Harshith Doppalapudi

✉ harshithdoppalapudi@gmail.com | 📱812-974-2104 | 💼 LinkedIn | GitHub

EDUCATION

Indiana University, Bloomington   **August 2023 – Present**
Master of Science in Data Science

- **Relevant Coursework:** Applying ML Techniques in NLP, Predictive Analysis and Data Mining, Data Visualization, Cloud-Based Analytics, Advanced Database Concepts, Statistics, Applied Algorithms, Exploratory Data Analysis, Applied Machine Learning, Usable AI.

EXPERIENCE

Research Data Analyst | Indiana University Bloomington   **January 2025 – Present**

- Developed a GPT-4o mini chatbot with poisoned data and circulated it via Qualtrics to 1,000 users, collecting interaction data to study phishing behaviour.
- Analysed interaction data using Python, identifying a 42% engagement rate with deceptive content and high-risk behavioural groups.
- Segmented user risk profiles with clustering techniques, enabling strategies that reduced phishing susceptibility by 25% in follow-up tests.

Geospatial AI Researcher | Kelley School of Business   **January 2025 - Present**

- Engineered CNN, U-Net, and ResNet models in PyTorch for unauthorized land use detection, achieving 95% segmentation accuracy using hyperspectral, multispectral, SAR, and DSM imagery on High-Performance Computing (BigRed 200).
- Tackled multi-modal data integration by developing a fusion framework combining four satellite image types, boosting classification accuracy by 18% and improving detection of critical changes in high-risk areas.
- Refined UMAP and PCA for dimensionality reduction, cutting computational costs by 35% while enhancing anomaly detection and preserving key spatial patterns for improved geospatial intelligence analysis.

Machine Learning Research Assistant | Kelly School of Business   **September 2024 - Present**

- Analysed high-dimensional network intrusion packet datasets, applying UMAP dimensionality reduction and feature engineering to enhance data interpretability and achieve a 20% improvement in anomaly detection accuracy.
- Built and evaluated ML (Decision Tree, Random Forest, XGBoost, SVM) and DL (DNN, auto-encoder) models, achieving 99.8% accuracy and enhanced robustness against perturbations.
- Designed reinforcement learning agents to perturb classifier inputs and restore perturbed data to original values, improving model resilience and providing actionable insights into sophisticated attack patterns.
- Collaborated with cross-functional teams under Ankit to translate complex cybersecurity findings into actionable business insights, driving a 25% improvement in decision-making accuracy.

Data Science Intern | Amrita Vishwa Vidyapeetham   **January 2023 – May 2023**

- Processed large image datasets of student behaviour using Python and advanced analytical techniques, reducing data processing time by 20% for efficient analysis.
- Created interactive dashboards with Tableau and Power BI to visualize behavioural insights from image data, improving stakeholder decision-making by 15%.
- Developed and fine-tuned CNN and Mobile Net models with TensorFlow, achieving 96% and 98% accuracy in identifying behavioural patterns from images.

PROJECTS

Analyse Political Messaging Trends with Power BI | *Python, NLP, Data Cleaning, Visualization* (***GitHub***)

- Scraped and analysed 15K+ political tweets using Beautiful Soup and Tweepy, uncovering sentiment trends and linguistic shifts pre/post-election with NLP. Built a Tableau dashboard to visualize key insights.
- Engineered features like TF-IDF and cosine similarity to extract unique political keywords, improving thematic differentiation by 30%. Provided actionable insights on voter behaviour and political influence.

Transform Sales Performance with Power BI | *Power BI, SQL, ETL, Data Visualization* (***GitHub***)

- Analyzed 48 months of sales data across 50+ cities, using exploratory data analysis (EDA) to uncover seasonal trends and regional patterns, which informed strategic decisions that contributed to revenue growth.
- Processed 10K sales records using advanced SQL techniques to establish reliable databases and created interactive Power BI dashboards with 5 key performance indicators (KPIs) to visualize sales trends and optimize decision-making.

TECHNICAL SKILLS

Programming & Databases:	Python, R, SQL (MySQL), Excel, Java, C, C++.
Data Analysis & Visualization:	Power BI, Tableau, Excel, SQL, Matplotlib, Seaborn.
Machine Learning & Statistical Modelling:	Scikit-learn, TensorFlow, PyTorch, Keras, NLP (NLTK, SpaCy), Anomaly Detection, Model Optimization.
Cloud & Platforms:	AWS, Google Cloud, Linux, Windows, Mac.
Domains & Expertise:	Data Analytics, Business Intelligence, Computer Vision, Natural Language Processing (NLP), Cybersecurity, Predictive Analytics.