

Raktimjyoti Parashar Data Scientist

✉ raktimjyoti@mywebemails.com ☎ (267)-221-1000 📍 San Jose, USA 🔗 LinkedIn 🐙 Github

Profile

Skilled data scientist with 5 years of experience in machine learning algorithms, AI model development, and Python programming. Proficient in using TensorFlow, PyTorch, and Pandas libraries, with expertise in NLP, statistical methods, ETL data processes, RESTful APIs, and Tableau for data visualization. Illustrates strong problem-solving, critical thinking, and teamwork skills to deliver clever solutions for complex data challenges.

Skills

Programming and Scripting — Python (NumPy, Pandas, scikit-learn, TensorFlow, PyTorch), SQL, MATLAB, Scala

Data Manipulation and Analysis — Data cleaning, visualization and preprocessing, EDA, Statistical analysis, Feature engineering, Time series analysis

Big Data Technologies — Apache (Hadoop, Spark, Kafka), Distributed computing frameworks (MapReduce)

Artificial Intelligence & Machine Learning — Machine Learning Algorithms, Random Forest, Linear Regression, SVM, Decision Tree, ARIMA, Tensor Flow, Keras, Sci-kit Learn, Deep Learning Algorithms, NLP

Frameworks & Tools — Flask, MS SQL, MySQL, PostgreSQL, MongoDB, Git, Docker, AWS, Azure, Power BI, Tableau

Professional Experience

Data Scientist, Capgemini 01/2023 – Present | Remote, USA

- Developed a 6-DOF, occlusion-aware object tracking pipeline using computer vision models, which improved automation capabilities by 60% and enabled computing optimal grasp poses for robotic handling of novel objects.
- Engineered a vital pipeline for revamping precise 3D models from raw point cloud data using Log-GPIS and Open3D, improved the accuracy of robot perception using NumPy & SciPy libraries, boosting finer control and interaction.
- Integrated Machine Learning techniques to enhance object tracking accuracy, employing supervised learning models for better feature extraction and classification.
- Developed a natural language processing module using LLMs (Large Language Models) to enable real-time command recognition and processing, enhancing the interaction between the robotic system and users.
- United with the team to create a NeRF model of indoor spaces using TensorFlow and PyTorch for accurate modelling and navigation around obstacles using vision-based systems only.
- Worked on LiDAR-Depth Camera Sensor Fusion onboard AgileX SCOUT using ROS (Robot Operating System) and created Python scripts for data acquisition, preprocessing, and fusion algorithms, enhancing accuracy by 80%.

Data Scientist, Genpact 11/2017 – 07/2021 | Remote, India

- Designed a scalable multi-cloud solution using Snowflake and AWS for Cancer patient length-of-stay data analysis, incorporating Python libraries and large language models for advanced analytics and insights.
- Developed data-driven methods integrating Cancer patient insights with supply chain needs, predicting shortages, and optimizing inventory levels for efficient hospital supply chain management.
- Consolidated data management in biotech research using MongoDB, and PyMongo for analysis, and machine learning integration for scientific insights. Automated key findings extraction from literature with LLMs for research capabilities.
- Optimized hospital logistics by implementing bioinformatics workflows, developing data pipelines for large dataset analysis, and integrating ML algorithms for real-time cancer patient outcome predictions and resource allocation.
- Implemented robust data control frameworks, secure data storage/transfer protocols, and LLMs for automated compliance checks and report generation to ensure HIPAA compliance.

Education

Master of Science, University Of Pennsylvania 08/2021 – 05/2023 | Philadelphia, USA
Mechanical Engineering and Applied Mechanics

Bachelor of Technology, Manipal Institute Of Technology 06/2013 – 09/2017 | Karnataka, India
Mechanical Engineering and Manufacturing Technology

Projects

Search Engine

- Engineered a sophisticated machine learning application to analyze linguistic relationships and patterns in medical terminology using NLTK, Scikit-Learn, Pandas, Numpy, Streamlit. Developed a smart search engine capable of querying complex medical records for specific terms, enhancing the retrieval of critical data. Successfully deployed and scaled the solution using Azure Machine Learning pipelines (Azure Data Factory, Azure Blob Storage, Azure Databricks), facilitating real-time data processing and significantly improving research and clinical decision-making efficiency.

Enhancing Business Engagement, Advanced AI and LLM for Detoxifying and Moderating Hate Speech in Online Communities

- Leveraged advanced AI and Large Language Models (LLM) to combat toxicity in online communities. Our findings revealed how sophisticated models like FLAN-T5, enhanced through Proximal Policy Optimization (PPO), can significantly reduce hate speech and improve digital interactions. End-to-end deep learning project to classify disease in potato plants as either early blight or late blight.