linkedin.com/in/diegobravoguerrero

Email: dbravo27@gmail.com

Location: Plano, Texas

DR. DIEGO BRAVO

OVERVIEW

With over 7 years of work experience in Machine Learning after receiving a PhD in Mathematics, my day-to-day tasks involve working with large and complex data sets. I solve a wide array of challenging problems using Data Engineering, Machine Learning, Cloud Computing, Big Data, Statistical Modeling, and other IT tools. Partnering with cross-functional teams, I inform, influence, support, and execute business strategies.

SKILLS

Programming Languages: Python, SQL, PySpark, HTML, CSS

Machine Learning: Scikit-Learn, XGBoost, LightGBM, CatBoost, spaCy, HuggingFace, MLflow, AutoML, Prophet, Optuna

Deep Learning: Pytorch, Tensorflow, Keras, OpenCV, YOLO, Convolutional Neural Networks, Transformers, BERT

Data Wrangling: Pandas, Numpy, Matplotlib, Seaborn, Scipy, Plotly, Statistical methods

Cloud Platforms: Azure, AWS, GCP

Big Data & Data Engineering: Databricks, DVC, Apache Airflow,

Databases: PostgreSQL, SQL Server, Azure Databases, Azure Data Studio

Deployment: Docker, FastAPI, GitHub Actions, Streamlit, Gradio, Azure Web Apps.

Project Management: Git, Github, Jira, GitHubProjects, AzureDevops, Bitbucket, Slack lists.

Generative AI: Azure Foundry, RAG, Azure OpenAI, OpenAI, Gemini API, LLama, LangChain, HuggingFace, Agents

Research: Active researcher with 11 published research papers

EXPERIENCE

Cresteo | Chicago, IL – *AI Leader and Senior Data Scientist*

JUL 2023 - PRESENT

1. Lead Data Scientist developing production-grade Recommender System for multi-tenant Learning Management System (2.8M interactions, 267K users, 395K courses). Architected hybrid recommendation pipeline combining ALS collaborative filtering, NLP-based content analysis, and popularity models using PySpark/Databricks on Azure. Implemented intelligent content deduplication, cold-start handling, and achieved 95% user coverage.

2. Pricing-Optimization Engine for a Top U.S. Manufacturing Company – Designed and deployed a machine-learning platform that slashed quote turnaround from 24 h to 2 h (97 % on-time). Leveraged advanced feature engineering, XGBoost models, and automated hyper-parameter optimization to recommend optimal prices, delivering an estimated 12% profit lift and ~20 % productivity gain while enhancing client relationships through faster, more accurate quotes.

3. Developed an AI-powered sales agent with up-to-date knowledge of publicly traded companies, generating high-quality 20+ page financial reports and enabling interactive chat via text, audio, and open-mic conversations. This

led to a 25% increase in client engagement, supporting targeted strategies for a leading Management and AI consultancy in Minnesota. Leveraged state-of-the-art Generative AI models from Azure OpenAI and Google, utilizing tools like Azure Cognitive Services, Gemini API, Yahoo Finance API, SEC Downloader, Tavily, Selenium, FastAPI, Docker, Azure Web Apps, and GitHub Workflows to deliver advanced customer insights.

4. Achieved a 20% increase in revenue and a 40% improvement in operational efficiency through real-time optimization strategies, by developing end-to-end AI models in the truck logistics sector,

- Enhancing operational techniques using Azure OpenAI and LangChain, integrating Big Data analytics for scalable and robust solutions.
- Conducting simulations with real-world constraints and A/B testing to ensure model reliability and effectiveness in operational environments.
- Leading the development of data pipelines and machine learning algorithms, utilizing Microsoft Azure, Databricks, Python, PySpark, Pandas, SQL, Cloud-based Databases.

5. Developed a data science-driven site selection tool to objectively evaluate clinical sites using a scoring mechanism, resulting in a 30% improvement in site selection accuracy and streamlining negotiation processes for clinical trials at a Healthcare Data Management Platform Company. Leveraged tools and frameworks for data manipulation and analysis using Python, Pandas, Matplotlib, Scikit-learn, XGBoost, LightGBM, TensorFlow, FastAPI, Docker, and MLflow.

Cencosud – *Principal Data Scientist*

JUL 2023 - NOV 2023

Driving AI innovation, I spearheaded initiatives and led a team of over 30 data scientists to align with stakeholder expectations, positioning South America's largest retail company at the forefront of AI advancements.

1. Advanced Customer Feedback Analysis with NLP: Developed an NLP pipeline using BERT for sentiment analysis and BERTopic for topic modeling, extracting insights from customer feedback across e-commerce platforms. This enhanced product performance metrics and guided strategic improvements while staying within budget and timeline.

2. Supermarket Forecasting and Inventory Optimization in Peru: Refactored forecasting and inventory management using machine learning models such as Prophet and XGBoost, reducing stockouts by 5% and waste by 10%. Optimized cloud infrastructure with serverless computing, cutting costs by 30% (~$40K annually). Implemented real-time ETL pipelines, improving operational efficiency and ensuring product availability.

3. MLOps Transformation and Efficiency Gains: Led the transformation of MLOps processes, automating model retraining, validation, and deployment, elevating operations from MLOps Level 0 to Level 1. Integrated CI/CD pipelines, containerized components, and continuous monitoring, reducing cloud costs by 15% (~$120K annually) while improving agility and system reliability.

4. AI Strategy and Vendor Evaluation: Developed an AI roadmap using OKRs and multi-criteria analysis, prioritizing initiatives such as GenAI for e-commerce and computer vision for in-store analytics. Evaluated external vendors, leading to successful implementations like customer behavior tracking in malls while avoiding redundant investments, saving ~$50K annually. Oversaw AI consulting projects, ensuring knowledge retention and optimal system integration.

Tryolabs – *Senior Data Scientist*

AUG 2020 - MAR 2023

Successfully deployed and managed comprehensive machine learning projects, leading teams of data scientists and machine learning engineers to drive innovation and excellence. I ensured that client deadlines were met, budget constraints adhered to, and resources optimally utilized, working across multiple cloud platforms and employing a wide range of artificial intelligence technology tools and algorithms.

Top 5 Projects:

1. Led an InnerSource initiative for a top tier Oil & Gas US company's AI and Data Science department, fostering collaboration and knowledge sharing among 400+ data scientists. Designed and developed an extensive InnerSource program, creating over 30 Python and Data Science courses. The initiative also established best-practice guidelines and hands-on workshops, enabling seamless skill development, standardization, and cross-team innovation in AI and data-driven projects.

2. Boosted AB InBev, a global beverages firm's profits by $10M in just 6 months by developing an assortment optimization API. This API, leveraging Discrete Choice Models, now supports 300,000 stores in the Mexico region, handling datasets exceeding 500GB.

3. Boosted FinTech loan approval efficiency by 70% developing an NLP API, achieving a 90% model performance. Streamlined legal document review, customer data matching, and information extraction through OCR and Named Entity Recognition with BERT, Regex, and spaCy.

4. Developed a 99% accurate image classification API in the retail fashion industry, using CNNs, PyTorch, Torchvision, Pillow, and Python libraries. This project, part of a larger initiative, labeled large datasets to analyze customer preferences for images of mannequins versus human models in various poses.

5. Transformed pricing strategies for a US real estate firm by developing a price forecasting API, leading to a 20% increase in sales conversions and a 3% uplift in profit margins within the first year of implementation. A\B testing techniques were applied.

Engineering School - University of the Republic – *Professor & Researcher*

`MAY 2014 - AUG 2020`

Courses Taught: Statistics, Statistical modeling, Data Mining, Linear Algebra, Calculus, Discrete Mathematics,, Spectral Graph Theory.

Advisor of 1 Doctoral and 2 Master theses.

Authored 11 research papers.

Speaker in 20 International conferences, in 10 different countries.

Andres Bello Catholic University – *Professor & Data Scientist*

`SEP 2008 - APR 2014`

Data Science projects

1. Significantly improved budget planning accuracy within the Finance department by forecasting student drop-offs and future enrollments, using advanced ARIMA and SARIMA models developed with Time Series Analysis in R.

2. Improved student retention strategies by identifying key factors influencing engagement and drop-off, facilitating the Provost's office in implementing targeted initiatives across undergraduate colleges.

3. Led the "Forecasting with Time Series" weekly seminar, focusing on discussions around book chapters, recent research papers, and practical use cases in Data Mining

Courses taught: Statistics 1, Statistics 2, Calculus 1, 2 and 3. Linear algebra. Differential Equations. Problem solving strategies.

Creator and Director of the Mathematical Olympiads of the Catholic University for 4 consecutive years with +600 participants from +30 High Schools and Colleges.

EDUCATION

Universidad de la República, Uruguay — *Postdoctoral Fellow*

Mar 2014 - Feb 2016

Focus: Representation Theory of Algebras

Universidad de Los Andes, Venezuela — *PhD in Mathematics*

OCT 2006 - SEP 2010

Thesis: Cotorsion Pairs and Adjoint Functors in C(R-Mod)

Advisor: Dr. Edgar Enochs, University of Kentucky.

Specialization: Homological Algebra

Universidad de Los Andes, Venezuela — *Master of Science in Mathematics*

SEP 2005 - SEP 2006

Thesis: Spectral Properties of Unicyclic Directed Graphs

Advisor: Dr. Juan Rada.

Specialization: Spectral Graph Theory

Universidad de Los Andes, Venezuela — *Bachelor of Science in Mathematics*

SEP 2000 - AUG 2005

Summa Cum Laude

CERTIFICATIONS

- Microsoft Certified
- Databricks Certified
- Python Certified
- Large Language Models in Azure OpenAI
- Git & GitHub

AWARDS

Summa Cum Laude: Graduated top of the class. Bachelor's degree.

Bronze medal at National Mathematics Olympiads

Best Student of the College of Sciences: Awarded for 3 consecutive years.

LIST OF PUBLICATIONS

1. Some families of digraphs determined by the complementarity spectrum

https://www.sciencedirect.com/science/article/abs/pii/S0024379525000229

2. Complementarity spectrum of digraphs

https://www.sciencedirect.com/science/article/abs/pii/S0024379521002354?via%3Dihub

3. Characterization of digraphs with three complementarity eigenvalues

https://link.springer.com/article/10.1007/s10801-023-01218-6

4. Generalised Igusa-Todorov functions and Lat-Igusa-Todorov algebras

https://www.sciencedirect.com/science/article/pii/S002186932100171X

5. Idempotent reduction for the finitistic dimension conjecture

https://www.ams.org/journals/proc/2020-148-05/S0002-9939-2020-14945-3/home.html

6. Jahangir Graphs

https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/30769

7. Energy of matrices

https://www.sciencedirect.com/science/article/abs/pii/S0096300317303636

8. Pullback diagrams, syzygy finite classes and Igusa–Todorov algebras

https://www.sciencedirect.com/science/article/pii/S002240491930026X

9. Cotorsion pairs in C(R-Mod)

https://projecteuclid.org/journals/rocky-mountain-journal-of-mathematics/volume-42/issue-6/Cotorsion-pairs-in-CR-Mod/10.1216/RMJ-2012-42-6-1787.full

10. Nonderogatory Unicyclic Digraphs

https://onlinelibrary.wiley.com/doi/10.1155/2007/46851

11. Coalescence of Difans and Diwheels

https://www.emis.de/journals/BMMSS/vol30_1_6.htm