Hide menu

Case Study

Reading: Classification Analysis Case Study - Demo
2h

Reading: Classification Analysis Case Study
2h

Discussion Prompt: Classification Analysis Exploration Exercise
2h

Graded Assignment: Self Reflection
Started

Self Reflection

Review Learning Objectives

Assignment details

Due
Mar 10, 11:59 PM IST

Attempts
Unlimited

Your grade
You haven't submitted this yet. We keep your highest score.

--

Like Dislike Report an issue

Instructions

1. Reflecting on the case study, what was the most challenging aspect of applying classification analysis to solve the real-world problem? How did you overcome this challenge, and what did you learn from it?

1 point

The most challenging part is feature engineering, preprocess the data to compute the right set of features and then select the right classification model. A few feature selection models and feature extraction models (e.g., PCA) can be used, along with standard set of preprocessing techniques (e.g, z-score normalization, imputation for the missing values) applied prior to applying the model.

Your answer cannot be more than 10000 characters.

Resume

2. Describe a situation where you encountered technical problems while applying classification analysis to the case study data. How did you troubleshoot and resolve these issues to ensure accurate results?

1 point

Classifiers such as k-NN had low accuracy (~66%) on the held-out test dataset (using 80-20 validation) without z-score normalization, but increased significantly (~85%) after the preprocessing step was applied.

Your answer cannot be more than 10000 characters.

3. Reflect on the interpretation of the classification model results. How did you derive actionable insights from the model outcomes to make data-driven decisions for the real-world problem?

1 point

Gaussian Process	0.8688524590163934	0.875	0.875	0.875
Decision Tree	0.819672131147541	0.8181818181818182	0.84375	0.8307692307692308
Random Forest	0.8688524590163934	0.875	0.875	0.875
Neural Net	0.8688524590163934	0.875	0.875	0.875
AdaBoost	0.819672131147541	0.8888888888888888	0.75	0.8135593220338982
Naive Bayes	0.8688524590163934	0.9	0.84375	0.8709677419354844
QDA	0.8360655737704918	0.9230769230769231	0.75	0.8275862068965517

As can be seen, the classifiers k-NN and AdaBoost have low recall and moderately high precision, whereas the classifiers Linear SVM and Naive VBayes have high recall and moderately high precision. The liear SVM has the best F1-score. With decision tree we can learn interesting (actionable) rules such as the following one:

if cp <= 0.5 and trtbps <= 109 and caa <= 0.5, a patient has low risk of getting heart attack (2 out of 52 patients).

Your answer cannot be more than 10000 characters.

4. How did the application of classification analysis in the case study scenario enhance your critical thinking and problem-solving skills? Provide specific examples of how classification analysis aided you in making informed decisions [Practice this question as if you were in an interview!]

1 point

We can take proactive decisions, for example, from the patterns / rules learnt by the decision tree. If a patient matches any of the patterns leading to high risk heart attack, we can advise the patient and start early treatment to stop heart failure. Also, ensemble models like random forest / adaboost output most important features that are responsible for heart attack, so we can prioritize these features over the others and monitor them more closely.

Your answer cannot be more than 10000 characters.

5. What were the most valuable lessons you learned from completing the case study? How do you plan to apply these insights to further develop your classification analysis skills and grow as a data analyst?

1 point

Data is the key ingredient for data analysis. The better the quality of data is the better the model trained on the data is likely to capture interesting patterns and the better result of analysis is expected to be. Hence, we need to first run the cleaning / preprocessing steps on the data, before we can train models on it. Evaluating the performance of a machine learning model on unseen dataset is also very important, in order to make the model generalize and good at predicting unseen dataset, hence we need cross-validation steps (along with model selection with hyperparameter tuning). Also, since machine learning is an iterative process, we can reiterate the steps and rebuild the model (from scratch or incrementally), as and when new data is available, to keep the model up-to-date.

Your answer cannot be more than 10000 characters.

Coursera Honor Code Learn more

I, **SANDIPAN DEY**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit **Save draft**

Last saved on Feb 22, 4:14 PM IST

Like Dislike Report an issue