

Dataset Documentation (Data Sheet + Dictionary)

۴/۹/۱ مقدمه و هدف

را به طور کامل مستند `nobatnou.ir` این سند مجموعه داده‌های سیستم نوبتدهی می‌کند. هدف، شفافسازی در مورد **منبع، کیفیت، مالکیت، محدودیت‌ها و ویژگی‌های** داده‌ها است تا اطمینان حاصل شود که داده‌ها به عنوان یک «دارایی» مدیریت شده و از تبدیل شدن به «بدهی» (به دلیل مشکلات قانونی، کیفیتی یا اخلاقی) جلوگیری شود.

۴/۹/۲ کارت مجموعه داده (Data Sheet)**

A. اطلاعات کلی

نام مجموعه داده	NobatNou Appointment System Dataset	
نگارش (Version)	1.0	
تاریخ ایجاد	**۰۱/۰۷/۱۴۰۲	
تاریخ آخرین بهروزرسانی	**۱۴۰۳/۰۲/۲۰	
مالک (Owner)	`nobatnou.ir` شرکت نوبت نو	

| مالکیت خصوصی - استفاده داخلی تنها برای ارائه | * مجوز استفاده** (License) |

| سرویس نوبتدهی |

| زبان** | فارسی (با الفبای فارسی و اعداد انگلیسی)** |

**B. ترکیب مجموعه داده # ##

| نوع داده | تعداد رکوردها (تقریبی) | درصد از کل |

| ----- | ----- | ----- |

| کاربران ثبت‌نام شده | ۱,۲۰۰ | ۲۰٪ |

| نوبت‌های ثبت‌شده | ۴,۸۰۰ | ۸۰٪ |

| **۶,۰۰۰** | **۱۰۰٪** | **جمع** |

**C. منبع داده‌ها # ## #

- منبع: **داده‌ها به‌طور مستقیم از کاربران نهایی (بیماران) از طریق فرم‌های وب‌سایت ** جمع‌آوری می‌شوند.

- **روش جمع‌آوری**

.۱. فرم ثبت‌نام (نام کامل، شماره تلفن، رمز عبور).

.۲. فرم درخواست نوبت (نام بیمار، شماره تلفن).

.۳. تعاملات کاربر (ورود، لغو نوبت، مشاهده وضعیت).

- **.جمع‌آوری نمی‌شود (Third-Party) هیچ داده‌ای از منابع ثالث**

**D. مالکیت و مجوزها

مالکیت: کلیه داده‌ها متعلق به **کاربران** (صاحبان داده) است -

- **مجوز استفاده:** کاربران با پذیرش **سیاست حریم خصوصی** (Privacy Policy)، به شرکت نوبت نو مجوز (Terms of Service) و **شرایط استفاده** (Terms of Service) محدود و غیرانحصاری برای پردازش داده‌هایشان به منظور ارائه سرویس نوبتدهی را می‌دهند.

- **محدودیت‌های استفاده**

داده‌ها **فقط** برای ارائه سرویس اصلی (نوبتدهی، پیگیری وضعیت، بهبود تجربه - کاربری) استفاده می‌شوند.

فروش، اجاره یا اشتراک داده‌ها با شرکت‌های ثالث مطلقاً ممنوع است -

- کامل و تأیید اخلاقی anonymization استفاده برای **تحقیقات دانشگاهی** تنها با امکان‌پذیر است.

**E. حریم خصوصی و اخلاق

- **(Consent): رضایت**

رضایت صریح از طریق چکباکس «موافقت با سیاست حریم خصوصی» دریافت می‌شود.

رضایت مجزا برای هر استفاده ثانویه (مثلًا ارسال پیامک تبلیغاتی) دریافت خواهد شد -

- **Anonymization:** داده‌ها برای اهداف تحلیلی به صورت ناشناس (با حذف نام و شماره تلفن) درمی‌آیند.

- حقوق کاربران: کاربران می‌توانند درخواست دسترسی، تصحیح یا حذف داده‌های **(و قوانین ایران GDPR مطابق)** خود را دهنند.

** دیکشنری داده ۴/۹/۳ (Data Dictionary)**

** جدول: `wp_nobat_users`**

نام ستون	نوع داده	توضیح	مقادیر مجاز/نمونه	کلید	Nullable?
`id`	INT(11)	شناسه منحصر به فرد کاربر	۱, ۲, ۳, ...	Primary Key	خیر
`full_name`	VARCHAR(100)	نام کامل کاربر	"علی محمدی"	خیر	-
`phone_number`	VARCHAR(20)	شماره تلفن همراه (۱۱ رقم)	"۰۹۱۲۱۲۳۴۵۶۷"	Unique Key	خیر
`password`	VARCHAR(255)	رمز عبور هش شده (با bcrypt)	"\$2y\$10\$..."	خیر	-
`created_at`	DATETIME	زمان ایجاد حساب	"۱۵-۰۱-۲۰۲۴ ۱۴:۳۰:۰۰"	خیر	-

** جدول: `wp_nobat_system`**

نام ستون	نوع داده	توضیح	مقادیر مجاز/نمونه	کلید	Nullable?

	شناسه منحصر به فرد نوبت `id` INT (11) ۱ , ۲ , ۳ , ... خیر Primary Key				
نام بیمار `patient_name` VARCHAR(100) خیر (فاطمه احمدی)					
					-
شماره تلفن بیمار `phone_number` VARCHAR(20) ۰۹۳۵۱۲۳۴۵۶۷ خیر Index					
`status` VARCHAR(20) pending, in-progress, completed, cancelled خیر Index					
`position` INT (11) ۱ , ۲ , ۳ , ... خیر Index					
زمان ایجاد نوبت `created_at` DATETIME ۲۰۰۳-۲۰۲۴ خیر - ۰۹:۱۵:۰۰					

* ** ویژگی‌های محاسبه شده (Derived Features)**

نام ویژگی فرمول/منطق هدف		
15 * (قبل از نوبت فعلی pending تعداد نوبت‌های) * زمان تخمینی ویزیت *		
دقیقه پیش‌بینی زمان انتظار		
(completed برای نوبت‌های) `مدت انتظار` (زمان تکمیل - زمان ایجاد) *		
تحلیل عملکرد سیستم		
نرخ لغو * `(تعداد نوبت‌های لغو شده / کل نوبت‌ها) * ۱۰۰` سنجش رضایت *		
کاربر		

۴/۹/۴ کیفیت داده (Data Quality)

(الف) کاملبودن (Completeness) و مقادیر مفقوده (Missingness)

- | ستون | نرخ مقادیر مفقوده (%) | اقدام اصلاحی |
 - | ----- | ----- | ----- |
 - | `full_name` | ۰٪ | الزامی در فرم ثبتنام |
 - | `phone_number` | ۱۱ رقمی | الزامی، اعتبارسنگی |
 - | `patient_name` | ۰٪ | الزامی در فرم نوبت |
 - | `status` | ۰٪ | `pending` مقدار پیشفرض |
 - | `position` | ۰٪ | MAX(position)+1 | محاسبه خودکار با

مقادیر مفقوده: ** در حال حاضر هیچ مقداری مفقوده نیست (به دلیل الزامی بودن فیلدها در فرم).

(ب) صحت (Accuracy)

- با** شود `regex ^[0-9]{11}\$` اعتبارسنگی می‌شود.

- نام: ** فقط حروف فارسی و فاصله مجاز است (اعتبارسنجی سمت سرور) -
- محاسبه می شود تا از `'MAX(position)+1`` موقعیت نوبت: با کوئری** تکراری بودن جلوگیری شود.

(ج) سازگاری (Consistency)**

- همه ** زمانی: قالب زمانی timestamp `YYYY-MM-DD HH:MM:SS` ها در قالب زمانی: ** می شوند (IRST) و زمان محلی ایران.
- همه شماره ها با حذف `+` یا `۰` اول به صورت ۱۱ رقمی ذخیره ** می شوند (مانند `۰۹۱۲۱۲۳۴۵۶۷`).

(د) یکتاپی (Uniqueness)**

- در جدول کاربران یکتا است (یک شماره، یک حساب) `phone_number`
- در یک لحظه می تواند تکراری باشد (اگر دو نوبت همزمان ثبت شوند) - `position` این یک نقطه ضعف** است که باید با تراکنش دیتابیس حل شود.

در داده ها (Bias) اریبی (BiAS) **

** منابع اریبی شناسایی شده**

| نوع اریبی | منبع احتمالی | تأثیر بر سیستم | راهکار کاهش |

| ----- | ----- | ----- | ----- | ----- |
کاربران فقط از **شیراز** هستند (محدوده | گزینشی**) |
جغرافیایی) | مدل‌های تحلیلی برای شهرهای دیگر قابل تعمیم نیستند | گسترش خدمات به
| شهرهای دیگر

داده‌ها در ساعت کاری (۸ صبح تا ۴ عصر) (Temporal Bias)** زمانی** |
بیشتر جمع‌آوری می‌شوند | پیش‌بینی‌ها برای شبها و آخر هفته نادرست است | جمع‌آوری
| داده در همه ساعت

فقط کاربران با دسترسی به اینترنت و تلفن | فناوری**) (Technology Bias)** |
همراه | محرومیت از سالماندان یا افراد با دسترسی محدود | ارائه امکان نوبت‌گیری تلفنی

| فقط فارسی‌زبانان | عدم پشتیبانی از سایر زبانی*) (Language Bias)** |
| زبان‌ها | افزودن پشتیبانی چندزبانه

#: اقدامات کاهش اربیی**

. ثبت **متاداده** (مانند شهر، زمان ثبت) برای تحلیل اربیی -

- متنوعسازی روش‌های ثبت نوبت (تلفنی، حضوری) -

- نظرسنجی دوره‌ی از کاربران برای شناسایی گروه‌های محروم -

۴/۹/۶ (Data Drift) و پایش (Monitoring)** تغییرات داده

Drift انواع **قابل رخداد

| تعريف | شاخص پايش | Drift نوع |

| ----- | ----- | ----- |

| **Concept Drift** | تغيير در رابطه بين ويزگيها و نتيجه (مثلاً تغيير ميانگين زمان ويزيت) | ميانگين زمان انتظار در هفته های مختلف

| **Data Drift** | تغيير در توزيع داده های ورودی (مثلاً افزایش ناگهانی نوبت های شبانه) | توزيع ساعت ثبت نوبت

| **Label Drift** | تغيير در مقادير خروجي (مثلاً افزایش نرخ لغو نوبت) | نرخ لغو ماهانه

Drift: ابزار های پايش

در بازه های زمانی `status` و `created_at` داشبورد داخلی: **نودار های** -

هشدار خودکار: **اگر نرخ لغو از ۲۰٪ عبور کند، ايميل به تيم فني ارسال شود** -

از تغييرات داده quarterly report تحليل دوره‌ی: **بررسی** -

۴/۹/۷ به داده امنیت و دسترسی

سطوح دسترسی

| نقش | دسترسی به داده ها | محدودیت ها |

| ----- | ----- | ----- |

| کاربر عادی** | فقط داده‌های خود | از طریق داشبورد کاربری** |

| ثابت، لاگ‌گیری IP | پشتیبانی** | داده‌های همه کاربران (فقط خواندن)** |

| توسعه‌دهنده** | دسترسی کامل (برای عیب‌یابی) | تنها در محیط تست با داده‌های ** | ساختگی

| کوتاه FA، session timeout | همه داده‌ها + توان حذف/تغییر | **۲

: رمزنگاری

- **: در انتقال TLS 1.3.

- **: در ذخیره: رمز عبور هش‌شده با** bcrypt.

- برنامه آینده: phone_number` رمزنگاری ستون** - AES-256.

: Backup & Recovery) (پشتیبان‌گیری و بازیابی ۴/۹/۸

: سیاست پشتیبان‌گیری

| نوع | فرکانس | محل نگهداری | مدت نگهداری |

| ----- | ----- | ----- | ----- |

- | **Full Backup** | سرور جداگانه + کلود (encrypted) | روزانه (ساعت ۲ بامداد) |
- | **Incremental Backup** | سرور جداگانه | ۷ روز | هر ۶ ساعت |
- | همان سرور | روز | ۱۴ روز | **لاغ‌های تراکنش** | Real-time |

تست باز پایه، **

بازیابی کامل داده‌ها** هر فصل** یکبار تست می‌شود -

- هدف RTO (Recovery Time Objective) < ۴ ساعت، RPO (Recovery Point Objective) < ۱۵ دقیقه.

- 3 -

مستندات و ایسته **۴،۹،۹**

سند توضیح لینک/مرجع
----- ----- -----
سیاست حریم خصوصی** توضیح نحوه جمع‌آوری و استفاده از داده**
`https://nobatnou.ir/privacy`
قوانین استفاده از سرویس **(TOS) شرایط استفاده**
`https://nobatnou.ir/terms`
(در صورت وجود) داخلی endpoints API** راهنمای** مستندات

| گزارش کیفیت داده** | گزارش ماهانه کیفیت دادهها | داخلی** |

** جمع‌بندی و اقدامات آینده # ## # # # ۰۱/۹/۴

** نقاط قوت داده‌های فعلی # ## # # #

- کامل‌بودن هیچ (missing value).

- صحت بالا (اعتبارسنجی سمت سرور) .

- مالکیت شفاف (تمامی داده‌ها از کاربران با رضایت گرفته شده‌اند)

** نقاط ضعف و بھبودهای آینده # ## # # #

| ضعف | اقدام بهبود | مهلت |

| ----- | ----- | ----- |

| در `phone_number` عدم رمزنگاری داده‌های حساس** | رمزنگاری**

| دیتابیس | ۶ ماه

| اریبی جغرافیایی** | گسترش خدمات به شهرهای دیگر | ۱۲ ماه**

| نداشتن داده‌های جمعیتی** | جمع‌آوری اختیاری سن و جنسیت (با رضایت) | ۳ ماه**

| Data Drift** راهاندازی dashboard با Grafana | عدم پایش خودکار**

| ۴ ماه

*نتیجه‌گیری ۱۱**۴/۹/۱۱

با کیفیت بالا، شفافیت در مالکیت و رعایت حریم `nobatnou.ir` مجموعه داده‌های خصوصی جمع‌آوری شده‌اند. با این حال، *پایش مستمر کیفیت، اریبی و تغییرات داده ** ضروری است. این سند به عنوان مرجع اصلی برای همه ذینفعان (توسعه‌دهنگان، مدیران، ناظران) عمل کرده و تضمین می‌کند که داده‌های سیستم به عنوان یک **دارایی مطمئن و پایدار** مدیریت شوند.