TU 257 – Fundamentals of Data Science

Data Analytics

Lab 4 – Classification – Part - 1

Brendan Tierney

Agenda

- Assessment-A
  - Putting theory into practice
  - We haven't covered everything – But enough for you to get started
  - Up to and including Week 7 materials needed to complete
  - You don't need to know everything today, but over next few weeks you will

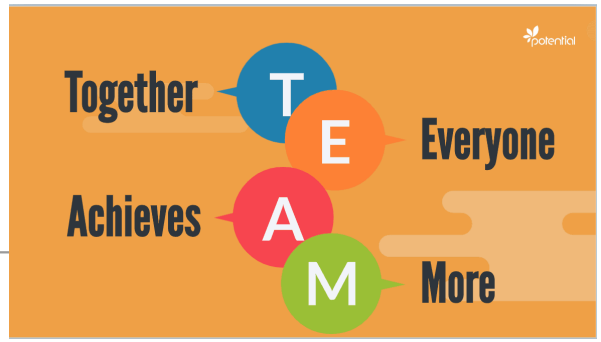
- We have 2 Demo notebooks - Both of these have Exercises
  - Exercise 1 – Demo Notebook 1 – Naïve Bayes
  - Exercise 2 – Demo Notebook 2 – Decision Trees

Assignment A

Assessment

- Assessment Time
- Groups of 3 people
  - Can be similar to your group in Data Wrangling module
- Based on Demos and Exercises from this week and next week
  - Builds upon your learning
  - Applying that learning to a different dataset
- See Assessment Handout in BrightSpace for more details
  - Read carefully and make sure you do everything as specified
- See Template for doc/notebook for completing the assignment.
  - This is what you will submit
  - You will be graded on what it contains
  - Add all code, comments, observations, discussions, explanations, etc.
  - Demonstrate understanding, discussion, additional exploration and insights

Assessment



- Start working/thinking about it early -> today, this week
- Arrange how often you will meet with your Team
- Review what we have covered to-date
- Think about - What can you do to showcase your work and understanding of the topics
- What can you add to the data analysis – What insights can you find?
- …

Exercise 1 – Demo Notebook 1

Exercise 1 – Demo Notebook 1

- Part 1 – Examples/Demo Notebook 1 – 2 examples of using Naïve Bayes
  - Download the Demo Notebook 1
  - Download the Dataset needed for Demo Notebook 1

  - Run all cells
  - Examine the output generated
  - Follow what is happening from cell-to-cell
  - Add extra annotations/descriptions based on your understanding
    - Enrich the descriptions to make them more meaningful for you
  - Complete the Exercises
    - These will require you to write a few lines of code
    - It's ok to look up the Documentation (RTFM)

Exercise 1 – Demo Notebook 2

Exercise 2 – Demo Notebook 2

- Examples/Demo Notebook 1 – 2 examples of using Decision Trees
  - Download the Demo Notebook
  - Download the Dataset needed for Demo Notebook

  - Run all cells
  - Examine the output generated

  - Run all cells
  - Examine the output generated
  - Follow what is happening from cell-to-cell
  - Add extra annotations/descriptions based on your understanding
    - Enrich the descriptions to make them more meaningful for you
  - Complete the Exercises
    - These will require you to write a few lines of code
    - It's ok to look up the Documentation (RTFM)

Exercise 2 – Demo Notebook 2

- See Questions at the end/throughout the Notebook

- Can you add additional code (taking a copy of the code earlier in the notebook)
  - Modify it slightly
  - Examine the results produced

- Compare the results with those found earlier in the Notebook

Exercise 2 – Demo Notebook 2

- Part 1 – Examples/Demo Notebook 2

  - Complete code to Create and Test models for
    - RandomForest
    - XGBoost

  - Compare all models
  - Which one would you recommend using?

Any Questions ?

What Now/Next ?

Complete **all** Lab Exercises before Next Week