

# Venkatesh Shanmugam

Virginia US | +1 (703) 216-2540 | [svenkatesh.js@gmail.com](mailto:svenkatesh.js@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

## SUMMARY

Data Scientist with 4+ years driving measurable business impact through advanced statistical modeling, A/B testing, and large-scale data analysis. Proven track record in end-to-end data science lifecycle experiment design, feature engineering, model deployment, and performance monitoring—achieving 78% ROC-AUC and delivering \$6K annual cost savings. Skilled in Python, SQL, Spark, LLMs, and MLOps, adept at translating complex data into insights that inform FAANG-level product and operational strategies.

## TECHNICAL SKILLS

- **Programming & Analytics:** Python(Pandas, NumPy), SQL(Hive, Presto, Spark), R
- **Data Science & ML:** Statistical Modeling, Hypothesis Testing, A/B Testing, Experimentation Design, Causal Inference, Predictive Modeling, Time Series Analysis, Deep Learning, NLP, Computer Vision, Graph-RAG, LoRA/PEFT
- **Cloud & MLOps:** AWS (S3, Glue, Athena, SageMaker), GCP (Vertex AI, Cloud Build), Docker, Kubernetes, MLflow, CI/CD
- **Big Data & Databases:** Apache Spark, Kafka, Airflow, PostgreSQL, MongoDB, Neo4j, Elasticsearch
- **Visualization & BI:** Tableau, PowerBI, Matplotlib, Seaborn, Data Storytelling, Product Analytics

## WORK EXPERIENCE

### ScriptChain Health

May 2024 - Present

Data Scientist

Washington DC

- Architected and deployed a predictive model for 30-day hospital readmission risk, achieving 78% ROC-AUC through advanced statistical modeling and feature engineering, enabling data-driven clinical interventions and reducing readmission-related costs by \$110K per patient annually
- Designed and executed MLflow A/B testing framework, improving precision-recall AUC by 6 percentage points and reducing false positives by 8%, driving evidence-based model iteration and enhancing healthcare outcomes.
- Developed agentic AI Graph-RAG recommendation system, boosting user engagement by 25% and relevance by 18% through behavioral analytics and adaptive retrieval, directly increasing platform adoption.
- Optimized Spark-based data pipelines and DeepSpeed multi-GPU training, cutting model training time by 67% (72 to 24 hrs) and deployment time by 94% (4 hrs to 15 min), while maintaining sub-200 ms inference at 1,000+ RPS for real-time analytics.

### Tata Consultancy Services

Apr 2021 - Aug 2023

Data Scientist

Chennai, India

- Architected and deployed data-driven automation solutions across 14+ client teams, reducing operational costs by 40% through statistical analysis of workflows and intelligent process optimization, enhancing business strategy
- Designed and implemented process optimization systems, achieving 100% client adoption by leveraging advanced data analytics to align automated delivery pipelines with key business performance indicators and support decision-making
- Developed and implemented A/B testing methodologies for evaluating the impact of automation changes, providing data-backed insights that informed strategic decisions and improved system effectiveness by 10%.

## EDUCATION

### George Washington University

Aug 2023 - May 2025

Master of Science, Computer Science

- **GPA:** 3.88/4.0

### SRM University

Aug 2016 - May 2020

Bachelor of Technology, Computer Science

- **GPA:** 3.5/4.0

## PROJECTS

### Agentic Graph RAG for Building codes | <https://vabuildingcode.netlify.app/>

- Built production-scale multi-agent AI system with LangGraph state machines and conditional routing, achieving 90% query accuracy uplift and 40% cost reduction through semantic analysis and data-driven optimization.
- Implemented observability with LangSmith tracing and Prometheus dashboards, ensuring 99.9% uptime and comprehensive performance monitoring.

### AI-Text Discriminator | <https://github.com/Venkat-Git98/AI-Text-Discriminator>

- Pioneered PEFT and LoRA fine-tuning, reducing trainable parameters by 90% while maintaining 97% classification accuracy on a 1.2M+ text corpus, demonstrating efficiency in large-scale NLP modeling.
- Developed scalable training pipeline achieving 3× faster training and 50% memory reduction, facilitating rapid iteration and resource optimization for text analytics.