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, 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, Pandas-profiling, Voila, Panel, ipywidgets, Xarray, Dask, Modin, Vaex, Snakeviz, Seaborn, Pydantic, Flask, FastAPI, Sphinx, Mypy, Pylint, Flake8, AutoPEP8, Hypothesis, Coverage, Pytest, Locust, Papermill, Elliot, Structlog), **R** (Tidyverse, CVXR, fitdistrplus, Caret, Plumber, Shiny, Rcpp, renv, reticulate, ggplot2, plotly, HTML widgets, Rmarkdown), **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.
- Utilised the latest Python scientific computing stack for robust & optimised code.
- Implemented QP, Non-negative Matrix Factorisation & 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 Lift solution.
- Developed & hosted containerised scalable REST API to serve model results using Python.
- Followed Agile methodology & git branching model for development 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.
- 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`

Projets

Panda Leans to Code : Trying to understand the Universe better with Feynman Technique — June, 2019

I have just started writing this (blog). This blog is developed using blogdown package. Hosted in GitHub & deployed through an automated CI/CD pipeline of Netlify. The motivation behind this blog is borrowed from a mental model called The Feynman Technique, named after Richard Feynman. This project is still under development.

`Feynman Technique` `Blogging`

Langues

English ●●●●●
Hindi ●●●●●
Bengali ●●●●●

Interests

- Cooking
- Blogging
- Reading

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

Detailed Work Experience

December, 2019 – August, 2019

Dashboard solution : Unified visualisation platform for all Nielsen Global Media products

Senior Research Executive, Product development team, The Nielsen Company, Bangalore, India

It is an interactive visualisation platform for Nielsen's media analytics services. It uses Power BI Embedded inside a custom app. DRD and final deliverables gets published in a client specific dashboard in within the application.

- Interacting with Power BI API using Python to automate the process. Authenticating the Power BI application using Azure.
- Single Power BI Pro license to produce multiple dashboards. Defining minimal data sufficient schema.
- **Developed an advanced ETL pipeline and validation of the data to ensure seamless automated integration within the app.**
- **Automatically creating and replicating visualisations, reports, dashboards and data refresh using Power BI API.**
- **Approached response curve aggregation, seasonality and flighting pattern. Presently working on a PoC for RC aggregation.**

Technologies used : **Python**, **Power BI Embedded**, Pandas, **Azure**, Docker

July, 2019 – March, 2019

Lift solution : Market level lift solution for Nielsen Global Connect

Senior Research Executive, Product development team, The Nielsen Company, Bangalore, India

It uses one-to-many store matching for creating a synthetic control group and measures lift due to promotional activity.

- **Developed an alternative methodology for store matching other than canonical ANCOVA approach.**
- For store matching, implimented non-negative least square regression with bound constraints using convex optimization.
- **Implementation of Estimating Ad Effectiveness using Geo Experiments in a Time-Based Regression Framework**
- PoC was built using shiny (web), electron.js (desktop), REST API (plumber) and production code was developed in Python.

Technologies used : **Pandas**, Shiny, Plumber, **Numpy**, **Tidyverse**, Flask, **CVXPY, Scipy**, Docker

February, 2019 – September 2017

Marketing Planner : Portfolio level optimization with target spend & goal

Senior Research Executive, Product development team, The Nielsen Company, Bangalore, India

It is a client-facing application for portfolio level marketing budget allocation with response curves. It is capable of allocating marketing budget to media and promotional activities (VMEpu) for a given set bounding box constraints. Mostly used in the expanded vertical (non-FMCG) to allocate budget for market-level studies.

- **Developed Batch Based Monte Carlo Grid Search for constraint based budget allocation for multi-product scenario.**
- **Implementation of SLSQP for budget allocation across the marketing tactics for a single product scenario.**
- **Coded the end to end algorithm of the app in Python using Numpy, Numba, SymPy, Pandas, Scikit Learn and Scipy.**
- **Pure Numpy and Numba code for simulation using Intel Math Kernel (MKL) and Short Vector Math Library (SVML).**
- Sales, Spend optimisation within the same integrated platform with response curve at a tactic level.
- Implementation of advanced logging such as structured logs, logging through web socket and coloured streaming logs.
- Developed data driven testing for scientific computing. Code profiling and optimisation using snakeviz and other profilers.
- Deployed of the application as REST API in RHEL servers using Flask and Python.
- Worked closely with the UI team of development of custom charts and JSON schema.
- **Attained workshops with client and UX team for designing of meaningful custom charts for the application.**

Technologies used : **Pandas**, **Numpy**, **Numba**, **Sklearn**, **Scipy**, Flask, Docker

July, 2017 – May, 2017

MMM Modelling Platform : Implementation of Nielsen Marketing Mix methodology for market-level data

Research Executive, Product development team, The Nielsen Company, Bangalore, India

Initially this application was build to server the growth market but at present it is available as web application for all the associates. The same version of the application is available in Nielsen Software Download centre for client specific need.

- **Developed ETL part for the modelling process and automated data extraction for Marketing Planner.**
- **Developed unit test, property based testing modules and load testing. Generated automated code coverage reports.**
- Documentation of code and implementation of use friendly readable logging messages.
- **Developed a PoC for automatic the optimal retention rate selection and parameter estimation using grid search.**
- **As part of the above mentioned PoC deployed the solution in cloud with Kubernetes using Helm chart .**

Technologies used : Python, **Pandas**, Flask, **Numpy**, **Scipy**, **Scikit Learn**, **Statmodels**, **Docker** and **Kubernetes**

Reference

Nilotpal Banerjee
The Nielsen Company,
@ nilotpal.banerjee@nielsen.com

Kerstin Petersen
The Nielsen Company,
@ kerstin.petersen@nielsen.com

Ronak Trivedi
Johnson & Johnson,
@ rrtrivedi@jnj.com