

Shivam Mishra

US | (812) 361-0960 | mishras0297@gmail.com | [GitHub](#) | [LinkedIn](#) | [Portfolio](#) | Open to relocation

WORK EXPERIENCE

Rocket Mortgage | Big Data Engineer Intern | Detroit, MI May 2023 - Aug 2023

- Implemented scalable ETL data pipelines using AWS Glue, and PySpark to automate daily ingestion of terabytes of finance, and sales datasets from sources like Snowflake and Redshift, achieving cost savings of \$10,000.
- Enhanced AWS S3 data monitoring project using Lambda, SQL and PySpark by adding critical logic for monitoring S3 buckets which improved data integrity and reduced the analytics pipeline failure by 11%.

School of Public Affairs | Data Research Assistant | Bloomington, IN Aug 2022 - May 2023

- Automated extraction and analysis of insurance claims and plans pricing data, from different sources, processing over 10TB datasets using Python and PySpark and Reduced data availability time from 5 days to 1 day.
- Gathered requirements from researchers to create informative Tableau Dashboards on insurance plan pricing data, presented key performance indicators contributing to research goals.

Ericsson | Data Analyst | Bangalore, IND Jun 2021 - Aug 2022

- Analyzed terabytes of production data using PySpark and SQL identifying errors, trends, and patterns in network reports, resulting in a 12% reduction in data errors.
- Optimized Jenkins pipeline stages for build, test, and deploy in CI/CD processes for open-source projects, including Kube State Metrics, improving product reliability and driving a 16% increase in the user base.
- Engineered Python validation scripts to validate thousands JSON and YAML files uploaded by users, preventing code failures in internal Kubernetes projects and reducing failure rates by 80%.

BlackRock | Data Analyst | Mumbai, IND Jan 2019 - Jun 2021

- Collaborated with Aladdin stakeholders to gather essential business requirements, ensuring a deep understanding of data needs and crafting effective business solutions for Aladdin clients.
- Optimized Snowflake SQL query plans and data models, reducing query execution time by 18% , supporting Aladdin data cloud and enhancing data governance.
- Automated portfolio analytics report generation with SQL and Python, reducing manual workload for senior managers.
- Developed Tableau dashboards for real-time compute farm resource visualization, increasing resource management efficiency by 25% and reducing report generation time.
- Generated statistical reports with Python and R to forecast SLA delays, guiding client teams, preventing SLA breaches, and contributing to \$50,000 savings.

PROJECTS

Stocks Recommendation System | Neural Networks, Natural Language Processing [[GitHub](#)] Aug 2023 - Dec 2023

- Developed an RNN-based stock recommendation system to predict stock prices and assist users in making better investment decisions to improve investment portfolio.
- Incorporated VADER analysis to analyze over 50,000 market news sentiments alongside prices for informed decisions. Integrated visualizations into web applications for enhanced user experience.

Creative Arts For Veterans | Python, Tableau, Hive, Qualtrics Sep 2023 - Mar 2024

- As part of Center for Rural Engagement, worked on designing Qualtrics Surveys for CAV books for veterans.
- Utilized Python and Tableau to analyze Veterans feedback data from Qualtrics Survey and MS Excel, demonstrating the positive impact of the CAV book on veterans' lives, resulting in securing a \$60,000 grant.

EDUCATION

Indiana University Bloomington Aug 2022 - May 2024

Master of Science In Data Science | GPA: 3.7 / 4

Coursework: Data Mining, Advance databases , Applied Algorithms, Statistics, Deep Learning Systems, Machine Learning

SKILLS

- **Programming Languages:** C, C++, Python, SQL, R, SAS, Perl, Java, JavaScript, HTML , Bash
- **Databases:** Sybase, Oracle, Microsoft SQL Server, PostgreSQL, MongoDB, Neo4
- **Business Intelligence:** Tableau, Power BI, MS Excel, Grafana, Databricks
- **Cloud Platforms :** AWS (S3, Glue ETL, Lambda, Athena), Azure (DevOps), Snowflake, GCP (BigQuery)
- **ETL Tools & ML:** Hadoop, PySpark, Pandas, Kafka, Kubernetes, Docker, Scikit-learn, MLlib, PyTorch

ACHIEVEMENTS

- Achieved finalist team status in Grant Thornton Idea Business Case Competition, awarded \$1000 prize.
- Secured the best project award out of 500+ university projects for pioneering smart farming with IoT.