ROHIT SUBRAMANIAN ARIVALAGAN

New Jersey, USA (973)2895653 rs269@njit.edu /rs269 /rs269

EDUCATION

New Jersey Institute of Technology ***Sept 2021 - May 2023***
MS in Data Science - Computational Track (4.0 CGPA)

Vellore Institute of Technology ***July 2017 - June 2021***
B.Tech in Computer Science and Engineering (3.7 CGPA)

SKILLS

Languages:	Python (Pandas, NumPy), SQL, R, Scala, C#, C++, Java
Tools & Frameworks:	TensorFlow, Keras, Scikit-learn, PyTorch, Hadoop, Spark, Kafka, Hive, HBase, Airflow, Power BI, Tableau, Azure Data Factory, Azure Synapse, Azure Data Lake , Databricks, Snowflake, Redshift, NoSQL (MongoDB, DynamoDB, Cosmos DB, Cassandra), RDBMS (MySQL, Postgres, MariaDB)
Cloud Technologies:	Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)
Industry Knowledge:	Machine Learning, Cloud Computing, Big Data, Data Mining, Data Warehousing, Data Modeling, Database Management Systems, Data Structures & Algorithms, Software Engineering, Statistics

WORK EXPERIENCE

Associate Data Engineer (Data Analytics) - GEP Worldwide ***Jun 2023 - Dec 2023***
- Performed data migration from V1 (ElasticSearch + Kafka) to V2 (MongoDB + Kafka) framework and SMART to LCX, based on domain model automation and created datedims, measures & crosscubes for report generation using SSIS/SSAS.
- Built and monitored end-to-end pipelines and scheduled triggers on ADF for ingestion, ETL, cube refresh and data extract jobs.
- Configured, tested and implemented Data Access Control rules on cubes, using Rest APIs to enforce data security and confidentiality of reports and worked on automation of the process, reducing manual efforts and implementation time by 60%.
- Enhanced, optimized and added functionalities to existing ADB frameworks, increasing stability and reducing bugs by 45%.
- Performed code cleanup to enable code sync and deployment between QC, UAT, SIT & PROD using Git and CI/CD Integration.

Data Engineer (Data Analytics) - GEP Worldwide ***July 2022 - Dec 2022***
- Ingested client data from Kafka and ElasticSearch, performed ETL and setup OLAP cubes on AAS for creating analytics reports.
- Transitioned data loading (streaming + historical data) from a full-load overwrite framework to ABC incremental framework for S2P and P2P modules, reducing load times by about 70% for each refresh cycle.

Data Science Research Assistant - New Jersey Institute of Technology ***May 2022 - July 2022***
- Performed EDA, compared various ML models to analyze existing features and identify patterns that cause significant delays, increasing the efficiency of Ship Maintenance Scheduling decisions for the US Office of Naval Research.
- Automated data extraction for dynamic querying from large US Navy's ships maintenance dataset, reducing query time by 30%.

Graduate Teaching Assistant - New Jersey Institute of Technology ***Jan 2022 - May 2023***
- Assisted professors with course materials creation, grading assignments and examinations for graduate level courses : Data Structures & Algorithms, Machine Learning, Deep Learning, Data Analytics with R Programming.

Data Analyst Intern - AVASOFT Inc ***May 2019 - July 2019***
- Performed ETL/ELT transformations and DAX functions on Avasoft's bug dataset present in the company's private TFS server.
- Developed Power BI dashboards to produce intuitive and insightful analysis of bug data, reducing debugging time by 45%.

RELEVANT PROJECTS

•**Amazon Product Review Analysis (Natural Language Processing):** Leveraged NLP and DL techniques to identify suspicious reviews (SVM), summarize reviews (Encoder-Decoder with 3-stacked LSTM) and predict user rating for the product based on sentiment classification of the review (Bidirectional RNN with LSTM – GRU) on a Python (Tkinter GUI) application.

• **Flight Data Analysis (Hadoop):** Developed an Oozie based workflow for Hadoop to execute mapreduce programs to analyze a large volume of flight data (12GB) using AWS VMs and compared performance analytics with varying number of VMs.

• **Online Computer Store (Database Systems):** Analyzed functional requirements, designed ER model & relational schema with integrity constraints to create and populate online computer store database (MySQL) with a Python menu-driven application.

• **Car Image Recognition and Wine Quality Prediction (AWS):** Created EC2 instance to process images present in S3 bucket and send indices of car images to another instance through SQS to filter images with text using Java & Maven. Trained ML model in parallel (4 worker + 1 master) using EMR and Docker to predict wine quality on training and validation datasets stored in S3.

• **H&M Personalized Fashion Recommendation System (Deep Learning):** Implemented a neural network based recommendation system to improve user shopping experience by recommending apparel based on past purchases (145 billion rows) and boosted results (15%) by employing ensemble methods combining various convolutional, sequential and filtering models.

CERTIFICATIONS

Microsoft Azure Data Engineer Associate **Databricks Lakehouse** Fundamentals **AWS** Cloud Practitioner