

# SHANMUGANATHAN SOMASHEKAR

| Blacksburg, Virginia | [shan.somashekaran@gmail.com](mailto:shan.somashekaran@gmail.com) | +15405585546 | [github.com/ShanSomashekaran](https://github.com/ShanSomashekaran) | [linkedin.com/in/s-somashekaran/](https://linkedin.com/in/s-somashekaran/)

## EDUCATION

### Virginia Tech, Blacksburg, Virginia

Master of Engineering, Computer Science

Jan 2024 - Dec 2025

Bachelor of Science, Computational Modeling and Data Analytics

Aug 2019 - May 2023

**Relevant Coursework:** Data Structures and Algorithms, Mathematical Modeling, Intermediate Data Analytics, Machine Learning, Experimental Designs, Bayesian Statistics, Data Analytics and Visualization, Regression Analysis, Linear Algebra

## WORK EXPERIENCE

### Data Engineer - Kasmo - Plano, Texas

May 2023 - Jan 2024

- Developed ETL data pipelines using **Snowflake** and **PySpark**, resulting in a 25% increase in data processing efficiency
- Optimized the **data warehouse**, reducing query response time by 18% and improving system performance
- Developed and executed data warehousing strategies using **SCD1** and **SCD2** materialized views in **Snowflake** and **Azure**, resulting in a 20% increase in data accuracy

### Data Scientist - Kearney and Company- Blacksburg, Virginia

Aug 2022 - Dec 2022

- Integrated **Python** script to scrape 400 web pages & parse PDF links to text, to collate a dataset of 2400 potential opportunities, saving 47 GB of storage
- Incorporated **FinBERT** and **NLTK** in model design architecture to evaluate opportunity scores of potential clients, resulting in 90% accuracy of potential opportunities
- Designed an interactive dashboard in **Tableau** displaying the findings of the opportunity score for every audit report, contributing to **700 new business opportunities**

### Undergraduate Researcher - Virginia Tech - Blacksburg, Virginia

Jan 2022 - May 2022

- Analyzed 2 years of Moss Art Center ticket sales data using **statistical models**, identifying a 15% increase in revenue potential for matinee performances
- Collaborated with the marketing team to create targeted promotions for underperforming categories, resulting in a 10% increase in ticket sales
- Developed an interactive **dashboard** to analyze ticket sales data by category, resulting in an 8-hour decrease in time spent on manual **data analysis**

### Data Analyst Intern - Choice School - Cochin, India

Jan 2021 - Dec 2021

- Streamlined administrative processes for Choice School by integrating G-Suite and Seesaw, reducing workload by 50%
- Analyzed and evaluated the student's academic performance and identified areas of improvement, reducing the drop-out rate by 8%
- Developed a comprehensive data **dashboard** that enabled the admissions team to review student data with accuracy, resulting in a 15% improvement in admissions decisions

## TECHNICAL SKILLS

**Programming:** Python, SQL, R, Matlab, C, C++, Java, HTML, CSS, Airflow, Spark, Hadoop, JMP

**Libraries:** Sci-kit-learn, NLTK, FinBert, spaCy, Pandas, NumPy, JSON, Beautiful Soup, Matplotlib, ddplyr, ggplot, knit

**Software/Tools:** Jupyter Notebook, RStudio, MATLAB, Git, Tableau, Power BI, AWS EC2, Linux, Minitab, Agile, Jira

**Machine Learning:** Hypothesis Testing, A/B Testing, Regression, Principal Component Analysis, Clustering, Classification, Statistical Modeling, ETL, Azure Data Factory, PySpark, Snowflake, Databricks, Redshift, MapReduce, PyTorch, TensorFlow, Database Management, RAG, LangChain, Vector Database

## CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate

## PROJECTS

### Zero-Shot Customer Feedback Insights Engine

- Built a zero-shot NLP pipeline using transformer models (T5/FinBERT) to extract sentiment, customer issues, and intent from unstructured feedback across Amazon, Yelp, and TripAdvisor.
- Integrated spaCy-based NER and keyword extraction with Snowflake and Tableau to visualize real-time insights by geography, sentiment trend, and product category.