VENKATESH GOPINATH BOGEM

Seattle, WA (open to relocation) | +1 (617) 466 6066 | bogemvenkateshgopinath@gmail.com | linkedin.com/in/venkatesh-gopinath/

EDUCATION

MASTER OF SCIENCE IN DATA ANALYTICS
Northeastern University

- Coursework: Database Management System, Data Warehousing, Cloud Computing and Distributed Systems, Advanced SQL, ML and Adv NLP

WORK EXPERIENCE

DATA ENGINEER **May 2024 – Present**
Abecedarian, Remote - Seattle, Washington

- Building scalable data pipelines to integrate **Gen AI techniques** for automated data ingestion, transformation, and synthetic data generation.
- Implementing **custom data models** and **ETL processes** to optimize data flow and support advanced analytics and machine learning tasks.
- Leveraging cloud infrastructure (**AWS**, **GCP**) to deploy, monitor, and maintain scalable data solutions for AI-driven data analysis.
- Utilizing **machine learning techniques** to automate data transformation and generate insights for strategic decision-making.
- Ensuring **data governance** and **security compliance** while managing large datasets and implementing **automated ETL processes**.

DATA ENGINEER / ANALYST **May 2023 – Dec 2023**
Datics, Charlotte, North Carolina

- Optimized **SQL** queries and integrated various databases like **AWS RDS**, **MYSQL**, **Snowflake**, **MongoDB**, **AWS DynamoDB**, etc improving data retrieval efficiency for large-scale projects.
- Enhanced **ETL** workflows by integrating **AWS Glue** with **S3** for data storage and **AWS Lambda** for automated triggers, reducing data processing time from raw data to insights by 50%.
- Developed **CI/CD** pipelines by containerizing with **Docker** and managing tasks using **Apache Airflow**, improving deployment efficiency.
- Managed the **data warehouse** ensuring data accuracy and availability for downstream, enabling a 40% improvement in reporting efficiency.

DATA ENGINEER I **Jun 2019 – Dec 2021**
ACCK Solutions, Hyderabad, India

- Engineered and optimized **ETL workflows** to process and transform over 1 million records daily from sources such as **Facebook Ads**, **Google Analytics**, and **Salesforce CRM**, ensuring data accuracy and timely availability for downstream analytics.
- Established data warehouse and data marts in **Star** and **Snowflake schema** via dimensional data modeling for MPP databases.
- Developed and optimized complex **SQL queries** and performance tuning through **indexes**, **partitions**, **stored procedures**, **triggers** and **aggregations**, resulting in faster generation of financial reports for stakeholders.
- Developed and optimized large-scale data processing workflows using **PySpark**, handling over 500GB of data, and performing complex transformations and aggregations.
- Developed interactive visualizations and dashboards using **Tableau, AWS QuickSight** and **PowerBI**, transforming complex datasets into actionable insights for stakeholders, which improved decision-making processes.
- Collaborated with cross-functional teams in an Agile environment, leveraging Jira for tracking, to develop and optimize robust data pipelines

MACHINE LEARNING ASSISTANT **Sep 2022 – May 2024**
Northeastern University, Boston, MA

- Automated the grading process using Python, **GitHub Actions** for consistency and scalability and also assisted in grading exams and assignments, providing personalized feedback to help students understand machine learning concepts and improve their coding skills.
- Mentored students in building a range of machine learning models, including **CNNs**, **RNNs**, etc, using **TensorFlow**, **PyTorch**, and **Keras**.

SKILLS

Technical Skills: Python, R, Java, Advanced SQL, NoSQL, PowerBI, Tableau, QuickSight, Looker, Numpy, Pandas, Matplotlib, Scikit-learn, TensorFlow, PyTorch, Apache Spark, Hadoop, BigQuery, Apache Airflow, Jenkins, Docker, Kubernetes, Terraform, SSIS, ODI, BODI, Informatica, Datastage, Git, Jira, MLflow, Kubeflow, Kafka, Alteryx, Data Catalog, Snowflake, PostgreSQL, Oracle, Flask, FastAPI, MS Office, Excel.
Cloud: AWS (S3, EMR, Kinesis, Glue, QuickSight, Redshift, Lambda, IAM, Lambda, Elasticsearch, IAM), similar services in Azure and GCP.

PROJECTS

Sentiment Analysis Of Online Product Reviews

- Utilized Python, NLTK, and Neural Networks for sentiment analysis on product reviews, applying Bag of Words, TF-IDF, and LSTM models to enhance positive sentiment prediction precision to 90% and overall accuracy to 85%, outperforming benchmarks by 5%.

MLOps Pipeline for Emotion Detection on GCP (Link)

- Developed end-to-end Emotion Detection MLOps pipeline leveraging Google Cloud Platform (GCP) infrastructure.
- Integrated tools such as TensorFlow, Airflow, MLflow and DVC to streamline model development, deployment, and monitoring processes.

Design and Development of Sports Database

- Built and normalized a comprehensive sports database, maintaining data integrity and reliability through strict table-level constraints; delivered a dynamic PowerBI analytics dashboard that effectively communicated insights through visualizations.