Cryptocurrency Pump-and-Dump Detection
from Pre-Spike On-Chain Signals

Gillian Schriever · Jiahui (Cecilia) Cai · Zhilin Chen · Jinghan Huang

Harvard John A. Paulson School of Engineering and Applied Sciences · AC297r Capstone · Spring 2026

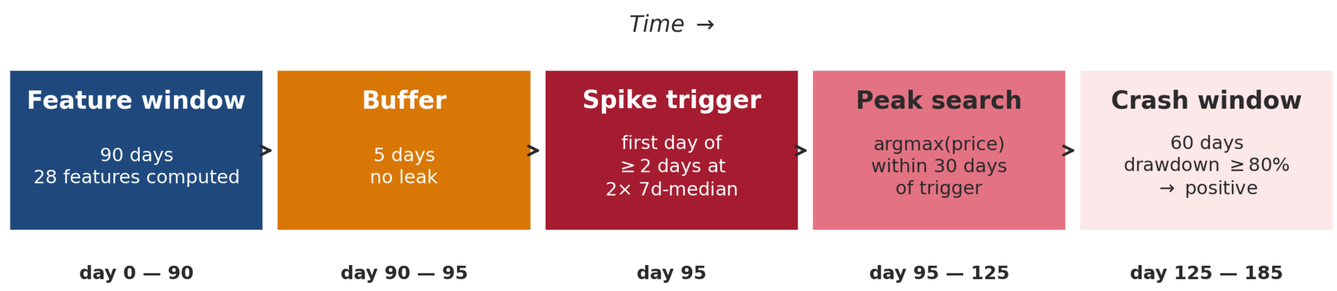
Why predict P&D?

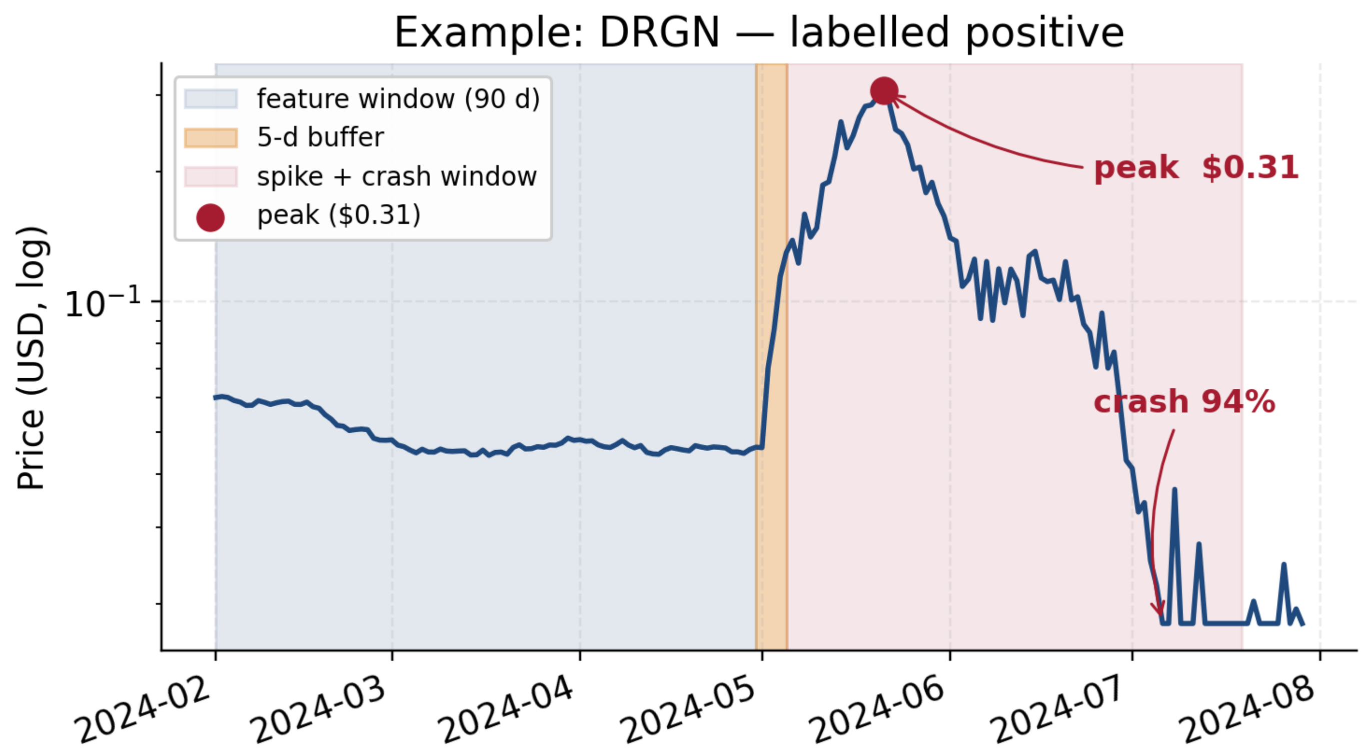
Pump-and-dump is the dominant fraud type in small-cap crypto. By the time a token is "obviously" a P&D, retail buyers have already lost money. We ask: using only *pre-spike* on-chain data — price, volume, wallet structure — can we flag the crash before it happens?

Working data: **2,040 small-cap Ethereum tokens** after deduplication and a 30-day pre-spike history filter — **1,580 P&D events** vs **460 controls** ($\sim 3.4{:}1$).

What is a P&D event?

We label as positive the first multi-day price spike whose peak is followed by a ≥ 80% drawdown within 60 days. Earlier "warmup" spikes that didn't crash are skipped — otherwise a token whose real P&D is the second event would be miscoded ambiguous.





Take-home

One sentence. A leakage-free supervised pipeline flags small-cap P&D events at AUC 0.886 from on-chain features that exist *before* the pump is visible.

What's new. Wallet-concentration + ETH-residual features push past volume-only baselines; a 5-day buffer + shared-window design keeps the score honest.

What's next. Network features (transfer cycles, deployer reputation) are the obvious lift. The pipeline is already production-shaped: Dune $\rightarrow$ batch pull $\rightarrow$ ranked candidates.

Engineering design choices

5-day buffer. The feature window ends 5 days before `spike_start`. The earliest days of a pump already reflect the manipulation; including them would leak the very signal we're trying to predict.

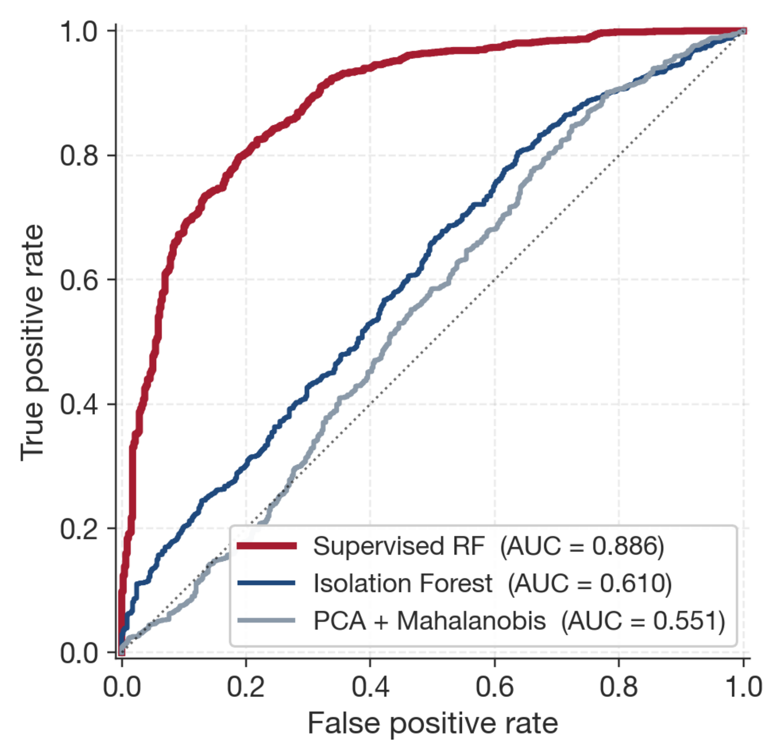
Beta normalization. Each token's daily log return is regressed on log ETH return: $r_{\text{tok}} = \alpha + \beta_{\text{ETH}}\, r_{\text{ETH}} + \varepsilon$. The residual block (volatility, max-gain, max-drawdown, Sharpe) isolates idiosyncratic — non-market-driven — movement. ETH-only, not BTC+ETH, because $\rho(\text{BTC}, \text{ETH}) \approx 0.79$ destabilises the joint regression.

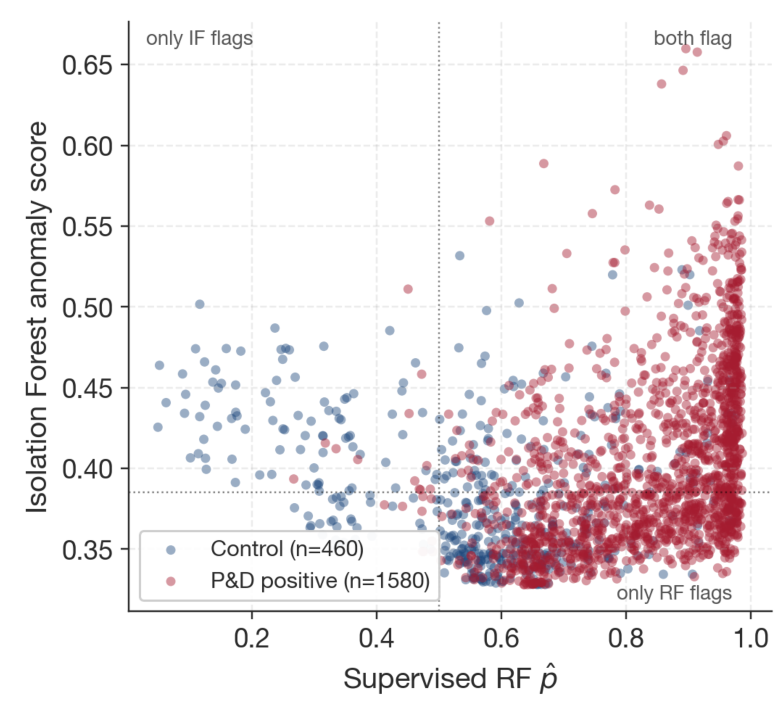
Shared time window. Candidates and controls are drawn from the same 2023–2025 window. Without this, the model could learn to distinguish bull from bear regimes rather than P&D from non-P&D.

Does it predict?

0.886

AUC-ROC · Random Forest · N = 2,040 · stratified 10-fold CV



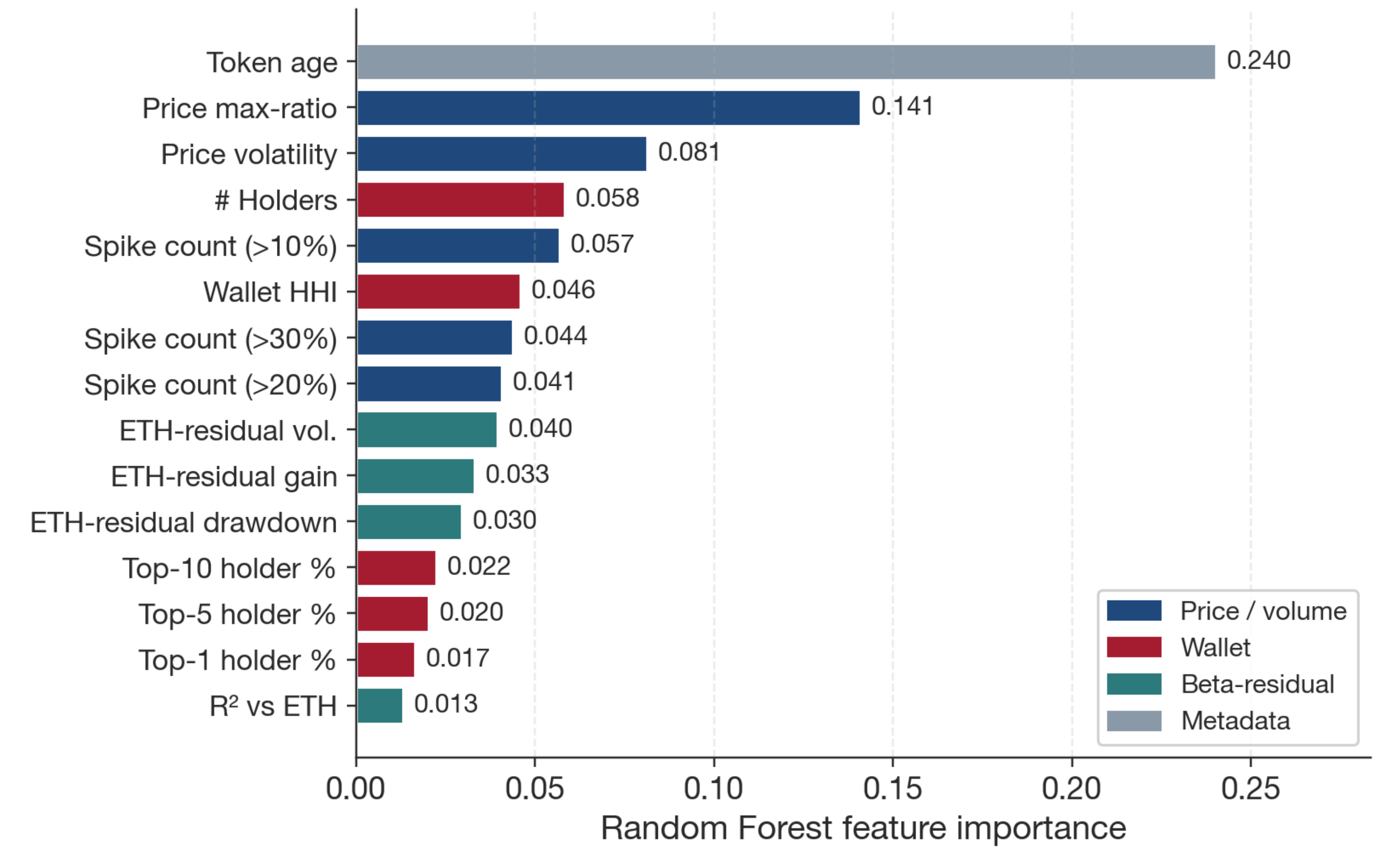


Both unsupervised baselines lag (Isolation Forest 0.610, PCA+Mahalanobis 0.551). Concordance with the supervised model is moderate ($\rho = 0.41$): the methods agree on the most obvious cases (top-right) but the supervised model uniquely flags a band of positives in the lower-right — tokens whose anomaly is subtle on volume features but visible in wallet HHI and beta-residuals.

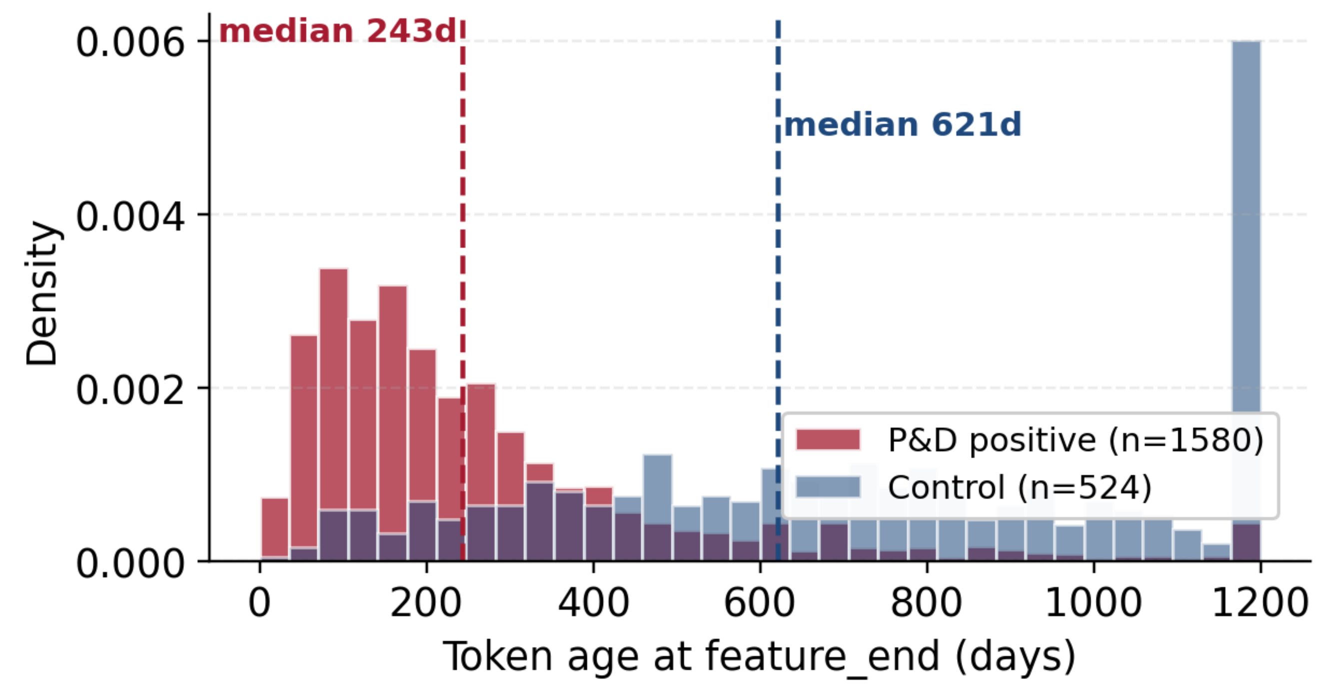
Highest-risk predictions. All top-five are confirmed P&D events; the worst false alarm is a control at $\hat{p} = 0.975$.

#	Token	True label	$\hat{p}$
1	TLOS	✓ P&D	0.986
2	UBEX	✓ P&D	0.986
3	FOUR	✓ P&D	0.985
4	DRGN	✓ P&D *(panel left)*	0.985
5	GCR	✓ P&D	0.985
worst FP	*COPYBOT*	control	0.975

What drives the prediction?



Token age alone explains 24% of the model's variance — younger tokens are far more likely to be P&D substrates. The histogram below confirms it:



Pre-spike **price max-ratio** and **volatility** round out the top three. The beta-residual block (idiosyncratic price action) contributes ~ 10%, supporting the prior that P&D price moves depart systematically from ETH-correlated drift.

Limitations & next steps

Our label catches the symptom, not the intent. A labelled positive could be a benign collapse, and a labelled control could be a slow manipulation. Network features such as transfer-graph cycles and developer reputation are some promising ways to improve AUC.