Statistical Learning Project

1st Milestone

G08

List of Group Members:
- Riccardo Ceccaroni,
- Giusy Beatrice Colarusso,
- Gabriele Giannotta,
- Francesco Lauro.

Research Title

Air Quality bla bla bla for bla bla bla with bla bla bla

Abstract

L'inquinamento atmosferico è uno dei grandi problemi che affligge le aree metropolitane di tutto il mondo. Il traffico e le industrie svolgono un ruolo significativo. Abbiamo bisogno di implementare modelli che registrino informazioni sulle concentrazioni di inquinanti atmosferici (SO_2, NO_2, ecc.) poichè la deposizione di questi gas nocivi nell'aria sta influenzando la qualità della vita delle persone.

La crescita della disponibilità di dati e l'avanzamento delle tecnologie computazionali stanno rendendo possibile la previsione e l'analisi della qualità dell'aria, fornendo informazioni estremamente utili per controllare l'inquinamento atmosferico.

Main research aim & framework

In questo studio... *Describe the main goal(s) of your project, why you are interested in it, and how you came up with this idea. If there are also secondary goals, list them too.*

Please cite any paper, webpage, video, other kind of support material that help in understading the feasibility and relevance of your work. List all the references in the last Section of this document.

Data source(s)

Air Quality Programmatic APIs

Differently from last year, the Science Journal App is not central to this project. Hence, please describe in details all the software/hardware/tools/platforms/apps/devices/etc you are plannig to use in order to get the data you need.

Data collection

Explain how you are plannig to actually collect the data. How many data-points do you (realistically) hope to collect? Do you foresee any difficulty in the data collection process? Approximatively, how "heavy" do you think your dataset will be in the end?

Model & Methods

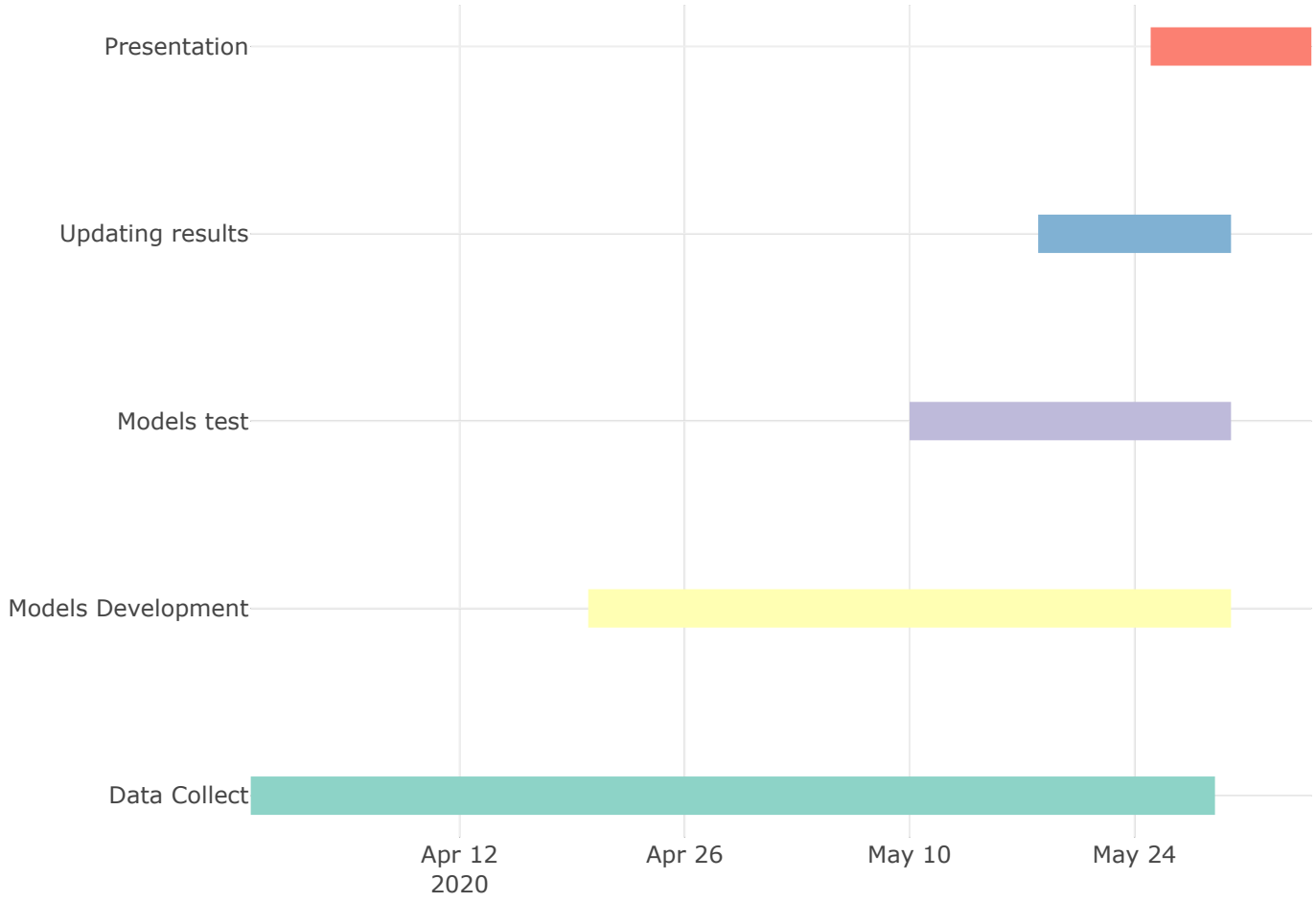
To the best of your current knowledge, what type of statistical/data analysis tools do you feel will be relevant? Explain briefly why.

If there's some cutting edge methods you may like to implement, cite it here adding the relevant paper(s)/book(s) in the last Section of this document.

Software/Hardware Toolkit

Tell me something about the software, programming language(s), package(s), module(s), framework(s) you're planning to use to handle/model/analyse your data and why. Are you also planning to use or develop dedicated hardware for this task? If you feel your home-computer/laptop may not be enough to handle everything, explain to me why and what kind of resources you may instead need.

Project Timeline



References

List here all the reference cited above. If you know what it is, it will be very much appreciated if you use and then upload on Moodle a .bib file. As an example:

- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag.
- Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag.
- Shen, X. and Wasserman, L. (2001). *Rates of convergence of posterior distributions.* The Annals of Statistics, **29**, pp. 687–714.
