# STATS 780/CSE 780, Winter Semester 2018

# ASSIGNMENT 3

**Submit directly to Prof. Sharon McNicholas.**
**Due at 2.30pm on Wednesday, February 14, 2018.**

**Your assignment must conform to the Assignment Standards listed below.**

**Assignments submitted up to 24 hours late will incur a 30% penalty.**
**Assignments submitted more than 24 hours late will receive a zero grade.**

Find a data set that is suitable as a topic of a regression tree or classification tree analysis; you may <u>not</u> use a data set that was discussed in class on January 31 or February 7, 2018. Briefly describe your chosen data set and clearly explain where it was sourced. Carry out a thorough analysis of your chosen data set. Your analysis must include bagging, random forests, and boosting. Provide a clear and concise description of the results. Clearly state what conclusions can be drawn from your analysis — these conclusions should be cast in the context of your chosen data set.

---

<u>Assignment Standards</u>
- LaTeX is strongly recommended but not strictly required.
- Eleven-point font (times or similar) must be used with 1.5 line spacing and margins of at least 1 inch all around.
- Your report may not exceed **three pages**, inclusive of tables, figures, and bibliography.
- Do not include a title page. The title and your name should be printed at the top of the first of the three pages.
- Assignments must be submitted in hard copy, using **two-sided printing** (i.e., printing on both sides of the page), so that only two pieces of paper should be submitted.
- The writing and referencing should be appropriate to the graduate level.
- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.

---