

Sergio Villordo

Data Scientist, PhD

Sant Antoni Maria Claret 87, 08025
Barcelona, Spain
NIE:Y7607842P
☎ +34 (689) 350 029
✉ villordosergiomanuel@gmail.com
📁 sergiovillordo.github.io/webpage/
🌐 sergio-manuel-villordo-7154b0a6
gitlab.com/sergio08



"Creative and enthusiastic Data Scientist with a huge experience in mining and modelling of complex data"

Core Competencies

- 10+ years working in multidisciplinary data analytics environments.
- 4+ years experience at high-tech companies, filling the gap between customers' requirements and the cutting-edge Artificial Intelligence algorithms.
- Strong team-work skills with experience in consulting and business partnering at highly international environments.
- Wide experience in developing end-to-end data analytics and machine learning solutions:
 - Identification of stakeholder needs, business understanding and scrum methodology adoption.
 - Data exploration and valuable insights extraction.
 - Feature engineering and development of prototype models using python and R.
 - Deployment of the most appropriate solution using docker or moving to scala/java.
 - Testing, benchmarking, performance analysis (A/B testing) and future improvement.
- Strong interpersonal and communication skills with ability to clearly communicate technical content to non-experts.

Key data analytics skills: "Customer Lifetime Value (CLV)", "Churn / Attrition", "Forecast", "Price Optimization", "Image Content Moderation", "Computer Vision", "Customer Segmentation", "Customer Satisfaction Surveys Analysis".

Software and Frameworks Skills

Python and R: My go-to languages for data exploration, visualization, hypothesis formulation and machine learning prototyping.

Machine learning and data exploration tools: I am proficient using python packages (Scikit-Learn, numpy, scipy, pandas, matplotlib, seaborn, NLTK, etc.), R libraries (tidyverse, datatable, dplyr, ggplot2, plotly, tidytext, tm, etc.). Also, I have used Spark MLlib, matlab, SAS, H2O, SPSS, pentaho, openrefine and tesseract for specific tasks.

Deep learning tools: Hands-on experience using DL architectures and transfer learning. Proficiency using high-level DL APIs (keras, Mxnet and fast.ai) and intermediate knowledge on low-level Tensorflow and pytorch APIs. Also, I have used theano, caffe and deeplearning4j in specific projects.

Computer vision tools: Hands-on experience with opencv, imageJ, skimage, PIL.

Scala and java: I have used them on several occasions to deploy code at production stage.

Spark: Hands-on experience using spark (through pySpark or Scala) in Big Data projects.

Docker: I have used containers to deploy python models at production stage.

Databases: Wide experience in the management of data stored at MySQL, PostgreSQL, presto SQL, mongoDB, cassandra, redshift.

Cloud environments: Hands-on experience using virtual machines and storing data at AWS.

Project management: Git tools (gitlab, github), Agile tools (JIRA, FishEye, trello).

Reporting: Proficient use of Shiny, dash, d3.js, markdowns, \LaTeX for reporting, and I have used QlikView and Tableau for specific tasks.

Experience at High Tech Companies

Despegar (www.despegar.com)

2017–2019 **Full-Stack Data Scientist**, Buenos Aires, Argentina.

With millions of daily active users, Despegar is one of the leading online-travel companies at LATAM.

At Despegar, I was responsible to drove end-to-end Data Science solutions in three different projects:

- **Image Content moderation**: In this project I worked as the lead of a interdisciplinary team. We developed a set of micro-services and packages for image scene labelling, image quality filtering and removal of duplicates using a combination of deep learning and traditional computer vision approaches.
- **Customer Lifetime Value analysis**: I worked together with a data engineer and we developed a api-rest service to provide a real time estimation of users CLV using xgboost.
- **Customer satisfaction surveys analysis**: I Developed a dashboard tool to do NLP analysis from customer opinions using R and Shiny.

Etermax (www.etermax.com)

2016–2017 **Senior Data Scientist**, Buenos Aires, Argentina.

Etermax is a mobile game development company, its best known game is trivia-crack.

At Etermax I was part of a multidisciplinary team, and we worked in the following tasks:

- **Image Content moderation**: In this project we developed models based on deep learning and decision trees for automatic image filtering.
- **Customer Lifetime Value analysis**: We developed probabilistics and ML models for generate several CLV estimations and the results were reported through a dashboards builded with Shiny.

Globant (www.globant.com)

2015–2016 **Semi/Senior Data Scientist** at Globant, Buenos Aires, Argentina.

Globant is an IT and Software Development company which mainly serves clients from the United States and Europe.

At Globant, I worked in several interdisciplinary groups of professionals (Data Engineers, architects, BI specialist, etc.) focused on a wide range of task: Price Optimization, stock forecasting, Churn/Attrition, year-to-year crop yield forecasting and topics modelling.

Experience at the Academy

Fundación Instituto Leloir (www.leloir.org.ar)

2015–2016 **Master's degree Thesis**, Buenos Aires, Argentina.

The master's Thesis work was performed at the Molecular Virology Department and the objective was the development of new Data Analysis tools for the determination of virus variants using New Generation Sequencing (NGS) technologies. The results of these works have been published as part of a scientific paper in a high impact peer-review Journal.

Fundación Instituto Leloir (www.leloir.org.ar)

2010–2015 **PhD Thesis**, Buenos Aires, Argentina.

The objective of this thesis was the discovery of new insights into dengue virus replication using bioinformatics and Molecular Biochemistry approaches. As a result of this work I gained good experience in scientific research, High Throughput Sequencing Data Analysis, Bioinformatics and Biochemistry and hands-on in Molecular Laboratory lab. The results of the these findings were published as part of 9 original scientific articles in high impact indexed journals and defended in several international conferences.

Commonly used Machine Learning Algorithms:

- **Classification / Regression:** Xgboost, RandomForest, SVM, logistic Regression, Ridge Regression, Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), probabilistics methods, Association Rules, Correspondence Analysis, survival analysis.
- **Dimensionality Reduction:** PCA, MCA, tSNE, Deep Autoencoders.
- **Clustering:** Kmeans, DBSCAN, hierarchical clustering, kNN.
- **Natural Language Processing:** Tokenization and n-grams analysis, topic modelling, sentiment analysis, words representation and transfer learning for text classification.

Commonly used Computer Vision Algorithms:

- **Image Processing:** Equalization, segmentation (Thresholding, Watershed), transformation using kernels (canny, sobel, blurring, sharpening), Fourier Transform applications, perceptual hashing, interpolation.
- **Image Classification:** Traditional feature extraction algorithms (SIFT, HOG, LBP, GIST). Convolutional Neural networks (CNNs), mask-RCNN, fast-RCNN. Use of pre-trained models and transfer learning. Model Ensemble and combination of Deep learning and traditional computer vision approaches.

Education

- 2015–2016 **Master's Degree in Data Science and Knowledge Discovery**, University of Buenos Aires, Argentina.
- 2010–2015 **PhD in Molecular Biology and Bioinformatics**, University of Buenos Aires, Argentina.
- 2012–2012 **Specialist in Scientific Journalism**, Universitat Pompeu Fabra, Argentina headquarter.
- 2002–2009 **Bachelor of Science in Genetics**, University of Misiones, Argentina.

Languages

Spanish	Native
English	Intermediate to advanced
Portuguese	Intermediate

Latest teaching experience and conferences

- 2018–2019 **Course:** An Introduction to Artificial Intelligence for managers, **Professor**, Universidad Argentina de la Empresa, UADE, Argentina.
- 2016–2019 **Course:** Clustering Approaches for Data Scientist, **Profesor**, Escuela Argentina de las Nuevas Tecnologías, EANT, Argentina.
- 2018 **Conference:** Getting started with AI tools for viral variants characterization, **Speaker**, Winter School in Molecular Biology. Federal University of Latin-American Integration, Foz do Iguaçu, Paraná, Brazil.
- 2018 **Conference:** A guide to start with RNA structure analysis, **Speaker**, Winter School in Molecular Biology. Federal University of Latin-American Integration, Foz do Iguaçu, Paraná, Brazil.

Latest publications at indexed peer-review Journals

- [1] L. de Borba, S. M. Villordo, F. L. Marsico, J. M. Carballeda, C. V. Filomatori, L. G. Gebhard, H. M. Pallares, S. Lequime, L. Lambrechts, I. Sanchez Vargas, C. D. Blair, and A. V. Gamarnik. RNA Structure Duplication in the Dengue Virus 3' UTR: Redundancy or Host Specificity? *MBio*, 10(1), Jan 2019.
- [2] C. V. Filomatori, J. M. Carballeda, S. M. Villordo, S. Aguirre, H. M. Pallares, A. M. Maestre, I. Sanchez-Vargas, C. D. Blair, C. Fabri, M. A. Morales, A. Fernandez-Sesma, and A. V. Gamarnik. Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. *PLoS Pathog.*, 13(3):e1006265, 03 2017.
- [3] S. M. Villordo, J. M. Carballeda, C. V. Filomatori, and A. V. Gamarnik. RNA Structure Duplications and Flavivirus Host Adaptation. *Trends Microbiol.*, 24(4):270–283, Apr 2016.
- [4] S. M. Villordo, C. V. Filomatori, I. Sanchez-Vargas, C. D. Blair, and A. V. Gamarnik. Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog.*, 11(1):e1004604, Jan 2015.

Please, find a complete list of publications at indexed journals following the link:

<https://www.ncbi.nlm.nih.gov/pubmed/?term=Villordo+SM>

Contacts for references

Gonzalo Zarza, PhD Gonzalo Zarza is the Technical Director of Big-Data Studio at Globant.
Email: gonzalo.zarza@globant.com

Please find more information about Gonzalo at the following link:

<https://www.linkedin.com/in/gonzalozarza/?originalSubdomain=es>

Andrea Gamarnik, PhD Andrea is the head of the Molecular Virology lab at Fundación Instituto Leloir and she was my PhD thesis director. Today I continue doing research collaboration with the Andrea's lab.

Email: agamarnik@leloir.org.ar

Please find more information about Andrea at the following link:

<http://www.leloir.org.ar/gamarnik-en/head-of-laboratory/>