Shanmuganathan Somashekar

Data Scientist

+15405585546 | shansomahekar@gmail.com | linkedin | github

SUMMARY

Data Scientist with 3 years of experience in developing end-to-end data solutions, predictive models, and analytics-driven insights. Proficient in Python and SQL, with a strong focus on building automated data pipelines and conducting in-depth analysis to drive user engagement and retention. Proven ability to transform complex data into actionable business strategies, optimize operations, and improve decision-making processes, contributing to measurable outcomes.

PROFESSIONAL EXPERIENCE

Data Scientist, **Kasmo** | **Plano, Texas** **May 2023 - Jan 2024**

- Developed regression and time-series models on 1TB+ multi-source operational data, reducing decision latency by 25% and enabling faster revenue-impacting insights for user engagement and retention.
- Engineered features across 10M+ records using PCA, correlation pruning, and statistical techniques, boosting model predictive accuracy by 18% and SLA adherence to 99%, contributing to growth forecasting.
- Designed automated Python & Airflow pipelines for data validation and anomaly detection, ensuring 99.8% reliability and saving $12K annually in reprocessing costs, streamlining analysis efficiency.
- Implemented model monitoring and A/B testing dashboards to communicate actionable insights that improved operational efficiency and supported ML-driven business decisions for 500+ users, influencing product strategy.

Data Scientist, **Kearney and Company** | **Blacksburg, Virginia** **Aug 2022 - Dec 2022**

- Automated extraction of 47GB from 400+ websites and 2,400 PDFs using Python, saving 40+ analyst hours per week and enabling near real-time business insights across multiple divisions.
- Developed NLP pipelines with FinBERT and NLTK to classify audit reports, achieving 90% accuracy and uncovering revenue-impacting contract discrepancies across multiple federal engagements.
- Designed and validated a lead scoring framework for business development, improving opportunity ranking efficiency by 3x and enhancing regional sales outreach effectiveness.
- Created interactive Tableau dashboards visualizing industry trends for 700+ executives, facilitating faster, data-driven investment decisions and reducing reporting cycle time by 50%.
- Conducted in-depth analysis to uncover insights into feature usage and identify drivers of engagement, informing product strategy and roadmap.

Undergraduate Researcher, **Virginia Tech** | **Blacksburg, Virginia** **Jan 2022 - May 2022**

- Utilized Python and R to analyze 50,000 ticket sales records over two years, uncovering seasonal trends and consumer behavior that led to a 15% increase in matinee revenue through targeted marketing initiatives.
- Developed regression and time-series models for user engagement and sales forecasting, generating $8K additional quarterly revenue through targeted pricing strategies.
- Collaborated with marketing to design data-driven promotions, boosting underperforming segment ticket sales by 10% and increasing audience engagement.
- Created R Shiny dashboards to automate reporting workflows, streamlining analysis processes and saving 8+ analyst hours weekly, thus accelerating decision-making.

Data Scientist Intern, **Choice School** | **Cochin, India** **Jan 2021 - Dec 2021**

- Automated academic workflows using Python to integrate G-Suite & Seesaw APIs, reducing workload by 50% and eliminating 5,000+ redundant records across multiple academic terms.
- Cleaned and standardized multi-year student datasets in SQL, enabling faster trend analysis and generating actionable insights for administrators supporting data-driven planning.
- Built Python-based predictive models analyzing multi-year student performance, accurately identifying at-risk students and enabling interventions that reduced overall dropout rates by 8%, improving retention.
- Developed interactive Power BI dashboards for visualizing admissions evaluation metrics, improving decision accuracy by 15% and shortening review time by 35% for senior academic panels.
- Delivered automated weekly performance reports, saving 10+ staff hours per week and ensuring timely interventions and data-driven academic decisions across administrative teams.

EDUCATION

Virginia Tech, Blacksburg, Virginia - *Master of Engineering*, *Computer Science* **Jan 2024 - May 2025**

Virginia Tech, Blacksburg, Virginia - *Bachelor of Science*, *Computational Modeling and Data Analytics* **Aug 2019 - May 2023**

PROJECTS

Zero-Shot Customer Feedback Insights Engine

- Built zero-shot NLP pipeline using T5/FinBERT and spaCy to process 10K+ daily customer feedback entries, integrating Snowflake & Tableau, reducing manual analysis by 80%.

Vehicle Firmware Integration Using Honda Dataset

- Designed and deployed production-quality vehicle firmware using Python and Honda real-world datasets, improving real-time system performance and ensuring safety-critical compliance across multiple vehicle modules.

Reinforcement Learning for Alarm Detection Grouping

- Applied reinforcement learning and anomaly detection models to identify system faults, optimizing AI-driven monitoring pipelines and reducing alarm handling time by 30% for large-scale sensor networks.

SKILLS

Programming & Scripting : Python, R, SQL, Matlab, C, C++, Java, HTML, CSS

Machine Learning & AI : Regression, Classification, Clustering, PCA, NLP, Reinforcement Learning, Time-Series Forecasting, Statistical Modeling, Predictive Analytics, RAG, LangChain, Vector Databases, PyTorch, TensorFlow, Scikit-learn, A/B Testing

Data Engineering & ETL : PySpark, Airflow, Snowflake, Databricks, Azure Data Factory, Hadoop, Redshift, Spark, G-Suite, Seesaw API integrations

Visualization & BI : Tableau, Power BI, Matplotlib, ggplot, d3.js, R Shiny

Cloud & Tools : Azure (Data Engineer Associate), AWS EC2, Git, Linux, Jira, Agile/Scrum

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate, Microsoft