

PRATYUSH SHUKLA

 [the-praxs](#) |  [hawk-praxs](#) |  [linkedin.com/in/praxs](#) |  pratyush.shukla@outlook.in |  +91-8853934890

EDUCATION

New York University | *M.S. Computer Engineering, GPA - 3.333*

May 2023

Lovely Professional University | *B.Tech (Hons.) Computer Science and Engineering, CGPA - 8.020*

July 2021

SKILLS

Languages: Python, C/C++17, SQL, NoSQL, MATLAB, \LaTeX

Tools: Git/GitHub, Docker, PostgreSQL, MongoDB, MySQL, GitHub Actions CI, Amazon Web Services (AWS), Excel, Word

Frameworks: Flask, PySpark, FastAPI, PyTorch, STL

Libraries: NumPy, SciPy, pandas, sklearn, Beautiful Soup, Camelot, XGBoost, CatBoost, Pydantic, SQLAlchemy, OpenCV, Boost

EXPERIENCE

OpenBB Finance, Remote | *Junior Software Engineer - Platform*

July 2023 – March 2024

- Designed and integrated data pipelines for four securities using FMP, Intrinio, and Polygon APIs, enhancing Terminal Pro's data availability and quality
- Maintained the data standardization from over five providers for four securities, ensuring high data quality and consistency
- Engineered and rolled out two automation scripts for local Platform package installations, optimizing developer efficiency and reducing setup time by a substantial 70%
- Addressed and resolved breaking changes in GitHub CI linters by updating 700+ files, eliminating major development bottlenecks
- Overhauled contributor guidelines and data pipeline development documentation, increasing accessibility and readability by 30%
- Developed and launched over ten extensions to access data from over five providers, expanding data coverage by 40% on the Platform
- Introduced a feature for adding custom deprecation messages in extensions, markedly improving user interface and experience
- Innovated an intermediate JSON structure for extension information, achieving a 100% sync in data accuracy and reducing manual processing by half
- Automated the generation of website documentation using the JSON structure, enhancing developer productivity by 80% and ensuring consistency across all platform documentation

Miko, Remote | *Data Engineering Intern*

May 2023 – August 2023

- Deployed an end-to-end data pipeline using the Mixpanel API, enabling real-time data storage in a PostgreSQL instance, streamlining data analysis processes
- Conducted thorough data validation and analysis to identify and rectify inconsistencies, ensuring high data integrity and authenticity
- Developed a proof of concept for a payment data pipeline utilizing the Chargebee API, facilitating efficient subscription data collection and analysis

Rosenblatt Securities, New York, NY | *Software Developer Intern*

March 2022 – April 2022

- Developed and optimized Python scripts for interfacing with DeFiLlama and CoinGecko APIs, boosting data extraction efficiency by 30% and enhancing analysis capabilities
- Implemented database optimizations that reduced replication lag by 80% and increased data synchronization speeds by 30%, improving data reliability four-fold
- Automated data updating and verification processes, ensuring high levels of data accuracy and integrity

Agile Social Learning 360, Remote | *Software Developer Intern*

January 2021 – June 2021

- Optimized AWS CloudFormation and VPC configurations to deploy user-specific virtual instances, reducing deployment times by 25% & cost per deployment by 20%, and improving cloud resource efficiency
- Enhanced security and connectivity within AWS environments by leveraging AWS VPC, Lambda, and Identity and Access Management, maintaining a zero-incident security record
- Managed a data warehouse integration, streamlining data loading from AWS S3 to AWS Redshift and boosting data retrieval speeds by 42%, enhancing data analysis capabilities and saving \$100 per month in data streaming costs

Lovely Professional University, India | *Research Assistant*

February 2019 – December 2020

- Extracted and transformed a comprehensive dataset from Zia's research paper into CSV format, facilitating efficient offline analysis and contributing to a 10% improvement in model predictions
- Implemented and tuned 15+ machine learning models using Python, sklearn, and XGBoost, benchmarking performance against the augmented dataset
- Advanced the application of optimization techniques in machine learning by porting and benchmarking the Satin Bowerbird Optimization algorithm from MATLAB to Python, leading to significant performance improvements

PROJECTS

DeFiLlama-Curl | *Python library to fetch DeFiLlama data via PycURL having over 13K+ downloads on PyPi and 600+ on Anaconda*

- Incorporated PycURL library to fetch data from DeFiLlama REST APIs improving overall performance by 80%
- Improved data retrieval speed via multithreading thus reducing fetch time by 50%

Adversarial Attack on Image Captioning Models | *Attack models to ascertain the security of image captioning models*

- Developed an image captioning model using an Encoder-Decoder-based Attention mechanism with Beam Search
- Implemented adversarial attack models using ResNet-50, ResNet-101, and ResNet-152 models as the encoder
- Trained the developed models on the MSCOCO 2014 and Flickr8k dataset
- Evaluated white-box attacks on every model, achieving a 69% success rate for the attack
- Evaluated black-box attacks on the CLIPCap captioning model reducing the BLEU-4 score from 32 to 25

Benchmarking Deep Learning Ecosystems on HPC | *Performance analysis of Flux and PyTorch framework on HPCs*

- Developed and trained the LeNet and ResNet-50 models on the MNIST and FashionMNIST datasets
- Benchmarked Total Training Accuracy against the batch size and the number of workers in both models
- Engineered 10 test suites for CUDA and non-CUDA HPCs to benchmark performance
- Concluded that the Flux framework offers faster performance but lacks against PyTorch on community support

Blockchain Protocol Analysis and Prediction | *Analyzing and predicting trends of popular blockchain protocols*

- Deployed a Python script to store data from DeFiLlama in a database for continuous data retrieval
- Performed EDA to observe market trends of top-performing protocols, traded volume, and the value of top chains
- Employed Prophet Library to predict the market trend for the next 3 months with 95% accuracy

Insurance Charges Prediction | *Foreseeing insurance charges based on analyzing given data*

- Performed EDA by analyzing factors affecting insurance charges using 3 metrics to obtain insights
- Performed Feature Engineering by clustering similar families, creating 2 features to improve accuracy by 15%
- Compared Linear, Polynomial, Ridge, Gradient Boost, XGBoost, and CatBoost regression models performance
- Tuned Gradient Boost hyperparameters among the compared models to obtain 85.44% max accuracy

Stock Price Movement | *Evaluate the impact of quarterly earnings report on stock price movement*

- Engineered a CLI application using STD framework to download, calculate, and plot the required data using GNUPlot
- Implemented liburl to fetch the adjusted close prices for selected Russell 3000 stocks and IWV via EODHD
- Analyzed the Russell 3000 stocks based on earnings and Zacks consensus estimate to sort into 3 groups
- Developed a Bootstrap class to calculate the AAR, AAR-STD, CAAR, and CAAR-STD for each group of stocks

Dog Breed Classifier | *Deep Learning model to classify a given face into top-n matching dog breeds*

- Augmented Stanford Dogs dataset from 120 breeds to 134 breeds, enhancing data size and data quality
- Developed a human face and a dog face detector using a ResNet-50 pre-trained model with 100% accuracy
- Trained a transfer learning-based classifier model to 87% accuracy, surpassing the ResNet-50+FT+DA model
- Deployed the trained classifier model using AWS SageMaker and AWS Lambda as APIs for easier end-user access

PUBLICATIONS

Attacking Compressed Vision Transformers

Taxonomy of Adaptive Neuro-Fuzzy Inference Systems in Modern Engineering Sciences, Hindawi, 2021

Nature-inspired optimization techniques, Nature-inspired Optimization Algorithms, De-Gruyter, 2021

A Novel Technique of Software Clone Detection based on Evaluation on Parameters, Journal of Emerging Technologies and Innovative Research, 2020

Biometric-based Individual Verification Frameworks, Journal of Emerging Technologies and Innovative Research, 2020

CERTIFICATIONS

Machine Learning Engineer Nanodegree, Udacity

Applied Data Science Module Unit 2 (Honors) - Machine Learning and Statistical Analysis

Applied Data Science Module Unit 1 (Honors)