Manali Vijay Shelar

+1 (812) 553-1456 | manalivshelar@gmail.com | LinkedIn | Github

Education

Master of Science in Data Science, Indiana University Bloomington Aug 2022 – May 2024

- **Coursework:** Statistics, Advanced Database Concepts, Data Mining, Usable AI, Exploratory Data Analysis, Data Visualization.

Bachelor of Engineering in Information Technology, University of Pune (India) Aug 2016 – June 2020

- **Coursework:** Data Structure and Algorithms, Machine Learning and Application, Cloud Computing, Software Testing.

Experience

Data Scientist | Observatory on Social Media (OSoMe), Indiana University Oct 2023 – Present

- Led the integration of **Large Language Models (e.g., GPT-4o)** into Helios-Web, a network visualization tool, significantly enhancing user engagement and experiences by leveraging **advanced NLP** techniques.
- Spearheaded performance evaluations and iterative trials of the integrated AI model, testing various hypotheses to enhance system accuracy and drive model improvements based on experimental findings.

Research and Analytics Intern | Institutional Analytics, Indiana University Jan 2024 – May 2024

- Designed **predictive** models using **machine learning algorithms** to assess factors impacting student retention, employing **Python, SQL, and Scikit-Learn** for large-scale data manipulation and **deep learning** approaches to refine outcomes.
- Transformed raw data into actionable insights, enabling data-informed decisions and improving retention strategies.
- Visualized insights with **Tableau dashboards**, integrating **SHapley Additive exPlanations (SHAP)** values for **feature importance** and enhancing **model interpretability** with **data-driven storytelling.**
- Performed **ad-hoc analyses** to support business and model risk functions, refining retention strategies with data-led insights.
- Collaborated with cross-functional teams to propose targeted interventions based on **prescriptive analytics**, recommending strategies to increase student retention through actionable measures.

Associate Consultant | Capgemini July 2020 – June 2022

- Spearheaded the development of **tailored software solutions** across **healthcare and supply chain** sectors, collaborating with cross-functional teams and business stakeholders to translate analytical insights into cost-saving operational improvements.
- Led the Solutions team in tackling day-to-day operational challenges across projects, creating **4 packaged solutions** adopted by **10+ clients**, streamlining their processes, and **reducing costs by an average of 15%**.
- Managed secure **deployments** of client solutions in **AWS environments** and performed **A/B testing** to ensure improvements in efficiency and performance, aligning the solutions with client business objectives and operational needs.

Senior Analyst Intern | Capgemini Jan 2020 – May 2020

- Collaborated extensively on projects involving **Java, SQL, and other relational databases** to analyze and optimize data for business insights, developing custom queries and integrating data from sources like **AWS, MsSQL, and SQL servers**.

Projects

Loan Default Prediction System

- Engineered a **scalable machine learning pipeline** for predicting loan defaults, utilizing **TensorFlow and Scikit-Learn** for improved accuracy in **credit risk assessment.** Leveraged **PySpark for distributed processing** of **large-scale complex datasets**, **fine-tuning XGBoost** models through **hyperparameter tuning** to improve prediction accuracy.

Recession Prediction System

- Developed a **Flask-based application** to assess the probability of U.S. recessions, analyzing past **100+ years** of data from the **Federal Reserve Economic Data** on key economic indicators like Treasury Bond Rates, GDP, CPI, and Inflation Rates.
- Utilized **statistical and machine learning** models, like **Logistic Regression, SVM, Decision Trees, and Time Series Analysis**, to predict economic downturns, improving prediction accuracy by **20%** and enhancing early warning capabilities.

Purchase Prediction for Big Organics

- Completed an **SAS** case study on Purchase Prediction for Big Organics LTD, leveraging **advanced statistical techniques and machine learning** to **segment potential customers** for the supermarket's new line of organic products.

Fuel Economy Analysis

- Conducted extensive **exploratory data analysis** using **descriptive statistics** alongside R packages like **ggplot2, dplyr, and tidyverse** to examine relationships between vehicle attributes and fuel economy. Also identified **patterns and trends** in fuel efficiency across different vehicle types, using **multivariate analysis** to uncover underlying factors affecting performance.

Skills

Programming Languages: Python, R, SAS, Java, JavaScript, Scala, C++, Angular, Spring Boot, React, Linux Bash.

Data Analysis & Machine Learning: Numpy, Pandas, Scikit-learn, NLTK, SciPy, Matplotlib, TensorFlow, PyTorch, Keras.

Data Preparation & Wrangling: Data Cleaning and Profiling, Feature Engineering, Outlier Detection, Feature Scaling.

Database Technologies & Visualization Tools: SQL, PostgreSQL, NoSQL, Cassandra, Neo4j, Power BI, Tableau, Looker.

Cloud Technologies & DevOps Stack: AWS, Azure, Databricks, GCP, Spark, Hadoop, Docker, Kubernetes, Git, Agile.

Certifications: Microsoft Azure AI Fundamentals (Aug 2024), AWS Solutions Architect(Jul 2024).