

Data mining i aprenentatge automàtic

Laia Subirats Maté

26 de maig de 2022

Introducció

- William Edwards Deming (1900-1993)
- *"Without data, you're just another person with an opinion"*



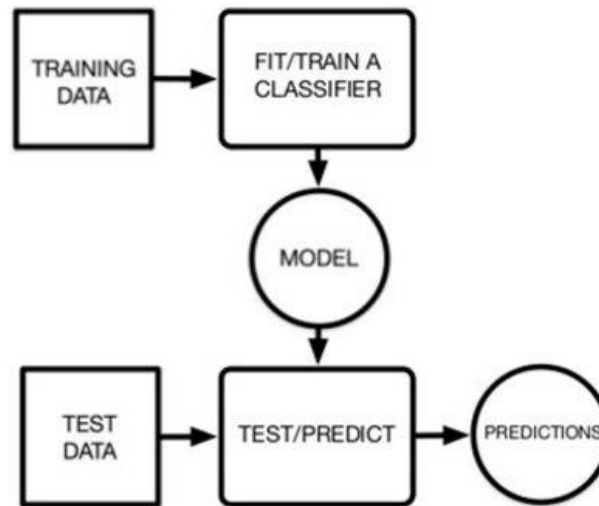
Data mining

La tasca d'anàlisi de dades es pot fer diferents maneres:

- Anàlisi estadístic clàssic: va bé per conjunts de dades amb pocs atributs.
- Regles de decisió: per poques variables
- Aprenentatge automàtic: casos complexos

Què és l'aprenentatge automàtic?

- Aprendre a partir de l'experiència, a partir d'un conjunt de dades complex on no és possible fer servir mètodes tradicionals
- En anglès machine learning (ML)



Data mining

- **Passos:**

- 1.Plantejar-se una pregunta**

- 2.Aconseguir dades**

- 3.Processar les dades**

- 4.Anàlisi de dades**

- 5.Prendre una decisió i actuar**

Camps d'aplicació

eurecat
Centre Tecnològic de Catalunya

Eurecat Àmbits de Coneixement Sectors Serveis Projectes Actualitat Contacte in o t f Català Q



Aeronàutica



Alimentació



Automoció



Comerç i Turisme



Construcció



Energia i Recursos



Esport



Ferrovial



Finances i assegurances



Indústries de Base i Bens d'Equip



Indústries Culturals i Creatives



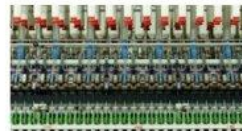
Química



Salut



Sector Públic



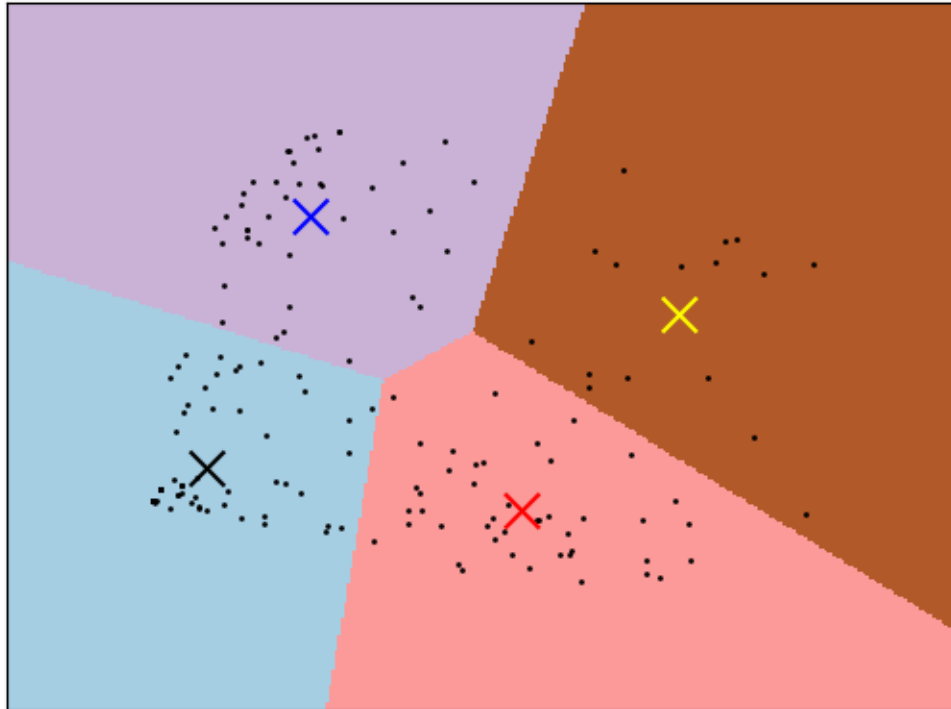
Tèxtil



Tecnologies de la informació i de les Comunicacions (TIC)

Aprenentatge no supervisat: clustering

K-means clustering on the Roman dataset (PCA-reduced data)
Centroids are marked with white cross



Aprenentatge supervisat: Regressió

Predicció de vendes de comestibles corporació favorita (Gabriel Kreplak 17% dels millors participants de Kaggle 269/1675)
Algorisme LightGBM

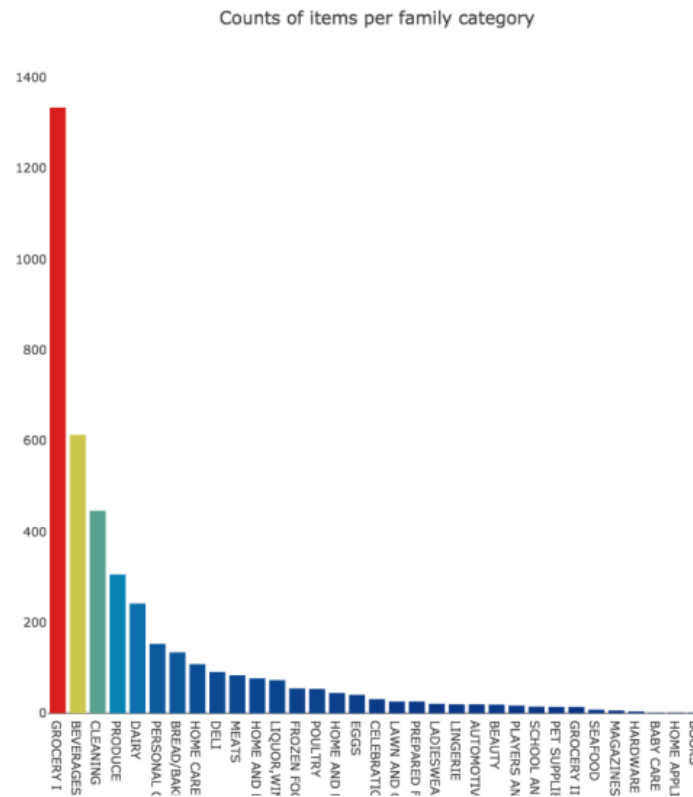
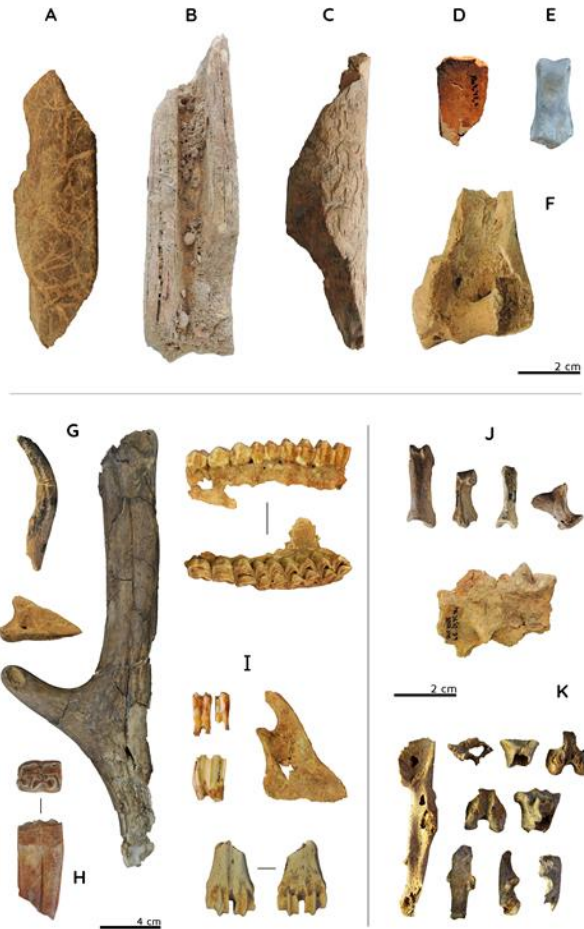


Figura 5: Número de productos por cada categoría

Aprenentatge supervisat: classificació



Aprenentatge supervisat – classificació



Intelligent Tutor System for archaeological sites

Warning: The name and surname form fields are required.

Name*:

What's your name?

Surname*:

What's your surname?

X

Insert value

Y

Insert value

Z

Insert value

Hueso: {0:'Ca', 1:'T', 2:'Cx', 3:'F', 4:'H', 5:'U', 6:'Fa1', 7:'Mtp', 8:'Mt2', 9:'Mc4', 10:'Mt4', 11:'Cr', 12:'VI', 13:'Vcd', 14:'Mc2', 15:'Ta', 16:'Hem', 17:'I/I', 18:'Fa', 19:'Mt3', 20:'Vc', 21:'I/I', 22:'Mx', 23:'As', 24:'Pa', 25:'R', 26:'Es', 27:'Fa2', 28:'Fa3', 29:'Mt5', 30: nan, 31:'Mc5', 32:'Ct', 33:'Vs', 34:'Vt', 35:'Mc3', 36:'V', 37:'Das', 38:'Dal', 39:'Da', 40:'Esq', 41:'Esp', 42:'Lt1', 43:'L', 44:'In', 45:'Lt2', 46:'Pi', 47:'Mt', 48:'M3', 49:'M1', 50:'Hy', 51:'Com', 52:'Lt3', 53:'Mc', 54:'P2', 55:'P4-M1', 56:'Art', 57:'M1', 58:'Cc', 59:'PT', 60:'M2', 61:'M/P', 62:'M1-M2', 63:'P4', 64:'M3', 65:'M1-M2', 66:'P2', 67:'Et', 68:'I', 69:'M/X', 70:'P/M', 71:'R/U', 72:'Co', 73:'P4', 74:'P3-P4', 75:'Se', 76:'dp2', 77:'P3', 78:'M/x', 79:'P4-P3', 80:'M3', 81:'PTPer8', 82:'PTPie', 83:'PTPL', 84:'P3', 85:'M', 86:'M2', 87:'Epl', 88:'M', 89:'Fa1-V', 90:'Fa2-II', 91:'Fa2-IV', 92:'Carpal', 93:'I/C', 94:'T-Ta', 95:'Fi', 96:'Ta-Mt', 97:'Asta', 98:'Coprolito', 99:'I/3', 100:'I4', 101:'d/3', 102:'Mc2-3', 103:'I/2', 104:'Ca/Ta'}

Insert value

Grupo anatomico: {0:'Mp', 1:'Ma', 2:'E', 3:'C', 4:'A', 5: nan, 6:'In', 7:'ES', 8:'Cr', 9:'PT', 10:'PTPL', 11:'PTPI'}

Insert value

GRCA

Insert value

Dientes

Insert value

Fragmento de hueso

Insert value

Longitud hueso largo

Insert value

<https://archaeologytutor.herokuapp.com>

Conjunts de dades

- Podeu buscar conjunts de dades a <https://datasetsearch.research.google.com>

Perfils data science



<https://todobi.com/tipos-de-roles-en-analytics-business>

- Harvard Business review
- Sexiest job of the 21st century

Comencem!

- <https://ciutadania.cat>

Gràcies per la vostra col·laboració!

Laia.Subirats@eurecat.org