WILLIAM GEMBA DATA SCIENTIST

Miami, FL | Cell: +1 917.588.1702 | Email: gemba.william@gmail.com | Social Media: linkedin.com/in/william-gemba/
Website: https://wgemba.github.io/webportfolio/

Data Scientist with Extensive Consulting Experience and Strong Foundation in Machine Learning and Data Mining Techniques

Results-driven Data Scientist with a robust background in client services and consulting, demonstrating a proven track record in driving data pipeline remediation efforts, conducting user acceptance testing, and effectively managing stakeholders in high-profile projects for large multinational companies. Proficient in utilizing machine learning methods to derive actionable insights. Demonstrates adaptability, self-motivation, and a keen ability to swiftly learn and master new tools, methodologies, and concepts.

Skills: Data Pipeline Remediation ▪ User Acceptance Testing (UAT) ▪ Extract, Transform, Load (ETL) ▪ Stakeholder Relationship Management ▪ Python ▪ SQL ▪ Data Mining ▪ Predictive Analysis ▪ Machine Learning ▪ Jupyter Notebook ▪ Excel ▪ Project Management ▪ Data Analytics ▪ Data Wrangling ▪ Data Visualization ▪ Feature Engineering ▪ Business Development ▪ Pandas ▪ NumPy ▪ Scikit-Learn ▪ Matplotlib ▪ Seaborn ▪ TensorFlow 2.0 ▪ Relational Databases ▪ Client Consulting ▪ Neural Networks ▪ Natural Language Processing (NLP)

CAREER OVERVIEW

Ernst & Young (EY) — **New York, NY**
Data Analyst, Consultant — **2021 - 2023**

- Spearheaded data pipeline remediation efforts for major U.S. multinational financial services organizations, ensuring compliance with the latest federal liquidity risk management standards.
- Conducted comprehensive user acceptance testing using Oracle SQL; querying, identifying and rectifying millions of dollars in imbalances stemming from incorrect architectural mapping logic.
- Orchestrated collaborative efforts with internal and external stakeholders to address liquidity risk management challenges, successfully identifying and resolving data sourcing issues. Instituted new quality controls to enhance data integrity.
- Streamlined audit and remediation processes by identifying and resolving ETL logic problems across multiple lines of business, covering hundreds of attributes.
- Applied advanced Microsoft Excel functions, including Pivot Tables and VLOOKUP, to streamline process improvement documentation, enhancing efficiency and accuracy in project workflows.

PROJECTS

Project Portfolio Link: https://github.com/wgemba

Wine Recommendation System Using Unsupervised Content-Based Filtering

- Performed exploratory data analysis of and created visualizations for a dataset of 130,000+ wine taster reviews.
- Utilized natural language processing (NLP) to parse through reviews and engineer a taste profile for each wine reviewed.
- Implemented a complex neural network architecture using TensorFlow 2.0 to predict the rating each user gave to each wine, achieving a final training loss of 0.003, and validation and test loss of 0.007.
- *Tools used*: *Python, Jupyter Notebook, Matplotlib, Seaborn, Scikit-Learn, Natural Language Toolkit (NLTK), TensorFlow 2.0.*

U.S. Small Business Association (SBA) Loan Default Classification Using an Optimized Artificial Neural Network Model

- Analyzed a dataset from the SBA consisting of 900,000 instances and 27 features, performing heavy feature engineering.
- Created a function to test 64 permutations of hidden layer width, selecting the model with the lowest training loss and validation loss. Top performing model was able to achieve an accuracy of 92%.
- *Tools used*: *Python, Jupyter Notebook, Matplotlib, Seaborn, Plotly, Scikit-Learn, TensorFlow 2.0.*

EDUCATION

Master of Science in Data Science, Fordham University — **New York, NY**
Additional Accreditation: **Advanced Certificate in Financial Econometrics and Data Analysis**
Graduated 2021 (GPA: 3.9)

Bachelor of Science in Global Finance & Business Economics, Fordham University — **New York, NY**
Graduated 2020 (GPA: 3.7)

CERTIFICATIONS

- Machine Learning Specialization | Stanford University / DeepLearning.AI | Issued: *December 2023*