

## PROFESSIONAL EXPERIENCE

### Guidehouse Inc. | New York, USA

#### Senior Consultant – Data Science

Nov 2020 - present

- Developing a robust and reusable next-gen risk XGBoost ML model pipeline to detect fraud and money laundering activities reaching 10M+ customers with 100M+ transactional data
- Identified minimum thresholds for suspected financial crime/fraud using statistical methods, thereby reduced alerts on monthly basis by 40% and decreased the time required for Fin-Crime COPS to evaluate frauds/mules from a week to 12 hours
- Introduced new fraud rules by decommissioning existing ones, thereby improved True Positive rate from 5% to 34% in final model
- Gave meaningful insights about data. Prepared reports, written analyses, quantitative exhibits, & other client deliverables on time

### PricewaterhouseCoopers (PwC) | Columbia, USA

#### Associate

Mar 2020 – Nov 2020

- Extracted KYC and transaction data of 1.5 million accounts using SQL. Cleaned and transformed the raw data for further analysis
- Developed a predictive analytics model to predict customers with high financial crime risk concern and evaluated best predictors using Logistic Regression and Anomaly Detection. Accuracy further improved by 8% using Regularization and Decision Trees
- Automated the loan approval process by creating Natural Language Processing (NLP) tool that resulted in increased efficiency by saving 20 human hours per week and speeded up the approval cycle by 65%

### HSBC | New York, USA

#### Data Scientist Intern

Feb 2020 – March 2020

- Developed a multilabel text classifier to identify different types of risk (12 types) associated with daily news using NLP techniques. Results were used by Auditors to classify news into risks, thereby reducing time effort by 60%
- Utilized Beautiful Soup to scrape 40000 news articles, visualized the same on Word-Cloud to recognize the most common words
- Realized a 15% increase in accuracy over baseline with Naïve Bayes and Random Forest by leveraging custom tokenizer with TFIDF

### Cognizant Technology Solutions | Pune, India

#### Associate

Jan 2014 – May 2018

- Created a forecasting model to estimate next day's sales using time series (ARIMA) that helped client to know their inventory level
  - Improved model robustness by engineering new variables based on trend and seasonality to uncover the patterns of sales data, thereby reducing human efforts by 12 hours per week
  - Developed a marketing mix modeling pipeline using linear and regularized regression in the retail media sector. Increased Total ROI by 18% and secured enhancements for 3 similar additional projects.
  - Wrote sustainable SQL scripts & streamlined data validation process by automation to mitigate cross-functional dependency
- Achievement: "PILLAR of the Month" Award twice from the client at Cognizant**

## TECHNICAL SKILLS

**Languages:** SQL, Python [Pandas|Numpy], R, Snowflake, Big Query, C, C++

**Frameworks:** Sckit-learn, Matplotlib, plotly, TensorFlow, Keras, RShiny, Flask, Hadoop

**Applications:** GIT, Tableau, Power BI, Looker, Docker, Google Analytics, Unix, Agile Methodology, Jira, Bloomberg, Microsoft Office Suite

**Statistics and Machine Learning:** Regression-Linear, Logistic, Ridge, Lasso | Decision Trees | Boosting-AdaBoost, Gradient Boosting, XG-Boost | Bagging-Random Forest | SVM | Naïve Bayes | K-Nearest Neighbors [KNN] | Clustering-K-means | Principal Component Analysis [PCA] | Analysis of Variance | Natural Language Processing [NLP] | A/B Testing | Hypothesis Testing | Time Series Analysis

## EDUCATION

### Rutgers Business School | Newark, USA

Aug 2018 – Dec 2019

Master of Science: Quantitative Finance

Courses: Statistics and Machine Learning, Optimization Methods, Econometrics, Business Intelligence, Database Management, Python

### Pune Institute of Computer Technology | Pune, India

Aug 2009 – May 2013

Bachelor of Engineering: Electronics and Telecommunication, Recipient of Central Sector Scholarship throughout 4 years

## PROJECTS

**Hackathon - Bank Term Deposit Subscription:** Designed a predictive model to classify whether customers will subscribe to bank term deposit or not. Improved model accuracy with parameter tuning by 8% using XG-Boost and Identified key drivers for term deposit

**Credit Card Fraud Detection:** Developed a predictive model to detect whether a credit card transaction is fraudulent or not. Achieved minimum false negatives with Random Forest. Accuracy further improved by 7% using SMOTE with higher recall metrics