Pranesh Mukhopadhyay

Mumbai, India

praneshmukherjee7@gmail.com | 96747-70912

github.com/Mukhopadhyay   linkedin.com/in/praneshmukhopadhyay   k kaggle.com/praneshmukhopadhyay

INTRO

I'm an accomplished data scientist with 3.5 years of experience, adept at leveraging statistical analysis and machine learning techniques to drive data-driven insights. Proficient in Python, Pandas, and Numpy for robust data manipulation and analysis, I've successfully delivered impactful solutions across diverse projects. I also have hands-on expertise with TensorFlow and PyTorch for building and training cutting-edge deep learning models. Known for my strong communication skills, I excel at translating complex concepts to non-technical stakeholders, facilitating seamless collaboration. Passionate about staying at the forefront of industry advancements, I'm eager to contribute to innovative and impactful projects.

WORK 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 **MLFlow** 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 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.

SKILLS SUMMARY

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 • Apache
Airflow • MLFlow

EDUCATION

ST. XAVIER'S COLLEGE | M.Sc in Computer Science
Aug 2020 | Kolkata, WB, India
Passed with First-Class in Honour
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)
Passed with First-Class in Honour
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 . 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 .
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

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