

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 @ You Order Me
- Próximo mes: Senior Machine Learning Engineer @ Bain & Company
- Investigación en el área de sistemas recomendadores



Vicente Dominguez

SEGUIR

Pontificia Universidad Católica de Chile

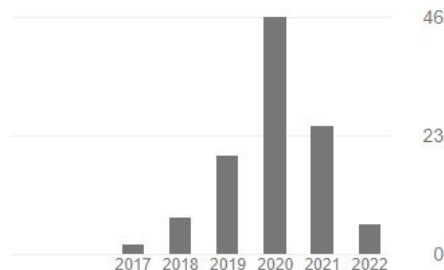
Dirección de correo verificada de uc.cl - [Página principal](#)

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	42	2019
<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 ...	26	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	16	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	10	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	5	2018
<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

Citado por

	Total	Desde 2017
Citas	106	106
Índice h	5	5
Índice i10	4	4



Coautores

EDITAR



Denis Parra
Pontificia Universidad Católica d...



Christoph Trattner
Professor & Centre Director of M...



Alvaro Soto
Professor Universidad Católica d...



Ivania Donoso Guzmán



Equipo Docente

Ayudantes:

- Daniela Concha (daconcha@uc.cl)
- Francisca Ibarra (faibarra1@uc.cl)
- Lucas Valenzuela (lucas.valenzuela@uc.cl)
- Sebastián Veliz (saveliz@uc.cl)
- Sebastián Zabala (szabala@uc.cl)
- Sofía Olmedo (sofia.olmedo.s@uc.cl)
- Tomás Cortes (ticortes@uc.cl)
- Tomás González (tgonzalezv@uc.cl)

Metodología

- Las clases serán solo los martes.
- **No hay les ni Examen.** Se harán 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	Actividad	Tarea (Miércoles)
I	7-mar	8-mar	10-mar	Intro Administrativo, El problema de aprender	Manejo de datos	Ayudantía Pandas y librerías		
II	14-mar	15-mar	17-mar	Regresión Lineal	Regresión Lineal		Actividad	
III	21-mar	22-mar	24-mar	Regresión Logística	Regresión Logística	Ayudantía Tarea		Tarea
IV	28-mar	29-mar	31-mar	Clasificadores - Evaluación	KNN			
V	4-abr	5-abr	7-abr	Árbol de decisión	Árbol de decisión		Actividad	
VI	11-abr	12-abr	14-abr	Ensamble	Random Forest, Gradient Boosted Trees	FERIADO (JUEVES SANTO)		Tarea
VII	18-abr	19-abr	21-abr	Naive Bayes	Inferencia	Ayudantía Tarea		
VIII	25-abr	26-abr	28-abr	Aprendizaje no supervisado	K-means, DBSCAN, Mean shift		Actividad	
IX	2-may	3-may	5-may	Mixture of Gaussians	Mixture of Gaussians	Ayudantía Tarea	NO EVALUACIONES	Tarea
X	9-may	10-may	12-may	Evaluación de Clustering	Evaluación de Clustering			
XI	16-may	17-may	19-may	Reducción de Dimensionalidad	PCA		Actividad	
XII	23-may	24-may	26-may	Auto encoders	Auto encoders	Ayudantía Tarea		Tarea
XIII	30-may	31-may	2-jun	Procesamiento de Texto	Procesamiento de Texto			
XIV	6-jun	7-jun	9-jun	LSA	LSA		Actividad	
XV	13-jun	14-jun	16-jun	Word2Vec	Word2Vec	Ayudantía Tarea		Tarea
XVI	20-jun	21-jun	23-jun	FERIADO	PUEBLOS INDÍGENAS			
XVII	27-jun	28-jun	30-jun					
XVIII	4-jul	5-jul	7-jul					
XIX	11-jul	12-jul						
Fechas Importantes								
Inicio de clases	7 de marzo							
Último plazo para botar cursos	9 de mayo							
Fin de clases	24 de junio							
Subida de Notas	12 de julio							

Herramientas del curso

- Python 3.8 o superior
- Jupyter Notebook
- Github Classroom
- Canvas
- 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-2022-1>

Faltas a la ética

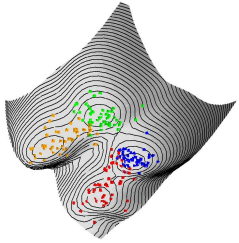
- Cualquier situación de **copia o falta a la ética** detectada en alguna evaluación tendrá como sanción un **1.1 final en el curso**.
- Se enviarán los antecedentes a la comisión de ética de la universidad y se aplicarán las sanciones dependiendo de la unidad académica.
- Si una evaluación es individual,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

Minería de Datos

IIC2433

Introducción

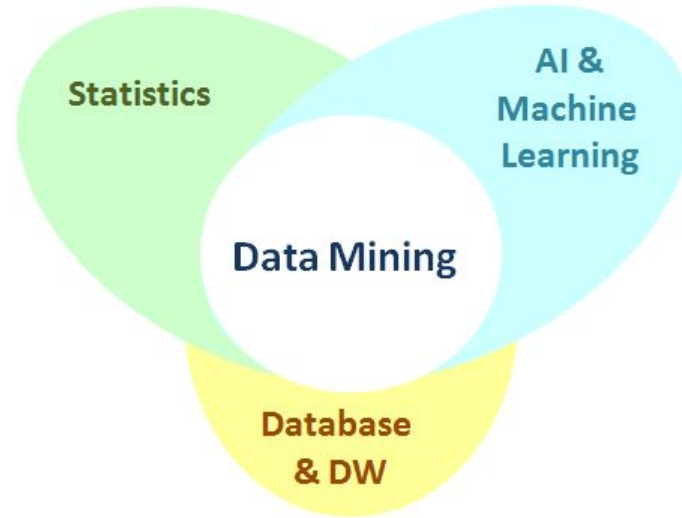
Vicente Domínguez



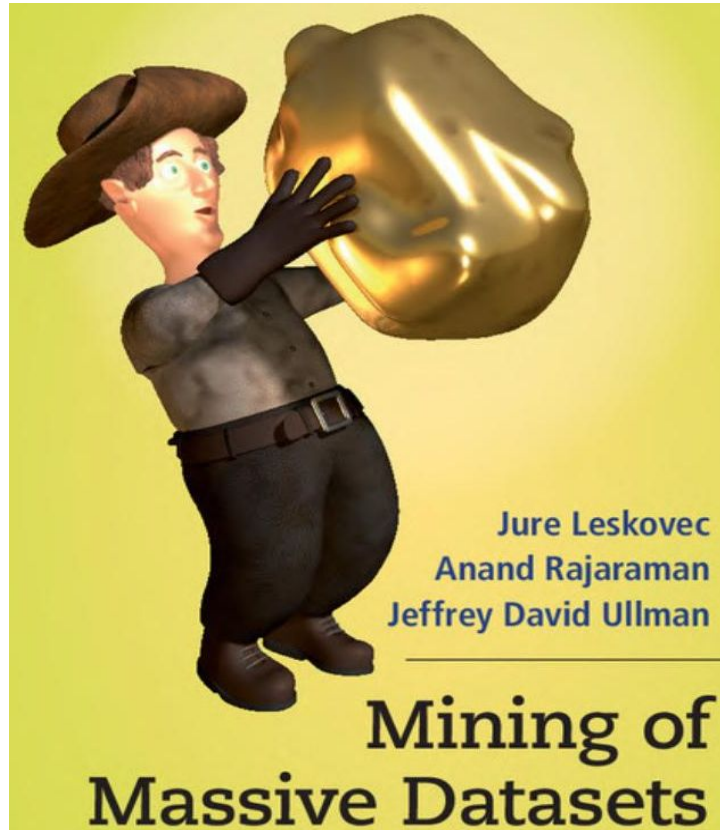
Minería de Datos

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

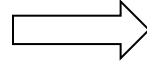


Minería ~ Mining



Minería ~ Mining

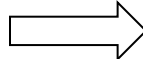
“Minar” oro desde rocas



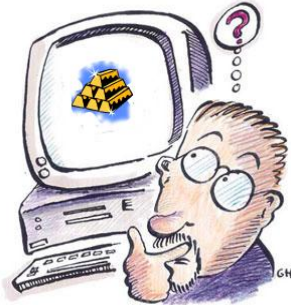
Minería de Oro



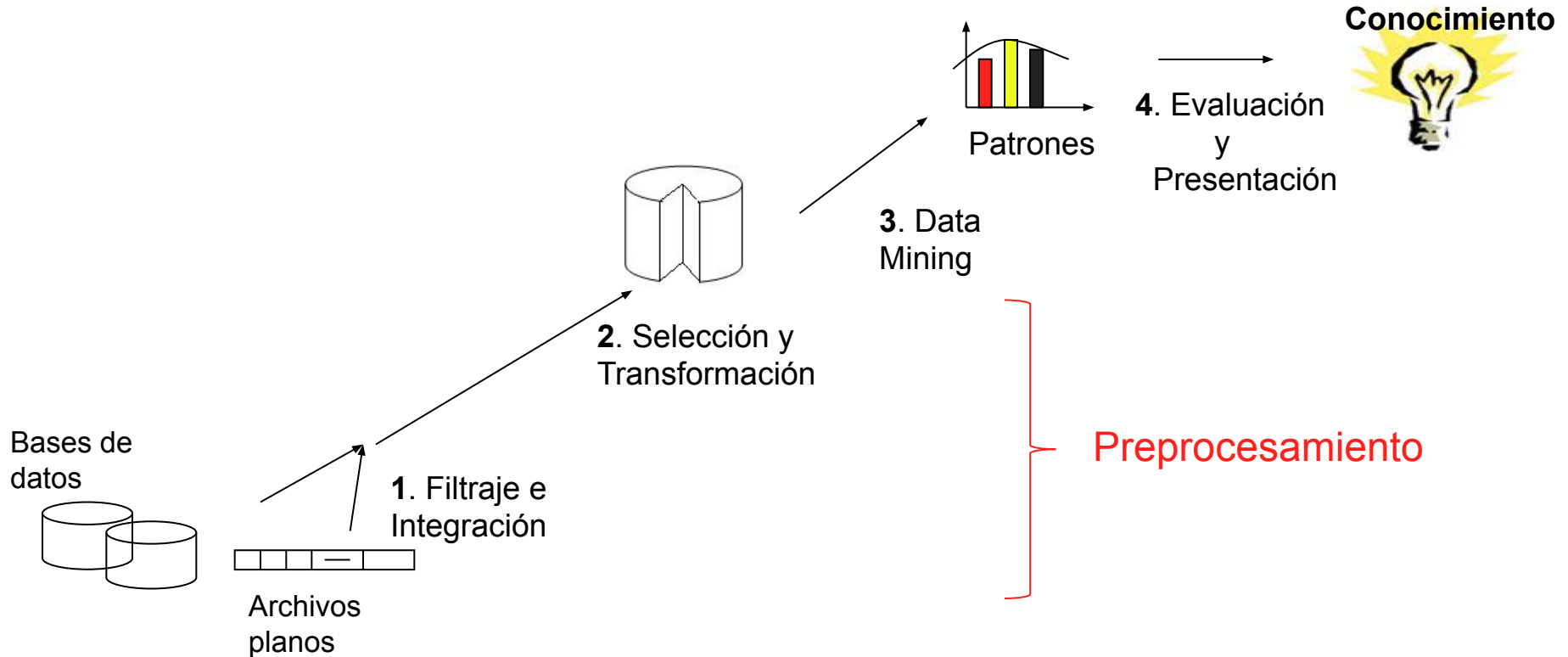
“Minar” datos desde
fuentes de información



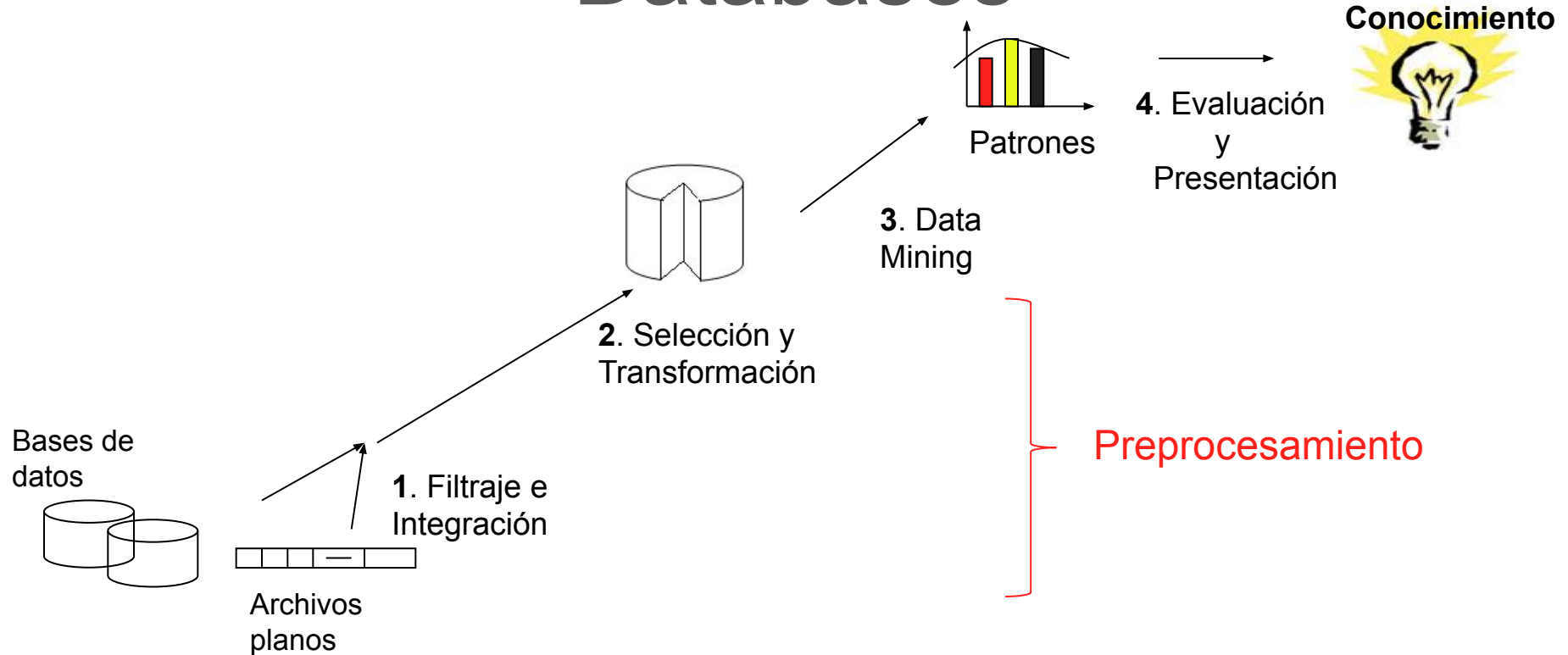
Minería de Datos



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.

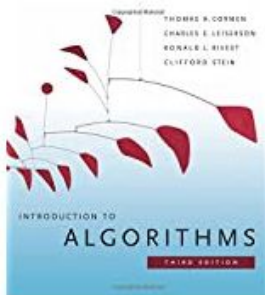
Sistemas Recomendadores



[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~~

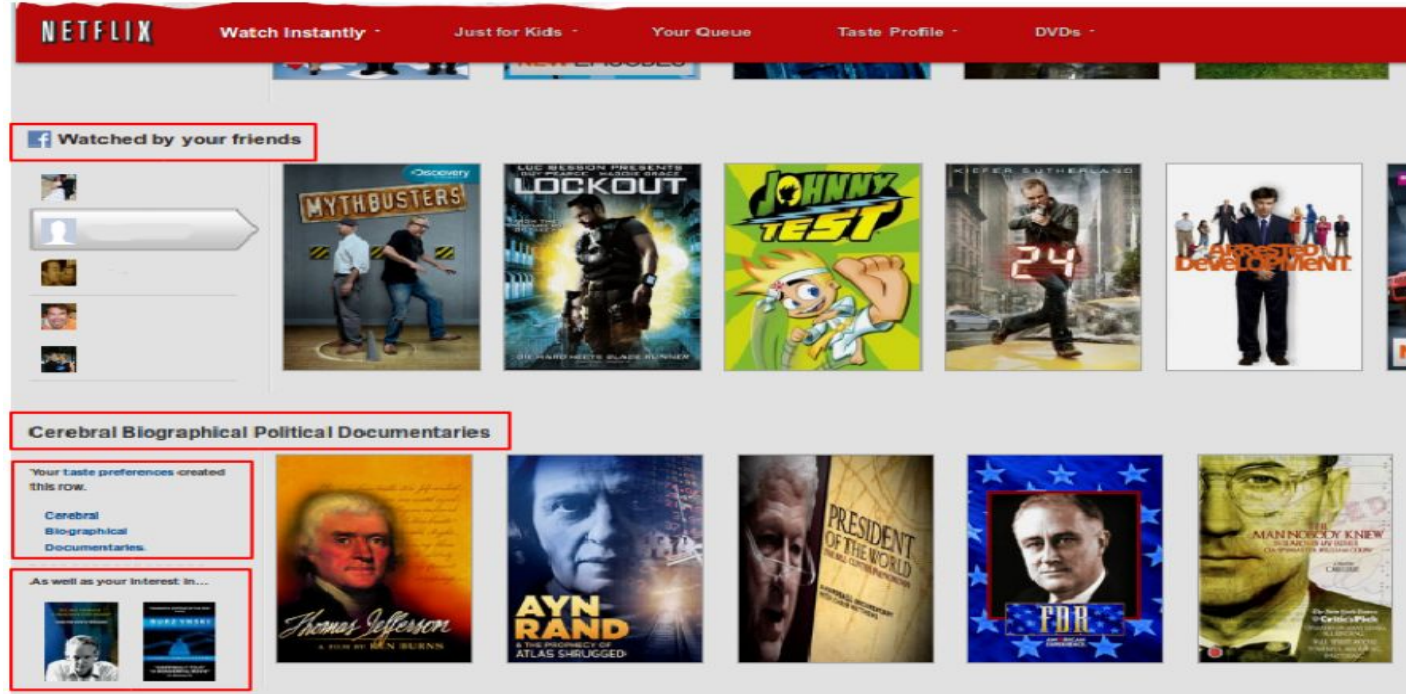
Price: **\$65.33**

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



Santiago

8 - 13 de mar

1 huésped



Hazte anfitrión



Precio

Tipo de alojamiento

Piscina

Wifi

Estacionamiento gratuito

Cocina

Lavadora

Aire acondicionado

Gimnasio

Acceso sin restricción de horario

Zona de trabajo

Secadora

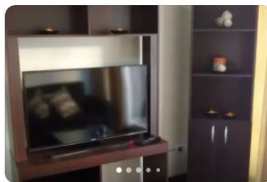
Plancha

Filtros

Más de 300 alojamientos en Santiago



Parece que vas a viajar muy pronto. Utiliza el filtro de Reserva inmediata para encontrar alojamientos que puedas reservar al momento.



Vivienda rentada entero en Santiago
Espacio Urbano San Diego

3 huéspedes · 1 habitación · 1 cama · 1 baño
Wifi · Piscina · Cocina

★ 4,75 (8 reseñas)

\$32,507 CLP / noche

\$177,536 CLP en total, impuestos no incluidos



SUPERANFITRIÓN

Vivienda rentada entero en San Miguel
Depto. Estudio Nuevo sin GAS, C/parking, Stgo

2 huéspedes · 1 habitación · 1 cama · 1 baño
Estacionamiento gratuito · Wifi · Piscina · Cocina

★ 4,90 (51 reseñas)

\$35,712 CLP / noche

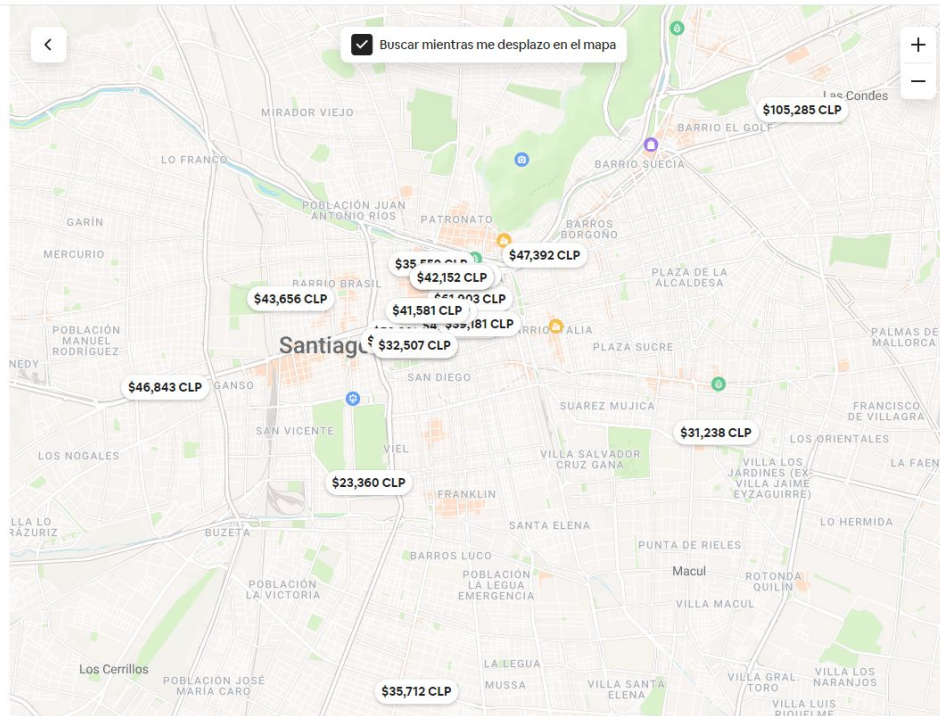
\$198,560 CLP en total, impuestos no incluidos



SUPERANFITRIÓN

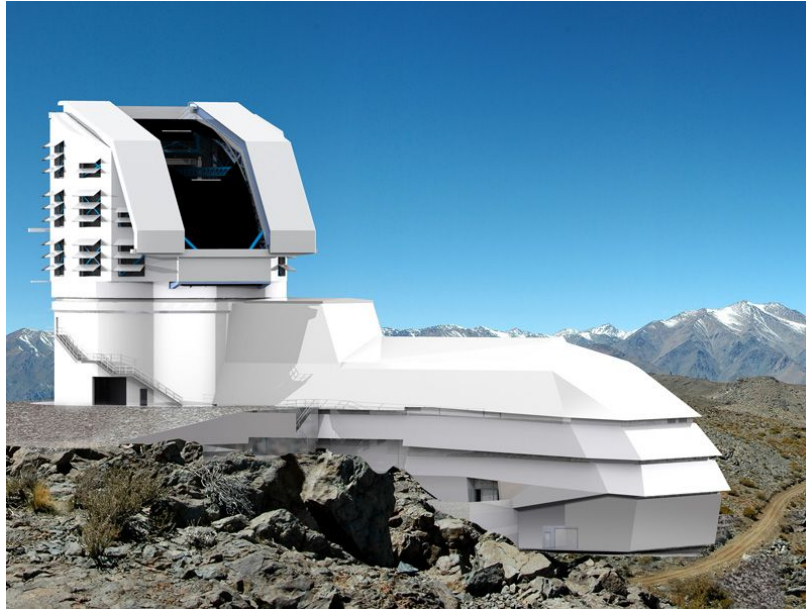
Conjunto residencial entero en Santiago
Hermoso, remodelado, todo nuevo, Metro u de Chile

2 huéspedes · 1 habitación · 1 cama · 1 baño
Wifi · Cocina



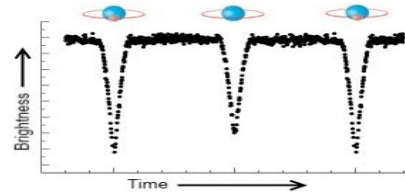
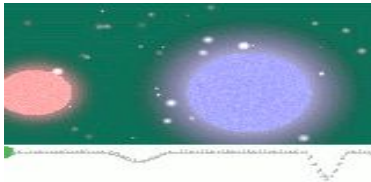
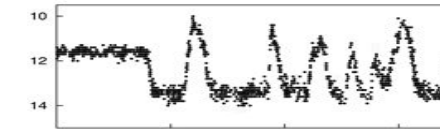
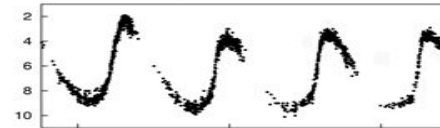
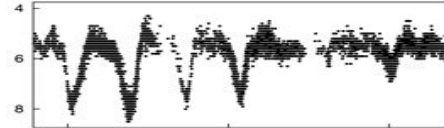
Minería de datos en Astronomía

Aplicación científica
en Chile

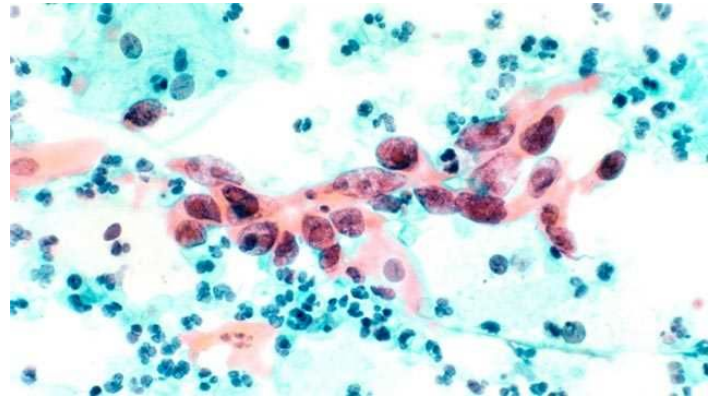


Minería de datos en Astronomía

Clasificación de Estrellas Variables



Biotecnología



- Análisis de documentos:
 - Detección de SPAM
 - Respuestas en texto libre (chats de atención al cliente, etc)



Análisis de Redes Sociales

