**FICHA DE ASIGNATURA**

**Título:** Monitorización y Data Mining

**Descripción**: Esta asignatura trata las tendencias que impulsan la necesidad de un tratamiento correcto de evidencias que alimentan la inteligencia operacional, examinando los beneficios de implementar sistemas para apoyarla, explorando diversos casos de uso y revisando algunas de las tecnologías disponibles. Igualmente, 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***:* Obligartoria

**Créditos ECTS:** 3

**Contextualización**: La inteligencia operacional se refiere a una categoría de métodos y tecnologías que permiten dar visibilidad al negocio y descubrimiento de conocimientos para las TI en toda la organización. La inteligencia operacional no es una consecuencia de la inteligencia empresarial (BI), sino un nuevo enfoque basado en fuentes de información no típicamente en el ámbito de las soluciones de BI. Detrás de cada infraestructura de TI, detrás de los sistemas que ejecutan su negocio, se están generando masivamente flujos de datos generados por las máquinas. Las principales organizaciones se dan cuenta de que estos datos pueden ser increíblemente valiosos para mejorar la eficiencia no solo de TI, sino también de otras partes del negocio. La inteligencia operacional está diseñada específicamente para abordar esta oportunidad. Se presentará 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:**

Parte1. Monitorización

1. Introducción y registros propios del sistema
    - Introducción y registros propios del sistema
    - Filtrado de eventos en Windows
    - Auditoría de borrado de documentos
    - Alarma "para pobres" (tareas + datos de evento + script)
    - Búsqueda en otros registros de eventos
    - Configuración de Syslog para aceptar logs de SSH
    - Centralización de logs mediante Syslog remoto
    - Envío de eventos de Windows a Syslog
    - Herramientas:
    - Visor de Eventos, Editor de directivas, Programador de tareas
    - Editor del registro, Firewall de Windows

- EventCreate, EventToSyslog
- Logrotate, Syslog (Rsyslog), Logger, OpenSSH

2. Sistemas de detección y monitorización
    - Sistemas de detección y monitorización
    - Jugando con Snort (instalación, sniffer, NIDS, contenido)
    - Jugando con OSSEC (instalación, WebUI, agentes, integridad)
    - Herramientas:
    - Snort, PulledPork
    - OSSEC (OSSEC, OSSEC WUI, agentes Linux/Windows)

3. Fuentes heterogéneas de datos y correlación de logs:
    - Fuentes heterogéneas de datos
    - Acceso a la Deep Web mediante TorBrowser
    - Uso de Maltego (instalación, Twitter, Shodan)
    - Correlación de logs
    - Jugando con OSSIM (configuración, activos, vulnerabilidades, OSSEC)
    - Jugando con Splunk (instalación, datos locales y remotos, búsquedas)
    - Herramientas:
        - Tor Browser, Maltego, Deep Web, Pastebin, Twitter, CVE
        - OSSIM (OSSIM, Collectors, nmap, Nagios, OpenVAS, OSSEC)
        - Splunk (Splunk, Forwarders, Apps)

Parte 2. Data Mining.

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

**Competencias Específicas:**

CE2 – Diseñar el despliegue de sistemas de vigilancia, análisis y protección de sistemas complejos de tratamiento, almacenamiento y transmisión de datos.

CE3 – Implementar soluciones de análisis de información relevante para la ciberseguridad basados en tecnologías emergentes de tratamiento de datos.

CE8 - Crear informes de carácter ejecutivo y de carácter científico-técnico que respondan a las necesidades de comunicación en términos de ciberseguridad dentro de una organización.

**Actividades Formativas**

| Actividad Formativa | Horas | Presencialidad |
|---|---|---|
| Sesiones síncronas | 15 | 50% |
| Ejercicios pràcticos | 6 | |

**Metodologías docentes**

- Clases síncronas
- Vídeos con píldoras de conceptos teóricos
- Caso práctico
- Soporte a consultas

**Normativa específica**:

**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