

# Rating Rules (Aktuální řešení)

Tento dokument popisuje aktuálně používaný výpočet leaderboardu v:

- `scripts/calculate_rating.py` (načtení dat, identita týmu, deduplikace)
- `scripts/calculate_rating_live_final.py` (finální veřejný rating)

Platí stav po zavedení:

- live okno **2,5 roku** (`LIVE_WINDOW_DAYS = 913`)
- kvalifikační úprava skóre **Podium boost**

## 1. Vstupní data a identita týmu

---

### 1.1 Zdroje

Do výpočtu vstupují normalizovaná CSV z `data/**/_results.csv` jen pro soutěže registrované v `COMPETITIONS`.

### 1.2 Týmová identita

Tým je interně identifikován jako:

- `team_id = normalize_handler(handler) + "|||" + normalized_dog_id`

Kde `normalized_dog_id` je preferenčně call name, jinak registered name.

### 1.3 Normalizace jmen

Používají se pravidla z `scripts/calculate_rating.py`:

- odstranění diakritiky pro interní matching
- sjednocení Last, First vs First Last
- aliasy handlerů (např. Katka Terčová -> Kateřina Terčová interně)
- aliasy psích jmen a registered name
- mapování registered name -> call name (např. A3Ch Libby Granting Pleasure -> Pšeník)
- fuzzy merge variant stejného psa u stejného handlera

Výsledek: jeden pes/psovod se má v ratingu chovat jako jedna entita i při různých zápisech.

## 2. Výběr běhů do live výpočtu

---

### 2.1 Časové okno

- `latest_date` = nejnovější datum soutěže v datasetu
- `cutoff_date` = `latest_date` - 913 dní
- do live výpočtu jdou jen běhy s `comp_date`  $\geq$  `cutoff_date`

### 2.2 Minimální velikost pole

Kolo se započítá jen pokud má po deduplikaci minimálně:

- `MIN_FIELD_SIZE` = 6 týmů

### 2.3 Deduplikace v kole

V rámci jednoho kola (`round_key`) se každý `team_id` bere maximálně jednou.

### 2.4 Zpracování eliminací

- neeliminovaní s validním rank jsou seřazeni dle pořadí
- eliminovaní sdílí společné poslední pořadí (`last_rank` = `len(clean)+1`)

## 3. Core rating model (OpenSkill)

---

### 3.1 Model

- PlackettLuce (OpenSkill)

### 3.2 Váha velkých akcí

- pokud je soutěž `tier == 1`, použije se váha 1.2
- jinak váha 1.0

### 3.3 Stabilizace nejistoty

Po každém update:

- `sigma` = `max(SIGMA_MIN, sigma * LIVE_SIGMA_DECAY)`
- `LIVE_SIGMA_DECAY` = 0.99
- `SIGMA_MIN` = 1.5 (z `calculate_rating.py`)

## 4. Veřejné skóre (Rating)

---

### 4.1 Base rating

Nejdřív se spočítá základ:

- $\text{rating\_base} = \text{DISPLAY\_BASE} + \text{DISPLAY\_SCALE} * (\mu - \text{RATING\_SIGMA\_MULTIPLIER} * \sigma)$
- aktuálně:  $\text{DISPLAY\_BASE} = 1000$ ,  $\text{DISPLAY\_SCALE} = 40$ ,  $\text{RATING\_SIGMA\_MULTIPLIER} = 1.0$

### 4.2 Podium boost (kvalitativní korekce)

Na `rating_base` se aplikuje faktor podle kvality výsledků:

- $\text{quality\_score} = 0.45 * \text{clean\_pct} + 0.35 * \text{top10\_pct} + 0.20 * \min(100, \text{top3\_pct} * 3)$
- $\text{quality\_norm} = \text{clamp}(\text{quality\_score} / 70.0, 0, 1)$
- $\text{quality\_factor} = 0.82 + 0.18 * \text{quality\_norm}$
- $\text{rating} = \text{rating\_base} * \text{quality\_factor}$

Interpretace:

- týmy s lepší kombinací čistoty + TOP10 + TOP3 mají faktor blíž 1.00
- týmy s horší kvalitou mají faktor blíž 0.82
- tím se omezuje efekt „jen hodně běhů“ bez dobrých umístění

## 5. Statistika týmů

---

Pro každý tým se v live okně sleduje:

- `num_runs`
- `clean_runs`, `clean_pct`
- `top3_runs`, `top3_pct`
- `top10_runs`, `top10_pct`
- `last_comp` (poslední soutěž v okně)

## 6. Tier labely a PROV

### 6.1 Tier labely (po velikostech)

Po výpočtu rating se v každé size kategorii spočítají percentily:

- Elite: top 2 %
- Champion: top 10 %
- Expert: top 30 %
- jinak Competitor

Výpočet používá jen týmy s `num_runs >= MIN_RUNS_FOR_LIVE_RANKING`.

### 6.2 Provisional

- PROV pokud `sigma >= LIVE_PROVISIONAL_SIGMA_THRESHOLD`
- aktuálně `LIVE_PROVISIONAL_SIGMA_THRESHOLD = 7.8`

## 7. Výstup a řazení

Generují se:

- `output/ratings_live_final.csv`
- `output/ratings_live_final.html`

Řazení v leaderboardu je v rámci každé size:

1. podle rating sestupně
2. zobrazují se jen týmy s `num_runs >= 5`

## 8. Aktuální konfigurační přehled

Parametr	Hodnota
LIVE_WINDOW_DAYS	913
MIN_RUNS_FOR_LIVE_RANKING	5
MIN_FIELD_SIZE	6
ENABLE_MAJOR_EVENT_WEIGHTING	True

Parametr	Hodnota
MAJOR_EVENT_WEIGHT	1.20
LIVE_SIGMA_DECAY	0.99
SIGMA_MIN	1.5
RATING_SIGMA_MULTIPLIER	1.0
ENABLE_PODIUM_BOOST	True
PODIUM_BOOST_BASE	0.82
PODIUM_BOOST_RANGE	0.18
PODIUM_BOOST_TARGET	70.0
LIVE_PROVISIONAL_SIGMA_THRESHOLD	7.8
ELITE_TOP_PERCENT	0.02
CHAMPION_TOP_PERCENT	0.10
EXPERT_TOP_PERCENT	0.30