

Université de technologie de Troyes

Expert Big Data Engineer (MS)

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC N° 1 : ANALYSER L'EXISTANT ET PROPOSER DES SOLUTIONS LIÉES A LA DATA			
<p>A.1.1 Analyse des besoins et des problématiques :</p> <ul style="list-style-type: none"> - Référencement et regroupement de l'ensemble des sources de données structurées ou non structurées nécessaires et pertinentes à l'analyse - analyse de la problématique liée aux données (<i>data</i>) - conception d'un projet <i>data</i> avec les parties prenantes - prise en compte de l'outil informatique existant (serveurs de stockage). 	<p>C.1.1 Analyser les problématiques au sein d'une organisation privée/publique en mettant en place un déploiement d'une architecture ou d'une structure, en interagissant avec des interlocuteurs différents et concernés par la problématique.</p>	<p>Réalisation de Projets : en équipe, les candidats devront répondre à une problématique client fictive. En relation avec des experts, ils doivent étudier un système d'informations donné, appréhender le volume de données et les risques afin d'apporter des recommandations.</p> <p>Un rapport détaillant les différentes étapes du processus de traitement de la problématique est attendu de chaque groupe, suivie d'une présentation orale.</p>	<ul style="list-style-type: none"> - La collecte des éléments (l'existant) est suffisante - L'interaction avec les acteurs du projet est fluide, le candidat est compris par l'auditoire - La synthèse de l'existant reprend les éléments essentiels pour cadre les besoins et les enjeux - L'état des lieux de l'outil informatique permet de cartographier en profondeur l'intégralité du système d'information, les priorités sont définies. - L'évaluation de la pertinence de l'outil

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

		<p>Contrôle continu sous forme d'évaluation écrite : Ces évaluations se réalisent sous la forme de questionnaires ou d'exercices. Ils sont d'une durée d'une heure à deux heures au maximum. Il s'agit de questions portant sur la réglementation de la gestion de données massives et sur la méthodologie de l'analyse des besoins et des problématiques.</p>	<p>informatique est mise en place</p> <ul style="list-style-type: none"> - Le diagnostic de l'outil permet d'émettre des recommandations - La veille technologique sur l'open data est réalisée et comprendre l'actualité des données publiques, administratives et réglementaires - Les données trouvées sont fiables (ex : météo France, sites gouvernementaux...) - La restitution individuelle des concepts exposés lors des cours et mis en œuvre lors des séances de travaux pratiques.
<p>A.1.2 Identification des données de l'organisation</p> <ul style="list-style-type: none"> - supervision d'une réunion d'équipe - management de projet du team data - collecte des données auprès des différents services/responsables métiers 	<p>C.1.2 Animer et mener une réunion avec les acteurs principaux et secondaires concernés par une problématique existante ou nouvelle afin de trouver les éventuelles solutions pour la résoudre en tenant compte de l'avis des différents acteurs et afin de s'immerger dans l'écosystème de la structure dans lequel on intervient.</p>	<p>Mise en situation professionnelle avec un jeu de données en équipe : il s'agit d'un travail en équipe sur des sites, afin de trouver des algorithmes. Ce travail se déroule en fil rouge. Les candidats sont dans une situation d'un travail en équipe, avec un chef</p>	<ul style="list-style-type: none"> - Les interrogations du candidat sont pertinentes - L'écoute est active - Les besoins du client sont correctement reformulés et ses enjeux pris en compte- Les difficultés et conflits éventuels sont bien gérés

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>- cartographie des données</p>		<p>d'équipe qui répartit les tâches au sein du groupe (vérification de la base, travail des algorithmes, rédaction d'un rapport...).</p> <p>Le candidat est mis en situation dans un groupe d'étudiants dont il anime une réunion en tant que chef de projet.</p> <p>Chaque équipe doit rendre une feuille de route de projet, détaillant les jalons définis, les acteurs et leur rôle, le tableau de bord de suivi du travail, et le reporting de la cartographie des données pour l'animation de la réunion.</p>	<ul style="list-style-type: none"> - La détermination du candidat pour l'obtention des données est notoire -Les collaborateurs externes sont convaincus et adhèrent à la démarche proposée, grâce à la présentation du prototype. - Le candidat s'est bien adapté à son auditoire, son langage est proportionné - Le projet respecte les jalons de départ - Les données recueillies sont bien retranscrites et visibles dans la cartographie - Les données sont anonymes et respectent la RGPD - La cartographie permet de : <ul style="list-style-type: none"> • visualiser le volume et la variété des données, • la vitesse d'acquisition, les moyens d'accès • et la requête automatique des données - La réglementation en vigueur sur la confidentialité des données
-----------------------------------	--	--	--

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

			(CNIL, RGPD) est prise en compte - La veille réglementaire inhérente à la confidentialité est judicieuse
<p>A.1.3 Formalisation du besoin et des solutions techniques :</p> <ul style="list-style-type: none"> - formalisation du besoin et des enjeux - analyse de manière qualitative et quantitative le projet dans son ensemble - analyse des données (data) et des données non structurées - réalisation d'un cahier des charges 	<p>C.1.3 Formaliser le besoin émanant de la problématique au sein d'un cahier des charges dans le respect du budget alloué afin de proposer des solutions réalisables et réalistes avec les données et ressources propres possédées.</p>	<p>Réalisation de Projets : rédaction d'un cahier des charges présentant le besoin et les contraintes (budget, délai, technique) et le présenter à l'oral.</p> <p>En plus du rapport de projet, l'étudiant devra rendre un suivi écrit retraçant les éventuelles difficultés rencontrées et liées à la réalisation de l'objectif du projet.</p>	<ul style="list-style-type: none"> - Le cahier des charges réalisé reprend bien tous les enjeux et objectifs auxquels le commanditaire est soumis. - Le cahier des charges reprend bien les fonctions, les critères et niveaux d'appréciation - Les contraintes liées au développement sont prises en compte - Les solutions de développement proposées sont souples et basées sur les méthodes Agiles, etc. - Le diagramme de Gantt est réalisé pour planifier de façon optimale et communiquer sur le planning établi et les choix qu'il impose.

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

			<p>Le diagramme réalisé permet :</p> <ul style="list-style-type: none"> * de déterminer les dates de réalisation du projet ; * d'identifier les marges existantes sur certaines tâches ; * de visualiser d'un seul coup d'œil le retard ou l'avancement des travaux. <p>- Les recommandations tiennent compte de l'utilisation d'outils appropriés pour le développement du projet</p>
<p>A.1.4 Préconisations de stratégies en matière de possession, gestion et transformation des données :</p> <ul style="list-style-type: none"> - proposition d'une stratégie Data - prise en compte du droit des données - prise en compte du cloud computing (serveurs, stockage, mise en réseau, logiciels) de l'entreprise/collectivité dans sa stratégie - prise en compte des risques liés à la donnée 	<p>C.1.4 Présenter un rapport détaillé de faisabilité en tenant compte des paramètres extérieurs tels que le développement durable, la réglementation en vigueur liée à la confidentialité des données, l'éthique afin de les rendre compatibles avec les contraintes inhérentes à l'organisation.</p>	<p>Étude de cas : Il est remis à l'étudiant un scénario, dont il doit faire un rapport détaillé en tenant compte des différents aspects nécessaires au respect de la procédure réglementaire. L'étudiant doit rendre le rapport et faire une présentation orale devant les membres du jury composé de l'enseignant et 3 étudiants.</p>	<ul style="list-style-type: none"> - La stratégie Data proposée est pertinente - La proposition respecte le droit des données - Les solutions proposées sont modernes et techniquement cohérentes avec le besoin - La prise en compte des différents risques liés à la donnée (risque d'image, risque technique, qualité du matériel, sensibilisation des collaborateurs, crash et

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>- proposition de serveurs auprès des utilisateurs de l'organisation privée/publique.</p>		<p>Contrôle continu sous forme d'évaluations écrites : Elles se réalisent sous la forme de questionnaires ou d'exercices. Elles sont d'une durée d'une heure à deux heures au maximum. Il s'agit de questions portant sur la réglementation de la gestion de données massive et sur les stratégies en matière de possession, gestion et transformation des données.</p>	<p>pannes, sauvegarde des données...) - La veille technologique est de bonne qualité - Les outils proposés sont en adéquation avec le niveau technique des utilisateurs potentiels - La restitution individuelle des concepts exposés lors des cours et mis en œuvre lors des séances de travaux pratiques. - La réglementation RGPD est maîtrisée</p>
<p>Bloc de compétences capitalisable, obligatoire pour l'obtention du titre (validité du bloc illimitée). L'acquisition du bloc fait l'objet de remise d'un certificat. Le cas échéant, description de tout autre document constitutif de la certification professionnelle. La thèse professionnelle, modalité d'évaluation globale et transversale, doit être validée au même titre que les 4 blocs de compétences pour obtenir la certification.</p>			

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC N° 2 : DÉPLOYER DES INFRASTRUCTURES INFORMATIQUES POUR ANALYSER LA DATA			
A.2.1 Définition de l'architecture de stockage - identification des solutions de stockage possibles des données - utilisation d'un logiciel de stockage des données (Hadoop) - proposition de solutions de stockage de données (stockage interne, en ligne) en fonction de l'organisation - explication et documentation du choix de la solution proposée - prise en compte de la législation des données selon la situation des serveurs (UE/hors UE)	C.2.1 Identifier la façon dont l'organisation assure le stockage et le format des données internes, structurées ou non structurées , en tenant compte des données manquantes, erronées ou perdues, des différents processus en fonction des pays, des entités qui vont produire les flux de données et de la réglementation en vigueur sur la confidentialité des données (RGPD) afin d'avoir un jeu de données adaptées (ex : Data Frame Python, R/Tibble).	Réalisation de Projets : étude sur la mise en place d'un projet data en intégrant les contraintes techniques et législatives. Soutenance orale du projet. Chaque étudiant doit rendre sa feuille de route de projet définissant la charte du projet, le calendrier prévisionnel, la présentation des résultats en tenant compte des aspects législatifs.	- Les différentes solutions de stockage sont correctement décrites. - Les outils à mettre en place en fonction de la solution choisie sont expliqués : <ul style="list-style-type: none"> • établir des connexions solides • gérer non seulement la sécurité des périphériques, mais également celle du réseau. - Les solutions et inconvénients inhérents à chaque solution sont décrits avec précision. - Le stockage proposé est adapté en fonction du budget, des besoins et capacités de l'entreprise et en fonction de l'activité.

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>A.2.2 Tests et mise en place de protocoles</p> <ul style="list-style-type: none"> - réalisation de tests sur les web services - rédaction des protocoles à destination des utilisateurs de ces outils - communication auprès des équipes sur les protocoles existants et sur leur utilisatio 	<p>C.2.2 Mettre en place les protocoles nécessaires à l'utilisation des web services au sein de la structure en réalisant des tests sur les services et rédigeant une documentation et en communiquant auprès des équipes pour leur permettre de les utiliser de façon optimale.</p>	<p>Réalisation de Projets : le projet porte sur la rédaction de protocoles / tutoriels. L'étudiant doit préparer des exemples types, et fournir des solutions adaptées à ces derniers. La mise en place d'un forum et wiki de discussion autour de l'outil mis en place est aussi requise. L'étudiant doit rendre son rapport avec les exemples types et les solutions sélectionnées.</p>	<ul style="list-style-type: none"> - Les tests pour la validation du fonctionnement de l'outil sont réalisés - Les tests de montée en charge de l'outil (par rapport au volume des données traitées, et à la qualité de requête des données...) sont réalisés - Les programmes de solutions de l'outil sont rédigés - La nécessité ou non de l'intégration API (<i>Application Programming Interface</i>) pour l'analyse du besoin - La création d'une API (<i>Application Programming Interface</i>) - La rédaction de protocoles et tutoriels est adaptée à l'outil et aux futurs utilisateurs (ex. : forum, wiki autour de l'outil mis en place) - Des exemples types et de solutions inhérentes sont proposés

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>A.2.3 Extraction d'information (Information Retrieval) :</p> <ul style="list-style-type: none"> - utilisation d'outils de visualisation de tendances en grandes dimensions - mise en place d'enrichissement des données - calcul d'indicateurs - mise en œuvre d'interconnexion et visualisation de graphes sociaux distribués 	<p>C.2.3 Mettre en œuvre des techniques d'extraction d'information intrinsèque ou par recoupement d'interconnexion en utilisant des outils permettant la visualisation de tendances en grandes dimensions afin d'extraire des data pertinentes dans les flux de données.</p>	<p>Réalisation de Projets : mise en place d'outils de visualisation des données et de calculs d'indicateurs par exemple à l'aide des bibliothèques pandas de Python et tidyverse de R. L'étudiant doit rendre un rapport et un script (une succession de commandes commentées soit sur Python/R)</p>	<ul style="list-style-type: none"> - Les outils de visualisation classique (d3.js, Tableau, Python, R, Infogram, FusionCharts Power BI, Jupyter...) sont maîtrisés et bien adaptés au besoin - Les outils utilisés permettent le traitement de données massives- La mise en relation des bases de données (enrichissement des données) internes et externes est faite. - Les outils de modélisation (Régression, réseau de neurones...), de classification (supervisés ou non) et de réduction de dimension (analyse factorielle...) sont utilisés
--	---	---	---

Bloc de compétences capitalisable, obligatoire pour l'obtention du titre (validité du bloc illimitée).

L'acquisition du bloc fait l'objet de remise d'un certificat.

Le cas échéant, description de tout autre document constitutif de la certification professionnelle.

La thèse professionnelle, modalité d'évaluation globale et transversale, doit être validée au même titre que les 4 blocs de compétences pour obtenir la certification.

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC N° 3 : TRAITER ET VISUALISER DE LA DATA			
<p>A.3.1 Mise en œuvre de langages et outils de programmation :</p> <ul style="list-style-type: none"> - utilisation d'une programmation pour l'analyse décisionnelle - utilisation de langages et outils de programmation : <i>ElasticSearch Java mapReduce, Python, R, SCALA,</i> - mise en œuvre des algorithmes de traitement de grosses volumétries de données 	<p>C.3.1 Programmer en utilisant des langages de programmation adaptés à la donnée récoltée au sein de l'organisation et de la donnée provenant de sources extérieures afin de mettre en place des algorithmes gérant le volume important des données.</p>	<p>Réalisation de Projets : le projet est centré sur les aspects techniques de la résolution d'un ou plusieurs problèmes réels. Le projet se fait en petit groupe (chef de projet, programmeur et testeur). Les étudiants devront répondre à une problématique donnée à leur groupe. Un retour d'expérience des autres groupes est également proposé.</p> <p>Chaque groupe devra rendre son rapport de projet avant la soutenance devant le jury composé de l'enseignant expert et trois étudiants. Le retour d'expérience se fera à la fin des soutenances, chaque groupe choisira un</p>	<ul style="list-style-type: none"> - Le choix est adapté avec le problème à résoudre - Le code est écrit dans différents langages - Le code réalisé est correct - Le code est bien documenté - Le code respecte les conventions d'écriture usuelles du langage (ex. : PEP 8 pour Python)

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

		orateur qui prendre la parole pendant 5 minutes pour donner leur impression par rapport à la vie de groupe.	
<p>A.3.2 Réalisation d'une visualisation dynamique :</p> <ul style="list-style-type: none"> - choix d'une solution de visualisation adaptée (<i>Chart.js DataHero, Dygraphs, InstantAtlas Plotly, Kibana, Raw ...</i>) - utilisation d'outils de visualisation des données 	<p>C.3.2 Utiliser des outils avancés pour la visualisation des données en choisissant les outils adaptés, en respectant les méthodes de représentation graphique afin d'obtenir une visualisation dynamique des données.</p>	<p>Réalisation de Projets : réalisation et présentation d'une visualisation dynamique dans le cadre de son projet.</p> <p>Les étudiants devront trouver par eux-mêmes une base de données qui nécessite la mise en place d'une visualisation approfondie et dynamique.</p> <p>Les étudiants en binôme devront rendre un rapport de projet lors de l'évaluation.</p>	<ul style="list-style-type: none"> - Le choix d'une solution de visualisation est en cohérence avec la nature des données à représenter - Le choix d'une solution de visualisation a été choisie en cohérence avec la masse de données à représenter - Les outils de visualisation sont judicieusement choisis - Les outils de visualisation sont fonctionnelles -La visualisation est simple, fonctionnelle et adaptée aux futurs utilisateurs
<p>A.3.3 Analyse des données et conseil auprès du commanditaire :</p> <ul style="list-style-type: none"> - analyse et visualisation spatio-temporelles - utilisation de <i>Scalable MultiMedia Analytics</i> 	<p>C.3.3 Analyser les données représentées graphiquement en identifiant les tendances et les valeurs inhabituelles dans ces données afin de prendre des décisions en s'appuyant sur ces données et d'apporter ainsi une solution aux utilisateurs.</p>	<p>Réalisation de Projets : Une base de données issue de tous les domaines possibles est donnée aux étudiants avec une problématique bien définie et des questions ouvertes.</p>	<ul style="list-style-type: none"> - Les conditions d'utilisation des outils (vérification des hypothèses) sont validées

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<ul style="list-style-type: none"> - visualisation des données permettant de découvrir un phénomène non connu / de confirmer ou d'affirmer une information ou une hypothèse de travail - interprétation des graphiques obtenus - formalisation de rapports et de conseils (stratégiques, économiques, organisationnels...) auprès des utilisateurs dédiés 		<p>Des groupes d'étudiants procèdent à l'analyse des données. Il est attendu une présentation orale des résultats obtenus.</p> <p>Cette présentation est suivie d'échanges et de questions avec le professeur et deux rapporteurs choisis parmi les étudiants. Le reporting de présentation est requis</p>	<ul style="list-style-type: none"> - Les explications sur les résultats numériques et graphiques obtenus sont cohérentes - Les questions et les objections sur les résultats obtenus sont cohérents - Les outils utilisés sont maîtrisés et leurs limites connues - Les conseils sont formalisés suivant la problématique initiale - Les conseils préconisés sont réalistes : adaptés aux interlocuteurs, à la structure et aux moyens (technique, fonctionnel, budget) - La présentation orale est fluide, claire, précise, concise, efficace et correspond au temps imparti
<p>Bloc de compétences capitalisable, obligatoire pour l'obtention du titre (validité du bloc illimitée). L'acquisition du bloc fait l'objet de remise d'un certificat. Le cas échéant, description de tout autre document constitutif de la certification professionnelle. La thèse professionnelle, modalité d'évaluation globale et transversale, doit être validée au même titre que les 4 blocs de compétences pour obtenir la certification.</p>			

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
BLOC N° 4 : ANALYSER ET MODELISER LA DATA			
A.4.1 Utilisation de données en grande dimension : - utilisation d'un système de gestion de bases de données (Bigtable, HyperTable NoSQL, SQL ...) - transformation de la donnée - recherche de cibles (<i>patterns</i>) enfouies dans les données - segmentation des données afin d'évaluer des probabilités - Mise en place d'un algorithme issu du <i>Machine Learning</i> pour prédire ou comprendre et expliquer.	C.4.1 Utiliser des outils permettant l'analyse et la classification de données sous différents angles afin d'identifier les liaisons ou les patterns (motifs) entre les nombreuses bases de données relationnelles.	Réalisation de Projets : Le projet demande dans un premier temps d'expliquer pourquoi nous sommes dans une problématique de grande dimension. La deuxième partie réclame la mise en application des techniques de grande dimension. La problématique est tournée soit vers un problème de classification, soit vers un problème de prédiction. Le rendu du projet se fera sous forme d'une présentation orale, devant un jury composé du professeur et de deux étudiants.	- Le système de gestion de bases de données en place permet le choix selon le volume, la nature et la variété des données - Le traitement des données est anticipé voire prévu - Les données provenant d'autres bases de données différentes sont agrégées - Les données sont formalisées - La gestion des valeurs manquantes (ex. : informations manquantes ou partiellement remplies) est établie. - Les valeurs manquantes sont identifiées. - Le traitement des valeurs manquantes proposé par le candidat est pertinent par rapport au besoin.

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

			<ul style="list-style-type: none">-L'extraction de connaissances (ex. : règles d'association) sur des jeux de données massives ou de grande dimension est mise en place.- Les outils de classification des données (supervisés ou non supervisés) sont en place et permettent : une classification ascendante hiérarchique, des nuées dynamiques, une analyse discriminante...- Le choix d'un algorithme de Machine Learning est adapté à la problématique- L'entraînement de l'algorithme de Machine Learning (SVM – Support Vector Machine, réseaux de neurones, cartes de Kohonen, algorithmes génétiques, forêts aléatoires...) est adapté et correct- Le codage de l'algorithme de Machine Learning sur un grand jeu de données est adéquat.

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>A.4.2 Déploiement et communication sur le modèle d'apprentissage profond (Deep Learning) :</p> <ul style="list-style-type: none"> - mise en place d'application de <i>Deep Learning</i>. - déploiement d'un modèle via une interface de programmation (API) dans le web. - réalisation d'un tableau de bord (Dashboard) pour présenter son travail de modélisation - Communication sur le modèle d'apprentissage auprès des utilisateurs - prise en compte de la consommation d'énergie dans son choix des outils ou prestataires. 	<p>C.4.2 Déployer un modèle d'apprentissage profond (Deep Learning) en développant des modèles, des algorithmes via une interface de programmation/graphique, en réalisant un document méthodologique/technique et en communiquant oralement sur ce modèle afin de favoriser l'intégration sa démarche de modélisation auprès de ses utilisateurs.</p>	<p>Réalisation de Projets :</p> <p>Le projet est tourné vers deux axes. Le premier axe concerne la prédiction. Pour qualifier cette prédiction, des indicateurs seront précisés et discutés (leur pertinence). Un deuxième axe est orienté vers la consommation de ressources et l'impact environnemental.</p> <p>L'étudiant devra rendre un rapport écrit en ne négligeant pas l'impact environnemental. En effet, il devra convaincre le lecteur au déploiement de son algorithme de <i>Deep Learning</i>.</p>	<ul style="list-style-type: none"> -L'identification de la structure du réseau de <i>Deep Learning</i> est appropriée. - La solution de logiciel a été correctement utilisée. - L'identification de l'architecture matérielle sur laquelle faire tourner l'algorithme/l'apprentissage est établie - Un algorithme de prédiction à partir d'un modèle ayant été convenablement entraîné est créé - L'encapsulation du modèle dans une API web permettant son utilisation à distance (entraînement et prédiction) est établi. - La mise en place d'un tableau de bord permet le pilotage du modèle et de l'évaluation de ses capacités d'ajustement et de prédiction- La qualité de la documentation (annexes techniques adaptées) et de la présentation de la démarche suivie pour choisir la méthodologie
---	---	---	---

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

			<p>applicable est cohérente et pertinente.</p> <ul style="list-style-type: none"> - La documentation (annexes techniques adaptées) et la communication sont également adaptées à une utilisation future et facile pour les utilisateurs concernés et potentiellement non spécialistes du domaine. - La présentation orale est fluide, claire, précise, concise, efficace et correspond au temps imparti - Le modèle de Deep-Learning choisi est un bon compromis entre efficacité et coût énergétique.
<p>A.4.3 Traitements de volumes variés de données :</p> <ul style="list-style-type: none"> - mise en œuvre avec le logiciel des applications sur des machines standards - traitement de grandes quantités de données 	<p>C.4.3 Assurer le traitement de volumes variés (gros ou non) de données avec un langage de programmation/logiciel dédié gérant des algorithmes permettant à la structure d'opérer plus vite.</p>	<p>Réalisation de Projets :</p> <p>Le projet est axé principalement sur l'explication détaillée et argumentée par l'étudiant du traitement de la masse des données. Il ne négligera pas la prise en compte</p>	<ul style="list-style-type: none"> - Les prestataires choisis sont écologiquement responsables. - Le choix du langage est adapté au volume des données (ex. : choix d'un langage traditionnel ou spécifique aux données massives)

ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>- prise en compte de la consommation d'énergie dans son choix des outils ou prestataires</p>		<p>de l'énergie dépensée pour traiter la masse de données. Il devra mettre en concurrence plusieurs prestataires pour résoudre son problème d'énergie.</p> <p>Un rapport écrit est demandé avec la mise en pratique du traitement de la masse des données et les bons arguments. Il expliquera comment il a choisi et motivera ses décisions.</p> <p>La présentation se fera à l'oral devant un jury composé du professeur et deux professionnels extérieurs.</p>	<ul style="list-style-type: none"> - le fonctionnement des algorithmes (parallélisation, utilisation de grilles de calcul, Map Reduce,...) est optimal - Des outils spécifiques au domaine (ex. : Spark, Hadoop, l'algorithme Map Reduce, AWS ...) d'expertise sont utilisés. - Le modèle choisi est un bon compromis entre efficacité et coût énergétique.
<p>Bloc de compétences capitalisable, obligatoire pour l'obtention du titre (validité du bloc illimitée). L'acquisition du bloc fait l'objet de remise d'un certificat. Le cas échéant, description de tout autre document constitutif de la certification professionnelle. La thèse professionnelle, modalité d'évaluation globale et transversale, doit être validée au même titre que les 4 blocs de compétences pour obtenir la certification.</p>			