SHUBENDU BISWAS

+91 8968683332 ⋄ Chandigarh, IN

biswasshubendu4@gmail.com ⋄ Linkedin ⋄ Portfolio ⋄ Github

SUMMARY

With over 4+ years of experience, I excel as a highly skilled Data Scientist specializing in AI engineering, particularly in Generative AI, NLP, and machine learning. I have a demonstrated history of deploying sophisticated ML models in production settings and exhibit adeptness in both Django Rest Framework (DRF) and Flask REST APIs.

EXPERIENCE

Data Engineer/ Data Scientist   April 2022 - October 2023
Grazitti Interactive   *Quark City, Mohali, IN*

- Accomplished efficient data retrieval from a community platform by hitting 30 APIs daily using CRON Jobs. Employed Python and Pandas for data transformation and successfully stored the processed data in BigQuery. Utilized NLP techniques to analyze user comments and determine sentiment, enhancing insights for further decision-making.

- Successfully deployed fine-tuned LLM models (Llama2 and Mistral) using Hugging Face, Flask API, and Docker on AWS EC2. Achieved a remarkable 40% cost reduction and 30% latency reduction by employing quantized versions of the models.

- Advanced utilization of Hugging Face Transformers and open-source NLP/NLG modules to architect sophisticated language models and innovative applications.

- Leveraging PySpark, created data pipeline using AWS GLUE, processed semi-structured data from Marketo APIs, seamlessly integrating it into PostgreSQL database which reduced manual workload by 20% monthly.

- **Tech:** Python, Machine Learning, GCP, AWS GLUE, S3, EC2, Lambda, LLM, HuggingFace, NLP, Data Pipelines, PySpark, Generative AI

Machine Learning Engineer   January 2020 - March 2022
GenerativeAPIs   *Noida, IN*

- Designed and developed RESTful APIs using Django Rest Framework, seamlessly integrating and deploying deep learning endpoints to enable advanced computer vision tasks.

- Optimized database performance with Redis caching, reducing query response times by 40%, and implemented Celery for distributed processing, which increased in system scalability.

- Implemented Nginx as a reverse proxy and load balancer for DRF APIs, significantly boosting the platform's capacity to handle 30% more concurrent requests.

- Utilized Docker for streamlined application deployment and effectively managed AWS EC2 instances, achieving cost savings through efficient resource utilization.

- **Tech:** Python, Deep Learning, nginx, docker, S3, EC2, celery, redis, MySql, ORM, Selenium.

Machine Learning Engineer - Intern   July 2019 - December 2019
GenerativeAPIs   *Noida, IN*

- **Designed and developed Image Scrapper** Created a Flask API utilizing Selenium to scrape images from a designated website, subsequently storing them in S3. Managed the storage of all associated metadata in an SQLite database.

SKILLS

Languages and Libraries	Python, C++, SQL, Tensorflow, Pytorch, PySpark, DRF, Flask API
Technologies and Frameworks	Numpy, Pandas, OCR, NLP, Docker, Nginx, Celery, Redis, Kubernetes
Cloud and AI technologies	AWS ec2, GLUE ETL, s3, sagemaker, GCP, HuggingFace, LLM

PROJECTS

AMAZON FASHION RECOMMENDATION ENGINE:
– Created CNN-based model for image feature extraction and compared using Euclidean distance.
– Also used BOW and avg W2V-TFIDF on metadata of images.
– Built using Django, deployed on Heroku. (Github)

MAHINDRA FIRST CHOICE SERVICES:
– Identifying car ownership patterns and spending behavior across the country.
– Analyzed car make, model, service time, and type for location-specific insights.
– Employed ARIMA Time Series Analysis to predict Customer Lifetime Value and revenue forecast. (Github)

YOUTUBE BOOKMARK PLAYLIST:
– Using Django, I implemented login authentication for secure user access.
– Utilized PostgreSQL with the Django ORM for efficient database management and performed all CRUD operations seamlessly.
– Integrated AJAX for improved user experience and dynamic interactions. (Github)

SENTIMENT ANALYSIS OF AMAZON FINE FOOD REVIEWS:
– NLP-based text data preprocessing with word2vec conversion.
– Utilized supervised ML algorithms (KNN, Logistic Regression, SVM, Naive Bayes, Decision Tree, Random Forest, XgBoost) with AUC ROC metric hyperparameter tuning.
– Applied unsupervised techniques (SVD, PCA, T-SNE, K-means clustering) for data exploration. (Github)

EDUCATION

Master of Computer Applications, Jaipur,IN   July 2021 - July 2023
Manipal University, Jaipur

Full Stack Data Science Engineering, Mumbai,IN   July 2019 - Nov 2020
Grey Atom School of Data Science

Bachelor Of Computer Applications (Hons.), Jalandhar, IN   July 2016 - June 2019
Lovely Professional University

AWARDS AND EXTRA-CURRICULAR ACTIVITIES

- Achieved 1st place in the ChatGPT hackathon at Grazitti Interactive, surpassing 300+ participants, showcasing top-notch innovation and problem-solving abilities.

- Recognized as the "Best Employee" for outstanding contributions and dedication, setting a benchmark for excellence and commitment within the organization.

- Won first place in machine learning hackathon conducted by Grey Atom.   *Mumbai, Jan 2020*

- Successfully completed coursework on Natural Language Processing with deep learning certificate conducted by Grey Atom.   *Mumbai, December 2021*

- Actively writing blog posts about machine learning and deep learning SOTA techniques