

تحلیل مهاجرت استعدادهای فنی جهانی

۱) مقدمه و هدف پروژه

در این پروژه هدف ما پیش‌بینی متغیر **Migration_Status** است؛ یعنی تشخیص اینکه آیا فرد مهاجرت کاری انجام می‌دهد یا نه. داده شامل ۵۰ هزار رکورد از متخصصان فنی است و متغیرهایی مثل فعالیت گیت‌هاب، تعداد استناد پژوهشی، سابقه صنعتی، و ویژگی‌های دسته‌ای مانند سطح تحصیلات دارد. نکته کلیدی این پروژه این است که فقط «یک مدل با عدد خوب» نمی‌خواهیم. ما باید یک سیستم تصمیم‌یار قابل اعتماد بسازیم که:

- از نظر فنی درست باشد (leakage-safe)
- از نظر آماری معتبر باشد (استنباط و کالیبراسیون)
- از نظر عملیاتی قابل پایش باشد (temporal robustness و drift)
- از نظر اخلاقی و عدالت قابل دفاع باشد

۲) پرسش‌های اصلی پروژه

این گزارش به چهار سؤال محوری جواب می‌دهد:

- آیا مدل نهایی بدون نشت داده و با کیفیت قابل قبول ساخته شده؟
- آیا احتمال‌های مدل قابل اعتماد هستند (کالیبره‌اند)؟
- آیا مدل در زمان و زیر drift پایدار می‌ماند؟
- آیا می‌توان عدالت را با افت کارایی قابل قبول بهتر کرد؟

۳) تحلیل داده و تصاویر اصلی (EDA + Diagnostics)

در این بخش، هر تصویر دقیق توضیح داده می‌شود.

تصویر ۱ – توزیع کلاس هدف (**Migration_Status**)

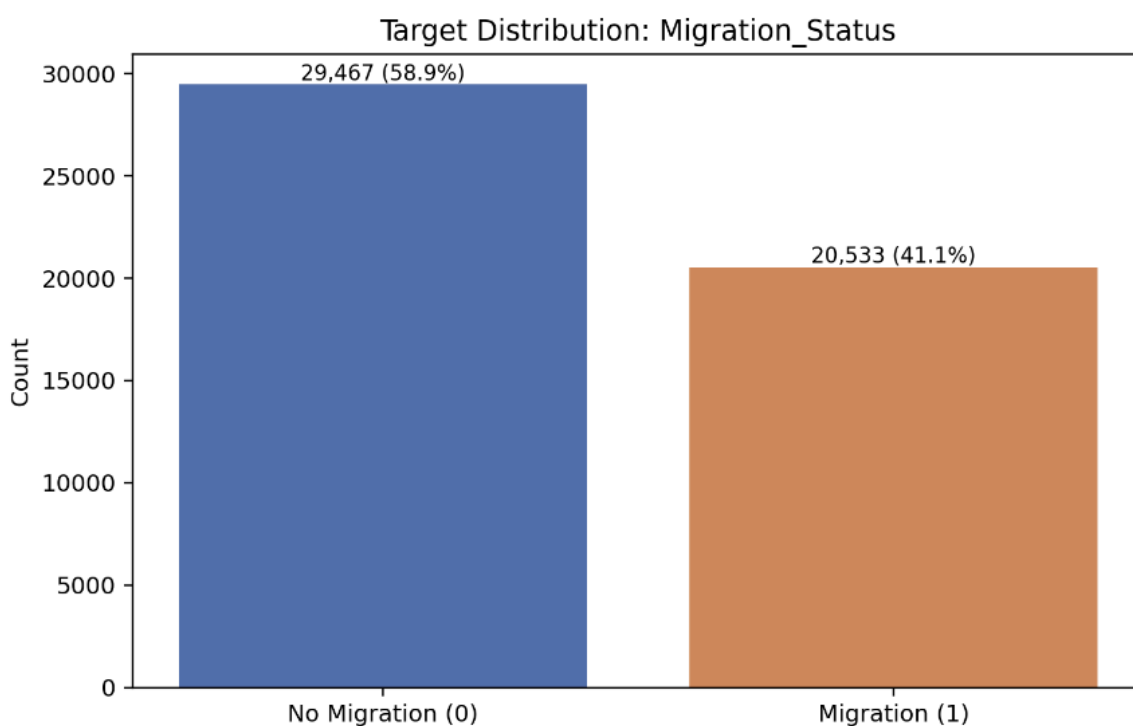


Fig. 1. Class distribution for Migration_Status.

این تصویر چه چیزی نشان می‌دهد؟

نمودار تعداد/درصد کلاس‌های ۰ و ۱ را نشان می‌دهد. اگر یکی از کلاس‌ها به‌وضوح غالب باشد، داده نامتوازن است.

چطور باید خوانده شود؟

- اگر فاصله ستون‌ها کم باشد → توازن خوب
- اگر یک ستون خیلی بلندتر باشد → نامتوازنی معنادار

نتیجه تحلیلی

در این پروژه توازن کامل نیست و کلاس هدف «نامتوازن متوسط» است. یعنی اگر فقط Accuracy گزارش کنیم ممکن است گمراه‌کننده باشد.

اثر بر تصمیم مدل‌سازی

- معیارهای اصلی باید: **ROC-AUC**، **F1**، **Precision**، **Recall** باشد.
- در مرحله آستانه‌گذاری، هزینه FN/FP باید جداگانه دیده شود.

محدودیت

نسبت کلاس در داده تاریخی ممکن است در داده آینده عوض شود؛ پس همین نسبت را ثابت فرض نمی‌کنیم.

تصویر ۲ – پروفایل گمشده‌ها (Top Missingness)

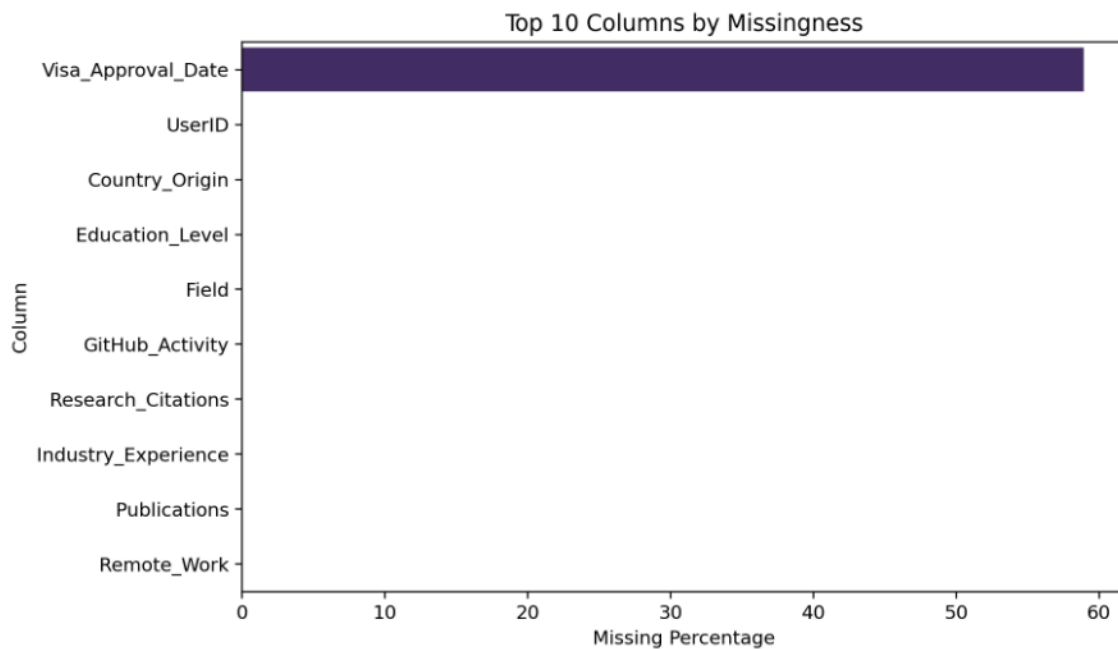


Fig. 2. Top-10 missingness rates across columns.

این تصویر چه چیزی نشان می‌دهد؟

درصد مقادیر گمشده برای ستون‌هایی که بیشترین missingness را دارند.

چطور باید خوانده شود؟

- ستون‌هایی با missing بالا باید با دقت بیشتری بررسی شوند یا حذف شوند.
- اگر ستون‌های اداری/ویزا گمشدگی بالا دارند، احتمال نشت یا فرآیندی بودن داده زیاد است.

نتیجه تحلیلی

گمشدگی‌ها معمولاً حول ویژگی‌های عملیاتی و ویزا جمع شده‌اند. این الگو اغلب تصادفی نیست و می‌تواند به روندهای اداری پس‌از-تصمیم مرتبط باشد.

اثر بر تصمیم

- این ستون‌ها وارد ممیزی نشت می‌شوند.
- برای ایمپیوت، روش یکسان برای همه ستون‌ها کافی نیست.
- گزارشی از «دلیل انتخاب روش ایمپیوت» باید ثبت شود.

محدودیت

اگر مکانیزم گمشدگی **MNAR** باشد، ایمپیوت ساده می‌تواند بایاس ایجاد کند.

تصویر ۳ – نقشه همبستگی ویژگی‌های عددی

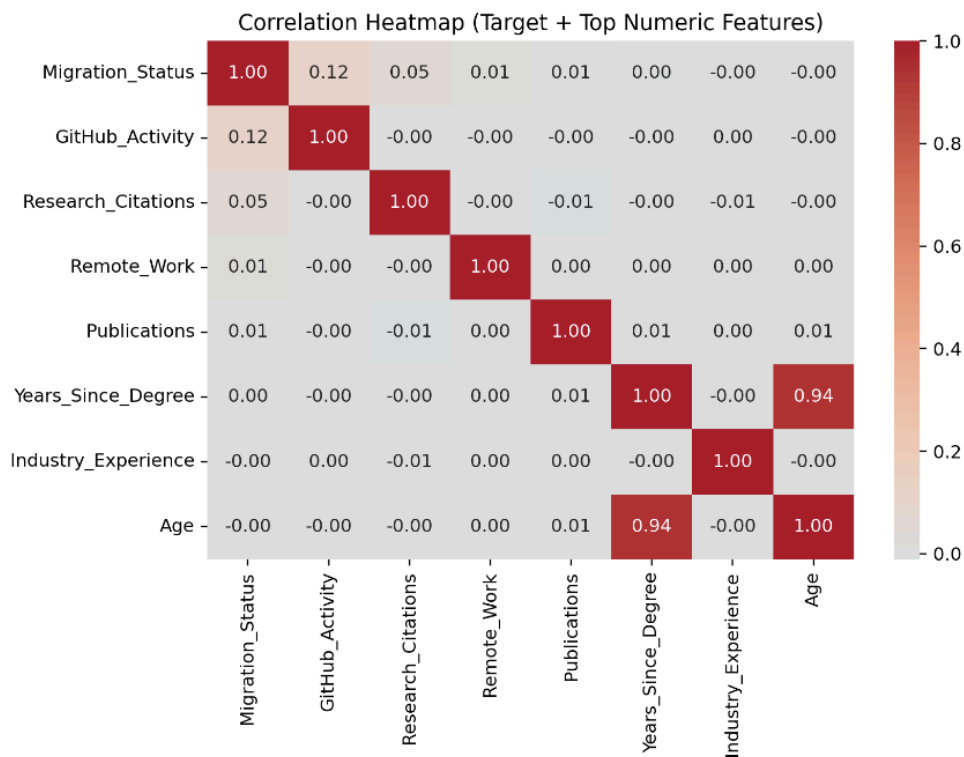


Fig. 3. Correlation heatmap for key numeric predictors and target.

این تصویر چه چیزی نشان می‌دهد؟

شدت و جهت همبستگی خطی بین ویژگی‌های عددی و همچنین ارتباطشان با هدف.

چطور باید خوانده شود؟

- رنگ/عدد نزدیک +1 یا -1 → همبستگی قوی‌تر
- نزدیک 0 → ارتباط خطی ضعیف
- بلوک‌های همبستگی بالا بین ویژگی‌ها → احتمال چندهمخطی

نتیجه تحلیلی

سیگنال‌ها بیشتر ضعیف تا متوسط هستند و یک «ابر-ویژگی» غالب نداریم. پس مدل‌های چندمتغیره و حتی غیرخطی منطقی‌اند.

اثر بر تصمیم

- استفاده از مدل‌های ensemble/boosting توجه دارد.
- در مدل خطی، regularization (مثل Elastic Net) ضروری می‌شود.

محدودیت

همبستگی \neq علیت.
ممکن است متغیر پنهان عامل اصلی باشد.

تصویر ۴ – نرخ مهاجرت در سطح کشور

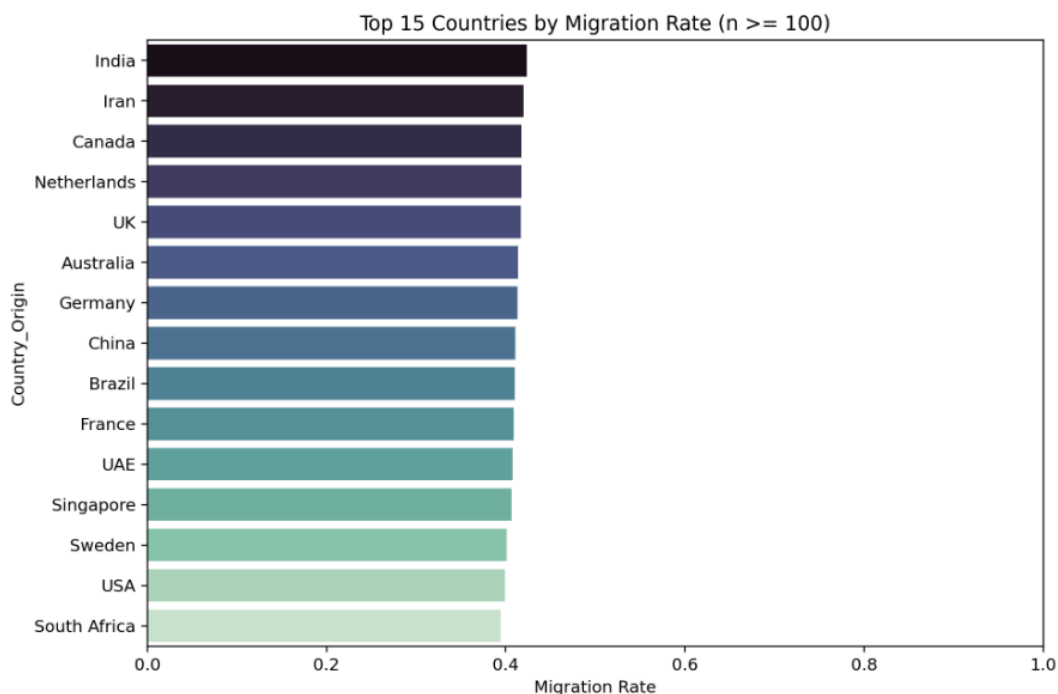


Fig. 4. Country-level migration rate ranking under minimum support filter.

این تصویر چه چیزی نشان می‌دهد؟

نرخ outcome (مهاجرت) به تفکیک کشور، معمولاً بعد از اعمال حداقل اندازه نمونه.

چطور باید خوانده شود؟

- کشورهایی با نرخ بسیار بالا/پایین را شناسایی می‌کنیم.
- اختلافات شدید، هشدار برای تحلیل عدالت و سوگیری است.

نتیجه تحلیلی

تفاوت بین کشورها معنادار است.

این تفاوت می‌تواند بازتاب سیاست ویزا، اقتصاد یا دسترسی باشد؛ نه فقط ویژگی فردی.

اثر بر تصمیم

- ارزیابی fairness به تفکیک زیرگروه‌ها اجباری است.
- تصمیم نهایی مدل بدون تحلیل subgroup نباید وارد سیاست‌گذاری شود.

محدودیت

ممکن است برخی کشورها نمونه کم داشته باشند و نرخشان ناپایدار باشد.

۴) تصاویر مدل‌سازی هسته (Q3 تا Q6)

تصویر ۵ – مسیر بهینه‌سازها روی تابع Ravine

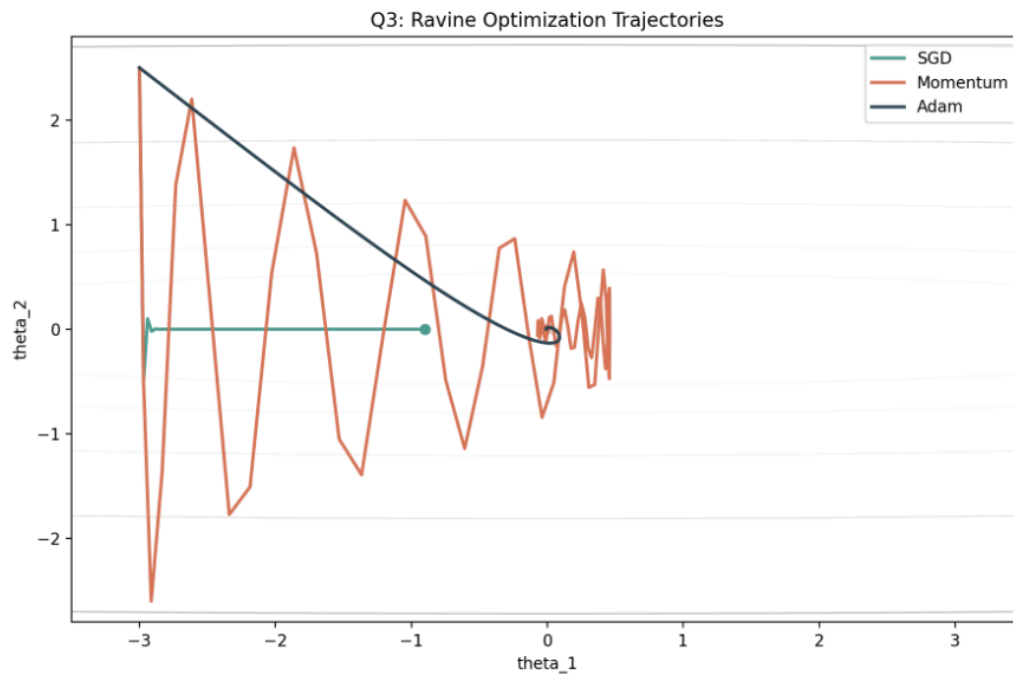


Fig. 5. SGD vs Momentum vs Adam trajectories on a ravine objective.

این تصویر چه چیزی نشان می‌دهد؟

مسیر حرکت SGD، Momentum و Adam روی یک تابع دره‌ای که در یک راستا شیب تند و در راستای دیگر شیب ملایم دارد.

چطور باید خوانده شود؟

- SGD: رفت‌وبرگشت زیاد (oscillation) در جهت شیب تند
- Momentum: حرکت نرم‌تر و سریع‌تر در جهت درست
- Adam: همگرایی پایدارتر در مقیاس‌های ناهمگن

نتیجه تحلیلی

Adam و Momentum معمولاً نسبت به SGD ساده پایدارتر و سریع‌تر همگرا می‌شوند.

اثر بر تصمیم

برای داده‌هایی با مقیاس‌های متفاوت یا فضای پارامتر بدشروط، Adam/Momentum گزینه بهتری هستند.

محدودیت

رفتار روی تابع مصنوعی ravine الزاماً عیناً به همه مسائل واقعی تعمیم نمی‌یابد.

تصویر ۶ – حساسیت SVM-RBF نسبت به گاما

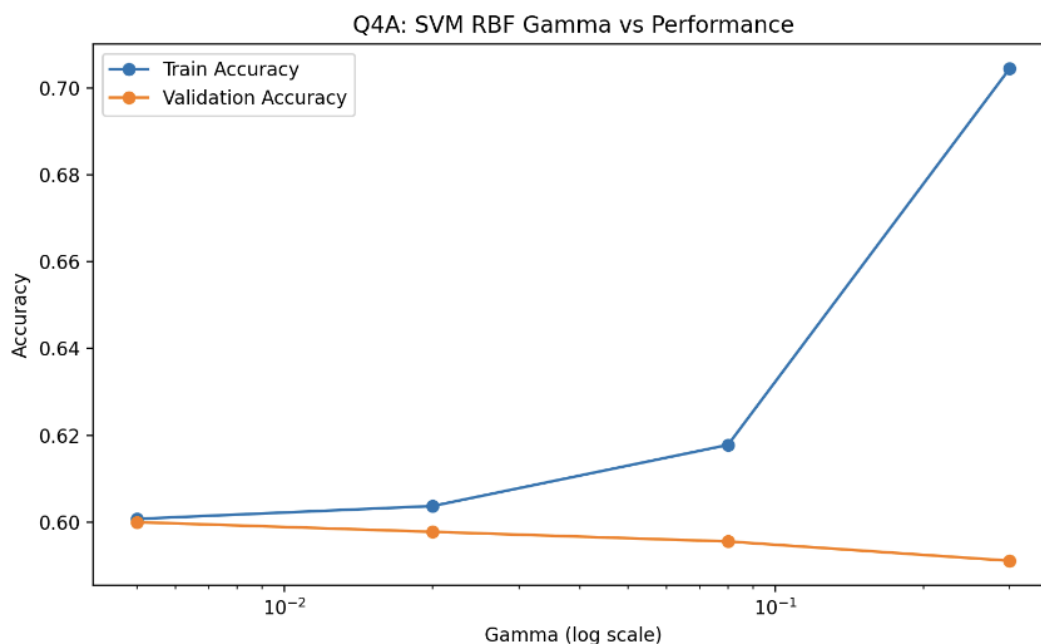


Fig. 6. Validation sensitivity to RBF kernel width (γ).

این تصویر چه چیزی نشان می‌دهد؟

تغییر عملکرد اعتبارسنجی مدل RBF-SVM در مقادیر مختلف γ .

چطور باید خوانده شود؟

- γ خیلی بزرگ \rightarrow مرز تصمیم بسیار پیچیده، ریسک overfit
- γ خیلی کوچک \rightarrow مرز بیش از حد صاف، ریسک underfit
- نقطه بهینه معمولاً وسط این بازه است.

نتیجه تحلیلی

تنظیم γ باید همراه با CCC و با CV انجام شود.

اثر بر تصمیم

اگر overfit دیدیم، اولین اقدام معمول: کاهش γ و بازتنظیم CCC.

محدودیت

حساسیت به γ وابسته به scaling ویژگی‌هاست؛ بدون scaling نتیجه قابل اعتماد نیست.

تصویر ۷ – منحنی هرس درخت تصمیم (Cost-Complexity)

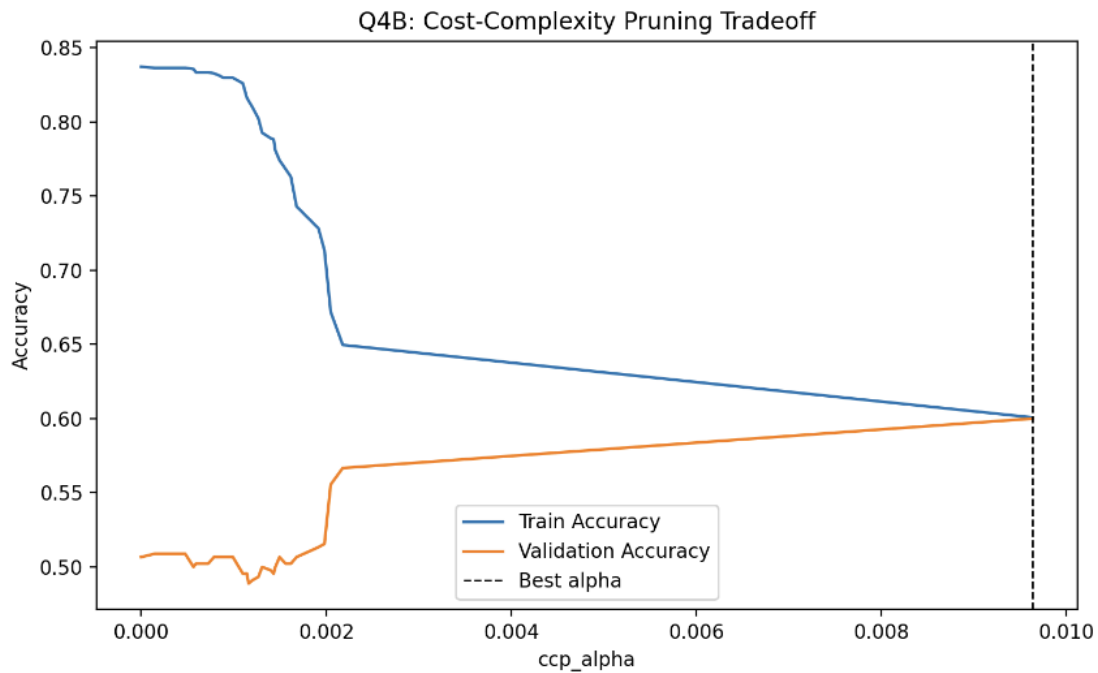


Fig. 7. Cost-complexity pruning tradeoff for CART.

این تصویر چه چیزی نشان می‌دهد؟

رابطه پارامتر هرس α با عملکرد/پیچیدگی درخت.

چطور باید خوانده شود؟

- α کوچک \rightarrow درخت بزرگ‌تر، واریانس بیشتر
- α بزرگ \rightarrow درخت کوچک‌تر، بایاس بیشتر
- نقطه مناسب جایی است که تعمیم‌پذیری بهتر شود، نه صرفاً آموزش.

نتیجه تحلیلی

هرس مناسب کمک می‌کند مدل از حفظ‌کردن نویز دور شود.

اثر بر تصمیم

انتخاب α باید با اعتبارسنجی انجام شود، نه بر اساس خطای train.

محدودیت

نقطه بهینه می‌تواند بین split های مختلف کمی جابه‌جا شود.

تصویر ۸ – K-Means در Elbow

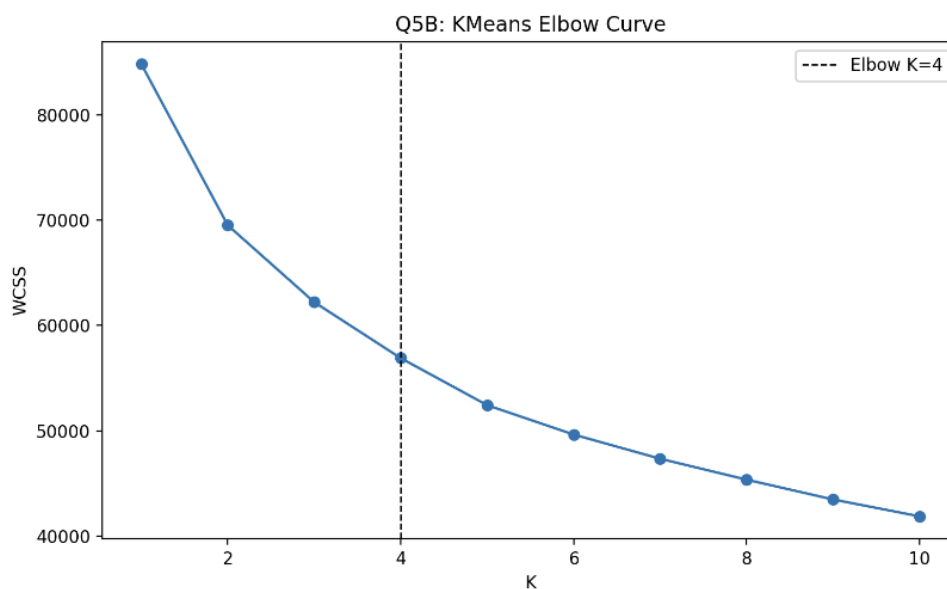


Fig. 8. WCSS elbow curve for model order selection in K-Means.

این تصویر چه چیزی نشان می‌دهد؟

کاهش WCSS با افزایش تعداد خوشه K.

چطور باید خوانده شود؟

- WCSS همیشه با افزایش KKK کم می‌شود.
- نقطه elbow جایی است که سود کاهش WCSS ناگهان کمتر می‌شود.

نتیجه تحلیلی

KKK انتخابی باید هم از نظر عددی خوب باشد و هم تفسیرپذیر.

اثر بر تصمیم

بعد از انتخاب اولیه با elbow، با silhouette و معنانشناسی خوشه‌ها تأیید نهایی می‌دهیم.

محدودیت

گاهی elbow واضح نیست؛ در آن حالت تصمیم باید با معیارهای مکمل گرفته شود.

تصویر ۹ – SHAP محلی (Local Explanation)



Fig. 9. Local SHAP explanation for a selected high-citation candidate.

این تصویر چه چیزی نشان می‌دهد؟

برای یک نمونه مشخص (مثلاً فرد با citation بالا)، سهم هر ویژگی در رساندن خروجی از `base_value` به `output_value`.

چطور باید خوانده شود؟

- ویژگی‌های مثبت خروجی را به سمت کلاس مهاجرت می‌برند.
- ویژگی‌های منفی آن را به سمت عدم مهاجرت می‌برند.
- جمع سهم‌ها = اختلاف output با base.

نتیجه تحلیلی

ممکن است با وجود citation بالا، چند عامل منفی قوی‌تر تصمیم مدل را به «عدم مهاجرت» ببرد.

اثر بر تصمیم

بازبینی انسانی می‌تواند روی همان چند عامل غالب تمرکز کند.

محدودیت

این توضیح «رفتار مدل» را توصیف می‌کند، نه علت واقعی رخداد در جهان.

تصویر ۱۰ – SHAP سراسری (Global Summary)

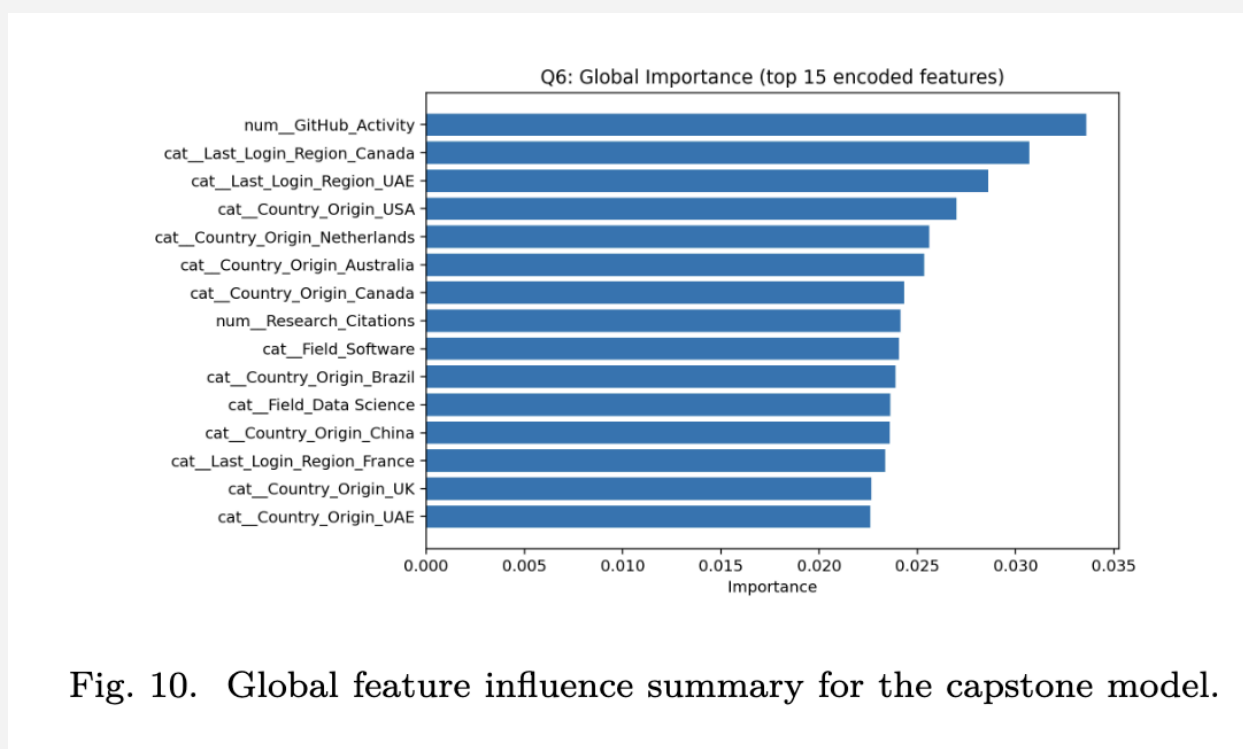


Fig. 10. Global feature influence summary for the capstone model.

این تصویر چه چیزی نشان می‌دهد؟

اهمیت و جهت اثر ویژگی‌ها در کل داده.

چطور باید خوانده شود؟

- ویژگی‌های بالاتر در نمودار → اثر کلی بیشتر

- پراکندگی رنگ/مقدار → ناهمگنی اثر

نتیجه تحلیلی

ویژگی‌های پژوهشی و فعالیت فنی معمولاً در درایورهای اصلی دیده می‌شوند.

اثر بر تصمیم

کیفیت داده و کنترل سیاستی برای ویژگی‌های top باید سخت‌گیرانه‌تر باشد.

محدودیت

یک اهمیت سراسری واحد می‌تواند تفاوت‌های زیرگروهی را پنهان کند.

۵) تصاویر بلوک تولیدمحور (Q15 تا Q20)

تصویر ۱۱ – منحنی کالیبراسیون (Reliability Curve)

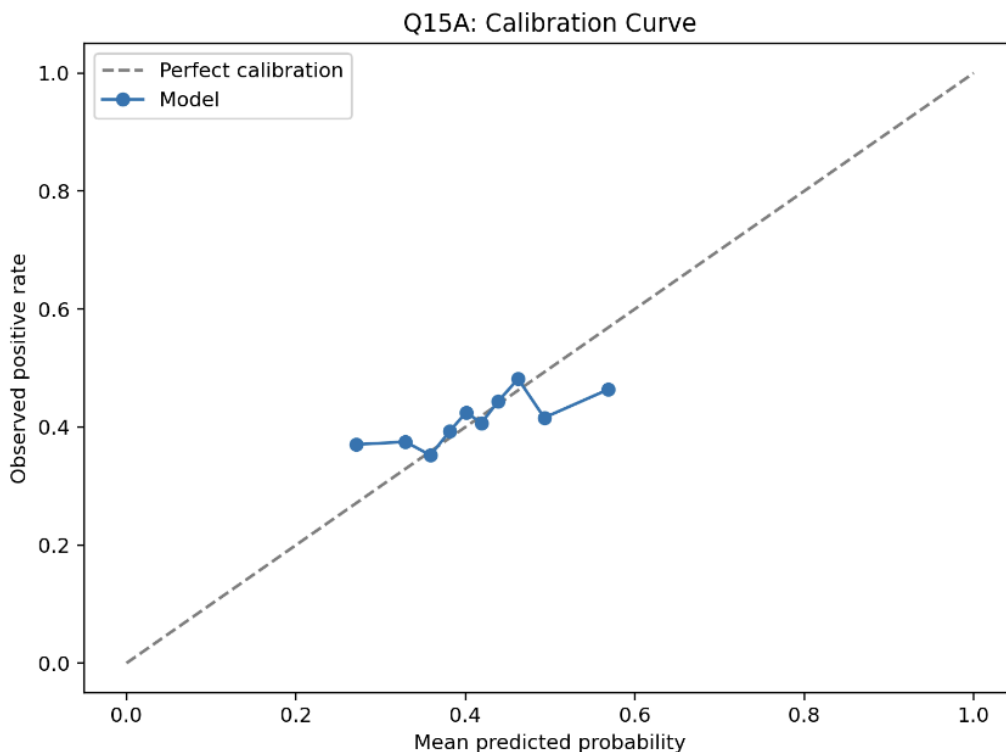


Fig. 11. Calibration reliability curve for the capstone model.

این تصویر چه چیزی نشان می‌دهد؟

رابطه بین احتمال پیش‌بینی‌شده و فراوانی واقعی رخداد.

چطور باید خوانده شود؟

- نزدیکی منحنی به خط قطر → کالیبراسیون بهتر
- انحراف زیاد → احتمال‌ها قابل اتکا نیستند

نتیجه تحلیلی

حتی مدل با AUC خوب ممکن است بد کالیبره باشد.

اثر بر تصمیم

آستانه‌گذاری سیاستی باید روی احتمال‌های کالیبره انجام شود.

محدودیت

کالیبراسیون با drift زمانی خراب می‌شود؛ بازکالیبراسیون دوره‌ای لازم است.

تصویر ۱۲ – نمودار Trade-off آستانه

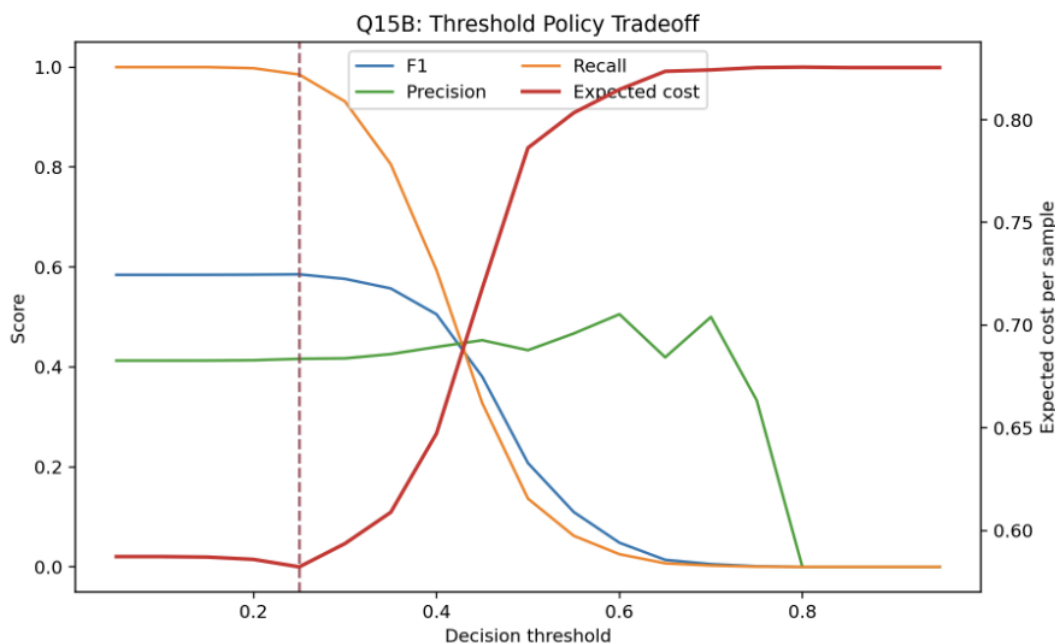


Fig. 12. Threshold tradeoff between F1 components and expected decision cost.

این تصویر چه چیزی نشان می‌دهد؟

تغییر Precision/Recall/F1 و هزینه مورد انتظار با تغییر threshold.

چطور باید خوانده شود؟

- با افزایش threshold معمولاً precision بالا و recall پایین می‌آید.
- نقطه عملیاتی باید با هزینه FN/FP انتخاب شود، نه صرفاً بیشینه F1.

نتیجه تحلیلی

یک آستانه «جهانی» برای همه سناریوها نداریم.

اثر بر تصمیم

دو آستانه تعریف می‌شود:

(۱) آستانه‌ی بهترین F1

(۲) آستانه‌ی هزینه-محور برای سیاست واقعی

محدودیت

اگر ماتریس هزینه درست تعریف نشود، آستانه انتخابی هم گمراه‌کننده می‌شود.

تصویر ۱۳ – رتبه‌بندی PSI با Drift

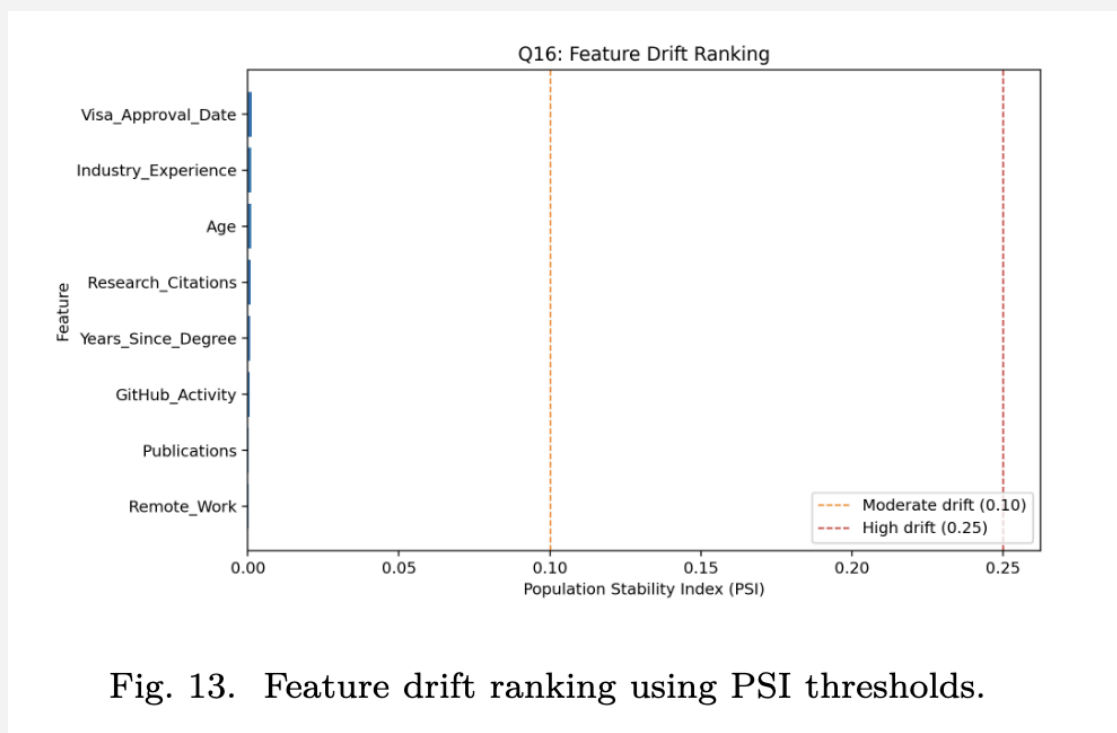


Fig. 13. Feature drift ranking using PSI thresholds.

این تصویر چه چیزی نشان می‌دهد؟

شدت تغییر توزیع ویژگی‌ها بین پنجره مرجع و پنجره جاری.

چطور باید خوانده شود؟

- PSI کم: پایدار

- PSI متوسط: هشدار

- PSI بالا: بحرانی

نتیجه تحلیلی

برخی ویژگی‌ها ناپایداری قابل توجه دارند و باید مانیتور شوند.

اثر بر تصمیم

برای ویژگی‌های بحرانی، trigger بازآموزی یا بازبینی مدل فعال می‌شود.

محدودیت

PSI فقط drift ویژگی را می‌بیند، نه لزوماً تغییر رابطه ویژگی با هدف (concept drift).

تصویر ۱۴ – تلاش ریکورس (Median Deltas)

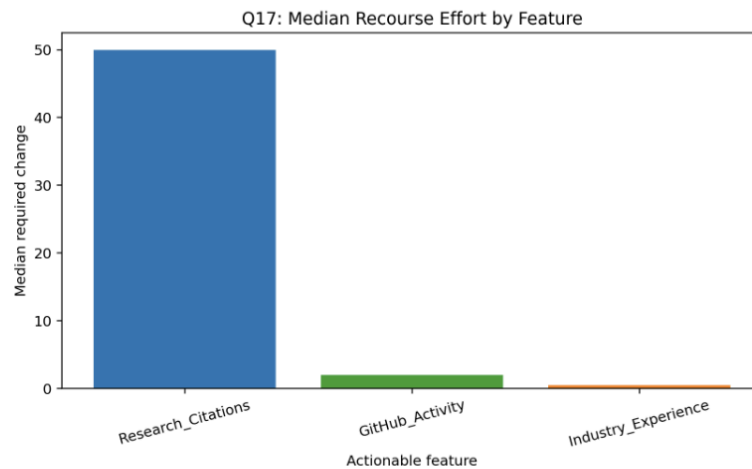


Fig. 14. Median actionable effort needed to flip near-boundary predictions.

این تصویر چه چیزی نشان می‌دهد؟

میزان تغییر لازم در ویژگی‌های قابل مداخله برای تغییر خروجی مدل.

چطور باید خوانده شود؟

- مقدار کمتر → مداخله عملی‌تر
- مقدار خیلی زیاد → پیشنهاد غیرواقعی

نتیجه تحلیلی

بار مداخله برای ویژگی‌ها یکسان نیست؛ بعضی مسیرهای ریکورس قابل‌اجراترند.

اثر بر تصمیم

توصیه‌های سیستم باید روی مداخلات feasible و کم‌هزینه متمرکز شود.

محدودیت

امکان‌پذیری واقعی (زمان/هزینه/محدودیت ساختاری) ممکن است کامل در داده نباشد.

تصویر ۱۵ – افت عملکرد زمانی (Temporal Degradation)

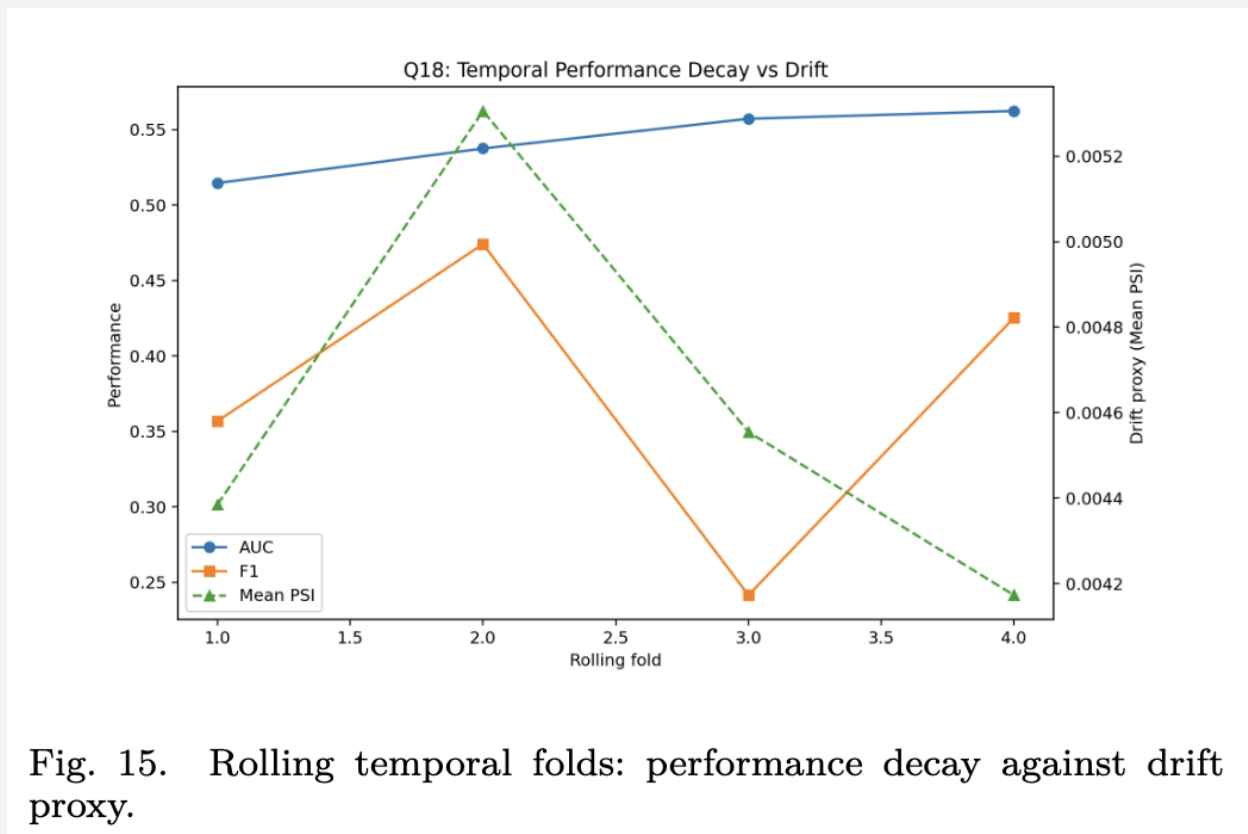


Fig. 15. Rolling temporal folds: performance decay against drift proxy.

این تصویر چه چیزی نشان می‌دهد؟

تغییر عملکرد مدل در foldهای زمانی و مقایسه آن با شاخص drift.

چطور باید خوانده شود؟

- افت پیوسته AUC/F1 در foldهای جلوتر → فرسایش زمانی مدل
- همزمانی افت با افزایش drift → نشانه قوی‌تر

نتیجه تحلیلی

مدل ممکن است روی آینده ضعیفتر عمل کند، حتی اگر روی split تصادفی خوب باشد.

اثر بر تصمیم

قبل از استقرار، ارزیابی زمانی باید الزامی باشد.

محدودیت

اگر timestamp واقعی نداریم و fallback می‌زنیم، قطعیت تفسیر کمتر می‌شود.

تصویر ۱۶ – پوشش عدم قطعیت (Coverage vs Confidence)

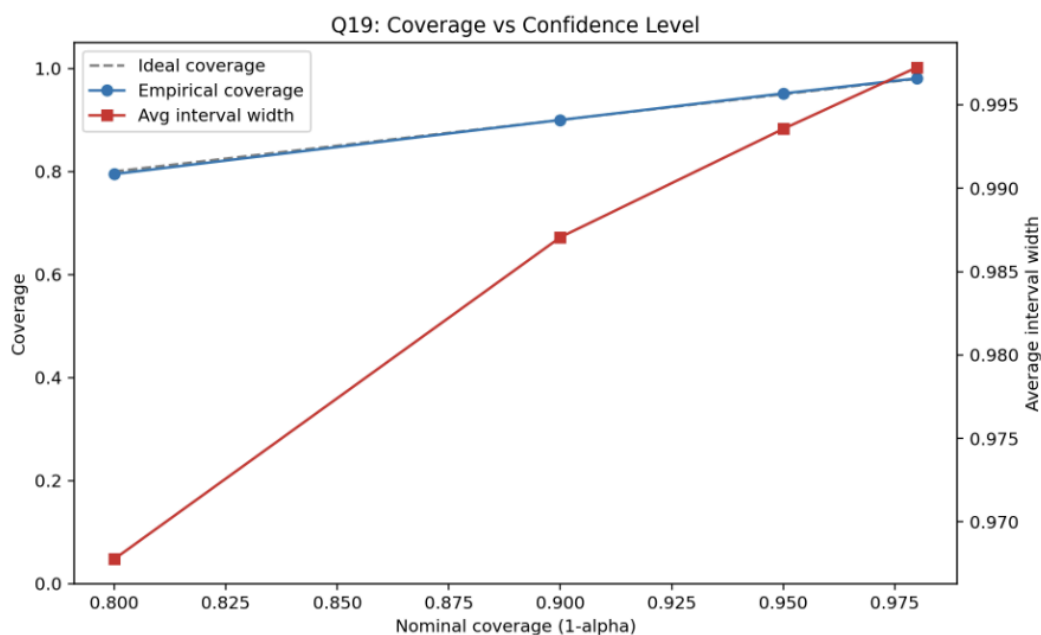


Fig. 16. Nominal vs empirical coverage and interval width across confidence levels.

این تصویر چه چیزی نشان می‌دهد؟

پوشش اسمی در برابر پوشش تجربی و تغییر پهنای بازه‌های اطمینان.

چطور باید خوانده شود؟

- پوشش تجربی نزدیک پوشش اسمی → قابل اعتماد

- under-coverage زیاد → ریسک بالای اعتماد کاذب

نتیجه تحلیلی

مدیریت عدم قطعیت باید بخشی از سیاست تصمیم باشد، نه گزینه اختیاری.

اثر بر تصمیم

نمونه‌های کم‌اعتماد به مسیر «بازبینی انسانی» هدایت می‌شوند.

محدودیت

زیر drift شدید، ضمانت‌های پوشش ممکن است تضعیف شود.

تصویر ۱۷ – Trade-off عدالت و کارایی

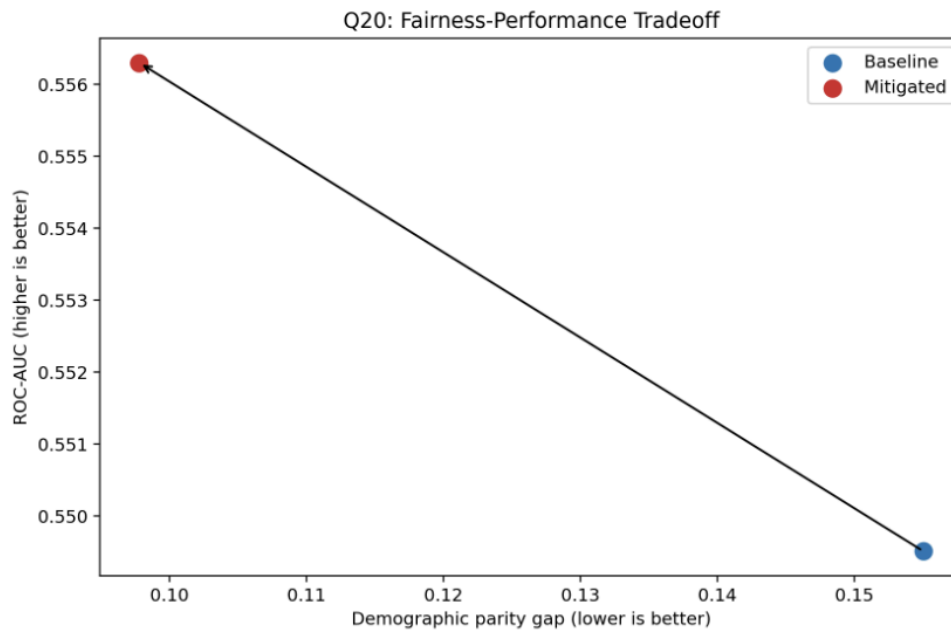


Fig. 17. Fairness-performance tradeoff from baseline to mitigated model.

این تصویر چه چیزی نشان می‌دهد؟

مقایسه مدل پایه و مدل مداخله‌شده روی صفحه fairness-performance.

چطور باید خوانده شود؟

- بهبود fairness معمولاً با کمی افت utility همراه است.
- باید دید آیا این افت داخل قید سیاستی تعریف‌شده هست یا نه.

نتیجه تحلیلی

مداخله عدالت وقتی قابل پذیرش است که کاهش تبعیض واقعی و افت کارایی کنترل‌شده باشد.

اثر بر تصمیم

تصمیم استقرار باید policy-driven باشد:
مثلاً «اگر AUC بیش از X افت کند، استقرار رد می‌شود».

محدودیت

بهبود یک معیار عدالت (مثلاً DP gap) ممکن است سایر ابعاد عدالت را کامل پوشش ندهد.

۶) خلاصه نتایج کلیدی (متریک‌ها)

در نسخه نهایی، مقادیر این شاخص‌ها باید از خروجی pipeline پر شوند:

- مدل نهایی کیستون
- AUC / Accuracy / F1
- Brier / ECE
- آستانه بهینه F1
- ویژگی با بیشترین drift و مقدار PSI
- نرخ موفقیت recourse
- میانگین AUC زمانی و میزان افت
- Coverage@90 و بیشترین under-coverage
- DP gap قبل/بعد از مداخله
- عبور یا عدم عبور از قید سیاستی

۷) بازتولیدپذیری و خروجی‌های ممیزی

اجرای کامل پروژه با:

```
{python code/scripts/full_solution_pipeline.py --profile {fast,balanced,heavy
```

خروجی‌های کلیدی:

- تصاویر: `/code/figures` یا `/figures/..`
- جداول و نتایج: `/code/solutions`
- خلاصه ماشینی: `run_summary.json`
- متریک‌های گزارش: `latex_metrics.json` و `latex_metrics.tex`
- CSV های Q18 تا Q20 برای ممیزی

۸) جمع‌بندی نهایی گزارش

این پروژه یک جریان کامل «آموزشی + نزدیک به تولید» ارائه می‌کند:

- مدل‌سازی بدون نشت داده
- سیاست تصمیم‌گیری مبتنی بر احتمال کالیبره
- پایش drift و تحلیل فرسایش زمانی

- مدیریت عدم قطعیت با ارجاع انسانی

- بهبود عدالت با قیود سیاستی روشن

پس خروجی نهایی فقط یک مدل نیست؛

یک چارچوب تصمیم‌یار قابل دفاع است که از نظر فنی، آماری، اخلاقی و مهندسی قابل ارزیابی و حسابرسی است.
