Abir Datta
Email: abirdatta4u@gmail.com Mob: +44-(0)7459 355 239

Career Objective

I wish to seek challenging assignments and responsibilities, with an opportunity for growth and career advancement as successful achievements through a continuous learning process and keep myself dynamic, visionary and competitive with the changing scenario of the world.

Professional Précis

- Working as Senior Data Scientist in Cognizant Technology Solutions Ltd in Insurance, Financial Services & Digital Analytics verticals with around 8 years of experience into Analytics & Data Science managing a team of 12 junior data scientists both at off-shore and on-shore locations.
- Key strengths include client interaction, requirement analysis, process improvements and ability to deliver critical business challenges within anticipated business timelines in addition to client facing responsibilities.
- Experience in designing & developing end-to-end Big Data integrated solutions for different verticals to cater to various data science problems.
- Developed algorithms to identify the latent characteristics of the customer-base for taking channelized strategic decisions for much more effective business successes.
- Hands-on Experience in R, Python, Big Data technologies in Windows and UNIX environment.
- Hands-on experience in different statistical modelling algorithms like Linear Regression, Logistic Regression, Severity-Frequency Modelling, Machine Learning algorithms, pricing & risk models for clients working in Banking, Financial Services & Insurance domain.

Technical Skills

- **Software Products:** R, Python, Pyspark, MATLAB/Octave, Hadoop **-** Hive, Mahout & RHadoop
- **Databases:** Oracle, SQL SERVER, PostgreSQL
- **Operating System:** Windows, UNIX
- **Applications:** MS Office

Qualifications

- Completed **Bachelor's in Science** with **Statistics** as Honours from **Calcutta University** in 2010.
- Completed **ISC (10+2)** in **Science Stream** from **CISCE** in 2006.
- Completed **ICSE** in 2004.

Career profile

1. Cognizant Technology Solutions - March'2014 - till date)

Project Title: Risk Model Developement Engagement

Client: One of the largest Global Banks

Role: Senior Data Scientist

Project details & Responsibilities:

- Develop risk models for delinquency prediction for multiple banking products like Credit Card, Unsecured Loans etc
  - Developed POC frameworks for enhancement of the existing risk models based on machine learning algorithms both in terms of performance and accuracy
  - Design and development of the risk modeling life-cycle
  - Design and develop the end-to-end framework for production environment implementation
- Develop a framework for Deep Customer Segmentation to generate a Customer 360 degree view
  - Leverage Pyspark and other Big Data technologies to create a segmentation framework for understanding the customer's spend patterns
  - Create a reusable analytics platform structure that can be leveraged to map the customer overall journey and provide end-to-end analytical insights
- Manage a team of 6 off-shore data scientists along with 4 data scientists from the client to cater to the business problems

Project Title: Data Science Engagement

Client: One of the largest Global Re-Insurance Service provider

Role: Senior Data Scientist

Project details & Responsibilities:

- Manage a team of 12 junior data scientists to cater to the multiple use-cases/business challenges brought up to the table by the client including the designing and development of the solution frameworks
- Leveraging Geospatial analysis to understand the potential market distribution for the Catastrophic Insurance
  - Geospatial Analysis of the US Flood Maps to understand potential impact of Floods and Storms on the House Valuation correlated with analysis of US Census Data and Zillow Home valuation data to understand the potential market for Home Insurance product against natural catastrophe.
  - Development of market penetration models using OLS, MARS and Bayesian LR to understand the parameters driving the rate of penetration in multiple Counties and Communities across different US States
- Algorithm development for Peer Ranking based on proximity score and Entity Classification
  - Peer Ranking Algorithm **-**
    - Peer ranking algorithm is defined as an ensemble of Hamming Distance and Euclidean Distance measures to calculate the proximity score for each entity against all entities.
    - Top **n** lowest scores are assigned as Top **n** peers for each entities and validated against a benchmark
    - Algorithm is defined as ensemble of 5 classification models developed using machine learning algorithms like C5.0, Random Forest & Extreme Gradient Boosted Trees.
    - Post selection of the final models based on the Kappa Statistics and the F1 Score, the entity dataset is scored to get the final probabilities for each class henceforth a voting framework is defined to assign the classes based on the maximum probability available for each class given for an entity
- Algorithm development for Rule-based Claims Management
  - Leverage Modern statistical methods to analyze and derive insights from historical claims data to identify key drivers for claims cost
    - Develop advanced modeling on claims data to identify targeted actions to address account deviations, based on hypothesis for key data points for identification of case and disease management
    - Risk stratification: identify chronic diseases and risk factors that are the biggest drivers for total costs & stratify patients using claim algorithms based on logic rules
  - Integrated tool for case management based on claims analytics to automatically flag high risk incoming claims for case management to ensure quality of care
    - Utilize the key findings from claim analytics to develop a rule base scoring system that could be used for case and disease management program
- Storm Event Risk Framework
  - Solution involves defining, analyzing and acquiring data available in the open-source data domain which included the details for the storm events including the locations and fatalities from NOAA database
    - Leveraged GLM, Time-Series and Machine Learning models to integrate key drivers for Storm Events in a single view along with forecasting future events occurrences and thereby driving the approach strategy through an algorithmic structure
    - Utilized US Government published data for the storm events to develop predictive framework to come up with risk bases scoring from severity, exposure and frequency
- Claims Reserve Estimation
  - Solution involves defining, analyzing and acquiring data available in the Swiss Re environment for Auto Insurance claims which included the details for the historical losses recorded
    - Developed traditional claims reserve models leveraging Chain-Ladder and Benktander methods estimate the ultimate losses
    - Developing a machine learning algorithm to estimate the incremental losses for each development period against each contract year and to predict the ultimate losses

Project Title: Design & development of CRM Daily Planner Algorithm

Client: One of the largest US-based Insurance carrier

Role: Data Scientist - Solution Architect

Project details & Responsibilities:

- Project involves in creating an analytical automated solution for the Regional Sales Consultants (RSCs) to optimize their calls and thereby enhance sales performance and profiteering
  - Analysis was conducted as a two-stream challenge and a composite solution as developed with surveying RSCs using a MaxDiff approach and calculated segment-level importance scores & identifying call type importance and compared to a paired comparison, MaxDiff
  - Assessed two analytical approached to determine scores with multi-logit results and equated segment-level scores and finalized algorithm & determined call type importance scores application to the data to calculate agent-level scores while using call type eligibility flags and importance scores to calculate scores

- The final algorithm is a combination of the call type importance scores and flags indicating if an agent is eligible for a particular call type

Project Title: Real-time Fraudulent Transaction Evaluation Solution leveraging Graph Network
Client: One of the largest US-based Insurance carrier
Role: Data Scientist - Solution Architect
Project details & Responsibilities:

- Project involves in developing an end-to-end real-time predictive fraud detection framework encompassing the broad spectrum of the investigative approach of fraud detection.
  - Providing a framework which can be deployed to score the transactions for the validation
  - Providing an early-warning system to the users of the potential cases of frauds during a transaction while the processing is ongoing

Project Title: Design & development of Smart Quotes Management System
Client: One of the largest NA-based Insurance carrier
Role: Data Scientist - Solution Architect
Project details & Responsibilities:

- Project involves in developing an analytics induced smart quote management system for P&C insurance industry
  - Design and develop the analytical component of the framework and test the functionality of the same.
  - Used logistic regression and random forest algorithms for identifying the likelihood of a quote getting accepted by the end-customer and thereby provide useful insights on the metrics driving the likelihood of the quote acceptance.

Project Title: Developing pricing risk models to predict the pure premium
Client: One of the largest P&C Insurance carriers in US
Role: Associate **-** Actuarial Analytics (Offshore)
Project details & Responsibilities:

- Pure premium modelling using Tweedie error structure modelling framework for Auto Insurance industry for one of the largest P&C Insurance carrier in US.
  - Worked on building Generalized Linear Models(GLMs) to estimate the pure premium for all coverage assuming a Tweedie error structure and using logarithmic link function in Emblem
  - Validating and scoring the data on a test sample and a separate hold-out sample using Residual testing & Time stability and consistency checks and lift metrics such gini coefficient
- Developing a static CLTV model to understand the factors impacting the CLV of a client using Decision Tree and develop rules for assigning a risk bucket based on their CLV

2. Fractal Analytics Ltd. - October 2012 - March 2014

Project title: Developing algorithm to identify the behavioral patterns of the customer-base for predicting the next action
Client: One of the Top 5 Institutional Investors based out of US
Role: Analyst
Project details & Responsibilities:

- Worked in developing algorithm for identifying the most profitable set of Financial Advisors for one of the largest financial company by
- Identifying hidden traits or characteristics from a given set of population of financial advisors which would drive the contact strategy for our client.
- Using the hidden traits as input to develop the next set of complex rules to predict the behavioral patterns for the financial advisors in the future and thereby help clients to take strategic business decisions

Project Title: Developing risk model to identify the factors driving the corporate insurance claims for the client
Client: One of the largest corporate Financial Services providers in US.
Role: Analyst
Project details & Responsibilities:

- Project involved in developing a multivariate predictive model to help client identify drivers of customer risk and develop a risk metric for each client by line of business and providing a robust framework to Pricing Review Team to appropriately bucket customers into different risk segments.

3. Tata Consultancy Services Ltd. - July 2011 - October 2012 .

Project Title: Strategic Recommendations for Retail & FMCG markets in APAC regions

Client: Global leader in Data Service provider

Role: Associate

Responsibilities:

- Ad-Hoc reporting to reason with client query on various Product-Level Data and Market-Share Distribution.
- Identifying the Halo/Cannibalization impact on the major products in the market along with Competitor Analysis with the final deliverable including strategic recommendations, price elasticity and promotional impact.

Achievements

- Consecutively awarded top ratings and have been praised for subjective matter and technical know-how by the Senior Leadership for multiple clients and internal stakeholders
- **Awarded Rock Star within the Data Science practice for contributio**n towards the enhancement of the practice
- Have been awarded Pinnacle Awards for being an outstanding performer.
- Member of the content review team for PACKT Publishing house and have been credited for the review of the **title *Machine Learning with R Cookbook* authored by David Chiu.**

This is to acknowledge that the above provided information is true and correct to the best of my knowledge. Any misrepresentation of facts or data will be entirely liable upon me.

Place: Northampton, UK

Date: 2nd December, 2018 Abir Datta