Manuel Alejandro Matías Astorga

Mexico city, Mexico. ZIP Code: 07750
alejandromatiasastorga@gmail.com
(+52) 951.310.47.07
alexmatiasas.github.io
linkedin.com/in/alexmatiasastorga
kaggle.com/alejandromatias
github.com/alexmatiasas
alexmatiasas.github.io/projects

DATA SCIENTIST

Data Scientist with a PhD in Physics and strong experience in machine learning, statistical modeling, and real-world data projects. Focused on delivering business value through applied analytics, NLP, and deployment-ready pipelines.

INTERESTS

Remote-first roles | Applied Data Science | Open to relocation (Europe / Canada)

PROFESSIONAL EXPERIENCE

PARTICLE PHYSICS RESEARCHER |CINVESTAV — Physics Department.
2022—2024| CDMX, México.
As part of my research project, I developed new models to obtain particle masses. Furthermore, based on lattice quantum chromodynamics (LQCD), the MIT bag model and Tsallis statistics, I found a new way to obtain pressures inside hadrons.

- Developed symbolic and numerical simulations using Python (pandas, NumPy, scikit-learn, Matplotlib) to model pressure distributions inside hadrons.
- Analyzed thermodynamic observables under Tsallis formalism with academic publishing in progress.

HIGHLIGHTED PROJECTS

SENTIMENT ANALYSIS ON IMDB REVIEWS (R-BASED) 2024—Ongoing

- Conducted full EDA on 50,000 IMDB movie reviews using R and RMarkdown.
- Preprocessed text with HTML tag removal, punctuation handling, stemming, and lemmatization (SnowballC, textstem, UDPipe).
- Visualized word distributions, bigrams, and sentiment polarity with 'ggplot2'.
- Documented findings in a reproducible '.Rmd' report with interactive tables and lollipop plots.
- Planning model development using Python and 'scikit-learn'/'PyTorch' for sentiment classification.
  [Portfolio] | [GitHub]

FRAUD DETECTION WITH MACHINE LEARNING (PYTHON) 2024—Ongoing

- Developing a binary classification pipeline to detect fraudulent transactions using real-world datasets.
- Handling imbalanced data through resampling techniques (SMOTE, undersampling) and performance metric analysis (ROC-AUC, precision/recall).
- Designing modular code structure with 'pandas', 'NumPy', and 'scikit-learn', separating preprocessing, model training, and evaluation.
- Aiming to benchmark different classifiers (Random Forest, Logistic Regression, XGBoost) using cross-validation.

STATISTICAL PHYSICS OF HADRONS USING TSALLIS FORMALISM (PH.D. RESEARCH) 2022—2024

- Modeled hadronic systems as non-extensive quark–gluon gases using Tsallis statistics.
- Derived pressure, entropy, and energy distributions in relativistic thermodynamic frameworks.
- Conducted symbolic and numerical analysis with 'NumPy', 'SymPy', 'Matplotlib', and 'SciPy' in Jupyter Notebooks.
- Visualized pressure profiles and compared results with experimental lattice QCD constraints.

MODIFIED LOTKA–VOLTERRA MODEL WITH FRACTIONAL CALCULUS (BACHELOR'S THESIS) 2019

- Developed a nonlinear population dynamics model using fractional differential equations in a fractal metric space.
- Implemented simulations in Python using Finite Element and Finite Difference Methods via the 'FEniCS' library.
- Explored chaotic behavior in predator–prey systems and compared numerical results against analytical benchmarks.
- Documented results using Jupyter Notebooks and visualized time-series dynamics.

EDUCATION

- **PHD IN PHYSICS** |CINVESTAV — Physics Department
  2019—2023 | Mexico city, Mexico
  Thesis: Pressure distribution in hadrons using non-extensive Tsallis statistics.
- **B.SC. IN ENGINEERING PHYSICS** |Mixteca Technological University — Physics Institute

2013—2019 | Huajuapan de León, Oaxaca, Mexico
Thesis: Population dynamics with fractional calculus and finite element methods (Python).

SKILLS

• TECHNICAL SKILLS

• Programming Languages:
Python (Advanced for Data Analysis and Scripting, Intermediate for Machine Learning), R (EDA and NLP), SQL (Basic querying)
• Scientific & Data Libraries:
Pandas, NumPy, SciPy, SymPy, Matplotlib, Seaborn, tidyverse, ggplot2, tidytext
• Machine Learning & NLP:
Scikit-learn (classification pipelines, model evaluation), TensorFlow (basic), PyTorch (planned), NLP (tokenization, stemming, lemmatization, POS tagging, n-grams)
• Tools & Environments:
Jupyter Notebooks, RMarkdown, Git (GitHub), Docker (basic), VSCode
• Cloud Platforms:
AWS (S3, Lambda, EC2 – exploratory use)
• Databases:
PostgreSQL, MySQL, SQLite (academic and project-based use)
• Operating Systems:
Linux (Fedora, Ubuntu), macOS

•LANGUAGES

•Native
Spanish
•Advanced
English (C1)
•Beginer
German (A2) & Japanese (A1 / N5)

•SOFT SKILLS

•Strong analytical thinking and scientific rigor
•Effective communication with non-technical stakeholders
•Agile collaboration with cross-functional teams
•Proactivity and solution-oriented mindset
•Capacity to handle complex projects independently

CERTIFICATIONS

• **THE COMPLETE PYTHON BOOTCAMP FROM ZERO TO HERO IN PYTHON** |Udemy—Por José Portilla, Pieran Training
2022—2023| Udemy.
THE COMPLETE SQL BOOTCAMP: GO FROM ZERO TO HERO |Udemy—Por José Portilla, Pieran Training
2022—2023| Udemy.
MACHINE LEARNING A-Z: AI, PYTHON AND R |Udemy—Por Kirill Eremenko, Hadelin Ponteves (SuperDataScience Team, Ligency Team)
2023—2024| Udemy.
ARTIFICIAL INTELLIGENCE A-Z 2024: BUILD 7 AI + LLM |Udemy—Por Kirill Eremenko, Hadelin Ponteves (SuperDataScience Team, Ligency Team)
2023—2024| Udemy.