

**CERTIFICATION PROFESSIONNELLE EXPERT EN INFORMATIQUE ET SYSTEME
D'INFORMATION RNCP35584****BLOC 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)

PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF

- **Objectif** : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Moyen** : L'équipe utilise un support de présentation
- **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).
- **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.

Electio-Analytics est une start-up créée en 2017 par Jean-Édouard de la Motte-Rouge. Elle se spécialise dans le conseil stratégique dédié aux campagnes électorales. Son ambition est d'apporter à ses clients – candidats, partis politiques et cabinets de conseil – des analyses prospectives précises grâce à l'exploitation de données publiques et privées.

L'équipe se compose d'un expert en analyse politique (doctorat en science politique et expérience terrain sur plusieurs scrutins nationaux), d'un business developer (master en marketing digital, réseau de contacts auprès de partis et d'organismes de financement) et d'une assistante (formation en gestion administrative). Le business developer s'occupe de la prospection, de l'élaboration d'offres et de la recherche de subventions, tandis que l'assistante assure la coordination du projet et le suivi administratif.

Le besoin stratégique d'Electio-Analytics est de mettre en place une capacité de prévision des tendances électorales à moyen terme (un à trois ans) 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.

II - CAHIER DES CHARGES

Avant d'engager des investissements lourds, la société veut valider cette approche par une preuve de concept (POC). Elle fait appel une société de service, pour laquelle vous travaillez, pour établir ce POC.

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

Le projet doit d'abord définir un périmètre géographique unique – une commune, un arrondissement ou une circonscription législative – afin de limiter la volumétrie des données et d'assurer la traçabilité des résultats.

Ensuite, il faut collecter les jeux de données publics disponibles : résultats électoraux historiques, indicateurs de sécurité, d'emploi, de démographie, d'activité économique, de pauvreté, etc. Il est possible d'enrichir le corpus avec des enquêtes d'opinion, des flux de réseaux sociaux ou les dépenses publiques locales.

Les données seront nettoyées, normalisées et stockées dans une base structurée. Un pipeline ETL automatisé (extraction, transformation, chargement) devra être mis en place, avec un schéma de base de données clairement nommé.

Une phase d'analyse exploratoire devra produire des visualisations descriptives (cartes, histogrammes, heat-maps) afin d'identifier les corrélations potentielles entre chaque indicateur et les résultats électoraux passés.

Le cœur du projet consiste à construire 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.

Les prédictions devront être présentées sous forme de graphiques illustrant les scénarios à un, deux et trois ans (courbes temporelles, cartes de chaleur, diagrammes de probabilité).

Enfin, toute la démarche, les résultats et les recommandations seront consignés dans un rapport complet, conformément aux livrables décrits ci-dessus.

III - BESOINS EXPRIMÉS PAR ELECTIO-ANALYTICS

L'entreprise attend une justification claire du choix géographique, basée sur la disponibilité des données, la représentativité et la taille exploitable. Elle souhaite également que les indicateurs retenus soient pertinents et démontrent une forte corrélation avec les résultats électoraux (sécurité, emploi, pauvreté, vie associative, densité de population, nombre d'entreprises, etc.).

Un pipeline de traitement automatisé, fiable et traçable est indispensable pour garantir la qualité des données et la reproductibilité du processus.

Les visualisations doivent être claires, accessibles à des non-techniciens et permettre une prise de décision stratégique rapide.

Une documentation exhaustive (rapport, schéma conceptuel de données, code commenté, jeu de données nettoyé) est requise pour répondre aux exigences d'un éventuel appel à financement et assurer la transférabilité du projet.

Enfin, les livrables doivent être fournis dans des formats compatibles avec les outils internes (SQL, CSV, notebooks Jupyter, PowerBI) pour faciliter le passage du POC à la phase de production.

Afin de vous guider dans votre démarche, Electio-Analytics vous prodigue les conseils suivants :

- 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)

Vous trouverez ci-dessous des exemples 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

À l'issue de la mission, Electio-Analytics attend une série de livrables clairement identifiés, qui permettront à la fois de valider la pertinence du travail réalisé et de préparer l'intégration de la preuve de concept dans l'écosystème global de la start-up.

Ces livrables doivent être produits selon une logique industrielle: ils seront entièrement documentés, reproductibles et immédiatement exploitables par les différentes parties prenantes (développeurs, data-scientists, chefs de produit et responsables de la stratégie électorale). Ainsi, chaque artefact fourni servira de référence fiable pour la continuité du projet, la mise à l'échelle ultérieure et la justification des demandes de financement ou d'aides à l'innovation.

1. Un dossier de synthèse répondant aux différentes attentes de l'entreprise Electio-Analytics, à savoir :

- 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).

2. Un jeu de données nettoyé, normalisé et optimisé

Le choix du format est libre : SQL ou NoSQL

3. Un code, propre et commenté

4. Support de soutenance

En complément des livrables, l'équipe projet devra préparer un support de présentation destiné à la soutenance finale devant le client (public technique). Ce support devra synthétiser les principaux éléments du travail réalisé.

Il est important de souligner que l'évaluation de cette MSPR repose sur la combinaison des trois éléments suivants :

- la qualité du travail réalisé au cours du projet,
- la pertinence et l'exhaustivité des livrables remis
- et la capacité de l'équipe à présenter, justifier et valoriser ce travail lors de la soutenance orale.

Les équipes devront donc s'assurer que la soutenance reflète bien **l'ensemble des compétences attendues** (cf. page 1), en démontrant à la fois la maîtrise technique et la capacité à communiquer efficacement auprès d'un client professionnel.

IV- RESSOURCES FOURNIES

Les jeux de données disponibles sont (liste non exhaustive) :

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

Assistance et périmètre

Dans le cadre de ce projet pédagogique, l'équipe projet n'aura aucun contact direct avec Electio-Analytics. Le cahier des charges constitue la seule expression officielle du besoin. Toute demande de clarification devra être traitée avec l'encadrant pédagogique, jouant le rôle du client.