IDUPULAPATI GAYATHRI SAI YASASWI

928, 800 W Renner Road, Richardson, TX

+1 (945) 527-4904 ● yasaswiigsms@gmail.com ● LinkedIn

Innovative Data Engineer with expertise in GCP, Python, and ETL pipeline development. Proven track record of optimizing data processes at Tenet Health and Lam Research, with strong skills in Big Query, Dataflow, and cloud-based data solutions. Combines technical proficiency with a data science background to deliver high-impact engineering solutions.

PROFESSIONAL EXPERIENCE

Data Engineer Intern, *Tenet Health Corporation, Dallas, United States* **June 2024 – Present**

- Built and optimized ETL pipelines using Google Cloud Platform (GCP) services like Big Query, Dataflow, and Cloud Storage.
- Developed a file ingestion framework in Python on GCP for seamless integration of data from various sources into Big Query.
- Led data migration process from T-SQL to BQ, to create reporting tables for Power BI dashboards, using Python-based scripts.
- Supported data science team with Python-based tools and infrastructure, ensuring integration with GCP analytics solutions.

Data Engineer, *Lam Research, Bangalore, India* **Aug 2021 - Aug 2023**

- Designed and maintained scalable data pipelines to process unstructured data for analytics and reporting, using Python and SQL.
- Migrated legacy data systems to cloud-based solutions, utilizing GCP services like Big Query and Dataflow, improving efficiency.
- Developed data ingestion frameworks integrating diverse data sources into centralized data warehouses, enhancing accessibility.
- Built automated data quality monitoring scripts to ensure reliable and consistent data pipelines.
- Implemented distributed data processing systems using Apache Hadoop and Spark, improving processing speed and efficiency.
- Collaborated with cross-functional teams to deliver optimized data solutions, reducing report generation time.

Software Engineer Intern, *Lam Research, Bangalore, India* **Jan 2021 - Jul 2021**

- Engineered a real-time data visualization platform to monitor system health and metrics, leveraging Python and Flask.
- Developed and optimized RESTful APIs for data integration and analytics, reducing backend latency.
- Designed scalable ETL workflows to handle data ingestion and transformation for operational reporting.
- Built a dynamic dashboard using Angular and JavaScript for visualizing key performance metrics of production systems.

ACADEMIC PROJECTS AND PUBLICATIONS

NIKE Stock Price Prediction using LSTM, *UT Dallas, Richardson, TX* **Mar-May 2024**

- Implemented LSTM using **pyspark and mapreduce** to predict the stock prices with efficiency of 96% accuracy.
- Performed a comparative study with **GBT regressor** and highlighted the model's strength and weakness in terms of efficacy.

Argon DB Engine, *UT Dallas, Richardson, TX* **Aug - Dec 2023**

- Implemented a DB Engine in **JAVA,** using file-per-table approach for efficient storage of table files, index files at the OS level.
- The system effortlessly manages **DDL, DML, DQL, and SDL** command lines, alongside API calls, with an accuracy rate of 99%.

Football Player Move Prediction, *RIT, Bangalore, India* **Aug - Mar 2021**

- Built a system integrating computer vision **(EDSR), CNN Dense Net** framework to predict players next moves over a video data.
- Evaluated **deep learning** frameworks like **ResNet152, DenseNet161, and ResNet162** for optimal predictive model architectures.

Hybrid Decision Support System to Enhance crop productivity, https://link.springer.com/chapter/10.1007 **Jan - May 2020**

- Designed a sophisticated system integrating **hybrid ML algorithms (ARIMA, LSTM)** to forecast seasonal crop prices and yields.
- Optimized the hybrid algorithm by 20% and trained the system for incorporating advanced predictive analytics techniques.
- Developed a web application to generate predictive crop forecasts, enhancing data representation through dynamic graphs.

Document Retrieval Through Cover Density Ranking, https://ieeexplore.ieee.org/document/9622541 **Aug - Dec 2019**

- Developed **Cover Density Ranking (CDR)** an enhanced information retrieval method that overcomes the limitations of **Vector Space Model (VSM)** in python with better computational power of 85%.

TECHNICAL SKILLS

- **Programming:** Python, Java, JavaScript, C, C++, TypeScript
- **OS**: Unix, Linux, Windows
- **Full-Stack Development:** Angular, React, Node.js, Spring Boot, Bootstrap, Flask, HTML, CSS, Restful API
- **Frameworks/Libraries:** Hadoop, Apache, Pyspark, Scrapy, Scikit Learn, PyTorch, Pandas, Numpy, Keras, Tensorflow
- **Database Management:** MySQL, MongoDB, PostgressSQL, SQL, NoSQL
- **Additional Skills:** Git, GCP, Jira, Microservices, Azure, Big Query, UNIX, Docker, Kubernetes, DevOps, AWS

EDUCATION

The University of Texas at Dallas, Richardson, TX **Aug 2023 - May 25**

Master of Science in Computer Science, specialization in Data Science

Ramaiah Institute of Technology, Bangalore, India **Aug 2017 - May 21**

Bachelor of Engineering in Computer Science