Saurabh Ghadge

Data Science Professional | NLP & ML Enthusiast

+917507653730 saurabhghadge84@gmail.com linkedin.com/in/nameis-saurabh github.com/nameissaurabh

Profile Summary

Data Scientist with around 3 years of experience crafting and deploying cutting-edge **machine learning and NLP** solutions tailored for the BFSI sector, specializing in Anti-Money Laundering (AML) and risk assessment. Proficient in harnessing the power of **Generative AI, Large Language Models (LLMs), and advanced statistical and machine learning techniques** to transform compliance frameworks, streamline operations, and deliver actionable intelligence.

Recognized for designing and validating impactful ML and NLP models that enhance regulatory adherence, boost operational efficiency, and effectively mitigate risks. Passionate about **leveraging Generative AI expertise & data-driven insights to pioneer innovative solutions**, enabling organizations to navigate complex regulatory landscapes with confidence & agility.

Education

Master's in Applied Statistics and Informatics **August 2020 – July 2022**
Shivaji University, Kolhapur, Maharashtra

Bachelor of Science in Statistics **June 2017 – May 2020**
K.T.H.M. College Nashik, Maharashtra

Technical Skills

Programming Languages: Python, R, SQL, PySpark

Tools: VS Code, Anaconda, Google-Colab, Kaggle, Langchain, Langsmith, Hugging Face, Github, AWS (S3, EC2, SageMaker), Snowflake, Databricks, My-SQL, aisuite

Skills: Machine-Learning, ML-Ops, Statistical Modeling, Object-Oriented Programming (OOP), **NLP, NLTK, Large Language Models, Tensorflow, Pytorch, PySpark,LLM Fine tuning, PEFT,** Supervised-Unsupervised Machine Learning, Feature Engineering, Auto-Encoders, Data Analysis, Data Visualization, DSA, Model/Process Documentation

Professional Experience

Data Scientist and AML Analytics **April 2024-present**
Solytics Partners Pune, Maharashtra

- Actively engaged in **Fine-Tuning** and optimizing **Large language models** (Generative AI applications) to enhance AML processes of financial institute specifically in TMS alert management and AML screening, driving advancements in automation and compliance workflows.
- Validated advanced **AI/ML-based** Correspondent Banks Money Laundering (CBML), Trade-Based Money Laundering (TBML) and risk models for a bank, focusing on **ETL validation, feature engineering, and model monitoring, developing challenger models and data drift metrics** to ensure compliance and improve model reliability.
- Developed production-level optimized **supervised ML models** for alert risk scoring and dynamic customer risk rating, delivering a more **adaptive and data-driven alternative to proprietary traditional models**, improving accuracy and business decision-making.
- Developed **unsupervised ML models for transaction related anomaly detection to identify suspicious patterns** in large-scale financial datasets, enhancing fraud detection and reducing false negatives up-to 20%.

Associate Data Scientist and Quantitative Analyst **February 2023 – March 2024**
Solytics Partners Pune, Maharashtra

- **Developed Generative AI applications** for Adverse Media Screening (AMS), leveraging power of **large language models** to enhance automation in alert management and compliance processes.
- Built optimized challenger **machine learning models for FI's proprietary risk models** using supervised ML techniques **(kernel methods, ensemble models, ANN/DNN)** to benchmark and improve classification model performance, increasing recall and SAR detection efficiency.
- Designed and optimized AML scenarios for **large financial datasets using advanced machine learning and PySpark**, reduced false positives by upto 30% and improving the detection of Suspicious Activity Reports (SARs) based on risk levels.
- **Validated AML TM models (SAS, KIA, Actimize, proprietary systems) by ensuring the accuracy of ETL layers and the robustness of key Machine Learning models**, including customer segmentation, risk rating, and alert prioritization. Enhanced risk capture and adherence to regulatory standards.

Quantitative Analyst Intern **July 2022 – January 2023**
Solytics Partners Pune, Maharashtra

- Developed Anti-Money Laundering (AML) scenarios for Middle Eastern banks using advanced machine learning and rule based statistical methodologies, adhering to Central Bank of the UAE guidelines, enhancing regulatory compliance.
- Executed stress testing of screening tools through text manipulation using **Natural Language Processing (NLP)** to evaluate system reliability and sensitivity, ensuring robustness in Know Your Customer and Customer Due Diligence processes.

Projects

Large Language Model Pipeline for Adverse Media Screening

- ***Technologies: Python, Hugging Face, Transformers, LangChain, Google BERT, RAG, Docker, PySpark, NLP***
- Developed and successfully implemented a **large language model pipeline** designed and **fine-tuned to summarize entity-related information** fetched from proprietary adverse media screening alerts and **classify entities as adverse media respect to AMS keywords** using Retrieval-Augmented Generation (RAG).
- This pipeline became instrumental in reducing false positives for AML screening, achieving a reduction of nearly 25-30% in false hits, significantly improving efficiency.
- Developed **alert description generation** using fine-tuned BERT-based LLMs, enhancing actionability and clarity for compliance teams in managing real-time TM alerts.

Orion: Real-Time Adverse Media Screening Application

- ***Technologies: Generative AI, LangChain, LangSmith, Serper, Hugging Face, Google Gemini-Flash, LLM***
- Built **Orion**, a Gen-AI application for real-time adverse media screening and compliance checks.
- Developed a complete Gen AI-driven pipeline using **Langsmith and Langchain** that Integrated real-time internet data retrieval and validation with regulatory and sanction databases.
- **Fine-tuned** LLMs to generate concise, risk-focused summaries of adverse screened alerts with **supporting rationales and references.** ***Click Here***.

Customer Risk and Alert Prioritization Models Supervised Models

- ***Technologies: Python, Machine Learning, AWS(S3, EC2, SageMaker), PySpark, TensorFlow***
- Developed and deployed advanced **Customer Risk Rating (CRR), Alert Risk Screening** and **Dynamic Customer Risk Assessment (DCRA)** models for Middle Eastern banks, enhancing risk prioritization of alerts in transaction monitoring.
- Applied ML techniques **(kernel-based, probability-based, tree-based, ANN)** and conducted **exhaustive feature engineering** within transaction behaviors, using historical SARs, and demographics attributes.
- Improved risk model reliability and scalability through **automated data pipelines** using AWS services.
- Enhanced alert prioritization models, achieving measurable improvements in risk mitigation and compliance workflows.

Customer Segmentation and anomaly detections for Corporate Banks

- ***Technologies: Python, Clustering, PCA, Auto-Encoders, Feature Engineering, anomaly detection***
- Designed and implemented **unsupervised models** for **customer segmentation and transaction anomaly detection** in UAE based corporate FI's.
- Conducted detailed **feature engineering** on transaction behavior and demographics, leveraging dimensionality reduction techniques like **feature selection, PCA, t-SNE, UMAP auto-encoders** to create a fine-grained and highly expressive feature set, enhancing model performance and interpretability.
- Improved clustering accuracy and enabled **targeted marketing strategies**, better customer risk assessments, and insights.

AML Scenario Design and Optimization

- ***Technologies: Python, PySpark, Machine Learning, Statistical Modeling***
- Designed, validated, and **optimized** AML scenarios for transaction monitoring in Middle Eastern and US banks.
- Reduced false positives upto 30% by refining alert breaching thresholds using advanced **statistical techniques and ML algorithms** with stabilized trade-offs between SAR's and False Positives .
- Governance documentation for validation processes, findings, and recommendations to improve system reliability.

AML Screening and Transaction Monitoring Validations

- ***Technologies: Python, SQL, ETL, Data Visualization, Statistical Modeling, NLP***
- Validated AML screening and transaction monitoring systems, ensuring **accurate ETL processes** and robust data flow.
- Enhanced **system reliability** by identifying and resolving data inconsistencies through collaboration with engineering teams. Also, Conducted stress testing on Screening systems for UAE banks using **NLP and statistical simulations**, improving robustness under variable conditions.
- Authored comprehensive governance documentation outlining testing methodologies, results, and actionable recommendations to improve system reliability and compliance with KYC and AML standards.

Achievements

- Consecutively awarded the Spotlight Award in recognition of outstanding contributions to the organization for two consecutive years (***2023***,***2024***). (Click on years to see awards)

Certifications

- Continuously engaged in learning through platforms like Coursera and Udemy, with relevant course completion particularly in **Generative AI, Machine Learning, and LLM's** fine tuning with certificates attached to my LinkedIn profile. TO view certifications - ***Click Here***.

Extracurricular

- Actively engage in solving coding challenges on LeetCode, Kaggle competitions to enhance problem-solving skills and algorithmic thinking. Regularly read analytics blogs related to data science, as a continuous eagerness to learn and understand new developments in the field.