

Sayan Mukhopadhyay

Kolkata, India | +91 801 723 2371 | sayan_jdp@yahoo.co.in

[LinkedIn](#) | [Facebook](#) | [Youtube](#) | [Github](#)

Algorithm Design | Product Development | Solution Architecture

- Strategic professional with over **15 years** of experience in **Data Science and Machine Learning** techniques in **Big Data and Cloud** environment; expertise in stimulating business process steps, analyzing current state business processes to identify opportunities for value added change and to improve future state business processes
- Highly proficient in working on Machine Learning algorithms applied in **Algorithmic Trading, Natural Language Processing, and Computer Vision**, proficient in advanced analytics techniques like Probabilistic Graphical Model, Goal oriented dialog system, recommendation in scalable and fault tolerant environment, Information retrieval, knowledge graph, automate call center operation using **LLM** and **RAG**
- Rich exposure in **FinTech Industry**; proficient in developing several types of **financial models** to predict risk in investment portfolio in share market, generate credit score from bank transaction or social media data; performing due diligence, research, analysis, and documentation of live transactions
- A seasoned **Solution Architect**, can decode business requirements to architectural concern from **Networking Protocols** to **Design Patterns**; proficient in tracking & gathering industry information, interacting with promoters, guiding on pitch preparation, and providing constructive feedback

Education

MTech (Research) in **Computational & Data Science**, **Indian Institute of Science**, India, 2014 ([Video](#))

BEng. in Instrumentation & Electronics, **Jadavpur University**, **India**, CSIR Fellow in Energy Research, 2005

Publications: *A parallel algorithm for molecular dynamics simulation*, would be a PhD thesis

Variance of difference as a distance like measure in synchronous time series microarray data clustering ([link](#))

Advance Data Analytics using Python, Apress, Sayan Mukhopadhyay (Book) [1st Ed](#) [2nd Ed](#)

Certification: SAS Certified Base Programmer, Cryptography from Coventry University, NCC B Certificate

Technical Skill

- **Languages:** Python, Java, C++, Shell, C# .net, SQL, Node-Angular-React, Scala, Elixir, Go, Rust, Unity, UML, VHDL
- **Tools:** Socket, IPC, RPC, ACE, Poco, HTTP, REST, SOAP, Play, Spring, Flask, Falcon, Django, WCF Unix (GDB, Gprof, Makefile), HPC (MPI, Open MP, POSIX, CUDA, FPGA) Selenium, Active- MQ, Tibco-EMS, Zero-MQ, git, svn, Open-bravo
- **Analytics:** MATLAB, R, Octave, Oracle, Sybase, MySQL, SQLite, Elastic Search, MongoDB, neo4j Hadoop, MapReduce, Hive, Mahoot, Storm, Spark, MS BI, Pentaho, D3, C3 Google Visualization API, Map API, Kibana, Tableau, Tensor-flow, Keras, Pytorch, Open-CV, Textblob, NLTK, pandas. Snowflake, Airflow, dbt, sk-learn, XGboost, Lightfm, dialo gpt, chat gpt, bert
- **System Administration:** Linux, Windows, Solaris, Virtual Box, Docker, Kubernetes, Router, Switch, SevOne, Nikson, Corvil, Ganglia, Spectrum, E-Health, Nagios, Cactil, Lucierna, Asterisknow, Socat, NMAP, SQL map, Kali Linux,,Suricata AWS, GCP, Azure; terraform, combodo-itop, vivado(fpga)

Work History

REMOTE CONSULTANT – Machine Learning, Big Data, Cloud, LLM (Nov, 2016 – Present)

- Spearheading the creation of new data sciences capabilities for the business by envisioning and executing strategies that will influence improvement of the business's performance by enabling informed decision making.
- Contributing to designs and launches innovative and complex analytic models, utilizing blend of contemporary & traditional data mining techniques, which are applied to both structured & unstructured datasets.
- Establishing scalable, efficient, automated processes for large scale data analyses, model development, model validation and model deployment.

Key Projects:

Client: Turing Working for RLHF of chatGPT for client Microsoft.

Client: McDonald as Data Science Engineer

Synopsis: Pricing Theory, Time Series Analysis, Validation of a recommendation system

Tool: Python, Pydash, Nixtla Framework, Nhit, Nbeat, , XG boost, Lightfm, Angular Js, influx db

Client: Ulventech as ML Engineer

Synopsis: chatbot, recommendation system, algorithmic trading, generate trade signal from market data

Tool: Python, AWS, Market Darta API, Dialo GPT,Bert, Chromadb,langchain, Tf-idf Vectorization, Text Clustering

Sayan Mukhopadhyay

Page 2

Client: Gotoko as Lead Data Scientist

Synopsis: Collaborative filter-based clustering, deep learning-based classifier, linear programming

Tools: python, GCP, Big Query, K-mean clustering, Tensorflow, Cloud Function, TPU, Elastic Search, dbt, sk-learn,

Client: SymphonyAI

Synopsis: Customer Churn Prediction

Tools: Snowflake, Sql Server, Azure Databricks, pyspark, tensorflow, keras

Client: Pubmatic as Lead Algorithm Analyst

Synopsis: Applying six sigma principles on ML model. Design experiment with trading algorithm

Tools: Go, Scala, Python, R, Spark, docker, Airflow, scikit-learn

Client: Future today

Synopsis: Video Recommendations System for smart TV channel

Tools: Collaborative Filtering, Naive Bayes Classifier, Page Rank Algorithm, pySpark, Mysql, Mongo, Roku Device

Client: Crossover: Attended a paid boot camp as Chief Software Architect

Project: ReverseAd Keyword Algorithm System – ReverseAd

Technology: Python, AWS, Redshift, Name Entity Recognition, Stanford NLP, NLTK, Mongo, EC-2, Flask

Project: Whizhack Anomaly Detection System – Whizhack

Technology: Python, AWS, Azure, Suricata, Tensorflow, Keras, Synapsys DB, S3, Kali-Linux

Responsibilities: built the product from scratch from REST API to ML model

Project: Sulvo Ad Price Predictor System – Sulvo

Technology: Python, falcon, GCP, Node js, Tensor-flow, Redis, Neural Networks, Big Query, App Engie, Cloud store

Abzooba(UST), Kolkata, India

Manager – Data Science, 6/2015 – 11/2016

Tools: Java, Play, Mongo, Python, Neo4j, Angularjs, Stanford NLP, Beautiful Soap, tensorflow, mem2n, AWS

Promoted to the role of Data Science Manager and Project Lead for key client accounts. Lead a team of seven staff members to deliver sophisticated projects within budget and ahead of schedule. Direct the entire project lifecycle from the initial design and planning phases, through to progress / risk monitoring and implementation.

Key Projects:

- Successfully designed analytic solutions (*GUI to ML model*) within a specified client framework in addition to establishing a credit score model using social media data.
- Designed and implemented optimal systems, such as dialog, security, auto-complete and feedback modules for retail e-commerce operations and a Topical Crawler for web harvesting.
- Developed and deployed a Health Insurance Claim Status Prediction System via HIPAA 835 & 837 data as well as vendor selection and price optimisation systems utilising transactional data.

TCG Digital, Kolkata, India

Data Scientist / Sr. Consultant - Analytics & Big data, 7/2014 – 6/2015

Tools: Python, Java, Mahout, Hadoop, Hive, Elastic Search, Kibana, Sentiment Analysis, Neural Network, Asterisk

Critical role, applying data science methods, processes and systems to troubleshoot analytics and complex bigdata issues for diverse clients. Built positive client relationships, with direct involvement in pre-sales activities.

Key Projects:

- Played a lead role in a sentiment analysis project focused on brand development as well as a Neural Network based Predictor for passenger loads for two clients in the aviation sector.
- Delivered a solution to address data errors for a leading electronics manufacturer / distributor, via the parallel implementation of KMean clustering using Needleman Wouch sequence matching score.
- Implement a voice call mechanism in a Java based ticketing system using asterisk server

Mphasis (HP Group Company), Pune, India

Technical Architect – Security Tool Group, 10/2013 – 7/2014

Tools: Nagios, Cacti, Ninja, Merlin, Lucierna, Combodo-itop, Jasper-Report, Sumologic, Event correlation, Topology Auto-discovery, SNMP-walk, netflow

Sayan Mukhopadhyay

Page 3

Fully accountable for the build, development and deployment of a complex transaction monitoring system based on available open-source stack. Provided optimal service to nine key clients, including Coca Cola, Verion and RPG. Performed requirements gathering and feasibility analysis, then developed robust project plans and roadmaps. Managed, motivated and coached the technical team, with 12 members, to achieve all objectives.

Key Projects:

- Designed and implemented enhanced system architecture, including a framework for Nagios Active Checks, recursive approval method for change management in ITSM
- Integration of **netflow** and Nagios core via netflow collector and Nagios check plug-in. Swiftly identified / fixed bugs, such as a data miss issue of cacti with '*polar table in not empty*' exception.

Wedoria (ABP Group Company), Kolkata, India

Lead Architect - Data Mining, 5/2013 – 8/2013

Tools: Bass Model, Collaborative filtering, Mechanism design, MySql, PHP, Java, Python, Storm

Tasked with creating a revenue model using a stochastic Bass model of user growth / influence. Formulated architecture to support a recommender system and design mechanism incorporating collaborative filtering and linear programming

Pubmatic, Pune, India

Senior Technical Analyst - Machine Learning, 5/2012 – 4/2013

Tools: Java, Perl, Shell, Python, Hadoop, MapReduce, Hive, Storm, MySql

Applied predictive approximation techniques for big data queries to enhance the Hadoop platform in addition to designing and deploying a new estimation and prediction framework for query handling. Conducted R&D, established a unit testing framework and applied Agile methodologies to ensure seamless project execution.

CA Technology, Hyderabad, India

Senior Software Engineer - Network Management System / Data Centre Operations, 9/2011 – 4/2012

Tools: C++, Java, Perl, SNMP, Barclay DB, Gmake, Tomcat, Rest-Full Web Service, RTC, SNMP, ICMP

Vital role as senior member of the Spectrum Backlog Reduction team, tasked with the development, maintenance and support of a large-scale, complex data centre operation domain. Performed bug fixing, code reviews and testing, then designed and deployed features to enhance performance, quality and security standards. Delivered 2nd Level support to clients and mentored junior members of the team.

Key Projects:

- Successfully implemented numerous product features, such as an ignore character case in seek command enabling users to find a model handler with given condition. Swiftly resolved complex bugs, including a GUI fault due to race condition between front end thread for servlet request and backend thread of server response.

FairFest Media Limited, Kolkata, India

Technical Manager - Patent Analysis / Loyalty Reward Program, 11/2010 – 9/2011

Tools: ASP/C# .net, Nhibernate, WCF, C++, Python, Poco, NVP/SOAP/Rest, AJAX

Conducted patent analysis, marketing and development activities within the poverty, environment and security areas. Acted as a focal point of contact on technical issues

PayPal (through CSC), Chennai, India

Senior Consultant - Application Development, 3/2010 – 11/2010

Tools: Apache, Red-hat Linux, C++, Swig, Python, Java, ClearCase, Active MQ

Strategically managed and developed core payment technology, including the optimisation of system architecture, framework and functionality. Oversaw code review, debugging, profiling and unit testing activities, with four direct reports. Promoted an environment of continual improvement and best practice.

Key Projects:

Established an AMQ framework that recorded corresponding error numbers in the Common Application Logger based on errors related to SQL. Developed a remove range method in Vector class for the PayPal developer community and applied swig generate python / java interface of admin-auth module to cache authentication data in a local server.

Credit-Suisse (through FCS), Pune, India

Technical Consultant – Investment Banking, Risk Analysis, Data Centre Operations, 6/2008 – 2/2010

Sayan Mukhopadhyay

Page 4

Tools: C++, C# .net, Perl, Shell, Python, Oracle, Sybase, Linux, Solaris, Tibco-EMS

Multifaceted role, providing technical consultancy across risk analysis, algorithmic trading and infrastructure development as well as acting as Rapid Application Developer for high-frequency trading.

Key Projects:

- Delivered an application support project involving the modification and enhancement of a trading time risk analysis tool which greatly improved risk management and control. Successfully designed and implemented a distributed, divide and conquer Real-time Log-file Parsing Framework for use in high-frequency trading activities.

2000-2003: IIT Math, Physics Faculty in various organizations

Details of Analytics Projects

Project:	Future Today Content Recommender System – Future Today (02/2018 – Till Date)
Technology:	Python, Spark, Falcon, Mysql, Mongo, Naive Bayes Classifier, Page Rank, Roku device
Details:	Future Today Recommender is updatable, i.e., in each training iteration, it will update the stored model not replace it. It is possible because model is a function of the count. Model is exposed as a Rest API to integrate with other systems. To make the REST API response fast, we stored all information required for the recommendation in shared data structured in RAM. To respond the request API does not need to fire any query in other than RAM data structure. We also implemented parallel execution of each query in one response. The training process is scalable with the size of input raw log data as it is implemented in Spark. Training system can also handle input data size more than it's RAM size. It implemented for the system which has 40 million user views per month and it response below 150ms for prediction request.
Responsibilities:	Building the product from scratch from REST API to ML model
Project:	Sulvo Ad Price Predictor System – Sulvo (11/2016 – 11/2017)
Technology:	Python, falcon, Google cloud platform, Node js, Tensor-flow, AWS, Redis, Neural Networks
Details:	The system has three major components – collector, real-time trainer, and predictor. It listens to the programmatic ad selling information for each impression through a collector and sends the predicted floor price for each impression to ad server through predictor. Real-time trainer builds the model on recently collected data and saves it, and predictor uses that model for prediction. Sulvo ad server is hosted in AWS cloud, and real-time trainer and predictor are hosted in Goggle cloud platform. The collector which is hosted in AWS and exposes a REST API to collect data receives information from the ad server and push it to a Redis message queue. On de-queue message from Redis server; the message is pushed to a Google Big Query instance. Real-time trainer which is hosted in a Google Compute Engine fetches the latest information from Big Query and builds a multi-layer CNN model using tensor-flow and save the model in Google Data Store. Predictor server which is a Falcon-based REST API is hosted on Google App Engine, receives prediction requests from Sulvo ad server and response back floor value calculated on basis stored data in the Data Store. The system is deep learning based, able to train the model with real-time data and also high performance, assured to send a response to ad server within 300ms and auto-scaled to serve 22 million prediction daily.
Responsibilities:	Building the product from scratch from data collector to ML model
Project:	Topical Crawler for Web Harvesting - Abzooba (9/2015-12/2015)
Technology:	Python, Beautiful Soup, Graph traversal Algorithms
Details:	Project for a real estate client to automate manual reporting processes detailing data for High-Net-Worth clients. Developed a two step approach; gathered the URL with relevant information via a Tropical Crawling approach as well as extracting name, organisation and location from output URL via a name entity recognition algorithm. In terms of Topical Crawling, two sets of URL's were used for input, the first being seed URL's where crawling will begin and the second target URL's as reference URL's and a known subset of relevant URL's. Utilised Jaccard Distance as a similarity measure to compare content of new URL's with targets and select the most similar. This step was repeated to make the selected URL as the seed URL, which also considered all output URL's in previous steps and excluded those already selected. Reductions in search space were achieved by establishing a threshold in the similarity measure URL, where an average similarity is among the target URL's. A set

Sayan Mukhopadhyay

Page 5

of keywords was created, i.e. 'profile', 'biography', for the relevant URL string. Using a Name Entity Recognition algorithm enabled the extraction of name, organisation and location in the selected output URL contents. BeautifulSoup library was used to parse the html pages and Stanford NER NLTK library for Name Entity recognition.

- Responsibilities:** Problem Formulation, Design Solution Architecture, Mentor Implementation, Client Liaison
- Project:** Vendor Recommendation using Cognitive Scale framework – Abzooba (2/2016-4/2016)
- Technology:** **Python, neo4j, angular js, Logistic Regression, Naive Bayes, Random Forest,sk-learn**
- Details:** Project to recommend the best vendors for a particular product ID using purchase order and invoice information. To address the problem this was divided into two sub-areas. The first area involved the prediction of probability of successful purchases. Three approaches were used in this calculation; a regression-based approach using Logistic Regression, a rule-based approach using Random Forest Algorithm and a Bayes-theory approach by Naive based classifier. The second phase involved grouping of data by VEN_ID and calculation of average probability of a successful purchase for each vendor. Vendors were then sorted according to this probability. Rule-based engines were then applied, such as vendor with minimum delivery time or vendor with maximum number of deliveries, and the most suitable vendor was recommended. Rules were prioritised by predicting aspects of user query using classification.
- Responsibilities:** Problem Formulation, Design Solution Architecture, Mentor Implementation, Client Liaison
- Project:** MCUBE Sentiment Analyzer - TCG Aviation Clients (7/2014 - 12/2014)
- Technology:** **Python, Java, MapReduce, Text Classification, Naive Bayes Algorithm, Elastic Search, Kibana, Mahout, Hive**
- Details:** Application text classification algorithm in brand reputation and promotion management, with input from various textual data in web, crm, mail server through crawler script or third party api and store data in HDFS. Classified sentences according to degree of sentiment (*Positive / Negative / Neutral*) as well as by brand dimension (*Food / Staff / Entertainment*); providing trend summaries / KPI information across individual brands to Data Brand Owners / Market Analysts. Selected naive bayes algorithm as text classifier following a literature survey, implemented via Mahout and visualised big data via Kibana; keeping the data in elastic search.
- Responsibilities:** Requirement Analysis, Application Architecture Design, Database Schema Design, Unit Testing, Integrated Testing, Planning for Release Strategy, POC.
- Project:** Nostradamus Approximation Framework - Inventory Estimation - Pubmatic (6/2012 - 12/2012)
- Technology:** **Java, Hadoop, MapReduce, Perl, MySql, Least Square Method, Correlation Analysis**
- Details:** Generic framework for estimation and forecasting of quantitative features from a huge historical data set; 100M orders per day and project to support 3 months of historical data. Proprietary implementation of Machine Learning methods enabled realisation of quantitative target features (*No. of Impression, ECPM*) as a linear regression of set of base features (*Geo, Site, AdSize, Frequency*). Applied linear regression to implement the Least Square method. Overall average accuracy improved by 60%+ and selected premium sites in smaller data sizes gave 80% accuracy.
- Responsibilities:** Requirement Analysis, Application Architecture Design, Database Schema Design, Unit Testing, Integrated Testing, Planning for Release Strategy, POC.
- Project:** METAL - Trading Time Risk Analysis Tool - Credit-Suisse (10/2009 - 2/2010)
- Technology:** **C# .net (WinForm), Market Data API, Position management System, WCF, Gemfire API, Oracle, Sql Server**
- Details:** Application Support project to modify and enhance an in-house position management, portfolio valuation, risk summary product. The application pulls market data through Reuters API and trade information from Fidesa, calculates various portfolio valuations and risk analysis parameters, then visualizes information to Traders.
- Responsibilities:** Front / Back-end Design, Development, Unit Testing & Profiling, Beta Calculation, Average Volume Calculation, Industry Wise Index Decomposition, Exception Management, Risk Analysis Group Member (*Imagine*) of CSFB Prop-IT.

Sayan Mukhopadhyay

Page 6

Details of Engineering Project

Project:	Real Time Latency Monitoring in High Frequency Trading - Credit-Suisse (9/2008 - 10/2009)
Technology:	C++, Shell Scripting, Unix Kernel Programming, Tibco EMS, Socat, FIX protocol, Corvil, Memory map file, ARM, Kronos, My-ARM
Details:	Developed a virtual order management system and simulated to a high transaction environment. Instrumented by ARM, MY-ARM and Kronos API and compared performance characteristics, such as accuracy, instrumentation overload in execution time, memory and processor usage. Parsed log file in trading time, then designed and implemented a distributed, divide and conquer architecture for Real-time Log-file Parsing Framework. Data of potential interest is transmitted to a central server through a robust and highly tuneable transport layer. Completed heavy weight parsing in the central server separated to the production host. Divided order latency into two phases (<i>inside the box / outside the box</i>) to enable further analysis; inside the box latency measured by Kernel program / Outside the box latency collected by Corvil API using WSDL.
Responsibilities:	Design, Development, Performance Testing and Production Release of Socat base shell script, C++ base log Development.
Project:	TRIANA Dashboard - Credit-Suisse (10/2008 - 9/2009)
Technology:	C# .net (WPF), WSDL client, Nikson, SevOne, Ganglia, Oracle, Corvil
Details:	Project to develop a Trader's Infrastructure Application Network Monitoring Tool; integrating network, server, application alerts in a common dash board for all devices under a particular prop trader. Front end is a grid view of WPF application. Back ends are different adapters those run in parallel threads. For network alerts we use SevOne API through WSDL client. For server alerts we use Ganglia API through telnet client. For application alerts it gets data from web server which stores the application alert through Log-file parsing project. When a device name is entitled in dash board connected device names also come through firing query to configuration database. It is an in-house enterprise product of Credit-Suisse.
Responsibilities:	Design, Develop, Unit Testing, Profiling Data Adapter parts (Ganglia, SevOne, Application alerts), Front End Development a explorer view bind to a database through Xml Document