Akshat Porwal

1050 Galatyn Pkwy., Apt 2060, Richardson, TX - 75082

734-934-5967 akshatp@umich.edu akshatporwal akshatporwal akshatporwal.github.io

Education

University of Michigan, Ann Arbor **May 2024**

M.S. in Data Science - *Statistical Inference, Regression & Multivariate Analysis, Machine Learning, NLP* *CGPA: 3.8 / 4.0*

Symbiosis University of Applied Sciences, India **July 2022**

B.Tech. in Computer Science & Information Technology - *Database Systems, Data Mining* *CGPA: 4.0 / 4.0*

Technical Skills

Programming Languages: Python, R, SQL, SAS, AWS, PySpark

Databases: SQL Server, MySQL, SQLite, Oracle

Data Science & Machine Learning Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, PyTorch, ggplot2, Plotly, NLTK, OpenCV

Machine Learning & Statistical Modeling: Linear Regression, Logistic Regression, Decision Trees, SVM, KNN, K Means, CNN, Natural Language Processing (NLP), Statistical Modeling

Visualization & Business Intelligence: Tableau, Power BI, Alteryx

Development & Collaboration Tools: Git, MS Excel, JIRA, APIs, Visual Studio Code, PyCharm, Jupyter Notebook

Work Experience

Research & Teaching Assistant - Univ. of Michigan, Ann Arbor **April 2023 – Present**

- Expanding research with Prof. Ceren Budak since Winter '24 to analyze subscription and donation behaviors across 5K+ news platforms, developing data strategies to understand user engagement and financial trends.
- Developed and deployed a real-time classifier to identify articles on local elections, leveraging XGBoost with TF-IDF and numeric features after optimizing a BERT-based model. Implemented paragraph-level aggregation for article classification and coordinated production handoff, enabling seamless use by newsroom partners.
- Served as a Teaching Assistant for Information Visualization with Prof. Eytan Adar, managing a cohort of 90.
- In Summer '23, conducted Bayesian analysis with Prof. Dallas Card on 1,000+ graduate student publications, identifying trends in computer science academic productivity.

Data Scientist - Freddie Mac **January 2024 – Present**

- Spearheaded initiatives in statistical and predictive modeling across various phases of modeling pipelines, significantly enhancing the applications of machine learning.
- Utilized Python and SQL to conduct regression and multivariate data analyses, tackling complex marketing analytics challenges, which led to a 15% increase in conversion rates and provided actionable insights through KPIs and metrics.
- Directed the deployment and management of data and ML infrastructure on AWS, achieving a 30% reduction in deployment time, while maintaining version control and overseeing testing via Jenkins CI/CD pipelines within the SDLC framework.
- Investigated customer behaviors to unearth valuable insights, employing clustering algorithms for effective customer segmentation to refine marketing strategies.
- Performed data integrity checks, cleansing, exploratory analysis, and feature engineering using Python, enhancing data visualization with Matplotlib and Seaborn.
- Generated compelling visualizations and reports using Power BI, Matplotlib, Seaborn, and SQL, contributing to a 25% increase in the communicative value of marketing campaigns.
- Orchestrated the implementation of Snowflake's retail solutions for customer analytics, inventory optimization, and fraud detection, achieving a 15% boost in customer satisfaction and a 20% improvement in inventory accuracy.
- Collaborated with database engineers to streamline ETL processes, optimizing SQL queries for enhanced data extraction and integration from SQL Server databases.

Data Scientist - Aspire Fintech, India **April 2023 – August 2023**

- Initiated process mining with pm4py to identify and mitigate application bottlenecks, optimizing loan workflows which enhanced customer satisfaction by 30% and increased processing efficiency by 25%.
- Conducted risk profile analyses using metadata and data elements, employing the K-Means Algorithm to effectively segment customers and precisely evaluate model performance.
- Processed and cleansed data utilizing missing value treatment and standardization techniques in pandas, followed by visualizing results with matplotlib to create bar charts, heat maps, and histograms.

- Executed classification models using supervised learning algorithms such as Logistic Regression, Decision Trees, KNN, and Naive Bayes, enhancing predictive accuracy.
- Led A/B testing initiatives for the take-up rate model by modifying credit terms for targeted users and analyzing outcomes with logistic regression, which resulted in a 15% increase in product adoption rates and refined targeting strategies.

Data Scientist - Hexaware Technologies, India **July 2020 – August 2022**

- Incorporated various classification algorithms such as Support Vector Machines (SVM), Random Forests, Naive Bayes, and K-Nearest Neighbors (KNN) for pattern recognition and categorization.
- Employed BERT-based methods to extract structured data from unstructured documents, standardizing fields like diagnosis and medication across 500K records, reducing missing data by 30% and boosting strategic decision-making efficacy by 20%.
- Orchestrated streamlined ETL processes, meticulously overseeing data extraction and transformation workflows, resulting in a notable 20% decrease in processing duration.
- Hands-on experience with Pandas, NumPy, SciPy, Matplotlib, Scikit-learn, and NLTK in Python for developing machine learning algorithms to assess risk and pricing, and forecast key metrics.
- Orchestrated log analysis workflows leveraging Kafka and Spark, combining logs from microservices to provide dashboards for monitoring systems and detecting anomalies.
- Established a modular pipeline for transforming 2.5+ GB of claims and form data stored across MySQL Server, PostgreSQL, and Excel (VLOOKUP, pivot tables) sources into analytics-ready datasets using Python and SQL, ensuring quality and accuracy for advanced ML models.
- Developed a robust data strategy on AWS, introducing cost-efficient cloud solutions that reduced infrastructure expenses by 20%.
- Ensured strict adherence to industry standards, compliance regulations, and best practices, thereby maintaining the utmost reliability and security of SQL Server databases.

Project Experience

Age Prediction From Naming Trends | *Bayesian Statistics, Time Series Analysis* **January 2024 - Present**

- Utilized a Bayesian approach on a century-long U.S. baby names and life expectancy dataset to uncover age-indicative naming trends, highlighting the impact of societal factors on naming practices.
- Conducted a detailed exploratory data analysis to pinpoint historical peaks in name popularity, providing valuable insights for sociolinguistic research without the creation of a predictive model.

Anime Odyssey | *Recommender Systems, SVD, Cluster Analysis* **October 2023 - December 2023**

- Tackled anime recommendation challenges by analyzing genre trends across 18K titles using the Jaccard index and K-means, enhancing content personalization and user engagement metrics.
- Devised SVD and KNN models in a recommendation system for anime viewership, achieving MAEs of 0.87 and 0.95 respectively, illustrating the system's accuracy in predicting user preferences among thousands of titles.

Flight Predictive Analysis: ARIMA & Neural Networks | *ARIMA, LSTM* **October 2023 - December 2023**

- Assessed ARIMA and LSTM models for forecasting flight departures, finding ARIMA's short-term prediction more accurate with a 159.96 RMSE and LSTM's long-term predictions reliable with a 174.27 RMSE.
- Implemented Dense Neural Networks, achieving the best RMSE of 121.15, demonstrating effective forecast accuracy with a trade-off in computation time.

Early Prediction of Sepsis | *Classification, Ensemble Methods* **February 2023 - April 2023**

- Formulated machine learning models including XGBoost, AdaBoost, and Histogram-Based Gradient Boosting, to analyze over 21,634 patient records for early sepsis detection in ICU patients.
- Achieved an AUC score of 0.89 with the developed models, addressing the data's class imbalance problem.

Social Media Analysis of Bitcoin | *Data Visualization, Spark SQL* **October 2022 - November 2022**

- Analyzed over 16 million tweets and 4 million Reddit comments to assess the impact of social sentiment on Bitcoin's price, identifying significant trends correlating public discourse with market fluctuations.
- Utilized Python and SparkSQL to process large datasets, discovering seven key periods where verified users' tweets aligned with Bitcoin price changes, shedding light on influential factors in cryptocurrency valuation.

Student Election System | *SQL, ASP.NET, Scala* **April 2020 - June 2020**

- Developed a system for college elections, facilitating the selection of club presidents, student representatives, etc., by capturing voter information and displaying all candidates, along with identifying the leading candidate with the highest votes.
- Engineered the system using ASP.NET, ADO.NET, HTML/CSS, and JavaScript to ensure seamless user interaction and accurate vote management.