FICHA DE ASIGNATURA

Título: Minería de Datos y Seguridad (Data Driven Security)

Descripción: El análisis de datos es una parte importante a la hora de gestionar la seguridad en sistemas. Tanto como para descubrir amenazas y fallos, así como para transmitir de forma inteligible la información forense adquirida. El análisis de datos en seguridad requiere el uso de las herramientas y métodos adecuados para tratar grandes volúmenes de datos, así como de su correcta visualización.

Carácter: Básica

Créditos ECTS: 3

Contextualización: El objetivo de este curso consiste en presentar el estado del arte en procesado de datos, los métodos de minería de datos, el lenguaje de programación R para datos y estadística, las herramientas gráficas de R, y herramientas para importar datos, exportar resultados, y reproducir experimentación.

Modalidad: Online

Temario:

- Tema 1: Introducción: Data Science: Cloud Computing & Big Data, Data Science, Estado del arte.
- Tema 2: Minería de Datos y R: Introducción a R, Tipos de Datos, Operaciones Básicas y Estructuras de Control, Vectorización, Funciones y Visibilidad
- Tema 3: Datos elegantes: Origen de los Datos, Lectura y Escritura de Datos (raw, xml, json, db, web…), Datos Elegantes (operaciones con Data Frames)
- Tema 4: Análisis de datos: Análisis de Datos Estructurados, Gráficos Analíticos, Clústering de Datos
- Tema 5: Visualización de datos: Conceptos Básicos, Gráficos en R (qplot y ggplot2), Gráficos con Mapas
- Tema 6: Investigación reproducible: Herramientas de Markdown, Herramientas Notebook, Librerías de R y Herramientas Git
- Tema 7: Herramientas de Seguridad en R: Exploración y Análisis de Logs (Syslog, Eventlog…), Análisis de información de Amenazas

Competencias*:*

Generales

- CG8 – Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos.

Básicas

- CB6 – Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- CB7 – Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o

limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de estos.

- CB8 – Que los estudiantes sepan comunicar sus conclusiones y conocimientos, así como las razones que las sustentan, a públicos especializados y no especializados de una forma clara y sin ambigüedades.
- CB9 – Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de una manera auto-dirigida o autónoma en gran medida.

Transversales

- CT4 – Gestionar la adquisición, estructuración, análisis y visualización de datos e información (en el ámbito informático y de seguridad), y valorar de forma crítica los resultados de esta gestión.
- CT12 – Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (en el ámbito informático y de seguridad) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

Específicas

- CE9 – Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

Actividades Formativas:

Actividad Formativa	Horas	Presencialidad
Clases Magistrales (Video-Sesiones)	6	100
Clases de Tutoría	6	100
Ejercicios prácticos Individuales	20	0
Estudio autónomo	30	0

Metodologías docentes:

- Clase magistral / método expositivo

- Plataforma virtual de aprendizaje

- Aprendizaje Cooperativo (realización de trabajos)

- Aprendizaje Basado en Problemas (ABP)

- Entornos de simulación (recreación de problemas reales)

Sistema de Evaluación:

Sistemas de evaluación	Ponderación mínima	Ponderación máxima
Presentación de trabajos y/o proyectos	20.0	40.0
Examen escrito/oral (prueba objetiva, prueba de respuesta corta y/o prueba de desarrollo).	40.0	60.0

Bibliografía:

Minería de Datos y Aprendizaje Automático:

- T. Hastie, R. Tibshirani, J. Friedman (2009) *"The elements of statistical learning: data mining, inference, and prediction"*, Springer, 2009, ISBN: 9780387848570.
- J.H. Maindonald, J. Braun (2010), *"Data analysis and graphics using R: an example-based approach"*, Cambridge University, 2010, ISBN: 9780521762939.
- R.O. Duda, P.E. Hart, D.G. Stork (2001), *"Pattern classification"* , John Wiley & Sons , 2001 , ISBN: 0-471-05669-3.

Herramientas de Minería de Datos:

- KDnuggets, (2016) *"Software para Minería de Datos"*. http://www.kdnuggets.com
- R, (2016) *"Comprehensive R Archive Network"*. http://www.cran.es.r-project.org
- Herramientas de red, (2016) *"R-Net-Tools"* https://github.com/r-net-tools

Análisis de Datos en Seguridad:

- David García (2016). *"Recopilación de Logs y Proxy"* http://www.securityartwork.es/2015/02/26/recopilacion-de-informacion-information-gathering-sobre-logs-de-proxy-i/ Securityatwork.com
- Dzidorius Martinaitis (2016). *"Data mining for Network security and Intrusion Detection"* https://www.r-bloggers.com/data-mining-for-network-security-and-intrusion-detection R-bloggers.

Uso de R avanzado:

- Hadley Wickham (2014), *"Advanced R"*, CRC Press.
- Hadley Wickham (2009), *"Plyr tutorial"* http://plyr.had.co.nz/09-user/ useR!
- Christopher Bare (2016), *"MySQL + R"* http://www.r-bloggers.com/mysql-and-r/
- Stacompute (2016), *"MongoDB + R"* https://www.r-bloggers.com/r-and-mongodb/
- SAPE research group (2016), *"ggplot2 reference"* http://sape.inf.usi.ch/quick-reference/ggplot2

Documentación sobre Markdown y Notebooks:

- John Gruber (2016), *"Markdown Basics"* http://daringfireball.net/projects/markdown/basics
- Jupyter Project, (2016) *"Jupyter Notebook QuickStart"* https://jupyter.readthedocs.io/en/latest/content-quickstart.html