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, MongoDB, Firebase, Spark, Hadoop
Libraries: NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, NLTK, Plotly
Framework: AWS Lambda, S3, Databricks, PySpark, BigQuery, Airflow, Mage
Tools: Tableau, PowerBI, Pipeline Pilot, Jupyter Notebook, Git, MS Office 365, Advanced Excel, Jenkins, UFT
Domains: Data Analysis, Data Management, Data Scraping, Data Visualization, DevOps, Machine Learning, 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)
- Engineered 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 25 hours per month.
- Built an interactive Jenkins monitoring dashboard, to visualize metrics, including built statuses, job execution time, and resource usage, leading to 30% improvement in visibility of job status across teams.
- Created a text parser using python to extract key metrics from Jenkins logs, effectively integrating them into MySQL database and analytics platforms with 99.9% uptime, which helped improve data accuracy and usability 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 framework 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, RegEx, Statistical Analysis, JavaScript, HTML/CSS, Pipeline Pilot, Time Series Analysis, Notebook)
- Established data pipelines to automate manual data analytics workflow across various data sources. Developed python scripts to automate experiment result validation, saving 20 hours per week for scientific team.
- Created interactive dashboard to visualize time series plots of COVID-19 data enabling informed decision-making for scientists.
- Successfully applied ARIMA modeling techniques to produce forecasts with minimal error and accuracy of 85%
- Formulated strategies for expanding parsing protocol's capabilities by deploying it as web services for critical stakeholders. Enhanced user experience of application, achieving reduced latency of under 60 milliseconds.

ACADEMIC PROJECTS

Uber Data Analytics Oct 2023 – Present
- Built an end-to-end automated ETL pipeline using mage on Google Cloud Platform. Performed dimensional modeling on large volume of customer data and utilized Big Query for analyzing trends in high traffic areas.
- Designed dashboard using Looker Studio to visualize the insights from data.

YouTube Trending Video Analysis Oct 2022- Dec 2022
- Developed a big-data system in python and Scala, to ingest, store, and preprocess data utilizing YouTube API.
- Integrated system with Apache Spark, and applied regression algorithms from Spark MLlib to forecast like/view counts based on trending genres.

Financial Stock Analysis Jul 2023
- Conducted in-depth analysis of Tesla stock by leveraging machine learning algorithms including regression, 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.

Indian Premier League Prediction Apr 2022
- 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.

Movie Recommender Systems Dec 2021
- Implemented movie recommendation system using cosine similarity to identify closely related movies based on movie descriptions, genres, titles, and related attributes. Top 5 movies with 80% similarities more are selected.
- Successfully deployed and hosted the system on Heroku.