VEDAANT AGRAWAL

vedaantagrawal17@gmail.com | (413) 404-6243 | linkedin.com/in/vedaant-agrawal | github.com/VedaantAgrawal
New York, NY

SUMMARY

Data scientist with production experience across the full DS pipeline from data wrangling and feature engineering to model deployment and stakeholder reporting. At BlackBox, cut document retrieval time 98% across 50,000+ docs and reduced hiring screening time 60% across 500+ applications using statistical scoring and NLP. At PwC, hit 92% forecast accuracy on $20M/month petrochemical trades using gradient-boosted regression with scikit-learn. Strong in statistical modeling, time-series forecasting, A/B testing, and ETL automation. UMass CS, 3.82 GPA, 6x Dean's List; Azure DP-100 + AWS AI Practitioner certified.

TECHNICAL SKILLS

- **Statistical Modeling & Analytics**: Machine Learning, Statistical Modeling, Predictive Modeling, Time-Series Forecasting, A/B Testing, Hypothesis Testing, Causal Inference, Feature Engineering, Regression, Classification, XGBoost, Random Forest, Deep Learning (PyTorch)
- **Languages & Frameworks**: Python, SQL, Java, JavaScript, TypeScript, React, Node.js, Flask, FastAPI, pandas, NumPy, scikit-learn, scipy, statsmodels, NLP (TF-IDF, K-means, t-SNE)
- **Cloud, Data & Analytics**: AWS (SageMaker, Lambda, ECS), Azure, Docker, Kubernetes, PostgreSQL, Snowflake, Spark, ETL/ELT Pipelines, CI/CD, Git, Tableau
- **Visualisation**: Matplotlib, Plotly, Seaborn
- **Certifications**: Microsoft Azure DP-100 (Data Scientist), AWS AI Practitioner

EXPERIENCE

BlackBox | *Data Science & AI Engineer Intern* | Jun 2025 - Sept 2025

- Reduced document retrieval time by 98% (30 min to <1 min) across 50,000+ documents by engineering a statistical ranking and keyword extraction pipeline with ML-based relevance scoring, enabling on-demand trend analysis for cross-functional teams.
- Decreased hiring screening time by 60% across ~500 applications per cycle by developing an NLP model using TF-IDF vectorization, feature engineering, and weighted statistical analysis to automate fitment scoring.
- Applied statistical clustering to workforce query patterns (2,000+ employees) to identify high-frequency pain points, reducing resolution time 40% - delivered as automated SNOW/SAP workflow triggers adopted org-wide

PricewaterhouseCoopers (PwC) | *Data Analytics & Cloud Engineer Intern* | Jul 2024 - Aug 2024

- Achieved 92% forecast accuracy on petrochemical pricing (Paraxylene) - reducing forecast error from 45% to 28% - by developing gradient-boosted regression (XGBoost) and random forest models with statistical cross-validation using pandas and scikit-learn.
- Accelerated data preparation from 2 hours to 30 minutes by building an automated ETL pipeline with Python and AWS Lambda, enabling daily data visualization and model refresh for ~8 analysts.
- Scaled real-time inference to 10,000+ records per day by containerizing models with Docker and deploying on AWS SageMaker, supporting commodity trading decisions on ~$20M in monthly spend.

PROJECTS

EquiSight | *Stock Analysis and Forecasting Tool* | Oct 2024 - Apr 2025

- Ran an A/B test across 50+ equities comparing ARIMA vs. gradient-boosted forecasting — achieved 18% RMSE reduction at $p < 0.05$; validated assumptions via hypothesis testing (ADF, Ljung-Box) across 5+ years of historical price data.
- Engineered a Flask/FastAPI backend with automated PostgreSQL pipelines and a React/Plotly dashboard, cutting API response time by 57% (2.3s to 1s) while maintaining 99.9% uptime across daily real-time inference.

COVID-19 Research Literature Clustering | *Lumiere Education Research* | Aug 2024 - Oct 2024

- Built an NLP pipeline (TF-IDF, t-SNE, K-means) to cluster 32,000+ research papers from the CORD-19 dataset, preserving 95% variance and reducing multi-thousand-paper corpora to navigable topic groups via LDA topic modeling.

EDUCATION

University of Massachusetts Amherst | *B.Sc. Computer Science, Minor in Business* | May 2026

- GPA: 3.82 | 6x Dean's List, Chancellor's Award | Coursework: Data Science, AI/ML, Data Management, Advanced DSA, Linear Algebra