

Ingénieur diplômé de l'École nationale de la statistique et de l'analyse de l'information du Groupe des écoles nationales d'économie et statistique

RNCP40694 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Sigle	Titre ingénieur — Titre ingénieur
Code(s) NSF	114b, 122b, 326m
Formacode(s)	11017, 13149, 31026, 32062, 15099
Code(s) ROME	C1202, H1502, M1806, K2402, M1403
Type d'enregistrement	Enregistrement de droit
Date d'échéance	31/08/2027

Certificateur(s)

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE · SIRET : 11004401300040

GROUPE DES ECOLES NATIONALES D ECONOMIE ET STATISTIQUE · SIRET : 13001422800089

Résumé de la certification

OBJECTIFS ET CONTEXTE

Dès qu'une entreprise collecte un grand nombre de données (data) via les outils numériques afin de développer ses relations avec ses clients, utilisateurs et usagers ; qu'elle soit de type TPE-PME ou plus grande ; et dans tous les secteurs d'activité ; elle a besoin des compétences de Data-Scientists pour collecter, nettoyer, analyser et interpréter ces données de manière à en tirer des connaissances exploitables selon sa stratégie.

Les premiers précurseurs sur le travail des données remontent aux années 1960 et 1970, lorsque des statisticiens et des informaticiens ont commencé à travailler sur des méthodes pour analyser et extraire des informations utiles à partir de grandes quantités de données.

Puis, avec la croissance exponentielle des données et l'avènement des technologies de l'information et de la communication, le rôle du data scientist a évolué et s'est développé. En effet, l'explosion de l'Internet, des médias sociaux, des appareils connectés, des transactions en ligne et d'autres sources de données génèrent une quantité massive de données qui nécessitent des outils de traitement adaptés.

Également, dans de nombreux domaines, la rapidité de traitement des données est devenue cruciale. Les entreprises doivent être en mesure de prendre des décisions en temps réel, de détecter les anomalies ou les fraudes rapidement, et de réagir aux événements en temps opportun. Le traitement en temps réel et les technologies de streaming permettent de répondre à ces besoins de vitesse.

Ce nouveau contexte a donc créé auprès des entreprises, un besoin croissant de professionnels capables de collecter, nettoyer, analyser et interpréter ces données de manière à en tirer des connaissances exploitables.

Depuis lors, le domaine de la data science a connu une forte croissance, et le rôle du data scientist est devenu essentiel dans de nombreux secteurs, notamment l'industrie, la finance, la santé, le marketing, et bien d'autres.

ACTIVITÉS VISÉES

Conception d'un projet de collecte (répondant aux besoins d'un commanditaire)

- Définition des données à collecter
- Description et choix des méthodes de collecte
- Réalisation d'enquêtes statistiques et de procédures automatisées de collecte des données.

Mise en œuvre un système d'information DATA exploitable par les parties prenantes de l'organisation.

- Création de l'environnement applicatif
- Traitement des données brutes
- Documentation de l'accès à l'application ainsi que les données
- Mise à disposition des données auprès des parties prenantes

Proposition d'un modèle mathématique, statistique et/ou économique permettant de représenter au mieux les relations et les structures dans les données.

- Présentation de la variabilité d'un phénomène statistique
- Mesure des effets de variables en vue d'expliquer un phénomène observé.

Réalisation d'analyse prédictive

- Ajustement des paramètres du modèle sur des données d'entraînement
- Formalisation d'analyse statistique
- Actions d'optimisation

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. Formalisation des éléments suivants :

- Informations descriptives permettant de comprendre les tendances, identifier les patterns, ou évaluer les performances.
- Les informations prédictives permettent de prévoir des événements futurs, pour prendre des décisions stratégiques, identifier les risques, ou optimiser les opérations.
- Informations prescriptives indiquant les actions à entreprendre pour atteindre un objectif : améliorer les performances, optimiser les ressources, ou résoudre des problèmes.

Ces informations peuvent prendre la forme d'indicateurs statistiques, de tableaux de bords, de graphiques, de bases de données, de nouvelles variables caractérisant les indicateurs étudiés. La forme est adaptée à l'interlocuteur notamment en cas de situation de handicap spécifique afin de veiller à l'égalité de traitement en matière d'information.

- Restitution des résultats des modèles mis en œuvre

Piloter des projets DATA

- Promotion de l'usage de la donnée auprès des différents métiers de l'entreprise
- Pilotage des projets Data en assurant le cadrage et en animant les différentes phases du projet
- Garantie du bon usage de la data dans le respect du cadre légal, déontologique, éthique.

Spécialisation de l'ingénieur en Datascience, activités visées selon les options :

Option 1 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine du marketing

- Développement des modèles d'analyse prédictive pour la segmentation des clients, la prévision des ventes ou l'optimisation des campagnes marketing
- Développement des services de marketing digital (système de recommandation, qualification des leads, ciblage, chatbot, personnalisation des messages)

Option 2 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la biostatistique

- Analyse des données cliniques et populationnelles pour identifier des tendances et des associations dans le domaine de la santé
- Analyse du volume de données caractérisées par un très grand nombre de variables et peu d'observations

Option 3 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la gestion des risques

- Analyse de données bancaires pour identifier des tendances et des associations
- Développement de modèles d'analyse prédictive pour la gestion des risques de crédit et des risques de marché
- Evaluation des prix de produits financiers complexes

Option 4 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine des entreprises industrielles (filière génie statistique)

- Résoudre des problèmes réels et proposer des solutions IA (maintenance prédictive, optimisation des process, création de produits...)

Option 5 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la modélisation économique et santé

- Aide à la décision et évaluation des politiques et des actions dans le secteur de la santé, de l'environnement, des territoires et des populations

Option 6 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de l'ingénierie des données

- Création d'un environnement applicatif (le cas échéant adapté au big data) en sélectionnant les plateformes et systèmes logiciels permettant l'accès aux données aux différentes parties prenantes de l'organisation
- Création d'un pipeline de collecte et de traitement des données dans un cadre MLOps
- Optimisation du stockage des données de façon sécurisée et efficiente.
- Maintenir le système informatique de données massives.

COMPÉTENCES ATTESTÉES

L'ingénieur diplômé de l'École nationale de la statistique et de l'analyse de l'information du Groupe des écoles nationales d'économie et statistique pilote et met en œuvre des projets de collecte de données. Il peut ainsi concevoir et mettre en œuvre des enquêtes statistiques ou un système de production de données statistiques à partir de sources diverses (données des SI de l'entreprise, du web). Il conçoit et met en œuvre le prétraitement des données et leur mise à disposition en tenant compte de la finalité et des contraintes techniques, réglementaires et environnementales.

L'ingénieur diplômé de l'École nationale de la statistique et de l'analyse de l'information du Groupe des écoles nationales d'économie et statistique met en œuvre des modélisations et des analyses prédictives sur données structurées ou non structurées en s'appuyant sur une démarche scientifique. Il maîtrise en particulier les différentes méthodes économétriques et d'apprentissage.

L'ingénieur diplômé de l'École nationale de la statistique et de l'analyse de l'information du Groupe des écoles nationales d'économie et statistique met en forme les résultats, les interprète et leur donne du sens. Il communique aussi bien à l'écrit qu'à l'oral avec les parties prenantes de l'entreprise en s'adaptant au public visé.

L'ingénieur diplômé de l'École nationale de la statistique et de l'analyse de l'information du Groupe des écoles nationales d'économie et statistique participe aux projets de l'entreprise en valorisant et optimisant l'apport des sciences de la donnée dans le respect des cadres déontologiques, éthiques et de développement durable. Il réalise une veille technologique et scientifique sur son domaine d'expertise.

Liste détaillées des compétences attestées :

- Cartographier les besoins data en mettant en œuvre une démarche scientifique (ateliers d'expression des besoins, états des lieux du SI, enquête auprès des équipes métiers, ...) pour répondre aux enjeux stratégiques et organisationnels du commanditaire et en prenant en compte les aspects légaux, éthiques et multiculturels
- Modéliser l'application permettant l'accès aux données définies dans l'étape de cartographie des besoins afin de définir un projet de collecte adapté intégrant l'ensemble des sources de données (données internes, accessibles sur internet via webscraping ou collectées)
- Définir un plan d'échantillonnage, un mode de collecte et des procédures adaptées afin de se prémunir de tout biais ou erreurs pendant le processus de collecte (par enquêtes ou à l'aide de procédures automatisées)
- Déployer la procédure de collecte et les programmes en réalisant des tests afin de rendre accessible l'accès aux données utiles aux parties prenantes de l'organisation
- Définir des règles de gestion de nettoyage des bases de données (formatage, suppression des doublons...) afin de garantir la traçabilité et la reproductibilité de l'ensemble des opérations réalisées
- Prétraiter les données d'enquêtes (analyse des non-répondants, repondération...) afin d'en établir des éléments statistiques utiles à la prise de décision
- Organiser la base de données en s'appuyant sur des règles de gestion de structuration des différentes bases de données en vue de mettre à disposition les données
- Communiquer les solutions concernant les accès, la documentation sur les données aux équipes métiers afin de faciliter leur compréhension et leurs usages potentiellement dans un contexte multiculturel et/ou international
- Sélectionner les variables permettant de répondre à la problématique donnée en les catégorisant par type de variable (variable d'intérêt, variable explicative, modificateur d'effet, facteur de confusion, variable de contexte) afin de mettre en œuvre le modèle pertinent.
- Explorer des données non étiquetées de manière automatique en mettant en œuvre des modèles ou algorithmes d'apprentissage non supervisé afin de réaliser la segmentation de données, la réduction de dimensionnalité, les détections d'anomalies, ou encore la génération de caractéristiques latentes.
- Choisir la stratégie d'imputation des valeurs manquantes à partir des caractéristiques des données en s'appuyant sur différentes méthodes afin de proposer les résultats le plus fiables possibles
- Modéliser une variable à expliquer en 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 prédire des

valeurs futures.

- Estimer de manière causale l'hétérogénéité des effets d'un traitement (politique publique, campagne marketing...) en mettant en oeuvre les techniques économétriques adaptées (variables instrumentales, score de propension, double différence, machine learning) et en mobilisant les principaux concepts de l'économie (micro et macro-économie) afin de corriger l'endogénéité des variables explicatives
- Modéliser une variable en utilisant les modèles supervisés non paramétriques et en s'appuyant sur les méthodes (arbres de décision, forêts aléatoires, méthodes à noyaux, machines à vecteur de support avec noyaux non linéaires) afin de prédire un comportement.
- Réaliser une analyse prédictive en utilisant les réseaux de neurones profonds (deep learning) afin d'extraire les caractéristiques des données dans les situations complexes (données massives, analyse textuelles, traitement de l'image...)
- Adapter le modèle aux données déséquilibrées pour améliorer la précision des prédictions, réduire le biais, améliorer la compréhension du modèle.
- Évaluer les performances des algorithmes des modèles d'apprentissage selon différents critères (scores, temps d'entraînement GPU, etc.) en adaptant les paramètres afin de choisir le modèle le plus performant pour la problématique métier.
- Traiter les données textuelles avec les outils du NLP (traitement automatique du langage nature) afin de réaliser la classification de texte, l'extraction d'informations, et l'analyse de sentiments permettant la compréhension et la prise de décision.
- Optimiser les temps de traitement en choisissant les solutions informatiques les plus adaptées à la volumétrie des données et à la complexité des données (traitement données massives, cloud computing, calcul distribué) afin de gagner en efficacité opérationnelle et réduire les coûts énergétiques.
- Modéliser les évolutions d'une caractéristique, en fonction du temps et à partir de l'analyse des séries chronologiques (en mettant en oeuvre des modèles ARIMA, SARIMA, Holt-Winters, Réseaux de neurones récurrents) afin de faire des prévisions de valeurs futures et d'agir sur certains déterminants
- Modéliser la probabilité de la survenue d'un événement en analysant des données de survie (durée) permettant la compréhension et la prise de décision.
- Fournir les éléments qualitatifs et quantitatifs rendant les résultats des modèles intelligibles et interprétables afin d'améliorer la confiance et l'acceptabilité des modèles, et limiter le risque de biais et de discrimination en compte du contexte multiculturel et/ou international.
- Réaliser des analyses descriptives et exploratoires en mettant en oeuvre les procédures statistiques (logiciels métier, langage informatique...) afin de décrire et qualifier les différentes données
- Produire des indicateurs répondant aux problématiques métiers en précisant leurs propriétés métrologiques et leurs limites d'interprétation afin d'accompagner la prise de décisions et la prise en compte des contraintes environnementales et du développement durable.
- Représenter visuellement les données de manière à les rendre plus compréhensibles et accessibles.
- Construire avec les équipes métiers une maquette des informations qui seront produites et/ou des tableaux de bords afin d'accompagner la prise de décision
- 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.
- Réaliser la présentation orale d'une démarche de modélisation à un client interne/externe afin de partager les résultats et faciliter la prise de décision de l'interlocuteur et tenant compte d'un contexte multiculturel et/ou international
- Rédiger une note méthodologique explicitant les algorithmes utilisés afin de convaincre de la justesse et de la fiabilité du travail engagé.
- Participer aux ateliers d'expression des besoins internes ou aux réunions de services en s'appuyant sur sa capacité d'écoute active et de questionnement afin de collecter des éléments d'information pour nourrir sa mission et favoriser la stratégie d'utilisation de la data dans l'entreprise.
- Traduire de manière analytique les problématiques métiers afin de proposer des solutions data adaptées éventuellement créatives ou innovantes dans l'organisation au service de l'entrepreneuriat ou l'entreprenariat ou de la recherche
- Rédiger un cahier des charges adapté à la complexité des projets reprenant les éléments de la commande (objectifs et contraintes), les solutions proposées, les ressources, le planning, les livrables afin de formaliser

clairement la demande et d'assurer le bon suivi du projet

- Accompagner les projets « Data » en animant les différentes phases et en pilotant les actions de gestion de l'informations (informations techniques, ordre du jour, compte-rendu, synthèse) en s'appuyant sur les outils collaboratifs afin de communiquer sur les points d'avancement auprès des parties prenantes (groupes techniques, comité de pilotage, instances légitimes, ...)
- Déployer une stratégie de collecte, de stockage et de diffusion de la donnée qui respecte le cadre RGPD et qui s'appuie sur les recommandations internes de son DPO (délégué à la protection des données) afin de sécuriser la démarche
- Identifier les biais des algorithmes engagés par sa propre action afin de les corriger et ainsi de prévenir l'automatisation (invisible) de la discrimination produite.
- Réaliser une veille technologique et scientifique dans le domaine de la data science afin d'améliorer les solutions proposées pour l'entreprise et anticiper les risques juridiques, environnementaux et éthiques
- Réaliser une veille technologique dans le domaine de la data science en s'appuyant notamment sur les outils I.A. afin d'améliorer les solutions proposées pour l'entreprise et gagner en performance dans son activité et inscrire son action dans l'amélioration continue
- Diminuer l'impact carbone du traitement des données en sélectionnant les solutions de stockage et de traitement les moins énergivores, en choisissant les algorithmes et en diminuant la dimension des données sous contraintes d'une qualité satisfaisante des modèles.

OPTION 1 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine du marketing :

- Collecter les données de vente et de CRM (les données de comportement des clients), les données de réseaux sociaux, etc. en mobilisant les logiciels métiers et les outils du web scraping et web mining afin de les traiter et analyser pour les mettre à disposition des équipes marketing
- Développer des modèles d'analyse prédictive pour la segmentation des clients s'appuyant notamment sur les techniques de scoring afin de permettre la prévision des ventes ou l'optimisation des campagnes marketing
- Automatiser le traitement des flux d'informations à destination des prospects et clients en utilisant les outils du deep learning pour permettre à l'entreprise de mener des opérations de marketing digital
- Produire les informations qualifiant les algorithmes de façon à pouvoir apprécier les aspects discriminatoires de ceux-ci et les réguler le cas échéant

OPTION 2 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la biostatistique :

- Collecter les données (cliniques, épidémiologiques, suivi de cohortes etc.) afin de les traiter et analyser pour les commanditaires du domaine de la santé (laboratoire pharmaceutique, laboratoire de recherche, institution de santé publique)
- Réaliser des analyses d'association dans le champ de la santé en utilisant les outils statistiques afin de produire et d'interpréter des résultats ou des analyses permettant des analyses causales ou prédictives
- identifier les facteurs de risques et estimer des fractions attribuables en population afin de quantifier les effets d'un traitement ou d'une exposition en population et déterminer des leviers d'actions sur le plan sanitaire
- Gérer et analyser des données omiques (caractéristiques du génome) afin d'établir des associations entre un trait ou un phénotype et des caractéristiques génétiques

OPTION 3 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la gestion des risques :

- Analyser les données bancaires telles que les données de transactions, les données de crédit, les données de comportement des clients en mobilisant les outils de scoring ou de modélisation statistique afin de produire des éléments d'aide à la décision appuyant les équipes métiers dans leurs missions (ouverture de crédit, opérations frauduleuse, proposition commerciale d'autres produits bancaires...)
- Participer à la réalisation d'une analyse de risque de l'ensemble du dispositif bancaire en utilisant des modèles statistiques définis par le législateur avec d'autres modèles d'intelligence artificielle afin de sécuriser le processus
- Développer des modèles d'analyse s'appuyant sur les techniques de modélisation statistique, de machine Learning et d'intelligence artificielle spécifique à la gestion des risques bancaires afin de traiter des données volumineuses et complexes
- Développer des modèles d'analyse utilisés en ingénierie financière et finance de marché permettant d'estimer le prix de produits financiers complexes

OPTION 4 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine des entreprises industrielles (filière génie statistique) :

- Concevoir un dispositif de suivi en temps réel de l'activité en implémentant une solution technique et/ou logicielle et en collectant les données provenant de capteurs et objets connectés afin de déployer un processus d'efficacité opérationnelle en optimisant les consommations énergétiques tout au long du processus de production.
- Modéliser la fiabilité des systèmes au cours du temps de façon à organiser les opérations de maintenance
- Concevoir et/ou analyser un système soumis à des événements extrêmes de façon à s'en prévenir
- Qualifier le niveau de performance des processus de production afin de mettre en place des actions correctives notamment à partir des constats de gaspillages et de pertes pendant le processus de production.

OPTION 5 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de la modélisation économique et santé :

- Collecter les données médico-administratives susceptibles de répondre la demande en termes descriptifs et de modélisation afin de mettre à disposition des éléments utiles à la prise de décision
- Analyser les données grâce aux outils du machine learning et du deep learning et aux méthodes et outils de l'économétrie afin de mesurer l'impact des actions engagées
- Estimer les effets d'un programme afin d'identifier les leviers d'action des mesures de prévention sur le plan sanitaire et économique
- Bâtir une estimation des coûts économiques et sociaux d'expositions environnementales, du recours ou du non-recours à des traitements en santé afin d'optimiser l'efficacité économique des programmes en santé

OPTION 6 : spécialisation du data scientist dans un domaine d'activité, activités et compétences distinctives du domaine de l'ingénierie des données :

- Concevoir et développer une application multi-tiers en Python (ou autre langage informatique) en utilisant plusieurs services externes.
- Définir une stratégie Big Data en collaboration avec la direction SI de l'entreprise et en sélectionnant les plateformes et systèmes logiciels adaptés pour permettre la collecte et le traitement des données, en particulier les données massives et complexes permettant la mise en oeuvre de la stratégie de l'entreprise.
- Structurer une architecture de stockage de données (data warehouse, data lake...) tout en respectant la politique de sécurité des données afin de répondre aux besoins data des équipes métiers
- Automatiser des processus de déploiement, de tests et de maintenance curative et préventive afin de s'assurer de la fiabilité de la solution et de maintenir un haut niveau d'efficacité de la solution proposée

Blocs de compétences

RNCP40694BC01 — Concevoir et mettre en œuvre des projets de collecte d'informations en utilisant les outils informatiques associés, les techniques de production statistiques (base de données, enquête, données non structurées d'internet)

LISTE DE COMPÉTENCES

Cartographier les besoins data en mettant en œuvre une démarche scientifique (ateliers d'expression des besoins, états des lieux du SI, enquête auprès des équipes métiers, ...) pour répondre aux enjeux stratégiques et organisationnels du commanditaire et en prenant en compte les aspects légaux, éthiques et multiculturels

Modéliser l'application permettant l'accès aux données définies dans l'étape de cartographie des besoins afin de définir un projet de collecte adapté intégrant l'ensemble des sources de données (données internes, accessibles sur internet via webscraping ou collectées)

Définir un plan d'échantillonnage, un mode de collecte et des procédures adaptées afin de se prémunir de tout biais ou erreurs pendant le processus de collecte (par enquêtes ou à l'aide de procédures automatisées)

Déployer la procédure de collecte et les programmes en réalisant des tests afin de rendre accessible l'accès aux données utiles aux parties prenantes de l'organisation

Définir des règles de gestion de nettoyage des bases de données (formatage, suppression des doublons...) afin de garantir la traçabilité et la reproductibilité de l'ensemble des opérations réalisées

Prétraiter les données d'enquêtes (analyse des non-répondants, repondération...) afin d'en établir des éléments statistiques utiles à la prise de décision

Organiser la base de données en s'appuyant sur des règles de gestion de structuration des différentes bases de données en vue de mettre à disposition les données

Communiquer les solutions concernant les accès, la documentation sur les données aux équipes métiers afin de faciliter leur compréhension et leurs usages potentiellement dans un contexte multiculturel et/ou international

MODALITÉS D'ÉVALUATION

Projets statistiques et informatiques (rapport et soutenance)

Les élèves réunis en groupe doivent réaliser une étude statistique (projet statistique) ou un logiciel de traitement de données ou concevoir et développer une application multi-tiers qui produise des informations statistiques à partir d'une base de données (projet informatique.)

Contrôle continu : Devoirs sur table

- Contrôle continu de base de données et programmation sql, de programmation orientation objet et documentation du code-Évaluation des aptitudes à mobiliser les méthodes d'échantillonnages, les biais et les erreurs liées aux méthodes de collecte, évaluation de la capacité à démontrer mathématiquement les propriétés des méthodes et en particulier l'incertitude attendue des résultats
- Évaluation des aptitudes à mobiliser les stratégies de correction de la non-réponse et le calage sur marges

RNCP40694BC02 — Réaliser une modélisation ou une analyse prédictive à partir d'un modèle d'apprentissage avec des données structurées ou non structurées

LISTE DE COMPÉTENCES

Sélectionner les variables permettant de répondre à la problématique donnée en les catégorisant par type de variable (variable d'intérêt, variable explicative, modificateur d'effet, facteur de confusion, variable de contexte) afin de mettre en œuvre le modèle pertinent.

Explorer des données non étiquetées de manière automatique en mettant en œuvre des modèles ou algorithmes d'apprentissage non supervisé afin de réaliser la segmentation de données, la réduction de dimensionnalité, les détections d'anomalies, ou encore la génération de caractéristiques latentes.

Choisir la stratégie d'imputation des valeurs manquantes à partir des caractéristiques des données en s'appuyant sur différentes méthodes afin de proposer les résultats le plus fiables possibles

Modéliser une variable à expliquer en 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 prédire des valeurs futures.

Estimer de manière causale l'hétérogénéité des effets d'un traitement (politique publique, campagne marketing...) en mettant en œuvre les techniques économétriques adaptées (variables instrumentales, score de propension, double différence, machine learning) et en mobilisant les principaux concepts de l'économie (micro et macro-économie) afin de corriger l'endogénéité des variables explicatives

Modéliser une variable en utilisant les modèles supervisés non paramétriques et en s'appuyant sur les méthodes (arbres de décision, forêts aléatoires, méthodes à noyaux, machines à vecteur de support avec noyaux non linéaires) afin de prédire un comportement.

Réaliser une analyse prédictive en utilisant les réseaux de neurones profonds (deep learning) afin d'extraire les caractéristiques des données dans les situations complexes (données massives, analyse textuelles, traitement de l'image...)

Adapter le modèle aux données déséquilibrées pour améliorer la précision des prédictions, réduire le biais, améliorer la compréhension du modèle.

Évaluer les performances des algorithmes des modèles d'apprentissage selon différents critères (scores, temps d'entraînement GPU, etc.) en adaptant les paramètres afin de choisir le modèle le plus performant pour la problématique métier.

Traiter les données textuelles avec les outils du NLP (traitement automatique du langage naturel) afin de réaliser la classification de texte, l'extraction d'informations, et l'analyse de sentiments permettant la compréhension et la prise de décision.

Optimiser les temps de traitement en choisissant les solutions informatiques les plus adaptées à la volumétrie des données et à la complexité des données (traitement données massives, cloud computing, calcul distribué) afin de gagner en efficacité opérationnelle et réduire les coûts énergétiques.

Modéliser les évolutions d'une caractéristique, 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 et d'agir sur certains déterminants

Modéliser la probabilité de la survenue d'un événement en analysant des données de survie (durée) permettant la compréhension et la prise de décision.

Fournir les éléments qualitatifs et quantitatifs rendant les résultats des modèles intelligibles et interprétables afin d'améliorer la confiance et l'acceptabilité des modèles, et limiter le risque de biais et de discrimination en compte du contexte multiculturel et/ou international.

MODALITÉS D'ÉVALUATION

Projet d'économie / Projet de modélisation statistique / Stage scientifique / Projet et stage de fin d'études

Les élèves sont évalués sur la partie de leur rapport qui présente le modèle retenu, interprète les variables, explicite les liens logiques et causaux entre les variables et précise les méthodes de sélection des variables qui sont retenues dans le modèle. Des coefficients mesurant les liens entre variables et la capacité du modèle à expliquer un phénomène sont interprétés.

Contrôle continu : Devoir sur table et devoir maison

- Contrôle continu d'apprentissage statistique supervisé et non supervisé
- Contrôle continu de machine learning
- Évaluation des aptitudes à mobiliser les méthodes de modélisation statistique (régression linéaire, GLM), évaluation de la capacité à démontrer mathématiquement les propriétés des estimateurs, évaluation de l'interprétation des résultats à partir de cas d'usage
- Évaluation de la capacité à juger du compromis complexité/performance à partir de cas d'usage
- Étude de cas sur table de mise en œuvre du NLP
- Contrôle de connaissance, examen sur table et TP notés du cours de Big Data.
- Évaluation des connaissances sur les méthodes de séries temporelles, évaluation de la capacité à démontrer mathématiquement les propriétés des estimateurs, évaluation de la mise en œuvre de la méthode de Box-Jenkins à partir de cas d'usage
- Évaluation des connaissances en Données de survie et Etude de cas sur table sur le traitement des données de survie.

RNCP40694BC03 — Mettre en forme les résultats, les interpréter et leur donner du sens, les communiquer à l'écrit comme à l'oral auprès des parties prenantes de l'entreprise

LISTE DE COMPÉTENCES

Réaliser des analyses descriptives et exploratoires en mettant en œuvre les procédures statistiques (logiciels métier, langage informatique...) afin de décrire et qualifier les différentes données

Produire des indicateurs répondant aux problématiques métiers en précisant leurs propriétés métrologiques et leurs limites d'interprétation afin d'accompagner la prise de décisions et la prise en compte des contraintes environnementales et du développement durable.

Représenter visuellement les données de manière à les rendre plus compréhensibles et accessibles.

Construire avec les équipes métiers une maquette des informations qui seront produites et/ou des tableaux de bords afin d'accompagner la prise de décision

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.

Réaliser la présentation orale d'une démarche de modélisation à un client interne/externe afin de partager les résultats et faciliter la prise de décision de l'interlocuteur et tenant compte d'un contexte multiculturel et/ou international

Rédiger une note méthodologique explicitant les algorithmes utilisés afin de convaincre de la justesse et de la fiabilité du travail engagé.

MODALITÉS D'ÉVALUATION

Projets statistiques (note de synthèse, rapport des projets, soutenance)

Les élèves réunis par groupes doivent exploiter une base de données pour produire une étude statistique répondant à une problématique donnée.

Les élèves sont évalués sur la réalisation d'un poster scientifique qui doit présenter en une page de façon visuelle les principaux résultats de leur projet (distribution de la variable d'intérêt, liens remarquables entre variables, résultats des modélisations ou prédiction, discussion des résultats).

Lors de la soutenance les élèves sont évalués sur leur capacité à répondre de façon pertinente et pédagogique aux questions (y compris en anglais) d'explication et d'approfondissement de leurs travaux :

- Clarté du discours
- Niveau de langage en anglais compréhensible et vocabulaire adaptée

Définition des concepts de modélisation conforme

Contrôle continu

- Contrôle continu de micro et macro-économie
- Contrôle continu d'économétrie

- Contrôle continu des cours de statistique inférentielle, d'introduction aux tests statistiques, à travers des évaluations individualisées.

- Évaluation des connaissances en data visualisation

Stage scientifique et de fin d'études (rapport et soutenance)

Stage scientifique et projet de fin d'études : les élèves sont évalués sur la base d'un travail individuel de mise en œuvre des méthodes statistiques pour répondre à une commande d'un client.

Lors de la soutenance, les élèves sont évalués sur leur capacité à répondre de façon pertinente et pédagogique aux questions (y compris en anglais) d'explication et d'approfondissement de leurs travaux :

- Clarté du discours
- Niveau de langage en anglais compréhensible et vocabulaire adaptée
- Définition des concepts de modélisation conforme

RNCP40694BC04 — Participer aux projets de l'entreprise en valorisant et optimisant l'apport des sciences de la donnée dans le respect des cadres déontologiques, éthiques et de développement durable

LISTE DE COMPÉTENCES

Participer aux ateliers d'expression des besoins internes ou aux réunions de services en s'appuyant sur sa capacité d'écoute active et de questionnement afin de collecter des éléments d'information pour nourrir sa mission et favoriser la stratégie d'utilisation de la data dans l'entreprise.

Traduire de manière analytique les problématiques métiers afin de proposer des solutions data adaptées éventuellement créatives ou innovantes dans l'organisation au service de l'intrapreneuriat ou l'entrepreneuriat ou de la recherche

Rédiger un cahier des charges adapté à la complexité des projets reprenant les éléments de la commande (objectifs et contraintes), les solutions proposées, les ressources, le planning, les livrables afin de formaliser clairement la demande et d'assurer le bon suivi du projet

Accompagner les projets « Data » en animant les différentes phases et en pilotant les actions de gestion de l'information (informations techniques, ordre du jour, compte-rendu, synthèse) en s'appuyant sur les outils collaboratifs afin de communiquer sur les points d'avancement auprès des parties prenantes (groupes techniques, comité de pilotage, instances légitimes, ...)

Déployer une stratégie de collecte, de stockage et de diffusion de la donnée qui respecte le cadre RGPD et qui s'appuie sur les recommandations internes de son DPO (délégué à la protection des données) afin de sécuriser la démarche

Identifier les biais des algorithmes engagés par sa propre action afin de les corriger et ainsi de prévenir l'automatisation (invisible) de la discrimination produite.

Réaliser une veille technologique et scientifique dans le domaine de la data science afin d'améliorer les solutions proposées pour l'entreprise et anticiper les risques juridiques, environnementaux et éthiques

Réaliser une veille technologique dans le domaine de la data science en s'appuyant notamment sur les outils I.A. afin d'améliorer les solutions proposées pour l'entreprise, gagner en performance dans son activité et inscrire son action dans l'amélioration continue

Diminuer l'impact carbone du traitement des données en sélectionnant les solutions de stockage et de traitement les moins énergivores, en choisissant les algorithmes et en diminuant la dimension des données sous contrainte d'une qualité satisfaisante des modèles.

MODALITÉS D'ÉVALUATION

Rapport des stages scientifiques et de fin d'études

Au travers de l'analyse du rapport de stage, le candidat démontre une analyse contextualisée de la situation de l'entreprise et des enjeux de la donnée en son sein, et fournit une synthèse des éléments liés à son intégration. D'autres éléments illustrant son intégration sont transmis par les tuteurs interne à l'entreprise (compte rendu d'entretien, mail récapitulatif des actions d'intégration...)

Projets statistiques, économiques et informatiques (rapports et soutenances)

Travail d'étude statistique en groupes pour répondre aux besoins d'un « client » professionnel que l'on désigne tuteur du groupe.

Contrôle continu (devoir sur table)

- Évaluation des connaissances de Droit des données et d'Éthique de l'ingénieur et d'optimisation algorithmique
- Évaluation des connaissances au sein du cours de Développement durable et des cours d'Apprentissage supervisé, d'Optimisation algorithmique

Projet d'économie / Projet méthodologique / Projet et stage de fin d'études

Les élèves doivent réaliser une revue de la littérature et l'utiliser dans le cadre de leurs travaux d'études. Ils sont questionnés sur l'apport ces points lors de la soutenance du projet de fin d'études. Il présente les éléments suivants :

- Définition du périmètre de la revue, identification des sujets spécifiques couverts, y compris les débats éthiques
- Applications de la data science dans le domaine étudié et les enjeux éthiques associés
- Identification et qualification des sources d'information : qualité, fiabilité, pertinence, exhaustivité des connaissances disponibles

Synthèse critique des informations sélectionnées (points importants, tendances, risques notamment éthiques)

Projet informatique / Stage de fin d'études

Les élèves présentent une liste d'outils et de plateforme utile à la mise en œuvre d'une veille technologique dans leur projet (outils récent). Un tableau synthèse des différentes sources de veille est présenté.

RNCP40694BC05 — Mener des projets de collecte et de modélisation des données en appui à la connaissance client et au marketing (bloc optionnel)

LISTE DE COMPÉTENCES

Collecter les données de vente et de CRM (les données de comportement des clients), les données de réseaux sociaux, etc. en mobilisant les logiciels métiers et les outils du web scraping et web mining afin de les traiter et analyser pour les mettre à disposition des équipes marketing

Développer des modèles d'analyse prédictive pour la segmentation des clients s'appuyant notamment sur les techniques de scoring afin de permettre la prévision des ventes ou l'optimisation des campagnes marketing

Automatiser le traitement des flux d'informations à destination des prospects et clients en utilisant les outils du deep learning pour permettre à l'entreprise de mener des opérations de marketing digital

Produire les informations qualifiant les algorithmes de façon à pouvoir apprécier les aspects discriminatoires de ceux-ci et les réguler le cas échéant

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

Les connaissances, associées à des résolutions de problèmes, sont évaluées à travers les enseignements : Marketing expérientiel et digital, Gestion de la relation client, Méthodes de scoring, Pricing, Économétrie des données de panel, Modèles à équations structurelles, Modèles de régression bayésienne, Modélisation avancée des choix discrets.

Un exercice de simulation de stratégie tarifaire complète ces enseignements.

Projet et stage de fin d'études

L'élève est évalué sur ses connaissances des théories du comportement du consommateur, l'adéquation des choix techniques mis en œuvre pour assurer la prédiction automatisée à partir des informations collectées sur un prospect ou un client. Durant la soutenance, les élèves sont évalués sur l'interprétation et la discussion des résultats des modèles utilisés.

Les différents biais liés aux données et à leur traitement (sélection, déséquilibre en particulier) ont été estimés.

RNCP40694BC06 — Mener des projets de collecte et de modélisation de données biostatistiques en tenant compte des spécificités des données de santé (bloc optionnel)

LISTE DE COMPÉTENCES

Collecter les données (cliniques, épidémiologiques, suivi de cohortes etc.) afin de les traiter et analyser pour les commanditaires du domaine de la santé (laboratoire pharmaceutique, laboratoire de recherche, institution de santé publique)

Réaliser des analyses d'association dans le champ de la santé en utilisant les outils statistiques afin de produire et d'interpréter des résultats ou des analyses permettant des analyses causales ou prédictives

identifier les facteurs de risques et estimer des fractions attribuables en population afin de quantifier les effets d'un traitement ou d'une exposition en population et déterminer des leviers d'actions sur le plan sanitaire

Gérer et analyser des données omiques (caractéristiques du génome) afin d'établir des associations entre un trait ou un phénotype et des caractéristiques génétiques

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

- Évaluation des connaissances en Conception de questionnaire et en Épidémiologie
- Évaluation des connaissances en Inférence causale et Épidémiologie
- Évaluation des connaissances en Analyse de données omiques

Projet d'essai clinique

L'élève est évalué sur :

- Sa maîtrise des fondements de statistique théorique en particulier les processus stochastiques, et la statistique bayésienne

Sa capacité à mettre en évidence de façon argumentée et pertinente les associations entre différents facteurs (facteurs d'exposition, caractéristiques d'état du patient, facteurs comportementaux, modificateurs d'effets et médiations...)

Projet et stage de fin d'études

Le jury professionnel évalue l'élève sur la base d'un rapport et d'une soutenance orale sur les critères suivants :

- Délimitation du périmètre de l'étude correspond à la problématique métier
- Le choix des méthodes statistique est argumenté et pertinent
- Les solutions statistiques proposées sont resituées dans les contraintes de l'organisation (délai, coût, informations disponibles, infrastructure informatiques, législation, organisation interne)
- Le choix des paramètres des algorithmes est justifié

Les résultats sont interprétés et resitués dans la logique des métiers.

RNCP40694BC07 — Mener des projets de collecte et de modélisation de données bancaires dans un objectif de gestion des risques (bloc optionnel)

LISTE DE COMPÉTENCES

Analyser les données bancaires telles que les données de transactions, les données de crédit, les données de comportement des clients en mobilisant les outils de scoring ou de modélisation statistique afin de produire des éléments d'aide à la décision appuyant les équipes métiers dans leurs missions (ouverture de crédit, opérations frauduleuse, proposition commerciale d'autres produits bancaires...)

Participer à la réalisation d'une analyse de risque de l'ensemble du dispositif bancaire en utilisant des modèles statistiques définis par le législateur avec d'autres modèles d'intelligence artificielle afin de sécuriser le processus

Développer des modèles d'analyse s'appuyant sur les techniques de modélisation statistique, de machine Learning et d'intelligence artificielle spécifique à la gestion des risques bancaires afin de traiter des données volumineuses et complexes

Développer des modèles d'analyse utilisés en ingénierie financière et finance de marché permettant d'estimer le prix de produits financiers complexes

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

Les connaissances, associées à des résolutions de problèmes, sont évaluées à travers les enseignements : Gestion des risques bancaires, Techniques et méthodes de scoring, Gestion de la liquidité des actifs, Théorie des valeurs extrêmes, Statistique des risques extrêmes, Gestion des risques multiples, Séries temporelles avancées, Investissement socialement responsable, Théorie d'Évaluation des Actifs, Calcul stochastique, Calibration de processus stochastiques, Modèles de courbes de taux.

Projet scoring

Les travaux des élèves sont évalués sur la connaissance approfondie des algorithmes de scoring et sur l'interprétation des résultats.

- Délimitation du périmètre de l'étude correspond à la problématique métier
- Le choix des méthodes statistique est argumenté et pertinent
- Le choix des paramètres de l'algorithme est justifié
- Les résultats sont interprétés et restitués dans la logique des métiers.

Projet et stage de fin d'études

Le jury professionnel évalue l'élève sur la base d'un rapport et d'une soutenance orale sur les critères suivants :

- Délimitation du périmètre de l'étude correspond à la problématique métier
- Le choix des méthodes statistique est argumenté et pertinent
- Le choix des paramètres de l'algorithme est justifié

Les résultats sont interprétés et resitués dans la logique des métiers.

RNCP40694BC08 — Mener des projets de suivi et de modélisation des processus industriels à des fins de maintenance, d'optimisation et de prévention (bloc optionnel)

LISTE DE COMPÉTENCES

Concevoir un dispositif de suivi en temps réel de l'activité en implémentant une solution technique et/ou logicielle et

en collectant les données provenant de capteurs et objets connectés afin de déployer un processus d'efficacité opérationnelle en optimisant les consommations énergétiques tout au long du processus de production.
Modéliser la fiabilité des systèmes au cours du temps de façon à organiser les opérations de maintenance
Concevoir et/ou analyser un système soumis à des événements extrêmes de façon à s'en prévenir
Qualifier le niveau de performance des processus de production afin de mettre en place des actions correctives notamment à partir des constats de gaspillages et de pertes pendant le processus de production.

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

Évaluation des connaissances en Modélisation non-linéaire et Filtrage

Évaluation des connaissances en Statistique des processus et Théorie des valeurs extrêmes

Stage de fin d'études

Le jury composé de professionnel évalue l'élève sur sa compréhension et sa prise en compte des contraintes et de l'environnement (l'implémentation de la solution). L'élève justifie le choix des modélisations statistiques des phénomènes physiques en jeu et sait en interpréter les résultats et en apprécier la robustesse.

Le jury évalue la compréhension des principes fondamentaux des processus industriels, tels que la production, la logistique et la maintenance.

Il évalue notamment le candidat sur l'examen critique des conditions de production et de suivi industriel.

RNCP40694BC09 — Mener des projets de collecte et de modélisation de données médico-économiques à des fins d'évaluation de coût et d'impact (bloc optionnel)

LISTE DE COMPÉTENCES

Collecter les données médico-administratives susceptibles de répondre la demande en termes descriptifs et de modélisation afin de mettre à disposition des éléments utiles à la prise de décision

Analyser les données grâce aux outils du machine learning et du deep learning et aux méthodes et outils de l'économétrie afin de mesurer l'impact des actions engagées

Estimer les effets d'un programme afin d'identifier les leviers d'action des mesures de prévention sur le plan sanitaire et économique

Bâtir une estimation des coûts économiques et sociaux d'expositions environnementales, du recours ou du non-recours à des traitements en santé afin d'optimiser l'efficacité économique des programmes en santé

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

- Évaluation des connaissances en Analyse spatiale et de Data visualisation
- Évaluation des connaissances en Méthodes d'évaluation des programmes
- Évaluation des connaissances en Modélisation économique de la santé, Économie urbaine et Évaluation médico-économique

Projet de fin d'étude / Projet méthodologique

Le jury professionnel évalue l'élève sur la base d'un rapport et d'une soutenance orale sur les critères suivants :

- Démonstration de sa compréhension des enjeux
- Les méthodes économétriques adaptées aux données socio-économiques ont été mise en œuvre de façon conforme ;
- La théorie économique a été mobilisée pour interpréter les résultats de modélisation statistique et à éclairer la prise de décisions (économétrie de panel, statistique spatiale).

Réalisation d'une méta analyse

RNCP40694BC10 — Concevoir et mettre en oeuvre la stratégie data d'un système d'information (bloc optionnel)

LISTE DE COMPÉTENCES

Concevoir et développer une application multi-tiers en Python (ou autre langage informatique) en utilisant plusieurs services externes.

Définir une stratégie Big Data en collaboration avec la direction SI de l'entreprise et en sélectionnant les plateformes et systèmes logiciels adaptés pour permettre la collecte et le traitement des données, en particulier les données massives et complexes permettant la mise en œuvre de la stratégie de l'entreprise.

Structurer une architecture de stockage de données (data warehouse, data lake...) tout en respectant la politique de sécurité des données afin de répondre aux besoins data des équipes métiers

Automatiser des processus de déploiement, de tests et de maintenance curative et préventive afin de s'assurer de la

fiabilité de la solution et de maintenir un haut niveau d'efficacité de la solution proposée

MODALITÉS D'ÉVALUATION

Contrôle continu (devoir sur table)

Les connaissances, associées à des résolutions de problèmes, sont évaluées à travers les enseignements: Génie logiciel, Développement web, Dataops, Technologies NoSQL et Cloud, Indexation web, Publication de données respectueuse de la vie privée, Réseaux et systèmes d'exploitation, Systèmes Répartis, Sécurité des données.

Projet et stage de fin d'études / Projet informatique (rapport et soutenance)

Les élèves sont évalués sur la base d'un rapport comprenant l'étude fonctionnelle de l'application, la modélisation UML de l'application, le code informatique, le choix de la structure de la base de données et sa mise en œuvre.

La démonstration du fonctionnement des applications est présentée lors de la soutenance.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Le besoin de traitement de la donnée touche un nombre croissant d'entreprises dans tous les secteurs d'activité. Les spécialistes du traitement de l'information et les data scientists sont recherchés

tout autant dans des start-ups, dans des entreprises ayant digitalisé leurs activités que dans de grands groupes ou les administrations publiques et le service statistique publique pour lesquels l'exploitation de données est stratégique.

Les diplômés de l'école sont

principalement recrutés dans la banque, l'assurance, les sociétés de conseil, l'industrie pharmaceutique, l'industrie classique (EDF, SNCF, Automobile, ...), le secteur de l'énergie ou le commerce et la distribution. Depuis plusieurs années les demandes d'emplois se développent dans d'autres

domaines très divers : la cybersécurité, le sport, les sciences de l'éducation, la génétique, les neurosciences.

TYPE D'EMPLOIS ACCESSIBLES

- Ingénieur statisticien, statisticien, data scientist, biostatisticien, ingénieur R&D (Recherche et Développement)
- Chargé d'études statistiques, économiques ou marketing
- Analyste quantitatif en finance, Risk Manager
- Consultant statisticien
- Ingénieur en systèmes décisionnels (industrie, grande distribution, banque, finance...)
- Responsable du contrôle qualité ou d'études de Fiabilité
- Chef de projet base de données

RÈGLEMENTATIONS D'ACTIVITÉ

Le métier de data scientist n'est pas encadré par une réglementation spécifique. En revanche, un data scientist doit se conformer aux principes du RGPD (Règlement Général sur la Protection des Données) dans l'exercice de son métier.

En effet, le RGPD concerne toutes les entreprises publiques ou privées collectant ou exploitant des données personnelles, et ce, quelle que soit la taille de leur structure et leur domaine d'activité. Il s'adresse à toutes les organisations installées en U.E ou hors U.E, dès lors qu'elles ciblent des résidents européens.

Marché du travail

Indicateurs au niveau des familles professionnelles (FAP) des métiers visés, agrégés au niveau national — France Travail / DARES (open data).

Tension de recrutement (DARES 2024)	Très élevée (5/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 1 à 5 sur 5.
Facteurs principaux	Lien formation-emploi, Manque de main-d'œuvre disponible
Salaire médian (mensuel)	2 604 €
Diplôme le plus fréquent	Diplôme de niveau bac+5 ou plus (doctorat, master ou équivalent) en communication et information (35 % des actifs en emploi)

Familles professionnelles (FAP)	H0X40 — Ouvriers qualifiés du contrôle qualité et de laboratoire · tension 5/5 (Très élevée) H0X90 — Techniciens du contrôle qualité · tension 5/5 (Très élevée) H1X93 — Ingénieurs des méthodes de production, du contrôle qualité · tension 5/5 (Très élevée) L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques · tension 4/5 (Élevée) L5X93 — Chargés d'études socio-économiques · tension 3/5 (Modérée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée) N0X70 — Techniciens d'étude, recherche et développement · tension 5/5 (Très élevée) N1X91 — Chercheurs (sauf industrie et enseignement supérieur) · tension 2/5 (Faible) QAX02 — Experts de la banque · tension 1/5 (Très faible)
Projets de recrutement (BMO 2026)	21 193
Jugés difficiles à recruter	43 %
Emplois saisonniers	5 %
Familles professionnelles (FAP)	H0X40 — Ouvriers qualifiés du contrôle qualité et de laboratoire H0X90 — Techniciens du contrôle qualité H1X93 — Ingénieurs des méthodes de production, du contrôle qualité L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques L5X93 — Chargés d'études socio-économiques M2X93 — Experts et consultants en systèmes d'information N0X70 — Techniciens d'étude, recherche et développement N1X91 — Chercheurs (sauf industrie et enseignement supérieur) QAX02 — Experts de la banque

Voies d'accès

MODALITÉS D'ACCÈS

Formation initiale, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

L'Ensai accueille des élèves non fonctionnaires, dits élèves ingénieurs, admis par concours ou sur titres. Le directeur de l'école fixe chaque année, sur proposition du conseil d'École, le nombre maximal d'élèves ingénieurs à admettre. Les conditions d'admission par concours sont fixées par arrêté du ministre chargé de l'Économie. La liste des élèves admis au concours fait l'objet d'une décision signée par le directeur d'école.

Le directeur de l'école peut admettre sur titres, sur proposition du jury d'admission de la direction des études et dans la limite des places disponibles, les candidats justifiant de titres ou diplômes d'enseignement supérieur requis :

- soit justifier d'un diplôme ou d'une formation validée d'un niveau au moins équivalent à une licence dans une spécialité à dominante mathématique, informatique, économique ou statistique ;
- soit justifier d'un diplôme délivré ou d'une formation validée par les universités étrangères de nature et de niveau comparables aux précédents ;
- soit avoir obtenu le diplôme ou avoir satisfait aux examens de sortie d'une école membre de la Conférence des grandes écoles ;
- soit justifier d'un diplôme universitaire de technologie dans l'un des deux départements suivants : statistique et informatique décisionnelle, informatique.

Les équivalences de niveau sont appréciées par le jury d'admission, après en cas de besoin, avis du comité d'enseignement et de la recherche de l'Ensai.

D'autres élèves, français ou étrangers, sont admis par le directeur de l'École pour y suivre certains cours. L'École ne leur délivre aucun diplôme ; sur leur demande, ils peuvent subir les examens des cours suivis et recevoir une attestation faisant connaître les notes qu'ils ont obtenues. Entrent aussi dans cette catégorie les élèves admis à l'École dans le cadre des accords Erasmus ou recrutés via d'autres programmes.

Admission sur concours

- **Concours Mathématiques**

Le recrutement des candidats se fait sur les Concours communs INP (CCINP statistiques). Il s'adresse ainsi aux élèves de classes préparatoires aux grandes écoles (CPGE), filières MP, MPI, PC et PSI.

- **Concours Économie et sciences sociales (B/L)**

Le recrutement des candidats se fait à partir de la banque inter-ENS. Les épreuves écrites sont communes à celles du concours économie et sciences sociales de l'ENSAE Paris. Les épreuves orales se déroulent à l'ENSAI.

- **Concours Économie et gestion (D2)**

Les épreuves sont communes à celles du concours économie gestion option économique et de gestion du concours de l'ENS Cachan (prépa ENS Cachan D2), auxquelles s'ajoute une épreuve orale spécifique de mathématiques.

Par ailleurs, tous les candidats doivent être en situation régulière au regard du service national.

Admission sur titre

L'ENSAI recrute des élèves ingénieurs sur titres :

- ayant validé leur 2^e année de BUT (Stid ou informatique) ;
- ayant un niveau licence 3 (L3) ou master 1 (M1) dans les spécialités Miashs (Mathématiques informatique appliquées et sciences humaines et sociales), Mathématiques, Statistiques, Sciences économiques ;
- ayant un diplôme délivré par les universités et écoles étrangères de nature et de niveau comparables ;
- ? ou ayant obtenu le diplôme ou ayant satisfait aux examens de sortie d'une école membre de la conférence des grandes écoles.

Conditions d'admission

Admission sur concours : Le candidat s'inscrit au concours adapté à sa filière s'il est élève d'une classe préparatoire aux grandes écoles. Les modalités d'inscription et de déroulement des concours sont accessibles sur le site internet respectif des concours CCINP statistiques, B/L et D2.

Admission sur titres : Les demandes d'admission sur titre se font exclusivement en ligne sur le site de l'ENSAI à partir de novembre pour les candidats scolarisés à l'étranger et janvier pour les candidats scolarisés en France.

- • BUT
- • L3/M1/M2 scolarisés en France
- • Candidats scolarisés à l'étranger dans une institution partenaire de l'Ensaï

COMPOSITION DU JURY — FORMATION INITIALE

Le comité d'enseignement et de la recherche peut siéger en formation de jury de validation, de diplomation et d'admission sur titres dans les conditions fixées par l'article 19 du décret 2010-1670 du 28 décembre 2010.

Le comité d'enseignement et de la recherche est présidé par le directeur de l'ENSAI. Les membres votants pour la validation de diplôme sont :

- cinq membres de droit (ou leur représentant) : le directeur de l'école (Président), le directeur de l'Ensaie, le directeur des études, le responsable du Crest-Ensaï, le chef de la division mobilité et carrières de l'Insee.
- deux représentants du corps des enseignants chercheurs de l'école, ayant été élus dans le cadre des élections au Conseil d'Administration du Groupe ou au Conseil Scientifique du Groupe, ou à défaut ayant été désignés par le Directeur de l'école,
- un représentant des anciens élèves de l'école, désigné par l'association des anciens élèves,
- au moins deux membres experts faisant partie du corps des enseignants chercheurs

COMPOSITION DU JURY — VAE

La composition du jury est définie selon l'article D6412-6 du Code du travail. Ce jury est composé d'au moins 2 personnes et comprend, outre les enseignants-chercheurs qui en constituent la majorité, des personnes compétentes pour apprécier la nature des acquis, notamment professionnels, dont la validation est sollicitée. Les jurys sont composés de façon à concourir à une représentation équilibrée entre les femmes et les hommes. Le jury est présidé par le Directeur de l'ENSAI.

Base légale

Date d'échéance de l'enregistrement 31/08/2027

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
-------	-----------	----------	------------------	--------------

2022	97	0	98 %	100 %
2021	104	0	95 %	100 %

Liens avec d'autres certifications

ANCIENNES VERSIONS

RNCP21374

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

Pour plus d'informations consulter : <https://ensai.fr/2-cursus/cursus-ingenieur/>

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/40694/>