

RAPPORT DE STAGE :

**PREVISION DES FAILLITES DES BANQUES EN UTILISANT
LEURS ETATS FINANCIERS HISTORIQUES**

Romain Ribeiro

Encadrants :

Angelo Riva¹ et Jay Yun Jun Kim²

*1- European Business School, Paris School of Economics, Paris-Jourdan
Sciences Économiques - UMR8545, Paris*

*2- Axe de Mathématiques pour l'Ingénierie Scientifique et Financière -
Laboratoire de recherche ECE, Paris*

Avril 2021 – Aout 2021

SOMMAIRE

1- Résumé	3
2- Remerciements	4
3- Introduction	5
4- Présentation de l'entreprise	5
4.1 Présentation de l'école ECE Paris	6
4.2 Présentation du Laboratoire de recherche de l'ECE Paris	6
5- Contexte	7
5.1 Causes et conséquences des crises financières	7
5.1.1 Le cycle financier naturel	7
5.1.2 Eclatement de la bulle spéculative	8
5.1.3 Panique et faillites bancaires : la contagion	9
5.1.4 Conséquences	9
5.2- La cas particulier Français au travers de la crise de 1929	10
6- Méthodologie	10
6.1 Objectifs	10
6.2 Etat de l'art et recherche bibliographique.	11
6.2.1 Financial Crises : Theory and evidence. (Allen F, Babus A, and Carletti E)	11
6.2.2 Banking Crises : A review. (Laeven L)	11
6.2.3 How does capital affect bank performance during financial crises? (Berger A N & Bouwman CHS)	12
6.2.4 La crise bancaire de 1930 en France : une nouvelle analyse historique et quantitative des données du Crédit Lyonnais. (Vilpoux E)	13
6.3 Base de données	13
6.4 Choix du model	13
6.4.1 Kmean	13
6.4.2 Régression logistique	14
6.5 Etablissement du modèle	15
6.5.1 Problèmes	15
6.5.2 Solution	16
6.6 Bibliothèques et implémentation	17
7- Résultats	17
8- Conclusion et perspectives.	19
9- Références Bibliographiques	20

1- Résumé

Ce travail de recherche a été effectué au laboratoire de recherche de l'ECE en collaboration avec l'EBS. Ce projet vise à la fois à faciliter la compréhension de l'impact de la crise bancaire de 1929 en France et à permettre la prédiction et donc l'anticipation des faillites bancaires.

Les crises bancaires entraînant des conséquences lourdes, l'anticipation de celles-ci présente un fort intérêt. Ce stage avait notamment pour objectif d'implémenter un algorithme de Machine Learning qui pourrait prédire efficacement les futures faillites en se basant sur les données limitées des banques française dans le contexte de la Grande Dépression.

A l'aide de plusieurs bibliothèques, un algorithme a été réalisé et conçu dans l'optique d'être facile réutilisable et améliorable.

Ce rapport introduit les différents mécanismes qui régissent les crises bancaires et explique comment l'algorithme en question a été réalisé et quels résultats il obtient en passant par les difficultés rencontrées.

Réaliser en python, même s'il est d'une efficacité à relativiser, il montre néanmoins une corrélation entre les liquidités d'une banque et sa capacité à ne pas faire faillite.

2- Remerciements

Tout d'abord, je tiens à remercier Mme Assia Soukane pour m'avoir accueilli chaleureusement au sein de son laboratoire de recherche.

Je remercie chaleureusement Jae Yun JUN KIM, enseignant-chercheur à l'ECE dans l'équipe « mathématiques pour les sciences de l'ingénieur » pour m'avoir pris en charge tout au long de ce stage ainsi que Angelo Riva, professeur à l'EBS pour sa patience et m'avoir initié à la problématique des crises financières.

Enfin, merci à toute l'équipe pour sa gentillesse et sa bonne ambiance, ce qui m'a permis de travailler avec envie et rigueur.

3- Introduction

Mon projet de stage intitulé : « ' La prévision des faillites des banques à partir de l'analyse de leurs états financiers » est à l'interface entre mathématiques et finances.

Les banques sont au cœur de l'économie mondiale. De tout temps, il a existé des crises bancaires qui ont des répercussions sur l'ensemble de la population et qui vont au-delà du chômage, comme l'éducation et la santé, voire le déclenchement de guerres.

Pouvoir prédire les faillites bancaires est donc un enjeu non seulement économique mais aussi politique. Depuis les années 2000 les acteurs de la finance cherchent mieux connaître les mécanismes des faillites bancaires pour pouvoir les prédire et anticiper leurs effets délétères.

Un enjeu important est donc de modéliser mathématiquement ces crises par un algorithme de prédiction.

4- Présentation de l'entreprise

4.1 Présentation de l'école ECE Paris

L'ECE est une école d'ingénieur généraliste, fondée en 1919 par Eugène Poirot, sous le nom d'École centrale de la TSF. La TSF est alors une technologie naissante, et l'école reste longtemps la seule à former les opérateurs radio civils et militaires. En 1960, son nom devient *École centrale de TSF et d'électronique*, puis, en 1963, *École centrale d'électronique*. Elle ajoute à son cursus l'application des transistors. Aujourd'hui, L'ECE forme principalement dans les technologies de l'information. Elle propose également des formations axées sur la santé et son lien avec la technologie, sur les transports et la mobilité, sur les énergies et l'environnement.

L'école possède deux campus situés à Paris et à Lyon et compte en 2019 un total de 2500 élèves et 398 enseignants dont 82 permanents. Son directeur actuel est Mr François Stephan. Elle propose 8 spécialités : 1- Santé et Technologie, 2- Nouvelles Energies et Environnement, 3- Objets Connectés, Réseaux et Services, 4- Véhicules Connectés et Autonomie, 5- Systèmes Embarqués, Aéronautique et Robotique, 6- Big Data et Analytics, 7- Système d'Information et Cyber Sécurité Défensive, 8- Finance et Ingénierie Quantitative.

L'école est membre du groupe INSEEC U regroupant seize grandes écoles et écoles spécialisées en management, sciences de l'ingénieur, communication et sciences politiques. Elle est aussi membre de la Conférence des grandes écoles (CGE), de Campus France et de l'Union des grandes écoles indépendantes.

4.2 Présentation du Laboratoire de recherche de l'ECE Paris

Le Laboratoire de recherche de l'ECE Paris est composé de 3 groupes axes autour de trois thématiques majeures : Systèmes intelligents et communicants, Nanosciences et nanotechnologies pour la santé et l'énergie, Mathématiques pour les sciences de l'ingénieur (Figure 1). Ces thématiques sont en lien avec celles enseignées dans le cursus ingénieur de l'ECE, comme la santé, l'énergie, les transports et la finance.

Les chercheurs cultivent des collaborations avec des organismes de recherche nationaux et internationaux. Ils créent des interactions avec les étudiants, en lien avec leurs activités de recherche. Le laboratoire compte 25 membres et est dirigé par Mme Assia Soukane (Figure 1).

Mon stage s'est déroulé dans l'axe « Mathématiques pour les sciences de l'ingénieur ».

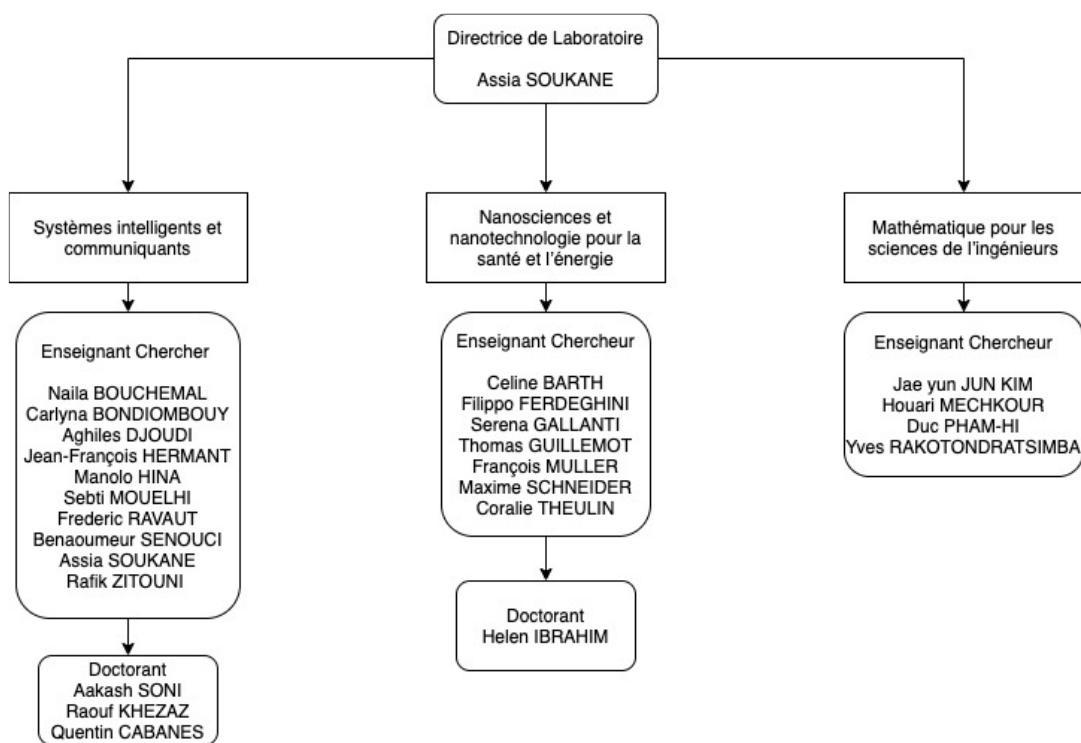


Figure 1: Diagramme de l'organisation du laboratoire de recherche de l'ECE

5- Contexte

Les crises bancaires mettent souvent un terme à une période d'exubérance et ruinent l'économie réelle en réduisant le crédit et en provoquant des liquidations. Les crises bancaires sont un phénomène courant tout au long de l'histoire. Ainsi, il a été dénombré 268 crises bancaires sur la période 1800 à 2008, et montré que la fréquence des crises bancaires a augmenté au cours des dernières décennies suite à libéralisation financière dans les années 1980 pour atteindre un niveau jamais vu depuis la Grande Dépression en les années (Leaven L).

5.1 Causes et conséquences des crises financières

Les banques et plus largement les institutions financières jouent un rôle majeur dans le financement de l'économie. En effet, le premier rôle des banques est l'octroi de crédits à des agents économiques que la banque choisit en fonction de leur capacité à rembourser.

Les crises financières sont dues principalement au cycle financier naturel, à l'éclatement de bulles spéculatives et au phénomène de panique

5.1.1 Le cycle financier naturel

Le cycle financier naturel peut se décomposer en 5 phases (Figure 2).

Phase 1 : essor. La phase d'essor se caractérise par une croissance économique après un ralentissement ou une récession. Cette phase de croissance peut être générée par une innovation, un changement technologique ou institutionnel ou encore alimentée par un investissement productif. Cette croissance engendre une expansion du crédit qui favorise une hausse du prix des actifs.

Phase 2 : euphorie. Cette phase se caractérise par un cercle auto-entretenu entre emballement du crédit et hausse du prix de certains actifs. Les risques sont sous-évalués, ce qui aggrave la fragilité des structures financières.

Phase 3 : paroxysme et retournement. Cette phase est un point de retournement, le passage d'une période d'euphorie à une période de décroissance. Les anticipations des agents économiques se renversent. Après une période de hausse régulière, on assiste à une chute brutale du prix des actifs. Ainsi les crises bancaires systémiques sont généralement précédées de booms du crédit et de bulles spéculatives

Phase 4 : reflux et pessimisme. Cette phase correspond ici au début de la crise. Les prix des actifs baissent brutalement.

Phase 5 : Déflation de la dette et restructuration du bilan. Les investisseurs surendettés doivent se désendetter. Ils peuvent être contraints de vendre leurs actifs pour faire face à leur besoin de liquidité, déclenchant une spirale de baisse auto-entretenue du prix de ces actifs et un assèchement de la liquidité. Cette situation se propage de marché en marché, d'un continent à l'autre. Ce phénomène de contagion contraint les banques à resserrer leur offre de crédit. La dépense privée se contracte (volonté de désendettement de la part des agents) et pèse sur les revenus et donc sur la croissance qui finalement freine le désendettement.

LES CYCLES FINANCIERS

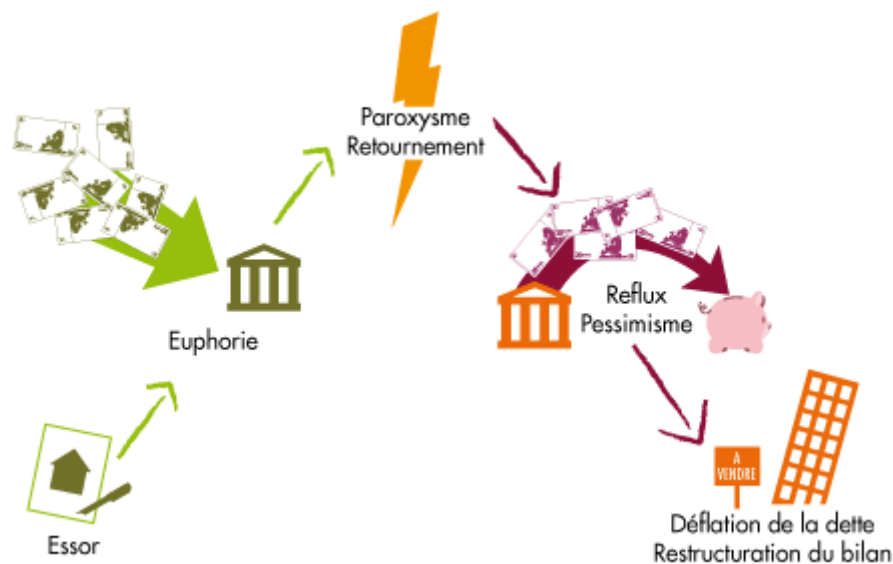


Figure 2 : Les cycles financiers naturels (d'après <https://www.lafinancepourtous.com/decryptages/crises-economiques/crise-des-subprimes/crise-financiere/les-cycles-financiers/>)

5.1.2 Eclatement de la bulle spéculative

Une bulle spéculative est une situation dans laquelle le prix de marché d'un actif financier s'éloigne de sa valeur fondamentale. Il est « surévalué ». De plus des comportements mimétiques alimentent le gonflement des bulles spéculatives par anticipation (*Aghion P et Menger P-M*).

Pour une raison ou pour une autre et à partir d'un certain moment — dont la date n'est jamais aisée à prévoir — une majorité d'agents détenteurs d'actifs financiers renversent leurs anticipations. Une fois les anticipations retournées à la baisse, c'est-à-dire que la majorité des agents sont devenus vendeurs, la demande de titres diminue, l'offre augmente et le prix de l'actif financier diminue effectivement : la bulle éclate (*Aghion P et Menger P-M*).

5.1.3 Panique et faillites bancaires : la contagion

Les deux causes de faillite correspondent, en schématisant, à deux des principales sources de paniques bancaires: la demande contagieuse de conversion des dépôts en espèces d'une part, et la détérioration de la qualité des créances bancaires par défaillance de leurs débiteurs d'autre part (*Aghion P et Menger P-M*). Quelle qu'en soit la source, les paniques bancaires se produisent plus fréquemment et violemment lorsque le crédit s'est emballé dans la période précédente. Les raisons de ces paniques peuvent être multiples mais elles ont toutes un point commun : la montée de l'incertitude concernant l'activité économique en général.

Un choc subi par une banque a de très fortes chances de se transmettre rapidement aux autres car elles sont, du fait de la nature de leurs activités, tour à tour créancières les unes des autres. Par conséquent, quelle que soit l'origine de la panique qui touche la première banque, l'interconnexion très forte entre les banques, conduit à une contagion rapide d'une banque à l'autre. C'est ainsi qu'une panique bancaire peut conduire à des faillites bancaires en chaîne (*Aghion P et Menger P-M*).

5.1.4 Conséquences

La crise de 2008, comme celle des années 1930 et la plupart des crises financières, qui se manifestent notamment par la baisse du prix des actifs financiers, des faillites bancaires et la disparition de la confiance, a des conséquences sur l'économie réelle, entraînant notamment une baisse de la production et une montée du chômage.

En termes de prévention et de résolution des crises, les régulations sur les bilans des banques ont été mises en place au fur et à mesure des crises. Pendant les crises, les gouvernements prennent le contrôle en forçant des injections de liquidités ou des fusions par exemple.

5.2- La cas particulier Français au travers de la crise de 1929

La Grande Dépression ou crise économique des années 1930 va du krach de 1929 aux États-Unis jusqu'à la Seconde Guerre mondiale. Ces conséquences se sont étendues à l'échelle mondiale et ont perduré longtemps.

La France connaît une crise bancaire significative dont l'ampleur est aujourd'hui sujette à discussion. Elle se matérialise notamment par la chute de la Banque Nationale de Crédit, quatrième banque de dépôts du pays et seule grande banque du pays à avoir fait faillite. Les autres grands établissements bancaires ayant résisté à celle-ci donnent le sentiment que le secteur bancaire français n'a que peu été atteint par la crise mondiale et qu'il n'a lui-même pas été source de dérèglements économiques. Par ailleurs, la prédominance des banques de dépôts en France par rapport à l'étranger a participé à retarder la crise économique (*Vilpoux E*).

Les nombreuses faillites bancaires au cours de la période ont certes été repérées, mais leur rôle macroéconomique n'a pas été apprécié, notamment à cause de l'absence de statistiques capables de fournir une image complète de la crise. Les banques n'étant pas réglementées en France avant 1941, leurs bilans n'étaient ni enregistrés, ni centralisés par une autorité de contrôle. En l'absence de telles statistiques, la méthode habituelle de calcul des séries de crédits et de dépôts bancaires consistait à s'appuyer sur les bilans des quatre plus grandes banques commerciales et à supposer que ces banques représentaient de manière stable la moitié du secteur bancaire. Étant donné que ces quatre grandes banques n'ont pas connu de difficultés et que leurs dépôts n'ont pas diminué en 1930 et 1931, la littérature existante n'a vu aucun indice d'une crise bancaire majeure en France (*Baubeau P et al*).

6- Méthodologie

6.1 Objectifs

A partir des données disponibles, l'objectif était d'implémenter un algorithme de Machine Learning qui pourrait prédire les faillites bancaires en fonction de leurs bilans financiers.

La première étape, consistant en une recherche bibliographique, avait pour but de comprendre les mécanismes qui régissent le fonctionnement des crises bancaire et

plus particulièrement ceux de la crise de 1929 en France ainsi que d'observer les corrélations mises en évidence entre les bilans financiers et la santé des banques.

La deuxième étape avait pour but d'implémenter les algorithmes nécessaires pour pouvoir observer l'existence ou non de corrélations similaires sur ces données ainsi que de prédire les potentielles faillites de ces banques.

6.2 Etat de l'art et recherche bibliographique.

Lors de cette première partie de stage, je me suis concentré sur quatre publications principales portant à la fois sur les causes et les conséquences des crises ainsi que sur des recherches portant sur un sujet similaire au mien.

6.2.1 Financial Crises : Theory and evidence. (Allen F, Babus A, and Carletti E)

Dans cette revue, les auteurs détaillent les causes des crises bancaires. Il en ressort que les crises s'expliquent majoritairement par les paniques bancaires, ou des retraits de dépôts en chaîne placent les banques en défaut de liquidités et par les bulles spéculatives, faisant partie inhérente du cycle financier. Ce texte met également en avant l'importance du mécanisme de contagion, puisque le rôle du marché interbancaire est de redistribuer les liquidités entre celles-ci, les difficultés liées aux crises se propagent à l'ensemble des banques.

6.2.2 Banking Crises : A review. (Laeven L)

Revenant sur les causes à l'origine des crises bancaires, ce texte publié dans « The Annual Review of Financial Economics, volume 3 », s'attarde principalement sur les conséquences des crises ainsi que sur les tentatives de préventions et de résolutions de celles-ci. Il y est expliqué que de par leur place primordiale au cœur du système financier notamment en tant que principale source de financement pour les entreprises, les crises bancaires affectent toute l'économie mais également les particuliers au-delà de l'augmentation du taux de chômage par exemple. Malgré les interventions gouvernementales, obligeant les fusions par exemple, en tant de crise, les conséquences restent massives.

Ainsi pour tenter de d'empêcher les futures crises, diverses réglementations sur les activités bancaires ont été mises en place au fil des différentes crises. Que ce soit par des taux de fonds propres minimaux ou par la limitation de certaines activités bancaires,

ces réglementations laissent certaines opportunités aux banques de les contourner limitant ainsi leur efficacité.

6.2.3 How does capital affect bank performance during financial crises? (Berger A N & Bouwman CHS)

Cet article met en évidence une corrélation entre la quantité de fonds propres dont dispose une banque et sa capacité à survivre à une crise. À l'aide d'une régression logistique, les auteurs ont montré qu'un au taux de fonds propres agissant favorablement sur la survie d'une banque en temps de crises et que ce facteur était moins impactant sur les très grosses banques. En plus de prendre en compte les périodes étudiées de manières distinctes, ils permettent de valider la pertinence des réglementations au niveau des ratios de fonds propres.

6.2.4 La crise bancaire de 1930 en France : une nouvelle analyse historique et quantitative des données du Crédit Lyonnais. (Vilpoux E)

Dans cette thèse à laquelle mon stage fait suite, l'auteur étudie la période de la crise de 1929 en France. Il y met notamment en avant le manque de données relatives à cette période pour ce pays ainsi que la littérature relativement peu abondante sur le sujet notamment dues aux manques de réglementations sur les banques françaises de cette époque.

À l'aide des données présentes dans l'album du Crédit Lyonnais, l'auteur a observé que la crise de 1929 en France était bien présente contrairement à ce que veut une idée répandue selon laquelle le secteur bancaire français aurait mieux résisté à la crise que les autres.

Avec une régression logistique et une régression probit, il établit à partir des bilans à la veille de la crise qu'un fort ratio de liquidités implique une plus grosse probabilité de survie.

6.3 Base de données

Les données relatives à ce contexte précis se résument à un album publié par le Crédit Lyonnais, il fallait que ces données soient homogénéisées afin que les données aient toutes la même signification économique. Une fois ces données traitées elles présentent des informations facilement utilisables comme diverses valeurs aux passifs

ou aux actifs des bilans, ainsi que des variables indiquant par exemple si une banque fait partie d'une des plus grosses banques du pays ou s'il s'agit d'une banque populaire.

Elle présente également des variables indiquant si une banque fait faillite, f1, f2 et f3 respectivement égales à un si une banque fait faillite ou subit une liquidation judiciaire, si elle subit une fusion ou une dissolution, si une banque est retirée de l'album sans qu'une raison soit mentionnée.

Afin d'analyser ces données j'ai utilisé plusieurs ratios de capitaux et de liquidités définis comme tel :

$$\text{Ratio de capital : } \frac{\text{Capital + Réserves}}{\text{Actifs Totaux}}$$

$$\text{Ratio de liquidité 1 : } \frac{\text{Cash + Portefeuille Commercial}}{\text{Actifs Totaux}}$$

$$\text{Ratio de liquidité 2 : } \frac{\text{Total des Dépôts}}{\text{Actifs Totaux}}$$

$$\text{Ratio de liquidité 3 : } \frac{\text{Total des Crédits à court terme}}{\text{Actifs Totaux}}$$

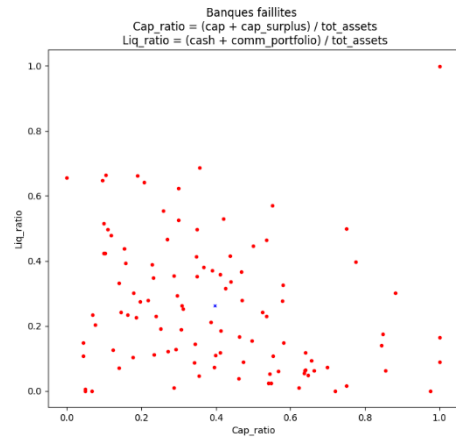
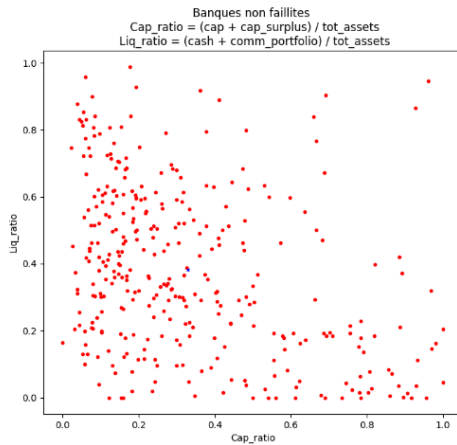
6.4 Choix du model

Dans un premier temps, il convient d'analyser la problématique afin de déterminer quel model il faut implémenter. Premièrement comme nous disposons de données pour entrainer notre algorithme il s'agira d'un model supervisé. Deuxièmement, c'est un problème de classification. Le but du model serait d'attribuer une classe, ici faillite ou non faillite, à chaque banque en fonction de ses bilans financiers.

Il existe plusieurs algorithmes pour résoudre ce genre de problème.

6.4.1 Kmean

Dans un premier temps j'ai mis en place un algorithme « K mean » pour voir s'il existait une corrélation évidente entre les ratios de liquidités et de capital et les chances de faillite bancaire. Malheureusement cet algorithme ne parvenait pas à établir de lien direct et les données étaient réparties de manières hétérogène comme nous pouvons le voir sur les graphiques suivants :

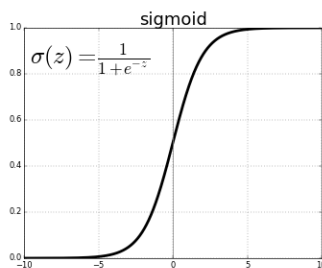


Ici par exemple nous avons la répartition des banques en faillites et en non-faillites en fonctions du ratio de capital et du ratio de liquidité 1. Il en était de même avec les autres ratios de liquidité et les autres faillites.

6.4.2 Régression logistique

Un autre algorithme répondant au problème de classification est la régression logistique. En plus d'être un des plus utilisés, il est aussi par E. Vilpoux dans ses travaux. Il me semblait donc être une bonne option.

Cet algorithme consiste à paramétrer la probabilité que la variable de sortie qui nous intéresse (ici f_1, f_2, f_3) en fonction de nos variables. A chaque variable est attribué un poids. La somme de ces variables multipliée par leur poids respectifs donne une valeur continue. Afin que cette valeur soit comprise entre 0 et 1, une fonction d'activation est utilisée, ici la fonction sigmoïde.



Ensuite grâce à un seuil de décision, ces valeurs continues sont converties en valeurs binaires : si après la fonction d'activation la valeur est strictement supérieure à 0.5 alors la valeur 1 est attribuée.

L'algorithme est entraîné sur une partie des données appelée *train set* et ses performances sont testées sur un *test set*.

Alors que précédemment, l'analyse des ratios a été effectuée pour seulement l'année en cours, j'ai implémenté des séries temporelles afin de prendre en compte plusieurs années en amont pour disposer de plus de variables.

Là où dans un modèle de régression linéaire nous pourrions utiliser la précision pour attester de la performance d'un modèle, cela n'est pas forcément pertinent pour un problème de classification. Pour évaluer les performances des modèles, j'ai donc utilisé des indicateurs de performance qui sont fréquemment utilisés dans des problèmes de classification. Ils se définissent comme tel dans notre cas :

Recall : Proportion des banques en faillites correctement classées.

Precision : proportion des banques classées en faillites l'étant réellement.

F1 : Une combinaison des deux

6.5 Etablissement du modèle

L'idée des séries temporelles est de pouvoir augmenter le nombre de variables et ainsi faciliter l'apprentissage de l'algorithme. En prenant les ratios pour i années précédant l'année étudiée on obtenait en effet de meilleurs résultats. Par exemple pour une banque ne faisant pas faillite en 1930, nous avons les i ratios de capital et les i ratios de liquidités qui la précèdent. Ici avec les 6 années précédentes par exemple :

0,178472	0,169044	0,165631	0,154712	0,139653	0,154427	0,451757	0,433175	0,512422	0,431317	0,431655	0,425532	0	0	0
0,169044	0,165631	0,154712	0,139653	0,154427	0,174689	0,433175	0,512422	0,431317	0,431655	0,425532	0,423137	0	0	0
0,165631	0,154712	0,139653	0,154427	0,174689	0,176583	0,512422	0,431317	0,431655	0,425532	0,423137	0,353167	0	0	0
0,154712	0,139653	0,154427	0,174689	0,176583	0,169903	0,431317	0,431655	0,425532	0,423137	0,353167	0,412621	0	0	0
0,139653	0,154427	0,174689	0,176583	0,169903	0,343243	0,431655	0,425532	0,423137	0,353167	0,412621	0,31379	0	0	0
0,154427	0,174689	0,176583	0,169903	0,343243	0,354055	0,425532	0,423137	0,353167	0,412621	0,31379	0,327494	0	0	0

6.5.1 Problèmes

Plusieurs problèmes furent mis en évidence au cours de l'implémentation de l'algorithme. Dans un premier temps du nombre d'entrées était inversement proportionnel à la taille des séries temporelles. En effet, pour certaines banques, il n'était pas possible de concevoir des séries de taille i , je les ai donc enlevées du jeu de données. Comme il se trouve que si un modèle a trop de variables par rapport à son

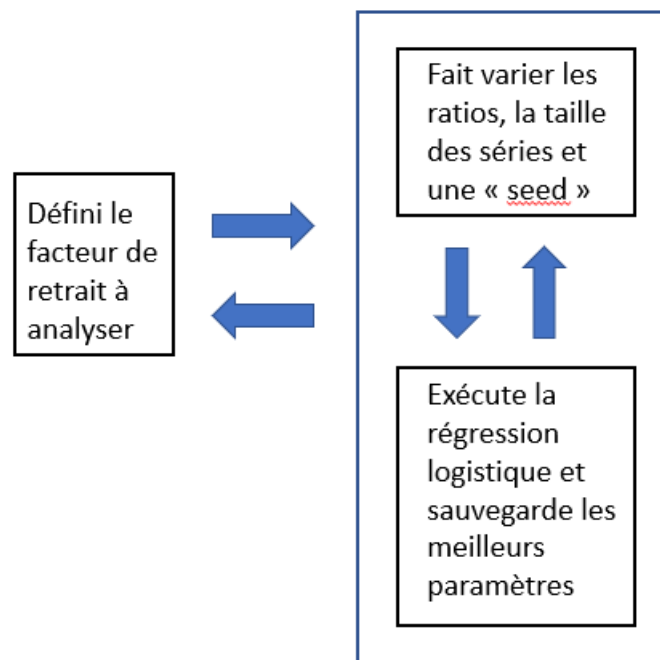
nombre d'entrées ou qu'il a trop d'entrées par rapport à son nombre de variables, l'algorithme aurait vite pu peiner à apprendre.

De plus il se trouve que la proportion de banques faisant faillites est extrêmement faible. Alors il aurait suffi à l'algorithme de toujours prédire une non-faillite pour obtenir des scores de précision élevée. Ainsi cela aurait faussé son apprentissage.

6.5.2 Solution

L'idée est donc de faire varier la taille des séries temporelles, le ratio de liquidités utilisé comme variable d'entrée, la variable de sortie que l'on souhaite prédire ainsi que la répartition des données dans le train-set et le test-set à l'aide d'une « seed » paramétrant la répartition aléatoire. Tout ceci se fait au travers de boucles qui conservent les paramètres ayant obtenus les meilleurs résultats.

Afin de contrôler la répartition dans les train- et test-set, j'ai implémenté l'algorithme de telle sorte qu'une proportion élevée de banques en faillite soient présentes dans le train-set et que toutes les autres en faillite dans le test set. J'ai également réparti les six banques dites « to big to fail » à raison de quatre dans le train-set et de 2 dans le test set. Ensuite les deux sets sont remplis aléatoirement avec les banques restantes jusqu'à atteindre un rapport de taille 70/30 entre le train-set et le test-set.



6.6 Bibliothèques et implémentation

Dans le but d'implémenter cette solution en langage Python, j'ai utilisé diverses bibliothèques :

Scikit-Learn : Cette bibliothèque python offre plusieurs outils pour le Machine Learning ainsi que diverses analyses statistiques. Pour ce travail, j'ai utilisé l'algorithme de régression logistique de cette bibliothèque.

Pyplot : Elle permet l'affichage de multiples graphes et je l'ai utilisé pour monter les performances de l'algorithme que j'ai implémenté

Pandas : Sûrement la bibliothèque qui permet de manipuler des données avec le plus d'efficacité et de simplicité. Je l'ai utilisé pour tous les traitements des données notamment pour la définition des ratios et des séries temporelles.

Le code final se présente sous la forme de méthodes utilisées par des scripts afin d'entraîner les modèles et d'en afficher les performances.

Comme il faut que le code soit réutilisable et améliorable par une autre personne, il convenait donc de coder les divers fichiers python de manière claire et les organiser de manière à ce que cela soit le plus simple possible. J'ai également établi un manuel qui explique le rôle de chaque méthode python que j'ai défini et des scripts.

A titre personnel, il me semble intéressant de noter que cet aspect du stage s'est en fait révélé coriace dans le sens où de par mon expérience majoritairement scolaire je n'ai jamais eu à penser de cette manière. Bien que déboussolant pour ma manière de coder, j'en garde une expérience enrichissante qui me permettra à l'avenir de coder différemment et sûrement plus efficacement.

7- Résultats

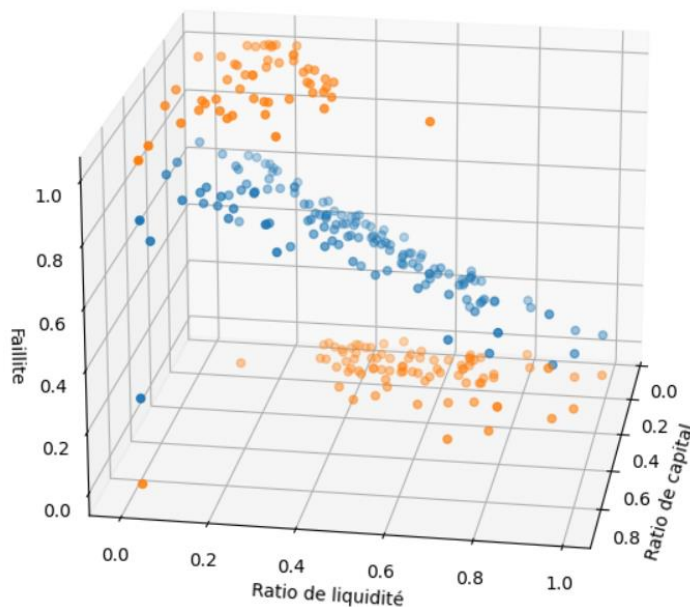
Une fois l'algorithme exécuté, il en ressort que peu importe la variable de sortie étudiée, c'est toujours le ratio de liquidités, défini comme la somme du cash et du portefeuille commercial sur les actifs totaux, qui permet d'obtenir les meilleurs résultats.

Dans le cas présenté ci-dessous il y a une proportion de 44.5% de banques faisant faillites dans le train-set et de 17.3% dans le test-set. L'algorithme arrive donc bien à répartir les banques de manière à obtenir une proportion de banques faisant faillite élevée dans le train-set.

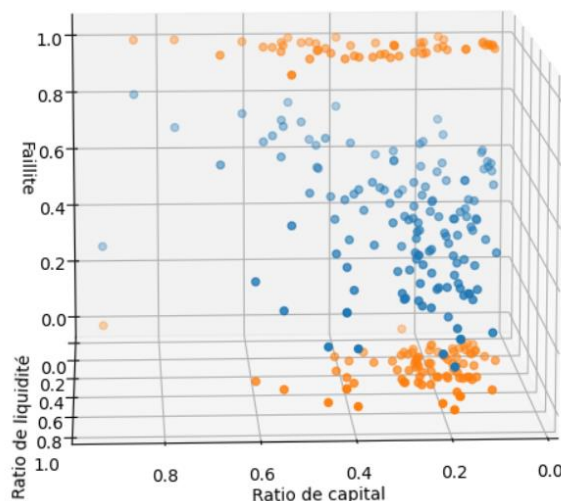
Les meilleurs résultats sont ceux obtenus en essayant de prédire f_1 et sont les suivant :

Sur les graphiques en 3 dimensions présentés, nous pouvons observer les faillites réelles (en orange) et la probabilité de faillite estimée par l'algorithme (en bleu), en fonction des ratios de liquidités et de capital.

Dans le graphique ci-dessous, nous pouvons observer qu'une corrélation entre le ratio de liquidités et les faillites semble se dessiner et que l'algorithme semble également suivre cette tendance



A l'inverse, nous n'observons pas de corrélation évidente entre le ratio de capital et les chances de faillite, comme présenté ci-dessous.



De plus, l'algorithme obtient un score f1 de 0.653, ce qui peut sembler relativement faible mais en contrepartie, il obtient un score de recalé de 1 ce qui met en avant sa capacité conservatrice et prudente.

		Predicted	
		Actual Negative	Actual Positive
Actual class	Actual Negative	True Negative 28	False Positive 10
	Actual Positive	False Negative 0	True Positive 8

Ici nous avons la matrice de confusion des prédictions du model sur le train-test.

8- Conclusion et perspectives.

Les résultats obtenus réussissent à mettre en évidence une corrélation bien connue du monde de la finance, celle entre les ratios de liquidités et les chances de faillites. Néanmoins, si le ratio de liquidité 1 permet à l'algorithme d'obtenir de meilleurs résultats, rien ne dit que ce soit la meilleure définition du ratio de liquidités.

En termes de pistes d'amélioration, je pense tout d'abord à valider ou non la pertinence du modèle choisi en le testant sur des jeux de données différents. L'intérêt d'ajouter plus de données permettrait de favoriser la précision du modèle. En effet, comme on limite le train-set à une certaine proportion de banques en faillites, on réduit la taille des sets. Ainsi ajouter plus de données bancaires relatives à des banques en faillites pourrait permettre d'ajouter plus de « features » sans que le modèle s'en retrouve affaibli.

Encore une fois, dans l'optique d'augmenter le nombre d'entrées, on pourrait simuler l'évolution des différents ratios sur diverses périodes afin d'avoir par exemple plusieurs bilans par année. Etant donné que beaucoup de choses peuvent se passer en un an, cette approche me semble pertinente.

De plus il faudrait permettre à l'algorithme de tester différentes répartitions de banques en faillites dans le train-set et le test-set.

Toujours dans l'optique d'améliorer ce modèle, la mise en place de variables de contrôles comme celles qui indiquent la localisation géographique des banques, une banque est située à Paris ou si elle est une banque régionale par exemple, pourrait permettre d'améliorer celui-ci.

On pourrait également choisir d'autres algorithmes de classification, notamment un arbre de décision qui permettrait de facilement intégrer des variables de contrôle que je n'ai pas eu le temps d'implémenter dans mon modèle.

Il me semble intéressant de noter que les 360 + itérations de l'algorithme prennent presque deux heures pour s'exécuter. L'accélération de l'algorithme semble être aussi un point crucial dans une potentielle amélioration future de celui-ci.

Même si un des objectifs à long terme de cette étude aurait été de mettre en place un réseau de neurones, qui sont connus pour obtenir de meilleurs résultats que les régressions simples, dans le cas ici présent il n'est pas sûr que ce soit possible étant donné le peu de données disponibles.

9- Références Bibliographiques

- Aghion P et Menger P-M, "Campus de l'Innovation pour les Lycées", Comment expliquer les crises financières et réguler le système financier ?)
- Allen, F., Babus, A., and Carletti, E. (2009), Financial crises: theory and evidence, Annual Review of Financial Economics, vol. 1, pp. 97-116
- Baubeau, P., Monet, E., Riva, A., and Ungaro, S. (2021), Flight-to-safety and the credit crunch: a new history of the banking crises in France during the Great Depression, The Economic History Review, vol. 74 no. 1, pp. 223-250
- Berger, A. N., and Bouwman, C.H.S. (2013), How does capital affect bank performance during financial crises?, Journal of Financial Economics, vol. 109, pp 146-176
- Laeven, L. (2011), Banking crises: A review, The Annual Review of Financial Economics, vol. 3, no.1, pp. 17-40
- Vilpoux, E. (2017), La crise bancaire de 1930 en France : une nouvelle analyse historique et quantitative des données du Crédit Lyonnais, Master Thesis, Paris School of Economics