Sanjay Krishna Kantamani
Data Scientist
512-564 -4813 | sanjay.k@mycvscout.com | Richardson, TX | LinkedIn

SUMMARY

- Data Scientist with around 4 years of expertise in data extraction, pre-processing, validation, exploratory data analysis, feature engineering, data wrangling, statistical analysis, machine learning, and data visualization.
- Competent in using a range of data science and machine learning tools, including Scikit-learn, TensorFlow, XGBoost, PyTorch, and cloud technologies like AWS to build scalable and efficient solutions.
- Leveraging deep learning expertise in implementing Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) with LSTM to tackle complex problems.
- Proficient in building an NLP text summarization model that reduced document processing time, allowing for information consumption.
- Optimizing SQL Server databases for real-time data ingestion and retrieval, supporting high-throughput applications with frequently accessed data, and maintaining SQL scripts to transform datasets for Machine Learning model training.

EXPERIENCE

Capital One Financial, TX **Aug 2023 – Present**
Data Scientist

- Boosted customer lifetime value prediction accuracy by 25% through data analysis and advanced customer segmentation/behavior modeling using Python (Scikit-learn, NumPy, pandas) and Spark (MLlib, PySpear)
- Designed an ML model using (Random Forest, and XGBoost) for customer churn prediction, achieving 92% accuracy and reducing churn by 4%, driving $100,000 in annual revenue retention.
- Developed a LangChain-powered model to extract and summarize key findings from diverse reports, accelerating knowledge discovery by 50% through efficient information distillation.
- Attained a significant 30% improvement in overall model performance by strategically combining Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) in a hybrid architecture for complex data analysis tasks.
- Leveraged AWS EMR to create highly scalable machine learning pipelines, capable of handling training datasets exceeding terabytes in size and achieving faster model deployment times

Zensar Technologies, India **Jun 2020 – Nov 2021**
Data Scientist - II

- Worked on K-Means clustering and PCA to develop a robust customer segmentation model, identifying high-value customer segments. This led to a 25% increase in targeted marketing campaign effectiveness, optimizing customer acquisition and retention strategies.
- Developed anomaly detection models using statistical and machine learning techniques to identify outliers and anomalies in large-scale datasets, leading to a 40% reduction in false positives and an improvement in anomaly detection accuracy.
- Engineered custom embeddings for semantic similarity and topic modeling in Azure Databricks, utilizing spaCy, Scikit-learn, and Sentence Transformers, accelerating data processing by 40% and enhancing feature extraction.
- Improved complex SQL queries by employing techniques such as partitioning, clustering, and materialized views, resulting in a 35% reduction in query execution time. These optimizations significantly improved database performance and responsiveness.
- Built deep learning models with PyTorch for diverse tasks like image recognition (object detection), NLP tasks (sentiment analysis, topic modeling), and time series forecasting, achieving validation accuracy of 85%.

Zensar Technologies, India **Aug 2019 – May 2020**
Data Scientist - I

- Accomplished a computer vision model to accurately detect product defects on assembly lines, and employed deep learning architectures and image processing techniques to achieve 95% accuracy in defect classification. Reduced false positives by 20%, leading to a 15% increase in production efficiency and cost savings.
- Optimized the performance of various machine learning and deep learning models through regularization techniques and advanced hyperparameter tuning techniques, including RandomSearchCV, GridSearchCV and K-Fold Cross-Validation.
- Enhanced SQL queries and adjusted resource allocation, maintained SQL pipelines that reduced data preparation time for ML Model by 20% while ensuring data integrity and better decision-making for the organization.
- Directed real-time monitoring dashboard using Python and Tableau, providing actionable insights and early warning signals to maintenance teams, leading to a 15% improvement in response time to equipment issues

SKILLS & CERTIFICATION

- **Language/ IDE's:** Python, R, SQL, Jupyter Notebook, Google Colab
- **Machine Learning:** Linear, Logistic Regression, Decision Trees, Random Forests, Naive Bayes, SVM
- **Deep Learning:** CNN, ANN, RNN, LSTM, Natural Language Processing(NLP), Large Language Model(LLM), LangChain
- **Cloud/Visualizations:** AWS, Azure, Tableau, Power BI, Looker
- **Packages and Frameworks:** NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, Keras, NLTK, XGBoost, PyTorch
- **Database and Tools**: MySQL, PostgreSQL, MongoDB, SQL Server
- **Certification:** Google Data Analytics

EDUCATION

Master of Science Information Technology and Management | The University of Texas at Dallas, TX

B.Tech in Computer Science and Engineering | GITAM Deemed to be University (Visakhapatnam), India