

**Grado en Ingeniería Informática de Gestión y Sistemas de
Información**

**Abnormal Behavior Identification through
Deep Learning: TensorFlow**

Estudio de viabilidad

Robert Torrell Belzach
TUTOR: Xevi Font

Curso 2021 - 2022

Planificación inicial y presupuesto

Planificación inicial

El siguiente diagrama de Gantt describirán las diferentes tareas a realizar durante todo el proyecto:



En azul claro, se marcan los entregables que se irán escribiendo a lo largo del periodo, en rojo, el período de los diferentes ciclos de desarrollo, y en púrpura, la preparación de la Defensa ante el jurado.

Fíjese que la duración de las tareas no es siempre la misma, especialmente en el caso de los ciclos de desarrollo. Los primeros ciclos son muy breves debido a que no hará falta mucho esfuerzo para implementarlos, además de que existen librerías de código que ya implementan a la perfección estos cálculos. Sobre todo en esos ciclos, el mayor reto sería el de la manipulación de los datos, pero con ese ligero margen de tiempo, ya debería de ser suficiente.

La redacción de los entregables, así como el desarrollo del proyecto, se realizan simultáneamente, pero una vez entregado el último documento, se dedicarán cuatro semanas solamente a la presentación y defensa del proyecto ante jurado.

Presupuesto

El proyecto se realiza desde el mes de Diciembre, ya comenzando desde la documentación sobre el campo de estudio a aplicar al problema, hasta Julio en que se presentará el proyecto ante el jurado. Si durante este periodo se trabaja en el proyecto durante tres horas diarias, a lo largo de todo el proyecto se habrán invertido unas cuatrocientas cincuenta horas.

Si el salario medio de un ingeniero Informático en España es de trece euros y medio por hora [19], el salario que se ha de invertir a cada miembro del equipo sería aproximadamente de seis mil setenta y cinco euros.

El coste del equipo informático en el que se ha de desarrollar el proyecto no es negligible. Las herramientas y librerías para desarrollar algoritmos de Inteligencia Artificial se benefician del uso de una tarjeta gráfica potente, así como de grandes cantidades de almacenamiento, memoria y un chip CPU de alto rendimiento. El día en que se ha redactado este documento, un ejemplo de un equipo informático óptimo para el desarrollo del proyecto valdría aproximadamente de mil setecientos euros (<https://www.pccomponentes.com/pccom-gold-amd-ryzen-5-5600x-16gb-1tbssd-rtx3060ti>)

Así pues, la inversión final del proyecto es de ocho mil euros, si este es realizado por tan solo un desarrollador.

Alternativamente, se puede contratar un servicio de desarrollo y ejecución Cloud que admita un entorno de trabajo y ejecución de las tecnologías necesarias para llevar a cabo el proyecto. Si, por ejemplo, usamos la solución Amazon SageMaker de AWS, nos costaría aproximadamente trescientos cincuenta euros al mes.

Análisis de viabilidad

Viabilidad Técnica

Hoy en día vivimos en una “edad de oro” en la investigación de la Inteligencia Artificial. Muchas de las técnicas y algoritmos más pioneros que existen, aunque se definieron a partir de los 1980, no han podido ser aplicadas hasta los últimos quince años, gracias las capacidades de cómputo y memoria de los equipos de hoy en día [20], las estructuras de computación paralelizada y arquitecturas de hardware SIMD comunes en GPUs [21] . Además, el creciente interés tanto por profesionales, investigadores e inversores en este campo de estudio, forman en la actualidad un entorno de investigación abierto, repleto de recursos, proyectos y herramientas de código libre.

Para poder desarrollar el software del proyecto, tenemos a nuestra disposición un gran número de lenguajes de programación, librerías y entornos de desarrollo, todos de gran calidad y utilizados efectivamente en numerosos proyectos exitosos.

Los dos lenguajes comunes para estos proyectos son Python y R [22], y las librerías más populares son...

... para Python: Numpy, Scipy, Matplotlib, Tensorflow, Keras, Pytorch.

... para R: Dplyr, Ggplot2, DataExplorer, Caret, DataExplorer, Superml.

Todas las tecnologías expuestas anteriormente son libres y gratuitas, fáciles de obtener y de comenzar a usar. Además, hay mucha documentación extensa sobre el software en Internet y en la literatura académica y didáctica.

La base teórica en la que se basa el proyecto es compleja y muy variada. Hace falta varios meses de investigación para poder adentrarse al proyecto desde los conocimientos de principiante sobre este campo de estudio. Ello conlleva una dificultad elevada, pero como se ha dicho previamente, tener tanta información y herramientas accesibles ayudan mucho al proceso de aprendizaje.

Viabilidad Técnica

La industria de la manufacturación es inmensa, muy presente en España y no para de crecer. Cada año, estas industrias contribuyen cada vez más al valor del Producto Interno Bruto del estado [23], y cada vez hay más inversión en el Desarrollo y la Investigación [24].

Debido a que el gasto económico más elevado que hay en el proyecto es el salario del desarrollador que realice el proyecto, más una suma de no más de dos mil euros por todo el Hardware + gasto energético, este proyecto requiere una inversión muy baja en comparación de los potenciales beneficios en aumento de fiabilidad y productividad de los procesos de manufacturación, así como el potencial interés que hay en el mercado por la contratación y uso de soluciones que asistan directamente a estas industrias.

Viabilidad medioambiental

Este proyecto, aunque apenas requiere de materiales, si necesita un fuerte consumo energético. De la misma manera que otras tecnologías como el Blockchain han sido fuertemente acusadas de causar graves daños en el ecosistema [25], si se le da uso a hardware SIMD como las GPU, se puede acabar por causar daños similares, sobre todo si se aplica esta solución en centros de grandes dimensiones.

Realizar el desarrollo en un terminal de trabajo que posea un consumo reducido (como un ordenador personal) y siempre asegurar que se entrenan y calcularán los modelos predictivos cuando haga falta y con métodos óptimos formará parte de la responsabilidad del desarrollador. Uno de los algoritmos ha sido desarrollado y ejecutado en un entorno “Cloud”, pero se ha ejecutado el código siempre con el programador presente ante el equipo, y nunca dejando que el proceso tarde mas de dos horas en terminar su ejecucion.

Aspectos legales

Este proyecto da uso de tecnologías que exclusivamente posean licencias libres y/o abiertas, de uso gratuito y accesibles para todos.

Python está protegido bajo la Python Software Foundation License, Tensorflow bajo Apache License 2.0 y Keras bajo MIT License.

Aunque todas estas licencias son diferentes, a la vez todas están inspiradas bajo los mismos principios del Código Abierto, que permiten el uso, estudio, redistribución y modificación de estas herramientas y el código fuente de estos, permitiéndonos trabajar con ellas sin apenas restricciones legales.