Gregg Rich
Fairfax, VA | 240-756-1025 | linkedin.com/in/greggjrich

Experienced data scientist and technology professional with expertise in machine learning, data analytics, database management, and cybersecurity. Strong background in statistical modeling, machine learning algorithms, data visualization, natural language processing, and feature engineering. Proven ability to lead cross-functional teams and collaborate on data-driven decision-making. Background in education and research enhances ability to communicate complex technical concepts effectively.

Technology Skills

- **Data Science & Machine Learning:** AI/ML algorithm implementation and fine tuning, model interpretability, classification models, predictive analytics, time series forecasting, feature engineering, natural language processing (NLP), sentiment analysis
- **Programming & Tools:** SQL, Python (pandas, numpy, scikit-learn, matplotlib, seaborn), R (tidyverse, ggplot2, plotly, Rshiny), VS Code, Anaconda
- **Data Management & Infrastructure:** AWS, Azure, Google Cloud, Databricks, Jupyter notebooks, Google Colab, relational database design and operation, ETL pipelines, and data wrangling
- **Data Visualization & Reporting:** Tableau, Power BI, MSPowerPlatform, RShiny, ggplot2, Matplotlib
- **Project Management & Collaboration:** Agile methodology, Git repositories, YouTrack, metadata standards, technical documentation

Professional Experience

Government Data Engineering and Analysis

Data Scientist, U.S. Digital Corps Fellow
Cybersecurity Infrastructure Security Agency (CISA) | August 2024 – Present

- Led project efforts to acquire a data storage and analytics platform as part of a division-wide data management solution for the Cybersecurity Division (CSD) to ensure accurate reliable data reflecting the health and effectiveness of the CSD service portfolio to protect crucial Federal government network infrastructure.
- Collaborated with cross-functional teams, including the Office of Chief Information Officer (OCIO), data scientists, policy analysts, and cybersecurity experts, regarding technical specifications and architecture requirements to support data modernization initiatives for enhanced decision-making.
- Analyzed available data sources and collection processes to streamline and automate data ingestion.
- Mapped data pipeline from sources to relational database fields to produce an entity relationship diagram (ERD).
- Participated in *DHSChat* pilot to use AI/LLM to analyze text and unstructured data within a secure environment.
- Led Digital Corps training on implementing GSA GenAI policy and overview of use cases.

Research Assistant, KRA Corporation, Silver Spring, MD | October 2001 – May 2004

- Provided quantitative research support for government contract projects for HUD and HHS, with a focus on economic development programs.
- Developed data processing protocols for accessing HUD's IDIS database through SQL queries and processing raw data into reports, including using linear regression to infer missing values as well as predict the loan performance based on a set of features identified in the database, as part of a larger study on CDBG and Section 108 funds for economic development.
- Evaluated U.S. Department of Labor grant proposals from both small and large community-based and faith-based organizations for projects involving community development, economic development, and job training and other support services for immigrants, older persons, or disabled persons.
- Co-authored reports on state disadvantaged worker training programs and an experimental grant to fund worker training programs and an evaluation of a high school-based youth vocational training program.
- Awarded Project Success Award in recognition of outstanding effort and results on national survey of child protective services agencies.

Math, Data, STEM Education

Oakton High School, Vienna, VA | August 2022 – June 2024

- Taught AP Statistics and Geometry, integrating data analysis and visualization techniques.
- Designed project-based learning modules incorporating statistical software.
- Selected as AP reader to evaluate free-response sections of the AP Statistics exam.

Fusion Academy, Tysons Corner, VA | November 2021 – July 2022

- Provided one-on-one instruction in mathematics, science, and Spanish, leveraging personalized learning strategies.
- Wrote daily reports detailing student progress and areas for improvement.

St. Anne's School, Asunción, Paraguay | August 2009 – December 2021

- Taught IB Math HL, Business Management HL, Physics, and Economics.
- Coached colleagues in technology integration for classroom teaching and administrative tasks.
- Led technology integration initiatives to enhance remote learning capabilities.
- Chaired Cognia accreditation review committee and coached math competition teams.

International Development

Peace Corps Volunteer, Paraguay | May 2004 – August 2006

- Designed and conducted adult education programs in business development and IT literacy.
- Managed economic training programs for cooperative members and students.

Education

Master of Science: **Data Analytics Engineering** (GPA: 3.97) Jan. 2022 - May 2024

George Mason University, Volgenau School of Engineering, Fairfax, Virginia

Relevant Coursework: Analytics: Big Data to Information, Principles of Data Management and Mining, Analytics & Decision Analysis, Applied Statistics and Visualization for Analytics, Applied Predictive Analytics, Decision and Risk Analysis, Intro. to Natural Language Processing, Knowledge Mining from Big Data, Applied Machine Learning

Capstone Project - "*Wing Beats to Data Feats: Forecasting and Visualizing Migratory Trends*" (April 2024): Worked with the US Fish & Wildlife Service - Div. of Migratory Bird Management to engineer an automated data pipeline and workflow process to analyze migratory bird survey data, uncovering temporal patterns and enabling predictive forecasting. Used Tableau to seamlessly integrate project datasets for an interactive visualization experience, allowing stakeholders to efficiently view and manage bird conservation efforts. Wrote clear product and project documentation to include metadata, error handling, and pipeline automation.

Other projects:

- Predicting Stock Market Price Movements Using an LSTM Time Series Forecasting Model (Dec. 2023): Using ARIMA and NeuralProphet as baseline models, project developed an LSTM model with a sequential neural network set up within TensorFlow's Keras API, using multivariate time series data to predict short-term stock price movements.
- Sentiment Analysis of Movie Reviews: Predicting Positive or Negative Sentiments (June 2023): Implementation of Natural Language Processing algorithms to classify movie reviews by sentiment using Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) to extract features from textual data and multiple machine learning algorithms, including XGBoost, Naive Bayes, Logistic Regression, Random Forest, SVM, Random Forest, K-Nearest Neighbors, and a pre-trained BERT transformer model, to perform the classification.
- Predicting Passenger Dimensional Transport (Apr. 2023): Analysis of a dataset from a Kaggle competition involving a fictitious case of a "Spaceship Titanic" to predict which passengers were removed from the ship. The project used RStudio and packages from caret and the tidyverse, as well as corrplot, randomForest, pROC, and mice, to implement Logistic Regression and Random Forest for classification, with k-Nearest Neighbors, Neural Net, MDA, and Naive Bayes as predictive algorithms.

- Decision making algorithm for university selection (Feb. 2023): Implemented decision making algorithms in the Logical Decisions v8.0 software package to produce an Affinity Diagram, generate Single Dimension Utility Functions, and elicit weights by rank sum and trade-off methods through interviews with students in the process of applying for university admissions. Focus on project management and documentation.
- Analysis of the IPEDS to Predict Graduation Rates (Dec. 2022): Analysis of university data using lasso regression and random forest regression machine learning algorithms to build a predictive model with a focus on visualizations in RStudio, with choropleth mapping and other plots to show relationships between variables.
- Inflation and its Trends in the US (May 2022): Time series analysis of factors affecting inflation using Random Forest, ARIMA, and NeuralProphet algorithms to train an explanatory/predictive model of inflation by sector, drawing from BLS and IMF published data. Focus on data standardization, and ETL for time series analysis. Project followed the Agile development process with weekly sprints.
- Global Economy and COVID-19 (Feb. 2022): Multivariable regression study on international trade, imports/exports, COVID-19 cases and death rates. Focus on data cleaning/wrangling, joining data sets, applying python libraries: pandas, numpy, sklearn, matplotlib, scipy, and seaborn.

Bachelor of Arts: International Relations-Political Economy May 2001

Honors: Dean's List, Member of Phi Sigma Pi National Honor Fraternity

College of William & Mary, Williamsburg, Virginia

Relevant Coursework: Econometrics, Principles of Statistics, Symbolic Logic

Other Courses/certifications:

* Statistics with Python specialization (Oct. 2021 – U. Michigan)
* Intro. to Machine Learning (Sept. 2021 – Duke University)
* Matrix Algebra for Engineers (May 2021 – The Hong Kong University of Science and Technology)
* Python 3 programming specialization (Apr. 2021 – U. Michigan)
* Python for Data Science, AI & Development (Mar. 2021 – IBM)
* Mathematics for Machine Learning specialization (Mar. 2021 – Imperial College London)
* Data Science Methodology (Feb. 2021 – IBM)
* Tools for Data Science (Jan. 2021 – IBM)
* Vector Calculus for Engineers (Jan. 2021 – The Hong Kong University of Science and Technology)
* The Data Scientist's Toolbox (Dec. 2020 – Johns Hopkins University)
* Managing Machine Learning Projects with Google Cloud (Nov. 2020 – Google Cloud)
* VDOE Collegiate Teaching License (Math and Physics Endorsement)
* Google Level 1 and Level 2 Educator certification
* International Baccalaureate training for math exploration: project design and supervision

---

Languages

- Spanish (Fluent)
- Guaraní (Upper-Intermediate)
