

## Appendix External References and Dependencies

### 1. API Services and Documentation

#### CoinMarketCap (CMC)

- Documentation URL: <https://coinmarketcap.com/api/documentation/v1/>
- API Endpoint: <https://pro-api.coinmarketcap.com/v2>
- Purpose: Primary cryptocurrency market data, pricing, and volume information
- Usage: Real-time and historical cryptocurrency data collection
- Tier: \$79/month recommended for comprehensive data access
- Key Features: OHLCV data, market cap rankings, global metrics

#### Polygon.io

- Documentation URLs:
  - Crypto: <https://polygon.io/docs/rest/crypto/overview>
  - Indices: <https://polygon.io/docs/rest/indices/overview>
  - Stocks: <https://polygon.io/docs/rest/stocks/overview>
  - Forex: <https://polygon.io/docs/rest/forex/overview>
  - Economy: <https://polygon.io/docs/rest/economy/overview>
  - REST API: <https://polygon.io/docs/rest/quickstart>
  - S3 Flat Files: <https://polygon.io/docs/flat-files/quickstart>
- S3 Endpoint: <https://files.polygon.io>
- API Endpoint: <https://api.polygon.io>
- Purpose: Multi-asset class financial data provider
- Usage: Primary data source for cryptocurrency, stocks, forex historical data
- Key Features: S3 bulk data access, real-time WebSocket feeds, aggregated bars

#### Bybit Exchange

- Documentation URL: <https://bybit-exchange.github.io/docs/>
- Purpose: Cryptocurrency derivatives exchange API
- Usage: Trading execution, order management, market data for derivatives
- Key Features: Futures, perpetuals, options trading capabilities

#### Santiment/Sanbase

- Documentation URL: <https://academy.santiment.net/sanapi/>
- Purpose: On-chain analytics and social sentiment data
- Usage: Enhanced features using blockchain metrics and social signals
- Key Features: Network activity, developer activity, social volume metrics

#### Yahoo Finance (yfinance)

- Note: Marked as unreliable since 2024 due to rate limiting
- Purpose: Free historical market data
- Usage: Fallback data source when primary APIs fail
- Limitations: Frequent rate limiting, data quality issues

---

## 2. Data Sources

### Primary Data Sources

Source	Type	Reliability	Cost	Use Case
Polygon.io	REST/S3	High	Paid	Primary historical data
CoinMarketCap	REST	High	\$79/month	Real-time crypto data
Bybit	WebSocket/REST	High	Free (trading fees)	Derivatives data
Sanbase	REST	Medium	Paid	On-chain metrics
yfinance	REST	Low	Free	Emergency fallback

#### Data Types Collected

- OHLCV: Open, High, Low, Close, Volume bars
- Tick Data: Trade-by-trade data for high-frequency analysis
- Order Book: Market depth and liquidity metrics
- On-chain: Network metrics, wallet movements, mining data
- Social: Reddit, Twitter sentiment indicators

---

### 3. External Libraries

#### Core Data Processing

Library	Version	Purpose
pandas	2.2.2	DataFrames, time series manipulation
numpy	>=1.26,<2.0	Numerical computations, array operations
pyarrow	>=16.0.0,<17	Parquet file format, efficient storage
python-dateutil	2.9.0.post0	Date/time parsing and manipulation

#### API Clients

Library	Version	Purpose
ccxt	>=4.1.0,<5	Unified cryptocurrency exchange interface

Library	Version	Purpose
finance	>=0.2.37,<0.3	Yahoo Finance data fetcher
requests	2.32.3	HTTP client for REST APIs
boto3	>=1.34,<2	AWS S3 client for Polygon flat files
Machine Learning		
Library	Usage	Purpose
scikit-learn	Extensive	Classical ML models, preprocessing, metrics
PyTorch	Preferred	Deep learning research and experimentation
TensorFlow	Alternative	Production deployment option
XGBoost	Ensemble	Gradient boosting for tabular data
LightGBM	Ensemble	Fast gradient boosting alternative
Technical Analysis		
Library	Purpose	
TA-Lib	Technical indicators (RSI, MACD, Bollinger Bands)	
ta	Pure Python technical analysis library	
vectorbt	Backtesting and portfolio optimization	
Visualization		
Library	Purpose	
matplotlib	Static plotting and charts	
seaborn	Statistical visualizations (Set2 palette)	
plotly	Interactive dashboards and 3D plots	
dash	Web-based interactive dashboards	
Utilities		

Library	Version	Purpose
python-dotenv	>=1.0.1,<2	Environment variable management
pydantic	>=2.7,<3	Data validation and settings
tenacity	>=8.2.3,<9	Retry logic for API calls
tqdm	>=4.66,<5	Progress bars for long operations
hydra-core	Config	Configuration management
omegaconf	Config	YAML configuration parsing

---

#### 4. Academic and Research References

##### Research Foundations

- Walk-Forward Optimization: Based on Pardo (1992), enhanced with modern ML techniques
- Diebold-Mariano Test: Statistical test for forecast comparison
- Time Series Cross-Validation: Bergmeir & Benitez (2012) methodology

##### Key Research Areas

- Market Microstructure: Liquidity, spread, depth analysis
- Portfolio Optimization: Markowitz framework with crypto adaptations
- Risk Management: VaR, CVaR, maximum drawdown metrics
- Feature Engineering: Technical, fundamental, and alternative data features

##### Citation Format

Project: Cryptocurrency ML Trading Research

Year: 2023-2024

Type: Research and Educational Purpose

License: Academic/Research Use

---

#### 5. Development Tools and Platforms

##### Cloud Platforms

Platform	Purpose	Configuration
AWS EC2	GPU instances	g4dn.xlarge recommended

Platform	Purpose	Configuration
Paperspace Gradient	Free GPU tier	Development and testing
Google Colab	Free notebooks	Prototyping and education

#### GPU Acceleration

- CUDA: NVIDIA GPU acceleration for deep learning
- cuDF/RAPIDS: GPU-accelerated pandas operations
- Requirements: CUDA 11.0+ for PyTorch GPU support

#### Container and Deployment

- Docker: TensorFlow and model deployment containers
- GitHub Actions: CI/CD pipeline configuration
- Watchman: Facebook's file watching service for development

#### Version Control

- Repository: <https://github.com/ManavA/claude.git>
- Branch Strategy: main branch for stable code
- Git LFS: Large file storage for model weights

---

## 6. Web Resources and URLs

### Package Repositories

- PyPI: <https://pypi.org/simple> (Python Package Index)
- NVIDIA PyPI: <https://pypi.nvidia.com/> (CUDA packages)
- Conda Forge: Alternative package source for scientific computing

### External Tools

- TA-Lib Source: <http://prdownloads.sourceforge.net/ta-lib/ta-lib-0.4.0-src.tar.gz>
- TA-Lib GitHub: <https://github.com/mrjbq7/ta-lib#installation>
- Dashboard: <http://localhost:8050> (Local Dash/Plotly server)

### Communication APIs

- Telegram Bot API: <https://api.telegram.org/bot{token}/sendMessage>
- Purpose: Trading alerts and notifications

### Social Data Sources (Mentioned but not implemented)

- Reddit: r/cryptocurrency, r/bitcoin for sentiment
- Twitter/X: Crypto influencer sentiment tracking
- Discord/Slack: Community sentiment indicators
- Medium: Technical analysis articles

---

## Usage Notes

### Priority Stack

1. Data Collection: Polygon.io (primary), CoinMarketCap (secondary)
2. Processing: pandas/numpy with pyarrow for storage
3. ML Framework: PyTorch for research, scikit-learn for baselines
4. Backtesting: Custom engine with vectorbt validation

5. Visualization: matplotlib/seaborn for reports, plotly for interactive

#### API Key Management

- All keys stored in `.env` file (gitignored)
- Environment variables loaded via python-dotenv
- Fallback logic implemented for API failures
- Rate limiting handled with tenacity retry logic

#### Performance Optimization

- S3 bulk downloads preferred over REST APIs
- Parquet format for efficient storage
- GPU acceleration available for deep learning
- Caching layer at `data/cache/` directory

#### Reliability Hierarchy

1. Most Reliable: Polygon S3 flat files (bulk historical)
2. Reliable: CoinMarketCap, Bybit (with rate limits)
3. Moderate: Sanbase (on-chain data delays)
4. Unreliable: yfinance (frequent failures)

---

#### Appendix Notes

- API key security (environment variables)