

# Pranesh Mukhopadhyay

Mumbai, India

praneshmukherjee7@gmail.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

I'm a data scientist with three years of experience. I have a strong background in statistics and machine learning, with a focus on data manipulation, analysis, and modeling. I've worked on diverse projects, utilizing Python, Pandas, and Numpy for data preprocessing and analysis. I also have hands-on experience with TensorFlow and PyTorch for building and training deep learning models. With excellent communication skills, I effectively translate complex concepts to non-technical stakeholders. I'm passionate about staying updated with the latest advancements and contributing to impactful projects.

## EXPERIENCE

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

Dec 2021 – Present | Navi Mumbai

#### ORDER ASSIGNMENT SYSTEM

Python • FastAPI • Celery • Docker compose • Redis

- Real-time order assignment system made using **Python**, **FastAPI**, and hosted and container orchestration handled using **Docker compose**, supporting Webhook listener & dynamic trip creation.
- Intelligent algorithm used to dynamically perform order batching based on delivery partner availability and order volumes. Maintaining a fair and unbiased delivery agent ranking system.
- Added support for centralized logging using **MongoDB** and end-to-end unit testing using **Pytest** & **Flake8**

#### PRODUCT SEARCH API

Python • FastAPI • Scikit Learn • Bag of Words model

- An incrementally trainable search API, which can index products even with their attributes, developed using **Python** & **Scikit-Learn** using the Bag of Characters representation of the data
- Hosted using **Gunicorn** for maximizing concurrent requests
- 80%+ Unit test coverage using **Pytest** & Passed swarm test conducted using Locust

#### ORDER ETA CALCULATION API

Python • FastAPI • TensorFlow • Pandas • Numpy

- A fast, interpretable stateless microservice for providing real-time order delivery time estimates.
- Built using **Python**, **FastAPI** & **TensorFlow** and hosted using Docker Compose for easy container orchestration.
- Maintaining model lifecycles using **MLOps** and testing the service using **Pytest** and **Locust**

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

Jul 2020 – Dec 2021 | Navi Mumbai

#### Spiritzone Product Recommender System

Python • Flask • Collaborative Filtering • Content based recommender system

- A hybrid recommender system trained using user's product browsing patterns and feedback path drawn from customer engagement platforms where the models are incrementally trained using the customer interaction data.
- Highly scalable and easily maintainable stateless microservices developed using Python and Flask. Capable of handling 30+ requests per second asynchronously.
- End to end unit & integration tested using python's built-in unit test library

#### Market Basket Analysis

Python • FastAPI • Apriori Algorithm • Docker compose • FP-Growth Algorithm

- Developed using the **Apriori** and **FP-Growth** algorithm for efficient, scalable and maintainable stateless microservice.
- Used FastAPI and Docker to make the deployment process easier alongside a complete build pipeline, capable of training the models on a daily basis.
- Unit tests with 90%+ code coverage & successfully handled load testing with 30+ concurrent requests per second (conducted using Locust).

#### Clickstream Data Analytics Pipeline

Azure Data Explorer • Python • Grafana

- A data pipeline using azure's python SDK and Azure Data Explorer to move huge volume of data incrementally.
- Scheduled reports that gets sent out everytime data ingestion process has been finished.

- Dynamically controlling frequency of ingestion without losing any data.

## EDUCATION

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

Aug 2020 | Kolkata, WB • Cum. GPA: N/A

#### Curriculum

Artificial Intelligence • Data Structures • Algorithms - Design & Analysis • Data Warehousing • Compiler Designing • Database Management System • Languages: Java, C++ • Distributed System

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

June 2018 | Kolkata, WB (Asutosh College)

Dean's List (All Semesters) • Cum. GPA: 3.92 / 4.0 • Major GPA: 3.94 / 4.0

#### Curriculum

Data Structures • Algorithms • Database Management System • Automata • Operating Systems • Computer Networking • Object Oriented Programming • Linux Programming (Shell scripting) • Languages - C, C++

## OPEN SOURCE PROJECTS

### MONOLG | AN EASY TO USE CENTRALIZED MONGO DB LOGGING LIBRARY

Python • MongoDB • Github Actions • Pytest

A Simple easy to use centralized logging library for Python using MongoDB. Deployed at Python Packaging Index aka. PyPI at [pypi.org/project/monolg](https://pypi.org/project/monolg). The library was written so that any Python developer who uses the default logging library can get started with Monolg as easily. With complete documentation and user guide hosted in Read the docs at [monolg.readthedocs.io](https://monolg.readthedocs.io).

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

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

Dataset containing 2500+ unique popular YouTube video subtitle, with manually annotated video category, that can be used as the dependent variable. Alongside the video categories we have various different video related attributes such as duration, channel name, subscriber count etc.

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

## SKILLS

#### Programming

Python • C++ • JavaScript

#### Libraries & frameworks

TensorFlow • PyTorch • Pandas • Numpy • Scipy • FastAPI

#### Databases

MSSQL • PostgreSQL • MongoDB • MySQL

#### DevOps and other tools

Docker • Kubernetes • Github Actions • Azure Pipelines • Grafana

## CERTIFICATES

### Machine Learning by Stanford University

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

### Data Science Math Skills by Duke University

🌟 [Coursera](#) • Certificate ID: AFQSU5CAT4HD

### Introduction to TensorFlow for Artificial Intelligence by DeepLearning.AI

🌟 [Coursera](#) • Certificate ID: E3XDATA99F3Y