VISMAYA MOHANRAM

vm@hawk.iit.edu • (872)-258-8900 • https://www.linkedin.com/in/vismayam/ • Portfolio

PROFESSIONAL SUMMARY

Experienced Data Engineer around 5 years of expertise in data migration, ETL development, and database management across cloud platforms like AWS and Azure. Proficient in SQL, Python, and automation, with a strong ability to design, optimize, and implement scalable data pipelines for seamless system integration. Adept at data modeling, validation, and performance tuning, ensuring data quality, compliance, and system efficiency. Collaborative team player, skilled in working with stakeholders, developers, and business teams to drive data-driven solutions and improve analytics infrastructure. Passionate about building robust data architectures, enhancing system performance, and solving complex data challenges in fast-paced environments.

EDUCATION

Illinois Institute of Technology, *Chicago, IL* ***Aug 2022 - May 2024***
Master of Science in Computer Science – Data Analytics | GPA: 3.6/4.0
Relevant Coursework: *Big Data Technologies, Data Preparation & Analysis, Mobile Application Development, Machine Learning, Computer Networks, Advance Database Organization, Online Social Network Analysis, Software Project Management*

The National Institute of Engineering, *Mysore, India* ***Aug 2014 - Jun 2018***
Bachelor of Engineering in Computer Science and Engineering | GPA: 3.71/4.0

WORK EXPERIENCE

Software Developer Intern | Flexday AI, *Chicago, IL* ***Sept 2024 – Present***

speciphicConnect – AI-Driven Data Migration & Optimization

- Designed and developed an AI-powered data pipeline to streamline data migration, transformation, and integration for enterprise customer support systems, ensuring seamless data conversion and compliance with industry standards.
- Led end-to-end data migration, mapping legacy system data using SQL and Python, achieving 95%+ accuracy in AI-driven query handling.
- Optimized ETL workflows, automating data ingestion and transformation processes, reducing manual intervention by 60% and enhancing system efficiency.
- Built automated data validation scripts, improving data integrity, compliance, and reducing errors by 40%.
- Integrated AI-driven sentiment analysis, structuring unstructured customer data for intelligent query classification and routing.
- Developed real-time data pipelines and interactive dashboards, delivering actionable insights and enhancing operational decision-making.
- Collaborated with cross-functional teams, ensuring scalable architecture and seamless integration with business intelligence and analytics platforms.

HR Data Sync – BambooHR & Merge.dev Integration

- Developed a Proof of Concept (PoC) to automate employee data migration, transformation, and visualization by integrating BambooHR with Merge.dev, enhancing HR data accessibility and operational efficiency.
- Migrated & structured 10,000+ employee records, ensuring 99% data accuracy through automated mapping and transformation.
- Built dynamic CRUD operations, reducing manual data entry by 50%, improving HR processing speed.
- Designed an interactive organization chart, enhancing workforce visibility and reporting accuracy by 100%.
- Implemented data validation & encryption, reducing data inconsistencies by 35% and ensuring full PII compliance.
- Optimized API integration, cutting latency by 30% and improving real-time data synchronization.
- Aligned data workflows with business needs, boosting HR operational efficiency by 20% through seamless automation.

Infosys Limited (Client: Swiss Re), *Bengaluru, India*

Data Engineer ***June 2020 – Aug 2022***

- Spearheaded the migration of over 10 terabytes of reinsurance data from on-premises servers to Azure Data Lake, resulting in a 30% reduction in data retrieval times and improved data accessibility for analytics teams.
- Designed Azure Data Factory pipelines to automate ETL processes, reducing data processing time by 60%.
- Ensured seamless integration and efficient operation of Databricks, optimizing the performance of data pipelines.
- Experienced in migrating and structuring data within Teradata and Snowflake, enhancing data accessibility and ensuring accuracy.
- Constructed complex SQL queries to extract and consolidate data from multiple sources, streamlining reporting through ETL techniques.
- Implemented advanced CI/CD techniques to streamline Python dependency updates, boosting data pipeline performance by 35% and saving the team 30 hours monthly in manual checks.
- Collaborated with over 15+ stakeholders from the product and design teams to address and resolve data-related technical challenges.

Data Analyst ***July 2018 – June 2020***

- Refined master data and executed automated clean-up processes, which enhanced data quality improvements for 10+ business processes and internal reporting.
- Implemented ETL pipelines using Apache Kafka and Apache Spark, processing over 5 million records to ensure real-time data streaming and transformation.
- Applied knowledge of SQL Server Integration Services (SSIS) for data integration and ETL processes, enhancing data pipeline efficiency.
- Improved data accuracy and reliability through SQL and Data Modeling for effective metric tracking and reporting.
- Developed interactive Power BI dashboards using expertise in DAX & Power Query. These dashboards translated complex business requirements into clear visualizations of over 20+ vital KPIs.
- Collaborated with cross-functional teams to gather requirements and generate detailed SQL queries, improving data accessibility by 35%.

TECHNICAL SKILLS

Programming Languages	**:** Python, SQL, Java
Database	**:** MySQL, PostgreSQL, Oracle, MongoDB, NoSQL
Cloud Technologies	**:** AWS (EC2, Redshift, Lambda, EMR), Azure (Databricks, Blob), S3 bucket
ML libraries and frameworks	**:** Pandas, NumPy, TensorFlow, Matplotlib, Scikit-Learn, Keras, Pytorch, Seaborn
Data Analytics	**:** Microsoft Power BI, Tableau, MS Excel, VLOOKUP, ETL pipelines, Apache Spark
Developer Tools and Platforms	**:** Git, Docker, JIRA, Agile, MS Excel, Jupyter Notebook

ACADEMIC PROJECTS (GitHub)

Social media Analysis - Reddit | *Illinois Institute of Technology* ***Jan 2024 – May 2024***

- Developed a social network analysis project using Python to analyze Reddit data. Utilized Reddit's API to extract data on relationships between the author of the post and author of comments on the post.
- Created interactive visualizations in Tableau to explore the data and leveraged PRAW library for efficient data retrieval from Reddit.

Credit Card Defaulters using Machine Learning | *Illinois Institute of Technology* ***Aug 2023 – Dec 2023***

- Developed a ML model to predict credit card defaults using consumer demographics, credit history, and payment habits data. The models are evaluated based on five performance metrics: accuracy, F-1 score, precision, sensitivity, and specificity.
- Implemented a voting classifier model achieving high precision, sensitivity, and specificity, offering a strong alternative to individual models. Evaluated various ML models to identify the most effective model for predicting default risk, voting classifier performed well in predicting credit card defaults and Random Forest performed well among the individual models achieving an accuracy of 81.3%.

Fake News Classification | *Illinois Institute of Technology* ***Jan 2023 – May 2023***

- Developed a machine learning model to accurately classify news articles as genuine or fake and developed a reliable tool for identifying and filtering misinformation.
- Analyzed performance metrics to identify the Siamese Network model as the most effective, achieving an exceptional accuracy of 95.44%.

Exploring Big Data Analytics for COVID-19 Dataset | *Illinois Institute of Technology* ***Aug 2022 – Dec 2022***

- Implemented a comparative analysis of ML algorithms on a WHO dataset, evaluating performance, speed, and accuracy.
- Investigated the potential of TensorFlow to enhance accuracy in machine learning applications. Developed and implemented a pipeline utilizing the TPOT classifier for automated feature selection, aiming to achieve superior model performance.
- Leveraged comprehensive data analysis, visualization, and diverse machine learning techniques to achieve significant model accuracy, with the Neural Network demonstrating the highest accuracy.