

Satyam Mishra

satyamgusmishra@gmail.com • 857-701-3819 • Boston, MA • [Linkedin](#) [Github](#) [Portfolio](#)

EDUCATION

Northeastern University - Boston, USA

Aug 2020 - May 2022

Master of Science, Data Analytics Engineering

Relevant Coursework: Data Mining, Data Management, Statistics, Machine Learning in Finance, Deep Learning

Medicaps University - Indore, India

Aug 2016 - Jul 2020

Bachelor of Technology, Computer Science Engineering

Relevant Coursework: Python Programming, NLP, Data Visualization, Discrete Mathematics, Economics & Finance

TECHNICAL SKILLS

Programming: Python (Pandas, Numpy, Matplotlib, Seaborn, NLTK, PySpark, Flask), R, C Programming

Tools/Frameworks: AWS, GCP, Cloud SQL, EC2, Lambda, Vault, Hadoop, Data Modeling, Git, CI/CD, Docker, Alteryx

Databases: MySQL, Snowflake, Redshift, S3, Hadoop, NoSQL, PySpark, Kafka, SSMS, Airflow, OLAP, MapReduce

Analytics: Tableau, EDA, Hypothesis Testing, A/B Testing, Market-Basket Analysis, Statistics, Dimensional Modeling

PROFESSIONAL EXPERIENCE

TRANSUNION

Aug 2022 - Feb 2024

Data Engineer

Boston, Massachusetts

- Implemented 7 scalable **ETL** pipelines with 99.8% uptime using **Python**, **Kinesis**, **Snowflake**, **Redshift** and **EC2**.
- Optimized **Spark** (databricks) pipelines using executor optimizations, reducing ETL time by 80 minutes.
- Developed and deployed custom **Kafka** consumers to process high-velocity credit data streams from various sources, including banks, lenders, and agencies.
- Programmed **AWS Lambda** with API gateway in Python to initiate triggers to calculate results on 3 events, reducing time taken for insights extraction by 25%.
- Engaged in **System Design** and **Dimensional Modeling** huddles, and documented them for 4 big data clients.
- Managed a data pipeline while complying with GLBA and CCPR for a data source catering 300+ million users.
- Wrote **optimized SQL** queries with normalizations, CTEs, stored procedures for **OLAP** scripts.

UNUM GROUP

Jul 2021 - Dec 2021

Data Engineer and Analytics Co-op

Portland, Maine

- Automated a pipeline to stage data from **AWS S3** to Redshift data warehouse and connecting to Tableau for real-time analysis of user finance health based on PHQ-9 and GAD-7 scores, aiding business intelligence.
- Assisted in writing cron jobs in **Python** (DAG) to trigger operations every 12 hours with **Airflow**, saving 20% time.
- Leveraged Python (NumPy and Pandas) for data transformation during ELT and ETL pipeline stages.
- Developed **REST API** integration POCs for seamless data transfer and integration, saving 8 hours weekly.
- Performed **data quality checks** using Pandas, then EDA in Tableau & statistical analysis on 540000 records.

ELETECH ENGINEERING

Jan 2020 - Jun 2020

Data Science Intern

Indore, India

- Engineered 5 KPIs on logistics data having 1 Million records and created 4 **Tableau** dashboards, saving Rs 210k.
- Harnessed **K-Means clustering** for segregating faulty components on impact metrics, improving it by 18%.
- Implemented Mean Variance Optimization and Conditional Value at Risk algorithm using **SciPy** module.
- Developed interactive Data studio reports with **GCP Cloud SQL**, used Sharpe ratio and Value at Risk as metrics.

AAM AADMI PARTY

Jul 2019 - Sep 2019

Data Analysis and Machine Learning Intern

Delhi, India

- Coded BeautifulSoup and Selenium script to extract data of 70 assemblies from Election Commission Website.
- Devised XGBoost and Ridge Regression models on electoral data of 15 years with an accuracy of 84% and 76%.

AWARDS AND ACHIEVEMENTS

- First Runners Up in Void Hacks Hackathon in Data Science Track, SVVV (April 2019)
- Second Runners Up in Version Beta Hackathon, NIT Bhopal (October 2018)
- First Runners Up in Convolution Coding Zonals, UIT RGPV (February 2018)