

creditsafe⁺

Creditsafe France

Synthèse

Construction des cartes de scores

Group Analytics

Mars 2021



1. Executive summary

1.1 - Observations au sujet des cartes de score Creditsafe France

Dans l'objectif de renforcer et d'étendre la présence de Creditsafe sur le marché français, un nouvel ensemble de cartes de score a été développé en interne par l'équipe *Creditsafe Group Analytics*, basée en Suède.

Creditsafe s'engage dans une démarche d'amélioration continue de ses outils *Score*, et utilise les dernières connaissances dans le domaine des *méthodes de scoring statistiques*. Ce afin d'optimiser et de maximiser l'exploitation de sa base de données, en constante expansion. Cette posture permet, par différents procédés, la réception de données toujours plus pertinentes, et assure à Creditsafe d'utiliser les informations les plus à jour et les plus appropriées parmi les informations disponibles au sein du marché *business intelligence* français. Ainsi, la façon dont Creditsafe *score* les entreprises s'améliore, et aboutit à des évaluations plus prédictives sur un nombre croissant d'entreprises, y compris sur des entreprises qui présentent moins d'informations disponibles.

1.2 - Cartes de score et Segmentation

L'approche de *scoring* Creditsafe a pour principe fondamental de prédire avec justesse le comportement d'affaire d'une entreprise au cours des 12 prochains mois, à travers une dichotomie *good performance / bad performance*. Cette approche s'appuie sur un ensemble de caractéristiques, qui identifient clairement les facteurs de risque élevé et les facteurs de risque faible.

Afin de conduire une discrimination optimale des entreprises, la population a été segmentée. L'objectif de cette segmentation est de définir une série de sous-populations avec, pour chacune d'entre elles, un modèle statistique dédié. Ainsi, l'utilisation d'un ensemble de modèles dédiés permet une évaluation du risque plus efficace que l'utilisation d'un modèle statistique unique qui serait appliqué à l'ensemble de la population. La base de cette segmentation repose sur l'*obligation de déclaration des comptes annuels*. Les entreprises *Avec obligation* sont catégorisées par taille quand des comptes sont disponibles, par âge quand les comptes sont absents. Deux cartes de score ont été créées spécifiquement pour les entreprises *Sans obligation*, qui s'avèrent moins pourvues en données prédictives. Ces deux cartes de scores séparent les entreprises par la forme juridique.

Pour les MICRO, SMALL et MED-LARGE, les tailles sont attribuées dès lors que de 2 des 3 critères sont atteints.

Avec obligation de déclaration des comptes

- 1- Nouvelles : Comptes absents. Entreprise de moins de 24 mois.
- 2- Anciennes : Comptes absents. Entreprise de plus de 24 mois.
- 3- Microentreprises : Avec comptes
- 4- Petites entreprises : Avec comptes
- 5- Moyennes et Grandes entreprises : Avec comptes

Sans obligation de déclaration des comptes

- 6- Structures sans obligation de déclaration : Associations déclarées, Sociétés civiles immobilières, Sociétés civiles professionnelles...etc
- 7- Entrepreneurs individuels : Affaires personnelles, Artisans, Commerçants...etc



1.3 – Correspondance entre les Probabilités de défaut individuelles et le Score de 1 à 100.

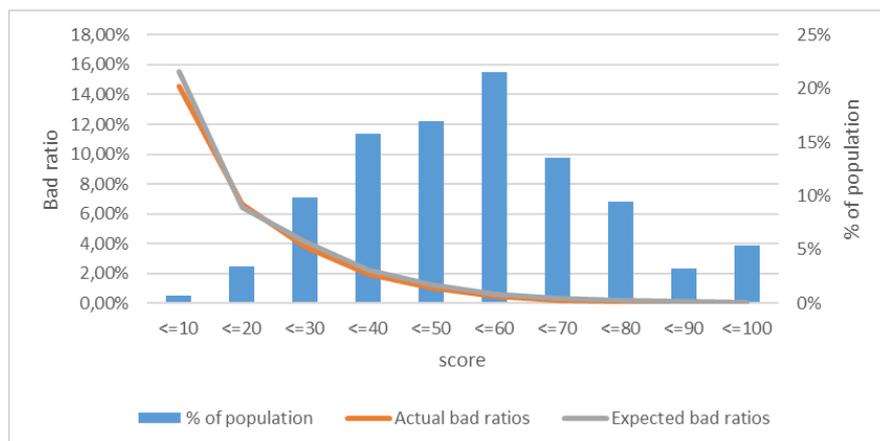
Les cartes de score Creditsafe produisent pour chaque entreprise un score entre 1 et 100. Ces nombres représentent respectivement le risque le plus haut et le risque le plus bas. L'échelle de 1 à 100 est construite à partir des probabilités de défaillance calculées par les modèles statistiques. Elle vise à octroyer aux utilisateurs Creditsafe une grille de lecture simple pour accepter, ou décliner, la demande d'en-cours d'une entreprise. Grâce au lancement des nouvelles cartes de score, le score Creditsafe, en tant qu'indicateur, sera désormais accompagné d'une probabilité de défaillance propre à chaque entreprise, établie par combinaison des risques associés à chacune de ses caractéristiques. Concernant la conversion, dans la mesure où chaque probabilité de défaillance s'avère propre à chaque entreprise, chaque score représente un intervalle de probabilité de défaillances. Par exemple, un score de 30 représente une probabilité de défaillance entre 3.03% et 3.11%.

1.4 – Synthèse des résultats

Matrice de répartition des entreprises AVEC COMPTES

Score band	Min PD	Max PD	Min Score	Max Score	% of population	Bad ratio
A	0,01%	0,25%	71	100	18%	0,09%
B	0,25%	0,91%	51	70	35%	0,36%
C	0,91%	3,12%	30	50	33%	1,49%
D	3,12%	99,99%	1	29	14%	5,06%

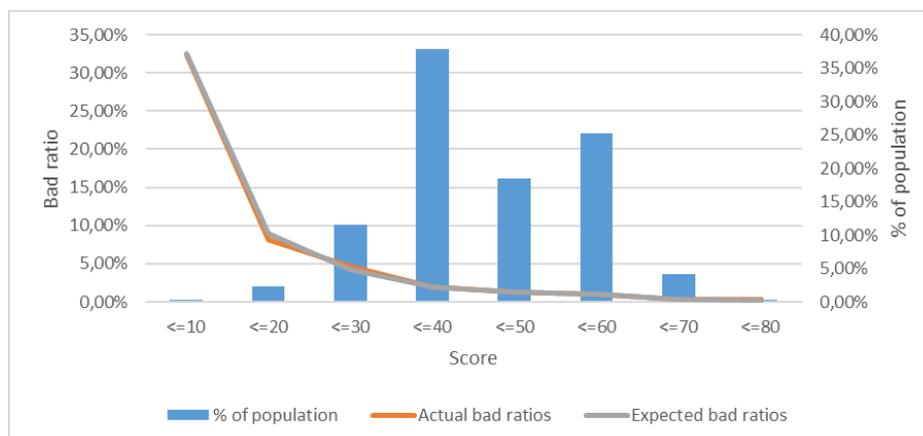
Distribution des entreprises AVEC COMPTES



Matrice de répartition des entreprises SANS COMPTES

Score band	Min PD	Max PD	Min Score	Max Score	% of population	Bad ratio
A	0,01%	0,25%	71	100	0,25%	0,25%
B	0,25%	0,91%	51	70	29,30%	0,89%
C	0,91%	3,12%	30	50	56,43%	1,71%
D	3,12%	99,99%	1	29	14,02%	5,96%

Distribution des entreprises SANS COMPTES





Coefficient de Gini

Le coefficient de Gini montre une discrimination pertinente à travers l'ensemble des segments de la population des entreprises françaises. Afin de confirmer la solidité des cartes de score, Creditsafe les a validées par *validation hors-délai*. Les résultats ont montré que l'ensemble des paramètres correspondent aux attentes sur la précision du modèle. Creditsafe assure une surveillance et un examen continus des cartes de score, afin de maintenir leur fiabilité. Le coefficient de Gini des cartes de score 2021 varie en fonction du segment étudié, principalement en raison de la disparité de données disponibles d'un segment à l'autre, et atteint jusqu'à 73%.

2. Préparation des données et Définition de la population

2.1 – Définition de l'échantillon

Les cartes de score ont été développées à partir d'un échantillon extrait des données françaises de Creditsafe. La sélection de cet *échantillon de développement* a rempli les conditions suivantes :

- L'échantillon s'appuie sur des entreprises économiquement actives. La population doit être suffisamment nombreuse pour développer un modèle fiable.
- Chaque entreprise a une période d'exposition de 12 mois, estimée suffisante pour voir survenir une éventuelle défaillance. Cette période couvre une année entière afin de neutraliser la saisonnalité.
- La fenêtre d'observation de l'échantillon s'avère suffisamment récente pour être *représentative* de la future population française.

La période d'observation de l'*échantillon de développement* s'étend entre 01 Juin 2015 et le 31 Mai 2017. La période d'exposition, durant laquelle est déterminée la *good performance* ou la *bad performance*, se trouve entre le 01 Juin 2017 et le 31 Mai 2018.

2.2 - Définition de la défaillance

La définition de la *good performance* et la *bad performance* apparaissent ci-dessous. Une entreprise est considérée BAD dès lors qu'elle présente un jugement classé BAD dans le tableau ci-dessous.

Qualification	Jugement
BAD	Clôture pour insuffisance d'actif Conversion de la sauvegarde en LJ Conversion de la sauvegarde en RJ Faillite personnelle Jugement de clôture après cession Liquidation judiciaire Liquidation judiciaire simplifiée Redressement judiciaire Redressement judiciaire simplifié Interdiction de gérer Extension de la procédure de redressement judiciaire Extension de la procédure de liquidation judiciaire Cessation de paiement
GOOD	Absence de l'un des jugements cités ci-dessus

3. Développement des cartes de score

Méthodologie de la modélisation

Les cartes de score ont été développées par *régression logistique automatisée*. Il s'agit de la méthode retenue au sein du Groupe Creditsafe. La régression logistique a l'avantage d'aboutir à la prédiction d'une probabilité de *good performance*, cela permet la création d'un score précis de prédiction de *bad performance*.

Confrontation à la *logique des affaires*

Même si une variable s'avère fortement prédictive, il demeure nécessaire de vérifier si son occurrence est logique et attendue vis-à-vis du résultat *good/bad*. La première confrontation à la *logique des affaires* est effectuée lors de l'étude initiale menée sur l'ensemble des variables, durant laquelle sont écartées les variables considérées comme dépourvues de sens d'un point de vue *business*.

Par ailleurs, une seconde confrontation est requise après construction du modèle résultant de la régression. Si les variables justes ont été sélectionnées à l'étude initiale, le modèle sera prédictif d'un point de vue mathématique. Cependant, une confrontation à la logique des affaires demeure nécessaire.

En particulier, il est crucial de vérifier que le poids d'une caractéristique dans la détermination du score soit en cohérence avec le taux de défaillance associé à cette caractéristique. Il convient également de s'assurer que le sens, positif ou négatif, de l'impact d'une caractéristique correspond à ce qu'attend la logique *business*.

Calibrage des cartes de score

Le modèle de régression produit une probabilité de défaillance individuelle, propre à chaque entreprise de chaque modèle. L'ensemble de ces probabilités de défaillances a été réparti sur une échelle de 1 à 100, de sorte que le risque de défaillance double à chaque fois qu'un intervalle de points défini est passé. Ce facteur a été déterminé afin que le signal renvoyé par le score concorde avec le risque de défaillance, et aussi afin de permettre la distribution des scores attendue pour les entreprises françaises. Le score est donc le reflet direct de la probabilité de défaillance d'une entreprise.

Le tableau ci-dessous détaille les intervalles de probabilités de défaillance pour chaque score de 1 à 100.

Score	Min_PD	Max_PD	Band	Score	Min_PD	Max_PD	Band
100	0	0,0461%	A	50	0,9080%	0,9649%	C
99	0,0461%	0,0488%	A	49	0,9649%	1,0253%	C
98	0,0488%	0,0517%	A	48	1,0253%	1,0895%	C
97	0,0517%	0,0548%	A	47	1,0895%	1,1576%	C
96	0,0548%	0,0580%	A	46	1,1576%	1,2299%	C
95	0,0580%	0,0615%	A	45	1,2299%	1,3067%	C
94	0,0615%	0,0651%	A	44	1,3067%	1,3882%	C
93	0,0651%	0,0690%	A	43	1,3882%	1,4748%	C
92	0,0690%	0,0731%	A	42	1,4748%	1,5666%	C
91	0,0731%	0,0774%	A	41	1,5666%	1,6640%	C
90	0,0774%	0,0821%	A	40	1,6640%	1,7674%	C
89	0,0821%	0,0869%	A	39	1,7674%	1,8772%	C
88	0,0869%	0,0921%	A	38	1,8772%	1,9935%	C
87	0,0921%	0,0976%	A	37	1,9935%	2,1170%	C
86	0,0976%	0,1034%	A	36	2,1170%	2,2479%	C
85	0,1034%	0,1095%	A	35	2,2479%	2,3867%	C
84	0,1095%	0,1160%	A	34	2,3867%	2,5339%	C
83	0,1160%	0,1229%	A	33	2,5339%	2,6899%	C
82	0,1229%	0,1302%	A	32	2,6899%	2,8552%	C
81	0,1302%	0,1379%	A	31	2,8552%	3,0303%	C
80	0,1379%	0,1461%	A	30	3,0303%	3,1154%	C
79	0,1461%	0,1548%	A	29	3,1154%	3,3385%	D
78	0,1548%	0,1640%	A	28	3,3385%	3,5769%	D
77	0,1640%	0,1737%	A	27	3,5769%	3,8318%	D
76	0,1737%	0,1840%	A	26	3,8318%	4,1039%	D
75	0,1840%	0,1949%	A	25	4,1039%	4,3946%	D
74	0,1949%	0,2065%	A	24	4,3946%	4,7048%	D
73	0,2065%	0,2188%	A	23	4,7048%	5,0357%	D
72	0,2188%	0,2317%	A	22	5,0357%	5,3886%	D
71	0,2317%	0,2455%	A	21	5,3886%	5,7648%	D
70	0,2455%	0,2600%	B	20	5,7648%	6,1655%	D
69	0,2600%	0,2755%	B	19	6,1655%	6,5921%	D
68	0,2755%	0,2918%	B	18	6,5921%	7,0460%	D
67	0,2918%	0,3091%	B	17	7,0460%	7,5286%	D
66	0,3091%	0,3274%	B	16	7,5286%	8,0414%	D
65	0,3274%	0,3468%	B	15	8,0414%	8,5859%	D
64	0,3468%	0,3673%	B	14	8,5859%	9,1636%	D
63	0,3673%	0,3891%	B	13	9,1636%	9,7761%	D
62	0,3891%	0,4121%	B	12	9,7761%	10,4247%	D
61	0,4121%	0,4365%	B	11	10,4247%	11,1111%	D
60	0,4365%	0,4624%	B	10	11,1111%	11,8367%	D
59	0,4624%	0,4897%	B	9	11,8367%	12,6030%	D
58	0,4897%	0,5187%	B	8	12,6030%	13,4114%	D
57	0,5187%	0,5494%	B	7	13,4114%	14,2631%	D
56	0,5494%	0,5819%	B	6	14,2631%	15,1595%	D
55	0,5819%	0,6163%	B	5	15,1595%	16,1016%	D
54	0,6163%	0,6527%	B	4	16,1016%	17,0904%	D
53	0,6527%	0,6912%	B	3	17,0904%	18,1269%	D
52	0,6912%	0,7320%	B	2	18,1269%	19,2117%	D
51	0,7320%	0,9080%	B	1	19,2117%	1	D