

Aritra BISWAS

ML Engineering Lead

[in linkedin.com/in/pandalearnstocode](https://www.linkedin.com/in/pandalearnstocode) github.com/pandalearnstocode

+91-9051237918 @ pandalearnstocode@gmail.com

940, 3rd Main Rd, AECS Layout - D Block, AECS Layout, Marathahalli, Bengaluru, Karnataka 560037

<https://www.aritro.in/>



I am a seasoned Machine Learning Engineer with 6+ years of experience, specializing in Marketing Analytics, B2B RecSys, LLMs, ML infrastructure & MLOps. My expertise includes distributed computing, distributed model training and ML product development. I am skilled in auto-scaling, infra selection, optimization & designing large-scale ML systems. I focus on building efficient architectures for ML workflows & life cycle, optimizing system performance & maximizing resource utilization. I am proficient in programming languages such as Python, R & C, with hands-on experience in Azure. Currently, I lead a team of ML Engineers at AB InBev.

COMPETENCIES

Programming languages	Python (TensorFlow, Numpy, Numba, CuPy, Ray, Dask, Modin, Pyspark, Pandas, Scikit-learn, CVXPY, Scipy, Statsmodels, FastAPI, LangChain), R (Tidyverse, Tidymodels, Plumber, Shiny), C
Other tools	Docker, Git, VS Code, Jupyter Ecosystem, RStudio, Minikube, Docker Compose, Markdown
Cloud technologies	Azure Machine Learning, Azure Data Bricks, Azure Kubernetes Service, GitHub actions
Domain knowledge	RecSys, Market Mix Modelling, Media Budget allocation, Promo Optimization, Lift solutions

EDUCATION

M.Sc Statistics, Delhi University	Probability Theory, Linear Algebra, Stochastic Processes, Statistical Inference, Multivariate Analysis, Generalized Linear Models, Econometrics & Time Series Analysis, Reliability Theory, Bayesian Inference.
B.Sc Statistics, Calcutta University	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

ML Engineering Lead, AB InBev, GAC, Bangalore, India

- **ML Platform:** Successfully optimized architecture & infrastructure selection, leading to efficient, scalable & reliable machine learning workflows, enhancing lifecycle management & overall performance.
- **Recommendation Engine:** Developed a versatile python ML library for custom-curated algorithms, catering to B2B use cases & supporting parallel processing backends. This library has successfully scaled cross-sell, recent purchase and forgotten item use cases in a production serving 13+ countries. Define KPIs for data drift, quality check and lineage tracking, model drift for the end-to-end solution.
- **Generative AI - Social Listening:** Leveraged Brand Watch API services to extract social media data, enabling prompt development, text summarization, sentiment classification & identification of inflection points in user sentiment. Established automation using ADB clusters for streamlined operations.
- **Marketing Mix Modelling:** Enhanced linear mixed-effect marketing models, transitioning from notebook-driven development to analytical products, accelerating time-to-market and reducing execution time in 20+ markets. Employed a comprehensive approach in architecture, API design, library creation, product development, and deployment, fostering data-driven marketing strategies.
- **Marketing Budget Allocator:** Developed a multi-objective constrained optimization framework for media investment allocation at brand, vehicle & sub-vehicle levels. Utilized visual bounds, response curves & media investment simulation tools for optimization. Designed architecture diagrams for web-based applications, ML workloads & RESTful endpoints to deliver optimization results at scale.
- **Life Solution:** Executed causal inference-based lift solution for 100+ campaigns across 3+ markets, automating end-to-end processes. Optimized DTW calculations for country-level store matching.
- **Research and Collaboration:** Collaborated with MIT BudLabs research group to enhance existing MMM methodology. Conducted experimenting & benchmarking with existing open-source alternatives like Facebook Robyn & Google's Lightweight MMM to improve and upgrade existing features.

Python TensorFlow Numpy Ray Numba Kubernetes Fast API Azure GitHub actions

March, 2020
May, 2017

Data Scientist, THE NIELSEN COMPANY, Bangalore, India

- **Rapid Suite of Products:** Developed, optimized & maintained production code, algorithms (goal seek optimization, quick optimize using Monte Carlo simulation, Non-negative Matrix Factorization & NNLS) & ML SDKs, ensuring quality, efficiency, & enhanced performance (JIT-compiled Numba code).
- **Nielsen StoryBoarder:** Utilized Azure cloud services, Power BI embedding & integrated CI/CD for scalable API delivery, streamlined deployment, & improved data visualization in custom applications.
- **Collaboration & Inner Sourcing:** Conducted multiple training sessions, training workshops, product roadshows. Collaborated with stakeholders, external partners to foster knowledge sharing, collaboration, & alignment in UI/UX, design decision, architecture & requirement gathering processes.

Python Numpy Numba Pandas R tidyverse Docker Flask FastAPI Azure

PRESENT FEB, 2023	<p>AlgoValue Python SDK, ML platform, GenAI Products ML engineering lead, GAC, Bangalore, India</p>
	<p>Currently leading development of AlgoVault, a library of recommendation algorithms for B2B use cases. Working with ML platforms to select, optimize and implement infrastructure. Additionally, I oversee architecture, infrastructure, and automation for GenAI products.</p> <ul style="list-style-type: none"> > Leading development team. Design overall architecture for AlgoVault. Continuously enhance & expand the algorithm library. > Enhance algorithm performance & scalability. Integration of AlgoVault with ML platforms. Infra selection for deployment. > Backend development in for social media monitoring tool. Insight mining using OpenAI models. ROI & scalability analysis. > Designed the architecture of an enterprise-grade employee assistant to mine structured and unstructured data using RAG.
AUG, 2022 NOVEMBER, 2020	<p>Tech stack: Python, Ray, LangChain, OpenAI, Scikit Learn, Huggingface Transformers, AML, Azure Cognitive Services MROI Products: Media Mix Model, Media & Marketing Budget Allocator ML engineering lead, MARTECH ANALYTICS, GAC, Bangalore, India</p>
	<p>MMM & MBA are used to measure media/marketing performance & allocate budget across different vehicles. These are cloud native turn key ML products, connected with multiple data sources to execute automated ML pipeline with minimal human intervention.</p> <ul style="list-style-type: none"> > Leading a team to develop, maintain the core machine learning algorithm & associated cloud infrastructure for MMM & MBA. > Building & maintaining pymmm (core python library) using CI/CD pipeline to package, distribute & execute ML workflow in Azure. > Automating hyper-parameter (ad-stock, saturation & effective frequency) tuning using hierarchically regularized regression. > Addressing long term impact of media advertising on sales/profit/revenue. Cannibalization between due to advertising. > Architected a micro-front-end & micro-service based template to re-use front-end & back-end services across multiple products. > Multi-objective constrained optimization between short & long term sales, profile, revenue, brand 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 through an research initiative between AB InBev & MIT.
OCTOBER, 2020 APRIL, 2020	<p>Tech stack: Python, TensorFlow, Numpy, Pandas, Scipy, CVXPY, Azure Machine Learning, Azure Kubernetes Service Project Multiplier: Lift solution to analyse in-store impact digital media campaigns Data Scientist, MARTECH ANALYTICS, GAC, Bangalore, India</p>
	<p>Multiplier measures performance of digital marketing campaigns using in-store sales lift as metric. The app is deployed in azure to handle high volume of daily store level data. This solution has been released and used by two major markets in 2020.</p> <ul style="list-style-type: none"> > 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. > Development of 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 an optimal digital campaign strategy. > Using R Azure Batch SDK for parallel computation of Linear Mixed Effect model leveraging Azure Low Priority VMs.
MARCH, APRIL AUGUST, 2019	<p>Tech stack: Python, Numpy, Numba, Pandas, R, Tidyverse, SparkR, LME4, Azure Data Bricks, Azure Batch Nielsen StoryBoarder: BI solution leveraging Power BI Embedded & Azure Data scientist, PRODUCT DEVELOPMENT TEAM, NIELSEN, Bangalore, India</p>
	<p>Nielsen StoryBoarder is one of the largest shared service in purview of the marketing effectiveness organization serving \$95MM worth of business across Lift & MMM.</p> <ul style="list-style-type: none"> > 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.
JULY, 2019 MARCH, 2019	<p>Tech stack: Python, Power BI Embedded, Pandas, Azure, Docker Rapid MPA: Aggregated market level lift solution leveraging granular store level data Data Scientist, PRODUCT DEVELOPMENT TEAM, NIELSEN, Bangalore, India</p>
	<p>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.</p> <ul style="list-style-type: none"> > 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.
FEBRUARY, 2019 SEPTEMBER 2017	<p>Methodology: Campaign Lift, Match Panel Analysis, Synthetic Controlled Matching, Time & Geography based regression. Rapid MMM Suite: An integrated market level end to end MROI solution Associate Data Scientist, PRODUCT DEVELOPMENT TEAM, NIELSEN, Bangalore, India</p>
	<p>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.</p> <ul style="list-style-type: none"> > 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 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. <p>Tech stack: Python, Pandas, Numpy, Numba, Scikit-Learn, Scipy, Flask, Docker</p>