

Machine learning for tex

Carlos Andres Hernandez Gomezm

1 de mayo de 2019



Índice general

1. Introduccion	5
1.1. Machine learning for text	5



Capítulo 1

Introduccion

Este documento presenta los apuntes realizados del libro *Machine Learning for Text* con ISBN 978-3-319-73531-3 (eBook, 2018) de Charu C. Aggarwal un distinguido investigador de IBM; El motivo de redactar este documento es aprender sobre el procesamiento del lenguaje natural para ser aplicado en el trabajo terminal (TT) **Recolector y clasificador de noticias** con número 2019-B013, de la **Escuela superior de cómputo (ESCOM)** del **Instituto politécnico nacional (IPN)**.

1.1. Machine learning for text

La extracción de información útil con varios tipos de algoritmos estadísticos es denominado **Extracción de datos**(*Text mining*), **Analítica de texto** (*Text analytics*) o **Aprendizaje automático para texto** (*Machine learning for text*), en este documento se utilizará de forma indistinta. En los últimos años este campo ha incrementado por el desarrollo de la web, redes sociales, correos electrónicos, bibliotecas virtuales. Algunos ejemplos para obtener información son:

- **Bibliotecas digitales:** El uso de la información electrónica ha superado la producción de libros y publicaciones impresas, este fenómeno ha proliferado la producción de bibliotecas digitales, estas pueden ser almacenadas y ser usadas para extraer información útil.
- **Noticias electrónicas:** Existe un movimiento masivo para pasar las noticias impresas hacia la publicación electrónica, esto permite que sean almacenadas para su análisis y extracción de información sobre eventos y perspectivas importantes. Sitios como *Google news* etiquen las noticias para hacer recomendaciones al lector basado en su anterior comportamiento o intereses específicos.

- **Web and Web-enabled applications:** La web contiene una gran cantidad de información en hipertexto, con links y otro tipo de recursos, la cual puede ser utilizada par el proceso de descubrimiento de nuevo conocimiento, al igual las *Web-enabled applications*¹ permiten obtener información que puede ser analizada.

¹ **Web and Web-enabled applications:** Aplicaciones de escritorio que son accedidas remotamente des de un buscador como internet explorer.
