SEC201 - IAML : IA et du ML pour la cybersécurité

Mis à jour le 04-12-2024

Code : SEC201

Unité d'enseignement de type cours

6 crédits

Volume horaire de référence (+/- 10%) : **50 heures**

Responsabilité nationale :
EPN05 - Informatique / 1

Contact national :

EPN05-Informatique

2 rue Conté
33.1.10A
75003 Paris

Marlène DEFFON
marlene.deffon@lecnam.net

Présentation

Prérequis

Bac+ 4 informatique

Il est fortement conseillé d'avoir suivi les unités d'enseignement suivantes : SEC102, SEC105, RCP101 ou RCP105, SEC107,

De même, la connaissance des langages python ou tout autre langage de programmation

Enfin il est recommandé de ne suivre qu'une UE 200 par semestre.

Objectifs pédagogiques

L'objectif pédagogique du cours sera d'apprendre à modéliser et concevoir des moteurs d'apprentissage artificiel simples (ML), supervisés et non supervisés susceptibles d'être utilisés dans un centre de sécurité opérationnel (SOC) en complément d'outils de gestion des informations de sécurité (SIEM). Il permettra de mettre en place une gestion des connaissances cyber (KM), à partir d'ontologies ou de graphes de connaissances. Il vous permettra également d'explorer des techniques intéressantes pour la cybersécurité comme le "process mining" (PM).

Enfin, dans un contexte où les hautes technologies évoluent rapidement, il est difficile de faire des choix structurants face à une problématique de traitement de données massives. Le cours vous "apprendra à apprendre" à maitriser ces "deep tech" à partir du module de recherche bibliographique, qui vous apprendra à avoir une démarche scientifique pour connaitre et évaluer l'état de l'art.

Compétences

Le cours vise l'acquisition de compétences élevées qui permettront de mener des activités d'extraction, d'analyses et de présentation sur les données massives présentes dans les centres de sécurité opérationnelle (SOC) à des fins d'investigation (forensic) ou d'anticipation de la menace (CTI-Hunting).

- Appliquer des prétraitements sur les donnés collectées, structurées ou non, dans un centre de sécurité opérationnelle (journaux d'évènements, états du système, base de cve, ...),
  - Prétraiter et analyser des données structurées pour répondre à un problème métier,
  - Prétraiter et analyser des données non structurées (texte, images) pour obtenir un jeu données exploitable
- Développement d'algorithmes basés sur des méthodes de machine learning ou de modélisation des connaissances, en sachant rédiger une spécification des besoins,
- Entraîner un modèle d'apprentissage :
  - supervisé pour réaliser une analyse prédictive, en l'appliquant par exemple à des moteurs de détection comportementale.
  - non supervisé pour la segmentation réduction de données en l'appliquant aux journaux d'évent collectés dans un centre opérationnel de sécurité.
- Déployer un modèle d'apprentissage automatique à l'échelle technologies du Big data (appliqué aux journaux d'évènements)
- Présenter et déployer un modèle d'apprentissage automatique auprès d'utilisateurs finaux.

Ces compétences (listées ci-dessous et issues d'offres d'emplois) sont demandées à un ingénieur informatique parcours cybersécurité :

- la compétence "Participer à la veille sur les nouveaux mécanismes de détection ainsi qu'aux nouvelles méthodes d'investigation"
- la compétence "Effectuer, à partir des scénarios d'agressions redoutés, les activités de mise sous surveillance, la traduction en règle de corrélation, la construction de la collecte des données nécessaires, la définition des réponses à incident, le pilotage de la mise en oeuvre

et la recette,
- la compétence "mettre en place des outillages d'ingénierie de la connaissance cyber visant à structurer et automatiser les phases de collecte des données puis d'extraction, de modélisation et d'enrichissement de la connaissance d'intérêt cyber à des fins de capitalisation."

Ces exemples de compétences font appel aux savoirs de conception, analyse, développement d'un prototype impliquant du machine learning (ML), de la gestion des connaissances (knowledge management (KM)) ou du process mining (PM).

Programme

Contenu

Le déploiement des enseignements s'effectue à raison d'un volume de 12 unités temps (UT).

Temps 1 : IAML pour la cyber

(IA/ML 1 UT*)

Histoire, enjeux et champ disciplinaire de l'intelligence artificielle.

Techniques de l'intelligence artificielle au service de la cybersécurité.

Fondamentaux de la détection d’anomalie à partir des données.

Typologie des données de sécurité traitées pour l'apprentissage (hétérogénité, structures, ..).

Modèle général du traitement automatique des logs.

Temps 2 : KM

(KM : 4 UT*)

Fondamentaux pour la gestion des connaissances

Langages semi-formels : ontologies et web sémantique

Temps 3 : ML

(ML : 4 UT*)

Classifications statistiques : supervisées, semi-supervisées, non supervisées

Fondamentaux pour l'apprentissage artificiel

Techniques du machine learning (Réseaux de neurone, Deep learning).

Temps 4 : PM

(PM : 1 UT*)

Généralités sur le Process Mining

Temps 5 : RB : IA/ML pour la cyber

RB: 4 UT*)

Lien avec les applications actuelles en cybersécurité au travers d'une étude bibliographique tutorée par un enseignant chercheur,

Outils de cybersécurité à base de machine learning, knowledge management et IA.

Remarques

*Par semaine, 1 UT comprend deux heures de cours, deux heures de travaux pratiques, attend quatre heures à minima de travail personnel. Chaque UT est donc espacée d'une semaine, ce rythme doit être pris en compte dans la planification des enseignements

Modalités de validation

- Contrôle continu
- Projet(s)
- Mémoire

Description des modalités de validation

Controle continu

Recherche bibliographique avec une note individuelle