نموذج توقع انسحاب المستخدمين (Customer Churn)

يوضح المستند خطوات تطوير نموذج لتوقع انسحاب المستخدمين (Customer Churn) من منصة إلكترونية تعتمد على نموذج الاشتراكات. يعتمد المشروع على سجلات نشاط المستخدمين في خدمة بث موسيقا. وبناءً على سلوك المستخدمين في استخدام المنصة يتنبأ نموذج التعلم الآلي على المستخدمين الذين من المرجّح أن يلغوا اشتراكهم.

Pipeline

# أولًا: خدمة تنظيف البيانات وفهم تحدياتها

خدمة تنظيف البيانات

## تحليل البيانات:

أهم

# ثانيًا: خدمة بناء Features

## هندسة الخصائص:

ما هي أهم المزايا

سبع

## تحدي عدم توازن الفئات (Class Imbalance):

كيف يمكن أن يتم حله

## معالجة تسرب بيانات (Data Leakage).

ما هو وماذا يعني وكيف يحدث

# ثالثًا: خدمة تدريب النموذج

(شرح النموذج)

توثيق جميع التجارب على النماذج المختلفة في ML Flow

## اختيار النموذج:

اختيار نموذج ملائم للمشكلة وشرحه بشكل واضح

## تدريب النموذج:

اختيار نموذج ملائم للمشكلة وشرحه بشكل واضح

# رابعًا: خدمة تقييم واختيار أفضل نموذج

## تقييم النموذج:

تقييم الأداء باستخدام المقاييس المناسبة.

## تحليل أخطاء النموذج

تحليل الأخطاء

# خامسًا: خدمة مراقبة الأداء وانحراف البيانات

خدمة لمراقبة أداء النموذج وانحراف البيانات

# سادسًا: خدمة إعادة التدريب

## بناء نظام لإعادة التدريب الدوري للنموذج لمعالجة التغير في البيانات.

عند إطلاق نموذج للتعامل مع بيانات في (production) قد تختلف البيانات في بيئة (production) مع مرور الوقت عن البيانات التي دُرب عليها النموذج هذا الاختلاف قد يكون في توزيع البيانات (Data Drift) وقد يكون في طبيعية العلاقات بين المدخلات والمخرجات (Concept Drift) بمعنى أن يقل تأثير بعض المزايا على المخرجات وقد يزداد تأثير مزايا أخرى.

يضمن وجود هذا النظام وجود نسخة مختلفة من النموذج تدرب بشكل تلقائي على البيانات الجديدة بحيث يظل النموذج متوافقًا مع طبيعة البيانات الجديدة، يتم هذا عن طريق:

* تحديد آلية Triggering قد تكون دورية مثلًا كل شهر أو عند عتبه محددة
* تجهيز Pipeline لتنظيف وتحليل البيانات ومن ثمَّ تدريب النموذج عليها
* آلية لحفظ نسخة مختلفة من النموذج باستخدام MLFlow
* اختبار النموذج الجديد واستبداله مع النموذج الحالي

# سابعًا: خدمة Inferencing

تهيئة API باستخدام FastAPI

## تقديم النموذج عبر (API) FastAPI.

# تحديات تقنية

# تحسينات للمشروع

## ملاحظات:

* تم تعريب جميع المصطلحات التقنية في المستند باستخدام [معجم البيانات والذكاء الاصطناعي](https://siwar.ksaa.gov.sa/public-dict-information/4cd164a7-7160-4de8-af5c-34382f5da657)

وهو معجم تم بناءه من قبل مَجمع الملك سلمان العالمي للغة العربية بالتعاون مع الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا)