RISHABH PATIL

+1 929-424-7773 • rbp5812@nyu.edu • linkedin.com/in/rishabhbhaskarpatil • github.com/rishswish • rishswish.github.io

EDUCATION

New York University, Center for Data Science September 2024 – May 2026
M.S. in Data Science GPA 3.78/4.0
Coursework: Natural Language Understanding,Big Data,Machine Learning,Computational Linear Algebra

University of Mumbai July 2020 – May 2024
B.Tech in Computer Science and Engineering (Data Science) with Honors in Computational Finance GPA 3.96/4.0
Coursework: Time-Series,Cloud Computing,Deep Learning,Database Management,Computer Vision,Reinforcement Learning

SKILLS

Technologies: Python, R, SQL, Git, TensorFlow, PyTorch, SciKit-Learn, OpenCV, PySpark, Hadoop, Dask, Kafka, BigQuery, CUDA, Cassandra, Kubernetes, Snowflake, Jenkins, Docker, PowerBI, Tableau, SQLite, AWS, MongoDB, MS Excel, YOLO

PROFESSIONAL EXPERIENCE

Solar Secure Solution, Karnataka, India| Generative AI Intern February 2023 - April 2023

- Engineered an intelligent RAG chatbot with **GPT-3, LangChain, Docker, and Kubernetes**, then wired into a **CI/CD pipeline–accelerating** end-to-end software-review cycles by **40%** and **cutting development costs 10%**.
- Architected **prompt-engineering modules** that **auto-generate** charts, slide decks, and multilingual summaries; integrated these as **Jenkins** build steps that **eliminated 75%** of manual review effort and **expanded** cross-team visibility.
- Designed and embedded a **12-Factor compliance checker** into the CI pipeline to **audit** for statelessness, configuration, and logging standards, **blocking 95%** of non-compliant builds and **streamlining** cloud-migration readiness decisions.

Acmegrade Pvt Ltd, Karnataka, India | Machine Learning Intern July 2022 - September 2022

- Developed a Python-based **Computer Vision + NLP pipeline** using **Azure OCR** to auto-extract key data fields from PDFs, screenshots, and log files—**reducing processing time 25%** while increasing downstream analytics throughput.
- Spearheaded a company-wide ML upskilling program for **100 engineers**, with curated notebooks, Dockerized environments, and cloud GPUs; **lifted skills-assessment scores 20%** and achieved an **85% lab-completion rate.**
- Co-developed a collaborative recommendation engine (PySpark + ALS) that delivered personalized product suggestions across 50+ SMB clients, raising partner revenue 15 % and expanding user engagement metrics.

ACADEMIC PROJECTS

MovieLens Recommendation & Segmentation| Github January 2025 – May 2025

- Deployed a terabyte-scale **PySpark + Hadoop HDFS** pipeline that ingested the complete **MovieLens corpus—330 K users** / **86 K movies**—enabling interactive analytics and large-batch model training across a multi-node cluster.
- Segmented users via a **MinHash + LSH** workflow that trimmed billions of pairwise checks to sub-second latency; the resulting top-100 “movie-twin” pairs showed a **2.3 × stronger preference alignment** over random matches.
- Engineered two recommendation engines on temporal splits: a **Spark ALS** collaborative filter that delivered **+30 % Precision@100** on cohorts with $>$ **20 % rating coverage**, and a **bias-corrected popularity model** tuned for **90 % sparsity**, achieving a **+66 % MAP** lift over the naive baseline and **20 × higher MAP** than ALS in ultra-sparse segments.

Progressive Learning in LLMs with Structured Grammar Books| Github January 2025 – May 2025

- Curated a **345-lesson curriculum** from *New Concept English* using Tesseract OCR and Stanza, generating **1.7 K** syntax feature vectors (POS, DEP, NER, morphology) that fuel progressive, syntax-aware LLM training.
- Built Transformer variants **(SyntaxGPT, SyntaxT5)** by concatenating token + syntax embeddings and running a curriculum$\rightarrow$ fine-tune pipeline in PyTorch/Hugging Face, cutting pre-training time from **2.5 days to 3 hours (-95 % compute)**.
- Validated on the TREC question-classification benchmark: **SyntaxT5 hit 87 % accuracy**, delivering **52 % faster inference (236 s $\rightarrow$ 114 s)** over baseline models while ensuring smoother convergence and stronger generalization in low-resource settings.

Personalized Recipe Recommendation System| Github July 2023 – May 2024

- Built a **Flask** interface backed by a **GPT-4 + text-embedding-ada-002 + LanceDB** RAG pipeline, driving a **35 % jump in user engagement**, **40 % higher recipe-match accuracy**, and **25 % fewer irrelevant suggestions.**
- Designed an allergy-aware cosine-similarity scorer plus real-time feedback loop that **eliminated cold-start & hallucination issues** and **boosted user-satisfaction scores by 50 %**.
- Orchestrated cloud workflows with **LangChain, Pandas, NumPy**, serving real-time recommendations to **10 K+ sessions**.

Driver Drowsiness Detection System| Github January 2022 – January 2024

- Trained dual **YOLOv5** models (eye-closure & yawning) on **1.2 K+ annotated images**, achieving **85 % accuracy** and **30 % faster alert-response** via real-time probability scores and voice alarms.
- Integrated a CNN-fusion layer that **reduced false positives by 20 %** and **improved accuracy by 15 %**, processing **1 M+ video frames** under low-light and occlusion conditions.
- Awarded **3rd Prize** in the **AI & Deep Learning track at ICDMAI 2024**; findings published in Springer LNNS 998 *Article* .