DEEPIKA BALASUBRAMANIAN
Boston, MA | balasubramanian.de@northeastern.edu | (857) 260-8562 | GitHub | Portfolio | LinkedIn

EDUCATION

Northeastern University *Boston, MA*
Master of Science in Information Systems (Data Analytics and Machine Learning) *Dec 2023*
Relevant Courses: Data Science, Data Management Database Design, Big Data Systems, Machine Learning in Fintech

TECHNICAL SKILLS

Languages:	Python, SQL, R, T-SQL, Shell/Unix Scripting, JavaScript, Java, Flask, Scala, HTML, CSS, Restful API
Databases:	Relational Database: MySQL, Microsoft SQL Server, PostgreSQL, NoSQL: MongoDB
Libraries:	NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, NLTK, Plotly, Tensor-flow, Beautiful Soup, PyTorch
Framework:	AWS, GCP, Databricks, PySpark, BigQuery, Airflow, Snowflake, Spark, Jupyter Notebook, Git, Jenkins
Data Visualization:	Tableau, PowerBI, Quicksight, Google Data Studio (Looker), Pipeline Pilot, MS Office
Domains:	Data Architecture, Data Warehousing, Data Wrangling, Data Lake, Data Mining, Data Science, Prompt Engineering, Artificial Intelligence, Predictive Modeling, LLM, MLOps, A/B Testing

WORK EXPERIENCE

Dassault Systemes *Waltham, USA*
Software Engineer Intern *Jan 2023 – Aug 2023*
(Pipelines, Dashboarding, Automation, Python, JavaScript, Jenkins, Full-Stack, Linux, GitHub, Debugging, Performance Testing)

- Implemented a new monitoring and reporting system into CI/CD workflow, enabling real-time generation of operational Key Performance Indicators (KPIs) resulting in improved data visibility and decision-making across teams.
- Created interactive dashboard integrated with data pipeline, generating daily reports defining 20+ key indicators, leading to faster reporting from 1+ hour to a few seconds and comprehensible results.
- Automated data migration process within Jenkins by configuring shell scripts, ensuring efficient transfer of log files from source to target systems by conforming to hierarchical structures.
- Engineered and optimized python scripts to parse unstructured data to effectively store them into MySQL database and analytical tools which improved root cause analysis process by 50%.
- Developed an internal product delivery dashboard using JavaScript, Python, integrated with LDAP server to enable seamless tracking of product metrics thus saving 1 week of manual work and successfully deploying it on a Red Hat Linux server.
- Built robust backend API services by creating functions, views and stored procedures using T-SQL that reduced code redundancy and increased data reusability from MySQL database.

Dassault Systemes *Chennai, India*
Industry Process Consultant *Sept 2020 – Aug 2021*
(Automation, Machine Learning, Time Series Analysis, Data Analysis, Statistical Inference, Data integration, Communication)

- Automated complex statistical data analysis process by orchestrating ETL data pipelines using python that extracted, analyzed, and reported drug-related data from raw texts, saving 35 manual hours per week for Bio pharmacist team.
- Implemented Time Series Forecasting for COVID-19 data of 6 months using ARIMA, SARIMAX, and prophet contributing to resource allocation and planning. Created reports to visualize patterns, seasonality, and data insights.
- Optimized backend infrastructure of drug discovery tool by developing and integrating a python script that validated observational data, thereby improving decision-making capabilities by 50%.
- Collaborated with cross-functional teams to streamline report generation processes, reducing report turnaround time by 30% and ensuring prompt response to 100% of ad-hoc requests from critical stakeholders.

ACADEMIC PROJECTS

Cloud Data Architecture *(ETL Development, AWS [EC2, S3, Bucket, Lambda, Redshift, IAM, Quicksight], Data Visualization)*

- Built and automated a python ETL process using Airflow on EC2 instance that extracts data from Zillow Rapid API. Utilized Amazon S3 Bucket and Redshift to load and transform data by triggering series of lambda function.
- Connected Redshift cluster to Quicksight to visualize real-time property sales factor.

YouTube Spark Big Data Analysis *(Data Architecture, Data Modelling, Data Integration, Data Modeling, Business Intelligence)*

- Developed a big-data processing system in python and Scala, to ingest, store, and preprocess large volumes of data from YouTube API. Combined the streamed data with Kaggle datasets for representing key data insights.
- Integrated system with Apache Spark, and applied regression algorithms from Spark MLlib to forecast like/view counts based on trending genres.

LLM Enhanced Student Catalog *(LLM, Google Palm, Langchain, FAISS, Prompt Engineering)*

- Optimized Google Palm LLM to develop teaching resources for answering questions from students for a particular course.
- Hugging Face Instructor embeddings were used and utilized FAISS toolkit to store embeddings in Vector Database for later use.

Financial Stock Analysis *(Financial Data Analysis, Classification, Regression, Predictive modeling, Statistics, Tableau)*

- Conducted in-depth analysis of Tesla stock by leveraging machine learning algorithms including Linear, Logistic Regression, XGBoost, decision trees, along with Kalman, Fama-French filtering techniques, achieving 98% accuracy in predictive modeling.
- Utilized moving average strategies to produce buy and sell trading signals, yielding a 15% profit increase.
- Visualized the findings and model performance using Tableau dashboard for informed decision-making.