Hamza M Zubair
Senior Lead Data Scientist

+91 7600953379 | hamzamohdzubair@gmail.com

Data scientist with **8**+ years of experience in deploying advanced ML solutions in a variety of industry verticals such as **Retail**, **BFSI** and **QSR**. Successfully productized and deployed **ML** models involving Supervised Learning, Reinforcement Learning (**RL**), Deep Learning (**DL**) Natural Language Processing (**NLP**), Computer Vision, Markov Models, Time Series, Forecasting, **LSTM** models and Large Language Models (**LLM**).

Experience

Data Science and AI Consultant, XCCELER (*Freelance*)   *Apr 2025 - Present*
- Provided strategic Data Science consulting focused on optimizing client marketing and CRM data.
- Implemented predictive modeling and segmentation strategies to identify customers most likely to purchase, significantly reducing marketing ad spend for clients through more targeted campaigns.
- Advised on effective use of customer data analytics to improve ROI on marketing campaigns for multiple clients.

Data Scientist, *Turing.com*   *Aug 2024 - Apr 2025*
- *Part of the team working on training **Gemini** using **SFT**.*
- *__Complex Chart Production Automation Tool__ (Team size: 1): Sole designer and developer of the tool, that reduced the chart generation time, by **93.3%** from 45 min to 3 min. The tool uses LLM pipelines with a costing of around 0.4 rupee per chart.*

Senior Lead Data Scientist, *Cartesian Consulting*   *Aug 2018 - Aug 2024*
- **Campaign Prioritizer** *(Team size: 1)*: Built using contextual multi-armed bandits (**CMAB**), **Thompson Sampling** and **UCB** Algorithm. Deployed in 2 major clients (**Pharma Retail and Brokerage**), leading to improved click-through rates and revenue growth by ensuring that the right marketing campaign is sent to the right customer.
- **Automated MLOps Platform** *(Team size: 2)*: Built a model builder platform that allowed users to configure, maintain and deploy supervised ML models like **XGBoost**, **LightGBM**, etc., automatically performing feature engineering, model tuning, calibration, retraining and scoring, feature importance computation etc. Implemented in nearly all Solus clients.
- **SMS Composer** (*Team size: 1*): Single-handedly responsible for building the generative NLP module and UI for marketing outreaches through SMS. The tool breaks down message structure based on purpose, tonality and the presence of personalization fields, and is capable of generating several variants of the same message depending on a given prompt.
- **Multi-Touch Attribution** (*Team size: 1*): Built a tool for multi-touch attribution of campaign responses. The tool adapts known techniques (Markov models) for attribution used in digital marketing to the CRM and direct marketing domain.
- Open Rate Predictor (Team Size: 3): Led the team that built an NLP pipeline to analyse email subject lines and predict their open rate. The tool also included an **RNN** generator to compose the most optimum Email Subject Line using Semantically Controlled **LSTM**, and a semi-structured prompt as input. The tool was deployed for a major Indian bank.
- Automatic Product Segmentation Tool (Team size: 1): Built using **NLP** Techniques like Latent Dirichlet Allocation (**LDA**) and Topic Modelling. It streamlined product categorization, enabling targeted marketing efforts leading to improved customer segmentation, and enhanced overall campaign effectiveness.

Founding Employee & AI Lead, *Untrodden Labs (Defunct)*   *Jan 2017 - Aug 2018*
- *Founded as a Bootstrapped Startup by a group of friends in an informal setting to build an Alexa-like personal desktop bot with voice commands, vision, face tracking and an LED screen that displayed emotions using a game engine. Now dissolved.*
- **Face Tracking and Speech Understanding System** (*Team size: 5*): Lead the team in architecting and building multiple modules: Speech Intent Classifier module using **SVM** and Random Forests (**RF**), a Novelty Detection module using **OneClassSVM**, Information Extraction module using **Grammar-Based Parsing**, **Dependency Parsing** and **SyntaxNet**, Object-Detection module to detect over 90 objects in live-streaming video using **YOLO, MobileNet SSD**, and **ResNetv2.**

Achievements
- **Awards**: *Employee of the Quarter* (*Cartesian*, Q1, 2020), *Innovation Award for work in NLP* (*Cartesian*, 2020**)**
- **Talks**: St. Claret College (2025), *Hackertech* (VIT Vellore 2019), *GIST Tech Connect* (American Center, Delhi, 2017)

Education

Woolf University   *2023- 2025*
    MS in Computer Science specialization in ML/AI

IISC Bangalore   *2019 - 2020*
    Certification in Reinforcement Learning

IIT Gandhinagar   *2013 - 2015*
    MSc in Cognitive Science

VIT Vellore   *2009 - 2013*
    BTech in Bioinformatics

Skill
- **Languages:** *Python, Rust, SQL, R*
- **Frameworks:** *numpy, pandas, scipy, keras, tensorflow, pytorch, vowpalwabbit*