

# Giridharan Chandrababu

giri1208srinivas@gmail.com

+14379933641

Toronto, ON

## Summary

Data Scientist with 6 years of experience specializing in data analysis, machine learning, and NLP. Proficient in Python, SQL, and cloud platforms (GCP, AWS). Expertise in developing scalable data solutions, including ETL pipelines and recommendation systems. Skilled in AI and NLP technologies (GPT, Hugging Face, LLMs, LangChain, RAG, BERT) and adept with Python frameworks (Flask, FastAPI), Pandas, Terraform, Docker, and databases (PostgreSQL, MongoDB, Typesense, BigQuery). Experienced in utilizing data visualization tools (Matplotlib, Seaborn, Plotly, Tableau, Metabase, Apache Superset, Power BI) to deliver impactful business insights.

## Experience

### Data Scientist

DataFabricX

09/2023 - 09/2024

- **API Development:** Developed RESTful APIs using Python(FastAPI) to enable real-time data retrieval and batch processing; implemented advanced search and filtering to enhance AI search usability.
- **Data Pipeline Automation:** Designed and automated ETL pipelines with Python and SQL to integrate data from multiple sources into the vector database, optimizing data ingestion for AI-powered search.
- **Machine Learning Pipeline:**
  - **Data Processing:** Developed data synchronization process including embedding creation, data chunking, and security mapping.
  - **OCR Processing:** Implemented Optical Character Recognition (OCR) for efficient document extraction.
- **AI Integration:** Built on top of Large Language Models (LLMs), LangChain, FAISS, and Typesense to provide advanced search capabilities.
  - Enabled AI-driven features allowing users to interact with documents effectively.
- **Deployment:** Created Windows service packages using batch scripts and Docker packages for containerized deployment, ensuring efficient and scalable releases across environments.

### Software Engineer(Data science)

Quickplay Media

01/2022 - 09/2023

- **End-to-End Data Pipeline Development:** Built comprehensive data pipelines and ETL processes on Google Cloud Platform (GCP) for data ingestion, cleaning, and transformation, preparing datasets for machine learning models and analytics.
- **ETL Workflows & Data Migration:** Led the design and deployment of ETL workflows for seamless data migration from AWS to Snowflake, Cloud Storage, and BigQuery, ensuring secure and efficient data transfer.
- **Infrastructure Automation:** Implemented Terraform for GCP resource deployment, reducing setup time by 30% and enhancing infrastructure scalability.
- **Automated Data Collection:** Developed custom Python scripts to automate data collection from multiple sources, integrated with GCP services for improved processing efficiency.
- **Data Marts & Business Intelligence:** Collaborated with cross-functional teams to develop data marts and Looker dashboards, supporting diverse business needs.
- **Recommendation Systems:** Designed and implemented content-based and user-based recommendation systems, significantly enhancing user engagement and personalization.
- **Audit Trail System Development:** Created an Audit Trail system to monitor and analyze user interactions across storefront applications, providing insights for optimizing user experience and security.  
Automated Reporting: Streamlined reporting by automating report generation in Tableau, improving data accessibility for stakeholders.

### Data Scientist

DataFabricX

05/2019 - 01/2022

- **ML Document Classification Application:** Developed a machine learning application to organize and classify enterprise documents, improving accessibility and structure before migrating to a Document Management System (DMS).
- **NLP Document Processing:** Applied NLP techniques for in-depth text analysis, enabling precise categorization based on content and metadata.
  - **Supervised Learning Models:** Trained models using Random Forest and SVM algorithms for high-accuracy document classification within the DMS.
  - **Unsupervised Clustering:** Utilized K-means clustering to group similar documents, streamlining organization and retrieval processes.
- **Data Pipeline Development:** Designed and maintained ETL pipelines from DMS to Elasticsearch, optimizing workflows and ensuring data consistency.
- **Workflow Automation :** Automated repetitive data processing tasks with Python scripts, reducing manual workload and increasing efficiency.
- **Reporting:** Created a 360-degree document reporting dashboard using Kibana, providing real-time insights into document metadata and categorization.

## Jr. Data Scientist

Trusfi Technologies

10/2018 - 12/2019

- **Text Recommendation System:** Designed and developed a sophisticated text recommendation system using Keras and NLP techniques, enhancing user engagement and content discovery.
- **Nudity Detection System:** Developed a nudity detection system employing advanced image processing and machine learning techniques, ensuring safe content delivery within the application.
- **Deep Learning Models:** Leveraged Keras for building and training neural network models, optimizing performance and accuracy in recommendation tasks.
- **Image Processing Techniques:** Implemented convolutional neural networks (CNNs) for robust nudity detection, ensuring compliance with content policies.

---

## Skills

- **Programming Languages:** Python, SQL, React, Go (Beginner)
- **Cloud Platforms:** Google Cloud Platform (BigQuery, Dataproc, Composer, Cloud Run, Cloud Functions, Cloud Storage, DataStore, GCP Recommendations ), AWS
- **Infrastructure & Automation:** Terraform, Docker, GitOps
- **Data Engineering:** ETL Pipeline Development, Data Ingestion, Data Transformation, Data Quality
- **Machine Learning & NLP:** Scikit-learn, TensorFlow, PyTorch, Hugging Face, GPT, LangChain, LLM, BERT, Deepseek
- **Databases:** PostgreSQL, MongoDB, Typesense, BigQuery
- **Data Visualization & BI:** Streamlit, pandas, Jupyter, Tabulae, Metabase, Looker
- **Tools:** PyTorch, FastAPI, Flask, Jupyter Notebooks

---

## Certificates

Python (Hacker Rank), SQL Advanced( Hacker Rank)

---

## Education

### Technology

St.Peters College of Engineering

12/2016

### Technology

Anna University-MIT Campus

12/2018