

ARITRA KUMAR LAHIRI

+1 480-277-2882 || Detroit, MI || lahiri.aritra@gmail.com

LinkedIn: <https://www.linkedin.com/in/aritrakumarahiri/> GitHub: <https://github.com/Aritra23>

Portfolio: <https://aritra23.github.io/>

Summary

4+ years of experience in developing scalable web applications using Java, TDD, Agile principles. Proficient in Python with hands on experience in Data Science domain applying Machine Learning algorithms for predictive data analysis.

Technical Skills

Programming Languages: Java, C#, C++, C, JavaScript, Python, R

DBMS: Oracle, MS-SQL, MySQL, MongoDB, Postgres, Hive

Technologies: J2EE, D3, HTML, CSS, Sass, Angular 6, XML, SOAP, REST, Bootstrap, JSP, JSON, WSDL, SOAP, JAX-WS

Tools/Frameworks/Libraries: Spring MVC, Hibernate, Spring Boot, AWS S3, Selenium, ASP.NET, Spark, Node.JS, Git, Tomcat, Anaconda, Jupyter Notebook, Pivotal Cloud Foundry, LDAP, ADFS, OAuth2, Pandas, NumPy, Matplotlib, Docker

Professional Experience

Data Engineer - Ford Motor Company

May 2019 – Till Date

- Developed data ingestion pipelines to process over 300 TB of data utilizing HDFS, Hive, Pig, Spring Batch
- Optimized existing ETL pipelines and initiated fine tuning of Hive queries
- Scheduled cron jobs using Spring Scheduler for handling incoming data streams, Rest Api for integrating Python ML model with Alteryx workflows and micro-services for response sharing with callback url. Deployed containerized app with Docker.
- Productionized application with IIOT platform saving \$6M per year and won Tech award for innovation as part of team

Software Development Engineer – NCS Pearson

June 2016 – May 2019

- Implemented Service Oriented Applications using web services and developed REST APIs utilizing Spring and Hibernate
- Designed and implemented Spring Boot Microservices for search and indexing data into backend MongoDB.
- Analyzed and implemented POC to gauge regional expenditures in K-12 schools utilizing Python libraries.
- Automated test scripts in Java using Selenium Web Driver, Gradle, TestNG drastically reducing manual resource hours.

Full Stack Developer Intern – Biodesign Institute, Arizona State University

May 2015 – Aug 2015

- Prototyped a health recommender MEAN stack application that maps disease risk factors with biosensor device features and generated a mapping algorithm to prescribe devices based on patient profile.

Academic and Research Projects

Sentiment Analysis on Game of Thrones Season 8 Tweets

May 2019

Analyzed sentiments of different characters by calculating polarity scores and predicted season outcome. **Tools:** Python, Pandas, Tweepy API, Textblob, Scikit learn ML library, NLTK, TF-IDF vectorizer

Spotify Data Visualization

Apr 2019

Analyzed playlist Dataset to extract, visualize top artists, tracks by each region. **Tools:** JvectorMaps, Python - Pandas, Plotly.JS

Manipulate Document Object Model to filter dataset based single or multiple search categories using Python, JS.

Mar 2019

Mission to Mars

Mar 2019

Implemented Flask App to render scraped Mars related data from NASA website and display the information in a single page application. **Tools:** BeautifulSoup, Pandas, Splinter, PyMongo and Flask.

Explored bellybutton biodiversity dataset to build Interactive dashboard using Plotly.JS, D3 and Flask App

Feb 2019

Inspection Ratings of Chicago Restaurants using ETL pipeline

Feb 2019

ETL on Chicago food inspections dataset using Pandas - data munging, SQL, Python – transformation, Mongo, Flask App – load

Data Analysis on Video Game and Election Voting Dataset to derive trends and anomalies using Python

Dec 2018

Geo-Spatial Operations using Apache Spark Cluster in HDFS

Jan 2015 - May 2015

Performed geo spatial operations like Spatial range, join query, convex hull in distributed Apache Spark framework using HDFS. Experimented across multiple nodes with large datasets. **Tools:** Java, PostgreSQL, Spark Java API and Java RDDs, Ganglia

Implementation of Data Partitioning techniques, Query processing operations

Jan 2015 - May 2015

Analyzed Movie-Lens dataset with 10M records to implement range and RR partitioning and implemented parallel sort and parallel join operations on partitioned tables. **Tools:** Python and PostgreSQL

Education

Data Science Analytics (Certificate Course) University of Arizona, Phoenix, AZ, USA

May 2019

MS in Computer Science – Arizona State University, Tempe, AZ, USA

May 2016

B-Tech in Computer Science and Engineering, West Bengal University of Technology, India

July 2013