Saurabh Ghadge

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

Objective

Data Scientist with around three years of experience in developing and **implementing advanced machine learning solutions in BFSI** for Anti-Money Laundering (AML) and risk assessment. Skilled in leveraging **Generative AI, Large Language Models (LLMs), Statistical/Machine Learning Modeling**, and data analysis to enhance compliance frameworks, optimize processes, and generate actionable insights. Proven ability to design, validate, and implement impactful machine learning models that drive operational efficiency and mitigate risk. Eager to apply Generative AI expertise and data science skills to innovative projects that support regulatory adherence and drive forward-thinking solutions.

Education

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

Bachelor's 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, Airflow, Databricks

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

Professional Experience

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

- Continuously engaged in **fine-tuning** and optimizing **large language models** (Generative AI applications) In AML, 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 monitoring model, developing challenger models and data drift metrics** to ensure compliance and improve model reliability.
- Developed production-level **supervised ML models** for alert risk scoring and dynamic customer risk rating, delivering a more **adaptive and data-driven alternative to existing models**, improving accuracy and business decision-making.
- Developed **unsupervised ML models for anomaly detection and transaction clustering to identify suspicious patterns** in large-scale financial datasets, enhancing fraud detection and reducing false negatives up-to 20%.

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

- **Designed and optimized AML scenarios for large financial datasets using advanced machine learning and PySpark**, reducing false positives by upto 30% and improving the detection of Suspicious Activity Reports (SARs) based on risk levels.
- Built 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.
- **Developed and fine-tuned Generative AI applications** for Adverse Media Screening (AMS) and Anti-Money Laundering, leveraging advanced language models to enhance automation and compliance processes.
- Developed challenger models using supervised machine learning techniques such as **Validated AML 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.
- Collaborated with clients to continuously improve both projects by incorporating feedback and aligning with evolving compliance regulations.

Projects

Large Language Model Pipeline for Adverse Media Screening

- **Technologies: Python, Hugging Face, Transformers, LangChain, Google BERT, RAG, Docker, PySpark**
- 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) using **BERT.**
- 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.

Orion: Real-Time Adverse Media Screening Application (Self-Learning Project)

- **Technologies: Generative AI, LangChain, LangSmith, Serper, Hugging Face, Google Gemini-Flash**
- 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 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 Feature Engineering**
- 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 in transaction monitoring.
- Applied ML techniques **(kernel-based, probability-based, tree-based, ANN)** and conducted **exhaustive feature engineering** using transaction data, historical SARs, and demographics.
- 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 Using Unsupervised Clustering for Corporate Banks

- **Technologies: Python, Clustering, PCA, Auto-Encoders,Feature Engineering**
- Designed and implemented **unsupervised clustering models** for customer segmentation in corporate banking.
- Conducted detailed **feature engineering** on transaction behavior and demographics, applying dimensionality reduction techniques like **PCA** and **auto-encoders.**
- 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 breaching thresholds using advanced **statistical techniques and ML algorithms.**
- 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 (2022-2024).

Certifications

- Continuously engage 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. You can view my 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.
- Enjoy playing Sudoku, and cricket, my favorite leisure activity, fostering teamwork and strategic thinking.