

**CERTIFICATION PROFESSIONNELLE EXPERT EN INFORMATIQUE ET SYSTEME
D'INFORMATION RNCP - RNCP N°35584**

BLOC E7.3 – Piloter l'informatique décisionnel d'un S.I
(Big Data & Business Intelligence)

Cahier des Charges de la MSPR « Big Data et Analyse de données »

COMPÉTENCES ÉVALUÉES :

- Collecter les besoins en données des directions métiers de l'entreprise afin d'avoir une vision structurée de l'ensemble des données du système d'information et partager la stratégie Data globale avec le comité de direction.
- Définir une architecture business intelligence à partir des orientations stratégiques arrêtées avec le comité de direction afin de mettre à disposition des utilisateurs métiers les données structurées d'un S.I.
- Définir une stratégie big data (de la collecte aux traitements des données) à partir des orientations stratégiques arrêtées avec le comité de direction afin d'aider l'entreprise à mieux comprendre ses clients et à créer de nouveaux services.
- Proposer des modèles statistiques et de data science (machine learning) à mettre en pratique aux directions métiers afin de détecter des nouveaux services, anticiper des besoins et résoudre des problématiques métiers de l'entreprise.
- Organiser les sources de données sous forme de résultats exploitables (data visualisation) pour alimenter les outils décisionnels et visualiser les résultats de façon compréhensible permettant d'aider les directions métiers à la prise de décision.
- Définir les données de référence de l'entreprise à partir des données utilisées pour créer un référentiel de données afin d'assurer la mise à disposition de données cohérentes aux directions métiers.
- Créer un entrepôt unique à partir du référentiel de données établi pour centraliser les informations stratégiques de l'entreprise et répondre rapidement aux besoins métiers.
- Assurer la qualité des données en utilisant les outils de gestion de la qualité de données pour garantir l'exactitude, la cohérence, la synchronisation et la traçabilité des données afin de satisfaire les besoins d'accessibilité des utilisateurs métiers.
- Appliquer les procédures de sécurité établies par le / la RSSI de l'entreprise afin d'assurer la confidentialité et la sécurité des données et garantir une mise en conformité avec les obligations légales du RGPD.

PHASE 1 : PRÉPARATION DE CETTE MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE

- Durée de préparation :
 - 25 heures
- Mise en œuvre :
 - Travail d'équipe constituée de 4 apprenants-candidats (5 maximum si groupe impair)
- Résultat attendu :
 - Produire un dossier de synthèse répondant aux différentes attentes de l'entreprise Elexxion.

PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF

- **Durée totale par groupe** : 50 mn se décomposant comme suit :
 - 20 mn de soutenance orale par l'équipe.
 - 30 mn d'entretien collectif avec le jury (questionnement complémentaire).
 - Objectif : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Jury d'évaluation** : 2 personnes (binôme d'évaluateurs) par jury – Ces évaluateurs ne sont pas intervenus durant la période de formation et ne connaissent pas les apprenants à évaluer.

I - PRÉSENTATION DE L'ENTREPRISE ET DE SON ACTIVITÉ

Préambule : L'entreprise choisie pour cette MSPR est fictive, les prénoms sont fictifs, toute ressemblance à un cas réel serait purement fortuite.

Jean-Edouard de la Motte Rouge a créé une start-up spécialisée dans le conseil sur la thématique des campagnes électorales.

La start-up comprend un.e expert.e en analyse politique, un.e business developer, et un.e assistant.e.

Il souhaite pouvoir prédire, grâce à l'intelligence artificielle, les tendances des élections à venir, en se basant sur un certain nombre d'indicateurs, comme la sécurité, l'emploi, la vie associative, la population, la vie économique (nombre d'entreprises), la pauvreté...

Vous pouvez, si cela vous paraît pertinent, ajouter des critères.

Cela lui donnerait un avantage concurrentiel important dans son activité.

Avant d'investir dans une infrastructure et une politique de recherche et développement en ce sens, et dans l'idée de demander des aides à l'innovation, il fait appel à votre groupe pour établir une preuve de concept (POC).

La POC devra être conçue pour un secteur géographique restreint (ville, arrondissement, circonscription, département...) et unique.

Afin de répondre à sa commande, il conviendra donc de :

1. Sélectionner un secteur géographique sur lequel sera établie la POC
2. Sélectionner différents jeux de données parmi ceux disponibles ci-dessous et/ou de votre choix, afin de rechercher des corrélations entre résultats des élections passées et contexte
3. Etablir un petit nombre de visualisations graphiques montrant les jeux de données choisis pour leur donner de la lisibilité
4. Etablir un modèle prédictif supervisé, en découpant les données à disposition en un ou plusieurs jeux d'entraînement et jeux de tests afin de vérifier la fiabilité de votre modèle
5. Proposer une visualisation graphique de votre modèle et des prédictions à 1 an, 2 ans et 3 ans.

Jeux de données disponibles

- <https://www.data.gouv.fr/fr/pages/donnees-des-elections/>
- <https://www.data.gouv.fr/fr/pages/donnees-securite/>
- https://www.data.gouv.fr/fr/pages/donnees_emploi/
- https://www.data.gouv.fr/fr/organizations/institut-national-de-la-statistique-et-des-etudes-economiques-insee/?datasets_page=7#organization-datasets

II- MÉTHODOLOGIE

- Proposer un schéma de traitement des données (flux de données)
- Utiliser un outil de normalisation des données pour le mettre en oeuvre (ETL)
- Définir une architecture de données adaptée aux traitements que vous ferez ensuite, en nommant spécifiquement et de manière pertinente vos tables et vos champs
- Ouvrir les données avec un langage ou un outil adapté (Python - Pandas, R...)
- A partir des données passées qui constitueront un jeu d'entraînement (apprentissage supervisé), modèle prédictif
- Proposez des visualisations claires, en utilisant soit un logiciel (PowerBI), soit une librairie python (Matplotlib)

III- EXEMPLE D'INDICATEURS D'ANALYSE

- Parmi les données que vous avez sélectionnées, laquelle est la plus corrélée aux résultats des élections?
- Définissez le principe d'un apprentissage supervisé
- Comment définissez-vous le degré de précision (accuracy) de votre modèle?

IV- LIVRABLES

- Dossier contenant :
 - Justification du choix de la zone géographique
 - Choix des critères, justification
 - La démarche suivie et les méthodes employées
 - Un Modèle Conceptuel de Données
 - Les modèles testés
 - Les résultats du modèle choisi
 - Les visualisations
 - Accuracy (pouvoir prédictif du modèle)
 - Les réponses aux questions posées dans les exemples d'indicateurs d'analyse ci-dessus (vous pouvez également ajouter d'autres indicateurs pertinents d'analyse).
- Un jeu de données nettoyé, normalisé et optimisé
 - Le choix du format est libre (SQL ou NoSQL)
- Votre code, propre et commenté

Les compétences évaluées durant cette MSPR :

- Collecter les besoins en données des directions métiers de l'entreprise afin d'avoir une vision structurée de l'ensemble des données du système d'information et partager la stratégie Data globale avec le comité de direction.
- Définir une architecture business intelligence à partir des orientations stratégiques arrêtées avec le comité de direction afin de mettre à disposition des utilisateurs métiers les données structurées d'un S.I.
- Définir une stratégie big data (de la collecte aux traitements des données) à partir des orientations stratégiques arrêtées avec le comité de direction afin d'aider l'entreprise à mieux comprendre ses clients et à créer de nouveaux services.
- Proposer des modèles statistiques et de data science (machine learning) à mettre en pratique aux directions métiers afin de détecter des nouveaux services, anticiper des besoins et résoudre des problématiques métiers de l'entreprise.
- Organiser les sources de données sous forme de résultats exploitables (data visualisation) pour alimenter les outils décisionnels et visualiser les résultats de façon compréhensible permettant d'aider les directions métiers à la prise de décision.
- Définir les données de référence de l'entreprise à partir des données utilisées pour créer un référentiel de données afin d'assurer la mise à disposition de données cohérentes aux directions métiers.
- Créer un entrepôt unique à partir du référentiel de données établi pour centraliser les informations stratégiques de l'entreprise et répondre rapidement aux besoins métiers.
- Assurer la qualité des données en utilisant les outils de gestion de la qualité de données pour garantir l'exactitude, la cohérence, la synchronisation et la traçabilité des données afin de satisfaire les besoins d'accessibilité des utilisateurs métiers.
- Appliquer les procédures de sécurité établies par le / la RSSI de l'entreprise afin d'assurer la confidentialité et la sécurité des données et garantir une mise en conformité avec les obligations légales du RGPD.