Vinay Shirole

Portfolio **930-333-2927** **vshirole@iu.edu** **LinkedIn** **GitHub**

EDUCATION

Masters in Data Science, Indiana University, *Bloomington, Indiana, USA* Aug 2023 - May 2025
Current Coursework: Natural Language Processing, Data Mining, Machine Learning, Big Data, Database Concepts.

Bachelors in Computer Engineering, University of Mumbai, *Mumbai, India* Jul 2017 - Jun 2021
Coursework: Data Structures and Algorithms, Software Engineering, Database Design and Warehousing, Computer Networks.

SKILLS AND EXPERTISE

Programming/Database : Python, Java, C/C++, Javascript, PostgreSQL, Flask, Oracle Database, MongoDB, MySQL.
Business Intelligence : R, Power BI, Tableau, Matplotlib, Seaborn, Plotly, SAS, Google Data Studio, Excel.
ETL and Data Pipeline : Azure Data Factory, Selenium, BeautifulSoup, Google Cloud Dataflow, PL/SQL.
Cloud and DevOps : Azure Blob Storage, Azure Databricks, CI/CD Pipelines, Azure Data Lake, Agile, Spark.
Machine Learning : Natural Language Processing, Pytorch, Machine Learning Algorithms, Scikit-Learn.

WORK EXPERIENCE

Graduate Research Assistant, Luddy School of Informatics, *Bloomington, Indiana, USA* Nov 2024 - Present
- Designing a Convolutional Neural Network (CNN) model to predict flood-prone areas using NOAA satellite imagery and geospatial features such as elevation data, aiming to achieve an accuracy of **85%+** for enhanced disaster management strategies.
- Processed and stitched large-scale satellite images of more than **100 GBs** using Python's GDAL library, optimizing image rendering and analysis for geospatial applications.
- Developed a custom web app using Node.js and Flask to visualize and annotate images with QGIS, marking flooded regions in red and non-flooded regions in blue to generate **250+** high-quality training images.

Graduate Research Assistant, eBay, *Bloomington, Indiana, USA* May 2024 - Jul 2024
- Refactored ETL pipelines using T-SQL and PySpark to process **10+** billion ad campaign records, enhancing data pipeline efficiency and computational performance.
- Conducted causal inference techniques to assess the impact of eBay's Promoted Ads on seller performance, analyzing **10,000+** adopters and non-adopters to generate strategic insights using R.
- Constructed foundational datasets for eBay's Ads Ranking team using Zeta, enabling data-driven ad campaign optimizations, which led to a **20%** increase in ROI and improved budget allocation.

Sr. Analyst/Software Engineer, Capgemini, *Navi Mumbai, India* Jul 2021 - Jul 2023
- Migrated over **5** TB of on-premises data to Azure Data Lake using ADF and Databricks, automating processes with **CI/CD** pipelines and resolving **10+** issues. troubleshooting issues weekly, ensuring data integrity and a smooth transition to the cloud.
- Improved operational efficiency by **50%** in data management operations by implementing an automated record maintenance system using Power Automate, effectively streamlining data processing and BI tasks, thereby reducing manual effort.
- Led a team of **3** IT professionals, providing **24/7** client support and ensuring service availability, while improving team performance and delivery quality through knowledge transfer sessions and technical training, resulting in higher client satisfaction.

PROJECTS

Spotify Recommendation system (Course project for Data Mining)
- Designed and implemented a sophisticated Spotify recommendation system utilizing data mining techniques on a dataset exceeding **100,000** entries, applying content-based filtering, and clustering methods like K-means to enhance recommendations.
- Created an algorithm leveraging cosine similarity paired with clustering techniques that enriched playlist diversity by offering over **1,000** new personalized track combinations based solely on individual listening habits.

Happiness Predictive modeling (Course project for Applied Machine Learning)
- Executed thorough testing protocols including hyperparameter tuning which resulted in consistent performance enhancements observed during validation phases; achieved an impressive rate of accurate predictions closely aligning at approximately **90%**.
- Implemented rigorous testing methodologies, including cross-validation, Grid Search Cross-Validation, and hyperparameter tuning, achieving a **95%** accuracy rate and ensuring model resilience and optimal performance.

GCP-Enhanced Airline Satisfaction Prediction (Course project for Management Access)
- Implemented an airline customer satisfaction analytics platform on Google Cloud, integrating Python-based ML models in Jupyter Notebook on a **VM** to predict satisfaction with **90%+** accuracy using **1M+** reviews.
- Managed cloud storage and data processing with Google BigQuery, achieving over **98%** prediction accuracy using RandomForestClassifier. Delivered actionable insights through data visualization, enhancing decision-making for service improvements.

CERTIFICATIONS

- Microsoft Certified: Azure Data Fundamentals: Demonstrates understanding of core data concepts and services for analytics.
- Microsoft Certified: **Power BI** Data Analyst Associate: Data preparation, modeling, and visualization skills in Power BI.

PUBLICATIONS

- Student Micro Loan Management System: Automated Web Application for Streamlined Student Loan Management **(ISSN No. 2455-2143**), published in International Journal of Engineering Applied Sciences and Technology, Vol. 5, Issue 12, 2021.b