

# MANSI BOROLE

[mansi.borole@rutgers.edu](mailto:mansi.borole@rutgers.edu) | +1 732-374-5143 | [Linkedin/mansi-borole](https://www.linkedin.com/in/mansi-borole) | [github/MansiGit](https://github.com/MansiGit) | [mansigit.github.io](https://mansigit.github.io) | New Brunswick, NJ

## EDUCATION

**Rutgers University**, New Brunswick, New Jersey

Aug 2022 - May 2024

Master of Science in Computer Science

GPA: 3.85/4.0

Coursework - Algorithms, Linear Algebra, Massive Data Mining, Artificial Intelligence, Design of Internet Services, Deep Learning

**University of Pune**

Aug 2015 - Jul 2019

Bachelor of Engineering in Computer Science

GPA: 8.61/10

## SKILLS

**Programming Proficiency:** Python, SQL, Java, C++, Shell Scripting

**Big Data Hadoop:** HDFS, Apache Spark, Pyspark, Scoop, ETL, AWS, Enterprise Data Ingestion and Warehousing

**Database platforms:** PostgreSQL, MySQL, Apache Hive, Oracle

**Data Visualization:** Tableau, Grafana, Matplotlib, Microsoft Excel

**Libraries/Tools:** Numpy, Pandas, NoSQL, Tensorflow, Pytorch, Node.js, Docker, Ansible, Powershell, git

**Other Skills:** Stakeholder management, Agile Workflow, JIRA, Problem-solving, Technical writing, Presentation

## PROFESSIONAL WORK EXPERIENCE

**HSBC Bank | Software Engineer (Big Data Analytics)**

Jul 2019 - Aug 2022

- 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.
- Data Visualization and Reporting: **Developed a BI tool "BigData360"** using **Grafana** for real-time monitoring and management of 58 Hadoop production and dev clusters. This provided visibility and control over HDFS data and cluster utilization, allowing clients to optimize usage and make informed decisions.
- Cluster Virtualization:** Migrated Hadoop Master services (Name nodes, YARN, Solr, Zookeeper, Hive) from on-premise servers to Virtualized machines using **ESXi hypervisor**, achieving a **cost avoidance of \$0.7M** on buying new expensive servers while saving 400 man hours in maintenance upgrades using Ansible Tower
- Developed **ETL pipelines** and scripts in python to Load Data and archive data. Leveraged Erasure coding to achieve savings of 50% in total storage space
- 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 200+ servers
- Ansible Automation:** Created Ansible playbooks and python scripts to build new RedHat servers with Hadoop capabilities
- As a communication team lead, led a team of 4 to deliver impactful newsletters, highlighting project success KPIs, working with global Ex Co members.

## RELEVANT PROJECTS

**Exploring Business Intelligence - Data Visualization** [Tableau Public Portfolio](#) ↗

- Analyzed and created interactive dashboards for case studies in Credit Card Complaints Trends, Netflix movies analysis, Salary Insights in NY State showcasing the Trends, Predictions, Density maps, Scatter Plots and Timeseries Charts. Used features such as filters, drilldowns, and tool tips to enhance the UX and make it easy for them to interact with the data

**Machine Learning**

**Pytorch, Tensorflow, Keras, Pandas**

- [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%
- Top 10 Friend recommendations with pySpark using MapReduce jobs on a dataset of 49995 records.

**Fast Trajectory Replanning using AI Agent**

**Python, Matplotlib**

- Evaluated and compared the model performances (paths) generated by 4 flavors of A-star search algorithms like Repeated A-Star, forward/backward A-star, Adaptive A-star [Project Link](#)

**Data Analytics**

**HTML, NodeJS, Gulp, JavaScript**

- Implemented Customer engagement tracking using location heatmaps and website interaction using Google Analytics/Google Tag Manager [Project Link](#)

## CERTIFICATIONS

- Tableau Certified Desktop Specialist (2022)**
- Certified Associate in Python Programming (2021) PCAP**