CSV structure

id, label, x0, x1, ..., x999, time_*, freq_*, acf_*, ar_*

- **id** – unique segment identifier
- **label** – anomaly label (`0` = non-anomalous, `1` = anomalous)
- **x0 ... x999** – raw time-series samples (e.g. 10 s @ 100 Hz)
- **time_*** – time-domain features (mean, std, skewness, kurtosis, etc.)
- **freq_*** – frequency-domain features (centroid, bandwidth, entropy, etc.)
- **acf_*** – autocorrelation-based features (lag, decay, integral time, etc.)
- **ar_*** – autoregressive (AR(1)) features (coefficient, residual variance, R^2)

Time domain

Name	Formula	Description
time_mean	$\mu = \frac{1}{N} \sum x_i$	Mean value of the signal
time_median	$\tilde{x} = \mathrm{median}(x)$	Middle value after sorting
time_min / time_max	$\min(x), \max(x)$	Minimum / maximum value
time_p5 / p25 / p75 / p95	$Q_p = \mathrm{percentile}(x, p)$	Percentile values
time_ptp	$\max(x) - \min(x)$	Peak-to-peak amplitude range
time_std / time_var	$\sigma = \sqrt{\frac{1}{N} \sum (x_i - \mu)^2}, \sigma^2$	Standard deviation and variance
time_rms	$\sqrt{\frac{1}{N} \sum x_i^2}$	Root-mean-square (RMS) value

Name	Formula	Description
time_iqr	$Q_{0.75} - Q_{0.25}$	Interquartile range
time_range_ratio	$\frac{\max(x) - \min(x)}{\mu + \varepsilon}$	Range normalized by mean
time_skew	$\frac{\frac{1}{N}\sum(x_i - \mu)^3}{\sigma^3}$	Skewness (asymmetry)
time_kurtosis	$\frac{\frac{1}{N}\sum(x_i - \mu)^4}{\sigma^4}$	Kurtosis (peakedness / flatness)

Name	Formula	Description
acf_lag_dom	$\tau_{dom} = \frac{k_{max}}{f_s}$	Dominant lag (in seconds); time delay at which autocorrelation (excluding lag 0) reaches its maximum.
acf_val_dom	$r(\tau_{dom}) = \frac{1}{N}\sum(x_t - \mu)(x_{t-k} - \mu)/\sigma^2$	Autocorrelation value at the dominant lag.
acf_time_decay	$t_{decay} = \frac{\min\{k : r_k < e^{-1}\}}{f_s}$	Decay time; first lag where autocorrelation falls below e^{-1}, indicating how quickly the signal loses its temporal correlation.
acf_integral_time	$\tau_{int} = 1 + 2\sum_{k=1}^{K} r_k$	Integral time; total area under the autocorrelation curve up to the first zero crossing.
acf_mean_abs	$mean\_abs = \frac{1}{K}\sum_{k=1}^{K} abs(r_k)$	Mean absolute autocorrelation; measures overall periodicity and temporal structure of the signal.

Name	Formula	Description
ar1_coef	$x_t = a_1 x_{t-1} + \epsilon_t,\ \text{feature} = a_1$	AR(1) coefficient — quantifies linear dependency between consecutive samples.
ar_resid_var	$Var(\epsilon_t) = \frac{1}{N} \sum \epsilon_t^2$	Variance of AR(1) residuals — represents the unpredictable or noise component of the signal.
ar_r2	$R^2 = 1 - \frac{Var(\epsilon_t)}{Var(x_t)}$	Coefficient of determination — proportion of signal variance explained by the AR(1) model.

Frequency domain

- $P(f) = \frac{|S(f)|}{\sum |S(f)|}$ — normalized power spectrum after RFFT
- $\varepsilon = 10^{-8}$ — small constant to prevent division by zero
- All features are computed on ≈ 10 s segments (~1000 samples @ 100 Hz)

$$S(f_k) = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi f_k n/N}$$

Symbol	Name	Meaning
S(f)	Complex spectrum	Representation of frequencies in the signal
|S(f)|	Amplitude spectrum	Magnitude of frequency components (without phase)
$\|S(f)\|^2$	Power spectrum	Energy of each frequency component
P(f)	Normalized spectrum	Probability-like distribution of spectral energy

Name	Formula	Description
freq_centroid	$f_c = \sum f \cdot P(f)$	Spectral centroid (energy center of mass)
freq_bandwidth	$\sqrt{\sum (f - f_c)^2 P(f)}$	Effective spectral bandwidth
freq_peak_freq	$\arg\max P(f)$	Frequency of the strongest amplitude peak
freq_entropy	$-\sum P(f) \log(P(f) + \varepsilon)$	Spectral entropy (energy dispersion)

Name	Formula	Description
freq_median_freq	$\sum_{f \leq f_{0.5}} P(f) = 0.5$	Median frequency (50 % of total power)
freq_rolloff_95	$\sum_{f \leq f_{0.95}} P(f) = 0.95$	Frequency containing 95 % of total energy
freq_top1_amp	$\max P(f)$	Amplitude of the strongest spectral peak
freq_top1_freq	$\arg\max_f P(f)$	Frequency location of that spectral peak