الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة سطيف1
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم العلوم الاقتصادية

مـذكـرة

مقدمة ضمن متطلبات الحصول على شهادة الماستر في العلوم الاقتصادية

تخصص اقتصاد كمي

الموضوع:

المفاضلة بين الانحدار اللوجستي وأشجار القرار في التنبؤ بمنح القروض

دراسة حالة الوكالة الوطنية لتسيير القرض المصغر ANGEM

من إعداد الطالبة:
- مهني زينب

تحت اشراف:
-معاش فيصل

تاريخ المناقشة: 2025/06/29

لجنة المناقشة		
الرئيس	دلفوف سفيان	أستاذ
المشرف	معاش فيصل	أستاذ -ب-
المناقش	شلاعي فاتح	أستاذ -أ-

السنة الجامعية 2025/2024

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة سطيف1
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم العلوم الاقتصادية

مـذكـرة

مقدمة ضمن متطلبات الحصول على شهادة الماستر في العلوم الاقتصادية

تخصص اقتصاد كمي

الموضوع:

المفاضلة بين الانحدار اللوجستي وأشجار القرار في التنبؤ بمنح القروض

دراسة حالة الوكالة الوطنية لتسيير القرض المصغر ANGEM

من إعداد الطالبة:
- مهني زينب

تحت اشراف:
-معاش فيصل

تاريخ المناقشة: 2025/06/29

لجنة المناقشة		
الرئيس	دلفوف سفيان	أستاذ
المشرف	معاش فيصل	أستاذ -ب-
المناقش	شلاعي فاتح	أستاذ -أ-

السنة الجامعية 2025/2024

شكر وعرفان

الحمد لله الذي هدانا لهذا وما كنا نهتدي لولا أن هدانا الله

اللهم صل وسلم وبارك على أشرف الخلق سيدنا محمد وعلى آله الطيبين الطاهرين

شكر يبتهج به اللسان دائما و أبدا لله سبحانه وتعالى الذي أعانني لإنجاز هذه المذكرة

أتقدم بالشكر للأستاذ المشرف "معاش فيصل"

و أتقدم بجزيل الشكر لوالدي حفظهما الله وأطال في عمرهما وجميع عائلتي

كما أشكر من ساعدتني في إنجاز هذا العمل

وكان لها الفضل في وصولي إلى هذه المرحلة صديقتي "مطروني امينة " النور الذي رافقني في ظلمات التعب.

الإهداء

بسم الله الرحمان الرحيم

"نرفع درجات من نشاء وفوق كل ذي علم عليم"

صدق الله العظيم

الحمد لله الذي وفقني وألهمني الصبر والتبات حتى نهاية هذا المشوار.

إلى أبي... منبع عزي، وبصمتي الأولى في درب النجاح،

يكفيني فخرا أنني ابنتك، ويكفيني عزا أنك ظهري.

وإلى أمي... عظمتي التي لا توصف، ونبع الحنان الذي لا ينضب،

منك تعلمت أن الصبر قوة، وأن الحب دعامة لا تنهار.

إلى أستاذي المشرف، الدكتور معاش فيصل، كنت بوصلتي حيت احتجت

التوجيه، شكرا لعلمك وثقتك.

إلى شهيناز، سارة وعبد الرحمان إخوتي ورفقاء الدرب،

لكم محبتي وامتناني، فقد كنتم السند في كل لحظة.

إلى خالي فرحات، شكرا لوجودك الجميل الذي لا ينسى

وإلى نفسي... لقد وصلت لأنك لم تتراجعي، وآمنت بأنك تستحقين.

اللهم بارك هذا الجهد، واجعله بداية لما هو أجمل.

فهرس المحتويات

فهرس المحتويات:

العنوان	الصفحة
شكر وعرفان	
اهداء	
فهرس المحتويات	
فهرس الجداول	
فهرس الأشكال	
الفصل الأول: الإطار العام للدراسة	-
1- تمهيد	2-1
2- مشكلة الدراسة	2
3- فرضيات الدراسة	3-2
4- أهداف الدراسة	3
5- أهمية الدراسة	4-3
6- حدود الدراسة	4
7- صعوبات الدراسة	5-4
8- تعريفات إجرائية للمتغيرات	6-5
الفصل الثاني: أدبيات الدراسة	-
تمهيد	8
أولا: الانحدار اللوجستي	9
1- نشأة نموذج الانحدار اللوجستي	9
2- مفهوم الانحدار اللوجستي	9
3- أنواع الانحدار اللوجستي	11-10
4- شروط تطبيق نموذج الانحدار اللوجستي	13-12
5- مزايا وعيوب الانحدار اللوجستي	14-13
ثانيا: أشجار القرار	14
1- مفهوم شجرة القرار	14
2- مكونات شجرة القرار	14
3-خوارزميات شجرة القرار	15

4- مزايا وعيوب شجرة القرار	17-16
5- أدوات تقييم ائتمان العملاء	17
ثالثا: الدراسات السابقة	23-18
رابعا: الفجوة البحثية	23
الفصل الثالث: نتائج الدراسة	-
1- منهج الدراسة	25
2- عينة الدراسة	25
3- أداة الدراسة	25
4- مصادر جمع البيانات	25
5- المعالجة الإحصائية المستخدمة	26-25
6- الدراسة الوصفية	32-26
7- أشجار القرار	33-32
8- تحليل مصفوفة الارتباك لنموذج تصنيف القروض	37-34
9- الانحدار اللوجستي	42-38
10-المقارنة	44-43
1.تقييم أداء الانحدار اللوجستي وأشجار القرار	45-44
الفصل الرابع: مناقشة النتائج والتوصيات	-
1-مناقشة نتائج الدراسة	48-47
2-استنتاجات أساسية	48
3- التوصيات وآفاق الدراسة	49-48
قائمة المصادر والمراجع	53-51
الملاحق	-
الملحق 01: أشجار القرار الجزء 1	55
الملحق02: أشجار القرار الجزء 2	56
الملحق03: أشجار القرار الجزء 3	57
الملحق04: أشجار القرار الجزء 4	58
الملحق05: الانحدار اللوجستي الجزء1	59
الملحق06: الانحدار اللوجستي الجزء2	60
ملخص	-

فـهـرس الـجـداول

فهرس الجداول:

الرقم	العنوان	الصفحة
01	الفرق بين الانحدار اللوجستي والانحدار الخطي	14
02	مخرجات نموذج أشجار القرار بعد تطبيق تقنية SMOTE	31
03	نموذج الانحدار اللوجستي بعد تطبيق FSS	37
04	مقارنة بين أداء نموذج أشجار القرار ونموذج الانحدار اللوجستي	41

فـهـرس الأشكـال

فهرس الأشكال

الرقم	العنوان	الصفحة
01	منحنى لوجستي F(z)	9
02	مخطط شجرة القرار	14
03	حالات القبول والرفض لطلبات القروض.	24
04	توزيع أعمار المتقدمين لطلب القروض من الوكالة	25
05	توزيع المستوى الدراسي للعملاء	26
06	توزيع المستفيدين وغير المستفيدين من القروض حسب مجال النشاط الاقتصادي	27
07	نسبة جنس العملاء لدى الوكالة	28
08	نسبة المتقدمين الحاصلين على شهادة تعليمية وغير الحاصلين	29
09	توزيع الفئات قبل وبعد تطبيق تقنية SMOTE	32
10	مصفوفة الارتباك لنموذج أشجار القرار بعد تطبيق تقنية SMOTE	30
11	منحنى ROC لتقييم أداء نموذج أشجار القرار	31
12	شجرة القرار	32
13	أهم المتغيرات المؤثرة في القبول والرفض باستخدام اشجار القرار	32
14	مصفوفة الارتباك لنموذج الانحدار اللوجستي	38
15	أهم المتغيرات المؤثرة في القبول والرفض باستخدام الانحدار اللوجستي	39
16	منحنى ROC لتقييم أداء الانحدار اللوجستي	40
17	منحنى ROC	42

الفصل الأول:
الإطار العام
للدراسة

1. تمهيد

في عالم تتسارع فيه وتيرة التغيرات الاقتصادية والاجتماعية، لم يعد اتخاد القرار يعتمد فقط على الخبرة والتجربة، بل أصبح مرهونا بقدرة المؤسسة على استخراج أنماط خفية من كميات هائلة من المعطيات.

وبينما تتسارع وتيرة التحول الرقمي في شتى القطاعات، برز التنبؤ كأحد أهم أدوات دعم القرار، لما يتيحه من نظرة استباقية تمكن من تقليل المخاطر وتعزيز الفعالية.

ومن بين أهم فروع التنبؤ، يبرز التنبؤ التصنيفي (classification)، الذي يهدف إلى تحديد الفئة أو القرار الأنسب استنادا إلى خصائص محددة وتكمن أهمية هذا النوع من التنبؤ في قدرته على تحويل البيانات إلى قرارات عملية، سواء تعلق الأمر بتشخيص طبي أو فرز طلبات توظيف أو تقييم أهلية مالية، وقد تطورت في هذا المجال عدة نماذج رياضية، تختلف في بنيتها وأداءها من بينها الانحدار اللوجستي (Logistic Regression) وأشجار القرار (Decision Trees)، آلات الدعم النقطي (SVM)، الغابات العشوائية (Random Forest) والشبكات العصبية الاصطناعية (ANN)، ويعد الانحدار اللوجستي وأشجار القرار من أكثر النماذج شيوعا، خاصة في المؤسسات التي تبحث عن نماذج التي تجمع بين الدقة وسهولة التفسير.

لقد أثبتت هذه النماذج فعاليتها في مجالات متعددة، ففي القطاع الطبي تستخدم لتحديد احتمالات الإصابة بأمراض مزمنة، وفي الزراعة الذكية تساعد على تصنيف جودة المحاصيل وتحديد توقيت الحصاد الأمثل، أما في القطاع المالي فقط أصبحت الأداة المفضلة في تقييم الجدارة الائتمانية، وتشير دراسة صادرة عن جامعة ميسوري (2019) إلى أن اعتماد البنوك الأمريكية على الذكاء الاصطناعي بما في ذلك نماذج التصنيف في معالجة قرارات الإقراض ارتفع من 14% في 2017 إلى 43% في 2019، ما يؤكد التحول العميق نحو قرارات قائمة على التحليل الكمي بدل القرارات التقليدية.

ويزداد هذا التحول أهمية في سياق مؤسسات القرض المصغر، التي تلعب دورا اجتماعيا حساسا في تمويل الفئات الشابة في المجتمع ودعم المشاريع الصغيرة ومتناهية الصغر.

ومع ارتفاع الطلب على هذه القروض وتنوع الخلفيات الاجتماعية والاقتصادية لطالبيها بات من الضروري توفير أدوات دقيقة ومرنة تساعد على التنبؤ في قرارات الإقراض خاصة أن هذه المؤسسات تعتمد في الغالب على بيانات غير تقليدية يصعب تقييمها بالوسائل الكلاسيكية.

انطلاقا من هذا الواقع تهدف هذه الدراسة إلى المفاضلة بين نموذجي الانحدار اللوجستي وأشجار القرار في التنبؤ بمنح القروض المصغرة من خلال تطبيق ميداني على وكالة القرض المصغر لولاية سطيف وذلك بهدف تحديد النموذج الأكثر كفاءة في دعم قرار وتوجيه منح التمويلات نحو المستحقين بما يعزز فعالية المؤسسة ويقلل من مخاطر التعثر.

2. مشكلة الدراسة

استنادا على السياق السابق وجدنا من بين هذه النماذج يستخدم كل من الانحدار اللوجستي وأشجار القرار بشكل واسع في تحليل التصنيف واتخاذ القرار، غير أن أداء كل منهما يختلف بحسب طبيعة البيانات وسياق التطبيق.

انطلاقاً من ذلك، تتمثل الإشكالية العامة للدراسة في السؤال الآتي:

أي النموذجين، الانحدار اللوجستي أم أشجار القرار يظهر أداء تنبؤي أفضل ودقة أكثر في تصنيف طلبات القروض المصغرة ضمن وكالة سطيف؟

الأسئلة الفرعية

انطلاقا من الإشكالية الرئيسية يتم طرح الأسئلة الفرعية التالية:

- ما هي الفروق النظرية والمنهجية بين نموذج الانحدار اللوجستي وأشجار القرار في سياق التنبؤ؟
- كيف يختلف أداء كل من النموذجين الانحدار اللوجستي وأشجار القرار في التنبؤ بمنح القروض من حيث مؤشرات الأداء الاحصائي (مثل الدقة، الاستدعاء، F1-Score)؟
- ما هي العوامل التي تحدد اختيار النموذج الأنسب في التطبيق العملي داخل المؤسسات المالية؟

3. فرضيات الدراسة

الفرضية الرئيسية

ومن خلال الدراسات السابقة التي تناولت الموضوع فإن دراستنا تنطلق من الفرضية الرئيسية التالية:

تتمتع أشجار القرار بفعالية تنبؤية أكثر ومؤشرات أداء أفضل مما يجعلها الأنسب في معالجة مشكلة التنبؤ بمنح القروض.

الفرضيات الفرعية:

- توجد فروق جوهرية بين النموذجين من حيث الطبيعة الرياضية، متطلبات البيانات وطبيعة كل منهما على التعامل مع العلاقات المعقدة بين المتغيرات.
- يظهر نموذج أشجار القرار تفوقا من حيث مؤشرات الأداء الاحصائي مثل الدقة، الاستدعاء، F1-Score عند استخدامه على بيانات القروض.
- يتأثر اختيار النموذج الأفضل في المؤسسات المالية بعوامل عديدة مثل توفر البيانات، بساطة التطبيق، قابلية التغيير ومستوى التحكم في فرط الملاءمة (Overfitting).

4. أهداف الدراسة

➢ الهدف الرئيسي:

تقييم أداء كل من نموذج أشجار القرار والانحدار اللوجستي من خلال مقاييس الأداء كالدقة (Accuracy)، الاستدعاء (Recall) و(F1-Score) وذلك انطلاقا من بيانات حقيقية متعلقة بقرارات منح القروض.

➢ الأهداف الفرعية

- تحليل الفروق المنهجية بين النموذجين من حيث البنية الرياضية، القدرة التفسيرية وسهولة التطبيق داخل بيئات القرار المالي.
- تحديد العوامل العملية والمؤسسية التي تؤثر في اختيار النموذج التنبؤي الأنسب، مثل وضوح النتائج، قابلية التفسير وسهولة الاستخدام من قبل غير المختصين.
- تقديم توصيات علمية وعملية للقائمين على برامج تمويل القروض المصغرة، تحثهم على اعتماد آليات اتخاذ القرار المبنية على البيانات (Data-Driven Decision Making).

5. أهمية الدراسة

تنبع أهمية هذه الدراسة من كونها تتناول إحدى الإشكاليات الأساسية التي تواجه مؤسسات التمويل المصغر، والمتمثلة في دقة اتخاذ القرار الائتماني، بالنظر إلى الاقتصادية والاجتماعية المترتبة عنه، إذ أن أي تقدير خاطئ قد يؤدي إلى إقصاء مستحقين فعليين من فرص تحسين أوضاعهم المعيشية، أو إلى تمويل غير مؤهلين، وهو ما يشكل خطرا مباشرا على استدامة المؤسسة، وتحقيق أهدافها التنموية والمالية.

وفي هذا السياق، تقترح هذه الدراسة مقاربة علمية مبنية على استخدام أدوات التحليل التنبئي الحديثة، بهدف تعزيز موضوعية القرار الائتماني ورفع كفاءته. وتتمثل القيمة المضافة لهذه الدراسة في مقارنة نموذجين من أكثر النماذج شيوعًا في مجال التصنيف والتنبؤ، وهما الانحدار اللوجستي وأشجار القرار، من خلال دراسة حالة تطبيقية واقعية على مستوى وكالة سطيف.

ولا تقتصر الدراسة على تحليل الأداء الإحصائي لكل نموذج فحسب، بل تمتد أيضًا إلى دراسة القابلية التطبيقية، والقدرة التفسيرية لكل نموذج، ومدى تأثر أدائه بطبيعة البيانات المتاحة، مما يمنح هذه الدراسة بعدًا عمليًا يسهم في تحسين جودة القرار داخل مؤسسات التمويل المصغر.

6. حدود الدراسة

- **الحدود المكانية:** ركزت الدراسة على الوكالة الوطنية لتسيير القرض المصغر ANGEM.
- **الحدود الزمنية:** اقتصرت هذه الدراسة من حيث الحدود الزمنية على البيانات والمعطيات المتاحة من سنة 2024-2025.
- **الحدود الموضوعية:** تركز الدراسة على مقارنة الانحدار اللوجستي وأشجار القرار باستخدام مؤشرات إحصائية فقط وبالاعتماد على بيانات وكالة ANGEM بسطيف.

7. صعوبات الدراسة

خلال مراحل إعداد هذا البحث التطبيقي، واجهنا مجموعة من الصعوبات التي أثرت بشكل مباشر على سيرورة العمل، ابتداءً من جمع البيانات ووصولاً إلى تحليل النتائج وتفسيرها، فيما يلي أبرزها:

- محدودية الوصول إلى بيانات مفصلة من وكالة القروض المصغرة بسطيف: صعوبة الحصول على بيانات دقيقة ومفصلة بسبب حساسية المعلومات المالية، وهو ما تطلب:
  - ✓ الحصول على موافقات إدارية مسبقة.
  - ✓ القيام بعدة زيارات ميدانية للوكالة لاستكمال البيانات الضرورية.
- صعوبة تنظيف البيانات ومعالجتها مسبقًا: تطلّب التعامل مع البيانات الأولية إجراء معالجة دقيقة من خلال ضمان ترميز المتغيرات النوعية لتكون قابلة للاستخدام في النماذج الإحصائية وخوارزميات تعلم الآلة.

- صعوبة تطبيق النماذج باستخدام لغة برمجة Python: على الرغم من دراسة بعض المفاهيم النظرية في الجامعة المتعلقة ببرمجة Python، فإن غياب التطبيق الفعلي حال دون استيعابها بالشكل الكافي والتمرس فيها.

8. تعريفات إجرائية للمتغيرات

- **الانحدار اللوجيستي (Logistic Regression):** نموذج إحصائي يستخدم لتقدير احتمال انتماء ملاحظة إلى فئة معينة، بناءً على متغيرات مستقلة.
- **أشجار القرار (Decision Tree):** نموذج يعتمد على تقسيم متكرر للبيانات إلى فروع بناءً على خصائصها للوصول إلى قرارات.
- **الفرط في التعميم (Overfitting):** حالة تحدث عندما يتعلم النموذج التفاصيل والضوضاء الخاصة بعينة التدريب لدرجة تؤثر على دقته في التعميم على بيانات جديدة.
- **البيانات غير المتوازنة (Imbalanced Data):** عندما تكون إحدى الفئات (مثلاً: "قبول القرض") أقل تمثيلًا بكثير من الأخرى، مما يسبب انحيازًا في التنبؤ.
- **منحنى ROC:** هو رسم بياني يستخدم لتقييم أداء نماذج التصنيف، يوضح العلاقة بين معدل الإيجابيات الحقيقية (TPR) ومعدل الإيجابيات الخاطئة (FPR).
- **Auc (AUC - Area Under the Curve) المساحة تحت المنحنى:** يشير هذا المصطلح إلى المساحة تحت المنحنى وهو قياس يستخدم في تقييم أداء نماذج التصنيف، خاصة عند استخدام منحنى ROC (Receiver Operating Characterisict Curve).
- **الترميز الوهمي (Dummies / One-Hot Encoding):** هو تقنية لتحويل المتغيرات النوعية (مثل الجنس أو القطاع) إلى متغيرات رقمية (0 و1) لتمكين النماذج الإحصائية من معالجتها.

- **الدقة (Accuracy):** هي نسبة جميع التصنيفات الصحيحة، سواءً كانت إيجابية أو سلبية.
- **الضبط (Precision):** يقيس دقة التنبؤات الإيجابية. ويجيب على السؤال: "من بين جميع العناصر التي صنفها النموذج على أنها إيجابية، كم منها كانت إيجابية بالفعل؟".
- **الاستدعاء (Recall):** تقيس قدرة النموذج على إيجاد جميع الحالات الإيجابية، تجيب على السؤال: "من بين جميع العناصر الإيجابية الفعلية، كم منها حددها النموذج بشكل صحيح؟".
- **F1 Score:** متوسط كل من الضبط والاستدعاء، ويوازن المقياسين في رقم واحد، مما يجعلها مفيدة بشكل خاص عند وجود تعارض بينهما.

- **مصفوفة الارتباك (Confusion Matrix)**: هي أداة من أدوات تقييم أداء نماذج التعلم الآلي، تعرض المصفوفة عدد النتائج الإيجابية الصحيحة (TP)، والسلبية الصحيحة(TN)، والإيجابية الخاطئة(FP)، والسلبية الخاطئة(FN). تساعد هذه المصفوفة في تحليل أداء النموذج، وتحديد التصنيفات الخاطئة وتحسين دقة التنبؤ.
- **الانحدار التدريجي(Stepwise Regression)**: هو تقنية تستخدم لبناء نموذج انحدار بسيط وفعال، من خلال إضافة أو إزالة المتغيرات تدريجيا بهدف الاحتفاظ فقط بالعوامل الأكثر تأثيرا، ينقسم إلى نوعين رئيسيين:

✓ **الاختيار الأمامي(Forward Selection)**: يبدأ بنموذج فارغ وتضاف المتغيرات واحدا تلو الآخر حسب أهميتها، إلى غاية نهاية التحسينات.

✓ **الاختيار/ الاقصاء العكسي (Backward Elimination)** : تبدأ الخوارزمية بنموذج يتضمن جميع المتغيرات من ثم تتم إزالة المتغيرات بشكل متكرر إلى غاية نهاية التحسينات.

- **Oversampling:** هي أحد التقنيات التي تتعامل مع مشكلة عدم توازن البيانات (Data Imbalance) عن طريق إما تكرار البيانات الخاصة بفئة الأقلية (Minority Class) أو توليد عينات جديدة بشكل مصطنع (Synthetic) وأحد أشهر أنواعها تقنية SMOTE و ADASYN.

الفصل الثاني:
أدبيات الدراسة

تمهيد

أصبح مفهوم تصنيف البيانات من المفاهيم المهمة التي تدخل في أغلب الاتجاهات في الحاسبات بصورة عامة وفي مجال تعلم الآلة بصورة خاصة، يتمثل التصنيف في التفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى، هذه الخاصية هي عموما كيفية، ومع تطور وتقدم أساليب التنبؤ أصبح هناك العديد من الطرق والتحاليل الإحصائية التي تساعد على دراسة المتغيرات، ومن هذه الطرق الانحدار اللوجستي وأشجار القرار وهما يهدفان لمعرفة أكثر المتغيرات تأثيرا على المتغير المصنف أو المعتمد وتصنيف الحالات الجديدة اعتمادا على البيانات السابقة.

أولا: الانحدار اللوجستي

في كثير من الأحيان نكون أمام دراسة متغيرات لها قيم ثنائية تتمثل في وجود أو انعدام خاصية معينة، مثل إصابة بمرض ما أو عدم الإصابة، الاستجابة أو عدم الاستجابة لعرض منتج جديد، والوفاء بالالتزامات أو عدم الوفاء، وعجز أو عدم عجز المؤسسات. لذلك يتم استخدام طرق إحصائية مغايرة عن الطرق المستخدمة في حالة المتغيرات الكمية، ومن بين هذه الطرق نجد الانحدار اللوجستي الذي يعتبر من أشهر الطرق المستخدمة في هذا المجال. (إبراهيم عدلي، 2015، ص1)

1. نشأة نموذج الانحدار اللوجستي

يعود الفضل الأول في استخدام النموذج اللوجستي إلى الأعمال التي قام بها العالم الرياضي البلجيكي الأصل Pierre François Verhulst، حيث قام هذا الأخير بعدة أبحاث معمقة في مجال قوانين الزيادة في عدد السكان، وقد تمكن Verhulst سنة 1844 من صياغة دالة رياضية أطلق عليها اسم "الدالة اللوجستية"، حيث تمكنت هذه الأخيرة من التنبؤ من تطور عدد السكان من خلال المعرفة البسيطة للعوامل المؤثرة على ثلاث تواريخ زمنية مختلفة. (Martial , 1804 -1849, P.542)

2. مفهوم نموذج الانحدار اللوجستي

يعتبر نموذج تحليل الانحدار اللوجستي أحد النماذج الاحتمالية غير الخطية الذي يحلل ويفسر العلاقة الموجودة بين متغير تابع وهو المتغير الذي يأخذ إحدى القيميتين (0، 1) وعدة متغيرات مستقلة. (رضوى، 2017، ص 81)

الانحدار عبارة عن أداة أكثر قوة لأنه يقدم اختبارا لمعنوية المعاملات، كما أنه يعطي الباحث فكرة عن مقدار تأثير المتغير المستقل في المتغير التابع النوعي ثنائي القيمة، كما انه يرتب تأثير المتغيرات المستقلة، التي تسمح للباحث باستنتاج أن متغيرا ما يعتبر أقوى من المتغير الاخر في فهم ظهور النتيجة المطلوبة. (علي، 2012، ص 238)

الانحدار اللوجستي ما هو إلا نوع من أنواع الانحدار يكون فيه المتغير التابع متغير نوعي قد يأخذ قيمتين (الانحدار اللوجستي الثنائي) وقد يأخذ أكثر من قيمتين (الانحدار اللوجستي المتعدد)، إلا أننا في الانحدار اللوجستي يكون هدفنا ليس تفسير التغير في قيم المتغير التابع وإنما تفسير احتمال حدوث وعدم حدوث الظاهرة محل الدراسة، ومن مزايا الانحدار اللوجستي أنه أقل حساسية تجاه الانحرافات عن التوزيع الطبيعي لمتغيرات الدراسة، وذلك مقارنة بأساليب إحصائية أخرى، بالإضافة الى أنه يستطيع أن يتجاوز لعديد من الافتراضات المقيدة لاستخدام طريقة المربعات الصغرى OLS في الانحدار الخطي. (سهيلة، 2014، ص54)

3. أنواع الانحدار اللوجستي

✓ **الانحدار اللوجستي الترتيبي**

يتميز هذا النوع بتفسير أثر المتغيرات المنبئة (المستقلة) باختلاف مستويات قياسها على الاستجابات الرتبية بمعنى أن يكون المتغير التابع متغيرا ترتيبيا. (دعيش، ساري، 2017، ص126).

✓ **الانحدار اللوجستي الثنائي الحدين**

الانحدار اللوجستي الثنائي نوع من تحليل الانحدار عندما يكون المتغير التابع كيفي ثنائي، أو هو نموذج يستخدم بيانات ثنائية للمتغير التابع، أي أنه الأسلوب الاحصائي المستخدم لفحص وتوفيق العلاقة بين المتغير التابع النوعي ثنائي القيمة ومتغير واحد أو أكثر من المتغيرات المستقلة أيا كان نوعها. (محمد نجيب، 2009، ص184) يقوم نموذج الانحدار اللوجستي على فرض أساسي وهو أن متغير تابع (Y) يأخذ القيمة (1) إذا حدث معين باحتمال (P)، ويأخذ القيمة (0) إذا لم يحدث ذلك الحدث باحتمال (P-1)، أي حدوث الاستجابة وعدم حدوثها، وبما أن المتغير التابع في النموذج المقدر هو متغير برنولي يأخذ أحد القيمتين (1 و0)، (إبراهيم عدلي، 2014، ص2)

والدالة اللوجستية تأخذ الصيغة التالية:

$$P_i = E\left(\frac{Y_i}{X_i}\right) = \frac{e^{(B_0+B_1X_1+\cdots+B_kX_k)}}{1+e^{(B_0+B_1X_1+\cdots+B_kX_k)}} = \frac{e^{C_{(x)}}}{1+e^{C_{(x)}}}$$

حيث:

P_i : تمثل احتمال حدوث الفئة أو السمة i من المتغير التابع

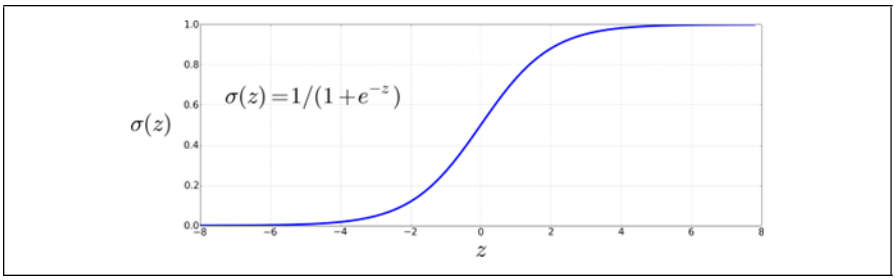
B: تمثل معالم النموذج

من خلال المعادلة يتضح أن العلاقة بين المتغير التابع والمتغيرات المستقلة علاقة غير خطية، لذلك هناك عدة تحويلات يمكن القيام بها لجعل العلاقة خطية ومن أشهر هذه التحويلات تحويل اللوجت (logit)، فتصبح المعادلة السابقة تأخذ الشكل التالي:

$$L = l_n\left(\frac{P_i}{1-P_i}\right) = B_0 + B_1 X_1 + \cdots + B_kX_k$$

ويتم توفيق البيانات في الانحدار اللوجستي بشكل منحنى لوجستي كما هو موضح في الشكل التالي:

الشكل رقم 01: منحنى لوجستي F(z)



المصدر: Jurafsky, & Martin, 2025, p 3

تظهر آلية عمل الانحدار اللوجستي من خلال منحنى دالة سيجمويد التي تعتبر العنصر الأساسي في تحويل المخرجات الخطية للنموذج إلى احتمالات تتراوح بين 0 و1.

على المحور الأفقي يظهر المتغير z وهو المزيج الخطي للمدخلات مع الأوزان $z = w_1x_1 + w_2x_2 + \cdots b$

بينما المحور العمودي يمثل ناتج دالة السيجمويد $\sigma(z)$ أي الاحتمال المتوقع لانتماء العينة إلى الفئة 1، الشكل يظهر كيف أن القيم الصغيرة لـ z تقود إلى احتمالات قريبة من 0 (تصنيف0)، والقيم الكبيرة تقود إلى احتمالات قريبة من 1 (تصنيف1)، بينما تكون نقطة الحسم عند z=0 حيث يكون الاحتمال 0,5.

هذا المنحنى يوضح بدقة كيف يقوم الانحدار اللوجستي بتحويل القيم الخطية إلى احتمالات تستخدم لاحقا لاتخاذ قرار التصنيف النهائي بين الفئتين.

- ✓ **الانحدار اللوجستي متعدد الحدود**

وهو أحد أنواع الانحدار اللوجستي، حيث يعتبر امتدادا بسيطا للانحدار اللوجستي الثنائي، يتم استخدامه في حالة كان المتغير التابع يتكون من أكثر من فئتين تصنيفيتين أو اسميتين. (Anass Bayaga, 2010, p298)

يشترك أسلوب ثنائي الحدين وأسلوب متعدد الحدود في مجموعة من الحدود أهمها: أن كلاهما يعتبر نوعا من أنواع الانحدار اللوجستي والذي يتميز بأن يكون المتغير التابع متغيرا تصنيفيا أو إسميا، كذلك يعتبر كلاهما من التقنيات التي تستعمل بكثرة في الدراسات الطبية، وعموما الدراسات التي تحاول تحديد وتحليل عوامل الخطورة لظاهرة ما، أما أوجه الاختلاف بينهما فتكمن في أن الانحدار اللوجيستي الثنائي يستعمل في حالة كان المتغير التابع المتنبئ به تصنيفيا ثنائيا ويظهر هذا على اسمه، بعكس الانحدار اللوجستي الذي يستعمل في حالة كان المتغير التابع تصنيفيا يحتوي على أكثر من فئتين أو صنفين، وبالضرورة يسكون هناك اختلاف رياضي في بناء معادلة كل أسلوب، كذلك في توظيف كل منهما لنوع خاص من الطرق الإحصائية ومعايير المطابقة النمذجية. (محمد أمين، محمد، 2017، ص4)

- ✓ **الفرق بين الانحدار اللوجستي والانحدار الخطي**

الجدول رقم01: الفرق بين الانحدار اللوجستي والانحدار الخطي

	الانحدار الخطي	الانحدار اللوجستي
الهدف	التنبؤ بقيم عددية مستمرة	التنبؤ باحتمالات النتائج الثنائية أو متعددة الفئات
نوع المتغير التابع	متغير تابع عددي مستمر	متغير تابع ثنائي
الافتراضات	العلاقة خطية بين المتغيرات المستقلة والتابعة	العلاقة بين المتغيرات المستقلة والتابعة غير خطية
القيم المتنبئ بها	قيم عددية مستمرة	قيم احتمالية تتراوح بين 0 و1
طريقة التقدير	استخدام طريقة المربعات الصغرى	استخدام تقنية الحد الأقصى للأرجحية
التفسير	المعاملات تمثل التأثير المباشر لكل متغير مستقل على المتغير التابع	المعاملات تمثل التأثير النسبي على احتمال النتيجة
عدد المتغيرات المستقلة	يمكن التعامل مع عدد كبير من المتغيرات المستقلة	يمكن التعامل مع عدد كبير من المتغيرات المستقلة مع قيود معينة

المصدر: (كرار، 2024، ص19، ص20)

4. شروط تطبيق نموذج تحليل الانحدار اللوجستي

لا بد من تحقيق بعض الشروط عند استخدام الانحدار اللوجستي وتمثلت في:

- ✓ **الجوانب النظرية:** تتفق جميع البحوث على تأكيد أهمية اختيار متغيرات التنبؤ وفقا للنموذج النظري، وفي الانحدار اللوجيستي كغيره من استراتيجيات النمذجة يتم جمع عدد كبير من متغيرات التنبؤ، ثم حذف المتغيرات التي تفتقد الدلالة الإحصائية. (عائشة بنت مربع، 2016، ص25)
- ✓ **الجوانب التطبيقية:** تأتي الافتراضات وبعض الخصائص التي يجب توفرها على البيانات المستخدمة عند تطبيق أسلوب تحليل الانحدار اللوجستي كالآتي: (عبد الحميد، 2011، ص-ص 1-2)

- المتغير التابع متغير وصفي ثنائي أو متعدد، التوقع الشرطي لهذا لمتغير $E\left(\frac{x}{y}\right)$ عبارة عن متغير محدودو بالفترة (1،0)، أما المتغيرات التفسيرية فيمكن أن تكون مستمرة أو متقطعة، وصفية ثنائية أو متعددة، كما يفترض أن جميع المتغيرات تقاس بدون أخطاء.
- هناك علاقة دالية بين المتغير التابع والمتغيرات التفسيرية تأخذ الصورة التالية:

$$P_r\left(Y_k = \frac{1}{x}\right) = \frac{e^{(\beta x)}}{1 + e^{(\beta x)}}$$

- القيمة المتوقعة للخطأ العشوائي تساوي صفر، كما أن تباين الخطأ لعشوائي يساوي (pr(x) (1-pr(x) وأن الخطأ العشوائي (Uk) يتبع توزيع ذو الحدين باحتمال يتحدد على أساس المتوسط الشرطي.
- لا يوجد ارتباط بين الأخطاء العشوائية (استقلال الأخطاء).
- لا يوجد ارتباط بين الأخطاء العشوائية والمتغيرات التفسيرية.
- لا يوجد ارتباط بين المتغيرات التفسيرية بعضها البعض بصورة كاملة، إذ يجب حذف المتغيرات التي توجد بينهما علاقة ارتباط تامة (ضرورة تجاوز مشكلة التعدد الخطي بين المتغيرات المستقلة).
- لا يحدد النموذج إلى أي المجتمعات تنتمي المشاهدات الجديدة فقط ولكنه يحدد أيضا احتمال هذا الانتماء.
- لا يشترط النموذج أن تكون المتغيرات المستقلة الداخلة في النموذج تتبع التوزيع الطبيعي، (سهيلة، 2014، ص57) وتظهر أهمية ذلك في صعوبة توافر هذا الشرط في حالات كثيرة ففي هذه الحالة يكون استخدام الانحدار اللوجستي أفضل من تحليل التمايز. (رضوى عبد الواحد، 2017، ص-ص 81-82)
- كما يشترط تطبيق نموذج تحليل الانحدار اللوجستي توفر حجم مناسب للعينة المستخدمة. (جولي، 2015، ص 212)

5. **مزايا وعيوب الانحدار اللوجستي**

❖ **مزايا الانحدار اللوجستي:** بناء على عدة عوامل يتميز الانحدار اللوجستي بـ: (كرار، 2024، ص17)

- **سهولة التفسير:** يمكن فهم الانحدار اللوجستي بسهولة، حيث يعطي كل عامل مؤثر قيمة أو معامل يشير إلى مدى تأثيره على احتمالية وقوع النتيجة.
- **التعامل مع العديد من المتغيرات التوضيحية:** يمكن استخدام الانحدار اللوجستي مع العديد من المتغيرات التوضيحية المستمرة أو الفئوية.
- **مناسب للبيانات ذات النتائج الثنائية:** يعتبر الانحدار اللوجستي خيارا جيدا للتنبؤ بالنتائج الثنائية مثل تحديد ما إذا كان المريض مصابا بمرض ما أو لا.

- **أقل تعقيدا من التعلم العميق:** يعتبر الانحدار اللوجستي أقل تعقيدا وأقل كثافة في الحوسبة من بعض تقنيات التعلم العميق.
- **شفافية الحسابات:** يمكن التحقق من حسابات الانحدار اللوجستي وتعديلها بسهولة من قبل المطورين عكس التعلم العميق. (كرار، 2024، ص17)

❖ **عيوب الانحدار اللوجستي**

- **لا يتعامل بشكل جيد مع العلاقات غير الخطية:** يفترض الانحدار اللوجستي علاقة خطية بين المتغيرات التوضيحية والنتيجة ولهذا لا يكون جيدا مع العلاقات غير الخطية.
- **حساسية للقيم المتطرفة:** يمكن أن تتأثر نتائج الانحدار اللوجستي بالقيم المتطرفة في البيانات.
- **افتراض الاستقلالية وتجانس التباين:** يفترض الانحدار اللوجستي الاستقلالية بين الأخطاء وتجانس التباين وهو ما قد لا يتحقق دائما في البيانات الواقعية.
- **عدم القدرة على التنبؤ بالقيم المستمرة:** لا يمكن استخدام الانحدار اللوجستي للتنبؤ بالقيم المستمرة بل هو مناسب فقط للقيم الفئوية أو الثنائية. (كرار، 2024، ص17)

ثانيا: أشجار القرار

تعتبر أشجار اتخاذ القرار طريقة بيانية تساعد متخذي القرار على اتخاذ القرار الأمثل من بين البدائل المتاحة فيما يتعلق بمختلف الحالات المستقبلية، حيث يفضل الاعتماد على شجرة اتخاذ القرار في المواضيع المتعلقة بالدراسات المالية والاستثمارية والمبيعات والإنتاج وكل ما يرتبط بدراسة الخسائر والأرباح المحتملة.

1. **مفهوم أشجار القرار**

أشجار القرار هي بنية بيانات هرمية تطبق استراتيجية القسمة والتغلب(divide and conquer) وهي طريقة معلمية فعالة، يمكن استخدامها لكل من التصنيف والانحدار، وهي نموذج هرمي للتعلم المراقب حيث يتم تحديد المنطقة المحلية بسلسلة من التقسيمات التكرارية في عدد أقل من الخطوات. (Ethem, 2014, p213)

2. **مكونات أشجار القرار**

تصنف أشجار القرار الحالات بفرزها من الجذر إلى عقدة فرعية، مما يعطي تصنيفًا للحالة، تحدد كل عقدة في الشجرة اختبارا لسمة من سمات الحالة، ويتوافق كل فرع نازل من تلك العقدة مع إحدى القيم المحتملة لهذه السمة، يصنف كل حالة بالبدء من العقدة الجذرية للشجرة، واختبار السمة التي تحددها هذه العقدة، ثم الانتقال إلى فرع الشجرة المقابل لقيمة السمة في المثال المعطى. تكرر هذه العملية بعد ذلك للشجرة الفرعية التي جذرها العقدة الجديدة. (عطية، بن زخروفة، 2015، ص471)

3. خوارزميات أشجار القرار

تعد الخوارزميات المستخدمة في إنشاء أشجار القرار من بين أكثر طرائق التعلم الآلي شهرة واستخداما عل نطاق واسع، من ضمن خوارزميات أشجار القرار ID3، C4.5،C5.0 وخوارزمية CART، وهذه الخوارزميات تعتبر الأكثر شيوعا.

- ✓ **خوارزمية C4.5**

في عام 1986 وضع الباحث روس كوينلان في مجال تعلم الآلة خوارزمية الانقسام التكراري ID3،(Quinlan J.R, 1986, P 85) ولكن بسبب العيوب التي تعاني منها هذه الخوارزمية حيث أنها تتعامل مع المتغيرات الاسمية فقط، طور كوينلان عام 1993 خوارزمية ID3 إلى خوارزمية C4.5 التي تتعامل مع المتغيرات الاسمية والكمية،(Alka, Ravindra, 2009, P203) حيث أن كلتا الخوارزميتين تعتمدان على مبدأ المعلومات المكتسبة كمعيار في تحديد الانقسامات وبناء شجرة القرار من خلال حساب كمية المعلومات اعتمادا على مفهوم الأنتروبي Entropy الذي طوره Claude Shannon لوصف كمية المعلومات. (غسان، عداس، 2015)

- ✓ **خوارزمية C5.0**

تعد خوارزمية أشجار القرار C5.0 خوارزمية محسنة لـ C4.5، طورت أيضا من قبل كوينلان، إن إنشاء أشجار قرار هذه الخوارزمية يشبه إنشاء أشجار قرار في C4.5، مع الاحتفاظ بجميع وظائف C4.5، تقدم C5.0 المزيد من التقنيات الجديدة يعد تعزيز Boosting أهمها، بالإضافة إلى أنها أفضل من C4.5 من حيث السرعة والذاكرة والكفاءة، يعمل نموذج C5.0 عن طريق تقسيم العينة بناء على أعلى كسب للمعلومات. (يمان، غسان، 2022، ص 12)

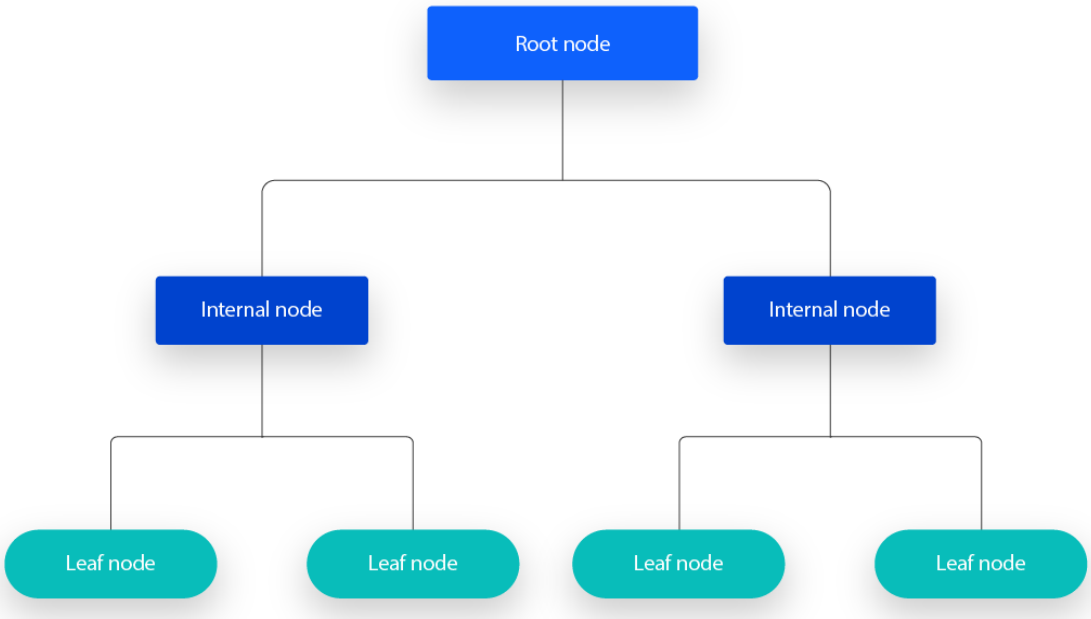
- ✓ **خوارزمية CART**

تم نشر خوارزمية CART لأول مرة بواسطة Leo Breiman في عام 1984، كما يوحي اسمها تستخدم هذه الخوارزمية في هرمية القرار، وهي أداة شائعة لدعم القرار في التعلم الآلي، مما يعني إعطائها مزيدا من المعلومات للتعلم منها في شكل قواعد أساسية للقرار، يمكن بناء نموذجها من خلال بناء كافة الفروع وكلما زاد عدد الفروع الموجودة فيها كان أفضل بحيث يناسب النموذج والغرض منه. (عطية، بن زخروفة، 2015، ص477)

وهي في الأساس عبارة عن هيكل هرمي بسيط نوعا ما يتكون من ثلاثة أنواع مختلفة من العناصر وهي:

- عقدة جذرية واحدة وهي نقطة البداية بما في ذلك جميع عينات التدريب.
- عقد اتخاذ القرار المتعددة حيث تقسم بياناتها باستخدام قواعد بسيطة القرار.
- عقد طرفية متعددة حيث نقوم النهاية بتعيين الفئات لغرض التصنيف الخاص بنا.

الشكل رقم 02: مخطط شجرة القرار



المصدر: موقع شركة IBM (www.ibm.com)

4. مزايا وعيوب أشجار القرار

يستخدم هذا الأسلوب عادة لتوضيح الرؤية المتعلقة بمدة زمنية ليست بالقصيرة وفي الظروف المبنية على حالات المخاطرة، حيث أن لها مزايا وعيوب:

✓ **مزايا أشجار القرار:**

تتمثل في: (Ville, 2006, P160)

- سهلة الفهم والتفسير ويمكن ذلك من خلال تصور الأشجار، وينتج عنها مجموعة من القواعد.
- قادرة على التعامل مع البيانات العددية والفئوية.
- قادرة على التعامل مع المهام التي يكون فيها عدد المخرجات متعددة
- يمكن التحقق من صحة النموذج باستخدام الاختبارات الإحصائية، وهذ يجعل من الممكن حساب كفاءة النموذج بدقة عالية.
- تؤدي الخوارزمية أداء جيد حتى لو تم انتهاك افتراضاتها إلى حد ما بواسطة النموذج الحقيقي الذي تم إنشاء البيانات منه.
- تتبع هذه الخوارزمية نفس النهج الذي يتبعه البشر في اتخاذ القرارات.
- عدد المعلمات الفائقة المراد ضبطها تكاد تكون خالية.

✓ **عيوب أشجار القرار**

تمثلت في: (عطية، بن زخروفة، 2015)

- يمكن إنشاء أشجار شديدة التعقيد لا نعمم البيانات بشكل جيد، وهذه المشكلة تسمى الافراط في التخصيص.
- يمكن أن تكون أشجار القرار غير مستقرة لأن الاختلافات الصغيرة في البيانات قد تؤدي إلى إنشاء شجرة مختلفة تماما، ويتم التخفيف من هذه المشكلة باستخدام أشجار القرار خلال مجموعات.
- تنبؤات أشجار القرار ليست سلسلة أو مستمرة، ولكنها تقديرات تقريبية ثابتة متعددة المستويات، لذلك فهذه الخوارزمية لا تجيد الاستقراء.
- خوارزمية أشجار القرار بشكل عام، تعطي دقة تنبؤ منخفضة لمجموعة البيانات مقارنة بخوارزميات التعلم الآلي الأخرى.
- يمكن أن تصبح الحسابات معقدة عندما يكون هناك العديد من تصنيفات الفئة.

5. أدوات تقييم ائتمان العملاء

✓ **طريقة التحليل المالي**

يعمل التحليل المالي على تحليل البيانات المالية للشركة لتحديد وضعها المالي وأدائها وإمكاناتها. يستخدم المستثمرون والشركات هذه البيانات، المستمدة من البيانات المالية وغيرها من التقارير، لاتخاذ قرارات استراتيجية. (Liberto, 2025)

✓ **طريقة التنقيط الائتماني**

التنقيط الائتماني هو تحليل إحصائي تُجريه الجهات المُقرضة والمؤسسات المالية لتحديد الجدارة الائتمانية للفرد أو الشركة الصغيرة المملوكة. يستخدم المُقرضون التقييم الائتماني للمساعدة في اتخاذ قرار بشأن منح الائتمان أو رفضه. يمكن أن يؤثر التقييم الائتماني على قدرتك على التأهل للحصول على منتجات مالية مثل الرهن العقاري، وقروض السيارات، وبطاقات الائتمان، والقروض الشخصية. (Investopedia, 2023)

✓ **طريقة أنظمة الخبراء**

تُعدّ أنظمة الخبراء (ES)، والتي يُشار إليها أحيانًا باسم أنظمة دعم القرار القائمة على المعرفة، أحد تطبيقات الذكاء الاصطناعي التي تُساعد صُنّاع القرار على اتخاذ قرارات أفضل. (Mahmoud et ali, 2008, p 03)

ثالثا: الدراسات السابقة

شهدت تقنيات تعلم الآلة اهتماما متزايدا في مجالات التصنيف والتنبؤ، خاصة باستخدام أشجار القرار ونماذج الانحدار، وقد تناولت العديد من الدراسات السابقة هذه الخوارزميات من حيث المفاهيم النظرية او التطبيقات العملية، إضافة إلى مقارنتها مع تطبيقات أخرى لتحسين الأداء ودقة النتائج.

يمثل هذا العرض خلاصة لأبرز الأعمال السابقة التي تشكل الأساس العملي لهذه الدراسة.

- **دراسة Thomas, L.C, Edelman, D. B. et Crook, J.N (2002) بعنوان: استخدام أشجار القرار في تصنيف القروض**

في كتابهم Credit Scoring a,d Its Aapplications حيث يناقش فيه النماذج الإحصائية والتعلم الآلي المستخدمة في تقييم الجدارة الائتمانية، مع التركيز على فعالية كل نموذج حسب طبيعة البيانات.

يبرز المؤلفون أن الانحدار اللوجستي يعد النموذج الاحصائي الأكثر استخداما في تصنيف القروض بسبب بساطته وسهولة تفسير معاملاته، إلا أن هذا النموذج يفترض علاقة خطية بين المتغيرات المستقلة والنتيجة مما يقيد دقته في الحالات الأكثر تعقيدا.

في المقابل يشير الكتاب إلى أن أشجار القرار تقدم مزايا مهمة، خصوصا في القدرة على معالجة العلاقات غير الخطية بين المتغيرات، وكذلك عند التعامل مع البيانات غير المتوازنة أو تلك التي تحتوي على تفاعلات متداخلة بين السمات، كما أن الشكل الهرمي لشجرة القرار يتيح تفسيرا مرئيا واضحا لمسار اتخاذ القرار، ما يجعلها مفضلة في سياقات تتطلب الشفافية في التقييم، مثل قرارات منح القروض.

- **دراسة Kotsiantis, S.B (2007) بعنوان: A Review of Classification Techniques تفوق أشجار القرار في بعض سياقات التصنيف**

يشير Kotsiantis في مقاله إلى أن اختيار خوارزمية التصنيف يعتمد بدرجة كبيرة على طبيعة البيانات والمشكلة محل الدراسة.

أبرز الباحث في دراسته أن الانحدار اللوجستي يقدم أداء جيدا عندما تكون العلاقة بين المتغيرات المستقلة والمتغير التابع خطية، لكنه يفقد فعاليته في حالات البيانات ذات البنية المعقدة أو غير الخطية.

من جهة أخرى تتميز أشجار القرار بالقدرة على التعامل مع العلاقات الخطية، ومع البيانات التي تحتوي تفاعلات مع المتغيرات، كما أنها لا تتطلب افتراضات قوية حول توزيع البيانات، ويوضح أن سهولة تفسير نتائج شجرة القرار ووضوح منطقتها يجعلها أكثر ملاءمة في التطبيقات التي تتطلب الشفافية في اتخاذ القرار، مثل تصنيف القروض أو تقييم المخاطر.

- **دراسة Joze Bencina (2011) بعنوان: Fuzzy Decision Trees As a Decision-Making Framework in the Public Sector أشجار القرار الضبابية لصنع القرار في القطاع العام.**

ذكرت الدراسة في مقال وقد استخدمت منهج التصنيف والتحليل في استخدام أشجار القرار والمنطق الضبابي، لدراسة حالة الاستثمار وتقيم المشاريع ضمن قسم البلدية في سلوفينيا، بالإضافة لتقييم كفاءة وفعالية المؤسسات العامة وذلك بإيجاد حلول لها خصائص معدلة خصيصا للاستخدام في القطاع العام، ويهدف البحث إلى جعل تطبيقات التقييم تقدم تقييما دقيقا وبشكل أفضل، وتكون سهلة الوصول والاستخدام، ويمكن تطبيق النموذج للتقييم في كافة المجالات ضمن القطاعين العام وغير العام، وتم استخدام مكتبة برمجة خاصة لبناء النموذج وكانت نتائج البحث دقيقة وجيدة لمساعدة متخذي القرار للحصول على قرار دقيق وواضح.

- **دراسة Renuka D, Suryawanshi, D.M.Thakore (2012) بعنوان: Implementation of Decision Tree for Classification with Fuzzy Logic تنفيذ شجرة القرار للتصنيف مع استخدام المنطق الضبابي.**

جاء في المقال أن الدراسة استخدمت المنهج الوصفي، تم ضمن هذه الدراسة بناء شجرة القرار التي باستخراج البيانات وتصنيفها بشكل مناسب، ثم تم اختيار المتغيرات ذات التأثير الأكبر من غيرها واعتبارها مدخلات المنطق الضبابي للوصول على النتائج المطلوبة، بعد ذلك تم مقارنة النتائج في حال استخدام شجرة القرار فقط أو شجرة القرار مع المنطق الضبابي لمعرفة الطريقة الأفضل بينهما وذلك باستخدام بيانات سوق الأوراق المالية والحصول على البيانات من قاعدة بيانات UCI مفتوحة المصدر، حيث كانت دقة النموذج 98% وهي دقة عالية.

- **دراسة Nwachuckwu, Kabaril (2013) بعنوان: Credit Risk Evaluaating System Using Decision Tree- Neuro Based Model تطوير نموذج دعم قرار باستخدام شجرة القرار والشبكات العصبية لدعم نظام منح القروض البنكية.**

في هذا المقال تم استخدام المنهج الوصفي التحليلي حيث أصبحت اتخاذ القرار في بيئة معقدة ومتغيرة في يومنا هذا بحاجة إلى استخدام تقنيات جديدة في الذكاء الاصطناعي للحصول على نظام دعم قرار ذكي، وقد تم تطبيق النموذج المقترح لمعالجة نظام دعم القرار على طلبات القروض الموجهة للبنوك والوصول إلى القرار المناسب لنمح القروض للزبائن، وتم الاعتماد على بيانات البنك في نيجيريا، وتم استخدام ضمن الدراسة لغة برمجية غرضية التوجه (OO-AD) ضمن الأداة Mat lab لتنفيذ الدراسة المقترحة وكان معدل نجاح النموذج 88% ونتائج القرارات جيد وتساعد متخذي القرار على الوصول إلى القرار بشكل أفضل وأسرع.

- **دراسة Masud R, Rashedur M (2013) بعنوان: Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing مقارنة أسلوب شجرة القرار وأسلوب بايو في التسويق المصرفي المباشر.**

جاء في المقال أن هذه الدراسة استخدمت نهج المقارنة وخوارزمية C4.5 لبناء شجرة القرار وأسلوب بايز حيث تم تحليل 4521 سجلا تم اختيارهم بشكل عشوائي من بيانات UCI(موقع البيانات مفتوحة المصدر) المتاحة لجمهور التدريب والاختبار، استخدمت هذه الدراسة 15 متغيرا تتعلق المتغيرات ببينات العملاء الشخصية والبيانات المصرفية للعملاء وبيانات تتعلق بالاتصالات بالعميل، عملت الدراسة أيضا على مقارنة الأسلوبين فقط، باستخدام المقاييس والتصنيفات الصحيحة والخاطئة وذلك من خلال الاعتماد على أداة Weka، كما ركزت الدراسة على استنباط المعرفة الموجودة في البيانات، وقد بينت الدراسة أن نموذج شجرة القرار المبني باستخدام خوارزمية C4.5 هو الأفضل.

- **دراسة سهيلة حمود عبد الله الفرهود (2014) بعنوان: استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم (دراسة تطبيقية على سوق الكويت للأوراق المالية).**

تناول هذا البحث استخدام أسلوب الانحدار اللوجيستي الثنائي لدراسة أهم المؤشرات المالية المؤثرة على أداء الأسهم في سوق الكويت للأوراق المالية، وقد توصلت الدراسة إلى مجموعة من النتائج أهمها كفاءة النموذج المستخلص بشكل كلي للتنبؤ بأداء أسهم الشركات قيد البحث إلى نوعين (جيد ورديء) وذلك بدقة عالية بلغت 74.2% وكان ترتيب النسب المالية المؤثرة معنويا عل أداء الأسهم بالسوق الكويتي وفقا لأهميتها النسبية كالتالي: ربحية لسهم، نسبة الدين إلى حقوق الملكية، نسبة السعر السوقي إلى القيمة الدفترية، نسبة التداول، نسبة صافي الإرادات البنكية إلى الأصول الكلية وأخيرا العائد على حقوق الملكية.

- **دراسة Abbas S (2015) بعنوان: Deposit Subscribe Prediction Using Data Mining Techniques Based Real Marketing Dataset مقارنة أساليب شجرة القرار مع أسلوب المجموعات التقريبية اعتماد على بيانات العملاء المشتركين.**

تناول هذا المقال منهج التصنيف وقامت بالمقارنة بين أسلوب شجرة القرار المبنية بخوارزمية C4.5، وأسلوب المجموعات التقريبية Rough Set Theory، اعتمد الدراسة علة 4521 سجلا من بيانات بنك في برتغاليا، تحوي هذه البيانات على 17 متغيرا تتعلق المتغيرات ببيانات العملاء الشخصية، ومعلومات التواصل، واستخدمت أداة Weka بإصدار 3-6-11 وركزت الدراسة على القواعد المستخرجة وتحليل النموذج دون التركيز على النموذج، وكانت النتائج تعبر عن أن القواعد التي تم استخراجها من النموذج تساعد متخذي القرار على أداء القرار بشكل أفضل وأسرع بمنح القروض للأفراد بإضافة إلى إدارة المخاطر.

- **دراسة Vaidehi R (2016) بعنوان: Comparison of Three Different Techniques Generalized Linear Model, Decision Tree Algorithm C4.5, and Rule-Based Modeling مقارنة بين ثلاث تقنيات مختلفة النموذج الخطي العام وأشجار القرار خوارزمية C4.5 والنمذجة القاعدية.**

جاء في هذا المقال أن ا هذه الدراسة استخدمت المنهج المقترح لتعزيز العملية التنبؤية، وتتناول البيانات الحقيقية المقدمة من آلية UCI (موقع البيانات مفتوحة المصدر)، تم جمع مجموعة البيانات من بنك التجزئة البرتغالي، من أيار 2008 إلى حزيران 2013، ترتبط مجموعة البيانات بحملات التسويق المباشر لمؤسسة مصرفية برتغالية، وقد تم استخدام الأداة R وتطبيقها على الدراسة حيث تم تحليل 45211 سجلا واستخدمت 16 متغيرا تتعلق ببيانات العميل الشخصية وبيانات العميل المصرفية، وبيانات تتعلق بالاتصالات بالعميل، وخلصت الدراسة إلى تحديد 6 متغيرات تكون أفضل في بناء النموذج، وركزت الدراسة على المقارنة بين الأسلوب الأول باستخدام كامل المتغيرات، والأسلوب المقترح باستخدام متغيرات محددة، وخلصت الدراسة إلى أن المتغيرات التي تم تحديدها وهي 6 متغيرات فقط واختبرت بنموذج المصنفات القاعدية وكانت الأفضل.

- **دراسة علاق وليد (2016) بعنوان: "تحليل المؤشرات المالية للتوقع بالتعثر المالي للمؤسسات دراسة تطبيقية على عينة من المؤسسات بولاية ورقلة".**

هدفت هذه الدراسة إلى التوصل لأفضل مجموعة من المؤشرات المالية التي يمكن استخدامها للتنبؤ بتعثر المؤسسات الاقتصادية في قطاع ورقلة، فالتنبؤ بالتعثر هذه المؤسسات في وقت مبكر يعطي مؤشر للجهات المعنية للتدخل واتخاذ الإجراءات التصحيحية المناسبة قبل أن تفشل هذه المؤسسات نهائيا.

ولتحقيق ذلك قمنا باحتساب سبعة عشرة نسبة مالية لفترة 2012 على عينة من المؤسسات والمتكونة من 20 مؤسسة (5 منها متعثرة والباقي سليمة)، ومن ثم تحليل هذه النسب المالية التي يمكن لها التمييز بين المؤسسات المتعثرة والغير المتعثرة وهي على الترتيب التالي:

- ✓ نسب الهيكلية(A1): الأصول الجارية / إجمالي الأصول.
- ✓ نسب الهيكلية(A2): الأصول الغير الجارية / إجمالي الأصول.
- ✓ نسب المردودية (A7): النتيجة الصافية / رقم الأعمال.

- **في دراسة لجامعة Kennesaw State University(2017) بعنوان: A Comparison of Decision Tree with Logistic Regression Model for Commercial Credit Risk Prediction**

جاء في المقال أنه اعتمدت هذه الدراسة على منهج المقارنة الكمية بين نموذجين شائعين في التنبؤ بالمخاطر الائتمانية، وهما الانحدار اللوجستي وأشجار القرار وذلك في سياق تقييم القروض التجارية، تم استخدام مجموعة بيانات حقيقية من مؤسسة مالية، وشملت خصائص مالية وتجارية مرتبطة بالمقترضين.

كشفت النتائج أن الانحدار اللوجستي قدم أداء أعلى من حيث الدقة والقدرة التنبؤية مقارنة بأشجار القرار، إضافة إلى تفوقه في قابلية التفسير وتحديد المتغيرات المؤثرة بشكل مباشر على قرار منح القرض، كما لوحظ أن

أشجار لقرار كانت أكثر عرضة للتقلب والتكيف الزائد مع البيانات (Overfitting) مما قلل من موثوقيتها في سياق التنبؤ المستقبلي.

خلصت الدراسة إلى أن الانحدار اللوجستي هو النموذج الأكثر ملائمة في البيئات التي تتطلب قرارات دقيقة ومستندة إلى تفسير واضح للمتغيرات، خصوصا في حالة القروض التجارية التي تنطوي على مخاطر مالية عالية.

- **دراسة Andrea et al Brunello (2019) بعنوان: A Novel Decision Tree Approach for the Handling of Sequential and Time Series Data**

معالجة البيانات المتسلسلة والوقتية للحالات المرضية باستخدام أشجار القرار.

تم استخدام المنهج التحليلي ضمن الدراسة والاعتماد على المتغيرات الطبية لمعالجة حالة مرضية، والتنبؤ من خلال البيانات المتسلسلة والوقتية باستخدام أشجار القرار والوصول إلى القرار والنتائج المطلوبة للمرض بسنبة كبيرة كما توصلت النتائج إلى أنه يمكن معرفة المريض من خلال معرفة فصيلة الدم والجنس والسمات الرقمية (العمر والوزن والطول) والبيانات المتسلسلة (تاريخ الأعراض التي تعرض لها، الوصفات الطبية التي أعطيت له، والسلسة الزمنية تطورت مع مرور الوقت من ضغط دمه، ومعدل ضربات القلب له) ضمن فترات زمنية معينة لكل مريض لدراسة الحالة، قد يكون اتجاه البحث هو تمديد النموذج من أجل التعامل مع صيغ المنطق الزمني، استخدام البحث البرامج الإحصائية بالإضافة إلى رسم شجرة القرار ضمن برنامج Weka.

- **دراسة Zhang et al SHS Web of Conferences (2024) بعنوان: Research on Loan Approval and Credit Risk Based on the Comparison of Machine Learning Models**

اعتمدت هذه الدراسة منهجا تجريبيا للمقارنة بين عدة نماذج من تقنيات التعلم الآلي في سياق الموافقة على القروض وتقييم المخاطر الائتمانية، وشملت النماذج التالية: الانحدار اللوجستي، أشجار القرار، الغابات العشوائية وXGBoost، حيث تم استخدام بيانات حقيقية من منصة إقراض صينية تحتوي على خصائص ديمغرافية ومالية للمقرضين.

أظهرت النتائج أن الانحدار اللوجستي حقق أداء جيدا من حيث الدقة والاتساق في التنبؤ، بالإضافة على قدرته على تفسير تأثير المتغيرات بسهولة، ما يجعله خيرا مناسبا للمؤسسات التي تعطي الأولوية للوضوح والشفافية في قراراتها، أما أشجار القرار فرغم تقديمها لمخرجات بصرية وتفسيرية إلا أنها أظهرت ضعفا نسبيا في الأداء في حالة البيانات الكبيرة أو غير المتوازنة.

خلصت الدراسة إلى أن الانحدار اللوجستي يعد أكثر ملائمة في التطبيقات التي تتطلب قرارات موثوقة وسهلة التفسير، مقارنة بأشجار القرار التي قد تكون أكثر عرضة للتكيف المفرط أو التباين في الأداء.

- دراسة Shih-JiunLin (2024) بعنوان **Prediction Models Using Decision Tree and Logistic Regression Method for Predicting Hospital Revisits in Peritoneal Dialysis Patients.**

جاءت هذه الدراسة في مجلة Diagnostic وتم تطبيق نموذجين تنبؤيين هما شجرة القرار والانحدار اللوجستي بهدف التنبؤ بزيارات المستشفى المتكررة لمرضى غسيل الكلى البريتوني، تشير الدراسة إلى أن اختبار النموذج الأنسب يعتمد على طبيعة البيانات وتعقيد العلاقة بين المتغيرات المستقلة والثابتة.

يظهر الانحدار اللوجستي الأداء الجيد في الحالات التي تتميز بعلاقات خطية بين المتغيرات، إذ يفترض هذا النموذج وجود علاقة خطية بين المدخلات واحتمالية النتيجة، لكن في كثير من الحالات مثل بيانات مرضى غسيل الكلى، تكون العلاقات معقدة وغير خطية، وهو ما يحد من كفاءة الانحدار اللوجستي في التنبؤ بدقة عالية.

في المقابل تتمتع أشجار القرار بقدرة عالية في التعامل مع العلاقات غير الخطية والتفاعلات المعقدة بين المتغيرات دون الحاجة لافتراضات مسبقة عن توزيع البيانات، كما توفر أشجار القرار نتائج سهلة التفسير وذات شفافية عالية في اتخاذ القرار، مما يجعلها مناسبة جدا في السياقات الطبية حيث الشرح الواضح للنتائج أمر ضروري.

وبناء على ذلك فإن أشجار القرار تعد خيارا متفوقا أو مكملا للانحدار اللوجستي في التنبؤ بزيارات المستشفى المتكررة لمرضى غسيل الكلى البريتوني، خصوصا في ظل تعقيد وتداخل عوامل المرضى.

رابعا: الفجوة البحثية

على ضوء هذه الدراسات نلاحظ أن موضوع تصنيف العملاء في المؤسسات المالية حسب الملاءة المالية موضوع واسع التطبيق خاصة باستخدام نموذجي أشجار القرار والانحدار اللوجيستي، مع نتائج تارة تظهر تفوق الأشجار القرار وتارة نموذج الانحدار اللوجيستي حسب بيانات كل دراسة، غير أننا نلاحظ نقصا واضحا في تطبيق مثل هذه النماذج الحديثة المستندة على الذكاء الاصطناعي في السياق الجزائري، ففي حال تم تطبيقها اقتصرت على البنوك مع ندرتها على العموم، خاصة في إطار وكالات تمويل المشاريع، لذلك جاءت هذه الدراسة كمحاولة لسد الفجوة الموجودة في تطبيق نماذج تعلم الآلة في التنبؤ بالملاءة المالية للعملاء في وكالات التمويل الجزائرية، وإن وجدت فهي تستخدم بشكل بسيط دون تحضير معمق للبيانات، مثل: تقنية SMOTE للتعامل مع مشكلة عدم توازن البيانات وكذا تحسين دقة المتغيرات الأكثر أهمية المتحصل عليها من خلال تقنية Stepwise وهذا ما ستعتمد على ادراجه هذه الدراسة.

الفصل الثالث:
نتائج الدراسة

1. منهج الدراسة

من أجل تصنيف عملاء وكالة سطيف لتسيير القرض المصغر ANGEM إلى مقبولين أو مرفوضين، اعتمدت هذه الدراسة على المنهج الوصفي التحليلي باعتباره الأنسب لطبيعة الموضوع، حيث يهدف إلى وصف واقع قرارات منح القروض المصغرة وتحليل العوامل المؤثرة فيها من خلال معالجة بيانات فعلية من الوكالة، وتم تعزيز هذا المنهج باستخدام أدوات إحصائية وتقنيات تعلم آلي (Machine Learning) لاستكشاف الأنماط المؤثرة في اتخاذ قرار القبول أو الرفض، اعتمدت الدراسة على مجموعة بيانات كمية وكذا كيفية تضم معلومات متنوعة عن الأفراد المتقدمين بطلبات القروض. وقد اشتملت هذه البيانات على عدد من المتغيرات مثل: الاسم، اللقب، البلدية، الولاية، وقيمة القرض المطلوب، إلا أن هذه المتغيرات لم تُظهر فاعلية في بناء النماذج التنبؤية، فتم استبعادها والتركيز على المتغيرات الأكثر تأثيرًا، وهي: الجنس، العمر، المستوى الدراسي، الشهادة. أما المتغير التابع فتمثل في قرار القبول أو الرفض لطلب القرض.

2. عينة الدراسة

بلغ حجم العينة 546 أي 546 متقدم للوكالة.

3. أداة الدراسة

تمثلت أداة الدراسة في مجموعة بيانات رقمية تضم معلومات عن الأفراد المتقدمين بطلبات قروض تشمل المتغيرات المستقلة: النوع، السن، المستوى الدراسي، الشهادة، فيما كان المتغير التابع هو قرار القبول أو الرفض.

4. مصادرجمع البيانات

تم جمع هذه البيانات إما من قواعد بيانات مؤسسة تمويل مصغر أو تم إنشاؤها بشكل يحاكي الواقع العملي.

5. المعالجة الإحصائية المستخدمة

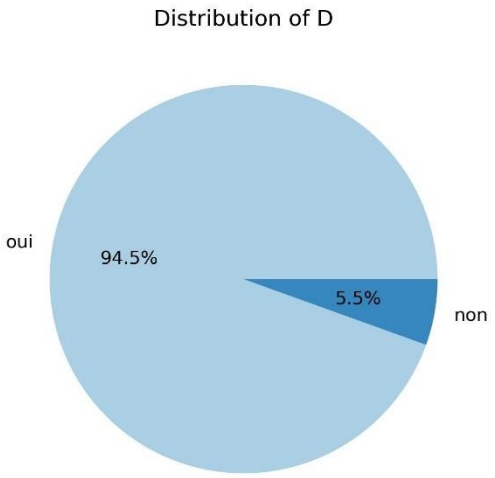
تمت المعالجة الإحصائية باستخدام لغة البرمجة Python، وذلك بالاعتماد على مكتبات pandas لمعالجة البيانات scikit-learn لبناء النماذج التنبؤية.

تم تطبيق نموذج الانحدار اللوجستي ونموذج أشجار القرار، مع استخدام تقنية الموازنة (SMOTE) لمعالجة مشكلة عدم توازن الفئات المستهدفة. ومن ثم تم تقييم أداء النماذج بناءً على مؤشرات إحصائية معتمدة، وهي: الدقة (Accuracy)، الاسترجاع (Recall)، الضبط (Precision) وF1-Score.

6. الدراسة الوصفية

تعد الدراسة الوصفية مرحلة محورية في التحليل الميداني، إذ تتيح فهما أوليا لبيئة الدراسة من خلال استعراض الخصائص العامة للمستفيدين من القروض المصغرة لدى وكالة « ANGEM» بسطيف، ويشكل هذا الوصف قاعدة أساسية لفهم السياق الاجتماعي والاقتصادي للعينة، مما يمهد لبناء نماذج تنبؤية أكثر دقة وواقعية.

الشكل رقم03: حالات القبول والرفض لطلبات القروض.

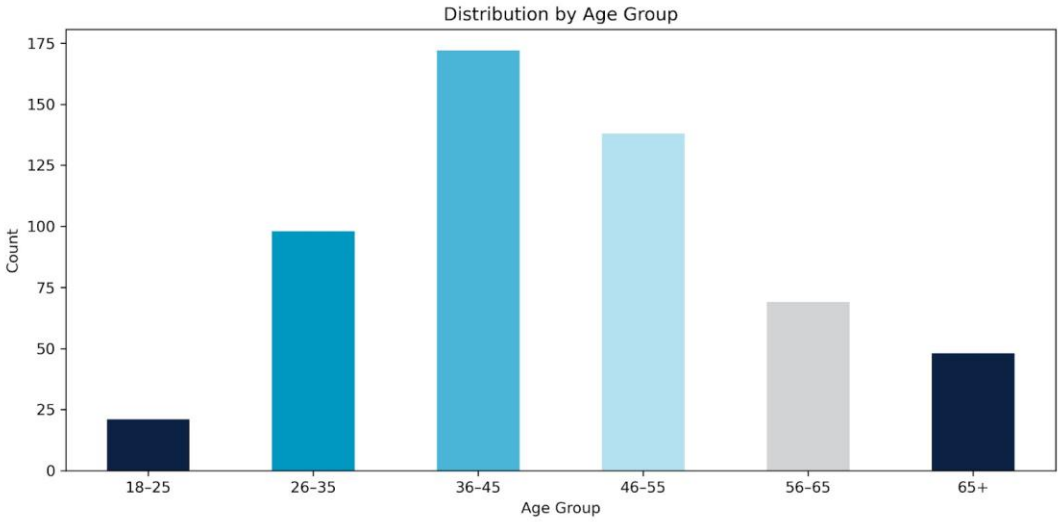


المصدر: من إعداد الطالبة باستخدام لغة برمجة بايثون

يمثل الشكل رقم 03 توزيع قرارات منح القروض، حيث تظهر النتائج أن 5.94% من الطلبات تم قبولها مقابل 5.5% فقط تم رفضها، هذا يشير إلى أن الغالبية العظمى من المتقدمين يحصلون على الموافقة، مما قد يعكس سياسة تمويل مرنة أو استهداف شريحة ذات موثوقية مالية عالية، غير أن هذا التوزيع غير المتوازن قد يشكل

تحديا في حال استخدامها لتدريب نماذج تنبؤية، حيث قد نواجه مشكلة "البيانات غير المتوازنة" (Imbalanced Data)، مما قد يؤدي إلى انحياز النموذج نحو فئة القبول على حساب الدقة في التنبؤ بحالات الرفض.

الشكل رقم 04: توزيع أعمار المتقدمين لطلب القروض من الوكالة.



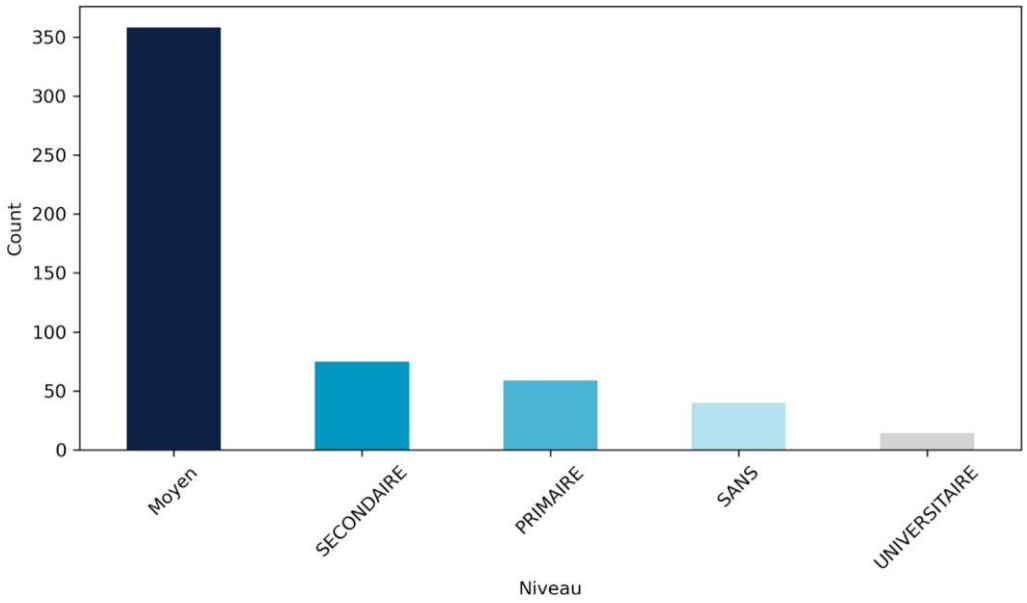
المصدر: من إعداد الطالبة باستخدام لغة برمجة بايثون

يبين مخطط الأعمدة في الشكل رقم 04 أن الفئة العمرية الأكثر إقبالا على الاقتراض من مؤسسة تمويل القرض المصغر ANGEM هي فئة الأشخاص الذين تتراوح أعمارهم بين 34 و45سنة، تليها الفئة الممتدة من 46 إلى 56سنة، في المقابل تعد فئة الشباب الذين تتراوح أعمارهم بين 18 و25سنة الأقل لجوء إلى القروض المصغرة. يعكس هذا التوزيع أن الأشخاص في منتصف العمر هم لأكثر استعدادًا لبدء مشاريعهم الخاصة، وربما يعود ذلك إلى استقرارهم المهني وامتلاكهم خبرة كافية في سوق العمل، ما يدفعهم للسعي نحو الاستقلال المالي ومزاولة نشاط حر بدلًا من الوظيفة التقليدية، خاصة وأنهم لا يزالون في سن تسمح لهم بالنشاط والمردودية.

أما انخفاض إقبال الفئة الشابة فقد يعزى إلى حداثة عهدهم بسوق الشغل، أو لانشغالهم بالدراسة الجامعية، ما يجعلهم أكثر ترددًا في خوض تجربة ريادة الأعمال بسبب قلة الخبرة، في حين أن الفئة المتقدمة في

السن، خصوصًا من تجاوزوا 60 سنة، قد يعزفون عن الاقتراض نتيجة ضعف القدرة الانتاجية أو التقاعد، مما يقلل من دافعهم لبدء مشروع جديد.

الشكل رقم 05: توزيع المستوى الدراسي للعملاء



المصدر: من إعداد الطالبة باستخدام لغة برمجة بايثون

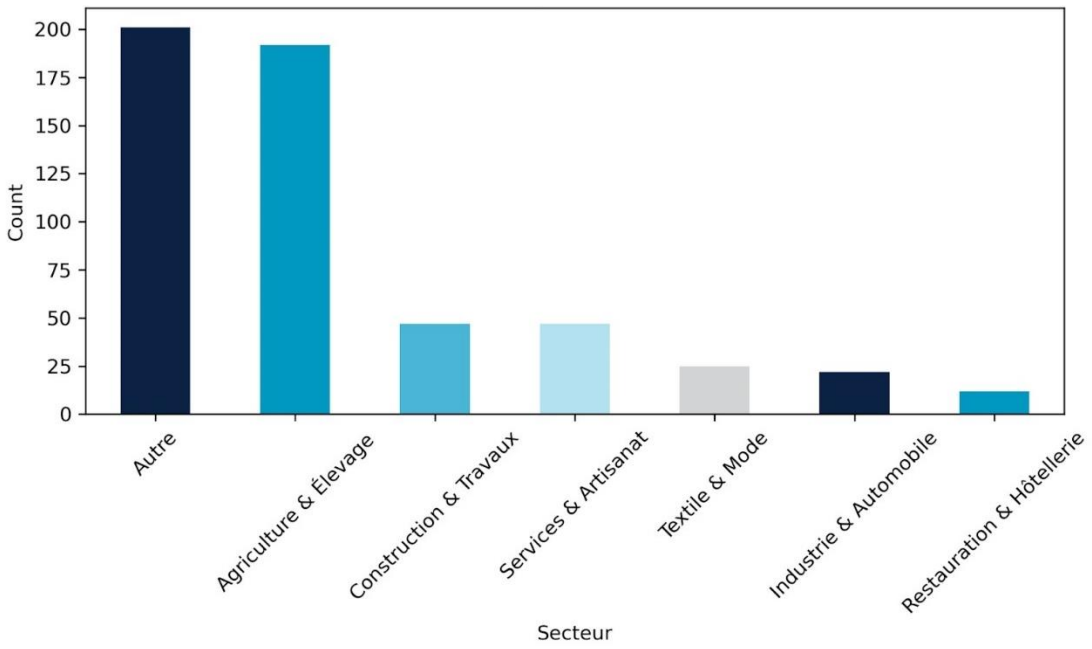
يظهر الرسم البياني تفوّقًا ملحوظًا لفئة "بدون تعليم " من حيث عدد المستفيدين من القروض المصغرة، تليها فئتا التعليم الابتدائي والثانوي.

هذا التوزيع يشير إلى أن الغالبية العظمى من المستفيدين هم من ذوي المستوى التعليمي لمحدود أو الأميين، وهي فئة غالبًا ما تواجهها صعوبات في الولوج إلى سوق العمل الرسمي نظرًا لضعف مؤهلاتها الأكاديمية والمهنية.

إن لجوء هذه الفئة إلى القروض المصغرة يمكن تفسيره برغبتها في تحسين أوضاعها الاقتصادية والاجتماعية، إذ تُعد هذه القروض وسيلة بديلة تمكنها من ولوج عالم الشغل من خلال ممارسة أنشطة مدرّة للدخل بشكل مستقل، خصوصًا في القطاعات غير الرسمية أو الحرفية التي لا تتطلب تأهيلا تعليميا عاليا.

كما يمكن اعتبار ذلك مؤشرًا على الدور الاجتماعي المهم الذي تلعبها مؤسسات التمويل المصغر، حيث تساهم في إدماج الفئات الهشة وتمكينها اقتصاديا، ما ينعكس ايجابيا على محاربة البطالة والتقليص من الفوارق الاجتماعية.

الشكل رقم 06: توزيع المستفيدين وغير المستفيدين من القروض حسب مجال النشاط الاقتصادي



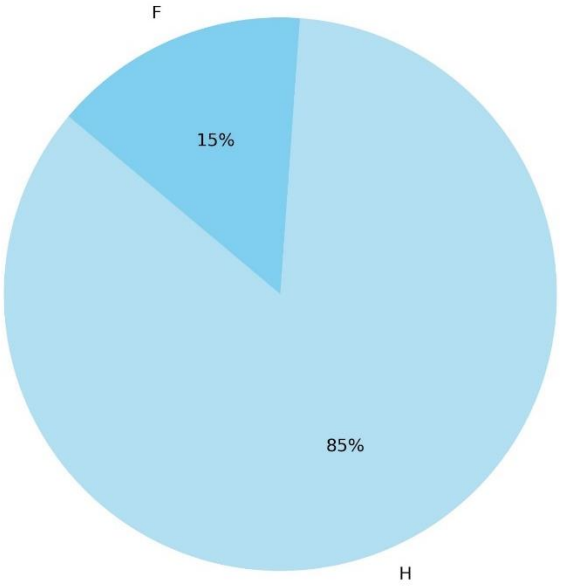
المصدر: من إعداد الطالبة باستخدام لغة برمجة بايثون

يوضح الرسم البياني في الشكل رقم 06 أن معظم المستفيدين من القروض المصغرة ينشطون في قطاعي الزراعة والحرف التقليدية، بالإضافة إلى أنشطة اقتصادية بسيطة أخرى، في حين تسجّل باقي القطاعات نسبًا أقل بكثير، وتُعد المطاعم والفندقة من بين أضعفها تمثيلا.

يعكس هذا التوزيع طبيعة القروض المصغرة، التي غالبًا ما توجها نحو تمويل المشاريع الصغيرة ذات المتطلبات المالية المحدودة، مثل الأنشطة الزراعية والحرف اليدوية، والتي لا تحتاج إلى تجهيزات أو رأس مال كبير. كما أن هذه القطاعات تتناسب مع مستوى المستفيدين، الذين غالبًا ما يكونون من ذوي التعليم المحدود أو دون تأهيل أكاديمي، ما يجعل هذه الأنشطة ملاذًا مناسبًا لهم للدخول في عالم الشغل.

في المقابل، القطاعات الأخرى كالفندقة والمطاعم تتطلب عادة تمويلا أكبر، ومعرفة إدارية وتقنية أكثر تعقيدا، ما يفسر ضعف الإقبال عليها في إطار القروض المصغرة، خاصةً من طرف الفئات الاجتماعية الهشة التي تفتقر إلى الخبرة أو الإمكانيات الكافية.

الشكل رقم07: نسبة جنس العملاء لدى الوكالة.



المصدر: من اعداد الطالبة باستخدام لغة برمجة بايثون.

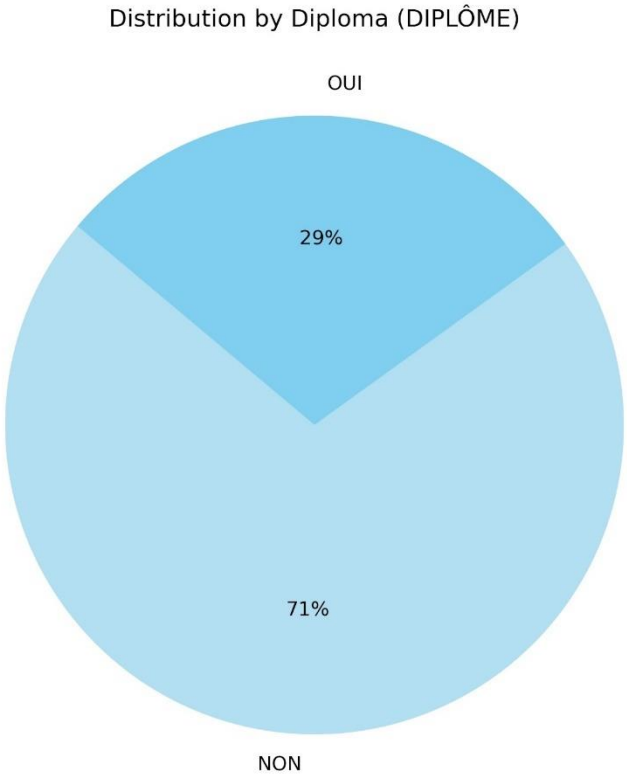
تظهر الدائرة النسبية أن الذكور يمثلون النسبة الأكبر من المستفيدين من القروض المصغرة، حيث يشكلون حوالي 85% من إجمالي المستفيدين، مقابل 15% فقط من الإناث.

يعكس هذا التفاوت الواضح في نسب الاستفادة وجود فوارق في المشاركة الاقتصادية بين الجنسين، إذ يسجل حضور أقوى للرجال في مختلف الأنشطة الاقتصادية، ولا سيما في القطاع الفلاحي والحرفي، وهي مجالات تستقطب النسبة الأكبر من مشاريع القروض المصغرة.

أما انخفاض نسبة الإناث فيرجع في الغالب إلى قيود اجتماعية وعائلية تحد من مشاركة المرأة في سوق العمل خاصة في المناطق الريفية، بالإضافة إلى الطبيعة البدنية لبعض الأنشطة التي تتطلب جهدا جسديا أكبر، ما

يجعل النساء أقل إقبالاً على طلب هذا النوع من القروض، رغم وجود برامج تحفيزية موجهة لتمكين المرأة اقتصاديا.

الشكل رقم 08: نسبة المتقدمين الحاصلين على شهادة تعليمية وغير الحاصلين



المصدر: من اعداد الطالبة باستخدام لغة برمجة بايثون.

تبين الدائرة النسبية أن حوالي 71% من المستفيدين من القروض المصغرة لا يمتلكون أي شهادة، ما يشير إلى أن الغالبية العظمى من المستفيدين هم من ذوي المستوى التعليمي المنخفض أو المنعدم.

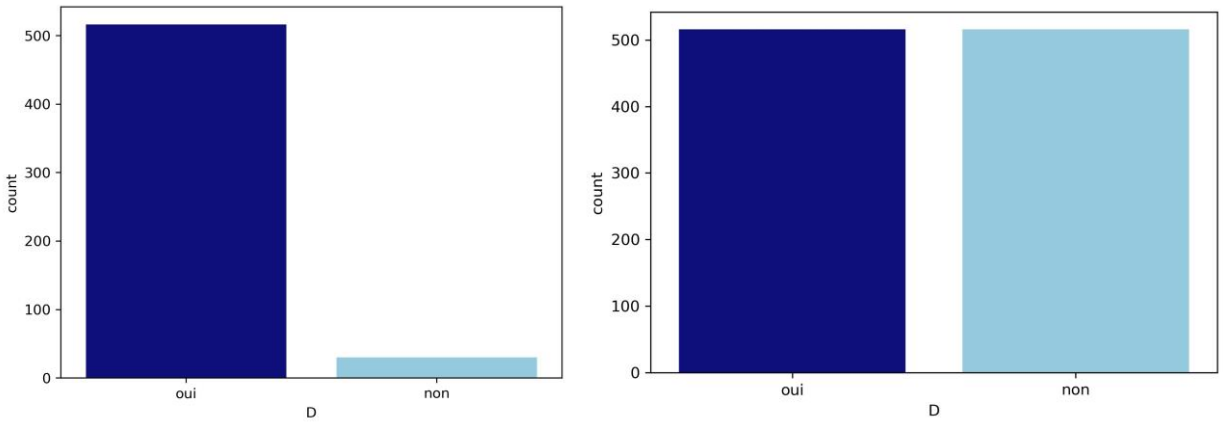
يعكس هذا الواقع أن آلية التمويل المصغر تُوجها أساسًا نحو الفئات الهشة والضعيفة التكوين، من أجل دعم ادماجها المهني والاقتصادي، فهذه الفئة التي غالبًا ما تُستبعد من سوق العمل الرسمي بسبب محدودية

مؤهلاتها، تجد في القروض المصغرة فرصة لإنشاء مشاريع بسيطة ومدرّة للدخل، ما يساهم في الحد من البطالة والفقر وتعزيز الإدماج الاجتماعي.

من أجل دراسة تأثير العوامل الاجتماعية والاقتصادية على منح القروض المصغّرة باستخدام أشجار القرار، واجهنا في البداية مشكلة عدم توازن البيانات كما هو موضح في الشكل 02، حيث كانت إحدى الفئتين (مثلاً: العملاء غير الحاصلين على قرض) تمثل نسبة صغيرة مقارنة بالفئة الأخرى (العملاء الحاصلين على قرض). ولمعالجة هذا الخلل، تم استخدام تقنية SMOTE (اختصارًا لـ Synthetic Minority Over-sampling Technique)، وهي تقنية تهدف إلى زيادة عينات الفئة الصغرى بشكل اصطناعي، مما يسمح للنموذج بالتعلم بشكل أكثر عدلاً دون الانحياز نحو الفئة الكبرى.

بعد تطبيق SMOTE، أصبحت البيانات متوازنة بين الفئتين، كما يظهر في الشكل 09، وهو ما يعد خطوة أساسية لتحسين أداء نموذج التصنيف وتقليل الانحياز الناتج عن عدم التوازن.

الشكل رقم 09: توزيع الفئات قبل وبعد تطبيق تقنية SMOTE



(ب) (أ)

المصدر: من مخرجات لغة برمجة بايثون

7. أشجار القرار

يعد نموذج أشجار القرار من أبرز خوارزميات التعلم الآلي الاشرافية المستخدمة في تصنيف البيانات واتخاذ القرارات، لما يتميز به من بساطة في التفسير وفعالية في معالجة البيانات المعقدة، في سياق هذه الدراسة تم اعتماد هذا النموذج لتقييم مدى أهلية الأفراد للحصول على قروض مصغرة، مع تحسين أداءه من خلال تطبيق تقنية SMOTE لمعالجة اختلال توازن البيانات بين الفئتين (المقبولين والمرفوضين).

فيما يلي نستعرض مخرجات النموذج وتحليل نتائجه بناء على عدة أدوات تقييم، بما في ذلك المقاييس الإحصائية، مصفوفة الارتباك، منحنى ROC وبنية شجرة القرار ذاتها إلى جانب تحليل أهمية المتغيرات التفسيرية.

الجدول رقم 02: مخرجات نموذج أشجار القرار بعد تطبيق تقنية SMOTE

Metrics/Class	Accuracy	precision	recall	f1-score	support
0	0.73	0.68	0.89	0.77	516
1	0.73	0.84	0.58	0.68	516

المصدر: من مخرجات لغة برمجة بايثون

يمثل الجدول رقم 02 نتائج تقييم أداء نموذج أشجار القرار بعد تطبيق تقنية SMOTE، حيث ساهم هذا التعديل في تحقيق توازن نسبي بين الصنفين (المقبولين والمرفوضين)، مما عزز من قدرة النموذج على التمييز بين الفئات بشكل أكثر فعالية.

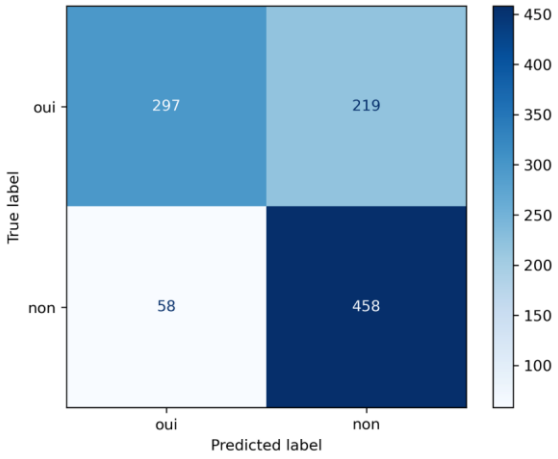
وقد بلغت دقة النموذج (Accuracy) حوالي 73%، وهي نسبة مرتفعة نسبيا تعكس أداءً جيدا في التنبؤ بالحالات الصحيحة.

من حيث الدقة الإيجابية (Precision)، سجل النموذج 68.0 للصنف 0 (المقبولين)، ما يعني أن 68% من الحالات التي تم التنبؤ بقبولها كانت صحيحة فعلاً، مقابل 84.0 للصنف 1 (المرفوضين)، مما يشير إلى موثوقية عالية في رفض منح القرض لغير المستحقين، أما فيما يتعلق بمقياس الاسترجاع (Recall)، فقد بلغ 89.0 للصنف 0، وهو ما يدل على قدرة جيدة للنموذج في التعرف على العملاء المقبولين ، مقابل 58.0 فقط للصنف 1، مما محدودية نسبية في التعرف على غير المقبولين ، إذ يفشل النموذج في اكتشاف حوالي 42% منهم، ما قد يؤدي إلى قرارات منح غير دقيقة، وأخيرا أظهر المعدل التوافقي (F1-Score) توازنًا نسبيا في أداء النموذج بين الصنفين، حيث سجل 77.0 للصنف 0 و68.0 للصنف 1، مع ميل طفيف نحو أداء أفضل في تصنيف العملاء المقبولين.

بشكل عام، يعتبر أداء النموذج جيدا ومقبولًا خاصة بعد تطبيق تقنية SMOTE، من حيث توجيه لقروض للعملاء المناسبين.

8. تحليل مصفوفة الارتباك لنموذج تصنيف القروض:

الشكل رقم 10: مصفوفة الارتباك لنموذج أشجار القرار بعد تطبيق تقنية SMOTE



المصدر: من مخرجات لغة برمجة بايثون

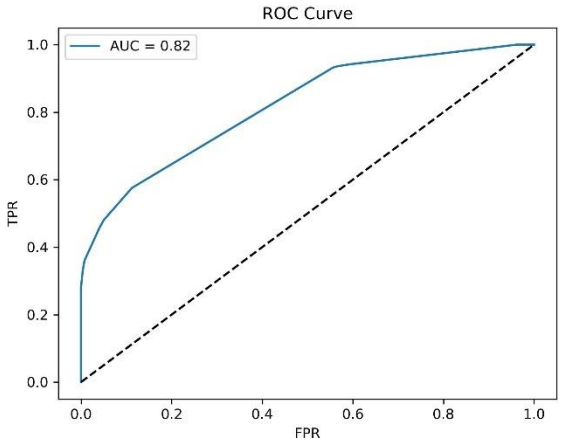
تمثل مصفوفة الارتباك في الشكل رقم 10 أداة محورية لتقييم أداء نموذج التصنيف، إذ تتيح مقارنة دقيقة بين القيم الحقيقية (هل العميل يستحق القرض أم لا) والتوقعات التي قدمها النموذج.

وقد أظهرت النتائج أن النموذج نجح في التنبؤ بشكل صحيح بحالة 297 عميلا ممن يستحقون القرض فعليًا، وهو ما يعرف بالحالات الإيجابية الصحيحة (True Positives)، مما يعكس قدرة النموذج على تحديد الفئة المستهدفة بكفاءة، وفي المقابل أخطأ النموذج في تصنيف 58 عميلا كأشخاص مؤهلين للحصول على القرض رغم أنهم غير مؤهلين في الواقع، وهو ما يعرف بالحالات الايجابية الخاطئة (False Positives)، ويعد هذا النوع من الأخطاء خطيرا من الناحية المالية لأنه يؤدي إلى منح قروض غير مضمونة السداد.

كما رفض النموذج 219 عميلا مؤهلين للحصول على القرض، وهي حالات سلبية خاطئة (False Negatives) تمثل فرصًا ضائعة قد تكون لمشاريع ناجحة، مما يؤثر سلبًا على فعالية القرار التمويلي، في المقابل تمكن النموذج من التنبؤ بدقة بعدم أهلية 458 عميلا، وهم الحالات السلبية الصحيحة (True Negatives)، ما يدل على قدرتها الجيدة في تجنب المخاطر.

وبشكل عام، تشير هذه النتائج إلى أن النموذج استطاع تصنيف نحو 73% من الحالات بشكل صحيح، ما يعكس دقة مقبولة، إلا أن هناك حاجة لتحسين أداء النموذج، لا سيما في تقليل الأخطاء التي تؤدي إلى منح قروض لعملاء غير مؤهلين.

الشكل رقم11: منحنى ROC لتقييم أداء نموذج أشجار القرار

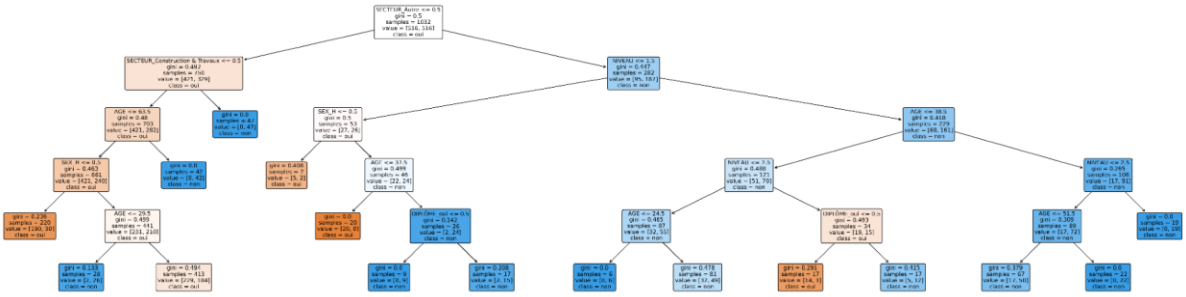


المصدر: من مخرجات لغة برمجة بايثون

الشكل رقم 11 يمثل منحنى (ROC Receiver Operating Characteristic) أداة تقييم مهمة لأداء النموذج في التمييز بين الحالات المقبولة والمرفوضة.

في هذا المنحنى، يشير المحور الأفقي (FPR) إلى معدل الايجابيات الكاذبة، أي نسبة الحالات التي تم رفضها رغم أنها تستحق القبول، بينما يمثل المحور العمودي (TPR) معدل الإيجابيات الحقيقية، أي نسبة الحالات التي تم قبولها بشكل صحيح يظهر المنحنى الأزرق أداء النموذج، بينما يشير الخط الأسود المتقطع إلى الأداء العشوائي بـ AUC = 5.0، في هذا المثال، بلغت قيمة AUC حوالي 82.0، ما يدل على أن النموذج يتمتع بقدرة جيدة جداً على التمييز بين الفئتين. فكلما اقترب المنحنى من الزاوية العلوية اليسرى، زادت كفاءة النموذج، وتشير قيمة AUC = 82.0 إلى أن النموذج قادر على اتخاذ قرارات صحيحة بنسبة 82% فيما يخص التمييز بين الأفراد الذين يستحقون القبول للحصول على قرض مصغر وأولئك الذين لا يستحقونها.

الشكل رقم 12: شجرة القرار

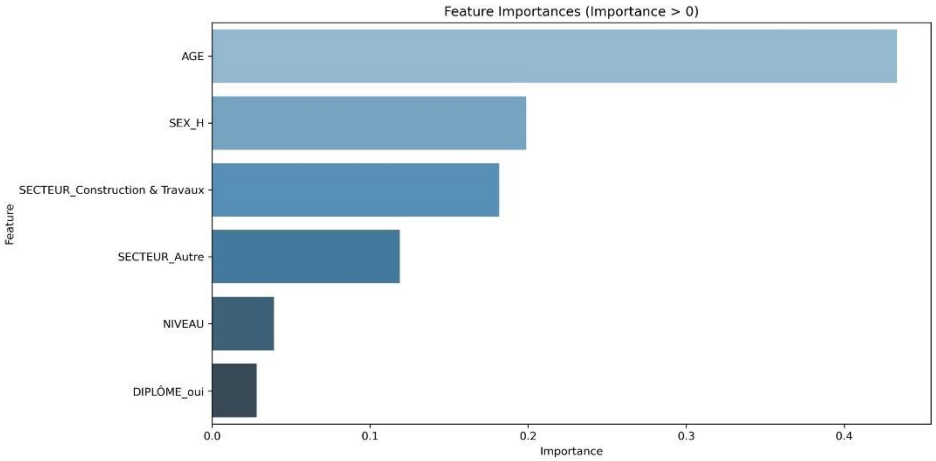


المصدر: من مخرجات لغة برمجة بايثون

الشكل رقم 12 يظهر شجرة القرار حيث، أن النموذج يتبع تسلسلاً منطقيا لاتخاذ قرار منح القرض، بدءًا من العامل الأهم وهو القطاع المهني (SECTEUR)، حيث يعتبر العمل في "Construction & Travaux" مؤشرا ايجابيا على الاستقرار المالي، إذا لم يكن الشخص ينتمي لهذا القطاع، ينتقل النموذج لتحليل العمر، إذ يفضل الأشخاص دون سن 5.63 سنة، مع فرصة أعلى للقبول كلما كان العمر أصغر، خاصة أقل من 30 سنة. في حالة العمل داخل قطاع البناء، يعاد تحليل العمر مرة أخرى، حيث يعتبر العمر أقل من 5.24 سنة مع تعليم منخفض مؤشراً على مخاطرة، مما يؤدي غالباً إلى الرفض. وبالمقابل فإن توفر شهادة وارتفاع العمر بشكل معتدل قد يغير القرار نحو القبول.

التحليل يظهر أن أهم العوامل المؤثرة هي القطاع، ثم العمر (باعتباره مرتبطاً بالمخاطر الائتمانية بشكل كبير)، ثم التعليم بينما للجنس والدخل الإضافي تأثير محدود، النموذج يفضل فئات محددة: الشباب من قطاع البناء وذوي الشهادات لديهم فرص قبول أعلى، في حين أن كبار السن (فوق 60 سنة) أو من خارج هذا القطاع فرصهم اقل.

الشكل رقم 13: أهم المتغيرات المؤثرة في القبول والرفض باستخدام اشجار القرار



المصدر: من مخرجات لغة برمجة بايثون

من الرسم البياني رقم 13 تظهر مخرجات نموذج شجرة القرار، التي تقيم أهمية الميزات في التنبؤ بقرار قبول أو رفض منح القروض المصغرة، أن العمر (AGE) هو المتغير الأكثر تأثيرا بشكل حاسم، حيث بلغت أهميتها حوالي 42.0، مما يشير إلى دوره المركزي في تحديد أهلية المقترضين، يليها مباشرة كل من الجنس (SEX_H) والقطاع الاقتصادي (SECTEUR_Construction & Travaux)، بأهمية تتراوح بين 18.0 و20.0، مما يؤكد أن العوامل الديمغرافية والمهنية تشكل المحركات الرئيسية لقرارات النموذج، بينما تظهر ميزة "قطاع آخر" (SECTEUR_Autre) أهمية متوسطة، فإن ميزتي "المستوى" (NIVEAU) و**"امتلاك شهادة /دبلوم" (DIPLÔME_oui)** تساهمان بقدر ضئيل في التنبؤ، مما يشير إلى تأثيرهما المحدود على احتمالية السداد في سياق هذه البيانات.

هذه النتائج تؤكد ضرورة تركيز المؤسسات المالية على تحليل دقيق للعمر، الجنس، والقطاع المهني للمقترضين عند تقييم طلبات القروض المصغرة، مع الأخذ في الاعتبار أهمية فهم العلاقات الكامنة (مثلاً، كيف يؤثر العمر بالضبط)، وإدارة أي تحديات أخلاقية أو قانونية قد تنشأ عن استخدام متغيرات مثل الجنس في عملية اتخاذ القرار الانحدار اللوجستي.

في هذه الدراسة، تم اعتماد تقنية الاختيار التدريجي نحو الأمام (Forward Stepwise Selection) كإحدى الأدوات الإحصائية الهادفة إلى بناء نموذج تنبؤي فعال ومبسط، تُستخدم هذه التقنية لاختيار المتغيرات الأكثر تأثيرا ضمن مجموعة من المتغيرات التفسيرية المحتملة، خصوصا في سياق نماذج الانحدار.

9. الانحدار اللوجستي

Forward Stepwise Selection وتتمثل آلية عملها في البدء بنموذج فارغ يحتوي فقط على الحد الثابت، ثم يتم اختبار جميع المتغيرات المرشحة واحدة تلو الأخرى، مع اختيار المتغير الذي يحقق أكبر تحسن في أداء النموذج استنادًا إلى معايير إحصائية كقيمة p أو مؤشر AIC، تستمر العملية بشكل تكراري بإعادة التقييم بعد كل إضافة، إلى أن لا يعود هناك متغير جديد يحسن النموذج بشكل ملحوظ، يسهم هذ النهج في تقليل التعقيد البنيوي للنموذج، وتحسين قدرتها التنبؤية، مع الحفاظ على تفسير منطقي لنتائجها.

دالة الانحدار

$$\text{logit}(p) = -6.441 + 1.197 \times \text{NIVEAU} + 0.059 \times \text{AGE} + 2.354 \times \text{SEX_H} + 3.037 \times \text{SECTEUR_Textile\&Mode}$$

المصدر: من مخرجات لغة برمجة بايثون .

تظهر نتائج نموذج الانحدار اللوجستي أن كل من المتغيرات المدروسة (المستوى التعليمي، العمر، الجنس، القطاع الاقتصادي) لها تأثير متفاوت على احتمال منح القرض المصغر، فالمعامل الخاص بمتغير NIVEAU بلغ +197.1، ما يدل على وجود علاقة طردية بين المستوى التعليمي واحتمال الحصول على القرض، حيث أن ارتفاع المستوى الدراسي للمستفيد يعزز من فرص قبولها من طرف جهة التمويل، ويعزى ذلك إلى أن التعليم غالبًا ما يرتبط بالكفاءة والقدرة على تسيير المشروع، أما معامل AGE، فقد بلغت قيمتها +059.0، وهو معامل ضعيف نسبيا ، ما يعني أن التقدم في السن لها تأثير محدود لكنه موجب، إذ كلما تقدم الفرد في العمر، زادت احتمالية حصولها على القرض ولو بشكل طفيف، ويحتمل أن يكون ذلك بسبب الخبرة أو الاستقرار الاجتماعي. فيما يخص متغير SEX_H، والذي يميز بين الجنسين (1 = ذكر، 0 = أنثى)، فقد بلغت قيمته +354.2، وهو ما يشير إلى أن الذكور يتمتعون بفرص أكبر للحصول على التمويل مقارنة بالإناث، وقد يعكس هذا المعامل انحيازا اجتماعيا أو مؤسساتيا، أو يعزى إلى تصورات مسبقة حول قدرة الرجل على تسيير المشاريع الإنتاجية، أما بالنسبة لمتغير SECTEUR_Textile & Mode، فقد سجل أعلى معامل في النموذج، حيث بلغ +037.3، وهو ما يبرز الأثر الإيجابي الكبير للانتماء إلى قطاع النسيج والموضة على قرار منح القرض. ويمكن تفسير ذلك بكون هذا القطاع يحظى بأهمية خاصة في سياسات التنمية المحلية أو ينظر إليها كمجال ناجح وواعد، ما يدفع الجهات الممولة إلى تشجيعها وتمويل مستفيديها.

بصفة عامة، تعكس معاملات النموذج أهمية الخصائص الفردية والقطاعية في عملية اتخاذ القرار التمويلي، وتشير إلى وجود تباينات في فرص الحصول على القروض تبعًا للمستوى التعليمي والجنس والقطاع الاقتصادي، ما يدعو إلى التفكير في سياسات أكثر توازناً وشمولاً.

جدول رقم 03: نموذج الانحدار اللوجستي بعد تطبيق FSS

Metrics/Class	Accuracy	precision	Recall	f1-score	support
0	0.75	0.76	0.73	0.75	516
1	0.75	0.74	0.76	0.75	516

المصدر: من مخرجات لغة برمجة بايثون

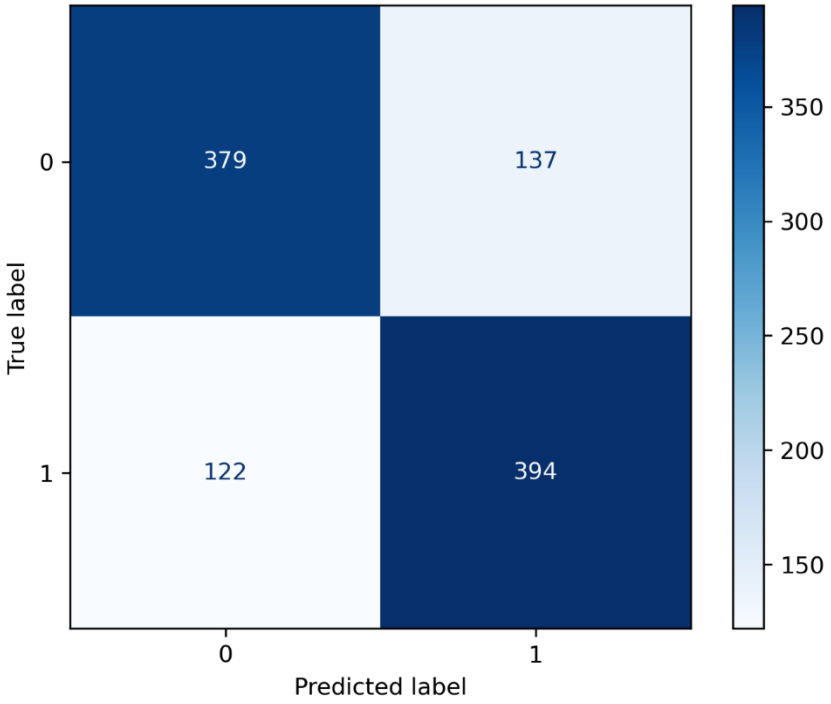
من خلال النتائج الإحصائية المستخرجة من نموذج الانحدار اللوجستي، يتضح أن النموذج يحقق دقة اجمالية (Accuracy) قدرها 75%، ما يدل على كفاءتها المقبولة في التنبؤ بقرار منح القروض، عند النظر إلى الأداء حسب الفئتين المستهدفتين (0 = رفض القرض، 1 = منح القرض)، نلاحظ أن الـ F1-score متساوية تقريبا لكلا الفئتين (75.0)، مما يعكس توازنًا جيدا بين كل من الدقة (Precision) والاسترجاع (Recall).

وبالنسبة للفئة 0 (المرفوضين)، فقد بلغ الاسترجاع 73.0، ما يشير إلى قدرة النموذج على اكتشاف أغلب الحالات غير المؤهلة للحصول على قرض، أما الدقة (76.0)، فتشير إلى أن أغلب الحالات التي تم رفضها فعليًا كانت صحيحة، وهو امر إيجابي من حيث تقليل الأخطاء من النوع الأول (قبول غير مستحقين).

وفي المقابل، بالنسبة للفئة 1 (مقبولين)، يظهر النموذج أداء مشابها، حيث بلغت الدقة 74.0 والاسترجاع 76.0، وهو ما يعكس قدرة النموذج على تحديد المستفيدين الحقيقيين من القرض مع هامش خطأ مقبول.

بشكل عام، تظهر هذه النتائج أن النموذج يتمتع بتوازن جيد في التنبؤ بكلا الفئتين، ما يجعلها أداة فعّالة يمكن الاعتماد عليها في اتخاذ القرار بشأن منح القروض المصغرة، مع الأخذ بعين الاعتبار أن هناك مجالًا لتحسين الأداء باستخدام تقنيات إضافية مثل تحسين اختيار المتغيرات أو المعالجة المتقدمة لعدم توازن البيانات.

الشكل رقم14: مصفوفة الارتباك لنموذج الانحدار اللوجستي



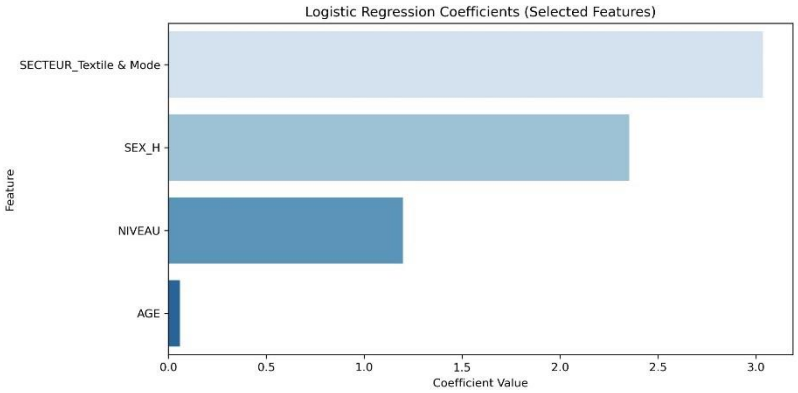
المصدر: من مخرجات لغة برمجة بايثون

مصفوفة الالتباس تعد أداة أساسية لتقييم دقة نموذج التصنيف من خلال مقارنة النتائج الحقيقية بتوقعات النموذج حول قبول أو رفض العملاء لمنح القرض.

أظهرت النتائج أن النموذج تمكن من التعرف بدقة على 379 عميلا مقبولا فعليا، أي حالات إيجابية صحيحة، مما يعكس فعالية جيدة في تصنيف العملاء المقبولين، في المقابل أخطأ النموذج بتصنيف 122 عميلا كقبول رغم أنهم في الواقع مرفوضون، وهو ما يعرف بالحالات الايجابية الخاطئة، وهذا يشكل مخاطرة مالية كبيرة بسبب احتمال عدم سداد القروض، كما رفض النموذج 137 عميلا كانوا فعليا مؤهلين للقبول، أي حالات سلبية خاطئة، مما يمثل فرصا ضائعة قد تؤثر سلبا على نجاح عملية الإقراض، ومن جهة أخرى نجح النموذج في تصنيف 394 عميلا مرفوضين فعليًا بشكل صحيح ، وهي حالات سلبية صحيحة تعكس قدرتها على تقليل المخاطر.

بشكل عام، حقق النموذج تصنيفا صحيحا لحوالي 73% من الحالات، وهو أداء جيد لكنه يحتاج إلى تحسين خاصة في تقليل الأخطاء التي تؤدي إلى منح قروض لعملاء مرفوضين.

الشكل رقم 15: أهم المتغيرات المؤثرة في القبول والرفض باستخدام الانحدار اللوجستي



المصدر: من مخرجات لغة برمجة بايثون

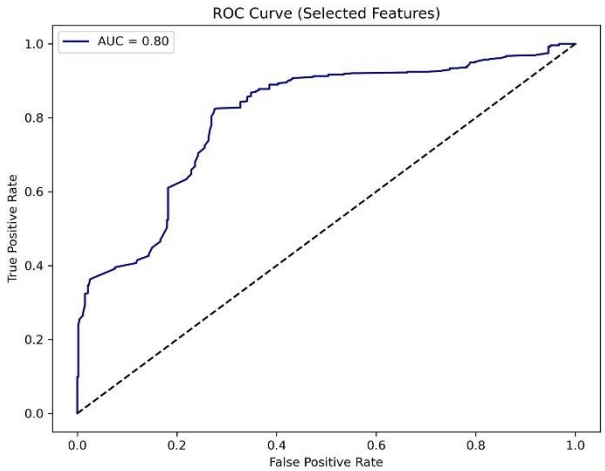
بالاعتماد على نتائج تحليل الانحدار اللوجستي الموضحة في الرسم البياني رقم 15، يمكن تقديم تفسير موسع حول العوامل المؤثرة في قرار القبول أو الرفض، مع توضيح أهمية وتأثير كل متغير على النموذج التنبؤي كما يلي:

بعض العوامل تلعب دورا حاسما في تحديد احتمالية القبول أو الرفض، بدرجات متفاوتة من التأثير، العامل الأكثر تأثيرا هو الانتماء إلى (SECTEUR_Textile & Mode) كالنسيج والخياطة، والذي أظهر أعلى قيمة معامل، مما يعني أن المتقدمين من هذا القطاع يتمتعون بفرص أكبر بكثير للقبول مقارنة بغيرهم، يلي ذلك عامل الجنس (SEX_H)، حيث تشير النتائج إلى أن كون المتقدم ذكرا يزيد بشكل ملحوظ من احتمال القبول، مما يسلط الضوء على وجود انحياز محتمل يجب التعامل معه بحذر.

أما المستوى الدراسي (NIVEAU)، فقد جاء في مرتبة وسطى من حيث التأثير، حيث يظهر أن ارتفاع المستوى الدراسي يرتبط بزيادة معتدلة في احتمالية القبول، لكن هذا التأثير يظل أقل وضوحا مقارنة بالعوامل المهنية أو الاجتماعية، في المقابل يعد العمر (AGE) العامل الأقل تأثيرا في النموذج، حيث أن تأثيره على القرار التنبؤي كان شبه معدوم، وهو ما يشير إلى حيادية هذا المتغير أو إلى أن بقية العوامل تغطي فعاليته ضمن السياق المدروس.

بشكل عام، تبرز نتائج النموذج ضرورة مراجعة مدى عدالة وحيادية استخدام بعض المتغيرات (خصوصًا الجنس والقطاع المهني) عند اتخاذ قرارات تؤثر على الأفراد، مع توصية بإعادة تقييم أهمية الكفاءة التعليمية والعمر، وذلك لضمان نموذج أكثر توازنا وموضوعية.

الشكل رقم 16: منحنى Roc لتقييم أداء الانحدار اللوجستي



المصدر: من مخرجات لغة برمجة بايثون

منحنى ROC يعكس قدرة النموذج على التمييز بفعالية بين العملاء المقبولين والمرفوضين، في هذا المنحنى، يمثل الخط الأزرق أداء النموذج الفعلي، في حين يشير الخط الأسود المتقطع إلى أداء عشوائي يفتقر إلى أي قدرة تمييز.

أظهرت النتائج أن قيمة مساحة تحت المنحنى (AUC) وصلت إلى حوالي 80.0، ما يعكس أداء جيدا جدًا للنموذج، هذه النتيجة تعني أن النموذج قادر على اتخاذ قرارات صحيحة في حوالي 80% من الحالات عند التمييز بين العملاء الذين يستحقون القرض وأولئك الذين لا يستحقونها، وهو مؤشر واضح على فعالية النموذج في دعم عملية اتخاذ القرار.

10. المقارنة

جدول04: مقارنة بين أداء نموذج أشجار القرار ونموذج الانحدار اللوجستي

Decision Tree	Logistic Regression	Metric	Class
0.77	0.75	F1-score	0
0.89	0.73	Recall	
0.68	**0.76**	Precision	
0.73	**0.75**	Accuracy	
0.68	**0.75**	F1-score	1
0.58	**0.76**	Recall	
0.84	0.74	Precision	
0.73	**0.75**	Accuracy	

المصدر: من مخرجات لغة برمجة بايثون

بالاعتماد على الجدول رقم تمت مقارنة أداء نموذج الانحدار اللوجستي ونموذج شجرة القرار في تصنيف طلبات القروض إلى فئتين: مقبولين ومرفوضين.

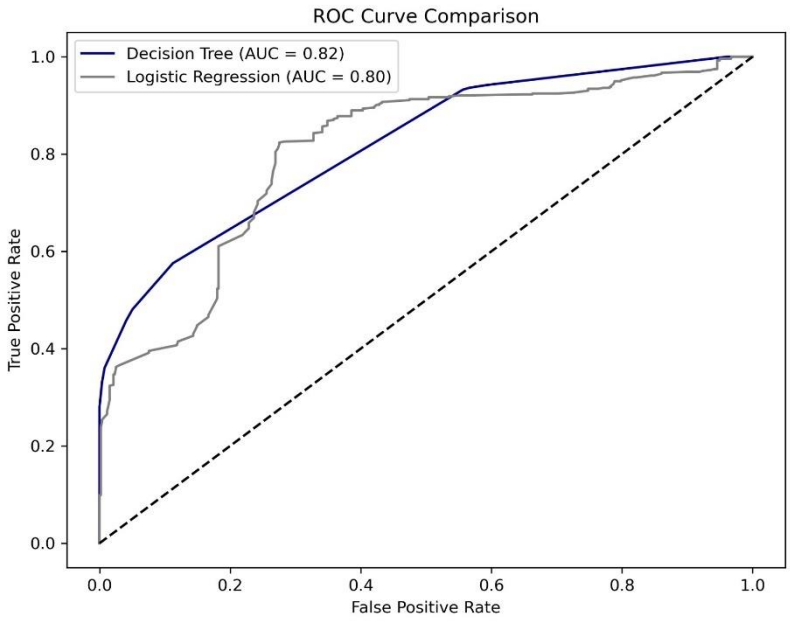
أظهرت النتائج أن شجرة القرار حققت أداء أفضل في تصنيف المقبولين، حيث بلغ معدل الاسترجاع 89% (Recall)، ما يشير إلى قدرتها العالية على اكتشاف معظم الأفراد المستحقين للقرض، حتى وإن كان ذلك على حساب بعض الدقة (Precision) التي سجلت 68%، بالمقابل تفوق الانحدار اللوجستي في تصنيف المرفوضين، حيث سجل معدل استرجاع 76%، وهو أعلى من شجرة القرار، مما يعني أنها أكثر كفاءة في اكتشاف الأفراد غير المؤهلين.

بناءً على ذلك، يمكن القول إن شجرة القرار مناسبة أكثر لتقليل حالات الرفض الخاطئ لأشخاص مستحقين للقرض، بينما يفضل استخدام الانحدار اللوجستي لضمان دقة رفض الملفات غير المؤهلة، ويوصي بدمج

النموذجين ضمن آلية هجينة تجمع بين نقاط القوة في كل منهما، أو اعتماد مراجعة بشرية في الحالات التي يختلف فيها النموذجان، من أجل تعزيز موثوقية القرار.

1. **تقييم أداء الانحدار اللوجستي وأشجار القرار**

الشكل رقم 17: منحنى ROC



المصدر: من مخرجات لغة برمجة بايثون

من الشكل رقم 17 يتضح أن نموذج شجرة القرار يقدم أداء أفضل بشكل ملحوظ مما يجعلها الخيار المفضل لتطبيقنا، فعلى الرغم من أن كلا النموذجين يظهران قدرة تمييزية جيدة جدًا بشكل عام (AUC=0.82 لشجرة القرار، وللانحدار اللوجستي AUC = 80.0)، إلا أن التفوق الطفيف لشجرة القرار في قيمة AUC يشير إلى كفاءة أعلى في التمييز بين المقترضين المقبولين والمرفوضين، والأهم من ذلك عند تحليل شكل المنحنيات، تبرز شجرة القرار بوضوح في المناطق ذات معدل الايجابيات الكاذبة (False Positive Rate) المنخفض جدًا، حيث ترتفع بسرعة أكبر لتحقق معدل إيجابيات حقيقية (True Positive Rate) أعلى، هذا يعني أن شجرة القرار قادرة على تحديد نسبة أكبر من المقترضين المستحقين فعليًا مع تقليل المخاطر والخسائر الناتجة عن منح قروض لمن لن يسددوا، وهو أمر بالغ الأهمية في مجال الإقراض للمشاريع المصغرة، إضافة إلى ذلك تتميز أشجار القرار

بقابليتها للتفسير عبر رسم شجرة القرار (Visualization)، مما يسهل فمه آليات اتخاذ القرار، وهو ما يضيف إلى فعاليتها كأداة للتنبؤ والمساعدة في اتخاذ قرارات منح القروض.

الفصل الرابع: مناقشة النتائج والتوصيات

مناقشة نتائج الدراسة

بعد الانتهاء من التحليل التطبيقي ومعالجة البيانات باستخدام تقنيات تعلم الآلة، نسعى في هذا المحور إلى مناقشة النتائج المتوصل إليها في ضوء الفرضيات التي انطلقت منها الدراسة، وذلك بالاعتماد على المؤشرات الإحصائية المحصلة من النماذج وربطها بالسياق النظري والدراسات السابقة ذات الصلة.

الفرضية الرئيسية: "تتمتع أشجار القرار بفعالية تنبؤية أكثر ومؤشرات أداء أفضل، مما يجعلها الأنسب في معالجة مشكلة التنبؤ بمنح القروض".

تشير النتائج التطبيقية بوضوح إلى إثبات هذه الفرضية، حيث حقق نموذج أشجار القرار دقة تنبؤية بلغت 73% مقارنة بـ 70% فقط لنموذج الانحدار اللوجستي، كما تفوق بوضوح من حيث معامل الاسترجاع (Recall) لفئة القبول، والذي بلغ 0,89 وهو مؤشر جوهري في سياق منح القروض، مقابل 0,73 فقط لنموذج الانحدار، إضافة إلى ذلك سجل F1-Score أعلى ما يدل على توازن جيد في الدقة والاسترجاع.

كما تميز النموذج ببنية تفسيرية واضحة تسهل فهم نتائج التصنيف من طرف غير المختصين، وهو ما يعزز صلاحيته كأداة عملية لاتخاذ القرار داخل مؤسسات التمويل المصغر.

الفرضية الفرعية الأولى: "توجد فروق جوهرية بين النموذجين من حيث الطبيعة الرياضية، متطلبات البيانات، وطبيعة كل منهما في التعامل مع العلاقات المعقدة بين المتغيرات".

تم اثبات هذه الفرضية استنادا إلى الخصائص البنيوية لكلا النموذجين، فالانحدار اللوجستي يفترض وجود علاقة خطية بين المتغيرات، ويشترط تحقق مجموعة من الفرضيات الإحصائية، مما يحد من مرونته في التعامل مع العلاقات غير الخطية أو البيانات المعقدة، أما نموذج أشجار القرار فيعتمد على مبدأ التقسيم التكراري للبيانات، ولا يتطلب افتراضات توزيع مسبقة، كما يسمح بالكشف عن تفاعلات متداخلة بين المتغيرات بطريقة مرئية وشاملة، ما يجعه أكثر تكيفا مع طبيعة بيانات القروض المصغرة.

الفرضية الفرعية الثانية: "يظهر نموذج أشجار القرار تفوقا من حيث مؤشرات الأداء الاحصائي مثل الدقة، الاستدعاء و F1-Score عند استخدامه على بيانات القروض".

أكدت النتائج لتطبيقية على صحة هده الفرضية بشكل واضح، حيث أظهر نموذج أشجار القرار أداء أفضل في عظم مؤشرات التقييم مقارنة بنموذج الانحدار اللوجستي، فقد حقق النموذج اعلى قيمة لمعامل الاسترجاع للفئة الإيجابية، مما يعكس قدرة أكبر على كشف الحالات المقبولة فعليا، وهي الفئة ذات الأهمية الأكبر في القرار الائتماني، كما سجل قيما معتبرة لكل من الدقة وF1-Score، مما يدل على موثوقية النموذج في التنبؤ بالنتائج الصحيحة في ظل بيانات غير متوازنة، خصوصا بعد استخدام تقنية SMOTE.

الفرضية الفرعية الثالثة: "يتأثر اختيار النموذج الأفضل في المؤسسات المالية بعوامل عديدة مثل توفر البيانات، بساطة التطبيق، قابلية التغيير ومستوى التحكم في فرط الملائمة (Overfitting)".

أثبتت النتائج أيضا صحة هذه الفرضية، فقد أظهر نموذج أشجار القرار قابلية أكبر للتفسير، وسهولة في الاستخدام من طرف الفاعلين داخل المؤسسة، خاصة في البيئات التي تتطلب خبرة تقنية عالية، إضافة إلى ذلك يوفر النموذج مرونة في التعديل والتحديث حسب المتغيرات الجديدة دون الحاجة إلى إعادة بناء النموذج كليا، وهو ماي جعله مناسبا في السياقات الديناميكية، كما يمكن التحكم في مشكلة فرط التخصيص (Overfitting) من خلال إجراءات تقليم الشجرة وضبط المعلمات، ما يعزز من استقرارية النموذج.

1. استنتاجات أساسية

انطلاقًا من الدراسة النظرية والتجريبية، يمكن تلخيص أهم ما توصّلنا إليه فيما يلي:

- أشجار القرار تمثل خيارا تنبؤيا أكثر كفاءة ومرونة في بيئة بيانات غير متوازنة مثل حالة وكالة سطيف.

إعادة التوازن باستخدام SMOTE حسّنت أداء النموذجين معًا، إلا أن أشجار القرار استفادت بشكل أكبر من هذه المعالجة.

- الانحدار اللوجيستي يظل مفيدًا حين تكون الحاجة للتفسير الشفاف ضرورية، لكن أداؤه محدود أمام أنماط أكثر تعقيدًا.

2. التوصيات و آفاق الدراسة

بناءً على النتائج المحصّلة، نقترح التوصيات والآفاق التالية:

- اعتماد أشجار القرار كنموذج أساسي داخل وكالات التمويل المصغر، خاصة في مرحلة ما قبل المعالجة الأولية للملفات.

- استكشاف نماذج أكثر تطورًا مثل الغابات العشوائية (Random Forest) وXGBoost، التي قد توفر أداء أعلى دون فقدان قابلية التفسير.
- دمج أدوات التفسير الحديثة مثل SHAP أو LIME لفهم أثر كل متغير على قرار القبول أو الرفض، وتعزيز شفافية النموذج.
- إعادة تطبيق النموذج على وكالات أخرى في مناطق مختلفة لاختبار مدى ثبات النتائج في سياقات جغرافية واجتماعية متعددة.
- توسيع الدراسة لتشمل متغيرات اجتماعية ونفسية إضافية (مثل التاريخ المالي، الدعم الأسري، مستوى المبادرة الفردية)، لتحسين فهم السلوك الائتماني للفئات المستهدفة.
- تحسين الاسقاط العملي لنماذج التعلم الالي من خلال استخدام تقنيات أخرى لاستخراج المتغيرات الأكثر أهمية (feature selection Techniqures) مثل تقنية SHAP، لمعرفة المتغيرات التي تؤثر واقعيا على قرارات الإقراض وليس فقط تقنيا.
- استخدام نماذج أكثر تقدما للتعامل مع مشكلة فرط الملائمة (Overfitting) مثل: الغابات العشوائية والشبكات العصبية العشوائية.

ختاما تبرز هذه الدراسة أهمية التقاء الجانب الأكاديمي التطبيقي مع الواقع العملي لمؤسسات التمويل المصغر، حيث لا يكفي أن تكون النماذج دقيقة من الناحية الإحصائية، بل يجب أن تكون قابلة للتطبيق، عادلة في قراراتها، وشفافة في منطقها. ومن خلال التمحيص المقارن، تبين أن أشجار القرار، في ظل بيئة بيانات غير متوازنة، تقدّم أداة فعالة لدعم اتخاذ القرار في منح القروض المصغرة، مما يجعلها مرشحة قوية للاعتماد المؤسسي مستقبلا.

ونأمل أن تُسهم هذه المذكرة في إثراء الأدبيات البحثية في مجال التمويل المصغّر وتطبيقات الذكاء الاصطناعي، وتكون منطلقًا لأبحاث لاحقة تتناول هذا المجال الحيوي من زوايا أكثر عمقًا واتساعًا.

قائمة المراجع والمصادر

قائمة المراجع:

المراجع باللغة العربية:

- **الكتب:**

1- محمد نجيب عبد الفتاح، أحمد عبد المنعم، التحليل المعمق للبيانات باستخدام حزمة البرامج الجاهزة spss، جامعة الدول العربية، 2009.

2- محمد الفاتح المغربي، الأساليب الكمية في إدارة الأعمال، دار الجنان، الطبعة الأولى، عمان، 2017.

3- جولي بالانت، التحليل الاحصائي باستخدام برنامج SPSS، دار الفاروق للاستثمارات الثقافية، ط1، الجيزة، مصر، 2015.

- **المذكرات والرسائل الجامعية:**

1 - إبراهيم عدلي، استشراف مستقبل الصناعة المعرفية، مداخلة ضمن يوم دراسي، كلية العلوم الاقتصادية جامعة العربي بن مهيدي أم البواقي، 21 أفريل 2015.

2- رضوى عبد الواحد عبد الحليم سويف، نموذج إحصائي للتنبؤ باحتياجات محدودي الدخل من الدعم السلعي في مصر، أطروحة دكتوراه، قسم الإحصاء والتأمين، كلية التجارة، جامعة عين الشمس، القاهرة، مصر، 2017.

3- عائشة ين مريع يحيى عسيري، دراسة مقارنة بين الانحدار اللوجستي والتحليل التمييزي في القدرة التنبؤية في ضوء أحجام عينات مختلفة، أطروحة دكتوراه، تخصص الإحصاء والبحوث، جامعة أم القرى، مكة، المملكة العربية السعودية، 2016.

4 - عبد الحميد محمد العباسي، الانحدار اللوجستي تطبيقات في العلوم الاجتماعية باستخدام SPSS، بحث مقدم لمعهد الدراسات والبحوث الإحصائية، قسم الإحصاء الحيوي والسكاني، جامعة القاهرة، مصر، 2011.

5 - بن زخروفة بوعلام، عطية العربي، دور شجرة القرار في تحسين القرارات التسويقية دراسة حالة شركة الأصنامية الخاصة بتصنيع وتسويق المشروبات الغازية، جامعة قتصدي مرابح، ورقلة، 2015.

- **المجلات والموسوعات:**

1- سهيلة حمود عبد الله الفرهود، استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم، مجلة الأزهر، 68-47، كلية الدراسات التجارية، قسم الاحصاء، الكويت، 2014.

2- كرار طعمه عبد عطية، الانحدار الخطي والانحدار اللوجستي، بحث مقدم إلى مجلة كلية التربية للعلوم الصرفة، جامعة بابل، العراق، 2024.

3- محمد أمين دعيش، محمد ساري، نموذج الانحدار اللوجستي، مجلة السراج في التربية وقضايا المجتمع، العدد الأول، 2017.

4 - سهيلة حمود عبد الله الفرهود، استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم (دراسة تطبيقية على سوق الكويت للأوراق المالية)، مجلة جامعة الأزهر (العلوم الطبيعية)، العدد 16، غزة، فلسطين، 2014.

5- ساكت غسان، عداس ضحى، استخدام أشجار القرار لدعم قرارات منح القروض، مجلة بحوث جامعة حلب، سلسلة العلوم الاقتصادية، العدد 14، حلب، 2015.

6- يمان بلال، سعد الدين العبد الله، غسان ساكت، محمد طاهر عنان، التصنيف باستخدام تحليل الانحدار اللوجستي وأشجار القرار دراسة مقارنة، مجلة بحوث جامعة حلب، سلسلة العلوم الأساسية، العدد 160، 2022.

المراجع باللغة الأجنبية:

1- Martial Schtickzelle, Pierre-François Verhulst (1804-1849) La première découverte de la fonction logistique, Population, Revue bimestrielle de l'institut nationale d'étude démographique, 36e année.
2- Jurafsky, D. & Martin, J.H. (2025) Speech and language processing, In Speech and language Processing.
3- Anass Bayaga, 2010, multinomial logistic regression: usage and appmication in risk analysis, journal of applied quantitative methods, vol 5.

4- Ethem Alpaydin, Introduction to Machine Learning, third edition, London, 2014.

5 -Mostafa Mahmoud, Najla Algadi, Ahmed Ali, Expert System for Banking Credit Decision, 2008.

6 -QUINLAN J.R. 1986 – Induction of Decision Trees. Kluwer Academic Publisher, Boston.

7 -ALKA G., RAVINDRA P., - Building Privacy-Preserving C4.5 Decision Tree Classifier on Multiparties, International Journal on Computer Science and Engineering, 2009.

8 -VILLE B., - Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner, SAS Institute, USA, 2006.

قـائـمـة الـمـلاحق

الملحق 01: أشجار القرار الجزء 1

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import OrdinalEncoder, StandardScaler, Normalizer
from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier, plot_tree
import seaborn as sns

# Load data
df = pd.read_excel(r"C:\Users\m\Downloads\memoire.xlsx")
# Ordinal encode 'NIVEAU'
import seaborn as sns
import matplotlib.pyplot as plt

# Count the values in df['D']
counts = df['DIPLÔME'].value_counts()

# Use seaborn's Blues palette
colors = sns.color_palette("Blues", n_colors=len(counts))

# Plot
plt.pie(counts, labels=counts.index, colors=colors, autopct='%1.1f%%')
plt.title('')
plt.savefig(r"C:\Users\m\Downloads\piee.png", dpi=300, bbox_inches='tight')

plt.show()

niveau_order = [['SANS', 'PRIMAIRE', 'Moyen', 'SECONDAIRE', 'UNIVERSITAIRE']]
ordinal_encoder = OrdinalEncoder(categories=niveau_order)
df['NIVEAU'] = ordinal_encoder.fit_transform(df[['NIVEAU']]).astype(int)

# One-hot encode other categorical variables
df = pd.get_dummies(df, columns=['SEX', 'SECTEUR', 'DIPLÔME'], drop_first=True, dtype=int)

# Separate features and target (target 'D' kept as strings 'oui'/'non')
X = df.drop(columns=['D'])
y = df['D']  # Make sure this column contains strings 'oui' and 'non'
```

الملحق 02: أشجار القرار الجزء 2

```
# Apply SMOTE with fixed seed
sm = SMOTE(random_state=7)
X_res, y_res = sm.fit_resample(X, y)

# Plot class distribution after SMOTE with your colors
plt.figure(figsize=(6, 4))
palette = {'oui': 'darkblue', 'non': 'skyblue'}
sns.countplot(x=y_res, palette=palette)
plt.xlabel("D")
plt.tight_layout()
plt.savefig(r"C:\Users\m\Downloads\smote_distribution.png", dpi=300, bbox_inches='tight')
plt.show()

# Train DecisionTreeClassifier with fixed seed
clf = DecisionTreeClassifier(
    max_depth=5,
    min_samples_leaf=3,
    max_features='sqrt',
    random_state=7
)
clf.fit(X_res, y_res)

# Predict on training data
y_pred = clf.predict(X_res)

# Classification report
print("Classification Report:\n", classification_report(y_res, y_pred))

# Confusion matrix
cm = confusion_matrix(y_res, y_pred, labels=['oui', 'non'])
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=['oui', 'non'])
disp.plot(cmap=plt.cm.Blues)
plt.title("Confusion Matrix")
plt.savefig(r"C:\Users\m\Downloads\confusion_matrix.png", dpi=300, bbox_inches='tight')
plt.show()
```

الملحق 03: أشجار القرار الجزء 3

```
# Plot decision tree
plt.figure(figsize=(40, 10))
plot_tree(
    clf,
    feature_names=X.columns,
    class_names=['oui', 'non'],
    filled=True,
    rounded=True,
    fontsize=10
)
plt.savefig(r"C:\Users\m\Downloads\decision_tree.png", dpi=300, bbox_inches='tight')
plt.title("Decision Tree")
plt.savefig(r"C:\Users\m\Downloads\decision_tree.png", dpi=300, bbox_inches='tight')
plt.show()
# Plot feature importances
importances = clf.feature_importances_
feature_names = X.columns

# Plot feature importances > 0
importances = clf.feature_importances_
feature_names = X.columns

# Create and filter DataFrame
feature_importance_df = pd.DataFrame({
    'Feature': feature_names,
    'Importance': importances
})
feature_importance_df = feature_importance_df[feature_importance_df['Importance'] > 0]
feature_importance_df = feature_importance_df.sort_values(by='Importance', ascending=False)

# Plot the bar chart
plt.figure(figsize=(12, 6))
sns.barplot(data=feature_importance_df, x='Importance', y='Feature', palette='Blues_d')
plt.title("Feature Importances (Importance > 0)")
plt.xlabel("Importance")
plt.ylabel("Feature")
plt.tight_layout()
plt.savefig(r"C:\Users\m\Downloads\feature_importance.png", dpi=300, bbox_inches='tight')
plt.show()
from sklearn.metrics import roc_curve, auc
```

الملحق 04: أشجار القرار الجزء 4

```
plt.plot(fpr, tpr, label=f'AUC = {roc_auc:.2f}')
plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC Curve')
plt.legend()
plt.savefig(r"C:\Users\m\Downloads\roc_curve.png", dpi=300)
plt.show()
```

الملحق 05: الانحدار اللوجستي الجزء 1

```python
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay, roc_curve, auc
from imblearn.over_sampling import SMOTE
from mlxtend.feature_selection import SequentialFeatureSelector as SFS

# 7. Define Logistic regression model
logreg = LogisticRegression(max_iter=1000, random_state=7, solver='lbfgs')

# 8. Sequential forward feature selection with logistic regression on SMOTE data
sfs = SFS(
    logreg,
    k_features=4,
    forward=True,
    scoring='accuracy',
    cv=5,
    n_jobs=-1,
    verbose=1
)
sfs = sfs.fit(X_res, y_res)

selected_features = list(sfs.k_feature_names_)
print("\nSelected Features:", selected_features)

# 9. Train logistic regression on SMOTE data with selected features only
clf_selected = LogisticRegression(max_iter=1000, random_state=7, solver='lbfgs')
clf_selected.fit(X_res[selected_features], y_res)

# 10. Predict and evaluate on training data (SMOTE)
y_pred = clf_selected.predict(X_res[selected_features])
print("\nClassification Report (Selected Features):")
print(classification_report(y_res, y_pred))

# 11. Confusion matrix
cm = confusion_matrix(y_res, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=[0, 1])
disp.plot(cmap=plt.cm.Blues)
plt.title("Confusion Matrix (Selected Features)")
plt.savefig(r"C:\Users\m\Downloads\confusion_matrix_selected.png", dpi=300, bbox_inches='tight')
plt.show()
```

الملحق 06: الانحدار اللوجستي الجزء 2

```
# 12. ROC curve and AUC
y_score = clf_selected.predict_proba(X_res[selected_features])[:, 1]
fpr, tpr, _ = roc_curve(y_res, y_score)
roc_auc = auc(fpr, tpr)

plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label=f'AUC = {roc_auc:.2f}', color='darkblue')
plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve (Selected Features)')
plt.legend()
plt.savefig(r"C:\Users\m\Downloads\roc_curve_selected.png", dpi=300)
plt.show()

# 13. Plot logistic regression coefficients for selected features
coef = clf_selected.coef_[0]
coef_df = pd.DataFrame({'Feature': selected_features, 'Coefficient': coef})
coef_df = coef_df.sort_values(by='Coefficient', ascending=False)

plt.figure(figsize=(10, 5))
sns.barplot(data=coef_df, x='Coefficient', y='Feature', palette='Blues')
plt.title("Logistic Regression Coefficients (Selected Features)")
plt.xlabel("Coefficient Value")
plt.ylabel("Feature")
plt.tight_layout()
plt.savefig(r"C:\Users\m\Downloads\selected_feature_importance.png", dpi=300, bbox_inches='tight')
plt.show()

# 14. Print Logistic regression equation (logit form)
intercept = clf_selected.intercept_[0]
terms = [f"{coef[i]:+.3f}x{selected_features[i]}" for i in range(len(selected_features))]
logit_eq = f"logit(p) = {intercept:+.3f} " + " ".join(terms)

print("\nLogistic Regression Equation (logit form):")
print(logit_eq)
```

الملخص

تسعى هذه الدراسة إلى استكشاف النماذج التنبؤية الحديثة في تحسين عملية اتخاذ القرار الائتماني في مجال القروض المصغرة، وذلك من خلال مقارنة تطبيقية بين نموذج الانحدار اللوجستي ونموذج أشجار القرار.

تم اعتماد بيانات واقعية مستمدة من الوكالة الوطنية لتسيير القرض المصغر« ANGEM » بسطيف، مع التركيز على معالجة مشكلة اختلال توازن البيانات باستعمال تقنية SMOTE، وقد أظهرت النتائج تباينا في الأداء بين النموذجين، حيث امتاز الانحدار اللوجستي بسهولة التفسير والوضوح الاحصائي، بينما برزت أشجار القرار بقدرتها على التعامل مع العلاقات غير الخطية والبيانات غير المتوازنة، خصوصا بعد تطبيق تقنيات المعالجة المسبقة.

توصلت الدراسة إلى أهمية التكامل بين الدقة الكمية والقدرة التفسيرية في النماذج التنبؤية، وأوصت بتبني مقاربات قائمة على الذكاء الاصطناعي لدعم قرارات التمويل، بما يعزز من عدالة وفعالية منح القروض المصغرة خاصة في السياقات المحلية ذات الطابع الاجتماعي والاقتصادي المتباين.

Summary

This study examines the effectiveness of modern predictive modeling techniques in optimizing credit decision processes within the realm of microcredit.

A comparative analysis is performed, contrasting logistic regression and decision tree models, grounded in empirical data sourced from the National Agency for the Management of Microcredit in Sétif. Particular emphasis is placed on addressing the challenges posed by data imbalance through the implementation of the SMOTE methodology,The findings reveal nuanced performance profiles for each model: logistic regression is notable for its interpretability and statistical rigor, whereas decision trees demonstrate marked proficiency in accommodating non-linear relationships and mitigating the effects of imbalanced datasets, particularly following preprocessing interventions.

The research emphasizes the imperative of synergizing quantitative accuracy with interpretive insight and advocates for the integration of AI-supported strategies to enhance financial decision-making, thereby fostering fairness and efficiency in microcredit resource allocation, especially within heterogeneous socio-economic contexts.