



# Rapport de projet **CRM**

Membres du groupe :

**Cheddadi Radja**

**Abichou Nour Elhouda**

2022/2023

# **Table des matières**

## **I. Introduction**

A. Contexte : Les enjeux des retards de vols

B. Objectifs du projet : Évaluer le coût des retards, améliorer la satisfaction des clients et proposer des solutions

## **II. Choix de la problématique**

**III. Périmètre des données** : analyse exploratoire et définition des données disponibles pour la modélisation

## **IV. Construction du modèle de prédiction des retards de vols**

A. Choix des algorithmes : Comparaison de différents algorithmes pour la prédiction des retards

B. Objectif du modèle : Représentativité et utilisation des résultats pour la prise de décision

C. Résultats des modèles : analyse des metrics de notation

## **V. Recommandations fonctionnelles**

A. Utilisation des prédictions de retards : Solutions envisagées pour répondre aux voyageurs, aux compagnies aériennes et aux aéroports

B. Omnicanalité et profil consommateur : Utilisation des prédictions de retards pour personnaliser l'expérience client et la fidélisation sur les sites de réservation en ligne

C. Opportunités commerciales : Identification de nouveaux services basés sur les prédictions de retards de vols

## **VI. Conclusion**

A. Récapitulation des objectifs et des résultats obtenus

B. Perspectives d'évolution et recommandations pour une utilisation optimale des prédictions de retards de vols

## **VII. Annexes**

## **I. Introduction**

A. Contexte : Les enjeux des retards de vols pour les compagnies aériennes, les voyageurs et les aéroports

Les retards de vols engendrent des coûts considérables pour les compagnies aériennes, entraînant des correspondances manquées, des retards supplémentaires sur les vols ultérieurs et des frais d'attente à l'aéroport. Ils suscitent également une grande frustration chez les voyageurs qui sont souvent très critiques envers les compagnies aériennes. Parallèlement, les aéroports doivent faire face à des défis logistiques causés par les retards et à des problèmes de congestion.

B. Objectifs du projet : Évaluer le coût des retards, améliorer la satisfaction des clients et proposer des solutions

Dans ce contexte, le projet vise à prédire les retards de vols, à améliorer la satisfaction des clients et à proposer des solutions pertinentes pour implémenter la solution. Pour ce faire, la compagnie IMSD 22-23, spécialisée dans l'analyse de données provenant du secteur aérien, nous a embauché comme data-scientist chargé de concevoir un nouvel algorithme permettant de prédire les retards de vols et de recommander des mesures pour améliorer la gestion de la relation client.

IMSD 22-23 compte parmi ses principaux clients de grandes compagnies aériennes telles que Bla Bla Jet, Creuse AIR LINE et LeBonAvion. Elle offre une gamme étendue de services liés à l'exploitation des données, notamment la visualisation, la prédiction des retards et des annulations de vols, ainsi que la gestion de la relation client.

Cependant, les performances actuelles du modèle utilisé pour les prédictions ne sont plus en phase avec la réalité. Afin de résoudre ce problème, IMSD 22-23 nous a recruté pour développer un nouvel algorithme de prédiction des retards et de recommander sa mise en œuvre pour améliorer la gestion de la relation client.

Pour mener à bien cette mission, IMSD 22-23 met à notre disposition une base de données comprenant des informations sur de nombreux vols. Cette base de données a été reconstruite à partir d'une sauvegarde de leur base principale et est fournie sous forme d'un fichier SQLite. Elle contient trois tables : aéroports, compagnies et vols. Notre objectif sera d'analyser cette base de données et de concevoir un nouvel algorithme permettant de prédire les retards de vols.

## **II. Choix de la problématique**

Pourquoi faisons-nous ce projet ?

Nous faisons cela pour améliorer la satisfaction des clients. Les retards de vols engendrent des coûts considérables pour les compagnies aériennes, tels que des correspondances manquées, des retards supplémentaires sur les vols ultérieurs et des frais d'attente à l'aéroport. Cela entraîne également une grande frustration chez les voyageurs. En évaluant les retards potentiels et en proposant des solutions pour les contrer, nous cherchons à améliorer la satisfaction globale des clients.

Pourquoi faire ce travail spécifiquement ?

Ce travail est pertinent car il permettra de prédire les retards de vols et d'améliorer la stratégie de gestion de la relation client (CRM) d'un nouveau client potentiel d'IMSD 22-23 qui est un site de réservation de vols en ligne (Booking.com). En utilisant les prédictions de retards de vols, le site de réservation pourra offrir des informations plus fiables et pertinentes aux clients lorsqu'ils choisissent leurs vols. Cela peut aider à réduire les désagréments causés par les retards de vols et à améliorer l'expérience des voyageurs.

Qu'est ce que les retards de vols peuvent engendrer ?

Les retards de vols peuvent également entraîner une baisse de la satisfaction des clients, une perte de confiance dans la compagnie aérienne et une diminution de la fidélité des clients. De plus, les retards de vols peuvent causer des problèmes de logistique et de congestion dans les aéroports.

Quelle cible choisir ?

La cible de notre projet est donc un site de réservation de vols en ligne, Booking.com. Nous choisissons cette cible car ces sites jouent un rôle crucial dans la réservation de vols pour de nombreux voyageurs. En améliorant la précision des prédictions de retards de vols et en fournissant des informations fiables aux clients, nous pouvons contribuer à améliorer leur expérience de réservation de vols et à réduire les inconvénients causés par les retards.

Comment délimiter le périmètre des données ?

Nous utiliserons une base de données fournie par IMSD 22-23, qui contient des informations sur de nombreux aspects de notre problématique. La base de données comprend des données sur les aéroports, les compagnies aériennes et les vols. Ces données seront utilisées pour entraîner notre algorithme de prédiction des retards de vols et pour fournir des informations pertinentes aux clients du site de réservation de vols en ligne. Nous aborderons le détail des données plus en profondeur dans les sections suivantes.

Prévoir des retards à l'arrivée ou au départ ?

L'un des aspects cruciaux de la prédiction des retards de vols est de choisir quelle variable représentera le retard dans notre modèle. Dans notre cas, nous disposons de deux variables pertinentes : le retard au départ et le retard à l'arrivée d'un avion. Cependant, en se

concentrant uniquement sur l'un de ces retards, nous pourrions négliger des informations importantes liées à l'autre retard. Afin de résoudre ce défi, nous proposons de calculer la moyenne des retards au départ et à l'arrivée pour créer une variable cible unique. Cette approche permet de combiner les deux informations et de capturer de manière plus globale les retards subis par les vols. En utilisant la moyenne des retards, nous prenons en compte à la fois les retards qui surviennent avant le décollage et ceux qui se produisent pendant le vol, jusqu'à l'atterrissage. Cette approche présente plusieurs avantages. Tout d'abord, elle nous permet d'obtenir une mesure plus complète du retard global subi par le vol, en prenant en considération les éventuels retards accumulés à différentes étapes du trajet. De plus, en utilisant une seule variable cible pour nos modèles de prédiction, nous simplifions le processus d'apprentissage et de généralisation, facilitant ainsi l'analyse et l'interprétation des résultats.

### III. Périmètre des données :

#### A. analyse exploratoire

La base de données mise à disposition contient les informations de nombreux vols, répartis en trois tables : aéroports, compagnies aériennes et vols. La table vols est la plus importante et contient toutes les informations relatives aux vols, telles que l'identifiant du vol, les aéroports de départ et d'arrivée, les horaires de départ et d'arrivée, la durée du vol, la distance parcourue, les retards éventuels et leur cause, etc. La table compagnies contient toutes les informations relatives aux compagnies aériennes, notamment leur nom et leur code. Enfin, la table aéroports contient toutes les informations relatives aux aéroports, tels que leur code IATA, leur nom, leur emplacement géographique, leur pays, leur altitude et leur tarification en cas de retard.

Pour la table aéroport, la variable "hauteur" (en pieds), la moyenne est de 711,28 pieds, avec une forte dispersion (écart-type élevé de 1232,28 pieds) (**annexe 02**). On remarque également que la valeur minimale est de 6 pieds et que la valeur maximale est de 7630 pieds, ce qui suggère qu'il y a une grande variabilité des hauteurs des aéroports dans cette base de données. Pour la variable "prix retard première 10 minutes", la moyenne est de 5167,19 dollars, avec un écart-type de 2820,14 dollars. On remarque également que la valeur minimale est de 105 dollars et que la valeur maximale est de 9877 dollars. Les quartiles indiquent que la moitié des observations sont comprises entre 2738 et 7489 dollars. Pour la variable "prix retard pour chaque minute après 10 minutes", la moyenne est de 502,95 dollars, avec un écart-type de 281,83 dollars. On remarque que la dispersion est similaire à celle de la variable précédente. Les quartiles indiquent que la moitié des observations sont comprises entre 267 et 738 dollars. Ces informations suggèrent qu'il y a une grande variabilité dans les hauteurs et les coûts de retard des aéroports de cette base de données.

L'analyse statistique des données de la table "vols" révèle des informations importantes pour comprendre les retards de vols (**annexe 03**). Sur les 3 millions d'entrées de la table, certaines variables clés sont disponibles pour évaluer les retards et les performances des vols.

En ce qui concerne les retards de départ, la moyenne observée est d'environ 9,46 minutes, ce qui indique que les vols ont généralement un léger retard par rapport à l'heure de départ programmée. Cependant, il est important de noter que la plage des retards de départ varie considérablement, allant de -82 minutes à 1 988 minutes, ce qui souligne l'existence de retards extrêmes dans certains cas.

Le temps de vol moyen enregistré est d'environ 116,63 minutes, avec une moyenne de 141,67 minutes pour le temps de vol programmé. Cela suggère que les vols ont tendance à prendre un peu plus de temps que prévu, ce qui peut potentiellement contribuer à des retards supplémentaires.

Concernant les retards à l'arrivée, la moyenne observée est d'environ 4,91 minutes, indiquant que les vols ont généralement de légers retards à leur arrivée prévue. Toutefois, il est important de noter que la plage des retards à l'arrivée est assez large, allant de -87 minutes à 1 971 minutes. Cela met en évidence la possibilité de retards significatifs dans certains cas, ce qui peut entraîner des perturbations importantes pour les voyageurs.

Lors de l'analyse des données de la table "vols", il est important de noter que les retards de départ et d'arrivée sont exprimés de manière conventionnelle, où un retard positif indique un retard réel et un retard négatif signifie que l'avion a de l'avance et est arrivé avant l'heure prévue. Ainsi, lorsque nous avons mentionné la moyenne des retards de départ ou d'arrivée dans l'analyse précédente, cela implique que les retards positifs sont inclus. Par exemple, une moyenne de 4,91 minutes pour les retards à l'arrivée signifie que les vols ont en moyenne un retard de 4,91 minutes.

Il convient également de noter la présence de variables telles que "DETOURNEMENT", "ANNULATION", "RETARD SYSTEM", "RETARD SECURITE", "RETARD COMPAGNIE", "RETARD AVION" et "RETARD METEO". Ces variables indiquent la présence de facteurs spécifiques qui peuvent contribuer aux retards de vols, tels que les déviations de vol, les annulations, les retards liés au système ou à la sécurité, les retards dus à la compagnie aérienne elle-même, les retards liés à l'avion ou les retards dus aux conditions météorologiques.

Enfin, le résultat de la table compagnies montre deux colonnes: "COMPAGNIE" et "CODE", avec des statistiques pour chaque colonne. Pour la colonne "COMPAGNIE", il y a 13 valeurs différentes (unique) sur un total de 13 (count).

## **B. Retraitement des données disponibles pour la modélisation**

### **1. Jointure LEFT JOIN :**

Les données du DataFrame 'vols' ont été fusionnées avec les données du DataFrame 'aéroports' en utilisant une jointure LEFT JOIN. La jointure a été effectuée en utilisant les colonnes 'AEROPORT DEPART' et 'CODE IATA' du DataFrame 'vols', et les colonnes correspondantes du DataFrame 'aéroports'. Et ceux afin de récupérer les données liées aux aéroports de départs présentes dans le data frame 'aéroports' et les lier au data frame 'vols'.

### **2. Renommage des colonnes :**

Certaines colonnes ont été renommées pour améliorer la clarté et la cohérence des noms de colonnes.

### **3. Nouvelle jointure LEFT JOIN :**

Les données du DataFrame résultant ont été fusionnées avec les données du DataFrame 'aéroports' en utilisant une jointure LEFT JOIN. La jointure a été effectuée en utilisant les colonnes 'AEROPORT ARRIVEE A' et 'CODE IATA' du DataFrame résultant, et les colonnes correspondantes du DataFrame 'aéroports'. Et ceux afin de récupérer les données liées aux aéroports d'arrivée présentes dans le data frame 'aéroports' et les lier au data frame 'vols'.

#### **4. Calcul de la moyenne des retards :**

Une nouvelle colonne 'RETARD MOYEN' a été créée en calculant la moyenne des colonnes 'RETARD DE DEPART' et 'RETARD A L'ARRIVEE'. En tant que data scientist travaillant sur la prédiction des retards des vols, l'une des décisions cruciales est de choisir quelle variable représentera le retard dans notre modèle. Dans notre cas, nous avons à disposition deux variables pertinentes : le retard au départ et le retard à l'arrivée d'un avion.

La prédiction des retards de vol est un problème complexe, car de nombreux facteurs peuvent influencer à la fois le retard au départ et le retard à l'arrivée. En prenant en compte uniquement l'un de ces retards, nous pourrions négliger des informations importantes liées à l'autre retard.

Afin de combiner ces deux informations et de capturer l'ensemble des retards subis par les vols, nous proposons de calculer la moyenne des retards au départ et à l'arrivée. Cette approche nous permet d'obtenir une variable cible unique, représentant de manière plus globale les retards des vols.

En utilisant la moyenne des retards, nous prenons en compte à la fois les retards qui surviennent avant le décollage et ceux qui se produisent pendant le vol, jusqu'à l'atterrissage. Cela nous permet d'obtenir une mesure plus complète du retard global subi par le vol, en prenant en considération les éventuels retards accumulés à différentes étapes du trajet.

De plus, en utilisant une seule variable cible pour nos modèles de prédiction, nous simplifions le processus d'apprentissage et de généralisation. Plutôt que de traiter séparément les retards au départ et à l'arrivée, notre modèle peut se concentrer sur la prédiction d'une seule valeur, facilitant ainsi l'analyse et l'interprétation des résultats.

#### **5. Conversion de la colonne 'DATE' en type de données de date/heure :**

La colonne 'DATE' a été convertie en type de données de date/heure à l'aide de la fonction 'pd.to\_datetime'.

#### **6. Encodage des variables catégorielles :**

Les colonnes catégorielles telles que 'RAISON D'ANNULATION', 'COMPAGNIE AERIENNE', 'CODE IATA DEPART' et 'CODE IATA ARRIVEE' ont été extraites dans un nouveau DataFrame appelé 'cat\_df'. Les valeurs des colonnes catégorielles ont été transformées en utilisant l'encodage LabelEncoder de la bibliothèque scikit-learn.

#### **7. Suppression des colonnes :**

Lors de la préparation des données et comme cité plus haut, pour notre modèle de prédiction de retards de vols, nous avons décidé de ne conserver que les colonnes 'CODE IATA DEPART' et 'CODE IATA ARRIVEE' parmi les variables catégorielles. Voici les raisons qui justifient ce choix :

**Pertinence de l'information :** Les codes IATA des aéroports de départ et d'arrivée contiennent l'essentiel de l'information nécessaire pour identifier les lieux de départ et d'arrivée des vols. Ces codes sont des identifiants normalisés utilisés dans l'industrie de l'aviation, ce qui les rend facilement interprétables et comparables. En conservant uniquement ces codes, nous pouvons conserver l'information pertinente tout en réduisant la redondance des autres colonnes.

**Redondance des autres colonnes :** Les autres colonnes telles que 'AEROPORT DEPART', 'AEROPORT ARRIVEE', 'LIEU DE DEPART', 'PAYS DE DEPART', 'LIEU DE ARRIVEE' et 'PAYS DARRIVEE' fournissent essentiellement la même information que les codes IATA des aéroports de départ et d'arrivée. Ces colonnes sont donc redondantes et n'apportent pas de valeur supplémentaire à la prédiction des retards de vols.

**Réduction de la dimensionnalité :** En éliminant ces colonnes redondantes, nous réduisons la dimensionnalité de notre jeu de données. Cela peut être avantageux car une dimensionnalité réduite peut faciliter l'entraînement des modèles, réduire les besoins en mémoire et améliorer les performances.

En résumé, en conservant uniquement les colonnes 'CODE IATA DEPART' et 'CODE IATA ARRIVEE', nous éliminons la redondance des informations tout en maintenant la pertinence nécessaire pour prédire les retards de vols. Cette approche permet une représentation plus concise et efficace des données pour notre modèle de prédiction.

## **8. Remplissage des valeurs manquantes :**

Il existe des valeurs manquantes des colonnes 'RETARD SYSTEM', 'RETARD SECURITE', 'RETARD COMPAGNIE', 'RETARD AVION' et 'RETARD METEO', ces dernières ont été remplies avec la valeur 0.

Ce que nous avons fait en remplaçant les NaN par des zéros dans les colonnes spécifiées est une opération de prétraitement des données cruciale. Cette opération est justifiée par le fait que dans notre contexte, la présence de valeurs manquantes (NaN) dans ces colonnes signifie que les avions n'ont pas fait de retard.

En remplaçant les NaN par des zéros, nous avons effectué une imputation des données manquantes en utilisant une valeur appropriée dans le contexte de notre problème. Nous avons pris la décision de considérer les NaN comme des retards nuls, ce qui est logique puisque l'absence de retard correspond à une valeur de zéro dans notre cas.

Cette imputation nous permet de conserver toutes les lignes de données et de fournir une valeur numérique valide pour chaque observation, ce qui est essentiel pour l'entraînement de nos modèles d'apprentissage automatique. Cela nous permet d'apprendre à associer les autres caractéristiques des vols aux retards, en prenant en compte le fait qu'un retard nul est représenté par un zéro.

Il est important de noter que cette imputation est une décision que nous avons prise en fonction de notre compréhension du domaine et du contexte de nos données. Elle est cohérente avec l'interprétation que nous avons donnée aux valeurs manquantes dans les colonnes concernées.



## 9. Suppression des ligne des valeurs manquantes :

Les lignes contenant des valeurs manquantes dans les colonnes 'HEURE DE DEPART', 'DECOLLAGE', 'TEMPS PASSE', 'TEMPS DE VOL', 'ATTERRISSAGE' et 'HEURE D'ARRIVEE' ont été supprimées du DataFrame résultant (**annexe 05**). Cette décision a été prise pour plusieurs raisons.

Premièrement, les colonnes 'HEURE DE DEPART', 'DECOLLAGE', 'TEMPS PASSE', 'TEMPS DE VOL', 'ATTERRISSAGE' et 'HEURE D'ARRIVEE' sont toutes des variables temporelles importantes pour notre modèle de prédiction des retards de vols. Les valeurs manquantes dans ces colonnes indiquent qu'il manque des informations cruciales sur les horaires de départ, d'arrivée et de vol. Conserver les lignes avec des valeurs manquantes dans ces colonnes aurait entraîné une perte de précision et de fiabilité dans notre modèle prédictif. Deuxièmement, les pourcentages de valeurs manquantes dans ces colonnes sont relativement faibles, allant de 1,64% à 1,99%. Éliminer les lignes avec des valeurs manquantes dans ces colonnes ne compromet pas significativement la taille et la représentativité de notre ensemble de données.

Enfin, la suppression des lignes avec des valeurs manquantes dans ces colonnes permet de garantir la cohérence et la validité des données utilisées pour l'analyse. En éliminant les valeurs manquantes, nous nous assurons que seules les observations complètes et fiables sont prises en compte dans notre modèle, ce qui renforce la qualité de nos résultats et de nos prédictions.

En résumé, la suppression des lignes avec des valeurs manquantes dans les colonnes 'HEURE DE DEPART', 'DECOLLAGE', 'TEMPS PASSE', 'TEMPS DE VOL', 'ATTERRISSAGE' et 'HEURE D'ARRIVEE' a été effectuée pour préserver la qualité et la cohérence des données utilisées dans notre modèle de prédiction des retards de vols.

## 10. Création de la variable cible "Retard" :

Notre objectif est de prédire si un vol sera en retard ou non en fonction de la valeur de la colonne "RETARD MOYEN". Pour cela, nous avons créé une nouvelle colonne appelée "Retard" qui servira de variable cible dans nos modèles de prédiction.

La variable cible "Retard" est binaire, avec deux classes : 1 pour les vols sans retard et 0 pour les vols avec retard. La logique derrière cette définition est que nous voulons prédire si un vol sera en retard ou non, donc nous avons besoin d'une variable qui représente cette information de manière binaire.

Pour attribuer les valeurs à la colonne "Retard", nous avons utilisé la condition suivante :

Si la valeur de la colonne "RETARD MOYEN" est strictement supérieure à 0, cela signifie qu'il y a un retard. Lorsque la valeur de la colonne "RETARD MOYEN" est inférieure ou égale à 0, nous avons attribué la valeur 1 à la colonne "Retard", ce qui indique l'absence de retard.

En résumé, la création de la variable cible "Retard" avec les classes 0 pour pas de retard et 1 pour retard est une étape essentielle pour entraîner notre modèle de prédiction qui pourra classer les vols en fonction de leur retard potentiel. Cette variable cible permettra au modèle d'apprendre à distinguer les caractéristiques des vols avec et sans retard afin de prédire avec précision si un vol donné sera en retard ou non.

## **IV. Construction du modèle de prédiction des retards de vols**

### **A. Choix des algorithmes : Comparaison de différents algorithmes pour la prédiction des retards**

Dans le processus de construction du modèle de prédiction des retards de vols, il est essentiel de comparer différents algorithmes pour déterminer celui qui offre les meilleures performances. Nous avons utilisé trois algorithmes différents : la régression logistique, le SVM (Support Vector Machine) et le réseau de neurones.

La régression logistique est un algorithme de classification largement utilisé qui modélise la relation entre les variables indépendantes et une variable dépendante binaire. Cet algorithme cherche à trouver les coefficients optimaux pour une fonction logistique qui est utilisée pour prédire la probabilité d'appartenance à une classe donnée. Dans notre projet, nous avons utilisé la régression logistique pour prédire les retards de vols.

Le SVM est un algorithme d'apprentissage supervisé qui peut être utilisé pour la classification et la régression. Il cherche à trouver un hyperplan qui sépare les exemples de différentes classes avec la plus grande marge possible. Dans notre cas, nous avons utilisé le SVM avec un noyau radial (RBF) et un paramètre de régularisation C de 1.0 pour prédire les retards de vols.

Enfin, le réseau de neurones est un modèle d'apprentissage profond qui est capable d'apprendre des représentations complexes à partir des données. Il est composé de multiples couches de neurones interconnectés qui effectuent des calculs non linéaires. Dans notre projet, nous avons utilisé un réseau de neurones avec deux couches cachées de 64 neurones chacune et une couche de sortie avec une fonction d'activation sigmoïde pour prédire les retards de vols.

### **B. Objectif du modèle : Représentativité et utilisation des résultats pour la prise de décision**

L'objectif du modèle de prédiction des retards de vols est de fournir des prédictions précises et représentatives pour accompagner la stratégie CRM des entreprises. En utilisant des techniques d'apprentissage automatique, le modèle vise à capturer les modèles et les relations dans les données historiques des vols afin de prédire les retards futurs.

La représentativité du modèle est essentielle pour garantir que les résultats sont généralisables et applicables dans différents contextes. Cela signifie que le modèle doit être entraîné et évalué sur des ensembles de données représentatifs, reflétant la diversité des scénarios réels auxquels il sera confronté. En veillant à la représentativité des données et en utilisant des techniques appropriées d'apprentissage automatique, le modèle peut fournir des prédictions fiables et utiles pour la prise de décision.

### **C. Résultats des modèles :**

Les résultats du modèle de régression logistique sur les données de test sont les suivants :

- Exactitude (Accuracy) : 0,8067
- Précision (Precision) : 0,7627
- Rappel (Recall) : 0,9970
- Score F1 : 0,8643

Ces mesures indiquent la performance du modèle sur les données de test. L'exactitude de 0,8067 signifie qu'environ 80,67 % des prédictions faites par le modèle sont correctes. Une précision de 0,7627 indique que parmi toutes les prédictions positives faites par le modèle, 76,27 % sont des vrais positifs. Un rappel de 0,9970 suggère que le modèle est capable d'identifier environ 99,70 % des instances positives réelles. Le score F1 de 0,8643 est la moyenne harmonique de la précision et du rappel, fournissant une mesure globale de la performance du modèle.

Ces résultats démontrent que le modèle de régression logistique présente une performance relativement bonne avec une haute précision, un rappel élevé et un score F1 équilibré. Cela signifie qu'il est capable de prédire avec précision la classe des instances positives et qu'il identifie la plupart des instances positives réelles. Cependant, il est important de noter que les performances peuvent varier en fonction des données spécifiques et des paramètres choisis pour le modèle.

Les résultats du modèle SVM sur les données de test sont les suivants :

- Exactitude (Accuracy) : 0,7218
- Précision (Precision) : 0,7258
- Rappel (Recall) : 0,8726
- Score F1 : 0,7925

Ces mesures indiquent la performance du modèle sur les données de test. L'exactitude de 0,7218 signifie qu'environ 72,18 % des prédictions faites par le modèle sont correctes. Une précision de 0,7258 indique que parmi toutes les prédictions positives faites par le modèle, 72,58 % sont des vrais positifs. Un rappel de 0,8726 suggère que le modèle est capable d'identifier environ 87,26 % des instances positives réelles. Le score F1 de 0,7925 est la moyenne harmonique de la précision et du rappel, fournissant une mesure globale de la performance du modèle.

En conclusion, les résultats du modèle SVM sur les données de test démontrent une performance encourageante et démontrent que le modèle a une capacité prometteuse à prédire les instances correctes et à identifier les cas positifs.

Les résultats du modèle de réseau de neurones sur les données de test sont les suivants :

- Exactitude (Accuracy) : 0,7218
- Précision (Precision) : 0,7258
- Rappel (Recall) : 0,8726
- Score F1 : 0,7925

Ces mesures indiquent la performance du modèle sur les données de test. L'exactitude de 0,7218 signifie qu'environ 72,18 % des prédictions faites par le modèle sont correctes. Une précision de 0,7258 indique que parmi toutes les prédictions positives faites par le modèle, 72,58 % sont des vrais positifs. Un rappel de 0,8726 suggère que le modèle est capable d'identifier environ 87,26 % des instances positives réelles. Le score F1 de 0,7925 est la moyenne harmonique de la précision et du rappel, fournissant une mesure globale de la performance du modèle.

En conclusion, les résultats du modèle de réseau de neurones sur les données de test démontrent une performance solide. Ces résultats mettent en évidence la capacité du modèle de réseau de neurones à bien généraliser sur les données de test.

Ces résultats permettent de comparer les performances des différents algorithmes et d'évaluer leur capacité à prédire les retards de vols. Dans ce cas, la régression logistique semble avoir la meilleure performance avec une plus haute précision et un score F1 plus élevé.

## **V. Recommandations fonctionnelles**

### **A. Utilisation des prédictions de retards : Solutions envisagées pour répondre aux voyageurs, aux compagnies aériennes et aux aéroports**

Les prédictions de retards de vols fournies par le modèle développé présentent plusieurs avantages et peuvent être utilisées pour proposer des solutions adaptées aux voyageurs, aux compagnies aériennes et aux aéroports.

Pour les voyageurs, ces prédictions de retards permettent d'anticiper les éventuels problèmes lors de leurs déplacements. En fournissant des informations fiables sur les retards probables, les voyageurs peuvent prendre des décisions éclairées telles que la modification de leur itinéraire, la recherche de vols alternatifs ou la préparation en conséquence de leur emploi du temps. Cela leur permet de minimiser les désagréments causés par les retards et d'optimiser leur expérience de voyage.

Les compagnies aériennes peuvent également bénéficier des prédictions de retards. En utilisant ces informations, elles peuvent prendre des mesures préventives pour atténuer les retards potentiels. Cela peut inclure la révision des horaires de vol, l'optimisation de la gestion des ressources et du personnel, ou même la communication proactive avec les passagers pour les tenir informés des retards et des solutions alternatives. Ces actions permettent aux compagnies aériennes de mieux gérer les retards, d'améliorer la satisfaction des clients et de réduire les coûts associés.

Du point de vue des aéroports, les prédictions de retards offrent la possibilité de planifier et de gérer plus efficacement les opérations aéroportuaires. En anticipant les retards, les aéroports peuvent ajuster leurs ressources, tels que les pistes d'atterrissage, les portes d'embarquement et le personnel de sécurité, pour minimiser les congestions et les temps d'attente. Cela permet une meilleure fluidité des opérations et une gestion plus efficace des infrastructures aéroportuaires.

En résumé, l'utilisation des prédictions de retards de vols offre des solutions potentielles pour répondre aux besoins des voyageurs, des compagnies aériennes et des aéroports. **Cela étant dit, l'objectif de notre projet est d'aller au-delà de ces cibles classiques et d'explorer comment les prédictions de retards de vols peuvent être utilisées de manière spécifique par les sites de réservation de vols en ligne.** Ces plateformes jouent un rôle essentiel dans la planification et la réservation des voyages pour de nombreux voyageurs.

L'utilisation des prédictions de retards de vols par les sites de réservation de vols en ligne ouvre de nouvelles possibilités pour améliorer l'expérience des voyageurs. En fournissant des informations précises et en proposant des solutions alternatives, ces sites peuvent aider les

voyageurs à prendre des décisions éclairées, à minimiser les perturbations causées par les retards et à renforcer leur satisfaction globale.

## **B. Omnicanalité et profil consommateur : Utilisation des prédictions de retards pour personnaliser l'expérience client et la fidélisation des site de réservation en ligne**

L'utilisation des prédictions de retards de vols dans le contexte de l'omnicanalité et de la personnalisation de l'expérience client offre de nouvelles possibilités pour les sites de réservation de vols en ligne. Voici quelques aspects importants à considérer :

- **Communication proactive et intégration d'alertes :**

En intégrant les prédictions de retards de vols dans les systèmes de communication des sites de réservation, il devient possible d'envoyer des alertes proactives aux clients concernés par les retards. Les sites peuvent configurer des notifications push, des messages texte ou des e-mails pour informer les voyageurs des retards prévus et des options disponibles. Par exemple, si un retard de vol est prévu, le site peut suggérer au client de reprogrammer son vol ou de choisir une alternative de vol pour éviter les désagréments. Cette communication proactive démontre l'engagement envers les clients et permet de renforcer leur confiance et leur fidélité.

- **Intégration dans une application web ou mobile :**

En développant une application web ou mobile dédiée, les clients peuvent accéder facilement aux prévisions de retard de leur vol. L'application peut être conçue pour permettre aux utilisateurs de saisir les détails de leur vol et afficher les prévisions de retard correspondantes. Cela offre une commodité aux voyageurs qui souhaitent connaître les prévisions de retard avant leur voyage, leur permettant ainsi de mieux planifier leurs déplacements et de prendre des décisions éclairées. Afin de développer une application web ou mobile dédiée intégrant les prévisions de retards de vols il faudrait suivre les étapes suivantes :

1. Développement de l'interface utilisateur : Engager une équipe de développement pour concevoir une interface utilisateur attrayante, conviviale et intuitive. L'application devrait permettre aux utilisateurs de saisir facilement les détails de leur vol, tels que la compagnie aérienne, le numéro de vol et la date de départ.
2. Intégration des données de prévisions de retard : Établir un partenariat avec des fournisseurs de données de prévisions de retards de vols fiables pour obtenir les flux de données nécessaires (IMSD 22-23). Intégrer ces données dans l'application afin de pouvoir afficher les prévisions de retard correspondantes pour chaque vol.
3. Fonctionnalités de notification : Mettre en place des fonctionnalités de notification pour informer les utilisateurs des prévisions de retard de leur vol. Envoyer des notifications en temps réel pour tenir les utilisateurs informés des changements éventuels et leur permettre de prendre des mesures appropriées.
4. Personnalisation de l'expérience utilisateur : Offrir aux utilisateurs la possibilité de personnaliser leur expérience en leur permettant de définir leurs préférences de notification. Par exemple, ils peuvent choisir de recevoir des notifications pour tous les retards ou uniquement pour les retards supérieurs à une certaine durée.
5. Fonctionnalités supplémentaires : Enrichir l'application avec des fonctionnalités complémentaires telles que des informations sur les aéroports, des conseils de voyage,

des options de reprogrammation de vol et des alternatives de voyage en cas de retard. Cela permettra aux utilisateurs d'avoir une vision complète et de prendre des décisions éclairées.

6. Intégration avec d'autres services de voyage : Explorer la possibilité d'intégrer l'application avec d'autres services de voyage, tels que les services de location de voitures, les réservations d'hôtels ou les services de transfert d'aéroport. Cela permettra aux utilisateurs de gérer facilement tous les aspects de leur voyage à partir d'une seule application.
7. Collecte de données et analyse : Mettre en place des mécanismes de collecte de données pour suivre l'utilisation de l'application, les préférences des utilisateurs et les performances de l'application. Analysez ces données pour obtenir des informations précieuses sur les comportements des utilisateurs et améliorer en continu l'expérience utilisateur.
8. Campagne de promotion : Mettre en place une campagne de marketing ciblée pour promouvoir l'application auprès des voyageurs. Utiliser des canaux de communication appropriés tels que les médias sociaux, les publicités en ligne et les partenariats avec des compagnies aériennes ou des agences de voyage pour atteindre votre public cible.

9.

- **Intégration dans les systèmes de réservation des sites :**

L'intégration des prédictions de retards de vols dans les systèmes de réservation des sites de réservation de vols en ligne offre de nombreux avantages en termes de personnalisation de l'itinéraire de voyage pour les clients concernés. Cette intégration permet d'anticiper les retards prévus et de proposer des solutions alternatives pour minimiser les désagréments et offrir une expérience de voyage plus fluide et personnalisée. Voici quelques points clés sur l'impact de cette intégration :

1. Proposition d'options de correspondance adaptées : Lorsqu'un vol risque d'être retardé, le système de réservation peut utiliser les prédictions de retards pour suggérer des options de correspondance plus adaptées. Par exemple, si le système détecte un retard probable sur un vol de correspondance, il peut proposer automatiquement une alternative avec une correspondance plus confortable ou avec un délai plus long pour éviter toute précipitation. Cela permet aux voyageurs de planifier leur itinéraire de manière plus efficace et d'éviter des situations stressantes.
2. Suggestions de changements d'aéroport : Si un retard important est prévu sur un vol, le système de réservation peut également proposer des changements d'aéroport pour minimiser les désagréments. Par exemple, si un aéroport de correspondance risque d'être congestionné en raison d'un retard, le système peut suggérer un changement vers un autre aéroport moins affecté. Cela permet d'éviter les retards supplémentaires et de réduire les contraintes liées aux retards.
3. Gestion proactive des retards : L'intégration des prédictions de retards permet aux systèmes de réservation de prendre une approche proactive pour gérer les retards. Plutôt que de simplement informer les voyageurs d'un retard, le système peut proposer des solutions alternatives en temps réel, telles que la réservation d'un vol ultérieur ou l'offre d'un hébergement temporaire en cas de retard important. Cela améliore considérablement l'expérience du voyageur et renforce la confiance envers la compagnie aérienne ou le site de réservation.
4. Informations en temps réel : L'intégration des prédictions de retards permet également aux systèmes de réservation de fournir des informations en temps réel aux voyageurs. Les clients peuvent être informés des retards prévus, des changements d'itinéraire ou des options de correspondance alternatives directement via le site de réservation ou

une application mobile. Cela permet aux voyageurs de rester informés et de prendre des décisions éclairées concernant leur voyage.

En intégrant les prédictions de retards de vols dans les systèmes de réservation des sites de réservation de vols en ligne, il devient possible d'améliorer considérablement l'expérience de voyage des clients concernés. Les voyageurs peuvent bénéficier de solutions alternatives adaptées à leurs besoins, minimisant ainsi les désagréments liés aux retards. Cette personnalisation de l'itinéraire de voyage renforce la satisfaction des clients et contribue à fidéliser leur confiance envers les compagnies aériennes et les sites de réservation.

- **Intégration dans le service client :**

En formant les équipes du service client à utiliser les prévisions de retard de vols, les agents peuvent mieux répondre aux questions et aux préoccupations des clients. Lorsqu'un client contacte le service client pour obtenir des informations sur son vol, les agents peuvent accéder aux prévisions de retard et fournir des informations précises et actualisées. Ils peuvent également offrir des conseils personnalisés sur les meilleures options disponibles en cas de retard, telles que l'assistance pour la reprogrammation du vol ou la réservation d'un hébergement à proximité de l'aéroport. Cela améliore la qualité du service client et renforce la satisfaction des clients.

L'intégration des prévisions de retard de vols dans le service client offre de nombreux avantages en termes de qualité du service et de satisfaction des clients. Voici comment cette intégration peut améliorer le service client :

1. Informations précises et actualisées : En formant les agents du service client à utiliser les prévisions de retard de vols, ils ont accès à des informations précises et actualisées concernant les retards prévus. Lorsqu'un client contacte le service client pour obtenir des informations sur son vol, les agents peuvent consulter ces prévisions et fournir des réponses précises et fiables. Cela permet d'éviter de donner des informations erronées ou dépassées, ce qui peut causer de la confusion et de la frustration chez les clients.
2. Conseils personnalisés : Grâce aux prévisions de retard, les agents du service client peuvent offrir des conseils personnalisés aux clients concernés. Par exemple, si un retard important est prévu, l'agent peut recommander des options alternatives telles que la reprogrammation du vol ou la réservation d'un hébergement à proximité de l'aéroport. Ces conseils personnalisés démontrent une attention particulière aux besoins du client et peuvent contribuer à minimiser les désagréments causés par les retards.
3. Réactivité accrue : En utilisant les prévisions de retard, les agents du service client peuvent anticiper les questions et les préoccupations des clients concernant les retards de vols. Ils peuvent ainsi se préparer à répondre de manière proactive à ces situations, plutôt que de réagir de manière passive. Cela permet d'améliorer la réactivité du service client et de démontrer un engagement envers la satisfaction du client.
4. Gestion des problèmes de manière efficace : L'intégration des prévisions de retard dans le service client permet aux agents d'apporter des solutions rapides et efficaces en cas de perturbations. Par exemple, si un client appelle pour signaler un retard, l'agent peut immédiatement consulter les prévisions et prendre des mesures appropriées, telles que la réservation d'un vol de correspondance ultérieur ou l'assistance pour les demandes de remboursement. Cela permet de résoudre les problèmes de manière proactive et d'offrir une assistance efficace aux clients.

5. Renforcement de la satisfaction client : Fournir un service client de qualité, basé sur des informations précises et des conseils personnalisés, renforce la satisfaction des clients. Lorsqu'ils se sentent écoutés, compris et bien pris en charge, les clients sont plus susceptibles de développer une confiance envers la compagnie aérienne ou le site de réservation. Cela peut favoriser la fidélité des clients et générer une image positive de l'entreprise.

En intégrant les prévisions de retard de vols dans le service client, les compagnies aériennes et les sites de réservation peuvent améliorer la qualité du service et la satisfaction des clients. Les agents formés peuvent offrir des informations précises, des conseils personnalisés et une réactivité accrue pour gérer efficacement les problèmes liés aux retards de vols. Cela contribue à renforcer la confiance des clients et à garantir une expérience positive tout au long de leur voyage.

- **Amélioration de la segmentation client :**

En utilisant les prévisions de retard de vols, il est possible de créer des segments clients plus précis et pertinents. Les sites de réservation peuvent identifier les clients fréquents qui sont souvent affectés par des retards et leur offrir des avantages ou des services spéciaux pour les fidéliser. Par exemple, ces clients pourraient bénéficier de priorité de réservation sur les vols moins sujets aux retards, de surclassements ou de services d'assistance supplémentaires. Cette approche contribue à renforcer la relation avec les clients les plus précieux et à accroître leur fidélité.

L'utilisation des prévisions de retard de vols pour améliorer la segmentation client offre plusieurs avantages en termes de fidélisation et de satisfaction des clients. Voici comment cette approche peut améliorer la segmentation client :

1. Identification des clients fréquemment touchés par des retards : En utilisant les prévisions de retard, les sites de réservation peuvent identifier les clients qui sont souvent affectés par des retards de vols. Cela permet de créer un segment spécifique de clients qui ont des besoins particuliers en termes de gestion des retards. Ces clients peuvent être identifiés en fonction de leur historique de réservation, de leur fréquence de voyage ou d'autres critères pertinents.
2. Offre d'avantages spéciaux : Une fois que les clients fréquemment touchés par des retards ont été identifiés, les sites de réservation peuvent leur offrir des avantages spéciaux pour les fidéliser. Par exemple, ces clients pourraient bénéficier de priorité de réservation sur des vols moins sujets aux retards, ce qui leur permettrait de minimiser les risques de perturbations. Ils pourraient également se voir proposer des surclassements ou des services d'assistance supplémentaires pour compenser les désagréments causés par les retards. Ces avantages spéciaux démontrent une attention particulière envers les clients et renforcent leur satisfaction et leur fidélité.
3. Personnalisation des offres et des communications : En utilisant les prévisions de retard, les sites de réservation peuvent personnaliser les offres et les communications pour les clients touchés par des retards. Par exemple, ils peuvent envoyer des informations spécifiques sur les retards prévus, des suggestions d'itinéraires alternatifs ou des offres promotionnelles pour des vols moins sujets aux perturbations. Cette personnalisation des offres et des communications montre aux clients que leur situation est prise en compte et qu'ils sont traités de manière spéciale.
4. Renforcement de la relation client : En offrant des avantages spéciaux et en personnalisant les communications, les sites de réservation renforcent la relation avec



les clients touchés par des retards. Les clients se sentent valorisés et pris en charge, ce qui contribue à renforcer leur fidélité à long terme. Ils sont plus susceptibles de choisir le site de réservation pour leurs voyages futurs et de recommander leurs services à d'autres personnes.

5. Optimisation des ressources : En segmentant les clients en fonction de leur susceptibilité aux retards, les sites de réservation peuvent optimiser l'utilisation des ressources. Par exemple, en offrant des avantages spéciaux aux clients fréquemment touchés par des retards, les compagnies aériennes peuvent mieux gérer la demande sur les vols moins sujets aux perturbations. Cela permet de maximiser l'efficacité opérationnelle tout en garantissant une expérience de voyage plus agréable pour les clients.

En utilisant les prévisions de retard de vols pour améliorer la segmentation client, les sites de réservation peuvent offrir des avantages spéciaux, personnaliser les offres et renforcer la relation avec les clients les plus précieux. Cette approche contribue à accroître la fidélité des clients, à optimiser les ressources et à garantir une expérience de voyage plus satisfaisante.

En conclusion, en intégrant les prédictions de retards de vols dans ces différentes facettes de leurs opérations, les sites de réservation de vols en ligne peuvent offrir une expérience client plus personnalisée, proactive et fiable. Cela permet non seulement d'améliorer la satisfaction des voyageurs, mais également de renforcer la fidélité à la plateforme. En utilisant ces prédictions, les sites de réservation peuvent mieux anticiper les besoins et les préférences des clients, proposer des alternatives en cas de retard et communiquer de manière proactive pour atténuer les désagréments liés aux retards. Cette approche favorise une relation de confiance avec les voyageurs, les incite à revenir et à recommander le site à d'autres utilisateurs, contribuant ainsi à la croissance et à la réussite de l'entreprise.

### **C. Opportunités commerciales : Identification de nouveaux services basés sur les prédictions de retards de vols**

Les prédictions de retards de vols offrent de nouvelles opportunités commerciales aux sites de réservation de vols en ligne. En exploitant ces prévisions, ils peuvent identifier de nouveaux services qui permettent d'améliorer l'expérience des voyageurs confrontés à des retards. Ces services spécifiques basés sur les prédictions de retards de vols offrent des avantages concrets aux clients et renforcent la confiance envers les sites de réservation.

1. Assurance retard de vol : Les sites de réservation de vols en ligne peuvent proposer une assurance spécifique couvrant les retards de vol. Cette assurance offrirait une compensation financière ou des avantages supplémentaires aux voyageurs en cas de retard de vol, ce qui contribuerait à renforcer la confiance des clients et à améliorer leur expérience de voyage.
2. Offres spéciales sur les services aéroportuaires : Les sites de réservation de vols en ligne pourraient négocier des partenariats avec des compagnies de services aéroportuaires (lounges, spa, restauration, etc.) pour offrir des réductions ou des avantages exclusifs aux voyageurs en cas de retard de vol. Cela créerait une valeur ajoutée pour les clients et renforcerait la satisfaction globale de l'expérience de voyage.
3. Programme de fidélité axé sur les retards de vol : Les sites de réservation de vols en ligne pourraient développer un programme de fidélité spécifique basé sur les retards de vol. Les voyageurs accumuleraient des points ou des avantages spéciaux chaque

fois qu'ils sont confrontés à un retard de vol, ce qui les inciterait à continuer d'utiliser la plateforme et à bénéficier de récompenses supplémentaires.

4. Service de préparation au retard de vol : Les sites de réservation pourraient proposer un service de préparation au retard de vol qui fournirait aux voyageurs des informations et des conseils pratiques sur la manière de gérer efficacement un retard de vol. Cela pourrait inclure des guides de survie, des listes de vérification et des recommandations sur les activités ou les services disponibles à l'aéroport pendant l'attente.

En explorant ces opportunités commerciales, les sites de réservation de vols en ligne peuvent diversifier leur offre de services, améliorer l'expérience client et se positionner comme des acteurs clés dans la gestion proactive des retards de vols.

## **VI. Conclusion**

### **A. Récapitulation des objectifs et des résultats obtenus**

Le projet visant à évaluer les retards de vols, améliorer la satisfaction des clients et proposer des solutions pertinentes a été mené avec succès. Les résultats des modèles de régression logistique, SVM et réseau de neurones ont démontré des performances encourageantes, avec une précision élevée, un rappel solide et un score F1 équilibré. Cependant, la régression logistique a présenté la meilleure performance globale avec une précision plus élevée et un score F1 supérieur.

Les prédictions de retards de vols offrent de nouvelles possibilités pour les sites de réservation de vols en ligne. Il est recommandé d'intégrer ces prédictions dans les systèmes de communication des sites afin de fournir des alertes proactives aux clients concernés, renforçant ainsi leur confiance et leur fidélité. Le développement d'une application web ou mobile dédiée permettrait aux voyageurs d'accéder facilement aux prévisions de retard de leur vol, favorisant une meilleure planification des déplacements.

Pour une utilisation optimale des prédictions de retards de vols, il est recommandé de développer une application en suivant les étapes telles que le développement d'une interface utilisateur attrayante, l'intégration des données de prévisions de retard, la mise en place de fonctionnalités de notification, la personnalisation de l'expérience utilisateur, l'intégration avec d'autres services de voyage, la collecte de données et l'analyse, ainsi que la mise en place d'une campagne de promotion ciblée.

L'intégration des prédictions de retards de vols dans les systèmes de réservation des sites de réservation de vols en ligne offre également de nombreux avantages, tels que la proposition d'options de correspondance adaptées, les suggestions de changements d'aéroport, la gestion proactive des retards et la fourniture d'informations en temps réel aux voyageurs.

En conclusion, l'utilisation optimale des prédictions de retards de vols permettra d'améliorer l'expérience des clients, de réduire les désagréments liés aux retards et de renforcer la confiance envers les compagnies aériennes et les sites de réservation. Il est essentiel de mettre en œuvre les recommandations fonctionnelles mentionnées pour exploiter pleinement le potentiel de ces prédictions et offrir une gestion de la relation client optimale dans le secteur aérien.

### **B. Perspectives d'évolution et recommandations pour une utilisation optimale des prédictions de retards de vols**

Les prédictions de retards de vols offrent de nombreuses perspectives d'évolution et des recommandations pour une utilisation optimale. Voici quelques perspectives et recommandations à considérer :

1. Amélioration continue des modèles prédictifs : Les modèles utilisés pour prédire les retards de vols peuvent être continuellement améliorés en utilisant des techniques avancées d'apprentissage automatique, telles que les réseaux de neurones profonds, les algorithmes génétiques ou l'apprentissage par renforcement. En investissant dans la recherche et le développement, il est possible d'obtenir des prédictions encore plus

précises et fiables, ce qui contribuera à améliorer la satisfaction des clients et à réduire les coûts associés aux retards de vols.

2. Intégration de données en temps réel : Pour améliorer l'exactitude des prédictions, il est recommandé d'intégrer des données en temps réel dans les modèles. Cela peut inclure des informations sur les conditions météorologiques actuelles, les problèmes opérationnels spécifiques à l'aéroport, les événements exceptionnels, etc. En utilisant des sources de données en temps réel et en mettant à jour régulièrement les modèles, il est possible de prendre en compte les facteurs changeants et d'obtenir des prédictions plus précises.
3. Collaboration avec les compagnies aériennes et les aéroports : Pour une utilisation optimale des prédictions de retards de vols, il est essentiel de collaborer étroitement avec les compagnies aériennes et les aéroports. Ces partenariats permettront d'échanger des données précieuses, d'identifier les principaux facteurs de retard et de mettre en place des mesures d'atténuation appropriées. En travaillant ensemble, il est possible de mettre en œuvre des solutions intégrées qui bénéficieront à l'ensemble de l'écosystème de l'aviation.

## VII. Annexes

### Annexe 01 : Canvas

Décisions	ML task	Value Proportion	Data Sources	Collecting Data
<p>les prédictions des retards des vols seront déployés sur les axes suivants:</p> <ul style="list-style-type: none"> <li>• Communication proactive et intégration d'alertes;</li> <li>• Intégration dans une application web ou mobile;</li> <li>• Intégration dans les systèmes de réservation des sites;</li> <li>• Intégration dans le service client;</li> <li>• Amélioration de la segmentation client.</li> </ul>	<p>Input: Données sur les vols et aéroports. Output: Prediction qu'un vol fasse du retard ou pas</p>	<p>Prédire si un vol choisi par un client sur un site de réservation en ligne fera du retard afin de contribuer à améliorer leur expérience de réservation de vols et à réduire les inconvénients causés par les retards.</p>	<p>Le fichier batch_1.db contient 3 tables :</p> <ul style="list-style-type: none"> <li>• aéroports</li> <li>• compagnies</li> <li>• vols</li> </ul>	<p>Extraction des données en effectuant les opérations suivantes :</p> <ul style="list-style-type: none"> <li>• Il importe les bibliothèques nécessaires : pandas et sqlite3.</li> <li>• Il se connecte à une base de données SQLite spécifiée (batch_1.db), en utilisant la fonction sqlite3.connect().</li> <li>• Il exécute une requête SQL pour récupérer les données de la table "aéroports" à partir de la base de données, en utilisant la fonction pd.read_sql_query(). Les données sont ensuite stockées dans un DataFrame appelé "aéroports".</li> </ul>
	<p><b>Offline Evaluation</b></p> <ul style="list-style-type: none"> <li>• Régression logistique</li> <li>• SVM</li> <li>• Réseau de neurones</li> <li>• Exactitude (Accuracy)</li> <li>• Précision (Precision)</li> <li>• Rappel (Recall)</li> <li>• Score F1</li> </ul>		<p><b>Features</b></p> <ul style="list-style-type: none"> <li>• IDENTIFIANT</li> <li>• VOL</li> <li>• DEPART PROGRAMME</li> <li>• HEURE DE DEPART</li> <li>• DECOLLAGE</li> <li>• TEMPS PROGRAMME</li> <li>• TEMPS PASSE</li> <li>• TEMPS DE VOL</li> <li>• DISTANCE</li> <li>• ATTERRISSAGE</li> <li>• ARRIVEE PROGRAMMEE</li> <li>• HEURE D'ARRIVEE</li> <li>• DETOURNEMENT</li> <li>• ANNULATION</li> <li>• RETARD SYSTEM</li> <li>• RETARD SECURITE</li> <li>• RETARD COMPAGNIE</li> <li>• RETARD AVION</li> <li>• RETARD METEO</li> <li>• DATE</li> <li>• NIVEAU DE SECURITE</li> <li>• PRIX RETARD PREMIERE 10 MINUTES DEPART</li> <li>• PRIX RETARD POUR CHAQUE MINUTE APRES 10 MINUTES DEPART</li> <li>• PRIX RETARD PREMIERE 10 MINUTES ARRIVEE</li> <li>• PRIX RETARD POUR CHAQUE MINUTE APRES 10 MINUTES ARRIVEE</li> <li>• RETARD MOYEN</li> <li>• RAISON D'ANNULATION</li> <li>• COMPAGNIE AERIENNE</li> <li>• CODE IATA DEPART</li> </ul>	<p><b>Building Models</b></p> <p>Mise à jour du training data set et fine tune des paramètres chaque mois.</p>

			<div><div></div><div>CODE IATA ARRIVEE</div><div></div><div>Retard</div></div>	
<div><div>Making Prediction</div><div>Faire des prédictions en fonction de la mise a jour de la base principale d' IMSD 22-23 qui est mensuelle.</div></div>	<div><div>Live Evaluation and Monitoring</div><div>Évaluation du modèle par la mise en place d'un monitoring des notes de scores.</div></div>			

Annexe 02 : tableau statistique de la table aéroports

	HAUTEUR	PRIX RETARD PREMIERE 10 MINUTES	PRIS RETARD POUR CHAQUE MINUTE APRES 10 MINUTES
count	321.000000	321.000000	321.000000
mean	711.283489	5167.186916	502.947040
std	1232.281912	2820.141750	281.834161
min	6.000000	105.000000	11.000000
25%	59.000000	2738.000000	267.000000
50%	220.000000	5499.000000	510.000000
75%	685.000000	7489.000000	738.000000
max	7630.000000	9877.000000	995.000000

Annexe 03 : tableau statistique de la table vols

	IDENTIFIANT	VOL	DEPART PROGRAMME	HEURE DE DEPART	RETARD DE DEPART	TEMPS DE DEPLACEMENT A TERRE AU DECOLLAGE	DECOLLAGE	TEMPS PROGRAMME	TEMPS PASSE	TEMPS DE VOL
count	3.000000e+06	3.000000e+06	3.000000e+06	2.950774e+06	2.950774e+06	2.948930e+06	2.948930e+06	2.999994e+06	2.940415e+06	2.940415e+06
mean	2.616279e+06	2.274759e+03	1.328864e+03	1.333119e+03	9.455701e+00	1.725790e+01	1.355781e+03	1.458505e+02	1.416684e+02	1.166334e+02
std	1.727447e+06	1.824416e+03	4.854302e+02	4.980278e+02	3.914365e+01	9.272093e+00	5.002325e+02	7.907005e+01	7.815676e+01	7.636377e+01
min	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	-8.200000e+01	1.000000e+00	1.000000e+00	1.800000e+01	1.500000e+01	7.000000e+00
25%	1.043038e+06	7.800000e+02	9.160000e+02	9.200000e+02	-5.000000e+00	1.200000e+01	9.360000e+02	8.800000e+01	8.500000e+01	6.000000e+01
50%	2.089116e+06	1.730000e+03	1.325000e+03	1.329000e+03	-2.000000e+00	1.500000e+01	1.342000e+03	1.250000e+02	1.220000e+02	9.600000e+01
75%	4.288817e+06	3.414000e+03	1.730000e+03	1.738000e+03	7.000000e+00	2.000000e+01	1.752000e+03	1.790000e+02	1.750000e+02	1.480000e+02
max	5.332913e+06	9.855000e+03	2.359000e+03	2.400000e+03	1.988000e+03	2.000000e+02	2.400000e+03	7.180000e+02	7.660000e+02	6.900000e+02

	DISTANCE	ATTERRISSAGE	TEMPS DE DEPLACEMENT A TERRE A L'ATTERRISSAGE	ARRIVEE PROGRAMMEE	HEURE D'ARRIVEE	RETARD A L'ARRIVEE	DETOURNEMENT	ANNULATION	RETARD SYSTEM	RETARD SECURITE
	3.000000e+06	2.947124e+06	2.947124e+06	3.000000e+06	2.947124e+06	2.940415e+06	3.000000e+06	3.000000e+06	560231.000000	560231.000000
	8.460305e+02	1.471310e+03	7.789882e+00	1.493656e+03	1.476963e+03	4.914760e+00	2.666333e-03	1.719533e-02	15.351337	0.08736
	6.436554e+02	5.199641e+02	5.981326e+00	5.052021e+02	5.240664e+02	4.140800e+01	5.156767e-02	1.299987e-01	30.331968	2.32786
	3.100000e+01	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	-8.700000e+01	0.000000e+00	0.000000e+00	0.000000	0.00000
	3.700000e+02	1.055000e+03	4.000000e+00	1.110000e+03	1.100000e+03	-1.400000e+01	0.000000e+00	0.000000e+00	0.000000	0.00000
	6.650000e+02	1.508000e+03	6.000000e+00	1.520000e+03	1.512000e+03	-5.000000e+00	0.000000e+00	0.000000e+00	4.000000	0.00000
	1.089000e+03	1.910000e+03	9.000000e+00	1.917000e+03	1.916000e+03	8.000000e+00	0.000000e+00	0.000000e+00	20.000000	0.00000

#### Annexe 04 : tableau statistique de la table campagnes

COMPAGNIE CODE		
count	13	13
unique	13	13
top	Try Hard Airlines	THA
freq	1	1

#### Annexe 05 : Pourcentage des valeurs vides dans chaque

IDENTIFIANT	0.000000
VOL	0.000000
DEPART PROGRAMME	0.000000
HEURE DE DEPART	1.640867
DECOLLAGES	1.702333
TEMPS PROGRAMME	0.000200
TEMPS PASSE	1.986167
TEMPS DE VOL	1.986167
DISTANCE	0.000000
ATTERRISSAGE	1.762533
ARRIVEE PROGRAMMEE	0.000000
HEURE D'ARRIVEE	1.762533
DETOURNEMENT	0.000000
ANNULATION	0.000000
RETARD SYSTEM	0.000000
RETARD SECURITE	0.000000
RETARD COMPAGNIE	0.000000
RETARD AVION	0.000000
RETARD METEO	0.000000
DATE	0.000000
NIVEAU DE SECURITE	0.000000
PRIX RETARD PREMIERE 10	0.000000
PRIS RETARD POUR CHAQUEMINUTES	0.000000
PRIX RETARD PREMIERE 10	0.000000
PRIS RETARD POUR CHAQUE	0.000000

RETARD MOYEN	1.986167
RAISON D'ANNULATION	0.000000
COMPAGNIE AERIENNE	0.000000
CODE IATA DEPART	0.000000
CODE IATA ARRIVEE	0.000000
Retard	0.000000