

Evaluating the Accuracy of Minimum Admission Score Threshold Projections Using Linear Regression and Shrinkage-to-Mean on SNBT Data

Royan Habibie Sukarna^{1*)}, Yulian Ansori²⁾, Fitri Damyati³⁾, Firdaus Satrio Utomo⁴⁾, Nisrina Putri Rajwa⁵⁾

¹⁾²⁾³⁾⁴⁾⁵⁾Program Studi Informatika, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa

^{*)}Correspondence author: royan@untirta.ac.id, Cilegon, Indonesia

DOI: <https://doi.org/10.37012/jtik.v12i2.3764>

Abstract

This study evaluates, out-of-sample, the accuracy of two methods for projecting minimum admission score thresholds for the SNBT selection track: per-series linear regression (program $\times$ degree-level combinations) and shrinkage-to-mean, using 2023–2025 historical data tested against actual 2026 values across 47 series. Pure linear regression achieves MAE 39.40, RMSE 52.34, and 71.5% pass/fail classification accuracy, but exhibits optimistic bias (recall for the "fail" class only 0.543) and a systematic failure pattern: all 13 series with monotonic three-year training trends (consistently rising or falling) missed 100% in the mean-reversion direction during the evaluation year. The shrinkage-to-mean method, which blends linear projections with historical averages weighted by in-sample R^2 , improves every metric (MAE 31.91, RMSE 43.58, accuracy 74.5%, "fail" recall 0.585) without additional data, yet does not change the bias direction and shows limitations for series with very short historical records (degenerate R^2). This study formulates and compares two training-data-only early risk-detection mechanisms, with an explicit trade-off between discriminative power and recalibration requirements. Main contributions: genuine out-of-sample evaluation, identification of a universally consistent mean-reversion pattern, and concrete recommendations for designing similar trend-based prediction features.

Keywords: Linear Regression, Shrinkage Estimator, Mean-Reversion, Out-Of-Sample Evaluation, Admission Score Threshold.

Abstrak

Penelitian ini mengevaluasi secara out-of-sample akurasi dua metode proyeksi ambang nilai minimum penerimaan jalur SNBT: regresi linear per seri (kombinasi program studi $\times$ jenjang) dan shrinkage-to-mean, menggunakan data historis 2023–2025 yang diuji terhadap nilai aktual 2026 pada 47 seri. Regresi linear murni mencapai MAE 39,40, RMSE 52,34, dan akurasi klasifikasi lolos/tidak-lolos 71,5%, namun menunjukkan bias optimis (recall kelas "tidak lolos" hanya 0,543) serta pola kegagalan sistematis: seluruh 13 seri dengan tren tiga tahun yang monoton (naik/turun terus) meleset 100% ke arah pembalikan tren (mean-reversion) pada tahun evaluasi. Metode shrinkage-to-mean, yang memadukan proyeksi linear dengan rata-rata historis berbobot R^2 in-sample, memperbaiki seluruh metrik (MAE 31,91, RMSE 43,58, akurasi 74,5%, recall "tidak lolos" 0,585) tanpa data tambahan, namun tidak mengubah arah bias tersebut dan menunjukkan keterbatasan pada seri berdata historis sangat pendek (R^2 degenerate). Penelitian ini merumuskan dan membandingkan dua mekanisme deteksi risiko dini berbasis data training semata, dengan trade-off eksplisit antara daya diskriminasi dan kebutuhan kalibrasi ulang. Kontribusi utama: evaluasi out-of-sample genuin, identifikasi pola mean-reversion yang berlaku universal, dan rekomendasi konkret untuk desain fitur prediksi berbasis tren serupa.

Kata Kunci: Regresi Linear, Shrinkage Estimator, Mean-Reversion, Evaluasi Out-Of-Sample, Ambang Nilai Penerimaan.

PENDAHULUAN

Ambang nilai minimum penerimaan pada seleksi masuk perguruan tinggi umumnya berubah dari tahun ke tahun, dipengaruhi oleh dinamika jumlah pendaftar, kebijakan kuota, dan tingkat kesulitan instrumen seleksi (Arnold, 2015; Qattan et al., 2023). Ketika sebuah institusi ingin menyediakan fitur pendukung keputusan yang memproyeksikan ambang nilai minimum tahun berjalan bagi calon pendaftar dalam kasus ini misalnya untuk membantu memperkirakan peluang lolos suatu program studi berdasarkan skor yang dimiliki dengan menggunakan pendekatan paling sederhana adalah mengekstrapolasi tren nilai minimum dari beberapa tahun sebelumnya. Pendekatan ini menarik karena implementasinya ringan dan tidak membutuhkan data tambahan di luar riwayat ambang nilai itu sendiri.

Proyeksi tren linear jangka pendek (tiga titik data historis per unit analisis) berisiko menghasilkan estimasi yang bias apabila pola sesungguhnya tidak murni linear, misalnya apabila ambang nilai bersifat mean-reverting (pergerakan tajam pada suatu periode cenderung dikoreksi kembali ke arah rata-rata pada periode berikutnya, bukan berlanjut searah). Tanpa evaluasi kuantitatif yang eksplisit, tidak diketahui seberapa besar risiko bias ini, maupun apakah tersedia metode alternatif yang lebih tahan terhadap risiko tersebut.

Regresi linear sederhana banyak dipakai untuk memproyeksikan deret waktu pendek per kategori (few observations per series), terutama ketika unit analisis berjumlah banyak namun titik data historis per unit sangat terbatas, ini merupakan kondisi yang umum dijumpai pada data agregat institusional (Iwata et al., 2025; Khan & Olivier, 2023). Salah satu risiko utama pendekatan ini adalah overfitting terhadap pergerakan jangka pendek yang sesungguhnya bersifat noise, bukan tren struktural yang merupakan sebuah gejala yang berkaitan erat dengan fenomena regression to the mean: nilai ekstrem pada satu periode pengamatan cenderung diikuti oleh nilai yang lebih mendekati rata-rata pada periode berikutnya (Khan & Olivier, 2023; Özkale & Mammadova, 2026).

Sebagai respons terhadap risiko tersebut, teknik shrinkage (penyusutan estimasi individual ke arah rata-rata kelompok atau rata-rata historisnya sendiri) dikenal sebagai cara meredam varians estimasi pada sampel kecil, dengan mengorbankan sedikit bias untuk mengurangi variance secara substansial (Ahmed & Raheem, 2012; Singh et al., 2024). Penerapan gagasan ini pada proyeksi ambang nilai penerimaan dan evaluasi formalnya secara out-of-sample terhadap data aktual belum banyak didokumentasikan secara eksplisit pada studi prediksi ambang seleksi perguruan tinggi (Florenzano Hernández et al., 2026; Li et al., 2023).

METODE PENELITIAN

Sumber Data

Satu model LinearRegression (tahun $\rightarrow$ nilai minimum) di-fit secara terpisah untuk setiap seri. Pendekatan model per-seri (bukan satu model global untuk seluruh data) dipilih karena setiap program studi memiliki level dan tren yang sangat berbeda (rentang nilai minimum pada data ini sekitar 160–730) — model global.

Data yang digunakan adalah nilai minimum kelulusan seleksi masuk perguruan tinggi jalur SNBT (Seleksi Nasional Berbasis Tes), tersedia per program studi dan jenjang (S1/D3), untuk tahun 2023–2026. Data bersifat agregat per program studi per tahun (bukan skor individual calon peserta).

Unit analisis adalah seri: satu kombinasi program studi $\times$ jenjang. Terdapat 50 seri pada jalur SNBT (48 program studi, dua di antaranya memiliki varian jenjang S1 dan D3 sekaligus). Distribusi jumlah titik tahun training (2023, 2024, 2025) yang tersedia tidak seragam antar seri:

Tabel 1.
Distribusi jumlah titik tahun training per seri (n = 50 seri)

# titik training tersedia	# seri
3	47
2	1
1	1
0	1

Tiga seri dikecualikan dari evaluasi, dengan alasan yang berbeda-beda (tidak semua terkait jumlah titik training):

Tabel 2.
Seri yang dikecualikan

Seri	# titik training	Alasan eksklusi
Administrasi Pajak (D3)	0	Tidak memiliki data historis sama sekali untuk dilatih
Keperawatan (D3)	1	Tidak memiliki data aktual tahun validasi untuk diuji
Perpajakan (D3)	3	Tidak memiliki data aktual tahun validasi untuk diuji dikecualikan bukan karena kekurangan data training, melainkan semata karena tidak ada data aktual sebagai pembanding

Dengan demikian, 47 dari 50 seri masuk evaluasi akhir. Dari 47 seri tersebut, 46 memiliki tiga titik training penuh dan satu seri (STATISTIKA, S1) hanya memiliki dua titik training — tetap disertakan dalam evaluasi karena regresi linear tetap dapat di-fit dengan dua titik

data, mengikuti aturan eligibilitas: seri dengan ≥ 2 titik training diproyeksikan dengan LinearRegression, seri dengan tepat 1 titik menggunakan proyeksi naif (nilai tahun tersebut dibawa apa adanya), dan seri dengan 0 titik tidak dapat diproyeksikan sama sekali. Tidak ada seri dalam evaluasi akhir yang menggunakan proyeksi naif satu-titik, karena satu-satunya seri berkategori itu (KEPERAWATAN, D3) justru dikecualikan akibat tidak memiliki data aktual validasi.

Perbedaan jumlah titik training ini relevan untuk interpretasi hasil pada bagian Hasil dan Pembahasan, khususnya terkait analisis pola arah tren, yang secara definisi hanya berlaku untuk seri dengan tiga titik training penuh.

Desain Evaluasi (Out-of-Sample)

- Data latih (training): nilai minimum tahun 2023, 2024, dan 2025 (sejumlah titik yang tersedia per seri, lihat Tabel 1).
- Data uji (validasi): nilai minimum aktual tahun berjalan berikutnya, bukan hasil proyeksi internal, sehingga evaluasi bersifat out-of-sample secara genuin terhadap data resmi.

Metode 1 – Regresi Linear per Seri

Satu model LinearRegression (tahun $\rightarrow$ nilai minimum) di-fit secara terpisah untuk setiap seri. Pendekatan model per-seri (bukan satu model global untuk seluruh data) dipilih karena setiap program studi memiliki level dan tren yang sangat berbeda (rentang nilai minimum pada data ini sekitar 160–730) model global akan membutuhkan program studi sebagai fitur kategorikal dengan hanya tiga observasi per kategori, terlalu sedikit untuk generalisasi yang berarti dibanding fit langsung per seri.

Metode 2 – Shrinkage-to-Mean

Metode kedua secara eksplisit meredam ekstrapolasi tren tajam dengan memadukan prediksi linear dan rata-rata historis, ditimbang oleh kualitas kecocokan (goodness-of-fit) tren linear itu sendiri:

$$\text{predicted} = r^2 \cdot \text{prediksi_linear} + (1 - r^2) \cdot \text{rata_rata_historis} \quad (1)$$

di mana r^2 adalah koefisien determinasi (R^2) in-sample dari hasil fit LinearRegression seri tersebut (diklip ke rentang $[0, 1]$). Tren yang rapi (R^2 tinggi, titik training nyaris membentuk garis lurus) tetap dipercaya penuh; tren yang tidak rapi (R^2 rendah) ditarik ke arah rata-rata historis, alih-alih diekstrapolasi secara agresif. Kasus tepi (seluruh titik training bernilai

identik, sehingga varians nol) ditangani secara eksplisit sebagai $R^2=1$ dengan prediksi sama dengan nilai tersebut, untuk menghindari ketidakpastian pembagian oleh nol.

Skema Evaluasi Klasifikasi

Karena data sumber hanya berisi nilai minimum agregat (bukan skor peserta individual), tidak tersedia label lolos/tidak-lolos yang sesungguhnya untuk membangun confusion matrix secara langsung. Skema evaluasi klasifikasi karena itu dibangun melalui grid skor sintetis: untuk setiap seri yang tervalidasi, diuji lima titik skor sintetis di sekitar ambang aktual tahun validasi:

$$skor_sintetis = actual_validasi + \{-30, -15, 0, +15, +30\}$$

$$prediksi_lolos = skor_sintetis \geq predicted_validasi$$

$$aktual_lolos = skor_sintetis \geq actual_validasi$$

Dengan 47 seri $\times$ 5 titik, diperoleh 235 titik evaluasi untuk setiap metode (regresi linear dan shrinkage-to-mean).

Alasan pemilihan rasional offset ± 15 dan ± 30 ini agar evaluasi berfokus pada calon peserta yang skornya dekat batas kelulusan pada kasus yang paling relevan untuk fitur pendukung keputusan semacam ini, karena skor yang jauh di atas/bawah ambang jarang membutuhkan "prediksi" sama sekali. Pemilihan besaran offset ini didukung oleh distribusi perubahan nilai minimum tahun-ke-tahun pada data training itu sendiri yang dihitung dari seluruh perubahan tahun-ke-tahun pada 47 seri berdata tiga-titik (94 nilai perubahan), median perubahan absolut adalah 22,17 poin, dengan kuartil pertama–ketiga 13,25–38,78 poin. Offset ± 15 berada dekat kuartil pertama, dan ± 30 berada di antara median dan kuartil ketiga, sehingga rentang offset yang dipilih sepadan dengan skala pergerakan tahunan yang sesungguhnya teramati pada data ini, bukan angka yang dipilih sembarang.

Mekanisme penanganan skor tidak masuk akal dengan melihat pada ambang nilai aktual terendah pada seluruh 47 seri yang dievaluasi adalah 213,09. Dengan offset terendah (-30), skor sintetis terendah yang mungkin dihasilkan adalah 183,09 ini masih berada pada rentang skor yang masuk akal secara empiris (tidak ada skor sintetis yang jatuh ke nilai negatif atau di luar rentang skor yang teramati pada data ini).

Metrik Evaluasi

- Regresi: Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) antara ambang prediksi dan ambang aktual, dihitung langsung pada skala nilai minimum.
- Klasifikasi: akurasi keseluruhan, confusion matrix, serta precision dan recall per kelas (lolos/tidak-lolos), dihitung dari 235 titik evaluasi grid skor sintetis.

Definisi Seri Bertren Monoton dan Identifikasi Pola Mean-Reversion

Untuk menguji apakah pola kegagalan berkaitan dengan karakteristik tren training, setiap seri dengan tiga titik training penuh diklasifikasikan berdasarkan arah dua segmen pergerakan tahun-ke-tahun:

$$d1 = \text{nilai}(\text{tahun ke-2}) - \text{nilai}(\text{tahun ke-1})$$

$$d2 = \text{nilai}(\text{tahun ke-3}) - \text{nilai}(\text{tahun ke-2})$$

turun_monoton jika $d1 < 0$ dan $d2 < 0$

naik_monoton jika $d1 > 0$ dan $d2 > 0$

reversal untuk kasus lainnya (arah berbalik di antara titik training)

Klasifikasi ini murni dihitung dari data training (tidak menyentuh data aktual tahun validasi), sehingga secara prinsip dapat dihitung sebelum hasil aktual diketahui, sifat yang relevan untuk digunakan sebagai mekanisme deteksi risiko dini (lihat bagian Hasil dan Pembahasan). Klasifikasi ini hanya terdefinisi untuk seri dengan tepat tiga titik training; seri dengan kurang dari tiga titik (lihat bagian Sumber Data) tidak dapat diklasifikasi arah trennya dan karenanya dikecualikan khusus dari analisis pola arah tren, meski tetap disertakan dalam evaluasi akurasi utama.

HASIL DAN PEMBAHASAN

Hasil Rgresi Linear (Baseline)

Tabel 3.
Error regresi mentah (47 seri)

Metrik	Nilai
MAE	39,40
RMSE	52,34

Skala nilai minimum pada data ini berkisar ~160–730, sehingga error rata-rata setara dengan ~7–11% dari rentang skala. Distribusi absolute error: median 30,80, standar deviasi 34,82, minimum 1,89, maksimum 152,87.

Tabel 4.

Confusion matrix — regresi linear (n = 235 titik evaluasi)

Prediksi: tidak lolos		Prediksi: lolos	
Aktual: tidak lolos	51	43	
Aktual: lolos	24	117	
Metrik	Kelas “tidak lolos”	Kelas “lolos”	Keseluruhan
Precision	0,680	0,731	-
Recall	0,543	0,830	-
F1-score	0,604	0,777	-
Akurasi	-	-	0,715

Recall kelas “tidak lolos” (0,543) jauh lebih rendah dibanding recall kelas “lolos” (0,830) yang artinya model baseline secara sistematis cenderung optimis. Model akan lebih sering salah memprediksi “lolos” untuk skor yang sebenarnya berada di bawah ambang, dibanding sebaliknya.

Hasil Shrinkage-to-Mean dengan Perbandingan Langsung

Tabel 5.

Baseline vs shrinkage-to-mean (47 seri, skema evaluasi identik)

Metrik	Regresi linear (baseline)	Shrinkage-to-mean
MAE	39,40	31,91
RMSE	52,34	43,58
Akurasi	0,715	0,745
Recall “tidak lolos”	0,543	0,585
Recall “lolos”	0,830	0,851

Shrinkage-to-mean unggul pada seluruh metrik yang diukur, termasuk recall “tidak lolos” yang menjadi perhatian utama pada baseline, meskipun perbaikannya tergolong sedang (0,543 → 0,585), bukan menghilangkan bias tersebut sepenuhnya. R^2 rata-rata pada seluruh 47 seri adalah 0,479 (std 0,354, rentang 0,0001–1,0000) — variasi besar antar seri, mengonfirmasi memang terdapat campuran tren yang rapi dan tren yang berantakan pada data ini.

Analisis Pola Kegagalan: Mean-Reversion pada Seri Bertren Monoton

Untuk menguji apakah pola kegagalan bersifat sistematis (bukan hanya anekdot dari kasus dengan error terbesar), 46 dari 47 seri evaluasi diklasifikasikan menurut arah tren training (satu seri dikecualikan dari analisis ini karena hanya memiliki dua titik training, lihat bagian Metode Penelitian - Sumber Data).

Tabel 6.

Signed error per arah tren training (n = 46 seri)

Arah tren	n	Mean <i>abs_error</i> (baseline)	Mean <i>abs_error</i> (shrinkage)	Konsistensi arah
turun_monoton	10	63,25	58,31	100% (10/10) under-predicted di kedua metode
naik_monoton	3	67,03	49,23	100% (3/3) over-predicted di kedua metode
reversal	33	29,77	22,22	79–82% (arah campur, bukan pola berarah)

Seluruh 13 seri dengan tren monoton pada training (10 turun-monoton, 3 naik-monoton), pada kedua metode evaluasi, meleset ke arah yang persis diprediksi hipotesis mean-reversion, ini nol pengecualian. Kelompok reversal (arah berbalik dalam training) sebaliknya menunjukkan arah yang campur (79–82%), konsisten dengan pola tersebut merupakan derau biasa, bukan bias sistematis. Mean absolute error kelompok monoton juga $2,0\text{--}2,3\times$ lebih tinggi dibanding kelompok reversal.

Tabel 7.

Contoh Seri Representatif

Seri	Arah Tren	Y(T-2)	Y(T-1)	Y(T)	Prediksi	Aktual	Error
Pendidikan Ipa (S1)	Turun_Monoton	463,98	421,01	333,71	275,96	428,83	–152,87 (Error Terbesar Di Dataset)
Pendidikan Guru							
Pendidikan Anak Usia Dini (S1)	Turun_Monoton	426,28	325,50	300,60	225,11	363,66	–138,55
Akuntansi (D3)	Naik_Monoton	363,14	510,42	529,09	633,50	526,65	+106,85

Pada ketiga contoh ini, tren training yang tampak konsisten (turun terus atau naik terus selama tiga tahun) ternyata dikoreksi tajam pada tahun validasi, persis pola yang diprediksi hipotesis mean-reversion, dan tidak terdeteksi oleh regresi linear murni yang mengasumsikan tren akan berlanjut.

Shrinkage-to-mean Membantu tapi Tidak Menghilangkan Bias

Shrinkage-to-mean mengurangi besaran (magnitude) error pada kelompok tren monoton ($63,25 \rightarrow 58,31$ dan $67,03 \rightarrow 49,23$), namun tidak mengubah arah bias sama sekali. Hal ini dapat terlihat dari persentase kesesuaian dengan arah hipotesis tetap 100% dari 100% pada kedua metode. Dengan kata lain, shrinkage meredam seberapa jauh ekstrapolasi menyimpang, bukan menghilangkan arah kesalahannya.

Kasus PENDIDIKAN IPA (S1) (*turun_monoton*, error terbesar pada baseline, 152,87) hanya membaik tipis menjadi 148,03 pada shrinkage, karena R^2 -nya 0,963 tren tiga titiknya memang rapi secara linear/monoton, sehingga shrinkage “mempercayai” tren tersebut dan nyaris tidak menariknya ke rata-rata historis, padahal nilai aktualnya rebound jauh dari ekstrapolasi tersebut. Temuan ini mengungkap batas metode shrinkage: R^2 mengukur kerapian garis, bukan risiko pembalikan arah, sehingga tren yang mulus secara statistik pun dapat salah total apabila penyebabnya adalah faktor eksternal yang tidak tertangkap tiga titik data historis semata.

Kasus degenerate: R^2 tidak informatif untuk seri berdata dua-titik. Satu seri di antara 47 yang dievaluasi (STATISTIKA, S1, lihat bagian Metode Penelitian) hanya memiliki dua titik data training. Pada kasus ini, R^2 in-sample dari regresi linear selalu bernilai tepat 1,0 secara matematis, garis dengan dua parameter (kemiringan dan intersep) pasti melewati tepat dua titik data yang tersedia, terlepas dari keandalan tren yang sesungguhnya. Akibatnya, metode shrinkage-to-mean memberi bobot penuh pada prediksi linear untuk seri ini (prediksi shrinkage identik dengan prediksi baseline, tidak ada penyusutan ke rata-rata historis sama sekali), justru pada seri dengan data historis paling sedikit di seluruh dataset, yang semestinya paling membutuhkan mitigasi ekstrapolasi. Pada evaluasi ini, dampaknya tidak ekstrem (absolute error seri ini berada tepat di atas median, peringkat pertengahan dari seluruh seri yang dievaluasi), namun ini merupakan keterbatasan struktural metode: R^2 tidak informatif untuk seri dengan tepat dua titik training, dan penerapan metode shrinkage semacam ini pada data dengan seri berdata historis pendek perlu mewaspadaai kasus degenerate ini secara eksplisit, alih-alih mengasumsikan R^2 tinggi selalu berarti tren yang benar-benar dapat diandalkan.

Mekanisme Deteksi Risiko Dini (*is_high_risk_series*)

Temuan pola mean-reversion yang konsisten pada seri bertren monoton mendorong perumusan mekanisme deteksi risiko yang dapat dihitung sebelum hasil aktual diketahui murni dari tiga titik data training. Dua kandidat definisi dievaluasi sebagai flag biner terhadap target “*high_error*” (kuartil-4 absolute error, base rate 0,261 pada $n=46$).

Tabel 8.
Perbandingan dua kandidat definisi flag risiko

Definisi	Logika	Precision (baseline / shrinkage)	Recall (baseline / shrinkage)	Kebutuhan kalibrasi
Arah tren monoton (<i>dipilih</i>)	turun_monoton ATAU naik_monoton (murni kategorikal, tanpa parameter ambang)	0,385 / 0,462	0,417 / 0,500	Tidak ada, selalu terdefinisi dari 3 titik data mana pun
Magnitude pergerakan (<i>swing</i>)	$ d1 + d2 \geq 77,4$ (ambang kuartil-4 dari distribusi <i>swing</i> pada data ini)	0,667 / 0,667	0,667 / 0,667	Perlu, ambang 77,4 dikalibrasi khusus pada n=46 seri ini, berpotensi perlu direkalibrasi untuk data/lingkup lain

Kedua definisi jelas di atas base rate (0,261), namun dengan trade-off yang berbeda. Definisi berbasis arah tren monoton dipilih sebagai mekanisme utama karena tidak membutuhkan parameter ambang yang harus dikalibrasi ulang — logikanya murni kategorikal (dua segmen pergerakan searah atau tidak) dan karenanya tetap terdefinisi secara konsisten pada data/lingkup mana pun di masa depan tanpa perlu penyesuaian. Namun demikian, daya diskriminasinya (precision 0,385–0,462) jauh lebih rendah dibanding definisi berbasis magnitude pergerakan (*swing*), yang mencapai precision/recall 0,667/0,667 — signifikan lebih baik, tetapi bergantung pada ambang numerik yang bersifat spesifik terhadap sebaran data n=46 pada penelitian ini, dan berisiko tidak transferabel apabila diterapkan pada lingkup data yang berbeda tanpa rekalisasi ulang.

Cara kerja mekanisme yang dipilih: untuk setiap seri, dihitung dua segmen pergerakan tahun-ke-tahun pada data training ($d1, d2$ – lihat bagian Metode Penelitian). Apabila kedua segmen bergerak searah (keduanya turun, atau keduanya naik), seri tersebut ditandai berisiko tinggi mengalami mean-reversion pada tahun prediksi. Mekanisme ini dimaksudkan untuk melunakkan output sistem pendukung keputusan bagi seri yang ditandai (misalnya dengan memperlebar pita ketidakpastian atau menambahkan disclaimer eksplisit), bukan untuk menolak memberikan proyeksi atau menyembunyikan pilihan tertentu dari pengguna. Karena logikanya dihitung murni dari data historis apa pun (bukan daftar nama kategori yang di-hardcode), mekanisme ini otomatis berlaku untuk seri/periode mana pun ke depan tanpa perlu pembaruan manual.

Perlu dicatat secara eksplisit: mekanisme deteksi risiko ini membantu mengurangi kasus terburuk, namun tidak menyelesaikan masalah recall “tidak lolos” yang rendah secara keseluruhan, mayoritas error tersebar pada seri dengan error sedang yang tidak akan

tertangkap oleh flag mana pun yang presisinya di bawah 1,0. Ini merupakan mitigasi tambahan, bukan pengganti perbaikan model.

Keterbatasan Penelitian

- Data agregat, bukan data individual. Confusion matrix dibangun dari grid skor sintetis di sekitar ambang, bukan dari skor peserta riil, sehingga angka akurasi/precision/recall mengukur “seberapa dekat prediksi ambang terhadap ambang aktual”, bukan “seberapa akurat prediksi terhadap peserta riil”. Apabila distribusi skor peserta riil menumpuk jauh dari ambang (tidak tersebar merata di kisaran ± 30), performa efektif di lapangan dapat berbeda dari angka yang dilaporkan di sini.
- Jumlah titik training kecil per seri (maksimal tiga titik tahun), regresi linear dengan tiga titik pada dasarnya hanya menghitung kemiringan dari dua segmen, sangat sensitif terhadap nilai ekstrem satu tahun. Ini bukan sekadar risiko teoretis: seluruh 13 seri dengan tren monoton pada training meleset 100% ke arah mean-reversion pada tahun validasi, pola yang berlaku umum pada data ini, bukan 1–2 kasus kebetulan.
- R^2 tidak informatif untuk seri berdata dua-titik, dibahas secara rinci di bagian Diskusi di atas; metode shrinkage tidak memberi perlindungan apa pun pada kasus ini karena R^2 selalu bernilai 1,0 secara matematis untuk regresi dua-titik.
- Kausalitas pola mean-reversion belum teruji. Penelitian ini membuktikan pola statistiknya (tren monoton $\rightarrow$ error berarah, konsisten 100%), namun penyebabnya (misalnya dugaan mengenai dinamika jumlah pendaftar dari tahun ke tahun) tidak dapat diuji dengan data yang tersedia, karena tidak ada kolom jumlah pendaftar/peminat/kuota per seri per tahun. Klaim kausal semacam itu tetap berstatus hipotesis yang belum diverifikasi.
- Offset grid (± 15 , ± 30), meski didukung oleh distribusi pergerakan tahun-ke-tahun pada data ini (bagian Metode Penelitian), tetap merupakan pilihan desain yang belum dikalibrasi terhadap sebaran skor riil calon peserta.
- Ruang lingkup hanya jalur SNBT, jalur seleksi lain tidak tercakup dalam penelitian ini. Generalisasi ke jalur seleksi lain (terutama yang ambangnya lebih dipengaruhi kebijakan institusional dibanding murni tren nilai ujian) membutuhkan validasi terpisah.
- Ukuran sampel 47 seri (46 untuk analisis pola arah tren), cukup untuk gambaran umum akurasi model secara agregat, namun terbatas untuk kesimpulan per-seri individual.

KESIMPULAN DAN REKOMENDASI

Penelitian ini mengevaluasi secara out-of-sample genuin akurasi dua pendekatan, yaitu regresi linear per seri dan shrinkage-to-mean. dalam memproyeksikan ambang nilai minimum penerimaan jalur SNBT dari tren tiga tahun sebelumnya. Tiga kontribusi utama dihasilkan:

1. Evaluasi yang diuji langsung terhadap data ambang aktual, bukan data buatan atau proyeksi internal;
2. Identifikasi pola kegagalan mean-reversion yang terbukti berlaku universal pada seluruh seri bertren monoton pada data ini, bukan sekadar anekdot;
3. Perumusan mekanisme deteksi risiko dini berbasis data training saja, disertai pemetaan eksplisit trade-off antara dua kandidat definisi yang tersedia.

Terhadap tujuan penelitian, hasil menunjukkan bahwa regresi linear murni memiliki nilai prediktif yang bermakna namun mengandung bias sistematis yang cukup besar, hal ini berarti model cenderung optimis dan gagal secara konsisten pada seri dengan tren training yang tampak “rapi” namun sesungguhnya rentan koreksi arah. Metode shrinkage-to-mean memperbaiki akurasi pada hampir seluruh metrik yang diukur dengan perubahan implementasi yang sederhana, namun tidak mengubah arah bias yang teridentifikasi, dan memiliki keterbatasan tersendiri pada seri dengan data historis yang sangat terbatas. Pola kegagalan yang teridentifikasi ternyata dapat dideteksi sejak dini menggunakan data training saja, meski dengan daya diskriminasi yang moderat.

Penulis menyampaikan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Sultan Ageng Tirtayasa atas pendanaan kegiatan ini melalui Skema Penelitian Dosen Pemula (PDP) tahun 2026. Apresiasi juga disampaikan kepada Unit Penunjang Akademik Teknologi Informasi dan Komunikasi (UPA TIK), Divisi PMB, atas bantuan dalam penyediaan data, serta kepada rekan-rekan yang telah memberikan dukungan dan saran berharga sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

1. Ahmed, S. E., & Raheem, S. E. (2012). Shrinkage and absolute penalty estimation in linear regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(6), 541–553. <https://doi.org/10.1002/wics.1232>

2. Arnold, I. J. M. (2015). Changing the guard: Staff turnover as a source of variation in test results. *Studies in Educational Evaluation*, 47, 12–18.
<https://doi.org/10.1016/j.stueduc.2015.05.002>
3. Florenzano Hernández, V., Alarcón-Bustamante, E., Varas, I. M., & Godoy, M. I. (2026). Beyond a Single Coefficient: A Model-Based Framework for Local Predictive Validity along the Score Scale. *Journal of Educational Measurement*, 63(3).
<https://doi.org/10.1111/jedm.70052>
4. Iwata, T., Kumagai, A., & Ida, Y. (2025). Meta-learning Task-specific Regularization Weights for Few-shot Linear Regression. *Proceedings of Machine Learning Research*, 258, 3052–3060.
<https://www.scopus.com/pages/publications/105014316105?origin=resultslist>
5. Khan, M., & Olivier, J. (2023). Regression to the mean: Estimation and adjustment under the bivariate normal distribution. *Communications in Statistics - Theory and Methods*, 52(19), 6972–6990. <https://doi.org/10.1080/03610926.2022.2037645>
6. Li, K. C., Wong, B. T.-M., & Chan, H. T. (2023). Predictive Analytics for University Student Admission: A Literature Review. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13978 LNCS, 250–259. https://doi.org/10.1007/978-3-031-35731-2_22
7. Özkale, M. R., & Mammadova, U. (2026). Performance analysis of shrinkage estimators in Conway-Maxwell-Poisson regression model. *Communications in Statistics - Theory and Methods*, 55(1), 63–91. <https://doi.org/10.1080/03610926.2025.2484449>
8. Qattan, G. A., Anwer, M. A., & Ali, A. M. (2023). Selecting Minimum Acceptable Student's Mark to Participate in Bologna Final Exam Using Machine Learning Approach. *Zanco Journal of Pure and Applied Sciences*, 35(4), 97–103.
<https://doi.org/10.21271/ZJPAS.35.4.09>
9. Singh, H. P., Vishwakarma, G. K., Joshi, H., & Gupta, S. (2024). Shrinkage estimation for square of location parameter of the exponential distribution with known coefficient of variation. *Journal of Computational and Applied Mathematics*, 437.
<https://doi.org/10.1016/j.cam.2023.115489>