

SIDDHANT TANDON



siddhant.tandon203@gmail.com



[LinkedIn - Siddhant Tandon](#)



Milan, Italy



[sid203.github.io](#)

About me

I am a data scientist with 6+ years of experience across various sectors including **Banking, Manufacturing and Telecom**. I have been thoroughly enjoying my professional journey so far which has been a good **balance of ML and software development**.

My experience consists of designing & implementing ML algorithms for business needs and R&D projects.

Over the last couple of years I have expanded my role to cover **complete lifecycle of application development for AI use cases**, gaining significant experience in **backend development**.

I have also been enjoying the privilege of collaboratively **shaping the business needs** & translating these into long term development roadmaps.

With every challenge I faced in my career I have learned to look beyond the backlog, questioning the feasibility & scope of objectives to fully grasp the broader context and business impact.

Work Experience

2023 - PRESENT | SCRATCH & SCREEN

Senior Data Scientist & Backend Dev

Currently I am leading product development of an **invoice management platform that allows brands to access B2C marketing insights**. My activity is backed by **active R&D**, aimed at improving software with use of **SoTA AI models**.

The future vision for the product development involves use of **LLM models like LayoutLMv3 and PaliGemma**, on which i've been conducting preliminary analysis. To do so, i've focused on improving our current **data gathering pipelines, setting up persistent storage and modernizing our deployment strategies via containers**.

Key part of my role is **stakeholders management** to achieve optimal understanding of business requirements, along with **prioritizing most valuable product increments** to be implemented.

Some of the key achievements in my current role:

- **Data gathering improvements** - developed & deployed python web service to parse unstructured invoice data, achieving reliable parsing for **~80%** of good quality images. Also developed & deployed deskewing algorithm to correct skewed invoice images which brings **~10%** improvement in data parsing.
- **MVP development & Client engagement** - Developed & deployed an MVP that enables user login, invoice upload & parsing, full text search & analytics via **ElasticSearch**. Also directly engaged in presenting the MVP to potential clients to seek collaboration & partnership.
- **Fully self hosted CI/CD & containerization** - Designed & implemented a secured jenkins CI/CD pipeline using docker & socat. Automated deployments via docker swarm including automated management of nginx config & SSL certificates.

2019 - 2023 | CGNAL SRL

Data Scientist

As a data scientist & consultant my role at Cgnal was highly focused on developing **tailored ML solutions for the client's use cases**. In this experience I collaborated with two major clients in the banking & manufacturing sector.

It was during this role that I expanded my expertise by actively contributing to **end-to-end application development for 2+ years**. Being part of a cross-functional team I gained immense **exposure to software principles such as TDD, DDD (domain driven design), Monads & Asynchronous programming**.

My longest and most important project created a **Refinitiv News Tagging application**, a portfolio of news ingested from multiple data providers such as Refinitiv and S&P.

Skills

Programming	Python, R, Java
Machine Learning	scikit-learn, pandas, numpy, tensorflow, pytorch
Databases & search	MongoDB, SQL, Elasticsearch, Kibana
Cloud & DevOps	Docker, Docker Swarm, Git, Jenkins, GCP (Cloud Run, GCE), Nginx
Software Eng Principles	TDD, OOP, Async, Monads, ORM

Languages

Hindi	Mothertongue
Italian	C1
English	C2

The project included **news tagging & topic enrichment** tailored to match the client's database of digital products. The application was equipped with a search engine powered by Tough-2-beat embeddings to lookup most relevant news articles in near-real time.

2018 - 2019 | NOKIA BELL LABS

Research Intern - Deep Reinforcement Learning

I was chosen as a research intern at Nokia Bell Labs under the Causal Data Analytics dept to write my masters thesis. The research was focused on exploring resource allocation strategies in cloud computing under unpredictable workloads via Deep Reinforcement Learning.

Specifically I was responsible for implementing **Deep Q-learning** to address efficient resource allocation problem. To further strengthen my experiments I was also responsible for evaluating this approach to strong theoretical baselines such as **Q-learning, Value Iteration & Policy Iteration**.

I was able to successfully develop a software written entirely in python and tensorflow that can simulate a Reinforcement Learning environment, generate appropriate rewards, setup theoretical baselines, train Deep Q-learning algorithm in tensorflow and generate reports on obtained rewards under simulated MDPs (markov decision processes).

Education

Masters in Data Science

2016 - 2018 | SAPIENZA UNIVERSITY OF ROME

Courses highlight: Information Retrieval, Parametric & Non parametric Bayesian modelling, Statistical Learning, Stochastic Processes

BTech in Electronics & communication engineering

2012 - 2016 | AMITY UNIVERSITY NOIDA, INDIA

Courses highlight: Numerical methods & optimization, Discrete mathematics, Digital Signal Processing