VARSHA VATTIKONDA

+1 (805)-743-9755 | varshavattikonda@gmail.com | LinkedIn

SKILLS

Languages	Python, SQL, C#, R
ML Frameworks & MLOps	PyTorch, Tensorflow, Scikit-learn, AWS SageMaker, GCP (Vertex AI), Airflow
Data Processing	Spark, Databricks, Pandas
Tools	MSSQL, Splunk, Power BI, Salesforce

WORK EXPERIENCE

Senior Data Scientist - JP Morgan Chase, New York City, USA Jun'23 - Curr

- Partnered with Sales team to drive **$5M** annual revenue uplift; a **Random Forest** model that identifies overlooked, high-propensity contacts, enabling a targeted outreach strategy, boosting conversions by **5%**
- Uncovered **30%** more qualified leads for follow-on offerings using an **XGBoost** model trained on 13F, trading, and CRM data, identifying non-obvious client opportunities
- Increased sales engagement by **25%** by developing a scoring model using **3-layer MLP** (**Neural Network**) to prioritize alerts for **300**+ global sales reps
- Developed a deep learning model using **PyTorch** for stock recommendation by analyzing **10GB**+ trade history; implemented an **LSTM** with **Attention** mechanism, achieving **80% Average Precision**
- Improved client meeting fulfillment by **18%** by formulating the allocation challenge as a constrained optimization problem
- Developed a **statistical** model to identify stocks with surging investor attention, garnered enough trust and has featured in the daily, client-facing articles

Data Scientist - MSCI Inc., Mumbai, India Aug'16 - Jul'21

- Generated **$1.2M** in ARR by developing a batch recommendation system using **PCA and K-Means** to cluster clients and generate cross-sell opportunities
- Reduced customer churn by **10 %** by developing an **XGBoost** model leveraging behavioral features (e.g., spikes, dips in user engagement KPIs) and hence enabling proactive retention campaigns
- Improved user experience by exposing simulation durations (**MAE of 5 minutes**)**;** two-stage model, first segmenting client portfolios with **K-Means** and then apply a **Linear Regression** model to each cluster
- Accelerated the data science lifecycle by **15 %** designing a centralized data framework that consolidated **10**+ **sources** (10GB/day), facilitating single source of truth for clean, feature-ready data
- Engineered a sentiment analysis pipeline that processed 10,000+ client emails using an **LSTM** model deployed on **Azure Functions**, providing real-time sentiment signals
- Championed a data-driven culture by training **50**+ employees in Power BI and data modeling, empowering teams with self-service analytics and increasing platform adoption by **40%**

RESEARCH & INTERNSHIPS

Research Assistant - New York University, USA Jan'23 - May'23

- Developed a novel 3-step **NLP pipeline** using **BERT**-based models for sequence classification and topic modeling to automatically extract causal relationships from text
- Designed a custom data annotation survey using **JavaScript** within Qualtrics, accelerating the labeling of training data by **40%**

Data Science Intern - MSCI Inc. ESG NYC, USA May'22 - Dec'22

- Secured a **$1M** deal in a strategic partnership with **Google** (**GCP & Earth Engine**); engineered a two-stage flood risk model (**KMeans Clustering** + **Random Forest**) that served as the foundational PoC
- Deployed the end-to-end pipeline on **Google Vertex AI**, enabling scalable batch-inference for risk assessment across 300,000+ assets and proving the technical feasibility of the joint offering of Google and MSCI

EDUCATION

M.S. Data Science | New York University, USA (3.9/4) Sep'21 - May'23

Coursework : Natural Language Processing, Deep Learning (Computer Vision), Big Data (Spark)

BTech & MTech Electrical Engineering | Indian Institute Of Technology Madras (IITM) Aug'11 - May'16

Coursework : Data Structures & Algorithms, Machine Learning, Linear Algebra, Probability and Statistics