



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería  
Informática**

**Sistema de Recomendación  
basado en Aprendizaje  
Profundo**



Presentado por Raúl Negro Carpintero  
en Universidad de Burgos — 24 de febrero  
de 2019

Tutor: Bruno Baruque Zanón







UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



D. Bruno Baruque Zanón, profesor del departamento de nombre departamento, área de nombre área.

Expone:

Que el alumno D. Raúl Negro Carpintero, con DNI 71305764Z, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado título de TFG.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 24 de febrero de 2019

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

D. nombre tutor

D. nombre co-tutor





## Resumen

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

## Descriptores

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android ...

## **Abstract**

A **brief** presentation of the topic addressed in the project.

## **Keywords**

keywords separated by commas.



---

# Índice general

---

Índice general	III
Índice de figuras	V
Índice de tablas	VI
Introducción	1
Objetivos del proyecto	3
2.1. Objetivos técnicos . . . . .	3
2.2. Objetivos personales . . . . .	3
Conceptos teóricos	5
3.1. Sistemas de recomendación . . . . .	5
3.2. Medidas de calidad . . . . .	6
3.3. Imágenes . . . . .	7
3.4. Tablas . . . . .	8
Técnicas y herramientas	11
4.1. Metodologías . . . . .	11
4.2. Control de versiones . . . . .	11
4.3. Repositorio . . . . .	11
4.4. Gestión del proyecto . . . . .	11
4.5. Entorno de Desarrollo Integrado (IDE) . . . . .	12
4.6. Librerías . . . . .	12
4.7. Datasets . . . . .	12
Aspectos relevantes del desarrollo del proyecto	15

5.1. Metodologías . . . . .	15
5.2. Formación . . . . .	15
5.3. Desarrollo del código . . . . .	16
5.4. Documentación . . . . .	16
<b>Trabajos relacionados</b>	<b>17</b>
<b>Conclusiones y Líneas de trabajo futuras</b>	<b>19</b>

---

# Índice de figuras

---

3.1. Autómata para una expresión vacía . . . . .	8
--	---

---

# Índice de tablas

---

3.1. Herramientas y tecnologías utilizadas en cada parte del proyecto	9
---	---

---

# Introducción

---

Descripción del contenido del trabajo y del estructura de la memoria y del resto de materiales entregados.



---

# Objetivos del proyecto

---

## 2.1. Objetivos técnicos

- Obtener un sistema de recomendación utilizando redes neuronales.
- Comparar el sistema obtenido con un modelo inicial.

## 2.2. Objetivos personales

- Poner en uso los conocimientos adquiridos durante la carrera-
- Ahondar más en el campo del aprendizaje profundo.





---

# Conceptos teóricos

---

Los conceptos teóricos más importantes del proyecto son los relacionados con los sistemas de recomendación y las redes neuronales.

## 3.1. Sistemas de recomendación

Los sistemas de recomendación son herramientas que generan recomendaciones sobre un determinado objeto de estudio, a partir de las preferencias y opiniones dadas por los usuarios.[?]

En un sistema de recomendación hay dos clases de entidades: usuarios e ítems. Los datos están representados en una matriz de utilidad de tal manera que los usuarios son filas y los ítems columnas. Para cada par hay un valor que representa el grado de preferencia de un usuario para un ítem. Se asume que la mayoría de los valores se desconocen. Es por ello, que el objetivo de un sistema de recomendación es rellenar esos espacios en blanco.

### Tipos de sistemas de recomendación

Existen dos tipos de sistemas de recomendación:

- Content-based systems
- Collaborative filtering systems

### Content-based systems

Examinan propiedades de los ítems recomendados. Solo tienen en cuenta los gustos de los usuarios.

En estos sistemas, para cada ítem hay que construir un perfil compuesto por las propiedades del mismo. Por ejemplo, si los ítem son películas, algunas de las propiedades serían género, director, fecha de estreno, actores...

## Collaborative filtering systems

Recomiendan ítems basándose en los gustos de usuarios similares.

En este caso, en lugar de usar el vector ítem-perfil de un ítem, usamos las filas de la matriz de utilidad. Los usuarios son parecidos si sus filas se acercan de acuerdo a alguna distancia.

## 3.2. Medidas de calidad

Podemos hacer uso de diferentes métricas para conocer cómo de bueno nuestro sistema de recomendación es. Las métricas elegidas son [?]:

- Cobertura
- Serendipia

### Cobertura

La cobertura o coverage es el dominio de ítems sobre los que el sistema de recomendación puede recomendar. Podemos distinguir dos tipos de cobertura: la cobertura de predicción (porcentaje de ítems sobre los que se puede recomendar) y la cobertura de catálogo (porcentaje de ítems que son recomendados).

La que vamos a utilizar es la primera, y su fórmula es:

$$PC = \frac{|Ip|}{|I|} \quad (3.1)$$

donde  $Ip$  es el conjunto de ítems sobre los que se puede hacer una predicción e  $I$  es el conjunto de todos los ítems disponibles.

### Serendipia

La serendipia o serendipity es una medida referente a si una recomendación es atractiva y sorprendente para el usuario. Para calcularla, primero tenemos

que saber cuáles son las recomendaciones que sorprenden al usuario. Esto se calcula de la siguiente manera:

$$UNEXP = RS \setminus PM \quad (3.2)$$

donde  $RS$  es una recomendación por nuestro sistema de recomendación y  $PM$  es una recomendación de un modelo primitivo. Una vez hecho esto, ya podemos calcular la serendipia como:

$$SRDP = \frac{\sum_{i=1}^N u(RSi)}{N} \quad (3.3)$$

donde  $RSi$  es un elemento de  $UNEXP$  y  $N$  es la longitud de  $UNEXP$ . Si  $u(RSi)$  es igual a 1, entonces la recomendación es atractiva; y si es 0, la recomendación no es atractiva y no vale.

### 3.3. Imágenes

Se pueden incluir imágenes con los comandos standard de  $\text{\LaTeX}$ , pero esta plantilla dispone de comandos propios como por ejemplo el siguiente:



Figura 3.1: Autómata para una expresión vacía

### 3.4. Tablas

Igualmente se pueden usar los comandos específicos de  $\text{\LaTeX}$  o bien usar alguno de los comandos de la plantilla.

Herramientas	App	AngularJS	API REST	BD	Memoria
HTML5		X			
CSS3		X			
BOOTSTRAP		X			
JavaScript		X			
AngularJS		X			
Bower		X			
PHP			X		
Karma + Jasmine		X			
Slim framework			X		
Idiorm			X		
Composer			X		
JSON		X	X		
PhpStorm		X	X		
MySQL				X	
PhpMyAdmin				X	
Git + BitBucket		X	X	X	X
MikTeX					X
TeXMaker					X
Astah					X
Balsamiq Mockups		X			
VersionOne		X	X	X	X

Tabla 3.1: Herramientas y tecnologías utilizadas en cada parte del proyecto



---

# Técnicas y herramientas

---

## 4.1. Metodologías

### Scrum

Descripción

## 4.2. Control de versiones

### Git

Descripción

## 4.3. Repositorio

### GitHub

Descripción

## 4.4. Gestión del proyecto

### ZenHub

Descripción

## 4.5. Entorno de Desarrollo Integrado (IDE)

### Python

- Herramientas consideradas: [Jupyter](#), [Spyder](#) y [PyCharm](#)
- Herramienta escogida: Jupyter [Jupyter](#)

Justificación

### Documentación

- Herramientas consideradas: [Texmaker](#) y [Apache OpenOffice](#)
- Herramienta escogida: [Texmaker](#)

Justificación

## 4.6. Librerías

### NumPy

[NumPy](#) descripción

### SciPy

[SciPy](#) descripción

### LightFM

[LightFM](#) descripción

## 4.7. Datasets

### Anime

Este conjunto de datos [?] contiene información relativa a las preferencias de 73.516 usuarios sobre 12.294 películas o series de anime. Cada usuario ha valorado unos 106 animes.

El conjunto de datos se encuentra disponible en: [Anime Dataset](#).



## Book-Crossing

Este conjunto de datos [?] contiene 1.149.780 valoraciones de 278.858 usuarios sobre 271.379 libros. Hay alrededor de 11 valoraciones por usuario.

El conjunto de datos se encuentra disponible en [Book-Crossing Dataset](#).

## MovieLens

Este conjunto de datos [?] contiene 100.000 valoraciones de 943 usuarios sobre 1.682 películas. Hay unas 106 valoraciones por usuario.

El conjunto de datos se encuentra disponible en los propios archivos descargados por la librería LightFM [?].

## Last.FM

Este conjunto de datos [?] contiene las veces que 1.892 usuarios han escuchado a 17.632 artistas. Cada usuario ha escuchado alrededor de unos 49 artistas.

El conjunto de datos se encuentra disponible en [Last.FM](#).

## Dating Agency

Este conjunto de datos [?] contiene 17.359.346 valoraciones de 135.359 usuarios sobre 168.791 perfiles de otros usuarios. Hay unas 128 valoraciones por usuario.

El conjunto de datos se encuentra disponible en [Dating Agency](#).



---

# Aspectos relevantes del desarrollo del proyecto

---

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto.

## 5.1. Metodologías

Para la gestión del proyecto se ha adaptado la metodología ágil Scrum. Las principales características han sido:

- La duración de los sprints fue de dos semanas aproximadamente.
- Los sprints finalizaban con una reunión en la que se revisaba el trabajo hecho y se planteaban las tareas del siguiente sprint.
- Se utilizó ZenHub para llevar el seguimiento de las tareas.

## 5.2. Formación

El proyecto ha requerido obtener una serie de conocimientos que no se tenían inicialmente. Se ha estudiado el campo de los sistemas de recomendación y la utilización de aprendizaje profundo en Python. Para la formación en los sistemas de recomendación se usó el libro:

- *Mining of Massive Datasets* (Jure Leskovec, Anand Rajaraman y Jeffrey D. Ullman).

Para la formación en aprendizaje profundo en Python se realizó el siguiente curso:

- *Practical Deep Learning For Coders, Part 1* (Jeremy Howard).

### 5.3. Desarrollo del código

El proyecto se ha centrado en dos aspectos fundamentales: obtener un modelo clásico y obtener un modelo basado en aprendizaje profundo. Para la obtención del modelo clásico se estudió, en primer lugar, hacer uso de la librería Crab. Esta opción se descartó debido a la cantidad de fallos que saltaron durante la instalación (no se pudo llegar a instalar) y a la falta de actividad por parte de sus autores en el repositorio (no se ha tocado desde hace 7 años). Por todo esto, se decidió utilizar la librería LightFM.

### 5.4. Documentación

Las opciones que se plantearon para realizar la documentación fueron Apache OpenOffice y Texmaker. Ya que el Trabajo Fin de Grado lo veo como una forma de aprender conceptos nuevos que se desmarcan un poco de lo visto durante la carrera, opté por Texmaker debido a la novedad y al querer aprender y usar  $\text{\LaTeX}$  (es la primera vez que lo utilizo).

---

## Trabajos relacionados

---

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.



---

## **Conclusiones y Líneas de trabajo futuras**

---

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.





---

## Bibliografía

---

- [1] Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proceedings of the 5th ACM conference on Recommender systems*, RecSys 2011, New York, NY, USA, 2011. ACM.
- [2] CopperUnion. Anime recommendations database. Technical report, Kaggle, 2017.
- [3] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *In RecSys '10*, page 257, 2010.
- [4] GroupLens. Movielens. In *MovieLens*, 1998.
- [5] Maciej Kula. Metadata embeddings for user and item cold-start recommendations. In Toine Bogers and Marijn Koolen, editors, *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 16-20, 2015.*, volume 1448 of *CEUR Workshop Proceedings*, pages 14–21. CEUR-WS.org, 2015.
- [6] Cai-Nicolas Ziegler. Book-crossing dataset. Technical report, IIF, 2004.