

Minería de Datos

IIC2433

Aspectos Administrativos
Vicente Domínguez

Equipo Docente

Profesor: Vicente Domínguez

Correo: vidominguez@ing.puc.cl

- Ingeniero Civil de Computación
- Magister en ciencias de la ingeniería
- Actualmente Data Scientist en You Order Me
- Investigación en el área de sistemas recomendadores



Vicente Dominguez

SEGUIR

Pontificia Universidad Católica de Chile

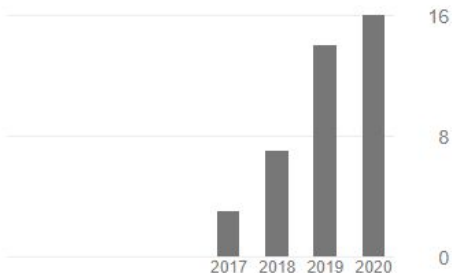
Dirección de correo verificada de uc.cl - [Página principal](#)

Machine Learning Recommender Systems

<input type="checkbox"/> TÍTULO	CITADO POR	AÑO
<input type="checkbox"/> Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features P Messina, V Dominguez, D Parra, C Trattner, A Soto User Modeling and User-Adapted Interaction 29 (2), 251-290	14	2019
<input type="checkbox"/> Comparing neural and attractiveness-based visual features for artwork recommendation V Dominguez, P Messina, D Parra, D Mery, C Trattner, A Soto Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, 55-59	7	2017
<input type="checkbox"/> pyreclab: A software library for quick prototyping of recommender systems G Sepulveda, V Dominguez, D Parra arXiv preprint arXiv:1706.06291	6	2017
<input type="checkbox"/> The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images V Dominguez, P Messina, I Donoso-Guzmán, D Parra Proceedings of the 24th International Conference on Intelligent User ...	5	2019
<input type="checkbox"/> Exploring Content-based Artwork Recommendation with Metadata and Visual Features P Messina, V Dominguez, D Parra, C Trattner, A Soto arXiv preprint arXiv:1706.05786	5	2017
<input type="checkbox"/> Do Better ImageNet Models Transfer Better... for Image Recommendation? F del Rio, P Messina, V Dominguez, D Parra arXiv preprint arXiv:1807.09870	2	2018
<input type="checkbox"/> Towards Explanations for Visual Recommender Systems of Artistic Images. V Dominguez, P Messina, C Trattner, D Parra IntRS@ RecSys, 69-73	1	2018
<input type="checkbox"/> Algorithmic and HCI aspects for explaining recommendations of artistic images V Dominguez, IN Donoso-Guzmán, P Messina, D Parra ACM Transactions on Interactive Intelligent Systems (TIIS)		2020

Citado por

	Total	Desde 2015
Citas	40	40
Índice h	5	5
Índice i10	1	1



Coautores

EDITAR

	Denis Parra Pontificia Universidad Catolica d...	>
	Christoph Trattner Associate Professor of Informati...	>
	Alvaro Soto Professor Universidad Catolica d...	>
	Ivania Donoso Guzmán	>
	Felipe del Rio Phd. student, Pontificia Universid...	>

Equipo Docente

Ayudantes:

- Vicente Castro (vvcastro@uc.cl)
- Federico Taladriz (fntaladriz@uc.cl)
- Daniela Flores (diflores@uc.cl)
- Ricardo Schilling (reschilling@uc.cl)
- Astrid San Martín (aesanmar@uc.cl)
- Sebastián Ricke (sricke@uc.cl)
- Hernan Valdivieso (hfvaldivieso@uc.cl)
- Javier Ruiz (jiruiz2@uc.cl)

Metodología

- Las clases serán solo los martes, excepto por esta semana. Que habrá el jueves.
- **No hay les ni Examen.** Se harán controles, tareas y un proyecto.
- Habrán actividades pero estas tendrán un carácter formativo.
- Cuando haya una clase doble los martes, habrá un descanso entre medio de alrededor de unos 15 minutos.

Calendario

Week	Fecha semana	Martes	Jueves	Clase Martes - 1	Clase Martes - 2	Ayudantía Jueves / Control
I	10-ago.	11-ago.	13-ago.	Intro Administrativo		Clase - Data Warehouse - OLAP
II	17-ago.	18-ago.	20-ago.	Web Scrapping	Actividad	Control Data WH
III	24-ago.	25-ago.	27-ago.	Data Prep	Pandas	Ayudantía Pandas y librerías
IV	31-ago.	1-sept.	3-sept.	Association Rules	Association Rules	Ayudantía Association Rules
V	7-sept.	8-sept.	10-sept.	PCA	Actividad	Control AR
VI	14-sept.	15-sept.	17-sept.	Regresiones	Actividad	Feriado
VII	21-sept.	22-sept.	24-sept.	Semana Receso	Semana Receso	Semana Receso
VIII	28-sept.	29-sept.	1-oct.	Reg log	Actividad	Control PCA y Reg
IX	5-oct.	6-oct.	8-oct.	KNN	Árboles de Decision	Ayudantía Knn y Árbol de Decisión
X	12-oct.	13-oct.	15-oct.	Random Forest	Random Forest	Control Reg log - Knn
XI	19-oct.	20-oct.	22-oct.	Naive Bayes	Actividad	Ayudantía RF y Naive
XII	26-oct.	27-oct.	29-oct.	Evaluacion Modelos	Actividad	Control Árboles - Naive
XIII	2-nov.	3-nov.	5-nov.	K - means - Meanshift	Clustering aggl	Ayudantía Clustering
XIV	9-nov.	10-nov.	12-nov.	Mixture of Gaussians	Mixture of Gaussians	Ayudantía Mixture of Gaussieans
XV	16-nov.	17-nov.	19-nov.	Eval clustering	Actividad	
XVI	23-nov.	24-nov.	26-nov.	Aplicaciones	Aplicaciones	Control Clustering
XVII	30-nov.	1-dic.	3-dic.	Avanzados	Avanzados	
XVIII	7-dic.					
XIX	14-dic.					
XX	21-dic.					

Herramientas del curso

- Python 3.8
- Jupyter Notebook
- Github Classroom
- Canvas
- Zoom
- Discord



GitHub
Classroom



Página del curso

- En un principio se subirá el material a Canvas y a Github, pero se espera finalmente que esté todo en Github.
- <https://github.com/IIC2433/Syllabus-2020-2>

Encuesta diagnóstico

- Antes de seguir por favor respondan esta encuesta.
- <https://forms.gle/N6abhXQyL8MsDziV8>

Faltas a la ética

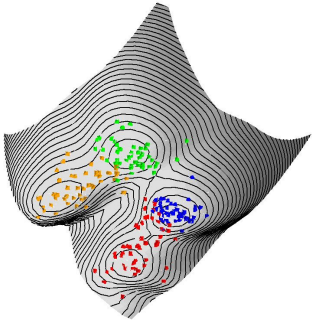
- Cualquier situación de **copia o falta a la ética** detectada en alguna evaluación tendrá como sanción un **1.1 final en el curso**.
- Se enviarán los antecedentes a la comisión de ética de la universidad y se aplicarán las sanciones dependiendo de la unidad académica.
- Si una evaluación es individual, que se desarrolle de forma individual.
- Compartir código es considerado como una falta a la ética también.

Minería de Datos

IIC2433

Introducción

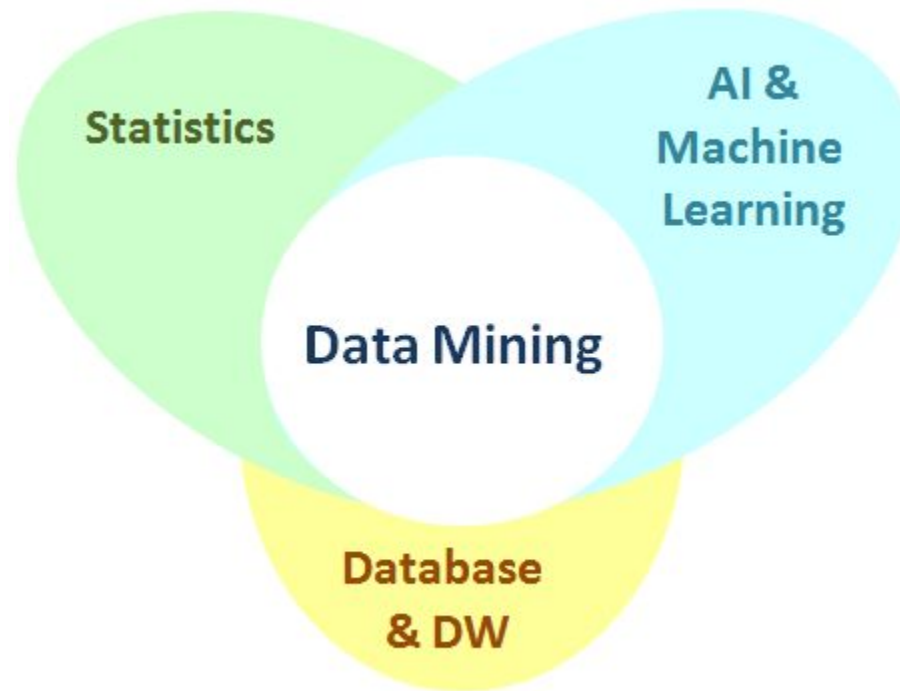
Vicente Domínguez



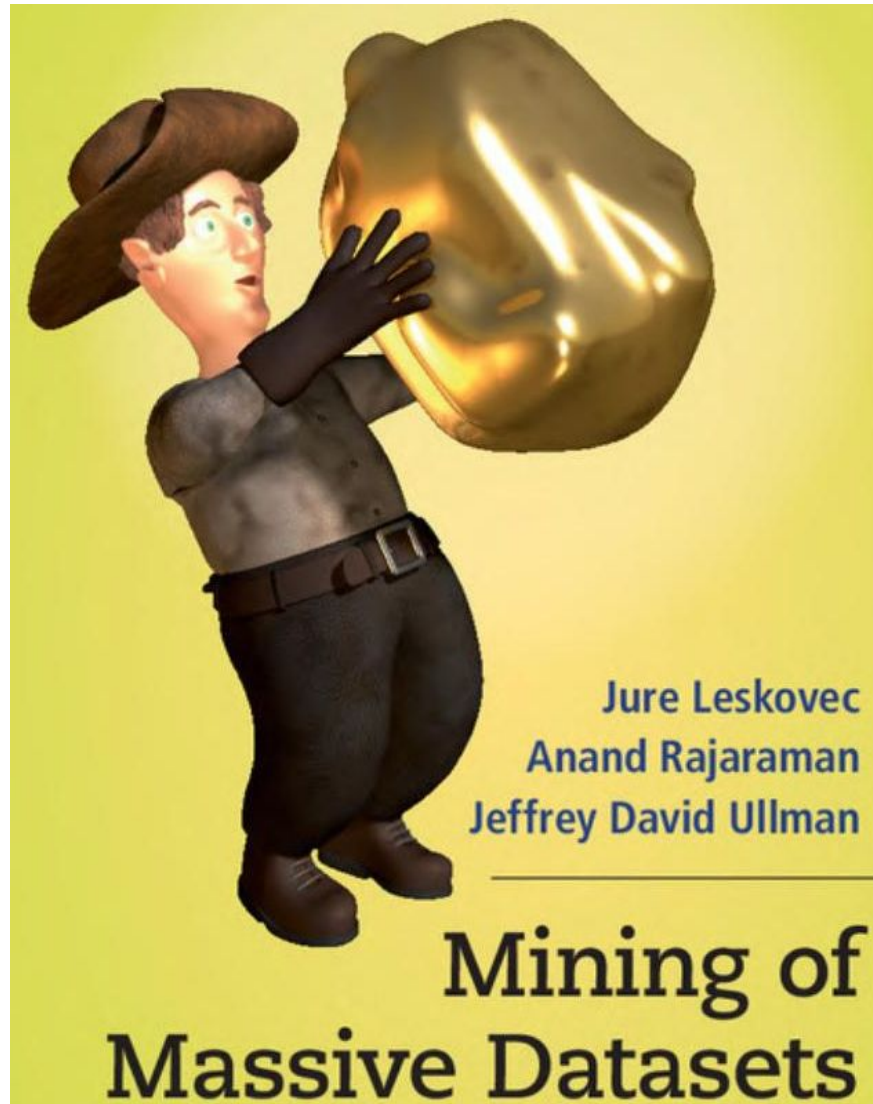
Minería de Datos

Área de Ciencia de la Computación y de la Estadística donde se estudian las distintas técnicas de cómo extraer conocimiento relevante desde grandes cantidades de información

Minería de Datos

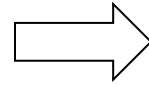


Minería ~ Mining



Minería ~ Mining

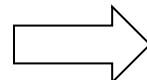
"Minar" oro desde rocas



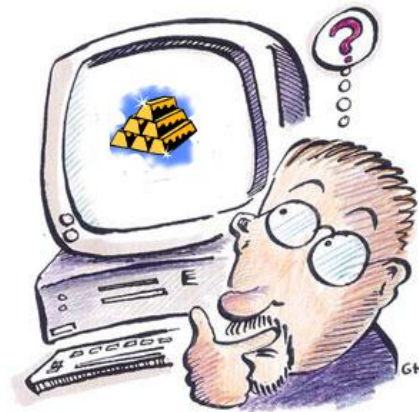
Minería de Oro



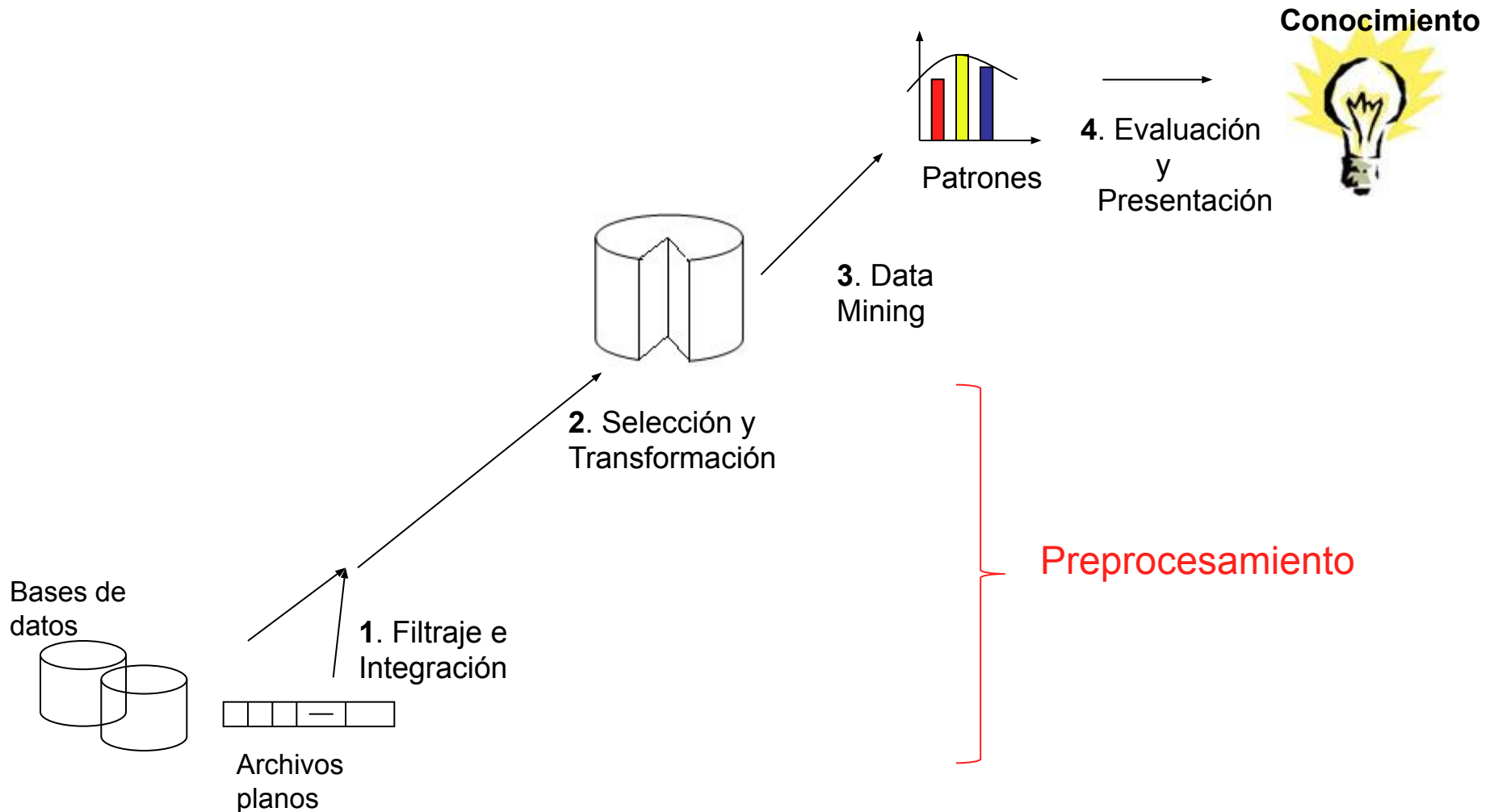
"Minar" datos desde
fuentes de información



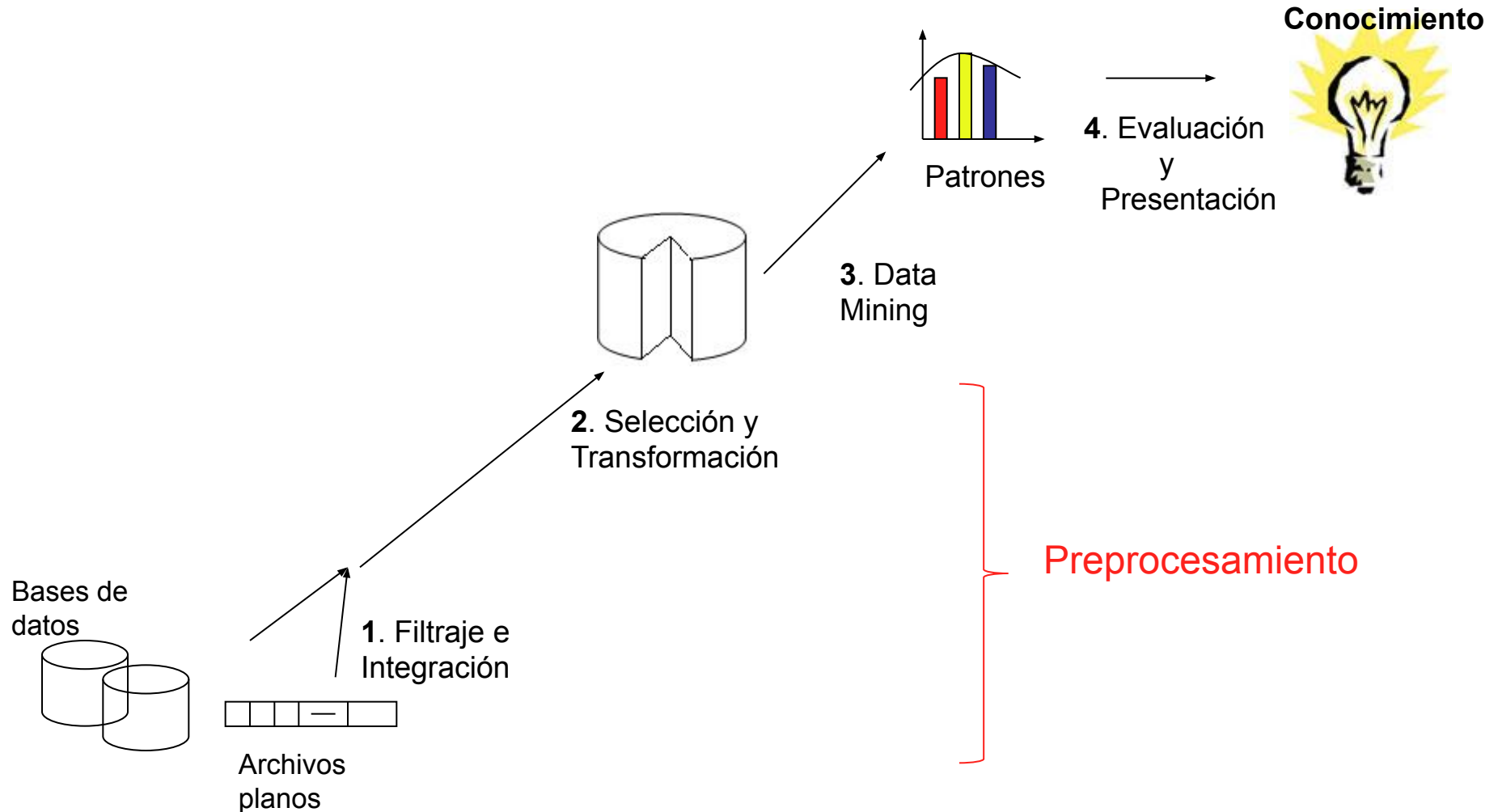
Minería de Datos



Desde otra perspectiva



Knowledge Discovery in Databases



Importancia de la Minería de Datos

- Bases de datos actuales no se pueden analizar manualmente.
- Necesidad de herramientas automáticas que realicen ciertas tareas.
- Surge la necesidad de científicos que puedan idear soluciones y algoritmos que cubran la necesidad anterior.

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Minor Data Science

- La idea del Minor es tratar de cubrir las habilidades necesarias para formar un Data Scientist.
- Este curso cumple un rol fundamental en esta formación.
- Por todo esto, tiene un carácter teórico práctico.

Aplicaciones de Minería de datos

- Hay una gran diversidad de aplicaciones.
- Hablaremos de algunas de las más importantes o relevantes dentro del mar de aplicaciones.

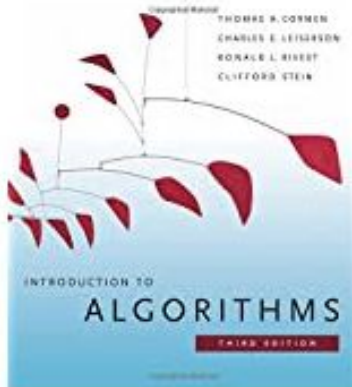
Sistemas Recomendadores



[Your Amazon.com](#) [Today's Deals](#) [See All Departments](#)

Vicente Domínguez,

Based on your recent activity, we thought you might be interested in this.



[Introduction to Algorithms, 3rd Edition](#)
by Thomas H. Cormen

List Price: ~~\$94.00~~

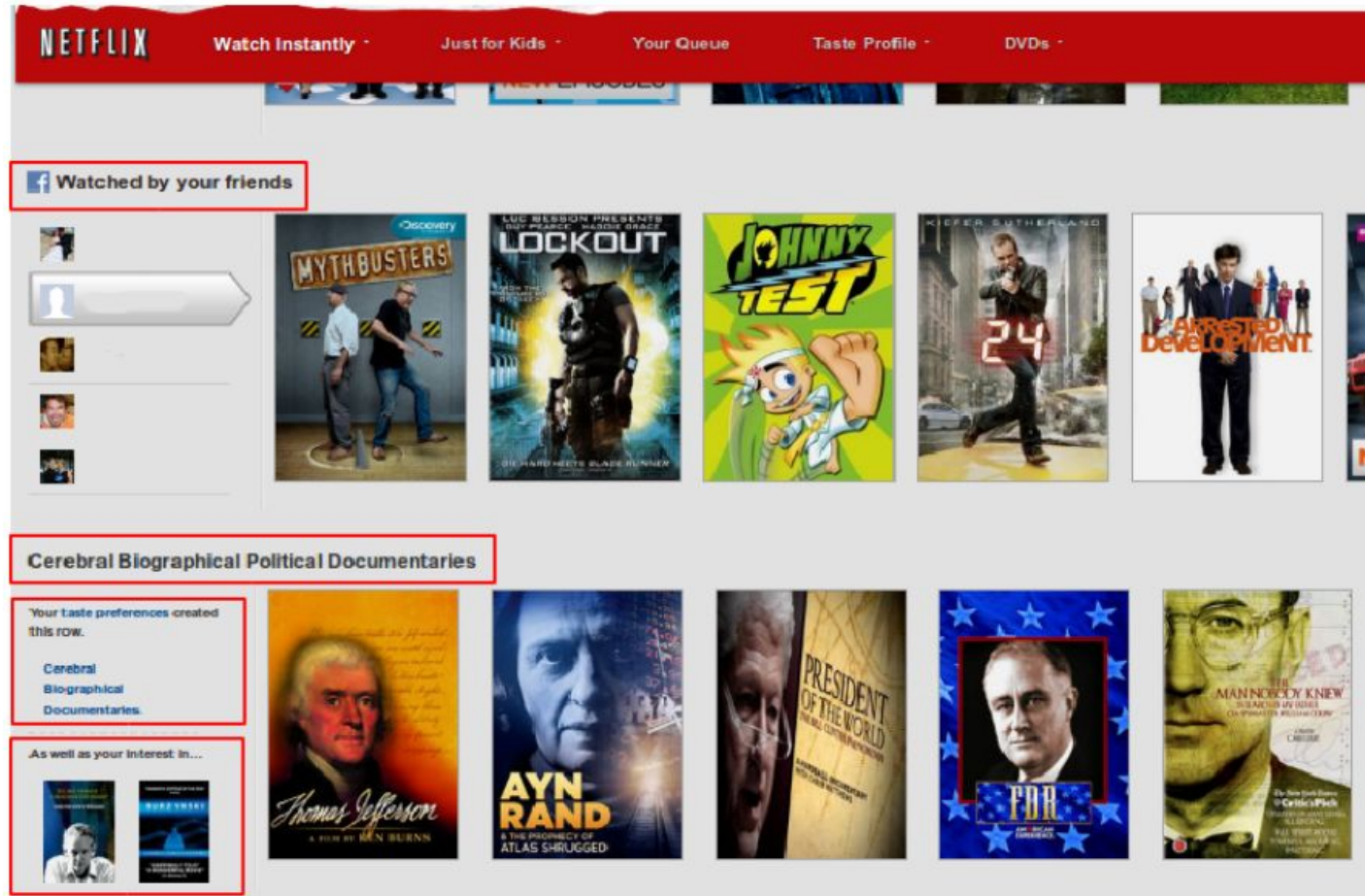
Price: **\$65.33**

You Save: **\$28.67 (31%)**

Some books on algorithms are rigorous but incomplete; others cover masses of material but lack rigor. Introduction to Algorithms uniquely combines rigor and comprehensiveness ... [Read More](#)

[Learn more](#)

Sistemas Recomendadores - II



Amatriain, X. (2013). Mining large streams of user data for personalized recommendations. ACM SIGKDD Explorations Newsletter, 14(2), 37-48.

Sistemas Recomendadores - II



NETFLIX

Netflix Prize

COMPLETED

Home Rules Leaderboard Update

Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Display top leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43

Congratulations!

The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.

On September 21, 2009 we awarded the \$1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about [their algorithm](#), checkout team scores on the [Leaderboard](#), and join the discussions on the [Forum](#).

We applaud all the contributors to this quest, which improves our ability to connect people to the movies they love.



Búsqueda + Personalización

goplacit Chile Venta Arriendo Proyectos Cómo funciona Ayuda Publicar 300

Favoritos

Provi-Nunoa

Los filtros de esta zona son:

Precio Venta: UF 4.000,00 - UF 6.500,00

Habitaciones: desde 3

Área: desde 80

Tipo: Departamento

Modalidad: Venta

Antigüedad: Usada

Calidad información: Buen contenido

Notificaciones: Cada 24 horas

Orden: Más reciente primero UF

UF 4.276,44
Departamento
3 habitaciones / 2 baños

UF 5.294,64
Departamento
3 habitaciones / 2 baños

UF 5.378,96
Departamento
3 habitaciones / 2 baños

UF 6.312,84
Departamento
3 habitaciones / 2 baños

UF 5.999,78

1 - 50 de 131

Servicios cercanos

Map Satellite

Estás viendo en esta zona

1 - 50 de 131 propiedades en venta

Map data ©2015 Google, Mapcity Terms of Use

Búsqueda + Personalización II

[GoPlaceit] nuevas propiedades te han encontrado en las últimas horas

domingo, marzo 8, 2015, 7:04 pm ● ★

De: "Goplaceit" <hello@goplaceit.com>

goplaceit.com



A: "Denis Parra" <dalpasa@yahoo.com>

Deja que las propiedades te encuentren

**Hay novedades en tus zonas,
estas propiedades te pueden interesar**

cabecados completos Vista Imprimible



Mi primera zona

2 nuevas propiedades

Mi primera zona

2

nuevas propiedades

Provi-Nunoa

1

nueva propiedad



Oficina en Arriendo en Providencia

Marchant Pereira/ Fidel Oteiza

\$ 340.000 (CLP)

1 Habitación / 1 Baño / 40 m²

Excelente Oficina Semi amoblada a pasos del Metro Pedro de Valdivia. Cuenta con un privado, un baño y kitchenette, piso flotante. Sin estacionamiento (estaciona...

[Ver propiedad](#)



Provi-Nunoa

1 nueva propiedad



Departamento en Venta en Ñuñoa

Hermoso depto en calle Jose Domingo

UF 4.276,60

3 Habitaciones / 2 Baños / 102 m²

Hermoso depto, en el corazón de Ñuñoa, ubicado en calle José Domingo Cañas, 3 amplias habitaciones, 2 baños completos ambos con luz natural, terraza en dormitorio...

[Ver propiedad](#)



Departamento en Arriendo en Providencia

José Manuel Infante 7, Providencia,

\$ 450.000 (CLP)

1 Habitación / 1 Baño / 35 m²

Arriendo departamento amoblado, por temporada, de preferencia a extranjeros que estén de paso en Chile, con espectacular vista a Fuente Bicentenario en Parque B...

[Ver propiedad](#)

Además, estos proyectos están en esta zona:



Don Hernando

Providencia

Desde 1 UF 3.680,00



Edificio Ferrer 24

Providencia

Desde 1 UF 5.287,00



Edificio LF 35

Providencia

Desde 1 UF 6.198,00

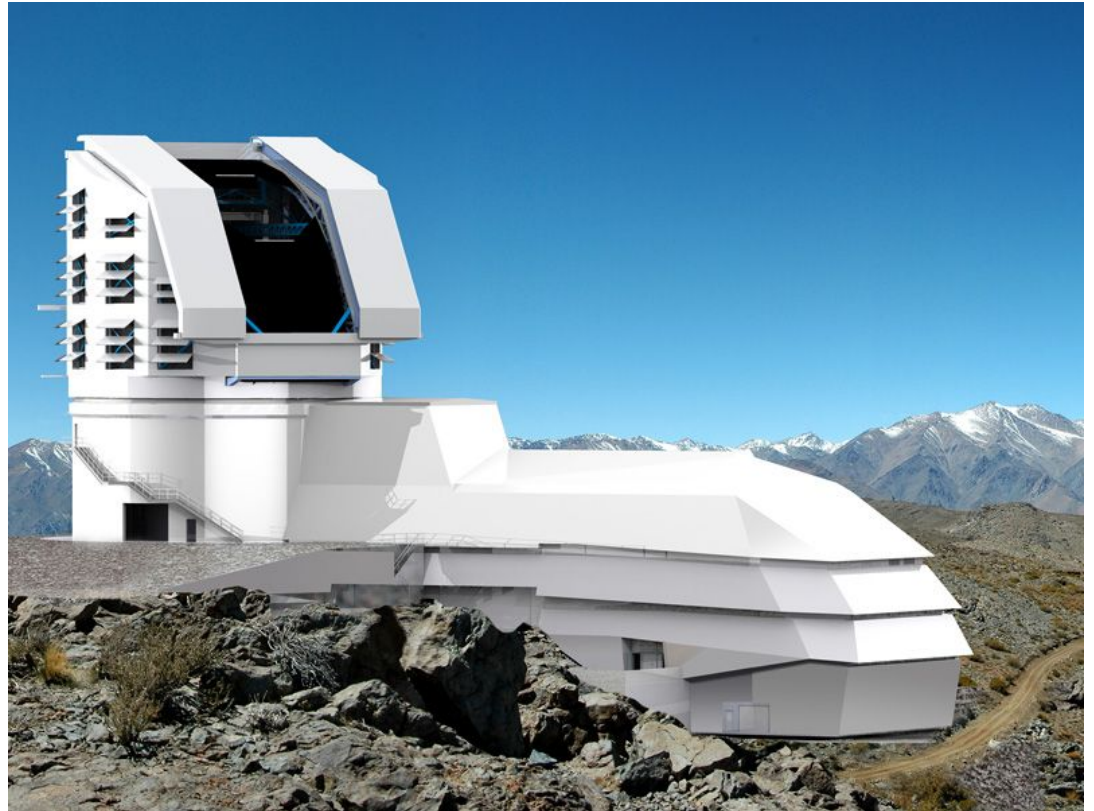


Reglas de Asociación:

$\{cebollas, vegetales\} \Rightarrow \{carne\}$

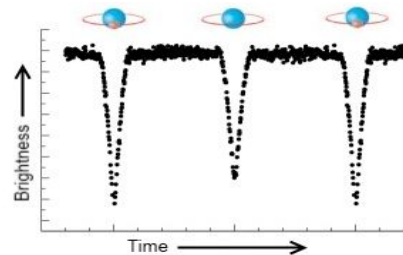
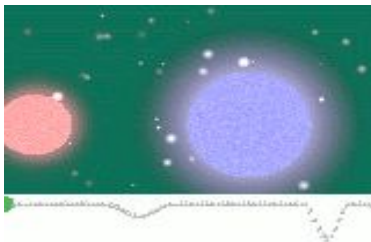
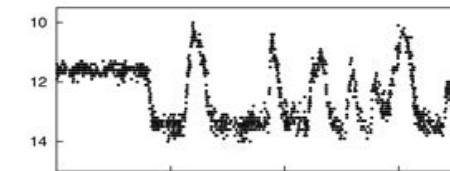
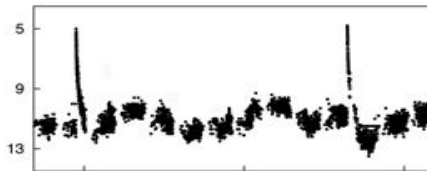
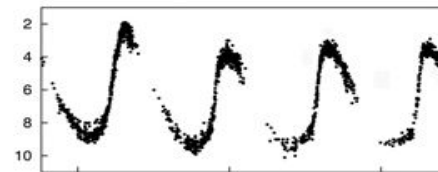
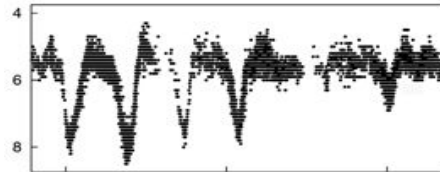
Minería de datos en Astronomía

Aplicación científica
en Chile

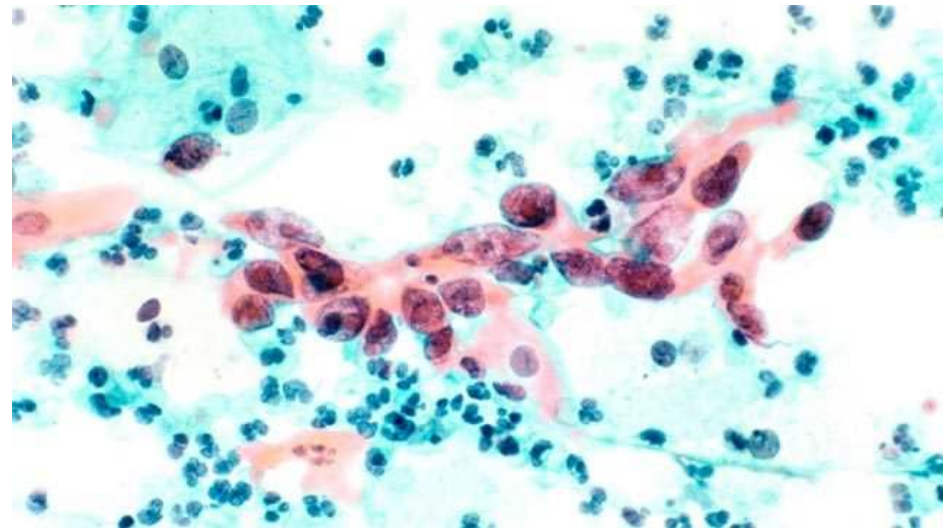
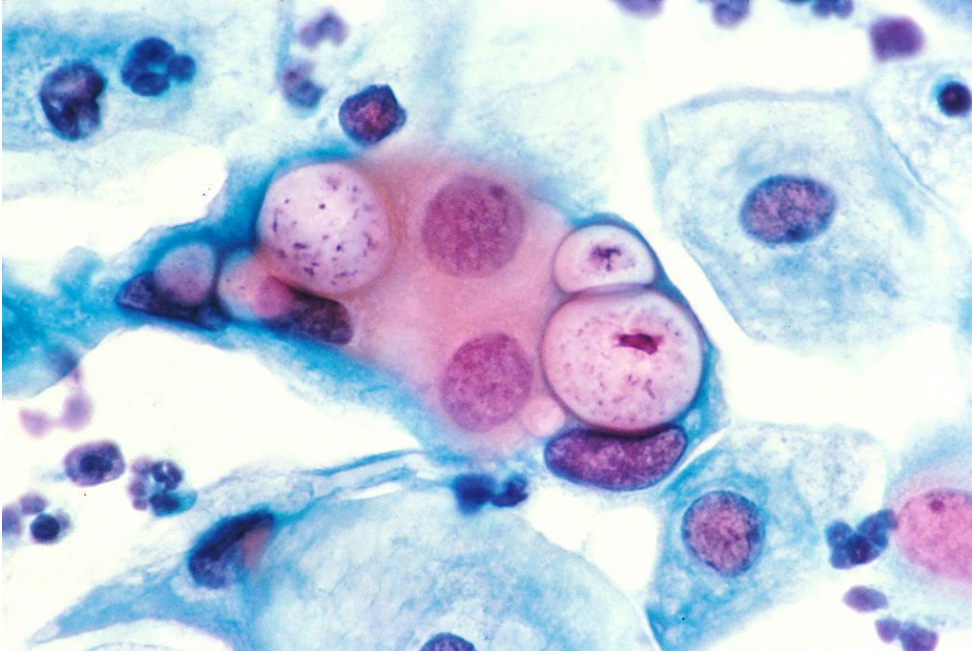


Minería de datos en Astronomía

Clasificación de Estrellas Variables



Biotechnología



- Análisis de documentos:

Detección de SPAM

Respuestas en texto libre (chats de atención al cliente, etc)



Análisis de Redes Sociales

