Shyamsundhar Yathirajam

Data Scientist | Machine Learning Engineer

California| 760-639-8222 | shyamsundhar@myemailjob.com |LinkedIn |Git Hub

SUMMARY

- Highly experienced Data Scientist with 5+ years of experience in Data Extraction, Data Modelling, Data Wrangling, Statistical Modeling, Mathematical Modelling, Data Mining, Machine Learning, Data Visualization, and Computer Vision applications.
- Experience working in Agile/Scrum Methodologies to accelerate Software Development iteration.
- Proficient in Machine Learning algorithms and Predictive Modeling, including Regression Models, Decision Trees, Random Forests, Sentiment Analysis, Naïve Bayes Classifier, SVM, and Ensemble Models.
- Proficient in AI and deep learning models, including neural networks, CNNs, and LSTMs, with significant awards and publications.
- Knowledge of Natural Language Processing (NLP) algorithm and Text Mining.
- Experience in deploying and maintaining machine learning models in production environments.
- Proficient in Python with SciPy Stack packages, including NumPy, Pandas, SciPy, Matplotlib, and Python.
- Knowledge of provisioning virtual clusters under the AWS cloud, which includes services like EC2, S3, and EMR.
- Skilled in optimizing machine learning models for performance and scalability.
- Well-versed in data visualization tools like Tableau, Python Matplotlib, and R Shiny to create visually powerful and actionable interactive reports and dashboards for storytelling through data analysis.
- Proficient in data analysis with sound knowledge of data extraction from various database sources like MySQL, MSSQL, Oracle, Teradata, and other database systems.

SKILLS

Methodologies: SDLC, Agile, Waterfall
Language: Python, R, SQL, SAS, C#, Scala, C/C++, Java, Javascript
IDEs: Visual Studio Code, PyCharm, Jupyter Notebook
Statistical Methods: Hypothetical Testing, ANOVA, Time Series
Machine Learning: Regression analysis, Bayesian Method, Decision Tree, Random Forests, Support Vector Machine, Neural Network, Sentiment Analysis, K-Means, KNN, Classification, SVM, Naive Bayes, NLP, LLM, CNN, XGBoost, Julia, Deep Learning, Predictive Models, LangChain, , LlamaIndex, Haystack, Pinecone, Cassandra, Qdrant, TensorRT.
Packages: NumPy, Pandas, Matplotlib, SciPy, ggplot2, Scikit-Learn, PyTorch (CUDA), TensorFlow, Keras, Spark
Visualization Tools: Tableau, Power BI, Microsoft Excel
Cloud Technologies: AWS, GCP(including Google BigQuery), Docker, Kubernetes, Azure, Data Warehousing
Database: MySQL, SQL Server, Oracle, MongoDB, SAP ECC, S/4 HANA Migration
Software/Other Skills: Jira, Data Cleaning, Data Wrangling, Critical Thinking, Communication Skills, Presentation Skills, Problem-solving, Decision-Making, EDA, Communication Skills, Databricks, Data Visualization, Predictive Analytics, Pattern Recognition, JMP, Data Integrity, Quantitative Data, Data Science, Statistics, Statistical Analysis, Data Analytics, Data Modeling, Big Query, Snowflake, Data Extraction, Data Loading, Data Mining, Data Reporting, Data Transformation

WORK EXPERIENCE

Ally Financial, USA | Data Scientist — Jan 2024 - Current

- Worked in Agile environments to deliver iterative solutions and adapt to changing project requirements.
- Provided insights into market trends by analyzing and transforming financial datasets into actionable business intelligence
- Employed ggplot2, a data visualization package in R, to create elegant and customizable graphics based on the grammar of graphics principles.
- Utilized machine learning algorithms such as linear regression, multivariate regression, Naive Bayes, Random Forests, K-means, & KNN
- Used AWS S3, DynamoDB, AWS lambda, AWS EC2 for data storage and models' deployment.
- Created and maintained reports to display the status and performance of the deployed model and algorithm with Tableau.
- Implemented, tuned, and deployed machine learning models using AWS Lambda and SageMaker.
- Worked with CI/CD pipelines to ensure seamless deployment of machine learning models.
- Integrated machine learning models with RESTful APIs for real-time predictions.
- Implemented, tuned, and tested the model on AWS Lambda with the best-performing algorithm and parameters.
- Identified and assessed available machine learning and statistical analysis libraries (including regressors, classifiers, statistical tests, and clustering algorithms).
- Leveraged NoSQL databases such as Redis and Neo4j to manage unstructured and semi-structured data, ensuring high scalability.
- Created customized SQL Queries using MS SQL Management Studio to pull specified data for analysis and report building in conjunction with Crystal Reports.
- Utilized Apache Spark to process and analyze massive datasets, significantly reducing processing time.
- Performed data cleaning and feature selection using the MLLib package in Spark and working with deep learning frameworks such as TensorFlow.
- Automated ETL workflows on AWS Glue to handle large-scale datasets efficiently for business analytics

California State University, USA | Research Data Scientist **Jun 2023 - Dec 2023**

- Enhanced data processing times by 55x using hyperdimensional computing models on both CPU and GPU (CUDA).
- Achieved 98% performance improvement of CNN algorithms by researching existing methods, coding new algorithms in Python.
- Applied Fourier transforms and signal processing methods to analyze wind farm oscillation data, improving feature extraction from complex datasets.
- Performed A/B testing to compare the effectiveness of different machine learning models for data-driven insights.
- Created customized ML pipelines for feature engineering, model tuning, and deployment to production environments.
- Published research findings in top-tier conferences, showcasing novel approaches in image segmentation and pattern recognition.
- Implemented 2+ advanced ML models for analyzing large wind farm data, addressing the significant research gap in windfarm oscillation detection, improving 2D CNN accuracy by 4%, and achieving a 98% recall rate.
- Leveraged GCP services, including Google Cloud Storage and BigQuery, for scalable data solutions and efficient processing of large datasets, while continuously monitoring and evaluating model performance.
- • Integrated computer vision models into edge devices for real-time analysis using TensorFlow Lite and NVIDIA Jetson platforms.

Rlogical Techsoft Pvt. Ltd, India | Data Scientist **Jan 2020 - July 2022**

- Conducted thorough testing and validation of data science models and solutions before deployment, following a structured and rigorous approach typical of Waterfall methodologies.
- Implemented automated machine learning (AutoML) techniques to optimize model selection and hyperparameters.
- Used pandas, NumPy, seaborn, SciPy, Matplotlib, scikit-learn, and NLTK in Python for developing various machine learning algorithms
- Data Manipulation and Aggregation from different sources using Nexus, Toad, Business Objects, Power BI, and Smart View.
- Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes
- Evaluated the performance of Various Classification and Regression algorithms using R language to predict the future power
- Collaborated with cross-functional teams to integrate Databricks solutions into existing data pipelines, ensuring seamless data flow.
- Implemented supervised and unsupervised learning algorithms, including LLM (Latent Dirichlet Allocation), to uncover patterns in large datasets.
- Involved in developing and data ingestion into a cluster in AWS environment (EC2 instances) using SparkSQL.
- Developed a deep learning model for the industrial decision-making process through Keras - Models such as RNN and LSTM were also implemented and deployed on a cloud environment like GCP for continuous monitoring of model performance.
- Implemented backup and recovery strategies for SQL Server databases, ensuring data availability and integrity in case of system failures or disasters.

Groovy Web, India | Data Analyst **Jan 2019 - Dec 2019**

- Involved in every phase of the SDLC process, starting from Inception to Transition.
- Performed exploratory data analysis (EDA) to find insight such as difference efficiency among different devices using ggplot2 in R.
- Worked on AWS Data Pipeline to configure data loads from S3 to Redshift.
- Completed a few hours of online training in advanced data analysis techniques, contributing to adopting best practices and a 15% increase in team productivity.
- Implemented automated data cleaning processes, resulting in a 30% reduction in data errors and a 25% increase in data validation efficiency.
- Maintained and optimized databases and wrote SQL queries for data extraction, transformation, and loading (ETL) processes.
- Deployed MS SQL Server platform to store, transform, and query data for data analyzing and reporting.
- Analyzed complex datasets using statistical models and machine learning algorithms, resulting in a 15% increase in customer retention rates.

EDUCATION

Master's in computer science | California State University San Marcos, San Marcos, CA

Bachelor's in computer science | Jawaharlal Nehru Technological University Hyderabad, Hyderabad, Telangana

Diploma in Computer Science | Osmania University, Hyderabad, Telangana

PUBLICATIONS

- Yathirajam Shyam, Peighambari Arash, Ruben Roberts, Sreedevi Gutta, Hamed Nademi, Morris Justin, and Ali Ahmadinia, “Windfarm Forced Oscillation Detection Using Hyper dimensional Computing” IEEE International Conference on Big Data (BigData), 2023.
- Yathirajam Shyam Sundhar, Sreedevi Gutta, “Efficient glioma grade prediction using learned features extracted from convolutional neural networks” in Journal of Medical Artificial Intelligence. 2024.
- Yathirajam Shyam Sundhar, Sreedevi Gutta, “Improved Glioma Grade Prediction using Mean Image Transformation” in AIME 2024: 22nd International Conference on Artificial Intelligence in Medicine, 2024.
- Yathirajam Shyam Sundhar, Sreedevi Gutta, “Glioma Grade Prediction using Hyper-dimensional Computing” in AIME 2024 Conference.
- Yathirajam Shyam Sundhar, Justin Morris, “NeuraLHDC:Neural Network and Learned Projections Based Hyperdimensional Computing”.

AWARDS

- Nominated for “**Dean’s Awards 2024”** at **CSUSM** for innovative research and publications in the field of AI in different domains.
- Awarded with **“Student Faculty Collaboration Award 2024”** at **CSUSM** for the research in Hyper dimensional computing and AI.