AISHWARYA MUCHANDI

DATA SCIENTIST

682-529-1882 | aishwarya.m@mycvtalent.com | Texas | GitHub | LinkedIn

SUMMARY

- **Data Scientist** with **3+ years** of expertise in **Data Extraction, preprocessing, Data Validation, Exploratory Data Analysis, Feature Engineering, Data Wrangling, Statistical Analysis, Machine Learning and Data Visualization**.
- Implemented deep learning algorithms such as Convolutional Neural Networks (**CNNs**), Recurrent Neural Networks (**RNNs**), **LSTMs**, and **Hugging Face Transformers**.
- Proficient in **Python** for data manipulation, including loading and extraction. Expertise in key libraries like **NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, PyTorch, Keras, OpenCV, Beautiful Soup, Generative AI, and NLTK.**

EDUCATION

Master of Science in Computer Science *May 2023*
The University of Texas at Arlington, Arlington, Texas

Bachelor in Bachelor in Computer Science & Engineering *August 2020*
University of Mumbai, India

SKILLS

Language/ IDE's:	Python, R, SQL, PyCharm, Jupyter Notebook, Google Colab
Machine Learning:	Linear, Logistic Regression, Decision Trees, Random Forests, Naive Bayes, SVM
Deep Learning:	ANN, CNN, RNN, Hugging Face Transformers (BERT, GPT-3), LSTM, GenAI, LLM
Visualizations:	Tableau, Power BI, Excel
Cloud:	AWS (DynamoDB, RDS, S3, EC2, Redshift), Azure (Azure Databricks, Azure DevOps)
Packages:	NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, Keras, NLTK, OpenCV, Beautiful Soup
Database and Tools:	SQL Server, MySQL, PostgreSQL, MongoDB, Docker, Jenkins, GitHub

EXPERIENCE

Capital One Financial, TX | Data Scientist *July 2023 - Present*

- Successfully leveraged **Generative AI methods** to overcome limitations in data availability for **customer churn prediction**, enabling the development of 20 % more accurate and more generalizable churn prediction models.
- Constructed an **LLM-powered system** that automatically generates data analysis reports, summarizing key findings and insights for stakeholders, leading to a reduction in report generation time by 30%.
- Integrated **PyTorch** to construct graphs, accelerated model and trim development iterations by 20%.
- Applied the **NLTK** library for Natural Language Processing (**NLP**) data processing and pattern recognition, resulting in a **15% improvement** in the **identification of churn risk factors** from customer feedback data
- Utilized **AWS Lambda and Amazon API Gateway** to achieve a serverless architecture, delivering a 40% reduction in infrastructure costs and an approximately 50% improvement in system scalability.
- Prepared preprocessed datasets using **SQL window functions** and advanced techniques, improving data quality by 10%.
- Employed **Hugging Face Transformers (BERT, GPT-3)** to achieve a 50% improvement in text classification accuracy for a document categorization project, resulting in faster and more efficient information retrieval.

Zimperium Inc, TX | Data Scientist *May 2022 - May 2023*

- Developed **Prediction ML model** and deployed with 96.42 % accuracy detecting anomaly behavior in (CVE IDs) Common Vulnerability and Exposure Ids; targeting various App Market Categories for fraudulent app detections and used **XG Boost**.
- Deployed Statistical ML model using **Python + Flask**, to predict the probability of CVE IDs (Security Flaws) in app market categories.
- Enhanced existing decision-making model with 30% accuracy using the above two models.
- Utilized **Machine Learning Techniques**: **k-NN, Naive Bayes, SVM, and Decision Forests Algorithms** to cluster CV EIds based on the probability occurrence in different App Market Categories in **(IOS, Android**)
- Built **data pipeline** in **S3**, **AWS Quick Sight** to create real-time dashboards resulting in 20% faster decision-making.
- Maintained **SQL** pipelines that reduced data preparation time for **ML Model** by 20% while ensuring data integrity.

Kalp Technolab, India | Data Scientist *March 2020 - July 2021*

- Operated **NLTK** to enhance the **NLP** functionality of a customer response application, enabling efficient resolved of 10+ daily interactions through automation.
- Applied **RNN** architectures to enhance anomaly detection efficiency in large datasets, resulting in a 30% reduction in false positives.
- Partnered with cross-functional teams to implement automated **dashboards** and reporting using **Power BI**, facilitated stakeholders with data-driven insights, and Increased workforce performance by 25%.
- Developed **Decision Tree, Random Forest, and XGBoost models**, resulting in a 10% increase in average customer lifetime value through targeted marketing campaigns.
- Analyzed **SQL** to explore over 10 large datasets (rows) for feature engineering and data cleaning, leading to improved model performance by 20%.
- Automated **CI/CD pipeline** using **Git** and **Jenkins**, achieving a 15% reduction in build and deployment times for big data projects on Azure.
- Increased **forecasting accuracy** by 80% on a time series prediction task using **LSTMs** through hyperparameter tuning of learning rate, batch size, and network architecture.