

EDUCATION

Master of Science in Business Analytics <i>University of Massachusetts, Amherst</i> <i>Relevant Coursework: Python for Data Science, Databases: Modelling and Theory, Data Science for Business, Healthcare Analytics</i>	Sep 2022 - Dec 2023 CGPA 3.7/4
Bachelor of Engineering in Information Technology <i>University of Pune, Pune Institute of Computer Technology, Pune, India</i>	Aug 2017 - Jun 2021 CGPA 9/10

PROFESSIONAL EXPERIENCE

Process Improvement Analyst, (Data Platform Team) – Sanford Health, Remote • Automated CI/CD workflows for ETL processes using Python & AWS Glue for data integration across sources, reducing deployment time by 40% • Designed ETL pipelines for migrating data from multiple ERPs (Tecsys, Lawson , etc.) to Redshift data warehouse, ensuring system-wide consistency • Designed multiple real-time data linkage processes for integrating data sources and enabling accurate and low-latency reporting for supply chain metrics	Jul 2024 - Present
Student Research Assistant – University of Massachusetts, Amherst • Led development & maintenance of a Python-based data monitoring pipeline using Airflow to track files received from UMass School of Agriculture • Automated archival of older files in the HANA transport folder using Apache Airflow, ensuring streamlined data hygiene & optimized storage	May 2023 – Aug 2023
Data Engineer – AK Fiserv, India • Built data pipelines for ingesting data from multiple sources(Workday, Peoplesoft, Salesforce, third-party vendor APIs) and load it into Snowflake using Spark , Python, AWS(S3, EC2, Lambda, Glue) , and Snowpipe , etc • Led cost-saving initiatives by implementing query tags in Airflow DAGs via Databricks & optimizing Snowflake queries, reducing monthly costs by 30% • Migrated processes from username-password to key-pair authentication for all Spark & Snowflake connectors that reduced vulnerabilities by 80% • Collaborated with cross-functional teams to discuss Data Profiling requirements and create Source-To-Target-Mapping (STTM) documents	Mar 2022 – Aug 2022
Data Engineer (GET)- Vodafone, India • Engineered scalable data pipelines for 30M+ real-time Kafka messages, transforming & loading into Snowflake tables and AWS storage with Prefect , cutting runtime by 65% • Implemented a proactive alerting system to swiftly identify server downtime and monitor crucial server metrics, reducing downtime by 80% • Engineered data transformation workflows to extract 15+ key performance indicators (KPIs), finding critical insights for customer churn cycle	Jul 2021 – Feb 2022
Machine Learning Intern – Yibe, India • Developed an image captioning model using encoder-decoder architecture and transfer learning in TensorFlow , achieving a BLEU-1 score of 0.66 on the Flickr30K dataset by integrating InceptionV3 for precise image feature encoding	Sep 2020 - Mar 2021

PROJECTS AND RESEARCH

Predicting Chess Games, Airflow, Spark, GCP, Databricks, MongoDB • Designed an automated ETL pipeline using Airflow to pull 3M rows of chess game data from Chess.com APIs, processing this data on GCP buckets • Transferred stored data from buckets to MongoDB clusters and connected it to a Databricks cluster to develop classification models, leveraging SparkMLlib to classify game outcomes and surpassing traditional ELO benchmark accuracy by 2.11%	Dec 2023
GenAI PDF Summarization and Interaction with ChatGPT, Python, OpenAI API, LangChain, RAG, Flask • Built GenAI PDF-GPT, a web application using OpenAI’s ChatGPT API and Retrieval-Augmented Generation (RAG) for advanced document analysis • Extracted and summarized critical insights from uploaded PDF files by using ChatGPT’s robust summarization capabilities for quicker data access • Implemented a chat interface for user interaction with extracted PDF information and used LangChain for efficient information retrieval from PDFs	May 2024
Multimodal Resume Analysis, Python, TensorFlow, PyTorch • Developed an advanced resume analyzer capable of extracting and analyzing text, images, and audio data from resumes and multimedia presentations • Engineered a TensorFlow -based pipeline to preprocess multimodal data for real-time classification , achieving processing of upto 500 resumes/minute	Aug 2023
Transliteration of Sanskrit Using Natural Language Processing, Python, NLP, Neo4j • Designed a platform that understood Sanskrit & performs NLP -based transliteration from English to Sanskrit & performs semantic relation analysis • Fabricated a complex WordNet using Graph Database for the storage of 65,000 Sanskrit words & built a transliteration module for converting given Sanskrit text in English script to its Devnagiri equivalent	Jan 2023

PUBLICATIONS

• "Behind the Scenes at Lyft: How Airflow and Flyte Power Data Workflows", Medium	Jan 2025
• "Applications of IOT in E-toilet Management System", International Journal of Emerging Technology and Research, ISSN:2349-5162	Sep 2021

CO-CURRICULARS

• Scholarship Reviewer, Grace Hopper Celebration (GHC) 2023 AnitaB.org	May 2023
• Volunteer , Voices of Data Science (CICS) – University of Massachusetts, Amherst	Aug 2023
• Volunteer , Isenberg Women in Business, University of Massachusetts, Boston	Oct 2023

CERTIFICATIONS

Google Data Analytics Professional Certificate, Google, Coursera | **Advanced Google Analytics**, Google Analytics | **Deep Learning Specialization**, deeplearning.ai, Coursera | **Python 3 Specialization**, University of Michigan, Coursera | **Introduction to Machine Learning**, IIT Madras, NPTEL

SKILLS

Languages	: Python, SQL, C++, R, HTML/CSS, JavaScript
Libraries	: Pandas, NumPy, Scikit-learn, NLTK, OpenCV, Matplotlib, Keras, Scipy
Frameworks & Tools	: Hadoop, Spark, Airflow, TensorFlow, PyTorch, Databricks, Kafka
Databases & DWs	: Snowflake, PostgreSQL, MongoDB, Neo4j, BigQuery, Delta Lake
Cloud Platforms	: AWS (S3, EC2, Lambda, SageMaker, Redshift, Glue), GCP
Others	: Docker, Jenkins, Tableau, Power BI, JIRA, dbt, LangChain, GenAI