

Harsh Anand

4849958218 | vyf8rj@virginia.edu | www.linkedin.com/in/harshanand007/ | 1904 JPA, Charlottesville, VA | www.harsh-anand.live

Data Scientist and Researcher with 8+ years leading and delivering advanced analytical solutions for global clients.

EDUCATION

University of Virginia, USA

PhD in Systems Engineering (Data Science and Operations Research concentration)

May 2021 - Aug 2024

GPA: **4.0/4.0**

Outstanding Academics Award, Outstanding Teaching Award, 10+ Q1 Research papers published

Pennsylvania State University, USA

MS in Data Science and Analytics

Aug 2019 - May 2021

GPA: **4.0/4.0**

Graduated Valedictorian with Outstanding Student Award

Manipal University, India

BS in Information Technology (Computer Science concentration)

Jun 2011 - May 2015

GPA: **3.67/4.0**

PROFESSIONAL EXPERIENCE

Doctoral Researcher and Project Leader - University of Virginia, Virginia, USA

May 2021 - Present

- Proposed a MOVSTAT-GAT technique that combines moving temporal statistics-based feature extraction and graph attention network-based multi-variate time series prediction to detect FDIA from the prediction error in an unsupervised manner.
- Led design and implementation of spatio-temporal predictive models (GCN, LSTM, CNN) for optimal data transmission and storage in large-scale wireless sensor networks under uncertainties, resulting in over 75% data reduction with 99% accuracy.
- Utilized ML and causal inference methods to estimate evacuation order effectiveness using large-scale mobility patterns across multiple hurricanes. Collaborated with emergency management stakeholders to enhance disaster response strategies.
- Devised scalable learning method to assess energy utilization patterns for policy development toward energy equity.
- Developed an innovative multi-fidelity deep Q-learning algorithm with an adaptive fidelity selection, achieving a 60% reduction in high-fidelity simulation needs and enhancing decision accuracy in complex environments.

Graduate Researcher and Team Leader - Pennsylvania State University, Pennsylvania, USA

Nov 2019 - May 2021

- Directed 2-students to develop an optimization model for multimodal freight transportation (critical infrastructure) resilience under natural disaster disruptions, now deployed by Penn DOT for economic impact assessments.
- Developed a systematic framework to model and analyze energy infrastructure resilience considering economic impact.

Data Science Intern - Swiss Re, New York, USA

Jun 2020 - Aug 2020

- Instituted company-wide standard claims data dictionary, subsequently coordinating with multiple stakeholders, resulting in a 25% reduction in data discrepancies and a 15% improvement in data processing efficiency.
- Designed and automated reinsurance use-case frameworks to explore and assess statistics, trends, and projections, over 2.2B transactions using machine learning, PySpark, and Foundry, yielding 30% faster analysis and 20% more accurate predictions.

Senior Data Scientist - A.T. Kearney, Mumbai, India

Jan 2017 - Aug 2019

- Automated level 5 categorization of transaction line items for spend analytics using machine learning and a rules-based categorization approach, building on over 3M transactions across 12 countries and \$2B spend, achieving 98% accuracy.
- Engineered a hybrid decision tree and support vector machine algorithm for K12 school evaluation and market research.
- Devised and executed a customer retention and migration strategy for three years by building a track & trace model using linear and non-linear optimization over 200K customers and 70 products, improving retention rate by 30% over 3 years.
- Implemented an automated leased lines inventory with a 12-month rolling forecast and reported current value leakage, including quantification and root cause analysis, resulting in estimated savings of 10% of overall costs per annum.
- Designed enterprise data lake for inventory management on Hadoop by analyzing and evaluating multiple data sources and business data workflows, and presented commercial effective tableau dashboards to stakeholders, boosting efficiency by 20%.

Machine Learning Engineer - TATA AI Research Lab, Kochi, India

Aug 2015 - Nov 2016

- Instituted time-series forecasting models for analyzing viewing patterns and anomaly detection in total viewership duration.
- Released strategic intelligence dashboard for PE clients to explore potential reach, impressions, and conversation size, mining critical real-time trending tweets using distributed platforms and topic modeling, enhancing engagement metrics by 15%.
- Improved product matching accuracy to 94% by devising a graph-based solution leveraging graph traversal algorithms.

GENERATIVE AI PROJECT EXPERIENCE

- Fine-tuned open-source language models, including Mistral and LLaMA, for content generation related to a non-profit financial institution using PERT (LoRA and qLoRA), improving task-specific performance by 25%.
- Developed and implemented a RAG evaluation framework using RAGA to reduce hallucinations in GPT-3.5, significantly enhancing the model's response accuracy and reliability for complex queries.
- Conducted multi-task fine-tuning using Mixtral with LoRA to perform both intent recognition and retrieval-augmented generation (RAG) for decision support applications in the education sector, enhancing model performance and task accuracy.
- Designed a RAG-based QA system for a biomedical start-up to assist healthcare professionals by combining neural retrieval with BERT and GPT-3, improving response accuracy by 30% and reducing incorrect answers by 15%.

SKILLS

Data Science:	Machine Learning, Supervised and Unsupervised Learning, Decision Trees, SVM, Gaussian Processes, Clustering, PCA, Deep Learning, CNN, RNN, Generative Adversarial Networks, Transformers, Graph Neural Networks, Large Language Models, Statistical Analysis, Multivariate Regression and Classification, Time Series Analysis, Sensitivity Analysis, Probabilistic Modeling, Optimization & Modeling, Multi-Fidelity Modeling, Physics-Informed Modeling, Numerical Optimization, Linear Programming, Reinforcement Learning, Multi-Criteria Decision Making, Risk and Resilience Assessment, Discrete Event Simulation, Agent-based Modeling
Generative AI:	GPT, BERT, Text Generation, Conversational AI, Content Creation, VAEs, Prompt Engineering, Seq2Seq Models, Zero-shot Learning, Few-shot Learning, Fine-tuning, RAG, Neural Retrieval, LoRA, qLoRa, RAGA
Programming:	Python, R, SQL, Java
Development:	Spark (PySpark, Spark SQL), Hadoop, Snowflake, Redshift, Neo4j, CI/CD Jenkins
Project Management:	Strategy, Project Planning, Agile Development, Leadership, Problem Solving
Visualization/Cloud:	Power BI, Tableau, Excel, ArcGIS, Azure, Minitab, AWS, IBM Bluemix, Palantir Foundry

ADDITIONAL INFORMATION

Leadership: President – Graduate Engineering Student Council, VP of Projects – Graduate Consulting Club
Case Competitions: Winner (2020, 2021, 2022) - *INFORMS Operations Research Case*, Finalist - *Duke-UNC-TMC 2022*