**Human AI for Human Development**

**2020 - Bogotá Summer School**

**Emmanuel Letouzé**

**Alvaro Riascos**

Version 3: June 22, 2020

**Professor:** Emmanuel Letouzé

Contact: e-mail: [eletouze@datapopalliance.org](about:blank)

Website: <https://datapopalliance.org/bio/emmanuel-letouze/>

**Professor:** Alvaro Riascos

Contact: e-mail: [ariascos@uniandes.edu.co](about:blank)

Website: [www.alvaroriascos.com](http://www.alvaroriascos.com)

1. **Summary and Course Objectives**

The rise of “Big Data” over the past decade and the more recent emergence of artificial intelligence (AI) have stirred many hopes and, increasingly, fears, about the fate of humankind in the “fourth industrial revolution”. Are we heading towards brighter or darker times? Do Big Data and AI pose existential threats to democracy, or do they offer the possibility of building a future where decisions will be more rational, policies more efficient, processes fairer, politicians more accountable?

This course will introduce you to some of the major applications and implications of Big Data and AI for society: how can personal data be tapped into in ethical, safe ways? How can we build consensus on what “ works” and does not in the age of alternative facts and deep fakes? What is the “good” end result of AI based policies? Is privacy dead, or under threat? Will decisions assisted by algorithms reduce or reproduce and perhaps reinforce discriminations and biases?

One major objective of this course is to introduce a formal framework to address some of these issues. In order to do so, we will introduce students to the most basic machine learning algorithms: linear regression, nearest neighborhood, trees, random forests, neural networks and deep neural networks, etc.) and to the statistical foundations that support most of machine learning applications: approximation vs estimation tradeoff (i.e., bias vs. variance), model validation methodologies (cross validation, bootstrapping, bagging, etc.), regularization techniques, etc.

With this basic machine learning toolbox, we will frame formally some of these challenges. We will learn the difference between prediction and causality and how, what actually works, is related to identifying causality relations. We will address algorithmic fairness in specific models and learn how to potentially fix them and we will also learn how privacy can be preserved through appropriate technical systems and governance standards. Across the lectures we will illustrate many some of these challenges using real world examples: crime prediction models, criminal justice applications, credit allocation, public health, etc.

This course will be taught half in English and half in Spanish.

1. **Contents**

|  |  |  |  |
| --- | --- | --- | --- |
| **Day** | **Topic** | **Professor** | **References** |
| **1** | **Theory:** The Big Data and AI Revolutions for Human Development: Context and Concepts    **Practical session:**  Data Science Ecosystems  Introduction to Python I | **Theory:**  Emmanuel Letouzé  **Practical session:**  Juan Moreno | Letouzé E. (2012), Big Data for Development:  Challenges and Opportunities, United Nations  Global Pulse.  Letouzé and Pentland (2018), Human AI for Human Development [https://www.itu.int/en/journal/002/Pages/15.aspx](about:blank)  A. Pentland,  <https://www.edge.org/conversation/alex_sandy_pentland-the-human-strategy>  Also:  https://www.edge.org/conversation/alex\_sa  ndy\_pentland-reinventing-society-in-thewake-  of-big-data. |
| **2** | **Theory:** Leveraging Data and AI for the SDGs, Official Statistics and M&E: Opportunities, Obstacles, and Requirements.  **Application:** Population density estimation algorithm development and calibration  **Practical session:**  Introduction to Python II | **Theory:**  Emmanuel Letouzé  **Application:** Till Koebe  **Practical session:**  Juan Moreno | J. Jutting and E. Letouzé “Big Data, Official Statistics and Human Development”  [https://paris21.org/sites/default/files/WPS\_OfficialStatistics\_June2015.pdf](about:blank)  Data-Pop Alliance: “Opportunities and Requirements for Leveraging Big Data for Official Statistics and the Sustainable Development Goals in Latin America” (2016)  [https://datapopalliance.org/wp-content/uploads/2016/05/Data-Pop-Alliance-LAC-NSO-EN.pdf](about:blank)  Data-Pop Alliance and Southern Voice: “Harnessing Innovative Data and Technology to Measure Development Effectiveness”: [http://southernvoice.org/wp-content/uploads/2019/08/190814-Ocassional-Paper-Series-No.-54\_final.pdf](about:blank) |
| **3** | **Theory:** The Politics, Economics and Ethics of Big Data, Open Algorithms and AI for Human Development  **Guest perspective:**  “The promise and limits of Big Data for Social Good”  **Practical session:**  Introduction to Python III | **Theory and Case study:**  Emmanuel Letouzé  **Guest perspective:**  Nuria Oliver  **Practical session:**  Juan Moreno | Data-Pop Alliance and Vodafone Institute: “Four requirements for private data sharing and use for public good” (2019): [https://datapopalliance.org/paper-sharing-is-caring-four-key-requirements-for-sustainable-private-data-sharing-and-use-for-public-good/](about:blank)  Lepri, Bruno; Oliver, Nuria; Letouze, Emmanuel F; Pentland, Alex Paul; Vinck, Patrick “Fair, Transparent, and Accountable Algorithmic Decision-making Processes” [https://dspace.mit.edu/bitstream/handle/1721.1/122933/13347\_2017\_279\_ReferencePDF.pdf?sequence=2&isAllowed=y](about:blank) |
| **4** | **Theory:** Basic Methods and Tools for Data Science  **Practical session:**  Web scrapping and satellite imagery analysis | **Theory**  Emmanuel Letouzé, Zinnya del Villar  **Practical session:**  Zinnya del Villar, Rodrigo Lara Molina |  |
| **5** | **Theory:** Statistical Learning Theory Fundamentals  **Application:** Prediction vs. Causality  **Practical session: ,** Sampling from distributions, estimation vs. variance trade-off | **Theory:**  Alvaro Riascos  **Practical session:** Hamadys Benavides | **Theory:**  [LS]  [JWHT]: Capítulo 1,2., [HTF]: Capítulo 1,2.  **Application:**  Athey, Imbens [2019].  Kleinberg, Ludwig, Mullainathan, Obermeyer [2015], Zhao y Hastie (2019) |
| **6** | **Theory:**  Basic techniques: Linear regression, nearest neighbor, logistic regression, trees, random forest, regularization cross validation, bagging, ROC curve  **Application:** Credit risk discrimination and moral hazard  **Practical session:**  The basic ML techniques in Python | **Theory:**  Alvaro Riascos  **Practical session:**  Hamadys Benavides | **Theory:**  [JWHT]: Chapters 3,4,5,6  **Application:** Fuster, Goldsmith-Pinkham, Ramadorai, Walther. [2018].  Sendhil Mullainathan and Ziad Obermeyer. [2017] |
| **7** | **Theory:**  Algortihmic Fairness, Bias, Feedbackloop, Discrimination  **Application:**  Crime prediction  **Practical sesssion:** Crime Models | **Theory:**  Alvaro Riascos  **Practical sesssion:**  Juan Sebastian Moreno | **Theory:**  Lum, Isaac [2016].  Ensign, Friedler, Scheidegger, Venkatasubramaniany. [2018]  **Application:**  Urcuqui, Moreno, Riascos [2019]  Dulce, Ramirez y Riascos [2018] |
| **8** | **Theory:**  Causality in Machine Learning*.*  **Application:**  TBA  **Practical session:**  Graphical models and partial dependece plots | **Theory:**  Alvaro Riascos    **Practical session:**  Hamadys Benavides | **Theory:**  [P]: |
| **9** | **Theory:**  Privacy  **Guest perspective:**  “Covid and the Privacy of Information”  **Practical session:**  A toy model of privacy preserving ML | **Theory:**  Alvaro Riascos  **Guest perspective:**  Roberto Rigobon (MIT)  **Practical session:**  Hamadys Benavides | **Theory:**  Catherine F. Higham, Desmond J. Higham. [2018] |

1. **Methodology**

This is an online course. We will use Zoom and all lectures will be recorded. Almost all sessions will run in blocks of 50 minutes of lectures and 10 minutes of break.

The course will be graded in the following way:

* Quiz 1 (15% of the final grade)
* Quiz 2 (15% of the final grade)
* Programming assignment (35% of the final grade)
* Final Exam (35% of the final grade)

All tests will be multiple choice. They will be managed using University of los Andes Sicua Platform. The assignment is a long fill-in-code assignment that students will answer daily but graded at the end.

Further logistic details will be given to registered students.

1. **Grading System**

Grades will be numeric between zero and five with two decimal places. For students of the university of los Andes grades will be mapped to and registered as Pass/Fail (Aprovado/Reprovado).

1. **University Rules and MAAD Protocol**

This course is subject to the University of los Andes general guidelines and principles as stated in: <https://secretariageneral.uniandes.edu.co/images/documents/reglamento-maestria-web-2020.pdf>

Any member of the community that is subject to, witnesses or has knowledge of a conduct of mistreatment, harassment, threat, discrimination, sexual or gender violence (MAAD) must report the case to the University. This, in order that institutional actions can be taken to handle the case, ensuring the well-being of the affected people.

For additional information please visit: <https://decanaturadeestudiantes.uniandes.edu.co/index.php/es/sobre-la-decanatura/827>

or write to: [lineamaad@uniandes.edu.co](mailto:lineamaad@uniandes.edu.co)

1. **Main References**

* [LS]: Luxburg, U., B. Scholkopf. 2008. Statistical Learning Theory: Models, Concepts and Results.  
  [http://arxiv.org/abs/0810.4752](about:blank)
* [JWHT]: Introduction to Statistical Learning with Applications in R.

[http://www-bcf.usc.edu/~gareth/ISL/](about:blank)

* [HTF]: Hastie, T., Tibshirani, R. y J. Hastie. 2009. The Elements of Statistical Learning: Data Minning, Inference and Prediction. Segunda Edición. Springer

[http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII\_print4.pdf](about:blank)

* [P]. Pearl, J. 2016. Causal Inference in Statistics - A Primer. Wiley.
* [AP]: Joshua Angrist and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics. Princeton University press.
* Athey,S.,and G. Imbens. 2019. Machine Learning Methods Economists Should Know About.
* Dulce, M., Ramirez, S. y A. Riascos (2018). Efficient allocation of law enforcement resources

using predictive police patrolling.   
[https://arxiv.org/pdf/1811.12880.pdf](about:blank)

* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. Prediction Policy Problems
* Lum, Isaac. 2016. To predict and to serve?
* Zhao, Q y T. Hastie (2019). Causal Interpretation of Black Box Models. [https://web.stanford.edu/~hastie/Papers/pdp\_zhao.pdf](about:blank)
* Ensign, Friedler, Scheidegger, Venkatasubramaniany. 2018. Runaway Feedback Loops in Predictive Policing
* Fuster, Goldsmith-Pinkham, Ramadorai, Walther. 2018. Predictably Unequal? The Effects of Machine Learning on Credit Markets.
* Sendhil Mullainathan and Ziad Obermeyer. 2017. Does Machine Learning Automate Moral Hazard and Error?
* Catherine F. Higham, Desmond J. Higham. 2018. Deep Learning: An Introduction for Applied Mathematicians.