

DOCUMENT DE TRANSPARENCE MÉTHODOLOGIQUE

## Insourcia Default Score (IDS) : méthode, données, performance

---

Probabilité d'ouverture d'une procédure collective<sup>1</sup> (sauvegarde, redressement ou liquidation judiciaire) à 12 mois, pour les entreprises françaises.

Modèle : Insourcia Default Score (IDS) - version 1.0, 2026 · mesure exposée : probabilité de défaut à 12 mois (PD)<sup>2</sup>, traduite en note AAA → D.

---

Ce document décrit comment le score est construit, sur quelles données, et comment il est validé en conditions réelles. Mesuré *out-of-time*<sup>3</sup> sur l'année 2025, il atteint une AUC<sup>4</sup> de 0,864 (intervalle de confiance à 95 % : 0,861–0,866) : dans les 10 % d'entreprises les plus mal notées, il capture 57 % des défaillances de l'année à venir. Un score de risque n'est exploitable que s'il est mesuré rigoureusement et explicable. L'ensemble de la méthode décrite ici est reproductible sur données publiques.

---

---

### 1. Le machine learning au service du risque de crédit

Le risque de crédit est un terrain naturel pour l'apprentissage automatique : de grands volumes de données hétérogènes, des relations non linéaires, et un objectif clair - estimer une probabilité de défaut (PD). Les modèles d'arbres de décision boostés (*gradient boosting*<sup>5</sup>) y sont aujourd'hui l'état de l'art sur données tabulaires : ils capturent des interactions subtiles que les modèles linéaires manquent, sans imposer d'hypothèses fortes sur la forme des relations.

Mais la performance d'un modèle dépend d'abord de sa méthodologie. Deux erreurs sont fréquentes : la fuite d'information du futur (utiliser, pour prédire le passé, une donnée indisponible à l'époque) et l'évaluation sur des données trop proches de l'apprentissage. Les deux produisent des scores élevés mais faux en production. Ce document explique comment nous les évitons.

---

## 2. Contexte et données

Notre objet : prédire l'ouverture d'une procédure collective pour une société française. Le tissu d'entreprises français est à la fois vaste et très hétérogène, allant de la TPE locale au grand groupe. De plus, une part importante des PME publie des comptes espacés et parfois parcellaires : un bon modèle doit rester robuste à cette rareté d'information.

Cible	Ouverture d'une procédure collective (sauvegarde, redressement ou liquidation), première occurrence (« avis initial »).
Horizon	12 mois à compter de la date d'observation.
Périmètre	Entreprises françaises disposant de comptes déposés exploitables à la date d'observation.
Couverture	867 072 sociétés notées (re-scorées au 30 mai 2026 sur données fraîches), extensible à l'ensemble des déposants ; les entités hors périmètre reçoivent le statut NR <sup>6</sup> (non notées) - jamais une note fabriquée par défaut.
Prévalence	≈ 2,1 % de défaillances sur 12 mois dans la population éligible (test 2025).

Le score repose exclusivement sur des sources officielles et publiques, recoupées et historisées : registre national des entreprises et greffes, comptes annuels déposés (plan comptable général), annonces légales (BODACC), répertoire SIRENE/INSEE, brevets, marchés publics, subventions, et indicateurs sectoriels et géographiques. Sa valeur vient du recoupement et de l'historisation *point-in-time*<sup>7</sup> de ces signaux : la combinaison des données financières et des signaux d'événements et de comportement est ce qui distingue un bon score d'un score purement comptable. Aucune donnée nominative confidentielle n'entre dans le score exposé.

---

## 3. Ingénierie et sélection des variables

Les données brutes sont pré-traitées en ratios financiers exprimant chaque dimension de risque (structure de capital, rentabilité, liquidité, levier, efficacité), complétés par des signaux d'événements et un contexte systémique (risque sectoriel et géographique). Nous privilégions des variables à bonne couverture sur l'ensemble du périmètre : une construction parcimonieuse maximise le nombre d'entreprises couvertes. Le modèle s'appuie sur 74 variables, dont le tableau 1 donne une vue par dimension de risque.

**Tableau 1 - Principales variables, par type et dimension de risque**

TYPE DE RISQUE	VARIABLES (EXEMPLES)	DIMENSION
Risque financier	Autonomie financière (capitaux propres / actif)	Structure de capital
Risque financier	Capacité de remboursement (dettes / CAF)	Service de la dette
Risque financier	Rentabilité de l'actif, marges (EBE, exploitation, nette)	Rentabilité
Risque financier	Trésorerie rapportée au bilan	Liquidité
Risque financier	Endettement net / capitaux propres, capitaux propres négatifs	Levier
Risque financier	BFR / CA, délai de paiement clients, années de pertes	Efficiency et dynamique
Risque financier	Délai de dépôt des comptes (retard de publication des bilans)	Discipline déclarative
Risque d'activité	Chiffre d'affaires, total bilan, effectifs	Taille
Risque d'activité	Ancienneté, forme juridique, secteur (NAF)	Structure
Signaux d'événements	Antécédent de procédure, conciliation récente, cessions de fonds	Détresse précoce
Signaux d'événements	Changements de dirigeants, dirigeants liés à des procédures passées	Gouvernance et réseau
Risque systémique	Taux de défaillance sectoriel x territorial (décalés d'un an)	Contexte macro
Activité et solidité	Brevets, marchés publics remportés, subventions reçues	Vitalité

Source : Insourcia, à partir de registres publics. À titre illustratif.

#### 4. Pas de fuite d'information du futur (point-in-time)

Chaque variable est calculée telle qu'elle était connue à la date d'observation T : un bilan n'est utilisé qu'à partir de sa date de dépôt réelle (et non de clôture) ; les événements légaux ne comptent qu'une fois réellement publiés (délai d'indexation) ; les indicateurs sectoriels sont décalés d'un an.

Un audit anti-fuite dédié a identifié et corrigé trois biais qui surévaluaient la performance dans nos versions initiales :

- un partage apprentissage/test qui laissait le modèle mémoriser les entreprises (-3,5 pts d'AUC une fois corrigé) ;
- une imputation trop optimiste des dates de bilan (-0,35 pt) ;
- un chevauchement des fenêtres d'observation entre apprentissage et test (-4,5 pts).

Le chiffre publié plus loin est celui obtenu après correction de ces trois biais. C'est cette rigueur qui garantit la fiabilité du score en conditions de production réelles.

---

## 5. Protocole de mesure : out-of-time

La validation *out-of-time* constitue l'unique standard de mesure robuste : l'entraînement s'effectue sur une période donnée, et l'évaluation sur une période postérieure et distincte - exactement comme en production réelle.

- Apprentissage : 1<sup>er</sup> semestre 2024, dont sont exclues toutes les sociétés qui font l'objet d'une défaillance dans la période de test - aucune entreprise défaillante n'est partagée entre apprentissage et test.
- Réglage et arrêt anticipé : sur un sous-ensemble d'entreprises distinctes de celles d'apprentissage - jamais le test.
- Test : 1<sup>er</sup> semestre 2025. Chaque défaillance mesurée concerne une entreprise que le modèle n'a jamais vue faire défaut à l'apprentissage - la seule mesure représentative d'un déploiement réel.

Le test n'est évalué qu'une seule fois. Les écarts de performance sont accompagnés d'intervalles de confiance (bootstrap<sup>8</sup> regroupé par entreprise) et de tests statistiques appariés (DeLong)<sup>9</sup> : nous ne déclarons un gain que s'il est statistiquement significatif. Un protocole plus exigeant que le classique partage in-sample / out-of-sample 90/10.

---

## 6. Les algorithmes comparés

La sélection d'un algorithme dépend du type de données, de l'interprétabilité et de la performance. Nous avons évalué plusieurs familles de techniques, du modèle de référence historique aux arbres boostés.

**Altman Z-score** - combinaison linéaire de ratios, standard de l'industrie depuis 1968. Référence de comparaison universelle.

**Régression logistique** - modèle statistique linéaire classique, transparent, mais limité face aux relations non linéaires.

**Arbre de décision** - un arbre unique (profondeur limitée) ; capture des non-linéarités simples mais reste sujet au surapprentissage.

**Gradient boosting d'arbres (LightGBM, XGBoost, CatBoost)** - état de l'art sur données tabulaires volumineuses. Trois implémentations indépendantes : leur convergence est en soi un gage de robustesse.

Le SVM (incalculable à l'échelle de plusieurs millions d'observations) et le classifieur naïf bayésien (inadapté aux distributions financières à queues lourdes) ont été testés puis écartés.

## 7. Résultats

Nous évaluons les modèles par la courbe ROC<sup>10</sup> et l'aire sous la courbe (AUC). Le tableau 2 compare, à jeu de variables identique, plusieurs familles d'algorithmes sur le test *out-of-time* 2025 - la mesure représentative d'un déploiement réel.

Tableau 2 - Pouvoir discriminant (AUC) par famille d'algorithmes, test out-of-time 2025

ALGORITHME	AUC (TEST OUT-OF-TIME)	LECTURE
Gradient boosting (LightGBM) - production	0,864	Bon pouvoir discriminant
Arbre de décision (simple)	0,824	Référence non linéaire
Régression logistique	0,810	Référence linéaire
Altman Z-score	≈ 0,75	Référence historique (littérature)

Source : Insourcia, test out-of-time 2025 (2,1 % de prévalence) ; modèles entraînés sur le même jeu de variables. Convention : AUC 0,70–0,80 « correct », 0,80–0,90 « bon », > 0,90 « excellent ».

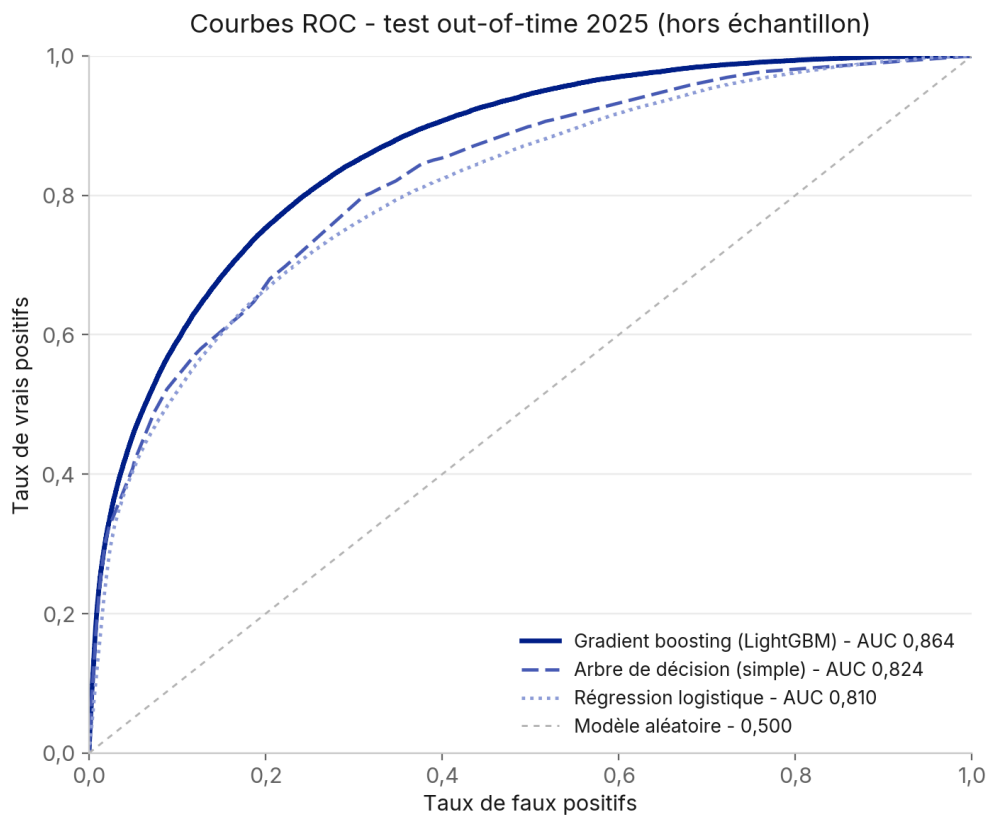


Figure 1 - Courbes ROC sur le test out-of-time 2025, par famille d'algorithmes. Plus la courbe s'élève vers le coin supérieur gauche, meilleur est le tri.

L'écart entre la régression logistique (0,810), l'arbre de décision simple (0,824) et le gradient boosting (0,864) mesure l'apport de la modélisation non linéaire - interactions et effets de seuil entre variables. Trois implémentations indépendantes de gradient boosting ont été comparées (LightGBM, CatBoost et

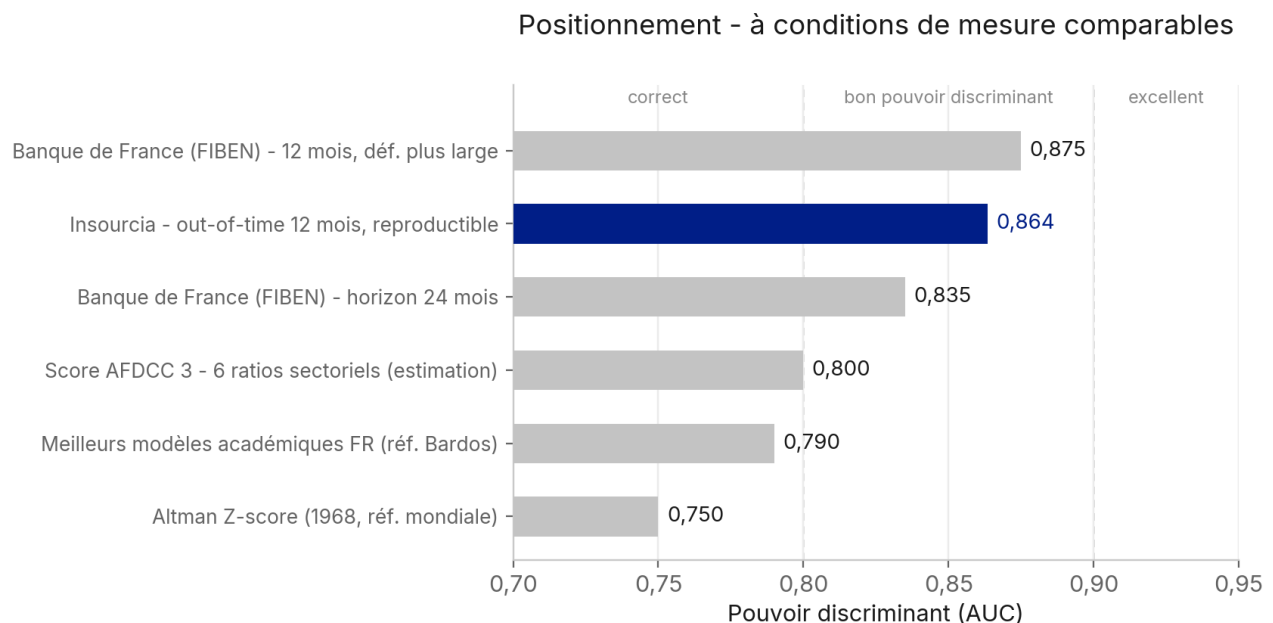
XGBoost) : leur convergence vers un même ordre de performance confirme que le résultat ne tient pas à un réglage particulier. LightGBM, le plus performant, est retenu comme modèle de production.

La performance se décline aussi par population :

POPULATION	AUC TEST 2025
Ensemble des entreprises	0,864
Entreprises sans antécédent (« nouvelles » défaillances)	0,820
Entreprises avec antécédent de procédure	0,677

## 8. Positionnement

Une performance ne se compare qu'à conditions de mesure égales. Notre AUC de 0,864 est mesurée *out-of-time* - le protocole le plus exigeant - et sur une cible stricte (procédure collective avérée). À ce niveau d'exigence, elle se situe au niveau de la cotation Banque de France sur périmètre comparable et des modèles internes bancaires de référence, nettement au-dessus des modèles académiques et des scores grand public.



**Figure 2** - Positionnement à conditions de mesure comparables. Bandes : « correct » / « bon » / « excellent ».

La cotation Banque de France publie un Gini<sup>11</sup> d'environ 77 % (AUC  $\approx$  0,885) à 12 mois sur une définition de défaut plus large (incidents de paiement inclus) ; ramenée à notre périmètre strict - la seule procédure collective - elle se situe autour de 0,85–0,86, soit le niveau de notre score. À horizon comparable de 24 mois, cette même cotation revient à une AUC d'environ 0,835, en deçà de notre score à 12 mois. Nous dépassons les modèles académiques de référence ( $\approx$  0,79), le score AFDCC 3 - outil de référence des credit managers français (6 ratios sectoriels notés sur 20, révisé en 2019), transparent mais linéaire ( $\approx$

0,80) - et le Z-score universel ( $\approx 0,75$ ) ; l'écart avec ces modèles linéaires mesure précisément l'apport de la non-linéarité et des variables non financières. Les deux acteurs de référence du marché français (Altares, Ellisphere) annoncent des AUC supérieures (0,90–0,92), mais sur des protocoles non auditable publiquement : notre différenciation est la transparence et la reproductibilité de la mesure.

Écart entre laboratoire et réalité. Il est fréquent d'observer sur le marché des scores théoriques dépassant 0,90 en environnement de test classique ; toutefois, l'absence de validation *out-of-time* conduit généralement à une surévaluation de la performance (sur données françaises, l'écart atteint couramment 0,10 point d'AUC). L'AUC de 0,864 rapportée ici est mesurée en conditions réelles - datée et reproductible.

---

## 9. Du modèle de validation au modèle de production

La performance d'un modèle ne se mesure honnêtement que sur un *backtest* figé : un modèle entraîné sur une période, puis évalué - une seule fois - sur une période postérieure dont l'issue est entièrement connue. Tous les chiffres de ce document proviennent de ce **modèle de validation** (apprentissage 1<sup>er</sup> semestre 2024, test 1<sup>er</sup> semestre 2025, sept mois d'écart). C'est le chiffre prudent et reproductible.

Mais un modèle qui sert en production ne doit pas rester figé sur une fenêtre ancienne : il doit rester aligné sur le régime économique courant. Insourcia distingue donc explicitement deux objets :

Modèle de validation	Figé sur le <i>backtest out-of-time</i> . Sert uniquement à mesurer la performance de façon auditable : c'est lui qui produit l'AUC de 0,864 et l'ensemble des statistiques publiées ici.
Modèle de production	Ré-entraîné sur une <b>fenêtre glissante de 12 mois</b> et re-scoré sur les données les plus récentes (au 30 mai 2026, 867 072 sociétés). Plus frais par construction : il intègre les derniers bilans et événements légaux et reste calé sur le régime de défaillance actuel.

Cette séparation est délibérée. Mesurer la performance sur une fenêtre glissante surévaluerait le chiffre (recouvrement des périodes) ; déployer un modèle figé de plusieurs mois dégraderait la pertinence réelle. Nous publions donc le chiffre du modèle de validation - conservateur - et déployons le modèle frais. Le modèle de production hérite exactement de la même architecture, des mêmes variables et des mêmes garde-fous anti-fuite ; seule sa fenêtre d'apprentissage est plus récente.

**Temporalité étendue.** Chaque note est datée et historisée : une société dispose d'une trajectoire de score dans le temps, qui matérialise une dégradation - ou une amélioration - souvent avant qu'elle ne se lise dans les comptes. Tout nouvel événement légal ou bilan est intégré au score en quelques jours.

## 10. Une probabilité réellement calibrée

Un bon classement (pouvoir discriminant) ne suffit pas : si le modèle annonce 10 % de probabilité de défaut, ce chiffre doit correspondre à 10 % de défauts observés dans la réalité. Nous calibrons<sup>12</sup> la probabilité sur la période de validation (jamais sur le test) par une transformation monotone qui préserve exactement le pouvoir discriminant tout en corrigeant le niveau. Sur l'essentiel de la population, la probabilité calibrée épouse la fréquence réelle ; pour les grades les plus risqués, la probabilité communiquée est ancrée directement sur les taux de défaut observés (section 11), garantissant la cohérence entre note et probabilité.

÷ 49

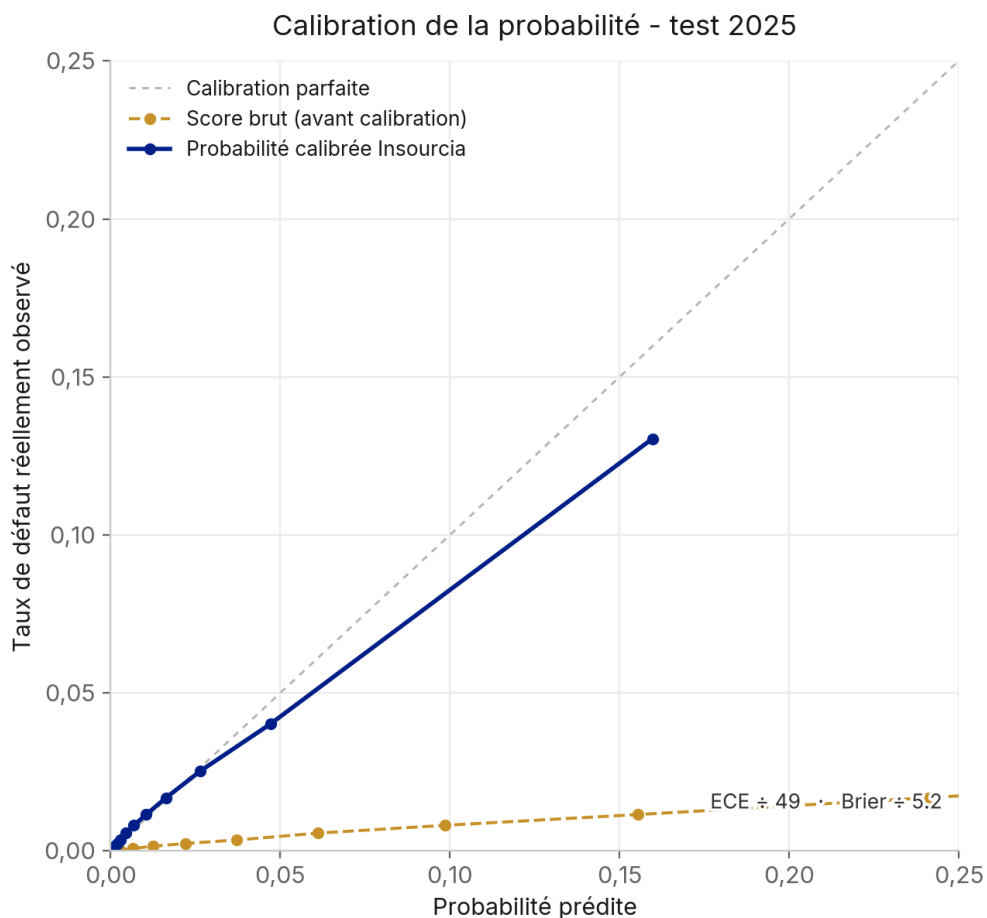
erreur de calibration (ECE) réduite sur le test

÷ 5,2

score de Brier réduit sur le test

2,3 %

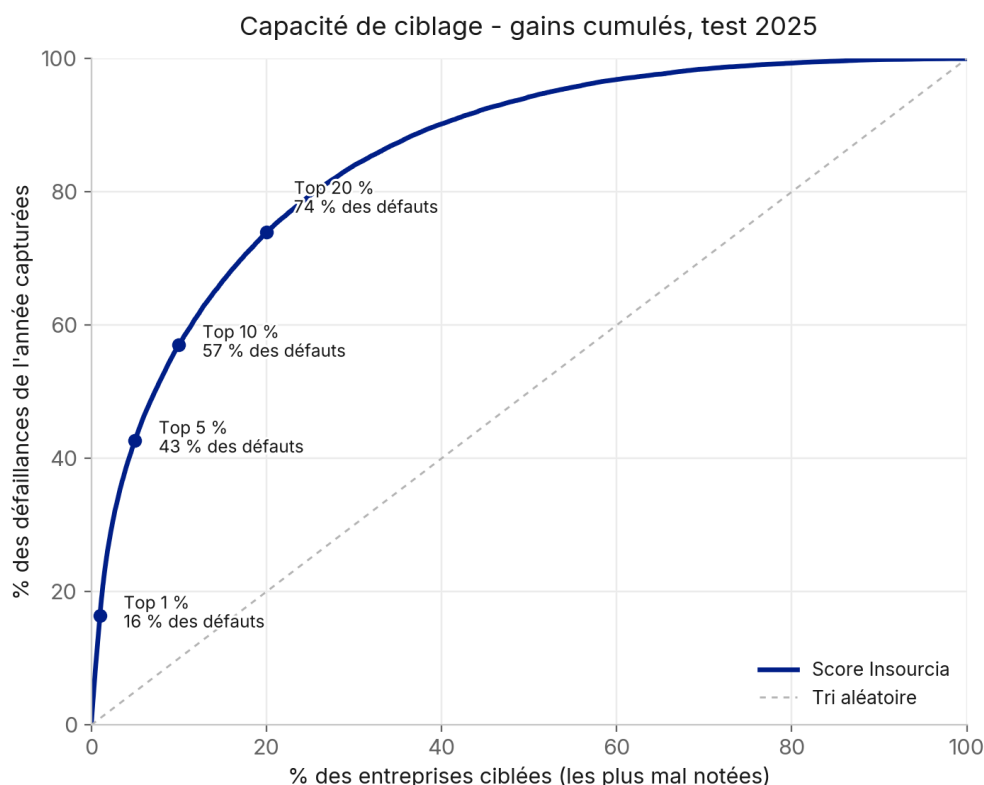
probabilité moyenne prédite, pour 2,1 % observé



**Figure 3** - Courbe de fiabilité. La probabilité calibrée épouse la diagonale. La probabilité communiquée est interprétable telle quelle.

## 11. Capacité de ciblage

L'usage opérationnel d'un score est la priorisation : concentrer l'attention sur les entreprises les plus à risque.



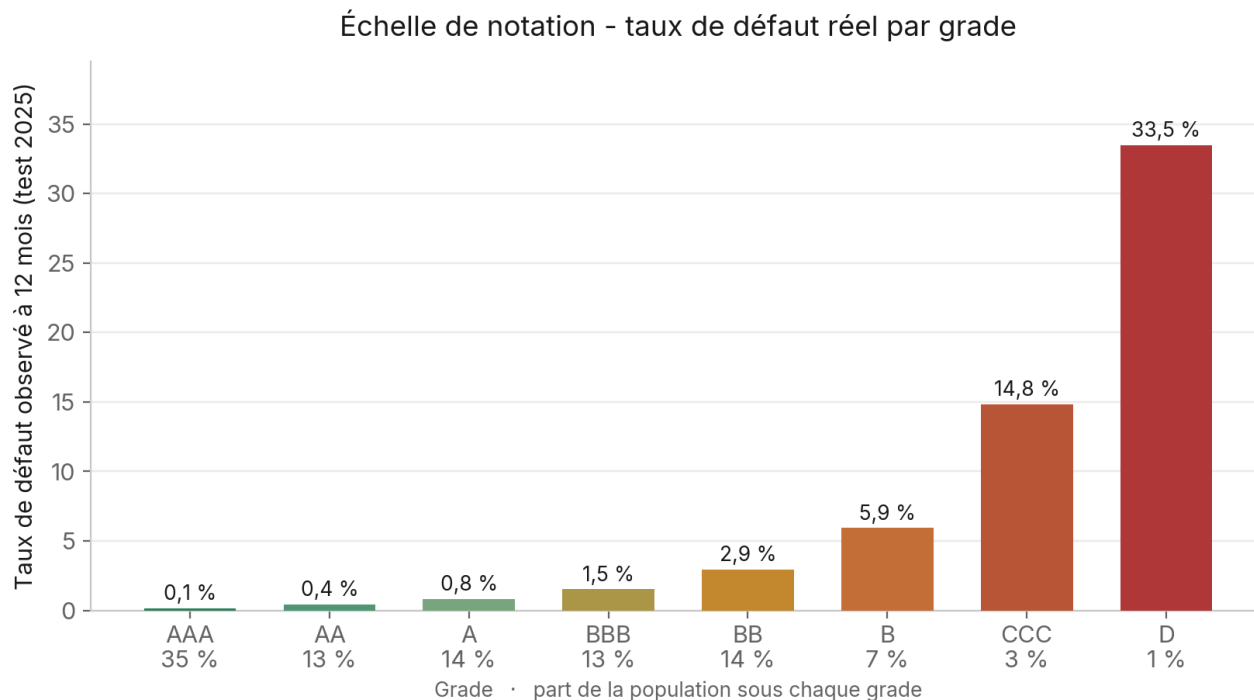
**Figure 4** - Gains cumulés (test 2025). En revoyant les 10 % les plus mal notés, on couvre 57 % des défaillances de l'année.

CIBLE (% DES PLUS RISQUÉES)	DÉFAUTS CAPTURÉS	VS HASARD
Top 1 %	16 %	x17
Top 5 %	43 %	x9
Top 10 %	57 %	x6
Top 20 %	74 %	x4

## 12. L'échelle de notation AAA → D

La probabilité calibrée est traduite en une échelle de grades à bornes fixes (stables dans le temps : une entreprise ne change de note que si son risque évolue, jamais parce que la population bouge). À chaque grade est attachée une probabilité de défaut égale à sa fréquence de défaut réellement observée sur la période de test hors échantillon - la convention des agences de notation. La probabilité communiquée

pour une note et le taux de défaut de cette note sont donc cohérents par construction. Les taux observés par grade sont strictement croissants, de 0,1 % pour AAA à 33,5 % pour D - preuve d'une échelle bien ordonnée.



**Figure 5** - Taux de défaut observé à 12 mois par grade (test 2025), avec la part de la population.

GRADE	TAUX DE DÉFAUT OBSERVÉ (12 M)	PART DE LA POPULATION
AAA	0,1 %	35 %
AA	0,4 %	13 %
A	0,8 %	14 %
BBB	1,5 %	13 %
BB	2,9 %	14 %
B	5,9 %	8 %
CCC	14,8 %	3 %
D	33,5 %	1 %

*Le grade D concentre un risque  $\approx 16$  fois supérieur à la moyenne. Le score interne brut n'est jamais exposé : seules la note et la probabilité calibrée le sont.*

### 13. Structure par terme : probabilité à 3, 6 et 12 mois

Au-delà du seul horizon 12 mois, le score se décline en une structure par terme : la probabilité de défaut à 3, 6 et 12 mois pour chaque grade. C'est le même classement de risque, calibré sur le taux de défaut effectivement observé à chacun de ces horizons. Cette granularité distingue une fragilité imminente - forte

probabilité dès 3 mois - d'un risque plus diffus, et alimente les dispositifs d'alerte précoce et de surveillance.

**Tableau 3 - Probabilité de défaut par grade et par horizon (test out-of-time 2025)**

GRADE	PD 3 MOIS	PD 6 MOIS	PD 12 MOIS
AAA	0,03 %	0,06 %	0,14 %
AA	0,10 %	0,20 %	0,44 %
A	0,19 %	0,36 %	0,83 %
BBB	0,38 %	0,72 %	1,52 %
BB	0,77 %	1,46 %	2,93 %
B	1,70 %	3,12 %	5,93 %
CCC	4,61 %	8,41 %	14,84 %
D	10,32 %	19,49 %	33,47 %

Source : Insourcia, taux de défaut observés par grade à chaque horizon (test 2025). Probabilités strictement croissantes avec l'horizon et avec le grade : une structure par terme bien ordonnée.

## 14. L'approche Insourcia : un modèle explicable

Le modèle de production - Insourcia Default Score (IDS), version 1.0 (2026) - est un gradient boosting (LightGBM) sur les 74 variables décrites. Chaque note est expliquée par une méthode d'attribution exacte (valeurs de Shapley<sup>13</sup>). Le score est livré avec ses 5 principaux facteurs, leur sens (aggravant ou atténuant) et un libellé clair, sans donnée confidentielle.

## 15. Exemple de sortie

La figure 6 illustre la sortie du score sur une société réelle (identité masquée). Le modèle attribue une probabilité de défaut élevée (grade CCC) et en explique l'origine : une autonomie financière dégradée, une trésorerie faible, des années de résultat net négatif et un contexte sectoriel défavorable. Chaque facteur est vérifiable par un analyste.

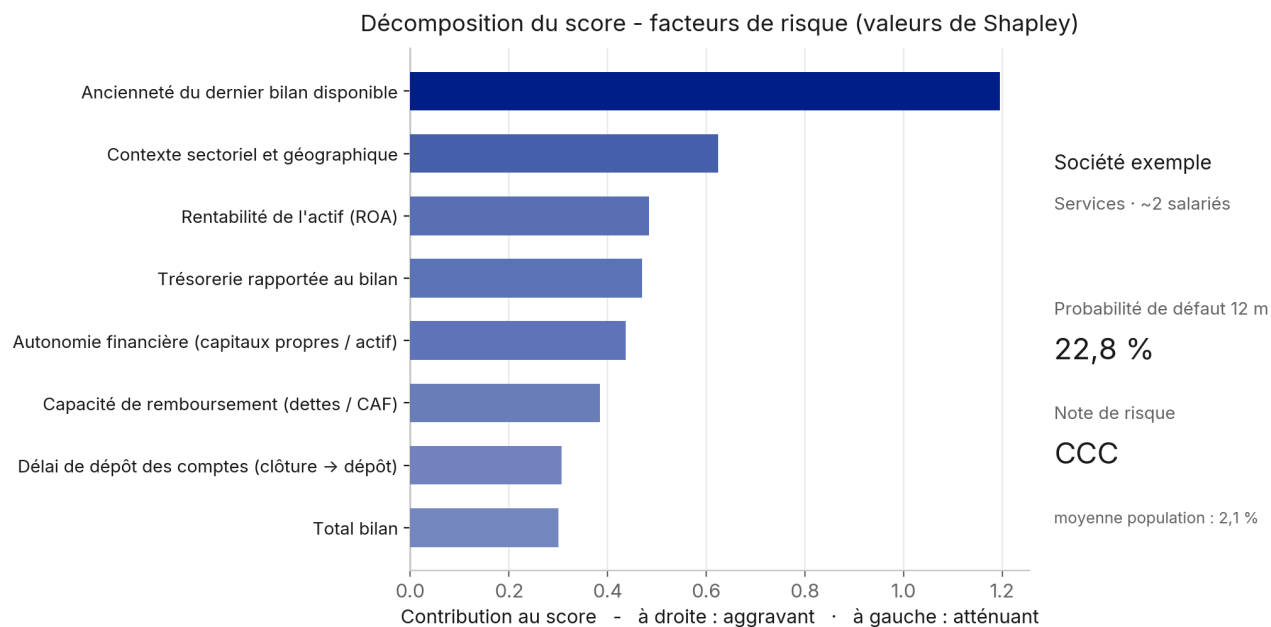


Figure 6 - Décomposition du score d'une société exemple (vrai positif : défaillance survenue en 2025).

## 16. Cas d'usage

- Origination et due diligence M&A : identification précoce de signaux faibles et analyse de la granularité du risque sur une cible et ses filiales (niveaux -1, -2) avant l'émission d'une LOI.
- Surveillance de portefeuille (PE, prêteurs, assurance-crédit) : alerte sur tout passage en zone de risque.
- Scoring de contreparties B2B : qualification clients/fournisseurs, priorisation du recouvrement.
- Origination *special situations* (retournement) : détection proactive d'actifs décotés ou de structures présentant des vulnérabilités de liquidité à court terme (grades B/CCC), offrant des opportunités de restructuration.

## 17. Périmètre et gouvernance

Nous ne notons que les entités disposant de comptes exploitables ; les autres reçoivent le statut NR, sans note par défaut. Cycle de vie du score : re-scoring quotidien (un nouvel événement légal ou bilan est intégré en quelques jours), surveillance de dérive (stabilité des variables et performance sur cohortes récentes), versionnage traçable de chaque note, réentraînement périodique.

---

## 18. En résumé

**Insourcia Default Score (IDS), version 1.0** - un score de défaillance à AUC 0,864 mesuré en conditions réelles out-of-time, calibré, explicable, sans fuite du futur, et reproductible sur données publiques. Nous publions la méthode pour permettre sa vérification.

Accès par API, par SIREN : score + grade + facteurs explicatifs, re-scoring quotidien.

Demandez une démonstration ou un échantillon de scoring sur votre portefeuille.

---

### NOTES & DÉFINITIONS

1. Procédure judiciaire de traitement des difficultés d'une entreprise : sauvegarde, redressement judiciaire ou liquidation judiciaire.
2. Probabilité de défaut : probabilité qu'un événement de défaut survienne sur l'horizon retenu (ici 12 mois).
3. Protocole de validation où l'apprentissage porte sur une période et le test sur une période strictement postérieure, comme en production réelle.
4. Aire sous la courbe ROC : probabilité que le modèle note plus mal une entreprise qui fait défaut qu'une entreprise saine. Convention d'interprétation : 0,70 à 0,80 « correct », 0,80 à 0,90 « bon pouvoir discriminant », au-delà de 0,90 « excellent ».
5. Technique d'apprentissage combinant un grand nombre d'arbres de décision, chacun corrigeant les erreurs des précédents ; état de l'art sur données tabulaires.
6. Not Rated : statut des entités hors périmètre (sans comptes exploitables), volontairement non notées plutôt que notées par défaut.
7. Principe consistant à n'utiliser, à une date d'observation T, que l'information réellement disponible à cette date ; aucune donnée postérieure n'entre dans le calcul.
8. Rééchantillonnage permettant d'estimer l'incertitude (intervalle de confiance) d'une mesure ; ici regroupé par entreprise, les observations d'un même SIREN n'étant pas indépendantes.
9. Test statistique de comparaison de deux AUC mesurées sur le même échantillon, pour valider qu'un écart de performance est significatif.
10. Courbe reliant le taux de vrais positifs au taux de faux positifs pour tous les seuils de décision possibles.
11. Indice de pouvoir discriminant équivalent à l'AUC :  $Gini = 2 \times AUC - 1$  (un Gini de 75 % équivaut à une AUC de 0,875).
12. Calibration : alignement de la probabilité prédite sur la fréquence de défaut réellement observée. Sa qualité se mesure par l'ECE (Expected Calibration Error) et le score de Brier - plus ils sont bas, meilleure est la calibration.
13. Méthode d'attribution issue de la théorie des jeux, décomposant exactement un score individuel en contributions additives de chaque variable.

---

Insourcia - Private Market Intelligence. Document de méthodologie à usage d'information. Les performances rapportées sont mesurées out-of-time sur un test 2025 évalué une seule fois ; elles sont indicatives et ne constituent ni une garantie de résultat ni une recommandation d'investissement. Le score est un outil d'aide à la décision et ne se substitue pas au jugement de l'utilisateur. Aucune donnée nominative confidentielle n'entre dans le score exposé. Sources : registres publics (RNE, BODACC, SIRENE/INSEE, comptes annuels, brevets, marchés publics, subventions). © 2026 Insourcia.

---

Copyright © 2026 Insourcia - Private Market Intelligence. Tous droits réservés.

**insourcia.io**