

COMP-598: Applied Machine Learning

Final project: Analyzing data from the city of Montreal

Due on November 27 (short oral presentation) and December 8 (written report)

Background:

There is a significant effort towards moving much of the data from the city of Montreal into an Open Data format. This data can be accessed here:

<http://donnees.ville.montreal.qc.ca/>

<http://donnees.ville.montreal.qc.ca/english-version-of-the-portail-des-donnees-ouvertes-de-la-ville-de-montreal/>

The goal of this project is to use this data to identify an interesting prediction question that can be tackled using machine learning methods, and solve the problem using appropriate machine learning algorithms and methodology. You are not restricted to using only this data (though you should use some of it). You can incorporate data from other sources, or collect additional data (e.g. new test set) if appropriate.

The choice of prediction task and dataset to use is open. Try to pick a prediction question that is relevant and important to the citizens or administrators of the city. Remember to design a prediction task that is well suited to your choice of dataset; and vice versa, pick the right data for tackling your prediction question.

There are no mandatory algorithms to implement for this project. You can implement your own algorithms, or use any existing toolbox or software, as long as you reference it appropriately in your report.

This project can be completed individually, or in a group of 2 or 3. You can work with any student in the class, including those with whom you worked on one of the previous mini-projects.

Requirements:

There are 2 components to submit:

- 1) Short oral presentation (3min / team) to be presented in class on November 27. One or all team members can speak. Slides can be used (pdf format only). Presentations will be timed precisely, so practice accordingly. The presentation should clearly identify the target prediction question, describe the data used, define the methods applied (or to be applied) and if possible, give preliminary results. Evaluation criteria: Importance of prediction question, pertinence of methods, discussion of results or expected results, quality of delivery, organization. Feedback will be provided by the instructor. Weight: 25% of final project grade.
- 2) Written report (1 / team). The report should have at most 8 pages. It should follow the same format as for the previous mini-projects (double-column, 10pt font, 1" margins or IEEE conference format). The report should clearly present the target prediction task, the data used, and present a clear and well-motivated methodology for analyzing the target question with the selected data. When appropriate, related work (e.g. attempts to tackle a similar prediction question, or analyzes similar data) should be briefly discussed and referenced. Weight: 75% of final project grade.

Submission instructions:

For the oral presentation, instructions will follow for submitting a copy of your slides (pdf only) in a dropbox folder. For the report, use the online conference management system: <https://easychair.org/conferences/?conf=comp598>