

## COMP4434 Big Data Analytics Team Project

Due Date: **23:59pm, Monday, 22 April 2019**

### Team Formation

- Form a project team with 5-6 members, preferred from the same lab groups. Send me email if you cannot find any team member and we will help you form a project team.
- Each group member should have a distinct role for some part of the project, i.e., to make sure it is done right (not mean you must do all the work for that part or only work solely on that part). Part of your individual grade will depend on how well you lead your team in this regard.
- Each project should have a project leader who ensures work distributed in a fair manner to each team member, the system developed smoothly, and project report and demonstration done properly. Project leaders should report your contact information, team members, project topic to TA (Mr. Qimai Li) in week 7.

### Dataset

This dataset includes crowd flows and meteorological data in Beijing from 2015/11/1 to 2016/4/10. The city is partitioned into a  $32 \times 32$  grid map based on the longitude and latitude. Time is divided into timeslots. The size of each timeslot is 30 minutes meaning there are 48 timeslots in a day. The datasets and source code to access the data can be downloaded from BB.

- *Crowd inflows/outflows*: for a grid  $(i, j)$  that lies at the  $i$ -th row and the  $j$ -th column, the inflow and outflow of the crowds are the total number of taxis that arrive and leave this grid during the timeslot, respectively.
- *Meteorological data*: includes weather (a one-hot vector), temperature (a continuous value), and wind speed (a continuous value) of each day.

### Problems

Choose ONE of the following topic for your project. You can use any models that we learned from class, or extend those models in your way as long as they can improve the performance.

- Design and test a model which applies last inflows/outflows of a grid to predict the flows in next timeslot. Using meteorological data may improve the prediction accuracy.
- Design and test models which classify daily inflows/outflows into groups with different dates (weekends/weekdays).
- Design and test models which classify daily inflows/outflows into groups with different weathers.

## Project Report

A project report should include all the information of this project, e.g.

- Background
- Problem definition
- Model design and considerations
- Solutions
- Performance evaluation and discussions
- Future development
- Reference
- Appendix
  - Documented source files and user manual
  - Contributions of team members

## Project Grading

The project is of 15% of the total subject assessment. There are 2 sets of deliverables throughout the semester.

- Project presentation (5%) will be scheduled in the last week of class
  - Demonstrate your solution
  - Prepare your presentation slides
- Prototype development and reports (10%) due on **22 April 2019**