

Hackathon 2024: *IESEG - Crédit Agricole Nord de France*

Présentation Finale - 27/05/2024

Présenté par :
Anais MELLAS
Kodangada Ketan Bopanna
Malek Ben Romdhane





Agenda

Contexte

Définition de la Problématique

Objectifs

Ressources

Stratégies

Modélisations

Limitations

Visualisations et Projection de la Carte

Contexte

Les principaux
chiffres relatifs
aux risques
naturels en France
avec un zoom sur
les inondations



1982-2023

17 500 événements à l'origine d'une reconnaissance de l'état de catastrophe naturelle dans au moins une commune.



L'inondation
représente 56 %
du risque.

Débordement de cours d'eau
10.5 millions de logements
28% du total des logements

Submersion Marine
1.3 millions de logements
3.5% du total des logements



1995-2019

un total de 14.4 Milliards d'euros d'indemnisation.

Définition de la Problématique

- ▶ Cette nette montée du risque d'inondation, représente une menace pour les actifs de la banque.



le 2ème en assurance habitation en France.

(Argus de l'assurance 23/05/23 sur chiffre d'affaires 2022).

- ▶ Intégration du risque d'inondation :
La stratégie d'assurance
Gestion des risques
Publications réglementaires

Objectifs



Définition d'un score de risque physique (Absent, Faible, Moyen, Élevé) lié au péril.



Estimation de la dépréciation du bien immobilier suite à la survenance du péril.

Ressources

Nord et Pas-de-Calais



Base nationale de Gestion Assistée
des Procédures Administratives relatives
aux Risques (GASPAR)

Base de données topographique (BDTOPO)

Donnés Météo

Demandes de Valeurs Foncières (DVF)

Base de Données Nationale des
Bâtiments (BDNB)

Base de données Observatoire National des
Risques Naturels (ONRN)

Communes de France

Stratégie



Définition d'un score de risque physique (Absent, Faible, Moyen, Élevé) lié au péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

3

Estimation du risque physique avec des paramètres liés à l'hydrographie, météo, Plans de Prévention de Risques Naturels, Atals des Zones Inondables et DDRM en plus de la variable cible dans un *modèle de classification*.

Stratégie



Définition d'un score de risque physique (Absent, Faible, Moyen, Élevé) lié au péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

3

Estimation du risque physique avec des paramètres liés à l'hydrographie, météo, Plans de Prévention de Risques Naturels, Atals des Zones Inondables et DDRM en plus de la variable cible dans un *modèle de classification*.



Sélection des bases de données nécessaires, les tables et les columns importantes.

Base nationale de Gestion Assistée des Procédures Administratives relatives aux Risques (GASPAR)

*Table Catnat_gaspar Table risq_gaspar
Table azi_gaspar Table pprn_gaspar

Météo Data - Données Climatiques de Base - Mensuelles

MENS_departement_59_période_1950-2024
MENS_departement_62_période_1950-2024

Base de données topographique (BDTOPO)

Table COURS_D_EAU
Table PLAN_D_EAU

Stratégie



Définition d'un score de risque physique (Absent, Faible, Moyen, Élevé) lié au péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

3

Estimation du risque physique avec des paramètres liés à l'hydrographie, météo, Plans de Prévention de Risques Naturels, Atals des Zones Inondables et DDRM en plus de la variable cible dans un *modèle de classification*.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

Analytic Hierarchy Process (AHP)

Objectif : Évaluer le risque physique en fonction de trois critères par ordre d'importance en paire. Fréquence, Durée moyenne et Récence du péril.

Fréq. vs. Durée : La fréq. est légèrement plus importante que la durée (score : 5).

Fréq. vs. Récence : La fréq. est modérément plus importante que la récence (score : 7).

Durée vs. Récence : La durée et la récence sont légèrement importantes (score : 3).

Critère	Fréquence	Durée	Récence
Fréquence	1	5	7
Durée	1/5	1	3
Récence	1/7	1/3	1

$$\text{Facteur de Risque Physique} = 0.73\mathbf{F} + 0.19\mathbf{D} + 0.08\mathbf{R}$$

Frequence

Durée Moyenne

Récence = 1 / Nbr de jours depuis le dernier péril

*Seule la Table Catnat_Gaspar contenant les données historiques des inondations a été utilisée pour cette étape. Nous avons extrait la fréquence, la durée et la récurrence annuelle pour chaque commune.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

Statistiques	Valeur
count	5673.0
Mean	2.975160
Std	6.613022
Min	0.920005
25%	1.110009
75%	3.205037
Max	185.980006

Fonction

Facteur du Risque Physique ≤ 1.68
Attribuer un score de risque 1

Facteur du Risque Physique ≤ 3.21
Attribuer un score de risque 2

Autre
Attribuer un score de risque 3

Total par score

Score de risque 1 2312
Score de risque 2 1943
Score de risque 3 1417



Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

Year	code commune	Durée moyenne (jours)	Fréquence	Récence(jours)	Risque Physique
1983	59052	1	1	14945	1
1983	59071	6	1	14910	2
....
2023	62906	39	3	147	3
2023	62908	23	2	191	3

Resultats: Chaque commune, qu'elle ait subi un péril ou non, se voit attribuer un score de risque.

- 0 : Absent
- 1 : Faible
- 2 : Moyen
- 3 : Élevé

Stratégie



Définition d'un score de risque physique (Absent, Faible, Moyen, Élevé) lié au péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible (Risque physique) en utilisant 3 paramètres de base à l'aide de Analytic Hierarchy Process (AHP).

3

Estimation du risque physique avec des paramètres liés à l'hydrographie, météo, Plans de Prévention de Risques Naturels, Atals des Zones Inondables et DDRM en plus de la variable cible à dans *modèle de classification*.

3

Estimation du risque physique avec des paramètres liés à l'hydrographie, météo, Plans de Prévention de Risques Naturels, Atals des Zones Inondables et DDRM en plus de la varibale cible dans un *modèle de classification*.

Year	code commune	RR	min dist. plan d'eau	min dist. cours d'eau	nbr cours d'eau	rsique submer. marine	Risque Physique
1983	59004	609.5	7736.61	470.85	...	7	1	2
1983	59005	695.8	302.93	48.69	...	14	1	0
....
2023	62905	1205.1	4444.5	2008.2	...	0	0	2
2023	62906	947.5	2566.4	256.6	...	18	0	3



*Les variables : fréquence, durée et récence ont été retirées.

Modélisation

Basetable

Plage temporelle: 1983 - 2023

Structure: 63181 lignes , 31 colonnes

Etudes longitudinales: Année , Code commune

Variable Cible: Risque Physique

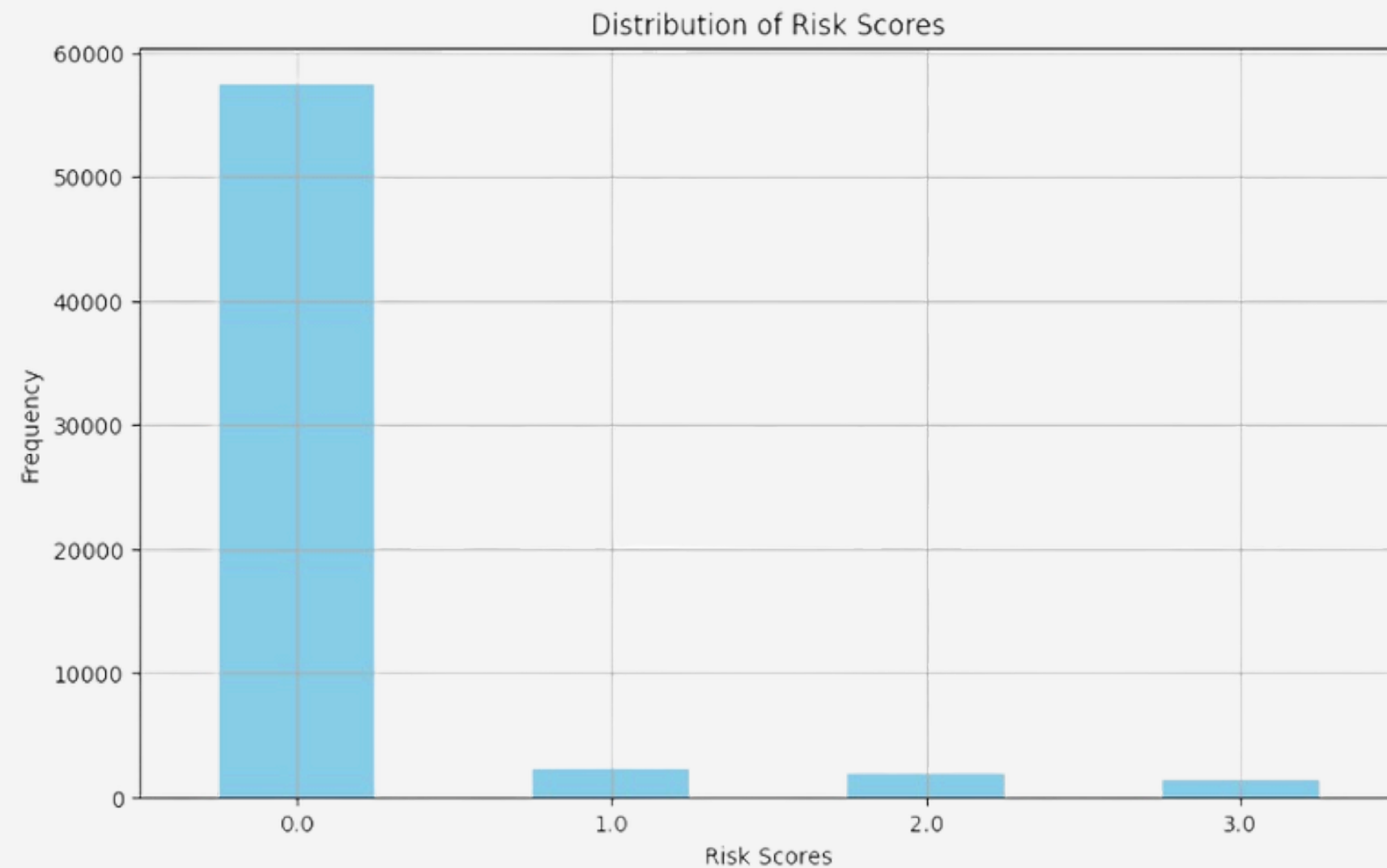
Variables BD TOPO: Importance du cours d'eau, importance du plan d'eau, nombre de plans d'eau, nombre de cours d'eau, distance du cours d'eau, distance du plan d'eau.

Variables BD Gaspar: Nombre de risques, nombre de programmes de prévention, risque remontée nappes, risque ruissellement et coulée de boue, risque de submersion marine, risque crue à débordement d'eau, risque rupture de barrage.

Variables BD météo: RR, NBRR, RRAB, RRABDAT, NBJRR1, NBJRR5, NBJRR10, NBJRR30, NBJRR50, NBJRR100.

Modélisation

Distribution des Classes



Note: la création de la nouvelle classe a causé un déséquilibre important dans la distribution des classes.

Score de risque 0: 57508

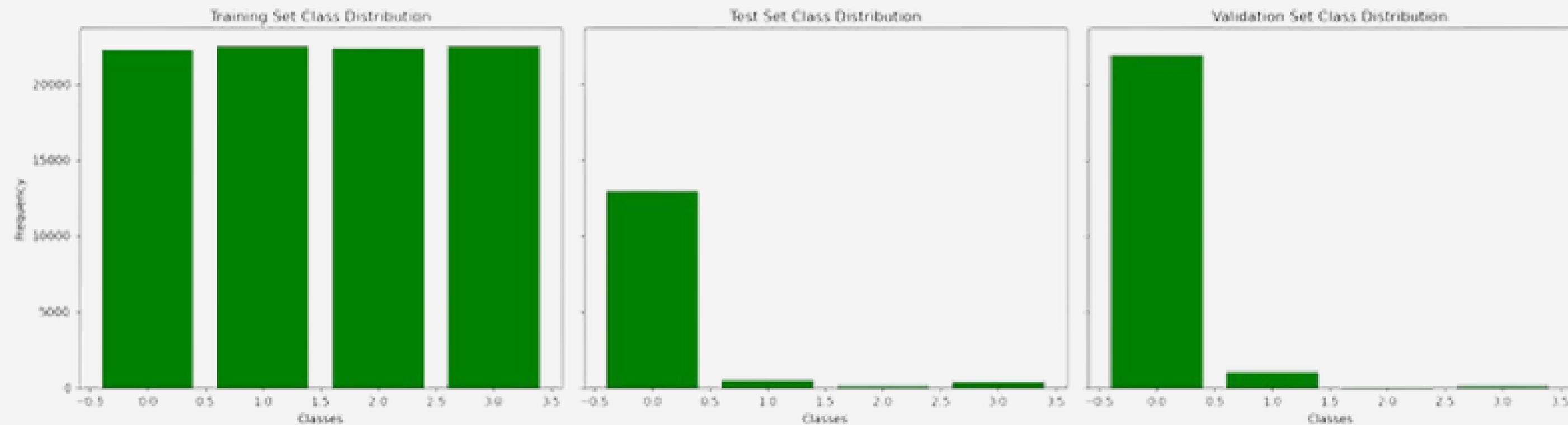
Score de risque 1: 2312

Score de risque 2: 1943

Score de risque 3: 1417

Modélisation

Technique de Traitement du Déséquilibre des Classes



Trois techniques ont été testées : OverSampling ,UnderSampling, **SMOTE**

Modélisation

Correlation entre les Paires de Variables

Variables Corrélées (> 0.8)

- Risque Par remontées de nappes - Risque Par une crue à débordement lent de cours d'eau : 1.0
- max importance cours d'eau - min importance cours d'eau: 0.99
- max importance plan d'eau - min importance plan d'eau: 0.97
- RR - RRAB: 0.85 RR - NBJRR5: 0.93
- RR - NBJRR1: 0.81 RR - NBJRR10: 0.91
- NBJRR5 - NBJRR1: 0.8 NBJRR5 - NBJRR10: 0.83

Variables Éliminées

- Risque Par une crue à débordement lent de cours d'eau
- min importance cours d'eau
- min importance plan d'eau
- NBJRR5
- NBJRR10

Modélisation

Liste des Modèles et leurs Paramètres

Modèle	Hypère Paramtères
Logistic Regression	max_iter=20000 , solver='saga', C=1.0
Decision Tree	max_depth=5, min_samples_leaf=4
Random Forest	n_estimators=100, max_depth=10, min_samples_leaf=4
Support Vector Machine	probability=True, C=1.0, gamma='scale'
KNN	n_neighbors=5
XGBoost HT	colsample_bytree: 0.8, learning_rate: 0.1, max_depth: 6, n_estimators: 200, reg_alpha: 1, reg_lambda: 1, subsample: 0.6



Note: StandardScaler a été utilisé pour mettre à l'échelle les ensembles de données.

Modélisation

Résultats du Modèle Sélectionné

Modèle	DataSet	Accuracy	AUC	Precision	Recall	F1 Score
Logistic Regression	Train	0.58	0.80	0.57	0.58	0.57
	Test	0.40	0.73	0.90	0.40	0.52
	Validation	0.30	0.68	0.93	0.30	0.44
XGBoost	Train	0.91	0.97	0.91	0.91	0.91
	Test	0.69	0.61	0.88	0.69	0.77
	Validation	0.67	0.55	0.90	0.67	0.77
XGBoost HT	Train	0.95	0.99	0.95	0.95	0.95
	Test	0.72	0.62	0.88	0.72	0.79
	Validation	0.71	0.58	0.90	0.71	0.80

Top 10 Feature Importance

- **NBJRR1**: Nombre de jour ou RR>1mm
- **NJRR30**: Nombre de jour ou RR>30mm
- **NBRR**: nombre de valeurs présentes de hauteur de précipitation quotidienne
- **Risque Inondation**
- **RRABDAT**: Jour du RRAB
- **NBJRR50**: Nombre de jour ou RR>50mm
- **RRAB**: précipitation maximale tombée en 24 heures au cours du mois
- **Risque remonté Nappes**
- **NBJRR100**: Nombre de jour ou RR>100mm
- **Max importance de cours d'eau**

Stratégie



Estimation de la dépréciation du bien immobilier suite à la survenance d'un péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible : la dépréciation d'un bien résidentiel

3

Estimation de la dépréciation des valeurs foncières avec des paramètres liés aux caractéristiques des bâtiments, l'hydrographie, les valeurs foncières des années précédentes, couts et indemnités versées pour assurer les inondations, et le score du risque physique.

Stratégie



Estimation de la dépréciation du bien immobilier suite à la survenance d'un péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible : la dépréciation d'un bien résidentiel

3

Estimation de la dépréciation des valeurs foncières avec des paramètres liés aux caractéristiques des bâtiments, l'hydrographie, les valeurs foncières des années précédentes, couts et indemnités versées pour assurer les inondations, et le score du risque physique.



Sélection des bases de données nécessaires, les tables et les columns importantes.

Base de Données Nationale des Bâtiments (BDNB)

Batiments groupes Adressses
Batiments construction Parcelles
Batiments DPE

Base de données topographique (BDTOPO)

Table COURS_D_EAU Table PLAN_D_EAU

Demandes de Valeurs Foncières (DVF)

DVF communes 2014 - 2023

Base de données ONRN

Table Indemnisation
Couts cumulées des inondations

BaseTable Modèle 1

Score du risque physique

Stratégie



Estimation de la dépréciation du bien immobilier suite à la survenance d'un péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible : la dépréciation d'un bien résidentiel

3

Estimation de la dépréciation des valeurs foncières avec des paramètres liés aux caractéristiques des bâtiments, l'hydrographie, les valeurs foncières des années précédentes, couts et indemnités versées pour assurer les inondations, et le score du risque physique.



Définition de la variable cible :
la dépréciation d'un bien résidentiel

La dépréciation = Prix moyen au m² de l'année courante - Prix moyen au m² de l'année précédente

Note: On a travaillé au niveau des communes par manque de données sur les bâtiments.

Stratégie



Estimation de la dépréciation du bien immobilier suite à la survenance d'un péril.

1

Sélection des bases de données nécessaires, les tables et les attributs importants.

2

Définition de la variable cible : la dépréciation d'un bien résidentiel

3

Estimation de la dépréciation des valeurs foncières avec des paramètres liés aux caractéristiques des bâtiments, l'hydrographie, les valeurs foncières des années précédentes, couts et indemnités versées pour assurer les inondations, et le score du risque physique.

3

Estimation de la dépréciation des valeurs foncières avec des paramètres liés aux batiments, l'hydrographie, les valeurs foncières des années précédentes, le score du risque physique en plus de la varibale cible à un *modèle de regression*.

Year	code commune	Risque Physique	Nbr cours d'eau	Nombre de batiments	age moyen des batiments	lagged_depreciation	Dépreciation (prix m² moyen)
2016	59001	0	2	187	...	36	-15.00	667.50
2016	59002	0	0	1432	...	59	-69.16	-109.58
....
2023	62908	3	9	643	...	43	-283.99	628.22
2023	62909	0	4	292	...	98	-239.29	1071.00

Modélisation

Basetable

Plage temporelle: 2016 - 2023

Etudes longitudinales: Année , Code commune

Structure: 12304 lignes , 42 colonnes

Variable Cible: Dépreciation (prix m² moyen)

Variables BD TOPO: Importance du cours d'eau, importance du plan d'eau, nombre de plans d'eau, nombre de cours d'eau, distance du cours d'eau, distance du plan d'eau.

Variables ONRN: Dépense Estimée sur inondations (€k)

Variables BDNB: Nombre de bâtiments, année construction, âge moyen bâtiments, matière murs, matière toit, Nombre de logements, mode aléa, somme surfaces parcelles, hauteur moyenne, altitude au sol moyenne, Nombre de niveaux moyen, surface habitable moyenne., surface vitrée moyenne, Nombre de balcons moyen, et % surface baie vitrée extérieur moyenne.

Ces paramètres permettent une analyse détaillée et complète des caractéristiques des bâtiments et de leur environnement dans chaque commune.

Variables DVF: Prix m² Moyen année précédente, dépreciation année précédente

Modélisation

Liste des Modèles et leurs Paramètres

Modèle	Hypère Paramtères
Linear Regression	/
Random Forest	n_estimators=100, max_depth=10, random_state=42
Gradient Boosting	n_estimators=100, learning_rate=0.1, max_depth=5, random_state=42
XGBoost	n_estimators=100, learning_rate=0.1, max_depth=5, random_state=42
catBoost	n_estimators=100, learning_rate=0.1, max_depth=5, random_state=42, verbose=0



Modélisation

Résultats des Modèles

Modèle	DataSet	MSE	RMSE	MAE	R2
Linear Regression	Train Validation	86229.91 116432.18	293.76 341.22	206.54 230.25	0.54 0.50
Random Forest	Train Validation	35183.34 93047.47	187.57 305.03	127.66 197.7	0.815 0.60
Gradient Boosting	Train Validation	35255.88 75345.81	187.76 274.49	127.23 189.93	0.85 0.67
XGBoost	Train Validation	30472.71 95452.87	174.56 308.95	131.23 197.71	0.83 0.60
CatBoost	Train Validation	48581.09 97290.22	220.47 311.91	161.51 197.11	0.74 0.58

Top 7 Feature Importance

- Prix m2 Moyen année précédente
- Dépreciation année précédente
- Nombre de logements
- Année construction
- % surface baie vitrée exterieur moyenne
- Altitude au sol moyenne
- Age moyen des batiments

Modélisation

Sénarios Optimiste:

1. Diminuer le niveau du score de risque de 1, sauf pour le niveau 0, pour 2024.
2. Réduire la dépense moyenne des sinistres de 25 % pour 2024.

Sénarios Modéré:

1. Conserver le même score de risque qu'en 2023.
2. Conserver les mêmes dépenses moyennes des sinistres qu'en 2023.

Sénarios Pessimiste:

1. Augmenter le score de risque d'un niveau, sauf pour le niveau 3, pour 2024.
2. Augmenter les dépenses moyennes des sinistres de 25 % pour 2024.

Limitations

La documentation de certaines bases de données est limitée.

Les données open source prennent du temps à explorer

Différence entre les plages temporelles à travers les bases de données.

Les données DVF accessibles au public ne fournissent que les dates de mutation des bâtiments, qui sont souvent éloignées de la date de survenue du péril, entraînant ainsi une grande différence entre ces dates

Il est difficile de déterminer l'impact du risque physique sur la dépréciation, en raison d'autres facteurs qui y influent tels que les facteurs économiques et socio-économiques

Les limitations des ressources CPU et GPU : Modélisations et la DataAPP

Visualisations et Cartes



Merci de Votre Attention

