

## پروژه: ابزار تحت وب Process Mining

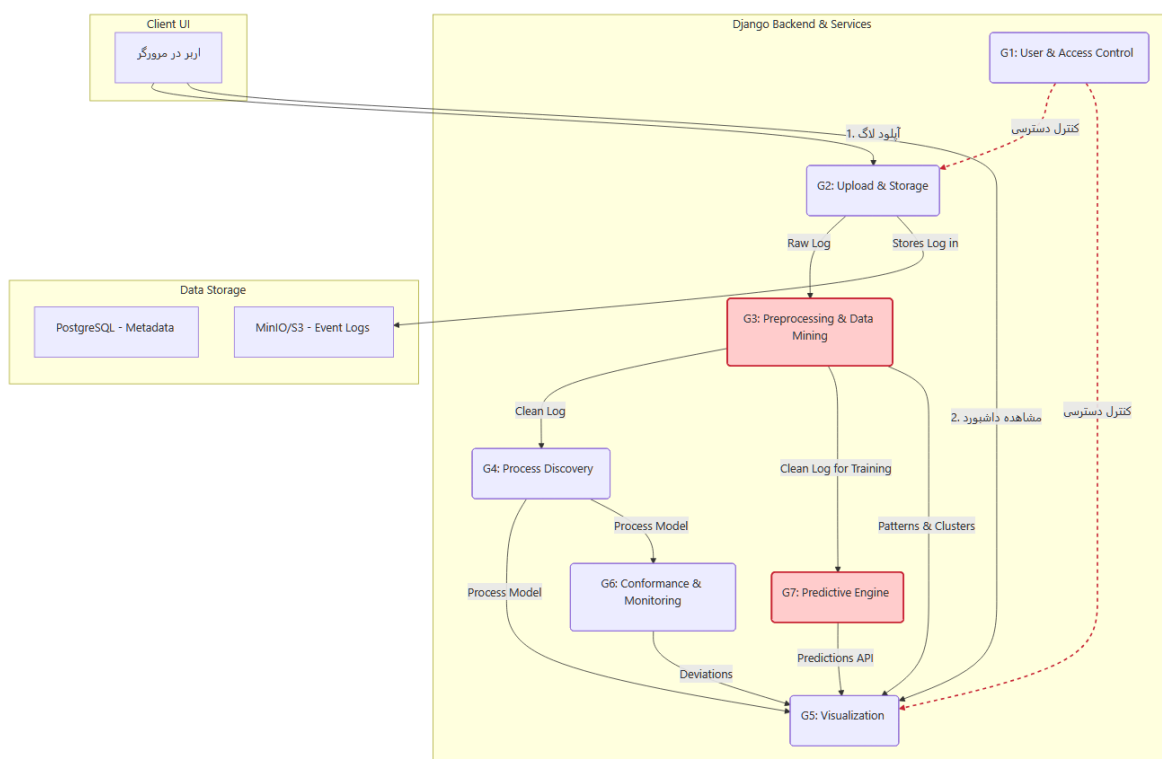
هدف از این پروژه ایجاد یک سیستم فرایند کاوی شبیه Disco است. دانشجویان در ۷ گروه دسته بندی میشوند و برخی گروه ها دو نفره و برخی سه نفره (بسته به سختی کار) است. پروژه در دو مقطع تحویل گیری میشود:

■ فاز ۱ (MVP - مهلت انتهای آبان ماه)

■ فاز ۲ (نسخه کامل - مهلت انتهای دی ماه)

در زیر معماری سیستم و وظایف دقیق هر گروه مشخص شده است. گروه بندی در کلاس انجام میشود باتوجه به علائق و توانمندی ها

### معماری سیستم



## تقسیم وظایف بین گروه ها

گروه	موضوع	تعداد نفرات	سطح پیچیدگی
گروه ۱	مدیریت کاربران و کنترل دسترسی	۲ نفره	استاندارد
گروه ۲	آپلود و ذخیره سازی Event Log	۲ نفره	استاندارد
گروه ۳	پیش پردازش و داده کاوی (Patterns & Clustering)	۳ نفره	بالا
گروه ۴	موتور کشف فرآیند (Process Discovery)	۲ نفره	استاندارد
گروه ۵	بصری سازی تعاملی (Interactive Visualization)	۲ نفره	متوسط
گروه ۶	بررسی انطباق و نظارت (Monitoring & Conformance)	۲ نفره	متوسط
گروه ۷	موتور پیش بینی فرآیند (Predictive Monitoring)	۳ نفره	بالا

## شرح دقیق وظایف هر گروه

### گروه ۱: مدیریت کاربران و کنترل دسترسی (۲ نفره)

- مسئولیت اصلی: پیاده سازی زیرساخت احراز هویت و مدیریت نقش ها.
- فاز ۱:
  - ساخت مدل های User, Role, (Admin, Analyst).
  - پیاده سازی صفحات ورود (Login)، ثبت نام (Register) و خروج (Logout).
- فاز ۲:
  - اعمال محدودیت دسترسی بر اساس نقش (مثلاً فقط Analyst بتواند لاگ آپلود کند).
  - ایجاد داشبورد ادمین برای مدیریت کاربران.
- تعامل: با تمام گروه هایی که نیاز به محدودیت دسترسی دارند (مخصوصاً ۲ و ۵).
- تکنولوژی ها: Django REST Framework, Django Authentication (برای API).

### گروه ۲: آپلود و ذخیره سازی Event Log (۲ نفره)

- مسئولیت اصلی: مدیریت آپلود، اعتبارسنجی اولیه و ذخیره سازی فایل های لاگ.
- فاز ۱:
  - ساخت فرم آپلود برای فایل های CSV و XES کوچک.
  - ذخیره فایل ها در فایل سیستم سرور و ثبت متادیتا در دیتابیس.
- فاز ۲:
  - بهینه سازی برای آپلود فایل های حجیم (Chunked Upload).
  - پشتیبانی از فرمت Parquet.
  - یکپارچه سازی با یک Object Storage مثل MinIO برای مقیاس پذیری.
- تعامل: داده خام را به گروه ۳ تحویل می دهد.
- تکنولوژی ها: Django Forms, Pandas, MiniO Client.

### گروه ۳: پیش‌پردازش و داده‌کاوی (۳ نفره – حجم کار بالا)

- مسئولیت اصلی: پاکسازی لاگ‌های خام و استخراج الگوهای پنهان با روش‌های داده‌کاوی.
- فاز ۱:
  - پیاده‌سازی pipeline پاکسازی داده با Pandas (مدیریت مقادیر گمشده، حذف نویز، مرتب‌سازی زمانی).
  - ارائه آمار توصیفی اولیه از لاگ (تعداد Case، فعالیت‌ها و ...).
- فاز ۲:
  - پیاده‌سازی الگوریتم Clustering برای گروه‌بندی مسیرهای مشابه فرآیند (Trace Clustering).
  - پیاده‌سازی الگوریتم Frequent Pattern Mining برای کشف توالی‌های پرتکرار فعالیت‌ها.
  - بهینه‌سازی پردازش‌ها با Dask برای کار با لاگ‌های بسیار بزرگ.
- تعامل: لاگ پاکسازی شده را به گروه ۴ و ۷ می‌دهد. نتایج کلاستر و الگوها را برای نمایش به گروه ۵ می‌دهد.
- تکنولوژی‌ها: Pandas, Dask, Scikit-learn (KMeans), Mlxtend.

### گروه ۴: موتور کشف فرآیند (۲ نفره)

- مسئولیت اصلی: تبدیل لاگ وقایع به مدل‌های فرآیندی قابل نمایش.
- فاز ۱:
  - دریافت لاگ پاک‌شده از گروه ۳.
  - پیاده‌سازی دو الگوریتم کشف فرآیند (Alpha Miner, Heuristics Miner) با استفاده از PM4Py.
- فاز ۲:
  - اضافه کردن الگوریتم‌های پیشرفته‌تر مانند Inductive Miner.
  - پشتیبانی از لاگ‌های OCEL برای تحلیل فرآیندهای پیچیده شیء-محور.
- تعامل: مدل فرآیند (BPMN/Petri Net) را به گروه ۵ و ۶ تحویل می‌دهد.
- تکنولوژی‌ها: PM4Py.

### گروه ۵: بصری‌سازی تعاملی (۲ نفره)

- مسئولیت اصلی: ساخت داشبورد کاربری برای نمایش نتایج تحلیل‌ها به صورت گرافیکی و تعاملی.
- فاز ۱:
  - نمایش مدل فرآیند دریافت شده از گروه ۴ به صورت یک گراف استاتیک.
  - نمایش آمار اولیه دریافت شده از گروه ۳.
- فاز ۲:
  - تبدیل گراف فرآیند به یک مدل تعاملی (Zoom, Pan, نمایش جزئیات روی یال‌ها و نودها).
  - نمایش نتایج کلاسترینگ (رنگ‌آمیزی مسیرها) و الگوهای پرتکرار از گروه ۳.
  - نمایش پیش‌بینی‌های زنده از گروه ۷ روی داشبورد (مثلاً نمایش زمان باقی‌مانده تخمینی برای یک Case فعال).
- تعامل: از گروه‌های ۳، ۴، ۶ و ۷ داده می‌گیرد و رابط کاربری نهایی را می‌سازد.
- تکنولوژی‌ها: D3.js, Plotly/Dash, Chart.js. یا کتابخانه‌های مشابه.

### گروه ۶: بررسی انطباق و نظارت (۲ نفره)

- مسئولیت اصلی: مقایسه لاگ واقعی با مدل فرآیند برای کشف انحرافات.
- فاز ۱:
  - پیاده‌سازی یک الگوریتم ساده Conformance Checking (مانند Token-based Replay).
  - گزارش انحرافات و نمایش Case های غیرمنطبق.
- فاز ۲:
  - محاسبه معیارهای انطباق پیشرفته‌تر (Fitness, Precision).

○ ساخت یک داشبورد نظارتی (Monitoring) برای نمایش وضعیت فرآیندها به صورت دوره‌ای.

• تعامل: مدل را از گروه ۴ و لاگ را از گروه ۳ می‌گیرد. نتایج انحرافات را به گروه ۵ می‌دهد.

• تکنولوژی‌ها: PM4Py (Conformance Checking modules).

**گروه ۷: موتور پیش‌بینی فرآیند (۳ نفره - حجم کار بالا)**

• مسئولیت اصلی: ساخت و ارائه مدل‌های یادگیری ماشین برای پیش‌بینی آینده یک فرآیند در حال اجرا.

• فاز ۱ (پیش‌بینی Outcome):

○ استفاده از لاگ پاک‌شده گروه ۳ برای آموزش یک مدل.

○ ساخت ویژگی از روی پیشوندهای فرآیند (Prefixes of cases).

○ آموزش یک مدل طبقه‌بندی (مانند Decision Tree یا Logistic Regression) برای پیش‌بینی نتیجه نهایی یک Case (مثلاً

Rejected/Approved) بر اساس فعالیت‌های اولیه.

• فاز ۲ (پیش‌بینی زمان و فعالیت بعدی):

○ پیاده‌سازی مدل‌های پیشرفته‌تر (مانند LSTM یا مدل‌های مبتنی بر توالی) برای پیش‌بینی زمان باقی‌مانده تا پایان Case یا فعالیت

بعدی محتمل.

○ ساخت یک API که case\_id یک فرآیند در حال اجرا را بگیرد و پیش‌بینی‌ها را برگرداند.

• تعامل: داده آموزشی را از گروه ۳ می‌گیرد. API پیش‌بینی را در اختیار گروه ۵ قرار می‌دهد تا در UI نمایش داده شود.

• تکنولوژی‌ها: Scikit-learn, Keras/TensorFlow, Pandas, Django REST Framework.