



---

**Universidad de Valladolid**

E.T.S Ingeniería Informática

Trabajo Fin de Grado

Grado en Ingeniería Informática,  
mención en Computación

Algoritmos para Big Data

Autor:  
Sergio García Prado



---

**Universidad de Valladolid**

E.T.S Ingeniería Informática

Trabajo Fin de Grado

Grado en Ingeniería Informática,  
mención en Computación

Algoritmos para Big Data

Autor:

Sergio García Prado

Tutor:

Manuel Barrio Solórzano

# Prefacio

Para entender el contenido de este documento así como la metodología seguida para su elaboración, se han de tener en cuenta diversos factores, entre los que se encuentran el contexto académico en que ha sido redactado el mismo, así como el tecnológico y social. Es por ello que a continuación se expone una breve descripción acerca de estos para tratar de facilitar la comprensión acerca del alcance de este documento.

Lo primero que se debe tener en cuenta es el contexto académico en que se ha llevado a cabo. Este se ha redactado para la asignatura de **Trabajo de Fin de Grado** para el *Grado de Ingeniería Informática* impartido en la *Universidad de Valladolid*. Dicha asignatura se caracteriza por ser necesario haber superado el resto de las asignaturas que componen los estudios de grado para su evaluación. Su carga de trabajo es de **12 créditos cts**, cuyo equivalente temporal es de *300 horas* de trabajo del alumno, que se ha llevado a cabo en un periodo de 4 meses.

La temática escogida para realizar dicho trabajo es **Algoritmos para Big Data**. El Big Data es la disciplina que se encarga de “todas las actividades relacionadas con los sistemas que manipulan grandes conjuntos de datos. Las dificultades más habituales vinculadas a la gestión de estas cantidades de datos se centran en la recolección y el almacenamiento, búsqueda, compartición, análisis, y visualización. La tendencia a manipular enormes cantidades de datos se debe a la necesidad en muchos casos de incluir dicha información para la creación de informes estadísticos y modelos predictivos utilizados en diversas materias”[1]

Uno de los puntos más importantes para entender la motivación por la que se ha escogido dicha temática es el contexto social en que nos encontramos. Debido a la importante evolución que están sufriendo otras disciplinas del mundo de la informática y las nuevas tecnologías cada vez es más sencillo y económico recoger una gran cantidad de información de cualquier proceso que se da en la vida real. Esto se debe a una gran cantidad de factores, entre los que se destacan los siguientes:

- **Reducción de costes derivados de la recolección de información:** Debido a la constante evolución tecnológica cada vez es más barato disponer de mecanismos (tanto a nivel de hardware como de software) a partir de los cuales se puede recabar datos sobre un determinado suceso.
- **Mayor capacidad de cómputo y almacenamiento:** La recolección y manipulación de las grandes cantidades de datos que se recogen a partir de sensores u otros métodos requieren por tanto del apoyo de altas capacidades de cómputo y almacenamiento. Actualmente esto se está llevando a cabo mediante técnicas de virtualización que permiten gestionar sistemas de gran tamaño ubicados en distintas zonas geográficas como una unidad.
- **Mejora de las telecomunicaciones:** Uno de los factores que ha permitido una gran disminución de problemática relacionada con la virtualización y la capacidad de respuesta ha sido el gran avance

a nivel global que han sufrido las telecomunicaciones en los últimos años permitiendo disminuir las barreras geográficas entre los sistemas tecnológicos.

[Hablar sobre técnicas de tratamiento de grandes conjuntos de datos]

[Hablar sobre algoritmos de manipulación de grandes conjuntos de datos]

# Índice general

1. Introducción	4
A. ¿Cómo se ha redactado este documento?	5

# Capítulo 1

## Introducción

## Apéndice A

**¿Cómo se ha redactado este documento?**

# Bibliografía

- [1] Wikipedia: Big Data,  
[https://es.wikipedia.org/wiki/Big\\_data](https://es.wikipedia.org/wiki/Big_data) 2017-02-03