Siva Kumar Chaini
Data Scientist
USA | Mail: sikuch06@gmail.com | Ph:**+1 (202) 381-6580** | LinkedIn

PROFESSIONAL SUMMARY:

- Data Scientist with around 4+ years of experience in financial services, specializing in machine learning, NLP, frauddetection and credit risk modelling across global institutions.
- Expertise in Basel III-compliant models (PD, LGD, EAD), stress testing (CCAR, DFAST) and regulatory reporting, ensuring compliance and audit readiness.
- Proficient in developing real-time fraud detection systems with Databricks, Apache Kafka, PySpark and Azure Synapse, achieving precision and reducing false positives significantly.
- Skilled in building predictive models (XGBoost, LightGBM, LSTMs) for loan default prediction, time-series forecasting and portfolio risk management, boosting accuracy.
- Strong experience in NLP (BERT, FinBERT, Hugging Face) for sentiment analysis, credit risk signals and entity classification, improving early warning systems and risk detection.
- Adept in data engineering and MLOps, including ETL pipelines (SQL , Airflow, NiFi, Spark, AWS Glue), model deployment (Docker, Kubernetes) and lifecycle automation with MLflow.
- Proven track record of delivering business impact through interactive dashboards (Tableau, Power BI, Streamlit) and model interpretability with SHAP values for stakeholder transparency.

TECHNICAL SKILLS:

Programming Languages:	Python, SQL
Operating Systems:	Windows, Linux, Mac OS
Machine Learning Libraries:	Scikit-learn, XGBoost, LightGBM, Optuna, SHAP, MLflow
Deep Learning & Time Series:	LSTM, Transformers
Data Processing & EDA:	Databricks, PySpark, Pandas, NumPy, Apache Spark
Natural Language Processing:	BERT Transformers, NLP (Named Entity Recognition, Sentiment Analysis)
Cloud & Orchestration:	Azure Synapse Analytics, AWS, Docker, Kubernetes
Model Deployment:	RESTful APIs, Docker, Kubernetes
MLOps & Experiment Tracking:	MLflow, Apache Airflow
Model Validation & Evaluation:	GridSearchCV, Optuna, SHAP, Monte Carlo Simulation
Data Storage:	SQL Databases, NoSQL Databases
Natural Language Processing:	NLP, Named Entity Recognition (NER), Sentiment Analysis, Text Classification
Business Intelligence Tools:	Tableau, Power BI, Excel
Development & Version Control:	Git, Agile, Scrum
Reporting & Compliance:	CCAR, DFAST, Basel III, Capital Adequacy Frameworks

JPMorgan Chase & Co., Tx, USA August 2024 – Present Data Scientist

- Implementing real-time fraud detection systems using Apache Kafka, and PySpark, detecting anomalous transactions with 97% precision and lowering false positives by 15%, significantly improving fraud prevention efficiency
- Developing Basel III-compliant credit risk models (PD, LGD, EAD) in Python using ensemble models and survival analysis, improving regulatory compliance accuracy by 20% and reducing quarterly reporting timelines.
- Enhancement in loan default prediction by building XGBoost and LightGBM models, optimized with GridSearchCV and Optuna, boosting model AUC by 13% compared to baseline approaches.
- Performed data quality checks, feature-level data profiling, and data modeling to ensure coverage, accuracy, and reliability of datasets used for model training and performance evaluation.
- Communicated analytical findings, model performance, and data insights to product managers, engineering, and non-technical stakeholders to support data-driven decision-making and model adoption.
- Built and optimized scalable data pipelines using Databricks, PySpark, and Apache Airflow, improving data processing performance and enabling efficient access to large-scale datasets for model development and analytics.
- Utilized SQL and Databricks for feature extraction, data validation, and model input preparation, improving data usability and supporting high-quality machine learning model development.
- Built LSTM-based time series forecasting models to predict loan delinquencies and assess macroeconomic stress, improving forecasting precision by 25% over traditional ARIMA methods.
- Leveraged Generative AI models such as OpenAI GPT-4, Hugging Face Transformers, and LangChain to automate extraction and summarization of financial filings, analyst reports, and market news—improving early risk-signal detection by 30%.
- Developed a Retrieval-Augmented Generation (RAG) pipeline using FAISS vector databases, OpenAI embeddings, and LangChain agents for contextual Q&A on 10-K filings, cutting manual document-review time by 40%.
- Applied **Prompt Engineering techniques** to fine-tune LLM outputs for financial tasks, optimizing prompts, context windows, and chain-of-thought flows to ensure factual accuracy and domain alignment
- Performed Monte Carlo simulations for stress testing fixed income portfolios, quantifying tail-risk exposures under adverse scenarios and validating capital adequacy against regulatory thresholds.
- Engineered domain-specific NLP systems using **FinBERT, NER, and Hugging Face pipelines** to classify financial entities and detect tone in analyst reports, improving risk-signal detection by 28%.

- Automated **ML and GenAI lifecycle management** using MLflow and Apache Airflow for hyperparameter tuning, experiment tracking, and retraining pipelines, improving model refresh efficiency by 30%.
- Built **explainable AI dashboards** in Streamlit and SHAP to visualize model impact and interpret GenAI-driven risk predictions, enabling transparent decision-making.

HCL Tech- India Feb 2020 - July 2023 Data Scientist

- Designed and deployment of advanced predictive models using XGBoost and Scikit-learn, improving loan approval accuracy by 28% and strengthening data-driven decision-making processes for the financial institution.
- Built and optimized scalable ETL and data processing pipelines using Databricks, PySpark, and Azure Data Factory, improving data availability and supporting high-quality feature engineering for machine learning models.
- Engineered and deployed a RESTful API for model serving with a sub-200ms response time, improving client-facing application performance and overall user experience.
- Performed exploratory data analysis (EDA) on complex financial datasets to identify key patterns and trends, supporting strategic, data-driven business initiatives.
- Developed, validated and fine-tuned machine learning models for predictive analytics, ensuring high precision in financial forecasting and directly supporting organizational strategic planning.
- Utilized SQL, Databricks, and NoSQL databases for data extraction, transformation, and feature preparation, improving data quality, consistency, and accessibility for analytics and model development.
- Designed and implemented natural language processing (NLP) models for sentiment analysis and text classification using TensorFlow and Hugging Face Transformers, improving customer sentiment interpretation accuracy by 22%, enabling targeted service enhancements.
- Containerized and deployed machine learning models using Docker and orchestrated them with Kubernetes, ensuring scalability, high availability and resilience in cloud environments.
- Designed a customer segmentation model with K-means clustering and RFM analysis, boosting cross-selling revenue by 18% and increasing customer retention by 20%.
- Designed interactive Tableau and Power BI dashboards for KPI visualization, enabling executive-level insights and enhancing reporting turnaround by 30%.
- Collaborated in Agile/Scrum environments, actively participating in sprint planning, daily stand-ups and retrospectives to ensure on-time delivery and continuous improvement.
- Established a scalable A/B testing framework for evaluating new product features, resulting in a 15% lift in product adoption.
- Integrated Lang Chain to develop contextual AI assistants capable of handling IT documentation, FAQs, and user queries with dynamic retrieval-augmented generation (RAG) workflows.
- Worked with advanced LLMs using Lang Chain for semantic search, summarization, and chatbot applications, enhancing automated responses accuracy by over 90% across IT support use cases.
- Built time-series forecasting models with time series models, enhancing investment decision accuracy and market trend predictions by 20%.
- Enhanced data storage and retrieval efficiency with Hadoop and HBase, reducing query response latency by 55% for largescale datasets.
- Engineered and deployed a credit scoring API using Flask on Azure Cloud, enabling real-time loan risk assessments and faster approval workflows.

EDUCATION:

Master of Computer Science - University of North Texas, Texas, USA
Bachelor of Computer Science - Raghu Engineering College, Visakhapatnam Andhra Pradesh, India

Certifications:

AWS Certified Cloud Practitioner – AWS
DevOps Beginner to Advance – Udemy
100 days of coding Python Pro Bootcamp - Udemy