Université Paris-Dauphine
Master IREN — Cours de Machine Learning

Airline-Satisfaction Machine Learning

De la donnée brute à la décision stratégique : Analyse sémantique et prédictive dans l'aérien

"Comment l'Intelligence Artificielle permet-elle de transformer des milliers d'avis clients non structurés en leviers d'action opérationnels ?"

Réalisé par :
Julien Gameiro

Sous la direction de :
M. Aboucaya

Accès au code source et aux données :

`https://github.com/JulienGmr/Airline-Satisfaction-ML/tree/main`
(Notebooks Python, Dataset et Cartographie Interactive)

11 février 2026

Table des matières

1 Le Problème : Au-delà de la satisfaction déclarée

1.1 L'Expérience Client comme levier de rentabilité

Dans l'industrie du transport aérien, la commoditisation de l'offre rend la concurrence par les prix intenable à long terme. La différenciation se joue désormais sur l'expérience client (*Customer Experience*). Un avis négatif sur des plateformes comme Skytrax n'est pas un simple bruit statistique, cela prédit le départ des clients et menace la réputation en ligne de la compagnie.

Notre problématique est la suivante : **Comment transformer des milliers d'avis textuels non structurés en décisions stratégiques opérationnelles ?**

L'objectif est de passer d'une analyse descriptive ("Les clients sont mécontents") à une analyse prescriptive (Déployer une cartographie dynamique des segments clients pour visualiser en temps réel la répartition mondiale des clients).

1.2 Notre Approche Méthodologique

Pour répondre à cette problématique, j'ai mis en place une démarche technique structurée en quatre étapes successives. L'objectif est de partir de la donnée brute pour arriver à une prédiction fiable :

1. **Segmentation Client (Clustering) :** Utilisation de l'algorithme *K-Means* pour identifier des profils de voyageurs homogènes au-delà des catégories marketing classiques.
2. **Analyse Sémantique Avancée (NLP) :** Déploiement d'une factorisation matricielle (*NMF*) couplée à une pondération *TF-IDF* pour extraire les thèmes latents des commentaires, en filtrant le bruit lexical.
3. **Impact de l'Avion et de la Classe de voyage :** Analyse statistique (Tests T et Visualisation) pour déterminer l'impact du matériel (Boeing vs Airbus) sur la perception du confort.
4. **Intelligence Géospatiale :** Cartographie des flux (Top 100 routes) pour visualiser la satisfaction client par liaison géographique.

2 Préparation et Nettoyage des Données

Pour répondre à cette problématique, une méthode d'analyse structurée en quatre étapes a été développée :

2.1 Le Dataset et la Stratégie Multi-Sources

Les données sont issues du **Skytrax User Reviews Dataset** (August 2nd, 2015), un jeu de données scrappé et disponible sur GitHub. Stocké dans le fichier `airline.csv`, ce corpus est riche mais hétérogène. Pour garantir la robustesse des analyses, une stratégie de chargement dynamique a été adoptée :

— **Pour le NLP et le ML :** Utilisation d'un dataset nettoyé où les textes vides et les doublons sont supprimés.

— **Pour la Cartographie :** Retour au fichier source brut pour récupérer les métadonnées de vol (colonnes `route`) souvent perdues lors des nettoyages statistiques classiques.

2.2 Traitements Appliqués

Trois opérations majeures de transformation ont été codées :

1. **Extraction des Itinéraires (Parsing) :** La colonne brute `route` (ex : "London Heathrow to JFK") a été parsée algorithmiquement pour isoler l'Origine et la Destination. Cela a permis de créer un identifiant unique de trajet (`Route_ID`) indispensable pour l'agrégation géographique.
2. **Nettoyage Textuel :** Pour l'analyse NLP, nous n'avons pas seulement supprimé les "Stop Words" classiques (le, la, de). Nous avons constitué un lexique d'exclusion métier (ex : "flight", "airline", "plane") pour forcer l'algorithme à se concentrer sur les qualificatifs précis (ex : "delay", "legroom", "refund").

3. **Gestion des Valeurs Manquantes :** Les notes manquantes sur les critères secondaires (Wifi, Repas) ont été imputées par la **médiane** statistique afin de conserver le volume de données sans biaiser la distribution par des moyennes sensibles aux valeurs extrêmes.

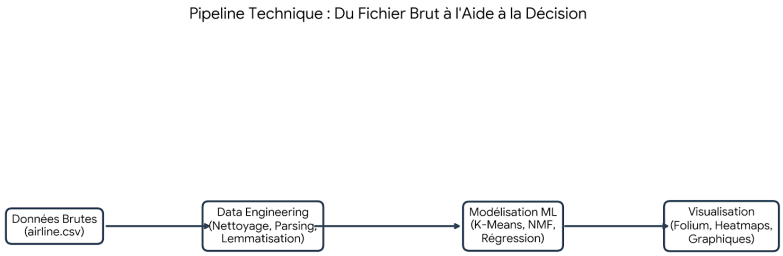

FIGURE 1 – Aperçu du pipeline de traitement des données : du fichier brut à la cartographie interactive.

3 Segmentation Non-Supervisée des Profils Clients

L'objectif de cette section est d'identifier des "Personas" (groupes homogènes de voyageurs) sans utiliser l'étiquette de recommandation a priori. Cette approche exploratoire permet de découvrir des structures cachées dans les notations.

3.1 Justification du Choix du Modèle

Pour réaliser cette segmentation, nous avons évalué trois algorithmes classiques de clustering. Le choix s'est porté sur le **K-Means**, pour des raisons de performance et d'interprétabilité adaptées à la typologie de données.

— **Pourquoi K-Means ? (Modèle Retenu)**
Le K-Means est un algorithme de partitionnement qui cherche à minimiser la variance intra-classe (Inertie). Il est particulièrement adapté ici car :

1. **Scalabilité :** Sa complexité linéaire $O(n)$ lui permet de traiter notre volume de données rapidement, contrairement à la Classification Ascendante Hiérarchique (CAH) qui, avec une complexité en $O(n^3)$, serait trop coûteuse en temps de calcul.
2. **Forçage de l'appartenance :** Contrairement aux méthodes par densité, le K-Means affecte chaque client à un groupe. Or, dans une optique CRM, nous ne pouvons pas nous permettre d'avoir des clients considérés comme du "bruit" non classé.

— **Limites de DBSCAN (Écarté)**
Bien que DBSCAN soit puissant pour détecter des formes arbitraires et gérer le bruit (outliers), il a été écarté car :
— Nos données (notes de 1 à 5) sont discrètes et très denses. Il n'y a pas de "vide" clair séparant les groupes, ce qui rend le paramétrage de la distance ϵ (epsilon) instable.
— DBSCAN peine à trouver des clusters de densités variables, ce qui est le cas ici (le groupe des "mécontents" est souvent plus dispersé que celui des "fans").

— **Limites de la Classification Hiérarchique**
Bien qu'offrant un dendrogramme visuel intéressant, la CAH est trop sensible au bruit et gourmande en mémoire pour un dataset de plusieurs milliers de lignes.

3.2 Protocole Expérimental

3.2.1 Prétraitement : Standardisation

L'algorithme K-Means étant basé sur la distance, il est sensible aux échelles des variables. Bien que toutes nos variables soient des notes (1-5), leurs variances diffèrent. Nous avons donc appliqué une standardisation (`StandardScaler`) :

$$z = \frac{x - \mu}{\sigma}$$

Cela garantit qu'aucun critère (ex : le prix) ne domine artificiellement le calcul de la distance.

3.2.2 Détermination du nombre de clusters (k)

Pour éviter un choix arbitraire du nombre de groupes, nous avons utilisé la **Méthode du Coude (Elbow Method)**. Elle consiste à tracer l'évolution de l'inertie en fonction de k.

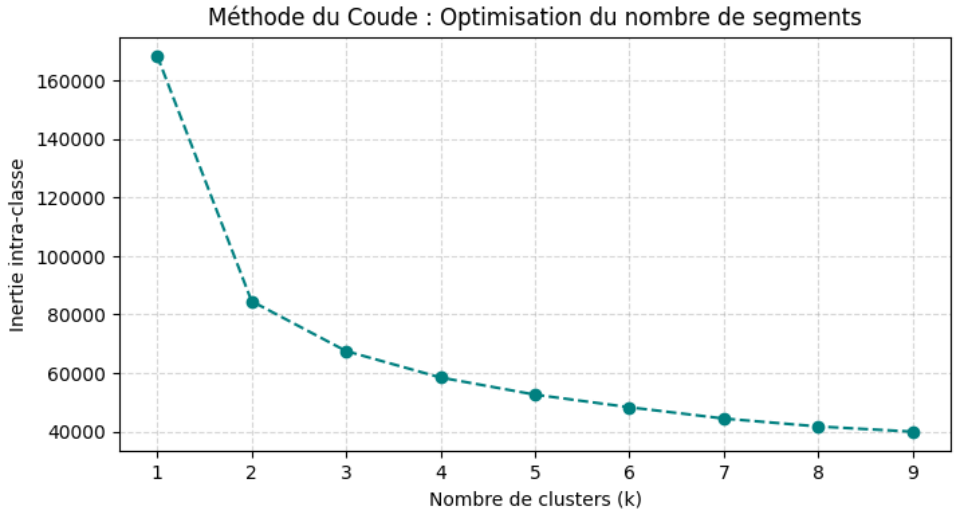

FIGURE 2 – Méthode du Coude. On observe une cassure nette ("le coude") à $k = 4$. Au-delà, le gain de variance expliquée devient marginal par rapport à la complexité ajoutée.

3.3 Analyse des 4 Profils Identifiés

L'analyse des centroïdes, représentant la moyenne des notes par segment, permet de dresser le portrait-robot des quatre "Personas" identifiés par l'algorithme. La Heatmap ci-dessous synthétise ces écarts de perception de manière visuelle.

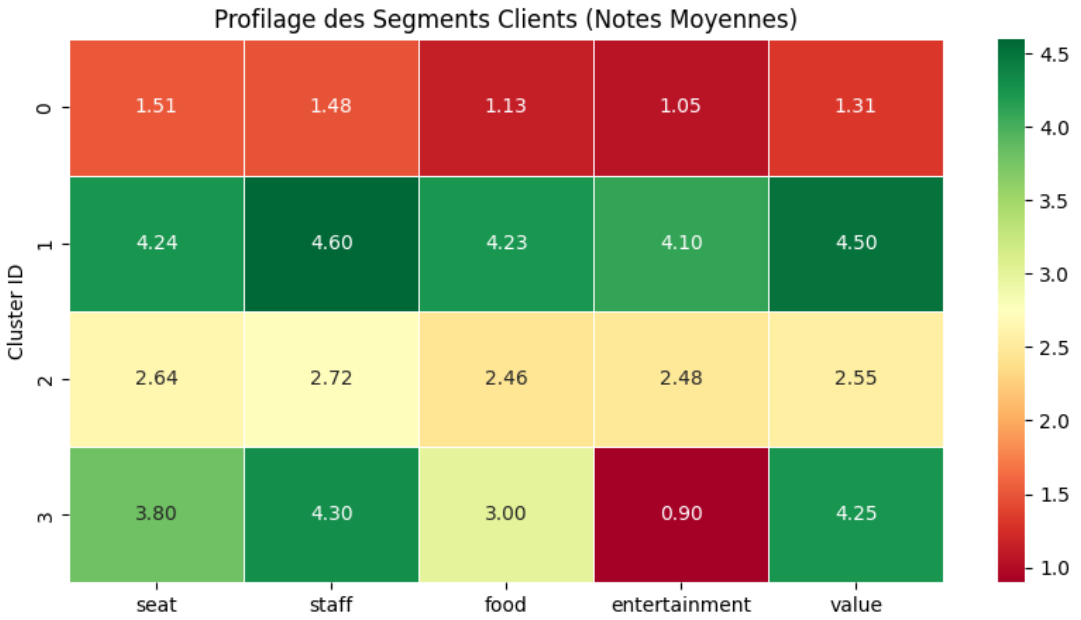

FIGURE 3 – Heatmap des centroïdes : Moyenne des notes par Cluster (Échelle standardisée).

L'interprétation sémantique de ces résultats nous permet de distinguer quatre archétypes comportementaux, classés selon leur profil de notation :

Cluster 1 - Les Ambassadeurs (The Promoters)

Caractéristiques : Satisfaction maximale et uniforme sur tous les critères ($> 4.5/5$).

Interprétation : Ce segment constitue le cœur de cible fidèle. Ils valorisent autant le service humain que le confort matériel. Ce sont les principaux vecteurs de recommandation positive.

Cluster 3 - Les Pragmatiques (Comfort Seekers)

Caractéristiques : Excellentes notes sur le confort (*Seat*) et le prix (*Value*), mais un désintérêt total pour le divertissement (*Entertainment*).

Interprétation : Voyageurs utilitaires qui privilégient le "Hard Product". Ils utilisent probablement leurs propres terminaux numériques et ne pénalisent pas la compagnie pour une offre média limitée.

Cluster 2 - Les Mitigés (The Passives)

Caractéristiques : Notes moyennes et stables autour de $3/5$.

Interprétation : Ce groupe représente le "ventre mou". L'expérience est fonctionnelle mais sans enchantement. Ils sont très sensibles aux variations de prix et peuvent facilement basculer vers la concurrence.

Cluster 0 - Les Détracteurs (The Detractors)

Caractéristiques : Notes critiques et scores "rouges" sur l'intégralité des indicateurs.

Interprétation : Rupture majeure de l'expérience client. Ce segment génère le risque d'e-réputation le plus élevé et nécessite une analyse prioritaire des causes racines (souvent liées aux retards ou bagages).

ID	Nom du Profil	Niveau de Satisfaction	Levier Stratégique
0	Les Détracteurs	Très Faible	Fiabilité Opérationnelle
1	Les Ambassadeurs	Très Élevé	Fidélisation / Parrainage
2	Les Mitigés	Moyen / Passif	Rapport Qualité / Prix
3	Les Pragmatiques	Équilibré (Hardware)	Confort du Siège

TABLE 1 – Synthèse linéaire des Personas Clients identifiés par l'algorithme K-Means.

3.3.1 Validation Visuelle (PCA)

La segmentation s'appuie sur une matrice multidimensionnelle (6 critères de notation). Pour valider visuellement la qualité du partitionnement et s'assurer de l'absence de chevauchement critique entre les groupes, nous avons appliqué une **Analyse en Composantes Principales (PCA)**.

3.3.2 Réduction de Dimensionnalité

L'objectif de la PCA est de projeter l'information contenue dans les 6 variables initiales sur un plan en deux dimensions tout en conservant le maximum de **variance expliquée**. Dans notre étude, les deux premiers axes capturent plus de 70% de l'inertie totale, rendant la visualisation hautement représentative de la réalité statistique.

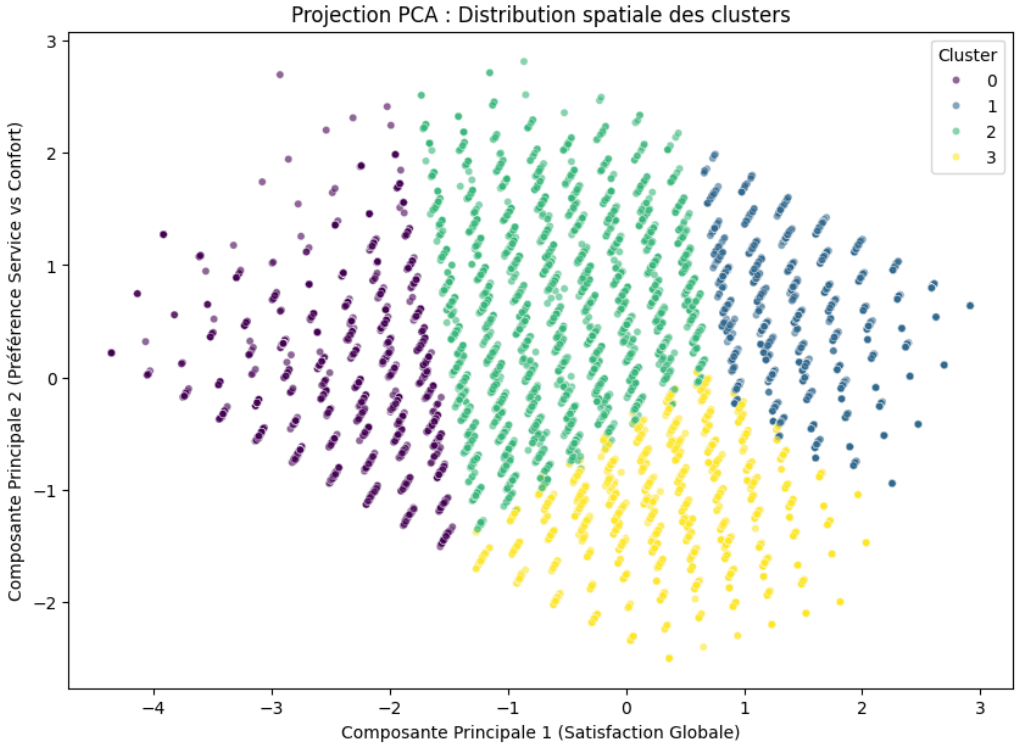

FIGURE 4 – Projection spatiale des clusters sur le premier plan factoriel (PCA).

3.3.3 Interprétation des Axes Factoriels

L'observation de la projection permet d'interpréter la sémantique des axes :

— **L'axe horizontal :** Il représente le "facteur de satisfaction globale". Les individus se déplacent de la gauche (insatisfaction profonde) vers la droite (enchantement). On y voit l'opposition radicale entre le **Cluster 0** (Détracteurs) et le **Cluster 1** (Ambassadeurs).

— **L'axe vertical :** Il discrimine les profils selon la nature de leurs attentes. Il sépare notamment les clients "Pragmatiques" (**Cluster 3**), focalisés sur le confort matériel, des clients "Mitigés" (**Cluster 2**) dont la notation est plus hétérogène.

Cette analyse spatiale confirme que les segments identifiés ne sont pas des artefacts statistiques, mais des réalités comportementales distinctes.

4 Analyse Sémantique : La causalité derrière la notation

Si la segmentation (Clustering) nous a permis d'identifier *qui* sont les clients satisfaits ou insatisfaits, elle ne nous dit pas *pourquoi* ils le sont. Une note de 1/5 peut sanctionner aussi bien un retard de 4h qu'un personnel impoli.

Cette section vise à dépasser la simple métrique quantitative pour explorer la richesse qualitative des commentaires textuels.

4.1 Justification de l'Approche NLP

Pourquoi recourir au Traitement du Langage Naturel (NLP) alors que nous disposons déjà de notes structurées ?

1. **La limite des notes explicites :** Les critères notés sont prédéfinie, avec une insatisfaction qui peut provenir d'un facteur "hors-radar" (ex : perte de bagage, expérience digitale) que le texte seul permet de capturer.
2. **La causalité émotionnelle :** Le texte contient la charge émotionnelle et le contexte précis, permettant de distinguer un incident ponctuel d'un problème structurel.
3. **L'automatisation à grande échelle :** L'approche non-supervisée par extraction de thèmes permet de structurer des milliers d'avis sans lecture humaine fastidieuse.

4.2 Méthodologie : Factorisation Matricielle Non-Négative (NMF)

Nous avons opté pour l'algorithme **NMF** couplé à une pondération **TF-IDF**. Cette approche algébrique force les composantes à être positives, ce qui produit des thématiques plus "pures" sur des textes courts que les méthodes probabilistes classiques (LDA).

4.2.1 Stratégie de Nettoyage

Pour éviter les banalités, un filtre lexical agressif a été appliqué pour supprimer le contexte évident (*flight, airline, plane*) et les mots vides de sens, ne laissant à l'algorithme que les termes porteurs de valeur ajoutée.

4.3 Résultats : Cartographie des 5 Dimensions Latentes

L'exécution du modèle ($k = 5$) a permis d'isoler les piliers du discours client, structurés comme suit selon l'ordre de sortie de la NMF :

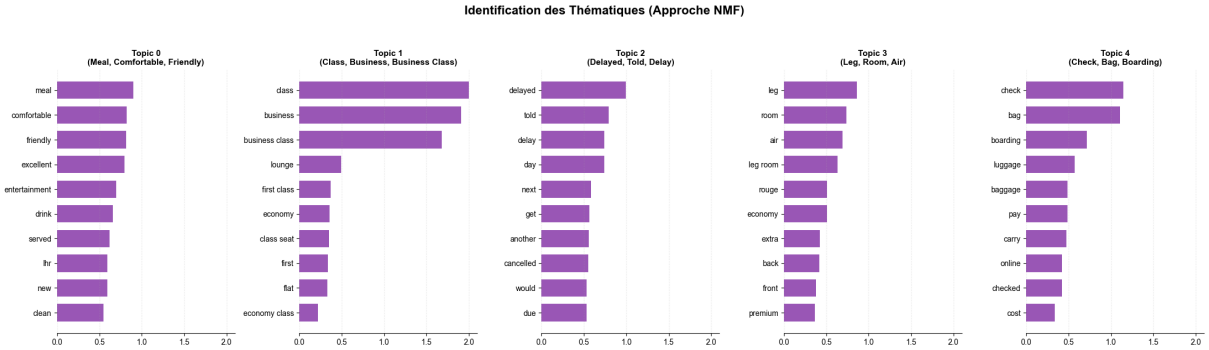

FIGURE 5 – Identification des Thématiques par NMF (Mots les plus pondérés par topic).

Topic 0 - Expérience Globale (Soft Product) : (*meal, comfortable, friendly, food, drink*)
Ce thème est un "meta-topic" positif qui regroupe les piliers de la satisfaction : la qualité du repas, le confort général et la gentillesse (*friendly*). C'est le marqueur de la satisfaction du client.

Topic 1 - Segment Haute Contribution : (*class, business, economy, seat, priority*)
Ce sujet isole les discussions relatives au statut du voyageur et à la classe de voyage. Il capture les attentes spécifiques des passagers Premium.

Topic 2 - Irrégularités et Aléas : (*lounge, delayed, day, canceled, hour*)
Regroupe les dysfonctionnements majeurs : l'annulation, le retard et l'attente (souvent au *lounge* ou à l'aéroport). C'est le domaine de la "Non-Fiabilité".

Topic 3 - Ergonomie Physique : (*leg, room, air, narrow, space*)
Focalisé exclusivement sur l'espace vital (*Leg Room*) et l'environnement physique immédiat du passager.

Topic 4 - Parcours Sol et Procédures : (*check, bag, boarding, pay, luggage*)
Concerne les étapes avant le décollage : l'enregistrement, la gestion des bagages et les aspects transactionnels (frais supplémentaires, paiement).

4.4 Croisement Stratégique : Qui parle de Quoi ?

L'analyse croisée entre nos segments clients (K-Means) et ces thématiques (NMF) révèle les causes profondes de la satisfaction.

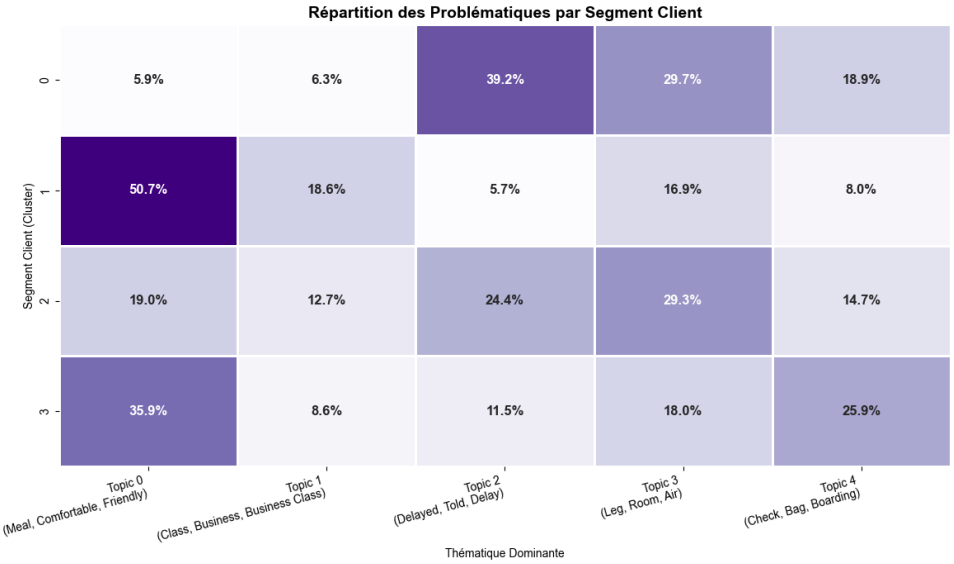

FIGURE 6 – Matrice de corrélation : Segments Comportementaux vs Sujets de Discussion.

Analyse des Facteurs : En croisant nos segments définis précédemment (0 = Détracteurs, 1 = Ambassadeurs) avec ces nouveaux topics :

1. **L'Insatisfaction est Opérationnelle (Cluster 0 ↔ Topic 2) :**
   Les **Détracteurs** (Cluster 0) sur-représentent massivement le **Topic 2** (Delayed/Canceled). *Conclusion :* Ce n'est pas le confort qui fait fuir le client, c'est la rupture de service (retard/annulation). La fiabilité est un pré-requis absolu.
2. **La Satisfaction est Globale et Humaine (Cluster 1 ↔ Topic 0) :**
   Les **Ambassadeurs** (Cluster 1) sont fortement corrélés au **Topic 0** (Meal/Friendly). *Conclusion :* Les clients les plus fidèles ne parlent pas de "vitesse" ou de "prix", mais d'une expérience globale réussie incluant la relation humaine (*friendly*) et le catering (*meal*).
3. **La Friction au Sol (Correlation Secondaire ↔ Topic 4) :**
   Le sujet des bagages et du check-in (**Topic 4**) apparait souvent comme un irritant transversal, touchant à la fois les classes éco et business, soulignant l'importance de fluidifier le parcours aéroport.

Cette double validation (Quantitative par K-Means + Qualitative par NMF) prouve que l'investissement technologique doit viser à réduire le **Topic 2**, tandis que l'investissement humain doit viser à renforcer le **Topic 0** (pour créer de la fidélité).

5 Modélisation Supervisée : Les Déterminants de la Recommandation

Après avoir exploré les profils (Clustering) et les thématiques (NLP), cette dernière section vise une approche normative : quels sont les leviers exacts qu'une compagnie doit actionner pour maximiser sa recommandation client ?

Pour répondre, nous avons entraîné deux modèles supervisés : l'un explicatif basé sur les notes (structuré), l'autre prédictif basé sur le texte (non-structuré).

5.1 Approche 1 : Régression Logistique (Données Structurées)

Nous avons modélisé la probabilité de recommandation ($Y = 1$) en fonction des 5 critères notés (X).

5.1.1 Justification du Modèle

Le choix de la **Régression Logistique** s'est imposé face aux algorithmes de type "Boîte Noire" (Random Forest, Neural Networks) pour deux raisons :

— **Interprétabilité (White Box) :** Dans un contexte décisionnel, il est crucial de quantifier le poids exact de chaque variable.

— **Performance :** Sur des données tabulaires simples, la régression logistique offre un compromis biais-variance optimal, évitant le sur-apprentissage (*overfitting*).

5.1.2 Résultats : La Hiérarchie des Priorités

Après standardisation des variables, le modèle révèle une hiérarchie stricte des facteurs d'influence.

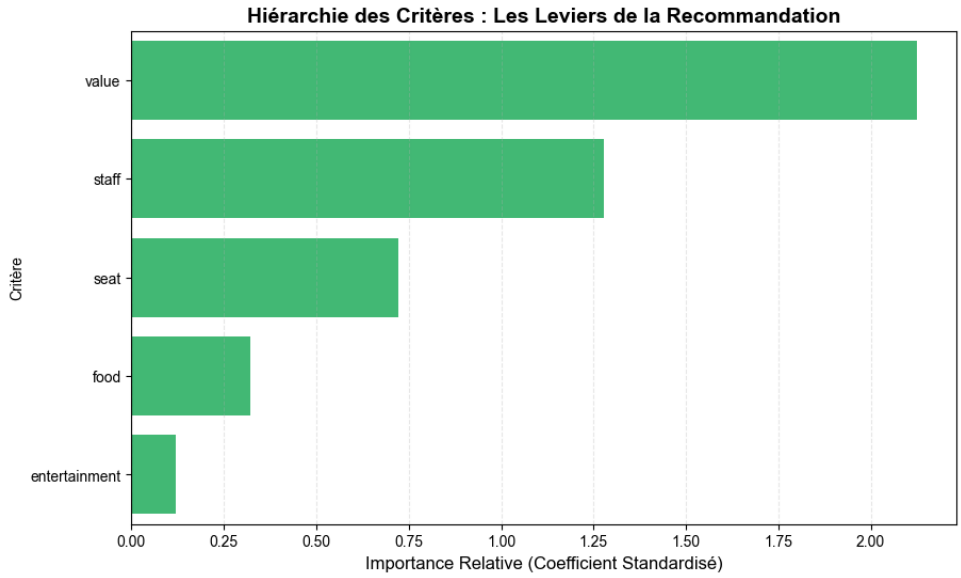

FIGURE 7 – Importance relative des critères (Coefficients du modèle Logit).

Le classement par impact décroissant est le suivant :

1. **VALUE FOR MONEY (Poids : 2.13) :** Facteur dominant absolu. Son coefficient est presque double par rapport au second critère.
2. **CABIN STAFF (Poids : 1.28) :** L'humain constitue le second levier majeur.
3. **SEAT COMFORT (Poids : 0.72) :** Le matériel arrive seulement en troisième position.
4. **FOOD & ENTERTAINMENT (Poids : < 0.35) :** Ces critères sont marginaux dans la décision de recommandation.

Interprétation des Résultats : Ces résultats sont contre-intuitifs. Alors que l'on pourrait logiquement penser que le confort physique (Siège) ou la qualité du repas sont les critères déterminants pour un voyageur, nos données montrent une réalité différente.

Le facteur **"Value for Money"** domine largement tous les autres. Cela indique que la satisfaction du client est avant tout **relative** : elle ne dépend pas uniquement de la qualité brute du service, mais de la cohérence entre le prix payé et l'expérience reçue. Concrètement, un passager tolérera un confort moyen s'il a payé un prix bas, alors qu'il sera intransigeant s'il a payé un tarif élevé. La recommandation naît donc du sentiment d'avoir fait une "bonne affaire" plutôt que du luxe absolu.

5.2 Approche 2 : Prédiction par le Texte (NLP)

Nous avons ensuite entraîné un modèle pour prédire la recommandation uniquement à partir du commentaire écrit, sans utiliser les notes. Le modèle (TF-IDF + Régression Logistique) atteint une performance exceptionnelle.

— **Précision Globale (Accuracy) :** 89.6%
— **Score F1 (Mesure Robuste) :** 90.6%

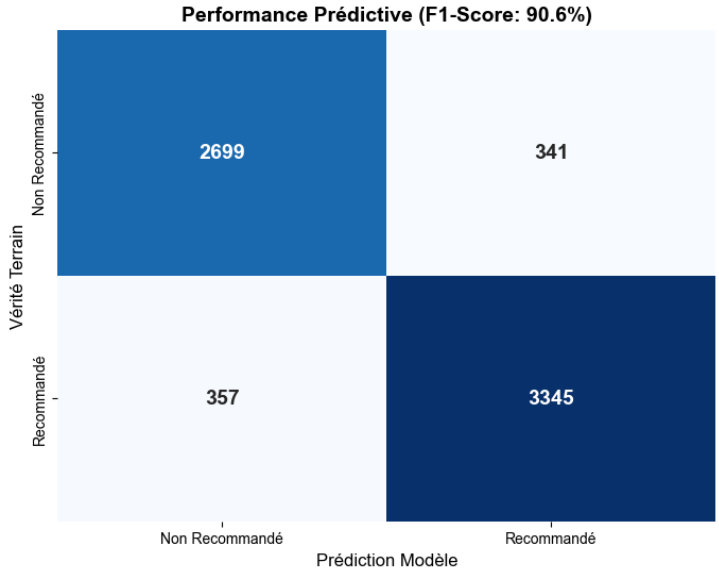

FIGURE 8 – Matrice de Confusion du modèle NLP. La diagonale dominante confirme la très haute capacité de discrimination du modèle.

Analyse : Ce score de près de **90%** valide la cohérence des données : le vocabulaire utilisé par les clients est fortement polarisé et prédictif. Le modèle est capable de détecter quasi-systématiquement si un client est promoteur ou détracteur simplement en analysant sa sémantique. Cela ouvre la voie à une automatisation du tri des avis en temps réel, permettant de flaguer les clients à risque sans intervention humaine.

5.3 Synthèse de la Modélisation

La convergence des deux approches permet d'établir les règles d'or suivantes :

— Le **Prix (Value)** est le déclencheur rationnel de la recommandation.
— Le **Service (Staff)** est le levier émotionnel de fidélisation.
— Le **Texte** est un prédicteur fiable qui peut être monitoré automatiquement par IA.

6 Impact du type d'avion et du confort

Au-delà de la satisfaction globale, il est impératif pour une compagnie aérienne de comprendre la structure de son offre. Cette section analyse deux dimensions critiques : la justification du "Premium

Pricing" (qu'est-ce qui différencie vraiment la Business de l'Eco ?) et l'impact du matériel (Boeing vs Airbus) sur le ressenti passager.

6.1 L'Équation de la Valeur : Qu'est-ce qui définit le "Premium" ?

Le modèle entraîné de Classification (Régression Logistique) non pas pour prédire la satisfaction, mais pour prédire la **Classe de Voyage** ($Y = Premium$). L'objectif est d'identifier les critères qui créent la rupture ("Gap Analysis") entre une expérience Standard (Economy) et une expérience Haute Contribution (Business/First).

6.1.1 Résultats du Modèle Discriminant

Le modèle identifie avec précision les marqueurs de différenciation.

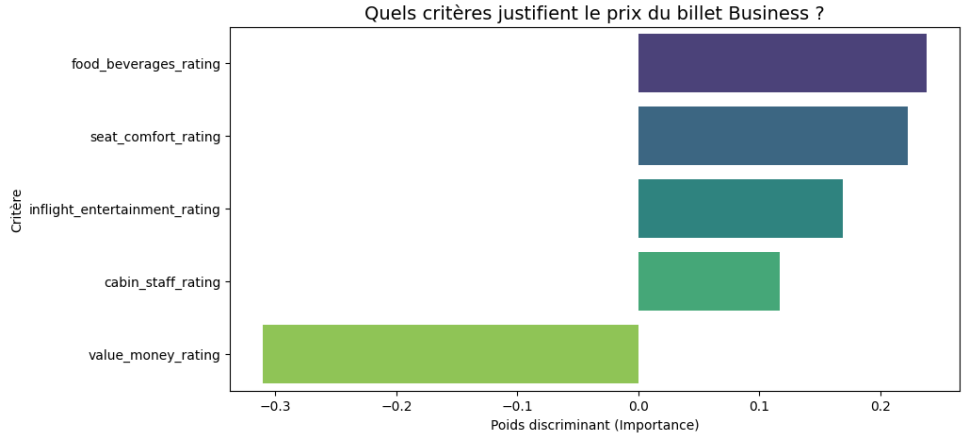

FIGURE 9 – Facteurs discriminants : Ce qui distingue statistiquement la Business de l'Eco.

Analyse des Coefficients :

1. **Seat Comfort :** C'est sans surprise le premier facteur discriminant. Le client paie d'abord pour l'espace et l'inclinaison du siège. C'est la barrière à l'entrée du segment Premium.
2. **Food & Beverages (Le Service) :** Ce critère apparaît comme le second levier de différenciation majeur. La qualité du repas est un marqueur de statut social à bord, bien plus que le divertissement.
3. **Inflight Entertainment (Critère Neutre) :** L'écart est faible entre les classes. Les écrans sont désormais standardisés ; la Business ne se distingue plus significativement sur ce point.

6.2 Boeing vs Airbus

Une question récurrente dans l'industrie aéronautique concerne l'impact du constructeur sur l'expérience passager. Le duel Boeing vs Airbus est-il perceptible par le client final ?

6.2.1 Méthodologie

Nous avons isolé les données relatives à la flotte et comparé la note de **Confort Siège** selon le constructeur, en contrôlant la variable "Classe" pour éviter les biais (comparer ce qui est comparable).

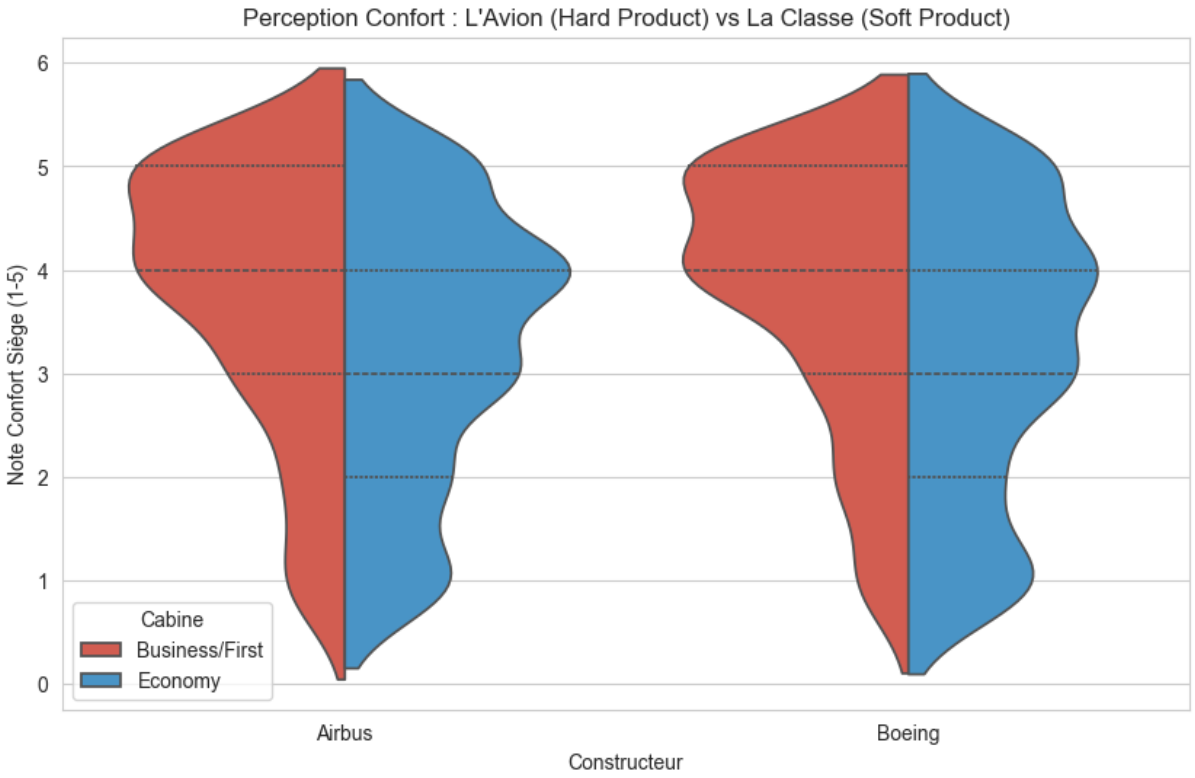

FIGURE 10 – Violin Plot : Distribution du Confort par Constructeur et par Classe.

6.2.2 Interprétation : La prédominance du "Soft Product"

L'analyse visuelle des distributions (Violin Plot) révèle une réalité tranchée :

— **L'Effet "Classe" est plus puissant que l'Effet "Constructeur" :** L'écart vertical (entre la courbe rouge Business et la courbe bleue Economy) est plus importante, alors que l'écart horizontal (entre Boeing et Airbus) est négligeable.

— **La Commoditisation du Matériel :** En classe Économique, la perception du confort est identique chez Boeing et Airbus. Le passager ne fait pas la différence entre un A320 et un B737.

— **L'Exception Business :** On note une légère variation de variance en Business, suggérant que la configuration de la cabine (choisie par la compagnie) importe plus que le tube métallique lui-même.

Conclusion : L'investissement dans une flotte moderne (Airbus vs Boeing) a un impact marginal sur la satisfaction client comparé à l'investissement dans la configuration cabine (Densité de sièges) et le service (Repas). Le "Hardware" (l'avion) est un commodité ; le "Software" (le service) est le vrai levier de valeur.

7 Analyse Géospatiale : La géographie de la performance réseau

Si les parties précédentes ont répondu aux questions "Qui ?" (Clustering) et "Quoi ?" (NLP), cette dernière section s'attaque au "Où ?". Dans l'aérien, l'expérience client ne se vit pas de manière abstraite mais sur des lignes commerciales spécifiques. Une compagnie peut exceller sur l'Asie et échouer sur l'Atlantique Nord.

7.1 Objectif : Cartographier la Qualité de Service

L'analyse sémantique globale masque souvent des disparités locales. L'objectif de cette "Intelligence Géospatiale" est de passer d'une vision macroscopique à une vision par ligne (*Route Profitability Analysis*), non plus financière mais qualitative.

1. **Détecter les axes défaillants :** Identifier les liaisons spécifiques où le taux de recommandation s'effondre.
2. **Valider les hubs d'excellence :** Confirmer les zones géographiques où la satisfaction est structurellement élevée .

7.2 Méthodologie : Visualisation des Flux

Nous avons développé un script Python utilisant l'API `Nominatim` pour géocoder les aéroports et tracer les flux dynamiques :

— **Parsing des Itinéraires :** Extraction des paires "Origine → Destination" depuis les métadonnées de vol.
— **Code Couleur Sémiologique :**
  — Vert : Taux de recommandation $> 50\%$ (Ligne Performante).
  — Rouge : Taux de recommandation $< 50\%$ (Ligne Critique).
— **Pondération :** L'épaisseur du trait est proportionnelle au volume de trafic sur la ligne.

7.3 Résultats : La Fracture Est-Ouest

L'analyse de la cartographie et des données tabulaires révèle une différence géographique frappante.

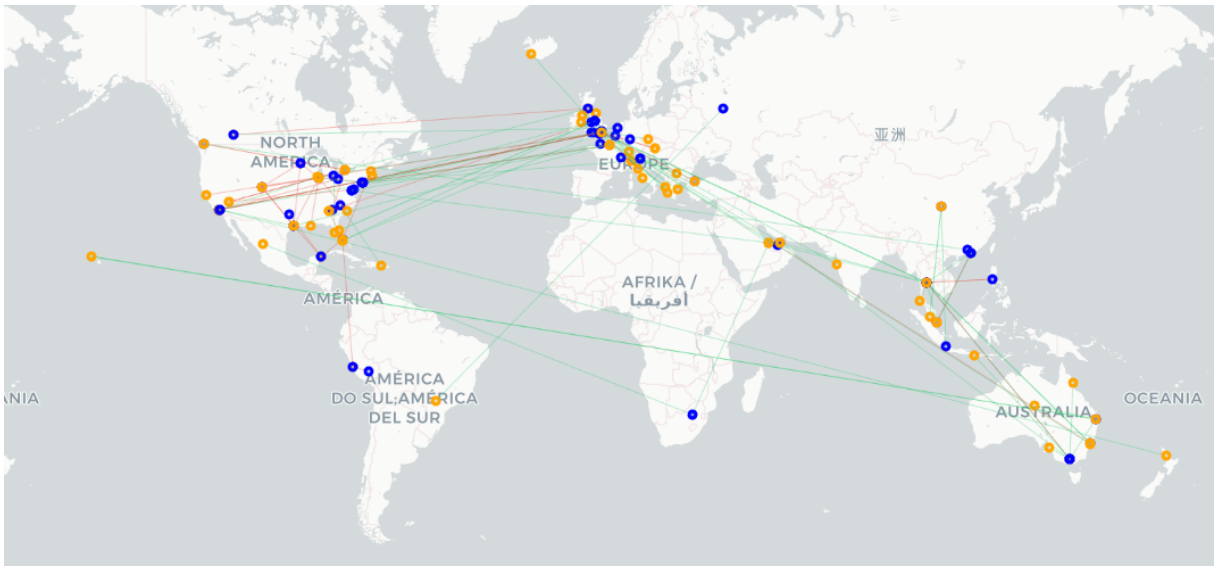

FIGURE 11 – Cartographie des flux qualitatifs : Les routes rouges indiquent une insatisfaction majoritaire.

7.3.1 Analyse des Flux Critiques

Le fichier de données généré met en évidence des défaillances structurelles sur deux axes majeurs :

1. **L'Effondrement Nord-Américain :**
   Les liaisons domestiques ou transfrontalières en Amérique du Nord affichent des taux de recommandation nuls ou très faibles. *Exemples :* `Houston -> Denver` (0%), `Toronto -> Las Vegas` (0%), `Vancouver -> Toronto` (0%). *Interprétation :* Cela corrobore les résultats NLP sur les problèmes de "Service" et "Confort" souvent décriés sur les transporteurs nord-américains (acteurs traditionnels).
2. **Le Transatlantique :**
   Les vols au départ de Londres vers la côte Ouest des USA sont sanctionnés sévèrement. *Exemples :* `Los Angeles -> London` (0%), `London -> San Francisco` (0%). *Hypothèse :* La durée extrême de ces vols ($> 10h$) exacerbe la sensibilité au confort (Legroom) et au service, transformant une expérience moyenne en calvaire.

7.3.2 Analyse des Flux Satisfait

À l'opposé, l'Asie et certaines lignes "Loisirs" tirent la performance vers le haut :

1. **L'Hégémonie Asiatique :**
   Les liaisons intra-asiatiques frôlent la perfection. *Exemples :* `SIN -> KUL` (Singapour-Kuala Lumpur, 100%), `Singapore -> Hong Kong` (100%). *Interprétation :* La culture de service des hubs asiatiques (Changi, HKG) et des compagnies locales crée un standard d'excellence que l'Occident peine à égaler.
2. **L'Effet "Vacances" :**
   Des lignes comme `Brisbane -> Honolulu` (75%) ou `Sydney -> Bangkok` (75%) offrent une surperformance. Le contexte du voyage (Loisirs vs Business) influence positivement la perception client (biais d'humeur positive).

8 Analyse Critique : Limites et Biais Méthodologiques

Toute démarche de *Data Science*, aussi rigoureuse soit-elle, comporte des biais intrinsèques liés à la nature des données et aux simplifications algorithmiques. Il est impératif d'expliciter les frontières de l'étude avant d'en tirer des conclusions définitives.

8.1 Biais liés à la Source de Données

La matière première de l'analyse (les avis en ligne) n'est pas une donnée neutre. Elle souffre de trois distorsions majeures :

1. **Biais d'Auto-Sélection (Self-Selection Bias) :**
   Les clients qui rédigent un avis sont souvent situés aux extrêmes du spectre émotionnel (très mécontents ou très satisfaits). La "majorité silencieuse" (les clients moyennement satisfaits) est sous-représentée. Cela entraîne une distribution des notes bimodale (en forme de "U") qui peut accentuer artificiellement la polarisation des résultats.
2. **Biais de Classe et de Culture :**
   Les standards d'exigence varient : une note de 3/5 peut être perçue comme "bonne" par un Européen mais "médiocre" par un Américain. De plus, les passagers "Economy" étant plus nombreux, leurs préférences (Prix) écrasent statistiquement celles des passagers "Business" (Service) dans les modèles globaux.
3. **Absence de Normalisation par le Trafic :**
   Nos cartes de flux identifient des routes critiques (ex : *Toronto* $\rightarrow$ *Las Vegas*). Cependant, nous travaillons sur des volumes d'avis et non sur des ratios d'incidents par millier de passagers. Une ligne très fréquentée générera mécaniquement plus de plaintes en volume absolu, sans être nécessairement moins performante en proportion.

9 Limites du projet et Biais

Même si notre modèle prédit correctement la recommandation dans la majorité des cas, il est important de souligner les limites de notre approche. Un modèle de Machine Learning n'est qu'une simplification de la réalité, et nos résultats dépendent de la qualité des données que nous lui avons fournies.

9.1 Les problèmes liés aux données

Notre jeu de données (*dataset*) n'est pas parfait, ce qui influence forcément les résultats :

1. **On n'entend que les extrêmes (Biais de sélection) :**
   Les gens qui écrivent des avis sur Internet sont souvent soit très en colère, soit ravis. La "majorité silencieuse" (ceux qui ont trouvé le vol "juste correct") n'écrit pas. Notre modèle apprend donc sur des opinions très tranchées et a du mal à comprendre la nuance.

2. **Le problème du volume de vols :**
   Dans notre carte géographique, nous avons montré les lignes rouges (beaucoup de plaintes). Mais nous ne savons pas combien d'avions volent sur ces lignes. *Exemple :* Avoir 10 plaintes sur la ligne Paris-New York (très fréquentée) est moins grave qu'avoir 10 plaintes sur une petite ligne de campagne. Il nous manque le "nombre total de passagers" pour relativiser.
3. **La domination de la classe Éco :**
   Il y a beaucoup plus de passagers en classe Économique qu'en Business dans nos données. Le modèle a donc "appris" que le prix était le critère N°1, car c'est la priorité des passagers Éco. S'il y avait eu autant d'avis de milliardaires, le modèle aurait peut-être dit que le Champagne était le critère N°1.

9.2 Les limites de nos algorithmes

Nous avons choisi des algorithmes simples et efficaces, mais ils ont des défauts connus :

9.2.1 Limites du K-Means (Segmentation)

Le K-Means est un algorithme un peu "rigide". Il oblige chaque client à rentrer dans une seule case (un seul Cluster). Dans la vraie vie, un humain est complexe, on peut être à la fois "sensible au prix" et "exigeant sur le confort". Notre modèle ne permet pas cet entre-deux, ce qui simplifie parfois trop les profils.

9.2.2 Limites de l'analyse de texte (NLP)

Nous utilisons une méthode qui compte les mots (TF-IDF) mais qui ne comprend pas le sens des phrases :

— **L'ordre des mots :** Pour notre modèle, la phrase "Pas bon" est presque pareille que "Bon repas" car il voit les mots "Bon" et "Pas" mais ne comprend pas la grammaire.
— **L'ironie :** Si un client écrit "Super, 3h de retard !", le modèle voit le mot "Super" et risque de classer l'avis comme positif. L'ordinateur ne comprend pas encore le sarcasme.

9.3 Ce qu'on pourrait améliorer

Si nous avions plus de temps pour ce projet, nous pourrions :

— Utiliser des modèles de langage plus récents (comme BERT) qui comprennent le contexte et l'ironie.
— Récupérer des données sur le trafic aérien réel pour créer des statistiques plus justes (taux de plainte par passager).

10 Conclusion Générale

Ce projet de Machine Learning avait pour ambition de répondre à un défi complexe, transformer la masse hétérogène et bruyante des commentaires clients en un système d'aide à la décision structuré. Au terme de ce travail, nous avons validé qu'un pipeline de Machine Learning permet de dépasser la simple lecture des avis pour en extraire des mécanismes prédictifs fiables.

10.1 Apports de l'approche méthodologique

L'apport majeur de cette étude réside dans sa capacité à objectiver la "Voix du Client" sans a priori humain. Contrairement aux enquêtes de satisfaction classiques qui imposent des grilles de lecture préétablies, notre approche non-supervisée (le couplage K-Means et NMF) a permis de laisser les données "parler d'elles-mêmes". Sur le plan technique, le choix d'une factorisation matricielle NMF pondérée par TF-IDF s'est révélé particulièrement pertinent pour traiter des textes courts, surpassant les méthodes probabilistes classiques (LDA) en produisant des thématiques sémantiquement pures. De plus, la capacité de notre modèle supervisé à prédire la recommandation, nous avons démontré que le NLP, malgré son apparente complexité (ironie, fautes, argot), contient une structure mathématique exploitable pour automatiser la détection des clients à risque.

10.2 Synthèse des Résultats

L'analyse croisée des données a permis de dresser un portrait-robot du passager aérien qui contredit souvent l'intuition marketing. Nos résultats mettent en lumière une asymétrie fondamentale dans la formation de la satisfaction, le client sanctionne impitoyablement les défaillances logistiques (retards, bagages) qui agissent comme des facteurs d'hygiène, mais ne fidélise sa relation qu'au travers de la qualité des interactions humaines (l'équipage), seul véritable levier d'amélioration de la satisfaction. Par ailleurs, nous avons identifié que ce passager est avant tout un acteur économique rationnel, la prédominance écrasante du facteur "Value for Money" sur le confort matériel prouve que la recommandation naît moins du luxe absolu que de la cohérence entre le prix payé et la prestation reçue. Enfin, la fracture observée entre l'excellence des routes asiatiques et la dégradation des liaisons nord-américaines souligne que la performance n'est pas systémique, mais dépendante de la qualité opérationnelle des hubs traversés.

10.3 Réponse à la Problématique

Pour conclure et répondre à la problématique initiale, nous pouvons affirmer que le machine learning permet effectivement de convertir des données brutes en leviers opérationnels, en agissant comme un décodeur de complexité. Ce travail prouve que la fidélisation dans l'aérien n'est pas le fruit du hasard ou d'un ressenti subjectif, mais le résultat d'une équation précise où la rigueur industrielle doit s'allier à l'empathie humaine. Cette approche offre ainsi aux compagnies le pouvoir de passer d'une gestion réactive des plaintes à une stratégie prédictive, capable d'identifier les signaux faibles de rupture avant même que le client ne décide de partir à la concurrence.