

# RINDHUJA TREESA JOHNSON

Baltimore, MD (Willing to Relocate) ◇ 914-746-5465 ◇ [rindhuj1@umbc.edu](mailto:rindhuj1@umbc.edu) ◇ [LinkedIn](#) ◇ [GitHub](#) ◇ [Portfolio](#)

## PROFESSIONAL SUMMARY

Machine Learning Engineer with experience in Python, TensorFlow, PyTorch, and Scikit-learn. Skilled in developing, deploying, and optimizing AI-driven solutions for demand forecasting and inventory optimization. Expertise in cloud computing (AWS), ML lifecycle management, and working with cross-functional teams to build scalable, production-ready enterprise applications.

## SKILLS

Tools	SQL, Python, C++, MS Power BI, Tableau, MS Excel, Streamlit, C#, .NET, HTML
Cloud & Big Data	Google Cloud Platform, Snowflake, AWS, Azure, Databricks, Spark, Hadoop
Database Management	MySQL, MS SQL Server, Azure Data Studio, PostgreSQL
IDE & Project Management	Docker, Jupyter NB, Visual Studio Code, GitHub, Google Colab, SharePoint
Python APIs/Lib	Pandas, NumPy, Matplotlib, Seaborn, StatsModels, Sci-Kit Learn, PySpark, TensorFlow, PyTorch, NLTK, Keras, LangChain, Transformers, HuggingFace
Expertise	ETL, Data Analysis, Data Visualization, Data Management, Machine Learning, Big Data, Cloud Computing, MLOps, Model Deployment, Data Pipelines, Predictive Analytics, Cross-functional Collaboration, Enterprise Applications, Supervised Learning, Unsupervised Learning, Generative AI, Natural Language Processing, Hypothesis-testing, Deep Learning, Scalable AI Reinforcement Learning with Human Feedback, Data Engineering, Supervised Fine-tuning

## EXPERIENCE

Junior Data Scientist, Leveragai Inc.	Dec 2023 - Present
<ul style="list-style-type: none"><li>Increased metadata generation efficiency by 30% and reduced manual intervention in image annotations by designing and deploying ML models for image captioning and context-aware retrieval using Salesforce BLIP, MS GIT, and ViT-GPT2 from Hugging Face.</li><li>Optimized LLM pipelines by implementing vector embeddings, Retrieval-Augmented Generation (RAG), and transformer-based models, significantly enhancing image-to-text system performance and ensuring scalability across large datasets.</li><li>Improved image feature extraction and caption generation accuracy by developing and fine-tuning deep learning models for multimodal tasks using PyTorch, Hugging Face Transformers, and TensorFlow.</li><li>Reduced model deployment time and ensured real-time inference by integrating and deploying vision-based models into Azure AI and AWS, leveraging MLOps best practices for scalable and efficient model serving.</li><li>Cut data processing time by 30% by designing end-to-end data pipelines for image and video processing, incorporating pre-processing, augmentation, and transformation techniques using OpenCV, Pandas, and NumPy, and optimizing training workflows for faster model convergence.</li></ul>	
Data Scientist, Smart Ecosystems Inc.	Jan 2024 - Present
<ul style="list-style-type: none"><li>Drove significant AI model improvements by leading a high-priority Reinforcement Learning with Human Feedback (RLHF) project, conducting in-depth qualitative analysis to identify and mitigate key loss categories, and boosting model refinement and performance.</li><li>Enhanced AI model adaptability and reasoning accuracy, leading to more precise decision-making and improved automation by leveraging advanced data pre-processing, feature engineering (Pandas, NumPy), machine learning (Scikit-Learn, TensorFlow, PyTorch), and NLP techniques (LangChain, Transformers, NLTK).</li><li>Accelerated decision-making by 20% by designing and deploying interactive Power BI dashboards that seamlessly integrated data from multiple sources, providing real-time visibility for cross-functional teams and enabling faster business insights.</li><li>Refined AI-generated content quality by developing ideal responses for Image Generation models, playing a pivotal role in the Supervised Fine-Tuning (SFT) of AI models, and ensuring more accurate, human-like outputs that aligned with user expectations.</li><li>Spearheaded quality control process improvements for AI model outputs, ensuring 95% accuracy in content generation through rigorous testing and validation using advanced QA frameworks over a 3-month period.</li></ul>	
Social Media Analyst, Redwood Algorithms	Jan 2021 - Dec 2022
<ul style="list-style-type: none"><li>Drove a 200% increase in customer engagement by leading data-driven ad campaigns for five business clients, utilizing predictive analytics, customer segmentation, and ML models to optimize targeting and personalize marketing strategies.</li><li>Improved lead generation and conversion rates by analyzing large-scale data sets from Google Analytics, Meta Ads, and Google Ads, applying A/B testing, time-series analysis, and clustering techniques to identify high-impact elements.</li><li>Enhanced marketing ROI by implementing SQL-based data pipelines to automate performance tracking, integrating data from Google Analytics, Meta Ads, and internal databases, and reducing manual reporting time by 40%.</li><li>Utilized data analysis and machine learning to enhance supply chain optimization, forecasting demand and reducing inventory costs by 15% over 6 months, leveraging tools like Python (Pandas, Scikit-Learn) and Tableau for insights.</li><li>Strengthened strategic decision-making by designing interactive Power BI dashboards, visualizing key digital marketing metrics such as click-through rates, engagement trends, and ad spend efficiency, enabling cross-functional teams to adapt campaigns in real-time.</li><li>Optimized campaign performance prediction models using Python (Pandas, NumPy, Scikit-Learn), machine learning algorithms, and NLP techniques, uncovering actionable insights that increased ad relevancy and customer retention.</li><li>Facilitated data-driven decision-making by documenting and presenting campaign performance insights via Microsoft Share-Point, collaborating with stakeholders to align marketing efforts with data-backed strategies.</li></ul>	

## PROJECTS

---

### CLV Prediction for Insurance Companies with Python, MS Power BI, AWS, SQL, Streamlit, and LangChain

- Achieved a 91% accurate CLV prediction model using Random Forest, implementing rigorous statistical hypothesis testing and regression analysis to optimize customer segmentation and personalized marketing strategies.
- Developed an interactive QnA Interface using LangChain and Google's Gemini LLM API, translating natural language queries into SQL for seamless non-technical access to complex data insights.
- Integrated AWS S3 storage with Python API access, establishing a scalable, cloud-based data pipeline to streamline CLV analytics and predictive modeling.
- Designed and deployed an interactive Power BI dashboard visualizing CLV for different insurance policies and premium tiers, enabling data-driven decision-making for targeted customer engagement strategies.
- Hosted a web-based CLV prediction platform, providing businesses with real-time forecasting tools to optimize customer retention and revenue growth.

### Breast Cancer Prediction with Computer Vision using PyTorch

- Developed and deployed a deep learning model leveraging Convolutional Neural Networks (CNNs) with TensorFlow and Keras to classify histopathological images for breast cancer detection, aiding in early diagnosis and treatment planning.
- Engineered a robust data pipeline for preprocessing medical images, including resizing, normalization, augmentation like rotation, flipping, contrast adjustment, and grayscale conversion, ensuring improved model generalization and performance.
- Trained the CNN model on high-resolution histopathological image datasets, fine-tuning hyperparameters (batch size, learning rate, dropout regularization) to optimize classification accuracy and reduce false negatives.
- Achieved a left-skewed probability distribution in model predictions, indicating a bias toward detecting malignant cases, prompting further model refinement through class-balancing techniques, weighted loss functions, and ensemble learning to improve sensitivity and specificity.
- Evaluated model performance using precision, recall, F1-score, and AUC-ROC metrics, analyzing classification thresholds to ensure a high recall rate, which is critical for minimizing missed cancerous cases.
- Deployed the trained model in a scalable cloud-based environment (Google Cloud/AWS) using Flask/Streamlit for a user-friendly interface, allowing medical professionals to upload images and receive real-time probability scores for malignancy.

### Automatic Traffic Light Management Application with AI sensors using C# in .NET framework

- Engineered an AI-powered traffic management system using .NET, C#, and MS SQL Server, achieving a 25% reduction in commute times, a 40% decrease in wait times, and a 20% drop in fuel consumption through adaptive signal optimization.
- Developed and deployed a scalable SQL Server database via Azure Data Studio, enabling real-time data storage, retrieval, and analytics, enhancing traffic signal responsiveness and congestion prediction.
- Implemented real-time traffic data collection using GPIO-connected sensors, facilitating dynamic traffic flow adjustments based on vehicle density and pedestrian activity.
- Leveraged Docker and DotNetEnv for containerized deployment, ensuring secure, modular, and scalable system implementation across diverse urban traffic environments.

### Steam Review Analysis on Big Data using Apache Spark and Hadoop in Python

- Led a team of three data scientists in analyzing an 8GB dataset with over 7 million records using HDFS and PySpark, identifying a 40% surge in user reviews on the Steam platform in November 2021, providing key insights into user engagement trends.
- Implemented extensive data cleaning, preprocessing, and feature engineering using SparkSQL and PySpark, leveraging Spark DataFrames and RDDs for scalable processing and real-time interactive visual analysis.
- Developed a personalized game recommendation model utilizing the Alternating Least Squares (ALS) algorithm from Spark MLlib, enabling tailored recommendations for new players based on their gaming history, improving recommendation accuracy and user retention.
- Optimized the model's hyperparameters (rank, regularization, iterations) to enhance prediction accuracy, employing cross-validation and evaluation metrics such as Root Mean Square Error (RMSE) to fine-tune recommendations.

### Portfolio Website Using HTML and CSS

- Escalated public visibility by 300% by developing a personal website, launched with Jekyll on GitHub that showcases career journey
- Drafted the website using HTML for content and CSS for styling and formatting

## EDUCATION

---

### University of Maryland Baltimore County (UMBC)

Master of Professional Studies in Data Science  
GPA: 4.0/4.0

Baltimore, MD

May 2024

### Pondicherry University

Master of Science in Physics  
GPA: 3.61/4.0 Valedictorian

Puducherry, India

May 2022