Mansi Borole

mansi.borole@rutgers.edu | +1 732-374-5143

linkedin/mansi-borole | github/MansiGit | mansigit.github.io

EDUCATION

Master of Science in Computer Science, Rutgers University, New Brunswick, New Jersey   *Sep 2022 - May 2024*
Coursework: Algorithms, Linear Algebra, Massive Data Mining, Artificial Intelligence   *GPA: 3.83/4.0*
Bachelor of Engineering in Computer Science, University of Pune   *Aug 2015 – Jul 2019*
Coursework: Data Structures, Database Management, Web Technology, Machine Learning   *CGPA: 8.61/10*

SKILLS

Programming Proficiency: Python, SQL, Java, Shell Scripting
Big Data Hadoop: HDFS, Apache Spark, Pyspark, NiFi, YARN, Cloudera Data Platform, Enterprise Data Ingestion and Warehousing
Database platforms: PostgreSQL, MySQL, Apache Hive, HBase
Data Visualization: Tableau, Grafana, QlikSense, Matplotlib, Microsoft Excel (Pivot Tables, Pivot Charts, V lookups)
Libraries/Frameworks: Flask, Django, Node.js, Keras, Tensorflow, Pytorch, Numpy, Pandas, Docker, git
Skills: Hadoop Administration, DevOps, Agile Development, Problem solving, Big Data Analytics, Stakeholder Management

---

PROFESSIONAL WORK EXPERIENCE

HSBC Bank | *Software Engineer (Big Data)*   *Jul 2019 – Aug 2022*

- **Cluster Virtualization**: Migrated Hadoop Master services (Name nodes, YARN, Solr, Zookeeper, Hive) from on-prem servers to Virtual Machines, achieving a **cost avoidance of $0.7M** on buying new expensive servers while saving 400 man hours in maintenance upgrades using Ansible Automation
- **Capacity Management**: Performed Hadoop/Linux administration provisioning space on clusters by building and decommissioning servers, Node repurposes from one cluster to another and installing Hadoop components and services on the nodes
- **Nifi Dataflow**: Set up a one-stop dashboard BigData360 and data-flows to monitor Hadoop services and alerts on clusters using Ambari Metrics/ Cloudera Metrics APIs that **reduced the mean time to repair failures by 1 hour**
- **Ansible Automation**: Created Ansible playbooks and **python scripts** to get daily FS image from clusters and find the daily organic growth of client's HDFS data on the clusters
- Created a MySQL tool to generate Chargeback reports to determine the cost of Physical Hardware, Licenses, and Supporting Staff for each of **2000+ servers** in the estate based on resource usage and used Qlik Sense to publish the reports.
- Wrote housekeeping scripts that would reduce replication factor of cold data on the cluster effectively saving 45 TB/month

HSBC Bank | *Communications Leader*   *Jul 2019 – Aug 2022*

- Lead a team of 4 colleagues delivered high-quality, impactful quarterly newsletters highlighting project success stories and KPIs from the 'Group Data Technology' Department, closely working with Ex Co members in US, UK and Asia-Pacific regions
- Published team's quarterly achievements and highlights via Newsletters to the entire department

---

CERTIFICATIONS & ACHIEVEMENTS

- **Tableau** Certified Desktop Specialist   2022
- Certified Associate in **Python** Programming   2021
- Statistics with Python – Specialization – University of Michigan   2021
- IBM Foundation Blockchain Developer   2019

RELEVANT PROJECTS

Tableau Data Visualization   2022

- Analyzed and visualized Credit Card Complaints Trends, Netflix Movies analysis, Salary Insights in NY State showcasing Trend predictions, Density maps, Scatter Plots, Timeseries. Checkout my portfolio on Tableau Public
- Animated running bar chart for demonstrating country wise COVID 19 positive cases over the span of last year using Tableau Desktop.

Machine Learning   2022

- AirBnB Price Prediction: Used linear regression, XGBoost and multivariate adaptive regression splines (MARS) to predict prices based on city, availability, reviews per month and room type.
- Image Similarity Search using a pretrained Convolutional Neural Network (CNN) model and then using KNN classifier to find similar images. Achieved an accuracy of 98%
- Plant Seedings Classification: Trained an Inception-ResNet v2 deep neural network using ADAM optimization with TensorFlow and Keras, achieving an accuracy of 97%.