

ML Phase 1 Methodology

Implementation Analysis

Based on: Phase 1 methodology.xlsx by Saleem Ahmad

Analysis Date: December 07, 2025

Project: AIAlgoTradeHits

AIAlgoTradeHits - ML Phase 1 Implementation Analysis

Executive Summary

Total Features in Phase 1 Methodology: 20

COMPLETE	PARTIAL	TO ADD
6 features (35%)	4 features (20%)	10 features (45%)

Training Data Available (Already Collected)

Symbol	Daily (10yr)	Hourly (1mo)	5-min (1wk)	Total
BTCUSD	3,651	720	2,304	6,675
QQQ	2,512	137	390	3,039
SPY	2,512	137	390	3,039
TOTAL	8,675	994	3,084	12,753

AIAlgoTradeHits - ML Phase 1 Implementation Analysis

Detailed Feature Analysis

#1	OHLCV + Timestamp	DONE	Effort: 0 min
----	-------------------	------	---------------

BigQuery: open, high, low, close, volume, datetime

Why: B

#2	Weekly Return (% change)	DONE	Effort: 0 min
----	--------------------------	------	---------------

BigQuery: percent_change, weekly_change_percent

Why: M

#3	Weekly Log Return	TO ADD	Effort: 5 min
----	-------------------	--------	---------------

BigQuery: weekly_log_return (FLOAT64)

Why: B

#4	Multi-lag Returns (2w/4w)	TO ADD	Effort: 10 min
----	---------------------------	--------	----------------

BigQuery: return_2w, return_4w (FLOAT64)

Why: M

#5	RSI(14)	DONE	Effort: 0 min
----	---------	------	---------------

BigQuery: rsi

Why: C

#6	RSI slope / z-score / flags	TO ADD	Effort: 15 min
----	-----------------------------	--------	----------------

BigQuery: rsi_slope, rsi_zscore, rsi_overbought_flag, rsi_oversold_flag

Why: D

#7	MACD(12,26,9)	DONE	Effort: 0 min
----	---------------	------	---------------

BigQuery: macd, macd_signal, macd_histogram

Why: M

#8	MACD Histogram + Cross flag	PARTIAL	Effort: 10 min
----	-----------------------------	---------	----------------

BigQuery: macd_histogram exists, need macd_cross_flag

Why: H

#9	SMA 20/50/200	DONE	Effort: 0 min
----	---------------	------	---------------

BigQuery: sma_20, sma_50, sma_200

Why: D

#10	EMA 20/50/200	PARTIAL	Effort: 10 min
-----	---------------	---------	----------------

BigQuery: ema_12, ema_26 exist; need ema_20, ema_50, ema_200

Why: E

AIAalgoTradeHits - ML Phase 1 Implementation Analysis

#11	MA Distance % (close vs MA)	TO ADD	Effort: 10 min
	BigQuery: close_vs_sma20_pct, close_vs_sma50_pct, close_vs_sma200_pct		Why: M
#12	EMA Slopes (20/50)	TO ADD	Effort: 10 min
	BigQuery: ema20_slope, ema50_slope (FLOAT64)		Why: C
#13	ATR(14) + ATR%	DONE	Effort: 0 min
	BigQuery: atr		Why: B
#14	ATR z-score / slope	TO ADD	Effort: 15 min
	BigQuery: atr_zscore, atr_slope (FLOAT64)		Why: N
#15	Bollinger Bands (20,2) + Width	PARTIAL	Effort: 5 min
	BigQuery: bollinger_upper/middle/lower exist; need bb_width		Why: D
#16	Volume z-score / ratio	TO ADD	Effort: 10 min
	BigQuery: volume_ratio, volume_zscore (FLOAT64)		Why: C
#17	ADX(14) + DI+/DI-	PARTIAL	Effort: 20 min
	BigQuery: adx exists; need di_plus_14, di_minus_14		Why: C
#18	Pivot High/Low flags	TO ADD	Effort: 20 min
	BigQuery: pivot_high_flag, pivot_low_flag, pivot_strength		Why: S
#19	Distance to last pivot	TO ADD	Effort: 15 min
	BigQuery: dist_to_last_pivot_high_pct, dist_to_last_pivot_low_pct		Why: G
#20	Numeric Regime State	TO ADD	Effort: 30 min
	BigQuery: regime_state (INT64), regime_confidence (FLOAT64)		Why: B

AIAlgoTradeHits - ML Phase 1 Implementation Analysis

AIAlgoTradeHits - ML Phase 1 Implementation Analysis

Implementation Timeline

Phase 1A: Quick Wins	30 minutes
-----------------------------	-------------------

- #3 Log Return
- #4 Multi-lag Returns
- #11 MA Distance %
- #15 BB Width

Phase 1B: Momentum Enhancements	45 minutes
--	-------------------

- #6 RSI derivatives
- #8 MACD Cross
- #10 EMAs
- #12 EMA Slopes
- #14 ATR derivatives
- #16 Volume z-score

Phase 1C: Advanced Features	60 minutes
------------------------------------	-------------------

- #17 ADX + DI
- #18 Pivot Points
- #19 Distance to Pivot
- #20 Regime State

TOTAL IMPLEMENTATION TIME: 135 minutes (~2.5 hours)

AIAlgoTradeHits - ML Phase 1 Implementation Analysis

Expected Model Accuracy

Phase	Features	Expected Accuracy
Current (7 complete)	35% of features	55-58%
Phase 1 Complete (20)	100% of features	58-63%
Phase 1.5 (+4 features)	24 features total	66-72%
High-Probability Setups	All features optimized	75-85%

Immediate Next Steps

1. Run ML_Training_Quick_Start.ipynb with existing 7 features
2. Add Phase 1A features (30 min) - quick wins
3. Add Phase 1B features (45 min) - momentum enhancements
4. Add Phase 1C features (60 min) - advanced features
5. Train ensemble model for 66-72% accuracy
6. Deploy to Vertex AI for production