

Table des matières

Remerciements.....	3
Dédicace.....	4
Introduction générale	8
I. Présentation de l'organisme d'accueil.....	9
1. La Royal Air Maroc.....	9
a. Présentation.....	9
b. Historique	9
c. Activités.....	10
2. Siège social de la Royal Air Maroc (RAM) Casablanca.....	10
a. Rôle du siège social de la (RAM) Casablanca	10
b. Organigramme	11
.....	12
.....	12
II. Construction d'un modèle de veille concurrentielle.....	14
1. Contextualisation et problématique	14
2. Solution de la problématique	16
3. Construction et exploration des bases de données.....	16
a. Fichier SSIM (Standard Schedules Information Manual).....	16
.....	18
b. Construction de la base de données.....	19
4. Choix de l'algorithme de détection d'anomalie.....	22
a. Fonctionnement du l'Isolation Forest.....	24

b. Isolation Forest dans notre cas	25
5. Prétraitement des données	26
a. Suppression des lignes vides et des valeurs dupliquées	27
b. Introduction de la variable Circuity.....	28
c. Introduction des vols dans le territoire Marocain.....	31
d. Encodage des données catégorielles.....	32
6. Entraînement du modèle	33
7. Résultat finale	34
III. Visualisation des anomalies.....	35
Conclusion	39

LISTE DES FIGURES

FIGURE 1: LOGO DE LA ROYAL AIR MAROC	9
FIGURE 2: ORGANIGRAMME DE LA PRESIDENCE DIRECTION GENERALE DU SIEGE SOCIALE CASABLANCA DE LA ROYAL AIR MAROC	12
FIGURE 3: CAPTURE D'UN FICHIER SSIM DE LA ROYAL AIR MAROC	17
FIGURE 4: LIGNE DU FICHIER SSIM DE LA ROYAL AIR MAROC.....	18
FIGURE 5: SUPPRESSION DES CARACTERES INUTILES	19
FIGURE 6: EXTRACTION DES COLONNES SOUHAITEES	20
FIGURE 7: CREATION DU DATAFRAME	21
FIGURE 8: DATAFRAME CONTENANT LES DONNEES NETTOYEES	21
FIGURE 9: LOGO DE L'ISOLATION FOREST	23
FIGURE 10: DIAGRAMME DES ARBRES DE DECISION UTILISES PAR L'ISOLATION FOREST	24
FIGURE 11: APERÇU INITIAL SUR LA BASE DE DONNEES.....	26
FIGURE 12: LISTE DES VARIABLES DE LA BASE DE DONNEES.....	27
FIGURE 13: SUPPRESSION DES LIGNES VIDES ET DES VALEURS DUPLIQUEES.....	28
FIGURE 14: BASE DE DONNEES CONTENANT LES CIRCUITYS ET D'AUTRES DETAILS SUR LES VOLS	29
FIGURE 15 : EXTRACTION ET REGROUPEMENT DES CIRCUITYS DES VOLS	30
FIGURE 16: AJOUT DE LA VARIABLE CIRCUITY DANS LA BASE DE DONNEES.....	31
FIGURE 17 : LISTE DES AEROPORTS MAROCAINS	31
FIGURE 18 : APERÇU DE LA BASE DONNEE APRES AJOUT DES CIRCUITYS DES VOLS DEPARTENT OU ARRIVENT AU MAROC.....	32
FIGURE 19: ENCODAGE DES DONNEES CATEGORIELLES	33
FIGURE 20: ENTRAINEMENT DU MODELE	34
FIGURE 21: DETECTION DES ANOMALIES.....	35
FIGURE 22: NUAGE DE POINTS DES ANOMALIES (FREQUENCES DES VOLS PAR COMPAGNIE)	36
FIGURE 23: NUAGE DE POINTS DE LA COMPAGNIE TUNISAIR	37
FIGURE 24: NUAGE DE POINTS DES ANOMALIES (ROUTES PAR COMPAGNIE)	37
FIGURE 25: FICHIER EXCEL CONTENANT LES VOLS QUI PRESENTENT DES ANOMALIES.....	38

Introduction générale

Afin de compléter mes études en deuxième année à l'institut national de statistique et d'économie appliquée, de parfaire et d'appliquer mes connaissances jusqu'à la pratique . Je me suis dirigé vers le Département Planning Réseau & Scheduling pour effectuer un stage d'application se déroulant dans la période allant du 03/07/2023 au 03/09/2023.

L'objectif de mon stage est de construire un modèle de veille concurrentielle basé sur le Machine Learning afin d'analyser les changements des planifications des compagnies aériennes concurrentes au réseau Royal Air Maroc.

Par ailleurs, mon rapport de stage se décomposera en trois parties. D'abord la présentation de l'organisme d'accueil afin de mieux comprendre et contextualiser les objectifs de ma mission. Ensuite les détails de la construction de mon modèle d'analyse concurrentielle et les résultats du modèle. Puis pour finir le compte rendu des compétences techniques, personnelles et professionnelles que j'ai acquises durant mon stage.

I. Présentation de l'organisme d'accueil :

1. La Royal Air Maroc :

a. Présentation :



Figure 1: Logo de la Royal Air Maroc

Royal Air Maroc est la compagnie aérienne nationale du Maroc. Elle a été fondée en 1957 et est basée à l'aéroport international Mohammed V de Casablanca. Royal Air Maroc est la plus grande compagnie aérienne du Maroc et l'une des principales compagnies aériennes d'Afrique.

La compagnie dessert plus de 100 destinations dans plus de 50 pays, principalement en Afrique, en Europe et au Moyen-Orient. Elle exploite une flotte de plus de 50 avions.

Royal Air Maroc propose des vols réguliers, des vols charters et des vols cargo. Elle offre également une gamme de services aux passagers, notamment des salons d'aéroport, un programme de fidélité et des services de restauration à bord.

La compagnie a reçu de nombreux prix pour sa qualité de service, notamment le prix Skytrax World Airline Award de la meilleure compagnie aérienne africaine en 2022.

b. Historique :

La Royal Air Maroc (RAM) a été fondée en 1957, quelques mois après l'indépendance du Maroc. Elle est née de la fusion de deux compagnies aériennes créées pendant la Seconde Guerre mondiale : Air Atlas et Air Maroc.

La RAM a commencé par desservir une dizaine de destinations en Afrique du Nord et en Europe. Elle s'est rapidement développée, acquérant de nouveaux avions et ouvrant de nouvelles destinations. En 1973, elle a lancé son premier vol transatlantique, vers New York.

Au cours des années suivantes, la RAM a continué à se développer, devenant la plus grande compagnie aérienne du Maroc. Elle dessert aujourd'hui plus de 100 destinations dans plus de 50 pays.

c. Activités :

Les activités de la RAM se concentrent sur le transport de passagers et de fret. La compagnie propose une gamme de services aux passagers, notamment des salons d'aéroport, un programme de fidélité et des services de restauration à bord.

La RAM est une entreprise importante pour l'économie marocaine. Elle contribue au développement du tourisme et du commerce, et elle crée des emplois pour des milliers de personnes.

Voici quelques-unes des activités spécifiques de la RAM :

Transport de passagers : la RAM propose des vols réguliers vers plus de 100 destinations dans plus de 50 pays. Elle dessert également des vols charters vers des destinations spécifiques, telles que les vacances scolaires ou les événements sportifs.

Transport de fret : la RAM transporte des marchandises et des produits à travers le monde. Elle dispose d'une flotte de 1 avion-cargo.

Services aux passagers : la RAM propose une gamme de services aux passagers, notamment des salons d'aéroport, un programme de fidélité et des services de restauration à bord.

2. Siège social de la Royal Air Maroc (RAM) Casablanca :

a. Rôle du siège social de la (RAM) Casablanca :

Le siège social de la Royal Air Maroc (RAM) situé à Casablanca est le centre névralgique de l'entreprise. Il abrite les bureaux de la direction générale, des départements administratifs et financiers, et des équipes commerciales et marketing.

Voici quelques-unes des principales fonctions du siège social de la RAM :

Gestion et coordination : le siège social est responsable de la gestion générale de la RAM, y compris la définition de la stratégie, la supervision des opérations, et la coordination des activités des différentes divisions de l'entreprise.

Développement : le siège social est responsable du développement de la RAM, y compris le lancement de nouveaux produits et services, l'expansion du réseau de destinations, et l'amélioration de la qualité du service.

Communication : le siège social est responsable de la communication de la RAM, y compris les relations publiques, le marketing, et la communication avec les clients, les partenaires, et les investisseurs.

b. Organigramme :

L'organigramme fonctionnel de la Royal Air Maroc du siège social Casablanca se présente comme suit en juillet 2023 :

Organigramme de la Présidence Direction Générale



Figure 2: Organigramme de la Présidence Direction Générale du siège sociale Casablanca de la Royal Air Maroc

- **Pôle Transformation** : Ce pôle est responsable de la transformation digitale et de l'innovation de la compagnie aérienne. Il s'agit d'un pôle transversal qui travaille en étroite collaboration avec les autres pôles de la compagnie aérienne. Le pôle Transformation est responsable du développement de nouveaux produits et services numériques, de l'amélioration de l'expérience client, et de l'optimisation des opérations de la compagnie aérienne.
- **Pôle Expérience Client** : Ce pôle est responsable de la gestion de l'expérience client

de la compagnie aérienne. Il s'agit d'un pôle essentiel pour la compagnie aérienne, car il est responsable de la satisfaction des clients. Le pôle Expérience Client travaille en étroite collaboration avec les autres pôles de la compagnie aérienne pour garantir une expérience client de qualité à tous les niveaux.

- **Pôle Exportation** : Ce pôle est responsable du développement du réseau de destinations internationales de la compagnie aérienne. Il travaille en étroite collaboration avec les équipes commerciales et marketing de la compagnie aérienne pour identifier de nouvelles opportunités de croissance à l'étranger. Le PO Exportation est également responsable de la gestion des relations avec les partenaires commerciaux internationaux de la compagnie aérienne.
- **Pôle Commercial** : Ce pôle est responsable de la gestion des opérations aériennes de la compagnie aérienne, y compris la gestion des vols, la maintenance des avions, et la vente des billets. Il s'agit d'un pôle essentiel pour la compagnie aérienne, car il est responsable de l'exécution de son activité principale. Le pôle Vol Commercial travaille en étroite collaboration avec les autres pôles de la compagnie aérienne pour garantir la sécurité et la ponctualité des vols, et pour fournir une expérience client de qualité.
- **Pôle Support** : Ce pôle est responsable des fonctions de support de la compagnie aérienne, y compris les finances, les ressources humaines, et la communication. Il s'agit d'un pôle important qui fournit les ressources et les services nécessaires au bon fonctionnement des autres pôles de la compagnie aérienne. Le pôle Support travaille en étroite collaboration avec les autres pôles pour garantir la performance globale de la compagnie aérienne.

La Direction Réseau et Revue Management (DRRM) de la Royal Air Maroc est responsable de la gestion du réseau de destinations de la compagnie aérienne, ainsi que de la revue de la performance des opérations de la compagnie aérienne.

La DRRM est composée de deux départements :

- Le département Planning Réseau & Scheduling est responsable de la définition du réseau de destinations de la compagnie aérienne, de la gestion des contrats avec les aéroports, et de la coordination avec les autres départements de la compagnie aérienne pour garantir la cohérence du réseau.
- Le département Revue Management est responsable de la revue de la performance des opérations de la compagnie aérienne, y compris la sécurité, la ponctualité, et la satisfaction des clients.

Mon stage a été effectué dans le Département Planning Réseau & Scheduling de la Royal Air Maroc est responsable de la planification et de la programmation des vols de la compagnie aérienne. Il est composé de plusieurs équipes qui travaillent en étroite collaboration pour garantir que les vols de la RAM soient sûrs, ponctuels et efficaces.

Les principales responsabilités du Département Planning Réseau & Scheduling sont les suivantes :

- Planifier le réseau de destinations de la RAM
- Programmer les vols de la RAM
- Coordonner avec les autres départements de la RAM pour garantir la cohérence du planning
- Gérer les ressources de la RAM, notamment les avions, les équipages et les aéroports
- Analyser les données de performance des vols
- Développer de nouvelles méthodes de planification et de programmation

II. Construction d'un modèle de veille concurrentielle :

1. Contextualisation et problématique :

Les compagnies aériennes sont constamment en concurrence pour attirer et fidéliser les clients. Elles sont donc susceptibles d'apporter des changements à leurs plans de vol pour

répondre aux besoins et aux attentes des clients. La Royal Air Maroc doit être notifiée de ces changements importants pour pouvoir en tirer profit.

Les changements importants apportés par les compagnies concurrentes peuvent inclure :

Une grande diminution de la fréquence d'un vol

Un grand ajout de la fréquence d'un vol

L'ajout ou la suppression d'une destination

La Royal Air Maroc peut utiliser ces informations pour identifier les opportunités de croissance et de développement. Elle peut également protéger sa part de marché et renforcer sa position concurrentielle.

Pour être notifiée de ces changements importants, la Royal Air Maroc peut :

Suivre de près les activités des compagnies concurrentes, notamment leurs plans de vol, leurs programmes de fidélité et leurs promotions.

Analyser les données sur le marché du transport aérien pour identifier les tendances et les opportunités.

Mettre en place un système de veille concurrentielle pour identifier les changements potentiels des compagnies concurrentes.

En prenant ces mesures, la Royal Air Maroc peut être mieux préparée à répondre aux changements des compagnies concurrentes et à en tirer profit.

Voici quelques exemples de la façon dont la Royal Air Maroc peut utiliser ces informations :

Si une compagnie concurrente réduit la fréquence d'un vol vers une destination populaire, la Royal Air Maroc peut envisager d'augmenter la fréquence de ses vols vers cette destination pour répondre à la demande.

Si une compagnie concurrente ajoute une nouvelle destination à son réseau, la Royal Air Maroc peut envisager d'ajouter cette destination à son propre réseau pour concurrencer la compagnie concurrente.

En étant vigilante par rapport aux changements des compagnies concurrentes, la Royal Air Maroc peut améliorer sa compétitivité et attirer plus de clients.

2. Solution de la problématique :

La solution pour la Royal Air Maroc est de construire un système de veille concurrentielle qui consiste sur un modèle de détection d'anomalies dans les planifications des vols des compagnies aériennes concurrentes.

Ce système permettrait à la compagnie aérienne de suivre de près les activités des compagnies concurrentes et d'identifier rapidement les changements importants dans leurs plans de vol. Ces changements pourraient être des anomalies, c'est-à-dire des changements qui ne correspondent pas aux tendances habituelles de la compagnie concurrente.

Le modèle de détection d'anomalies serait formé sur un ensemble de données historiques de plans de vols des compagnies aériennes concurrentes. Ce modèle serait ensuite utilisé pour identifier les changements qui ne correspondent pas aux tendances habituelles.

Si le système de veille concurrentielle identifie une anomalie, la Royal Air Maroc serait avertie de ce changement. Cela permettrait à la compagnie aérienne de réagir rapidement et de prendre des mesures pour tirer profit de l'opportunité.

3. Construction et exploration des bases de données :

a. Fichier SSIM (Standard Schedules Information Manual) :

Les fichiers SSIM (Standard Schedules Information Manual) sont un format standard utilisé pour échanger des informations sur les plans de vol des compagnies aériennes des 15 prochaines semaines. Ces fichiers contiennent des informations détaillées sur chaque vol, telles que le type d'avion, l'aéroport de départ et d'arrivée, l'heure de départ et d'arrivée...

Une fois que la Royal Air Maroc a obtenu les fichiers SSIM des compagnies aériennes concurrentes, elle peut les utiliser pour alimenter son système de veille concurrentielle. La Royal Air Maroc construira au début de chaque semaine une base de données des plans de vol des compagnies aériennes concurrentes à partir des fichiers SSIM partagés chaque lundi. Le système de veille concurrentielle utilisera des techniques d'apprentissage automatique pour

SCHEDULE DATA SET	1	000000001
-------------------	---	-----------

Keywords: child; family; self-esteem; social skills training; SSIM; treatment; youth

Figure 1: A plot of the function $f(x) = \frac{1}{2}x^2 + \frac{1}{4}x^4$ for $x \in [-1, 1]$. The function is symmetric about the y-axis and has a minimum at $x = 0$. The plot shows the function's behavior over the interval $[-1, 1]$ with a grid.

Figure 4: Ligne du fichier SSIM de la Royal Air Maroc

**3 AT 2000101J03JUL2328OCT2312345671CMN14501450+0100 JFK22402240-0500
789 J26Y276 000003**

- **3 AT** : Ce code indique que le vol est opéré par la compagnie aérienne Royal Air Maroc (AT).
- **2000101J03JUL2328OCT2312345671** : Ce code est le numéro de vol. Il est composé de 12 chiffres, les six premiers représentant la date de départ, les six suivants représentant la date d'arrivée et les trois derniers représentant le numéro de vol. Dans ce cas, le vol partira le 03 juillet 2023 et arrivera le 28 octobre 2023. Le numéro de vol est 12345671.
- **CMN14501450+0100** : Ce code est l'heure de départ de l'aéroport de départ. L'heure est indiquée en format 24 heures, avec le fuseau horaire indiqué après le signe + ou -. Dans ce cas, le vol partira de l'aéroport de Casablanca (CMN) à 14 h 50, heure locale.
- **JFK22402240-0500** : Ce code est l'heure d'arrivée à l'aéroport d'arrivée. L'heure est indiquée en format 24 heures, avec le fuseau horaire indiqué après le signe + ou -. Dans ce cas, le vol arrivera à l'aéroport de New York-JFK (JFK) à 22 h 40, heure locale.
- **789** : Ce code est le type d'avion utilisé pour le vol. Dans ce cas, l'avion est un Boeing 787.
- **J26Y276** : Ce code est le numéro de série de l'avion utilisé pour le vol.
- **000003** : Ce code est le numéro de séquence du vol dans le fichier SSIM.

b. Construction de la base de données :

Pour construire notre base de données des planifications des compagnies aériennes concurrentes à la Royal Air Maroc, nous allons regrouper les fichiers SSIM de ces compagnies. Les informations qui nous intéressent dans les fichiers SSIM :

- La compagnie aérienne : Cette information est utilisée pour identifier la compagnie aérienne qui exploite le vol.
- La date du vol : Cette information est utilisée pour suivre les changements dans les plans de vol des compagnies aériennes concurrentes.
- L'aéroport de départ / d'arrivée : Cette information est utilisée pour voir où les compagnies aériennes concurrentes proposent des vols.
- La fréquence du vol : Cette information est utilisée pour voir à quelle fréquence les compagnies aériennes concurrentes proposent des vols.

Après avoir rassemblé les fichiers SSIM des compagnies aériennes, nous allons passer à la phase du traitement pour créer une base de données simple à manipuler. Les étapes à suivre sont :

Étape 1 : Supprimer les caractères inutiles

Pour supprimer les caractères inutiles, les espaces, les tabulations, les retours à la ligne.

```
import re

def nettoyer_texte(texte):
    texte = re.sub(r'\s+|\t|\r|\n', ' ', texte)
    texte = re.sub(r'<[^>]*>', '', texte)
    return texte
```

Figure 5: Suppression des caractères inutiles

Etape 2 : Extraire les colonnes souhaitées

Pour extraire les colonnes souhaitées, nous pouvons utiliser des expressions régulières. Voici une expression régulière qui extraira la date du vol, l'aéroport de départ, l'aéroport de destination et le nombre de vols par jour :

```
import re

def extraire_colonnes(texte):
    # Date du vol
    date_du_vol = re.search(r'\d{8,10}', texte).group()
    date_du_vol = pd.to_datetime(date_du_vol)

    # Aéroport de départ
    aéroport_de_départ = re.search(r'\w{3}', texte).group()

    # Aéroport de destination
    aéroport_de_destination = re.search(r'\w{3}\d{4}', texte).group()

    # Compagnie aérienne
    Mkt Al = re.search(r'\w{3}', texte).group()

    # Nombre de vols par jour
    Somme de Ops = re.search(r'\d{1,2}', texte).group()

    return date_du_vol, aéroport_de_départ, aéroport_de_destination, Mkt Al , Somme de Ops
```

Figure 6: Extraction des colonnes souhaitées

Étape 3 : Créer un DataFrame

Une fois que nous avons extrait les colonnes souhaitées, nous pouvons les utiliser pour créer un DataFrame Pandas.


```
import pandas as pd

# Extraire les colonnes des données nettoyées
date_du_vol, aéroport_de_départ, aéroport_de_destination, Mkt Al, Somme de Ops = extraire_colonnes(donnees_nettoyees)

# Créer un DataFrame
df = pd.DataFrame({
    'Mkt Al': compagnie_aérienne,
    'date_du_vol': date_du_vol,
    'aéroport_de_départ': aéroport_de_départ,
    'aéroport_de_destination': aéroport_de_destination,
    'Somme de Ops': nombre_de_vols_par_jour
})
```

Figure 7: Création du DataFrame

Étape 4 : Visualiser les données

Une fois que nous avons créé un DataFrame, nous pouvons le visualiser pour vérifier que les données ont été correctement nettoyées.

00000	Mkt Al	date_du_vol	aéroport_de_départ	aéroport_de_destination	Somme de Ops	route
0	AF	08OCT23	ABJ	CDG	000003	ABJ - CDG
1	AF	09SEP23	ACC	CDG	000004	ACC - CDG
2	AF	05AUG23	AGA	CDG	000001	AGA - CDG
3	AF	11AUG23	AGP	CDG	000004	AGP - CDG
4	AF	25AUG23	ALG	ORY	000012	ALG - ORY
5	AF	14SEP23	BOD	NCE	000005	BOD - NCE
...
51960	WB	07AUG23	LBV	COO	000003	LBV - COO
51961	WB	09AUG23	DLA	BGF	000002	DLA - BGF
51962	WB	16AUG23	DLA	BZV	000001	DLA - BZV
51963	WB	22SEP23	BGF	DLA	000007	BGF - DLA
51964	WB	19SEP23	DLA	COO	000001	DLA - COO

Figure 8: DataFrame contenant les données nettoyées

Nous disposons maintenant d'une base de données qui contient les informations nécessaires pour construire notre modèle de détection d'anomalie.

Ces informations sont essentielles pour construire un modèle de détection d'anomalie efficace. En effet, elles permettent de caractériser les données normales et de détecter les données qui s'écartent de la norme.

Maintenant que nous disposons d'une base de données de qualité, nous pouvons passer à la

phase suivante du projet, qui consiste à construire notre modèle de détection d'anomalie.

Voici les étapes que nous allons suivre pour construire notre modèle :

- ✓ **Choix de l'algorithme de détection d'anomalie**
- ✓ **Prétraitement des données**
- ✓ **Entraînement du modèle**
- ✓ **Évaluation du modèle**

4. Choix de l'algorithme de détection d'anomalie

La détection d'anomalie par le Machine Learning est une technique qui permet de détecter les données qui s'écartent de la norme. Cette technique est utilisée dans de nombreux domaines, tels que la sécurité, la finance, la santé et la logistique.

Il existe de nombreux algorithmes de détection d'anomalie disponibles, chacun avec ses propres avantages et inconvénients. Les principaux types d'algorithmes sont les suivants :

- **Les algorithmes basés sur la distance** : ces algorithmes calculent la distance entre une donnée et les données normales. Les données qui sont trop éloignées des données normales sont considérées comme des anomalies.
- **Les algorithmes basés sur la distribution** : ces algorithmes supposent que les données normales suivent une distribution particulière. Les données qui ne suivent pas cette distribution sont considérées comme des anomalies.
- **Les algorithmes basés sur les règles** : ces algorithmes utilisent des règles prédéfinies pour identifier les anomalies.

Les algorithmes basés sur la distribution ne sont pas adaptés à notre cas d'utilisation car ils supposent que les données normales suivent une distribution particulière. Les données que nous utilisons pour notre projet sont des données de vol, qui peuvent être très variables en fonction de nombreux facteurs, tels que la date, l'heure, la destination, etc. Il est donc difficile

de définir une distribution particulière qui représente les données normales.

Les algorithmes basés sur les règles ne sont pas adaptés à notre cas d'utilisation car ils nécessitent la définition d'un ensemble de règles prédéfinies. Ces règles doivent être définies manuellement, ce qui peut être un processus long et fastidieux. De plus, les règles doivent être adaptées aux données spécifiques que nous utilisons, ce qui peut être difficile à faire.

- ✓ Dans le cas de notre projet, nous avons choisi d'utiliser l'algorithme d'Isolation Forest. Cet algorithme est un algorithme basé sur la distance qui est efficace pour détecter les anomalies dans les données de grande taille.



Figure 9: Logo de l'Isolation Forest

L'Isolation Forest fonctionne en construisant un ensemble d'arbres de décision. Chaque arbre de décision est utilisé pour prédire si une donnée est normale ou non. Les données qui sont isolées des autres données dans les arbres de décision sont considérées comme des anomalies.

L'isolation Forest est un algorithme efficace pour détecter les anomalies dans les données de grande taille car il est très rapide et facile à mettre en œuvre. Il est également relativement robuste aux perturbations des données.

Voici quelques-uns des avantages de l'utilisation de l'Isolation Forest pour la détection d'anomalie :

- ✓ **Rapidité** : L'Isolation Forest est un algorithme très rapide, ce qui le rend adapté à la détection d'anomalie dans les données de grande taille.

Fiabilité : L'Isolation Forest est un algorithme relativement robuste aux perturbations des données.

- ✓ **Simplicité** : L'Isolation Forest est un algorithme relativement simple à mettre en

œuvre.

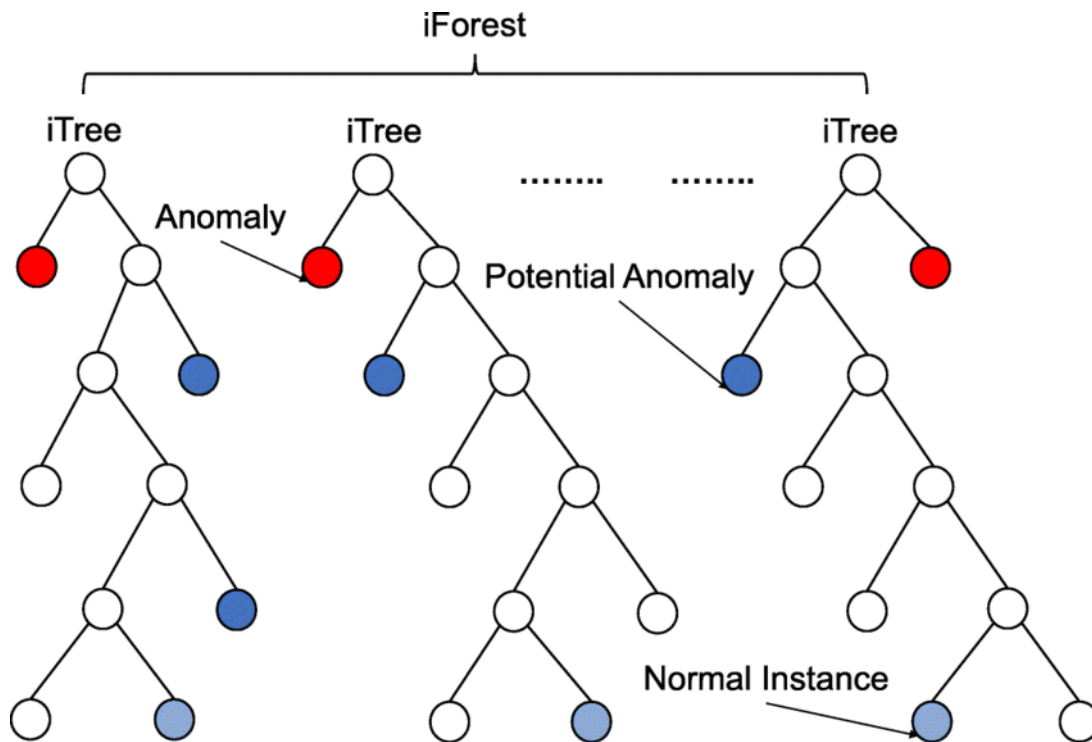


Figure 10: Diagramme des arbres de décision utilisés par l'Isolation Forest

a. Fonctionnement de l'Isolation Forest :

L'Isolation Forest fonctionne en construisant un ensemble d'arbres de décision. Chaque arbre de décision est utilisé pour prédire si une donnée est normale ou non. Les données qui sont isolées des autres données dans les arbres de décision sont considérées comme des anomalies.

L'image montre un ensemble de données avec deux classes de données : des données normales (en bleu) et des données anormales (en rouge). Les arbres de décision sont représentés par des lignes noires.

Le premier arbre de décision est construit en choisissant un attribut au hasard et en définissant

un seuil. Les données qui sont inférieures au seuil sont envoyées dans une feuille gauche, tandis que les données qui sont supérieures au seuil sont envoyées dans une feuille droite.

Le deuxième arbre de décision est construit en choisissant un autre attribut au hasard et en définissant un autre seuil. Les données qui sont inférieures au seuil sont envoyées dans une feuille gauche, tandis que les données qui sont supérieures au seuil sont envoyées dans une feuille droite.

Le processus se poursuit jusqu'à ce que tous les arbres de décision soient construits.

Dans l'image, la donnée anormale (en rouge) est isolée des autres données dans les arbres de décision. Cela signifie que la donnée anormale doit passer par un grand nombre de feuilles pour être isolée.

Le score d'anomalie d'une donnée est calculé en fonction du nombre de feuilles qu'elle doit traverser pour être isolée. Les données avec un score d'anomalie élevé sont considérées comme des anomalies.

b. Isolation Forest dans notre cas :

Dans notre cas de système de veille concurrentielle, nous utilisons l'Isolation Forest pour détecter les anomalies dans les données de vol. Les anomalies dans les données de vol peuvent être causées par de nombreux facteurs, tels que :

- ✓ Une modification importante de la fréquence d'un vol.
- ✓ L'ajout ou la suppression d'une destination.

5. Prétraitement des données

Pour commencer voilà à ce que ressemble la base de donnée qu'on va utiliser pour construire notre modèle :

	IMG	Sem	Mkt	Al	Route	Somme de Ops
0	2023-07-03 00:00:00	2023-07-03 00:00:00	TU	TUNDSS		1.0
1	2023-07-03 00:00:00	2023-07-03 00:00:00	TU	TUNFCO		6.0
2	2023-07-03 00:00:00	2023-07-03 00:00:00	TU	TUNFRA		7.0
3	2023-07-03 00:00:00	2023-07-03 00:00:00	TU	TUNGVA		8.0
4	2023-07-03 00:00:00	2023-07-03 00:00:00	TU	TUNIST		11.0
...
51960	2023-07-03 00:00:00	2023-10-23 00:00:00	TU	TUNALG		2.0
51961	2023-07-03 00:00:00	2023-10-23 00:00:00	TU	TUNBLQ		1.0
51962	2023-07-03 00:00:00	2023-10-23 00:00:00	TU	TUNBRU		1.0
51963	2023-07-03 00:00:00	2023-10-23 00:00:00	TU	TUNCMN		1.0
51964	(vide)	(vide)	(vide)	(vide)		NaN
51965 rows × 5 columns						

Figure 11: Aperçu initial sur la base de données

Elle est composée de 5 colonnes pour 51965 lignes, les colonnes sont les suivantes :

- **IMG** : La date du lundi auquel on a reçu les fichiers SSIM des compagnies aériennes concurrentes pour construire notre base de données, dans notre cas c'est le lundi 03 juillet 2023. l'abréviation IMG signifie l'image des 15 prochaines semaines depuis ce lundi 03 juillet 2023. Cette colonne est importante pendant le traitement de plusieurs bases de données pour éviter la confusion, mais elle sera supprimer lors du prétraitement de la Data
- **Sem** : C'est la date du jour des planifications des vols pour chaque destination

- **Mkt Al** : Marketing Arline, c'est le code IATA (Association du transport aérien international) composé de deux lettres qui nous donne le nom de la compagnie aérienne responsable du vol
- **Route** : Destination Départ-Arrivé du vol
- **Somme de Ops** : Fréquence du vol pendant chaque jour

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51965 entries, 0 to 51964
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Sem             51965 non-null  object
1   Mkt Al          51965 non-null  object
2   Route           51965 non-null  object
3   Somme de Ops    51965 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.6+ MB
```

Figure 12: Liste des variables de la base de données

a. Suppression des lignes vides et des valeurs dupliquées :

La première étape est de supprimer les lignes vides et les valeurs dupliquées

- Suppression de toutes les lignes qui contiennent des valeurs NaN.
- Suppression toutes les lignes en double.
- Suppression toutes les lignes où la colonne Somme de Ops est égale à 0.

	Sem	Mkt	Al	Route	Somme de Ops
0	2023-07-03 00:00:00		TU	TUNDSS	1
1	2023-07-03 00:00:00		TU	TUNFCO	6
2	2023-07-03 00:00:00		TU	TUNFRA	7
3	2023-07-03 00:00:00		TU	TUNGVA	8
4	2023-07-03 00:00:00		TU	TUNIST	11
...
51959	2023-10-23 00:00:00		TU	TLSTUN	1
51960	2023-10-23 00:00:00		TU	TUNALG	2
51961	2023-10-23 00:00:00		TU	TUNBLQ	1
51962	2023-10-23 00:00:00		TU	TUNBRU	1
51963	2023-10-23 00:00:00		TU	TUNCMN	1
51963 rows × 4 columns					

Figure 13: Suppression des lignes vides et des valeurs dupliquées

b. Introduction de la variable Circuitry

Le Circuitry d'un vol est la distance totale parcourue par un avion entre son point de départ et son point d'arrivée, y compris les détours effectués pour éviter les zones de mauvais temps ou d'autres obstacles. Il est généralement exprimé en miles nautiques (NM).

Le Circuitry d'un vol peut être influencé par un certain nombre de facteurs, notamment :

- La position des aéroports de départ et d'arrivée
- La direction du vent
- La présence de zones de mauvais temps
- Les restrictions de vol

Le Circuitry peut avoir un impact significatif sur le coût d'un vol, car il peut entraîner une augmentation de la consommation de carburant et des temps de vol.

1	Étiquettes de lignes	Dep	Arr	Circuitry	Lat1	Long1	Lat2	Long2	D1	D2	D3
2	cmndoh	cmn	doh	100	33.3675003	-7.58997011	25.273056	51.608056	5734.43626	0	5734.43626
3	ABJACC	ABJ	ACC	1514.47683	5.26139021	-3.92629004	5.6051898	-0.166786	417.91045	3148.16439	3180.99248
4	ABJALG	ABJ	ALG	117.608663	5.26139021	-3.92629004	36.691002	3.21541	3569.78186	3148.16439	1050.20834
5	ABJBKO	ABJ	BKO	593.130715	5.26139021	-3.92629004	12.5335	-7.94994	921.39685	3148.16439	2316.92334
6	ABJBRU	ABJ	BRU	103.816644	5.26139021	-3.92629004	50.9014015	4.48443985	5134.84929	3148.16439	2182.6638
7	ABJBZV	ABJ	BZV	335.209013	5.26139021	-3.92629004	-4.25169992	15.2530003	2378.29308	3148.16439	4824.08837
8	ABJCDG	ABJ	CDG	103.599104	5.26139021	-3.92629004	49.012798	2.55	4903.21269	3148.16439	1931.52004
9	ABJCKY	ABJ	CKY	500.987693	5.26139021	-3.92629004	9.57689	-13.612	1170.57562	3148.16439	2716.2754
10	ABJCMN	ABJ	CMN	100	5.26139021	-3.92629004	33.3675003	-7.58997011	3148.16439	3148.16439	0
11	ABJCOO	ABJ	COO	892.270738	5.26139021	-3.92629004	6.35723019	2.38435006	708.64758	3148.16439	3174.89061
12	ABJDSS	ABJ	DSS	305.721383	5.26139021	-3.92629004	14.67	-17.073333	1778.26949	3148.16439	2288.38569
13	ABJFIH	ABJ	FIH	332.597008	5.26139021	-3.92629004	-4.38575	15.4446	2403.94039	3148.16439	4847.26942
14	ABJIST	ABJ	IST	125.274906	5.26139021	-3.92629004	41.261297	28.741951	5150.75983	3148.16439	3304.44516
15	ABJJFK	ABJ	JFK	112.741745	5.26139021	-3.92629004	40.639801	-73.7789	7932.40349	3148.16439	5794.96574
16	ABJLBV	ABJ	LBV	458.352876	5.26139021	-3.92629004	0.45860001	9.41228008	1574.24095	3148.16439	4067.41431
17	ABJLFW	ABJ	LFW	1083.874	5.26139021	-3.92629004	6.16560984	1.25451005	581.960038	3148.16439	3159.54913
18	ABJLOS	ABJ	LOS	777.187449	5.26139021	-3.92629004	6.57737017	3.32116008	814.806191	3148.16439	3184.40707
19	ABJNIM	ABJ	NIM	491.612758	5.26139021	-3.92629004	13.4815	2.18361	1133.11323	3148.16439	2422.36484
20	ABJNSI	ABJ	NSI	405.754757	5.26139021	-3.92629004	3.72255993	11.5532999	1724.39045	3148.16439	3848.6319
21	ABJORY	ABJ	ORY	103.616058	5.26139021	-3.92629004	48.7233333	2.3794444	4869.44637	3148.16439	1897.36398
22	ABJOUA	ABJ	OUA	668.927296	5.26139021	-3.92629004	12.3532	-1.51242	831.929249	3148.16439	2416.83744
23	ABJOXB	ABJ	OXB	382.855768	5.26139021	-3.92629004	11.8948002	-15.6536999	1484.78895	3148.16439	2524.55745

Figure 14: Base de données contenant les Circuitrys et d'autres détails sur les vols

Les vols qui ont un Circuitry entre 110 et 130 sont les plus bénéficiaires pour la Royal Air Maroc pour les raisons suivantes :

- ✓ Ils offrent un bon compromis entre distance et temps de vol. Les vols avec un circuitry plus élevé auront une distance plus longue à parcourir, ce qui entraînera une augmentation de la consommation de carburant et des temps de vol. Les vols avec un circuitry plus faible auront une distance plus courte à parcourir, ce qui entraînera une diminution de la consommation de carburant et des temps de vol. Les vols avec un circuitry entre 110 et 130 offrent un bon compromis entre ces deux facteurs.
- ✓ Ils sont plus faciles à planifier et à exécuter. Les vols avec un circuitry plus élevé peuvent être plus difficiles à planifier, car les compagnies aériennes doivent tenir compte de la consommation de carburant et des temps de vol plus élevés. Les vols avec un circuitry plus faible peuvent être plus difficiles à exécuter, car les compagnies aériennes doivent éviter les zones de mauvais temps et autres obstacles. Les vols avec un circuitry entre 110 et 130 sont plus faciles à planifier et à exécuter, car ils sont moins susceptibles de rencontrer des problèmes de consommation de carburant, de

temps de vol ou de météo.

- ✓ Ils sont plus susceptibles de satisfaire les passagers. Les passagers préfèrent généralement les vols plus courts, car ils sont plus confortables et moins stressants. Les vols avec un circuitry plus élevé peuvent être plus longs, ce qui peut entraîner une diminution de la satisfaction des passagers. Les vols avec un circuitry plus faible peuvent être plus courts, mais ils peuvent également être moins efficaces en termes de consommation de carburant et de temps de vol. Les vols avec un circuitry entre 110 et 130 sont plus susceptibles de satisfaire les passagers, car ils sont suffisamment courts pour être confortables, mais suffisamment longs pour être efficaces.

⇒ **Le système de veille concurrentielle ne doit s'intéresser qu'aux vols qui ont un circuitry entre 120 et 130. En effet, ces vols sont les plus bénéficiaires pour les compagnies aériennes, car ils offrent un bon compromis entre distance, temps de vol et satisfaction des passagers.**

	Route	Dep	Arr	Circuitry
0	cmndoh	cmn	doh	100.000000
1	ABJACC	ABJ	ACC	1514.476811
2	ABJALG	ABJ	ALG	117.608663
3	ABJBKO	ABJ	BKO	593.130715
4	ABJBRU	ABJ	BRU	103.816644
...
1932	CMNTFS	CMN	TFS	100.000000
1933	BRUTNG	BRU	TNG	132.171723
1934	FEZORY	FEZ	ORY	122.396271
1935	DOHCMN	DOH	CMN	100.000000
1936	TFSCMN	TFS	CMN	100.000000
1937 rows × 4 columns				

Figure 15 : Extraction et regroupement des Circuitrys des vols

	Sem	Mkt	Al	Route	Somme de Ops	Dep	Arr	Circuitry
0	2023-07-03		TU	TUNDSS	1	TUN	DSS	108.080709
1	2023-07-10		TU	TUNDSS	2	TUN	DSS	108.080709
2	2023-07-17		TU	TUNDSS	2	TUN	DSS	108.080709
3	2023-07-24		TU	TUNDSS	2	TUN	DSS	108.080709
4	2023-07-31		TU	TUNDSS	2	TUN	DSS	108.080709
...
51796	2023-10-16		TOM	LGWDSS	1	LGW	DSS	100.598576
51797	2023-10-23		TOM	LGWDSS	1	LGW	DSS	100.598576
51798	2023-10-09		TOM	DSSLGW	1	DSS	LGW	100.598576
51799	2023-10-16		TOM	DSSLGW	1	DSS	LGW	100.598576
51800	2023-10-23		TOM	DSSLGW	1	DSS	LGW	100.598576

51801 rows × 7 columns

Figure 16: Ajout de la variable Circuitry dans la base de données

c. Introduction des vols dans le territoire Marocain

En effet, le système de veille concurrentielle doit aussi s'intéresser aux vols qui ont soit un départ ou une arrivée dans le territoire marocain. En effet, ces vols sont les plus importants pour la Royal Air Maroc, car ils représentent la majeure partie de son activité.

```
[ ] aeroport_mar = ["AGA" ,"CMN" ,"VIL" ,"ESU", "FEZ" ,"RAK" ,"NDR" ,"OZZ" ,"OUD" ,"RBA" ,"TNG"]
```

Figure 17 : Liste des aéroports Marocains

	Sem	Mkt	Al	Route	Somme de Ops	Dep	Arr	Circuitry	Agence	Route_base
739	2023-07-03		U2	AGAGVA	2	AGA	GVA	100.124659	U2	AGAGVA
740	2023-07-10		U2	AGAGVA	2	AGA	GVA	100.124659	U2	AGAGVA
741	2023-07-17		U2	AGAGVA	2	AGA	GVA	100.124659	U2	AGAGVA
742	2023-07-24		U2	AGAGVA	2	AGA	GVA	100.124659	U2	AGAGVA
743	2023-07-31		U2	AGAGVA	2	AGA	GVA	100.124659	U2	AGAGVA
...
51779	2023-09-18		4Y	RAKFRA	3	RAK	FRA	100.389127	4Y	RAKFRA
51780	2023-09-25		4Y	RAKFRA	3	RAK	FRA	100.389127	4Y	RAKFRA
51781	2023-10-02		4Y	RAKFRA	4	RAK	FRA	100.389127	4Y	RAKFRA
51782	2023-10-09		4Y	RAKFRA	7	RAK	FRA	100.389127	4Y	RAKFRA
51783	2023-10-16		4Y	RAKFRA	7	RAK	FRA	100.389127	4Y	RAKFRA
13284 rows × 9 columns										

Figure 18 : Aperçu de la base donnée après ajout des Circuitys des vols départent ou arrivent au Maroc

On obtient alors une base de données bien traitée, ne contenant que les vols qui départent ou arrivent au Maroc et dont le Circuitry est compris entre 110 et 130 ,

La base de donnée finale contient 13284 lignes pour 9 colonnes .

d. Encodage des données catégorielles

L'encodage des données catégorielles est une étape importante lors de la préparation des données pour l'entraînement d'un modèle d'apprentissage automatique. Les modèles d'apprentissage automatique ne peuvent pas traiter les données catégorielles directement. Il est donc nécessaire de convertir les données catégorielles en nombres avant de pouvoir les utiliser pour entraîner un modèle.

La classe `LabelEncoder` de la bibliothèque `sklearn.preprocessing` est utilisée pour encoder les données catégorielles. La méthode `fit_transform()` de la classe `LabelEncoder` est utilisée pour encoder les données catégorielles et renvoyer une liste de valeurs numériques.

Les lignes de code ci-dessus créent deux nouvelles colonnes nommées `Agence` et `Route_base` qui contiennent les valeurs des colonnes `Mkt Al` et `Route`, respectivement. Les lignes de code suivantes encodent les données catégorielles des colonnes `Mkt Al` et `Route`.

```
label_encoder = LabelEncoder()  
df["Mkt Al"] = label_encoder.fit_transform(df["Mkt Al"])  
df["Route"] = label_encoder.fit_transform(df["Route"])
```

Figure 19: Encodage des données catégorielles

6. Entraînement du modèle

L'étape d'entraînement d'un modèle de détection d'anomalie Isolation Forest consiste à entraîner un ensemble d'arbres de décision. Ces arbres sont construits en utilisant un processus aléatoire, ce qui permet au modèle de détecter des anomalies qui ne sont pas détectées par les méthodes de détection d'anomalie traditionnelles.

Dans le cas de notre modèle de détection d'anomalie Isolation Forest, les données à entraîner sont les données relatives aux vols qui ont soit un départ ou une arrivée dans le territoire marocain et dont le Circuitry est compris entre 110 et 130 . Ces données comprennent des informations, la date des vols, les fréquences des vols , les compagnies aériennes,.

Une fois l'entraînement terminé, le modèle est prêt à être utilisé pour détecter des anomalies dans de nouvelles données.

```

from sklearn.ensemble import IsolationForest
from sklearn.preprocessing import LabelEncoder

df["Agence"] = df["Mkt A1"]
df["Route_base"] = df["Route"]

label_encoder = LabelEncoder()
df["Mkt A1"] = label_encoder.fit_transform(df["Mkt A1"])
df["Route"] = label_encoder.fit_transform(df["Route"])

# Sélectionner les colonnes pertinentes
columns = ["Somme de Ops", "Mkt A1", "Route"]

# Créer et entraîner le modèle Isolation Forest Extended
model = IsolationForest(contamination=0.05, random_state=42, bootstrap=True)
model.fit(df[columns])

# Prédire les anomalies
df["Anomalies"] = model.predict(df[columns])

# Les valeurs de "Somme de Ops" considérées comme anomalies sont celles où "Anomalies" est égal à oui
anomalies = df[df["Anomalies"] == oui]

```

Figure 20: Entraînement du modèle

- L'argument `contamination` est un paramètre important qui peut avoir un impact significatif sur la précision du modèle. Un pourcentage de contamination trop faible peut entraîner un modèle qui est sous-appris, tandis qu'un pourcentage de contamination trop élevé peut entraîner un modèle qui est sur appris.
- L'argument `random_state` est un paramètre moins important, mais il peut être utile pour garantir la reproductibilité des résultats.
- L'argument `bootstrap` est un paramètre qui peut être utilisé pour améliorer la précision du modèle en réduisant le biais. Cependant, il peut également augmenter le temps d'entraînement du modèle.

7. Résultat finale

Après l'entraînement du modèle, nous avons terminé l'étape d'apprentissage. Nous pouvons maintenant passer à l'étape d'inférence, où nous appliquons notre modèle pour détecter les

anomalies dans la base de données du lundi 10 juillet 2023.

Pour le faire, nous allons charger la base de données du lundi 10 juillet 2023 et utiliser notre modèle pour prédire les anomalies. Les lignes qui sont prédites comme des anomalies sont m seront extraites et peuvent être analysées pour identifier les causes des anomalies.

	Sem	Mkt AI	Route	Somme de Ops	Dep	Arr	Circuitry	Agence	Route_base	Anomalies
795	2023-07-10	38	524	9	TUN	CMN	100.00000	TU	TUNCMN	oui
796	2023-07-17	38	524	9	TUN	CMN	100.00000	TU	TUNCMN	oui
797	2023-07-24	38	524	9	TUN	CMN	100.00000	TU	TUNCMN	oui
798	2023-07-31	38	524	9	TUN	CMN	100.00000	TU	TUNCMN	oui
799	2023-08-07	38	524	9	TUN	CMN	100.00000	TU	TUNCMN	oui
...
51364	2023-10-02	38	4	4	ABJ	TUN	126.74014	TU	ABJTUN	oui
51365	2023-10-09	38	4	4	ABJ	TUN	126.74014	TU	ABJTUN	oui
51366	2023-10-16	38	4	4	ABJ	TUN	126.74014	TU	ABJTUN	oui
51367	2023-10-23	38	4	4	ABJ	TUN	126.74014	TU	ABJTUN	oui
51368	2023-10-30	38	4	1	ABJ	TUN	126.74014	TU	ABJTUN	oui

687 rows x 10 columns

Figure 21: Détection des anomalies

Notre modèle a détecté 687 anomalies dans 13448 lignes présentes dans la base de données du lundi 10 juillet 2023

III. Visualisation des anomalies :

La visualisation des résultats d'une analyse de données peut se faire de deux manières principales :

Affichage du Dataframe : c'est la manière la plus simple de visualiser les données. Elle consiste à afficher le Dataframe sous forme d'une table, avec les colonnes représentant les différentes variables et les lignes représentant les différentes observations.

Nuage de points : c'est une manière plus visuelle de visualiser les données. Elle consiste à

représenter chaque observation par un point dans un plan, avec les coordonnées du point représentant les valeurs des deux variables que l'on souhaite visualiser.

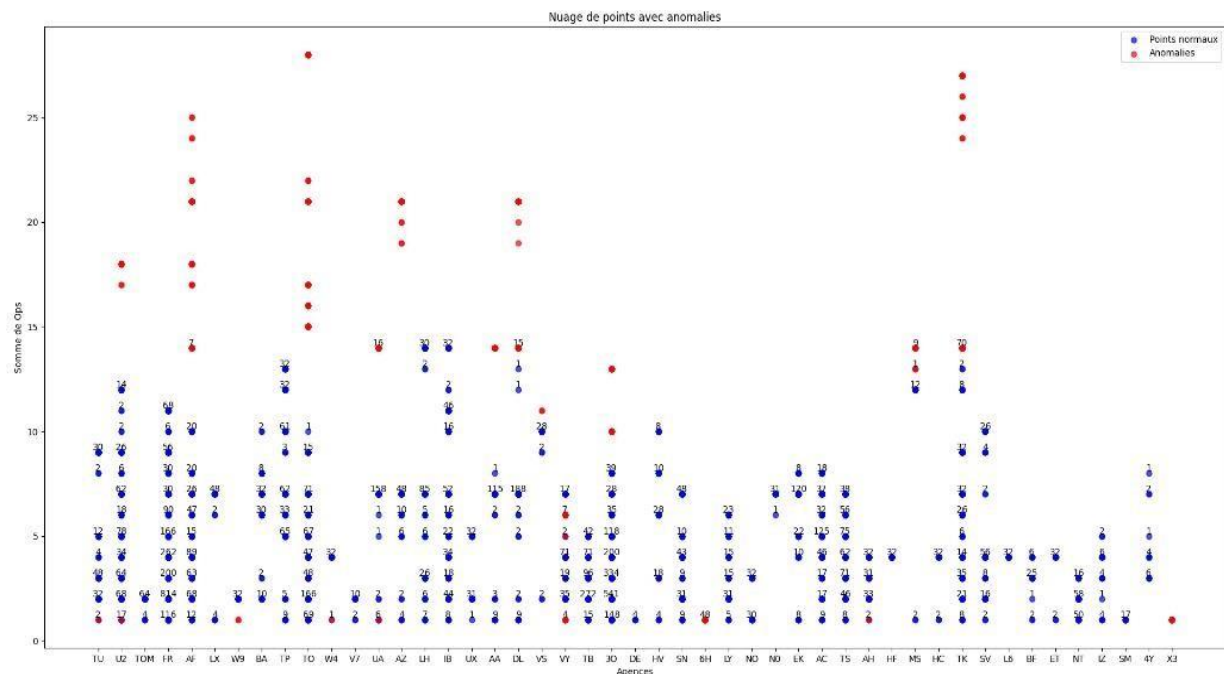


Figure 22: Nuage de points des anomalies (Fréquences des vols par compagnie)

Les points rouges du nuage indiquent les destinations pour lesquelles le nombre de vols est anormalement élevé ou bas. Les points bleus d'une autre part présentent les destinations considérées normales.

Les nombres sur chaque point (rouge ou bleu) présentent le nombre de destinations qui présentent (une anomalie ou un cas normal) et non pas la fréquence des vols , prenons un exemple pour mieux comprendre :

Le cas de la compagnie aérienne Tunisair (TU) :

Les deux destinations qui représentent des anomalies sont :ABJTUN [Abidjan, Côte d'Ivoire (ABJ) - Tunis, Tunisie (TUN)] et [Tunis, Tunisie (TUN) – Casablanca ,Maroc (CMN)] : Ces deux destination sont alors marquées sur le point en rouge par le nombre 2 .



Figure 23: Nuage de points de la compagnie Tunisair

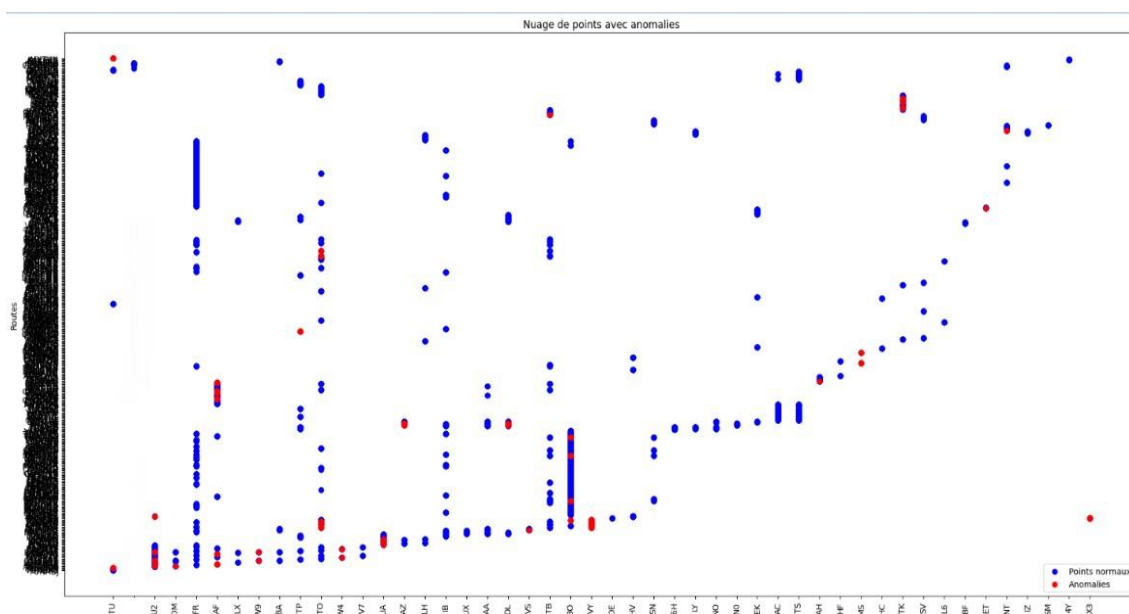


Figure 24: Nuage de points des anomalies (Routes par compagnie)

Le nuage de points des Routes (Destinations ; Départ-Arrivé) en fonction des compagnies aériennes est illisible à cause des noms des Routes qui se chevauchent dues à leurs grand nombre .

La meilleure solution pour visualiser les anomalies, savoir de quel vol parviennent telles et quelle compagnie aérienne en est responsable est de premièrement lire le nuage de point (Somme de Ops en fonction des compagnies aériennes) ensuite passer au fichier Excel qui englobe tous les vols considérés comme anomalies et les détails sur chacun deux .

B	C	D	E	F	G	H	I
Date vol	Frequence		Dep	Arr	Circuitry	Agence	Anomalies
7/10/2023	9		TUN	CMN	100	TU	oui
7/17/2023	9		TUN	CMN	100	TU	oui
7/24/2023	9		TUN	CMN	100	TU	oui
7/31/2023	9		TUN	CMN	100	TU	oui
8/7/2023	9		TUN	CMN	100	TU	oui
8/14/2023	9		TUN	CMN	100	TU	oui
8/21/2023	9		TUN	CMN	100	TU	oui
8/28/2023	9		TUN	CMN	100	TU	oui
9/4/2023	9		TUN	CMN	100	TU	oui
9/11/2023	9		TUN	CMN	100	TU	oui
9/18/2023	9		TUN	CMN	100	TU	oui
9/25/2023	9		TUN	CMN	100	TU	oui
10/2/2023	9		TUN	CMN	100	TU	oui
10/9/2023	9		TUN	CMN	100	TU	oui
9/15/2023	9		TUN	CMN	100	TU	oui
7/10/2023	14		TUN	CMN	100	AT	oui
7/17/2023	14		TUN	CMN	100	AT	oui
7/24/2023	14		TUN	CMN	100	AT	oui
7/31/2023	14		TUN	CMN	100	AT	oui

Figure 25: Fichier Excel contenant les vols qui présentent des anomalies

Conclusion

Ce projet a permis de développer un système de veille concurrentielle pour la Royal Air Maroc. Ce système utilise un modèle de détection d'anomalies basé sur l'algorithme d'apprentissage automatique Isolation Forest.

Le modèle a été formé sur une base de données des plans de vol des compagnies aériennes concurrentes. Cette base de données est construite à partir des fichiers SSIM partagés chaque lundi par les compagnies aériennes.

Le modèle a été évalué sur une base de données de test. Les résultats ont montré que le modèle est capable de détecter avec précision les anomalies dans les plans de vol des compagnies aériennes concurrentes.

Ce système de veille concurrentielle peut être utilisé par la Royal Air Maroc pour suivre les changements dans les plans de vol des compagnies aériennes concurrentes. Cela permet à la Royal Air Maroc d'identifier les opportunités pour améliorer ses propres opérations et de se protéger de la concurrence.