Final Project: Building a Rainfall Prediction Classifier

Legendary Performance! 🏆

Congratulations on successfully completing this assignment! Your grade has been recorded. Feel free to close this tab and return to the main course page.

Required passing grade: 70%

Status: **Passed**

Final Score: 20 / 20 (100%)

Provide Assignment Feedback

Question 1

Score: **2/2**

What is the True Positive rate of the RandomForestClassifier based on the confusion matrix from Exercise 13 in the assignment notebook?

Note: The answer should be expressed as a percentage, rounded to the nearest whole number.

51%

Great job! Your response correctly identifies the True Positive rate as 51%. Keep up the good work in accurately interpreting confusion matrices and calculating performance metrics.

Question 2

Score: **2/2**

Identify the most important feature for predicting whether it will rain based on the feature importance bar graph from the project lab.

- (●) Humidity3pm ✓
  Correct! Humidity3pm is the most important feature for predicting rain according to the feature importance bar graph
- ( ) Temp3pm
- ( ) Pressure3pm
- ( ) Sunshine

Correct! Humidity3pm is the most important feature for predicting rain according to the feature importance bar graph

Question 3

Score: **10/10**

Upload the duly completed Jupyter notebook for "Final Project : Building a Rainfall Prediction Classifier". Please ensure that you have completed all 14 coding exercises in

the ipynb file before sharing the same.

- FinalProject_AUSWeather.ipynb View Content

The Jupyter notebook is thoroughly completed with all 15 coding exercises, demonstrating exceptional understanding and application of the concepts. The code is well-organized, logically structured with clear headings, and follows best practices in coding style. Comprehensive comments and explanations are provided, enhancing readability and understanding. Visuals such as graphs and charts are effectively used to support the analysis. The document is free of grammatical errors, showcasing a high level of proficiency in both coding and communication. Excellent work!

Question 4

Score: **3/3**

Which of the following features would be inefficient in predicting whether it will rain tomorrow or not? (Select all that apply)

- [x] WindGustDir
  WindGustDir cannot be used efficiently since the direction of the strongest wind gust of the day cannot be known till the day is over. ✓
- [x] MaxTemp
  MaxTemp cannot be used efficiently since the maximum temperature cannot be known till the day is over. ✓
- [x] Evaporation
  Evaporation cannot be used efficiently since the total amount of water evaporated cannot be known till the day is over. ✓
- [ ] Humidity9am

WindGustDir cannot be used efficiently since the direction of the strongest wind gust of the day cannot be known till the day is over..
MaxTemp cannot be used efficiently since the maximum temperature

cannot be known till the day is over..
Evaporation cannot be used efficiently since the total amount of water evaporated cannot be known till the day is over..
You selected all correct options!

Question 5

Score: **3/3**

In bullets or as a numbered list, answer the following.

Comment on the accuracy of the LogisticRegression and Random Forest Classifier models.

Comment on the true positive rate of the two models.

Overall, which one of the two is a better predictor of whether it will rain tomorrow or not?

Random Forest:

- True Positive rate based on confusion matrix plot for Random Forest: 0.5139664804469274 or 51% when rounded
- Accuracy for Random Forest based on classification report is: 0.84 or 84%

Logistic Regression:

- True Positive rate based on confusion matrix plot for Logistic Regression: 0.5083798882681564 or 51% when rounded
- Accuracy for Random Forest based on classification report is: 0.83 or 83%

- The true positive rate for both models is essentially the same when the answer is rounded
- The accuracy scores are not different by lot either 83% for logistic regression and 84% for random forest classifier, so we cannot definitively state that one model is a much better predictor when compared to the other.

The student's response correctly identifies that the accuracy of both the Logistic Regression and Random Forest Classifier models is almost the same, with the Random Forest having a slight edge at 84% compared to 83% for Logistic Regression. The true positive rates are also noted to be essentially the same when rounded. Additionally, the student accurately concludes that the Random Forest Classifier model may be a slightly better predictor overall, given the marginally higher accuracy. Well done on providing a clear and correct analysis!

Report an issue

Retake Assignment