# Adith Prabukumar

📞 7169944271   ✉ adithprabu27@gmail.com   in Linkedin   ↖ Portfolio

## Professional Experience

**Data Engineer,** *Frontier Science and Technology*                 Mar 2023 | Remote, United States
- Designed **ETL Framework** using **dbt** and **Apache Spark**, to transfer data from hospitals to AWS cloud servers, leading in 60% in employee productivity
- Collaborated with Senior Enterprise Architects, Solution Architects, Product Managers, to understand the workflow and ensure standards and rationalization plans to improve visibility and insights of strategic outcomes to drive product **impact by around 70%**
- Worked on real-time data transfer pipeline utilizing Kafka, transferring over **1M patient data** from local server to **AWS Redshift** data warehouse, ensuring ease of access to data and increasing productivity by 80%
- Played a pivotal role in migration of complex data sets to AWS deploying **AWS DMS**, establishing easy access to data for more than 3000 employees across the different parts of the country
- Developed a customized **NLP chatbot** for employee inquiries utilizing **Large Language Models (LLMs)** and integrated Snowflake Security tool for data access control based on clearance levels, boosting productivity by 60%
- Worked on unit tests using JUnit and performed quality checks to achieve software functionality, glitch elimination and update maintenance of access control, leading to **70% increase in performance.**
- Employed **Apache Airflow** to execute DAGs for data collection, reducing data processing time by 67%.
- Spearheaded a design and implementation of robust **CI/CD pipelines** using **GitHub Actions** and other version control methodologies across all projects.

**Data Analyst,** *AINQA Group*                 Aug 2020 – Jul 2021 | Mumbai, India
- Built a website using **Python** and **Flask** capable of parsing CSV files and pushing to the company's database.
- Leveraged **OCR technology** using **OpenCV** to extract structured data from clinical files from pharmacies leading to the creation of a comprehensive data set, increasing the productivity of the team by 40%.
- Developed and maintained interactive **Tableau** dashboards that visualized critical business metrics for executive decision-making, which was presented to Senior Product Managers and Company Vice President.
- Employed diverse **KPI trackers** and vendor graphs to refine the codebase, resulting in optimized performance while creating **user-friendly documentation** for analysts

## Skills

**Languages:** (Python, R, Go, C, Java, SQL (PostgreSQL, MySQL), Git)

**Big Data:** (Apache Hive, Kafka, Hadoop, Apache Spark, Apache Airflow, Hadoop)

**Certifications:** (DataBricks Certified Data Engineer Associate)

## Education

**Masters of Science in Data Science,** *University at Buffalo*                 Aug 2021 – Feb 2023 | Buffalo, NY
  **Cumulative GPA**: 3.8/ 4.0 - *Magna Cum Laude*

**Bachelor of Engineering,** *Anna University*                 Aug 2017 – Mar 2021 | Tamil Nadu, India
  **Cumulative GPA:** 3.7/ 4.0

## Projects

**News Summarizer** ⇗
- Built a user-friendly website that uses an **NLP-based BERT** transformer model to summarize news articles.
- We developed the site with the **Streamlit** package and hosted it on an **AWS EC2** Instance.
- The entire project is packaged in **Docker** for easy deployment, and our deployment process is automated using **GitHub Actions** in a CI/CD pipeline.

**Audiolytics: A Speech to Text Conversion Tool** ⇗
- Developed an **end-to-end ML model** that implements speech-to-text on audio files and returns transcribed text along with timestamps of each word.
- Applications involve podcasts and audio book editing. Model made using **wav2vec2** and endpoint with **Flask**. Packaged using **Docker** and maintained using **GitHub Actions**.