Manasi Dalvi

Greater Boston MA 02120 |manasidalvi14@gmail.com|857-320-9099|github.com/dalvimanasi

SKILLS

Languages	Java, SQL, Python, Beginner in Scala, JavaScript
Web Technologies	HTML5, CSS3, Bootstrap, JQuery
Database	MySQL, Neo4j, MongoDB
Frameworks/Libraries	HDFS, MapReduce, Spark, Pandas, Numpy, Scipy, Airflow, Luigi
Tools	Eclipse IDE, IntelliJ, AWS, Docker, PowerBI, Tableau, Azure ML, Git

EDUCATION

Northeastern University, Boston, MA — *Master of Science: Information Systems* — **Dec 2017**

University of Mumbai, India — *Bachelor of Engineering: Information Technology* — **July 2012**

ACADEMIC PROJECTS

Network Analysis of YouTube Video Network **Aug 2017**

- Created a network of YouTube videos on Neo4j. Implemented PageRank and Centrality Algorithms using Apoc libraries. Used Spark on AWS EC2 cluster and containerized the application using Docker.
- Performed views estimation for videos using Random Forest and formed profile based clusters of uploaders using *k-means* clustering.
- Performed Map-Reduce using Hadoop for aggregating user profile, deployed Flask app on IBM Bluemix.

Prediction and Classification (Freddie Mac's Single-Family Loans dataset) **July 2017**

- Predicted interest rates with Random Forest and classified loans (delinquents) with Logistic regression
- Performed feature selection using Recursive Feature Elimination, tested models during financial crisis, and economic boom period.
- Deployed a Flask app on IBM Bluemix, displayed the results as machine learning as service using Microsoft Azure ML Studio

Exploratory Analysis and Database as Service (ETL Pipeline) **June 2017**

- Performed web scraping using BeautifulSoup, and developed rest API for Zillow dataset, provided location based and parameter based searching options. Built Airflow pipeline to extract, transform and load data.
- Built pipelines using Airflow and Luigi to automate data ingestion in MongoDB cloud(MLABs) and persisting backup in S3. Exploratory analysis using numpy, matplotlib, scikit-learn, pandas libraries.

Santander Product Recommender using Zeppelin, Scala and MLlib in Spark to recommend products to existing customers according to their current products and profile**. (**Kaggle completion). **Oct 2016**

Machine Learning project on the US-pollution and US-drought datasets, to find a correlation between the pollutant type and the drought type, using Hadoop and Map Reduce to summarize, aggregate and join the two datasets to perform exploratory data analysis.

UNSW network data classification as malicious and non-malicious using Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes and K-Nearest Neighbours **Nov 2017**

WORK EXPERIENCE

Project Coordinator, **Shezartech Pvt Ltd, India** (E-Learning Solutions company) **Jan 2015 –Aug 2015**

- Assisted the project managers to plan, design and execute projects. Worked as a single point of contact from initiation to completion of the project. Procured low cost vendors, increasing the cost efficiency by 12%.
- Analyzed the overall data of our company, using SQL, and derived patterns which helped my team to identify bottlenecks, resulting in efficient and easy progress for the projects.

Junior Bigdata Engineer, IntellectBrains Inc, Chicago **Jan 2016 – May 2016**

- Assisted in the ETL pipeline creation to download and load datasets from BigQuery, into Hadoop system.
- Performed data transformation and cleaning using Hive and Spark and created a Zeppelin notebook to showcase the results.