SAURABH **KULKARNI**

+1 (862) 270-7140 | sckulkarni99@gmail.com | www.linkedin.com/in/sckulkarni | https://github.com/sk2062

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

PROFESSIONAL EXPERIENCE

PricewaterhouseCoopers (PwC) | Columbia, USA

Modeling Associate **Mar 2020 - present**

- 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 Science 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
- Measured the effectiveness and ROI (Return on Investment) of various promotional channels in the mix to inform and aid brand strategy. Enhanced brand performance (~$3million) by determining optimal investments using Linear and Regularized Regression
- Determined inter-tactic relationships between promotional channels that extended engagement to 2 additional geographies
- 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

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 of Term Deposit subscription

Recommendation Engine

- Recommended top 5 similar movies to a user with the help of item-based and user-based collaborative filtering model using KNN
- Analyzed users & movie data, built popularity based recommendation system and compared the same with collaborating filtering

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

Customer Churn Prediction

- Built a predictive model to analyze the behavior of customers to retain them
- Performed exploratory data analysis, determined key factors affecting customer attrition using SVM and Decision Tree classifier

TECHNICAL SKILLS

Programming: SQL, Python [Pandas| Numpy | Scikit-learn | Matplotlib], R

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

Applications: Tableau, Power BI, Hadoop, Unix, Agile Methodology, Jira, Bloomberg, Microsoft Office Suite