ADS4293 LAB8

Chapter 12: Supervised Data Mining: k-Nearest Neighbors and Naive Bayes

TA: Malak Al-Hanafi

October30 , 2025

LAB SETUP

- Excel file **Ch12** dataset
- Submission:
  - Submit your updated **Excel file** on Brightspace (Owl)
- Today’s lab has 1 **question. Happy Halloween & Reading Week :)**

QUESTION 1

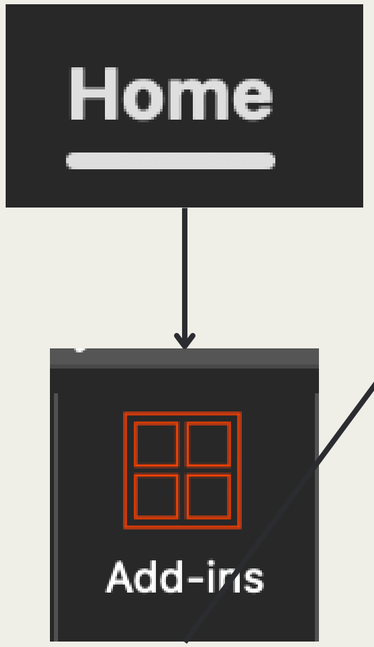
9. FILE ***CFA.*** The Chartered Financial Analyst (CFA) designation is the de facto professional certification for the financial industry. Employers encourage their prospective employees to complete the CFA exam. Daniella Campos, an HR manager at SolidRock Investment, is reviewing 10 job applications. Given the low pass rate for CFA Level 1, Daniella wants to know whether or not the 10 prospective employees will be able to pass the CFA Level 1 exam. Historically, the pass rate is higher for those with work experience and a good GPA in college. With this insight, she compiles the accompanying data file with information on 63 current employees who took the CFA Level I exam last year, including the employee's success on the exam (1 for pass, 0 for fail), the employee's college GPA, and years of work experience.

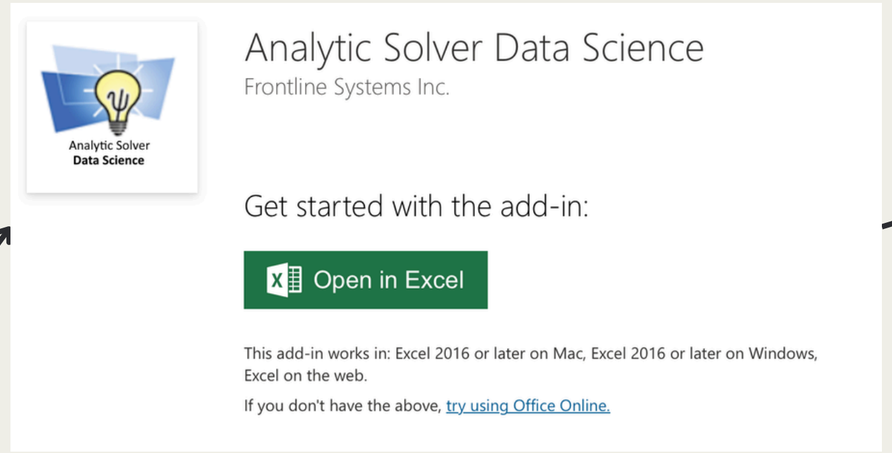
   a. Perform KNN analysis to estimate a classification model for the CFA Level 1 exam using the CFA_Data worksheet and score the 10 job applicants in the CFA_Score worksheet. What are the misclassification rates for $k = 3$, 4, and 5?

   b. What is the optimal value of *k?* Report the accuracy, specificity, sensitivity, and precision rates for the test data set (for Analytic Solver and Python) or validation data set (for R).

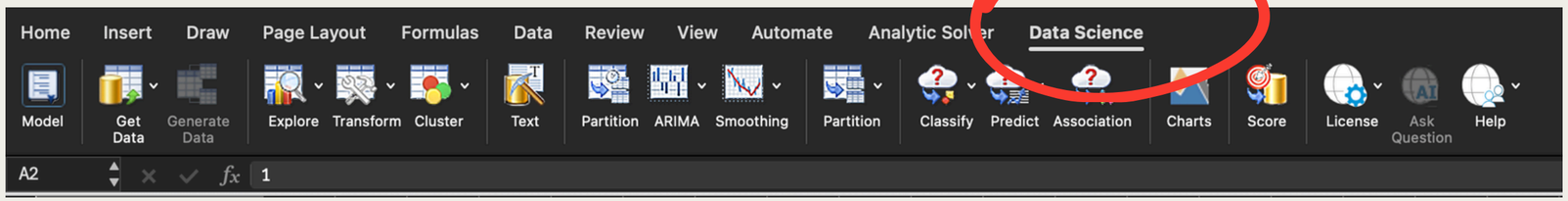
   c. What is the predicted CFA Level 1 outcome for the first job applicant?

REQUIREMENTS





Follow steps, make sure to log in with your school info

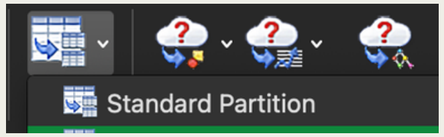


Q 1 A

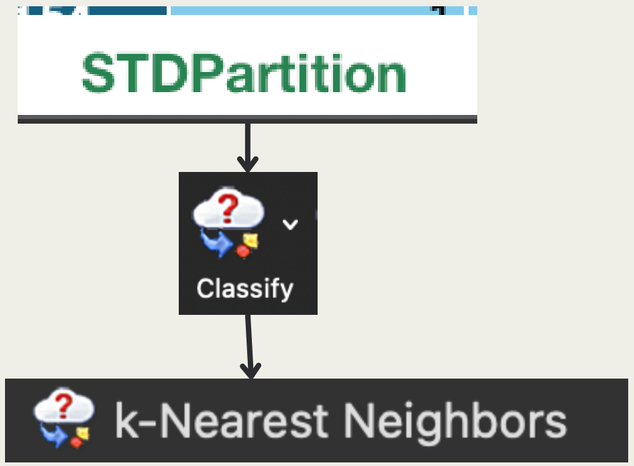
9. FILE ***CFA.*** The Chartered Financial Analyst (CFA) designation is the de facto professional certification for the financial industry. Employers encourage their prospective employees to complete the CFA exam. Daniella Campos, an HR manager at SolidRock Investment, is reviewing 10 job applications. Given the low pass rate for CFA Level 1, Daniella wants to know whether or not the 10 prospective employees will be able to pass the CFA Level 1 exam. Historically, the pass rate is higher for those with work experience and a good GPA in college. With this insight, she compiles the accompanying data file with information on 63 current employees who took the CFA Level I exam last year, including the employee's success on the exam (1 for pass, 0 for fail), the employee's college GPA, and years of work experience.

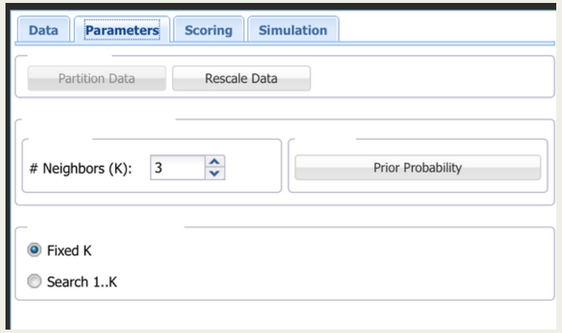
   a. Perform KNN analysis to estimate a classification model for the CFA Level 1 exam using the CFA_Data worksheet and score the 10 job applicants in the CFA_Score worksheet. What are the misclassification rates for k = 3, 4, and 5?

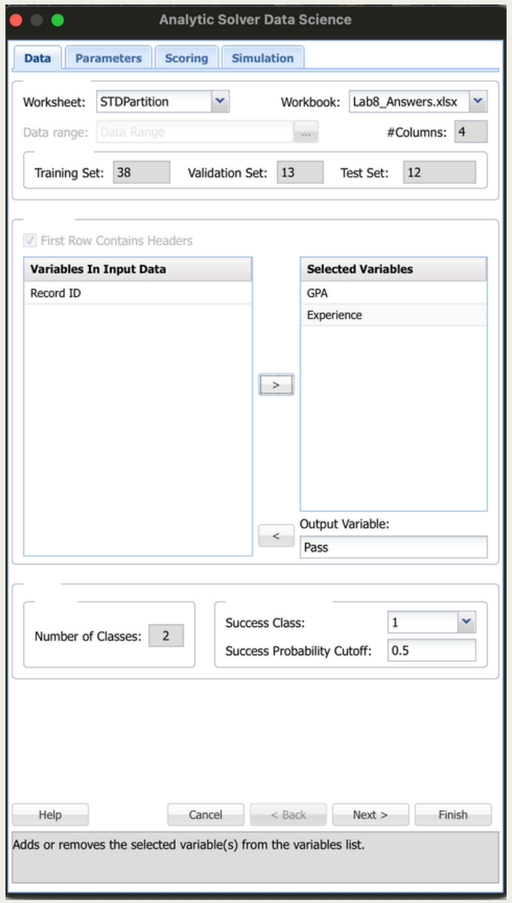
1. **Select the whole table in CFA_Data (include headers)**
2. **Partition the data.** **Data Science → Partition → Standard Partition**
3. **Fill in the following:**
   a. **Variables**: Pass, GPA, Experience, click the ">" button to move them to the right box
   b. **Partition Method:** Random
   c. **Specify Percentages:** Training %: 60, Validation %: 20, Test %: 20
   d. **Set Seed:** 12345

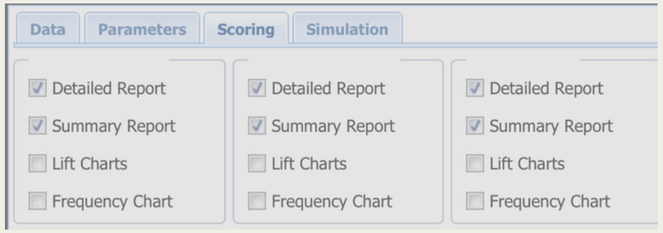


Q 1 A









- Find the misclassification rate for k = 4, and 5 from the Error Report, go back to the KNN model dialog (under Data Science → Classify → k-Nearest Neighbors).
- Change the parameter # Neighbors (K) to 4, click Finish, and rerun.
- The optimal value of k is the one with the **lowest misclassification rate.**

Q 1 B

b. What is the optimal value of *k*? Report the accuracy, specificity, sensitivity, and precision rates for the test data set (for Analytic Solver and Python) or validation data set (for R).

KNNC_TestScore

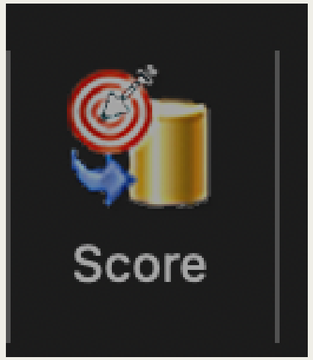
In that sheet, look at the “Testing: Classification Summary” section.

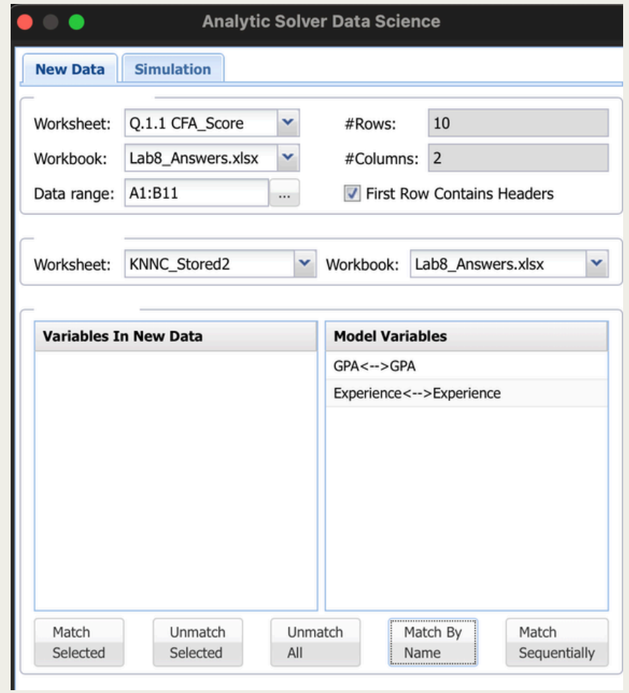
Q 1 C

c. What is the predicted CFA Level 1 outcome for the first job applicant?

- After training the model, we use it to predict for a new applicant (or new observation) and report the predicted class (0 = Fail, 1 = Pass).
  a. Select the data in your Q1.1
  b. Go to Analytic Solver → Data Science → Score → Score New Data
  c. Under Model Worksheet, choose your stored KNN model from before (the best one).
  d. Then click Match by Name to align the variables

The predicted CFA Level 1 outcome for the first new applicant is ??





Thank you!