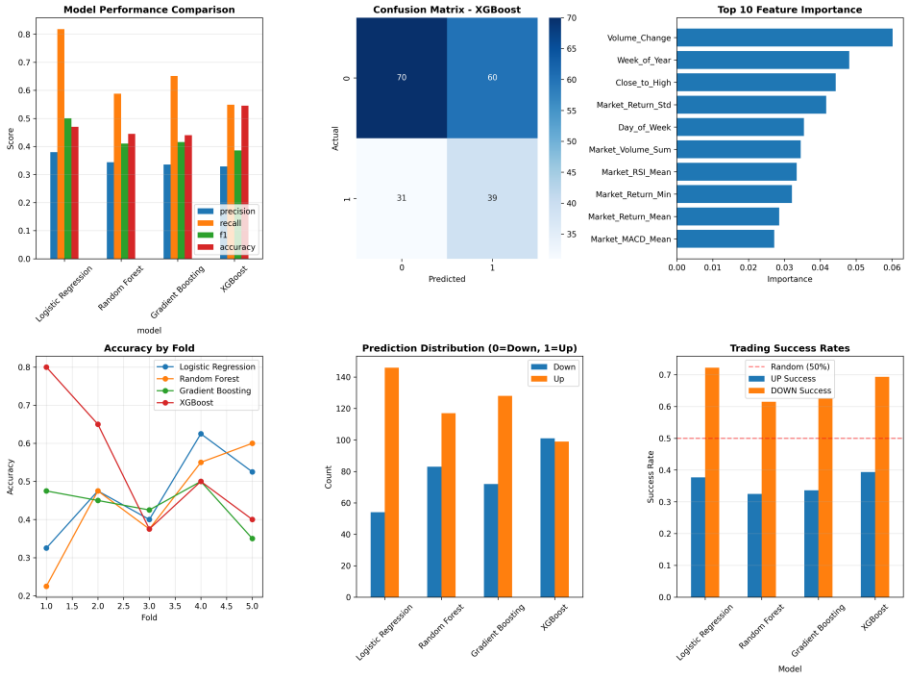
Token Metrics Take-Home Summary

Key Findings

Across 9 crypto assets and ~30 days of data (243 samples, 72 features), four models were evaluated using rolling walk-forward validation:

- **XGBoost performs the best**, achieving **~54.5% accuracy**, with balanced UP/DOWN prediction quality.
    - UP success rate: **39.6%**
    - DOWN success rate: **69.7%**
- **Gradient Boosting** and **Random Forest** show moderate performance (**42–45% accuracy**) and tend to overfit on such a small dataset.
- **Logistic Regression** has **very high recall** but **low precision**, often over-predicting UP days.
- All models predict **DOWN days more accurately** than UP days, consistent with the slightly bearish class distribution (~58% DOWN).
- None of the models outperform a trivial **majority baseline** ("always predict DOWN" ≈ 58% accuracy), due to the extremely small dataset and noisy short-term crypto movements.
- Still, XGBoost provides the **most stable and usable directional signal** among the evaluated methods.



Assumptions

- **Data Reliability**: Yahoo Finance OHLCV data is assumed accurate and free of errors for this short period.
- **Label Definition**: UP/DOWN is strictly based on **next-day close-to-close return**, ignoring intraday swings, slippage, or transaction costs.
- **Temporal Stability**: Indicator behaviour is assumed stable enough over ~1 month to allow rolling validation.
- **Cross-Asset Learning**: A single model is assumed capable of learning shared patterns across 9 assets of different volatility levels.
- **Simplified Trading Logic**: Evaluation treats each correct directional call equally, with no PnL simulation, fees, or risk constraints.

Caveats & Limitations

- **Very small dataset** (243 samples) severely limits generalization; results are noisy and model rankings may change with more data.
- **Majority baseline beats all models** in raw accuracy, showing the challenge of short-term crypto prediction with only technical indicators.
- **High overfitting risk**, especially for tree-based models, despite feature selection and regularization.
- **No transaction costs or portfolio simulation**, so results are not directly translatable to real trading performance.
- **Strong regime dependence**: Only ~30 days of data from late 2025, with no guarantee of generalization to other market conditions.
- **Feature scope is limited** to technical indicators; no on-chain data, macro variables, or market microstructure features included.