Aritra Biswas

Data Scientist

linkedin.com/in/pandalearnstocode github.com/pandalearnstocode
+91.9051237918 pandalearnstocode@gmail.com
186, 18th B main, HAL 2nd Stage, Indiranagar, Bangalore 560038, Karnakata, India
https://pandalearnstocode.in/

Data scientist with an academic background in Statistics & experience in Python, R, C/C++, Julia to build & deploy large scale distributed data-driven solutions. Passionate about reproducible research & product development, focused on algorithms, computational statistics, profiling & optimization of code.

Competencies

Programming Languages	**Python** (Numpy, Numba, NumExpr, Scikit-learn, CVXPY, Scipy, SymPy, statsmodels, Pandas, Dask, Modin, Vaex, Snakeviz, Seaborn, Pydantic, Flask, FastAPI, Sphinx, Mypy, Pylint, Auto-PEP8, Hypothesis, Pytest, Locust, Papermill, Elliot, Structlog), **R** (Tidyverse, CVXR, fitdistrplus, Caret, Plumber, Shiny, Rcpp, ggplot2, plotly, HTML widgets), **C/C++, Julia**, LaTeX, Markdown
Visualization & Other tools	**Docker, Kubernetes, Git, MongoDB, Power BI**, Swagger UI, Jupyter, Rstudio, VSC
Cloud & Operating Systems	**Cent OS**, Ubuntu, **RHLE, Alpine, Windows Server**, Azure Apps, Functions, ACR, AKS
Technical skill	Generalised Linear Models, Monte Carlo Simulation, Inference, Mathematical Optimization

Education

M.Sc Statistics, University of Delhi	Analysis, Probability Theory, Linear Algebra, Stochastic Processes, Statistical Inference, Multivariate Analysis, Generalized Linear Models, Operational Research, Econometrics and Time Series Analysis, Statistical Quality Control and Reliability, Bayesian Inference, Advanced Statistical Computing and Data Mining.
B.Sc Statistics, Calcutta University	Probability Theory, Linear Algebra, Mathematical Methods, Sampling Distributions and Statistical Inference, Multivariate Analysis and Large Sample Theory, Time Series Analysis, Numerical methods, Population studies, Monte Carlo Simulation, Statistical Computing with C, Minitab, Advanced Excel.

Work Experience

Present – May, 2017 | **The Nielsen Company | Media | MMM Product Development, Global Markets, Bangalore, India**

- Delivered 8,000+ lines of production quality analytical code for three products.
- Developed JIT-compiled code for C level speed using N-D labeled arrays.
- Implemented structured logs in math code for better debugging & validation.
- Documentation of API, Python module & methodology using LaTeX, Sphinx & Swagger.
- Utilized the latest Python scientific computing stack for robust & optimized code.
- Implemented QP, Non-negative Matrix Factorization & Non-negative Least Square in Python.
- Written unit test, property base test and load test to ensure code quality.
- Developed automated test coverage report for the mathematical part of the application.
- Implemented Power BI Embedding in a custom application using Python backend in Azure.
- Received Simply Excellent award for contribution in the development of Rapid MPA.
- Developed & hosted containerized scalable REST API to serve model results using Python.
- Followed Agile methodology & git branching model for developmet and CI/CD pipeline.
- Developed multiple PoC applications for User Acceptance Testing (UAT) and demonstration.
- Conducted multiple in-person training sessions in Baroda, Bangalore & Warsaw CoE.
- Visited Poland, France for requirement gathering, training, & workshops.

Python | Numpy | Pandas | R | tidyverse | Docker | Flask | FastAPI | Azure

June, 2016 – August, 2016 | **Blue Copper, Business Analyst (Intern), Blue Copper Technologies, Calcutta, India**

- Developed interactive visualization dashboard using Shiny, HTML Widget & Rmarkdown.
- Document classification using NLP & developing domain-specific lexicon using web-scrapping.
- Parsing deeply nested JSON files & advanced ELT to converted nested data to flatten data.
- Computing MLE of right censored life distributions using profile likelihood estimation.

R | htmlwidget | Shiny | Regression | NLP | Web Scraping

Career interest

Computation Statistics :	R package development, Functional Programming , **Profiling and code optimization**
Product Development :	REST API, Parallel Computing, Distributed computing, Reliability engineering
Theoretical Statistics :	Linear Algebra, Reliability Theory, Statistical Learning, Mathematical Optimization