

Assignment 3

Due Feb 27 by 11:59pm **Points** 100 **Submitting** a file upload

Available Feb 15 at 12am - Mar 5 at 11:59pm

This assignment was locked Mar 5 at 11:59pm.

Assignment 3

Before working on this assignment, please read this tutorial –

<https://towardsdatascience.com/deep-learning-with-python-neural-networks-complete-tutorial-6b53c0b06af0>  [_ \(https://towardsdatascience.com/deep-learning-with-python-neural-networks-complete-tutorial-6b53c0b06af0\)](https://towardsdatascience.com/deep-learning-with-python-neural-networks-complete-tutorial-6b53c0b06af0)

Note that you don't have to use deep learning for this assignment. However, if you want to challenge yourself (e.g., you have AI, ML background), please give a try.

In this assignment, you will use the **Student Grade Prediction dataset**

[_ \(https://canvas.wpi.edu/courses/44874/files/5414823?wrap=1\)](https://canvas.wpi.edu/courses/44874/files/5414823?wrap=1) 

[_ \(https://canvas.wpi.edu/courses/44874/files/5414823/download?download_frd=1\)](https://canvas.wpi.edu/courses/44874/files/5414823/download?download_frd=1) to predict student performance. The data carries the following features:

1. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2. sex - student's sex (binary: 'F' - female or 'M' - male)
3. age - student's age (numeric: from 15 to 22)
4. address - student's home address type (binary: 'U' - urban or 'R' - rural)
5. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
9. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
11. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13. traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

14. studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15. failures - number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
16. schoolsup - extra educational support (binary: yes or no)
17. famsup - family educational support (binary: yes or no)
18. paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19. activities - extra-curricular activities (binary: yes or no)
20. nursery - attended nursery school (binary: yes or no)
21. higher - wants to take higher education (binary: yes or no)
22. internet - Internet access at home (binary: yes or no)
23. romantic - with a romantic relationship (binary: yes or no)
24. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25. freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26. goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29. health - current health status (numeric: from 1 - very bad to 5 - very good)
30. absences - number of school absences (numeric: from 0 to 93)

For this assignment, you will need to predict student performance “Performance

”(High, Low, Normal). You can use whatever algorithms for prediction (e.g., decision tree, naive bayes, SVM, kNN, deep learning...), and you can choose to use all features or a subset of features (e.g., by using feature selection algorithms). The requirements are:

1. You need to pick three different algorithms for prediction.
2. You need to pick your own evaluation metrics, e.g., precision, recall, or accuracy, to assess your algorithm performance.
3. You can use all features or a subset of features for prediction.

Please provide codes (with comments), and a brief report (no more than 2 pages) for this assignment. In the report, I would like to know – data processing method, why you choose these algorithms, evaluation metrics and results (in a table), conclusions, limitations, and future works...