Universitat Oberta de Catalunya (UOC)

Máster Universitario en Ciencia de Datos (*Data Science*)

TRABAJO FINAL DE MÁSTER

Área: Minería de datos y Machine Learning

# Detección de anomalías en entorno del Internet de las cosas

---

Autor: Gonzalo Pedro Mellizo-Soto Díaz

Tutor: Carlos Hernández Gañán

Profesor: Jordi Casas Roma

---

Madrid, 17 de marzo de 2019

# Copyright

Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada 3.0 España de CreativeCommons.

# FICHA DEL TRABAJO FINAL

| | |
|---|---|
| Título del trabajo: | Detección de anomalías en el entorno del Internet de las cosas |
| Nombre del autor: | Gonzalo Pedro Mellizo-Soto Díaz |
| Nombre del colaborador/a docente: | Carlos Hernández Gañán |
| Nombre del PRA: | Jordi Casas Roma |
| Fecha de entrega (mm/aaaa): | 06/2019 |
| Titulación o programa: | Máster Universitario en Ciencia de Datos |
| Área del Trabajo Final: | Minería de datos y Machine Learning |
| Idioma del trabajo: | Español |
| Palabras clave | Machine Learning, IOT, Anomaly Detection |

# Cita

*"Nuestro lema es: más humanos que los humanos"*

Eldon Tyrell, *Blade Runner*

# Agradecimientos

TO BE DEFINED

Si se considera oportuno, mencionar a las personas, empresas o instituciones que hayan contribuido en la realización de este proyecto.

# Abstract

In recent years the amount of connected devices has greatly increased, with an increasing number of applications in the industry each day. This devices can be subject of attacks causing instability or data leaks that can be dangerous both for the users and the enterprises, in order to avoid or confront them, security and early detection are becoming a must in a connected world. The focus is the monitoring and detection of the attacks in Internet of Things devices using state of the art Machine Learning techniques. Models such as SVM, DBScan or Isolation Forests have been used and assembled in order to identify with a better accuracy when an attack is happening. With this assembly, attack detection has increased up to 15 % comparing to traditional methods and individual model usages and times have been considerably reduced. An active use of Machine Learning models has shown a great improvement at anomaly detection by securing the devices and decreasing the reaction times when facing attacks.

Durante los últimos años se encuentra una creciente cantidad de dispositivos conectados entre sí, cada vez con más aplicaciones en la industria. Estos dispositivos pueden ser atacados y provocar inestabilidad o una fuga de datos, por lo tanto la protección y la pronta detección de ataques y/o anomalías es vital en un mundo cada vez más conectado. El objetivo es la monitorización y detección de estos ataques en dispositivos del *Internet of Things* utilizando técnicas del estado del arte de Machine Learning para su detección y poder así responder con una mayor rapidez a los ataques. Para la detección se han utilizado modelos estadísticos, como SVM, DBScan o Isolation Forests, que en su conjunto permitan identificar con mayor precisión cuando se está produciendo un ataque. El conjunto de la clusterización con la clasificación de puntos anómalos muestra una mayor robustez, frente al uso individual de cada uno de los modelos aumentando la detección en hasta un 15 %. Se demuestra cómo el uso de los modelos permite proteger los dispositivos y mejorar la seguridad al disminuir los tiempos de reacción frente a los ataques.

**Palabras clave**: Machine Learning, IOT, Anomaly Detection

# Índice general

# Índice de figuras

# Índice de cuadros

# Capítulo 1

# Introducción

## 1.1. Contexto y justificación del Trabajo

Cada vez se encuentran más dispositivos conectados entre sí, no solo en la industria, sino también en los hogares, esta conexión entre dispositivos los hace vulnerables a ataques informáticos que pueden afectar en gran medida a los usuarios, no solo pueden provocar un mal funcionamiento de los mismos, sino que también puede provocar la fuga de datos de distinta sensibilidad. La previsión en los futuros años es de cada vez estar más conectados y una pronta detección de los ataques puedo ayudar a evitar los problemas derivados de los mismos, mediante una pronta reacción o frente a la previsión de un ataque.

Actualmente se utilizan distintas medidas de seguridad como control de acceso físico al dispositivo, encriptación de datos, firewalls, securización de red, etc... Sin embargo, en muchos casos la detección de anomalías aplican un umbral estacionario, provocando que en el caso de un ataque ya sea demasiado tarde para reaccionar o no sean capaz de identificar patrones extraños previos al ataque. Con la aplicación de nuevas técnicas se pretende mejorar los tiempos de reacción e incluso predecir cuándo puede suceder un ataque.

## 1.2. Explicación de la motivación personal

La razón principal de la selección del proyecto es la posibilidad de aprender y utilizar técnicas de Machine Learning en un sector desconocido que permita diversificar conocimientos. Cada vez más todo se encuentra conectado e investigar cómo clasificar cuando se está produciendo un ataque permite profundizar en conocimientos de seguridad y aplicarlos en un problema real.

La aplicación de técnicas en un problema real ayuda a comprender mejor el uso de las herramientas y el por qué y cuándo se deben de utilizar. De este modo, se añade un nuevo conocimiento que puede aportar en el ámbito profesional y puede utilizarse también en un entorno privado.

## 1.3. Objetivos del Trabajo

Los objetivos del trabajo son los siguientes:

- Adquisición de conocimiento del sector y de los dispositivos IOT
- Lectura y comprensión de las técnicas del estado del arte en detección de anomalías en dispositivos conectados.
- Obtención de datos reales de conexiones a dispositivos IOT, en caso de no ser posible, generación de datos sintéticos.
- Prueba de las técnicas encontradas y evaluación en el problema actual.
- Investigación de algoritmos tradicionales de Machine Learning y su utilidad en el problema actual.
- Desarrollo de la solución utilizando los algoritmos más aptos.
- Evaluación de la detección de patrones y ataques de las técnicas utilizadas.
- Comparar los resultados obtenidos con el estado del arte y probar si existen mejoras frente a los métodos actuales.

## 1.4. Descripción general del problema

En la actualidad, la cantidad de dispositivos informáticos existentes que pueden ser víctimas de un ataques es masiva, por lo que securizar bien los dispositivos es fundamental. A pesar de el uso de distintos protocolos de seguridad, se generan nuevos tipos de ataque todos los días, por lo tanto la detección es una necesidad para evitar los problemas derivados.

La pérdida de control de los dispositivos o la fuga de información de los mismos puede provocar pérdidas millonarias a las empresas o provocar una gran inseguridad a los usuarios de productos IOT, así como provocar posibles daño a infraestructuras o personas. Por otro lado, muchos de los dispositivos pueden no tener la capacidad de computación necesaria para incluir una capa de seguridad robusta, que puede compensarse con una detección temprana.

## 1.5. Enfoque y método seguido

El enfoque es aplicado al conocer el problema y se intenta dar respuesta a preguntas específicas, mediante la aplicación de los conocimientos obtenidos y la evaluación de la propia aplicación de los mismos.

El método seguido del desarrollo se va a implementar dentro del marco de trabajo *Agile* pensando el desarrollo de la solución como un producto, de este modo se podrá iterar sobre una arquitectura definida y enfocarse en el desarrollo del producto en su total, frente a un modelo tradicional de cascada donde cada fase de desarrollo recae en una sola parte del total.

El formato de entrega será mediante un *minimum viable product* (MVP) en sprints de dos semanas durante el periodo de desarrollo y evaluación de las necesidades según la evolución del producto.

## 1.6. Planificación del Trabajo

Para la planificación del trabajo se han subdividido las tareas principales y se han mostrado en el siguiente diagrama de Gantt 1.1.

- Pec01 - Definición
- Pec01 - Planificación
- Pec02 - Búsqueda de fuentes del estado del arte
- Pec02 - Lectura de estado del arte
- Pec02 - Justificación de estado del arte
- Pec02 - Redacción
- Pec02 - Refinamiento de objetivos
- Pec03 - Planificación de Sprints
- Pec03 - Sprint 1
- Pec03 - Sprint 2
- Pec03 - Sprint 3
- Pec03 - Refinamiento/Redacción
- Pec04 - Revisión apartados anteriores
- Pec04 - Redacción nuevos apartados
- Pec05 - Presentación y defensa

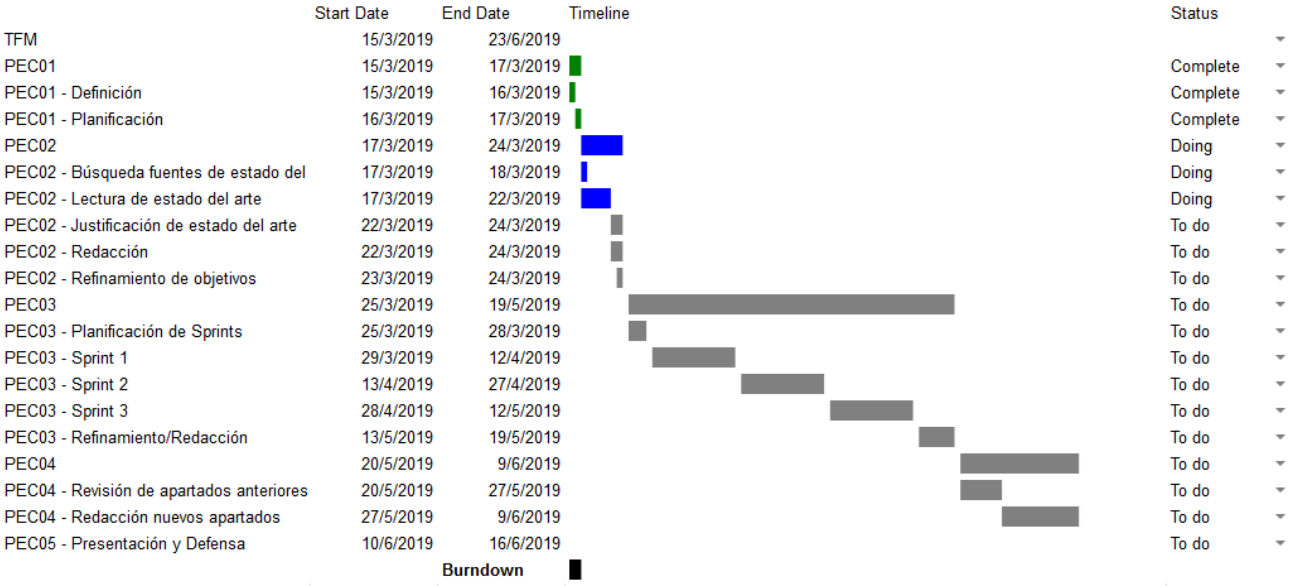


Figura 1.1: Planificación de tareas

# Bibliografía