

LIVE CODING : Les Fondamentaux du Machine Learning

De la Donnée à la Décision : Votre Première Expédition en IA

NFALY SYLLA

CODE "MATRIX" | Élève Ingénieur Génie Informatique, ENSA Tanger

Samedi 28 Juin 2025

Notre Programme de la Soirée

Grandes Étapes de notre Voyage

- 1 **La Grande Image** : L'écosystème de la Data et les types de problèmes en IA.
- 2 **Application 1 - Découverte** : Notre premier modèle de prédiction avec un mini-dataset.
- 3 **Application 2 - Plongée** : Analyse et modélisation d'un vrai dataset Kaggle.
- 4 **Conclusion & Prochaines Étapes** : Comment continuer votre aventure.

L'objectif : démystifier, pratiquer, et inspirer !

Ce que nous allons (et n'allons pas) faire

✓ Objectifs

- Démystifier le Machine Learning.
- Explorer des données de A à Z.
- **Coder en direct** deux modèles.
- Comprendre les intuitions derrière les algos.
- Vous donner les clés pour continuer.

✗ Hors-Périmètre

- Devenir des experts en 2h.
- Plonger dans les maths complexes.
- Parler de déploiement avancé (MLOps).

Les Types de Problèmes en Machine Learning

Régression

Question : Combien ?

Ex : Prédire le prix d'une maison.

Classification

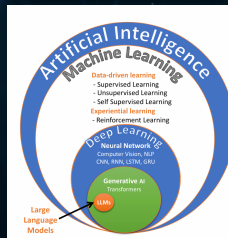
Question : Lequel/Laquelle ?

Ex : Déterminer si un e-mail est un spam.

Clustering

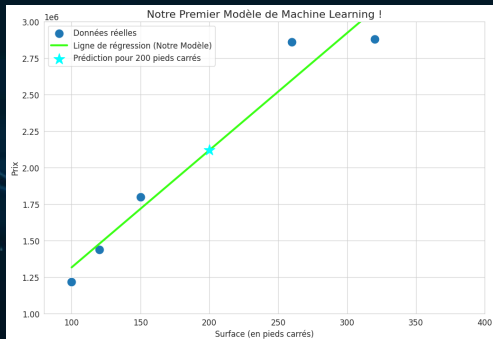
Question : Quels groupes ?

Ex : Segmenter des clients.



Ce soir, nous nous concentrons sur la Régression !

Notre Premier Défi : 5 Maisons



Objectif

Comprendre l'intuition de la Régression Linéaire de manière simple et visuelle.

- Visualiser
- Entraîner
- Prédire

Live Coding : Le Code pour notre 1er Modèle

Préparation et Entraînement

```
1 # Dataset simple
2 df_simple = pd.read_csv('Housing_Simple.csv')
3
4 # Préparation des données X et y
5 X_simple = df_simple[['area']]
6 y_simple = df_simple['price']
7
8 # Entraînement du modèle
9 model_simple = LinearRegression()
10 model_simple.fit(X_simple, y_simple)
```

Live Coding : Le Code pour notre 1er Modèle

Préparation et Entraînement

```
1 # Dataset simple
2 df_simple = pd.read_csv('Housing_Simple.csv')
3
4 # Préparation des données X et y
5 X_simple = df_simple[['area']]
6 y_simple = df_simple['price']
7
8 # Entraînement du modèle
9 model_simple = LinearRegression()
10 model_simple.fit(X_simple, y_simple)
```

Prédiction

Pour une maison de 200m², notre modèle prédit : **211, 9869 €**

Le Défi Kaggle : "House Prices"



Mission : Prédire le prix de vente final

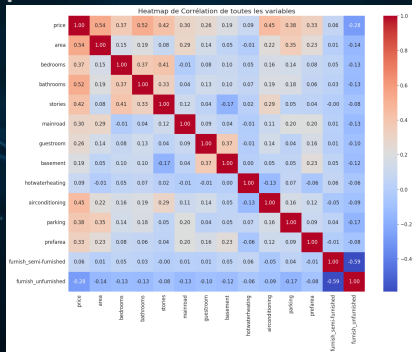
Dataset réel et complexe avec :

- 545 lignes d'entraînement
- 13 caractéristiques (colonnes)
- Données manquantes, textuelles, numériques...

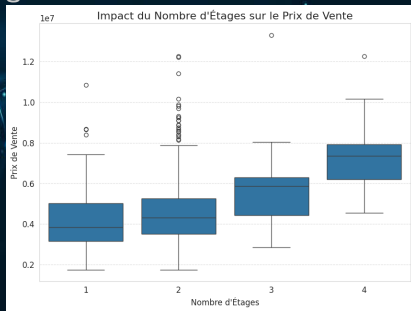
On passe du jouet à l'outil professionnel.

Les "insights" clés

Variables les plus corrélées au prix :



Importance de la qualité générale :



Le Workflow Complet en Pratique

- 1 **Data Cleaning** : Gérer les valeurs manquantes.
- 2 **Feature Engineering** : Transformer les données textuelles en numériques (One-Hot Encoding).
- 3 **Évaluation** : Mesurer la performance avec des métriques adaptées.
- 4 **Interprétation** : Analyser l'importance des caractéristiques (Feature Importance).

Tout cela sera détaillé dans le notebook Colab partagé !

Votre Aventure ne fait que commencer...

Ce que vous avez accompli ce soir

Exploré, nettoyé, visualisé, modélisé, et interprété des données. Vous avez vu le workflow complet d'un projet de Data Science.

Prochaines étapes pour devenir un Padawan de l'IA

- **Pratiquer** : Participer à des compétitions sur **Kaggle**.
- **Apprendre** : Suivre des cours en ligne (Coursera, edX, fast.ai).
- **Construire** : Créer vos propres projets et les partager sur GitHub.

Merci pour votre attention !

Place à vos questions.
Restons connectés et continuons la discussion !



/in/nfaly-sylla



/rmss00-2synf/aids-repo

Les ressources sont sur GitHub.



@sylla.rmss