Shakayla Mosely

CSCI 4525

Machine Learning Report

Tic-Tac-Toe Endgame

Data Set:

The data set chosen is called “Tic-Tac-Toe Endgame” written by David W. Aha who is from the University of California, School of Information and Computer Science. The data set entails all the possible ways of winning a Tic-Tac-Toe game using the attributes of the game board: top left, middle, and right square; middle left, middle right square; bottom left, middle, right square. It also uses the variables of ‘x’ and ‘o’ to define the players of the game, and ‘b’ if they players did not use the square on the board. The data base is also defined to the possible wins of ‘x’ as if ‘x’ were to play first.

Baseline Classifier:

The baseline classifier used is named ZeroR, is a straightforward rule-based classifier that consistently predicts the majority class in the training set. To put it another way, it ignores all of the input features and predicts the class that occurs the most frequently in the training set of data. Accuracy (or other pertinent performance indicators) are provided in the baseline classifier's output. The baseline classifier accurately identified 626 out of 958 instances in the dataset, with an accuracy of 65.3445%. There were 332 cases with inaccurate classifications, or 34.6555% of the dataset.. At least one example of an observation that the baseline classifier failed to classify is given, and the impact of this failure on the problem is stated: The baseline classifier failed to classify any of the negative instances correctly. This means that if the classifier were to be used in practice, it would always predict the positive class, and would not be useful for predicting negative instances. The impact of this failure on the problem is that the baseline classifier is not suitable for this dataset, since it is important to be able to accurately predict both positive and negative instances.

Intelligent Classifier:

The Naive Bayes classifier, a probabilistic machine learning technique founded on Bayes' theorem, is the classifier I used for this data set. Since the Naive Bayes classifier is known to work well on datasets with a lot of features and can accept missing values, I employed it. It is used to forecast whether player X will win or lose a game of tic-tac-toe. The classifier makes the assumption that the characteristics (the locations of X and O on the board), given the class (win or loss), are conditionally independent. This is a simplification that enables the classifier to compute efficiently by estimating the probability of each attribute separately.

Based on the training data, the model is built by predicting the probability of each attribute (top-left-square, top-middle-square, etc.) given the class (win or loss). For instance, the number of times X appears in that position in winning games divided by the total number of winning games is used to estimate the likelihood of the top-left square being X given that the game resulted in a win for X. The likelihood that O will appear in the top-left square in a game where X lost is calculated similarly by dividing the total number of losing games by the frequency with which O appears in that position. Given the attribute values, the Naive Bayes classifier determines the likelihood of each class (win or lose, the positions of X and O on the board) utilizing the Bayes theorem, the prediction is made for the class with the highest probability. Using 10-fold cross-validation, the Naive Bayes classifier's overall accuracy on the test data was 69.62%. With an F-measure of 0.78 for positive outcomes and 0.49 for negative outcomes, the classifier performs better at predicting positive outcomes (wins for X) than negative outcomes (losses for X).

Compare & Contrast:

The majority class is predicted as the outcome for each input in the first model, the ZeroR. The dominant class in this instance is "positive." According to the findings of the 10-fold cross-validation, this model has a classification accuracy of 65.34%, correctly classifying 626 out of 958 instances. The classification of the remaining 332 occurrences is flawed. The model did not categorize any instances as "negative," which is the minority class in the sample, according to the confusion matrix.

The classifier model of the second model, the Naive Bayes classifier, displays the probabilities for each class given the values of the characteristics. According to the 10-fold cross-validation findings, this model has a classification accuracy of 69.62%, correctly classifying 667 out of 958 examples. The remaining 291 occurrences have incorrect classifications. The confusion matrix demonstrates that the model did not label any instances as "negative."

Work Cited

Lichman, M. (2013). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Retrieved from http://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame#