

2024

PROJET PREDICTION ET OPTIMISATION DE LA MAINTENANCE AERONAUTIQUE



Andréa, Guillaume, Patrick & Ulrich

TOP GUN

06/06/2024

PLAN DE PROJET TOP GUN

NOM DU PROJET	CHEF DE PROJET
TOP GUN	ANDREA

RESUME DU PROJET

PREDICTION ET OPTIMISATION DE LA MAINTENANCE AERONAUTIQUE

OBJECTIFS DU PROJET

Nous travaillons pour une compagnie aérienne qui cherche à optimiser ses opérations de maintenance. Notre mission est d'analyser les données de vol et de maintenance pour prédire les besoins de maintenance futurs et leurs coûts associés.

PORTEE DES TRAVAUX

- **Définition des rôles et des tâches**
- **Etude de l'intégration des données :**
 - Rassembler les datasets aéronefs, composants, logs des vols et dégradations.
 - Choix d'une infrastructure centralisée pour le stockage des données
 - Définir un process pour l'intégration des données logs et dégradation mis à jour chaque jour dans une infrastructure centralisée.
- **Préparation de l'automatisation des collectes des données et de leur nettoyage :**
 - Mise en place du système d'automatisation
 - Nettoyer les données pour éliminer les erreurs, les valeurs manquantes et les duplicatas existantes.
 - Normaliser et transformer les données pour les rendre compatibles avec les modèles de machine Learning.
- **Analyse exploratoire des données :**
 - Effectuer des analyses pour comprendre les caractéristiques des données, identifier les tendances, les motifs et les anomalies.
 - Visualiser les données pour aider à la compréhension.
 - Calcul des coûts prenant en compte le coût des réparations (coût des composants nécessaires) et de l'immobilisation
 - Calculer à partir de quel taux d'usure le voyant s'affiche
- **Développement de modèles prédictifs :**
 - Sélectionner et tester différents algorithmes de machine learning
 - Entraîner les modèles en utilisant des ensembles de données historiques.
 - Valider et évaluer les performances des modèles à l'aide de métriques appropriées.
- **Développement d'un dashboard :**
 - Connecter le dashboard aux bases de données
 - Assurer la mise à jour en temps réel des données affichées
 - Développer des graphiques interactifs pour afficher les KPI (et les prédictions ?).
 - Utiliser des diagrammes et d'autres éléments visuels pertinents.
- **Développement de l'application :** Concevoir une interface utilisateur intuitive sur streamlit
- **Déploiement des modèles dans l'application :**
 - Intégrer les modèles prédictifs et d'optimisation dans l'application

	<ul style="list-style-type: none"> - Mettre en place des pipelines automatisés pour l'ingestion des nouvelles données et la mise à jour des modèles. <ul style="list-style-type: none"> • Surveillance et mise à jour des modèles : <ul style="list-style-type: none"> - Tester et surveiller les performances des modèles en production pour détecter les anomalies - Mettre à jour les modèles en fonction des nouvelles données suite à d'éventuelles corrections d'anomalies constatées. • Documentation : Documenter tous les aspects du projet, y compris les méthodologies, les modèles, les processus de données et les résultats.
PRINCIPAUX LIVRABLES	<ul style="list-style-type: none"> • Prédire l'état des voyants d'avertissement avec des données de vol actualisées. • Estimer les économies réalisées en anticipant les maintenances.
INDICATEURS DE CLES DE PERFORMANCES PROJET EQUIPE	<ul style="list-style-type: none"> • Précision du modèle et des prédictions • Réduction des temps d'immobilisation et des coûts de maintenance évités <p><u>ANDREA : CHEF DE PROJET</u></p> <ul style="list-style-type: none"> • Planifie : élabore un plan détaillé, établir un calendrier, et déterminer les ressources nécessaires. • Coordination des équipes : Collaborer avec le data scientists, d'analystes, de développeurs pour assurer une bonne communication et une coopération efficace. • Suivi et contrôle : Surveille l'avancement du projet, assurer le respect des délais, et ajuster les plans en cas de déviations. • Qualité des données : S'assurer que les données collectées sont précises et complètes. • Support des services : Offre une aide directe aux membres de l'équipes pour résoudre des problèmes techniques et pour éviter la surcharge de travail selon le planning • Documentation : Documenter les processus, les méthodologies et les résultats pour faciliter la compréhension et la reproductibilité des projets futurs. • Gestion des risques : Identifier les risques potentiels, évaluer leur impact et mettre en place des stratégies de mitigation. • Gestion du git : merge et vérifier en fonctions des tâches les codes <p><u>GUILLAUME & PATRICK : DATA ENGINEER & DEVELOPPEUR</u></p> <ul style="list-style-type: none"> • Conception et maintenance des architectures de données : Concevoir, construire le stockage des données sur une base SQL évolutives, incluant des bases de données, et des fonctions de traitement de données. • Collecte et ingestion de données : Développer des processus pour collecter et ingérer des données de manière efficace et fiable. • Nettoyage et transformation des données : Nettoyer, transformer et préparer les données pour les rendre exploitables par le data scientists. Cela inclut le traitement des données manquantes, l'élimination des duplicatas (lié à la coquille de l'aéronefs en doublon) et la normalisation des formats de données. • Automatisation des workflows : Automatiser les workflows de traitement de données pour assurer la continuité et l'efficacité des opérations de données, y compris l'automatisation des tâches répétitives, la mise en place de systèmes de surveillance pour détecter et corriger les erreurs, le transfert de la base SQL en python pour alimenter le machine learning.

	<ul style="list-style-type: none"> • Conception sur streamlit de l'application • Documentation : Documenter les processus, les architectures et les flux de données pour garantir une bonne compréhension et faciliter la maintenance et l'évolution des systèmes de données.
	<u>ULRICH: DATA SCIENTIST & DATA ANALYST</u>
	<ul style="list-style-type: none"> • Analyse exploratoire des données : Effectuer une analyse exploratoire des données (EDA) pour identifier des motifs, des tendances et des anomalies. Cette étape permet de mieux comprendre les caractéristiques des données et d'informer la sélection des modèles. • Modélisation prédictive : Développer des modèles prédictifs en utilisant des techniques de machine learning et de statistiques pour anticiper les pannes et les besoins de maintenance. Cela peut inclure des algorithmes de régression, des arbres de décision, des réseaux de neurones, des forêts aléatoires, etc. • Validation et évaluation des modèles : Évaluer les performances des modèles prédictifs en utilisant des métriques appropriées (comme la précision, le rappel, la F1-score et la matrice de confusion). Valider les modèles avec des données de test et s'assurer qu'ils généralisent bien à des données nouvelles. • Déploiement des modèles : Mettre en production les modèles prédictifs et d'optimisation dans les systèmes de maintenance existants, en collaborant avec les équipes d'ingénieurs en données et les développeurs pour assurer une intégration fluide. • Surveillance et maintenance des modèles : Surveiller les performances des modèles déployés et effectuer des mises à jour régulières pour s'assurer qu'ils restent précis et efficaces au fil du temps. Adapter les modèles en fonction des nouvelles données.

CHRONOLOGIE DU PROJET

SEMAINE 1		
CHEF DE PROJET (Andréa)	SERVICE DATA ENGINEER (Guillaume et Patrick)	SERVICE DATA SCIENTIST (Ulrich)
<ul style="list-style-type: none"> - Définir les rôles et le plan pour attendre les objectifs attendus - Rédaction du plan de projet et mise en place des outils des organisations - Rédaction bibliothèque base de données - Support auprès du service DATA SCIENTIST 	<ul style="list-style-type: none"> - <u>Guillaume & Patrick</u> : Nettoyage de la base aeronef - <u>Patrick</u> : Préparation de l'automatisation - <u>Guillaume</u> : Test de la fonction de la collecte pour intégrer les données sur la future base SQL 	<ul style="list-style-type: none"> - Etude de la base des données : création visu de la base donnée sur dbdiagram - Première analyses des datasets

PROBLEMATIQUES

Choix base de données : SQL car moins lourd que le csv et interface mieux adaptés pour les bases de données lourde (plus rapide, simplification pour mise à jour d'un dashboard)

Stockage des données ; incrémentation des nouvelles bases de données

Voir pour le machine learning passer de SQL à python pour lancer traitement modèle

Calculer à partir de quel taux d'usure le voyant s'affiche

Dacttb : librairie pour transfert de sql à python

Il y a tous les jours 10530 lignes qui sont mises à jour ou pas si il y a eu un vol : A analyser puis nettoyer si besoin