



پایش بهره‌وری آموزش سامانه‌ی هوش تجاری برای پلتفرم آموزش آنلاین

کشف نقاط ناکارآمد با نگاه مهندسی صنایع: پلتفرم آموزشی به مثابه
یک خط تولید – گلوگاه، ضایعات، دوباره‌کاری و ظرفیت هدررفته.

پروژه‌ی پایانی | درس هوش تجاری (Business Intelligence)

دانشگاه صنعتی شریف – دانشکده‌ی مهندسی صنایع

دانشگاه

موضوع 21 – پلتفرم آموزش آنلاین

موضوع پروژه

پایگاه داده و ERD • تولید داده • داشبورد Looker Studio • گزارش‌های
سه‌نقشی • چت‌بات داده‌محور

بخش‌های نمره‌دار

PostgreSQL 16 • Python • FastAPI • React • Tremor • Looker Studio •
Docker

پشته‌ی فناوری

معصومه یحیویان • شماره‌ی دانشجویی: 401155565

نگارنده

دکتر محمدرضا قدوسی

استاد درس

مرداد 1405

تاریخ تدوین

bi.packup.ir

سامانه‌ی زنده

فهرست مطالب

2	1 صورت مسئله و ذی‌نفعان
3	2 مدل داده و ERD
7	3 تولید داده مصنوعی
9	4 لایه معنایی – شانزده ویوی تحلیلی
10	5 یافته‌ها – پنج اقدام برتر
12	6 داشبورد Looker Studio
16	7 سامانه وب سه‌نقشی
17	8 چت‌بات داده‌محور
19	9 راستی‌آزمایی
20	10 جمع‌بندی و محدودیت‌ها

چکیده

این پروژه یک پلتفرم آموزش آنلاین را **مثل یک کارخانه** می‌بیند: دانش‌آموز ماده اولیه نیست، سفارش است؛ هر دوره یک **خط تولید** با ایستگاه‌های پشت‌سرهم؛ رهاکردن دوره **ضایعات**؛ سؤال معیوب آزمون **دوباره‌کاری**؛ و صندلی خالی جلسه **ظرفیت هدررفته**. با این قاب، سامانه به‌جای رسم نمودار خام، **خودش تشخیص می‌دهد**: خروجی اصلی یک فهرست اولویت‌دار از ناکارآمدی‌هاست که هر سطرش سه چیز دارد – **شاهد عددی، علت، و اثر ریالی تخمینی**.

روی یک پایگاه داده PostgreSQL با 26 جدول، 43 کلید خارجی و 65 قید CHECK، حدود 232,000 رکورد مصنوعی سازگار تولید شد و روی آن 16 ویوی تحلیلی نشست. سه مصرف‌کننده از همین ویوها تغذیه می‌شوند: داشبورد Looker Studio، یک سامانه وب سه‌نقشی، و یک چت‌بات داده‌محور که هم داده می‌گیرد و ثبت می‌کند و هم به پرسش‌های تحلیلی با **نمودار** پاسخ می‌دهد.

1 صورت مسئله و ذی‌نفعان

مسئله

پلتفرم‌های آموزش آنلاین داده‌ی فراوان تولید می‌کنند و تصمیم کم. گزارش‌های متعارف به «چند نفر ثبت‌نام کردند» و «چقدر فروش داشتیم» ختم می‌شوند؛ این اعداد وضعیت را می‌گویند ولی کار بعدی را نه. مسئله‌ی این پروژه دقیقاً همین شکاف است:

از میان هزاران رویداد ثبت‌شده، کدام چند ناکارآمدی بیشترین پول را می‌سوزانند، شاهد عددی‌شان چیست، و اقدام مشخص هفته‌ی آینده کدام است؟

چرا قاب مهندسی صنایع

یک دوره آموزشی از نظر ساختاری با یک خط تولید یکی است: ورودی (ثبت‌نام)، ایستگاه‌های متوالی (جلسه‌ها)، بازرسی (آزمون)، و خروجی (گواهی). این تشابه چهار سنجۀ کلاسیک صنایع را مستقیماً قابل استفاده می‌کند:

جلسه‌ای که بیشترین ریزش را می‌سازد؛ ظرفیت کل خط را همان تعیین می‌کند.

گلوگاه (TOC)

ثبت‌نامی که به گواهی نمی‌رسد؛ هزینه‌اش پرداخت شده ولی ارزشی تحویل نشده.

ضایعات (Scrap)

سؤالی با ضریب تمیز منفی؛ نمره را خراب می‌کند و بازآزمون می‌سازد.

دوباره‌کاری (Rework)

صندلی خالی جلسه؛ منتور حاضر است ولی ظرفیتش مصرف نمی‌شود.

بهره‌برداری
(Utilization)

ذی‌نفعان و پرسش هرکدام

نقش	پرسش اصلی	تصمیمی که می‌گیرد
مدیر	پول کجا برمی‌گردد و کجا می‌سوزد؟	بودجه‌ی بازاریابی، حذف یا بازسازی دوره
کارشناس عملیات	کدام جلسه ریزش می‌سازد؟	بازطراحی محتوا، بازآموزی منتور
دانش‌آموز و خانواده	کجا عقبم و چقدر پیشرفت کرده‌ام؟	برنامه‌ی مطالعه، انتخاب دوره بعدی
پشتیبانی	کدام درخواست از SLA گذشته؟	اولویت‌دهی به صف تیکت

2 مدل داده و ERD

اصول طراحی

اسکیما در فضای نام `edu` ساخته شد تا از `public` جدا بماند. سه قاعده در کل طراحی رعایت شد:

- هیچ حقیقتی دوبار ذخیره نمی‌شود. نرخ تکمیل، درآمد و ریزش هیچ‌کدام ستون نیستند؛ همه از رویدادهای خام محاسبه می‌شوند. یعنی داده هرگز با خودش ناسازگار نمی‌شود.
- قید در پایگاه داده، نه در برنامه. 65 قید CHECK جلوی داده‌ی بی‌معنا را می‌گیرد – نمره‌ی منفی، پایان قبل از شروع، درصد بالای صد.
- زمان محلی است. پایگاه داده روی `Asia/Tehran` تنظیم شد؛ وگرنه «جلسه‌ی ۹ صبح» در گزارش‌ها ۴:۳۰ بامداد دیده می‌شد.

ابعاد اسکیما

جدول	کلید خارجی	قید CHECK	نمایه	ویوی تحلیلی
26	43	65	22	16

نمودار ERD

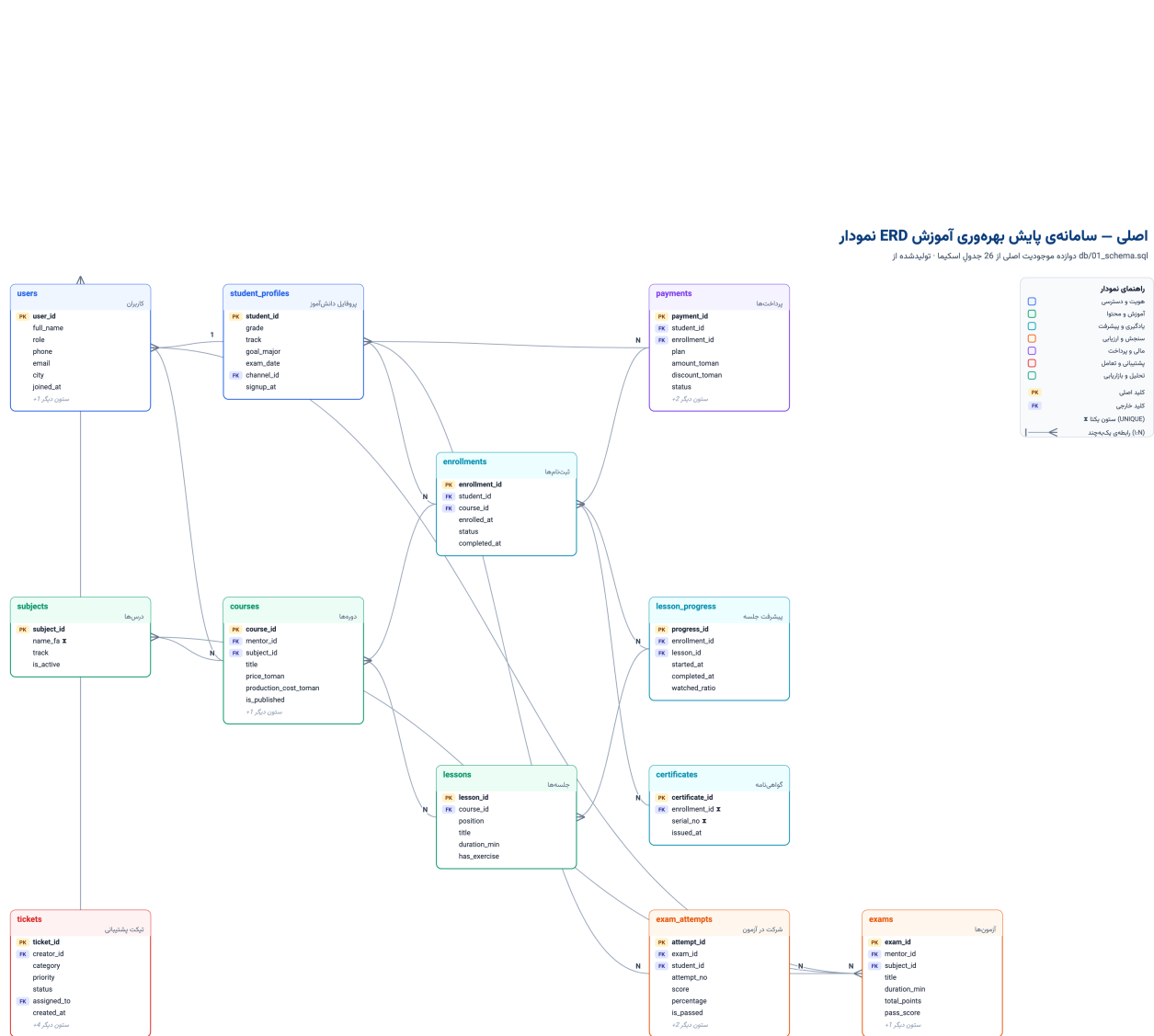
نمودار زیر دوازده موجودیت اصلی سامانه را نشان می‌دهد – همان مسیری که کسب‌وکار از آن می‌گذرد: از کاربر و پروفایل دانش‌آموز تا ثبت‌نام، پیشرفت جلسه، سنجش، پرداخت و پشتیبانی. در هر جعبه نام جدول، برچسب فارسی، کلید اصلی (PK)، کلیدهای خارجی (FK) و ستون‌های یکتا مشخص‌اند؛ روی هر خط، کاردینالیتی دو سر رابطه (1 و N) نوشته شده و پایانه سه‌شاخه سمت چند را نشان می‌دهد.

نمودار از روی اسکیما ساخته می‌شود، نه با دست

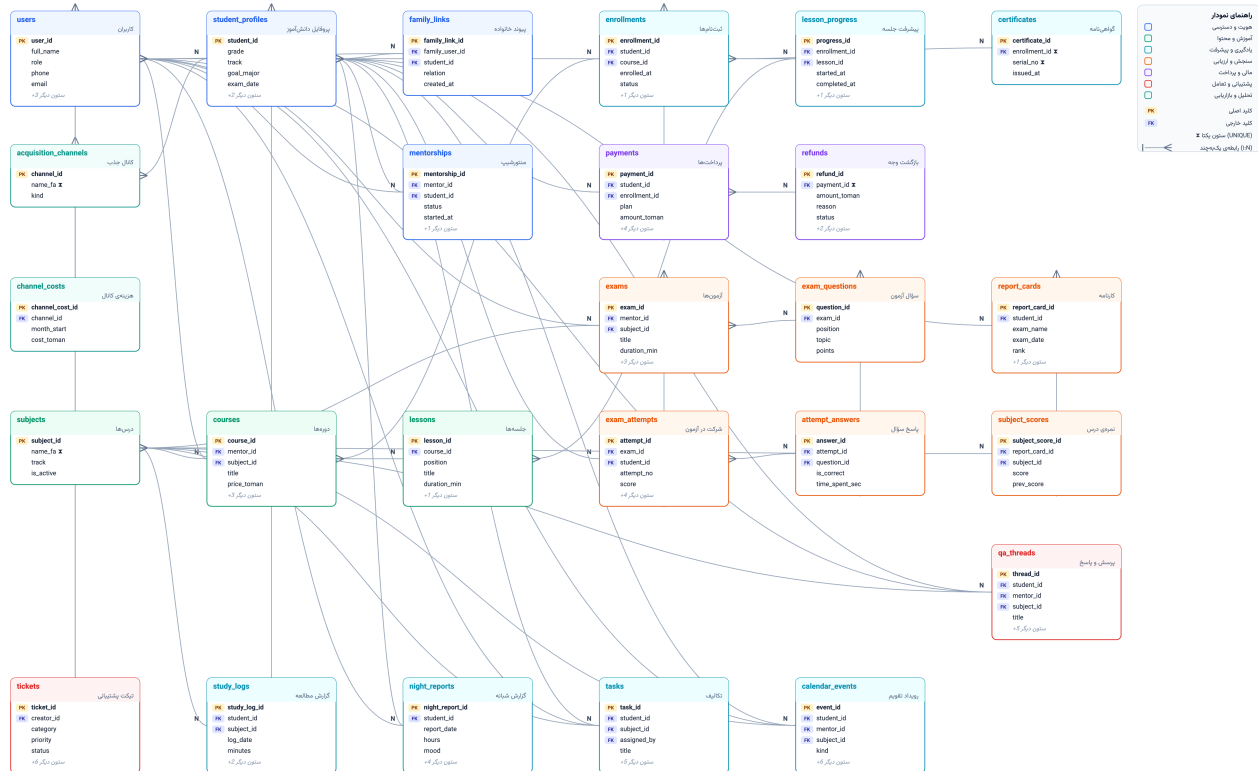
هر دو نمودار این فصل خروجی اسکریپت `docs/erd/generate_erd.py` هستند که مستقیماً `db/01_schema.sql` را می‌خواند و جدول‌ها، کلیدها و روابط را استخراج می‌کند. اگر اسکیما تغییر کند، اجرای دوباره اسکریپت نمودار را هم‌گام می‌کند؛ و اگر جدولی در چیدمان باشد که در اسکیما نیست، اسکریپت خطا می‌دهد. یعنی نمودار هیچ‌وقت از پایگاه داده جدا نمی‌افتد – اتفاقی که برای ERDهای دستی رایج است.

روابط اصلی و معنای کسب‌وکاری‌شان

رابطه	کاردینالیتی	قاعده کسب‌وکار
<code>users</code> → <code>student_profiles</code>	1:1	کلید اصلی پروفایل خودش کلید خارجی است، پس رابطه اجباراً یک‌به‌یک است
<code>lessons</code> → <code>courses</code>	1:N	دوره از چند جلسه مرتب ساخته شده – «ایستگاه‌های خط تولید»
<code>enrollments</code> → <code>lesson_progress</code>	1:N	قلب محاسبه قیف ریزش
<code>enrollments</code> → <code>certificates</code>	1:1	گواهی حداکثر یکی، با قید UNIQUE
<code>exam_attempts</code> → <code>exams</code>	1:N	<code>attempt_no > 1</code> یعنی دوباره‌کاری
<code>exam_attempts</code> → <code>attempt_answers</code>	1:N	پایه محاسبه ضریب تمیز سؤال
<code>refunds</code> → <code>payments</code>	1:1	هر پرداخت حداکثر یک بازگشت وجه
<code>tickets</code> → <code>users</code>	1:N (دوگانه)	یک بار <code>creator_id</code> ، یک بار <code>assigned_to</code>



شکل 1: نمودار ERD اصلی – دوازده موجودیت کلیدی، با کلید اصلی، کلید خارجی و کاردینالیتی

کامل – همه‌ی جدول‌های اسکیمای نمودار
db/01_schema.sql 26 جدول، تولیدشده از 26

شکل 2: نمودار ERD کامل – هر 26 جدول اسکیمای با 43 کلید خارجی

خوشه‌های موجودیت

جدول‌ها در شش خوشه‌ی معنایی گروه می‌شوند:

users ، student_profiles ، family_links ، mentorships

هویت و نقش

subjects ، courses ، lessons ، enrollments ، lesson_progress ، certificates

محتوا و ثبت‌نام

study_logs ، night_reports ، tasks ، calendar_events

فعالیت روزانه

exams ، exam_questions ، exam_attempts ، attempt_answers ، subject_scores ، report_cards

سنجش

payments ، refunds ، acquisition_channels ، channel_costs

مالی و جذب

tickets ، qa_threads

پشتیبانی

تصمیم طراحی قابل دفاع

جدول attempt_answers پاسخ تک‌تک سؤال‌ها را نگه می‌دارد، نه فقط نمره نهایی را. این باعث بزرگ‌شدن جدول شد (54,080 رکورد) ولی بدون آن، تحلیل سؤال ممکن نبود – و همان تحلیل است که یکی از پنج یافته اصلی را می‌سازد. جزئیاتی که ذخیره نکنی، تحلیلی است که هرگز نخواهی داشت.

3 تولید داده مصنوعی

چرا مولد برنامه‌ای، نه داده دستی

داده با یک اسکریپت پایتون قطعی تولید شد: بذر ثابت (seed = 1405)، بازه 21 اسفند 1403 تا 19 مرداد 1405. هر اجرا دقیقاً همان 232,258 رکورد را می‌سازد. این قطعیت سه فایده داشت: نتایج گزارش بازتولیدپذیر است، خطاها قابل ردیابی‌اند، و می‌شد مولد را چند بار تنظیم کرد تا اعداد واقع‌نما شوند.

دستورهای تولید داده – «با چه منطقی به این داده رسیدیم»

برای اینکه مسیر رسیدن به داده قابل بازرسی باشد، مشخصاتی که به مولد داده شد در ادامه آمده و بلافاصله زیر هر مشخصات، کد واقعی‌ای که آن را اجرا می‌کند نقل شده است.

یک تذکر صادقانه

داده با مدل زبانی تولید نشده؛ با یک اسکریپت پایتون قطعی ساخته شده است. آنچه «دستور» نامیده می‌شود همان مشخصات تولید است. دلیل این انتخاب قطعیت است: مدل زبانی هر بار داده متفاوت می‌دهد و رعایت قیدهای FOREIGN KEY و CHECK را تضمین نمی‌کند.

دستور رفتار مطالعه

احتمال ثبت مطالعه در هر روز از پشتکار ثابت دانش‌آموز می‌آید، جمع‌ها کم می‌شود و هرچه به کنکور نزدیک‌تر شویم بالا می‌رود:

```
p_log = clamp(0.30 + 0.55 * diligence, 0.10, 0.92) * weekend
weekend = 0.55 if friday else 1.0
pressure = clamp(1.9 - days_to_exam / 330, 0.75, 1.9)
minutes = clamp(gauss(125 * diligence * pressure, 45), 10, 480)
```

دستور قیف یادگیری و ریزش

ریزش فقط در لحظه «شروع جلسه» رخ می‌دهد؛ جلسه طولانی بدون تمرین احتمال شروع را تقریباً نصف می‌کند – همان گلوگاه کاشته‌شده:

```
p_start = 0.96 if position == 1 else keep_base
if duration >= 78 and not has_exercise: p_start *= 0.55
elif duration >= 60: p_start *= 0.93
watched_ratio = clamp(betavariate(8, 1.2), 0.05, 1.0)
completed = watched_ratio >= 0.65
```

دستور کیفیت منتور

یک منتور عمداً ضعیف ساخته شد: کم بازبینی می‌کند، دیر پاسخ می‌دهد و امتیاز پایین می‌گیرد:

```
p_review = 0.35 if bad_mentor else clamp(0.70 + 0.25*(quality-1), 0.5, 0.95)
lag_hours = randint(20, 60) if bad_mentor else randint(1, 14)
rating = randint(1, 3) if bad_mentor else randint(3, 5)
```

شرح کامل هر پنج دستور به همراه سطرهای متناظر کد در docs/DATA.md بخش ۶ آمده است.

حجم داده

ثبت‌نام	لاگ مطالعه	پیشرفت جلسه	پاسخ سؤال	پرداخت	تیکت
2,937	56,979	12,596	54,080	3,357	2,601

واقع‌نمایی: سه بار تنظیم مولد

نسخه‌ی اول داده‌ی سازگار تولید کرد ولی باورپذیر نبود. سه اصلاح لازم شد:

قیف شکسته	حلقه‌ی تولید هر بار که نسبت تماشا زیر 0.85 می‌شد کلاً می‌شکست، پس فقط یک گواهی صادر شد. «رهاکردن دوره» از «تکمیل جلسه» جدا شد؛ نرخ تکمیل به 13.7% رسید که برای آموزش آنلاین عدد متعارفی است.
CAC غیرواقعی	هزینه‌ی ماهانه‌ی کانال‌ها حدود پنج برابر بالا بود و CAC به 12.85 میلیون تومان به‌ازای هر دانش‌آموز می‌رسید. با مقیاس‌گذاری و وابسته‌کردن تعداد ثبت‌نام به پشتکار دانش‌آموز، LTV بین کانال‌ها تنوع گرفت.
منطقه‌ی زمانی	EXTRACT(hour) روی UTC کار می‌کرد و ساعت ۸ صبح، ۴ بامداد دیده می‌شد. با <code>-c timezone=Asia/Tehran</code> روی خود سرور پایگاه داده حل شد.

یافته‌های کاشته‌شده

برای اینکه داشبورد چیزی برای کشف‌کردن داشته باشد، پنج ناکارآمدی عمداً در داده کاشته شد – نه به‌صورت برجسته، بلکه به‌صورت الگوی رفتاری که فقط با تحلیل پیدا می‌شود. هر پنج مورد بعداً با SQL مستقل راستی‌آزمایی شد.

#	نوع	چطور کاشته شد
1	گلوگاه محتوا	یک جلسه‌ی مشخص طولانی‌تر و بدون تمرین ساخته شد.
2	منتور ضعیف	یک منتور با نرخ بازبینی و سرعت پاسخ‌دهی پایین.
3	کانال گران	یک کانال جذب با هزینه‌ی بالا و ماندگاری کم.
4	سؤال معیوب	در یک آزمون، چهار سؤال با پاسخ کلید اشتباه.
5	ساعت بد	جلسه‌های صبح زود با حضور بسیار پایین.

4 لایه معنایی – شانزده ویوی تحلیلی

چرا ویو، نه کوئری در برنامه

هر سه مصرف‌کننده (داشبورد Looker، سامانه وب، چت‌بات) از همین شانزده ویو تغذیه می‌شوند. یعنی «نرخ تکمیل» فقط یک تعریف دارد و در هر سه جا همان عدد دیده می‌شود. اگر هر مصرف‌کننده کوئری خودش را می‌نوشت، سه تعریف کمی متفاوت پیدا می‌کردیم – کلاسیک‌ترین راه بی‌اعتبار شدن یک داشبورد.

پرسشی که پاسخ می‌دهد

ویو

هشت شاخص کلان سامانه در یک سطر	v_kpi_overview
پنج اقدام برتر با شاهد و اثر – کارت امضای پروژه	v_top_actions_full
کیف جلسه‌به‌جلسه و افت نسبت به جلسه قبل	v_funnel_lesson
ضریب دشواری و ضریب تمیز هر سؤال	v_exam_item_analysis
بهره‌وری هر منتور به ازای 100 ساعت مطالعه	v_mentor_performance
CAC، LTV و نسبتشان برای هر کانال جذب	v_channel_economics
نرخ حضور بر حسب ساعت و بازه روز	v_session_capacity
ماندگاری هر کوهورت در ماه‌های بعد	v_cohort_retention
نرخ تکمیل و سود ناخالص هر دوره	v_course_health
نمای کامل یک دانش‌آموز	v_student_360
رابطه ساعت مطالعه با بهبود نمره	v_productivity
زمان پاسخ اول و وضعیت SLA هر تیکت	v_ticket_sla

تاریخ جلالی در پایگاه داده

یک تابع PL/pgSQL به نام `to_jalali` نوشته شد و جدول `dim_date` برچسب ماه شمسی را از پیش محاسبه می‌کند. دلیلش عملی است: اگر تبدیل تاریخ در لایه نمایش انجام شود، Looker Studio نمی‌تواند بر اساس ماه شمسی مرتب یا گروه‌بندی کند.

دو سنجه محوری

$$\text{Discrimination}_q = P(\text{correct} \mid \text{top}27\%) - P(\text{correct} \mid \text{bottom}27\%)$$

ضریب تمیز سؤال. مقدار منفی یعنی دانش‌آموز قوی بیشتر غلط زده – نشانه قطعی ایراد در طراحی یا کلید سؤال، نه ضعف دانش‌آموز.

$$\text{Gain}_{100h} = \frac{\overline{\Delta \text{score}}}{\text{study_hours}} \times 100$$

بهره‌وری منتور: بهبود نمره به ازای هر 100 ساعت مطالعه دانش‌آموزانش. برخلاف «میانگین نمره»، این سنجه منتوری را که دانش‌آموزان قوی گرفته پاداش نمی‌دهد.

5 یافته‌ها – پنج اقدام برتر

این بخش خروجی اصلی پروژه است. هر سطر از ویوی `v_top_actions_full` می‌آید و همان چیزی است که در صدر داشبورد مدیر دیده می‌شود.

یافته ۱ – گلوگاه محتوا

دوره‌ی «نکته و تست ریاضی» – جلسه‌ی ۴. ریزش 59% (49 نفر) در همین یک جلسه؛ مدت‌ش 94 دقیقه و بدون هیچ تمرینی.
اقدام: تقسیم جلسه و افزودن تمرین. **اثر تخمینی:** 8 میلیون تومان درآمد حفظ‌شده.
 به زبان صنایع: این ایستگاه گند خط است. تا وقتی باز نشود، ظرفیت کل دوره را همین تعیین می‌کند.

یافته ۲ – کیفیت منتورینگ

منتور: حسین میرزایی. رضایت 1.77 از 5؛ نرخ بازبینی 38.3%؛ میانگین پاسخ‌دهی 52 ساعت. بهره‌وری او 0.21 است در برابر 0.37 برای منتورهای بعدی – یعنی حدود نصف.
اقدام: بازآموزی یا بازتوزیع دانش‌آموزان. **اثر:** 22 دانش‌آموز در معرض ریسک.

یافته ۳ – بازده بازاریابی

کانال: تبلیغات اینستاگرام. نسبت LTV/CAC برابر 1.51 در حالی که آستانه متعارف سلامت 3 است؛ CAC برابر 2,304 هزار تومان.
اقدام: کاهش بودجه و انتقال به کانال‌های پربازده. **اثر:** 558 میلیون تومان بودجه قابل بازتخصیص.

کانال جذب	دانش‌آموز	CAC (هزار ت)	LTV/CAC
تبلیغات اینستاگرام	242	2,304	1.51
تبلیغات یوتیوب/آپارات	118	2,095	2.08
همکاری با مدارس	94	2,011	2.45
کانال تلگرام	104	987	4.70
جست‌وجوی گوگل	139	441	10.30
معرفی دوستان	103	293	18.81

کانالی که بیشترین دانش‌آموز را می‌آورد، کم‌بازده‌ترین هم هست. «تعداد جذب» بدون «هزینه جذب» گمراه‌کننده است.

یافته ۴ – کیفیت سنجش

آزمون زیست‌شناسی – سری ۴. سه سؤال با ضریب تمیز منفی (-0.38، -0.19، -0.19). دانش‌آموز قوی بیشتر از ضعیف غلط زده.
اقدام: بازطراحی سؤال‌های معیوب پیش از آزمون بعدی. **اثر:** اعتبار نمره 64 دانش‌آموز.
 این «دوباره‌کاری» است: نمره‌ای که تولید شده معیوب است و باید دور ریخته یا اصلاح شود.

یافته ۵ – بهره‌وری ظرفیت

جلسه‌های ساعت ۹ صبح. نرخ حضور 21.0% در 830 جلسه – در حالی که میانگین کل سامانه 52.5% است. ساعت ۸ صبح هم با 22.6% وضع مشابهی دارد.

اقدام: جابه‌جایی جلسه‌های صبح به بازه 19 تا 21. اثر: 656 جلسه هدررفته.

چرا این خروجی با «داشبورد» فرق دارد

یک داشبورد متعارف نمودار «نرخ حضور بر حسب ساعت» را رسم می‌کند و کار را تمام‌شده می‌داند. اینجا سامانه یک قدم جلوتر می‌رود: خودش پایین‌ترین ستون را پیدا می‌کند، فاصله‌اش با میانگین را می‌سنجد، و آن را با تعداد جلسه ضرب می‌کند تا به عدد قابل تصمیم برسد. تفاوت، تفاوت گزارش و تشخیص است.

6 داشبورد Looker Studio

اتصال

داشبورد مستقیماً به همان PostgreSQL روی سرور آمستردام وصل می‌شود – نه به یک فایل صادرشده. یعنی هر تغییری در داده بلافاصله در گزارش دیده می‌شود. برای این کار یک نقش **فقط خواندنی** به نام `looker_ro` ساخته شد که فقط `SELECT` روی اسکیمای `edu` دارد؛ با پروپ مستقیم تأیید شد که `CREATE TABLE` و `DELETE` هر دو با «permission denied» رد می‌شوند.

دو نکته عملی که وقت می‌گیرند

یک: گزینه `TABLE` کار نمی‌کند، چون ویوها در اسکیمای `edu` هستند و فهرست جدول‌های Looker فقط `public` را نشان می‌دهد. باید `CUSTOM QUERY` انتخاب شود.

دو: Looker Studio نام ستون فارسی را نمی‌پذیرد («invalid characters in its field names»). هر 103 ستون ویوها به انگلیسی تغییر نام دادند و برچسب فارسی به صورت نمایشی در خود Looker تنظیم شد. نگاشت کامل برچسب‌ها در `docs/FIELD_LABELS.md` است.

ساختار سه صفحه‌ای

هر صفحه برای یک ذی‌نفع طراحی شده و فقط به پرسش همان نقش پاسخ می‌دهد:

صفحه	عناصر	محتوا
۱ – نمای مدیر	11	چهار کارت شاخص • جدول پنج اقدام برتر • میله‌ای بازده کانال (LTV/CAC) • دونات وضعیت ثبت‌نام • سری زمانی درآمد ماهانه • قیف جلسه‌به‌جلسه • جدول سؤال‌های معیوب • جدول عملکرد منتور • میله‌ای حضور بر بازه روز
۲ – دانش‌آموز و خانواده	8	کنترل کشویی انتخاب دانش‌آموز • چهار کارت میانگین (ساعت مطالعه، دوره ثبت‌نامی، نمره، پایبندی به تکالیف) • سری زمانی مطالعه • مطالعه به تفکیک درس • مطالعه بر حسب روز هفته • روند نمره به تفکیک درس
۳ – پشتیبانی و SLA	4	کارت شمار تیکت • سری زمانی تیکت • تیکت بر حسب دسته • میانگین زمان اولین پاسخ بر حسب دسته

حداقل خواسته‌شده درس

این پروژه

۳ صفحه

۲۲ عنصر

۶ کنترل

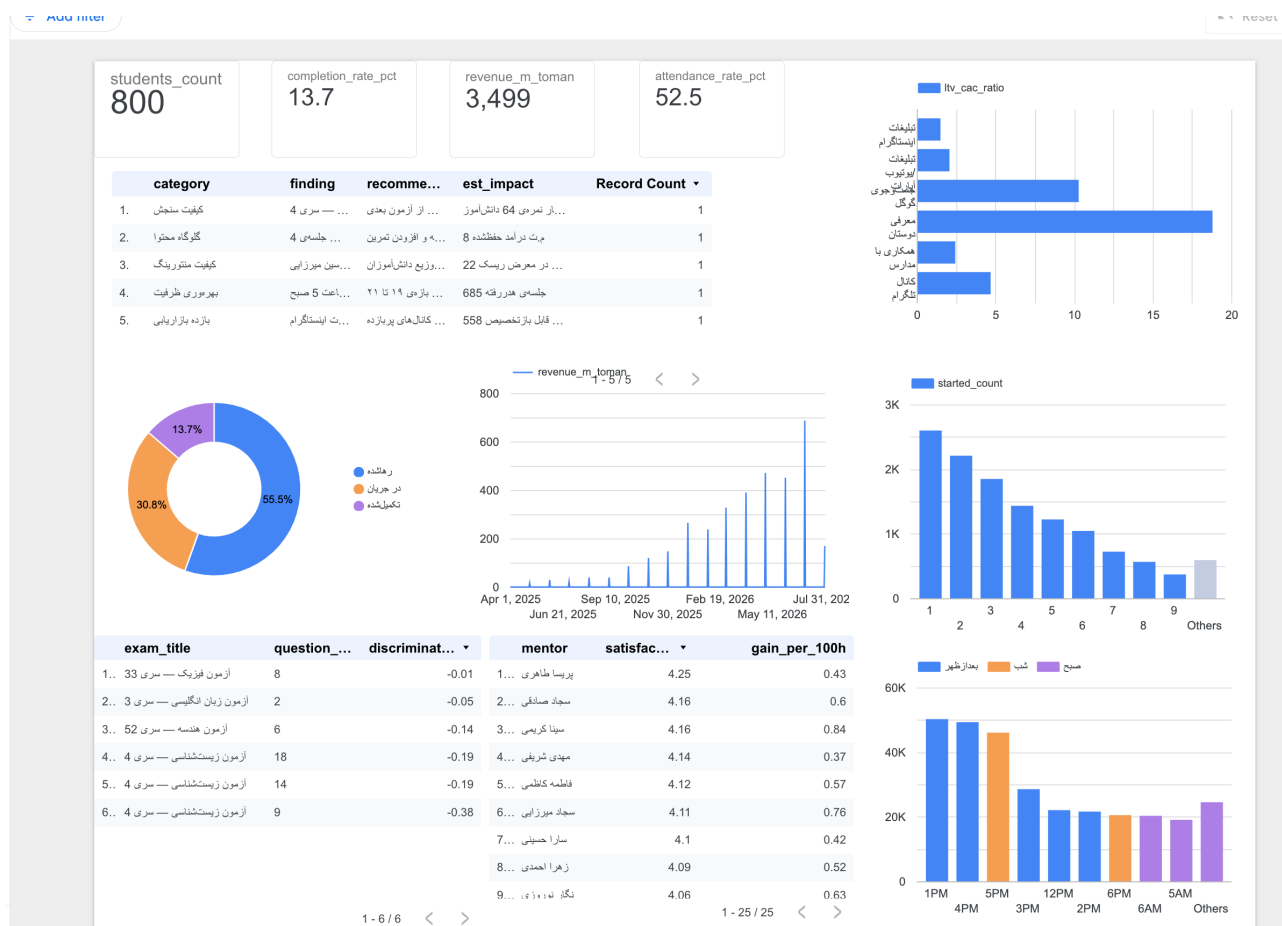
هر ۶ نوع

دست‌کم ۳ صفحه (مدیر، کارشناس، کاربر بیرونی)

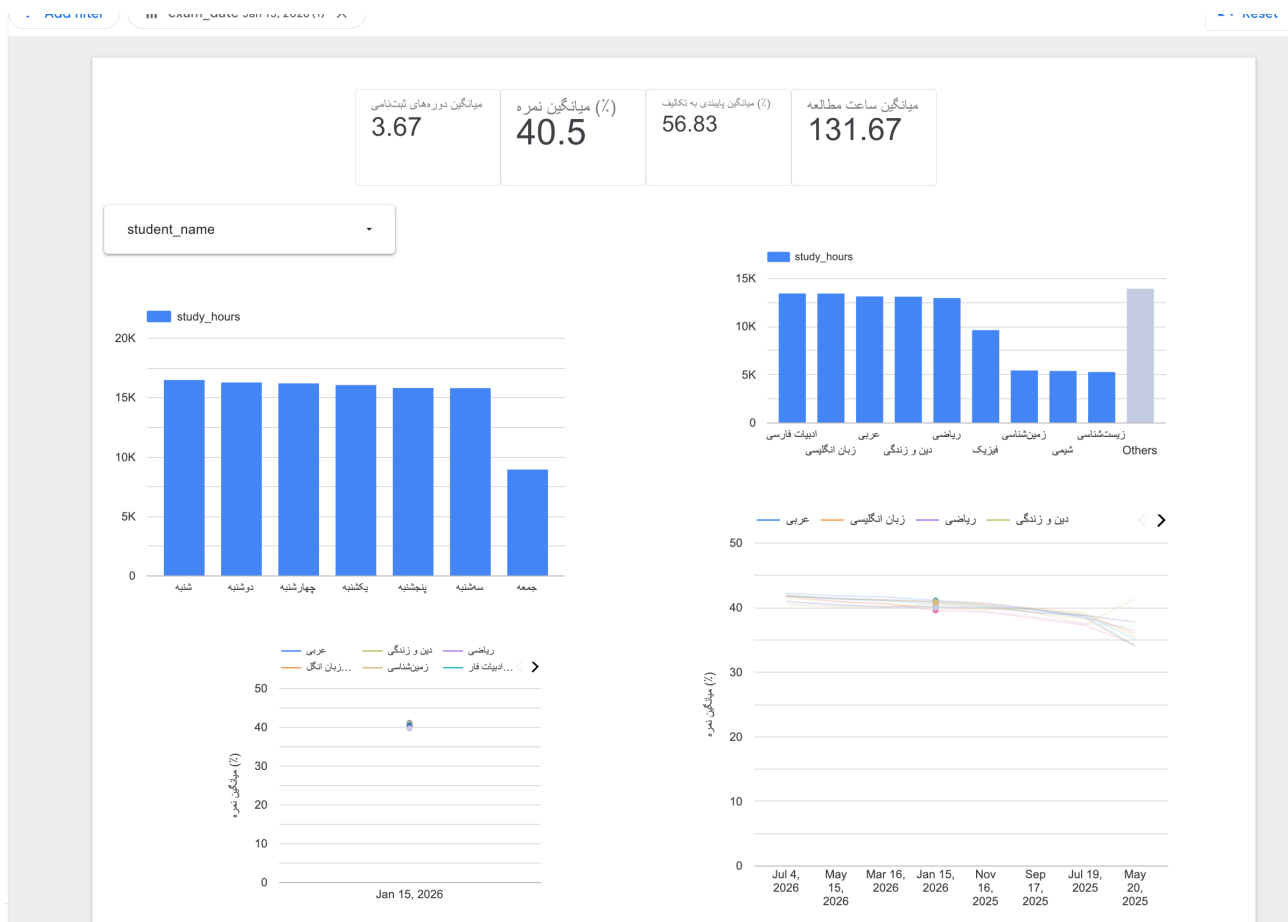
دست‌کم ۸ نمودار یا کارت KPI معنادار

فیلتر تاریخ، دسته‌بندی، واحد یا نقش

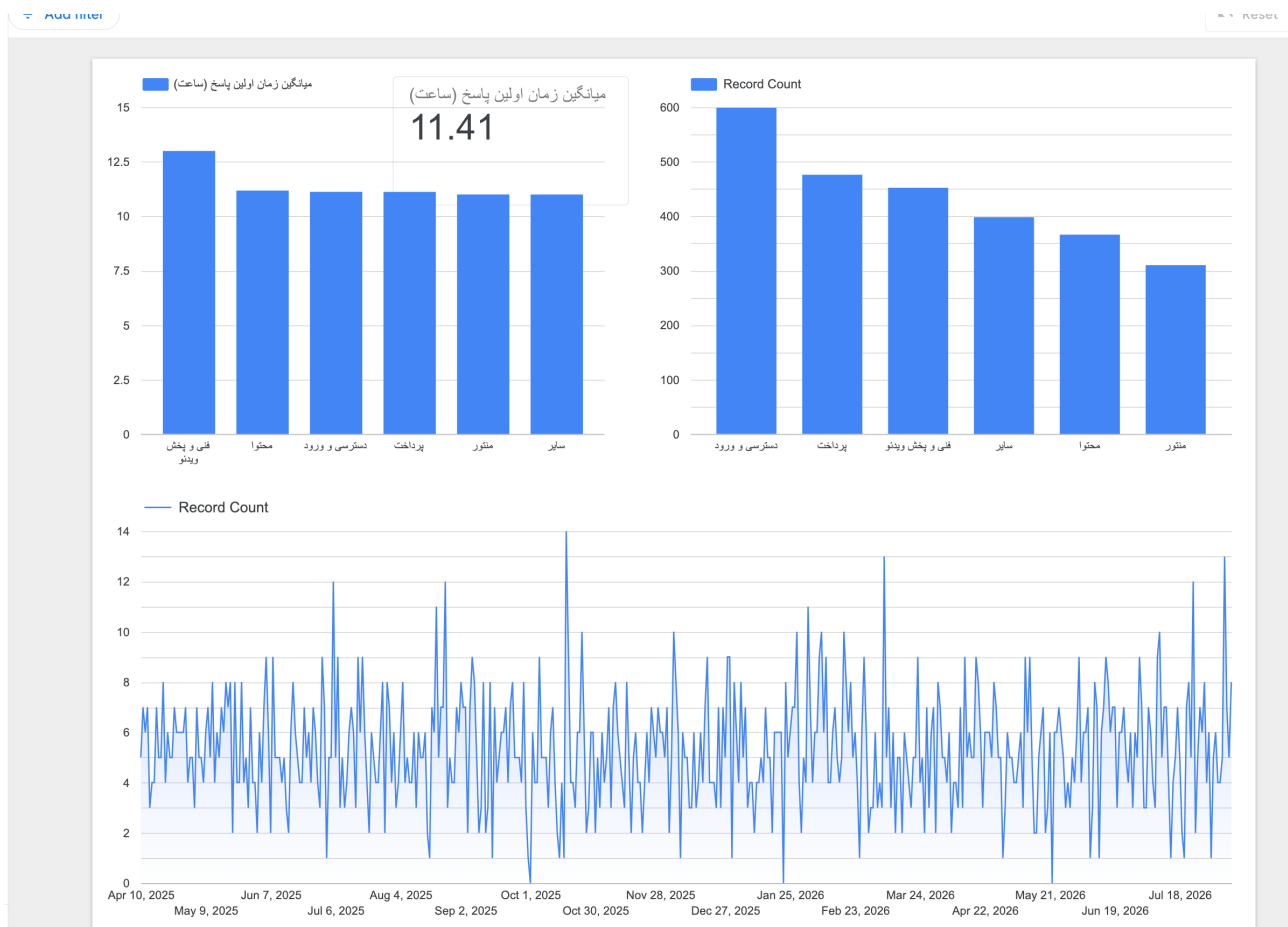
تنوع نوع نمودار



شکل 3: صفحه ۱ داشبورد – نمای مدیر: چهار کارت شاخص، جدول پنج اقدام برتر، بازده کانال‌ها، دونا ت وضعیت، قیف جلسه به جلسه، تحلیل سؤال و عملکرد منتور – همه با اتصال زنده به پایگاه داده



شکل 4: صفحه ۲ – دانش‌آموز و خانواده: کنترل کشویی هر هشت عنصر را هم‌زمان فیلتر می‌کند



شکل 5: صفحه ۳ – پشتیبانی و SLA: میانگین زمان اولین پاسخ بر حسب دسته، همان سنج‌های که ضعف SLA را نشان می‌دهد

7 سامانه وب سه‌نقشی

علاوه بر داشبورد Looker، یک برنامه تک‌صفحه‌ای React روی bi.packup.ir مستقر شد که از همان ویوها تغذیه می‌شود. دلیل وجودش این است که Looker Studio دو کار را نمی‌تواند: نوشتن در پایگاه داده (که چت‌بات لازم دارد)، و راست‌چینی درست فارسی.

معماری

شرح

لایه

PostgreSQL 16، اسکیمای edu، منطقه زمانی Asia/Tehran

پایگاه داده

FastAPI با psycpg3؛ استخراج اتصال با دو مسیر جدا: q() فقط خواندنی برای گزارش‌ها و نوشتن که فقط چت‌بات از آن استفاده می‌کند qw()

بک‌اند

React 18 + Vite + TypeScript + Tremor؛ دارک‌مود، واکنش‌گرا از 320 پیکسل

فرانت‌اند

Docker Compose روی سرور آمستردام؛ Caddy با گواهی خودکار Let's Encrypt

استقرار

راست‌چینی متن دوجته

متن فارسی که واژه لاتین یا عدد در خود دارد، اگر ایزوله نشود ترتیبش به هم می‌ریزد. الگوی به‌کاررفته دو بخش دارد: تابع detectDir جهت هر رشته را با شمارش حروف غالب تعیین می‌کند (لاتین باید اکثریت قاطع داشته باشد وگرنه راست‌به‌چپ برنده است)، و هر عدد در <bdi> با unicode-bidi: isolate پیچیده می‌شود.

اشکالی که فقط با اندازه‌گیری پیدا شد

کامپوننت Card در کتابخانه Tremor کلاس text-left را سخت‌کد می‌کند. در سند راست‌به‌چپ این یعنی همه متن فارسی داخل کارت به لبه چپ می‌چسبد. چشم این را در عرض کم نمی‌گیرد؛ با اندازه‌گیری مستقیم جعبه متن معلوم شد: صفر پیکسل فاصله از چپ و 730 پیکسل خالی از راست. راه حل text-align: start است نه right – چون start جهت را از سند می‌گیرد و داخل بوم چپ‌به‌راست نمودارها همچنان چپ می‌ماند.

8 چت‌بات داده‌محور

چت‌بات دو کار کاملاً متفاوت می‌کند: از کاربر داده می‌گیرد و ثبت می‌کند، و به مدیر پاسخ تحلیلی با نمودار می‌دهد. هر دو بدون مدل زبانی پیاده شدند – منطق قاعده‌محور است و هر پاسخ از پایگاه داده می‌آید، نه از حافظه مدل.

نقش‌ها و سناریوها

نقش	سناریو	خروجی
دانش‌آموز	ثبت تیکت	شماره پیگیری TKT-1405-XXXXXX
دانش‌آموز	درخواست بازگشت وجه	تأیید یا رد عددی مستدل
کارشناس پشتیبانی	ثبت جلسه زنده	بررسی تداخل زمانی
مدیر	پرسش تحلیلی	پاسخ متنی به همراه نمودار

پنج لایه اعتبارسنجی

داده پیش از رسیدن به پایگاه داده از پنج صافی می‌گذرد:

1. قالب – ارقام فارسی به لاتین تبدیل می‌شوند؛ کد ملی با الگوریتم رقم کنترلی بررسی می‌شود، نه فقط طولش.
2. دامنه – مقدار باید در بازه مجاز باشد (ساعت 0 تا 23، نمره 0 تا 20).
3. کامل بودن – هیچ فیلد اجباری خالی نمی‌ماند؛ بات تا گرفتن آن پیش نمی‌رود.
4. تطابق ارجاعی – شناسه دوره یا دانش‌آموز باید واقعاً در پایگاه داده وجود داشته باشد.
5. قاعده کسب و کار – شرط‌های دامنه؛ مثلاً بازگشت وجه فقط تا 7 روز و با پیشرفت کمتر از 20%.

نمونه اجرای واقعی (راستی‌آزمایی شده روی سامانه زنده)

تیکت TKT-1405-002601 واقعاً ساخته شد و وجودش در پایگاه داده تأیید شد. در تلاش دوم، تشخیص تکراری فعال شد و از ثبت دوباره همان مشکل در بازه 24 ساعت جلوگیری کرد. درخواست بازگشت وجه هم با دلیل عددی رد شد – نه با پیام عمومی، بلکه با ذکر روز سپری شده و درصد پیشرفت.

نمودار درون‌گفتگو

پاسخ تحلیلی علاوه بر متن، یک مشخصه نمودار برمی‌گرداند و کلاینت آن را با Tremor کنار همان پاسخ رسم می‌کند. هفت پرسش متداول نمودار متناظر خود را دارند:

پرسش	نمودار
گلوگاه کجاست؟	کیف افت همان دوره
کدام کانال بازده ندارد؟	میلای افقی LTV/CAC
وضعیت منتورها چطور است؟	میلای رضایت
کدام سؤال معیوب است؟	میلای ضریب تمیز
ظرفیت کجا هدر می‌رود؟	نرخ حضور بر ساعت
وضعیت پشتیبانی؟	تیکت به تفکیک دسته
نمای کلی	مساحتی روند درآمد

چرا قاعده‌محور و نه مدل زبانی

یک مدل زبانی می‌تواند عددی بسازد که وجود ندارد. در گزارشی که قرار است مبنای تصمیم باشد، این ریسک پذیرفتنی نیست. اینجا هر عدد نتیجه یک کوئری مشخص روی ویوی مشخص است – قابل ردیابی و قابل بازتولید. هزینه‌اش این است که بات فقط به پرسش‌های پیش‌بینی‌شده پاسخ می‌دهد؛ سودش این است که هیچ‌وقت دروغ نمی‌گوید.

9 راستی‌آزمایی

هر ادعای عددی این گزارش با کوئری مستقل روی همان پایگاه داده بررسی شد. اسکریپت db/99_verify.sql هر پنج یافته کاشته‌شده را دوباره پیدا می‌کند – یعنی زنجیره «تولید داده → ویدئو → تشخیص» سالم است.

بررسی	انتظار	نتیجه
پنج یافته کاشته‌شده دوباره پیدا می‌شوند	هر پنج	تأیید
نقش <code>looker_ro</code> نمی‌تواند بنویسد	رد شدن	تأیید
اتصال Looker به پایگاه داده	برقرار	تأیید
تیکت ساخته‌شده در پایگاه داده هست	موجود	تأیید
تشخیص تیکت تکراری	فعال شود	تأیید
قاعده بازگشت وجه	رد مستدل	تأیید
نمودار چت‌بات داده برمی‌گرداند	شش نقطه	تأیید

شاخص‌های نهایی سامانه

دانش‌آموز	ثبت‌نام	نرخ تکمیل	درآمد (م.ت)	هزینه جذب	نرخ حضور
800	2,937	13.7%	3,499	1,188	52.5%

10 جمع‌بندی و محدودیت‌ها

آنچه ساخته شد

یک زنجیره کامل از داده خام تا تصمیم: پایگاه داده قیددار، مولد داده قطعی، لایه معنایی مشترک، و سه مصرف‌کننده که همگی از یک منبع حقیقت تغذیه می‌شوند. خروجی نهایی نمودار نیست – فهرست اولویت‌دار کارهایی است که هفته آینده باید انجام شوند، هرکدام با شاهد و اثر ریالی.

محدودیت‌ها – صادقانه

- داده مصنوعی است. الگوها واقع‌نما تنظیم شده‌اند ولی از رفتار واقعی کاربر نیامده‌اند. اعداد اثر (مثلاً «8 میلیون تومان») تخمین‌اند، نه اندازه‌گیری.
- اثر تخمینی، علیت را ثابت نمی‌کند. اینکه جلسه ۴ بیشترین ریزش را دارد نشان می‌دهد کجا مشکل است، نه اینکه اصلاحش حتماً همان مقدار درآمد را برمی‌گرداند. برای آن، آزمون A/B لازم است.
- ترافیک پایگاه داده رمزنگاری نشده. سرور گواهی TLS ندارد، پس اتصال Looker بدون SSL است. ریسک با نقش فقط خواندنی و مصنوعی بودن داده محدود شده، ولی در محیط واقعی باید حتماً برطرف شود.
- چت‌بات فقط به پرسش‌های پیش‌بینی‌شده پاسخ می‌دهد. این انتخاب آگاهانه بود (بخش ۸)، ولی محدودیت است.

کارهای بعدی

1. آزمون A/B روی جلسه گلوگاهی برای سنجش اثر واقعی اصلاح.
2. مدل پیش‌بینی ریزش برای هشدار زودهنگام، پیش از آنکه دانش‌آموز رها کند.
3. خودکارسازی به‌روزرسانی ویوها به صورت `MATERIALIZED VIEW` برای دیتاست بزرگ‌تر.
4. TLS روی پایگاه داده و محدودکردن فایروال به بازه آی‌پی Looker Studio.

مراجع و پیوندها

- سامانه زنده: bi.packup.ir • دروازه پروژه‌ها: packup.ir
- ساختار دامنه از پلتفرم این‌مسیر (panel.inmasir.ir) الگوبرداری شده – محصولی که دوستان نگارنده در دانشگاه صنعتی شریف توسعه داده‌اند.
- مستندات فنی در مخزن: `docs/FIELD_LABELS.md` ، `docs/DASHBOARD.md` ، `docs/DATA.md` ، `docs/ERD.md`
- Goldratt, E. M. – The Goal (نظریه محدودیت‌ها؛ مبنای مفهومی «گلوگاه» در این گزارش)
- Kelley, T. L. (1939) – مبنای آماری ضریب تمیز سؤال با گروه‌های 27% بالا و پایین