HARSHITA MOHANRAM

Macungie, PA | Ph no: (469)-(924)-3583   Harshitamohanram13@gmail.com | https://www.linkedin.com/in/harshitamohan137/ |
https://github.com/harshita-mohan | https://harshita-mohan.github.io/HarshitaMohanram/portfolio/ |

3+ years of professional programming, problem solving and analytics experience in the field of data. Proficient in SQL, Python for querying, scripting and visualization tool such as Tableau. Hands on experience in Big Data Tools like Hive, Impala, Airflow. Ability to work independently and as a Team, creating and maintaining professional relationships. Demonstrated expertise in both quantitative and qualitative data analysis, creative problem-solving and a proactive approach to process improvements.

SKILLS

- **Programming**: Python. SQL , R, SAS, UNIX, LINUX, google analytics, Business Intelligence, Qlikview , JDA, Salesforce, Fivetran
- **Frameworks** : MLFlow, Flask, Django, GIT, Keras, Agile, Apache Spark, Hadoop, Alteryx, Linux, RStudio, MySQL , ArchiMate, JSON
- **Database Management** : SQL, Mongo DB, Cassandra DB, NoSQL, MS SQL Server, Spark, SSMS, SSIS, SSRS. A/B Testing, SQL scripts
- **Machine Learning**: : Regression, Random Forest, XGBoost, Bayesian Analysis, Decision Tres, Clustering Analysis, SVM, Microsoft Office Outlook
- **Cloud:** Azure ML Studio, AWS Sage maker, Azure Databricks, Azure DevOps, Synapse, Azure Machine Learning, Azure Data Factory Software Tools: PyCharm, VSCode, Docker, Azure Data Studio, MS Excel, MS PowerPoint, MS Word, JIRA, Tableau, Gitlab, GCP,

WORK EXPERIENCE

Logistics Data Analyst Intern - Volvo Group, Pennsylvania   **Sept 2023- Present**
- Built **Power BI** dashboards story by integrating data from **relational database** like **SQL** server, Metadata, **MS Access**, and **Terraform** (1TB) using **DAX** custom calculations to visualize business **KPIs** and optimized data workflows using **PowerApps** saving the team 10+ hours per week.
- Implementing **time series forecasting** to improve demand planning accuracy by 20%, leading to better inventory with **SDLC**.
- Implemented **machine learning models** and **Hypothesis testing** resulting in a remarkable 30% improvement in processing speed for Chassis to deliver production orders. Conducted **A/B testing**, leading to a substantial 15% increase in funnel conversion rate.
- Conducted extensive **data mapping** exercises to understand data sources, relationships, and integration points for enhanced reporting and analysis.
- Designed a system using Raspberry Pi and Jetson platform, integrating **YOLOv5**, and used Neural Network like TensorFlow, Pytorch, Etc., for real-time object detection in inventory and created website using **XML**, **JavaScript**, **CSS**.
- Created bots leveraging **Selenium** for web scraping and **REST API** integration to extract information from disparate sources, enhancing data.
- Conducted thorough **exploratory data analysis (EDA)** using advanced **statistical analysis** including **t-tests** and **Chi-squared tests** to find the root cause and **regression** analysis on large datasets for **predictive data modeling**, resulting in accurate forecasts and improved business planning.
- Worked closely with cross-functional teams to design **data structure**, **data migration** and **Agile** teams to utilize sprint techniques.

Data Engineer Intern - Digital Factory, Chicago   **Jan 2023 - May 2023**
- Migration of on premise data (**Oracle/ Teradata**) to **Azure Data Lake Storage** (**ADLS**) using **Azure Data Factory**(**ADF** V1/V2).
- Designed and developed robust data ingestion **ETL** processes using **Glue**, ensuring accurate and timely data transformations; used **S3** for data pipeline using **Java** performance and stored in data lake resulting in a 30% reduction in processing time.
- Implemented and maintained robust **data governance** policies to ensure the accuracy, integrity, and security of organizational data, while also developing and delivering **ad hoc** reports to provide insights for decision-making for an ecommerce client.
- Utilized analytical tools such as **SPSS**, **data management** and **data manipulation** techniques to extract insights from complex datasets.
- Orchestrated **streamlining reporting** workflows and ensured continuous data availability by providing expert support for data platforms, leveraging **PowerShell** and wrote **airflow** using **Python scripts** for process automation, resulting in a 30% increase in team productivity.
- Designed and deployed 15 serverless functions using **AWS Lambda**, streamlining data processing tasks and boosting system efficiency by 30%.
- Used **Azure DevOps** & **Jenkins** pipelines to build and deploy different resources (Infrastructure as code) in Azure. Ensure the developed solutions are data documented and signed off by business.
- Implemented **Git** for version control, streamlining collaboration across a team of 15 developers and reducing code conflicts by 60%.

Senior Supply Chain Data Analyst- ABB, India   **April 2020- Dec 2021**
- Sourcing using **SAP HANA** system and used **data visualization** tool to create 40+ interactive dashboards using **Tableau** for reporting to the analysis team and conducted **data mining** techniques to uncover trends, resulting in a satisfaction rating from 92% to 98% for stakeholders.
- Optimized **PySpark** and **Spark-SQL** applications for data extraction, data cleaning, and transformation, with large scale data tables resulting in a 40% reduction in processing time and improved **data accuracy**, **data quality** and **data integration.**
- Leveraged **Macros** using **VBA** to automate data processing tasks, reducing manual effort by 30% and increasing accuracy in reporting.
- Utilized **PowerPivot, VLOOKUP** for complex data modeling to support **OLAP** processes ensuring accurate datasets for analysis.
- Implemented and maintained **CI/CD pipelines** for automated deployment and continuous integration of data pipelines, reducing deployment time and increasing overall pipeline efficiency.
- Implemented **PaaS** solutions in **Snowflake**, resulting in a 20% improvement in the utilization of cloud resources
- Developed and implemented ETL pipelines utilizing Azure Data Factory, enabling efficient extraction, transformation, and loading of 2TB data from **MongoDB** to Azure Synapse warehouse, improving data processing efficiency by 40% and reducing latency by 50%.
- Migrated data from **SQL** server to **Hadoop** ecosystem using **Kafka** and **Spark** streaming, resulting in the seamless ingestion and processing of clickstream events; Implemented efficient **data warehouse structure** to optimized data storage in **HDFS**, reducing storage costs by 30% while maintaining high data availability and reliability with **Scala**.

PROJECTS

1) **Twitter Case Study- Python, Machine Learning [Scikit-Learn, Pandas, NumPy, Matplotlib] , EDA, Visualization, Statistics, Graph**
- Built a ETL Pipeline for data transformation using Azure ML and Azure Databricks. Implemented data transformations on 30GB of gaming data using Spark, creating an interactive dashboard in Tableau that visualized gaming trends.

EDUCATION

Master of Science - *Information Technology & Management, University of Texas at Dallas*   *Dec 2023, GPA-3.6*
Bachelor of Technology – *Electronics and Communication Engineer, SJB Institute of Technology*

CERTIFICATIONS & ACTIVITIES

Certified cloud engineer - Microsoft Certified: Azure Data Engineer Associate   Links
1st place in Salesforce project- Salesforce: Gods own project