VAMSHI KRISHNA GAREGA

TX, USA | +1 (830) 359-9463 | Email | LinkedIn | Portfolio

SUMMARY

Data Engineer and AI Researcher seeking full-time opportunities, skilled in Python, SQL, Big Data, ETL/ELT, and cloud platforms like AWS. Strong foundation in core computer science concepts, including Data Structures and Algorithms, with a focus on building scalable data solutions.

EDUCATION

Master's in Computer Science CGPA: 3.75/4.0 ***Aug 2022 – May 2024***
Texas State University *San Marcos, TX*
Bachelor of Technology in Computer Science (AI Specialization) CGPA: 3.76/4.0 ***July 2018 – June 2022***
Koneru Lakshmaiah Education Foundation *Hyderabad, India*

TECHNICAL SKILLS

Programming Languages: Python, Java, C++, Scala, PHP, R, Perl
Data Technologies: Hadoop, Spark, Flink, MySQL, PL/SQL, Oracle, MongoDB, Hive, Treasure Data
Tools and Environments: PyCharm, Eclipse, VS Code, Jupyter, AWS, EC2, TensorFlow, Jenkins, Docker, Git
Software and Web Development: HTML, CSS, JavaScript, NodeJS, RESTful APIs, Azure Data Lake, Snowflake
Data Analysis and Visualization: Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI

PROFESSIONAL BACKGROUND

Braintrust Partners LLC *Austin, Texas*
Associate Data Engineer ***July 2024 – Present***

- Built, managed, and optimized ETL/ELT pipelines using Treasure Data, processing customer data across multiple stages to ensure clean, consistent data flow for a leading Texas bank.
- Developed automated workflows using Python and SQL to update audience segments, integrate data with HubSpot for marketing campaigns, and streamline data processing operations.
- Supported SQL query optimization for data analysts, enabling efficient churn analysis and customer segmentation by optimizing queries for large-scale datasets. Leveraged Apache Spark for distributed data processing, optimizing SparkSQL queries to enhance performance and reduce data processing time for large datasets.
- Worked with Databricks and AWS S3 to store, retrieve, and analyze structured and unstructured data, ensuring scalable, secure, and efficient data handling. Developed unit tests to ensure the reliability of production-level code and used Git for version control and code deployment, ensuring smooth project rollouts.
- Collaborated with data scientists and analysts to define and prepare data for AI/ML models by creating robust data pipelines and real-time streaming workflows using Kafka.
- Assisted with data migration from on-premises systems to cloud-based environments such as AWS Redshift, Azure Data Lake, and Snowflake, ensuring data integrity and accuracy during transitions.
- Responsible for monitoring and resolving data pipeline issues in real-time, leveraging Databricks and AWS EMR for debugging and fine-tuning pipeline performance.
- Integrated customer data with 3rd-party platforms like Facebook and LinkedIn for multi-channel marketing campaigns, ensuring seamless data activation and timely updates.
- Participated in weekly client deliverables, collaborating with the Chief Data Architect and stakeholders to ensure the smooth execution of data engineering projects and the resolution of support tickets using Click Up.

TEXAS STATE UNIVERSITY *San Marcos, Texas*
Graduate Research Assistant ***Aug 2022 – May 2024***

- Conducted advanced research on time series forecasting algorithms by integrating Adversarial Statistical Decision Theory (ASDT) and statistical models for ERCOT house load data, addressing real-world challenges in energy consumption prediction.
- Developed and optimized machine learning models using techniques such as ARIMA, SARIMA, and HMM for predictive analytics, improving forecasting accuracy by 20%. Assisted in the design, implementation, and maintenance of data pipelines and data platforms, ensuring efficient data flow and integration for research projects.
- Applied AI/ML algorithms for predictive modeling and collaborated on the Joint All Domain Command and Control (JADC2) framework for the United States Air Force Office of Scientific Research to enhance secure decision-making under data manipulation threats in distributed environments.

- Preprocessed, analyzed, and cleaned large datasets using Python and libraries such as Pandas, NumPy, SciPy, and Scikit-Learn, ensuring accurate input for machine learning models by optimizing model performance through hyperparameter tuning and cross-validation
- Designed and implemented scalable time series models and statistical analysis techniques, optimizing the performance of machine learning models to provide actionable insights for energy resource optimization and consumption prediction.
- Implemented CI/CD pipelines using Jenkins and Git, automating the workflow for model development, testing, and deployment, enhancing efficiency and reducing development time.
- Visualized model performance and time series forecasting results using Matplotlib, Seaborn, and Plotly, providing clear insights to stakeholders and research collaborators.
- Published research findings in academic journals, demonstrating the application of AI and statistical models in secure time series forecasting for energy and defense sectors.

ERNST AND YOUNG (EY) *Bengaluru, India*
Senior Software Analyst ***Feb 2022 – July 2022***

- Developed and maintained enterprise-level Java-based applications, focusing on business-critical legacy systems, ensuring stability, performance, and scalability. Collaborated with cross-functional teams to resolve bug fixes and improve system reliability by performing root-cause analysis of application issues, incorporating debugging and testing using.
- Integrated API services for third-party platforms, writing RESTful APIs to extend legacy systems, ensuring smooth interoperability between old and new software components. Participated in the entire SDLC, from requirements gathering and system design to testing and deployment, applying agile and waterfall methodologies based on project needs.
- Optimized algorithms and data processing workflows by applying principles of data structures, algorithms, and time complexity analysis, ensuring efficient code execution in high-load environments.
- Worked with legacy Unix/Linux environments to deploy, test, and manage Java applications, leveraging shell scripting for automation and system task management.

PUBLICATIONS

Title: "Command and Control with Poisoned Temporal Batch Data" Published in SPIE 12538 on June 12, 2023. DOI: 10.1117/12.2663

- This publication delves into the realm of command-and-control systems, specifically addressing the challenges posed by poisoned temporal batch data. This paper contributes to the advancement of strategies and frameworks for secure decision-making in dynamic environments. Published in a reputable SPIE journal, showcasing the impact of the research on the broader scientific and technological landscape.

ACADEMIC PROJECTS

Text Summarizer:

- Developed a sophisticated web application using Django, incorporating Natural Language Processing (NLP) and Machine Learning (ML) techniques to generate concise and meaningful summaries of web pages based on user-provided URLs.
- Implemented advanced algorithms for information extraction and abstraction, demonstrating a strong grasp of distributed computing principles to handle real-time summarization demands.

DotCom E-commerce Website:

- Led the full-stack development of a robust E-commerce platform using Python and Django, showcasing proficiency in backend and frontend technologies.
- Implemented features for order management and inventory tracking, utilizing distributed computing principles to ensure seamless performance and responsiveness.

Note Taking Application:

- Designed and implemented an Android application using React Native and NodeJS, providing users with an intuitive platform for efficient recording of important daily events and notes.
- Integrated Google Firebase for real-time data storage and synchronization, highlighting API integration skills.
- Utilized Bootstrap for responsive and user-friendly design, demonstrating a keen eye for user experience and frontend development.

HONORS AND AWARDS

United States Airforce Office of Scientific Research Student Fellowship Grant (~$16,000)
Texas State University Student Government Scholarship (Academic Year 2023 - 2024)
Texas State University Computer Science Graduate Academic Excellence Award (Academic Year 2023 – 2024)
Best Graduate Research Poster for 2024 (TXST Center for Analytics and Data Science)