

SAYAN SANYAL

(510) · 309 · 1139 ◊ sayans@berkeley.edu ◊ github.com/sayan91

I help business users make sense of their data

EDUCATION

University of California, Berkeley 2018 (Expected)

Masters in Information Management & Systems; *Focus Area: Scalable Data Science*

RV College of Engineering, Bangalore 2013

B.Eng in Computer Science & Engineering; *First Class with Distinction*

TECHNICAL SKILLS

Languages	Python, SQL, C#, R, C++, VB for Applications, Unix Shell Scripting
Data Frameworks	SQL Server, Apache Spark, Hadoop, Alteryx, AWS EMR, Teradata, Impala, ERwin
Data Visualisation	Tableau, D3.js, Matplotlib, Bokeh, Microsoft PowerBI
Methods	Machine Learning, Multivariate Analysis, Natural Language Processing, MapReduce
Libraries	Scikit-Learn, NLTK, SpaCy, Keras, TensorFlow, Spark MLlib

EXPERIENCE

Thorogood Associates, BI & Analytics Consultant, London & Bangalore July 2013 - June 2016

- ***Analytics Lead:*** Blended disease and demographic data from WHO and World Bank indicators with internal campaign information to create “story telling” visual journeys for one of the largest global sustainability programs. Provided analysis and recommendation that prompted redeployment of \$45M to maximize impact. Technology: Tableau, R, SQL Server 2016
- ***Solution Architect:*** Led the design of a custom SQL query engine generating customized monthly financial decks for 43 countries, resulting in \$3.8M of savings in operating costs over 5 years. Technology: C#, JavaScript, SQL Server
- ***Data Modelling Lead:*** Designed a scalable data model for one of the largest global CPG company’s sales department that blended aggregated Nielsen ePOS data with granular internal sales data. Analysis enabled by the model resulted in redesigning the incentive structures for sales representatives.
- ***Technical Lead:*** Led Data Services team responsible for delivering insights into retailer profitability of large European retailers for products sold to them by our client. Redesigned the ETL pipelines in place for Germany that saw a 410% increase in the efficiency of insight delivery.
- ***Big Data Capability Team:*** Created a Hadoop integration PoC with a Teradata data warehouse, building Spark based ETL pipelines to create an ‘active archive’ to enable ad-hoc querying from Tableau which resulted in a 120% decrease in reporting cycles. Added reporting attributes using clustering methods using MLlib.

Indian Institute of Management Bangalore, Technical Research Assistant December 2012 - June 2016

- Developed a Python based (BeautifulSoup, Mechanize) web crawler that captured data from US Patent office, the FDA website, LinkedIn and other online sources to cleanse and curate data
- Gathered, cleansed and performed feature engineering for regression models in publication and patent data comparing the trends in innovation between the Indian and Chinese Pharmaceutical industries.

PROJECTS

Orwellian Rules November 2016

- Used techniques in Natural Language Processing and Machine Learning to create computational classifiers for George Orwell’s Rules of Effective Writing

Poverty Mapping in Rwanda February 2017

- Developing an integrated model to predict poverty in Rwanda combining day and night time satellite imagery by building transfer learning Convolutional Neural Networks and Ridge Regression models.