

**Bloc 1 : Collecter des données ou des informations.**

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 1 : Mise en œuvre de processus d'exploration des données et de collecte d'information :</b></p> <ul style="list-style-type: none"> <li>- Analyse du patrimoine informationnel de données de l'entreprise et du système existant</li> <li>- Extraction et stockage de données dans le respect de la réglementation</li> <li>- Automatisation de l'extraction en mobilisant les outils de web scraping (extraction des données du web)</li> <li>- Qualification des données client et pré-traitement</li> </ul>	<p>C1. Explorer le besoin d'un commanditaire interne en produisant une note ou un cahier des charges précisant les questions statistiques qui seront étudiées, les sources d'informations mobilisées et l'analyse des informations produite afin de valider la bonne compréhension de la demande.</p>	<p><i>Epreuve 1 : Rapport d'analyse de besoin et de faisabilité</i></p> <p>A partir de l'analyse d'un besoin connaissance client d'une entreprise, le candidat est chargé de réaliser un rapport écrit de type étude de faisabilité sur la réalisation d'un travail statistique répondant à ce besoin.</p> <p>Ce rapport comprend a minima :</p> <ul style="list-style-type: none"> <li>-Une présentation de l'entreprise</li> <li>-Identification d'une problématique entreprise.</li> <li>-Proposition d'une approche statistique</li> </ul> <p>Le rapport comprendra une partie spécifique décrivant un enrichissement possible des données de l'entreprise par des données collectées sur le site internet de l'entreprise ou d'autres sites (web scraping).</p>	<p>Cr1a : la présentation de l'entreprise comprend :</p> <ul style="list-style-type: none"> <li>-la dimension économique et sociale</li> <li>- l'organisation générale (organigramme)</li> <li>-l'évolution de la place de la donnée dans la stratégie de l'entreprise et l'organisation du travail de la data (organigramme précis).</li> <li>-les enjeux marketing déclinés de manière opérationnelle</li> </ul> <p>Cr1b : le rapport décrit une question précise en la situant dans les enjeux précités.</p> <p>Il propose une approche statistique permettant de répondre à cette question précise : les variables d'intérêt, les variables explicatives et de contexte qui seront mobilisées.</p> <p>Cr2 : le rapport comprend une proposition de cartographie du SI de l'entreprise. Elle concerne l'architecture globale du SI et contient un zoom sur l'organisation et les contenus des bases de données clients.</p> <p>Cr3 : le rapport comprend un schéma logique de requête du SI permettant l'accès aux données</p> <p>Cr4. Une liste des sites sources à mobiliser pour du web scraping est fournie ; les codes des programmes de web scraping sont inclus dans le rapport ; les programmes proposés sont fonctionnels et permettent l'extraction de données.</p> <p>Cr5 : Le rapport contient une présentation d'indicateurs sur la qualité des variables :</p> <ul style="list-style-type: none"> <li>- proposition de traitements mis en œuvre pour améliorer les données (doublons, valeurs</li> </ul>
	<p>C2. Auditer le système d'information en ayant réalisé une cartographie au préalable afin d'identifier les informations utiles à la réponse aux besoins et les possibilités d'amélioration des systèmes d'information.</p>		
	<p>C3. Requête le système d'information à partir des outils de requête des bases de données (SQL) et des API de l'entreprise pour collecter les données qui seront utilisées pour répondre à la demande du commanditaire.</p>		
	<p>C4. Collecter des données sur le web en mobilisant les outils de web scraping afin d'automatiser les extractions permettant de répondre à la demande d'une part et d'autre part d'enrichir le système d'information de façon pérenne.</p>		
	<p>C5. Prétraiter les données collectées en procédant à des vérifications statistiques (qualité des données, valeurs manquantes, aberrantes, complétude des données) et en réalisant le cas échéant des recoupements afin de qualifier les caractéristiques des données et aider à la documentation des données.</p>		
	<p>C6. Documenter les données extraites en produisant l'information nécessaire à son utilisation (sources, traitements réalisés, vigilance sur les biais d'observation, de sélection, les valeurs manquantes, aberrantes...,) dans un objectif de conservation et de réutilisation et dans le respect du cadre déontologique.</p>		

			<p>manquantes, valeurs aberrantes ou singulières) et permettant d'apprécier la qualité des variables,</p> <ul style="list-style-type: none"><li>-les principaux biais sont listés et discutés (biais de collecte, biais de sélection, biais d'observation)</li></ul> <p>Cr6 : Le rapport présente une partie de documentation des données claire et facile d'usage permettant une utilisation du travail par d'autres et sa reproductibilité.</p>
--	--	--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Bloc 2 : Réaliser une modélisation ou une analyse prédictive à partir d'un modèle d'apprentissage avec des données structurées ou non structurées**

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'EVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 2 : Proposition de modèle mathématique, statistique et/ou économique permettant de représenter au mieux les relations et les structures dans les données, et approfondir la connaissance client</b></p> <ul style="list-style-type: none"> <li>- Compréhension du comportement d'achat, de réceptibilité à la communication, dans le cadre plus général de la relation client.</li> <li>- Développement des modèles d'analyse prédictive pour la segmentation des clients, la prévision des ventes ou l'optimisation des campagnes marketing (scoring).</li> <li>- Mise en œuvre des outils et méthodes statistiques permettant le développement des services de marketing digital (système de recommandation, qualification des leads, ciblage,</li> </ul>	C1. Réaliser des analyses de données en mettant en œuvre les procédures statistiques descriptives et exploratoires afin de décrire et qualifier les différentes données caractérisant les clients et leurs comportements, détecter les incohérences dans les données et comprendre les corrélations entre les variables.	<p><i>Epreuve 2 : rapport de méthodologie</i></p> <p>Le rapport présente la comparaison de 2 modèles concurrents (un modèle de régression et un modèle de machine learning) permettant de répondre à une même question statistique. Le rapport devra comprendre également une proposition de méthode d'analyse textuelle et une proposition de modélisation d'analyse du comportement client en fonction du temps sur la même thématique.</p>	<p>Cr1 : le rapport contient la mise en œuvre d'une analyse multifactorielle sur un jeu de données de l'entreprise -les résultats sont commentés et interprétés par rapport à la connaissance clients ; -des conclusions statistiques sont tirées des résultats pour préparer l'étape de modélisation.</p> <p>Cr2 : le rapport comprend :</p> <ul style="list-style-type: none"> <li>- une présentation du fonctionnement des modèles ou des algorithmes ;</li> <li>-une présentation des principales options et des hyperparamètres des modèles en précisant les cas d'usages ;</li> <li>- une application réelle sur un jeu de données de l'entreprise permettant la production et l'analyse de résultats ;</li> <li>-une comparaison des modèles permettant de réaliser un choix raisonné, en particulier les indicateurs quantitatifs de qualité du modèle sont calculés et commentés.</li> </ul> <p>Cr3 -Le rapport présente une analyse de données textuelles -Description des données textuelles travaillées -Présentation des étapes de pré processing de données textuelles -Mise en œuvre des méthodes NLP</p> <p>Cr4 Le rapport comprend l'analyse d'un comportement client dépendant du temps</p>
	C2. Modéliser un comportement donné d'un client ciblé mettant en œuvre les modèles de régression y compris dans les cas complexes (grand nombre de variables, multi colinéarité) afin d'identifier les relations entre les variables et mieux comprendre les actes de consommation du client.		
	C3. Prédire un comportement donné d'un client cible en utilisant les modèles supervisés non paramétriques et en s'appuyant sur les méthodes du machine learning (arbres de décision, forêts aléatoires, méthodes à noyau, machines à vecteur de support avec noyau non linéaires) afin d'optimiser la connaissance et la relation client.		
	C4.Traiter les données textuelles recueillis auprès des clients avec les outils du NLP (traitement automatique du langage naturel) afin de réaliser la classification de texte, l'extraction d'informations, et permettre la mise en œuvre des méthodes dites d'analyse de sentiments permettant la compréhension du client.		
	C5. Modéliser les évolutions d'un comportement client, en fonction du temps et à partir de l'analyse des séries chronologiques (en mettant en œuvre des modèles ARIMA, SARIMA, Holt-Winters, Réseaux de neurones récurrents) afin de faire des prévisions de valeurs futures en fonction de certains déterminants, en particulier ceux sur lesquels une action est possible.		

<p>chatbot, personnalisation des messages).</p>	<p>C6. Produire les modèles adaptés aux données déséquilibrées et aux valeurs manquantes en paramétrant les procédures statistiques utilisées afin d'obtenir une meilleure fiabilité des informations produites.</p>		<p>(mise en œuvre de modèles de durée ou de séries chronologiques). Cr5 Le rapport comprend une partie présentant les principales méthodes qui pallient aux données déséquilibrées ou manquantes.</p>
-------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Bloc 3 : Formaliser et visualiser les résultats des traitements et analyses statistiques**

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 3 : Visualisation de la donnée et interprétation des données pertinentes et utile à la connaissance client</b></p> <ul style="list-style-type: none"> <li>- Production d'informations à l'intention des équipes métiers (y compris publics en situation de handicap) permettant l'évaluation, l'organisation, la prise de décision</li> <li>- Identification des indicateurs utilisés en marketing, ou production d'indicateurs nouveaux à partir d'une recherche</li> <li>- Représentations graphiques</li> </ul>	<p>C1. Produire des indicateurs répondant aux problématiques métiers en précisant leurs propriétés métrologiques et leurs limites d'interprétation afin de répondre à la demande des directions métiers et d'accompagner la prise de décision.</p>	<p><i>Épreuve 3 : La thèse professionnelle.</i></p> <p>Elle donne lieu à un rapport écrit et une soutenance orale.</p> <p>Le rapport est un document de 60 à 80 pages sur l'exploration d'un besoin identifié d'un service métier (marketing, connaissance client ...) présentant une proposition argumentée de réponse mobilisant les outils et méthodes de traitement de l'information.</p> <p>La thèse doit permettre de développer une démarche scientifique adaptée au travail de la donnée et une utilisation professionnelle des outils de la statistique : en particulier modélisation, prédiction.</p> <p>Le rapport aborde notamment les points suivants :</p> <ul style="list-style-type: none"> <li>- La problématique</li> <li>- La description des données : recueil, qualité</li> <li>- Des statistiques descriptives</li> <li>- Les méthodes</li> <li>- Les résultats et leur discussion</li> </ul>	<p>Cr1 : la thèse comprend une description détaillée des enjeux métiers resituée dans la stratégie de l'entreprise qui permet de proposer une solution data pertinente à ces enjeux.</p> <p>Cr2 : la thèse comprend une revue de la littérature qui permet de réaliser un état de l'art sur la question posée et de reformuler la problématique. La revue doit citer a minima 5 articles de revues scientifiques.</p> <p>Cr3 : Un choix raisonné de la variable d'intérêt est présenté, en particulier en discutant des alternatives, et comprend les éléments statistiques permettant d'apprécier sa pertinence par rapport à la problématique.</p> <p>Cr4 : les indicateurs sont calculés, justes et présentés dans le rapport.</p> <p>Cr5 : La thèse comprend des représentations graphiques.</p> <ul style="list-style-type: none"> <li>-Les graphiques mettent en évidence les informations utiles à la compréhension des principaux constats de l'étude</li> <li>-Les titres des graphiques respectent la syntaxe statistique</li> <li>-Les légendes des graphiques sont explicites et justes.</li> </ul> <p>Des notes de lecture guident le lecteur dans la compréhension du graphique.</p> <p>Les graphiques respectent les règles concernant les handicaps visuels.</p> <p>Cr6 La thèse comprend une proposition de tableau de bord et décrit la démarche utilisée pour le concevoir ; le tableau de bord proposé</p>
	<p>C2. Réaliser les représentations graphiques les plus pertinentes possibles en vérifiant les règles de la sémiologie graphique et en déterminant les formes graphiques les plus adaptées afin de rendre accessible la compréhension des données au client de façon claire et pédagogique, et ce en intégrant la prise en compte des handicaps visuels et en utilisant les outils de datavisualisation.</p>		
	<p>C3. Concevoir des dashboard (tableaux de bord) en interaction avec les équipes en ayant documenté les données en amont et en s'assurant de leur automatisation possible afin de compiler les informations essentielles sur les moyens, le suivi de l'activité et la performance et ce, en tenant compte des ressources de l'entreprise. (data notamment)</p>		
	<p>C4. Construire un discours narratif permettant de susciter l'intérêt de son auditoire afin de mettre en avant les principaux messages de son analyse tout en favorisant sa mémorisation et ce, en s'appuyant sur les formes produites.</p>		

		<p>- Des conclusions en forme de préconisations</p> <p>La thèse devra également inclure une proposition de tableau de bord sur la thématique traitée et un compte-rendu de la démarche utilisée pour le concevoir.</p>	<p>tient en 1 page, contient au maximum 6 graphiques ; les graphiques choisis répondent à une logique métier permettant de suivre l'activité ou le comportement étudié ; le tableau de bord proposé a été construit en interaction avec les équipes métier</p> <p>Cr7 : le rapport est rédigé dans un style narratif correspondant aux codes du storytelling :</p> <ul style="list-style-type: none"> <li>-La problématique est claire et a été traduite en questions de façon à impliquer le lecteur.</li> <li>- Le rapport est construit comme une suite de problèmes à résoudre.</li> <li>- il est clair et fait usage d'un vocabulaire adapté (vulgarisation scientifique)</li> <li>- la syntaxe statistique est rigoureuse</li> </ul>
--	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Bloc 4 : Participer à la mise en œuvre de projets de l'entreprise**

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'EVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 4 : Management des projets de l'entreprise en valorisant et optimisant l'apport des sciences de la donnée dans le respect des cadres déontologiques, éthiques et du développement durable</b></p> <ul style="list-style-type: none"> <li>- Gestion et planification du projet de campagne marketing ou d'analyse de données client</li> <li>- Pilotage des ressources du projet</li> </ul> <p><b>Activité 5 : Partage de la culture de la donnée au sein de l'entreprise :</b></p> <ul style="list-style-type: none"> <li>- Valorisation de l'utilisation de la donnée</li> <li>- Prévention des risques réglementaires</li> <li>- Veille technologique (outil IA, data science....)</li> </ul>	<p>C1. Définir le cadre de gestion de projet le plus adapté en tenant compte des contraintes de délais, de ressources et d'objectifs afin de cadrer son projet et gagner en efficacité pendant son déploiement.</p>	<p><i>Epreuve 4 : rapport d'expérience professionnelle</i></p> <p>Un rapport de 5 à 10 pages présentant de 1 à 3 projets professionnels (en fonction de la complexité) auquel le candidat a participé. Cela peut être un projet majeur pour le service impliquant une organisation lourde ou 2 ou 3 projets de moindre ampleur. Ce sont des projets dans lesquels l'apprenant a eu un rôle significatif comme participant et/ou éventuellement comme pilote. Le rapport est structuré de façon à mettre en évidence les points clés de la gestion de projet et du partage de la culture de la donnée au sein de l'entreprise.</p>	<p>Cr1 : le rapport contient une présentation du ou des projets et de ses objectifs Le candidat propose un cadre de gestion de projet de type GANTT. Cr2 : le rapport contient une description du pilotage, de la coordination, de la communication : -livrables intermédiaires et finaux attendus ; -organisation dans le temps, jalons ; - rôles et missions des principales ressources humaines intégrées au projet. Cr3 : le rapport contient une description des contraintes des projets en termes de respect des cadres déontologiques, éthiques et de développement durable. Il évoque des textes de lois applicables à son activité. Cr4 le rapport contient une description des rôles et missions de chaque métier qui est impacté par le projet. Cr5 : le rapport décrit les moyens de communication employés vers les professionnels d'autres métiers : compte-rendu de réunion, document de présentation de son métier, de son activité, ou de son projet autour des données. Les documents sont rédigés pour un public non formé à la data science. Cr6 : Le rapport comprend une liste d'outils ou de sources utiles pour maintenir une veille technologique et réglementaire. Les sources</p>
	<p>C2. Piloter un projet dans le respect du cadre donné par une feuille de route, incluant la bonne gestion des ressources humaines du projet dans le respect des besoins individuels (notamment en cas de situation de handicap), le respect des différents jalons afin de s'assurer de la bonne qualité du livrable et du respect de la demande du client et le cas échéant, du cahier des charges.</p>		
	<p>C3. Diffuser l'usage des nouvelles méthodes d'exploitation des données, notamment l'IA, et leur cadre éthique auprès des directions commerciale et marketing en s'appuyant sur une veille active, et en apportant des conseils sur les processus et procédures à mettre en place à tous les stades de l'utilisation des données afin de développer des bonnes pratiques et garantir le respect de la vie privée, la confidentialité des données et le respect des règles de sécurité de l'information.</p>		
	<p>C4. Partager ses besoins et ses contraintes, ses objectifs, sa culture métier dans le cadre de dispositifs collectifs existant ou, le cas échéant, créés à son initiative afin de mieux comprendre les problématiques rencontrées par les différents professionnels des métiers connexes dans l'entreprise autour du traitement de l'information et construire un écosystème de la data plus efficace et rendre plus fluide la collaboration entre les services et métiers.</p>		
	<p>C5. Concevoir un processus structuré de veille technologique et réglementaire en réalisant une sélection d'outils performants par rapport à l'objet des recherches à effectuer afin de s'assurer du maintien de son efficacité opérationnelle et de ses compétences technologiques (notamment en matière d'IA, outils datascience) et assurer le respect du cadre déontologique et éthique.</p>		

	<p>C6. Réaliser des notes explicatives à l'attention des parties prenantes et lecteurs non data scientists permettant une prise de décision éclairée à partir d'une bonne compréhension des résultats et de leur limites (interprétabilité, intelligibilité) et tenant compte d'un contexte multiculturel et/ou international ou de la présence d'un éventuel profil en situation de handicap.</p>		<p>sont de qualité et s'appuient sur des éléments vérifiés.</p> <p>Cr7 Le rapport est structuré, le vocabulaire est adapté (vulgarisation scientifique), la rigueur de la syntaxe statistique.</p> <p>Cr8 le contenu est adapté aux différentes typologies de handicap :</p> <ul style="list-style-type: none"> <li>- Le candidat précise l'existence de 6 typologies.</li> <li>- Le candidat présente des modalités d'adaptation de son contenu de sorte à ce qu'il puisse répondre aux besoins d'adaptation d'un trouble spécifique de handicap a minima. Il s'agit notamment des handicaps cognitifs de type troubles dys (dyslexie, dysorthographe, dyscalculie, dyspraxie, etc.), trouble de déficit de l'attention avec ou sans hyperactivité (TDAH)... ou encore des handicaps sensoriels, notamment pour exemple une déficience visuelle, déficience auditive, daltonisme, hyper ou hypo-sensibilité (lumière, texture, son...).</li> </ul> <p>Il réalise une proposition permettant l'accès à l'information au public en situation de handicap</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------