

Abhishek Das

Econometrics, PhD Finance, Data Scientist

Sydney, NSW

+61 423 874 530

factorwonk@protonmail.com

<https://factorwonk.github.io>

Academically trained researcher and ex-quantitative portfolio manager with machine learning skills in Natural Language Processing, Computer Vision and Deep Learning. Passionate about solving challenging business problems with proven track record of providing data-driven, action-oriented insights and quantifiable results. Relied upon to engage with and clearly communicate complex results to stakeholders.

Data Science Experience

KPMG SOLUTION 49x

July 2018 – Present

Senior Consultant (Machine Learning)

Canberra ACT • Sydney NSW

Consulting with Australian Department of Defense on developing an augmented decision-making process for their security vetting process. Used Bayesian modeling to develop an inference engine to flag security concerns around potential candidates and Natural Language Processing for identifying conflicts in candidate interview notes and timelines. Developed optical character recognition workflow to convert typed interview notes to machine readable format for ingestion to downstream insight generation processes.

Responsibilities:

- Understanding the business case from discussions with key stakeholders on client side and designing technical prototypes to showcase augmented decision-making capabilities
- Working with solution architects, automation and data engineers to transform 'as-is' current state processes to 'to-be' future state processes
- Lead data scientist responsible for developing open source NLP and OCR capabilities and provided handover and support for prototypes to client-side developers

Achievements:

- Built optical character recognition-based workflows to automate consumption of typed or printed interview notes and associated checkboxes
- Created an NLP based text contradiction prototype to find discrepancies in user histories including conflicting timelines and personal networks, summarizing results into an overlay report
- Created a Discrete Bayesian Network model using pomegranate (python library) to identify the cause of a candidate application being accepted or rejected based on several flags during the vetting process

BOSTON CONSULTING GROUP DIGITAL VENTURES

January 2018 – June 2018

Data Scientist (Engineering)

Sydney NSW

Working in an agile software engineering team, building a RESTful API that uses Natural Language Processing and Convolutional Neural Networks to query unstructured document data. Building, testing and deploying at-scale using AWS such as Redis, S3 and Aurora deployed across Kubernetes clusters.

Responsibilities:

- Worked in software development environment to deliver production ready Natural Language Processing (NLP) model to identify fields of interest from scanned images of utility bills
- Focused on NLP model refinement to identify key fields from a scanned invoice

- Delivered end-of-month data science updates to corporate partner venture board

Achievements:

- Built production ready API services to create model features and train models. Conducted POSTMAN tests to ensure services were operational
- Trained first-pass ensemble model in sandbox environment to identify fields of interest in a training set of 1000 documents

ZURICH MERCURIEN MOTOR FLEET INSURANCE

July 2017 – December 2017

Data Scientist

Sydney NSW

Consulting with an insurance startup to determine the probability of 'lease for ride share' drivers (Uber) having accidents. Created Nearest Neighbors and CART based classification algorithms, based on driver cognition and driving telemetry data, to identify 'good' versus 'bad' drivers. Realized savings of \$55 per driver hired by identifying drivers less likely to incur accidents.

Responsibilities:

- Data Science lead for establishing the link between driver cognition and safety
- Extract Transform and Load (ETL) Mercurien's driver cognition, telematics and accident claims data from unorganized csv files to PostgreSQL database
- Prepared weekly analysis reports for Mercurien CEO and COO

Achievements:

- Created a classification model to identify drivers more likely to have accidents based on driving speed and cognitive processing ability
- Used K-Nearest Neighbors, Decision Tree and Neural Networks trained on driver cognition, empathy, personality types and driver behavior
- Realized savings of \$55 per driver hired by screening for drivers more likely to have an accident

Quantitative Finance Experience

Challenger Investment Solutions

July 2016 – July 2017

Quantitative Analyst

Sydney NSW

Created automated performance attribution systems for Challenger's internal machine learning based Equity Market Neutral fund. Using a combination of Total Return Equity derivatives data and Bloomberg's Python API. Constructed daily, weekly and monthly factor return breakouts for value, momentum and low beta including trading costs.

Responsibilities:

- Co-author for monthly client reports on portfolio performance and trading ideas
- Trade execution and booking of total return swaps to provide global equity exposure for superannuation clients
- Researched a news event-based model that used twitter data to create sentiment-based equity trading signals

Achievements:

- Back-tested real-time event driven equity trading strategy using Twitter data and PySpark

- Created options back-testing tool used in monthly client facing reports
- Created Excel-VBA based portfolio attribution system for long-short EMN fund from TR Swaps database providing daily market level performance
- Created factor attribution system for systematic long-short EMN fund using data sourced from Bloomberg's Python API

Dimensional Fund Advisors

May 2015 – June 2016

Quantitative Portfolio Manager

Austin TX • Sydney NSW

Managed multi-billion-dollar quantitative equity portfolios for pension clients based on systematic, evidence based financial research.

Responsibilities:

- Managed US\$ 3 Billion in Australian and Global equity market portfolios run using supervised machine learning models
- Generated FX ordered to manage currency spending, overdrafts and negative interest currencies
- Undertook qualitative security reviews for universe eligibility and global corporate actions
- Portfolio analysis, attribution and performance write-ups for clients

Achievements:

- Created options back-testing tool used in monthly client facing reports
- Created Excel-VBA based portfolio attribution system for long-short EMN fund from TR Swaps database providing daily market level performance
- Created factor attribution system for systematic long-short EMN fund using data sourced from Bloomberg's Python API

Standard & Poor's Global Market Intelligence

Nov 2010 – June 2016

Quantitative Specialist

New York NY • Sydney NSW

Australian specialist for S&P supervised learning driven investment models using proprietary financials data. Consulting with quantitative hedge funds to feature engineer supervised sentiment, value, quality and low volatility machine learning models.

Responsibilities:

- Responsible for cross-validation of supervised machine trading models including quantitative value, momentum-driven, mean-reversion, analyst-estimate, value versus growth and volatility driven strategies
- Data specialist for S&P proprietary RDBMS databases

Achievements:

- Used dimensionality reduction (PCA) to create hybrid risk model for Chinese equity market
- Co-authored white papers and seminar presentations on systematic value research

Machine Learning Stack

Data Analysis: Webscraping, API Controllers, Network Analysis, Bayesian Modeling, Recommendation Engines

Machine Learning: Natural Language Processing, Naïve Bayes, K-means, Linear Regression, Multinomial Logistic Regression, Regularization, Bootstrapping, Cross-validation, Model Selection, Model Validation, Hyper-parameter Tuning and Monte Carlo Simulation

Deep Learning Tools: Tensorflow, Tensorboard, Keras and Pytorch

Data Mining: K-Means, PCA, DB Scan, Cosine Distance and Elbow method

Databases: SQL, AWS Redis (NoSQL) AWS AuroraDB (PostgreSQL), MySQL, and Spark

Languages: Python, R, C++, Excel/VBA, bash scripting and Scala

Cloud Computing: Amazon Web Services (AWS), EC2, S3, Apache Server, iPython Notebook Server, Google Cloud, Google Big Query, Hadoop and Microsoft Azure

Data Visualization: seaborn, matplotlib, bokeh, RShiny, ggplot2, Tableau and D3.js

Operating Systems: Ubuntu, OSX and Windows

Education & Training

Data Science Immersive

GENERAL ASSEMBLY

Doctor of Philosophy (PhD) Finance

Thesis Topic: Essays on Quantitative Hedge Fund Strategies

THE UNIVERSITY OF SYDNEY

Bachelor of Economics (First Class Honors)

Majors: Econometrics and Finance

THE UNIVERSITY OF SYDNEY