

Web Development

Johan Madroñero Cuervo

2023 - 1S

Contents

1	Lecture	3
1.1	Introducción	3
1.2	Motivación	3
1.3	Pandas	3
1.4	Webscrapping	3
1.5	Análisis exploratorio de datos	3
1.6	Técnicas de Machine Learning	4
1.7	Sciki-learn	4
1.8	Análisis de sentimientos	4
1.9	Regresión lineal simple	4
1.10	Fundamentos de Analítica	5
1.10.1	Caso 1:	5
1.10.2	Sistemas de Bases de Datos Relacionales	5
1.10.3	Data Mining & Business Intelligence 1989	5
1.10.4	Método Crisp	6
1.10.5	Business Intelligence 2.0 (1996)	6
1.10.6	Data Science	6
1.10.7	Servicios Web	6
1.10.8	Ingeniería de Datos, Análisis de Datos	6
1.10.9	Tradicional Analytics	6
1.10.10	Datos semi-estructurados y bases de datos NoSQL (2009)	7
1.10.11	Open Data Science & Modern Analytics (2009)	7

1 Lecture

1.1 Introducción

- Profesor: Alberto Espinosa Bedoya
- Correo: aespinosa@unal.edu.co
- Oficina: M8A-204
- Horario de atención: Lunes 9 - 12, lunes 14 - 15:30
- **10** laboratorios en el semestre.
- No es posible que sean secciones virtuales debido a la facultad.

1.2 Motivación

En la actualidad los datos son considerados activos dentro de las organizaciones.

1.3 Pandas

Librería usada para la gestión y manejo de datos. Permite acciones como:

- Limpieza de datos
- Adecuación de datos
- Integrarlo con Matplotlib \implies graficación de datos.

1.4 Webscrapping

Extraer información de páginas web, depurar, adecuar y procesar los datos.

- **Seaborn:** Herramienta complementaria en el proceso de *webscrapping*.

1.5 Análisis exploratorio de datos

Se busca identificar:

- Relaciones entre datos.
- Tendencias.
- Patrones.

Esto se conoce como modelos.

1.6 Técnicas de Machine Learning

- Redes neuronales.
- Redes convolucionales.
- Máquinas de soporte vectorial.
- Árboles de decisión.
- Random Forrest.

1.7 Sciki-learn

Librería de Python para procesos de Machine Learning.

- Procesamiento de datos i.e imágenes, texto.
- Transformaciones de datos.
- Técnicas basadas de machine Learning.

1.8 Análisis de sentimientos

Técnica usada para identificar y clasificar comentarios. Por ejemplo, reseñas, trinos, comentarios y el *Sistema* debe conseguir identificar si hay comentarios positivos o negativos en base a las palabras y expresiones.

1.9 Regresión lineal simple

Definida en Estadística y popularmente usada como introducción.

1.10 Fundamentos de Analítica

1.10.1 Caso 1:

Se desea clasificar de manera automática el estilo de conducción de una persona (pasivo, agresivo) a partir de las señales fisiológicas del conductor a través de un *Smartwatch*. Dado que se ha encontrado correlación en la conducción agresiva y la accidentalidad.

- Ritmo cardíaco.
- Presión arterial.
- Ciclo de sueño.

La cantidad de datos generada por el *Smartwatch* se convierte en un problema para el **humano** a la hora de analizar la información a partir de esta información de busca identificar patrones. *Patrones es conocimiento*. Procesos de aprendizaje automático \Rightarrow predicción, recomendación.

Definición 1. Conocimiento explícito: Procesos fórmulas, algoritmos.

Definición 2. Conocimiento implícito: Datos, historia, experiencias.

Smartwatch \Rightarrow Tabla de datos \Rightarrow Problemas de conexión \Rightarrow depuración \Rightarrow Análisis de datos \Rightarrow Construcción de un modelo \Rightarrow Entrenamiento y verificación del modelo \Rightarrow Implementación de modelo en una **App**.

1.10.2 Sistemas de Bases de Datos Relacionales

El Sistema de datos relaciones me permite crear tablas y relaciones entre las tablas, como las relaciones entre estudiantes y cursos.

- **Reporte 1:** Mostrar los estudiantes matriculados en el curso de cálculo diferencial.
- **Reporte 2:** Mostrar todos los estudiantes y cursos.

El modelo relacional presenta problemas al almacenar muchos datos al hacerse muy lentos. Claramente se ha buscado solución al problema con otras técnicas. Como el *SQL* y los *DataMarts*, *Data warehouse*.

1.10.3 Data Mining & Business Intelligence 1989

Aplicar inteligencia de negocios a datos, Recuperar gran cantidad de datos.

1.10.4 Método Crisp

Infiere los problemas del de una empresa en base a los datos.

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Distribución

1.10.5 Business Intelligence 2.0 (1996)

Software para entender el *business*.

1.10.6 Data Science

Áreas de conocimiento de la ciencia de los datos:

- Aprendizaje de máquinas
- Ingeniería de datos (Bases de datos, data lakes...)
- Análisis de datos (patrones, covarianza, métricas y verificadores...)

1.10.7 Servicios Web

Computación local, cloud computing y servicios proveedores.

1.10.8 Ingeniería de Datos, Análisis de Datos

Definición 3. La **identificación de características** es el proceso mediante el cual se extrae información de los datos que "*los describen*" de una manera apropiada.

Hadoop Permite el manejo de grandes volúmenes de datos.

- Velocidad.
- Eficiencia.

1.10.9 Tradicional Analytics

Arrancamos desde entender el problema hasta poder implementar una solución real de análisis de los datos.

1.10.10 Datos semi-estructurados y bases de datos NoSQL (2009)

Formato *JSON*, *YAML*, *Hash Maps*,...

1.10.11 Open Data Science & Modern Analytics (2009)

Profesional requerido en la actualidad y que debe conocer sobre múltiples campos.