Data Science In FinTech

Landing a data science role hinges on mastering the fundamentals - **coding, statistics, and machine learning.** These pillars **represent about 80% of what it takes to secure a role in data science.** For any aspiring data scientist, honing these skills should be the primary focus.

However, once these fundamentals are solid, candidates can shorten their job search by adding another crucial element: **business domain-specific projects**. Tailoring your portfolio to showcase expertise in industry-relevant problems not only highlights your practical application skills but also sets you apart in a competitive job market.

This guide will deep dive into **Data Science projects in the Finance Industry.**

Finance has multiple sub-fields and we will dive into 6 of the main sub-fields that hire Data Scientists (DS) or Data Analysts (DA). For each sub-field we will briefly cover the **business model** and **metrics** you should be aware of (especially when communicating during your interviews), in addition to **providing project ideas** you can build out.

High Frequency Trading

Utilizes ML to make real-time trading decisions, executing orders at high speeds based on market data.

Business Model

Companies in this space generate revenue through the spread (difference between buy and sell prices), arbitrage opportunities, and market-making activities. The emphasis is on high volume and low latency trades.

Companies

- **Jane Street**: A quantitative trading firm that heavily relies on ML for HFT.

- **Citadel Securities**: Known for using advanced algorithms in market-making and trading.

- **Two Sigma**: A firm that uses machine learning to predict market movements and execute trades.

Key Metrics

- **Sharpe Ratio**: Measures the performance of an investment relative to its risk.

- **Execution Speed**: The time it takes to execute a trade, crucial in HFT.

- **Order Fill Rate**: The percentage of trade orders that are successfully executed.

Some of the projects you can consider to stand out are -

Predictive Market Making

- **Objective**: Develop a model to predict bid and ask prices in real-time, allowing for more efficient market-making.

- **Dataset**: Kaggle High-Frequency Trading Dataset

- **ML Techniques**: Time Series Analysis, Reinforcement Learning

- **Metrics**: **Sharpe Ratio**, **Execution Speed**

Volatility Prediction for Trading Strategies

- **Objective**: Create a model to predict short-term market volatility, helping to optimize trading strategies.

- **Dataset**: Historical Intraday Stock Market Data

- **ML Techniques**: GARCH Models, LSTM Networks

- **Metrics**: **Prediction Accuracy**, **Volatility Forecast Error**

Arbitrage Opportunity Identification

- **Objective**: Identify arbitrage opportunities across multiple exchanges or assets using real-time data.

- **Dataset**: Crypto Arbitrage Dataset

- **ML Techniques**: Regression Models, Anomaly Detection

- **Metrics**: **Profit per Trade**, **Execution Latency**

Credit Scoring and Risk Management

ML models are used to assess creditworthiness, predict defaults, and manage risk portfolios.

Business Model

Revenue is generated through lending (interest on loans), credit products, and risk assessment services.

Companies

- **FICO**: Provides credit scoring and analytics for risk management.

- **Zest AI**: Uses machine learning to offer AI-powered credit underwriting.

- **Kabbage**: A fintech company that uses AI to offer small business loans with rapid credit decisions.

Key Metrics

- **Default Rate**: The percentage of loans that default.

- **Credit Score Accuracy**: How accurately the model predicts creditworthiness.

- **Loss Given Default (LGD)**: The amount of loss a lender incurs when a borrower defaults.

Some of the projects you can consider to stand out are -

Credit Risk Prediction

- **Objective**: Build a credit scoring model to predict the probability of default for loan applicants.

- **Dataset**: LendingClub Loan Data

- **ML Techniques**: Logistic Regression, Boosted Trees, SHAP/PDP/ICE for explanation

- **Metrics: AUC-ROC, Accuracy, F1 Score**

Dynamic Credit Limit Adjustment

- **Objective**: Develop a model to dynamically adjust credit limits based on user behavior, or predict risk profile, or predict default risk of customers.

- **Dataset**: Credit Card Data

- **ML Techniques:** Logistic Regression, Boosted Trees, SHAP/PDP/ICE for explanation

- **Metrics**: **Default Rate**, **Customer Retention Rate**

Loan Portfolio Risk Optimization

- **Objective**: Optimize the risk-return profile of a loan portfolio using machine learning.

- **Dataset**: Home Credit Default Risk Dataset

- **ML Techniques:** Portfolio Optimization, Monte Carlo Simulations

- **Metrics**: **Sharpe Ratio**, **Expected Shortfall**

Fraud Detection and Prevention

ML models are deployed to detect and prevent fraudulent activities, especially in transactions.

Business Model

Companies in this space earn through transaction fees, subscriptions for fraud detection services, and penalties recovered from fraud attempts.

Companies

- **Stripe**: Uses machine learning to detect and prevent payment fraud.

- **Feedzai**: Provides fraud prevention solutions powered by AI.

- **PayPal**: Employs advanced ML models to secure transactions and prevent fraud.

Key Metrics

- **False Positive Rate**: The percentage of legitimate transactions incorrectly flagged as fraudulent.

- **Fraud Detection Rate**: The percentage of fraudulent activities accurately identified.

- **Chargeback Rate**: The proportion of transactions that are disputed by customers due to fraud.

Some of the projects you can consider to stand out are -

Real-Time Transaction Fraud Detection

- **Objective**: Build a real-time fraud detection model to flag suspicious transactions as they occur.

- **Dataset**: Credit Card Fraud Detection Dataset

- **ML Techniques**: Anomaly Detection, Autoencoders, Isolation Forest

- **Metrics**: **Precision**, **Recall**, **False Positive Rate**

User Behavior Profiling for Fraud Prevention

- **Objective**: Develop user behavior profiles to detect unusual activities that might indicate fraud.

- **Dataset**: Synthetic Financial Datasets For Fraud Detection

- **ML Techniques**: Logistic Regression, Decision Trees, Gradient Boosting

- **Metrics**: **AUC-ROC**, **F1 Score**

Portfolio Management and Robo-Advisors

ML is used to optimize investment portfolios and automate financial advice.

Business Model

Revenue is typically generated through management fees, advisory fees, and performance-based fees.

Companies

- **Wealthfront**: A robo-advisor that uses ML to manage portfolios.

- **Betterment**: Utilizes AI to provide automated, personalized investment advice.

- **BlackRock**: Uses ML for portfolio management and financial market analysis.

Key Metrics

- **Portfolio Return**: The overall return on investment for the managed portfolio.

- **Risk-Adjusted Return**: Measures portfolio performance relative to the risk taken.

- **Customer Retention Rate**: The percentage of clients that continue to use the service over time.

Some of the projects you can consider to stand out are -

Personalized Portfolio Recommendation

- **Objective**: Build a recommendation system that suggests personalized investment portfolios based on user preferences and risk tolerance.

- **Dataset**: Portfolio Management Dataset (See this notebook for an example)

- **ML Techniques**: Collaborative Filtering, Matrix Factorization

- **Metrics: User Satisfaction Score**, **Portfolio Performance**

Portfolio Risk Prediction Using ML

- **Objective**: Develop a model to predict the risk associated with different portfolio compositions.

- **Dataset**: Create your own portfolio combination using Historical Stock Prices

- **ML Techniques**: Time Series Forecasting, Monte Carlo Simulations

- **Metrics: Value at Risk (VaR)**, **Expected Shortfall**

Automated Rebalancing of Portfolios

- **Objective**: Implement an automated system to rebalance investment portfolios based on market conditions.

- **Dataset**: Create your own portfolio combination using Historical Stock Prices

- **ML Techniques**: Reinforcement Learning, Optimization Algorithms

- **Metrics**: **Portfolio Return**, **Transaction Costs**

Financial Forecasting and Analytics

Involves using ML to forecast financial metrics like stock prices, economic indicators, and company performance.

Business Model

Companies provide forecasting tools, financial analytics software, and consultancy services.

Companies

- **Bloomberg**: Uses AI for financial news and market data analytics.

- **Refinitiv**: Offers financial data and analytics tools powered by AI.

- **Kensho Technologies**: Provides predictive analytics for financial markets.

Key Metrics

- **Prediction Accuracy**: How closely the forecasts match actual outcomes.

- **Mean Absolute Error (MAE)**: A common metric for evaluating forecast accuracy.

- **Economic Value Added (EVA)**: A measure of a company's financial performance based on residual wealth.

Some of the projects you can consider to stand out are -

Stock Price Prediction

- **Objective**: Build a model to predict stock prices using historical data and market indicators.

- **Dataset**: <u>S&P 500 Stock Data</u>

- **ML Techniques**: LSTM Networks, ARIMA, Gradient Boosting

- **Metrics**: **Mean Absolute Error (MAE)**, **R-Squared**

Economic Indicator Forecasting

- **Objective**: Predict key economic indicators such as GDP growth or unemployment rates using machine learning.

- **Dataset**: <u>World Bank Economic Indicators</u>

- **ML Techniques**: Time Series Forecasting, Bayesian Models

- **Metrics**: **Forecast Accuracy**, **Root Mean Square Error (RMSE)**

Earnings Call Sentiment Analysis

- **Objective**: Analyze the sentiment of earnings calls to predict stock price movements.

- **Dataset**: <u>Earnings Call Transcripts</u>

- **ML Techniques**: NLP, Sentiment Analysis, Text Classification

- **Metrics**: **Sentiment Prediction Accuracy**, **Impact on Stock Prices**

Customer Personalization and Marketing Analytics

ML models are used to personalize customer experiences, optimize marketing campaigns, and improve customer segmentation.

Business Model

Revenue is generated through enhanced customer retention, cross-selling, and personalized product recommendations.

Companies

- **Capital One**: Uses machine learning for customer segmentation and personalized offers.

- **JPMorgan Chase**: Employs AI for targeted marketing and customer relationship management.

- **American Express**: Uses ML to personalize customer interactions and optimize marketing strategies.

Key Metrics

- **Customer Lifetime Value (CLTV)**: The total value a customer is expected to bring over their relationship with the company.

- **Conversion Rate**: The percentage of marketing efforts that lead to successful customer actions.

- **Churn Rate**: The rate at which customers stop doing business with the company.

Some of the projects you can consider to stand out are -

Personalized Financial Product Recommendations

- **Objective**: Develop a recommendation system that suggests personalized financial products based on user behavior.

- **Dataset**: Bank Marketing Dataset

- **ML Techniques**: Collaborative Filtering, Content-Based Filtering

- **Metrics**: **Click-Through Rate (CTR)**, **Conversion Rate**

Churn Prediction in Banking

- **Objective**: Build a model to predict customer churn and suggest retention strategies.

- **Dataset**: Churn Modeling Dataset

- **ML Techniques**: Classification Models (Random Forest, SVM), Survival Analysis

- **Metrics**: **Churn Rate**, **Retention Rate**

Customer Segmentation for Targeted Marketing

- **Objective**: Segment customers based on transaction data for more targeted marketing campaigns.

- **Dataset**: Retail Banking Dataset

- **ML Techniques**: Clustering (K-means, Hierarchical Clustering)

- **Metrics**: **Segmentation Accuracy**, **Marketing Campaign ROI**

Hope this helps! Feel free to reach out to us for everything DS/ML -

LinkedIn - https://www.linkedin.com/company/buildml

Instagram - https://www.instagram.com/buildmledu/

YouTube - https://www.youtube.com/@buildmledu