Rapport de Mission de Scoring Comportemental

Nom de l'Organisation/Entreprise : SOCIETE GENERALE France(siège)

Titre de la Mission : Développement d'un modèle de Scoring Comportemental

Date : du 25 -30 /09/2023

Résumé Exécutif

Ce rapport présente les résultats et les expériences de notre mission, qui avait pour objectif de développer un modèle de Scoring comportemental qui évalue le comportement financier des clients et attribue un score en fonction de leurs antécédents de paiement, d'endettement, de dépenses et d'autres facteurs pertinents.

Au cours de la mission, des rencontres avec d'autres acteurs de la banque ont été organisées pour partager des informations et des meilleures pratiques,contribuant ainsi à enrichir notre connaissance.

Table des Matières

1. Introduction

2. Description de la Mission

4. Choix des Algorithmes de Machine Learning

5. Entraînement et Évaluation du Modèle

6. Résultats et Analyse

7. Rencontres avec d'autres acteurs de la banque

8. Conclusions

9. Recommandations

10. Annexes

1. Introduction

La mission avait pour but de développer un modèle de scoring comportemental qui permettrait à la banque d'évaluer le risque de comportement des clients dans le cadre de la gestion des crédits et des risques.

Les principaux objectifs de cette étude sont les suivants :

Développer un Modèle de Scoring comportemental : Concevoir et construire un modèle prédictif qui évalue le comportement financier des clients et attribue un score en fonction de leurs antécédents de paiement, d'endettement, de dépenses et d'autres facteurs pertinents.

1. Évaluer le Risque de Défaut :L'objectif principal du scoring comportemental est d'évaluer le risque de défaut d'un individu. Cela implique de développer un modèle qui peut prédire avec précision la probabilité qu'un emprunteur ou un client ne parvienne pas à rembourser ses dettes.

2. Identifer les Facteurs de Risque : Examiner les variables et les caractéristiques qui ont un impact significatif sur le comportement de remboursement. Cela peut inclure des variables telles que l'historique de paiement, le montant des dettes, l'ancienneté bancaire, etc.

3. Optimiser les Décisions de Crédit : Utiliser les scores comportementaux pour prendre des décisions éclairées en matière de crédit. Les scores peuvent aider les prêteurs à déterminer les limites de crédit appropriées, à évaluer la solvabilité et à personnaliser les offres de produits financiers.

4. Améliorer la Rentabilité : En réduisant les risques de défaut et en prenant des décisions plus précises en matière de crédit, les institutions financières peuvent améliorer leur rentabilité globale.

5. Validation et Ajustements du Modèle : Les objectifs incluent la validation et l'ajustement constants du modèle de scoring pour garantir sa précision et sa pertinence dans un environnement financier en évolution.

2. Description de la Mission

La mission a débuté par la collecte de données comportementales des clients, comprenant des informations telles que l'historique des transactions, les antécédents de paiement, et les habitudes de dépenses, entre autres. Ensuite, des traitements de data cleaning, des analyses descriptives, et des opérations de data engineering ont été réalisés. Ces données ont ensuite été exploitées pour créer un modèle de machine learning.

4. Choix des Algorithmes de Machine Learning

Plusieurs algorithmes de machine learning auraient pu être utilisés, cependant, le choix s'est porté sur la régression logistique.

Quant à la sélection des variables, elle s'est opérée à travers un processus basé sur divers critères, incluant la corrélation, la stabilité au cours de trois années consécutives de données, leur pertinence informationnelle, et d'autres facteurs significatifs.

5. Entraînement et Évaluation du Modèle

Le modèle final a été entraîné sur un ensemble de données d'apprentissage et évalué à l'aide de mesures de performance telles que la précision, le rappel, et la courbe ROC. Les résultats initiaux ont montré une promesse significative.

Le modèle a été entraîné sur un ensemble de données d'entraînement et évalué sur un ensemble de données de test. Les métriques d'évaluation comprenaient la précision, l’accuracy, la F-mesure et l'aire sous la courbe ROC.

7. Résultats et Analyse

Les résultats montrent que le modèle atteint une précision de 80% dans la classification du comportement des clients. Cependant, des ajustements sont nécessaires pour améliorer ce score.

6. Rencontres avec d'autres acteurs de la banque

Au cours de la mission, des rencontres fructueuses avec d'autres acteurs de la banque ont eu lieu. Ces rencontres ont permis d'échanger des informations sur leurs méthodes de travail, les outils utilisés, …. Les informations recueillies ont grandement enrichi notre compréhension du domaine et nous ont fourni des perspectives précieuses.

8. Conclusions

La mission a atteint son objectif initial de développement d'un un modèle de scoring comportemental, mais des améliorations sont nécessaires pour atteindre une plus grande précision. De plus, les rencontres avec d'autres acteurs de la banque ont renforcé notre expertise dans ce domaine.

9. Recommandations

Mettre en œuvre notre modèle sur un ensemble de données récent afin d'évaluer à la fois sa robustesse et ses performances.

10. Annexes

- Graphiques de performance du modèle.

- Rapports des rencontres avec les acteurs de la banque.

Ce rapport de mission de machine learning résume les étapes et les résultats de notre mission, mettant en lumière l'importance des rencontres avec d'autres acteurs de la banque.

Rapport sur les Rencontres avec les Acteurs de la Banque

* Rapport rencontre avec Sandrine

Au cours de nos rencontres Sandrine XXXX , fonction de AFMO ,,,nous avons découvert des initiatives intéressantes visant à améliorer la gestion des données et à exploiter leur potentiel pour la création de valeur. Voici un aperçu des principales discussions et développements :

1. DGLE Inno et la Data Leaders Community : Il a été noté que DGLE Inno a mis en place une initiative appelée "Data Leaders Community" qui organise des réunions trimestrielles pour discuter de la qualité du portefeuille, de la création de valeur, des projets en cours et des formations. Cette communauté sert de forum pour partager des idées et des meilleures pratiques en matière de gestion des données.

2. Lancement de Data IA par Afmo : En décembre 2019, Afmo a lancé l'initiative "Data IA" avec pour objectif d'acculturer ses collaborateurs à la science des données. Cela a suscité la mise en place d'une plateforme pour évaluer le niveau de connaissance de chaque collaborateur en matière de données, afin de proposer des parcours de formation adaptés.

3. Hackathon et Émulation: À la suite de ces initiatives, un hackathon a été organisé avec succès, suscitant un fort enthousiasme parmi les collaborateurs. Cette compétition a permis de stimuler la créativité et l'innovation au sein de l'entreprise d’où la naissance de No Caisse.

4. Création de Valeur Dissimulée chez Afmo: Il est devenu évident que chez Afmo, il existe des opportunités de création de valeur cachées au sein des différentes filiales. Cependant, ces ressources sont souvent sous-exploitées, entraînant une perte de valeur potentielle. Cela a conduit à une réflexion sur la nécessité de structurer, de mettre en place des démarches et des méthodologies, ainsi que des processus de travail pour dupliquer les solutions réussies dans d'autres filiales où elles pourraient être bénéfiques.

powerpoint voir

Le cycle de vie dun cas d usage ? comment sont valoriser les uses cases ?

Document de cadrage sur moana exemple

thématique pmo a couvrir avec les métiers et son equipe voir doc (- La description du projet Moana, y compris son objectif, sa portée, et ses objectifs.

- Les parties prenantes clés et leurs rôles.

- Les cas d'utilisation ou les fonctionnalités attendues de Moana.

- Les ressources nécessaires, y compris les acteurs impliqués et les budgets.

- Les calendriers et les jalons du projet.

- Les risques potentiels et les plans d'atténuation.

- Les indicateurs de performance clés (KPI) pour évaluer la réussite du projet.

)

Target operating model, comment de bout en bout, les acteurs a embarquer, leurs roles,quel est le livrables final a avoir (Pour le "Target Operating Model", il s'agit généralement d'un modèle qui décrit comment une organisation fonctionne de manière opérationnelle. Il couvre les aspects tels que les processus, les ressources, les technologies, les rôles et responsabilités, etc. Le processus de mise en œuvre de ce modèle nécessite l'identification des acteurs impliqués, la définition de leurs rôles, et la création de livrables tels que des manuels de procédures, des organigrammes, des plans de formation, etc.

)

En résumé, les rencontres avec les acteurs de la banque ont révélé un engagement important envers l'amélioration de la gestion des données et la recherche de nouvelles opportunités pour créer de la valeur. Les initiatives telles que la Data Leaders Community, Data IA et les hackathons ont renforcé la collaboration et l'émulation au sein de l'entreprise. La reconnaissance des opportunités de création de valeur dissimulée a incité à la mise en place de méthodologies et de processus visant à optimiser ces ressources au sein de l'ensemble de l'organisation.

Le cycle de vie d'un cas d'utilisation (ou cas d'usage) dans le contexte de la gestion de projet et de la transformation peut être divisé en plusieurs étapes :

1. \*\*Identification et Cadrage\*\* : Cette phase consiste à identifier les cas d'utilisation potentiels, à les documenter et à les évaluer en termes de leur pertinence pour le projet ou l'initiative.

2. \*\*Analyse et Conception\*\* : Dans cette étape, les cas d'utilisation sont analysés en détail pour comprendre les besoins métier sous-jacents. Les spécifications sont élaborées, et les scénarios d'utilisation sont modélisés.

3. \*\*Développement ou Mise en Œuvre\*\* : Les cas d'utilisation sont mis en œuvre, que ce soit sous forme de développement logiciel, de processus métier, ou d'autres solutions.

4. \*\*Test et Validation\*\* : Les cas d'utilisation sont testés pour s'assurer qu'ils fonctionnent conformément aux spécifications et répondent aux besoins métier.

5. \*\*Déploiement\*\* : Les cas d'utilisation sont déployés dans l'environnement opérationnel.

6. \*\*Exploitation et Suivi\*\* : Après le déploiement, les cas d'utilisation sont surveillés pour s'assurer qu'ils continuent de répondre aux besoins métier et sont maintenus en fonctionnement.

7. \*\*Optimisation et Évolution\*\* : Les cas d'utilisation peuvent nécessiter des ajustements, des améliorations, ou des évolutions pour s'adapter aux changements dans l'entreprise ou dans l'environnement.

Quant à la valorisation des cas d'utilisation, cela dépend du contexte de votre projet ou de votre initiative. La valeur des cas d'utilisation peut être mesurée en fonction de divers indicateurs tels que l'efficacité opérationnelle, la réduction des coûts, l'amélioration de la satisfaction des clients, etc.

* Plateforme de collecte de de visualisation de projets a voir se connecter avec sg connect
* Voir la valeur d un use case par projet
* Organiser un hackathon, parainage dans des écoles
* Portail group go & AI VOIr acces en mode lecture avec compte corp, on a 1280 tous les uses case. Yes 708 ml : detection de fraude en temps réel doc en mozaique à voir SGRF voir leur use case, ia generatif, lmm
* Participer au salon nc est pr simuler l ideation
* GLE INNOV

Scrum board dns teaam, backlog : Backlog, c'est en gros les idées qu'on peut avoir. Qui seront à développer dans peu à moyen terme.

Durent maximum de 3 jours parce qu'en gros, c'est une tâche qui dure une semaine, ça veut dire que tu peux la décomposer. En plus de ça, J et cetera. En gros, une tâche, c'est vraiment un truc sur lequel tu as bossé de lourds. Après tu n'es pas entre une demi-journée et 3 jours.

Github , faire un repo par kedro ou cookiecutter avec vscode

[Cookiecutter https://drivendata.github.io/cookiecutter-data-science/]( Cookiecutter   https://drivendata.github.io/cookiecutter-data-science/                                     )

https://medium.com/devcareers/cookie-cutter-organizing-data-science-projects-c3f13acd3247

Cookiecutter Data Science est un projet open source hébergé sur GitHub. Il s'agit d'un modèle de projet prêt à l'emploi qui facilite la création d'environnements de développement de science des données cohérents, organisés et reproductibles. Voici une explication plus détaillée de ce projet :

1. \*\*Modèle de Projet\*\* : Cookiecutter Data Science fournit un modèle de projet préconçu pour les projets de science des données. Ce modèle inclut une structure de répertoires organisée, des fichiers de configuration et des modèles de documentation, tout prêt à être utilisé.

2. \*\*Structure de Répertoires\*\* : Le modèle Cookiecutter Data Science définit une structure de répertoires standard pour votre projet de science des données. Il organise les répertoires pour les données, le code source, la documentation, les résultats, les modèles, etc.

3. \*\*Gestion des Dépendances\*\* : Il peut inclure des fichiers de configuration pour gérer les dépendances du projet, tels que des fichiers `requirements.txt` pour Python, facilitant la répétabilité de l'environnement.

4. \*\*Documentation\*\* : Le modèle inclut souvent des modèles de documentation prêts à l'emploi pour la documentation du projet, ce qui encourage une bonne pratique de documentation dans le domaine de la science des données.

5. \*\*Scripts et Workflow\*\* : Il peut contenir des scripts et des fichiers de configuration pour automatiser des tâches courantes telles que le nettoyage de données, la formation de modèles, la validation croisée, etc.

6. \*\*Gestion de Projet\*\* : Cookiecutter Data Science encourage également la gestion de projet efficace, y compris la gestion de versions, la collaboration en équipe et le suivi des tâches.

7. \*\*Répétabilité\*\* : En utilisant le modèle, vous assurez que votre projet de science des données peut être reproduit par d'autres. Cela favorise la transparence et la vérifiabilité des résultats.

8. \*\*Personnalisable\*\* : Bien que le modèle fournisse une structure de base, il est également personnalisable pour s'adapter aux besoins spécifiques de votre projet.

Pour utiliser Cookiecutter Data Science, vous pouvez cloner le modèle de projet depuis le référentiel GitHub et personnaliser la structure, les fichiers de configuration et la documentation en fonction de votre projet de science des données. Cela vous permet de gagner du temps en évitant de partir de zéro à chaque fois que vous démarrez un nouveau projet de science des données, tout en garantissant une organisation et une répétabilité cohérentes.