Isaac Vergara

isaacvergaram@hotmail.com +1 (919) 619-3370 linkedin.com/in/isaac-vergara github.com/isaacvm98

Education

Duke University, Durham, NC **May 2027**

Master of Science in Data Science, Quantitative Finance Concentration | **GPA: 4.0/4.0**
Relevant Coursework: Algorithmic Trading & Investment, Machine Learning, Deep Learning, Natural Language Processing, Stochastic Calculus, Financial Time Series, Risk Management & Derivatives
Awards: Merit Scholarship & Mexican National Science and Technology Scholarship (CONAHCyT)

Universidad Panamericana, Mexico City, Mexico **December 2021**

Bachelor of Economics | **GPA: 3.5/4.0**
Relevant Coursework: Econometrics (I, II, III), Stochastic Calculus, Game Theory, Risk Management, Spatial Econometrics, Differential Equations and Optimization, Inferential Statistics, Corporate Finance (I, II)
Study Abroad: London School of Economics - Political and Economic Summer Program (2018)
Awards: Merit Scholarship (4 years, maintained 8.5/10 GPA requirement)

Microsoft Certified: Azure Data Scientist Associate **2022**

Microsoft Learn - Professional certification in Azure Machine Learning Studio, model deployment, and cloud-based ML workflows

Professional Experience

Semi Sr. OT Data Scientist, Enable Global **July 2023 - August 2025**

- **Quantitative Analysis & Statistical Modeling:** Led comprehensive statistical analysis using ANOVA, anomaly detection algorithms, and non-parametric hypothesis testing to identify root causes of manufacturing defects in steel transformation processes, enabling data-driven decision making for $2M+ system optimization and installation of new heater systems
- **Production ML Systems:** Engineered automated data quality web application using Python, incorporating cosine similarity algorithms and clustering techniques for robust deduplication, leveraging Pydantic for data validation and schema enforcement; processed 1M+ master and spare parts records for multiple manufacturers
- **Executive Communication:** Presented complex quantitative findings and statistical models to C-suite stakeholders across manufacturing and energy sectors, translating technical analysis into actionable business insights that drove operational efficiency improvements
- **End-to-End Data Science Solutions:** Developed understandable dashboards with Dash as deliverables, creating complete data science solutions from problem definition through production deployment

Data Scientist & Product Manager, Entropia AI **August 2021 - July 2023**

- **Economic Theory Implementation:** Architected production-ready Almost Ideal Demand System (AIDS) model, translating advanced economic theory (Deaton & Muellbauer, 1980) into real-time quantitative production system with automated price elasticity estimation pipeline that updates upon availability of new economic indicators, ensuring model calibration with current market conditions
- **Time Series Forecasting & Optimization:** Implemented hierarchical multi-level SARIMAX forecasting system for pharmaceutical distribution, improving prediction accuracy from 88% to 94% through advanced statistical modeling across various business dimensions (SKU, store, region) to optimize budget allocation
- **Cloud Infrastructure & ML Deployment:** Deployed production models using Prefect for data pipeline workflow orchestration and AWS ElasticBeanstalk for inference service, enabling scalable sales and demand forecasting solutions
- **Cross-Functional Leadership:** Led teams of 5+ data analysts and scientists as Product Manager for multiple projects, delivering quantitative data science solutions on time and within budget across diverse industries including pharmaceuticals, manufacturing, and retail
- **State-of-the-Art Model Implementation:** Utilized cutting-edge models and cloud technologies (AWS, Azure) to deliver high-value solutions to clients, incorporating modern MLOps practices and production deployment patterns

Technical Co-Founder, Fiscal Mind **January 2025 - Present**

- **AI System Architecture:** Co-founded legal tech startup developing AI-powered tax advisory platform, combining domain expertise in Mexican fiscal law with quantitative modeling and multi-agent workflows
- **Advanced ML Pipeline Development:** Architected end-to-end machine learning pipeline using multi-agent workflows with knowledge graphs for intelligent legal document processing and tax advisory generation
- **Production System Optimization:** Built production FastAPI backend with PostgreSQL integration, implementing intelligent caching algorithms that reduced API costs by 60% while maintaining response quality and system reliability
- **Startup Leadership:** Managing full product lifecycle from initial concept through technical implementation and go-to-market strategy, balancing technical development with business needs

Consulting Intern, IQVIA **2020 - 2021**

- **Pharmaceutical Market Research:** Provided strategic support for pharmaceutical product launches, analyzing both quantitative and qualitative survey data from healthcare sector using statistical methods
- **Survey Design & Execution:** Designed and conducted structured interviews with healthcare professionals, managing end-to-end survey execution from questionnaire design through data collection and validation
- **Executive Reporting:** Synthesized complex market research findings into executive-ready PowerPoint presentations, translating data insights into clear, actionable recommendations for product launch strategies

Research & Academic Projects

Regime-Adaptive Multi-Strategy Portfolio System **Duke University - Personal Research**

- **Advanced Portfolio Construction:** Designed and backtested regime-aware multi-strategy portfolio combining (1) copula-based statistical arbitrage, (2) HMM-gated options overlay, and (3) filtered momentum; achieved 20.6% CAGR, 0.89 Sharpe ratio, 15.4% max drawdown over 2019-2026 with $49M estimated capacity on QuantConnect platform
- **Hidden Markov Model Implementation:** Developed two-state Hidden Markov Model with switching variance for SPY returns (μ_0=0.13%, σ_0=11.6% vs. μ_1=-0.16%, σ_1=35.6% annualized); optimized Expectation-Maximization estimation on rolling 3-year windows with 2-day persistence filters, achieving 97% COVID detection accuracy vs. 76% volatility baseline and 59% turbulence index

- **Copula-Based Pairs Trading:** Implemented Clayton copula (θ parameterized via Kendall's tau) for GDX/GLD pairs with rank-transformed returns, generating mispricing index $MI\_t = \partial C(u,v)/\partial v$; entered long/short spreads at MI<0.05/MI>0.95 thresholds with structural break exits when $\tau<0.2$ during crisis regimes
- **Options Overlay Strategy:** Engineered zero-cost collar overlay (long ATM put, short 1.02× OTM call, 1-40 DTE Friday expiries) activated exclusively in HMM high-variance states, hedging ⌊shares/100⌋ lots on long legs only; reduced theta decay while capping left-tail losses during 2020/2022 stress periods
- **Momentum System with Technical Filters:** Constructed triple-filter momentum system on QQQ (Price>SMA200 AND 120-day ROC>0 AND ADX_{14}>20) with dynamic capital rotation to SHV during filter failures; generated 32.3% CAGR in 2025 out-of-sample period with 0.80 beta and 3.2% max drawdown
- **Risk Management Framework:** Implemented comprehensive risk framework including rolling VaR_{95}/$CVaR_{99}$ monitoring on daily P&L, regime-conditional position sizing, and cointegration stability checks; validated across 15-year backtest (2010-2025) including 2011 correction, 2020 COVID crash, and 2022 bear market with consistent 0.68+ Sharpe ratio
- **Portfolio Optimization:** Optimized pair diversification through BAC/JPM financial sector allocation and regime-adaptive copula selection; conducted extensive ablation studies on ROC lookback (120 vs. 252 days), ADX thresholds, and leverage (1x to 2x) to maximize risk-adjusted returns

Advanced Portfolio Optimization Suite **Duke University - Personal Research**

- **Multi-Strategy Portfolio Framework:** Implemented comprehensive trading pipeline integrating Markowitz mean-variance optimization, Black-Litterman model with investor views, transaction cost modeling, and Hidden Markov Model regime detection for institutional-grade portfolio management
- **Risk Parity & Advanced Optimization:** Developed Equal Risk Contribution (ERC) optimization, Hierarchical Risk Parity (HRP) using hierarchical clustering, and robust Markowitz optimization under parameter uncertainty for diversified portfolio construction
- **VaR Optimization Framework:** Engineered Value-at-Risk optimization using CVXPY with Conditional VaR (CVaR) analysis and second-order cone programming constraints for tail risk management
- **Backtesting Infrastructure:** Built comprehensive backtesting engine analyzing 11 different portfolio strategies across multiple market regimes with detailed performance attribution, transaction cost analysis, and regime transition monitoring
- **Regime Detection System:** Implemented Hidden Markov Models for market regime identification (Bull/Bear/Crisis) with regime-adaptive allocation strategies and transition probability matrix analysis

Advanced Mortgage Credit Risk Modeling **Duke University - Fall 2025**

- **Large-Scale Survival Analysis:** Analyzed 1.1M Fannie Mae mortgage loans using survival analysis techniques to compare FICO and VantageScore 4.0 performance, revealing score update timing (30% gain) dominates score methodology choice (1.6% gain) in default prediction
- **Hybrid Statistical Learning:** Built innovative hybrid survival model combining Cox Proportional Hazards baseline with XGBoost residual modeling, achieving C-index of 0.998 with clear risk stratification (3.25x high/low default rate separation)
- **Credit Risk Reclassification Framework:** Designed three-way reclassification framework quantifying credit expansion opportunities (1,146 borrowers) and risk mitigation (5,389 loans flagged), demonstrating net conservative stance with improved access to credit
- **Advanced Feature Engineering:** Conducted rigorous feature engineering using PCA for dimensionality reduction, correlation analysis for multicollinearity removal, and domain-driven selection yielding 25 features from 117 raw variables
- **Technologies:** R (survival, xgboost, caret packages), Cox Proportional Hazards models, gradient boosting, statistical hypothesis testing

Geospatial Market Optimization Model **Andrea (Leading Wholesale Footwear Company) - Entropia AI**

- **Predictive Modeling & Geospatial Analysis:** Developed municipal-level sales prediction model using socioeconomic variables and distributor characteristics, combining API-based geolocation with demographic targeting to identify high-opportunity markets across Mexico
- **Location Intelligence Algorithm:** Built optimization framework to identify optimal zones for new distributor recruitment and retail locations, incorporating competitor density analysis (1km radius), target population demographics, and strategic proximity constraints
- **Market Expansion Strategy:** Generated actionable insights for national expansion strategy through spatial econometric modeling and geographic clustering analysis

SkillMiner - Full-Stack Data Engineering Platform **Duke University - Personal Project**

- **Automated ETL Pipelines:** Engineered automated ETL pipelines processing 30,000+ job postings to identify in-demand technical skills and curate educational resources from GitHub and YouTube APIs
- **Production Infrastructure:** Deployed production infrastructure on AWS with containerized Apache Airflow orchestrating weekly DAGs, RDS PostgreSQL database, and S3 storage with IAM role-based authentication
- **Content Scoring Algorithm:** Developed proprietary content scoring algorithm filtering educational resources using semantic analysis, improving content quality over basic keyword searches
- **RAG-Based Analysis:** Implemented Retrieval-Augmented Generation (RAG) based skill gap analysis matching resume PDFs against job descriptions to generate personalized learning recommendations
- **Semantic Search:** Built semantic search functionality using ChromaDB and OpenAI embeddings for intelligent resource discovery across 10,000+ curated learning materials
- **Technologies:** Python, AWS (EC2, RDS, S3), Apache Airflow, Docker, PostgreSQL, ChromaDB, FastAPI

Predictors of American AI Sentiment **Duke University - Fall 2024**

- **Survey Methodology & Causal Inference:** Implemented survey-weighted ordered logistic regression on nationally representative sample (N=5,400, Pew ATP Wave 152) to model AI sentiment as 5-level ordered outcome
- **Interaction Effects Analysis:** Identified and validated Race × Income interaction through Wald F-tests and visualization; addressed non-random missingness patterns using chi-square tests and complete case analysis with survey weights
- **Statistical Modeling:** Constructed predictive models showing gender (OR=0.64), ideology (Liberal OR=0.39), and age effects (50-64 OR=1.60) on AI optimism using logistic regression with complex survey design
- **Technologies:** R (svyolr, regTermTest packages), missing data analysis, interaction effects modeling, policy communication

Spatial Econometric Analysis of Poverty-Crime Relationships **Universidad Panamericana - Undergraduate Research**

- **Spatial Lag Model Development:** Developed Spatial Lag Model analyzing poverty-crime relationships across Mexico's 32 counties, incorporating spatial dependence through Moran's I statistics under guidance of Dr. Jose Miguel Torres

- **Spatial Econometrics:** Applied advanced spatial econometric techniques to account for geographic spillover effects and spatial autocorrelation in socioeconomic analysis

Competitions & Publications

Kalshi Prediction Markets Conference 2026 **Conference Submission (January 2026)**

"Empirical Validation of Epoch-Based Leverage Pricing in Prediction Markets"

- **Large-Scale Market Microstructure Analysis:** Conducted first comprehensive empirical test of Messari epoch framework across 2,129 prediction markets with 4.2M observations, validating theoretical pricing models in live market conditions
- **Monte Carlo Simulation Framework:** Designed and executed 1,000 Monte Carlo simulations to test strategy robustness across market conditions, parameter sensitivity, and execution scenarios
- **Quantitative Trading Strategy Performance:** Achieved 83.6% annualized returns with Sharpe ratio of 2.85 through systematic exploitation of mispricing in prediction market leverage products
- **Novel Research Contribution:** First empirical validation of epoch-based pricing theory in decentralized prediction markets, contributing to market microstructure literature

NCAA March Madness Kaggle Competition **12th Place (Top 1%) - Silver Medal**

- **Ensemble Model Development:** Developed sophisticated ensemble prediction model using XGBoost and Generalized Linear Mixed Effects models, ranking 12th out of 1,200 teams through advanced feature engineering, Bayesian optimization, and rigorous cross-validation
- **Hierarchical Modeling:** Implemented mixed effects models capturing team-level random effects and game-level fixed effects for tournament outcome prediction

"Analyzing NBA Timeouts" Research Publication **Medium Boost Platform**

- **Advanced Statistical Analysis:** Published research implementing multilevel logistic models to analyze timeout effectiveness using play-by-play data and statistical hypothesis testing
- **Editorial Recognition:** Selected by Medium's editors for broader distribution through Boost platform, demonstrating ability to communicate complex quantitative analysis to general audiences

NFL Big Data Bowl 2024 **Kaggle Competition - Participant**

- **Sports Analytics Metrics Development:** Developed methodology to evaluate tacklers using NFL tracking data, creating two novel metrics: Time to Tackle Opportunity Above Replacement (TTTOAR) and Tackles Above Replacement (TAR)
- **Player Tracking Analysis:** Analyzed high-frequency GPS tracking data to quantify defensive player performance and decision-making in real-time game situations

Player Profile - NBA Analytics Platform **Personal Project**

- **Full-Stack Analytics Application:** Created comprehensive NBA analytics platform (player-profile.onrender.com) built with Dash, featuring interactive dashboards for player evaluation and performance analysis
- **Advanced Metrics Development:** Incorporated custom Added Value metric relative to contract, play-type analysis, year-over-year performance comparisons, expected points from shot selection, and assist evaluation diagrams

Leadership & Activities

Research Director, Quantitative Finance Club, Duke University **August 2025 - Present**

- Direct quantitative research initiatives and publish weekly market analysis in club newsletter for 200+ members, focusing on algorithmic trading strategies, portfolio optimization techniques, and quantitative market insights
- Organize industry speaker events and coordinate research collaborations with faculty including Dr. David Ye (MIDS Quantitative Finance Director, former CRO) and Professor Massimo Cutuli (OCC Chief Financial Risk Officer, former Citadel CRO)

Cloud Club Co-Captain, Duke University **August 2025 - Present**

- Co-lead AWS Cloud Club initiatives focused on hands-on cloud computing projects, data analysis, and machine learning deployment on cloud infrastructure
- Organize workshops and technical sessions on cloud architecture, serverless computing, and MLOps best practices

Treasurer, Economics Student Society, Universidad Panamericana **2020 - 2021**

- Managed organizational budget and coordinated academic and professional development events for economics students
- Organized guest speaker series featuring industry professionals and academic researchers

Basketball Team Captain, Universidad del Valle de Mexico **2016 - 2017**

- Led team to National High School Championship title as Team Captain, demonstrating leadership and team coordination under pressure
- Maintained full-tuition athletic scholarship (awarded to only 5 players annually) through consistent academic performance and athletic excellence

Teaching Experience

Data Science Tutor, Coderhouse **2022 - 2023**

- Provided instruction and guidance to students in data manipulation, statistical analysis, and machine learning with Python
- Mentored students through hands-on projects covering pandas, scikit-learn, and data visualization best practices

Technical Skills

Programming Languages: Python (expert), R (advanced), SQL (advanced), MATLAB (intermediate)

Machine Learning & Statistics: XGBoost, neural networks, ensemble methods, time series forecasting (ARIMA, SARIMAX), survival analysis (Cox PH), Monte Carlo methods, Bayesian inference, ANOVA, hypothesis testing, clustering algorithms, dimensionality reduction (PCA), mixed effects models, spatial econometrics

Quantitative Finance: Portfolio optimization (Markowitz, Black-Litterman, risk parity, HRP), Hidden Markov Models, copula methods (Clayton, Gumbel, Frank), VaR/CVaR modeling, stochastic calculus, options pricing (Black-Scholes, Greeks), econometrics, cointegration, backtesting frameworks, regime detection, pairs trading

Python Libraries: scikit-learn, pandas, NumPy, TensorFlow, PyTorch, XGBoost, SpaCy, Keras, sktime, statsmodels, scipy, CVXPY, hmmlearn, copulas, QuantLib, yfinance, PyPortfolioOpt

R Libraries: survival, xgboost, caret, svyolr, regTermTest, tidyverse, ggplot2, spatial econometrics packages

Data Visualization: Matplotlib, Seaborn, Plotly, Dash, Streamlit, PowerBI, Tableau, ggplot2

Cloud & Databases: MongoDB, MS SQL, PostgreSQL, MySQL, Azure Machine Learning Studio, AWS (SageMaker, EC2, RDS, S3, ElasticBeanstalk, IAM), GCP Vertex AI

Development Tools & Frameworks: FastAPI, Docker, Prefect, Apache Airflow, Git, GitHub, Pydantic, ChromaDB, OpenAI API

Trading Platforms: QuantConnect, Bloomberg Terminal (familiar), options trading mechanics

Operating Systems: Windows, Linux

Languages: English (fluent), Spanish (native)

Certifications & Additional Training

Microsoft Certified: Azure Data Scientist Associate **2022**

Professional certification in Azure Machine Learning Studio, model deployment, and cloud-based ML workflows

Natural Language Processing with Python **2022**

Udemy - Comprehensive course covering NLP techniques, text processing, and language models

Data Politics **2021**

Universidad Panamericana - Course on ethical implications and policy considerations in data science