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

EDUCATION

Northeastern University *Boston, MA*
Master of Science (MS) in Information Systems Dec 2023
Relevant Courses: Data Science, Data Management Database Design, Big Data Systems, Machine Learning in Finance, Web Development, Application Development Engineering

Anna University *Chennai, India*
Bachelor of Engineering (BE) in Computer Science
Relevant Courses: DBMS, Information Retrieval, Algorithms, Distributed Systems, Networks, Cloud Computing

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, SQL, Shell/Unix Scripting, R, C++, Scala, Java
Web Technologies: Flask, REST, HTML5, CSS3, DOM, JSON
Databases: MySQL, Microsoft SQL Server Management Studio, PostgreSQL, Firebase, Spark, Hadoop
Libraries: NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, NLTK, Plotly
Framework: AWS Lambda, EC2, S3, RedShift, Databricks, PySpark, BigQuery, Airflow
Tools: Tableau, PowerBI, Pipeline Pilot, Jupyter Notebook, Git, MS Office 365, Advanced Excel, Jenkins, UFT
Domains: Data Analysis, Data Management, Data Wrangling, Data Scraping, Data Visualization, Data Science, Machine Learning, Statistical Inference, MLOps, A/B Testing, Hypothesis testing

WORK EXPERIENCE

Dassault Systemes - *Boston, USA* Jan 2023 – Aug 2023
Software Engineer Co-op
(Python, MySQL, JavaScript, HTML/CSS, Flask, Shell Scripting, Jenkins, Secure FTP, Linux, GitHub, Performance Testing)

- Built an end-to-end automated data pipeline for consolidating Jenkins job statuses in real time, thereby reducing 60% of the monthly manual workload and saving approximately 40 hours per month.
- Developed a Jenkins metrics dashboard, to analyze built statuses, job execution time, and resource utilization, leading to 30% improvement in visibility of job status across teams.
- Created a text parser using python to extract key performance metrics (KPI) from Jenkins logs, effectively stored them into MySQL database and analytics platforms with 99.9% uptime, which enhanced root cause analysis process by 20%
- Designed and developed release management application using Python Flask framework, integrating organization's LDAP server and deployed it on Red Hat Linux server using wsgi, achieving 40 hrs manual effort reduction for team.
- Implemented backend API services using Flask and SQL triggers that retrieves and reuses the most up-to-date release checklist from MySQL database, resulting in a savings of 20 hours of manual work.

Dassault Systemes - *Chennai, India* Sept 2020 – Aug 2021
Data Analyst
(Python, R, RegEx, Time Series Analysis, JavaScript, HTML/CSS, Pipeline Pilot, Notebook)

- Automated comprehensive biological 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 and seasonality.
- Optimized back-end infrastructure of drug discovery tool by developing and integrating a python parser that validated observational data, thereby improving decision-making capabilities by 50%
- Successfully 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 Pipeline *(AWS [EC2, S3 Bucket, Redshift, QuickSight], python)*

- 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.

Customer Churn Analysis *(SQL, Python, Tableau)*

- Implemented data cleaning, preparation and analysis on Maven Telecom data using SQL and identified 6 important customer retention strategies by performing thorough Exploratory Data Analysis.
- Utilized Tableau to visualize key churn indicators and churn profile which helped in data-driven decision making.

Indian Premier League Prediction *(Data Processing, Feature Engineering, Predive Modeling, Beautiful Soup, Hypothesis testing)*

- Performed Exploratory Data Analysis (EDA) for the IPL dataset with merged web-scraped player transfer data.
- Forecasted the winning probability of a team by analyzing their previous records and player's performance metrics.
- Performed Hypothesis testing on the statistical difference by considering alpha to be 0.05.