Structural Break - Alphabot's Team Solution

Humberto Brandão, Ph.D. (humberto.brandao@alphabot.com.br)

João Peinado (joao.peinado@alphabot.com.br)

Mario Filho (mario.filho@alphabot.com.br)

Rafael Alencar, M.Sc. (rafjaa@gmail.com)

October 2025

1 Introduction

This team is composed of Quantitative Traders from Alphabot, a Brazilian quantitative firm. Before describing our solution, we provide a brief introduction to the company:

Alphabot is a proprietary trading firm operating in the trillion-dollar global trading industry, currently focused on high-frequency trading (HFT) and relative-value arbitrage strategies in Brazil.

Our flagship HFT fund has delivered 300% annualized returns since inception in 2023, with a Sharpe ratio above 5, while our equity fund, launched in 2021, has already achieved a 12x multiple.

We operate with BRL 45M of proprietary capital, fully deployed in our own strategies and with no external funding to date. Our AI-driven, low-latency infrastructure enables us to scale trading volume and profitability more than 10x with no increase in cost base, and performance models indicate scalability of up to 50x without significant decay.

We are structured as a long-term partnership driven by data and research, where technical excellence, intellectual honesty, and continuous learning are core values. Our edge lies in cutting-edge research, scalable infrastructure, and relentless execution discipline in highly competitive markets.

Our team is world-class in technology and quantitative research, bringing together serial fintech founders, entrepreneurs, and PhDs or PhD candidates in Computer Science and AI applied to algorithmic trading. The broader group includes former Google engineers, Kaggle Masters, ACM ICPC World Finalists, and former professors. We have built a high-retention partnership, with a 25-person team averaging 3.4 years of tenure in a 5-year-old firm.

We are advised by the founders of Méliuz (B3:CASH3) and are currently exploring strategic partnerships to accelerate our expansion and unlock the next stage of growth.

2 Our approach

The Alphabot solution, like many classification systems we have built for real-world applications, as well as for other Kaggle competitions, is based on Stacking.

We usually allow team members to work independently in order to generate diversity in both the features and the models created. This way, we can combine them in a second layer of machine learning to improve the final score obtained.

Another important aspect of our solution was ensuring that our validation was correct, which allowed us to divide the workload among all four team members. This gave us confidence that no data leakage occurred between different layers of the stacking model.

This emphasis on diversity tends to work well when dealing with problems of this nature. The work is highly parallelized among all members, maximizing our individual productivity. At the end of the project or competition, we begin to exchange insights about the dataset.

If we exchange information too early in the project, it can harm the quality of our solutions, since team members' thought processes start to be influenced by one another's opinions, preventing new perspectives from emerging.

In this competition, we bet on our diversity and managed to achieve a strong solution in the public phase. We have high confidence in our validation and believe we achieved a solid score in the final phase, successfully avoiding validation overfitting.

In this context, each team member worked on an independent feature generation process and developed independent models so that, in the end, we could aggregate everything into a second-level learning stage within the stacking framework.

3 Feature Engineering

In this competition, our team chose to work independently at first, allowing each member to develop their own ideas and solutions before combining them into a single, more robust result. Based on this approach, we will present each team member's feature creation process in separate sections.

3.1 Humberto's Approach

Considering the need for feature diversity when building a stacking model, Humberto Brandão decided to construct a group of features that perform statistical and related tests.

These features were used in different classifiers within our stacking model, as will be shown in an image later in the report.

A total of 58 features were created in this block and are described below:

Feature Descriptions

f_01 — $p_{\textbf{local\_t\_abs\_50}}$

This feature probes a local change in dispersion or tail behavior by comparing absolute magnitudes around the pre/post boundary. Let $X_{\text{pre},-50} = \{|x_i|\}_{i=n_0-49}^{n_0}$ and $Y_{\text{post},+50} = \{|y_j|\}_{j=1}^{50}$. Structural breaks often manifest as abrupt shifts in stochastic volatility, which are captured by differences in $|z|$. Using Welch's two-sample t-test (robust to unequal variances),

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}, \qquad f_{01} = \Pr(|T_\nu| \geq |t|)\,,$$

where ν is Satterthwaite's degrees of freedom. Small f_{01} indicates a sharp local shift consistent with a structural break.

f_02 — $p_{\textbf{local\_t\_abs\_fisher}}$

To increase sensitivity to multi-scale deviations, Fisher aggregation combines p-values from spans $n \in \{50, 100\}$ computed as in f_01. If p_{50} and p_{100} denote Welch-t p-values on absolute values,

$$T = -2 \sum_{k \in \{50,100\}} \log p_k, \qquad f_{02} = 1 - F_{\chi^2_4}(T).$$

Low f_{02} signifies coordinated evidence of a break across adjacent window sizes, mitigating scale-specific noise.

f_03 — $cp_{\min\_dist\_to\_\tau}$

Let $\widehat{\tau}$ be a data-driven change-point (e.g., argmax of a contrast such as CUSUM/MOSUM) and let $\tau = n_0$ denote the known pre/post split. The normalized alignment

$$f_{03} = \frac{|\widehat{\tau} - \tau|}{n_0 + n_1}$$

quantifies how sharply the estimated break coincides with the boundary. Smaller values indicate boundary-aligned, abrupt changes.

f_04 — $v51\_C\_\textbf{full\_piecewise}\, dR^2_{\textbf{lin}\to\textbf{step}}$

On the cumulative path $C_t = \sum_{s \leq t} z_s$, consider $M_0 : C_t \sim 1 + t$ and $M_1 : C_t \sim 1 + t + \mathbf{1}\{t > \tau\}$. The feature measures improvement in explanatory power due to a step at τ:

$$f_{04} = R^2(M_1) - R^2(M_0).$$

A large f_{04} indicates a level shift in the cumulated process, characteristic of a structural break in drift.

f_05 — $p_{\textbf{Fligner}}$

Variance shifts are canonical indicators of structural change. The Fligner–Killeen test compares dispersions robustly under non-normality. With samples X_{pre} and X_{post},

$$f_{05} = \text{pval}_{\text{Fligner}}(X_{\text{pre}}, X_{\text{post}}).$$

Small values suggest a statistically significant heteroscedastic break.

f_06 — $\textbf{core}_{[25,75]}$ **slope small-k in-band**

Restrict both periods to their time "core" (25–75%) and further to observations satisfying $|z-\mu_0| \leq k\sigma_0$ to reduce outlier influence. Fit linear models to obtain slopes $\widehat{\beta}_1$:

$$f_{06} = |\widehat{\beta}_1^{\text{post}} - \widehat{\beta}_1^{\text{pre}}|.$$

A large value indicates a regime shift in deterministic trend not driven by extremes.

f_07 — $\text{diff\_entropy}_{q40\text{–}60}$

Compute entropy within the central interval $[q_{0.4}, q_{0.6}]$ for each period. Let $H(Z) = -\int p(z) \log p(z)\, dz$. The feature

$$f_{07} = H(Z_{\text{post}} \mid z \in [q_{0.4}, q_{0.6}]) - H(Z_{\text{pre}} \mid z \in [q_{0.4}, q_{0.6}])$$

captures central-shape changes (e.g., curvature around the median) often accompanying structural breaks.

f_08 — $\text{JSD}_{\textbf{posonly}, m=100}$

Restrict to positive values and form $m = 100$-bin histograms for pre/post. With P and Q the discrete densities and $M = (P+Q)/2$, the Jensen–Shannon divergence is

$$\text{JSD}(P,Q) = \tfrac{1}{2} D_{\text{KL}}(P\|M) + \tfrac{1}{2} D_{\text{KL}}(Q\|M), \qquad f_{08} = \text{JSD}(P^+_{\text{pre}}, P^+_{\text{post}}).$$

Higher values reveal asymmetric distributional shifts relevant to breaks.

f_09 — $-\log_{10} p(F\textbf{-variance test})$

Let $F = s^2_{\text{post}}/s^2_{\text{pre}}$ be the variance ratio and p its F-test p-value. Define

$$f_{09} = -\log_{10}(p),$$

a scale that expresses the strength of heteroscedastic evidence. Larger f_{09} points to volatility breaks.

f_10 — $nf_{\textbf{tw\_pre}, k=1}$

Two-tail exceedance around the pre mean/scale summarizes baseline tail heaviness:

$$f_{10} = P\big(|X_{\text{pre}} - \mu_0| > \sigma_0\big).$$

Discrepancies relative to post (captured elsewhere) indicate tail-structure changes associated with breaks.

f_11 — $\text{Hellinger}_{\textbf{posonly}, m=100}$

Hellinger distance on positive support is

$$H(P,Q) = \sqrt{1 - \sum_{b=1}^{m} \sqrt{p_b q_b}}, \qquad f_{11} = H(P^+_{\text{pre}}, P^+_{\text{post}}).$$

It detects broad, bounded differences in shape/scale that can be asymmetric across the boundary.

f_12 — $bds_{\min p}$

The BDS test probes independence/nonlinearity. Breaks often alter dependence (e.g., clustering). Over a small set of embeddings $\mathcal{M}$ and tolerance ϵ,

$$f_{12} = \min_{m \in \mathcal{M}} \text{pval}_{\text{BDS}}(X_{\text{pre}}, m, \epsilon).$$

Small f_{12} indicates strong evidence of dependence changes.

f_13 — $\sup F$ **(step/hinge scan)**

Scan candidate breakpoints τ comparing nested regressions (e.g., linear vs. linear+step or +hinge). The supremum F statistic is

$$f_{13} = \max_{\tau} \frac{(\text{RSS}_0 - \text{RSS}_1(\tau))/d}{\text{RSS}_1(\tau)/(n-p)}, \quad d = 1.$$

Large values suggest a sharp break at some τ.

f_14 — core$_{[45,55]}$ $\min p$ (in-band)

Within the central time decile and small k-bands around μ_0, compute two-sample test p-values across a set $\mathcal{K}$ and take the minimum:

$$f_{14} = \min_{k \in \mathcal{K}} \mathrm{pval}\Big(X^{(k)}_{\mathrm{core}}, Y^{(k)}_{\mathrm{core}} \Big) .$$

A small value flags a localized, robust difference near the boundary.

f_15 — absdiff$_{m=250,\ \textbf{zero},\ k=0.05,\ \textbf{in-band}}$

For equal-size windows of length m, compare in-band occupancy near zero (a proxy for quiescent states) using anchor scale s:

$$f_{15} = \big| P\big(|Y| \le 0.05\, s\big) - P\big(|X| \le 0.05\, s\big) \big| .$$

Large f_{15} marks pronounced shifts in near-zero behavior.

f_16 — $nf_{\textbf{ring\_pre},[1,2]}$

The pre-period ring share between 1 and 2 standard deviations is

$$f_{16} = P\left(1 \le \frac{|X_{\mathrm{pre}} - \mu_0|}{\sigma_0} \le 2 \right) .$$

It anchors moderate-extreme frequency; changes relative to post indicate dispersion-shape breaks.

f_17 — absdiff$_{m=100,\ \textbf{zero},\ k=0.5,\ \textbf{in-band}}$

As in f_15 but with shorter windows and a wider band, improving sensitivity to moderate-scale changes:

$$f_{17} = \big| P\big(|Y| \le 0.5\, s\big) - P\big(|X| \le 0.5\, s\big) \big| .$$

High values flag appreciable shifts in moderate-amplitude occupancy.

f_18 — Hellinger$_{q40–60}$

Compute Hellinger distance restricted to $[q_{0.4}, q_{0.6}]$:

$$f_{18} = H(P_{\mathrm{pre}} \mid z \in [q_{0.4}, q_{0.6}],\ P_{\mathrm{post}} \mid z \in [q_{0.4}, q_{0.6}]) .$$

This central-shape comparison isolates regime changes around the median.

f_19 — mean_absdiff$_{\textbf{zero},\,k=0.1,\textbf{in-band}}$

Mean absolute deviation from zero within a tight band compares local roughness around equilibrium:

$$f_{19} = E[|Y| \mid |Y| \le 0.1s] - E[|X| \mid |X| \le 0.1s]\,.$$

Nonzero values indicate altered local amplitude dynamics.

f_20 — $p_{\textbf{local\_}t\textbf{\_abs\_100}}$

Identical to f_01 but with $n = 100$ on both sides, yielding a more stable local test at the cost of immediacy. With $|X|$ and $|Y|$ computed on 100-point windows,

$$t = \frac{\overline{|X|} - \overline{|Y|}}{\sqrt{\frac{s^2_{|X|}}{100} + \frac{s^2_{|Y|}}{100}}}, \qquad f_{20} = \Pr(|T_\nu| \ge |t|)\,.$$

Small f_{20} supports a sustained local dispersion/shape shift near the boundary.

f_21 — skew_q05–95$^{(p0)}$

This feature measures the change in third standardized moment (skewness) restricted to the robust central band defined from pre-period quantiles. Let $q_{0.05}$ and $q_{0.95}$ be computed on X_{pre}. Define $X^\star = \{x \in X_{\text{pre}} : q_{0.05} \le x \le q_{0.95}\}$ and $Y^\star = \{y \in Y_{\text{post}} : q_{0.05} \le y \le q_{0.95}\}$. The skewness of a sample Z is $\text{skew}(Z) = \frac{1}{s^3}\frac{1}{|Z|}\sum_{z\in Z}(z-\bar{z})^3$ with s^2 the sample variance. The feature is

$$f_{21} = \text{skew}(Y^\star) - \text{skew}(X^\star).$$

Non-zero values indicate asymmetric deformation of the central mass around the break, typical in location/shape shifts.

f_22 — $v51\_G\_\textbf{full\_piecewise}\, dR^2_{\textbf{lin}\to\textbf{step}}$

On the log-cumulative path $G_t = \sum_{s\le t}\log(1+z_s)$, fit $M_0 : G_t \sim 1+t$ and $M_1 : G_t \sim 1+t+\mathbf{1}\{t > \tau\}$, where τ is the pre/post index. With $R^2(M)$ denoting the coefficient of determination,

$$f_{22} = R^2(M_1) - R^2(M_0).$$

A large increment reveals a discrete level shift in the integrated growth, consistent with a structural break in drift or compounding behavior.

f_23 — $p_{\textbf{local\_varF\_50}}$

Variance breaks are hallmarks of regime change. Using the last 50 pre points and first 50 post points, compute the two-sample F-test p-value:

$$F = \frac{s_Y^2}{s_X^2}, \qquad f_{23} = \Pr(F_{\nu_1,\nu_2} \geq F) \text{ (two-sided)}.$$

Small f_{23} indicates a local heteroscedastic jump at the boundary.

f_24 — mean_absdiff_**zero**, $k = 0.03$, **in-band**

Anchoring on the pre scale $s_0 = \mathrm{sd}(X_{\mathrm{pre}})$, restrict to the tight band $\mathcal{B} = \{|z| \leq 0.03\, s_0\}$. Define the conditional means $m_X = E[|X| \mid X \in \mathcal{B}]$ and $m_Y = E[|Y| \mid Y \in \mathcal{B}]$. The feature is

$$f_{24} = m_Y - m_X.$$

Deviations capture subtle changes in near-zero activity (micro-volatility) associated with incipient breaks.

f_25 — $h_{\textbf{pre}}$ **(Shannon entropy)**

Let $\widehat{p}$ be a histogram density of X_{pre} on a data-adaptive grid. The Shannon entropy

$$f_{25} = H(X_{\mathrm{pre}}) = -\sum_b \widehat{p}_b \log \widehat{p}_b$$

summarizes baseline uncertainty/dispersion. Under breaks, downstream features compare this baseline to post; f_{25} serves as a normalized anchor for information-theoretic changes.

f_26 — absdiff_$m = 100$, **zero**, $k = 0.1$, **in-band**

Over matched windows of length $m = 100$, compute the in-band probabilities around zero with pre scale s_0:

$$p_X = P(|X| \leq 0.1\, s_0), \quad p_Y = P(|Y| \leq 0.1\, s_0), \quad f_{26} = |p_Y - p_X|.$$

A spike in f_{26} reflects altered occupancy of the low-amplitude regime across the break.

f_27 — absdiff_CV

Coefficient of variation contrasts scale relative to mean: $\mathrm{CV}(Z) = \frac{\mathrm{sd}(Z)}{|EZ|}$. The feature

$$f_{27} = |\mathrm{CV}(Y_{\mathrm{post}}) - \mathrm{CV}(X_{\mathrm{pre}})|$$

detects proportional volatility shifts, robust to global scaling, often accompanying structural transitions.

f_28 — **core**$_{[45,55]}$ ring absdiff_**all**, μ_0, $k = 1.5$

Within the time "miolo" (indices 45–55 percentiles) and using pre anchors (μ_0, s_0), consider the ring indicator $I\{1.0 < |z - \mu_0|/s_0 \leq 1.5\}$. Let r_X, r_Y be ring shares for pre/post in that core. Define

$$f_{28} = |r_Y - r_X|.$$

It isolates moderate-amplitude structure changes local to the boundary, suppressing extreme-driven effects.

f_29 — $v51_G$_**full_auc_signed** Δ

On G_t, define deviations from its end-to-end line L_t; the signed area proxy is $\mathrm{AUC}_{\pm}(Z) = \frac{1}{n}\sum_t (Z_t - L_t)$. The feature compares pre/post:

$$f_{29} = \mathrm{AUC}_{\pm}(G_{\mathrm{post}}) - \mathrm{AUC}_{\pm}(G_{\mathrm{pre}}).$$

Large magnitude indicates a curvature/level deformation in the cumulative growth path at the break.

f_30 — absdiff_post vs preH1 CV

Let $X_{\mathrm{pre}}^{(H1)}$ denote the highest-variance contiguous half of the pre segment (selected by maximizing rolling variance). With CV as before,

$$f_{30} = \left|\mathrm{CV}(Y_{\mathrm{post}}) - \mathrm{CV}\left(X_{\mathrm{pre}}^{(H1)}\right)\right|.$$

It contrasts post against the most turbulent pre regime, enhancing sensitivity to volatility level-shifts.

f_31 — best_ratio

Over a small collection of robust ratios $\mathcal{R}$ (e.g., σ-ratio, IQR-ratio, trimmed-mean ratio), define

$$f_{31} = \max_{R \in \mathcal{R}} R\big(Y_{\text{post}}, X_{\text{pre}}\big),$$

after absolute value or log-symmetrization as appropriate. A large f_{31} signals the strongest standardized contrast across metrics, highlighting multi-faceted breaks.

f_32 — Fligner p

Identical in spirit to f_05 but used in different pipelines. For completeness, with groups $X_{\text{pre}}, Y_{\text{post}}$,

$$f_{32} = \text{pval}_{\text{Fligner}}(X_{\text{pre}}, Y_{\text{post}}).$$

Its robustness to non-Gaussian tails makes it reliable for variance breaks under heavy-tailed noise.

f_33 — $\mathbf{core}_{[45,55]}$ ring absdiff_**all**, **zero**, $k \in [0.1, 0.25]$

Let $\text{ring}_{[a,b]}(z) = I\{a \le |z|/s_0 \le b\}$ with s_0 from pre. In the central time core, compute

$$r_X = E[\text{ring}_{[0.1,0.25]}(X)], \quad r_Y = E[\text{ring}_{[0.1,0.25]}(Y)], \quad f_{33} = |r_Y - r_X|.$$

This focuses on low-to-moderate amplitudes where subtle regime changes may first appear.

f_34 — $|\rho_s(|\cdot|,\ t)|_{\mathbf{post}} - |\rho_s(|\cdot|,\ t)|_{\mathbf{pre}}$

Let $\rho_s(|Z|, t)$ be the Spearman correlation between absolute values and time. The feature is

$$f_{34} = |\rho_s(|Y|, t)| - |\rho_s(|X|, t)|.$$

It detects monotonic drifts in volatility (volatility trend), a signature of evolving regimes.

f_35 — $\mathbf{core}_{[25,75]}\ |\mathbf{\Delta slope}_{\log m,\ \textbf{in-band}}|$

Within the interquartile time core and restricted to $|z| \le 0.5\, s_0$, regress $\log(|z| + m)$ on time with a small offset $m > 0$ to stabilize near zero. With slopes $\widehat{\beta}_1^{\text{pre}}, \widehat{\beta}_1^{\text{post}}$,

$$f_{35} = |\widehat{\beta}_1^{\text{post}} - \widehat{\beta}_1^{\text{pre}}|.$$

It captures trend changes in local amplitude on a log scale, robust to rare spikes.

f_36 — p_t **Fisher aggregate**

Combine Welch-t p-values across a set of windows $\mathcal{N}$ (e.g., 50, 100, 200):

$$T = -2 \sum_{n \in \mathcal{N}} \log p_n, \qquad f_{36} = 1 - F_{\chi^2_{2|\mathcal{N}|}}(T).$$

Aggregation boosts power against a variety of local alternatives, characteristic of multi-scale breaks.

f_37 — $\rho_s(|X_{\mathbf{pre}}|, t)$

This is the Spearman correlation between pre absolute values and time:

$$f_{37} = \rho_s\big(|X_{\mathrm{pre}}|, t\big).$$

Nonzero f_{37} indicates temporal ordering in volatility pre-break; shifts of this dependence post-break strengthen break evidence.

f_38 — $\min p_t$

Over a predefined family of local Welch-t comparisons (e.g., multiple windows or offsets), define

$$f_{38} = \min_{w \in \mathcal{W}} p_t^{(w)}.$$

The minimum p-value serves as a sensitive detector of at least one local mean/scale change.

f_39 — CUSUMSQ **at** τ

Fit a global linear trend on the combined series, collect residuals e_t, and compute cumulative normalized sum of squares $S_k = \frac{\sum_{t \le k} e_t^2}{\sum_t e_t^2}$. With τ the boundary,

$$f_{39} = \left| S_\tau - \frac{\tau}{n} \right|.$$

A deviation from the baseline line indicates variance redistribution at the split, a break marker.

f_40 — $\log \frac{\mathrm{Var}(|Y|_{q05-95})}{\mathrm{Var}(|X|_{q05-95})}$

Using absolute values restricted to the central quantile band $[q_{0.05}, q_{0.95}]$ (computed on pre), define

$$f_{40} = \log \frac{\mathrm{Var}\big(|Y|\,\mathbf{1}_{[q_{0.05}, q_{0.95}]}\big)}{\mathrm{Var}\big(|X|\,\mathbf{1}_{[q_{0.05}, q_{0.95}]}\big)}.$$

It targets mid-core dispersion changes, robust to heavy tails.

f_41 — JSD_**shape**, $q40$–60

Within $[q_{0.40}, q_{0.60}]$ (pre-defined), compare shape-only by renormalizing densities. With discrete densities P, Q in that slab,

$$f_{41} = \tfrac{1}{2} \sum_b P_b \log \frac{P_b}{M_b} + \tfrac{1}{2} \sum_b Q_b \log \frac{Q_b}{M_b}, \quad M = \tfrac{1}{2}(P + Q).$$

It emphasizes curvature and modality changes near the median.

f_42 — ΔH_**shape**, $q40$–60

Let $H(\cdot)$ be Shannon entropy of the reweighted central slab densities (as in f_41). The feature

$$f_{42} = H(Y \mid q40\text{–}60) - H(X \mid q40\text{–}60)$$

isolates changes in concentration vs. spread of the central mass, a subtle sign of structural adjustment.

f_43 — $q_{0.35}(X_{\mathbf{pre}})$

A robust lower-central quantile from the pre period:

$$f_{43} = q_{0.35}(X_{\mathrm{pre}}).$$

It acts as a baseline structural anchor used by other comparative features; abrupt drift alters this anchor relative to post quantiles.

f_44 — $G\_n = 1000:\ \Delta\mathrm{AUC}_{\pm}$

On the local log-cumulative path windowed to 1000 points around the boundary, compute the signed area contrast as in f_29:

$$f_{44} = \mathrm{AUC}_{\pm}\big(G^{(1000)}_{\mathrm{post}}\big) - \mathrm{AUC}_{\pm}\big(G^{(1000)}_{\mathrm{pre}}\big).$$

It focuses the cumulative-shape shift at an intermediate scale.

f_45 — WD_1 **shape base**

The 1-Wasserstein distance between standardized (shape-only) versions of pre/post captures quantile-wise shifts:

$$f_{45} = \int_0^1 |Q_Y(u) - Q_X(u)|\, du,$$

where Q_Z is the quantile function of the standardized Z. It is sensitive to a broad class of distributional breaks.

f_46 — $p_{\mathbf{acf1}}$ **Fisher**

Compute Welch-type z-tests for differences in lag-1 autocorrelation across a small set of windows, producing p-values $\{p_k\}$. Aggregate via Fisher:

$$T = -2\sum_k \log p_k, \qquad f_{46} = 1 - F_{\chi^2_{2K}}(T).$$

Breaks often reconfigure short-memory structure; f_{46} amplifies evidence across scales.

f_47 — $\mathbf{core}_{[45,55]}\ p_{\mathbf{two\text{-}tails}}(k = 1.5)$

Using pre anchors (μ_0, s_0), evaluate a two-proportion test comparing $P(|Z - \mu_0| > 1.5\, s_0)$ in the time core:

$$f_{47} = \text{pval\_prop}\big(I\{|X - \mu_0| > 1.5 s_0\},\ I\{|Y - \mu_0| > 1.5 s_0\}\big).$$

It targets shifts in extreme-tail occupancy near the boundary.

f_48 — absdiff_$m = 1000$, **zero**, $k = 0.01$, **in-band**

As in f_26 but with $m = 1000$ and a very tight band:

$$f_{48} = \left|P(|Y| \le 0.01\, s_0) - P(|X| \le 0.01\, s_0)\right|.$$

It probes ultra-quiet regime occupancy, sensitive to microstructure changes.

f_49 — $\log \frac{\mathrm{CV}(Y)}{\mathrm{CV}(X)}$

A log-symmetrized CV ratio,

$$f_{49} = \log \frac{\mathrm{sd}(Y)/|EY|}{\mathrm{sd}(X)/|EX|}.$$

It provides a scale-free contrast of relative volatility, with symmetry that stabilizes estimation.

f_50 — sd_**preH1**

Select the highest-variance contiguous half of the pre series, $X^{(H1)}_{\mathrm{pre}}$, and compute

$$f_{50} = \mathrm{sd}\Big(X^{(H1)}_{\mathrm{pre}}\Big).$$

It characterizes the volatility ceiling present before the break, used by comparative features to gauge post-break escalation or dissipation.

f_51 — run_**pre, in-band**$(k = 0.1)$ **mean length**

Define in-band flags $b_t = I\{|x_t - \mu_0| \le 0.1\, s_0\}$ over pre and compute run-lengths of consecutive ones. The feature is

$$f_{51} = \text{mean run length}\big(\{b_t\}_{t\in\text{pre}}\big).$$

Breaks shift persistence of quiet periods; f_{51} captures pre-baseline persistence.

f_52 — $p_{\textbf{F-var}}$ **Fisher**

Aggregate F-test p-values for variance differences across a set of local windows $\{n\}$:

$$T = -2\sum_{n} \log p_F^{(n)}, \qquad f_{52} = 1 - F_{\chi^2_{2|\{n\}|}}(T).$$

It increases robustness to heteroscedastic breaks manifesting at different scales.

f_53 — $\min p_{\textbf{Wilcoxon}}$

Over a small family of local comparisons (windows/offsets), compute Wilcoxon rank-sum p-values $\{p_W^{(w)}\}$ and take

$$f_{53} = \min_{w} p_W^{(w)}.$$

It is sensitive to broad distributional shifts (median/shape) insensitive to outliers, as expected under regime changes.

f_54 — kurt_$q05$–$95^{(p1)}$

Using the post-period sample restricted to $[q_{0.05}, q_{0.95}]$ (pre-defined), compute excess kurtosis

$$\text{kurt}(Z) = \frac{E\big[(Z-\bar{Z})^4\big]}{E\big[(Z-\bar{Z})^2\big]^2} - 3, \qquad f_{54} = \text{kurt}(Y^\star)\,.$$

Elevated f_{54} suggests heavier central tails post-break relative to pre.

f_55 — $p_{\Delta\text{acf1}}$

Test equality of lag-1 autocorrelation via Fisher z-transform: $z = \tanh^{-1}(\widehat{\rho})$ with variance $(n-3)^{-1}$. With estimates $\widehat{\rho}_X, \widehat{\rho}_Y$,

$$Z = \frac{z_Y - z_X}{\sqrt{\frac{1}{n_Y-3} + \frac{1}{n_X-3}}}, \qquad f_{55} = 2\big(1 - \Phi(|Z|)\big).$$

Dependence reconfiguration across a break yields small f_{55}.

f_56 — JSD_**shape**, $q30$–70

Analogous to f_41 but on the wider slab $[q_{0.30}, q_{0.70}]$:

$$f_{56} = \mathrm{JSD}\big(P_{\text{pre}}^{(0.30,0.70)}, P_{\text{post}}^{(0.30,0.70)}\big).$$

The wider core increases power against medium-scale deformation of the central body.

f_57 — mean_absdiff_μ_0, $k = 0.1$, **in-band**

Condition on the pre-anchored band $\{|z - \mu_0| \le 0.1\, s_0\}$. Let $m_X = E[|X - \mu_0| \,|\, \cdot]$ and m_Y analogously. Define

$$f_{57} = m_Y - m_X.$$

It quantifies central roughness changes around the pre mean, a fine-grained indicator of regime shift.

f_58 — $\Delta\text{ES}_{0.95}$

Expected Shortfall at level 0.95 (left or right, depending on sign convention) for a sample Z is $\text{ES}_{0.95}(Z) = E[Z \mid Z \ge q_{0.95}(Z)]$. The feature compares post vs. pre:

$$f_{58} = \text{ES}_{0.95}(Y_{\text{post}}) - \text{ES}_{0.95}(X_{\text{pre}}).$$

Structural breaks frequently alter extreme risk; f_{58} highlights tail amplification/attenuation beyond the Value-at-Risk threshold.

3.2 Mario's approach

3.2.1 Overview

The feature extraction framework follows a two-stage hierarchical design aimed at capturing both *descriptive* and *cross-regime structural* information from time-series data.

In the first stage, the pipeline computes a comprehensive set of statistical, financial, and distributional descriptors for each segment of the series ("pre" and "post" regimes). In the second stage $(f_0, f_1, \ldots, f_{286})$, a set of *composite interaction features* is generated through a semi-automated iterative search process, in which randomly combined first-stage features are evaluated based on improvements in validation performance.

3.2.2 Stage 1 – Foundational Descriptors

The first stage constructs a rich representation of each regime by combining descriptive, autocorrelation, financial, and divergence-based metrics.

a. Descriptive Statistics. For each slice of the series, the following descriptive statistics are computed: mean, standard deviation, maximum, minimum, skewness, kurtosis and key quantile values (0.10, 0.25, 0.50, 0.75, 0.90).

b. Autocorrelation Structure. Temporal dependencies are captured through autocorrelation values at multiple lags:

$$\mathrm{ACF}(k) = \frac{\sum_t (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_t (x_t - \bar{x})^2}, \quad k \in \{1, 2, 5, 10, 20\}.$$

c. Financial and Risk Metrics. Micro-financial statistics are computed for each regime to capture risk and performance characteristics:

$$\text{Sharpe} = \frac{\mu}{\sigma}, \qquad \text{Sortino} = \frac{\mu}{\sigma_-}, \qquad \text{Profit Factor} = \frac{\sum x_+}{|\sum x_-|}.$$

In addition, the framework extracts Conditional Value-at-Risk (CVaR_α for $\alpha \in \{1\%, 2.5\%, 5\%, 10\%\}$), regression coefficients (R^2, slope, and intercept), the Gini ratio, and maximum drawdown.

d. Cross-Regime Fusions. To quantify structural changes between the pre and post regimes, each metric m is combined through a set of relational operators:

$$m_{\text{diff}} = m_{\text{post}} - m_{\text{pre}}, \qquad m_{\text{adiff}} = |m_{\text{post}} - m_{\text{pre}}|, \qquad m_{\text{prod}} = m_{\text{post}} \cdot m_{\text{pre}}, \qquad m_{\text{div}} = \frac{m_{\text{post}}}{m_{\text{pre}}}.$$

These fusions capture both the magnitude and direction of regime shifts.

e. Distance and Statistical Divergence Tests. Distributional differences between regimes are further quantified through:

- Kolmogorov–Smirnov test statistics ($D_{\mathrm{KS}}, p_{\mathrm{KS}}$, location, and sign);

- Wasserstein distance $W(p, q)$;
- Euclidean distance between clipped empirical distributions.

Together, these measures summarize both the shape and location shifts occurring across regime transitions.

3.2.3 Stage 2 – Composite and Interaction Features (f_i)

The second stage introduces 287 non-linear composite features $f_0, f_1, \ldots, f_{286}$, derived through an iterative validation-guided search process. In each iteration:

1. Two to four first-stage variables are selected at random.
2. Random arithmetic and non-linear transformations (e.g., addition, division, trigonometric, or logarithmic functions) are applied.
3. The resulting feature is retained only if it yields an improvement in validation performance.

A general form of the generated features can be expressed as:

$$f_i = \phi(a_1 \cdot g_1(x) \ \oplus \ a_2 \cdot g_2(y) + c)\,,$$

where $\phi \in \{\sin, \cos, \tanh, \log(1+x), \sqrt{\cdot}\}$, g_j are base feature transformations, $\oplus$ denotes a random operator $(+, -, \times, /)$, and a_j, c are empirical constants.

The composite features can be categorized into several broad groups:

- **Ratio and product features:** multiplicative or relative relationships between descriptors.
- **Non-linear transformations:** use of $\sin, \cos, \tanh, \log, \exp, \sqrt{\cdot}$, sign to introduce non-linear response behavior.
- **Cross-moment interactions:** relationships between higher-order statistical moments (e.g., skewness, kurtosis, Sharpe ratio).
- **Shift-sensitive interactions:** non-linear combinations of quantiles, autocorrelations, and divergence metrics that highlight regime transitions.

3.2.4 Transformations and Slicing Strategy

Each time series undergoes multiple transformations to emphasize different statistical aspects:

$$T(x) \in \{x, |x|, \log(1+x), \mathrm{cumsum}(x), \mathrm{cumprod}(1+x)\}.$$

These transformations serve to normalize scale, capture cumulative dynamics, and highlight multiplicative effects.

To focus on regime transition dynamics, the framework also considers partial slicing of pre and post segments:

$$(\text{pre\_tail}, \text{post\_head}) \in \{(\text{None}, \text{None}), (0.1, 0.1)\},$$

allowing both full-regime and boundary-sensitive features to be included.

3.2.5 Parallel Processing and Robustness

All computations are parallelized across time-series identifiers using a thread-based job scheduler. NaN-safe numerical operations are employed throughout the process to ensure stability. The resulting feature matrix is clipped to the range $[-10^{10}, 10^{10}]$ and sanitized to remove non-finite values, ensuring consistency across datasets and hardware configurations.

3.2.6 Summary

In summary, the feature-engineering pipeline integrates:

- Descriptive, temporal, and financial metrics computed separately for pre and post segments;
- Cross-regime fusion operators that quantify structural and statistical changes;
- Iteratively evolved non-linear composite features (f_i) that enhance model expressiveness and improve predictive robustness.

The resulting feature set captures both low-level statistical properties and high-level transitional dynamics, providing a comprehensive representation for the detection and modeling of regime shifts and structural breaks in time-series data.

3.3 Rafael's Approach

3.3.1 Overview

Rafael built a comprehensive, regime-aware feature pipeline aimed at capturing distributional shifts between `pre` (`period`=0) and `pos` (`period`=1) segments of each series. The process first creates multiple signal representations—raw, IQR-filtered (with absolute and sign-flipped variants), wavelet-denoised (Symlet-2, level 1), cumulative sums, and percentage changes. On these views, we compute global and windowed summary statistics (mean/median/extrema, variance/standard deviation, percentiles, and mode), then quantify `pre`–`pos` differences via a battery of distances (e.g., Canberra, Bray–Curtis, Jensen–Shannon) and two-sample KS tests across full and windowed cuts. Features are sanitized (coercion, $\pm\infty \to$ NaN, imputation, clipping) and expanded into hundreds

of composite interactions (arithmetic, powers, tanh, modulo, etc.) to form model-ready bases for both the main and alternate classifiers. This design emphasizes robustness to outliers/noise and sensitivity to regime changes while remaining lightweight enough for large-scale screening.

3.3.2 Data Slicing

For each series i, split the indexed observations by period into $\texttt{pre}_i = \{x_t : \texttt{period} = 0\}$ and $\texttt{pos}_i = \{x_t : \texttt{period} = 1\}$. Also form the concatenated series $\texttt{serie}_i = [\texttt{pre}_i, \texttt{pos}_i]$.

3.3.3 Transforms

- **Outlier filtering (IQR)**: For a vector v, remove values outside $[Q_1 - \lambda \cdot \text{IQR},\ Q_3 + \lambda \cdot \text{IQR}]$. We use $\lambda \in \{1.0,\ 1.75,\ 2.0\}$ on `pre`, `pos`, and `serie`, plus absolute value ($|\cdot|$) variants and a sign-flipped variant ($\times -1$).
- **Wavelet denoising**: Symlet-2 (`sym2`), level 1. Universal threshold $\hat{\sigma}\sqrt{2\log n}$ with $\hat{\sigma} = \text{median}(|\text{detail}|)/0.6745$, soft-thresholding.
- **Cumulative & relative changes**: cumulative sum (cumsum) and percentage change (pct_change) for `pre`, `pos`, and `serie`.

3.3.4 Summary Statistics (Feature Families)

We compute statistics on raw, transformed, pct-change, cumsum, and wavelet versions:

- **Global stats on `serie`**: mean, median, max, min, std, var, mode; percentiles $p \in \{10, 20, \ldots, 90\}$.
- **Same on `serie_transform`** (IQR+abs) and on `serie_pct`, `serie_cumsum`.
- **Separate stats for `pre` and `pos`** (raw, transform(1.75/1.0/2.0), wavelet).
- **Windowed stats**: for window sizes $i \in \{20, 40, 60, 80, 100, 120, 140, 160, 200, 220, 240\}$, compute mean/median/max/min/std/var over tail/head slices: $\texttt{pre}[-i:]$, $\texttt{pos}[:i]$, plus swapped-direction variants (suffix `2`).
- **Percentiles**: $p \in \{10, \ldots, 90\}$ for `pre`, `pos`, and all transform variants.

3.3.5 Distributional Distances (`pre` vs `pos`)

For cuts $c \in \{30, 120, 150, 170, 200\}$, build sorted tails/heads $\texttt{v\_pre} = \text{sort}(\texttt{pre\_transform}[-c:])$, $\texttt{v\_pos} = \text{sort}(\texttt{pos\_transform}[:c])$ and analogs for pct-change. Compute: Canberra, Bray–Curtis, Hamming, Jensen–Shannon, Rogers–Tanimoto, Dice, Yule, Cityblock.

3.3.6 Regime-Shift Tests

Two-sample Kolmogorov–Smirnov (KS) D and p:

- `pre` vs `pos` on full series and on transforms.
- Windowed comparisons: (`pre`$[-40:]$, `pos`$[:40]$), (`pre`$[-80:]$, `pos`$[:80]$), (`pre`$[-120:]$, `pos`$[:120]$), with raw and transformed variants.
- Within-`pre` splits: first half vs second half; first third vs last two thirds (raw and transformed).

3.3.7 Sanitization

All features are coerced to numeric; $\pm\infty \to$ `NaN`; `NaN`$\to -999{,}999$; clipping to $[-10^6, 10^6]$.

3.4 Model-Ready Composites

- **Base set**: 477 engineered interactions (`f1`–`f477`) combining the primitives above via arithmetic operators, tanh, $\sqrt{\cdot}$, powers, integer division/modulo, and sign/abs; includes the distance and KS features.
- **Alternate model set**: 22 additional composites (`f1_alt_model`–`f22_alt_model`).

3.4.1 Implementation Notes

- *Outlier filters*: $\lambda \in \{1.0, 1.75, 2.0\}$; absolute-value and sign-flip versions.
- *Wavelets*: `pywt.wavedec/waverec` with `sym2`, level 1, soft threshold.
- *Modes*: computed via `scipy.stats.mode`.
- *KS*: `scipy.stats.ks_2samp`.
- *Distances*: `scipy.spatial.distance`.

3.5 João's Approach

João developed features that combined the predictions of the first-level models with statistical descriptors extracted from the time series. This methodology allowed up to six features to be merged into a single one, enhancing predictive power—including that of the first-level models' outputs. There are many possible ways to combine these features; the most common were:

- **Subtraction:** $F_1 - F_2 - F_3$
- **Multiplication:** $F_1 \times F_2 \times F_3$

- **Division:** $\frac{F_1}{F_2 \times F_3}$
- **Average (Mean):** $\frac{F_1+F_2+\cdots+F_N}{N}$
- **Geometric Mean:** $\sqrt[N]{F_1 \times F_2 \times \cdots \times F_N}$
- **Harmonic Mean:** $\frac{N}{\frac{1}{F_1}+\frac{1}{F_2}+\cdots+\frac{1}{F_N}}$

4 Models

After completing the feature engineering phase, we proceeded to develop the first-level models. A total of eight models were constructed at this stage:

- **R-01:** XGBoost Classifier using Rafael's group-one features;
- **B-01:** XGBoost Classifier using Rafael's group-one features with five-seed bagging;
- **R-02:** XGBoost Regressor using Rafael's group-two features;
- **R-03:** XGBoost Regressor using all of Rafael's features;
- **H-01:** XGBoost Classifier using Humberto's features;
- **M-01:** XGBoost Classifier using Mario's features;
- **RF-01:** Random Forest Classifier using Mario's features combined with Rafael's group-two features;
- **RHM-01:** XGBoost Classifier using the combined features from Rafael, Mario, and Humberto.

As shown, all of our models are tree-based, which proved to be the most effective approach for capturing and combining the patterns produced during our feature engineering process.

5 Stacking

As mentioned earlier, we developed eight first-level models. The predictions from these models were then used as inputs for a second-level model and were also incorporated into João's feature-generation process. The final model ultimately consisted of approximately 100 features. Our stacking architecture is shown in Figure 1.

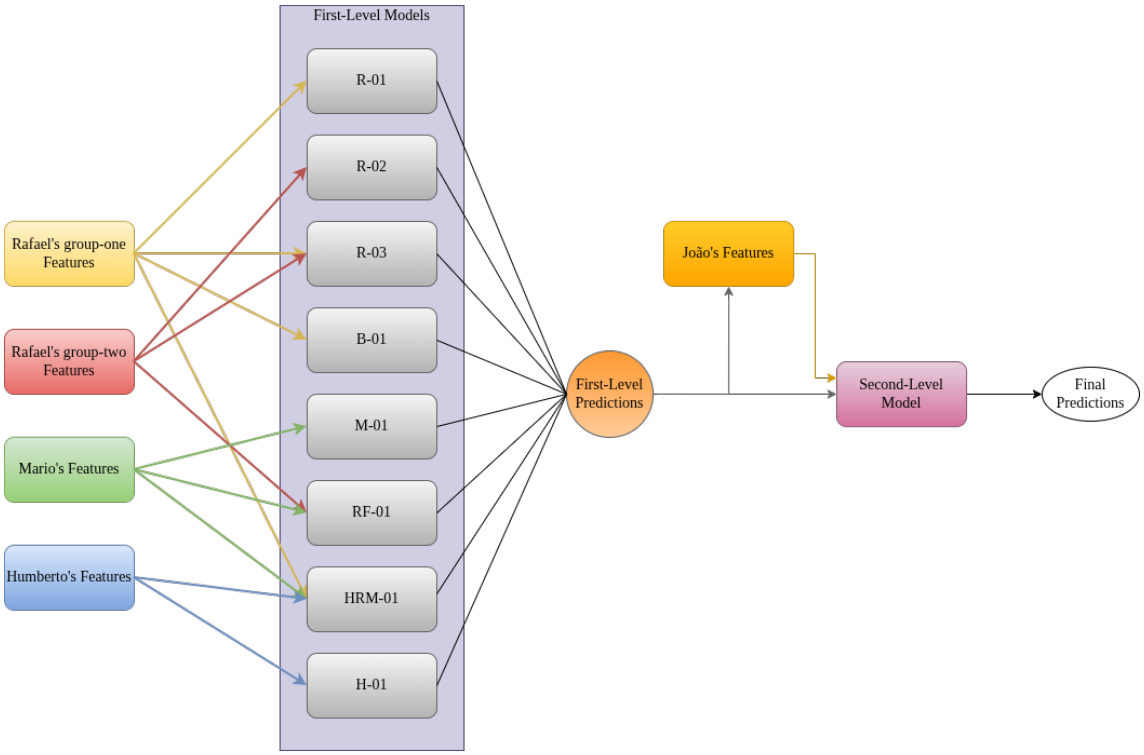

Figure 1: Stacking architecture used in the final solution.

6 Conclusion

The solution developed by the Alphabot team reflects a solid and well-structured approach to the challenge of structural break detection. The use of stacking enabled the exploration of diverse ideas and individual methodologies from each team member, resulting in a set of complementary and highly robust models.

The focus on maintaining independence between feature generation and model-building processes ensured not only the diversity required for the second-level learning stage but also the absence of data leakage between layers. This strategy proved effective in both the public and final phases of the competition, reinforcing the reliability of the validation employed.

Another noteworthy aspect was the wide spectrum of features developed, ranging from classical statistical tests to divergence, entropy, and probabilistic distance metrics. This variety allowed the models to capture different aspects of the pre- and post-break regimes, providing a comprehensive and multi-scale view of the problem.

Overall, the results demonstrate the strength of a collaborative process grounded in quantitative research, feature engineering, and rigorous validation. The experience reinforces our belief that diversity, both in ideas and in technical approaches, is one of the key factors driving success in competitions and real-world machine learning applications involving financial time series.