

RAHUL YADAV

B.Tech Mathematics & Computing · Machine Learning · Quantitative Finance · Generative AI
+91-7379609531 | rahul20806@gmail.com | [LinkedIn](#) | [GitHub: rahul20805](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Mathematics and Computing undergraduate with foundations in Machine Learning, Quantitative Finance, and Generative AI. Experienced in building end-to-end ML pipelines, financial data analysis, LLM fine-tuning (LoRA), Monte Carlo simulation, and Limit Order Book feature engineering for algorithmic trading. Proficient in probability theory, linear algebra, and statistical modeling. Seeking a Quantitative Research or Machine Learning internship to solve real-world problems through data-driven mathematical models.

TECHNICAL SKILLS

Programming Languages: Python, C++, SQL

Machine Learning: Supervised Learning, Regression, Classification, Decision Trees, Gradient Boosting, Feature Engineering, Model Evaluation, Cross-Validation

Quantitative Finance: Monte Carlo Simulation, Geometric Brownian Motion, Time Series Analysis, Risk Modeling, Limit Order Book (LOB) Analysis, HFT Strategy Evaluation, Option Pricing

Generative AI: LLM Fine-tuning (LoRA / PEFT), Prompt Engineering, AI Automation

Libraries & Tools: NumPy, Pandas, Scikit-learn, Matplotlib, Jupyter Notebook, Git, VS Code

Mathematics: Probability Theory, Linear Algebra, Statistics, Optimization, Numerical Methods

PROJECTS

Limit Order Book HFT — Beyond Accuracy: Profitability-Driven ML Evaluation *[Running]* [↗ GitHub](#)

- Designed an ML evaluation framework for High-Frequency Trading using Limit Order Book (LOB) features — shifting the performance metric from classification accuracy to real trading profitability indicators such as P&L, Sharpe ratio, and maximum drawdown.
- Engineered key LOB-derived features including order imbalance, bid-ask spread dynamics, queue depth, and trade flow toxicity to capture short-term price movement signals.
- Built and compared Gradient Boosting, Random Forest, and logistic classifiers; developed a backtesting engine to measure actual strategy returns, revealing significant divergence between model accuracy and market profitability.
- Technologies: Python, Pandas, NumPy, Scikit-learn, Backtesting Framework, Matplotlib.

Palm Leaf Manuscript Digitization & OCR [internship]-IIT TIRUPATI [↗ GitHub](#)

- Built an end-to-end AI pipeline to digitize and extract text from ancient palm leaf manuscripts, addressing the challenge of degraded, irregular, and hand-scribed characters.
- Applied advanced image preprocessing techniques — adaptive binarization, morphological noise removal, skew correction, and contrast enhancement — to prepare manuscript images for accurate OCR recognition.
- Integrated Tesseract OCR with custom configuration to recognize historical script characters, contributing to the digital preservation and searchability of rare cultural heritage documents.
- Technologies: Python, OpenCV, Tesseract OCR, PIL/Pillow.

Monte Carlo Simulation for Option Pricing-college project [↗ GitHub](#)

- Simulated thousands of stock price trajectories using Geometric Brownian Motion (GBM) to estimate fair values of European call and put options under stochastic price dynamics.
- Applied variance reduction techniques (antithetic variates, control variates) to significantly reduce simulation error and improve convergence speed.
- Validated simulated prices against Black-Scholes closed-form solutions; performed sensitivity analysis on volatility (σ), risk-free rate (r), and time-to-expiry (T) parameters.

Fine-tuning Large Language Models with LoRA (PEFT) [Demo project] [↗ GitHub](#)

- Applied Low-Rank Adaptation (LoRA), a parameter-efficient fine-tuning (PEFT) technique, to adapt large pretrained language models to domain-specific tasks without full retraining.
- Curated and preprocessed domain-specific datasets; evaluated model performance using BLEU, F1, and perplexity metrics before and after fine-tuning to quantify improvement.
- Optimized inference efficiency for deployment, reducing GPU memory footprint while maintaining task-specific accuracy gains.

AI Smart Train Traffic Control System [hackathon project] [↗ GitHub](#)

- Developed a supervised ML model to predict train delays using simulated railway operational data, enabling proactive scheduling adjustments to reduce cascading delays.
- Applied feature engineering on route, weather, and load variables; compared Decision Tree, Random Forest, and Logistic Regression models, achieving highest accuracy with ensemble methods.

EDUCATION

B.Tech — Mathematics and Computing

2024 – 2028

(SMVDU), Jammu & Kashmir

Coursework: Machine Learning, Probability & Statistics, Linear Algebra, Data Structures & Algorithms, Numerical Methods, Optimization

PCM — Physics, Chemistry, Mathematics

2021 – 2023

Central Hindu Boys School (BHU), Varanasi

CERTIFICATIONS

- Machine Learning Specialization (3-course series) — Stanford University & DeepLearning.AI
- Advanced Learning Algorithms · Supervised Machine Learning: Regression & Classification — Stanford University
- Goldman Sachs Risk Management · J.P. Morgan Investment Banking · Deloitte Data Analytics — Forage
- HackerRank Software Engineer Certification

ACHIEVEMENTS & EXTRACURRICULARS

- Published Author & Poet — authored and published original literary works in poetry and prose; recognized for creative writing in Hindi and English
- 5-Star Problem Solving Badge — HackerRank
- Campus Ambassador — E-Cell IIT Bombay
- Vice Secretary — Hindi Cell,
- Public Speaker & Storyteller at cultural and literary events

RESEARCH INTERESTS

Quantitative Finance · Algorithmic & High-Frequency Trading · Machine Learning for Finance · Generative AI · Financial Data Science · Natural Language Processing