

Membangun Kurva Tolok Ukur Lichess Stratifikasi untuk Elo+Chess

Catatan Penelitian Elo+Chess

31 Mei 2026

Abstract

Laporan benchmark Elo+Chess dibuat dari sampel bertingkat yang besar Game Lichess.org, dengan sampel benchmark yang digunakan dalam catatan ini berisi 483.974 game peluru 1+0, 479.974 game blitz 3+0, 925.636 game cepat 10+0, dan 660.540 permainan cepat yang lebih panjang. Untuk setiap jenis permainan utama, alurnya diperluas setiap permainan menjadi status papan sebelum langkah dan setelah langkah, menghitung ratusan variabel perantara dan permainan pemain, menggabungkan variabel-variabel tersebut menjadi melatih metrik, lalu membuat rata-rata metrik tersebut berdasarkan bucket Lichess Elo. Itu sampel sengaja dibuat bertingkat, bukan proporsional dengan pemain mentahnya pool: tujuannya adalah untuk memperkirakan kurva perilaku yang stabil di setiap tingkat peringkat, termasuk ember yang seharusnya jarang digunakan, bukan untuk meniru alam distribusi peringkat. Hasilnya adalah sekumpulan kurva patokan empiris: for setiap metrik, kita dapat menanyakan seberapa sering pemain di level Elo yang berbeda menunjukkan yang diberikan kebiasaan atau prinsip dalam permainan nyata. Catatan ini menjelaskan metode, skala perhitungan, dan beberapa contoh prinsip pembukaan yang melibatkan rokade, pengembangan bagian kecil, gerakan bidak pembuka yang berulang, dan gerakan ratu awal. Penekanannya adalah metodologis: Elo+Chess tidak berasumsi bahwa prinsip-prinsip catur bersifat universal linier di semua tingkat keterampilan. Sebaliknya, kurva acuan diperiksa oleh rentang peringkat, dan produk berfokus pada wilayah pemula hingga mahir awal di mana hubungannya paling kuat dan paling dapat ditafsirkan.

1 Tujuan

Tujuan dari sistem benchmark adalah untuk mengubah saran catur yang luas menjadi jumlah yang dapat diukur. Pernyataan seperti “kembangkan bagian kecil Anda” adalah berguna, tetapi menjadi lebih dapat ditindaklanjuti ketika kita dapat mengetahui berapa banyak bagian kecil a Pemain biasanya berkembang dengan gerakan 10, bagaimana jika dibandingkan dengan rekan-rekannya di gerakan yang sama kelompok pemeringkatan, dan bagaimana kebiasaan tersebut berubah di seluruh spektrum pemeringkatan.

Elo+Chess therefore builds a metric library from actual games. Each metric is computed for every player-game row, summarized by game type and Elo bucket, and used as a benchmark curve in the report. The report can then place a user on the same curve and show whether the user is above, below, or near the behavior observed among comparable players.

2 Sumber Data dan Stratifikasi

Sumber benchmark adalah database game Lichess.org publik. Elo+Chess menggunakan sampel bertingkat besar berdasarkan jenis permainan dan kategori peringkat, bukan berdasarkan sam-

pel kecil sampel kenyamanan. Artefak runtime produksi saat ini mencakup empat artefak utama kategori permainan:

- 1+0 peluru,
- serangan kilat 3+0,
- 10+0 cepat,
- permainan cepat lebih dari 10 menit.

Sasaran stratifikasi adalah luasnya tingkat pemeringkatan. Acak proporsional sampel akan efisien untuk menggambarkan keseluruhan kumpulan pemain, tetapi itu akan efisien menjadi tidak efisien untuk membangun kurva patokan. Kelompok pemeringkatan yang paling umum akan mendominasi sampel, sedangkan tail bucket dengan dukungan rendah akan memiliki terlalu sedikit permainan untuk rata-rata metrik yang stabil. Oleh karena itu Elo+Chess mengambil sampel berdasarkan permainan rata-rata Ember Elo. Dalam konstruksi sampel bersih saat ini, game mentah adalah yang pertama difilter ke game berperingkat valid dengan pemain yang dikenal, peringkat yang diketahui, hasil yang valid, dan kontrol waktu yang diketahui. Game kemudian disaring ke pemain dengan game yang cukup dalam kecepatan tersebut dan peringkat yang cukup stabil untuk mewakili tingkat keterampilan yang koheren: untuk setiap pemain, Elo+Chess menghitung rentang peringkat kecepatan dalam pemain ($\max(\text{Elo}) - \min(\text{Elo})$) di seluruh kumpulan sumber dan menyimpannya permainan hanya jika kedua pemain memiliki setidaknya jumlah permainan minimum yang dikonfigurasi dan tidak lebih dari 300 titik Elo penyimpangan dalam kecepatan. Filter penyimpangan tidak dimaksudkan untuk menghilangkan peningkatan atau varians normal; itu mengurangi kasus di mana pemain permainan akan menggabungkan tingkat keterampilan yang berbeda secara material, seperti akun sementara, pemain yang kembali, atau akun yang berpindah dengan cepat melalui beberapa keranjang selama periode sampel. Terakhir, game yang memenuhi syarat diberi peringkat dalam setiap rata-rata Elo bucket dengan hash deterministik dari id dan seed game, dan jumlah yang dibatasi permainan dipertahankan dari setiap keranjang.

Desain ini membuat kurva benchmark jauh lebih berguna. Setiap ember Elo menyumbangkan pengamatan yang cukup untuk memperkirakan nilai rata-rata setiap metrik, dan bucket langka dengan rating tinggi atau rendah tidak tenggelam oleh bentuk aslinya dari distribusi pemain aktif. Kurva benchmark yang digunakan oleh Elo+Chess adalah kemudian dibatasi pada rentang yang paling relevan dengan produk, kira-kira pemula melalui permainan lanjutan awal. Dalam laporan skala Lichess, rentang utama tersebut adalah 600–1600. Saat pengguna melihat laporan pada skala Chess.com, kelompok peringkat yang sesuai dipetakan ke Lichess dasar yang sama bucket benchmark menggunakan metode pemetaan rating lintas platform yang terpisah dijelaskan dalam makalah pendamping: [Estimating Cross-Platform Chess Rating Mappings with Regresi Modal](#).

3 Distribusi Peringkat Pemain Aktif

Jika kita melihat sebaran rating Elo pemain Lichess.org hingga akhir pada bulan Maret 2026, berdasarkan jenis permainan utama, kami melihat jumlah dan persentil berikut. Untuk tampilan distribusi ini, setiap pemain dihitung satu kali per jenis permainan yang tepat, menggunakan rating observasi terbaru pemain tersebut di raw yang sesuai baris pengatur waktu, setelah membutuhkan setidaknya lima game dalam game tersebut ketik.¹

¹We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1–4 tampilkan distribusi

Game type	Players	Median	75th pct.	80th pct.	% ≤ 1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
>10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: Distribusi rating pemain aktif Lichess terbaru yang diamati secara tepat jenis permainan setelah membutuhkan setidaknya lima permainan dalam jenis permainan yang sama.

Distribusi ini berbeda dengan sampel benchmark bertingkat. Distribusi ini menjelaskan kumpulan pemain Lichess aktif yang diamati dalam versi mentah memindai. Sampel patokan tersebut kemudian sengaja distratifikasi agar ember Elo memiliki dukungan yang cukup untuk kurva metrik yang stabil.

Untuk memetakan dari peringkat Elo Lichess.org untuk jenis permainan tertentu ke setara dengan peringkat Chess.com, Elo+Chess menggunakan modal khusus tipe game persamaan regresi diperkirakan dalam pendamping [Elo+Chess rating-mapping paper](#):

$$\begin{aligned}
\hat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,>10rapid}} &= -369.02 + 0.9374R_{\text{Lichess}}.
\end{aligned}$$

Persamaan tersebut memetakan nilai peringkat, bukan peringkat persentil. Peringkat persentil adalah didefinisikan di dalam kumpulan pemain tertentu, dan pemain Chess.com dan Lichess kolom mempunyai bentuk yang berbeda-beda. Dalam Tabel 2, untuk Jenis permainan cepat 10 menit, kolom pertama memberikan contoh kecil kemungkinan Nilai peringkat Chess.com Elo. Kolom kedua memberikan informasi yang sesuai Peringkat persentil Chess.com ditampilkan di Chess.com per 31 Mei 2026. A 71,7

peringkat di Lichess.org selama 10 menit cepat sesuai dengan peringkat Elo Lichess sebesar 1,644² dan nilai tersebut ditunjukkan di kolom ketiga. Jika kita menerapkan persamaan pemetaan Chess.com–Lichess, 1,644 pada Lichess setara dengan sekitar 1,271 di Chess.com. Dengan asumsi perkiraan persamaan pemetaan masuk akal, kesenjangan +522 antara peringkat Chess.com di kolom 1 dan rating yang dipetakan pada kolom 4 dapat dijelaskan dengan angka yang besar pemain kasual di kumpulan pemain Chess.com menggembungkan Chess.com peringkat persentil. Dengan kata lain, berada di persentil ke-72 pada Lichess.org adalah pencapaian yang lebih besar daripada berada di persentil ke-72 di Chess.com, karena peringkat persentil tidak terpetakan dengan rapi dari satu sistem ke sistem lainnya. Elo sebenarnya nilai dapat dipetakan menggunakan persamaan yang diperkirakan dari skor pemain yang sama kedua sistem tersebut.

Alasan kesenjangan dalam Tabel 2 adalah karena pemetaan memperkirakan peringkat kekuatan bermain yang setara, sedangkan persentil menjelaskan posisi dalam populasi satu platform itu sendiri. Pemain dengan persentil 71,7 aktif Oleh karena itu, Chess.com tidak secara otomatis menjadi pemutar persentil ke-71,7 di Lichess, dan sebaliknya. Angka-angka yang diamati konsisten dengan publik Chess.com kumpulan cepat yang memiliki massa pemain berperingkat rendah

dengan dan tanpa filter ini.

²Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Perbandingan ilustratif antara persentil cepat 10 menit Chess.com poin dan titik persentil yang sama di Lichess 10+0 cepat yang aktif distribusi setelah membutuhkan setidaknya lima game 10+0. “Chess.com dipetakan equiv.” menerapkan Elo+Chess saat ini 10+0 konversi skala peringkat cepat ke peringkat Lichess dengan persentil yang sama. Itu baris ekor tertinggi disertakan untuk menunjukkan perbedaan skala, tetapi dipasang garis peringkat paling dapat diandalkan dalam rentang pelatihan utama.

atau kasual yang jauh lebih besar, sedangkan kumpulan Lichess kontrol tepat aktif lebih terkonsentrasi di antara pengulangan yang berpengalaman pemain. Poin kuncinya bukanlah bahwa satu skala persentil itu “benar” dan lainnya adalah “salah”; hanya saja peringkat persentil tidak portabel di seluruh platform kecuali distribusi pemain yang mendasarinya memiliki bentuk yang sama.

Note on the 1500 spike. The visible jump near 1500 ada dalam data, bukan dibuat dengan pemulusan. Yang terbaru distribusi peringkat yang diamati, 1500 peringkat yang tepat menyumbang 13,1

pemain peluru, 5,6

17.6

jangkar awal atau sementara dalam sistem pemeringkatan.

Lonjakan tepat 1.500 sebagian besar disebabkan oleh akun yang sedikit aktif. Untuk periksa ini, kami menghitung ulang distribusi peringkat terbaru yang sama setelah memerlukannya setiap pemain memiliki setidaknya lima permainan dalam jenis permainan yang sama. Lonjakannya hampir menghilang di bawah filter ini: 1500 peringkat tepatnya turun dari 5,6–17,6

pemain dalam distribusi semua pemain menjadi 0,2–1,2

distribusi dilaporkan pada Tabel 1.

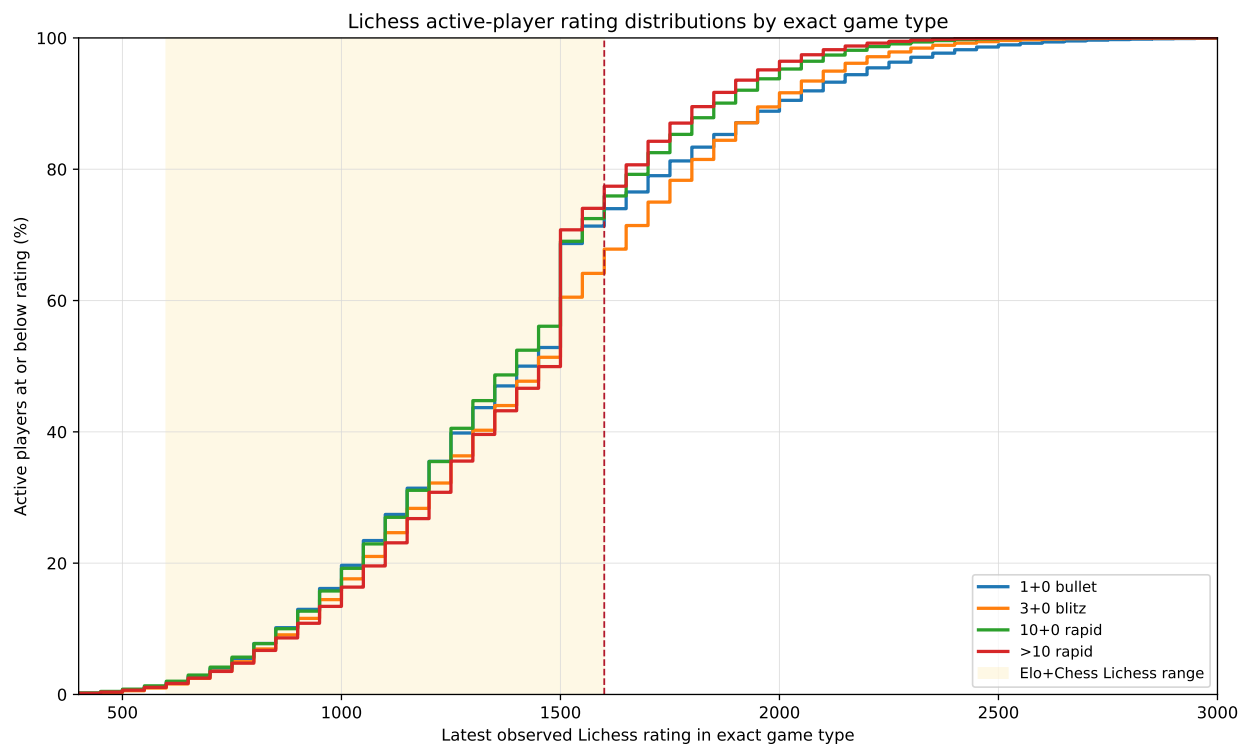


Figure 1: Distribusi peringkat pemain aktif kumulatif berdasarkan jenis permainan yang tepat. Itu pita kuning menandai kisaran 600–1600 Lichess yang digunakan oleh Elo+Chess utama tolok ukur pembinaan.

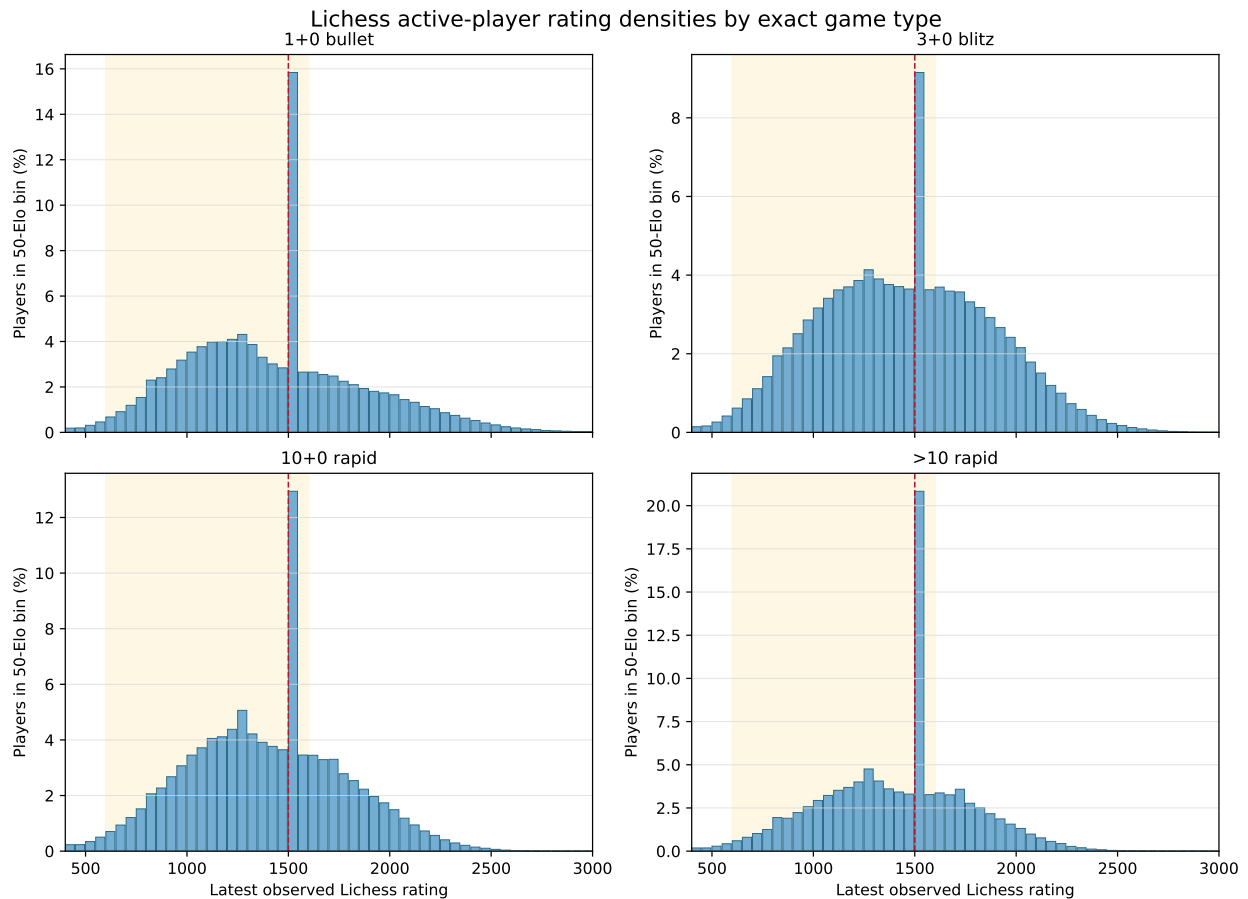


Figure 2: Tampilan kepadatan dari distribusi rating pemain aktif yang sama. Itu garis putus-putus menandai 1500 Lichess, di mana distribusi mentah dengan peringkat terbaru memiliki titik massa dengan rating pasti yang besar. Pita kuning menandai 600–1600 Lichess rentang yang digunakan oleh laporan metrik Elo+Chess utama. Lonjakan paling terlihat di kepadatan semua pemain karena banyak akun yang sedikit aktif tetap berada di 1500; Gambar 4 mengulangi plot kepadatan setelah membutuhkan setidaknya lima game dalam jenis game yang sama, yang sebagian besar akan dihilangkan artefak ini.

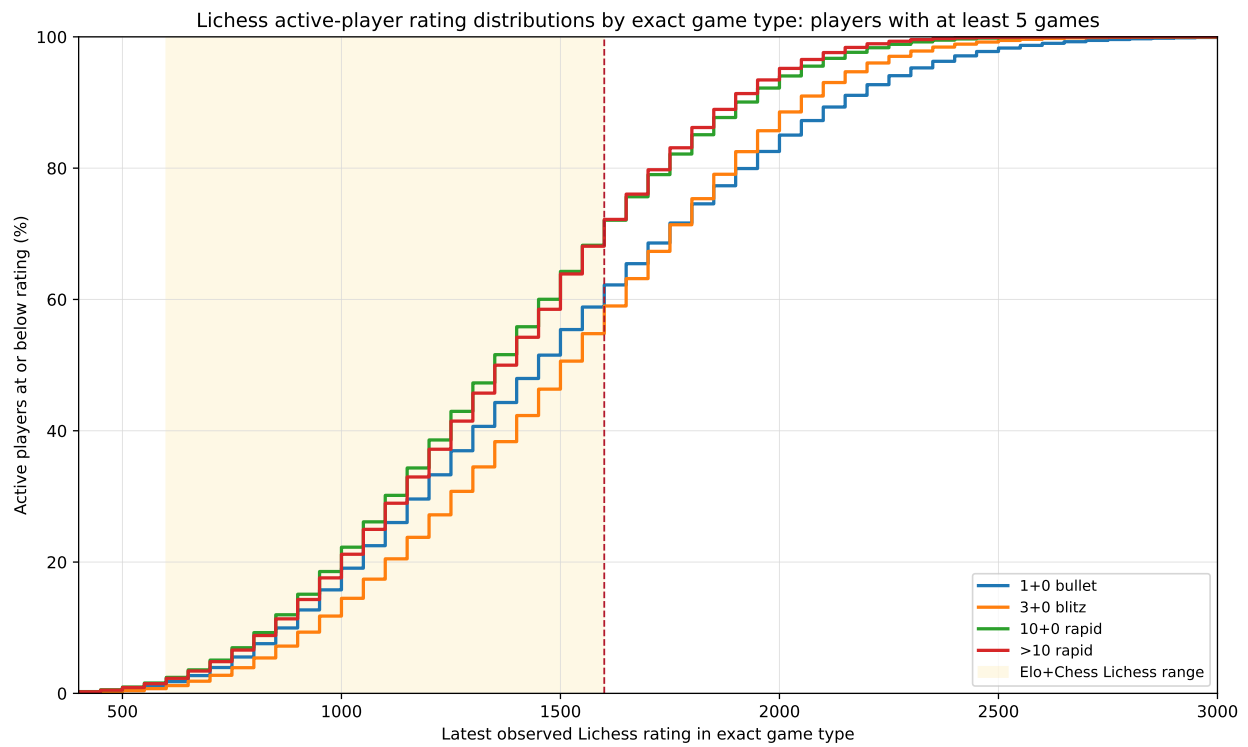


Figure 3: Distribusi rating pemain aktif kumulatif setelah dibutuhkan di setidaknya lima game dalam jenis game yang sama. Lompatan 1500 sangat berkurang.

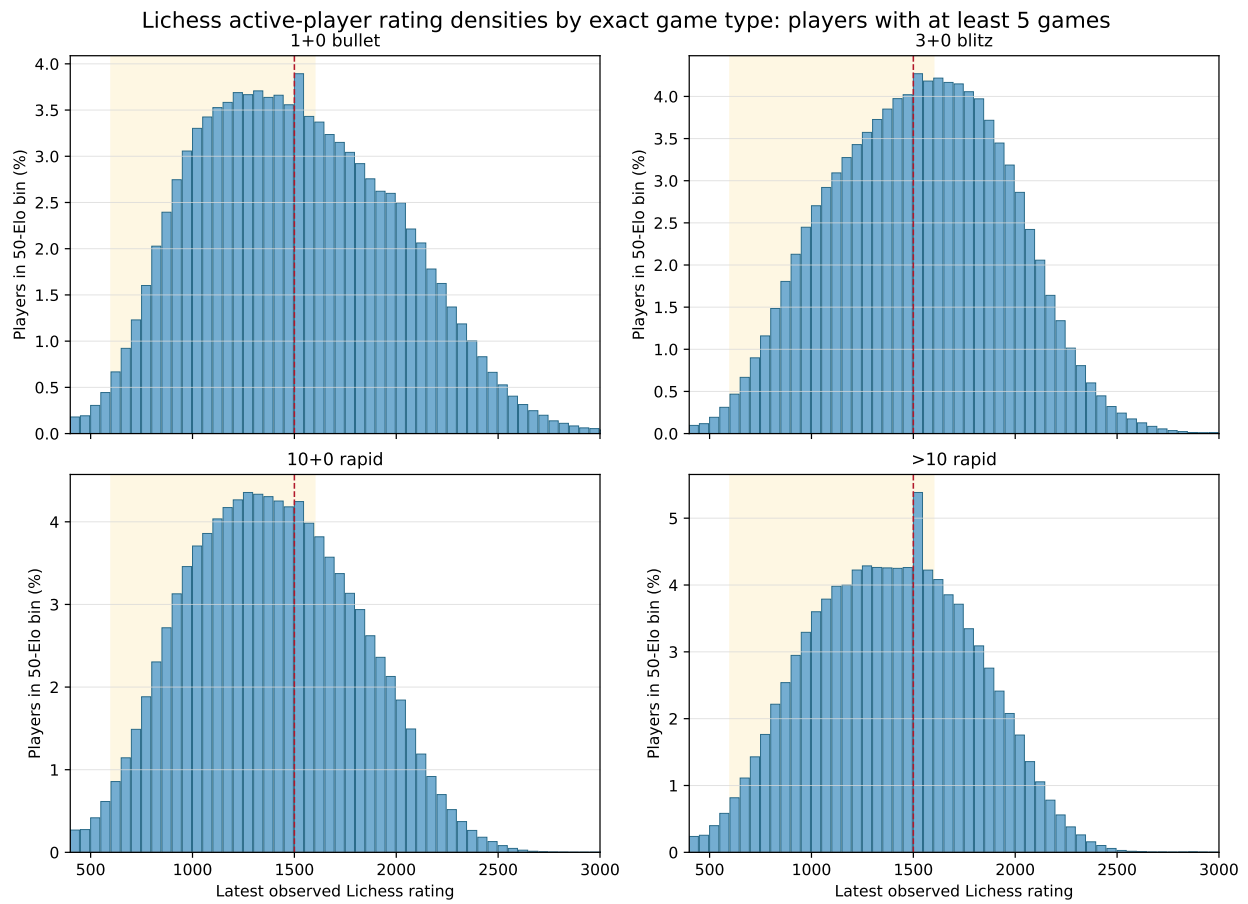


Figure 4: Tampilan kepadatan setelah membutuhkan setidaknya lima game dalam game yang tepat mengetik. Pandangan ini menegaskan bahwa 1500 titik massa besar di semua pemain distribusi sebagian besar didorong oleh akun yang sedikit aktif. Tanda pita kuning rentang 600–1600 Lichess yang digunakan oleh laporan metrik Elo+Chess utama.

4 Skala Artefak Tolok Ukur Saat Ini

Perhitungannya sengaja dibuat besar. Ini bukan goresan profil, itu kecil kumpulan PGN yang diunduh, atau kumpulan contoh instruktif yang dikurasi secara manual. Tiket Lichess awal mencakup file bulanan lengkap mulai Januari 2025 hingga Maret 2026. Tabel pemindaian yang tidak terstratifikasi berisi 1.382.178.087 game dan 2.764.356.174 baris rating game-pemain. Di seluruh baris tersebut, 6.623.438 unik nama pengguna pemain muncul setidaknya sekali.

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: Skala pemindaian Lichess yang tidak terstratifikasi sebelum pengambilan sampel benchmark. Pemindaian mencakup file riwayat permainan bulanan lengkap dari Januari 2025 hingga Maret 2026.

Karena artefak patokan produksi jauh lebih kecil daripada pemindaian mentah mereka dipilih untuk konstruksi metrik. Mereka masih besar: tolok ukurnya tahapnya adalah perluasan negara pindah dan pipa agregasi lebih dari ratusan ribuan game per jenis game.

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
>10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: Skala artefak patokan bertingkat setelah penyelarasan ke skema fitur runtime 277 kolom yang umum. SEBUAH “baris permainan pemain” berarti satu permainan dari sudut pandang satu pemain, jadi sebagian besar permainan menghasilkan dua baris permainan pemain.

Untuk 3+0 blitz saja, benchmark produksi saat ini menggunakan 479.974 game melibatkan 129.398 pemain unik dan 28.138.904 baris status pergerakan berbeda. Itu tabel fitur permainan pemain blitz yang selaras berisi 277 variabel turunan. Itu artefak pembuatan fitur cepat bahkan lebih luas pada tahap keadaan bergerak: penuh tabel status pergerakan cepat yang digunakan selama konstruksi fitur berisi 168 kolom status pindah sebelum digabungkan menjadi fitur permainan pemain 277 kolom meja. Artefak benchmark runtime frontend untuk keempat jenis game tersebut adalah kemudian diselaraskan dengan skema fitur permainan pemain 277 kolom yang sama, termasuk variabel terobosan taktik-taksonomi yang diperluas. Dari tabel fitur ini, laporan memilih dan menampilkan metrik pembinaan yang paling dapat ditafsirkan.

Jumlah ini penting untuk menafsirkan hasil. Kurva acuan bukanlah perkiraan dari beberapa ratus game yang ditinjau secara manual. Sebaliknya, mereka berasal dari puluhan juta negara bagian dewan yang dievaluasi dan lebih dari empat juta baris permainan pemain di empat kategori permainan produksi saat ini.

5 Ikhtisar Saluran Pipa

Pipa konstruksi acuan memiliki enam langkah utama:

1. Pilih game Lichess dengan rating di jenis game target dan strata rating.
2. Parsing setiap langkah permainan demi langkah dan simpan langkah sebelum dan sesudah negara bagian.
3. Hitung variabel keadaan perpindahan perantara dari papan, perpindahan, perspektif pemain, dan perspektif lawan.
4. Gabungkan variabel status pergerakan ke dalam variabel fitur permainan pemain.
5. Ubah variabel permainan pemain menjadi metrik laporan dengan jelas arah interpretasi: lebih tinggi biasanya lebih baik, lebih rendah biasanya lebih baik, atau deskriptif.
6. Rata-ratakan setiap metrik berdasarkan ember Elo dan periksa hasilnya secara manual hubungan.

Struktur ini penting karena banyak metrik laporan yang tidak secara langsung terlihat di header PGN. Mereka memerlukan rekonstruksi situasi dan definisi potongan mana yang dipindahkan dari kotak awal mana, membandingkan sebelum bergerak dan material dan struktur setelah pemindahan, dan kemudian menggabungkan hasil dari berpindah level ke level pemain-permainan.

6 Variabel Menengah

Variabel perantara dirancang untuk menggambarkan apa yang berubah di papan dan mengapa hal itu penting dari sudut pandang pelapor. Contohnya meliputi:

- nomor gerak, nomor gerak pemain, sisi yang bergerak, warna penggerak, dan pemain warna;
- dari persegi, ke persegi, jenis potongan bergerak, jenis potongan yang ditangkap, dan apakah langkah tersebut merupakan penangkapan, pemeriksaan, kastil, promosi, atau pasangan;
- neraca material sebelum dan sesudah pemindahan;
- apakah bidak kecil meninggalkan kotak awalnya dengan nomor perpindahan tertentu;
- apakah bidak yang sama berpindah pada gerakan pemain yang berurutan dalam pembukaan;
- apakah ratu pindah sebelum langkah 10;
- sisi rokade, nomor gerakan rokade, dan apakah benteng terhubung setelah rokade;
- bidak lepas, bidak gantung, bidak ganda, bidak terasing, dioper bidak, dan bidak mundur;
- status file benteng, file terbuka, file setengah terbuka, dan keamanan raja fitur;
- peluang taktis, taktik yang dilaksanakan, taktik tekanan material, jaminan keuntungan tindak lanjut, dan keuntungan yang direalisasikan.

Set perantara yang tepat bervariasi menurut tahap pembuatan. Basis data laporan runtime disimpan hanya kolom yang diperlukan untuk menyajikan laporan dengan cepat, sedangkan fitur lengkap dibuat artefak mencakup tabel status pergerakan yang lebih luas dan patch fitur yang digunakan untuk membuat baris fitur permainan pemain terakhir.

7 Contoh Metrik Prinsip Pembukaan

Prinsip pembukaan adalah contoh yang berguna karena mudah dilakukan oleh pengguna dipahami dan mudah dihubungkan dengan perhitungan keadaan bergerak.

7.1 Minor Pieces Developed by Move 10

Metrik ini menghitung berapa banyak dari empat ksatria dan uskup pemain yang tersisa kotak rumah mereka dengan langkah ke-10 pemain. Perhitungan menengah menggunakan warna penggerak, jenis bidak bergerak, kotak awal, dan gerakan pemain nomor. Untuk Putih, kotak rumahnya adalah b1, g1, c1, dan f1. Bagi Hitam, memang benar b8, g8, c8, dan f8. Untuk setiap permainan pemain, Elo+Chess menghitung rumah yang berbeda kotak di antara ksatria dan uskup yang digerakkan oleh gerakan pemain 10.

Ini adalah metrik yang lebih tinggi biasanya lebih baik. Ia tidak mengklaim bahwa setiap anak di bawah umur bidak harus selalu bergerak demi gerakan 10, tetapi di antara pemain yang sedang berkembang, yang empiris hubungannya jelas: pemain yang mencapai ember Elo yang lebih tinggi cenderung berkembang lebih banyak potongan kecil sebelumnya.

7.2 Repeated Piece Moves in the Opening

Metrik ini menghitung langkah awal ekstra yang digunakan untuk memindahkan bidak yang sama lagi untuk mengembangkan sesuatu yang baru. Perhitungan perantara menggunakan perhitungan sebelumnya kotak tujuan, kotak asal saat ini, dan nomor perpindahan pemain. Jika bidak yang sama dipindahkan pada gerakan pemain sebelumnya dan dipindahkan lagi di dalam 10 gerakan pemain pertama, jumlah gerakan berulang bertambah.

Ini adalah metrik yang lebih rendah biasanya lebih baik. Ini mengukur waktu pembukaan yang terbuang di a cara sederhana. Beberapa gerakan yang diulang-ulang secara taktis dapat dibenarkan, namun dalam skala besar kurva ember menangkap kecenderungan umum: pemain yang lebih kuat mencapai target level menghabiskan lebih sedikit gerakan awal untuk menggerakkan bidak yang sama berulang kali.

7.3 Queen Touched Before Move 10

Metrik ini menghitung gerakan ratu sebelum pemain bergerak 10. Perantara perhitungannya menggunakan jenis bidak bergerak dan nomor gerak pemain. Jika bagian yang bergerak adalah ratu dan jumlah gerakan pemain paling banyak 10, hitungannya bertambah.

Ini juga lebih rendah-biasanya-lebih baik untuk rentang target Elo+Chess. Awal gerakan ratu dapat memenangkan konsesi materi atau paksaan, tetapi banyak pemain berkembang mengeluarkan ratu terlalu dini, kehilangan waktu untuk menyerang ratu, dan menunda pengembangan bagian kecil dan keamanan raja.

7.4 Castling and King Safety

Metrik castling menggunakan bendera castling, sisi castling, nomor gerakan castling, dan keadaan papan setelah rokade. Contohnya termasuk apakah pemain melakukan rokade, apakah pemain melakukan rokade dengan gerakan 10, apakah pemain melakukan rokade di sisi raja atau sisi ratu, apakah benteng terhubung setelah rokade, dan apakah benteng file terbuka atau setengah terbuka setelah rokade.

Metrik ini tidak semuanya diinterpretasikan dengan cara yang sama. Penyelesaian kastil dan rokade demi gerakan 10 umumnya lebih tinggi-biasanya-lebih baik dalam kisaran target. Rokade Queenside atau rokade ke dalam file terbuka bisa lebih bergantung pada konteks. Itulah sebabnya Elo+Chess tidak sekadar mengkodekan aturan moral secara keras; itu membangun ember kurva dan memeriksa hubungan mana yang cukup stabil untuk pembinaan.

8 Kurva Bucket dan Inspeksi Manual

Setelah metrik permainan pemain dihitung, Elo+Chess mengelompokkannya berdasarkan Lichess Elo bucket dan menghitung rata-rata bucket. Rata-rata ember adalah empiris mentah titik acuan untuk metrik tersebut. Garis tren kemudian digunakan dalam laporan tetapkan nilai gaya Elo ke nilai pengguna.

Kurva diperiksa secara manual sebelum digunakan. Metrik lebih berguna ketika bucket berarti menunjukkan hubungan yang koheren di seluruh rentang peringkat dilayani oleh produk tersebut. Metrik kurang berguna jika berisik, datar, dan sangat tinggi non-monotonik, atau didorong oleh kasus-kasus khusus yang sulit dijelaskan kepada pengguna.

Langkah pemeriksaan manual ini sangat penting terutama pada bagian atas Elo+Chess pemo-tongan. Pemain yang lebih kuat tahu kapan harus melanggar prinsip pembukaan dan melakukannya secara efektif. Di atas rentang pemula hingga mahir awal, hubungannya antara kebiasaan sederhana dan penilaian sering kali mendatar atau berubah bentuk. Produk oleh karena itu menekankan rentang di mana hubungan empirisnya kuat dan berguna secara pedagogi.

9 Contoh Kurva Pembukaan Rentang Penuh

Gambar 5 menunjukkan empat kurva prinsip pembukaan blitz 3+0. Titik abu-abu menunjukkan rentang peringkat Lichess yang lebih luas. Titik biru dan tren biru garis menunjukkan kisaran 600–1600 Lichess yang digunakan oleh pelatihan utama Elo+Chess tolok ukur. Garis merah putus-putus menandai batas atas.

Pola yang sama terlihat pada kecepatan penuh 10+0 dan kecepatan lebih lama artefak. Keputusan pemodelan utama tidak berubah: gunakan secara empiris kisaran stabil untuk klaim pembinaan dan hindari berpura-pura bahwa kurva kebiasaan sederhana tetap sama informatifnya selamanya.

Pola yang sama muncul secara numerik di Tabel 5. Lereng adalah dilaporkan sebagai perubahan nilai metrik per 100 poin Elo. Castling, potongan kecil perkembangan, pergerakan bidak yang berulang, dan pergerakan ratu awal semuanya terlihat jelas pergerakan melintasi 600–1600. Di atas tahun 1600, lerengnya jauh lebih kecil contoh, yang konsisten dengan gagasan bahwa pemain dengan peringkat lebih tinggi akan kalah prinsip dasar secara lebih selektif dan lebih berhasil. Mekanisme kedua adalah saturasi. Untuk beberapa kebiasaan, metriknya mencapai batas praktis atau manis spot: pemain kuat sudah mempelajari kebiasaan itu, jadi tidak ada alasan untuk itu kurva untuk terus bergerak tajam ke atas. Lebih banyak juga tidak selalu lebih baik setelahnya titik itu. Oleh karena itu, data tersebut meratakan keduanya karena pemain yang lebih kuat dapat mematahkannya aturan dalam posisi konkret dan karena banyak kebiasaan dasar telah tercapai jangkauan operasinya yang berguna.

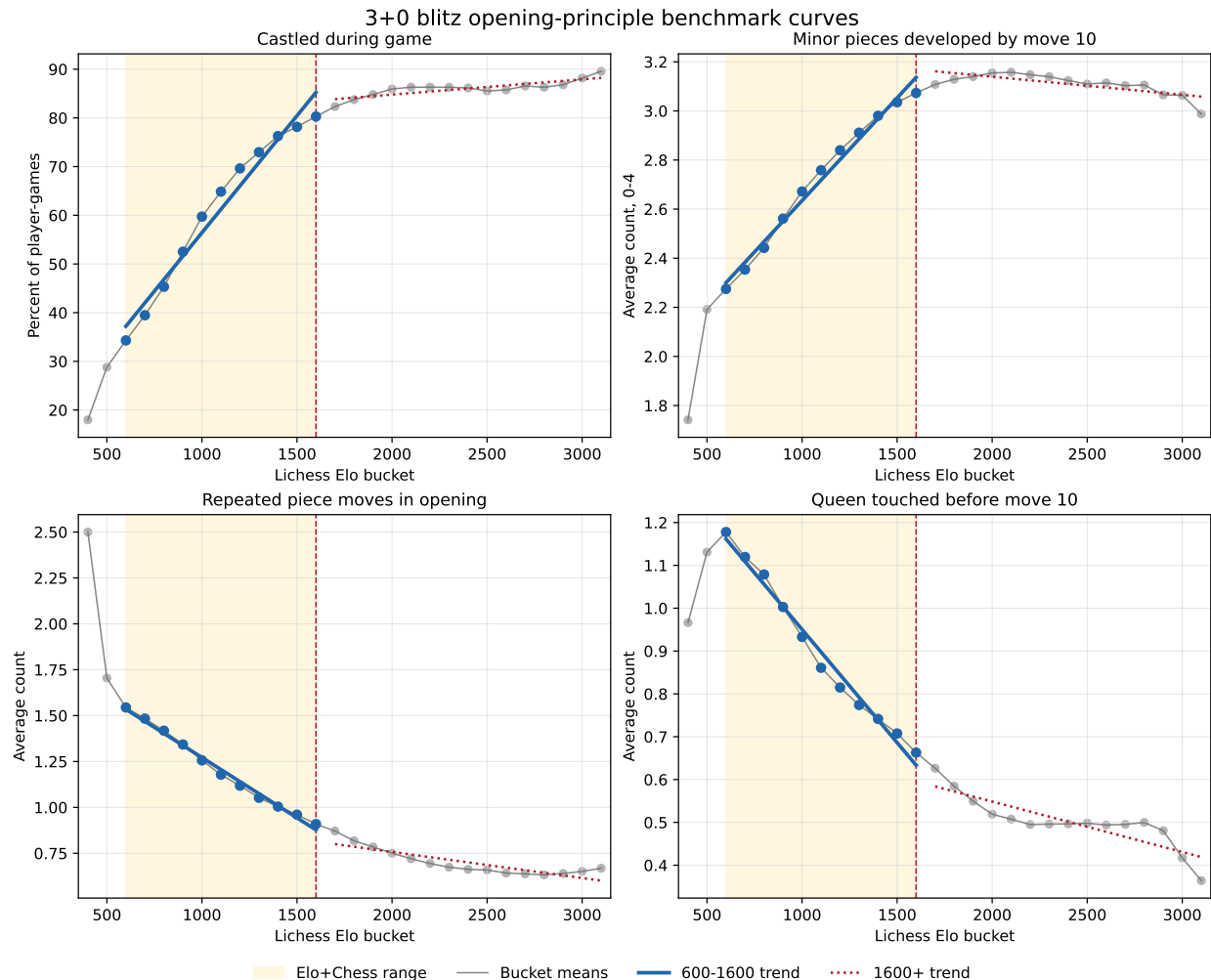


Figure 5: Kurva tolok ukur prinsip pembukaan yang dipilih untuk blitz 3+0. Itu hubungan yang paling kuat dan paling berguna dalam rangkaian pelatihan Elo+Chess. Di atas batas atas, beberapa hubungan menjadi datar secara substansial.

10 Mengapa Batas Atas Penting

Target pengguna Elo+Chess adalah pemain pemula hingga mahir awal. Untuk ini pemain, kebiasaan dasar sangat terkait dengan peringkat: mengembangkan bidak, hindari gerakan bidak awal yang berulang, hindari gerakan ratu awal yang tidak perlu, kastil tepat waktu, hubungkan benteng, kelola material, dan hindari membiarkan bagian lepas atau gantung.

Pada peringkat yang lebih tinggi, kebiasaan permukaan yang sama menjadi kurang diagnostik. Kuat pemain mungkin menunda rokade karena pusatnya ditutup, pindahkan ratu lebih awal karena posisinya mengandung taktik konkrit, atau menggerakkan bidak yang sama sebanyak dua kali karena ia memenangkan tempo atau memaksakan konsesi struktural. Kebiasaan mentahnya masih memiliki arti, namun hubungan antara kebiasaan dan rating menjadi lebih bersyarat. Inilah sebabnya mengapa Elo+Chess menggunakan batas atas untuk pembinaan utamanya klaim dan mengapa pengguna di atas kisaran tersebut diperingatkan bahwa ada hubungan yang erat antara kebiasaan berprinsip dan kerusakan peringkat.

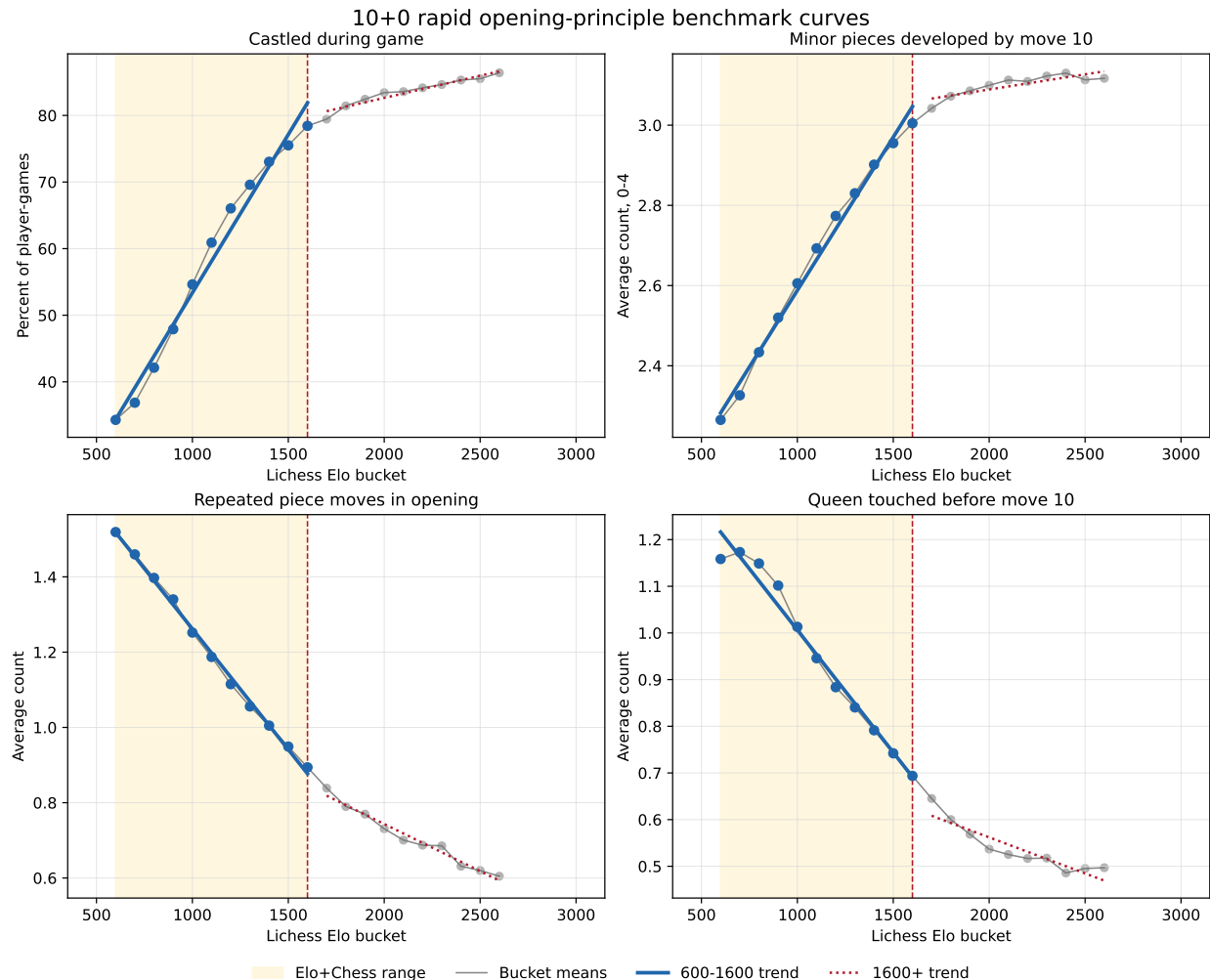


Figure 6: Kurva tolak ukur prinsip pembukaan yang dipilih untuk game cepat 10+0. Itu hubungan yang paling kuat dan paling berguna dalam rangkaian pelatihan Elo+Chess. Di atas batas atas, beberapa hubungan menjadi datar secara substansial.

11 Metrik Taktis Dua Sisi dan Target Persentil

Beberapa metrik memiliki bentuk yang berlawanan dengan contoh prinsip pembuka. Garpu adalah contoh yang berguna karena peristiwa tersebut pada dasarnya bersifat dua sisi. Seorang pemain yang menemukan dan mengubah garpu yang lebih berguna daripada rekan-rekannya melakukan sesuatu yang baik, tapi tingkat populasi garpu yang dikonversi turun seiring dengan peringkat karena lebih kuat lawan memberikan lebih sedikit peluang bercabang dan bertahan dari ancaman taktis lebih aktif.

Gambar 8 menunjukkan efek ini. Keduanya mengkonversi garpu dengan pemain laporan dan garpu yang dikonversi oleh lawan menjadi lebih jarang peringkat naik. Itu tidak berarti mengekskusi fork itu buruk. Artinya peristiwa mentah tingkat sebagian dikendalikan oleh tingkat kesalahan lawan dan tekstur permainan. Untuk metrik seperti ini, Elo+Chess menggunakan persentil riwayat rekan di dalamnya ember Elo milik pemain daripada hanya mengandalkan tren rata-rata ember global.

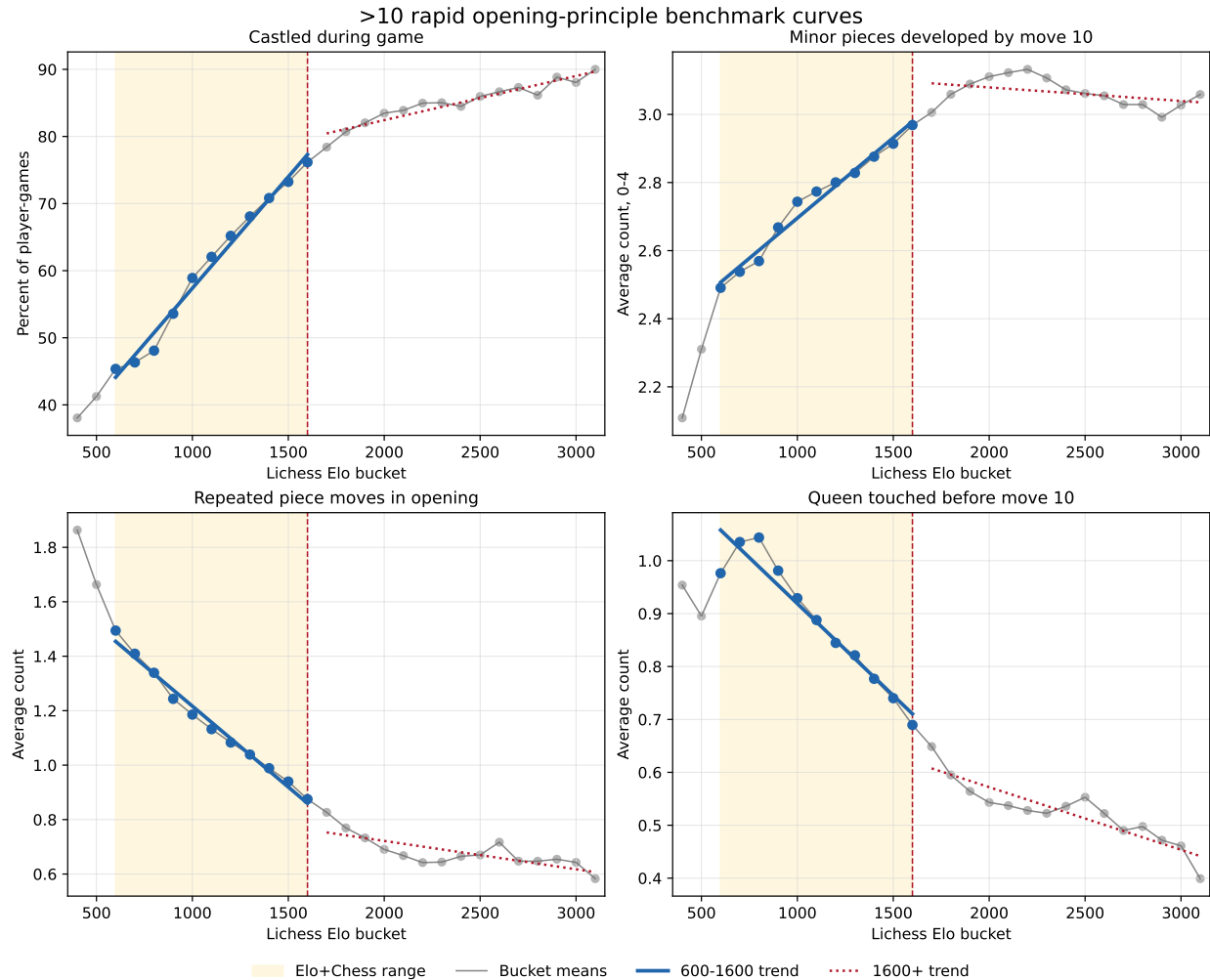


Figure 7: Kurva tolak ukur prinsip pembukaan yang dipilih untuk permainan cepat yang lebih lama.

Untuk membangun target persentil ini, Elo+Chess mengambil sampel pemain berdasarkan jenis permainan dan Elo bucket, lalu menggunakan game kualifikasi terkini untuk setiap pemain sampel. Itu kumpulan target saat ini menggunakan 11 keranjang dari 600 hingga 1600, 200 pemain per ember, dan hingga 50 game kualifikasi terbaru per pemain untuk setiap produksi jenis permainan. Ini memberikan 2.200 sampel pemain per jenis game dan sekitar 110.000 baris permainan pemain per jenis permainan untuk persentil dalam keranjang distribusi.

Gambar 9 mengilustrasikan target yang dihasilkan distribusi untuk tingkat fork yang dikonversi pemain sampel dalam serangan kilat 3+0. Setiap kotak adalah distribusi dalam kelompok di seluruh pemain yang dijadikan sampel, bukan rata-rata di seluruh kelompok permainan individu. Di UI laporan, nilai pengguna dibandingkan dengan nilai yang relevan kotak untuk ember Elo yang dipetakan pengguna. Ini adalah dasar untuk pernyataan seperti apakah tingkat konversi fork seorang pemain tinggi atau rendah dibandingkan dengan rekan-rekannya di pertandingan tersebut tingkat yang sama, bahkan ketika tingkat kejadian populasi secara keseluruhan menurun seiring dengan rating.

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: Perkiraan kemiringan kurva ember untuk empat contoh prinsip pembukaan metrik di seluruh jenis game produksi.

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: Kumpulan target persentil riwayat rekan saat ini. Ini adalah sampel riwayat pemain yang digunakan untuk memperkirakan distribusi dalam keranjang metrik gaya persentil.

12 Biaya Replikasi

Pada prinsipnya, metode ini dapat direproduksi, tetapi tidak ringan. Membangun kembali kurva acuan memerlukan:

- akses ke file riwayat permainan bulanan Lichess yang besar;
- pemfilteran bertingkat berdasarkan jenis game, kelompok peringkat, dan kelayakan aturan;
- parser langkah hukum dan evaluator dewan negara yang mampu berkembang puluhan juta negara bagian yang berpindah;
- logika fitur untuk pembukaan, material, struktur pion, keamanan raja, benteng aktivitas, taktik, manajemen waktu, dan hasil permainan;
- agregasi dari status perpindahan ke fitur permainan pemain;
- ringkasan tingkat keranjang untuk setiap metrik laporan;
- tinjauan manual terhadap setiap kurva benchmark kandidat sebelum digunakan laporan pelatihan.

Two-sided converted-fork rates by Elo bucket

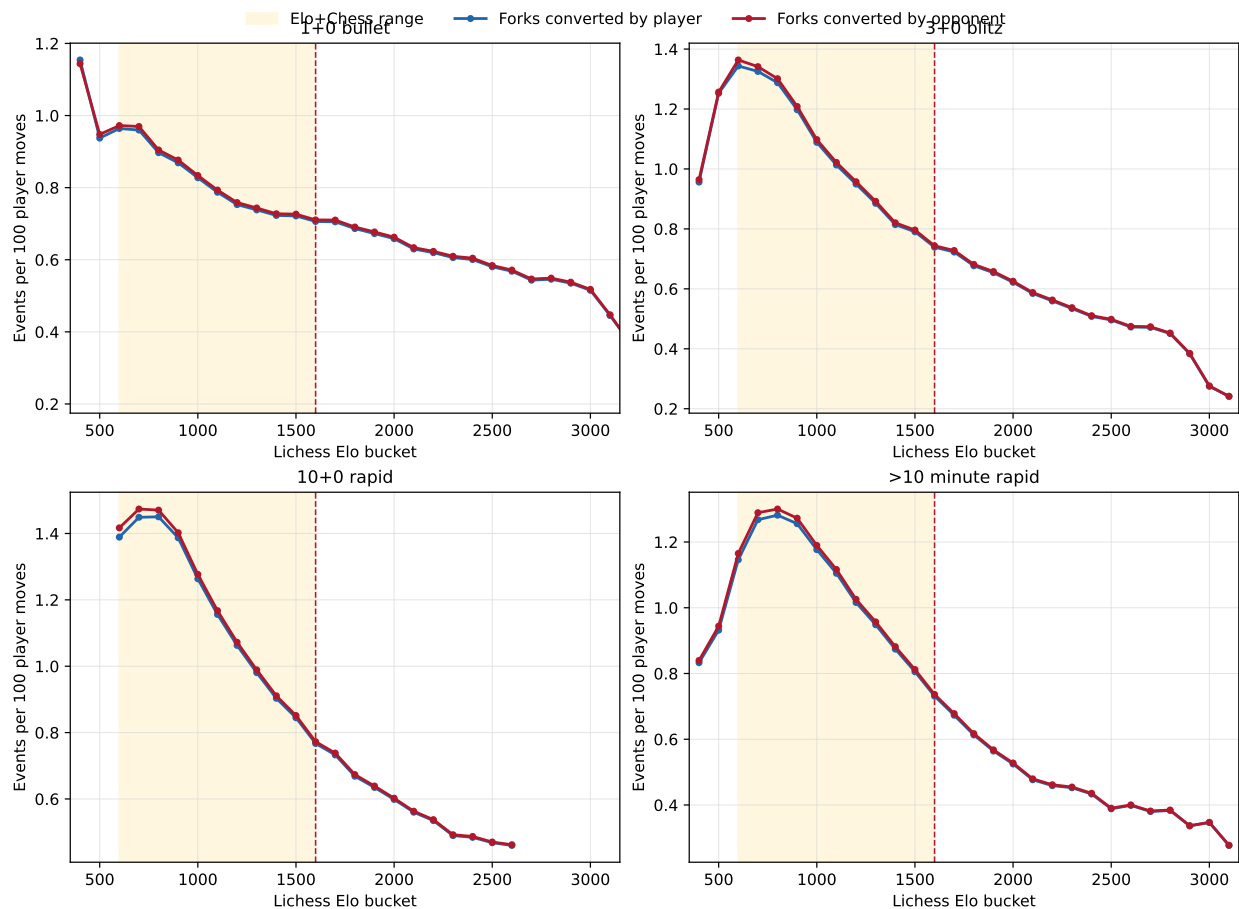


Figure 8: Tarif garpu konversi dua sisi dengan ember Elo. Acara fork yang dikonversi jatuh pada peringkat yang lebih tinggi karena kedua belah pihak bertahan lebih baik dan lebih sedikit clean fork konversi tersedia.

Untuk blitz 3+0, ini berarti hampir setengah juta game, hampir satu juta baris permainan pemain, dan lebih dari 28 juta baris keadaan bergerak. Di keempatnya jenis permainan benchmark, artefak benchmark saat ini berisi lebih dari 2,55 juta game, lebih dari 5,10 juta baris permainan, dan lebih dari 147 juta baris keadaan bergerak. Skala adalah pusat dari nilai benchmark: memungkinkan laporan menunjukkan kepada pengguna apa yang sebenarnya terjadi dalam game yang dimainkan oleh pemain yang mendekati levelnya daripada hanya mengandalkan slogan catur umum.

13 Keterbatasan

Kurva patokan adalah deskripsi empiris, bukan bukti sebab akibat. Seorang pemain tidak mendapatkan peringkat hanya dengan mencocokkan mean bucket. Metrik juga demikian ringkasan yang tidak sempurna dari posisi kompleks: gerakan ratu awal bisa menjadi hal yang sangat baik di satu posisi dan ceroboh di posisi lain. Tujuan dari kurva adalah untuk mengidentifikasi kebiasaan yang berulang dan membandingkannya dengan perilaku teman sebaya, bukan dengan menggantikan

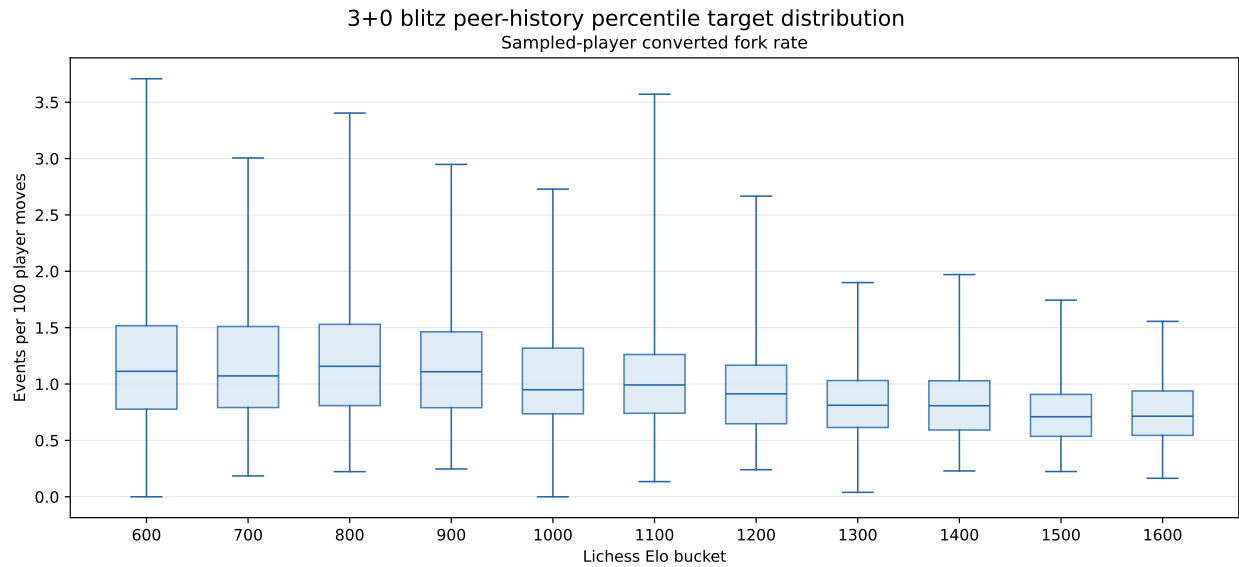


Figure 9: Contoh distribusi target histori sejawat kilat 3+0 untuk pemain sampel tingkat garpu yang dikonversi. Persentil dihitung dalam ember Elo, jadi laporannya membandingkan seorang pemain dengan rekan-rekannya yang menghadapi peluang taktis yang serupa tarif.

perhitungan atau penilaian pembinaan.

Kurvanya juga merupakan kurva berskala Lichess. Saat pengguna membawa game Chess.com sejarah, Elo+Chess memetakan kelompok peringkat yang relevan ke benchmark Lichess skala menggunakan model pemetaan lintas platform terpisah. Lapisan pemetaan itu adalah dijelaskan dalam catatan penelitian pendamping Elo+Chess di Lichess-to-Chess.com konversi peringkat: [Estimating Cross-Platform Chess Rating Mappings with Regresi Modal](#). Pemetaan ini tidak boleh dibaca sebagai mempertahankan persentil peringkat. Hal ini mempertahankan perkiraan kesetaraan peringkat antara kedua peringkat tersebut sistem; peringkat persentil tetap spesifik platform karena pemain aktif distribusinya berbeda.

14 Grafik Laporan Produksi

Laporan produksi mengubah artefak tolok ukur menjadi panel tingkat metrik yang dapat dibaca tanpa mengetahui pipeline build lengkap. Angka 10 and 11 tunjukkan tiga contoh dari desain laporan Elo+Chess langsung.

Panel prinsip pembuka pada Gambar 10 gunakan presentasi kurva benchmark biasa. Setiap kartu dimulai dengan keluarga metrik dan nama metrik, lalu memberikan aturan interpretasi singkat. Itu grade besar di pojok kanan atas sengaja dipisahkan dari yang mentah nilai metrik: ini menjawab pertanyaan yang dihadapi pengguna “bagaimana nilai kebiasaan ini terhadap patokan?” sementara baris di bawah mempertahankan nilai yang diukur. Empat baris ringkasan menunjukkan Seumur Hidup, 10 Terakhir, Kemenangan, dan Kalah. Hal ini membuat diagnostik panel, bukan sekadar dekoratif. Jika Seumur Hidup dan 10 Terakhir setuju, kebiasaan itu mungkin stabil. Jika kemenangan dan kekalahan berbeda, metriknya mungkin berbeda bergantung pada hasil atau terikat pada kondisi permainan, bukan kebiasaan sederhana.

Area grafik menggunakan latar belakang gelap, garis tren emas, dan berwarna penanda untuk

empat irisan. Anotasi tren secara eksplisit menyatakan apakah hubungan patokannya naik atau turun dengan Elo. Ini penting karena beberapa metrik lebih tinggi-lebih baik, seperti pengembangan bagian kecil, sementara itu yang lain lebih rendah-lebih-lebih, seperti gerakan ratu awal. Oleh karena itu, kartu itu tidak meminta pengguna untuk menyimpulkan arah hanya dari kemiringannya saja. Yang terakhir Baris “Apa artinya” dan “Tips pelatihan” menerjemahkan hasil benchmark menjadi instruksi catur yang konkret dengan tetap menjaga konteks empiris.

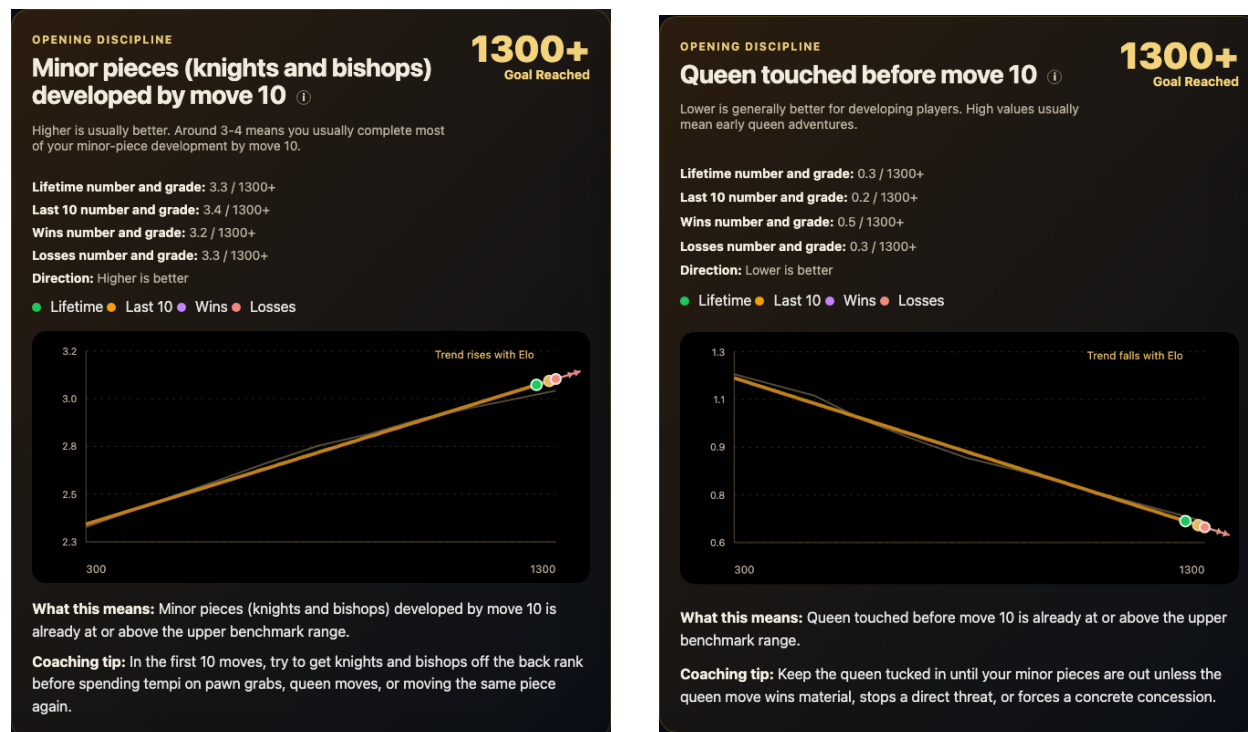


Figure 10: Panel prinsip pembukaan produksi. Tata bahasa visual yang sama menangani a metrik yang lebih tinggi adalah yang lebih baik, pengembangan bagian kecil, dan metrik yang lebih rendah adalah yang lebih baik, gerakan ratu awal. Kedua panel mengekspos Seumur Hidup, 10 Terakhir, Kemenangan, dan Kerugian sehingga laporan tersebut dapat memisahkan kebiasaan stabil dari kebiasaan terkini atau yang bergantung pada hasil pola.

Gambar 11 menunjukkan gaya persentil presentasi yang digunakan untuk metrik yang laju mentahnya tidak terwakili dengan baik oleh a tren global tunggal. Garpu yang menang adalah contoh taktisnya: tingkat populasi perubahan garpu yang dikonversi dengan keterampilan pemain dan tingkat kesalahan lawan. Itu Oleh karena itu, kartu membandingkan nilai pengguna dengan riwayat pemain rekan di memetakan ember Elo. Boxplot menampilkan distribusi rekan, termasuk jangkauan antarkuartil, median, dan kumis. Penanda pengguna diplot pada skala yang sama dan diberi label dengan nilai mentah dan peringkat persentil. Berdekatan keranjang perbandingan tetap terlihat dalam bentuk senyap sehingga pengguna dapat melihat bahwa peer bucket bersifat lokal, bukan skala universal.

Desain ini membuat metrik persentil dapat diaudit. Pengguna melihat nilai mentahnya, persentil, keranjang rekan yang digunakan untuk perbandingan, dan terjemahannya Pemetaan bucket Chess.com-ke-Lichess. Seumur Hidup yang sama, 10 Terakhir, Kemenangan, dan Barisan kekalahan dipertahankan sehingga skor persentil taktis masih dapat diperiksa untuk efek kekinian dan ketergantungan hasil.

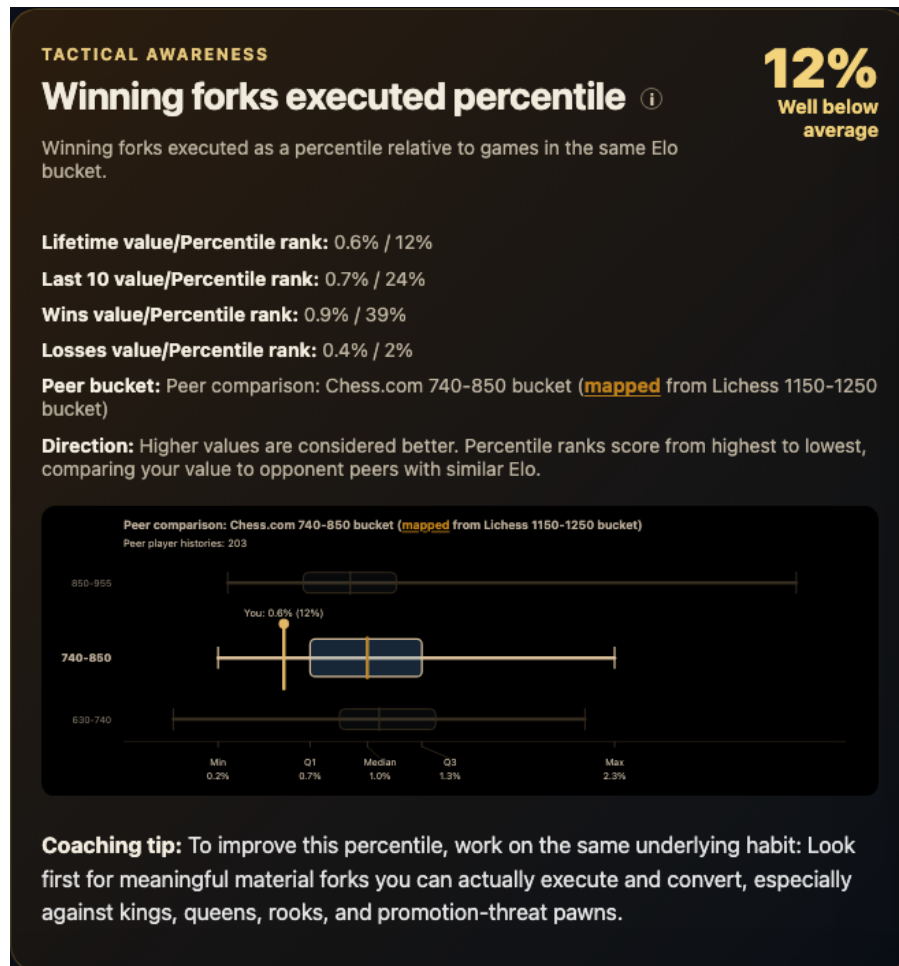


Figure 11: Panel persentil produksi untuk garpu pemenang dieksekusi. Plot kotak menunjukkan distribusi rekan keranjang yang sama yang digunakan oleh laporan, sedangkan emas penanda menunjukkan nilai mentah dan peringkat persentil pengguna.

15 Kesimpulan

Laporan benchmark Elo+Chess dibuat dari perhitungan status pergerakan yang besar sampel Lichess bertingkat. Setiap metrik dimulai sebagai satu set keadaan dewan yang konkret dan variabel keadaan gerak, menjadi fitur permainan pemain, dan kemudian dirata-ratakan oleh ember Elo untuk membentuk kurva patokan empiris. Prinsip pembuka contoh menunjukkan mengapa pendekatan ini berguna: kebiasaan catur yang sederhana mempunyai pengaruh yang kuat, hubungan yang terlihat dengan peringkat dalam kisaran target, sementara itu hubungan sering kali mendatar atau menjadi lebih kondisional di atas batas atas. Itulah alasan utama Elo+Chess memfokuskan klaim pembinaannya pada pemeringkatan rentang di mana data mendukungnya dengan paling jelas.