



Programme des enseignements de 3^e année

Filière Data Science & Ingénierie des Données

ANNÉE SCOLAIRE 2021 / 2022

FILIÈRE DATA SCIENCE & INGÉNIERIE DES DONNÉES

ANNÉE SCOLAIRE 2021/2022

DATA ENGINEERING & SCIENCE SPECIALIZATION

2021/2022 ACADEMIC YEAR

Table des matières

Présentation de la filière	4
Descriptifs des enseignements communs.....	7
UE : APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING)	8
APPRENTISSAGE AUTOMATIQUE.....	9
APPRENTISSAGE PROFOND.....	10
TRAITEMENT AUTOMATIQUE DU LANGAGE ET FOUILLE DU WEB	11
UE : COMPLÉMENTS D'APPRENTISSAGE AUTOMATIQUE (ADVANCED MACHINE LEARNING).....	12
APPRENTISSAGE PAR RENFORCEMENT	14
APPRENTISSAGE PROFOND AVANCÉ	15
APPRENTISSAGE À GRANDE ÉCHELLE	16
SYSTÈMES DE RECOMMANDATION	17
REGULARISATION.....	18
TRAITEMENT AUTOMATIQUE DU LANGAGE AVANCÉ	19
UE : PROJETS.....	20
PROJET METHODOLOGIQUE	21
PROJET DE FIN D'ETUDES.....	22
DATA CHALLENGE	23
ANGLAIS.....	24
Descriptifs des enseignements de la filière	25
UE SPECIFIQUES FILIERE INGENIERIE DES DONNEES.....	26
GENIE LOGICIEL	27
DEVELOPPEMENT WEB	29
DATAOPS	30
TECHNOLOGIES NOSQL ET CLOUD	31
INDEXATION WEB	32
PUBLICATION DE DONNEES RESPECTUEUSE DE LA VIE PRIVEE	35
INITIATION A UNIX	36
RESEAUX ET SYSTEMES D'EXPLOITATION	37
SYSTEMES REPARTIS	38
SECURITE DES DONNEES	39
UE : PROJET PROFESSIONNEL ET STAGES	41
DROIT DU TRAVAIL	42

Présentation de la filière

La formation d'ingénieur de l'Ensaï inclut 6 filières de spécialisation. Toutes ces filières forment aux métiers de la Data Science, avec une maîtrise des outils permettant l'extraction, l'analyse et la fouille de données et une capacité à choisir les modalités de traitements des données massives (Big Data) et des techniques d'apprentissage automatique (machine learning). Selon les spécialisations, ces compétences sont spécifiques à un domaine ou transversales. Ces compétences sont renforcées pour la filière Statistique et Ingénierie des Données (SID), qui s'est concentré sur ces thématiques depuis plusieurs années. L'ensemble des filières continue à former aux compétences transversales (soft skills) et à la valorisation des travaux menés dans un contexte professionnel et international. Lors des cours et du projet méthodologique en anglais, les élèves travaillent toutes les compétences linguistiques et communicationnelles et approfondissent leurs connaissances liées au monde de l'entreprise et de la recherche. La séquence de Tronc Commun mêlant enseignements scientifiques, projets et anglais conclut la formation à l'autonomie et la capacité à mettre en œuvre des analyses de données en situation complexe. Un stage de fin d'études est à réaliser à l'issue de la scolarité, qui permet de mettre en œuvre dans un cadre professionnel une démarche scientifique autour d'une problématique en lien avec les enseignements de la filière.

Les enseignements spécifiques de la filière se divisent en deux, les enseignements informatiques et les enseignements liés à l'ingénierie des données.

Le développement de grandes applications informatiques nécessite d'utiliser des méthodes d'aide à la conception. Elles peuvent être purement techniques, comme la maîtrise de patron de conception ou l'architecture multi-tiers ou organisationnel comme les méthodes de gestions de projet, aussi bien pour la spécification d'un besoin que le suivi. L'unité d'enseignement (UE) **développement d'application** apportera aux élèves ces compétences essentielles à de futurs chefs de projet informatique. L'UE de **système et réseaux** quant à elle se concentrera sur la conception, la mise en place et l'administration des systèmes complexes pour traiter des données.

En parallèle, le rôle du statisticien – informaticien est d'analyser des données à l'aide de méthodes statistiques et de l'outil informatique. Nous pouvons définir 4 phases dans le traitement informatique des données : la récupération, le stockage, l'analyse et la visualisation des résultats. L'UE de **Big Data** va rentrer va présenter les 3 dernières phases. Ils apprendront à manipuler de très grands volumes de données, à créer des entrepôts de données et à effectuer une analyse multi-dimensionnelle de ces données. Les cours de **machine learning** leur permettront d'extraire de la connaissance à partir de ces données. Ils découvriront comment inventer de nouvelles technologies de stockage et de gestion des données, dans le cadre du Big Data. Ils aborderont également les problèmes de sécurité associés à la diffusion de données.

Plusieurs projets sont réalisés au cours de l'année afin de mettre en pratique les divers enseignements dispensés au cours de cette année. Ils sont réalisés individuellement ou en groupe, permettant aux étudiants de vivre la réalité d'un développement d'application. Les projets sont de nature très diverse et peuvent être réalisés en partenariat avec des industriels ou des chercheurs. Ils ont tous pour but de mettre les étudiants en situation de statisticien ayant de bonnes compétences en informatique, ils sont aussi l'occasion de mettre en œuvre les compétences générales suivantes :

- Connaître les fondements des architectures distribuées, des réseaux aux systèmes répartis, des architectures logicielles big Data ;
- Avoir une ouverture sur les nouvelles technologies et savoir veiller à leurs évolutions ;
- Maîtriser les outils permettant l'extraction, l'intégration, l'analyse et la fouille de données (datamining);
- Maîtriser les outils permettant de manipuler de grands volumes de données et de créer des entrepôts de données ;
- Capacité à choisir et mettre en place le meilleur modèle de machine learning pour réaliser le traitement souhaité ;
- Capacité à concevoir et implémenter des systèmes de traitement de l'information ;
- Capacité à choisir, mettre en place et administrer la technologie de stockage la plus adaptée à un besoin précis (bases de données relationnelles, multidimensionnelles ou non relationnelles) ;

- Capacité à mobiliser les techniques de sécurisation des données à tous les niveaux, stockage, échange et diffusion ;
- Capacité à concevoir et mettre en œuvre un projet de développement d'applications informatiques (connaissances de base en architecture des systèmes, en réseaux, en sécurité, en méthodes de conduite de projet, méthodes de développement d'application).

Volume horaire					
	Cours	Ateliers	Projets	Total	Crédits
UE1 Machine learning					
Machine learning	18	12		30	2,5
Apprentissage profond	6	6		12	1
Webmining et traitement du langage	9	12		21	1,5
Total	33	30		63	5
UE1 Bis Compléments de Machine learning					
Apprentissage par renforcement	15			15	1
Apprentissage profond avancé	12			12	1
Apprentissage à grande échelle	6	6		12	1
Systèmes de recommandation	6	6		12	1
Régularisation	9	6		15	1
Traitement du langage avancé		12		12	1
Total	48	30		78	6
UE2 Développement d'application et DataOps					
Génie logiciel	24	24		48	3
Développement web	12	15		27	2
Ou Dataops	12	15		27	2
Total	12	15		75	5
UE3 Big data					
Technologies NoSQL et Cloud	18	12		30	2
Indexation web	9	6		15	1
Publication de données respectueuse de la vie privée	9			9	1
Total	36	18		54	4
UE4 Système et réseaux					
Initiation à Unix	9	6		15	0
Réseaux et systèmes d'exploitation	15	6		21	1,5
Systèmes Répartis	15	6		21	1,5
Sécurité des données	9	6		15	1
Total	48	24		72	4
UE Projets					
Projet méthodologique		9	18	27	1,5
Projet de fin d'étude		9	36	45	3,5
Data Challenge		12		12	0
Anglais		30		30	1
Total		60	54	114	6
UE Projet professionnel et stages					
Stage de fin d'études					25
Stage d'application					5
Séminaires et projet professionnels (dont droit du travail)	30			30	0
Total	30			30	30
Sport		30		30	0
TOTAL 2021	207	207	54	516	60

Descriptifs des enseignements communs

UE 1

UE : APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING)

Correspondant de l'UE : Adrien Saumard
 Nombre d'ECTS : 5

Volume horaire de travail élève : env. 125h
 (enseignements + travail personnel)
 Nombre d'heures d'enseignement : 63h

Finalité de l'UE :

L'apprentissage automatique (machine-learning) est un paradigme essentiellement différent des approches statistiques exploratoires (statistiques au sens classique) ou explicatives (économétrie). Il vise un objectif de prédiction dans la continuité des méthodes d'apprentissage statistique supervisé introduites lors des premières années de la formation d'ingénieur. Largement utilisé dans l'ensemble des professions statistiques à l'heure actuelle (les métiers de la *data science*), l'apprentissage statistique est incontournable dans la formation de l'ingénieur statisticien et trouve de nombreuses applications: prédiction des cours basés à partir d'articles de presse en finance, détection de maladie par imagerie médicale en santé, recommandation de produits en marketing, compression d'images ou encore modèles de traitement du langage, toutes ces applications reposent sur les mêmes bases.

Structuration de l'UE :

L'UE se compose de 3 matières : apprentissage statistique (*machine-learning*), apprentissage profond (*deep-learning*) et traitement automatique de la langue et fouille du web (*natural language processing and webmining*). L'ensemble de ces matières permettent de mettre en œuvre les techniques classiques, en développant un esprit critique sur leurs limites (sur-apprentissage, grande dimension, représentativité de l'échantillon) et en utilisant des données non structurées (texte, image...). L'UE est accompagnée de séminaires professionnels sur des sujets émergents dans le champ de l'apprentissage statistique.

Compétences ou acquis d'apprentissage à l'issue de l'UE :

Cette UE permet de maîtriser des méthodes et des outils de l'ingénieur (identification, modélisation et résolution de problèmes même non familiers et incomplètement définis, l'utilisation des approches numériques et des outils informatiques, l'analyse et la conception de systèmes) en développant l'aptitude à étudier et résoudre des problèmes complexes, à concevoir et mettre en œuvre des projets de collecte et d'analyse d'informations et à concevoir et mettre en œuvre des algorithmes prédictifs de *machine-learning*.

Les pré-requis de l'UE :

Modélisation statistique de 2^{ème} année, méthodes d'optimisation et d'algorithmique, aisance en R et Python.

UE Machine Learning

APPRENTISSAGE AUTOMATIQUE

Machine Learning

<i>Enseignant</i>	: François PORTIER (Ensay) et Brigitte GELEIN (Ensay)
<i>Nombre d'ECTS</i>	: 2,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Env. 60h
<i>Répartition des enseignements</i>	: 18h de cours et 12h d'ateliers
<i>Langue d'enseignement</i>	: Anglais (cours) et Français (ateliers)
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: régression, régression logistique, convergence de lois de probabilité, théorème central-limite, algèbre linéaire ; classification par arbre ; programmation avec R ; programmation orientée objet ; optimisation: maximisation d'une fonction, algorithme de Newton

Modalités d'évaluation :

- 1 compte-rendu de TP
- examen final

Acquis d'apprentissage (objectifs) :

Comprendre les différents modèles de l'état de l'art (modèle linéaire: régression et classification, pénalisation, méthodes locales, SVM, forêt) ; Connaître les cas d'usage de ces modèles ; Savoir comparer empiriquement différents modèles pour une tâche donnée ; Savoir implémenter les méthodes étudiées en Python ou R

Principales notions abordées :

Supervised learning; Regression; Classification; Empirical risk minimization; Model evaluation; Cross validation; Functional approximation; Model complexity; Large scale optimization; Stochastic gradient descent; Regularization; RIDGE and LASSO; Support Vector Machine; Kernel trick; Ensemble methods; Aggregation and Boosting; Random forest.

Références bibliographiques :

- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.), 2009
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013), *An introduction to statistical learning*, New York: springer.
- Statistical learning with sparsity: the lasso and generalizations, T Hastie, R Tibshirani, M Wainwright – 2019
Link: <https://web.stanford.edu/~hastie/StatLearnSparsity/>

UE Machine Learning

APPRENTISSAGE PROFOND

Deep Learning

<i>Enseignant</i>	: Romaric GAUDEL (Ensay)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: 6h de cours et 6h d'ateliers.
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python (tensorflow)
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: R, Python, modélisation statistique, apprentissage statistique, optimisation de fonctions

Modalités d'évaluation :

1 Quizz et 1 compte-rendu de TP

Acquis d'apprentissage (objectifs) :

- identifier une tâche particulièrement adaptée pour un réseau de neurones et/ou un réseau de neurones profond
- identifier et implémenter un réseau de neurone approprié pour un tâche d'apprentissage supervisé donnée (architecture, fonction de coût, méthode d'optimisation)
- utiliser et spécialiser un réseau de neurones pré-entraîné

Principales notions abordées :

Les réseaux de neurones profonds sont au cœur d'avancées rapides en traitement d'image et de la langue depuis les années 2010. Ce cours présente ces modèles, leur fonctionnement, ainsi que comment les utiliser.

- Principe des réseaux de neurones
- propriétés des réseaux de neurones simples
- descente de gradient
- réseaux de neurones profonds
- architectures particulières : réseaux à convolution ; réseaux récurrents.

Références bibliographiques :

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

UE Machine Learning

TRAITEMENT AUTOMATIQUE DU LANGAGE ET FOUILLE DU WEB

Natural language processing and webmining

<i>Enseignant</i>	: Guillaume Gravier (Irisa)
<i>Nombre d'ECTS</i>	: 1,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 42h
<i>Répartition des enseignements</i>	: Cours : 9h • Atelier : 12h
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: Support de cours, Supports de TP
<i>Pré-requis</i>	: Programmation avec Python, Apprentissage statistique

Modalités d'évaluation :

Projet

Acquis d'apprentissage (objectifs) :

- collecter des données, extraire de l'information et apparier des sources textuelles
- choisir une méthode de traitement automatique de la langue pour une tâche classique (classification, analyse de sentiment, détection > d'entités...)
- se repérer parmi le foisonnement des modèles d'étude de la langue

Principales notions abordées :

1. What's natural language and its processing
2. The representation of words
3. The representation and classification of documents
4. Language modeling and contextual word embedding
5. Sentence-level tagging (token level tasks)
6. Sequence to sequence models and transformers
7. Overview of standard NLP tasks today

Références bibliographiques :

- Daniel Jurafsky, James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edition, Prentice-Hall, 2009. Draft of the 3rd edition partly available at <https://web.stanford.edu/~jurafsky/slp3>.
- Yoav Goldberg. *Neural Network Methods for Natural Language Processing*. 2017. An earlier draft is freely available online at <http://u.cs.biu.ac.il/~yogo/nnlp.pdf>.
- Kevin Gimpel's lectures (Toyota Technological Institute at Chicago and UChicago) on Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31190-s18/index.html>) and on Advanced Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31210-s19/index.html>).

UE 1b

UE : COMPLÉMENTS D'APPRENTISSAGE AUTOMATIQUE (ADVANCED MACHINE LEARNING)

Correspondant de l'UE
Filières concernées

: Rémi Pépin
Ingénierie des données
Marketing
Génie statistique

Nombre d'ECTS

: 4-6 (cf. ci-dessous)

Volume horaire de travail élève
(enseignements + travail personnel)

: 100-150h (cf. ci-dessous)

Nombre d'heures d'enseignement

: 54h-78h (cf. ci-dessous)

Finalité de l'UE :

L'apprentissage automatique (machine-learning) possède de nombreux cas d'application nécessitant une théorie spécifique. C'est le cas par exemple des applications aux images ou à langue (réseaux de neurones convolutifs, récurrents ou avec mécanisme d'attention) ou aux systèmes autonomes (systèmes de recommandation, apprentissage par renforcement). Des problèmes comme le sur-apprentissage possèdent des solutions provenant de la théorie de la compression du signal (régularisation). Enfin, l'apprentissage sur des données suffisamment massives engendre un arbitrage entre qualité et temps de calcul (apprentissage à grande échelle).

Structuration de l'UE :

L'UE se compose de 6 matières, et seule la filière ingénierie des données la suit dans son intégralité. Le tableau ci-dessous résume les ECTS et horaires par filière :

	Heures de cours	Heures de travail	ECTS	Ingénierie des données	Marketing	Génie statistique
Régularisation	15	30	1	X	X	X
Apprentissage à grande échelle	12	24	1	X		X
Apprentissage profond avancé	12	24	1	X	X	X
Traitement automatique de la langue avancé	12	24	1	X	X	
Systèmes de recommandation	12	24	1	X	X	
Apprentissage par renforcement	15	30	1	X	X	X
Total				78 h cours 156 h perso 6 ECTS	60 h cours 123 h perso 5 ECTS	48h cours 105h perso 4 ECTS

Compétences ou acquis d'apprentissage à l'issue de l'UE :

Cette UE permet de consolider les acquis d'apprentissage de l'UE 1 Apprentissage statistique et de développer des compétences pour résoudre des problèmes spécifiques à des sous-disciplines d'intérêt (traitement de l'image, traitement du signal, traitement de la langue, systèmes autonomes, données massives...).

Les pré-requis de l'UE :

UE 1 Apprentissage statistique, méthodes d'optimisation et d'algorithmique, panorama du big data, aisance en Python.

UE Complément d'apprentissage statistique

APPRENTISSAGE PAR RENFORCEMENT

Re-inforcement learning

<i>Enseignants</i>	: Frédéric Logé (Sonio) Rémi Besson (Sonio)
<i>Filières concernées</i>	: Marketing Ingénierie des données Génie statistique
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 30h
<i>Répartition des enseignements</i>	: 15h de cours / ateliers
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: Apprentissage profond et Chaînes de Markov facilitent la compréhension

Modalités d'évaluation :

1 mini-projet

Acquis d'apprentissage (objectifs) :

- objectif1 : savoir modéliser une situation comme un problème d'apprentissage par renforcement (choix d'une fonction de récompense, d'un espace d'état, d'un espace d'actions...)
- bjectif2 : être capable d'apprendre une politique à partir de données et de l'évaluer
- objectif3 : comprendre les limites de la modélisation par apprentissage par renforcement

Principales notions abordées :

- modèles de bandits
- dilemme exploration-exploitation
- algorithmes de bandits
- regret
- processus de décision markovien (MDP)
- équation de Bellman
- fonction état-valeur
- algorithmes pour MDP

Références bibliographiques :

- Tor Lattimore and Csaba Szepesvári, Bandit Algorithms, 2020 (<https://torlattimore.com/downloads/book/book.pdf>)
- Sutton and Barto, Introduction to Reinforcement Learning, 2020 (<http://incompleteideas.net/book/RLbook2020.pdf>)
- David Silver course on RL: <https://www.davidsilver.uk/teaching/>
- GYM, OpenAI, <https://gym.openai.com/>

UE Complément d'apprentissage statistique

APPRENTISSAGE PROFOND AVANCÉ

Advanced Deep Learning

<i>Enseignant</i>	: Romaric GAUDEL (Ensaï)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: 12h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python (tensorflow)
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: R, Python, algèbre linéaire, optimisation de fonctions, apprentissage profond

Modalités d'évaluation :

1 Quizz et 1 compte-rendu de TP

Acquis d'apprentissage (objectifs) :

- Les réseaux de neurones profonds s'adaptent à de nombreuses tâches. Ce cours donne quelques exemples de ces adaptations.
 - objectif : adapter l'architecture d'un réseau de neurones pour des applications non-triviales (multi-vue, image vers texte, texte vers image, compression, génération d'adversaires, ...)
 - identifier une tâche particulièrement adaptée pour un réseau de neurones et/ou un réseau de neurones profond

Principales notions abordées :

- Quelques exemples d'architectures non-triviales
- Auto-encodeurs
- Réseaux adversariaux
- Réseaux de neurones bayésiens

Références bibliographiques :

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

UE Compléments d'apprentissage statistique

APPRENTISSAGE À GRANDE ÉCHELLE

Large-scale machine learning

<i>Enseignant</i>	: Romaric Gaudel (ENSAI)
<i>Filières concernées</i>	: Génie statistique Ingénierie des données
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: Cours : 6h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: support de cours, fiches de TP
<i>Pré-requis</i>	: panorama du big data, machine-learning, apprentissage statistique, optimisation

Modalités d'évaluation :

2 quizz en contrôle continu, 1 TP noté

Acquis d'apprentissage (objectifs) :

- Le passage à des bases de données à grande échelle modifie certains usages en apprentissage statistiques. Nous en verrons quelques exemples, avec les justifications théoriques sous-jacentes.
 - objectif 1 : être en mesure de choisir des approches appropriées pour un problème « de grande taille » donné, ce qui nécessite de...
 - Objectif 2 : connaître le comportement en termes de taille de stockage et/ou de temps de calcul et/ou de qualité d'estimation des approches présentées.

Principales notions abordées :

- la descente de gradient stochastique
- données et modèles parcimonieux (matrices de rang faible, attributs de Fourier aléatoires, distillation, ...)
- le calcul distribué

Références bibliographiques :

- Introduction to High Performance Computing for Scientists and Engineers, Georg Hager, Gerhard Wellein, CRC Press, 2010
- Introduction to High Performance Scientific Computing, Victor Eijkhout, Edmond Chow, Robert van de Geijn, 2014
- Bekkerman, Ron, Mikhail Bilenko, and John Langford. n.d. *Scaling up Machine Learning: Parallel and Distributed Approaches*.
- The MIT Press. n.d. "Large-Scale Kernel Machines." Accessed September 1, 2020. <https://mitpress.mit.edu/books/large-scale-kernel-machines>.

UE Complément d'apprentissage statistique

SYSTÈMES DE RECOMMANDATION

Recommander systems

<i>Enseignant</i>	: Romaric GAUDEL (Ensay)
<i>Filières concernées</i>	: Marketing Ingénierie des données
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: 6h de cours et 6h d'ateliers.
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python (tensorflow)
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: R, Python, algèbre linéaire, optimisation de fonctions, apprentissage statistique

Modalités d'évaluation :

1 Quizz et 1 compte-rendu de TP

Acquis d'apprentissage (objectifs) :

Les systèmes de recommandation choisissent les options à présenter à des utilisateurs parmi un grand nombre de possibilités. Ils permettent par exemple de recommander la prochaine vidéo à regarder, le prochain morceau à écouter, les photos à montrer... Le cours présentera les modèles utilisés pour construire de tels systèmes.

- Objectif : identifier le modèle de recommandation à appliquer dans un contexte donné
- Objectif : implémenter les modèles de recommandation standards
- Objectif : mesurer la qualité des recommandations d'un modèle

Principales notions abordées :

- Objectif des systèmes de recommandation
- État de l'art des systèmes de recommandation (plus proches voisins, filtrage collaboratif...)
- Évaluation des systèmes de recommandation
- Problèmes rencontrés par les systèmes de recommandation et solutions afférentes (démarrage à froid, compromis exploration-exploitation...)

Références bibliographiques :

- Statistical Methods for Recommender Systems. Deepak K. Agarwal, Bee-Chung Chen. 2016.
- Recommender Systems: The Textbook. Charu C. Aggarwal. Springer, 2016.
- Bandit algorithms for Website optimization. John Myles White. O'Reilly Media.
- Blog et tutoriels de Sebastien Bubeck's : <https://blogs.princeton.edu/imabandit/>
- Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. Sébastien Bubeck, Nicolò Cesa-Bianchi. <https://arxiv.org/abs/1204.5721>
- Bandit Algorithms. Tor Lattimore and Csaba Szepesvári. Cambridge University Press. <https://torlattimore.com/downloads/book/book.pdf>

UE Compléments d'apprentissage statistique

REGULARISATION

Regularization

<i>Enseignant</i>	: Cédric HERZET (Inria) & Clément ELVIRA (Inria)
<i>Filières concernées</i>	Ingénierie des données Marketing Génie statistique
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 30h
<i>Répartition des enseignements</i>	: Cours : 9h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: Supports de cours, Supports de TP, Bibliographie
<i>Pré-requis</i>	: Optimisation, Python, Algèbre

Modalités d'évaluation :

Un examen sur table de 2 heures avec questions de cours et résolution de problèmes, 1 compte-rendu de TP

Acquis d'apprentissage (objectifs) :

- identifier les cas d'apprentissage statistique et les problèmes inverses où la régularisation est utile : comprendre quels sont les motivations et les objectifs de la pénalisation
- connaître les modèles de régularisation les plus courants et savoir quelles caractéristiques de reconstruction ils favorisent choisir une régularisation parmi les méthodes les plus courantes et estimer un modèle régularisé par une méthode de descente de gradient
- connaître et comprendre les différents types de problèmes d'optimisation et les algorithmes qui permettent de les résoudre numériquement

Principales notions abordées :

- Ingrédients principaux des problèmes inverses et d'apprentissage statistique + exemples pratiques
- Motivations et objectifs de la régularisation
- Types de régularisation et fonctions de régularisation couramment rencontrées
- Caractérisation des problèmes pénalisés: existence de solution, unicité, conditions d'optimalité.
- Méthodes numériques de résolution de problèmes d'optimisation
- Conditions théoriques de reconstruction correcte

Références bibliographiques :

- Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press.
- C. Bishop. Pattern recognition and machine learning. Springer-Verlag New York, 2006.
- S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- D. P. Bertsekas. Nonlinear Programming. Athena Scientific, USA, 2003.

UE Compléments d'apprentissage statistique

TRAITEMENT AUTOMATIQUE DU LANGAGE AVANCÉ

Advanced natural language processing

<i>Enseignant</i>	: Guillaume Gravier (Irisa)
<i>Filières concernées</i>	: Marketing Ingénierie des données
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: 12h d'ateliers
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: Supports de TP
<i>Pré-requis</i>	: Programmation avec Python, Apprentissage statistique, Traitement automatique de la langue

Modalités d'évaluation :

Projet

Acquis d'apprentissage (objectifs) :

Entraîner un nouveau modèle de langue

Principales notions abordées :

1. Document preprocessing
2. Applied neural networks
3. Long- and short-term memory models (LSTM)
4. How to apply an embedding
7. Entity recognition as token-sequence labelling

Références bibliographiques :

- Daniel Jurafsky, James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edition, Prentice-Hall, 2009. Draft of the 3rd edition partly available at <https://web.stanford.edu/~jurafsky/slp3>.
- Yoav Goldberg. *Neural Network Methods for Natural Language Processing*. 2017. An earlier draft is freely available online at <http://u.cs.biu.ac.il/~yogo/nlp.pdf>.
- Kevin Gimpel's lectures (Toyota Technological Institute at Chicago and UChicago) on Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31190-s18/index.html>) and on Advanced Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31210-s19/index.html>).

UE Projets

UE : PROJETS

<i>Correspondant de l'UE</i>	: Basile de Loynes
<i>Nombre d'ECTS</i>	: 6
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 120h et 150h
<i>Nombre d'heures d'enseignement</i>	: Suivis réguliers avec les encadrants (114h)

Finalité de l'UE :

Les projets sont l'occasion pour les étudiants de mettre en œuvre leurs connaissances acquises à l'ENSAI sur des cas d'études concrets. Ils visent à mettre en œuvre les outils et connaissances acquises en statistique, en informatique et en économie, dans une démarche de résolution de problèmes concrets type ingénieur.

Les projets se déclinent en deux versions: le projet méthodologique, en langue anglaise, vise à approfondir une thématique centrée autour d'un ou plusieurs articles scientifiques ; le projet de fin d'études (dans toutes les filières sauf biostatistique), plus appliqué, nécessairement sur des données issues d'une collecte, vise à proposer une solution pratique à une problématique générale proposée par une entreprise ou un laboratoire de recherche. À eux deux, ces projets couvrent toute l'étendue d'une démarche de développement: diagnostique d'un problème nouveau, lecture de la littérature scientifique sur le sujet, résolution d'un problème en respectant un compromis entre les règles de l'art d'une part, et les contraintes humaines, financières et techniques de l'autre. Ils permettent par ailleurs aux élèves de mesurer l'utilité de toutes les notions acquises au cours des trois années de formation.

Selon les filières, la réalisation d'un Data Challenge complète ces cas d'études concrets, à travers la réalisation d'un projet sur un temps court et des contraintes spécifiques.

Structuration de l'UE :

Projet méthodologique: approfondissement d'une démarche de revue de littérature en langue anglaise ; constitue la partie théorique de recherche d'information dans une démarche de recherche et développement. Les élèves désireux de développer un projet de recherche académique peuvent privilégier un projet plus conséquent.

Projet de fin d'étude: approfondissement d'une démarche pratique, sachant composer avec des contraintes opposées, entre rigueur scientifique et nécessités pratiques ; constitue la partie implémentation dans une démarche de recherche et développement.

Data Challenge (optionnel, selon les filières) : rassembler sur une période très courte différentes équipes de profils variés afin de collaborer sur un projet.

Compétences ou acquis d'apprentissage à l'issue de l'UE :

Ces projets concluent la formation d'ingénieur de l'Ensaï, et mobilisent un ensemble de compétences de l'ingénieur : capacité à trouver l'information pertinente, à faire une veille scientifique, à prendre en compte les enjeux de l'entreprise, à travailler dans un contexte international, tout en mobilisant des compétences techniques pour résoudre des problèmes complexes, et mener une démarche scientifique.

Les pré-requis de l'UE :

Méthodes de travail des projets de 1^{ère} et 2^{ème} année.

UE Projets

PROJET METHODOLOGIQUE

Methodological project

<i>Enseignant</i>	: Divers intervenants
<i>Filières concernées</i>	: Toutes filières, avec des modalités différentes
<i>Nombre d'ECTS</i>	: 1,5 ou 2,5 selon modalités (sauf filière Biostatistique) 3 (filière Biostatistique)
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 40h et 75 h selon modalités
<i>Répartition des enseignements</i>	: 9h d'ateliers, et suivis réguliers
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Pré-requis</i>	: Connaissances générales en statistiques (1 ^e et 2 ^e années)

Modalités d'évaluation :

Le projet méthodologique consiste en la production d'un article de synthèse sur un sujet de recherche à choisir parmi un catalogue. L'évaluation tient compte de l'article rédigé et de la réalisation d'une soutenance.

Acquis d'apprentissage (objectifs) :

Les objectifs du projet méthodologique, et donc les compétences qui sont renforcées grâce à celui-ci, sont multiples:

- familiarisation avec la forme des productions académiques (articles notamment), en lecture comme en écriture
- capacité à faire une revue de littérature mélangeant ouvrages scientifiques et professionnels
- prise de conscience des enjeux autour de la reproductibilité des résultats de recherche
- communication sur des sujets techniques

Pour les étudiants désirant s'initier à la recherche, et pour l'ensemble des étudiants de la filière Biostatistique, le projet méthodologique prendra la forme de l'étude détaillée, la reproduction et/ou la mise en perspective d'un ou de plusieurs articles, au lieu d'une revue de littérature.

À cela s'ajoute les objectifs spécifiques à la production d'un travail technique en langue anglaise: mise en œuvre d'un projet complexe en langue anglaise, communication écrite et orale, acquisition d'un vocabulaire spécialisé, maîtrise de différents niveaux de langues en terme de style (oral vs. écrit) et de technicité (vulgarisation vs. spécialisation), mise en place de stratégies pour faire face à des difficultés linguistiques.

Principales notions abordées :

Travail de recherche en groupe suivi par un chercheur (env. 5 séances) et un professeur d'anglais (4 séances). Des séances de travail personnel sont réservées dans l'emploi du temps.

UE Projets

PROJET DE FIN D'ETUDES

Methodological project

<i>Enseignant</i>	: Divers intervenants
<i>Filières concernées</i>	Toutes filières sauf Biostatistique
<i>Nombre d'ECTS</i>	: 3,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 60h et 90 h
<i>Répartition des enseignements</i>	: 9h d'ateliers et suivis réguliers
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Pré-requis</i>	: Connaissance de statistiques des trois années de l'ENSAI

Modalités d'évaluation :

Le projet de fin d'études consiste en la production d'une étude statistique de niveau professionnel dans le monde de l'entreprise ou de la recherche, parmi un catalogue de sujet mis à disposition des élèves. Le projet est évalué à travers un rapport et une soutenance.

Les élèves désireux de s'initier à la recherche passant plus de temps sur le projet méthodologique, les attendus du projet de fin d'études sont ajustés pour tenir compte du moindre temps consacré que leurs camarades.

Acquis d'apprentissage (objectifs) :

Les objectifs du projet de fin d'études, et donc les compétences qui sont renforcées grâce à celui-ci, sont multiples:

- mise en situation professionnelle
- capacité à définir une stratégie d'étude en réponse à une demande client
- mobilisation des compétences techniques (statistiques, économiques, informatiques)
- compromis entre rigueur scientifique et contraintes pratiques (limitations financières, logicielles, cognitives, temporelles...)
- travail de groupe
- gestion d'un projet sur le temps long
- communication (écrite, orale) sur des sujets techniques

Principales notions abordées :

Travail autonome en groupe suivi par un professionnel de l'entreprise ou de la recherche (env. 5 séances).

UE Projets

DATA CHALLENGE

Data Challenge

<i>Enseignant</i>	: Divers intervenants industriels (correspondante : Salima El Kolei)
<i>Nombre d'ECTS</i>	: Pas d'attribution d'ECTS
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 2 journées
<i>Répartition des enseignements</i>	: 12h d'ateliers
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Pré-requis</i>	: Méthodes de travail des projets, Compétences statistiques et informatiques de 3ème année

Modalités d'évaluation :

Les élèves participent au data challenge proposé à l'Ensaï ouvert également aux élèves de deuxième année. Il n'y a pas d'évaluation.

Acquis d'apprentissage (objectifs) :

Le data challenge permet de rassembler sur une période très courte différentes équipes de profils variés afin de collaborer sur un projet. Cette expérience se rapproche des conditions réelles dans laquelle évoluent les datascientists au sein des entreprises. Il permet, à partir des mécanismes du jeu, de dynamiser et d'articuler la pédagogie autour d'un besoin concret d'entreprise et d'un événement qui s'achève par une évaluation objective. De nombreux challenges sont proposés autour de la Data ou présentant des problématiques Data importantes.

L'objectif de ce cours est de valoriser les compétences transversales acquises dans ce contexte opérationnel. Les compétences qui sont renforcées grâce à celui-ci sont multiples:

- Comprendre les problèmes à résoudre.
- Travailler en mode projet avec des contraintes.
- S'intégrer et s'adapter dans un contexte pluridisciplinaire. Selon les challenges, les compétences seront mobilisées à géométrie variable.
- S'adapter à la réalité de la Data d'entreprise (données non structurées, manquantes, volumétrie...)
- Communication orale des résultats (pitch...)

Principales notions abordées :

Travail en groupe sur un temps court.

UE Projets

ANGLAIS

English

<i>Enseignant</i>	: Divers intervenants (correspondant : Todd Donahue)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 45h
<i>Répartition des enseignements</i>	: 15h de cours, 15h d'aide au projet de fin d'études
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sous Moodle
<i>Pré-requis</i>	: Aucun

Modalités d'évaluation :

L'examen final prend la forme d'une simulation d'entretien d'embauche. Cet examen oral durera environ 25 minutes, sera noté, et permettra d'évaluer le niveau d'expression orale sur l'échelle CECRL (Cadre européen commun de référence pour les langues). Le CV et la lettre faite pour cet exercice seront évalués et feront partie de la note finale. L'anglais est également évalué à travers le rapport écrit et la soutenance orale du projet de fin d'études. Le niveau acquis apparaîtra sur le Supplément au diplôme. L'objectif de la CTI (Commission des Titres d'Ingénieur) pour tous les élèves ingénieurs est d'atteindre le niveau B2.

Acquis d'apprentissage (objectifs) :

- maîtriser une ou plusieurs langues étrangères
- savoir candidater et réussir un recrutement en langue anglaise
- contextualiser et prendre en compte les enjeux et les besoins de la société
- se connaître, s'auto-évaluer, gérer ses compétences, opérer ses choix professionnels
- s'intégrer et évoluer dans un groupe pour mener à bien un projet dans un contexte international et/ou pluriculturel
- savoir identifier les informations pertinentes, à les évaluer et à les exploiter

Principales notions abordées :

Pour les élèves qui n'ont pas eu un score d'au moins 785 au TOEIC : pendant les 5 premières séances, la plupart des cours seront basés sur la préparation à cet examen. Les ressources informatiques de l'École doivent aussi être mises à profit (pages Moodle, TOEIC Mastery), ainsi que les méthodes disponibles à la bibliothèque. Pour les autres élèves, les cours seront organisés par groupe de niveau et conçus afin de les préparer à affronter le monde professionnel sur le plan international. Les thèmes suivants seront traités : « Leading meetings », « Interviews », « Presentations », « Taking decisions », et « Negotiating deals », et « Cultural and Political Current Events ». Ensuite, les 5 dernières séances seront consacrées au travail de rédaction/correction des rapports faits en anglais dans chaque filière ainsi qu'à la préparation des soutenances orales. Chaque responsable de filière indiquera aux élèves, en début d'année, le projet concerné et les modalités de notation. Les élèves recevront des consignes détaillées avant de démarrer ces cinq séances, afin d'arriver à la première séance avec une première version ou extrait de leur rapport en anglais prêt pour correction et relecture. **Pour tout complément d'information, chaque élève peut consulter le Programme des enseignements : Langues étrangères, distribué au début de l'année académique.**

Références bibliographiques : Définies par chaque intervenant.

Descriptifs des enseignements de la filière

UE Spécifiques filière ID

UE SPECIFIQUES FILIERE INGENIERIE DES DONNEES

<i>Correspondant de l'UE</i>	: Rémi Pépin
<i>Nombre d'ECTS</i>	: 13/14
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: De 25 à 30h par ECTS
<i>Nombre d'heures d'enseignement</i>	: 285h

Finalité des UE :

Ces UE visent à renforcer les compétences informatiques des élèves ingénieurs, pour s'orienter vers des métiers de la Data Science directement en lien avec le génie logiciel, l'ingénierie des données ou l'inclusion de méthodes d'apprentissage statistique dans des architectures big data.

Structuration de l'UE :

La filière SID inclut 3 UE spécifiques : développement d'application, big data et système et réseaux.

Compétences ou acquis d'apprentissage à l'issue de l'UE :

- Connaître les fondements des architectures distribuées, des réseaux aux systèmes répartis, des architectures logicielles big Data ;
- Avoir une ouverture sur les nouvelles technologies et savoir veiller à leurs évolutions ;
- Maîtriser les outils permettant l'extraction, l'intégration, l'analyse et la fouille de données (datamining);
- Maîtriser les outils permettant de manipuler de grands volumes de données et de créer des entrepôts de données ;
- Capacité à choisir, mettre en place et administrer la technologie de stockage la plus adaptée à un besoin précis (bases de données relationnelles, multidimensionnelles ou non relationnelles) ;
- Capacité à mobiliser les techniques de sécurisation des données à tous les niveaux, stockage, échange et diffusion ;
- Capacité à concevoir et mettre en œuvre un projet de développement d'applications informatiques (connaissances de base en architecture des systèmes, en réseaux, en sécurité, en méthodes de conduite de projet, méthodes de développement d'application).

UE 2 – Développement d'application

GENIE LOGICIEL

Software engineering

<i>Enseignants</i>	: Djamel Eddine KHELLADI (Université Rennes 1) & Rémi PEPIN (Ensay)
<i>Nombre d'ECTS</i>	: 3
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 90h
<i>Répartition des enseignements</i>	: • Cours : 24h • Atelier : 24h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Notation UML, connaissance du langage JAVA

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif pédagogique :

L'objectif de ce cours est d'introduire les moyens de concevoir des applications informatiques de qualité (répondant aux besoins, évolutives et faciles à maintenir).

Il s'agit de présenter l'ingénierie dirigée par les modèles en positionnant la conception dans les cycles de développement, et en mettant l'accent sur les enjeux et les pièges à éviter.

Le cours présente une introduction aux modèles de conception classiques, base du génie logiciel autour des technologies objet, en proposant des applications pratiques au cours de travaux pratiques et en étudiant des patrons de conception développés en Java. Cet enseignement vise également à apprendre à développer et déployer un site Web dynamique en Java. Il permet de se familiariser avec les architectures n-tiers et les serveurs d'applications et de bien maîtriser les principaux outils et langages avancés de développement des applications Web/JavaEE.

Contenu de la matière :

I. Le génie logiciel

1. Introduction au génie logiciel et bonne pratique de conception.
2. Architecture logicielle et modèle en couche, Exemple sur GWT.
3. Principaux patrons de conception, principe et mise en œuvre en Java.
4. Le test logiciel et l'ingénierie des langages.
5. L'ingénierie dirigée par les modèles.

II. Programmation Client Serveur (JavaEE)

1. Architectures distribuées et plate-forme JavaEE
 - i. Les Technologies JavaEE et Spring
 - ii. Architecture : composants, services et communications
 - iii. Les problématiques des applications serveurs
2. API et frameworks JavaEE / Spring
3. La persistance avec JPA
 - i. Problématique du "mapping" objet-relationnel
 - ii. Les outils de mapping : JPA, Hibernate, Toplink
 - iii. Le mapping
 - iv. L'entity-manager
 - v. Le langage de requetage

4. Les services web, le cloud.

Modalités d'évaluation :

Un TP noté sera rendu.

Références bibliographiques :

- I. SOMMERVILLE, *Le Génie logiciel*, Addison Wesley-France, 1988
- B. BEIZER, *Software Testing Techniques*, Second Edition, Van Norstrand, 1990
- B.W. BOEHM, *Software Engineering Economics*, Prentice-Hall, 1981
- E. GAMMA, R. HELM, R. JOHNSON, J. VLISSIDES, *Design patterns, catalogue de modèles de conception réutilisables*, Vuibert, 2007

UE 2 – Développement d'application

DEVELOPPEMENT WEB

Software engineering

Enseignant	: Olivier CHANTREL (Orange)
Nombre d'ECTS	: 2
Volume horaire de travail élève (enseignements + travail personnel)	: 50h
Répartition des enseignements	: • Cours : 12h • Atelier : 15h
Langue d'enseignement	: Français
Logiciels	: Sans objet
Documents pédagogiques	: Slides/Manuscrit
Pré-requis	: Sans objet

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif pédagogique :

L'objectif de cet enseignement est d'effectuer un projet de bout en bout. Ce projet commence par une modélisation utilisant les méthodes et techniques vues en génie logiciel et se termine par une implémentation en JavaEE.

Ce cours vise à donner aux étudiants une vision détaillée du web d'aujourd'hui en présentant les technologies historiques du web (html, xml) et les technologies plus récentes du web 2.0 (css, javascript, HTML5, Ajax, Php, MySql).

Contenu de la matière :

- Encadrement en début de projet afin de préciser les besoins et pour déterminer l'architecture générale du programme.
- Encadrement distant sur des questions techniques ponctuelles
- Encadrement technique lors de la phase d'implémentation
- Soutenance des projets
- L'historique du web / XML et ses applications (html, xml, dtd, web services etc.)
- Le web actuel (css, javascript, HTML5, Ajax, MySql, Php)

Modalités d'évaluation :

Les étudiants sont évalués sur la base du rapport d'étude et d'une soutenance devant un jury, incluant une démonstration de leur réalisation.

Références bibliographiques :

- L. Roland, « Structurez vos données avec XML », 2014
- L Van Lancker, « AJAX - Développez pour le Web 2.0 - Entrez dans le code : JavaScript, XML, DOM, XMLHttpRequest2... (2ième édition) », 2015
- C Pierre de Geyer & E Daspect, « PHP5 avancé », 2012

UE 2 – Développement d'application

DATAOPS

DataOps

<i>Enseignants</i>	: Christophe Blefari (Kixel) & Larysa Visengeriyeva (INNOQ)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 50h
<i>Répartition des enseignements</i>	: • Course : 12h • Workshop : 15h
<i>Langue d'enseignement</i>	: English
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Familiar with Python and Unix, Cloud (AWS) basic concepts: S3, EC2, networking, RDS-git

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Learning objectives :

This classe aims to introduce to the DataOps skills set and philosophy. DataOps is a new concept that appeared around 2018. Because companies started to treat data as a product we needed to approach data in a more agile way. In order to simplify data usage for all, engineers need to treat data like an asset. This is DataOps.

During this classe students will see and create self-service data platform through concepts like Infrastructure as Code (IaC — Terraform, Ansible), CI/CD and Software Engineering abstraction.

This lecture will also help you navigate through the whole data tools ecosystem and will give you the decision framework to choose a tool over another.

Main subjects covered :

- Why DataOps, MLOps and DevOps?
- AWS Cloud platform, a reminder
- Infrastructure as Code
- CI/CD concepts
- Docker 101 and how it changed applications deployment
- Kubernetes light introduction
- Data pipelines and data observability

Course evaluation :

Written exam and short projet.

Bibliography references :

- DataOps 101 — DataTalks.Club podcast with Lars Albertsson ; <https://datatalks.club/podcast/s02e11-dataops.html>
- DataOps Manifesto ; <https://dataopsmanifesto.org/fr/>
- Getting DataOps Right — O'Reilly Media, Inc. ; <https://www.oreilly.com/library/view/getting-dataops-right/9781492031765/>

UE 3 – Big Data

TECHNOLOGIES NOSQL ET CLOUD

NoSQL Technologies and cloud

<i>Enseignants</i>	: David GROSS-AMBLARD (IRISA) & Gabriel ANTONIU (INRIA)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 50h
<i>Répartition des enseignements</i>	: • Cours : 18h • Atelier : 12h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Base de données relationnelles

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif pédagogique :

Cet enseignement vise à donner aux étudiants les connaissances de base sur les architectures distribuées spécialisées dans le stockage et le traitement du Big Data. Les hypothèses, concepts and algorithmes seront détaillés pour chacune d'entre elles.

L'objectif du cours est double. Il permettra aux élèves d'avoir une connaissance des systèmes sur Cloud disponibles actuellement et de pointer les directions futures de ces architectures et de savoir s'orienter dans la jungle des technologie NoSql pour trouver la solution qui répondra le plus à leurs besoins. Ces bases de données se distinguent des approches classiques relationnelles. Ces approches abandonnent la représentation matricielle de l'information ainsi que le langage SQL au profit d'une plus grande simplicité, d'une meilleure performance et d'une meilleure scalabilité.

Contenu de la matière :

- Système de fichiers distribué avec HDFS
- Traitement distribué avec Hadoop MapReduce
- Virtualisation et conteneurisation

Modalités d'évaluation :

Examen écrit

Références bibliographiques :

- G. PLOUIN, Cloud computing et SaaS, Editions Dunod
- Le livre blanc du Cloud, du SaaS et des Managed Services pour les partenaires IT et télécoms. Edition 2013
- R. HENNION, H. TOURNIER, E. BOURGEOIS, Cloud computing : Décider - Concevoir - Piloter - Améliorer, Editions Eyrolles, 2012

UE 3 – Big Data

INDEXATION WEB

Web datamining

Enseignant	: Nawfal TACHFINE (aramisauto)
Nombre d'ECTS	: 1
Volume horaire de travail élève (enseignements + travail personnel)	: 30h
Répartition des enseignements	: • Cours : 9h • Atelier : 6h
Langue d'enseignement	: Français
Logiciels	: Sans objet
Documents pédagogiques	: Slides/Manuscrit
Pré-requis	: SQL

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif pédagogique :

A l'issue de ce cours, les élèves devront savoir collecter des informations issues du web, connaître la notion d'Information Retrieval, savoir constituer des corpus, et les organiser à des fins d'analyse exploratoires. Ils devront maîtriser également l'algorithme qui permet de hiérarchiser les pages web (pagerank) et les techniques de classification de documents textuels.

Par ailleurs, ils devront avoir acquis les notions d'opinion mining (classification de textes, analyses de sentiments, évaluation de modèles).

Toutes les applications seront traitées en R.

Contenu de la matière :

Partie 1 – Information Retrieval : Preprocessing, Extraction and PageRank

Mots clés : twitter, R, pagerank, corpus, term-document matrix, Information retrieval, tf-idf, stemming, Regex, kmeans

Partie théorique (3h)

- Information Retrieval
 - o Concepts & Définitions
 - o Term Document Matrix
 - o Tf-idf, Cosine Index, jaccard Index
 - o Stemming
- Web Search : Google
 - o Google et le Page Rank
 - o Pages Jaunes (Notion de tri alpha)
 - o Notion de graphes et de vecteurs propres

Partie pratique (9h)

- TP1 : Introduction à R pour le Web Mining (3h)
 - o Installation de bibliothèques de textmining disponible dans R
 - o Collecter les informations issues du WEB : Twitter, Wikipedia
 - o Pre-processing : Stemmatisation, Lemmatisation,
 - o Parsing HTML, XML,
 - o Tokenization

- Introduction à la term-document matrix
- TP2 : Similarité de documents (Applications aux recherches utilisateurs sur le site pagesjaunes.fr (3h))
 - Indices de similarité : Tf, tf-idf Jaccard, Cosine
 - Distance de Damerau, Distance de jaro
 - Liens entre les recherches, Notion de graphe de recherche
- TP3 : Ordonnement des résultats d'une recherche (3h)
 - PageRank
 - Détecter les mots clés
 - Intro à la classification des docs sur mots clés

Partie 2 – Opinion Mining : Textmining, analyse de sentiments, classification et évaluation des modèles.

Mots clés : Facebook, R, opinion mining, corpus, sentiment analysis, annotation syntaxique.

Partie théorique (4h)

- Introduction
 - Quelles applications dans quels domaines d'activités
- État de l'art (opinion mining, sentiment analysis, affective computing)
 - Quels descripteurs pour quels types de données ?
 - Textuelles
 - Audio
 - Images
 - Sélection automatique de descripteurs (réduction de l'espace de recherche)
 - Quels algorithmes de classification dans quels cas ?
- Constitution du corpus
 - Réflexions générales sur la qualité des données et son impact
 - Annotation manuelle et automatique (schéma d'annotation, calcul d'un score d'agrément inter-annotateur,)
 - Répartition des données dans les classes
- Pre-processing (texte)
 - Quelle granularité pour mes données (mot, phrases, paragraphes)
 - Annotation syntaxique et sémantique (exemples de POS, WordNet-Affect, etc)
- Évaluation
 - Quelles mesures utiliser pour mesurer la qualité d'un modèle (rappel, précision, f-score, ROC, indices de confiance à 0.95)
- Les produits du marché (exemples)
 - Produit de la société TEMIS (cartouche sentiments)
 - Produit de la société Sinequa

Partie pratique (8h)

- TP1 : classification de la valence d'un texte littéraire (critiques de cinéma)
- TP2 : classification de la valence de textes issus de réseaux sociaux (twitter, facebook)
- TP3 : Fusion de modèles (à partir des modèles créés dans le TP2)
- **TP4 (optionnel) :** Constructions de modèles à partir d'indices multimodaux (texte + audio)

Modalités d'évaluation :

Projet par groupe d'élèves.

Références bibliographiques

Les * indiquent les lectures fortement conseillées.

- Web DataMining, Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer (Chapitre 6 à 13) (*)
- Information Retrieval, <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> (chapitres 1-3) (*)
- **package tm in R**, <http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf> (*)
- Infrastructure of Textmining with R, <http://www.jstatsoft.org/v25/i05/paper>

- Webmining plugging in R, <http://cran.r-project.org/web/packages/tm.plugin.webmining/vignettes/ShortIntro.pdf>
- PageRank, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
- Introduction to PageRank, <http://www.stat.berkeley.edu/~vigre/undergrad/reports/christensonNathan.pdf> (*)
- Mining the social web, <https://github.com/ptwobrussell/Mining-the-Social-Web>
- Pang B. and Lee L. (2008). "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval **2**(1-2).
- Dini L. and Mazinni G. (2002). Opinion classification through information extraction. CELI. Turin, Italy
- Cornuéjols A., Miclet L. and Kodratoff Y. (2002). Apprentissage artificiel : Concepts et algorithmes
- Ilieva L. (2004). Combining Pattern Classifiers : Methods and Algorithms (chapitre 1 "Fundamentals of Pattern Recognition", chapitre 4 "Fusion of Label Outputs") (*)

UE 3 – Big Data

PUBLICATION DE DONNEES RESPECTUEUSE DE LA VIE PRIVEE

Privacy-preserving data publishing

<i>Enseignants</i>	: Javier Rojas Balderrama
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 18h
<i>Répartition des enseignements</i>	: 9h de séminaire
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Connaissances de base en gestion de données, en algorithmique, et en probabilités et statistiques & Compétences de base dans un langage de programmation parmi Java, Python, ou R.

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif de la matière :

« Les données personnelles sont le nouveau pétrole d'Internet et la nouvelle monnaie du monde numérique » a déclaré M. Kouneva, commissaire européen à la protection des consommateurs en mars 2009. La valeur de l'analyse massive des données personnelles pour les industriels, les scientifiques et la société en général est largement reconnue aujourd'hui. Cependant, leur caractère personnel et potentiellement sensible est un obstacle majeur à leur partage à grande échelle. L'objectif des modèles et algorithmes de publication de données respectueuse de la vie privée est précisément d'offrir des garanties fortes de respect de la vie privée tout en autorisant un partage de qualité à des fins d'analyse. La tâche est loin d'être triviale comme l'ont démontré plusieurs scandales de ré-identification. L'objectif de ce cours est de présenter aux étudiants les principaux paradigmes et techniques de publication de données respectueuse de la vie privée.

L'accent sera particulièrement mis sur un modèle proéminent aujourd'hui : **la differential privacy**.

Contenu de la matière :

Présentations par divers intervenants extérieurs. Ces présentations vont exposer les problématiques et solutions de traitement et de publication de données individuelle dans leur cadre industriel.

Modalités d'évaluation :

QCM en fin de présentation

Références bibliographiques :

- B.-C. Chen, D. Kifer, K. LeFevre, et A. Machanavajjhala, Privacy-Preserving Data Publishing, Found. Trends databases, vol. 2, no 1-2, p. 1-167, 2009.
- C. Dwork et A. Roth, The Algorithmic Foundations of Differential Privacy, Found. Trends Theor. Comput. Sci., vol. 9, no 3-4, p. 211-407, 2014.
- B. C. M. Fung, K. Wang, R. Chen, et P. S. Yu, Privacy-preserving data publishing : A survey of recent developments, ACM Comput. Surv., vol. 42, no 4, p. 14:1-14:53, 2010.

UE 4 – Système et Réseaux

INITIATION A UNIX

Networks and Systems

<i>Enseignant</i>	: Romain BRECHET (Orange Consulting)
<i>Nombre d'ECTS</i>	: 0
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 21h
<i>Répartition des enseignements</i>	: • Cours : 9h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Sans objet

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données »

Objectif pédagogique :

Il s'agit d'un atelier intense pendant lequel les étudiants vont installer une version récente de Linux et apprendre à manipuler ce système d'exploitation afin de l'utiliser tout au long de l'année.

Linux est en particulier central pour utiliser et développer les technologies Big Data.

Contenu de la matière :

1. Présentation d'Unix
2. Installation d'une version
3. Découverte pratique d'Unix
4. Installation de logiciels

Modalités d'évaluation :

Aucune

Références bibliographiques :

C. PELISSIER, Unix, Editions Hermès

UE 4 – Système et Réseaux

RESEAUX ET SYSTEMES D'EXPLOITATION

Computer Networks and Operating System

<i>Enseignant</i>	: Jean-Baptiste LOISEL (Orange Consulting)
<i>Nombre d'ECTS</i>	: 1,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 42h
<i>Répartition des enseignements</i>	: • Cours : 15h • Atelier : 6h
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Sans objet

Enseignement destiné aux élèves de la filière « Data Science & Ingénierie des Données » et du Master Big Data

Course objectives :

This course aims to provide students with an understanding of the core principles of technologies constituting the foundation of the IT world: operating systems and computer networks.

In the first part, we will study the way an operating system organizes and facilitates the interaction of its key resources such as processor, memory, and file system in a multi-tasking and multi-user context.

The second part will focus on networks and will address various topics, such as network topology and technologies, Ethernet, ADSL, LAN, WAN, VLAN, Internet, Wifi and secure Wifi, TCP/IP layers, major protocols (DNS, SMTP...), network devices, architecture designs (dimensioning, redundancy, segmentation, DMZ...).

Implications for the security of the Information System will also be touched when addressing these topics, in order to raise awareness about inherent security risks and relevant countermeasures.

Course description :

Operating Systems

1. Operation Systems overview
2. Operation Systems overview
3. Processes
4. Inter-process communication
5. Memory management
6. Processes scheduling
7. File systems
8. Disk management systems (RAID)
9. Virtualization

Practicals will supplement the course.

Computer Networks

1. Introduction
2. Host-network layer
3. Internet layer
4. Transport layer
5. Application layer
6. Architecture review

Course evaluation :

Written exam

Bibliography :

- Modern Operating Systems. Andrew Tanenbaum. Pearson Education. 4th edition (2014). ISBN-13: 978-0133591620 ISBN-10: 013359162X
- Computer networks. Andrew Tanenbaum & David Wetherall. Pearson. 5th edition (2010). ISBN-13: 978-0132126953 ISBN-10: 0132126958

UE 4 – Systèmes et Réseaux

SYSTEMES REPARTIS

Networks

<i>Enseignants</i>	: David FREY & George GIAKKOUPIS (Inria Rennes)
<i>Nombre d'ECTS</i>	: 1,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 42h
<i>Répartition des enseignements</i>	: • Cours : 15h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Algorithmique, Programmation orientée Objet

Enseignement destiné aux élèves de la filière « Statistique et Ingénierie des Données »

Objectif pédagogique :

Cet enseignement vise à donner aux étudiants les connaissances de base sur les architectures distribuées, réparties sur différents sites partout dans le monde. Les trois architectures réparties à grande échelles les plus courantes seront présentées : grilles, systèmes peer-to-peer, et cloud. Les hypothèses, concepts and algorithmes seront détaillés pour chacune d'entre elles. L'objectif est d'avoir une connaissance des systèmes répartis disponibles actuellement et de pointer les directions futures de ces architectures.

Contenu de la matière :

1. Introduction aux architectures distribuées
2. Les concepts fondateurs (synchronisation, exclusion mutuelle, etc.)
3. Les approches centralisées et semi-centralisées (cloud, grilles, etc.)
4. Les approches décentralisées (systèmes P2P - structurés, non-structurés et hybrides)
5. Application aux systèmes de partage de fichiers et aux protocoles épidémiques

Modalités d'évaluation :

Un projet de mise en œuvre de système décentralisé, type épidémique

Références bibliographiques :

- Andrew S. Tanenbaum et Maarten Van Steen. Distributed Systems: Principles and Paradigms. Pearson New International Edition (2013)
- Kenneth Birman : Guide to Reliable Distributed Systems. Springer Verlag (2012)
- Fabrice Le Fessant et Jean-Marie Thomas. Le peer-to-peer : Comprendre et utiliser. Eyrolles (2011)
- Andrew S. Tanenbaum. Systèmes d'exploitation : Systèmes centralisés, systèmes distribués. Dunod (1999)

UE 4 – Systèmes et Réseaux

SECURITE DES DONNEES

Data Security

<i>Enseignant</i>	: Franck LANDELLE (DGA MI)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 24h
<i>Répartition des enseignements</i>	: • Cours : 9h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Sans objet

Enseignement destiné aux élèves de la filière « Statistique et Ingénierie des Données »

Objectif pédagogique :

La sécurité informatique fait actuellement l'objet d'une actualité particulièrement dynamique : attaques spectaculaires (virus, intrusion, ...), commerce électronique, évolutions de législations...

L'objet de ce cours est de présenter les grands principes de la sécurité informatique et les techniques de protection des données.

L'usage de la cryptographie est l'un des outils de protection contre la divulgation, la modification ou l'accès illégitime à des données ou moyens.

Les techniques cryptographiques qui permettent d'assurer les services de confidentialité, d'intégrité, de signature ou d'authentification.

Finalement, des systèmes utilisant ces techniques seront schématiquement décrits.

Contenu de la matière :

1. Introduction à la sécurité
 - 1.1. Besoins
 - 1.2. Menaces
2. Cryptographie
 - 2.1. Définitions générales
 - 2.2. Cryptographies à clés secrètes
 - 2.3. Cryptographies à clés publiques
 - 2.4. Protocoles cryptographiques
3. Systèmes utilisateurs
 - 3.1. Applications Web
 - 3.2. Carte bancaire
- 3.3. Application réseaux

Modalités d'évaluation :

Examen écrit

Références bibliographiques :

- Schneier, Cryptographie appliquée, Thomson Publishing, 1997
- Stinson, Cryptographie : Théorie et pratique, Vuibert 2003
- Menezes, Van Oorschot, Vanstone, Handbook of applied Cryptography, CRC Press, 1997 (version actualisée en ligne)
- Vergnaud, Exercice et problèmes de la cryptographie, Dunod, 2012.
- Singh, Histoire des codes secrets, JC Lattes, 1999

UE Stages

UE : PROJET PROFESSIONNEL ET STAGES

<i>Correspondant de l'UE</i>	: Patrick Gandubert
<i>Nombre d'ECTS</i>	: 30
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Travail en entreprise
<i>Nombre d'heures d'enseignement</i>	: 30h (séminaires)

Finalité de l'UE :

Cette UE correspond à des temps pédagogiques en lien direct avec les entreprises. Les séminaires professionnels ont pour objectif de présenter aux étudiants diverses problématiques auxquelles ils seront confrontés dans leur environnement professionnel. Il permet d'apporter des compléments par rapport à certains cours, et fait le lien entre les enseignements et les applications pratiques qui en découlent. Le projet professionnel permet de préparer les étudiants à leur entrée dans la vie professionnelle et aux stages, il est réalisé sur la 2ème et 3ème année de formation. Des simulations d'entretien de recrutement sont organisées en 3e année. Elles sont assurées par des recruteurs d'entreprises et d'organisations partenaires de l'Ensaï. Les stages (application en 2ème année, fin d'études en 3ème année) permettent aux élèves de mettre en pratique les enseignements de mathématiques appliquées, d'informatique et d'économie dans un cadre professionnel. Le stage de fin d'études, d'une durée de 20 semaines minimum, vise à appliquer les enseignements de 3ème année et à acquérir de l'expérience pour assurer la transition vers l'emploi. Il constitue une étape essentielle de mise en situation professionnelle pour le futur ingénieur qui dispose à ce stade de l'ensemble des bagages techniques de la formation.

Structuration de l'UE :

Le stage de fin d'études constitue la majeure partie de l'évaluation de cette UE (25 ECTS). L'Ensaï exige une forte adéquation entre le contenu du stage et la filière de spécialisation de 3e année. Il fait l'objet d'une procédure de validation par le responsable de filière et par le département des relations avec les entreprises. L'évaluation tient compte de la capacité d'intégration de l'étudiant dans l'entreprise, ses capacités d'initiative et de satisfaction au regard des objectifs du stage, et de la qualité du rapport et de la soutenance réalisée devant un jury composé d'un président, d'un vice-président, tous les deux issus du monde de l'entreprise et d'un permanent de l'école. Le stage d'application de 2ème année est pris en compte dans cette UE (5 ECTS). Les séminaires professionnels ne sont pas évalués.

Compétences ou acquis d'apprentissage à l'issue de l'UE :

Le stage de fin d'études (et l'UE) comprend un objectif technique - il s'agit de répondre à la commande, à la problématique inscrite dans le thème du stage à l'aide des connaissances acquises - et un objectif professionnel - il s'agit de parfaire la connaissance du monde du travail, de développer des capacités relationnelles et d'adopter une démarche d'insertion dans le monde professionnel.

Les pré-requis de l'UE :

Aucun

UE Projet Professionnel

DROIT DU TRAVAIL

Work Law

<i>Enseignant</i>	: Charlotte GRUNDMAN, Avocat au Barreau de Paris
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 15h
<i>Répartition des enseignements</i>	: Cours : 3h • Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Distribués pendant le cours
<i>Pré-requis</i>	: Aucun

Modalités d'évaluation :

Exposé d'un cas pratique réalisé lors des TD.

Acquis d'apprentissage (objectifs) :

La matière étant extrêmement vaste et complexe, il est ici proposé aux étudiants une approche didactique et vivante du sujet, l'objectif de l'enseignement étant de permettre aux étudiants qui travailleront dans un futur proche en entreprise d'avoir compris certaines notions pratiques essentielles en droit du travail.

Principales notions abordées :

Hormis le cours d'amphi, il sera systématiquement proposé aux étudiants, après l'étude d'une notion, un exercice visant à mettre en pratique la notion abordée. Le cours commun (3 heures) traite des notions suivantes : Comprendre d'où l'on vient pour savoir où on va (introduction historique au droit du travail, les sources du droit du travail, ordre public absolu et ordre public social), les instances de contrôle du droit du travail, formation et exécution du contrat de travail, la rupture du contrat à durée indéterminée. Pour les TD, la première heure de cours sera consacrée à l'étude d'un chapitre (la modification du contrat de travail, le recrutement, les droits fondamentaux du salarié). Cet exposé sera suivi d'une mise en situation pratique, où les étudiants devront par groupe répondre à un cas pratique. Un rapporteur sera désigné par groupe, et la notation se fera à cette occasion.