

Université du Québec à Montréal (UQAM)
Département d'informatique

Projet final – DIC9315
**Systèmes neuro-flous pour la prédiction du
risque de crédit**

Présenté par :
Abdoulaye Baldé

Professeur :
BOUKADOUM, A. Mounir

Session : Automne 2025

Table des matières

1	Introduction	2
2	Revue de littérature	4
2.1	Article 1 : ANFIS et risques bancaires	4
2.2	Article 2 : Machine learning responsable pour le crédit	5
2.3	Article 3 : Réseaux de neurones graphiques hybrides	6
2.4	Article 4 : Gen-AI versus modèles traditionnels	6
2.5	Synthèse et positionnement	7
3	Méthodologie expérimentale	9
3.1	Jeu de données	9
3.2	Prétraitement des données	9
3.3	Modèles étudiés	9
3.4	Configuration expérimentale	9
4	Résultats	10
4.1	Comparaison des performances	10
4.2	Analyse des erreurs	10
4.3	Interprétabilité des modèles	10
5	Discussion	11
6	Conclusion	11

1 Introduction

Le risque de crédit est un des enjeux les plus importants dans le domaine financier contemporain. Les institutions comme les banques, les coopératives d'épargne, ou encore les entreprises de micro-finance doivent constamment évaluer la probabilité qu'un client ne respecte pas ses engagements financiers. Cette problématique est particulièrement critique dans un contexte économique marqué par l'incertitude, la hausse des taux d'intérêt et la multiplication des produits financiers. Selon plusieurs auteurs, une mauvaise évaluation du risque peut entraîner des pertes majeures, voire des crises financières à grande échelle. Pour cette raison, le développement de modèles de prédiction fiables est devenu un sujet central autant dans la recherche académique que dans l'industrie.

Dans ce travail, nous nous intéressons à la prédiction du risque de crédit par l'utilisation d'un système neuro-flou, plus précisément le modèle ANFIS (Adaptive Neuro-Fuzzy Inference System). L'un des aspects intéressants de ce paradigme est qu'il combine deux approches habituellement séparées : d'un côté les réseaux de neurones, reconnus pour leur capacité d'apprentissage, et de l'autre la logique floue, qui permet de modéliser l'incertitude et le raisonnement linguistique humain. Comme le souligne MEHDI et al. (2025), les systèmes neuro-flous peuvent offrir une interprétabilité plus élevée que des modèles purement neuronaux, tout en conservant un bon niveau de performance prédictive dans des contextes financiers.

La motivation de ce projet vient aussi de mon expérience dans plusieurs compétitions de data science, notamment sur Kaggle et Zindi Africa, où j'ai eu l'occasion de travailler sur différents problèmes de classification et de modélisation du risque. À cela s'ajoute ma situation professionnelle actuelle, puisque je dois réaliser un stage en analyse du risque financier chez Desjardins. Cette expérience m'a déjà permis d'observer concrètement certains défis auxquels les institutions financières font face quand il s'agit de prédire le défaut de paiement.

Dans le contexte réel, un modèle ne doit pas seulement être performant : il doit aussi être explicable, stable et compréhensible, car il s'inscrit dans un environnement très réglementé. Par exemple, des modèles trop complexes, comme certaines architectures profondes de deep learning, peuvent obtenir de bons résultats mais deviennent difficiles à justifier auprès des équipes opérationnelles, des gestionnaires de risque ou même des régulateurs. À l'inverse, des approches plus transparentes, comme les systèmes neuro-flous, offrent un compromis intéressant entre performance et interprétabilité. Ce projet s'inscrit donc dans l'idée d'explorer des modèles qui pourraient répondre à la fois aux exigences pratiques d'une institution financière et aux besoins de performance dans un cadre académique.

Pour mieux situer notre approche, nous nous appuyons aussi sur les recommandations méthodologiques discutées par VALDRIGHI et al. (2024), qui insistent sur la nécessité d'utiliser des modèles responsables, équitables et bien calibrés dans le domaine du credit scoring. Le travail de ces auteurs nous rappelle que la prédiction du défaut de paiement n'est pas uniquement un problème mathématique, mais touche aussi des considérations éthiques, comme l'équité envers les clients ou la protection contre les biais dans les données. Cela renforce l'importance de comparer différents modèles dans notre étude et de comprendre en profondeur leurs comportements.

Dans ce projet, nous allons donc comparer la performance du modèle ANFIS à deux

autres modèles classiques du machine learning utilisés dans l'évaluation du risque de crédit : Random Forest et Support Vector Machine (SVM). Random Forest est souvent considéré comme une référence robuste et relativement facile à expliquer, ce qui le rend très prisé dans les institutions financières. Le SVM, quant à lui, est un modèle puissant pour la classification binaire, particulièrement efficace lorsque les frontières entre les classes sont complexes. Ces deux modèles serviront de points de comparaison pour évaluer si ANFIS peut constituer une alternative solide dans ce type de problème.

Le dataset utilisé pour nos expérimentations est le *UCI Credit Card Default Dataset*, un jeu de données bien connu qui comporte les informations de 30 000 clients de cartes de crédit à Taïwan. Ce dataset est largement utilisé dans la recherche en credit scoring, ce qui permet de comparer nos résultats à ceux obtenus dans la littérature. Son utilisation est également motivée par sa diversité de variables financières et comportementales, ce qui en fait un environnement approprié pour tester des modèles basés sur des règles floues.

Ainsi, les principaux objectifs de ce projet sont les suivants :

- étudier en profondeur le fonctionnement du modèle ANFIS et son application au risque de crédit ;
- analyser la littérature scientifique récente sur les approches ANFIS, GNN et Gen-AI appliquées à la finance ;
- développer un pipeline expérimental permettant de comparer ANFIS, Random Forest et SVM dans un cadre équitable ;
- évaluer la performance des différents modèles selon plusieurs métriques et analyser leurs forces et limites ;
- proposer une discussion critique sur la pertinence des systèmes neuro-flous dans un contexte bancaire moderne.

En résumé, ce travail vise à mieux comprendre comment les systèmes neuro-flous peuvent être utilisés pour prédire le risque de crédit et comment ils se comparent à des modèles plus traditionnels ou plus récents. L'approche adoptée cherche à être à la fois technique et pratique, en prenant en compte non seulement les performances numériques, mais aussi les considérations d'interprétabilité et d'applicabilité réelle dans une institution financière. Cette introduction pose ainsi les bases pour les sections suivantes, qui détailleront la revue des articles, la méthodologie expérimentale et les résultats obtenus.

Les principaux objectifs du projet sont :

- Étudier l'approche ANFIS dans un contexte de prédiction du risque de défaut.
- Comparer ANFIS à des modèles classiques (Random Forest, SVM).
- Analyser les performances obtenues sur un dataset financier réel.
- Discuter des avantages et limites des systèmes neuro-flous.

2 Revue de littérature

Cette section présente quatre articles récents (2024-2025) sur la prédiction du risque de crédit par l'intelligence artificielle. J'ai choisi ces articles pour couvrir différents paradigmes : les systèmes neuro-flous (ANFIS), les questions d'éthique en ML, les réseaux de neurones graphiques, et les modèles génératifs. Pour chaque article, je vais résumer leur contribution, identifier les points forts et limites, et expliquer le lien avec mon projet.

Le problème général du credit scoring peut se formuler mathématiquement comme un problème de classification binaire. Soit $\mathbf{x} = (x_1, x_2, \dots, x_n)$ un vecteur de caractéristiques financières et démographiques d'un client. L'objectif est de prédire $y \in \{0, 1\}$ où :

$$y = \begin{cases} 1 & \text{si défaut de paiement} \\ 0 & \text{si remboursement normal} \end{cases} \quad (1)$$

On cherche donc une fonction $f : R^n \rightarrow \{0, 1\}$ qui minimise l'erreur de prédiction sur un ensemble de données.

2.1 Article 1 : ANFIS et risques bancaires

MEHDI et al. (2025) utilisent ANFIS pour mesurer l'impact des risques bancaires (crédit, liquidité, capital) sur la performance de 62 banques du Moyen-Orient et d'Afrique entre 2013 et 2022. Leur modèle analyse comment sept variables de risque (NPL, PLR, LDR, LAR, CAR, TIER1, ETC) influencent trois indicateurs : le ROE, l'EPS et le PER.

Leurs résultats montrent que le risque de crédit (NPL) a l'impact le plus fort sur le ROE et l'EPS, tandis que le risque de liquidité domine pour le PER. C'est intéressant parce que ça montre que différents types de risques affectent différemment les mesures de performance.

Points forts : L'approche holistique qui considère plusieurs types de risques simultanément est un vrai plus. ANFIS combine bien l'apprentissage neuronal et la logique floue, ce qui devrait offrir plus d'interprétabilité qu'un réseau de neurones pur.

Le modèle ANFIS fonctionne avec des règles floues de type Takagi-Sugeno. Pour un système avec deux entrées x_1 et x_2 , une règle typique s'écrit :

$$\text{IF } x_1 \text{ est } A_i \text{ AND } x_2 \text{ est } B_j \text{ THEN } f_{ij} = p_{ij}x_1 + q_{ij}x_2 + r_{ij} \quad (2)$$

où A_i et B_j sont des ensembles flous définis par des fonctions d'appartenance $\mu_{A_i}(x_1)$ et $\mu_{B_j}(x_2)$. La sortie finale est calculée comme une moyenne pondérée :

$$y = \frac{\sum_{i,j} w_{ij} \cdot f_{ij}}{\sum_{i,j} w_{ij}} \quad \text{avec} \quad w_{ij} = \mu_{A_i}(x_1) \times \mu_{B_j}(x_2) \quad (3)$$

Leur méthode de perturbation pour calculer les sensibilités est robuste. Les résultats sont solides avec un VEcv (variance expliquée en validation croisée) supérieur à 88%.

Limites : Ils ne montrent pas vraiment les règles floues générées, ce qui réduit l'avantage d'interprétabilité. Surtout, ils ne comparent pas ANFIS avec d'autres méthodes comme Random Forest ou SVM sur leur propre dataset, donc c'est difficile d'évaluer la vraie valeur

ajoutée. La généralisation est aussi limitée vu qu'ils se concentrent sur une région spécifique. Finalement, ils ne considèrent pas vraiment l'aspect temporel des données.

Lien avec mon projet : Cet article justifie mon choix d'utiliser ANFIS pour le crédit. Par contre, moi je vais travailler au niveau individuel (risque de défaut des clients) plutôt qu'au niveau institutionnel. Contrairement à eux, je vais comparer ANFIS avec Random Forest et SVM pour vraiment évaluer ses performances relatives.

2.2 Article 2 : Machine learning responsable pour le crédit

VALDRIGHI et al. (2024) proposent un tutorial complet sur les meilleures pratiques pour développer des modèles ML responsables en credit scoring. Ils couvrent trois piliers : la fairness (équité), le reject inference, et l'explainability.

Sur la fairness, ils expliquent comment les modèles peuvent perpétuer des biais même sans utiliser directement les attributs sensibles. Ils présentent différentes définitions de l'équité. Par exemple, la *demographic parity* exige que :

$$P(\hat{y} = 1|A = 0) = P(\hat{y} = 1|A = 1) \quad (4)$$

où A est un attribut sensible (genre, ethnie, etc.) et \hat{y} est la prédiction. Une autre métrique importante est l'*equalized odds* qui exige :

$$P(\hat{y} = 1|y = k, A = 0) = P(\hat{y} = 1|y = k, A = 1) \quad \forall k \in \{0, 1\} \quad (5)$$

Ils présentent les méthodes pour atténuer les biais (pré-traitement, in-processing, post-traitement).

Pour le reject inference, ils traitent du problème souvent négligé que les modèles sont entraînés seulement sur les clients acceptés, créant un biais de sélection. Ils présentent des techniques pour incorporer les demandeurs rejetés mais reconnaissent que c'est un problème difficile sans solution parfaite.

Sur l'explainability, ils distinguent l'interprétabilité globale et locale, et présentent des techniques comme SHAP, LIME et les counterfactual explanations.

Points forts : Le caractère pédagogique et pratique avec du code Python sur GitHub est excellent. Ils sont honnêtes sur les compromis entre performance et équité. Le traitement du reject inference est particulièrement bien fait pour un sujet souvent ignoré.

Limites : C'est un travail de synthèse sans contributions méthodologiques nouvelles. Les expériences sont basiques sans comparaisons systématiques pour guider le choix des méthodes. Les aspects réglementaires restent superficiels. Pas de discussion sur les coûts et la faisabilité opérationnelle. Les datasets utilisés sont trop propres comparés aux données bancaires réelles.

Lien avec mon projet : Cet article me rappelle que maximiser l'accuracy n'est pas suffisant. Je dois aussi vérifier l'équité de mes modèles et leur explicabilité. Je pourrais utiliser SHAP pour comparer l'interprétabilité de mes trois modèles. Je dois aussi mentionner le biais de sélection dans mon dataset UCI.

2.3 Article 3 : Réseaux de neurones graphiques hybrides

XU et al. (2024) proposent des Graph Neural Networks hybrides pour le risque de crédit. Leur idée est que les emprunteurs sont liés par des relations (familiales, professionnelles, géographiques) qui contiennent des infos utiles. Ils transforment les données en graphe où les nœuds sont des emprunteurs et les arêtes sont des relations.

Leur architecture hybride combine des convolutions locales (sous-graphes) et globales (graphe entier) pour éviter le over-smoothing et mieux capturer les features. Pour un graphe $G = (V, E)$ avec matrice d'adjacence A et features de nœuds X , une couche de convolution graphique se formule comme :

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)} \right) \quad (6)$$

où $\tilde{A} = A + I$ (on ajoute les self-loops), \tilde{D} est la matrice de degré, $W^{(l)}$ sont les poids apprenables, et σ est une fonction d'activation.

Ils utilisent aussi un mécanisme d'attention pour identifier les relations les plus informatives, avec des poids d'attention α_{ij} entre nœuds i et j :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (7)$$

où e_{ij} mesure l'importance de la relation entre i et j .

Sur un dataset de 3000 emprunteurs (2500 non-défaillants, 500 défaillants), leur modèle atteint 68.90% d'accuracy, surpassant SVM (62.72%) et les arbres de décision (64.84%).

Points forts : La perspective relationnelle fait du sens – si mes voisins font défaut, ça indique peut-être des conditions locales défavorables. L'architecture multi-échelle est élégante. Les résultats expérimentaux sont convaincants. Le mécanisme d'attention pour sélectionner les relations importantes est une bonne idée.

Limites : La construction du graphe n'est pas bien détaillée et les infos relationnelles sont difficiles à obtenir en pratique. La scalabilité computationnelle pose problème pour de gros portefeuilles. L'interprétabilité est faible comparé aux modèles plus simples, ce qui est problématique pour la régulation. Le dataset est petit (3000 observations) et déséquilibré. Pas d'analyse d'ablation pour quantifier l'apport de chaque composante.

Lien avec mon projet : Les GNN représentent le cutting edge mais souffrent de problèmes d'interprétabilité et de complexité. Mon dataset UCI n'a pas d'infos relationnelles explicites. Cet article renforce la pertinence d'étudier ANFIS qui offre un meilleur compromis entre performance et transparence.

2.4 Article 4 : Gen-AI versus modèles traditionnels

LAVECCHIA et al. (2025) comparent les modèles d'IA générative (GPT-4) avec les méthodes traditionnelles pour le credit scoring. Sur le dataset German Credit, ils testent la régression logistique, les réseaux de neurones, Random Forest, Gradient Boosting, et GPT-4o avec différentes stratégies de prompting.

Les résultats sont clairs : GenAI (AUC 0.65-0.71) sous-performe significativement par rapport aux méthodes traditionnelles (régression logistique 0.76, Gradient Boosting 0.79).

L'AUC-ROC (Area Under the Receiver Operating Characteristic curve) mesure la capacité du modèle à distinguer les classes :

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(t)) dt \quad (8)$$

où TPR est le taux de vrais positifs et FPR le taux de faux positifs. D'autres métriques importantes incluent :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

Même en variant le nombre d'exemples (3 à 30) et la stratégie de sélection, l'écart persiste. Par contre, GenAI produit des prédictions légèrement plus équitables (score BRIO).

Points forts : Le pragmatisme et l'honnêteté dans une période d'enthousiasme excessif autour de GenAI. Méthodologie exhaustive testant différentes configurations. Évaluation multidimensionnelle incluant l'équité. Perspective constructive suggérant GenAI comme outil de preprocessing. Discussion rare sur les coûts d'utilisation.

Limites : Un seul dataset (German Credit, 1000 observations). Limité à GPT-4o sans tester d'autres LLM. N'explore pas vraiment les approches hybrides sophistiquées. Les LLM progressent vite donc les résultats pourraient changer rapidement. L'analyse d'équité reste superficielle.

Lien avec mon projet : Cet article montre que "nouveau" ne signifie pas "meilleur". Ça renforce la pertinence d'étudier ANFIS (années 90) face aux méthodes établies. Je dois être prudent et honnête dans mes conclusions. Même si ANFIS ne surpasse pas Random Forest ou SVM, sa simplicité et son interprétabilité peuvent compenser.

2.5 Synthèse et positionnement

Ces quatre articles couvrent un spectre large de la recherche actuelle en credit scoring. Plusieurs thèmes ressortent qui m'aident à positionner mon projet.

Le spectre performance-interprétabilité : Les GNN représentent "haute performance, faible interprétabilité" tandis que la régression logistique est "performance modérée, haute interprétabilité". ANFIS se situe au milieu avec sa capacité à capturer des relations non-linéaires tout en offrant des règles floues interprétables. Cette position intermédiaire pourrait être un "sweet spot" pour des applications pratiques.

L'importance du contexte : Il n'y a pas de "meilleur" modèle universel. Pour MEHDI et al. (2025), ANFIS marche bien au niveau institutionnel. Pour XU et al. (2024), les GNN sont appropriés avec des données relationnelles. Pour LAVECCHIA et al. (2025), les méthodes traditionnelles restent supérieures en production. Mon projet va contribuer à comprendre dans quelles conditions ANFIS excelle ou échoue.

Les enjeux éthiques : VALDRIGHI et al. (2024) rappellent qu'un modèle doit satisfaire pas juste des critères de performance mais aussi d'équité et d'explicabilité. Je dois donc évaluer mes modèles sur plusieurs dimensions, pas seulement l'AUC ou l'accuracy.

L'innovation technologique : LAVECCHIA et al. (2025) montrent que GenAI sous-performe face à des méthodes décennales. La nouveauté technologique ne garantit pas la

supériorité pratique. Je ne dois pas présupposer qu'ANFIS va nécessairement surpasser des approches plus simples – les résultats empiriques seuls pourront trancher.

Ma contribution : Mon projet s'inscrit dans ces conversations en fournissant une comparaison rigoureuse d'ANFIS face à Random Forest et SVM sur le dataset UCI Credit Card Default. Cette comparaison, avec des tests d'équité et des analyses d'explicabilité, devrait apporter des insights utiles pour les praticiens cherchant à sélectionner un modèle adapté à leurs besoins.

3 Méthodologie expérimentale

Cette section détaille l’approche expérimentale utilisée pour comparer trois modèles : ANFIS, Random Forest, et Support Vector Machine (SVM).

3.1 Jeu de données

Description du UCI Credit Card Default Dataset :

- 30 000 clients ;
- Données démographiques et financières ;
- Variable cible binaire (défaut vs non-défaut).

3.2 Prétraitement des données

Les étapes incluent :

- Nettoyage des valeurs aberrantes ;
- Normalisation des variables continues ;
- Séparation des ensembles train/test ;
- Gestion du déséquilibre (SMOTE ou pondération).

3.3 Modèles étudiés

Random Forest (baseline)

Modèle robuste utilisé comme référence.

Support Vector Machine (SVM)

Modèle classique performant pour la classification non linéaire avec noyau RBF.

ANFIS (modèle étudié)

- Définition des fonctions d’appartenance ;
- Construction des règles floues ;
- Entraînement hybride (descente + moindres carrés).

3.4 Configuration expérimentale

- Hyperparamètres optimisés par validation croisée ;
- Métriques : accuracy, F1-score, AUC-ROC ;
- Environnement : Python, scikit-learn, scikit-fuzzy.

4 Résultats

Les résultats comparent les trois modèles selon plusieurs métriques.

- Tableaux comparatifs des scores (RF vs SVM vs ANFIS).
- Courbes ROC pour les trois modèles.
- Matrices de confusion détaillées.
- Visualisation des fonctions d'appartenance ANFIS.
- Analyse critique et interprétabilité.

Des sous-sections pourront inclure :

4.1 Comparaison des performances

4.2 Analyse des erreurs

4.3 Interprétabilité des modèles

5 Discussion

Cette section discute :

- des implications pratiques pour le scoring de crédit ;
- des forces et limites de chaque paradigme ;
- du compromis entre performance et interprétabilité ;
- de la pertinence de ANFIS dans un contexte bancaire moderne.

6 Conclusion

La conclusion résume les résultats expérimentaux, discute des contributions du projet, et ouvre vers des pistes de recherche futures (autres datasets, modèles hybrides, deep ANFIS, etc.).

Références

Références

- LAVECCHIA, N., FADANELLI, S., RICCIUTI, F., ALOE, G., BAGLI, E., GIUFFRIDA, P., & VERGARI, D. (2025). Comparing Credit Risk Estimates in the Gen-AI Era [Retrieved from arXiv]. *arXiv preprint arXiv :2506.07754*.
- MEHDI, R., AHMED, I. E., & MOHAMED, E. A. (2025). Rating the Impact of Risks in Banking on Performance : Utilizing the Adaptive Neural Network-Based Fuzzy Inference System (ANFIS). *Risks*, 13(85). <https://doi.org/10.3390/risks13050085>
- VALDRIGHI, G., RIBEIRO, A. M., PEREIRA, J. S. B., GUARDIEIRO, V., HENDRICKS, A., MIRANDA FILHO, D., NIETO GARCIA, J. D., BOCCA, F. F., VERONESE, T. B., WANNER, L., & RAIMUNDO, M. M. (2024). Best Practices for Responsible Machine Learning in Credit Scoring [Tutorial paper, Machine Learning Fairness, UC Campinas]. *arXiv preprint arXiv :2409.20536*.
- XU, Z., SUN, Y., LIU, S., JIANG, M., SUN, M., & SUN, W. (2024). Applying Hybrid Graph Neural Networks to Strengthen Credit Risk Analysis [Retrieved from arXiv]. *arXiv preprint arXiv :2410.04283*.