

PRÉSENTATION DE MON PARCOURS



RÉSUMÉ COMPLET

- **Historique et Contexte du Projet**

Mon parcours en science des données a été marqué par une série d'exercices intensifs réalisés dans un environnement virtuel. L'objectif principal de ce parcours était de me former de manière approfondie aux différentes étapes du cycle de vie des données, en mettant l'accent sur la collecte, le traitement, l'analyse et la visualisation des données.

La science des données est devenue essentielle dans de nombreux domaines, offrant des perspectives nouvelles et profondes sur les problèmes complexes grâce à l'utilisation efficace des données. Pour moi, ce projet représentait une opportunité d'acquérir des compétences pratiques et techniques nécessaires pour devenir un data scientist accompli.



RÉSUMÉ COMPLET

- **Problèmes à Résoudre**

Au cours de ce parcours, j'ai abordé plusieurs défis spécifiques :

Collecter des données structurées et non structurées à partir de sources variées comme des APIs (par exemple, SpaceX) et le web (web scraping).

Nettoyer et prétraiter les données pour les rendre utilisables dans des analyses ultérieures.

Explorer et visualiser les données pour extraire des insights significatifs.

Construire et évaluer des modèles de machine learning pour résoudre des problèmes de classification et de prédiction.



MÉTHODOLOGIES

- **Collecte de Données - API SpaceX**

Pour la collecte de données à partir de l'API SpaceX, j'ai suivi un processus structuré :

- **Organigrammes et Phrases Clés** : J'ai utilisé des organigrammes pour illustrer le flux de travail, depuis la requête d'API jusqu'au stockage des données récupérées. Les phrases clés ont été utilisées pour décrire chaque étape critique, telles que l'authentification, la gestion des requêtes et le traitement des réponses.

- **Collecte de Données - Web Scraping**

Pour compléter mes données, j'ai également exploré l'utilisation du web scraping :

- **Organigrammes et Phrases Clés** : J'ai détaillé le processus de web scraping à l'aide d'organigrammes, mettant en lumière les outils utilisés (comme BeautifulSoup) et les techniques employées pour extraire des données structurées à partir de sources web variées.

MÉTHODOLOGIES

- **Méthodologie de Traitement des Données**

Le traitement des données a été une étape cruciale pour assurer la qualité et la cohérence des données collectées :

- **Organigrammes et Phrases Clés** : J'ai décrit comment j'ai nettoyé et prétraité les données pour éliminer les valeurs manquantes, gérer les doublons, normaliser les données, etc. Les organigrammes ont mis en lumière les étapes spécifiques du processus de nettoyage et de transformation des données.
- **EDA avec Visualisation des Données**

L'Exploratory Data Analysis (EDA) était essentielle pour comprendre les caractéristiques et les tendances des données collectées :

- **Résumé des Graphiques** : J'ai utilisé une variété de graphiques (histogrammes, nuages de points, diagrammes en barres, etc.) pour explorer et visualiser les données. Chaque type de graphique a été sélectionné en fonction de son aptitude à mettre en lumière des insights spécifiques sur les données.

MÉTHODOLOGIES

- **EDA avec SQL**

J'ai également exploré l'utilisation de SQL pour effectuer des analyses et extraire des informations spécifiques des données :

- **Résumé des Requêtes SQL :** J'ai récapitulé les requêtes SQL utilisées pour explorer les relations entre les différentes entités de données, calculer des agrégations et obtenir des résumés statistiques.

- **Analyse Visuelle Interactive**

Pour rendre l'analyse des données plus dynamique et interactive, j'ai utilisé des outils comme Folium pour la cartographie et Plotly Dash pour les applications interactives :

- **Résumé des Outils Utilisés :** J'ai expliqué comment j'ai utilisé Folium pour créer des cartes interactives, ainsi que l'utilisation de Plotly Dash pour développer des applications interactives permettant de visualiser les données de manière dynamique.

MÉTHODOLOGIES

- **Modèles de Classification**

Enfin, j'ai exploré la construction et l'évaluation de plusieurs modèles de classification pour résoudre des problèmes spécifiques :

- **Description des Modèles :** J'ai décrit les étapes de construction des modèles KNN, Régression Logistique, Arbre de Décision et SVM. J'ai également discuté des métriques de performance utilisées pour évaluer et comparer ces modèles.

RÉSULTATS

- **Analyses Spécifiques Demandées**

J'ai produit une série d'analyses spécifiques pour répondre aux objectifs du projet :

- **Diagrammes de Dispersion** : J'ai utilisé des diagrammes de dispersion pour visualiser les relations entre différentes variables, comme le nombre de vols par rapport au site de lancement et la charge utile par rapport au résultat du lancement.
- **Requêtes SQL** : J'ai réalisé diverses requêtes SQL pour extraire des informations précises, telles que la masse totale de la charge utile transportée ou le nombre de missions réussies et d'échecs.
- **Cartes Interactives** : À l'aide de Folium, j'ai cartographié les sites de lancement et les résultats de lancement par site, offrant ainsi une vue géographique des opérations de SpaceX.

RÉSULTATS

- **Performances des Modèles**

J'ai évalué les performances de plusieurs modèles de classification :

- **KNN, Régression Logistique, Arbre de Décision, SVM** : Chaque modèle a été évalué en termes de précision, rappel et F-mesure. Le modèle ayant obtenu les meilleures performances a été identifié et discuté en détail.
- **Matrice de Confusion** : J'ai présenté et expliqué la matrice de confusion du meilleur modèle, démontrant la capacité du modèle à prédire avec précision les classes d'intérêt.

RÉSULTATS

-
- **Conclusions et Discussions**
 - **Synthèse des Conclusions :** J'ai tiré des conclusions importantes de mon analyse, mettant en lumière les insights clés obtenus tout au long du parcours.
 - **Recommandations :** J'ai formulé des recommandations basées sur les résultats obtenus, visant à améliorer les performances futures des modèles ou à explorer davantage certains aspects des données.

PRÉSENTATION

Ajouts et Améliorations : Pour améliorer la présentation des résultats, j'ai intégré des animations, des graphiques supplémentaires et ajusté le style de police pour une meilleure lisibilité et compréhension.

TÂCHES SUPPLÉMENTAIRES

- **Tâches Complémentaires d'EDA** : Au-delà des exigences de base, j'ai exploré des analyses exploratoires supplémentaires pour approfondir ma compréhension des données.
- **Améliorations des Modèles** : J'ai discuté des techniques avancées utilisées pour améliorer et évaluer les modèles de classification, en explorant notamment des stratégies de sélection de variables et de réglage des hyperparamètres.
- **Composants Additionnels du Tableau de Bord** : J'ai enrichi mon tableau de bord interactif avec des fonctionnalités supplémentaires pour offrir une expérience utilisateur enrichie et intuitive.