

Projet final : Prix des maisons

Prof : Mr SY

Email : syibrahima31@outllok.com

Master 2 : Modélisation Statistique et Informatique (MSI)

Description du projet

Demandez à un acheteur de décrire la maison de ses rêves et il ne commencera probablement pas par la hauteur du plafond du sous-sol ou la proximité d'une voie ferrée . Mais l'ensemble de données de ce projet prouve que les négociations de prix sont influencées par bien d'autres éléments que le nombre de chambres à coucher ou une clôture à piquets blancs.

Avec **79** variables explicatives décrivant (presque) tous les aspects des maisons résidentielles à Ames, Iowa aux états unies , ce projet vous met au défi de développer un modèle de machine learning permettant de prédire le prix final de chaque maison.

But du projet

Le but de projet est d'élaborer un modèle de machine learning dont **l'erreur de généralisation** est la plus faible possible

Métrique

La métrique retenue pour mesurer l'erreur de généralisation(mesurer sur le testset) est le [Root-Mean-Squared-Error \(RMSE\)](#)

Modélisation

Durant la l'élaboration du modèle

- Penser à faire du `featur engineering` pour donner plus de capacité a votre modèle (si vous le trouvez nécessaire)
- Pour pouvoir faire une `réduction de dimension` (si vous le trouver nécessaire)
- Hormis les modèles utilisés en cours si vous le souhaitez vous pouvez utiliser d'autres modèles telque (`xgboost` , `random_forest` , `svm` etc)

Fichiers dans le dossier

- **train.csv** - l'ensemble d'entraînement
- **test.csv** - l'ensemble de test

Description des variables

Voici une brève version de ce que vous trouverez dans le fichier de description des données.

- **SalePrice** - le prix de vente de la propriété en dollars. C'est la variable cible que vous essayez de prédire.
- **MSSubClass** : La classe de construction
- **MSZoning** : La classification générale du zonage
- **LotFrontage** : Pieds linéaires de rue reliés à la propriété
- **LotArea** : Taille du terrain en pieds carrés
- **Rue** : Type d'accès routier
- **Allée** : Type d'accès à l'allée
- **LotShape** : Forme générale de la propriété
- **LandContour** : Planéité de la propriété
- **Utilitaires** : Type d'utilitaires disponibles
- **LotConfig** : Configuration du lot
- **LandSlope** : Pente de la propriété
- **Quartier** : Emplacements physiques dans les limites de la ville d'Ames
- **Condition1** : Proximité de la route principale ou de la voie ferrée
- **Condition2** : Proximité de la route principale ou de la voie ferrée (si une seconde est présente)
- **BldgType** : Type de logement
- **HouseStyle** : Style d'habitation
- **OverallQual** : **Qualité** globale du matériau et de la finition
- **OverallCond** : évaluation de l'état général
- **Année de construction** : Date de construction d'origine
- **YearRemodAdd** : Date de **refonte**
- **RoofStyle** : Type de toit
- **RoofMatl** : Matériau du toit
- **Exterior1st** : Revêtement extérieur de la maison
- **Extérieur2e** : Revêtement extérieur de la maison (si plus d'un matériau)
- **MasVnrType** : Type de placage de maçonnerie
- **MasVnrArea** : Surface de placage de maçonnerie en pieds carrés
- **ExterQual** : Qualité des matériaux extérieurs
- **ExterCond** : **Etat** actuel du matériel à l'extérieur
- **Fondation** : Type de fondation
- **BsmtQual** : Hauteur du sous-sol
- **BsmtCond** : État général du sous-sol
- **BsmtExposure** : Murs de sous-sol en rez-de-jardin ou en **sortie**
- **BsmtFinType1** : Qualité de la surface finie du sous-sol
- **BsmtFinSF1** : Pieds carrés finis de type 1
- **BsmtFinType2** : Qualité de la deuxième zone finie (si présente)
- **BsmtFinSF2** : Pieds carrés finis de type 2
- **BsmtUnfSF** : Pieds carrés non **finis** du sous-sol
- **TotalBsmtSF** : Total des pieds carrés de la superficie du sous-sol
- **Chauffage** : Type de chauffage

- **ChauffageQC** : Qualité et état du chauffage
- **CentralAir** : Air climatisé central
- **Électrique** : Système électrique
- **1stFlrSF** : Pieds carrés au premier étage
- **2ndFlrSF** : Pieds carrés au deuxième étage
- **LowQualFinSF** : Pieds carrés finis de basse qualité (tous les étages)
- **GrLivArea** : **Surface habitable hors** sol (sol) pieds carrés
- **BsmtFullBath** : **Salles de** bain complètes au sous-sol
- **BsmtHalfBath** : Demi-salles de bain au sous-sol
- **FullBath** : **Salles de** bain complètes au-dessus du niveau du **sol**
- **HalfBath** : Demi-bains au dessus du niveau du **sol**
- **Chambre à coucher** : Nombre de chambres au-dessus du sous-sol
- **Cuisine** : Nombre de cuisines
- **KitchenQual** : Qualité de la cuisine
- **TotRmsAbvGrd** : Total des chambres au-dessus du **sol** (n'inclut pas les salles de bain)
- **Fonctionnel** : Évaluation de la fonctionnalité de la maison
- **Cheminées** : Nombre de cheminées
- **FireplaceQu** : Cheminée de qualité
- **Type de garage** : Emplacement du garage
- **GarageYrBlt** : Année de construction du garage
- **GarageFinish** : Finition intérieure du garage
- **GarageCars** : Taille du garage en capacité de voiture
- **GarageArea** : Taille du garage en pieds carrés
- **GarageQual** : Qualité garage
- **GarageCond** : **Etat du** garage
- **PavedDrive** : Allée pavée
- **WoodDeckSF** : Superficie de la terrasse en bois en pieds carrés
- **OpenPorchSF** : Aire de porche ouverte en pieds carrés
- **EnclosedPorch** : Aire de porche fermée en pieds carrés
- **3SsnPorch** : Aire de porche trois saisons en pieds carrés
- **ScreenPorch** : Superficie du porche **moustiquaire** en pieds carrés
- **PoolArea** : Espace piscine en pieds carrés
- **PoolQC** : Qualité de la piscine
- **Clôture** : Qualité clôture
- **MiscFeature** : **Fonctionnalité diverse** non couverte dans d'autres catégories
- **MiscVal** : \$Value de la fonction diverse
- **MoSold** : Mois vendu
- **YrSold** : Année de vente
- **SaleType** : Type de vente
- **SaleCondition** : Condition de vente