

# Pranesh Mukhopadhyay

Mumbai, India

pranesh.mukhopadhyay@outlook.com | 96747-70912

[github.com/Mukhopadhyay](https://github.com/Mukhopadhyay) [linkedin.com/in/praneshmukhopadhyay](https://linkedin.com/in/praneshmukhopadhyay) [kaggle.com/praneshmukhopadhyay](https://kaggle.com/praneshmukhopadhyay)

## INTRO

Backend and applied ML engineer with 5.5+ years of experience designing and deploying production-grade Python systems. Specialized in FastAPI-based microservices, real-time decision systems, and ML-powered APIs used in high-throughput environments. Strong background in data pipelines, model lifecycle management, and scalable backend architecture, with hands-on ownership from design to deployment.

## EXPERIENCE

### BOXWISH MEDIA LLP. / SPIRITZONE | CONSULTANT - DATA ENGINEERING

Dec 2021 – Present | Mumbai

#### REAL-TIME ORDER ASSIGNMENT & DISPATCH SYSTEM

Python • FastAPI • Celery • Redis • Docker • PostgreSQL

- Designed and deployed an **asynchronous real-time order assignment system** for dynamic batching and delivery partner allocation based on availability, load, and geographic constraints.
- Implemented ranking and batching logic capable of handling 30+ orders per minute per instance, including real-time agent discovery and assignment while accounting for operational and geographic constraints.
- Built asynchronous workflows using **Celery** and **Redis**, with centralized logging and comprehensive testing to ensure reliability in production.

#### ORDER ETA PREDICTION SERVICE

Python • FastAPI • TensorFlow • Pandas • Numpy • MLFlow • Redis

- Built a **stateless ML service** for real-time delivery ETA prediction, exposed via a low-latency **FastAPI** endpoint.
- Developed feature pipelines using historical order data, delivery partner availability, and traffic signals, with scheduled retraining and lifecycle management via **MLflow**.
- Optimized inference performance using geofence-based **Redis caching** and improved ETA accuracy by 75%, reducing delivery-related customer support tickets.

#### AGENT-BASED SKU ONBOARDING & VALIDATION PIPELINE

Python • LangChain • FastAPI • PostgreSQL • Redis • Docker

- Designed an automated, agent-driven pipeline to onboard new SKUs, replacing manual, error-prone operational workflows.
- Used **LLM**-based agents (e.g., **GPT-4** / **GPT-3.5** -class models) with deterministic guardrails to normalize metadata, validate attributes, and detect duplicates before database insertion.
- Added end-to-end observability across validation stages to track failures, confidence scores, and processing latency.
- Reduced manual onboarding effort by **16-20 hours per week**, saving hundreds of operational hours annually while improving catalog quality.

#### PLATFORM RELIABILITY & ENGINEERING PRACTICES

- Introduced centralized logging and structured error reporting across services to improve debuggability and incident response.
- Established unit and integration testing standards using **Pytest**, achieving **80–90%+ coverage** across critical services.
- Supported **load** and **stress testing** using **Locust** to validate system behavior under concurrent traffic.

### QUOSPHERE INFOSOLUTIONS PVT. LTD. | JUNIOR DATA SCIENTIST

Jul 2020 – Dec 2021 | Navi Mumbai

#### SPIRITZONE PRODUCT RECOMMENDATION SYSTEM

Python • Flask • Collaborative Filtering • Content based Models

- Built and deployed a **hybrid recommendation system** combining collaborative filtering and content-based models using user browsing behavior and feedback signals.
- Implemented incremental training pipelines to update models from live customer interaction data continuously.
- Exposed recommendations via stateless Python services, designed for maintainability and production use.
- Established unit and integration testing to ensure reliability of model outputs in production.

## MARKET BASKET ANALYSIS SERVICE

Python • FastAPI • Apriori • FP-Growth • Docker

- Developed a market basket analysis service using **Apriori** and **FP-Growth** algorithms to uncover frequent item associations.
- Automated daily model training and redeployment through containerized pipelines using **FastAPI** and **Docker**.
- Achieved 90%+ test coverage and validated service performance through load testing under concurrent traffic.

## CLICKSTREAM DATA ANALYTICS PIPELINE

Azure Data Explorer • Python • Grafana

- Built a scalable **clickstream data ingestion pipeline** using **Azure Data Explorer** and Python SDKs to process large volumes of event data incrementally.
- Implemented scheduling and monitoring to trigger downstream reports upon successful ingestion.
- Designed adaptive ingestion controls to handle variable data rates without data loss.

## OPEN SOURCE PROJECTS

### MONOLG | CENTRALIZED LOGGING LIBRARY FOR PYTHON

Python • MongoDB • Github Actions • Pytest

- Built and published a plug-and-play centralized logging library compatible with Python's standard logging module, enabling easy adoption of structured logging across services.
- Distributed via PyPI with CI-backed testing and comprehensive documentation on Read the Docs.

 [github.com/Mukhopadhyay/monolg](https://github.com/Mukhopadhyay/monolg)

### YOUTUBE SUBTITLE DATASET | THE MOST VERSATILE YOUTUBE TRANSCRIPTION DATASET

- Curated and released a dataset of 2,500+ YouTube video subtitles with manually annotated categories and enriched metadata.
- Suitable for NLP, classification, and content analysis use cases. Distributed via Kaggle.

 [kaggle.com/datasets/praneshmukhopadhyay/youtubers-saying-things](https://kaggle.com/datasets/praneshmukhopadhyay/youtubers-saying-things)

## EDUCATION

### ST. XAVIER'S COLLEGE | M.Sc IN COMPUTER SCIENCE

Aug 2020 | Kolkata, WB

### UNIVERSITY OF CALCUTTA | B.Sc IN COMPUTER SCIENCE

June 2018 | Kolkata, WB (Asutosh College)

## SKILLS

### Languages

Python • JavaScript • TypeScript • C++

### Backend & APIs

FastAPI • Flask • RESTful APIs • Asynchronous processing • Celery • Pydantic

### DevOps and other tools

Docker • Kubernetes • CI/CD (Github Actions, Azure Pipelines) • Monitoring (Grafana)

### LLM & Agent Systems

LangChain • LangGraph • PydanticAI • Agent-based workflows • Validation pipelines • Retrieval-augmented systems (RAG)

### Databases & Caching

PostgreSQL • MongoDB • MySQL • Redis • Pinecone • pgvector • Qdrant

### Machine Learning & Data

TensorFlow • PyTorch • scikit-learn • Pandas • NumPy • Feature Engineering • Model Evaluation

### ML Systems & MLOps

MLflow • Weights & Biases • Experiment Tracking • Model Versioning • Scheduled Retraining • Inference Optimization

## CERTIFICATES

### Machine Learning by Stanford University

 [Coursera](#) • Certificate ID: 636WHFRBMH5V

### Data Science Math Skills by Duke University

 [Coursera](#) • Certificate ID: AFQSU5CAT4HD