

John Walsh

618.554.1495 | Houston, TX | johnrwalsh34@gmail.com

Technical Skills

DATA SCIENCE & MACHINE LEARNING: Time-series Analysis | Association Rules Mining | NLP | TensorFlow | Deep Learning | Random Forest | K-means | Gradient Descent | Neural Networks | Supervised and Unsupervised Learning | ApacheSpark | SQL | Text Classification | Hyperparameter Tuning | Feature Selection and Extraction | PCA | XML | HTML

PROGRAMMING LANGUAGES: Python | Bash Shell | R | Scala

INDUSTRY EXPERTISE: Mathematical Modeling | Python Programming | Probability Theory in Machine Learning | Probabilistic Models | Data Cleaning | Data Processing | Data Visualization | Statistical Analysis | Statistical Testing | Regex | Text Processing | Association Rules | Data Mining | Markdown | LaTeX | Algorithm Development | Sports Analytics | Operations Management | Presentations | Communication at all levels | ML Operations | Microsoft Excel

Education

Data Scientist Certification | The Data Incubator Fellowship Program | Nov 2023

BA in Interdisciplinary Mathematical Sciences (Honors) and Minor Concentration in Biochemistry | Florida Atlantic University, Recipient of Outstanding Thesis Award | May 2023

Professional Experience

24/7 Software, Business Analyst Intern

May 2022 – May 2023

24/7 Software develops Venue Operations Management software to improve operations, and reduce risk for Sports & Entertainment venues

SPATIOTEMPORAL DETERMINANTS OF FOOTBALL STADIUM INCIDENTS

- Initiated analysis of raw Stadium Incident Data (>1G) reported at one professional football stadium with the goal to quantitatively analyze stadium incident response times and predict incidents' occurrence and locations. Regularly presented work to senior leadership and extended initial analysis and modeling to include **20 professional sports stadiums**.
- Cleaned and processed the data, built regression models, and performed classification using Apriori Association Rules Mining (ARM) to analyze the spatiotemporal determinants of football stadium incidents.
- Constructed visualizations displaying the likelihood of incidents occurring in various stadium locations during various times.
- Results elucidated operational inefficiencies particularly related to security incidents and ADA accessibility requests.
https://github.com/jrw34/ThesisJW_PDF/blob/main/JW_Thesis_pdf2.pdf
- For example, it was determined that 25.6% of accessibility requests occurred when the stadium gates opened. Also found that 9% of fan code of conduct violations occurred in a specific section of the stadium during the 2nd quarter.
- Verified that data driven insights could be productionalized and introduced a model capable of productionalization. Currently, the company is developing a product that would improve venue operations based on time/location risk prediction.
(<https://www.247software.com/proactive-operations>)

Mathematics Tutor

Jan 2021 – May 2023

- Tutored all undergraduate mathematics courses, for example: Introductory Statistics, Matrix Theory, and Differential Equations.

Data Science Projects

Ingredient Identifier (Capstone Project for The Data Incubator)

Oct 2023 – Nov 2023

- Built a deployable application with Streamlit in Python capable of identifying branded food products registered with the FDA that match user-specified criteria such as ingredients the user either wants included or excluded from their diet. The dataset is cleaned and stored in a personally established Postgresql database. The application queried the database upon user request and produced a custom hierarchically directed acyclic graph to display query results, allowing users to find food items they want to incorporate into their diet.

Senior Thesis, an Outstanding Thesis Award, Florida Atlantic University & 24/7 Software

May 2022 – May 2023

- The Spatiotemporal Football Incidents model pro-actively enables the venue to determine optimal staff allocation.

Chemistry Laboratory Volunteer, Florida Atlantic University

May 2021 – Aug 2021

- Performed least squares optimization (analytically) using SVD in a Linux environment to project data points onto a best-fit plane in 3-D. As a result, enabled better assessment of molecular interactions in the dynamic model.