SIDDHARTH CHEVELLA

📞 +91 8074909647 | ✉ siddharth070504@gmail.com | in linkedin.com/in/siddharth-ch05 | ⚙ github.com/SiddharthCh7

Work Experience

Futurense Technologies   Jun 2024 – Sep 2024
Data Analyst Intern
- Streamlined data collection and reporting procedures, reducing data gathering time by **30%** and accelerating decision-making.
- Collaborated cross-functionally to gather requirements and define project scopes, improving delivery timelines while ensuring alignment with business objectives.
- Produced comprehensive **reports and presentations** summarizing findings and recommendations, leading to a **15%** improvement in stakeholder decision-making speed and facilitating clear communication with key business stakeholders.

Projects

Dynamic ELT Pipeline — Azure Data Factory (GitHub)   *Mar 2025 – Apr 2025*
- Built a dynamic, reusable **ELT pipeline** using **Azure Data Factory** to ingest data from **multiple APIs** with minimal configuration.
- Leveraged **pipeline parameterization** to automate ingestion of diverse datasets, enabling scalable data onboarding with a single pipeline.
- Reduced manual overhead by **80%** through dynamic dataset handling and linked service reuse across multiple API sources.
- Improved pipeline deployment time by **70%** and enhanced maintainability with **modular** pipeline design.

RAG-powered Automation Tool — LangChain, Chroma, Hugging Face (GitHub)   *Feb 2025 – Mar 2025*
- Developed an automated LinkedIn post generation tool that allows users to input a GitHub repository link or a topic, generating and posting content on their LinkedIn profile.
- Integrated **Beautiful Soup** for scraping GitHub/topic data and **LangChain** for processing, reducing research time by **85%**.
- Automated **5+** manual steps into a single workflow, reducing post creation time from **45 minutes** to **under 5 minutes (90% faster)** and enabling **10x** content output.

Hate Speech Classification — Polars, NLTK, Gensim, Scikit-learn, Azure (GitHub)   *Nov 2024 – Dec 2024*
- Achieved **83%** accuracy in hate speech detection using Scikit-learn and NLTK, with hyperparameter tuning reducing false positives by **22%** compared to baseline models.
- Optimized pipeline processed **700K+** samples in **under 2 minutes (5,800+ samples/sec)** via lazy evaluation and batch processing (Polars), reducing memory overhead by **35%**.
- Enabled efficient large-scale processing by implementing Gensim memory mapped I/O and NumPy sparse matrices, handling datasets **8x** larger than conventional approaches.

Certifications

DP-900: Azure Data Fundamentals   Oct 2024 – Oct 2025
Microsoft (Link)

Skills

Programming Languages: Python, Java
Frameworks & Libraries: PyTorch, FastAPI, NLTK, LangChain, Scikit-learn, Hugging Face, Docker, Azure, gRPC, Pandas, Matplotlib

Education

Lovely Professional University, Punjab   2022 – Present
Computer Science Engineering, CGPA: 7.5   *Jalandhar, Punjab*

L.F Junior College, Hyderabad   2021 – 2022
12th with science, CGPA: 8.5   *Hyderabad, Telangana*

Narayana High School, Hyderabad   2019 – 2020
10th with science, CGPA: 9.5   *Hyderabad, Telangana*