الجمهـوريـة العربيــة الســـورية

المعهد العالي للعلوم التطبيقية والتكنولوجيا

قسـم المعلوميات

العام الدراسي 2025/2024

**مشروع تخرج**

أعد لنيل درجة الإجازة في هندسة البرمجيات والذكاء الصنعي

منصة لخدمات الذكاء الصنعي موجهة لمواقع التجارة الإلكترونية

تقديم

آرام محمد

إشراف

د. رياض سنبل

5/8/2025

جدول المحتويات

Contents

[1 الفصل الأول 3](#_Toc204905235)

[1.1 مقدمة 3](#_Toc204905236)

[2.1 الهدف من المشروع 3](#_Toc204905237)

[3.1 المتطلبات الوظيفية 4](#_Toc204905238)

[4.1 المتطلبات غير الوظيفية 4](#_Toc204905239)

[2 الفصل الثاني 5](#_Toc204905240)

[1.2 توقع انسحاب العملاء Churn Prediction. 5](#_Toc204905241)

[2.2 المحولات 5](#_Toc204905242)

[3 الفصل الثالث 6](#_Toc204905243)

[1.3 مقدمة 6](#_Toc204905244)

[2.3 توقع انسحاب العملاء Churn Prediction. 6](#_Toc204905245)

[1.2.3 تطور تقنيات توقع انسحاب العملاء: من التقنيات التقليدية إلى التعلم العميق. 6](#_Toc204905246)

[2.2.3 استخدام المحولات في توقع انسحاب العملاء. 7](#_Toc204905247)

[3.2.3 استخدام المحولات في توقع انسحاب عملاء بناءً على سلوك الاستخدام اليومي. 7](#_Toc204905248)

[4.2.3 استخدام المحولات مع تمثيلات RFM الزمنية. 8](#_Toc204905249)

[5.2.3 تنبؤ وتمثيل مبني على سلاسل نشاط العملاء في مواقع التجارة الالكترونية. 9](#_Toc204905250)

[3.3 مجموعة بيانات في مجال التجارة الإلكترونية 9](#_Toc204905251)

[1.3.3 مجموعة بيانات E-commerce Customer Data For Behavior Analysis. 9](#_Toc204905252)

[2.3.3 مجموعة بيانات Brazilian E-Commerce Public Dataset by OLIST. 10](#_Toc204905253)

[4.3 دراسة التطبيقات المشابهة 11](#_Toc204905254)

[4 الفصل الرابع 12](#_Toc204905255)

[1.4 مخططات حالات الاستخدام 12](#_Toc204905256)

[2.4 السرد النصي لحالات الاستخدام 14](#_Toc204905257)

[1.2.4 إنشاء حساب جديد. 14](#_Toc204905258)

[2.2.4 رفع بيانات المستخدم عن طريق ملف csv. 15](#_Toc204905259)

[3.2.4 رفع البيانات عن طريق تقديم رابط ومفتاح اتصال API. 17](#_Toc204905260)

[4.2.4 استعراض العملاء المتوقع انسحابها. 18](#_Toc204905261)

[5.2.4 استعراض شرائح العملاء. 19](#_Toc204905262)

[6.2.4 استعراض معلومات عن القيمة الدائمة للعملاء. 19](#_Toc204905263)

[5 الفصل الخامس 21](#_Toc204905264)

[1.5 مقدمة 21](#_Toc204905265)

[2.5 النموذج المقترح لتوقع انسحاب العملاء 21](#_Toc204905266)

[1.2.5 بنية النموذج المقترح لتوقع انسحاب العملاء 22](#_Toc204905267)

[حيث كانت المعاملات الفوقية (Hyper-Parameters) هي: 23](#_Toc204905268)

[2.2.5 تحضير البيانات 23](#_Toc204905269)

[3.2.5 منهجية التدريب 24](#_Toc204905270)

[3.5 النموذج المقترح لتقسيم العملاء إلى شرائح 24](#_Toc204905271)

[4.5 النموذج المقترح لحساب القيمة الدائمة للعميل 25](#_Toc204905272)

[6 الفصل السادس 26](#_Toc204905273)

[1.6 مقدمة 26](#_Toc204905274)

[2.6 خدمة البيانات Data service 27](#_Toc204905275)

[3.6 خدمة المصادقة Auth service 27](#_Toc204905276)

[4.6 خدمة توقع انسحاب العملاء Churn service 28](#_Toc204905277)

[5.6 خدمة تقسيم العملاء إلى شرائح Customer Segmentation Service 28](#_Toc204905278)

[6.6 خدمة القيمة الدائمة للعملاء Customer Lifetime Value 28](#_Toc204905279)

[7.6 خدمة البوابة API-Gateway 29](#_Toc204905280)

[8.6 خدمة واجهة المستخدم Front-end 29](#_Toc204905281)

[9.6 مخطط تصميم النظام 30](#_Toc204905282)

[7 الفصل السابع 31](#_Toc204905283)

[1.7 PostgreSQL 31](#_Toc204905284)

[2.7 React 31](#_Toc204905285)

[3.7 Python 31](#_Toc204905286)

[4.7 FastAPI 31](#_Toc204905287)

[5.7 Json Web Token (JWT) 31](#_Toc204905288)

[6.7 Git 32](#_Toc204905289)

[7.7 GitHub 32](#_Toc204905290)

[8.7 Redis 32](#_Toc204905291)

[9.7 PyTorch 32](#_Toc204905292)

[10.7 Jupyter Notebook 32](#_Toc204905293)

[8 الفصل الثامن 33](#_Toc204905294)

[1.8 مقدمة 33](#_Toc204905295)

[2.8 بنية النظام ومكوناته 34](#_Toc204905296)

[3.8 خدمة توقع انسحاب العملاء 35](#_Toc204905297)

[1.3.8 تفاصيل التنجيز 35](#_Toc204905298)

[4.8 خدمة القيمة الدائمة للعميل 37](#_Toc204905299)

[1.4.8 تفاصيل التنجيز 37](#_Toc204905300)

[5.8 خدمة تقسيم البيانات إلى شرائح 39](#_Toc204905301)

[1.5.8 تفاصيل التنجيز 39](#_Toc204905302)

[6.8 خدمة البيانات 40](#_Toc204905303)

[1.6.8 تفاصيل التنجيز 40](#_Toc204905304)

[7.8 خدمة المصادقة 41](#_Toc204905305)

[1.7.8 تفاصيل التنجيز 41](#_Toc204905306)

[8.8 خدمة البوابة 41](#_Toc204905307)

[1.8.8 تفاصيل التنجيز 41](#_Toc204905308)

[9.8 خدمة وجهات المستخدم 42](#_Toc204905309)

[9 الفصل التاسع 44](#_Toc204905310)

[1.9 اختبار النموذج المقترح لتوقع انسحاب العملاء 44](#_Toc204905311)

[1.1.9 منهجية التقييم المتبعة 44](#_Toc204905312)

الاختصارات

|  |  |  |
| --- | --- | --- |
| التكرار والتردد والقيمة النقدية | Recency, Frequency, Monetary Value | RFM |
| التعلم الآلي | Machine Learning | ML |
| آلة المتجهات الداعمة | Support Vector Machine | SVM |
| الغابات العشوائية | Random Forests | RF |
| الشبكات العصبية | Neural Networks | NN |
| التعلم العميق | Deep Learning | DL |
| الشبكات العصبية المتكررة | Recurrent Neural Networks | RNN |
|  |  | LSTM |
| تطبيقات الصفحة الواحدة | Single Page Application | SPA |
|  |  | CLTV |

# الفصل الأول

التعريف بالمشروع

يتضمن هذا الفصل التعريف بالمشروع ومتطلباته.

## مقدمة

في العصر الحديث للتجارة الرقمية، أصبحت القدرة على فهم سلوك العملاء واتخاذ قرارات مدعومة بالبيانات عنصراً جوهرياً لنجاح مواقع التجارة الإلكترونية. وفي هذا السياق، تبرز منصات الذكاء الصنعي كحلول استراتيجية تمكّن هذه المواقع من تحسين أدائها وتعزيز ولاء العملاء. توفر هذه المنصات مجموعة من الخدمات الذكية، مثل التنبؤ بانقطاع العملاء، وتصنيفهم إلى شرائح ذات أنماط سلوكية متقاربة، وتقدير القيمة المستقبلية لكل عميل والعائد المتوقع منه. ومع ذلك، فإن الاستفادة من هذه القدرات غالباً ما تتطلب خبرات متقدمة في علوم البيانات وتحليلات الأعمال، مما قد يشكل حاجزاً أمام العديد من المواقع. ومن هنا تبرز الحاجة إلى منصة توفر هذه الخدمات بطريقة مؤتمتة وسهلة الاستخدام، لتتيح لمواقع التجارة الإلكترونية، بمختلف أحجامها، الاستفادة من الذكاء الاصطناعي في تعزيز استراتيجياتها التسويقية وتحقيق نمو مستدام.

## الهدف من المشروع

يندرج عملنا في هذا المشروع ضمن سياقين: السياق الأول نتطرق فيه للمسألة من منظور الذكاء الصنعي، حيث نحاول الاستفادة من آخر التطورات في مجال تعلم الآلة والتعلم العميق للوصول إلى نماذج ذكاء صنعي توفر خدمات تعزز استراتيجياتها وتزيد من أرباحها. والسياق الثاني نتطرق فيه إلى بناء تطبيق برمجي يستفيد من نماذج الذكاء الصنعي التي توصلنا إليها في السياق الأول للوصول إلى تطبيق مفيد عملياً، مع مراعاة أسس ومبادئ هندسة البرمجيات.

## المتطلبات الوظيفية

يجب أن يقدم النظام للمستخدم ما يلي:

1. السماح بإنشاء حساب جديد ضمن النظام
2. السماح بتسجيل الدخول من حساب مُنشأ سابقا.
3. السماح للمستخدم برفع بيانات عملاء موقع التجارة الإلكترونية الخاصة به بطرق متعددة.
4. السماح للمستخدم باستعراض بيانات تحليلية عامة حول العملاء الخاصة به.
5. السماح للمستخدم باستعراض عملاء موقع التجارة الالكترونية الخاصة به الذين يتوقع انسحابهم.
6. السماح للمستخدم باستعراض معلومات القيمة الدائمة للعميل لدى موقع التجارة الالكترونية الخاص به.
7. السماح للمستخدم باستعراض بيانات تقسيم العملاء إلى شرائح.

## المتطلبات غير الوظيفية

1. يجب أن يكون النظام آمناً، حيث يسمح فقط للمستخدمين المسجّلين باستخدامه.
2. يجب أن يوفر النظام واجهات سهلة الاستخدام وجيدة المظهر.
3. يجب أن يكون الكود البرمجي قابلاً للتعديل والصيانة.
4. يجب أن يكون النظام قابلاً للتوسع.
5. يجب أن يكون النظام قادراً على التعامل مع حجم بيانات كبير نسبيا مليون عملية شراء للعملاء.

# الفصل الثاني

الدراسة النظرية

يوضّح هذا الفصل مجموعة من المفاهيم النظرية المستخدمة ضمن العمل المقدم.

## توقع انسحاب العملاء Churn Prediction.

يعد انسحاب العملاء (Customer Churn) باعتبارها إحدى أبرز التحديات التي تواجه الشركات في مختلف القطاعات، لاسيما في بيئات الأعمال التنافسية مثل التجارة الإلكترونية والاتصالات. يُعد فهم أسباب انسحاب العملاء والتنبؤ به عاملاً حاسماً في صياغة استراتيجيات فعالة للاحتفاظ بالعملاء، خاصةً أن تكلفة اكتساب عميل جديد تُقدَّر بأنها أعلى بعشر مرات من تكلفة الاحتفاظ بعميل حالي [CASPER]. وقد دفع هذا المبدأ الاقتصادي العديد من الشركات إلى الاستثمار في تطوير نماذج تحليلية وتنبؤية تعتمد على تقنيات الذكاء الاصطناعي والتعلم الآلي، بهدف اكتشاف الأنماط السلوكية التي تسبق الانسحاب [a Data Driven approach].

وفي سياق التجارة الإلكترونية، يكتسب تحليل انسحاب العملاء أهمية متزايدة، نظراً لاعتماد هذه المنصات بشكل كبير على ولاء العملاء وتكرار عمليات الشراء. فقد يؤدي انسحاب عدد محدود من العملاء إلى تأثير كبير على الإيرادات، خاصة مع ارتفاع تكلفة الإعلانات والاستحواذ على عملاء جدد. ولهذا، تعتمد المتاجر الإلكترونية الناجحة على تتبع سلوك العملاء في الوقت الفعلي، وتحليل بياناتهم مثل تكرار الشراء، وزمن التفاعل، وسلال التسوق المهجورة، لتحديد المؤشرات المبكرة للانسحاب واتخاذ تدابير استباقية للحد منه. من خلال تقليل معدل churn، يمكن للشركات تحسين كفاءة الإنفاق التسويقي وزيادة العائد على الاستثمار، ما يجعل من الاحتفاظ بالعملاء أولوية استراتيجية لها تأثير مباشر على الربحية والاستدامة [data driven approach].

## الانتباه (Attention)

الانتباه هو آلية تعلم تمكن النماذج من التركيز على أجزاء معينة من بيانات الدخل عند توليد المخرجات، حيث تسمح للنموذج بوزن أهمية المدخلات المختلفة بشكل ديناميكي، مما يعزز قدرته على التقاط العلاقات والتبعيات داخل البيانات، بعض النظر عن المسافة بينها ضمن سلسلة الدخل. [attention is all you need]

## المحولات Transformers

هي بنية شبكة عصبونية تستخدم في المهام التي تتضمن معالجة بيانات متسلسلة، مثل معالجة اللغة الطبيعية، وهي تستخدم آليات الانتباه بشكل أساسي، حيث تسمح بمعالجة بيانات الإدخال بالتوازي بدلاً من التتابع. تُعد هذه البنية أكثر كفاءة من الشبكات التقليدية مثل RNN وLSTM، حيث تتفوق في فهم العلاقات الطويلة بين عناصر السلسلة وتُسهّل تدريب النماذج على نطاق واسع. [Attention is all you need]

## خوارزمية K-means

تُعد خوارزمية K-Means من أكثر خوارزميات التجميع (Clustering) استخداماً في تعلم الآلة، وخصوصاً في تحليل البيانات الاستكشافية وتقسيم العملاء في أنظمة دعم القرار. تهدف هذه الخوارزمية إلى تقسيم مجموعة من البيانات إلى K عنقود غير متداخلين، بحيث تكون البيانات ضمن كل عنقود متقاربة قدر الإمكان، وتكون المسافة إلى مركز العنقود (Centroid) أقل ما يمكن.

تبدأ الخوارزمية بتحديد عدد العناقيد 𝐾 واختيار مراكز أولية عشوائية لها، ثم تتكرر خطوتان رئيسيتان حتى يحدث الاستقرار:

* إسناد كل نقطة بيانات إلى أقرب مركز عنقود باستخدام مسافة إقليدية.
* تحديث مراكز العناقيد كمتوسط للنقاط المنتمية لها.

تستمر العملية إلى أن تتوقف مراكز العناقيد عن التغير أو يتم الوصول إلى عدد محدد من التكرارات.

تُستخدم K-Means في تطبيقات متنوعة تشمل: تحليل شرائح العملاء، تجزئة الصور، تصنيف المستندات، واكتشاف الأنماط السلوكية. وتُعد مناسبة بشكل خاص للبيانات العددية القابلة للقياس. [Data-clustering]

## القيمة الدائمة لحياة العميل Customer Lifetime Value

تُعد قيمة عمر العميل (Customer Lifetime Value - CLTV) من المفاهيم الأساسية في التسويق الحديث، وتُستخدم كأداة استراتيجية لتقدير العائد المالي المتوقع من العميل طوال فترة تعامله مع الشركة. تكتسب CLTV أهمية متزايدة في البيئات الرقمية مثل التجارة الإلكترونية، حيث تسهم في توجيه القرارات المتعلقة بالاستحواذ على العملاء، وتخصيص الميزانيات التسويقية، وتصميم استراتيجيات الاحتفاظ. [CLTV]

## نموذج RFM

نموذج RFM (Recency, Frequency, Monetary) هو أداة تحليلية تُستخدم لفهم سلوك العملاء في التسويق، حيث يتم تقسيم العملاء استناداً إلى ثلاثة معايير رئيسية: الحداثة (كم من الوقت مضى منذ آخر شراء)، التكرار (عدد مرات الشراء)، والقيمة النقدية (إجمالي المبالغ المنفقة). يساعد هذا النموذج الشركات على استهداف العملاء بشكل أكثر فاعلية، وتحسين الحملات التسويقية، وزيادة العوائد على الاستثمار عن طريق تخصيص العروض بناءً على سلوك العملاء. رغم فعاليته، يتطلب النموذج تحديثاً دورياً للبيانات لضمان دقة التصنيفات. [Marketing Management]

# الفصل الثالث

الدراسة المرجعية

يعرض هذا الفصل الأبحاث والبيانات المرتبطة بالعمل المقدم.

## مقدمة

إن مجال تحليل بيانات مستخدمين مواقع التجارة الالكترونية هو مجال واسع ويحتوي على العديد من التحاليل والمخططات التي تفيد مدراء هذه المواقع في فهم تعامل العملاء مع الموقع وتوقع الأرباح والقيام بالإجراءات الوقائية لتحقيق أكبر ربح، سنستعرض في الفقرات التالية بعض التحاليل والقياسات التي استخرجت من الأبحاث ضمن هذا المجال.

## توقع انسحاب العملاء Churn Prediction.

### تطور تقنيات توقع انسحاب العملاء: من التقنيات التقليدية إلى التعلم العميق.

في توقع انسحاب العملاء، غالبا ما تشمل الطرق التقليدية النماذج الإحصائية والخوارزميات مثل نموذج RFM أو سلاسل ماركوف ولكن غالبا ما تكون قدرتها التنبؤية محدودة، نتيجة لذلك تم تطبيق تقنيات التعلم الآلي Machine Learning في هذا المجال وساهمت بشكل كبير في تطويره، حيث أظهرت الدراسات فعالية خوارزميات مثل Support Vector Machine (SVM)، والغابات العشوائية Random Forests (RF)، والشبكات العصبية Neural Networks (NNs) وبالرغم من تفوقها إلا أنها غالبا ما تكون غير قادرة على نمذجة الأنماط الزمنية أو نمذجة الأنماط التتابعية في بيانات العملاء [Improved Customer].

ومع التطورات في مجال التعلم العميق Deep Learning أظهرت الشبكات العصبية التكرارية (RNNs) Recurrent Neural Networks ومحولات البيانات Transformers إمكانيات واعدة في نمذجة بيانات سلوك العملاء المتغيرة مع الزمن، وتم استخدامها لالتقاط الأنماط التتابعية في بيانات العملاء، مما أدى إلى تفوقها على الطرق التقليدية وطرق التعلم الآلي في العديد من الحالات [Exploiting RFM].

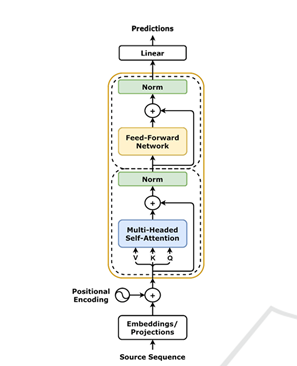
### استخدام المحولات في توقع انسحاب العملاء.

في توقع انسحاب العملاء، ظهرت بنى المحول Transformer كأداة قوية لتحليل بيانات سلوك العملاء المتسلسلة. ويأتي تفوق هذه البنى في التقاط الأنماط طويلة المدى ضمن بيانات السلاسل الزمنية دون الحاجة الى هندسة ميزات معقدة. وخصوصا نماذج المحولات التي تستخدم آليات الانتباه متعددة الرؤوس Multi-Head Attention، فهي فعالة عند التعامل مع بيانات مختلطة، تجمع بين مقاييس RFM المتغيرة زمنياً والمتغيرات الثابتة للعملاء. ان المعالجة المباشرة لبيانات السلاسل الزمنية بواسطة المحولات تقلل من الحاجة إلى هندسة ميزات معقدة. كما نماذج المحول توفر تقديرات لعدم اليقين، مما يساعد في حملات مخصصة للعملاء أكثر استهدافا، للحد من انسحاب العملاء وتقدير خطر انسحاب كل عميل [Exploit].

### استخدام المحولات في توقع انسحاب عملاء بناءً على سلوك الاستخدام اليومي.

جرى استخدام المحول في قطاع B2B، تحديداً في مجال الاتصالات، للتنبؤ بانسحاب العملاء بناءً على بيانات سلوك الاستخدام اليومية. يتميز هذا النموذج بقدرته على التعامل مع بيانات سلاسل زمنية متعددة المتغيرات، وهو ما يسمح بتحليل سلوك العملاء المتغير بشكل دقيق عبر الزمن [B2B].

#### بنية النموذج:



الشكل 1 المحولات في توقع انسحاب العملاء بناءً على سلوك الاستخدام اليومي [B2B]

* تم استبدال طبقة التضمين (Embedding) التقليدية بـ 1D Depth wise Convolution.
* تم استخدام 8 طبقات Transformer مع رؤوس الانتباه المتعدد (Multi-Head Attention).
* يتم استخدام Global Average Pooling لتقليل حجم البيانات الناتجة وتحويلها إلى تمثيل نهائي.
* تم تطبيق وظيفة تفعيل Sigmoid في الطبقة النهائية لإنتاج احتمالية انسحاب العميل [3].

#### طبيعة البيانات المستخدمة:

كانت هي بيانات يومية على مستوى الاشتراك مثل: طول فترة التعاقد، عدد المكالمات، استهلاك البيانات، وعدد الباقات [3].

### استخدام المحولات مع تمثيلات RFM الزمنية.

تم بتطبيق نماذج Transformers على مقاييس RFM المتغيرة عبر الزمن، بهدف تحسين التنبؤ بانسحاب العملاء في قطاع الخدمات المالية. يُعدّ هذا النهج مهماً لأن مقاييس RFM تعتبر من المؤشرات الأساسية لتقييم سلوك العملاء وقدرتهم على التنبؤ بالانسحاب [2].

#### بنية النموذج.

* دمج البيانات الثابتة والمتغيرة: تم دمج مقاييس RFM مع البيانات الثابتة باستخدام آلية Concatenation أو Merging.
* Attention Mechanism: تم استخدام آلية Multi-Head Attention لتعزيز قدرة النموذج على التركيز على أهم السمات الزمنية [2].

#### طبيعة البيانات المستخدمة.

مجموعة بيانات هي بيانات أحد كبار مزودي الخدمات المالية في فرنسا، وتحتوي على سجلات شهرية للعملاء. تتكون العينة المقدمة من عملاء تعتبر الشركة موفرهم الأساسي للخدمات المالية، وذلك لضمان جودة عالية في بيانات السلوك. تحتوي قاعدة البيانات على متغيرات تستخدم كثيراً في التنبؤ بانسحاب العميل، مثل الخصائص الديموغرافية والمعلومات السلوكية لحساب متغيرات RFM المتتابعة [2].

### تنبؤ وتمثيل مبني على سلاسل نشاط العملاء في مواقع التجارة الالكترونية.

جرى تحسين التنبؤ بسلوك العملاء من خلال معالجة بيانات النشاطات اليومية للعملاء كسلسلة زمنية. تم تكييف بنية المحولات لتحليل البيانات المجدولة أو السلوكية للعملاء. حيث توفر هذه الطريقة تمثيلاً عميقاً للبيانات الصعبة والمعقدة، دون الحاجة إلى هندسة خصائص مكثفة. اعتمدت هذه الطريقة في نموذج Customer Activity Sequence-based Prediction and Representation [4].

#### بنية النموذج.

* تحويل البيانات إلى سلاسل: تتحول البيانات المجدولة إلى سلسلة من الأحداث لكل عميل.
* تدريب النموذج باستخدام Transformer: تستخدم بنية Transformer لتحليل هذه السلاسل وتوليد تمثيلات متجهية (Embeddings) للعملاء.
* استخدام التمثيلات في المهام المختلفة: يمكن استخدام هذه التمثيلات في مهام متعددة مثل تنبؤ الانسحاب أو توصيات المنتجات.

#### طبيعة البيانات المستخدمة.

بيانات سجلات المشتريات أو الجلسات من مواقع مثل Microsoft Retail Stores او Instacart.

## استخدام خوارزمية K-Means في تقسيم عملاء متجر تجارة إلكترونية.

يعد تحليل شرائح العملاء (Customer Segmentation) من أهم التطبيقات العملية لخوارزمية K-Means في بيئة التجارة الإلكترونية. تهدف هذه العملية إلى تقسيم قاعدة العملاء إلى شرائح متجانسة بناءً على سلوكهم الشرائي وخصائصهم، مما يساعد المتاجر الإلكترونية على اتخاذ قرارات تسويقية مخصصة، وتحسين تجربة المستخدم، وزيادة ولاء العملاء.

تعتمد خوارزمية K-Means على تجميع العملاء وفقاً لسمات محددة مثل: عدد الطلبات، إجمالي الإنفاق، معدل التكرار، والمدة الزمنية منذ آخر عملية شراء. يتم تمثيل كل عميل كنقطة في فضاء متعدد الأبعاد، وتُقسم هذه النقاط إلى عناقيد باستخدام مبدأ تقليل المسافة بين العملاء ومراكز العناقيد. تساعد هذه الطريقة في تحديد أنماط مثل "عملاء مميزون"، "عملاء جدد"، أو "عملاء مهددون بالانسحاب"، وهو ما يمكّن الشركات من تخصيص استراتيجيات التسويق والعروض بشكل أكثر دقة.

وقد أظهرت فعالية K-Means نظراً لبساطتها وسرعتها في التعامل مع مجموعات بيانات كبيرة، مما يجعلها مناسبة للتطبيقات العملية في المتاجر الإلكترونية. كما يمكن دمج نتائج التقسيم مع تقنيات أخرى، مثل التوصية التنبؤية أو توقع الانسحاب، لتحقيق أداء أفضل في إدارة العلاقة مع العملاء. [data = clustering]

## مجموعات البيانات في مجال التجارة الإلكترونية

### مجموعة بيانات E-commerce Customer Data For Behavior Analysis.

مجموعة البيانات هذه تتكون من 250000 عملية شراء للعملاء تخدم أغراض التحليل الإحصائي والنمذجة التنبؤية في مجال التجارة الإلكترونية، تحتوي البيانات على مجموعة عمليات شراء لعدة مستخدمين، وهي تناسب تطبيقات متعددة مثل:

* التنبؤ بانسحاب العملاء (Churn Prediction).
* تحليل سلال التسوق (Market Basket Analysis).
* أنظمة التوصية (Recommendation Systems).

حيث أن كل سطر في قاعدة البيانات هذه هو عبارة عن عملية شراء قام بها أحد المستخدمين تضمن المعلومات عملية الشراء.

#### أعمدة قاعدة البيانات.

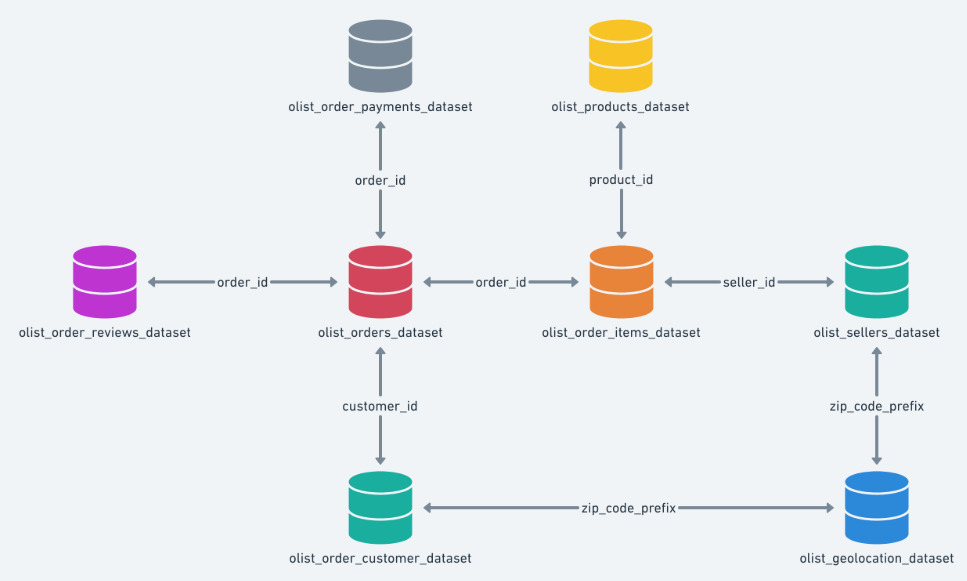
|  |  |  |
| --- | --- | --- |
| الاسم | وصف | نوع البيانات |
| Customer ID | رقم تعريفي لكل عميل | Integer |
| Customer Name | اسم العميل | String |
| Customer Age | عمر العميل | String |
| Gender | جنس العميل | String (Male,Female) |
| Purchase Date | تاريخ الشراء | Time stamp |
| Product Category | تصنيف المنتج الذي تم شراؤه | String |
| Product Price | السعر الفردي للمنتج | Integer |
| Quantity | عدد الوحدات المشتراة من المنتج | Integer |
| Total Purchase Amount | إجمالي مبلغ الشراء الذي تم دفعه في العملية | Integer |
| Payment Method | وسيلة الدفع المستخدمة | String |
| Returns | مؤشر يدل على إذا ما تم ارجاع المنتج أم لا | Boolean |
| Churn | مؤشر يدل على ما إذا كان العميل قد انسحب أم لا | Boolean |

جدول 1 أعمدة مجموعة بيانات E-commerce Customer Data For Behavior Analysis

### مجموعة بيانات Brazilian E-Commerce Public Dataset by OLIST.

مجموعة البيانات هذه هي قاعدة بيانات واقعية تتضمن معلومات تفصيلية حول أكثر من 100 ألف طلب شراء تم تنفيذها عبر متجر OLIST الإلكتروني والتي تعد من أكبر المتاجر ضمن الأسواق البرازيلية.

قسمت البيانات إلى عدة مجموعات لتسهيل الفهم والتنظيم حيث كانت البنية كما يلي:



الشكل 2 مجموعة بيانات OLIST

#### أعمدة مجموعة البيانات

قاعدة البيانات تحتوي على عدة أعمدة في مختلف المجموعات سنذكر أهمها:

|  |  |  |
| --- | --- | --- |
| الاسم | وصف | نوع البيانات |
| customer\_unique\_id | رقم تعريفي لكل عميل | Integer |
| order\_purchase\_timestamp | تاريخ الشراء | Time stamp |
| product\_category\_name\_english | تصنيف المنتج الذي تم شراؤه | String |
| price | السعر الفردي للمنتج | Integer |
| freight\_value | الرسوم اللوجستية (سعر الشحن) | Integer |
| payment\_type | وسيلة الدفع المستخدمة | String |

جدول 2 أهم أعمدة مجموعة بيانات OLIST

## دراسة التطبيقات المشابهة

# الفصل الرابع

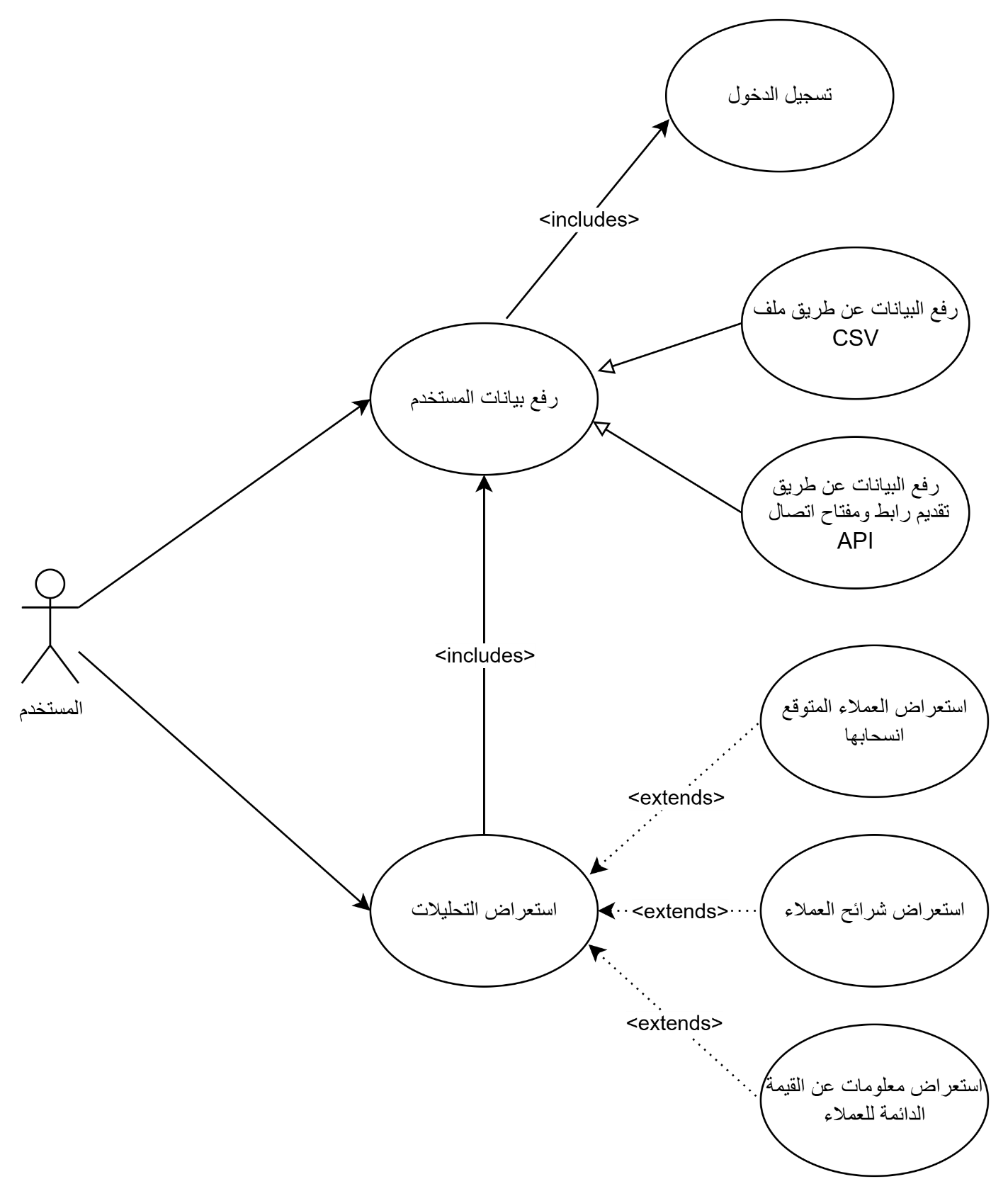
الدراسة التحليلية

يوضح هذا الفصل عمليّة تحليل النظام ودراسة متطلباته.

## مخططات حالات الاستخدام



الشكل 1 مخطط حالات استخدام إدارة حسابات المستخدمين



الشكل 2 مخطط استخدام إدارة عمليات النظام

## السرد النصي لحالات الاستخدام

### إنشاء حساب جديد.

|  |  |
| --- | --- |
| اسم الحالة: إنشاء حساب جديد | |
| الوصف Description | يقوم المستخدم بإنشاء حساب جديد خاص به. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | لا يوجد. |
| الشروط اللاحقة Postconditions | تم إنشاء حساب جديد خاص بالمستخدم ضمن النظام. |

جدول 3: حالة استخدم إنشاء حساب جديد.

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب إنشاء مشروع جديد |
| 1. يطلب النظام المعلومات التالية:  * اسم المستخدم * البريد الالكتروني الخاص بالمستخدم * كلمة المرور |  |
|  | 1. يحدد المستخدم المعلومات المطلوبة ويطلب تأكيد العمليّة |
| 1. يتحقق النظام من صحة المعلومات ويعيد للمستخدم رسالة توضح انتهاء العمليّة. |  |

جدول 4: السيناريو الناجح لحالة استخدام إنشاء حساب جديد.

**المسارات البديلة:**

لا يوجد

**مسارات الأخطاء:**

**E1:** في المرحلة رقم 2 في حال كان:

* اسم المستخدم غير صالح
* البريد الالكتروني غير صالح
* كلمة المرور ضعيفة

في حال تحقّقت إحدى الحالات السابقة، يتم استبدال الخطوة الثانية في السيناريو الأساسي بالخطوة التالية:

2. يعيد النظام رسالة توضّح سبب الخطأ ويطلب تحديد المعلومات من جديد.

**E2:** في المرحلة رقم 4 في حال كان:

* اسم المستخدم مسجل سابقاً ضمن النظام.
* البريد الإلكتروني مسجل سابقاً ضمن النظام.

في حال تحقّقت إحدى الحالات السابقة، يتم استبدال الخطوة الثانية في السيناريو الأساسي بالخطوة التالية:

4. يعيد النظام رسالة توضّح سبب الخطأ ويطلب تحديد المعلومات من جديد.

### رفع بيانات المستخدم عن طريق ملف csv.

|  |  |
| --- | --- |
| اسم الحالة: رفع بيانات المستخدم عن طريق ملف csv | |
| الوصف Description | يقوم المستخدم برفع البيانات الخاصة به عن طريق إرفاق ملف بصيغة csv. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | المستخدم مسجل في النظام سابقاً. |
| الشروط اللاحقة Postconditions | تم رفع بيانات المستخدم وحفظها ضمن النظام. |

جدول 5: حالة استخدام رفع بيانات المستخدم عن طريق ملف csv

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب رفع بياناته خاصة. |
| 1. يطلب النظام مسار الملف المراد رفعه. |  |
|  | 1. يحدد المعلومات المطلوبة ويطلب تأكيد العمليّة. |
| 1. يتحقق النظام من صحة المعلومات ويعيد للمستخدم رسالة توضح رفع الملف بنجاح. |  |
| 1. يطلب النظام ربط الأعمدة الخاصة ببيانات المستخدم بالأعمدة المعرفة ضمن النظام. |  |
|  | 1. يحدد ربط الأعمدة ويطلب تأكيد العمليّة. |
| 1. يتحقق النظام من صحة ربط الأعمدة ويعيد للمستخدم رسالة توضح صحة ربط الأعمدة. |  |
| 1. يطلب النظام من المستخدم تأكيد رفع البيانات. |  |
|  | 1. يطلب المستخدم تأكيد العمليّة. |
| 1. يتم رفع بيانات المستخدم ويعيد رسالة توضح انتهاء العمليّة. |  |

جدول 6: السيناريو الناجح لحالة استخدام رفع بيانات المستخدم عن طريق ملف csv.

**المسارات البديلة:**

لا يوجد

**مسارات الأخطاء:**

**E1:** في المرحلة رقم 2 في حال كان الملف غير موجود في المسار المشار إليه او الملف غير صالح.

في حال تحقّقت إحدى الحالات السابقة، يتم استبدال الخطوة الثانية في السيناريو الأساسي بالخطوة التالية:

4. يعيد النظام رسالة توضّح سبب الخطأ ويطلب تحديد المسار من جديد.

**E2:** في المرحلة رقم 6 في حال لم يحدد المستخدم أحد الأعمدة المطلوب ربطها. يتم استبدال الخطوة السادسة في السناريو الأساسي بالخطوة التالية:

1. يعد النظام رسالة توضح سبب الخطأ ويطلب تحديد ربط الأعمدة من جديد.

### رفع البيانات عن طريق تقديم رابط ومفتاح اتصال API.

|  |  |
| --- | --- |
| اسم الحالة: رفع البيانات عن طريق تقديم رابط ومفتاح اتصال API | |
| الوصف Description | يقوم المستخدم برفع البيانات الخاصة به عن طريق تقديم رابط إلى نقطة اتصال API endpoint في نظامه ومفتاح اتصال. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | المستخدم مسجل في النظام سابقاً. |
| الشروط اللاحقة Postconditions | تم رفع بيانات المستخدم وحفظها ضمن النظام. |

جدول 7: حالة استخدام رفع البيانات عن طريق تقديم رابط ومفتاح اتصال API.

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب رفع بياناته الخاصة. |
| 1. يطلب النظام المعلومات التالية:  * رابط نقطة الاتصال الخاصة بالمستخدم * مفتاح الاتصال الخاص بنقطة الاتصال |  |
|  | 1. يحدد المستخدم المعلومات المطلوبة ويطلب تأكيد العمليّة |
| 1. يتحقق النظام من صحة المعلومات ويقوم بطلب البيانات من نقطة الاتصال وتخزينها ضمن النظام ويعيد للمستخدم رسالة توضح انتهاء العمليّة. |  |

جدول 8: السيناريو الناجح لحالة رفع البيانات عن طريق تقديم رابط ومفتاح اتصال API.

**المسارات البديلة:**

لا يوجد

**مسارات الأخطاء:**

**E1:** في المرحلة رقم 4 في حال كان:

* رابط نقطة الاتصال غير صالح.
* مفتاح الاتصال غير صحيح.

يتم استبدال الخطوة الرابعة في السيناريو الأساسي بالخطوة التالية:

4. يعيد النظام رسالة توضّح سبب الخطأ ويطلب تحديد المعلومات من جديد.

### استعراض العملاء المتوقع انسحابها.

|  |  |
| --- | --- |
| اسم الحالة: استعراض العملاء المتوقع انسحابها | |
| الوصف Description | يقوم المستخدم باستعراض العملاء المتوقع انسحابها ومعلومات تحليلية عنها. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | المستخدم مسجل في النظام سابقاً، وقد رفع بياناته سابقا إلى النظام. |
| الشروط اللاحقة Postconditions | تم عرض النتائج الخاصة بالعملاء المتوقع انسحابها. |

جدول 9: حالة استخدام استعراض العملاء المتوقع انسحابها

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب استعراض العملاء المتوقع انسحابها. |
| 1. يقوم النظام بتحليل بيانات العملاء وعرض العملاء المتوقع انسحابها. |  |

جدول 10: السيناريو الناجح لحالة استعراض العملاء المتوقع انسحابها.

**المسارات البديلة:**

لا يوجد.

**مسارات الأخطاء:**

لا يوجد.

### استعراض شرائح العملاء.

|  |  |
| --- | --- |
| اسم الحالة: استعراض شرائح العملاء | |
| الوصف Description | يقوم المستخدم باستعراض شرائح العملاء. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | المستخدم مسجل في النظام سابقاً، وقد رفع بياناته سابقا إلى النظام. |
| الشروط اللاحقة Postconditions | تم عرض النتائج الخاصة بشرائح العملاء. |

جدول 11: حالة استخدام استعراض شرائح العملاء

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب استعراض شرائح العملاء. |
| 1. يقوم النظام بتحليل بيانات العملاء وتقسيميهم إلى شرائح وعرضها. |  |

جدول 12: السيناريو الناجح لحالة استخدام استعراض شرائح العملاء.

**المسارات البديلة:**

لا يوجد.

**مسارات الأخطاء:**

لا يوجد.

### استعراض معلومات عن القيمة الدائمة للعملاء.

|  |  |
| --- | --- |
| اسم الحالة: استعراض معلومات عن القيمة الدائمة للعملاء. | |
| الوصف Description | يقوم المستخدم باستعراض معلومات عن القيمة الدائمة للعملاء. |
| الفاعلين Actors | المستخدم |
| الشروط السابقة Preconditions | المستخدم مسجل في النظام سابقاً، وقد رفع بياناته سابقا إلى النظام. |
| الشروط اللاحقة Postconditions | تم عرض النتائج الخاصة بالقيمة الدائمة للعملاء. |

جدول 13: حالة استخدام استعراض معلومات عن الدائمة للعملاء

**سير الأحداث**

**السيناريو الأساسي الناجح:**

|  |  |
| --- | --- |
| النظام | المستخدم |
|  | 1. يطلب استعراض معلومات عن القيمة الدائمة للعملاء. |
| 1. يقوم النظام بتحليل بيانات العملاء وعرضها. |  |

جدول 14: السيناريو الناجح لحالة استخدام استعراض معلومات عن القيمة الدائمة للعملاء.

**المسارات البديلة:**

لا يوجد.

**مسارات الأخطاء:**

لا يوجد.

# الفصل الخامس

المنهجية المقترحة

يعرض هذا الفصل المنهجية المقترحة ضمن العمل المقدّم.

## مقدمة

بعد دراسة التطبيقات المشابهة تم استخلاص الميزات الأساسية التي سيتم تقديمها في التطبيق وكانت:

* توقع انسحاب العملاء.
* تقسيم العملاء إلى شرائح.
* حساب القيمة الدائمة للعملاء.

سنذكر في الفقرات القادمة المنهجية المقترحة لكل خدمة.

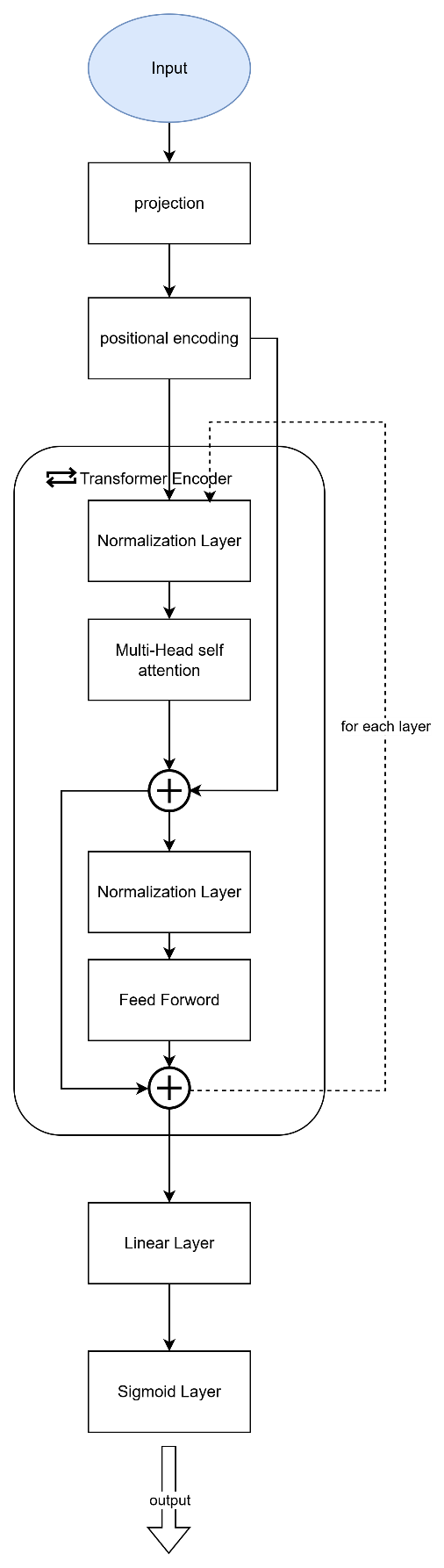
## النموذج المقترح لتوقع انسحاب العملاء

بعد النظر إلى الأبحاث في مجال توقع انسحاب العملاء جرى اتباع منهجية بناء نموذج مبني على المحولات (Transformers) وتدريبه على عمليات شراء العملاء للتوقع بانسحاب العملاء نظرا لقدرة المحولات على فهم سلسة عمليات وتمثيلها.

بداية نعرف دخل النموذج، حيث يتكون الدخل من مجموعة عمليات شراء للعملاء تحتوي على معلومات عن كل عملية شراء ومعلومات عن العميل حيث كانت:

* معلومات عن كل عملية شراء مثل: تاريخ عملية الشراء، الكمية، طريقة الدفع، المبلغ الإجمالي المدفوع، نوع المنتج ضمن عملية الشراء، سعر المنتج الفردي، هل تم إعادة المنتج.
* معلومات عن العميل مثل: عمر العميل وجنس العميل.

### بنية النموذج المقترح لتوقع انسحاب العملاء



الشكل 3بنية النموذج المقترح لتوقع انحساب العملاء

النموذج المقترح يكون من طبقتين من المحولات وكل طبقة تحتوي على 8 رؤوس من الانتباه الذاتي (Self-attention).

### حيث كانت المعاملات الفوقية (Hyper-Parameters) هي:

|  |  |
| --- | --- |
| المعامل | القيمة |
| الطبقة العرض projection | 128 |
| عدد الرؤوس Number of heads | 8 |
| عدد طبقات المحول  Number of Transformers level | 2 |
| طول سلسلة العمليات Sequence Length | 10 \* 14 |
| معامل ترك التعلم Dropout | 0.1 |

جدول 15 المعاملات الفوقية للنموذج المقترح

### تحضير البيانات

في مرحلة التدريب تم استخدام مجموعة بيانات E-commerce Customer Data for Behavior Analysis.

جرى تحضير هذه البيانات قبل تدريب النموذج المقترح لاختيار أفضل الميزات التي يمكن أن يستخلص النموذج منها المعلومات وتقديمها للنموذج بأفضل شكل.

حيث قمنا بفصل التاريخ إلى ثلاث ميزات (السنة، الشهر، اليوم) بدلا من ميزة تحتوي على المعلومات كلها. وايضاً تم اتباع منهجية One-Hot encoding لتحويل القيم الفئوية (Categorical values) إلى قيم عددية 0 او 1 بحيث يمكن للنموذج استخلاص معلومات منها. وبالإضافة إلى ذلك تمت معالجة البيانات الفارغة Null وتعبئتها (في مجموعة البيانات المستخدمة كانت لا تحتاج على معالجة كبيرة للبيانات الفارغة).

بعد تحضير البيانات تم تجميع عمليات شراء كل عميل على شكل سلسلة أو أكثر تتكون من 10 (الرقم يعتبر متحول يجب ملائمته للحصول على أفضل نتائج) عمليات شراء وبعدها تم تقييس البيانات (Scaling).

في بعض الحالات وغالبا في مسائل توقع انسحاب العملاء تكون البيانات غير متوازنة من حيث عدد العملاء المنسحبين لذلك جرى اتباع منهجية إغناء البيانات oversampling لمنع النموذج عند التدريب من الانحياز إلى تصنيف العملاء على انهم عملاء غير منسحبين أو العكس.

### منهجية التدريب

بعد تحضير البيانات، تم تقسيمها إلى ثلاث اقسام وهي:

* مجموعة التدريب Training set وكانت نسبتها 70% من البيانات الكلية.
* مجموعة التحقق Validation set وكانت نسبتها 15% من البيانات الكلية.
* مجموعة الاختبار Testing set وكانت نسبتها 15% من البيانات الكلية.

تم تدريب النموذج على فترات (Epochs) على بيانات التدريب، بعد كل فترة (epoch) يتم تقييم النموذج والتحقق من التدريب على مجموعة التحقق وعند الانتهاء من كل الفترات يتم اختبار النموذج على مجموعة الاختبار. تم اعتماد منهجية التوقف المبكر (Early Stopping) وترك التعلم (Dropout) للحد من انحياز النموذج خلال عملية التدريب على بيانات التدريب. حيث جرى استخدام BCELoss وهي Binary Cross-Entropy Loss وهي دالة خسارة مناسبة لمهام التصنيف الثنائي حيث يتم حساب الخطأ وبما أن النموذج المقترح يخرج قيمة احتمالية بعد مروره عبر الطبقة الأخيرة sigmoid والهدف هو تصنيف ثنائي 1 انسحاب العميل 0 عدم انسحابه فتكون قيمة الخسارة كبيرة كلما كانت التنبؤات بعيدة عن القيمة الحقيقية.

كما أنه تم تهيئة الأوزان قبل بدء التدريب باستخدام (Xavier uniform) من أجل تحسين بدء التدريب وتجنب مشاكل بطء التعلم والتدرجات غير المستمرة.

## النموذج المقترح لتقسيم العملاء إلى شرائح

تم اعتماد خوارزمية K-Means لتقسيم العملاء إلى شرائح بناء على خصائصهم وسلوكهم الشرائي، نظراً لقدرتها الفعالة على اكتشاف الأنماط ضمن مجموعات كبيرة من البيانات بشكل غير إشرافي (Unsupervised).

تتميز K-Means بالبساطة وسرعة الأداء، كما أنها مناسبة للمسائل التي يكون فيها الهدف هو تصنيف العملاء إلى مجموعات متجانسة اعتماداً على معايير مثل العمر، الإنفاق الكلي، وعدد المشتريات.

لضمان دقة التجميع، جرى تقييس البيانات (Scaling) لتوحيد الوحدات بين المتغيرات. ثم جرى استخدام طريقة "مخطط المرفق (Elbow Method)" لتحديد العدد الأمثل للشرائح (K). تعتمد هذه الطريقة على حساب مجموع التباين داخل المجموعات (Within-Cluster Sum of Squares) لكل قيمة ممكنة لـ K، ثم اختيار العدد الذي يحدث عنده انحناء، مما يشير إلى أن إضافة شرائح أخرى لن تحسن من جودة التصنيف بشكل كبير. بالإضافة إلى ذلك، تم الاستعانة بمؤشر Silhouette Score كأداة داعمة لقياس مدى تمايز الشرائح الناتجة، حيث تعكس القيم الأعلى جودة التقسيم.

## النموذج المقترح لحساب القيمة الدائمة للعميل

يعتمد النموذج المقترح لحساب القيمة الدائمة للعميل على تحليل البيانات التاريخية لسلوك العملاء بهدف تقدير الإيرادات المستقبلية المتوقعة من كل عميل على حدة. يقوم النموذج بحساب متوسط قيمة الطلب (AOV) من خلال قسمة إجمالي المشتريات على عدد الطلبات، ثم تحديد تكرار الشراء (Purchase Frequency PF) لكل عميل خلال فترة زمنية محددة. كما يتم تقدير فترة حياة العميل (Customer Lifespan CL) من خلال معدل الانسحاب (Churn Rate)، حيث يُفترض أن العمر المتوقع للعميل يعادل مقلوب معدل الانسحاب.

ويُدرج هامش الربح الإجمالي (Gross Margin GM) لتقديم تقدير دقيق للقيمة الفعلية المكتسبة من العميل. تُدمج هذه العناصر في المعادلة التالية:

# الفصل السادس

تصميم النظام

يعرض هذا الفصل القرارات التصميمة التي بني من خلالها النظام

## مقدمة

جرى اعتماد بنية الخدمات المصغرة (micro-services) لبناء التطبيق، اذ يتكون النظام من مجموعة من الخدمات التي تعمل مع بعضها بشكل متكامل لخدمة المستخدم النهائي. تقدم هذه البنية مجموعة من الفوائد أهمها:

* **التوسع المستقل:** يمكن توسيع كل خدمة مصغرة بشكل مستقل بناءً على الطلب، حيث تسمح هذه المرونة باستخدام الموارد بكفاءة، اذ يتم تسيع أجزاء التطبيق التي تحتاج إلى المزيد من الموارد فقط، بدا من النظام بأكمله.
* **المرونة في استخدام التقنيات:** يمكن استخدام لغات برمجة أطر عمل وقاعد بيانات وتقنيات أخرى مختلفة للخدمات المصغرة المختلفة، وتحسين كل خدمة بناءً على متطلباتها المحددة، إذ يمكن ان يؤدي هذا إلى أداء أفضل وحلول أكثر تخصيصا لكل خدمة.
* **النشر المستقل:** يمكن نشر الخدمات المصغرة بشكل مستقل، مما يسمح بالتكامل المستمر والتسليم المستمر (CI/CD) مما يعطي إمكانية إصدار التحديثات وإصلاح الأخطاء بسرعة دون انتظار دورة إصدار النظام الكاملة.

نعرض في الفقرات التالية شرحاً لتصميم كل خدمة.

## خدمة البيانات Data service

الهدف الأساسي من هذه الخدمة هو تخزين الموارد الخاصة بعملاء المستخدمين. لتزويد جميع الخدمات الأخرى بالمعلومات المطلوبة دون الحاجة إلى تكرار تنجيز الوصول إلى هذه البيانات في كل خدمة من الخدمات الأخرى. تقوم هذه الخدمة بجلب البيانات وتخزينها في قاعدة البيانات وتوفر واجهات تخاطب لطلب جلب البيانات أو تخزين بيانات جديدة من الخدمات الأخرى.

لعدم تعقيد التنجيز تم استخدام منهجية 3-tier التي تعتمد على وجود API يطلب خدمة منطق العمل Business logic التي تتخاطب بدورها مع قاعدة المعطيات لتخزين واسترجاع الموارد.

من خلال فصل هذه الخدمة في خدمة مستقلة نحقق قابلية التوسع من حيث عدد المستخدمين من جهةـ وقابلية التوسع في أنماط البيانات القابلة للتخزين وكيفية تخزينها من جهة أخرى. حيث معالجة أي نوع جديد من الموارد أو أي تعديل في بنية تخزين البيانات سيتم ضمن هذه الخدمة بمعزل عن بقية الخدمات، مما يزيد من قابلية الصيانة ويسرع عملية التطوير.

## خدمة المصادقة Auth service

الهدف الأساسي من هذه الخدمة هو المصادقة على طلبات المستخدمين، حيث يتم تزويد هذه الخدمة رمز الوصول (Access Token) الخاصة بالمستخدم الذي طلب هذا الطلب وتقوم هذه الخدمة بالمصادقة عليه، كما أن هذه الخدمة مسؤولة عن عمليات تسجيل دخول وإنشاء حسابات المستخدمين.

من خلال فصل هذه الخدمة على حدة، يمكن لكل الخدمات الأخرى التحقق من أي رمز وصول (Access Token) من أجل مصادقة المستخدمين وبذلك يتحقق أمن لبيانات المستخدمين بحيث لا يمكن لاي مستخدم الوصول إلى بيانات المستخدمين الآخرين.

## خدمة توقع انسحاب العملاء Churn service

الهدف الأساسي من هذه الخدمة هو استخدام نموذج ذكاء صنعي قادر على التنبؤ بالعملاء الذين يتوقع انسحابهم من منصة التجارة الإلكترونية الخاصة بالمستخدم. حيث تقوم هذه الخدمة بالتواصل مع خدمة البيانات للحصول على بيانات عملاء المستخدم وتقوم بتحليلها وتنبؤ العملاء المتوقع انسحابهم وعرضهم على المستخدم.

تم استخدام منهجية 3-tier في هذه الخدمة لتوفير خدمات التوقع بانسحاب العملاء. من خلال فصل هذه الخدمة نحقق قابلية التوسع وتزيد سرعة الاستجابة لأن من الممكن ان تكون هذه الخدمة هي عنق الزجاجة في النظام بسبب الزمن لازم لتوقع انسحاب العملاء وبذلك يمكن التوسع على عدد المخدمات أكبر لتحقيق أفضل استجابة للمستخدمين دون تأثير أحدهم على الآخر.

## خدمة تقسيم العملاء إلى شرائح Customer Segmentation Service

الهدف من هذه الخدمة هو تحديد الشريحة التي ينتمي إليها كل عميل وعرض معلومات عن شرائح العملاء للمستخدم. حيث تتواصل هذه الخدمة مع خدمة البيانات للحصول على بيانات المستخدمين.

تم استخدام منهجية 3-tier في هذه الخدمة أيضاً، ومن خلال فصل هذه الخدمة يمكن أيضا توسيعها ونشرها على أكثر من مخدم حسب الطلب لتحقيق قابلية التوسع.

## خدمة القيمة الدائمة للعملاء Customer Lifetime Value

الهدف من هذه الخدمة هو حساب القيمة الدائمة للعملاء الخاصة بالمستخدم، حيث تتواصل هذه الخدمة مع خدمة البيانات لجلب بيانات المستخدمين وتتواصل أيضاً مع خدمة توقع انسحاب العملاء للحصول على بعض القيم والمقاييس التي تفيد في حساب أفضل للقيمة الدائمة للعملاء.

تم استخدام منهجية 3-tiers في هذه الخدمة أيضاً، ومن خلال فصل هذه الخدمة يمكن توسيعها ونشرها على أكثر من مخدم وفي حال قمنا بتغيير المنهجية المتعبة في حساب القيمة الدائمة للعملاء أو إضافة ميزات جديدة يسهل اضافتها وتشغيلها دون الحاجة إلى التحقق من تأثير هذه الإضافات على النظام ككل.

## خدمة البوابة API-Gateway

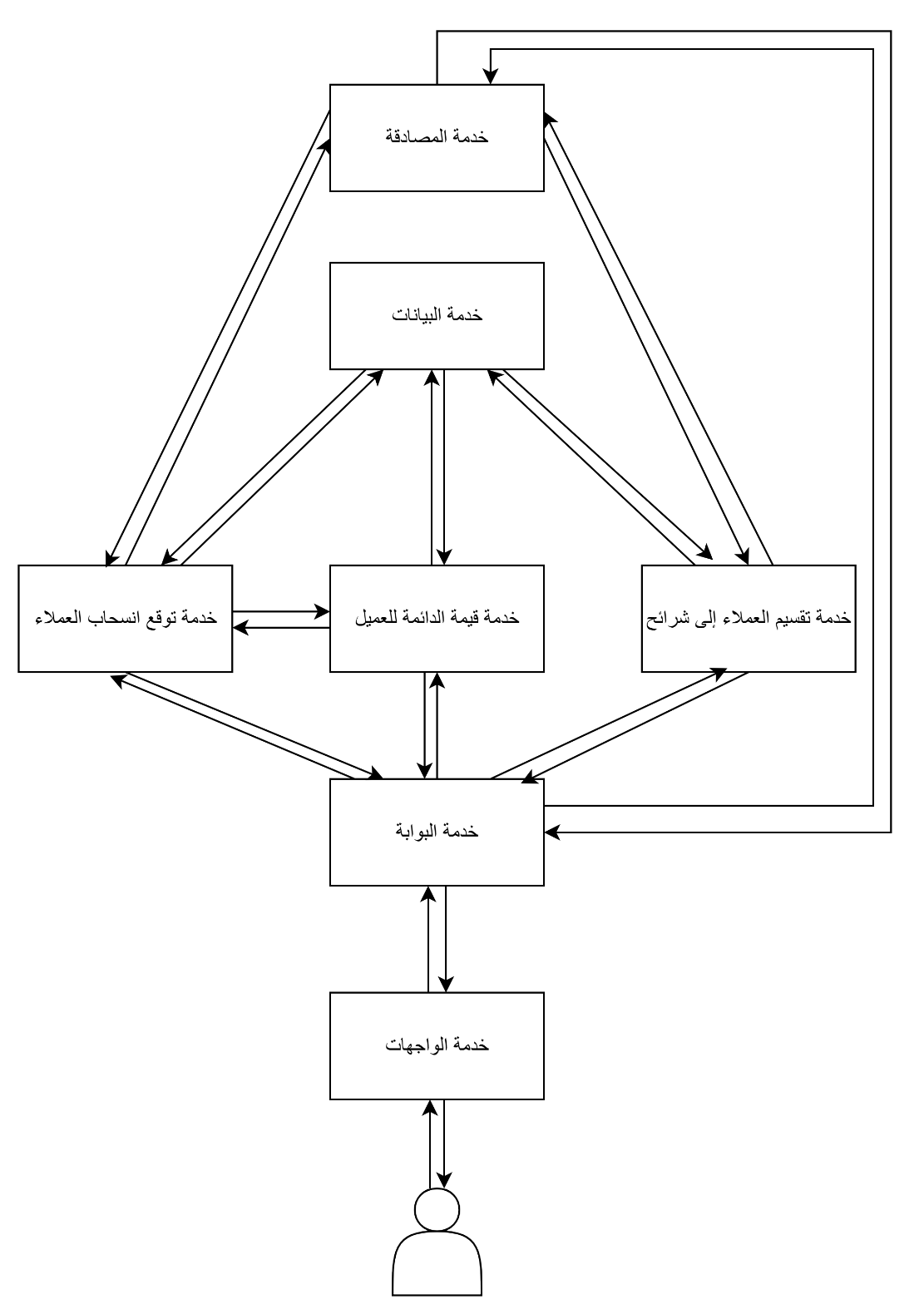
بعد النظر إلى الخدمات التي تكوّن النظام الكلي، نجد أن تواصل المستخدم مع كل خدمة على حدا يشكل مجموعة من التحديات مما يؤثر على سهولة صيانة وتشغيل النظام لذلك ظهرت الحاجة لوجود مخدم يعمل كواجهة تخاطب للنظام بحيث تقوم بتوجيه الطلبات على الخدمات الأخرى حسب الطلب.

## خدمة واجهة المستخدم Front-end

تم بناء واجهة المستخدم باستخدام بنية قائمة على المكونات مما يوفر العديد من الفوائد أهمها:

* النسقية (Modularity): تسمح بتقسيم موقع الويب إلى مكونات أصغر مستقلة. يغلف كل مكون وظيفة محددة، مما يجعله سهل الإدارة والفهم.
* قابلية إعادة الاستخدام: يمكن إعادة استخدام كل مكون من مكونات موقع الويب في أجزاء مختلفة ضمن الموقع أو حتى في مشاريع مختلفة لاحقة. مما يؤدي إلى تسريع عملية التطوير.
* سهولة تصحيح الأخطاء: بسبب عزل المكونات عن بعضها البعض تكون عملية تصحيح الأخطاء وتتبعها عملية بسيطة لا تتطلب الوقت والجهد في غالب الأحيان.

## مخطط تصميم النظام



الشكل 4 مخطط تصميم النظام

# الفصل السابع

الأدوات المستخدمة

يعرض هذا الفصل الأدوات المستخدمة لتنجيز العمل.

## PostgreSQL

PostgreSQL هو نظام إدارة قواعد بيانات علائقية مفتوح المصدر وقوي يدعم الاستعلامات المعقدة والمعاملات. يُستخدم على نطاق واسع في التطبيقات التي تتطلب استقراراً وأماناً عالياً.

## React

React مكتبة JavaScript تُستخدم لبناء واجهات المستخدم التفاعلية، تتيح إنشاء مكونات قابلة لإعادة الاستخدام وتحديث واجهة المستخدم بشكل فعّال عند تغيّر البيانات.

طورتها Meta وتُعد أساساً لتطبيقات (Single Page Application SPA).

## Python

Python هي لغة برمجة عالية المستوى، تتميز بالبساطة والوضوح وتُستخدم في تطوير الويب والذكاء الاصطناعي وتحليل البيانات.

## FastAPI

FastAPI إطار عمل Python حديث لإنشاء واجهات REST API عالية الأداء. يتميز بالسرعة ودعم توثيق API تلقائي باستخدام Swagger.

## Json Web Token (JWT)

هو معيار مفتوح يُستخدم لنقل المعلومات بين طرفين بشكل آمن كرمز مشفر (Token)، يُستخدم عادةً للتحقق من الهوية في تطبيقات الويب.

## Git

نظام تحكم في الإصدارات يُستخدم لتتبع التغييرات في الملفات البرمجية وتنسيق العمل الجماعي بين المطورين. ويدعم العمل الموزع.

## GitHub

GitHub منصة استضافة لمشاريع Git تسمح بالتعاون في تطوير البرمجيات، تشمل ميزات مثل طلبات السحب، تتبع المشاكل، وأتمتة CI/CD.

## Redis

Redis قاعدة بيانات NoSQL في الذاكرة تدعم هياكل بيانات مثل السلاسل والقوائم والمجموعات. تُستخدم في التخزين المؤقت (caching).

## PyTorch

مكتبة تعلم عميق طُورت بواسطة Meta، تُستخدم في بناء وتدريب الشبكات العصبية باستخدام واجهة Python مرنة وسهلة التجريب.

## Jupyter Notebook

Jupyter Notebook بيئة تفاعلية تُستخدم لكتابة وتشغيل كود Python وتحليل البيانات وتوثيق النتائج، خاصة في مجالات التعليم والبحث العلمي.

# الفصل الثامن

تنجيز النظام

يعرض هذا الفصل كيفية تنجيز النظام مع تفصيل كل جزء من أجزاءه.

## مقدمة

خلال عملية تطوير الخدمات يتم رفع الرماز المصدري الخاص بكل خدمة باستخدام أداة Git إلى مستودعات (Repositories) ضمن موقع GitHub حيث يتم تجميع الرماز المصدري ضمن فرع (Branch) أساسي اسمه main وعند إضافة ميزات جديدة يتم العمل على فروع فرعية بحيث يتم التطوير عليها وإجراء الاختبار دون التأثير على الفرع الأساسي الذي يحتوي على رماز التطبيق النهائي، عند الانتهاء من تطوير الميزة الجديدة يتم دمج الفرعين بحيث يتحدث الفرع الأساسي ويكون يحتوي على الميزة الجديدة.

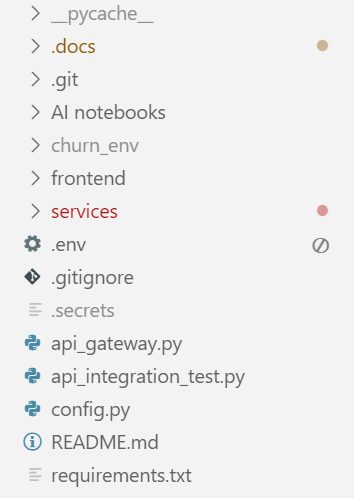
يساعد هذا الأسلوب في حفظ النسخ (Version Control) مما يمنع فقدان الرماز المصدري بشكل كامل ويسهل عملية التطوير ومراقبة التعديلات. وفي حال كان هنالك فريق من المطورين يعمل هذا تطوير التطبيق يساعد هذا الأسلوب في تنسيق عمل الفريق وعدم تداخل العمل بين المطورين مما يقلل الأخطاء ويسرع عملية التطوير.

تم بناء هذه الخدمات باستخدام إطار عمل FastAPI بلغة Python، تم اختيار Python عن غيرها من اللغات بسبب الدعم الكبير لها في مجال الذكاء الصنعي حيث تحتوي على العديد من المكتبات التي تساعد في بناء التطبيقات ونماذج الذكاء الصنعي مثل PyTocrh, TensorFlow, Sklearn مما يسهل ويسرع عملية التطوير.

تم الاعتماد على إطار العمل FastAPI لكونه إطار العمل الأكثر دعما من حيث المجتمعات والمكتبات.

نعرض في الفقرات التالية شرحاً تفصيليا عن كل خدمة.

## بنية النظام ومكوناته



الشكل 5 مجلدات النظام

تم تجميع النظام بكل تطبيقاته وخدماته ضمن بيئة واحدة لسهولة الوصول والتشغيل، حيث كانت المكونات النظام هي:

**ReadMe:** ملف توثيقي بصيغة Markdown يحتوي على شرح عام للمشروع، طريقة التثبيت، التشغيل، واستخدامه.

**.env:** ملف لتخزين متغيرات البيئة مثل روابط تشغيل الخدماتيانات متغيرة لا تُضمّن في الكود مباشرة.

**.secrets:** يُستخدم لتخزين معلومات سرّية مثل مفاتيح التشفير بشكل آمن وهو مستثنى من التتبع في Git.

**Churn\_env:** يُستخدم كبيئة افتراضية خاصة بالمشروع تحتوي على الحزم المثبتة للمشروع.

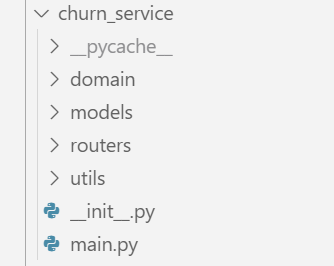
**frontend:** المجلد يحتوي على خدمة واجهات المستخدم.

**services:** مجلد يحتوي على الخدمات المقدمة في النظام.

**Requirements.txt:** يحتوي الملف على قائمة بجميع الحزم والمكتبات المطلوبة لتشغيل المشروع مع تحديد إصداراتها، يستخدم هذا الملف لتثبيت الاعتماديات عند البدء بالمشروع من أجل التوافقية.

## خدمة توقع انسحاب العملاء

### تفاصيل التنجيز



الشكل 6 مجلدات خدمة توقع انسحاب العملاء

تتكون هيكلية مجلدات خدمة انسحاب العملاء من المكونات التالية:

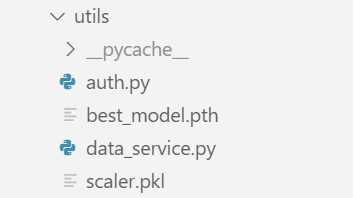
**Domain**: هو الجزء الذي يحتوي على منطق العمل Business Logic وتجيز التوابع المستخدمة في هذه الخدمة.

**Models**: يحتوي على الأنماط والصفوف التي ستتعامل معها الخدمة لتحقيق منطق العمل Business Logic.

**Routers**: تشمل تعريف نقاط النهاية (Endpoints) التي تتعامل مع الطلبات (Requests) والاستجابات (Responses) ضمن FastAPI.

**Utils**: تحتوي على وظائف أو أدوات مساعدة (Helper Functions) مثل الاتصال مع الخدمات الأخرى وتحتوي على نموذج الذكاء الصنعي وأداة التقييس.

ضمن مجلد utils يوجد نموذج مدرب للذكاء الصنعي model ومقيّس scaler مقاسة عليه البيانات سابقاً عند التدريب.



الشكل 7 مكونات مجلد utils في خدمة توقع انسحاب العملاء

**سيناريو العمل الأساسي للخدمة:** يتم تحميل الـ Model والـ Scaler عند بدء التطبيق، عند وصول طلب يتم جلب البيانات الخاصة بالمستخدم من خدمة البيانات ومعالجتها باستخدام التوابع الموجودة ضمن مجلد Domain الذي بدوره يقيّس البيانات باستخدام scaler وتمرير هذه البيانات على الـ model للحصول على التوقع لكل عميل.

جرى إضافة تخزين مؤقت باستخدام (Redis) حيث انه عندما يبدأ توقع انسحاب العملاء يتم انشاء مهمة Task خاصة بالطلب ويتم تخزين التقدم progress لمهمة التوقع مع نتائج التي قد تمت معالجتها ويتم طلب تحديث بالتقدم كل فترة زمنية لإعلام المستخدم النهائي بمقدار التقدم الحاصل في توقع انسحاب العملاء.

**تواصل خدمة توقع انسحاب العملاء مع الخدمات الأخرى**



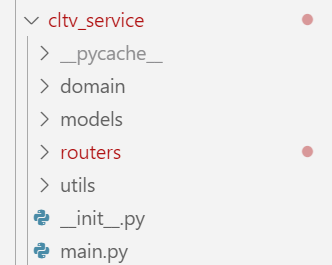
الشكل 8 تواصل خدمة انسحاب العملاء بالخدمات الأخرى

تستقبل خدمة توقع انسحاب العملاء من خدمة البوابة طلبات المستخدمين للحصول على التوقع بالعملاء المتوقع انسحابهم وتعيد لخدمة البوابة التوقعات بعد معالجتها وتحليلها للبيانات. عند وصول طلب تصادق الخدمة المستخدم الذي طلب الطلب عبر خدمة البوابة من خلال المصادقة مع خدمة المصادقة حيث يتم ارسال رمز الوصول إلى خدمة المصادقة وتعيد تأكيد بهوية المستخدم، في حال التأكيد تطلب خدمة توقع انسحاب العملاء بيانات المستخدم المصادق عليه من خدمة البيانات وتقوم بمعالجتها وتهيئتها بحيث يتم توقع الانسحاب باستخدام نموذج الذكاء الصنعي وتعيد الخدمة النتائج لخدمة البوابة.

أيضا تتواصل خدمة توقع انسحاب العملاء مع خدمة القيمة الدائمة للعميل حيث تطلب خدمة القيمة الدائمة للعميل حساب نسبة العملاء المنسحبين.

## خدمة القيمة الدائمة للعميل

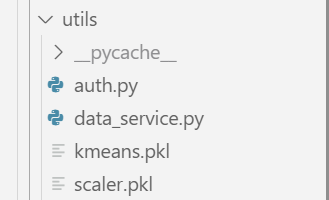
### تفاصيل التنجيز



الشكل 9 مجلدات خدمة القيمة الدائمة للعميل

هيكلية مجلدات خدمة القيمة الدائمة للعميل هي نفسها هيكلية ملفات خدمة انسحاب العملاء المذكورة سابقا ولكن يختلف التنجيز ضمنها لكي تتوافق مع الهدف الأساسي من الخدمة.

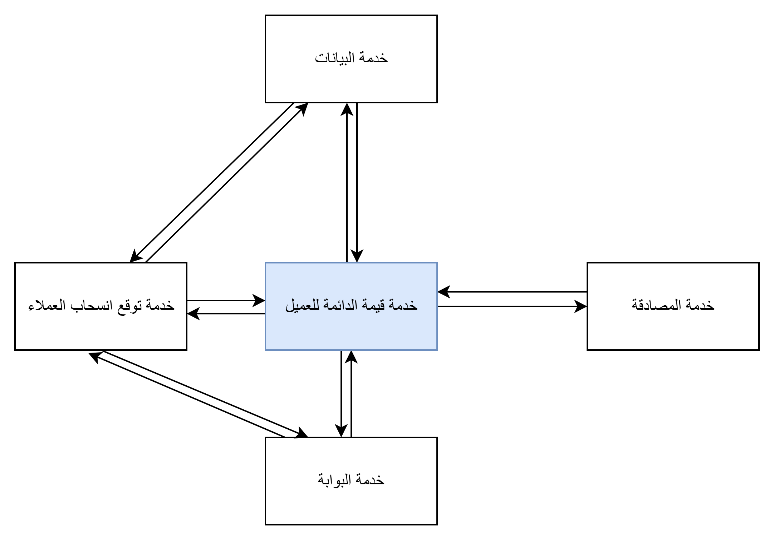
ضمن مجلد utils يوجد نموذج مدرب لخوارزمية K-means ومقيّس scaler مقاسة عليه البيانات سابقاً عند التدريب.



الشكل 10 مكونات مجلد utils في خدمة تقسيم العملاء إلى شرائح

**سيناريو العمل الأساسي للخدمة**: يتم تحميل الـ Kmeans والـ Scaler عند بدء التطبيق، عند وصول طلب يتم جلب البيانات الخاصة بالمستخدم من خدمة البيانات ومعالجتها باستخدام التوابع الموجودة ضمن مجلد Domain الذي بدوره يقيّس البيانات باستخدام scaler وتمرير هذه البيانات على الـ Kmeans للحصول على الشريحة التي ينتمي إليها كل عميل.

**تواصل خدمة القيمة الدائمة للعميل مع الخدمات الأخرى**

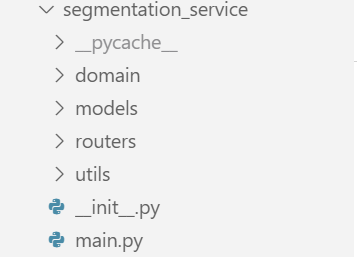


الشكل 11 تواصل خدمة القيمة الدائمة للعميل بالخدمات الأخرى

تستقبل خدمة القيمة الدائمة للعميل من خدمة البوابة طلبات المستخدمين للحصول على القيمة الدائمة للعملاء وتعيد لخدمة البوابة النتائج بعد معالجتها وتحليلها للبيانات. عند وصول طلب تصادق الخدمة المستخدم الذي طلب الطلب من خلال المصادقة مع خدمة المصادقة حيث يتم ارسال رمز الوصول إلى خدمة المصادقة وتعيد تأكيد بهوية المستخدم، في حال التأكيد تطلب الخدمة بيانات المستخدم المصادق عليه من خدمة البيانات وتقوم بمعالجتها وتهيئتها بحيث يتم حساب القيمة الدائمة للعملاء وتعيد الخدمة النتائج لخدمة البوابة.

## خدمة تقسيم البيانات إلى شرائح

### تفاصيل التنجيز

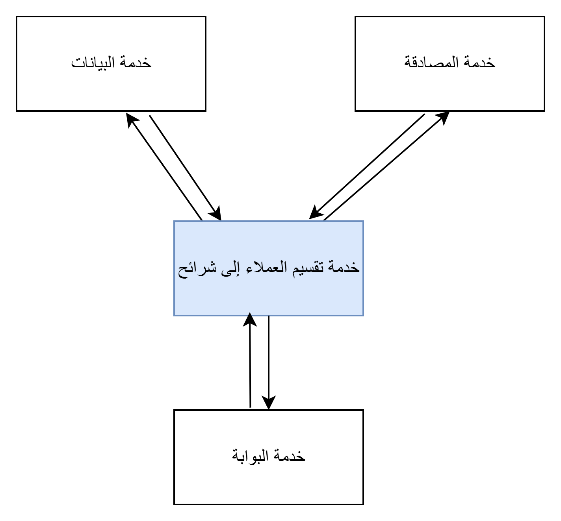


الشكل 12 مجلدات خدمة تقسيم العملاء إلى شرائح

هيكلية مجلدات خدمة تقسيم العملاء إلى شرائح هي نفسها هيكلية ملفات خدمة انسحاب العملاء المذكورة سابقا ولكن يختلف التنجيز ضمنها لكي تتوافق مع الهدف الأساسي من الخدمة.

**سيناريو العمل الأساسي للخدمة**: يتم طلب هذه الخدمة من خدمة البوابة للحصول على شرائح العملاء ومعلومات عنها تقوم الخدمة بطلب بيانات المستخدم وتقوم بإجراء العمليات الحسابية وتعيد نتائج القيمة الدائمة للعملاء.

**تواصل خدمة تقسيم العملاء إلى شرائح مع الخدمات الأخرى**

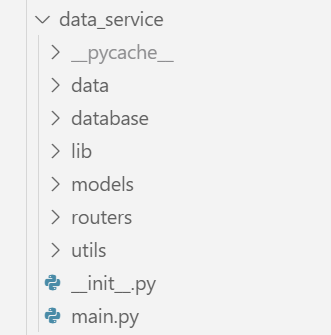


الشكل 13 تواصل خدمة القيمة الدائمة للعميل مع الخدمات الأخرى

تستقبل خدمة تقسيم العملاء إلى شرائح من خدمة البوابة طلبات المستخدمين للحصول على شرائح العملاء وتعيد لخدمة البوابة التوقعات بعد معالجتها وتحليلها للبيانات. عند وصول طلب تصادق الخدمة المستخدم الذي طلب الطلب من خلال المصادقة مع خدمة المصادقة حيث يتم ارسال رمز الوصول إلى خدمة المصادقة وتعيد تأكيد بهوية المستخدم، في حال التأكيد تطلب الخدمة بيانات المستخدم المصادق عليه من خدمة البيانات وتقوم بمعالجتها وتهيئتها بحيث يتم تقسيم العملاء إلى شرائح وتعيد الخدمة النتائج لخدمة البوابة.

## خدمة البيانات

### تفاصيل التنجيز



الشكل 14 مجلدات خدمة البيانات

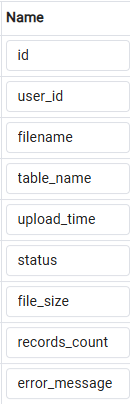
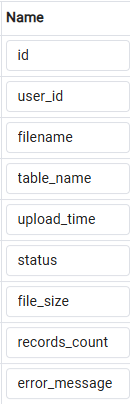
هيكلية مجلدات خدمة البيانات هي نفسها هيكلية ملفات خدمة انسحاب العملاء المذكورة سابقا ولكن يختلف التنجيز ضمنها لكي تتوافق مع الهدف الأساسي من الخدمة مع إضافة مجلد database الذي يحتوي على ملفات إعداد الاتصال بقاعدة البيانات وتعريف الجداول وتنجيز توابع لعمليات تخزين واسترجاع البيانات.

**الهدف الأساسي من هذه الخدمة**: هو ان يتم عزل تنجيز استرجاع البيانات من قاعدة البيانات عن الخدمات الأخرى بحيث لا يتم تكرار التنجيز ضمن الخدمات، إضافة إلى تحقيق معايير الأمن، حيث انه لا يمكن للمستخدمين الوصول إلى بيانات عملاء مستخدمين آخرين حيث ان تنجيز هذا موجود ضمن هذه الخدمة.

### جداول البيانات في قاعدة البيانات

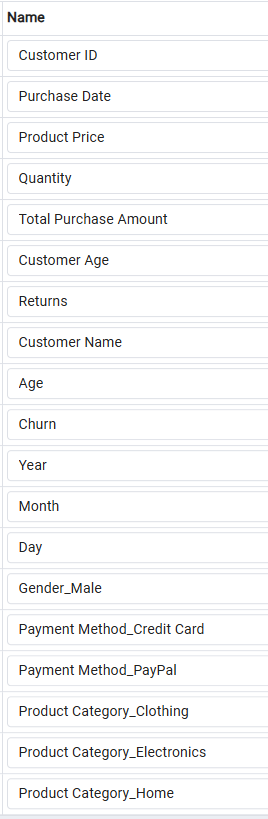
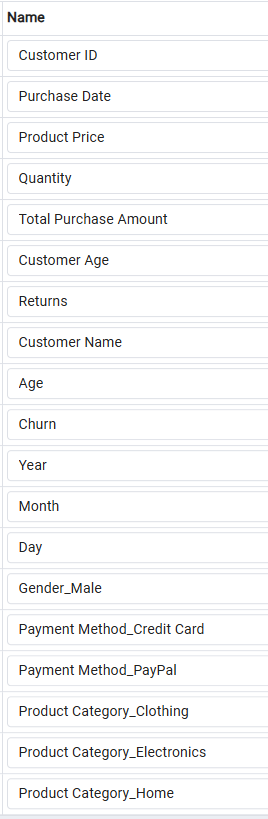
تحتوي قاعدة الجداول التالية:

* upload\_history: هو جدول يتم فيه تخزين معلومات عن عملية رفع بيانات المستحدمين.

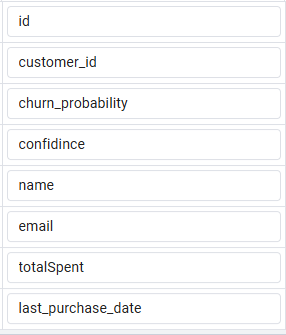
الشكل 15 أعمدة جدول upload\_history

* user\_data\_{id}: هو جدول يُنشئ لكل مستخدم من المستخدمين عند رفع البيانات لأول مرة يحتوي على بيانات عملاء المستخدم بعد معالجتها.



الشكل 16 أعمد جدول user\_data\_{id}

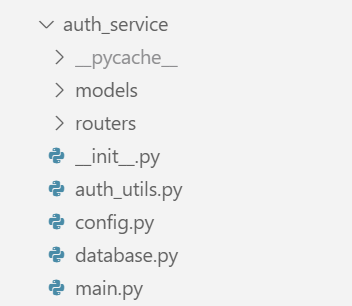
* user\_predictions\_{id}: هو جدول يُنشئ لكل مستخدم من المستخدمين ينشئ عند طلب توقع انحساب العملاء لأول مرة يحتوي على بيانات عملاء المستخدم بعد معالجتها.



الشكل 17 أعمدة جدول user\_predictions

## خدمة المصادقة

### تفاصيل التنجيز



الشكل 18 مجلدات خدمة المصادقة

هيكلية مجلدات خدمة المصادقة هي نفسها هيكلية ملفات خدمة انسحاب العملاء المذكورة سابقا ولكن يختلف التنجيز ضمنها لكي تتوافق مع الهدف الأساسي من الخدمة مع إضافة مجلد database الذي يحتوي على ملفات إعداد الاتصال بقاعدة البيانات وتعريف الجداول وتنجيز توابع لعمليات تخزين واسترجاع بيانات المستخدمين.

**الهدف الأساسي من هذه الخدمة**: هو إدارة المستخدمين وبهذا يتم عزل تنجيز مصادقة المستخدمين عن الخدمات الأخرى بحيث لا يتم تكرار التنجيز ضمن الخدمات.

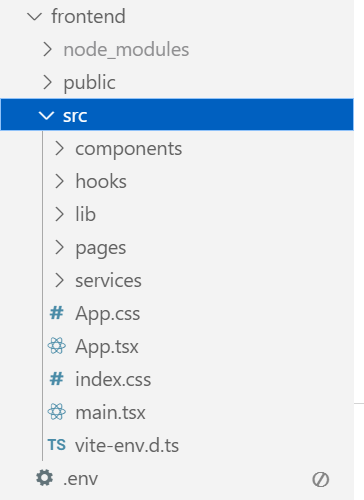
## خدمة البوابة

### تفاصيل التنجيز

تحتوي خدمة البوابة على تابع proxy يقوم بتنجيز مفهوم العمل الخاص بهذه الخدمة Business Logic، خدمة البوابة هي خدمة مسؤولة عن تحويل طلبات المستخدمين إلى باقي خدمات النظام بحيث توجه الطلبات حسب نوع الطلب والخدمة.

تم تنجيز هذه الخدمة باستخدام إطار عمل FastAPI عوضا عن استخدام أدوات جاهزة وذلك لسهولة التحكم بمنطق العمل وسهولة التوسعة عند الحاجة ولأنها تتكامل بشكل مباشر مع البنية الخلفية حيث يمكن إضافة سجل عمليات النظام ضمنها logging.

## خدمة وجهات المستخدم



الشكل 19 مجلدات خدمة واجهات المستخدم

تتكون هيكلية مجلدات خدمة واجهات المستخدم من المكونات التالية:

**Components**: يحتوي على مكونات واجهة المستخدم القابلة لإعادة الاستخدام.

**Pages**: يتضمن مكونات تمثل الصفحات الكاملة (views) التي تُعرض في مسارات (Routes) معينة.

**Services**: يضم ملفات لتنفيذ طلبات API والتفاعل مع الخادم (backend) مثل جلب البيانات، إرسال نماذج، أو تنفيذ مصادقة.

**Lib:** يحتوي على أدوات أو وظائف عامة

**Hooks:** يحتوي على Hooks مخصصة (Custom React Hooks) لتعزيز وإعادة استخدام المنطق البرمجي المرتبط بالحالة أو التأثيرات مثل Toast.

**الهدف الأساسي من هذه الخدمة**: هو تقديم واجهات للمستخدم يستطيع من خلالها التفاعل مع النظام.

تتواصل هذه الخدمة مع خدمة البوابة فقط حيث أن خدمة البوابة تقوم بتوجيه الطلبات الى الخدمات المطلوبة وتعيد الإجابة لواجهات المستخدم.

# الفصل التاسع

اختبارات النظام ومناقشة النتائج

يوضح هذا الفصل الاختبارات التي تم تطبيقها ونتائج هذا العمل.

## اختبار النموذج المقترح لتوقع انسحاب العملاء

### منهجية التقييم المتبعة

تم تقييم النموذج المقترح باستخدام مجموعتي بيانات وفق المنهجية المتبعة في البحث [4]. اعتمدت الورقة البحثية على تقييم أداء نموذج CASPR في التنبؤ بانسحاب العملاء عبر استخدام مقياسي AUROC وF1-score، وذلك بمقارنته مع نماذج تقليدية تعتمد على ميزات هندسية مثل RFM. تم تطبيق المنهجية على بيانات واقعية مثل KKBox وMicrosoft Retail، حيث تم اعتبار العميل منسحباً إذا لم يسجل أي تفاعل خلال فترة زمنية محددة. أظهرت النتائج أن تمثيلات CASPR حسّنت الأداء بشكل ملحوظ، خصوصاً في F1-score، ما يدل على قدرتها في التقاط أنماط سلوكية دقيقة دون الحاجة لهندسة ميزات معقدة.

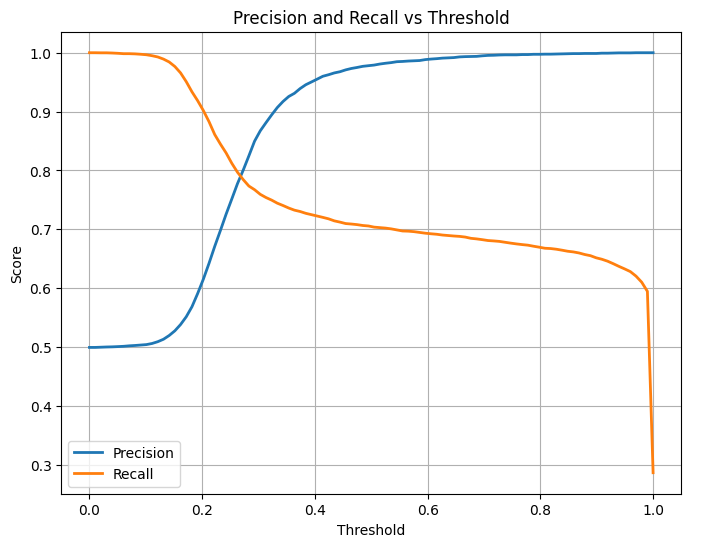
#### تقييم النموذج المقترح على مجموعة بيانات E-commerce Customer Data For Behavior Analysis

بعد تنفيذ النموذج المقترح وتقييمه على جزء من مجموعة البيانات التي تم التدريب عليها ولكن دون أن يتدرب النموذج على جزء الاختبار، كانت النتائج على الشكل التالي:

|  |  |
| --- | --- |
| Accuracy | 0.84 |
| Precision | 0.96 |
| Recall | 0.71 |
| F1 Score | 0.82 |
| AUROC | 0.87 |

جدول 16 النتائج وفق مجموعة بيانات E-commerce Customer Data For Behavior Analysis

يعرض الشكل التالي علاقة Precision وRecall مع العتبة Threshold:



الشكل 20علاقة precision recall مع العتبة Threshold على مجموعة بيانات E-commerce Customer Data For Behavior Analysis

تظهر النتائج التوازن بين الدقة Precision والاسترجاع Recall حيث كان F1 Score جيد يبلغ 0.82، بينما يشير AUROC البالغ 0.87 إلى قدرة تمييز عالية للنموذج عبر مختلف العتبات.

في الشكل 17، نلاحظ العلاقة العكسية بين الدقة والاسترجاع عند تغيير قيمة العتبة (Threshold): عند العتبات المنخفضة، يكون الاسترجاع مرتفعاً لكن الدقة منخفضة، بينما يحدث العكس عند رفع العتبة. يتقاطع المنحنيان تقريباً عند عتبة 0.25، وهي نقطة مناسبة للموازنة بين الدقة والاسترجاع، وتدعم اختيار عتبة تحقق توازناً جيداً حسب الهدف التجاري (مثلاً: تقليل خسارة العملاء أو تقليل الإنذارات الكاذبة).

#### تقييم النموذج المقترح على مجموعة بيانات Brazilian E-Commerce Public Dataset by OLIST.

**تحضير البيانات بالنسبة لمجموعة بيانات OLIST.**

مجموعة البيانات هذه كانت البيانات من غير طبيعة البيانات السابقة، ولم يتدرب عليها النموذج حيث أنها كانت لا تحتوي على بعض الميزات التي دُرب عليها النموذج سابقا مثل ميزة عمر العميل وجنسه وكمية العناصر المشتراة في عملية الشراء وما إذا كان العميل أعاد الطلب (لم يقبل بعملية الشراء) وأيضاً لم يكن هنالك ميزة ما إذا كان العميل منسحبا أم لا.

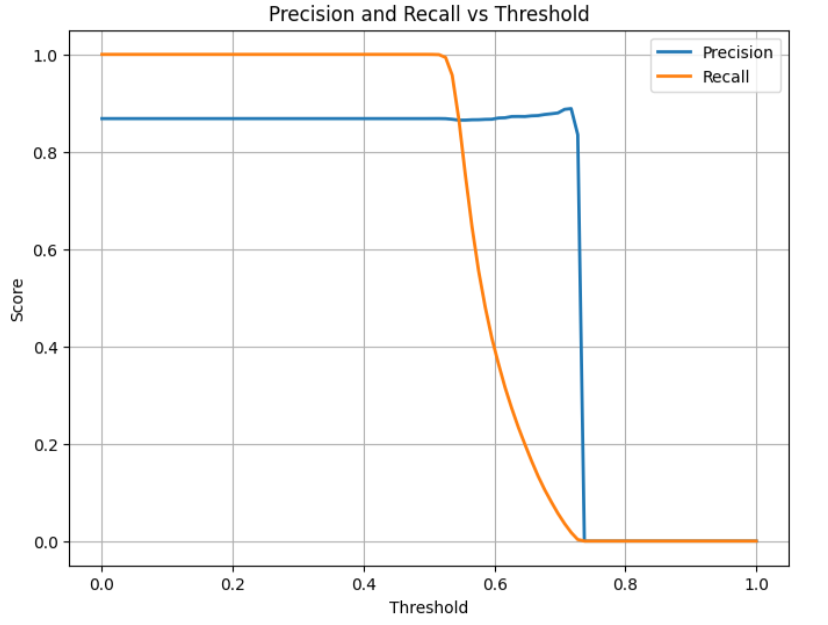
لمعالجة فقدان المعلومات تم وضع القيمة الوسطية في مجموعة البيانات السابقة للميزات المفقودة أما بالنسبة لتحديد ما إذا كان العميل منسحباً أم لا حيث تم اعتبار العميل منسحباً إذا لم يسجل أي تفاعل خلال فترة زمنية محددة.

بعد تنفيذ النموذج المقترح وتقييمه على مجموعة البيانات كانت النتائج على الشكل التالي

|  |  |
| --- | --- |
| Accuracy | 0.72 |
| Precision | 0.86 |
| Recall | 0.81 |
| F1 Score | 0.83 |
| AUROC | 0.49 |

جدول 17 النتائج وفق مجموعة بيانات OLIST

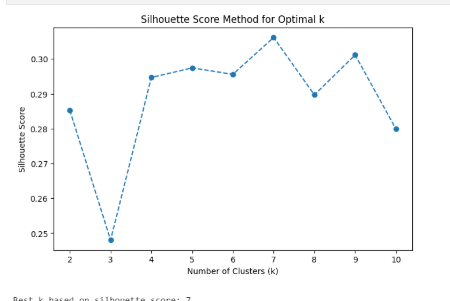
يعرض الشكل التالي علاقة Precision وRecall مع العتبة Threshold:



الشكل 21 علاقة Precision وRecall مع العتبة Threshold على مجموعة بيانات OLIST

تعكس نتائج النموذج على مجموعة بيانات OLIST أداءً مشجعاً في عدة جوانب مهمة. فقد حقق النموذج دقة إيجابية (Precision) مرتفعة بلغت 0.86، مما يدل على أن غالبية التنبؤات الإيجابية كانت صحيحة، وهو أمر بالغ الأهمية في تطبيقات تتطلب تقليل الإنذارات الكاذبة. كما أن الاسترجاع (Recall) الجيد عند 0.81 يشير إلى قدرة النموذج على اكتشاف عدد كبير من الحالات الإيجابية الحقيقية، مما يعزز موثوقيته في الكشف عن الأنماط الفعلية. بالإضافة إلى ذلك، فإن درجة F1 البالغة 0.83 تعكس توازناً ممتازاً بين الدقة والاسترجاع، وهي مؤشر واضح على فعالية النموذج في معالجة التحديات التنبؤية. ورغم أن قيمة AUROC كانت منخفضة، فإن النتائج الأخرى تثبت أن النموذج يقدم أداءً عملياً جيداً عند اختيار عتبة مناسبة، ويمكن تحسينه بسهولة بمزيد من الضبط أو التوسيع في البيانات.

## نتائج نموذج تقسيم العملاء إلى شرائح



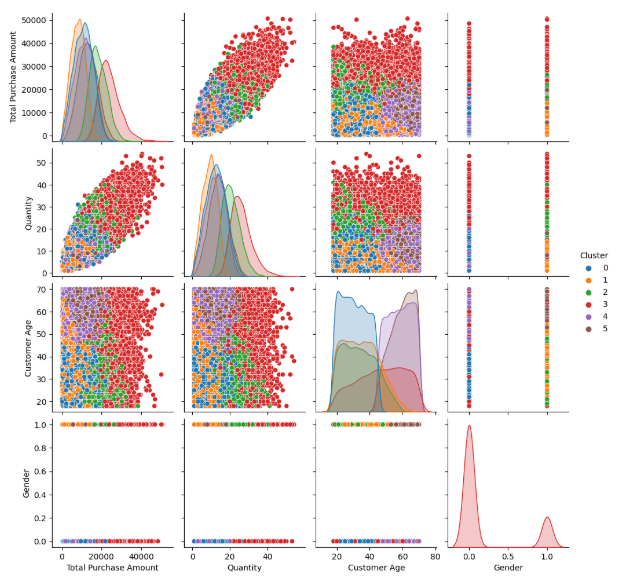
الشكل 22 مؤشر السيلويت لتحديد عدد العناقيد الأمثلي

بناءً على نتائج مؤشر السيلويت الموضحة في الشكل، نجد أن أفضل عدد للعناقيد هو سبعة عناقيد (k=7) حيث يحقق أعلى قيمة للمؤشر، مما يدل على أن تقسيم العملاء إلى سبع مجموعات يعطي أفضل تمايز بين الشرائح. ويشير ذلك إلى أن كل شريحة من العملاء لها خصائص مميزة بشكل كافٍ عن الشرائح الأخرى في هذا التقسيم. مع ذلك، يمكن ملاحظة أن قيم المؤشر ليست عالية جداً مما يعني أن هناك بعض التداخل بين الشرائح، وهذا أمر شائع في بيانات العملاء.

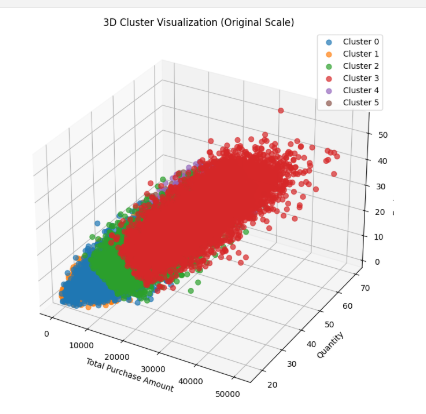
لكي تكون شرائح العملاء معبرة عن صفات مفهومة تم تقسيم العملاء إلى 6 شرائح عوضا عن 7 وهي كانت:

* صغار العمر الذكور ذوي الصرف الكبير وكميات شراء متوسطة.
* صغار العمر الإناث ذوي الصرف القليل وكميات شراء قليلة.
* صغار العمر الإناث ذوي الصرف الكبير وكميات شراء كبيرة.
* متوسطي العمر من الذكور والإناث ذوي الصرف الكبير وكميات شراء كبيرة.
* كبار العمر الذكور ذوي الصرف الكبير وكميات شراء كبيرة.
* كبار العمر الإناث ذوي الصرف المتوسط وكميات شراء قليلة.

وكانت العلاقة بين الشرائح وفق القيم كالتالي:



الشكل 23 العلاقة بين الميزات بدلالة بعضها البعض



الشكل 24 العلاقة بين عمر العميل وكمية الشراء والمبلغ الإجمالي

الخاتمة والآفاق المستقبلية

في هذا المشروع، قدّمنا منصة متكاملة تعتمد على تقنيات الذكاء الاصطناعي لتحليل سلوك العملاء في مواقع التجارة الإلكترونية، بهدف دعم اتخاذ القرار وتخصيص الخدمات التسويقية بشكل أكثر فعالية. تم استخدام نماذج مثل K-Means لتقسيم العملاء إلى شرائح سلوكية اعتماداً على بياناتهم، بالإضافة إلى إنشاء نموذج تنبؤي لتقدير احتمالية انسحاب العملاء. تتيح هذه المنصة لملاك المتاجر الإلكترونية فهم شرائح عملائهم بشكل أعمق، مما يسهم في تقليل معدل الانسحاب وزيادة العائد من الحملات التسويقية. حيث تم تغليف النماذج بطريقة قابلة للتوسع والاختبار بحيث يمكن مستقبلا إضافة خدمات إضافية.

أمّا بالنسبة للآفاق المستقبلية، بعد تقييم نتائج المنهجية المقترحة يمكننا استخلاص عدّة نقاط من شأنها تحسن الأداء أهمّها:

* تدريب النموذج على بيانات أكبر وذات صفات مخصصة.
* تخصيص النموذج بناءً على بيانات المستخدم من خلال ملائمة الأوزان.

ومن الممكن إضافة ميزات جديدة تؤدي إلى تقديم تجربة أفضل للمستخدم، أهمّها:

* تقديم تحليل للأسباب المؤدية إلى انسحاب العملاء.
* تقديم العروض والتوصيات مخصصة لشرائح العملاء.

المراجع