**Sharmisha Y**

**SUMMARY**

Accomplished Data Scientist with 5+ years of experience specializes in AI, Machine Learning, Deep Learning, Data Mining, and Natural Language Processing (NLP). Experienced in the generation of actionable insights using data visualization tools such as Tableau and Power BI and Python. Established proficiency in the management of real-world data and complex datasets, the development of predictive models, and the enhancement of business outcomes in the finance, retail, insurance, real estate, manufacturing, and healthcare sectors. Demonstrated ability to resolve issues and a commitment to ongoing professional development by delivering communication strategies that were clear, concise and actionable.

**EDUCATION**

**Rowan University September 2022 – December 2023**

Master of Science in Data Science

**CMR Institute of Technology July 2017 – June 2021**

Bachelor of Technology in Mechanical Engineering

**TECHNICAL SKILLS**

**Programming Language:** Python, R, SQL

**IDE’s:** Visual Studio Code, Anaconda, R Studio

**ML Algorithm:** Linear Regression, Logistic Regression, Decision Trees, Classification, Random Forests, Naive Bayes, K Means **Deep Learning:** NLP, ANN, CNN, RNN, LSTM, Attention, Transfer Learning, VAEs, GANs and Transformers (BERT, GPT) **Cloud and Big Data Technologies:** AWS (EC2, S3, Athena, Lambda, EMR, CloudWatch), PySpark and Databricks

**Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Ggplot2, PyTorch, Keras, Beautiful Soup, NLTK

**Visualization Tools:** Tableau, Power BI

**Other Tools:** Git, Google Analytics, Docker, SQL Server, Microsoft Excel (VLOOKUP, Pivot Tables, Data Analysis), Windows, Linux

**EXPERIENCE**

**STERIS, NJ, UNITED STATES September 2023 – Present**

**Data Scientist**

* Augmented sentiment analysis accuracy by 30% by implementing advanced NLP models using NLTK, improving customer insights.
* Strengthened model robustness by 30% by through Generative AI techniques like GANs and VAEs, improving dataset augmentation and generalization.
* Cut infrastructure costs by 40% and bolstered scalability by 50% via serverless architecture using AWS Lambda and Amazon
* API Gateway.
* Led cross-functional workshops with stakeholders and domain experts to define BERT application parameters; resulting in a
* 20% reduction in customer churn with personalized recommendations.
* Implemented automated data pipelines, reducing data processing time by 60% and enabling real-time analytics.

**MACRO SOFTWARE SOLUTIONS, HYDERABAD, INDIA August 2018 – July 2022**

**Data Scientist**

* Spearheaded the implementation of Convolutional Neural Networks (CNNs) which enhanced image classification accuracy by
* 20%, optimizing model efficiency across different datasets and improving overall system performance.
* Boosted customer churn prediction accuracy by 25% by orchestrating the application of advanced statistical modeling techniques with XGBoost in Python, enabling targeted retention strategies based on accurate predictions.
* Streamlined report generation, reducing report generation time by 50% through the development of an LLM-powered system that automatically generates data analysis reports, summarizing key findings and insights for stakeholders.
* Designed and deployed a recommendation engine that increased product cross-sell rates by 35%, leveraging collaborative filtering techniques.
* Conducted comprehensive data analysis to identify business opportunities, resulting in a 15% increase in revenue.

**PROJECT EXPERIENCE**

**Master's Capstone Project in Data Mining**

• Enhanced data quality by 40% through comprehensive data understanding and processing initiative, which involved detailed

data cleaning, transformation, and integration processes.

• Refined predictive modeling accuracy by 15% by leading classification tasks and performance evaluation, applying machine learning techniques to refine models and improve prediction accuracy.

• Boosted customer engagement by 25% by evaluating customer data using clustering techniques to uncover distinct segments

and implementing personalized marketing strategies based on these insights.

**Lossy Reconstructions of ImageNet100 Dataset Using Hierarchy of Autoencoders (HAE).**

• Led a codebase enhancement project for hierarchical autoencoders, boosting evaluation accuracy by 30% and overall performance, resulting in a highly compliant framework for lossy image reconstructions, classification, and evaluation.

**End-to-End Big Data Project: From Raw Dataset to Insightful Visualizations |** [YouTube Link](https://youtu.be/z5exVZgDlmI)

• Transformed crash reporting driver data into actionable insights using AWS S3, SQL, Databricks, and Tableau; implemented targeted safety campaigns reducing accident rates by 40%.

**Loan Transaction Analysis |** [Databricks Part 1 **|**](https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4696499163168464/3439751324214102/149976116126251/latest.html) [Databricks Part **2**](https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4696499163168464/2692732258955859/149976116126251/latest.html)

• Built a streamlined data pipeline on AWS S3 and Databricks, tackling missing data challenges and generating 3 impactful visualizations using SQL. This analysis demonstrates the expertise in data warehousing, ETL processes, and data analysis.

**CERTIFICATIONS**

• Analyze Survey Data using Principal Component Analysis, Coursera

• IBM Data Science Professional Certificate, Coursera

• Google IT Automation with Python Professional Certificate, Coursera