

SAI NIKHIL ARATIPAMULA

+1 (850) 339-3663 | snaratipamula@gmail.com | Tallahassee, FL, 32303 | www.linkedin.com/in/sainikhilaratipamula/

EDUCATION

Florida State University | CGPA: 3.6 / 4.00

08/2023 - 05/2025

Master of Science in Data Science

Tallahassee, FL

Relevant Coursework: Advance Data Science, Machine Learning, Applied Regression Methods, Mathematics for Data Science.

SRM Institute of Science and Technology | CGPA: 3.7 / 4.00

06/2018 - 05/2022

Bachelor of Technology in Computer Science Engineering

Chennai, IND

Relevant Coursework: Artificial Intelligence, Data Mining and Analytics, High Performance Computing, Distributed Operating Systems.

WORK EXPERIENCE

LTIMindtree

08/2022 - 07/2023

Data Engineer — Chennai, India

- Implemented Apache Spark, Hive, and Hadoop, enhancing the efficiency of large dataset management.
- Achieved a 40% increase in data efficiency and a 50% reduction in processing times through optimized data processing.
- Developed, tested, and debugged ETL code to improve data workflow efficiency.
- Initiated Unix scripts for job scheduling and automation, reducing manual tasks by 50% and processing time by 30%.
- Spearheaded cross-functional collaboration with users, developers, and stakeholders to optimize Off-Shore/On-Shore models.
- Enhanced data quality and reduced processing time by 30% by resolving issues within the Precertification tool.
- Organized interactive meetings to gather user and functional requirements effectively.
- Accelerated project development, reducing time by 30% and ensuring the successful delivery of solutions that meet customer needs.

Larsen and Toubro Infotech

06/2022 - 08/2022

Data Engineer Intern — Mumbai, India

- Orchestrated seamless integration of diverse data sources, enhancing end-to-end data processes from ingestion to analysis.
- Facilitated data-driven decision-making with impactful visualizations that influenced company-wide strategy.
- Acquired expertise in industry-standard tools such as PySpark, HDFS, Tableau, and AWS S3.
- Enabled efficient data storage and visualization, contributing to data-driven strategies and project success.

SKILLS

Programming: C, C++, **Python**, **SQL**, Java, R

Tools and Libraries : PySpark, Apache Spark, Hadoop, Hive, Kafka, **Kubernetes**, Unix, Oracle, ServiceNow, **ETL/ELT**, Apache Airflow, NumPy, Scikit learn, AWS QuickSight, Tectia, Power BI, **Data Pipelines**, Databricks, Tableau, Git, Github.

Additional Skills: Data Warehousing, **Data Modelling**, **Big Data Architecture**, **Data Analysis**, **Data Lake**, Advance Excel, SDLC, Information Technology, Information Systems, Technical Support, Operating Systems - Windows, Linux, Ubuntu, Web development - HTML, CSS, Cloud - **AWS**, **GCP**.

CERTIFICATIONS

Databricks Certified Associate Developer for Apache Spark 3.0, AWS Fundamentals: Going Cloud-Native, Image Augmentation with Keras, Programming for Everybody (Python), Introduction to Chatbot

PROJECTS

PREDICTIVE MODELING OF HOUSE PRICES IN AMES | R, ML, linear algebra, regression analysis

10/2023 – 12/2023

- Engineered an advanced housing price prediction model using the Ames Housing dataset, employing Linear models, Random Forest.
- Produced a 15% accuracy boost through innovative feature engineering and model optimization techniques.
- Applied Box-Cox and square root transformations to tackle non-constant variance issues.
- Showcased expertise in advanced statistical methodologies, enhancing predictive modeling performance for the Ames Dataset.

COVID-19 ANALYSIS | PySpark, Ubuntu, Tableau, Hive, Hadoop

06/2022 - 08/2022

- Utilized Spark to streamline healthcare datasets, converting unstructured data into a structured format for better accessibility.
- Merged data from various sources and formats, enhancing analysis capabilities in healthcare.
- Implemented a range of Spark transformations, significantly improving data processing efficiency.
- Assessed a 40% reduction in data processing time, facilitating real-time analytics in the healthcare sector.

SOCIAL DISTANCING AND FACE MASK DETECTION FOR COVID PREVENTION | Python, ML

01/2022 - 05/2022

- Created a machine learning system for monitoring social distancing and mask use, boosting compliance by 40% across locations.
- Exercised live feeds and recorded data to assess public health measures effectively, ensuring adherence to safety protocols.
- Improved object detection by enhancing MobileNetV2 and YOLO libraries with Python and OpenCV, focusing on public health.
- Analysed high precision and real-time data processing, critical for maintaining public health and safety standards.

CHATBOT FOR HEALTHCARE | Python, Natural Language Processing, MySQL, Web Technologies

08/2021 - 12/2021

- Designed a healthcare web app with a chatbot, leveraging NLP, data collection, and user authentication to enhance medical access.
- Utilized intent/entity recognition to streamline patient interactions, significantly reducing unnecessary hospital visits by 40%.
- Improved patient satisfaction through better access and privacy, leading to enhanced healthcare experiences.