Chesta Pruthi

cheshtapruthi2502@gmail.com| +91 7011521124 | linkedin.com/in/chesta-pruthi

Aspiring Data Engineer | Python | SQL | AWS | ETL | Power BI | Data Warehousing | ELT

Professional Summary

Detail-oriented Computer Science undergraduate with a strong foundation in Python and SQL, and hands-on experience in data analytics, data engineering, and cloud tools. Skilled in managing databases with MySQL and building data pipelines using Airflow, Spark, and AWS. Proficient in Power BI and Tableau for data visualization. Adept at version control with Git and working on platforms like Databricks and Azure Data Studio. Passionate about using data-driven insights to solve real-world problems and eager to continuously learn emerging technologies in data and software development.

Technical Skills

- **Programming Languages:** Python, SQL
- **Data Analytics Tools:** Power BI, Tableau, Excel
- **Databases:** MySQL
- **Tools & Platforms:** Git, Databricks, Azure Data Studio, AWS S3, Redshift, Apache Airflow, Spark, Kafka, Hive
- **Libraries/Frameworks:** pandas, NLTK, matplotlib, scikit-learn, Streamlit, Prophet, Statsmodels, PySpark

Skills Summary: Python, SQL, ETL, AWS, Airflow, Power BI, Tableau, MySQL, Spark, Hive, Databricks, Forecasting, NLP, Machine Learning, Git, ELT, Data Warehousing

Education

Bachelor of Technology in Computer Science
Lovely Professional University, India
Expected Graduation: 2027 | **CGPA:** 7.0/10

Certifications

- Microsoft Azure AI Fundamentals (AI-900)
- Microsoft Certified: Power BI Data Analyst Associate (PL-300) (In-Progress)

Projects

Sales Forecasting Pipeline for E-Commerce
Role: Data Engineer
Tools: Python (Advanced), SQL (Advanced), AWS S3, AWS Redshift, Apache Airflow, Prophet, ARIMA (Statsmodels), pandas, Tableau

- Designed and implemented an automated ETL pipeline using Apache Airflow to integrate data from AWS S3 into Redshift.
- Built time series forecasting models (Prophet, ARIMA) to predict multi-product sales across regions.
- Developed interactive Tableau dashboards to visualize forecast accuracy and support strategic inventory decisions.

Patient Readmission Risk Predictor
Role: Data Engineer
Tools: Python, SQL, PySpark, Apache Hive, scikit-learn, pandas, Streamlit

- Processed large-scale healthcare datasets using PySpark and Hive to build efficient machine learning pipelines.
- Trained classification models to predict patient readmission risk with over 85% accuracy.
- Built a Streamlit dashboard to help healthcare providers assess patient risk scores and make informed decisions.

Real-Time Twitter Sentiment Analyzer
Role: Data Engineer
Tools: Python, Apache Kafka, Apache Spark Streaming, PostgreSQL, Power BI, NLTK, pandas, matplotlib

- Built a real-time data pipeline using Kafka and Spark Streaming to process live tweets for sentiment analysis.
- Applied NLP techniques with NLTK and pandas for text preprocessing and sentiment scoring.
- Stored results in PostgreSQL and developed a Power BI dashboard to visualize trends by topic and region.