# PAVAN KUMAR DHARMOJU

+1 (224) 463-7625| pavandharmoju2024@u.northwestern.edu | pavankumardharmoju.me | linkedin.com/in/pavandharmoju/

## SUMMARY

Data Scientist with over 3 years of experience in analytics and data science, specializing in finance and marketing analytics. Proven expertise in leveraging data-driven insights to inform strategic decision-making and optimize business performance. Skilled in SQL, Python, statistical modeling, and data visualization. Strong expertise in A/B testing, statistical modeling, and translating complex analyses into actionable business insights. Experienced in working with cross-functional teams to drive product and marketing decisions.

## SKILLS

- **Data Analysis:** Python, R, JavaScript, SQL, Snowflake, Scala, Tableau, Power BI, Flourish, Excel
- **Programming Languages & Tools:** Hugging Face, AWS ECS, EC2, Kinesis, S3, Lambda, RDS, IAM, Docker, Mlflow
- **Machine Learning & Modeling:** Pandas, NumPy, Scikit-Learn, PyTorch, TensorFlow
- **Certifications:** Salesforce Certified Administrator, Tableau CRM & Einstein Discovery Consultant.
- **Frameworks:** Jenkins, Docker, Apache Airflow, LangChain
- **Methodologies:** Agile, Lean, Scrum, Kanban, Continuous Integration/Continuous Deployment (CI/CD)
- **Big Data:** Hadoop, Spark, MapReduce, Hive
- **Algorithms:** Decision Tree, Random Forest, Gradient Boosting, XGBoost, Logistic Regression, Linear Regression

## EDUCATION

**Northwestern University, Chicago -** M.S. in Artificial Intelligence. **August 2023- December 2024**
Coursework: *NLP, AI Practicum, Deep Learning, Design & Analysis of Algorithms, HCI, Data Science pipeline*
Projects:

- Scalable Sentiment Analysis for Content Publishers: Developed an NLP model to analyze over 1 million articles, providing actionable insights.
- Truck Fleet Optimization Using Reinforcement Learning: Built an RL model to optimize maintenance strategies, reducing operational costs.

**Indian Institute of Technology Madras -** Diploma in Data Science**.**
Publications: “Ranking System for All Tourism Related Industries Using NLP Approach”, IEEE ICCCNT, 2022. [Link], “Forecasting Electrical Demand for Residential Sector Using Deep Learning”, IEEE AIMV, 2021. [Link], “Ranking System for All Tourism Related Industries Using NLP Approach”, IEEE ICCCNT, 2022. [Link]

## WORK EXPERIENCE

**Intern, Chan Zuckerberg Biohub, Chicago, IL** June 2024 - September 2024

- Developed a Retrieval-Augmented Generation model using PubMed Central to enhance immunology research
- Fine-tuned LLaMA 3.1 and OpenAI models, integrating them into a web application for data analysis & generation.
- Designed scalable RESTful APIs using AWS, enabling large-scale data processing for bioinformatics research.
- Web-scraped 6,000,000+ files from multiple APIs improved their metadata and ingested them into the agent

**Data Scientist, Healee, Remote** September 2023 - April 2024

- Boosted lead generation by 23% through predictive modeling and A/B testing using SQL, Python, and Google Analytics.
- Built Marketing Mix Models (MMM) to optimize marketing spend using SQL and Python.
- Performed A/B testing and statistical analysis to evaluate marketing strategies and growth experiments.
- Automated ETL processes using AWS Lambda and Apache Airflow, reducing data processing time by 10%.

**Business Technology Analyst, Deloitte Consulting USI, Hyderabad,** August 2021 - September 2023
*Awards:* Earned SPOT Award for managing ML deployments and data pipelines while leading a 3-person data science team.

- Enhanced customer satisfaction by deploying ML models in Salesforce CRM for personalized recommendations.
- Streamlined deployment, cutting delivery time by 10% through automated CI/CD pipelines with Jenkins and Docker
- Eliminated 20 hours of manual work weekly by automating ETL workflows with Python, improving team efficiency.
- Optimized data processing by 18% by deploying predictive models on AWS, enabling faster client insights.

## TECHNICAL EXPERIENCE

**Real-Time Demand Forecasting System for Retail Supply Chain Optimization**

- Developed a forecasting system using LSTM and XGBoost, reducing stock outs by 18%. The system provided real-time updates and integrated with supply chain software for dynamic inventory management.
- Tech Stack: Python, TensorFlow, XGBoost, GCP, Docker, SQL.

**Product Categorization System for a Global Marketplace**

- Created a BERT-based model for multi-lingual product categorization, providing real-time insights and low-latency processing of millions of products.
- *Tech Stack*: Python, BERT, PyTorch, AKS, Azure Databricks, Docker, SQL.