the dataset created in step 3 and the process is continued until the data is classified into separate classes and achieves homogeneity.

RANDOM FOREST

A random forest is a supervised machine learning algorithm that combines several decision tree models. It can be used to solve problems involving classification and regression. It is based on ensemble learning, which evaluates the predictions of multiple decision trees and selects the output based on majority.

BAGGING (BOOTSTRAP AGGREGATION):
Bagging is an ensemble algorithm that improves the accuracy of machine learning algorithms. It predicts by taking the average of the output from many decision trees. Increasing the number of trees increases the precision of the outcome and reduces the issue of overfitting.

BOOSTING:
It creates sequential models and combines predictions from all the models such that the final model has the highest accuracy.

FEATURES:

1. Accuracy is greater than Decision tree as it is a combination of many decision trees.
2. Allows to handle the missing data instead of modifying it.
3. Helps in solving the issue of overfitting of the model.

SUPPORT VECTOR MACHINE

❖ KEY TERMS:

SUPPORT VECTOR:
They are the vectors that are in close proximity to the hyperplane and influence its position.

HYPERPLANE:
It is a linear decision boundary that separates two different classes.

MARGIN:
It refers to the distance between the hyperplane and the support vectors.

Support Vector Machines (SVM) aim to find an optimal hyperplane that maximizes the margin. By doing so, the two classes are properly classified and there isn't any confusion while classifying a new data point.

❖ HANDLING NON - LINEAR DATA USING KERNEL TRICK:

In the case of non-linearly separable data, it may not be possible to identify an optimal hyperplane in the initial feature space. In these situations, SVM employs the kernel trick.

The kernel trick permits SVM to map the data implicitly to a higher-dimensional space where a linear decision boundary may be feasible. SVM operates as if it is in a higher-dimensional space without explicitly computing the transformation by computing the inner product between pairs of data points in the original feature space using a kernel function.

Thus, by implementing the kernel trick, SVM can find an optimal hyperplane in the higher-dimensional feature space, which corresponds to a non-linear decision boundary in the original feature space.

K - NEAREST NEIGHBORS

K-Nearest Neighbors (KNN) operates based on the idea that similar data points belong to the same class. When a new data point needs to be classified, KNN finds its k nearest neighbors in the training data and makes predictions based on their labels.

CHOOSING k:

In the KNN algorithm, the k value specifies the number of neighbors that will be examined to determine the classification of a new data point. In order to avoid ties during classification, it is recommended that k be an odd number, and cross-validation techniques can aid in selecting the optimal k for the dataset.

METRIC FOR MEASURING THE PROXIMITY:

In order to determine which data points are closest to a given data point, the distance between the new data point and the other data points need to be calculated.

The following metrics are used to calculate the proximity:

Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$

Manhattan distance:

$$d(x, y) = \sum_{i=1}^{m} \left| x_i - y_i \right|$$

While classifying, the class label is assigned to the new data point based on the majority between the classes among the k neighbors.

NAIVE BAYES

The Naive Bayes classifier is based on Bayes Theorem. Naive Bayes classifier has the following assumptions :

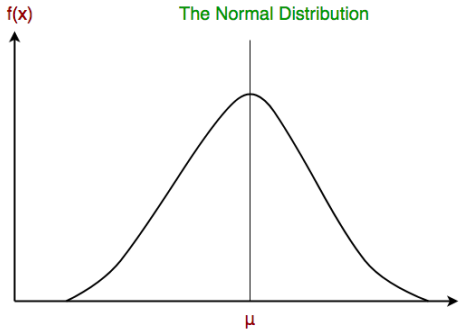
1. Every feature is independent of other features.
2. Every feature has an equal contribution to the prediction.

Bayes Theorem : Bayes Theorem finds the probability of an event occurring given that another event has already occurred.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

1. P(A) is termed as the prior probability.
2. P(A|B) is termed as the posterior probability.
3. P(B|A) is termed as the likelihood probability.

Gaussian Naive Bayes Classifier: In Gaussian Naive Bayes, it is presumed that continuous values associated with each feature follow a normal distribution. When plotted, a bell-shaped curve symmetric about the mean of the feature values is obtained.



WORKING OF NAIVE BAYES ALGORITHM:

1) The data set is converted into a frequency table.
2) This is in turn converted to a likelihood table by finding the probabilities.
3) Then the Bayes theorem is used to calculate the posterior probability of each class.
4) The class with the highest posterior probability is said to be the outcome of the prediction.

ABOUT EXPLAINABLE AI

Explainable AI is a collection of tools and frameworks that enable users to comprehend and interpret the predictions made by various machine learning models, thereby enhancing their trust in them.

The aim of Explainable AI (XAI) is to create machine learning techniques that develop more explainable models that maintain a high level of learning performance. It must also enable users to understand, trust, and manage new emerging machine learning and artificial intelligence models.

To achieve this objective, machine learning techniques are modified such that the predictions will be explained, the strengths and limitations of the model will be identified, and their future behavior will be predicted.

Benefits of XAI :

1. Helps in developing Trust in AI models
2. Helps to reduce biasing in AI models
3. Provides insights against Adversarial Attacks

EXPLAINABLE AI ALGORITHMS

The following explainable AI algorithms were used to explain the predictions made by the machine learning algorithms:

1. LIME
2. SHAP

LIME

LIME stands for **Local Interpretable Model-agnostic Explanations.** It is a visualization technique that helps in explaining individual predictions.

Local explanations:
It means that LIME explains the predictions of a single data point such that it is locally faithful in the environment it is being explained.

Interpretable:
It refers to the ability to understand and explain the predictions of a complex model in a human-understandable manner.

Model-agnostic:
It means that LIME can give explanations for any supervised learning model by treating it as a 'black box' separately. This means that LIME can be implemented on any model.

WORKING OF LIME:

LIME operates under the assumption that every complex model on a local scale is linear. LIME attempts to implement a simplified model centered on a single data point that mimics the local behavior of the global model. This basic model can then be used to explain the complex model's predictions locally.

EXPLANATION OF THE LIME ALGORITHM:

1. The data point to be explained is perturbed n times to create replicated data with slight value modifications. This perturbed data is a fake data created around the data point by LIME to build the local linear model.
2. The outcome is predicted for the perturbed data.
3. The distance from each perturbed data point to the original observation is calculated.
4. The distance is then converted to a similarity score.
5. From the perturbed data, m features are selected that best describe the predictions.
6. A simple model is fitted to the perturbed data for the selected features.
7. The feature weights of the simple model are the explanations of the observation.

SHAP

SHAP stands for **Shapley Additive Explanations**. This method aims to explain the prediction of an instance/observation by computing the contribution of each feature to the prediction.

Shapley values can help you in:

1. Global model interpretability
2. Local interpretability

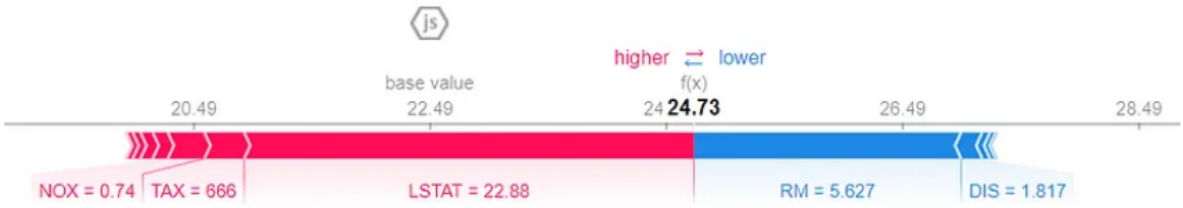
Different SHAP model explainers which are used to explain the predictions are:

TreeExplainer:
This method implements TreeSHAP algorithm and is useful for tree based algorithms such as Decision Tree, Random Forest etc.
DeepExplainer:
This method implements DeepLIFT algorithm and is used for deep learning models.
LinearExplainer:
As the name itself says, this method is ideal for linear models.

KernelExplainer:
This method is a model-agnostic method. Means it can be used to explain any model — linear models, tree models or deep learning models.

FORCE PLOT:

Force plots are used to explain the prediction of individual data points. One example of a Force plot is :



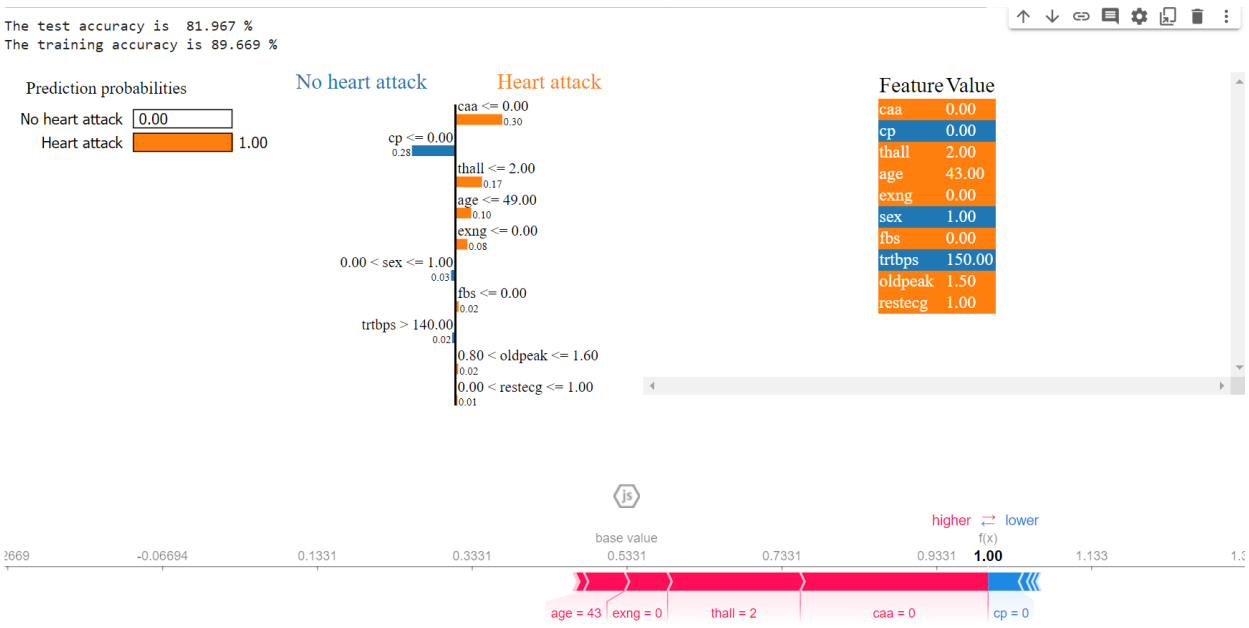
Base value is defined as the mean prediction over the entire testing dataset. It is the value that would be predicted in the absence of any features for the current output.

RESULTS

Abbreviations used in the dataset:

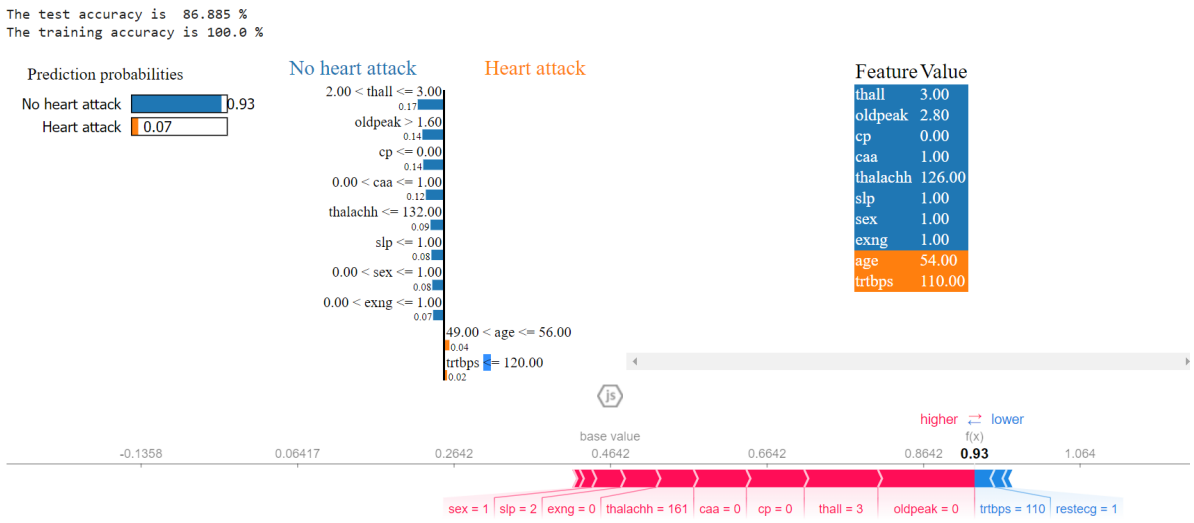
cp - chest pain type
trtbps - resting blood pressure
chol - serum cholesterol
fbs - fasting blood sugar
restecg - resting electrocardiographic results
thalachh - maximum heart rate achieved
exng - exercise induced angina
oldpeak - ST depression induced by exercise relative to rest
slp - slope of peak exercise ST segment
caa - number of major vessels colored by Fluoroscopy
thall - Thallium Stress Test

DECISION TREE:



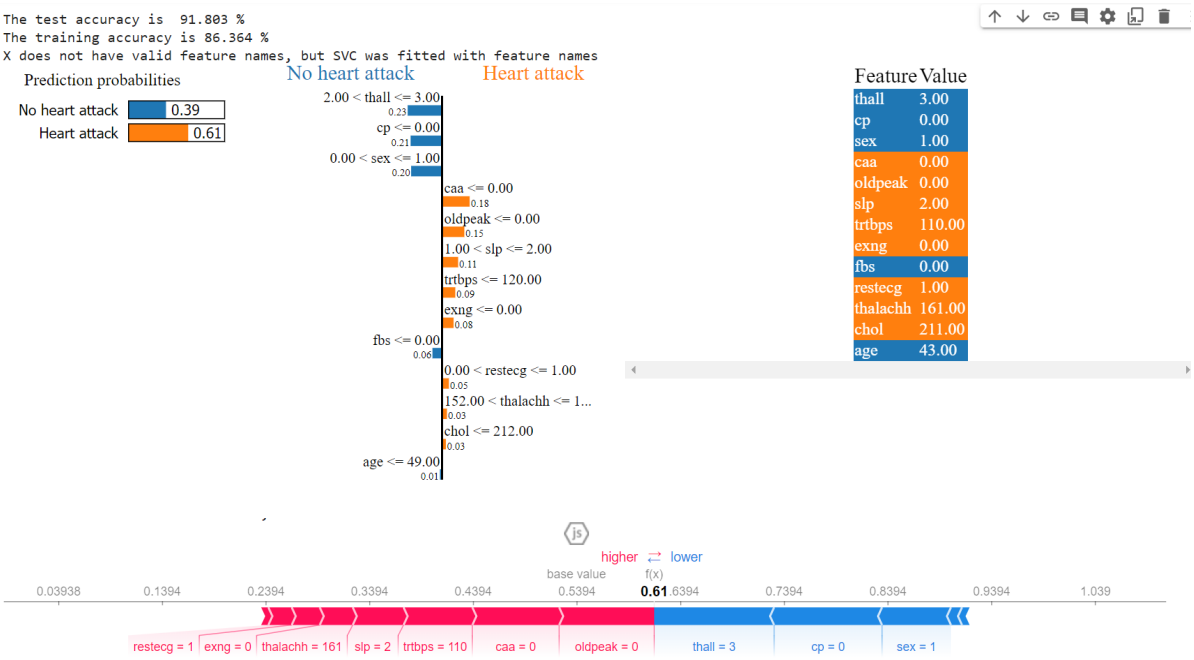
The test and training accuracy were approximately 82% and 90% respectively for the Decision Tree model. The model predicted the output “Heart attack” for the particular test case. According to the LIME algorithm, the features which influenced this prediction the most are caa, thall, age in decreasing order. According to SHAP algorithm, this prediction was due to caa, thall, exng in the decreasing order.

RANDOM FOREST:



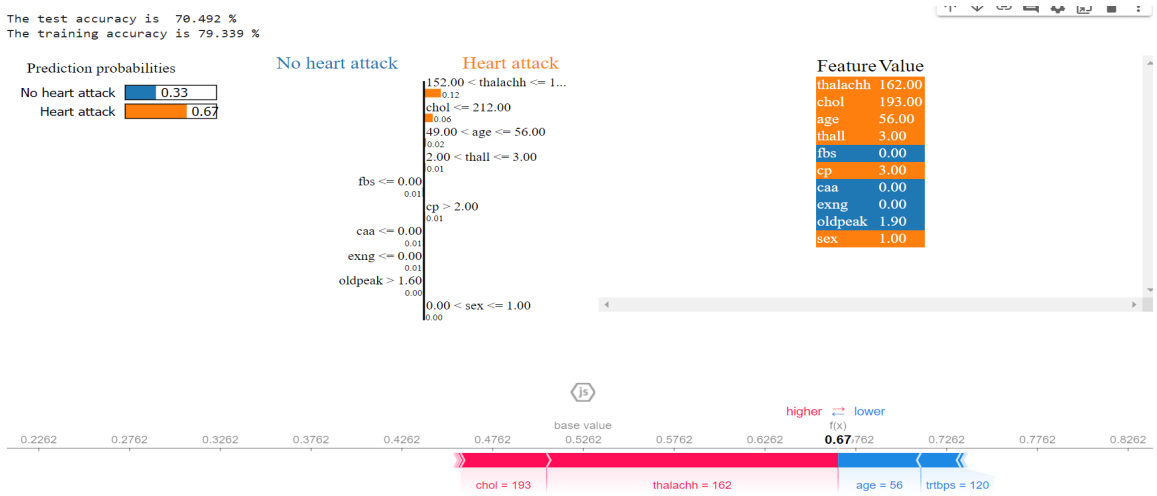
The Random Forest model makes the predictions with an accuracy of approximately 87%**.** The prediction made by the model is “No heart attack” which is based on the features thall, oldpeak, cp, etc.

SUPPORT VECTOR MACHINE:



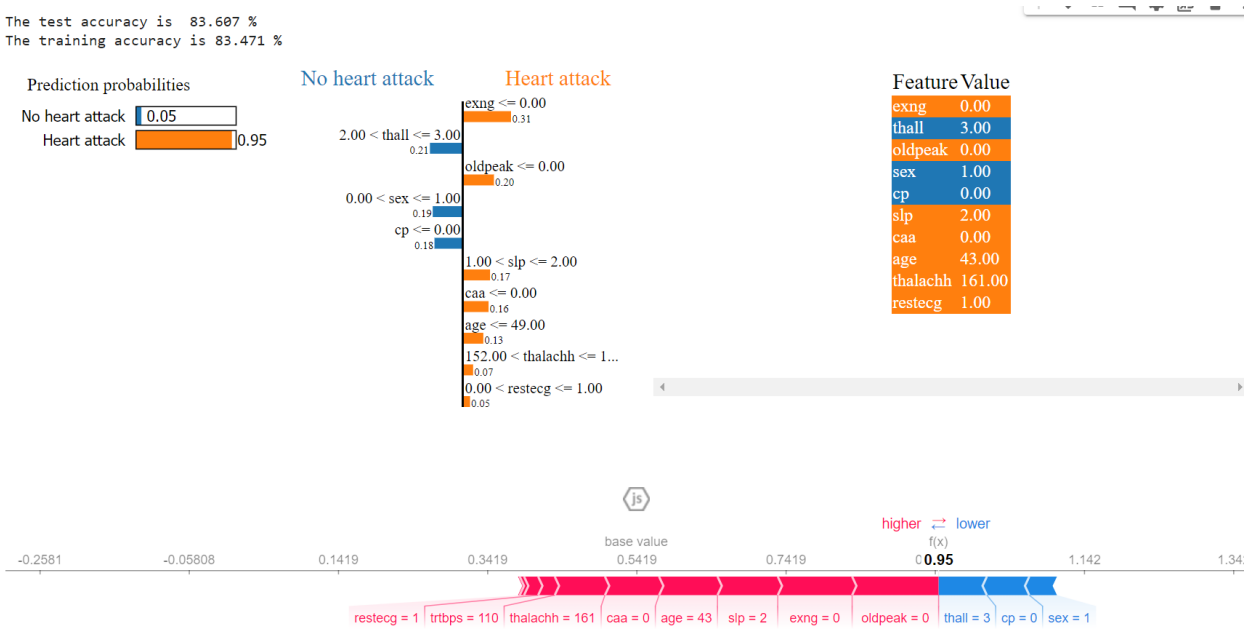
The test and training accuracy came out to be approximately 92% and 86% respectively for Support Vector Machine. The prediction came out to be “Heart attack” in this case which was supported by the features such as caa, old peak and slp.

K - NEAREST NEIGHBORS:



The predictions made by KNN were accurate with a test accuracy of approximately 70%. The prediction made by the model in the case above is “Heart attack” which is based on the features such as thalachh, chol, age etc.

NAIVE BAYES:



Naive Bayes performed quite well on the test dataset with a test accuracy of 84%. “Heart attack” was predicted by the model with a probability of 0.95. LIME was further used to explain the prediction which highlighted the key features affecting the output such as exng, oldpeak, slp etc.

CHAPTER - 2

OPTICAL CHARACTER RECOGNITION USING NEURAL NETWORK ALGORITHMS

CONTENTS

ABOUT THE PROJECT

Project Title - Optical character recognition using neural network algorithms

The project aims to recognize any numeric from 0 to 9 when the input is given in the form of an image. The model was implemented using two neural network algorithms namely, feedforward neural network and convolutional neural network. The model was trained and tested on the MNIST database with 60,000 and 10,000 images respectively. The model has also been given a feature where the user can input an image of a numeric and the model would be able to predict it. Furthermore, the test accuracy of the model has been calculated and a graph between training accuracy and validation accuracy has been plotted.

In the feedforward neural network, three layers were implemented; the input layer, the hidden layer, the output layer. Images of 28 x 28 were given in the input layer which had 64 filters. The hidden layer had 32 filters while the output layer had 10 filters. In the convolutional neural network, two convolutional layers with 32 and 64 filters respectively were implemented followed by which two max pooling layers were implemented using a 2 x 2 filter. Finally, the obtained image was flattened and was fed into a fully connected layer consisting of 80 neurons. The output softmax layer had 10 neurons which represent 10 digits from 0-9.

The trained model can be used to predict any handwritten digit or optical character with the help of softmax classifier. The digit is then classified into an appropriate class based on the prediction given.

NEURAL NETWORK

Neural networks are deep learning algorithms that are used to recognize different patterns from the given data and help in making future predictions. Neural networks have a similar architecture to that of the neurons in the human brain. A neural network consists of several layers which help in the recognition of various patterns and also train the parameters. Once the neural network learns some pattern, the hyperparameter tuning can be done to achieve the best results. Neural networks are widely used as they can handle image data, linear data as well as nonlinear data. There are different types of neural networks such as the feedforward neural network, convolutional neural network and recurrent neural networks. They are classified into various categories based on the input data used to make the predictions. It is also used to classify the output predictions into different classes such as a multiclass classification.

A neural network can have many layers depending on the complexity of the features to be extracted. Each layer computes some pattern or function and accordingly updates the parameters in order to learn the behavior of data. The number of neurons present in each layer can also vary depending on the amount of data and variance of the algorithm. These parameters can be adjusted later on in the validation set to achieve the best possible results. The model is fine-tuned on the validation data to obtain the best possible hyperparameter settings.

Neural networks have various applications such as optical character recognition, facial recognition, machine translation, instance segmentation, natural language processing, etc. These are also used along with machine learning and artificial intelligence techniques to develop some complex algorithms and functionalities.

FEED-FORWARD NEURAL NETWORK

A feedforward neural network is a type of deep learning algorithm in which all the nodes are connected to each other and the information is processed in only one direction. It is also known as an artificial neural network. It consists of many hidden layers with different numbers of neurons in each layer which help in recognizing different features or patterns from the data.

The feedforward neural network in its simplest form is also known as perceptron which consists of a single neuron. The input vector X is passed into the neuron and multiplied by some weight vector W. Furthermore, the sum of the weighted features is computed and some bias b is added to the sum. Then non-linearity is introduced into the model to ensure that the correct patterns are learned from the given data to make the right predictions. Generally, sigmoid nonlinearity is used in the output layer to get the probability of the prediction as its value lies between 0 and 1. The output of the perceptron is usually predicted as 0 or 1 depending on the threshold value which can be used for binary classification.

A deep neural network is usually a combination of many perceptrons arranged in various layers. The input data is fed into the input layer and computed in different layers by adding some weights and biases to the features. Finally, it is used to calculate a cost function. The main aim of the model is to find the correct set of parameters that helps in minimizing the cost function. Backpropagation is a technique that helps in calculating the gradients with the help of the chain rule. These gradients are used in the gradient descent which is an optimization algorithm, to update the parameters in the right direction and thus minimize the cost function.

Feedforward neural networks are used in various applications such as binary classification, spam filtering, optical character recognition, etc. Different hyperparameters can be fine-tuned to achieve the best possible results.

CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are also a type of deep learning architecture that can be used to handle image data and make predictions from it. It is used for object detection, classification, instance segmentation, face recognition, etc. Convolutional neural networks are used because it is difficult to handle a large amount of data in the feedforward neural network. These types of networks make use of kernels or filters to detect various features in the images such as vertical and horizontal edges, different objects, etc. The filters, which are usually of odd dimensions, slide on the images to compute various features, highlight some pixels or details, and also help to look into the depth of an image by increasing the number of channels.

A convolutional neural network consists of different layers such as convolutional layers, pooling layers and fully connected layers. Each of these layers makes use of some filters. The pooling layers are used to decrease the size of the image while keeping the same number of channels. It also provides some state invariance and helps in the stability of the algorithm. Different hyperparameters such as padding and stride can be optimized to achieve the best results.

Finally, after all the convolutional operations the image is flattened into a single column vector and is fed into a feedforward neural network. Softmax is used in the output layer which helps in multi-class classification. It also helps in categorizing the data into different classes based on the predictions given by conv-net. Sometimes instead of passing the whole data once at a time, it is passed in mini batches to reduce the computation time. It is also known as mini batch gradient descent. Various techniques such as batch normalization and regularization are used to reduce overfitting.

CNNs have various applications and are used in various fields such as neural style transfer, facial recognition, etc. They can also be used with recurrent neural networks to increase the efficiency of the model.

OPTICAL CHARACTER RECOGNITION

Optical character recognition is the process of converting a text image into a machine-understandable text format. It helps in the detection of various numeric images and then classifies them into different classes depending on the prediction. It can also be used to recognize handwritten digits and therefore, it is also known as a text extraction technique. OCR helps in simplifying the data entry process by creating effortless text searches, editing and storage.

Optical character recognition can be done with the help of both feedforward neural network as well as convolutional neural network. The handwritten images of the characters can be flattened in a single column vector which is then fed into a feedforward neural network for further processing. Every layer of the neural network recognizes some features such as edges, texture, etc. Different activation functions can be used in different layers to get good results. Finally, the cost function is computed and it is minimized by updating the weights and biases. Generally, batch norm and regularization techniques are used to reduce overfitting. Softmax classifier is used in the output layer to provide a multiclass classification.

On the other hand, the image data of the characters can be directly fed into a convolutional neural network for further processing. ConvNets such as Lenet-5 can be used for optical character recognition. It consists of various convolutional and pooling layers followed by some fully connected layers. Convolutional neural networks are generally preferred over feedforward neural networks as it reduces the computational cost significantly. Moreover, CNN helps in getting an in-depth analysis of the image.

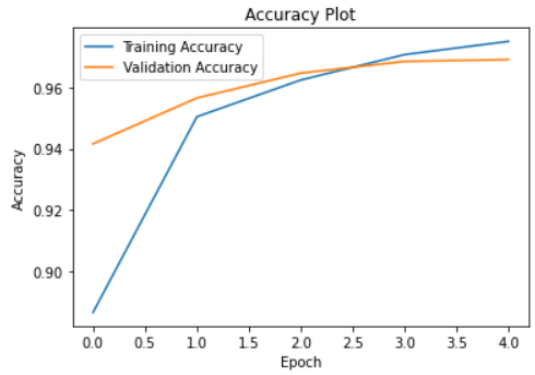
Once the model is trained and fine-tuned on the given data, the predictions are made by the model. Different handwritten digits are classified into various classes using the softmax classifier. OCR is of great interest as it is used in numerous applications such as word processing, legal documentation, etc.

RESULTS

FEED-FORWARD NEURAL NETWORK :

```
Epoch 1/5
469/469 [==============================] - 1s 2ms/step - loss: 0.4076 - accuracy: 0.8868 - val_loss: 0.1934 - val_accuracy: 0.9
416
Epoch 2/5
469/469 [==============================] - 1s 1ms/step - loss: 0.1689 - accuracy: 0.9504 - val_loss: 0.1450 - val_accuracy: 0.9
565
Epoch 3/5
469/469 [==============================] - 1s 2ms/step - loss: 0.1256 - accuracy: 0.9624 - val_loss: 0.1166 - val_accuracy: 0.9
646
Epoch 4/5
469/469 [==============================] - 1s 1ms/step - loss: 0.1008 - accuracy: 0.9706 - val_loss: 0.1063 - val_accuracy: 0.9
684
Epoch 5/5
469/469 [==============================] - 1s 1ms/step - loss: 0.0838 - accuracy: 0.9749 - val_loss: 0.1054 - val_accuracy: 0.9
690
```

```
Test Accuracy: 0.9690
```



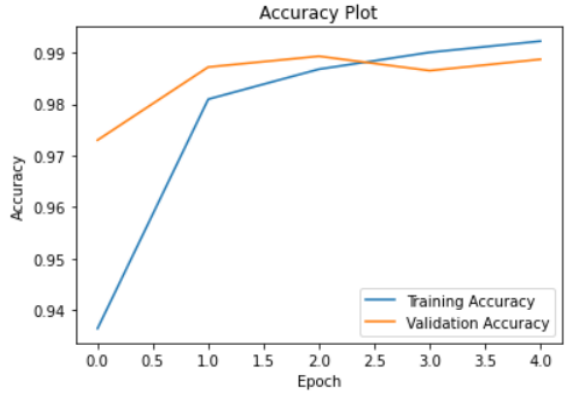

```
1/1 [==============================] - 0s 17ms/step
Predicted number: 3
Image from which the number was predicted:
```

The feed-forward neural network ran five epochs on the MNIST database and calculated the test accuracy to be approximately 96.9 %. A graph has been plotted between training accuracy and validation accuracy, and it is observed that validation accuracy is greater up to a certain point further which training accuracy is higher. An image of number three has been inputted into the model and it has rightly predicted the numeric to be 3.

CONVOLUTIONAL NEURAL NETWORK :

```
Epoch 1/5
469/469 [==============================] - 21s 40ms/step - loss: 0.2221 - accuracy: 0.9364 - val_loss: 0.0824 - val_accuracy:
0.9730
Epoch 2/5
469/469 [==============================] - 19s 41ms/step - loss: 0.0610 - accuracy: 0.9809 - val_loss: 0.0379 - val_accuracy:
0.9872
Epoch 3/5
469/469 [==============================] - 20s 43ms/step - loss: 0.0423 - accuracy: 0.9868 - val_loss: 0.0331 - val_accuracy:
0.9893
Epoch 4/5
469/469 [==============================] - 19s 41ms/step - loss: 0.0320 - accuracy: 0.9901 - val_loss: 0.0386 - val_accuracy:
0.9865
Epoch 5/5
469/469 [==============================] - 20s 42ms/step - loss: 0.0257 - accuracy: 0.9922 - val_loss: 0.0345 - val_accuracy:
0.9887
```

```
Test Accuracy: 0.9887
```




```
1/1 [==============================] - 0s 72ms/step
Predicted number: 5
Image from which the number was predicted:
```

The convolutional neural network has a test accuracy of 98.9 % after running five epochs. Validation accuracy is greater than training accuracy up to 2.5 epochs after which training accuracy is higher. The CNN model also predicted the number to be 5 from the input image.

CONCLUSION

It can be concluded from the project that explainable AI plays a crucial role in developing trust in various machine learning models so that they can be deployed for real world applications. The users can now get a clear idea as to which factors influence their model and can make any desired changes.

Throughout the project, various machine learning algorithms and their predictions were observed and among all the models, Support Vector Machine (SVM) had the best performance in predicting the correct output with a test accuracy of approximately **91 percent**. The predictions made by the Naive Bayes algorithm were also quite successful with a test accuracy of **83 percent.** Random forest outperformed the Decision tree model which is clearly evident as it is a combination of many decision trees and reduces the issue of overfitting of the model.

The predictions of various models were explained more accurately by LIME as it gives a clear understanding of various features used by the model in making the predictions and their contribution in making the final prediction.

The feedforward neural network was trained on the MNIST database by running five epochs much quickly compared to the convolutional neural network. CNN had a training accuracy of **99.2 percent** whereas FNN had an accuracy of **97.5 percent**. However, the test accuracy of the convolutional neural network was greater than that of the feedforward neural network. CNN also helps in reducing the computational cost by reducing the number of parameters to be trained.

BIBLIOGRAPHY

1. https://www.ibm.com/topics/machine-learning
2. https://www.lucidchart.com/pages/decision-tree
3. https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/
4. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
5. https://www.ibm.com/topichttps://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-networks/knn
6. https://www.geeksforgeeks.org/naive-bayes-classifiers/
7. https://www.darpa.mil/program/explainable-artificial-intelligence
8. https://www.geeksforgeeks.org/introduction-to-explainable-aixai-using-lime/
9. https://www.kaggle.com/code/bextuychiev/model-explainability-with-shap-only-guide-u-need
10. https://www.investopedia.com/terms/n/neuralnetwork.asp
11. https://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-network
12. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939