HARRISON GOOSEN

contact@harrisongoosen.com | 4059219417 | harrisongoosen.com

PROFESSIONAL EXPERIENCE

Oklahoma Department of Education
Data Analyst   Feb 2020 - Present

- Create data extraction tools for Oklahoma's Cost Accounting System using OCR, reducing reporting time by 50%.
- Build analysis and visualization tools for the State Systematic Improvement Plan, outperforming goals by 10%.
- Coordinate with Public Consulting Group, supporting the real-time data ingestion of the 700,000 student records.
- Develop data protocols, including assisting with analysis and creating custom Tableau dashboards for 550 School Districts.

PROJECTS

Malicious Binary Detection   2021

- Reviewed and examined 150 related works.
- Constructed a dataset consisting of 25,000 unlabeled malware samples from the VirusShare database.
- Extracted nine features based on the header and PE file structure, containing the most significant differences.
- The final model (Random Forest) identified malicious programs with 95% accuracy, with a considerable reduction in training time.

Finding Donors for Charities   2020

- Used supervised learning algorithms such as Decision Trees, SVMs, and Boosting.
- Optimized the best candidate model by hyperparameter tuning using Grid Search and Cross-Validation.
- The final model predicted an individual's probability of donating with 87% accuracy.

EDUCATION

University of Pennsylvania
Master of Computer and Information Technology (MCIT), Online   2023

Oklahoma State University
Bachelor of Science in Information Assurance and Data Science   2020

PROFESSIONAL DEVELOPMENT

Mathematics for Machine Learning Specialization
Imperial College London, Coursera - Eigenvectors, PCA, Multivariate Calculus   2021

Computer Science Fundamentals Specialization
University of Illinois Urbana-Champaign, Coursera - Data Structures, and Algorithms in C++   2021

Statistics with R Specialization
Duke University, Coursera - Bayesian Statistics, Linear Regression, Statistical Inference   2020

TECHNICAL SKILLS

Languages (proficient):	Python, SQL
Languages (intermediate):	Java, R, C++, JavaScript
Machine Learning:	scikit-learn, XGBoost, PyTorch, Tensorflow
Database:	SQL Server, MySQL, PostgreSQL, MongoDB
Data Visualization:	Tableau, Power BI
Cybersecurity:	Snort, Suricata, Nessus, Metasploit