Aritra **Biswas**

ML Engineer

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/

ML Engineer with academic background in Statistics. I have 4+ years of experience in product development domain. I am a Pythonista & R fanboy. I'm passionate about developing & improving algorithms, building & deploying large scale distributed data-driven solutions, research in ML domain, product development, cloud computing, code quality & travelling.

Competencies

Programming languages	**Python** (TensorFlow, Numpy, Numba, CuPy, Pandas, Scikit-learn, CVXPY, Scipy, Statsmodels, Matplotlib, FastAPI), **R** (Tidyverse, Tidymodels, Plumber, Shiny, GGPlot2, Plotly), **C, C++**
Other tools	**Docker, Git**, Visual Studio Code, Jupyter Ecosystem, Rstudio, LaTeX, Markdown
Cloud technologies	Azure Machine Learning, Data Bricks, Azure Container Instances, Azure Container Registry, Azure DevOps, Azure Data Factory, Azure Apps, Azure Functions, Azure Kubernetes Service
Domain knowledge	Marketing anlytics (MROI, Marketing Mix Models, Media budget allocation, Lift solutions), analytical product development, production quality ML code development & maintenance.

Education

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

Work Experience

Present – April, 2020 | **Machine Learning Engineer, AB InBev | MarTech Analytics, GAC, Bangalore, India**

- Developing methodology & Python SDK using cloud technologies to run automated MMM at scale.
- Developing multi-objective marketing budget allocator to find a optimal media spending strategy.
- Enhancement, development, cloud migration & maintenance of existing LMM based MMM tool.
- Collaboration & research of with MIT BudLabs to explore & enhance existing MMM methodology.
- Development of a cloud based Lift solution to measure in-store impact due to digital campaigns.
- Developing LLVM compiled DTW distance with Sakoe-Chiba band for Market Matching using Numba.
- Streamlining & automation of data pipeline to integrate with the existing products using Azure.

Python | TensorFlow 1.x | Numpy | Pandas | Numba | CVXPY | Azure ML | Data Bricks | R

March, 2020 – May, 2017 | **Data Scientist, The Nielsen Company | Media, Global Markets, Bangalore, India**

- Contributed & maintained 8,000+ lines of production quality code for three analytical products.
- Designed & developed JIT-compiled code optimized Numba to achieve C level speed in production.
- Designed custom algorithms to solve goal seek optimization, synthetic control group generation.
- Ensuring QA, QC, documentation, logging, validation, testing & monitoring for ML SDK & REST APIs.
- Optimizing run time for Non-negative Matrix Factorization & NNLS in case of large sparse matrices.
- Exploration & evaluation of appropriate components from Azure cloud services to serve APIs at scale.
- Implementation of Power BI Embedding in a custom application using Python SDK in Azure Functions.
- Development & deployment of containerized scalable REST APIs to serve ML models using FastAPI.
- Integration of Azure CI/CD pipeline for automatic deployment of ML APIs following git branching model.
- Conducted multiple in-person training sessions and workshops in Baroda, Bangalore, Warsaw & Paris.
- Organized multiple UI/UX, requirement gathering workshops/boot-camp with different stakeholders.

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

Projets

Panda Leans to Code : Trying to understand everything better with Feynman Technique — Since June, 2019

I have just started writing this (blog). This blog is developed using blogdown package & R. Hosted in GitHub & deployed using 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 is pet project & still under development.

Feynman Technique | Blogging

Research interest

Computational statistics :	Simulation, Cloud computing, Parallel computing, Distributed computing
Product development :	R & Python SDK development, Scalable ML API development Reliability engineering
Statistical machine learning :	Linear Algebra, Reliability Theory, Statistical Learning, Mathematical Optimization

Detailed Work Experience

Present
May, 2020

MROI Products : Media Mix Model & Marketing Budget Allocator

Machine Learning Engineer, MarTech Analytics, GAC, Bangalore, India

MMM & MBA are used to measure media performance & allocate budget across different media vehicles. These are cloud native turn key ML products, connected with multiple data sources to execute automated MLOps pipeline with minimal human intervention.

- Leading a team to develop, maintain the core machine learning algorithm & associated cloud infrastructure for MMM & MBA Tool.
- Building & maintaining PyMMM (Python SDK) using Azure CI/CD pipeline to package, distribute & execute ML workflow in Azure.
- Training teams & vendor partners about the MMM methodology & MMM product usage. Creating user guide with other members.
- Automating hyper-parameter (ad-stock, saturation & effective frequency) tuning using hierarchically regularized regression.
- Addressing long term impact of media advertising on sales. Cannibalization between own brands due to advertising.
- Detection of impacts due to sponsorship event, holidays, trend, seasonality, DoW & macro-economic factors in the MMM model.
- Developing multi-objective constrained optimization between short term & long term sales, power, equity, goal-seek.
- Automated model refresh, validation, data drift detection using cloud native technology. Scalable across different regions.
- Migration of VM based tool to cloud based, scalable architecture using AKS & other Azure cloud components.
- Collaboration with BudLabs, MIT to improve MMM methodology though an research initiative between AB InBev & MIT.

Tech stack : Python, TensorFlow, Numpy, Pandas, Scipy, CVXPY, Azure Machine Learning Studio, Azure Kubernetes Service

October, 2020
April, 2020

Project Multipler : Lift solution to analyse in-store impact digital media campaigns

Data Scientist, MarTech Analytics, GAC, Bangalore, India

Multiplier measures performance of digital marketing campings using in-store sales lift as metric. Multiplier is deployed in Azure to deals with high volume of daily store level data. This solution has been rolled out for two major markets in 2020.

- Designing & building end to end methodology & deployment in Azure Data Bricks for scalability.
- Monitoring, leading & managing vendors to analyse 100+ digital campaigns in a quarter for a major market.
- Developed LLVM compiled DTW distance with Sakoe-Chiba band for store matching between test & control group.
- Automated data validation pipeline, ETL & modeling in Azure Data Bricks using multiple nodes GPU cluster.
- Initial development of simulation and optimization module to come up with a optimal digital campaign strategy.
- Using R Azure Batch SDK for parallel computation of Linear Mixed Effect model leveraging Azure Low Priority VMs.

Tech stack : Python, Numpy, Numba, Pandas, R, Tidyverse, SparkR, LME4, Azure Data Bricks, Azure Batch

March, April
August, 2019

Nielsen StoryBoarder : BI solution leveraging Power BI Embedded & Azure

Data scientist, Product development team, Nielsen, Bangalore, India

Nielsen StoryBoarder is one of the largest shared service in purview of the marketing effectiveness organization serving $95MM worth of business across Lift & MMM.

- Interacting with Power BI API using Python to automate the process. Authenticating the Power BI application using Azure.
- Developed an advanced ETL pipeline & validation of the data to ensure seamless automated integration within the app.
- Automatically creating & replicating visualizations, reports, dashboards & data refresh using Power BI API.
- Developed a response curve visualization & aggregation framework to understand confounding effect of media execution.

Tech stack : Python, Power BI Embedded, Pandas, Azure, Docker

July, 2019
March, 2019

Rapid MPA : Aggregated market level lift solution leveraging granular store level data

Data Scientist, Product development team, Nielsen, Bangalore, India

Rapid MPA is used for one-to-many store matching, synthetic control group generation & measuring in-store lift due to promotional activities. Scaling DTW for daily store level data was a challenge. Solved this problem during project multiplier.

- Synthetic controlled matching for performing store match using non-negative least square regression with bound constraints.
- Implementation of estimating ad-effectiveness using geo-experiments in a time-based regression framework.
- Minimize runtime of SCM to less than a minute by designing a optimized LLVM compiled code using Numba & Intel MKL.

Methodology : Campaign Lift, Match Panel Analysis, Synthetic Controlled Matching, Time & Geography based regression.

February, 2019
September 2017

Rapid MMM Suite : An integrated market level end to end MROI solution

Data Scientist, Product development team, Nielsen, Bangalore, India

Served business worth of $25MM. Rapid modeler used hierarchical additive linear model methodology. Rapid Simulator used customized monte carlo grid search for portfolio level marketing budget allocation with vehicle level response curves. Both of tools deals with media vehicles, promotional activities, short term and long term sales impact.

- Rapid MMM Suite consists of MMM Modeler & Simulator with plans of extending the suite to integrate with other Nielsen products.
- Rapid Simulator enables sales & spend optimization within the same integrated platform with response curves at a tactic level.
- Developed batch based monte carlo grid search for constraint based budget allocation for multi-product scenario.
- Implementation & enhancement of SLSQP for budget allocation across the marketing tactics for a single product scenario.
- Developed Numpy & Numba code for simulation using Intel Math Kernel (MKL) & Short Vector Math Library (SVML) for production.
- Developed ETL part for the modelling process & automated data extraction for marketing planner.
- Developed data driven testing for scientific computing. Runtime, memory profiling & optimization of ML code.
- Deployed of the ML Code as REST API endpoint in RHEL servers using Flask & Python. Stress testing using Locust.
- Collaboration with UI/UX team, vendor partners, clients, users & other stakeholders for requirements capturing.

Tech stack : Python, Pandas, Numpy, Numba, Scikit-Learn, Scipy, Flask, Docker