

DISCOURS ORAL - DÉTECTION DE FAUX BILLETS

Timing: 12-15 minutes

SLIDE 1 - TITRE (30 sec)

"Bonjour à tous, je m'appelle Maryline. Aujourd'hui, je vais vous présenter un projet qui m'a fascinée : **comment utiliser l'intelligence artificielle et le machine learning pour détecter automatiquement les faux billets en euros.**

C'est un sujet qui touche à la sécurité monétaire de notre pays, et c'est exactement le genre de défi que nous, data analysts, pouvons résoudre avec les bonnes données et les bonnes techniques."

SLIDE 2 - CONTEXTE ET OBJECTIF (1 min 30)

"Commençons par le contexte. L'ONCFM, l'Organisation Nationale de Lutte contre le Faux-Monnayage, fait face à un problème croissant : **chaque année, les faux billets en circulation augmentent.**

Aujourd'hui, la détection se fait manuellement : des agents vérifient billet par billet. C'est **coûteux, lent, et peu fiable**. Il y a un besoin urgent d'automatiser ce processus.

C'est là qu'entre en jeu mon projet. **L'objectif est simple mais ambitieux** : développer un algorithme capable de dire si un billet est vrai ou faux... **juste en regardant ses dimensions géométriques** — sa longueur, ses hauteurs, ses marges, sa diagonale.

Ça peut sembler fou, mais vous verrez qu'on peut faire des miracles avec juste 6 mesures et un peu de machine learning."

SLIDE 3 - LES DONNÉES (1 min)

"Regardons les données qu'on a. C'est un dataset classique mais vraiment bien équilibré :

- **1500 billets au total**
- **1000 vrais billets**

- **500 faux billets**
- **6 variables géométriques** qui décrivent chaque billet

Ces données viennent d'un vrai cas d'usage — pas synthétique, pas inventé. C'est du vrai travail d'expert monétaire. Et vous verrez que c'est largement suffisant pour créer un modèle performant."

SLIDE 4 - PRÉPARATION ET EXPLORATION (1 min 30)

"Avant de construire nos modèles, j'ai fait ce qu'on appelle l'EDA — l'Exploratory Data Analysis.

J'ai nettoyé les données, j'ai regardé les statistiques descriptives : moyennes, écarts-types, valeurs extrêmes. Puis j'ai créé des visualisations — histogrammes, boîtes à moustaches — pour comprendre comment se distribuent les vraies mesures par rapport aux fausses.

Ce qui a émergé, c'est qu'il y avait **des patterns clairs** : les faux billets ne reproduisent pas parfaitement les dimensions des vrais. C'était bon signe pour nos modèles."

SLIDE 5 - CORRÉLATIONS ET PATTERNS (1 min)

"Un résultat clé que j'ai trouvé : **les dimensions géométriques sont des prédicteurs fiables**.

Concrètement, ça veut dire que les 6 variables qu'on a — longueur, hauteur gauche, hauteur droite, marge haute, marge basse, diagonale — sont fortement corrélées avec l'authenticité du billet.

Il n'y a pas besoin de variables complexes ou d'images haute résolution. **Les simples dimensions suffisent**. C'est élégant, c'est efficace, et c'est facilement déployable sur le terrain."

SLIDE 6 - MÉTHODOLOGIE ET MODÈLES (1 min 30)

"Passons à la méthodologie. J'ai testé plusieurs approches :

D'abord l'apprentissage supervisé — la régression logistique pour une classification binaire (vrai ou faux), plus KNN et Random Forest comme alternatives.

Puis l'approche non supervisée — K-means, pour voir si on pouvait retrouver naturellement 2 groupes dans les données.

Pour tous les modèles, j'ai suivi le même protocole :

- 80% des données pour l'entraînement
- 20% pour le test (pour pas tricher)
- Évaluation avec accuracy, précision, rappel, F1-score, et matrices de confusion.

C'est standard, c'est robuste, c'est traçable."

SLIDE 7 - RÉSULTATS (2 min)

"Et maintenant, les résultats — c'est là que ça devient intéressant.

[Pointer les chiffres]

Tous les modèles sont excellents — accuracy entre 98,3% et 99%. Mais regardez en détail :

- **Régression logistique : 99% d'accuracy** — c'est notre meilleure, avec **seulement 2 faux billets** pris pour vrais sur 300 testés. C'est critique pour l'ONCFM.
- **Random Forest : 98,6%** — aussi bon, mais 3 faux billets manqués.
- **KNN : 98,3%** — légèrement moins bon, 3 faux billets manqués aussi.
- **K-means : 98,6%** — performant, mais c'est non supervisé, donc utile pour explorer mais pas pour la production.

La régression logistique est clairement la gagnante. Et c'est intéressant parce que c'est le modèle le plus simple, le plus rapide à entraîner, et le plus facile à interpréter. C'est parfait pour une mise en production."

SLIDE 8 - RECOMMANDATIONS ET DÉPLOIEMENT (1 min 30)

"Voici donc mes recommandations officielles :

J'ai retenu la régression logistique comme modèle final. Pourquoi ?

1. **Seuls 2 faux billets** pris pour vrais sur 300 — c'est extraordinaire.
2. **Meilleure performance globale** — 99% d'accuracy.

3. **Elle minimise les faux positifs** — et c'est PRIORITAIRE pour l'ONCFM. Un faux billet qui passe inaperçu, c'est un désastre opérationnel.
4. **Le modèle est simple et interprétable** — l'ONCFM peut comprendre comment il prend ses décisions.
5. **Il est robuste et reproductible** — on peut le retrainner, l'auditer, le valider.

C'est un modèle production-ready."

SLIDE 9 - SCRIPT / MISE EN PLACE (1 min 30)

"Parlons d'implémentation. Concrètement, voici comment ça fonctionne :

Étape 1 : On charge le modèle sauvegardé en mémoire — c'est un fichier .joblib qu'on génère une fois pendant l'entraînement.

Étape 2 : On reçoit les données — les 6 mesures d'un nouveau billet.

Étape 3 : On exécute la prédiction. Le modèle nous dit :

- La classe prédite : « VRAI » ou « FAUX »
- La probabilité associée : par exemple, 99,8% de confiance que c'est un vrai billet
- Un rapport de confiance global

C'est rapide — millisecondes.

C'est fiable — on a testé ça 300 fois et ça marche à 99%.

C'est facile à intégrer — n'importe quel agent ou système automatisé peut l'utiliser."

SLIDE 10 - CONCLUSION (1 min)

"En résumé :

Ce projet a démontré qu'on **PEUT** créer un système automatisé et fiable pour détecter les faux billets.

Juste avec les dimensions géométriques.

La régression logistique, avec ses 99% d'accuracy et seulement 2 erreurs sur 300 billets, offre une solution **simple, rapide et prête à être déployée** par l'ONCFM.

Mais c'est juste le début. À l'avenir, on pourrait :

- Intégrer ce modèle dans des scanners physiques aux frontières
- Ajouter de nouvelles features comme la texture ou des images haute résolution
- Renforcer la robustesse du système face à des formes de contrefaçon nouvelles

C'est un exemple de comment le machine learning peut résoudre des problèmes réels, opérationnels, et impactants."

SLIDE 11 - THANK YOU (30 sec)

"Merci beaucoup de votre attention. Je suis maintenant disponible pour répondre à vos questions — sur la méthodologie, les modèles, ou les perspectives de déploiement."

⌘ TIMING TOTAL : 12-13 minutes (avec les slides lues à rythme naturel)