JAMES LEWIS

Data Scientist

James066lewis@gmail.com | linkedin.com/in/james-lewis3/ | github.com/KetchupJL

Data Scientist (Exeter MSc Data Science & Statistics, Distinction) with end-to-end expertise in exploratory analysis, feature engineering, model building and production deployment. I combine deep data diagnostics with a *simple first, complex when needed* modeling philosophy and robust cloud-native pipelines to deliver high-impact solutions in tech, fintech and blockchain - such as LightGBM trading bots yielding 15–20% simulated ROI and strategic insights from commercial datasets. Renowned for lateral thinking and tenacity, I transform complex data into scalable, revenue-driving models.

EDUCATION

University of Exeter | MSc Data Science and Statistics (Distinction, 78% avg)| 2024-25

- Key methods: supervised & unsupervised learning, Bayesian inference, time-series forecasting, hierarchical & spatial–temporal modelling.

Loughborough University | BSc (Hons) Economics | 2019-23

SKILLS & TECH STACK

Languages & Tools: Python | R | SQL | Git | Bash

Data Science & ML: Feature Engineering | EDA |Data Cleaning | Model Selection | Supervised & Unsupervised Learning | Cross-Validation | Hyperparameter Tuning | Time Series Analysis | Statistical Modelling | A/B Testing

Libraries: scikit-learn | LightGBM | XGBoost | CatBoost | statsmodels | TensorFlow | PyTorch | Keras

Databases & Cloud: PostgreSQL, BigQuery, Redshift, SQLite, Google Cloud Platform (BigQuery, Cloud Functions)

Visualisation: ggplot2 | Plotly | Tableau | Seaborn | Matplotlib | Dash | Streamlit

RELEVANT EXPERIENCE

Self-Employed | Quantitative Trader & Analyst| 2023–Present

- Built automated Python/TypeScript/SQL pipelines and engineered trading indicators for real-time ML strategies, delivering live alpha signals to Discord (<1s latency) and achieving 15–20% simulated ROI via robust backtesting and retraining workflows.

Hudl | Intern, Revenue Operations | Oct 2023–Apr 2024

- Leveraged SQL to aggregate and visualise 30+ product features, equipping sales with market-research insights
- Collaborated with cross-functional teams to define KPIs and measure feature impact; built a live-updating intranet spreadsheet that halved data-retrieval time (≈8 hrs/week saved).

Project Highlights — Full Data Science & ML Portfolio on GitHub

MSc Dissertation (in progress): Risk-Adjusted Interval Forecasting of 72-Hour Returns in Mid-Cap Solana Tokens

- Aggregated 6+ months of 12-hour OHLCV, order-book, on-chain, and sentiment data for tokens ≥ $50 M
- Engineered 12-h features: momentum, realised volatility, ATR, liquidity, wallet growth, social mentions
- Built and compared three interval-forecast models: Linear Quantile Regression, LightGBM + bootstrap, and QRF (10/50/90 % quantiles)
- Evaluated via rolling CV (pinball loss, coverage) and statistical tests; devised tail-risk-aware position-sizing rules

Podcast Listening Time Prediction (Kaggle) | Python, LightGBM, Feature Engineering

- Engineered time-series & user features; tuned a LightGBM model via time-based CV to finish Top 20% (RMSE 18.5), and delivered a fully reproducible pipeline on GitHub.

MLTradingBot: Real-Time Buy-Signal Classifier | Python, LightGBM, APIs, Pandas, PostgreSQL

- Designed high-throughput pipelines to process minute-level OHLCV data and extract advanced momentum/volatility signals for 1000+ Solana tokens.
- Developed, tuned, and deployed a LightGBM classifier with 12–15% simulated ROI and 74% backtest hit rate.
- Automated end-to-end ingestion, feature updates, and live buy-signal notifications with robust scheduling and failover handling.

CERTIFICATIONS & INTERESTS

Clubs & Leadership: University of Exeter Google Student Developer Club (Advanced Python workshops); Computer Science Society (Hackathons & AI projects); Finance & Investment Society

Interests: Brazilian Jiu-Jitsu; Hackathons; Crypto & Trading