

PGT Exam Performance Feedback

2018/2019 Semester 1

COMP60711 Data Engineering

John Keane
Sandra Sampaio

Comments Section A

In this year's exam paper, as in previous years, simple, bookwork questions were the ones in which students scored higher. As there were several such questions, overall, the marks were good, with plenty of average marks, a significant number of high marks, and a few very low marks.

It was striking to see that a significant number of the students memorised their answers to past exam papers, as they were giving memorised answers to questions that were never present in past exam papers, and thus lost marks for not answering what was being asked.

Section B

Please see the attached report from John Keane.

COMP60711 Data Engineering Section B. Review

Q2 a)

- i) Decision tree creation process generally well described. accurately. Splitting methods were mentioned without detail. Few mentioned how tree depth relates to overfitting.
- ii) Straightforward, most problems were encountered computing Specificity.
- iii) a) Many discussed the bias problem when creating models on imbalanced data; few gave an example.
b) Most answers explained an ensemble model, but not always the differences between bagging and boosting.

Q2 b)

- i) Most students described the algorithm, but not all defined the subset property.
- ii) a) and b) Few students found the all the correct itemsets in C5 or the discarded ones.

Q2 c)

- i) Most defined and contrasted column/row store and how they relate to OLAP/OLTP.
- ii) Few attempted to define data wrangling or discuss utility in future projects.
- iii) Answers were generally good. Few discussed ways to guarantee privacy of the data.