**Venkata** Medabala

London | +44-07466612493 | [venkat2200@gmail.com](mailto:venkat2200@gmail.com) | [linkedin.com/in/venkata-sm](https://www.linkedin.com/in/venkata-sm)

Highly-skilled Data Scientist with a proven track record of delivering data-driven solutions across diverse industries including Retail, Public Sector, Healthcare and Financial Services. Successfully transitioned from data science to ML Engineer and excel in developing impactful AI/ML-powered products from ideation to production. Expertise in Python, MLOps, Machine Learning, GenAI, and cloud technologies.

**Technical Skills**

* Languages: Python, Rust, R, SQL, NoSQL, BigQuery, SAS, bash
* Frameworks: PyTorch, TesnorFlow, LangChain, Llamaindex, transformers, spaCy, scikit-learn, Streamlit, FastAPI
* Machine Learning, LLMs, Generative AI, Deep Learning, Statistics, Bayesian Statistics, Forecasting
* MLOps: CI/CD, GitHub, GitLab, Docker, kubeflow, Vector DB, MLflow, Comet
* Cloud: Azure,GCP ,AWS, Qdrant
* Project Management/ Documentation: Jira, Confluence, Miro

**Experience**

**Machine Learning Engineer, UBS, London Aug 2024 - Present**

* **Implementing AI-driven semantic search and Conversational AI using Azure and LLMs to streamline legal processes, improve decision-making, and support legal teams.**

**Sr Data Scientist/ML Engineer, NIQ-GFK, London Apr 2022 – Jul 2024**

* **Collaborated** with stakeholders, cross functional teams to **define requirements,** optimize, maintain and monitor the pipelines.
* Led the price-promo team as a **squad lead**, managing 3 data scientists and overseeing **project plannin**g and development using **Jira**.
* Developed and Deployed **RAG based LLM application** to extract product, various promotional information (bundle, cashback, discount, etc.) from unstructured data. This resulted in a **9% increase in accuracy** and 1**2% wider product coverage**, enabling precise pricing insights across promotional uplift, baseline sales, and price elasticity.
* Fine-tuned llma2 7B model using LoRA, leading to optimized retrieval accuracy and efficiency in the RAG system, enhancing the quality of extraction process.
* Deployed multiple pricing insights and **ML models** (price elasticity, price promotion, price decomposition, competitor analysis) in GCP from research to production utilizing **MLOPs** practices.
* Demonstrated expertise in production-level **Python coding - OOP, SOLID principles, TDD,** adhering to software development best practices, and conducting successful code reviews to enhance code quality.
* Identified and **resolved bugs**, ensuring the stability and reliability of the deployed models.
* Effectively **presented** complex analyses and explained intricate topics to product owners in sprint reviews.

**Data Scientist, Cognizant, London Jun 2016 - Mar 2022**

* **Designed and deployed real-time, global food risk identification system for the UK Food Standards Agency leveraging LLMs to automate 90%+ of the previously manual process.**
* Built **web scraping modules** to scrape food recalls, outbreaks, border rejection from various data sources including non-English articles using Python
* Built **NLP pipelines** as **micro services** to handles Language translation, food article detection, Opinion articles detection, duplicate information flagging, food/feed/FCM classification using finetuned **BERT models**.
* Extract Food Name, Country of Origin and Hazard, manufacturer and other key information from the articles using **spaCy** & **sentence embedding** techniques.
* Setup pipelines as micro services and integrated with **CI/CD pipelines** using **GitHub** actions
* **Customer Segmentation & Customer Lifetime Value Prediction (Ecommerce)**
* Built models to **predict customer lifetime value** and future purchases using ensemble model random forest.
* **Customer segmentation** based on recency, frequency, monetary and demographic information to identify high value customers for better marketing strategies.
* **Three Level Text Categorization**
* Built an automated **text categorization** solution using NLP techniques **Word2Vec embeddings** that aims to categorize the user comments to open-ended survey questions into relevant product categories, subcategories, and conditions.
* **Patient Risk Stratification**
* Performed Risk Analysis on member data using electronic healthcare record and segment them into different risk levels (high risk, low risk, rising risk) using **fuzzy c clustering**.
* Identified the key features contributing to member transition between various stages (low risk to medium, medium to high risk) using **ensemble models**.
* Design & implemented a **Bayesian** framework to personalize treatment plans for patients with comorbidities

**Assistant Manager, HSBC, Bangalore**  **Oct 2014 - May 2016**

* **Credit Risk Scorecard Model**
* Developed acquisition & behavioral **scorecard models** for credit card & home loan portfolio using SAS
* Conducted score cut-off analysis, contributing to 7% reduction in credit risk.
* Documentation of model development and validations of scorecards
* Review & revise credit policy and customer segmentation risk profiling on a regular basis.

**Business Analyst, Hewlett Packard (HP), Bangalore**    **May 2011 - Sep 2014**

* **Warranty Fraud Detection using ML which has a business impact of $20M.**
* Segmentation of warranty claims into noncompliant, highly suspicious, suspicious and compliant claims employing advanced clustering algorithms such as fuzzy-c and k-means
* Developed predictive models to detect/predict noncompliant warranty claims by using machine learning techniques such as SVM, random forest, and naïve bayes classifier using R
* **Warranty Units & Customer Service Order Forecasting**
* Developed Forecast model to forecast customer service orders and warranty units using time series techniques and built a simulator tool in excel.
* **Customer Experience Key Driver Analysis**
* Built Structural Equation model (SEM) to estimate cause-and-effect relationships of satisfaction-loyalty chain and identified what is important to customer Satisfaction and Loyalty intentions.
* Developed SAS Macros for Shapley Value Regression, to measure the importance of attributes in key driver analysis.

**Research Assistant, Indian School of Business (ISB), Hyderabad**   **May 2008 - Apr 2011**

**Research Assistant, INSEAD, France**  **July 2010 - Aug 2010**

**Statistical Trainee, Indian Statistical Institute (ISI), Kolkata**   **Dec 2007 - Mar 2008**

**Education**

**Master of Science in Statistics-OR,** *University of Hyderabad, Hyderabad***,**   **May 2007**

**Bachelor of Science in Mathematics & Statistics,** *Andhra University,*  **May 2005**

**Conferences & Presentations**

* “Doc as Code – An Approach to Writing & Managing Documentation”*, we.innovate, GFK, London, April, 2024*
* “Static Code Analysis: Best Practices and Tools”*, innovate ,GFK, London, Sep 2023*
* “Measuring RelativeImpact of Key Drivers – A Game Theory Approach”,*4th IIMA International Conference, 2015.*
* “Predictive Modeling of Non-compliance Detection”, *3rd IIMA International Conference on Advanced Data Analysis, 2012.*

**Professional Development**

* Deep Learning Specialization by Andrew NG, Coursera.
* MLOps Specialization by DeepLearning.AI, Coursera
* Algorithms and Data Structures, MIT OCW