Machine Learning Engineer Nanodegree

Capstone Project Proposal

Sai Hitesh Chandra Mada

January 6th 2019

Proposal:

Tic-Tac-Toe Endgame

Domain Background:

History:

Tic-tac-toe is a simple, two-player game which, if played optimally by both players,

will always result in a tie. The game is also called noughts and crosses or Xs and Os.

Tic-tac-toe is a game that is traditionally played by being drawn on paper, and it can be played on a computer or on a variety of media. Other games, such as connect4, are based on this classic.

An early variation of the game was played in the Roman Empire, around the 1st century B.C. It was called *terni lapilli,* which means "three pebbles at a time." The game's grid markings have been found chalked all over Roman ruins.

A relatively simple game usually played on a grid of 3 x 3 squares, Tic-tac-toe can be made significantly more complex by increasing the size of the board to 4 x 4, 5 x 5, or even up to a 20 x 20 grid.

The goal of tic-tac-toe is to be the first player to get three in a row on a 3 x 3 grid, or four in a row in a 4 x 4 grid.

To start, one player draws a board, creating a grid of squares, usually 3 x 3 or 4 x 4.

The player who is playing "X" always goes first. Players alternate placing Xs and Os on the board until either one player has three in a row, horizontally, vertically, or diagonally; or all nine squares are filled.

The dataset was taken from UCI https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame

Relevant Papers:

Matheus, C.J., & Rendell, L.A. (1989). Constructive induction on decision trees. In Proceedings of the Eleventh International Joint Conference on Artificial Intelligence. (pp. 645--650). Detroit, MI: Morgan Kaufmann.
[Web Link]

Matheus, C.J. (1990). Adding domain knowledge to SBL through feature construction. In Proceedings of the Eighth National Conference on Artificial Intelligence (pp. 803--808). Boston, MA: AAAI Press.
[Web Link]

Problem Statement:

The main aim of my project is to predict the Binary classification task on possible

configurations of tic-tac-toe game. For this I selected the dataset from UCI

(https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame)

So, my goal is to predict the outcome of the game by using the given attributes and the class labels.

Here class labels are positive and negative. Here I am using classification modelsto find the accuracy of each model and select the model which will have high accuracy. Here the input parameter is given in the form of training data.

Datasets and input parameters:

Here the dataset will have categorical values. There will be total of 958 instances, and

the total number of attributes is 9. The dataset here is a multivariate.

Attribute information:

1. top-left-square: {x, o, b}
2. top-middle-square: {x, o, b}
3. top-right-square: {x, o, b}
4. middle-left-square: {x, o, b}
5. middle-middle-square: {x, o, b}
6. middle-right-square: {x, o, b}
7. bottom-left-square: {x, o, b}
8. bottom-middle-square: {x, o, b}

9. bottom-right-square: {x, o, b}
10. Class: {positive, negative}

Solution Statement:

Here I am trying to predict the outcome from the selected data. For doing so we want to use different classification models. Then we will find the accuracy score for each classification model. I explore the dataset with read _ cv, numpy, pandas, seaborn libraries in this project. By using visualization helps me to better understand the solution.

Benchmark model:

This step will be important because compare your final model with some of them and see if it got better, same or worse. We will choose Naive bayes as a benchmark model. We will try to find out a model whose accuracy is better than the benchmark model in this case Naive Bayes. Here Accuracy score will be compared between the models and select the best one.

Evaluation Metrics:

Here the classes are closely balanced hence I will use accuracy score as evaluation metric. I will be predicting the accuracy score for the selected model. The model with high accuracy score will be the best model out of the chosen models.

Project Design:

The project is composed of the following steps:

Pre-processing:

First task is to read the dataset and perform visualization on it to get some insights about the data. After reading the data clean the data that is removing unwanted data or replacing null values with some constant values or removing duplicates After data exploration, I will split the data into training set and testing set. Then applying the classification models and predicting the accuracy score to the selecting models.

Procedure for Preprocessing:

1. Read the data

2. Check whether there are any null values.

3. know the count of values in class attributes.

4.Create the dummy values for categorical values.

Training and testing the data:

After the completion of preprocessing, we will divide the dataset into testing data and training data. We will assign 70% of the whole data to training data and the remaining 30% to the testing data and we will carry-on with creation of our classification models.

I want to apply classification models of my own and use them on the data. I want to apply

Decision Tree, logistic regression, and random forest. Then I will find the accuracy score for the above-mentioned models. For this I will first train the algorithms with the training data, and then carry-on testing with the testing data that I split before Finally, I will declare the model which has the highest accuracy score out of all the chosen algorithms and declare it as the best one.