



Programme des enseignements 3^e année

Filière Statistique et ingénierie des données

ANNEE SCOLAIRE 2017 / 2018



École nationale
de la statistique
et de l'analyse
de l'information

FILIÈRE STATISTIQUE ET INGÉNIERIE DES DONNÉES

ANNÉE SCOLAIRE 2017/2018

Data Science and Big Data

2017/2018 ACADEMIC YEAR

Table des matières

Présentation de la filière	4
Descriptifs des enseignements du tronc commun	6
Simulation d'entretien	7
Droit des entreprises	8
Droit du Travail	10
Anglais	12
Sport	14
Descriptifs des enseignements de la filière	15
UE - Statistiques pour la fouille et le Big Data	16
Réseaux de neurones	16
Support vector machines	17
Modèles de régression pénalisée	18
Méthodes d'agrégation	19
UE – Développement d'application	21
Génie Logiciel	21
Indexation Web et Webdatamining	23
UE – Big Data	26
Technologies Sémantiques	26
Technologies NoSQL	27
Publication de données respectueuse de la vie privée	28
UE – Systèmes et Réseaux	30
Réseaux et systèmes d'exploitation	30
Initiation à Unix	31
Systèmes Répartis	32
Sécurité des données	33
Grandes masses de données sur Cloud	34
UE - Applications Statistiques et Informatiques	35
Projet de fin d'études	35
Projet WEB (JavaEE)	36
Veille sur les médias	37
Projet Anglophone : Applications Web	38
Séminaires professionnels	39
Séminaires professionnels	39

Présentation de la filière

L'objectif de la filière Statistique et Ingénierie des Données est d'offrir aux étudiants une double compétence recherchée sur le marché actuel du travail où il est impossible de se passer d'informatique et où le traitement de données de plus en plus volumineuses nécessite de sérieuses connaissances statistiques. Au cours de cette 3^{ème} année, les élèves vont acquérir des bases informatiques solides qui leur permettront de maîtriser la conception de systèmes. Ils pourront ainsi mieux définir leurs attentes afin d'exploiter, à l'aide d'outils statistiques, les nouvelles mines d'information disponibles par le monde. La filière SID ne fixe pas de domaine d'application, celui-ci peut être défini pendant l'année voire pendant le stage, cette filière apporte essentiellement des compétences complémentaires en informatique, elle garde ouverte les portes de la finance, de la biostat, du marketing ou tout autre domaine et permet bien entendu de poursuivre en thèse. Cette filière forme aux métiers de *data scientists* et *data analyst*.

Au cours des deux premières années à l'ENSAI, les étudiants ont acquis les bases de la conception orientée objet (Java), ils ont appris à utiliser des outils de développement spécialement adaptés (Eclipse par exemple), à modéliser des applications (UML) et à concevoir des bases de données. Ils ont également assimilé les principales méthodes statistiques. Toutes ces connaissances ont été appliquées et consolidées au cours de projets en deuxième année.

S'appuyant sur ces acquis, la filière s'articule autour de deux axes principaux : l'outil informatique d'une part, l'apprentissage statistique et l'ingénierie des données d'autre part.

Les outils informatiques

Le développement de grandes applications informatiques nécessite d'utiliser des méthodes d'aide à la conception. Les méthodes de développement d'application sont présentées dans le cours de Génie logiciel. Les architectures des grandes applications sont désormais de type multi-tiers, pour les appréhender, il est nécessaire d'avoir des connaissances de base en architecture des systèmes, en réseaux, en sécurité. Le développement d'applications est réalisé sur la plate-forme J2EE. Les technologies du Web Sémantique seront également abordées.

L'ingénierie des données

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. 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 d'apprentissage statistique 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 au traitement des données.

Plusieurs projets sont réalisés au cours de l'année afin de mettre en pratique les divers enseignements dispensés au cours de cette année. Ces projets sont réalisés individuellement ou en groupe, permettant aux étudiants de vivre la réalité d'un développement d'application. Les projets sont de nature très diverses 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.

Enseignements de 3^e année : Filière Statistique et ingénierie des données

Enseignements	Volume horaire				Crédits
	Cours	Atelier	Projet	Total	Coefficients
Tronc commun					
Simulation Entretien		3		3	0
Droit des entreprises	6			6	0,5
Droit du Travail	9			9	1
Anglais	30			30	2,5
Sport		30		30	0
Total	45	33	0	78	4
Statistiques pour la fouille et le Big Data					
Réseaux de neurones ²	6	3		9	1,5
Séparateurs à vaste marge ¹	18	0,0		18	1,5
Régression pénalisée ¹	24	0,0		24	1,5
Méthodes d'agrégation ¹	9	3,0		12	1,5
Total	57	6,0	0	63	6
Développement d'application					
Génie Logiciel	39	39		78	2
Indexation web	15	9		24	2
Total	54	48	0	102	4
Big Data					
Technologies Sémantiques	6	9		15	1
Technologies NoSQL	15	0		15	2
Publication de données respectueuse de la vie privée	15	6		21	1
Total	36	15	0	51	4
Systèmes et Réseaux					
Réseaux et systèmes d'exploitation	15	6		21	2
Initiation à Unix	9	6		15	0
Systèmes Répartis	15	6		21	1
Sécurité des données	15			15	1
Grandes masses de données sur Cloud	12	12		24	2
Total	66	30	0	96	6
Applications statistiques et informatiques					
Projet de Fin d'Etudes			9	9	2
Projet Web		3	9	12	2
Veille sur les médias		3		3	1
Projet anglophone : applications web	12,0	12	6	30	1
Total	12	18	24	54	6
Séminaires professionnels					
Python 4 data science		6			0
Spark		6			0
Elastic Stack		6			0
TensorFlow		6			0
Hive		6			0
Total	0	30	0	30	0
Total	270	180	24	474	30

¹ Enseignement commun à la filière *Marketing quantitatif et revenue management*² Enseignement commun à la filière *Génie Statistique**Partie stage*

UE Stages	Crédits 30
------------------	-----------------------------

Descriptifs des enseignements du tronc commun

UE – Tronc commun

Simulation d'entretien

Job Interviews, CVs and Cover Letters

Atelier : 3h

Enseignant : Divers intervenants

Correspondant : Patrick GANDUBERT

Cours facultatif

Objectif pédagogique

L'objectif est de préparer les étudiants aux entretiens d'embauche : stage ou premier emploi ainsi qu'aux entretiens qu'ils auront à passer lors du forum des entreprises de l'Ensaï.

Contenu de la matière

Au delà des savoir-faire technique, le caractère humain de l'entretien d'embauche

Travail sur le CV

Envoi de candidature : rédaction du mail et de la lettre de motivation

Mise en situation : candidat – RH – Responsable technique - Observateur

Pré-requis

Préparer son CV

Contrôle des connaissances

Cours non évalué

Références bibliographiques

Seront données en cours

Langue d'enseignement

Français

UE – Tronc commun

Droit des entreprises

Corporate Law

Cours : 6h

Enseignant : Divers intervenants

Correspondant : Ronan LE SAOUT

Objectif pédagogique

- Acquérir les connaissances de base en droit social concernant les relations du travail dans l'entreprise : sources du droit du travail, formation, exécution et rupture du contrat de travail, droits et devoirs du salarié ;
 - Acquérir quelques notions en droit commercial : typologie des entreprises, les sociétés commerciales, les entreprises en difficulté.
- Favoriser débats et échanges sur des problèmes concrets du domaine.

Contenu de la matière

Première partie : Notions essentielles de droit du travail

Chapitre 1 : Statut du travailleur français à l'étranger et du travailleur étranger en France

Section 1 : Le statut de salarié à l'étranger

Section 2 : Travailler en France lorsque l'on est étranger

Chapitre 2 : Les sources du droit du travail

Chapitre 3 : Formation, modification et contenu du contrat de travail

Section 1 : Le cas particulier des contrats à durée déterminée

Section 2 : La période d'essai

Section 3 : Les fonctions

Section 4 : Le lieu de travail

Section 5 : La rémunération

Section 6 : La durée du travail

Section 7 : La clause de non-concurrence

Section 8 : L'évolution du contrat

Chapitre 4 : Droits et devoirs au sein de l'entreprise

Section 1 : Liberté d'expression et vie privée

Section 2 : Protection contre les discriminations et le harcèlement

Section 3 : Protection de la femme enceinte, congés maternité et paternité

Section 4 : Congés payés

Section 5 : Droit à la formation

Section 6 : Maladie, accident de trajet, accident du travail

Section 7 : Les représentants du personnel

Chapitre 5 : La rupture du contrat de travail et ses suites

Section 1 : La démission et la rupture conventionnelle

Section 2 : Le licenciement pour motif personnel

Section 3 : Le licenciement pour motif économique

Section 4 : La prise en charge par Pôle Emploi

Deuxième partie : Notions essentielles de droit commercial

Chapitre 1 : L'auto-entrepreneur

Chapitre 2 : L'entreprise individuelle et l'entreprise en société

Chapitre 3 : Les sociétés commerciales

Chapitre 4 : Sauvegarde, redressement judiciaire, liquidation judiciaire

Pré-requis

Contrôle des connaissances

Examen final

Références bibliographiques

Langue d'enseignement

Français

UE – Tronc commun

Droit du Travail

Work Law

Cours : 9h

Enseignant : Charlotte Grundman, Avocat au Barreau de Paris. nants

Correspondant : Ronan LE SAOUT

Objectif pédagogique :

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.

A cette fin, et 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.

Afin de satisfaire le plus possible à cet objectif, il est ainsi proposé l'organisation suivante des cours :

Cours commun (3 heures) :

Chapitre 1 : Comprendre d'où l'on vient pour savoir où on va :

- Introduction historique au droit du travail
- Les sources du droit du travail
 - sources imposées,
 - sources négociées
- Ordre public absolu et ordre public social

Chapitre 2 : les instances de contrôle du droit du travail

- L'inspecteur du travail
- Les multiples juges du droit du travail
- Point sur la procédure prud'homale

Chapitre 3 : Formation et exécution du contrat de travail

- la qualification du contrat de travail : « faux artisans, faux auto-entrepreneurs et vrai salarié ».
- le contrat à durée indéterminée, norme juridique et sociale
- la période d'essai après la loi du 25 juin 2008 : définition, durée et rupture
- les principales clauses du contrat de travail :
 - la clause de mobilité
 - la clause de non-concurrence

Chapitre 4 : la rupture du contrat à durée indéterminée

- le licenciement pour motif personnel
- le licenciement pour motif économique
- la démission du salarié
- les autres modes de rupture

Les TD :

La première heure de cours sera consacrée à l'étude d'un chapitre. 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.

Chapitre 1 : La modification du contrat de travail

Modification du contrat de travail et changement des conditions de travail

- la durée du travail (focus sur le forfait-jour)
- le lieu de travail
- la rémunération

Chapitre 2 : Le recrutement

Chapitre 3 : les droits fondamentaux du salarié

- Le fait religieux en entreprise
- Vie personnelle et technologies de l'information et de la communication (TIC)
- La mise en place de moyens de contrôle via les TIC en entreprise
- Harcèlements
- Maladie et maternité du salarié

Langue d'enseignement

Français

UE – Tronc commun

Anglais

English

Cours : 30h (dont 15h d'aide au projet)

Enseignant : Divers intervenants

Correspondant : Esther LALAU-KERALY

Objectif pédagogique

Les élèves qui n'ont pas passé ou qui n'ont pas réussi le TOEIC l'année dernière auront progressé dans les compétences requises – c'est à dire, la compréhension orale, la reconnaissance des erreurs, les pièges grammaticaux, et la compréhension écrite. Les autres auront acquis les compétences nécessaires pour affronter le monde professionnel. Ils auront vu les tournures qui aident à diriger et à participer à des réunions, à prendre des décisions, et à négocier. Ils se seront entraînés à faire des présentations. Ils auront rédigé un projet en anglais et préparé la soutenance de ce projet.

Contenu de la matière

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'Ecole 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.

Pré-requis

Aucun

Contrôle des connaissances

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*. Le CV et la lettre faite pour cet exercice seront évalués et feront partie de la note finale. Le niveau acquis apparaîtra sur le Supplément au diplôme. L'objectif de la CTI[†] pour tous les élèves ingénieurs est d'atteindre le niveau B2.

* le Cadre européen commun de référence pour les langues.

[†] la Commission des Titres d'Ingénieur.

Références bibliographiques

- Arbogast, B., *30 Days to the TOEIC Test*, Canada: Peterson's, 2002.
- Schramper-Azar, B., *Understanding and Using English Grammar*, New York: Longman, 1999.
- Buckwalter, Elvis, et.al, *Boostez votre score au TOEIC-spécial étudiants*, Paris: Eyrolles, 2009.

- Gear, Jolene, *Cambridge Grammar and Vocabulary for the TOEIC Test*, Cambridge: Cambridge University Press, 2010.
- Lecomte, Stéphane, et. al, *La Grammaire au TOEIC et au TOEFL : Mode d'emploi*, Paris: Ophrys, 2008.
- Loughheed, Lin, *Tests complets pour le nouveau TOEIC (4^{ème} ed.)*, Paris: Pearson Education France, 2008.
- MBA Center, *New TOEIC Study Book*, Paris: MBA Center Publications, 2007.

Langue d'enseignement

Anglais

Pour tout complément d'information, chaque élève peut consulter le *Programme des enseignements : Langues étrangères*, distribué au début de l'année académique.

UE – Tronc commun
Sport
<i>Sport</i>
Atelier : 30h
Enseignant : Divers intervenants
Correspondant : Julien LEPAGE
<i>Cours facultatif</i>

Objectif de la matière

L'objectif est d'amener les élèves à maintenir un esprit sportif, sortir du strict cadre académique et développer leurs capacités physiques.

Contenu de la matière

8 activités sportives sont proposées par l'école :

- Badminton
- Basket
- Fitness
- Football
- Hand-ball
- Tennis de table
- Tennis débutant
- Volley-ball
- Course à pied/préparation physique/coaching sportif (nouveau)

Outre les entraînements, les élèves inscrits peuvent être amenés à participer à des compétitions.

Prise en compte dans la scolarité

La participation à une activité sportive peut donner lieu à l'attribution d'un bonus ajouté sur la moyenne du semestre concerné. Le niveau de ce bonus est précisé dans une circulaire d'application en début d'année académique. Il varie selon l'assiduité aux séances, l'engagement et la participation aux compétitions tout au long de l'année.

Pour être définitive, la liste des élèves bénéficiant de ces bonus doit être validée par le directeur des études.

Un bonus peut être exceptionnellement attribué en dehors des activités sportives réalisées dans le cadre Ensaï. Pour y prétendre, les élèves concernés doivent remplir les 3 conditions suivantes :

- pratiquer régulièrement une activité sportive et participer aux compétitions liées ;
- posséder un niveau national (voir très bon niveau régional suivant le sport en question) ;
- déposer une demande argumentée auprès de la direction des études et du service sport en début d'année scolaire, afin de faire valider le programme d'entraînement, des compétitions et les modalités de diffusion des performances.

Pour certains ayant des contraintes sportives, des aménagements horaires pourront d'ailleurs être ainsi envisagés si besoin.

Descriptifs des enseignements de la filière

UE - Statistiques pour la fouille et le Big Data

Réseaux de neurones

Neural Networks

Cours : 6h • Atelier : 3h

Enseignant : Romaric Gaudel (ENSAI)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves des filières SID et Génie statistique.

Objectif pédagogique

Les réseaux de neurones artificiels sont des outils inspirés par la structure du cerveau, créés à l'origine pour imiter certaines fonctions présentes dans le monde animal (pas seulement dans le comportement des humains). Ces outils ont connu un fort développement depuis, et ont étendu considérablement leurs champs d'applications. Aujourd'hui, ils sont présents dans énormément de domaines, et certains types de réseaux de neurones sont parmi les outils principaux pour la réalisation de tâches spécifiques. On peut les classer parmi les outils de la statistique ou comme méthodes de ce qu'on appelle l'Intelligence Artificielle en informatique.

Les réseaux de neurones sont utilisés pour implémenter des fonctions de type apprentissage, pour faire de la classification, pour prédire le comportement d'un système, pour optimiser un fonction complexe dans un contexte très combinatoire, pour des fonctions de mémorisation d'information, et pour d'autres tâches plus connectées avec certaines des fonctions cognitives que les humains sont capables de faire (perceptions, d'autres fonctions de la mémoire, raisonnements, etc.). Dans ce cours, nous aurons une introduction au sujet en illustrant l'utilisation de ces objets pour des applications représentatives de leur usage, en apprentissage supervisé et non supervisé, en optimisation, ou en prédiction. Ceci passera pour traiter la notion de neurone artificiel et celle de réseau de neurones, pour voir ensuite comment on utilise ces outils pour les applications sélectionnées.

Contenu de la matière

- Neurones artificiels et réseaux de neurones artificiels. Exemple de base : réseaux multicouches. Différentes variantes de ces concepts. Utilisation des réseaux multicouches en apprentissage supervisé.
- Utilisation des réseaux de neurones en apprentissage non supervisé, en optimisation, en prédiction. Illustrations de ces techniques dans différents domaines, en incluant les prédictions en finances.
- Etude détaillée d'un cas réel, avec introduction d'un type particulier de réseau et analyse pas à pas des différentes étapes à suivre pour résoudre un problème d'apprentissage supervisé (qui peut aussi être vu comme un problème de classification).
- Compléments : réseaux de neurones et big data. L'apprentissage profond. Quelques exemples supplémentaires de succès industriels récents de tous ces outils.

Contrôle des connaissances

Sera indiqué par l'enseignant.

Références bibliographiques

Langue d'enseignement

Français

UE Data science, machine learning

Support vector machines

Support vector machines

Cours : 12h

Enseignant : Fabien NAVARRO

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves des filières « Marketing quantitatif et revenue management » et « Statistique et ingénierie des données ».

Objectif pédagogique

Les séparateurs à vaste marge (ou support vector machines) sont des méthodes de machine learning de plus en plus utilisées dans un contexte de grande dimension. Elles sont souvent utilisées à des fins de prédiction. La mise en application de ces méthodes est faite en atelier sous R.

Contenu de la matière

Sera fourni par l'enseignant.

Pré-requis

Notions de statistique vues 1^{re} et 2^e année, algèbre linéaire, optimisation, bootstrap.

Contrôle des connaissances

Sera indiqué par l'enseignant.

Références bibliographiques

- P. BUHLMANN, B. YU. Boosting with the L2 loss: Regression and classification, 98.462, 2003
- N. CRISTIANI and J. SHAWE-TAYLOR. *An introduction to support vector machines*, 2000
- L. DEVROYE, L. GYORFI, and G. LUGOSI. *A Probabilistic Theory of Pattern Recognition*, 1996
- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.)*, 2009
- B. SCHOLKOPF and A. J. SMOLA. *Learning with Kernels*, Cambridge, 2002
- V. VAPNIK. *Statistical Learning Theory*, Wiley, 2004

Langue d'enseignement

Français

UE Data science, machine learning

Modèles de régression pénalisée

Sparse methods in regression

Cours : 14h

Enseignant : Denys POMMERET (université Aix-Marseille)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves des filières « Marketing quantitatif et revenue management » et « Statistique et ingénierie des données ».

Objectif pédagogique

Ce cours présente différents modèles de régression pénalisée. Ces méthodes sont mises en œuvre en présence d'un grand nombre de facteurs (risque très élevé de multicollinéarité). La mise en application de ces méthodes est faite en atelier sous R.

Contenu de la matière

Modèles de régression pénalisée.
Méthodes Lasso, ridge.

Le contenu sera plus détaillé par l'enseignant.

Pré-requis

Notions de statistique vues 1^{re} et 2^e année, algèbre linéaire, optimisation, bootstrap.

Contrôle des connaissances

Sera indiqué par l'enseignant.

Références bibliographiques

- P. BUHLMANN, B. YU. Boosting with the L2 loss: Regression and classification, 98.462, 2003
- L. DEVROYE, L. GYORFI, and G. LUGOSI. *A Probabilistic Theory of Pattern Recognition*, 1996
- Y. FREUND and E. SCHAPIRE, *Experiments with a new boosting algorithm*, 1996
- L. GYORFI, M. KOHLER, A. KRZYVAK, and H. WALK. *A distribution-free theory of non-parametric regression*, 2002
- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.)*, 2009
- V. VAPNIK. *Statistical Learning Theory*, Wiley, 2004

Langue d'enseignement

Français

UE - Statistiques pour la fouille et le Big Data

Méthodes d'agrégation

Statistical learning

Cours : 18h •

Enseignant : Brigitte GELEIN

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves des filières SID, « Marketing quantitatif et revenue management » et « Statistique et ingénierie des données »

Objectif pédagogique

Ce cours présente différentes méthodes d'agrégation de modèles. Les algorithmes décrits sont basés sur des stratégies adaptatives ou aléatoires. Ils permettent d'améliorer l'ajustement par une combinaison ou agrégation d'un grand nombre de modèles tout en évitant le sur-ajustement.

- Bagging and random forest : agrégation par moyenne pondérée ou vote des estimations de différents modèles du même type, construits sur des sous-ensembles de la table d'apprentissage.
- Boosting : agrégation par moyenne pondérée ou vote des estimations de différents modèles du même type construits sur toute la table d'apprentissage, chaque modèle est une version adaptée du précédent en donnant plus de poids, lors de l'estimation suivante, aux observations mal ajustées ou mal prédites lors de l'étape précédente.
- Stacking : agrégation de différents modèles de types différents par construction d'un modèle « superviseur » qui combine au mieux les prédictions des modèles primaires.

La mise en application de ces méthodes est faite en atelier sous R et Python principalement. Les procédures de SAS feront également l'objet d'une présentation.

Contenu de la matière

1. Modèles d'apprentissage statistique : quelques rappels et approfondissements (multiregression trees).
2. Boosting.
3. Bagging, forêts aléatoires.
4. Stacking

Pré-requis

Notions de statistique vues 1^{re} et 2^e année, optimisation, bootstrap et méthodes de discrimination.

Contrôle des connaissances

Examen sur table.

Références bibliographiques

- P. BARTLETT and M. TRASKIN. AdaBoost is consistent, *JMLR*, 2007
- G. Biau, A. Ficher, B. Guedj et J. D. Malley, COBRA : A Nonlinear Aggregation Strategy, *Journal of Multivariate Analysis* 2013
- L. BREIMAN. Bagging predictors, *Machine Learning*, 24.2, 1996
- L. BREIMAN., *Stacked Regression*, *Machine Learning*, 24.1, 1996
- P. BUHLMANN, B. YU. *Boosting with the L2 loss: Regression and classification*, 98.462, 2003

- Y. FREUND and E. SCHAPIRE, *Experiments with a new boosting algorithm*, 1996
- R. GENUER. *Forêts aléatoires : aspects théoriques, sélection de variables et applications*, Thèse de l'Université Paris XI, 2010
- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.)*, 2009
- E. Scornet, G. Biau et J. P. Vert, *Consistency of random forests*, The Annals of Statistics, 2015
- V. VAPNIK. *Statistical Learning Theory*, Wiley, 2004

Langue d'enseignement

Français

UE – Développement d'application

Génie Logiciel

Software engineering

Cours : 39h • Atelier : 39h

Enseignant : Mathieu ACHER, Johann BOURCIER , Olivier BARAIS, et Benoit COMBE-MALE (Université Rennes 1)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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

- Pourquoi ?
- Comment ?

Introduction au génie logiciel et bonnes pratiques de conception.

Architecture logicielle et modèle en couche, Exemple sur GWT.

Principaux patrons de conception, principe et mise en œuvre en Java.

Le test logiciel et l'ingénierie des langages.

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 requêtage
4. Les services web, le cloud.

Pré-requis

Notation UML, connaissance du langage JAVA.

Contrôle des connaissances

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

Langue d'enseignement

Français

UE – Développement d'application

Indexation Web et Webdatamining

Web datamining

Cours : 15h • Atelier : 9h

Enseignant : Nawfal TACHFINE (aramisauto)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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
 - Concepts & Définitions
 - Term Document Matrix
 - Tf-idf, Cosine Index, jaccard Index
 - Stemming
- Web Search : Google
 - Google et le Page Rank
 - Pages Jaunes (Notion de tri alpha)
 - Notion de graphes et de vecteurs propres

Partie pratique (9h)

- TP1 : Introduction à R pour le Web Mining (3h)
 - Installation de bibliothèques de textmining disponible dans R
 - Collecter les informations issues du WEB : Twitter, Wikipedia
 - Pre-processing : Stemmatisation, Lemmatisation,
 - Parsing HTML, XML,
 - 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
 - o Quelles applications dans quels domaines d'activités
- État de l'art (opinion mining, sentiment analysis, affective computing)
 - o Quels descripteurs pour quels types de données ?
 - Textuelles
 - Audio
 - Images
 - o Sélection automatique de descripteurs (réduction de l'espace de recherche)
 - o Quels algorithmes de classification dans quels cas ?
- Constitution du corpus
 - o Réflexions générales sur la qualité des données et son impact
 - o Annotation manuelle et automatique (schéma d'annotation, calcul d'un score d'agrément inter-annotateur,)
 - o Répartition des données dans les classes
- Pre-processing (texte)
 - o Quelle granularité pour mes données (mot, phrases, paragraphes)
 - o Annotation syntaxique et sémantique (exemples de POS, WordNet-Affect, etc)
- Évaluation
 - o 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)
 - o Produit de la société TEMIS (cartouche sentiments)
 - o 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)

Pré-requis

SQL.

Contrôle des connaissances

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") (*)

Langue d'enseignement

Français.

UE – Big Data

Technologies Sémantiques

Semantic Technologies

Cours : 6h • Atelier : 9h

Enseignants : Julien BLASS (Semsoft), Cédric LASBLEIZ (Semsoft)

Correspondant : Farah AIT SALAHT

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

Objectif pédagogique:

Le cours vise à donner aux étudiants une vision détaillée de la prochaine génération du web - le web sémantique -, qui introduit le sens de l'information dans les échanges de données pour permettre aux machines de traiter automatiquement l'information disponible sur le web. Le cours présente les standards du web sémantique et propose aux étudiants de manipuler des outils implémentant ces standards pour répondre à un cas d'usage concret.

Contenu de la matière:

- * Introduction au Web sémantique et au modèle de description RDF
- * Introduction à RDFS/OWL et au langage SPARQL
- * Le Web sémantique en pratique

Pré-requis:

- * Programmation Java
- * Connaissance des technologies du web: HTTP, HTML, XML

Contrôle des connaissances:

TP noté

Références bibliographiques:

- * Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer, 2nd Edition (Cooperative Information Systems), 2008

Langue d'enseignement:

Français

UE – Big Data

Technologies NoSQL

NoSQL Technologies

Cours : 12h • Atelier : 3h

Enseignant : Julien BALAS

Correspondant : Farah AIT SALAHT

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

Objectif pédagogique

Ce cours vis à présenter les différentes approches présentes dans le contexte des bases de données NoSQL. 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

Pré-requis

Bases de données relationnelles

Contrôle des connaissances

Références bibliographiques

Langue d'enseignement

Français

UE – Big Data

Publication de données respectueuse de la vie privée

Privacy-preserving data publishing

Cours : 15h • Atelier : 6h

Enseignant : Tristan Allard (Univ. Rennes 1)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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 prééminent aujourd'hui : la differential privacy.

Contenu de la matière :

- Introduction : motivation, défis, survol
- Paradigmes : non-informatif, differential privacy
- Publication interactive: modèles type differential privacy, mécanismes principaux de perturbation interactive (e.g., Laplace)
- *Perturbation locale : le mécanismes des réponses randomisés pour satisfaire la differential privacy
- Publication centralisée : mécanismes de génération de données synthétiques satisfaisant la differential privacy, survol des modèles basés sur le partitionnement (e.g., k-anonymat, l-diversité) et des mécanismes principaux pour les satisfaire (e.g., algorithme de Mondrian)
- Conclusion : les pratiques « dans le monde réel », questions ouvertes

Pré-requis :

- 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.

Contrôle des connaissances :

Contrôle continu et examen final.

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.

Langue d'enseignement:

UE – Systèmes et Réseaux

Réseaux et systèmes d'exploitation

Computer Networks and Operating System

Cours : 15h • Atelier : 6h

Enseignant : Jean-Baptiste LOISEL (Orange Consulting)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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 counter-measures.

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

Computer Networks

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

Practicals will supplement the course.

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

Langue d'enseignement

Anglais

UE – Systèmes et Réseaux

Initiation à Unix

Networks and Systems

Cours : 9h • Atelier : 6h

Enseignant : François Xavier BRU (Orange Consulting)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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

Pré-requis

Contrôle des connaissances

Le contrôle des connaissances s'effectue sur l'ensemble des matières de l'UE.

Références bibliographiques

C. PELISSIER, Unix, Editions Hermès

Langue d'enseignement

Français

UE – Systèmes et Réseaux

Systèmes Répartis

Networks

Cours : 15h • Atelier : 6h

Enseignant : ??

Correspondant : Farah AIT SALAHT

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

Pré-requis

Algorithmique, Programmation orientée Objet

Contrôle des connaissances

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)

Langue d'enseignement

Français

UE – Systèmes et Réseaux

Sécurité des données

Data Security

Cours : 9h • Atelier : 6h

Enseignant : Franck LANDELLE (DGA MI)

Correspondant : Farah AIT SALAHT

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

Pré-requis

Contrôle des connaissances

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

Langue d'enseignement

Français

UE – Systèmes et Réseaux

Grandes masses de données sur Cloud

Big Data and Cloud Computing,

Cours : 12h • Atelier : 12h

Enseignant : Gabriel ANTONIU et Shadi IBRAHIM (Inria Rennes)

Correspondant : Farah AIT SALAHT

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

Objectif de la matière

Cet enseignement vise à donner aux étudiants les connaissances de base sur les architectures distribuées spécialisées dans le traitement du Big Data. Les hypothèses, concepts and algorithmes seront détaillés pour chacune d'entre elles. L'objectif est d'avoir une connaissance des systèmes sur Cloud disponibles actuellement et de pointer les directions futures de ces architectures. L'objectif final est la mise en œuvre avec Hadoop et une base de données Big Data spécifique.

Contenu de la matière

- Introduction aux infrastructures distribuées : clusters, supercalculateurs, grilles, clouds
- Introduction au cloud computing
- Big Data: introduction, défis, enjeux
- Explicit Data management
- Transparent Data Management: NFS, Gfarm, Google File System:
- Introduction à MapReduce et Hadoop:
- Atelier pratique sur Hadoop
- Avenir de MapReduce: défis
- Approches post-MapReduce: Shark/Spark (Berkeley)
- Approches post-MapReduce: Stratosphere

Pré-requis

Système répartis, interrogation (SQL) de bases de données relationnelles.

Contrôle des connaissances

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

Langue d'enseignement

Français

UE - Applications Statistiques et Informatiques

Projet de fin d'études

End of study project

Projet : 9h

Enseignant : Divers intervenants industriels

Correspondant : Farah AIT SALAHT

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

Objectif pédagogique

Une part très importante des enseignements est consacrée à la mise en œuvre d'applications informatiques : génie logiciel, conception de systèmes d'information avec UML, bases de données, programmation objet en JAVA, applications réparties et applications web (Java EE, Hadoop). L'un des objectifs de ce projet est de mettre en œuvre toutes ces connaissances acquises au sein d'une même application. Ce projet permet également de mettre en pratique les méthodes vues dans le cours de gestion de projets.

L'apprentissage lié au projet ne se limite pas à la mise en pratique des cours théoriques mais a pour ambition de sensibiliser les élèves aux autres composantes nécessaires à la réussite d'un projet comme la communication entre les différents acteurs du projet : entre les étudiants et également entre le client demandeur du projet et le groupe d'étudiants.

Ce projet peut faire l'objet d'un soutien particulier, effectué par un spécialiste, selon les besoins des étudiants.

Contenu de la matière

- Encadrement en début de projet afin de préciser les besoins.
- Encadrement distant sur des questions techniques ponctuelles
- Encadrement technique lors de la phase d'implémentation
- Soutenance des projets

Pré-requis

Contrôle des connaissances

Ce projet donnera lieu à la remise d'un rapport et à une soutenance réalisée par l'ensemble des étudiants. Il est à réaliser par l'ensemble des étudiants de la filière, éventuellement en lien avec une entreprise ou un organisme extérieur.

UE – Applications Statistiques et Informatiques

Projet WEB (JavaEE)

Web applications (JavaEE)

Atelier : 3h, Projet : 9h

Enseignants : Olivier BARAIS (U. Rennes I)

Correspondant : Farah AIT SALAHT

Enseignement destiné aux élèves de la filière « Statistique et 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.

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

Pré-requis

Contrôle des connaissances

Soutenance et note de suivi

Langue d'enseignement

Français

Références bibliographiques

UE - Applications Statistiques et Informatiques

Veille sur les médias

Technology watch through medias

Atelier : 3h

Enseignants : Farah AIT SALAHT (ENSAI)

Correspondant : Farah AIT SALAHT

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

Objectif pédagogique

L'évolution du monde informatique est constante, tant au niveau des techniques de logiciel que des ressources matérielles. Il se développe sans cesse un « jargon » informatique, de nouveaux sigles apparaissent régulièrement dans la presse spécialisée, les sociétés travaillant dans le domaine de l'informatique sont l'objet de rachats, regroupements stratégiques... L'objectif de ce projet est d'habituer les étudiants à se maintenir en état de « veille » à travers la « grande » presse informatique pour être au courant des technologies qui arrivent, des nouvelles évolutions du marché des SSII par exemple, et pour pouvoir rapidement s'adapter à de nouveaux environnements.

Chacun des étudiants de la filière doit réaliser une étude sur un sujet qu'il choisit en accord avec l'encadrant :

- étudier comment une nouvelle technologie est vue à travers plusieurs medias
- étudier sur plusieurs semaines comment évolue l'information concernant une technologie dans un media donné
- approfondir un sujet exposé dans un media.

Contenu de la matière

Pré-requis

Contrôle des connaissances

Ce projet donnera lieu à une soutenance et des discussions avec l'ensemble des étudiants.

Références bibliographiques

Langue d'enseignement

Français

UE - Applications statistiques et informatiques

Projet Anglophone : Applications Web

English project : Web applications and languages

Cours 12h ; Atelier 12h

Enseignant : Olivier Chantrel (Orange)

Correspondant : Romaric LUDINARD

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

Objectif pédagogique

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

1. L'historique du web / XML et ses applications (html, xml, dtd, web services etc.)
2. Le web actuel (css, javascript, HTML5, Ajax, MySql, Php)

Pré-requis

Contrôle des connaissances

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

Langue d'enseignement

Anglais

Séminaires professionnels

Séminaires professionnels

Seminars

Projet : 24h

Enseignants : divers

Correspondant : Farah AIT SALAHT

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

Objectif pédagogique

Découvrir de façon ludique et sans contrôle de connaissance des technologies importantes qui ne peuvent pas être intégrées dans le cursus classique de la formation.

Ces technologies sont abordées par l'exemple, directement par la pratique des entreprises ou des laboratoires de recherche

Ces séminaires thématiques seront complétés, au cours de l'année, par des séminaires de présentation des problématiques actuelles de différentes entreprises.

Contenu de la matière

Spark

ElasticStack

TensorFlow

Hive

Pré-requis

Contrôle des connaissances

Aucun.

Références bibliographiques

Langue d'enseignement

Français