Shubham Goyal

Data Scientist

shubhamgoyal769@gmail.com
8700664182
shubham769.github.io/
linkedin.com/in/shubham-goyal-0946b7127
github.com/shubham769
medium.com/@shubhamgoyal769

Data Scientist with 3+ years of experience, working on solving some of the most complex business problems using data analytics, machine learning, and interactive data visualization. Domains of expertise include Anti-Money Laundering (AML), Fraud Detection, and Supply Chain Analytics. Skilled in Machine Learning using Predictive modeling techniques such as Linear, Logistic Regression, Decision Trees, Random Forest, KNN, and XGBoost.

WORK EXPERIENCE

Data Scientist
Tookitaki AI

10/2020 - Present *Bengaluru, India*

Tookitaki is an industry leading, cloud-based AML (Anti Money Laundering) Transaction Monitoring Software provider.

Achievements/Tasks

- Part of core customer success and research/development team to have engagement with clients and stakeholders to deploy cutting edge AI products (Anti Money Laundering & Reconciliation Suite) to produce results in Banks and Financial institutions.
- Develop algorithms to monitor and detect fraudulent Transactions, payments, money laundering activities in Financial Industry. Working closely with the business leaders; cross-functional collaboration for providing solutions
- Customer Clustering and Segmentation using HDBSCAN and then using decision trees on top for interpretability.
- Finding the anomalous customers based on the features and reporting suspicious transactions Using Mahalanobis distance to catch the anomalies because of it being better than isolation forests.
- Techniques used: Data Preparation (K-fold sampling, time-series transformations, binning etc.), Feature Engineering and Variable Selection, Ensemble Models (XGBoost), Tree models, Driver Analysis

Data Science Engineer
Knoldus Inc

01/2019 - 09/2020 *Noida, India*

Achievements/Tasks

- Worked in an Agile environment and different software engineering techniques with clients across the globe.
- Developing models and algorithms for Forecasting Product and Customer segments to cluster the customers into multiple segments to understand the behavior of the customers. Tech - AWS, Python, Spark, Python, Algorithms - XGBoost, Linear Regression, Tree models.
- Worked closely with the business leaders; cross-functional collaboration for understanding domain and designing & developing solutions as proof of concepts based on the problem statement across domains.
- Worked on Scala/python along with reactive technologies like Scala, Kafka, Akka, Lagom and NoSQL databases

EDUCATION

Bachelor of Computer Science and Engineering
Lovely Professional University

08/2015 - 06/2019

Courses

- Major - Data Science

SKILLS

Data Analytics | Python | Anti-Money Laundering | Linear Algebra , Statistics | Machine Learning | Quantitative Analysis | Business-oriented approach | PySpark | Hadoop | Hive | Text Analytics | Apache | SQL | NO SQL | Risk Modeling

PROJECTS

Transactions Monitoring System (10/2020 - Present)
- Develop Data transformations (ETL), Models, Algorithms for Primary and Secondary Transaction monitoring.
- The approach uses a proprietary semi-supervised technique that is based on a combination of multi-dimensional unsupervised techniques, network analysis, and supervised learning to detect complex money laundering structuring mechanisms.
- Deployed in various multinational banks and payment providers in APAC, North-America, European region.

Name Screening System (04/2021 - Present)
- Develop Models, Algorithms for Primary and Secondary Name Screening Solution.
- The approach uses proprietary multilayered NLP and supervised techniques which combine approaches in improved matching and techniques (handles typos, spelling errors, titles, prefix/suffix, etc.) and detailed analysis of secondary information obtained either from internal sources or externally available sources.

Forecasting Platform (04/2020 - 09/2020)
- The goal of the project is to create a collaborative platform for time series analysis and forecasting between data engineers and data scientists with help of Knime for decision-makers.
- A Guided Analytic platform is created, where user can do data preparation, data aggregation, data visualisations and can choose algorithm based on time series component's analysis.

Trailblazer (01/2019 - 05/2019)
- Developed a Microservice application that is capable to consume and process real-time temperature data collected via sensors to provide monitoring, alerting, and visualization over it.
- Technology Stack: Scala, Lagom, Kafka, Cassandra, Prometheus, Grafana, Docker, Kubernetes.

ACHIEVEMENTS

Knime certified (03/2020 - 06/2020)
Certified with knime level 1- level 4 certifications

Rising star Award (02/2020 - 02/2020)
For showing immense growth as an emerging leader in a staggeringly short period of time

LANL Earthquake Prediction, Kaggle (03/2019 - 05/2019)
Won silver medal and top 2% spot accross the globe

DeepLearning.AI TensorFlow Developer (06/2019 - 10/2019)
Tensorflow Deep learning certified developer