

Data Mining (Penambangan data)

Mulaab, S.Si. M.Kom

085733215507

Persiapan

- Join ke schoology
- <https://>

Sign up for Schoology [Back](#)

35K6-B2FC-Q3Z9Q

nama Lengkap **NPM**

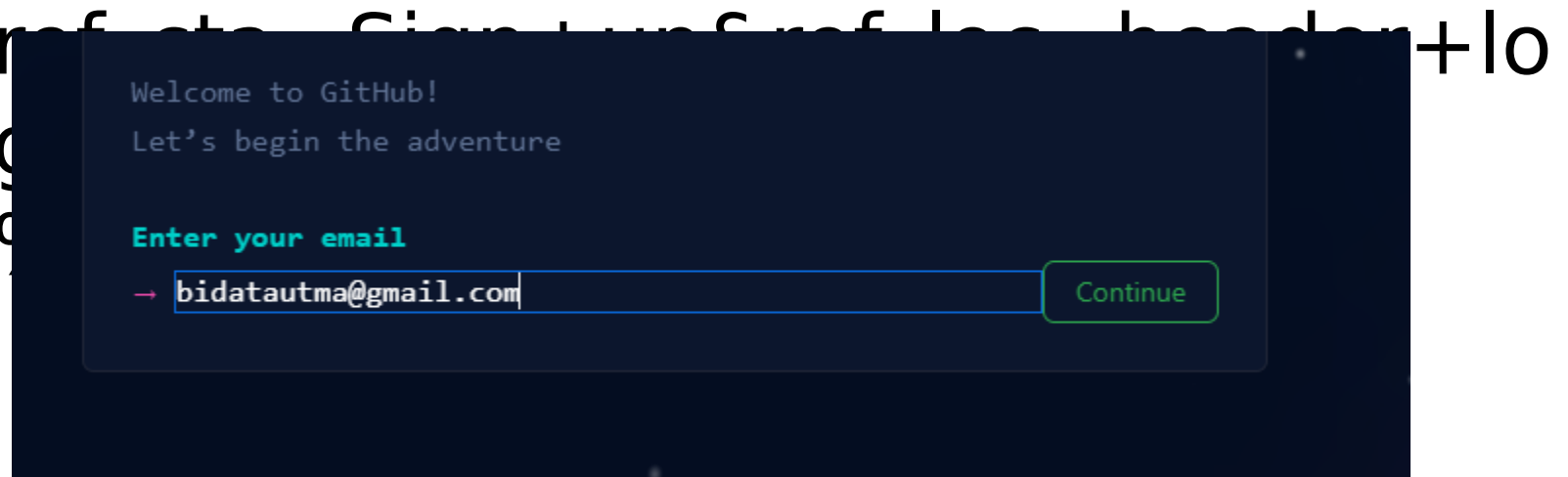
Birthdate: Month Day Year

Please read our [Privacy Policy](#) and [Terms of Use](#)

Sign up for Schoology

persiapan

- Buat akun email gmail (colab di google)
- Buat akun github
- [https://github.com/signup?](https://github.com/signup?ref=cta_Signup_Sref_Loc_header+lo)



Welcome to GitHub!
Let's begin the adventure

Enter your email

✓ bidatautma@gmail.com

Create a password

✓

Enter a username

✓ bidatautma

Would you like to receive product updates and announcements via email?

Type "y" for yes or "n" for no

→ n

Continue

- Tulis code yang telah dikirim ke email and

You're almost done!

We sent a launch code to `bidatautma@gmail.com`

→ Enter code

Eight input boxes for the code, with a vertical cursor in the sixth box.

Didn't get your email? [Resend the code](#) or [update your email address](#).

How many team members will be working with you?

This will help us guide you to the tools that are best suited for your projects.

Just me

2 - 5

5 - 10

10 - 20

20 - 50

50+

Are you a student or teacher?

Student

Teacher

Continue

Continue for free



Search or jump to...



[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)

Create your first project

Ready to start building? Create a repository for a new idea or bring over an existing repository to keep contributing to it.

[Create repository](#)

[Import repository](#)

Recent activity

When you take actions across GitHub, we'll provide links to that activity here.

The home for all developers — including you.

Welcome to your personal dashboard, where you can find an introduction to how GitHub works, tools to help you build software, and help merging your first lines of code.



Start writing code



Start a new repository

Collaborate on code with others and track your



Create your profile README


Create a file in a



Contribute to an existing repository

[Find repos that need](#)

Owner *

 bidautma ▼

Repository name *

/ datamining ✓

Great repository names are short and memorable. Need inspiration? How about [refactored-meme?](#)

Description (optional)

☒  **Public**

Anyone on the internet can see this repository. You choose who can commit.

☐  **Private**

You choose who can see

Initialize this repository with

Skip this step if you're import


☒ **Add a README file**

This is where you can write a l

Choose a license

A license tells others what they can and can't do with your code. [Learn more.](#)

License: None ▼

 You are creating a public repository in your personal account.

Create repository

General

Access

Collaborators

Moderation options ▼

Code and automation

Branches

Tags

Actions ▼

Webhooks


Environments

Pages

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is live at <https://bidatautma.github.io/datamining/>

Last deployed by  github-pages 1 minute ago

[Visit site](#)



Build and deployment

Source

Deploy from a branch ▼

Branch

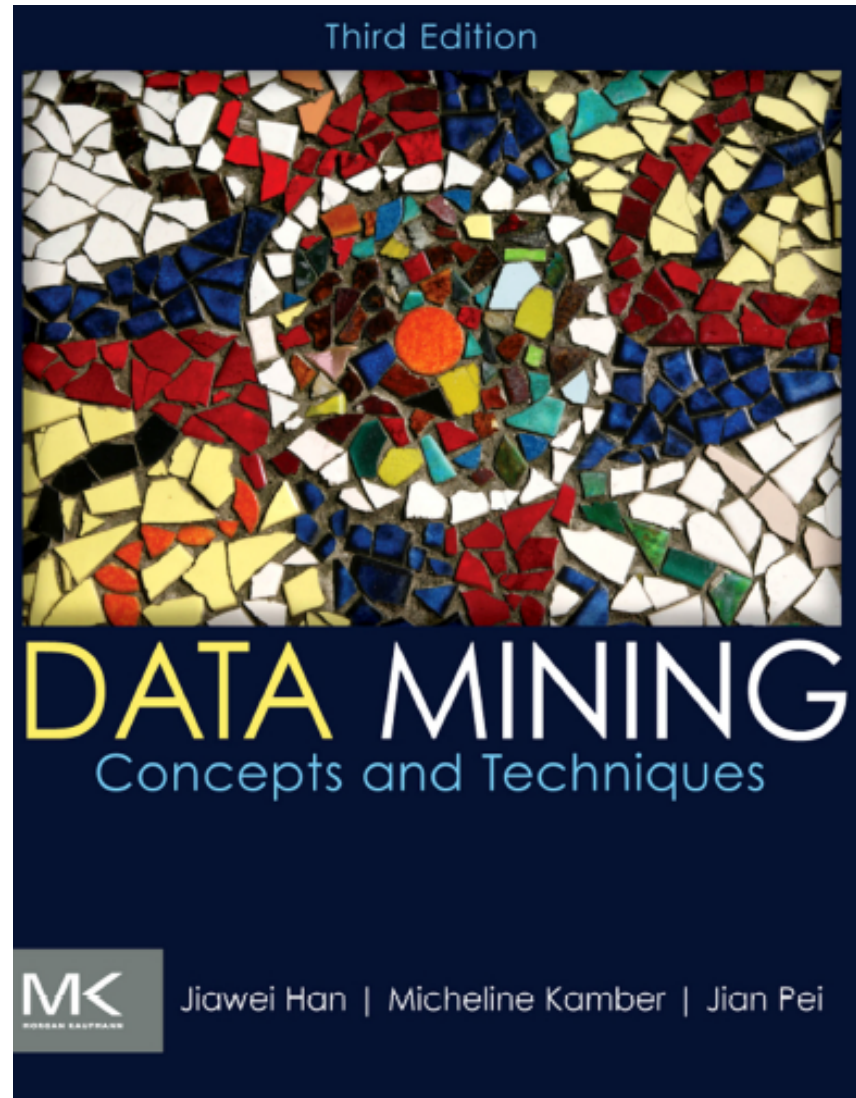
Your GitHub Pages site is currently being built from the `main` branch. [Learn more.](#)

 `main` ▼

 `/ (root)` ▼

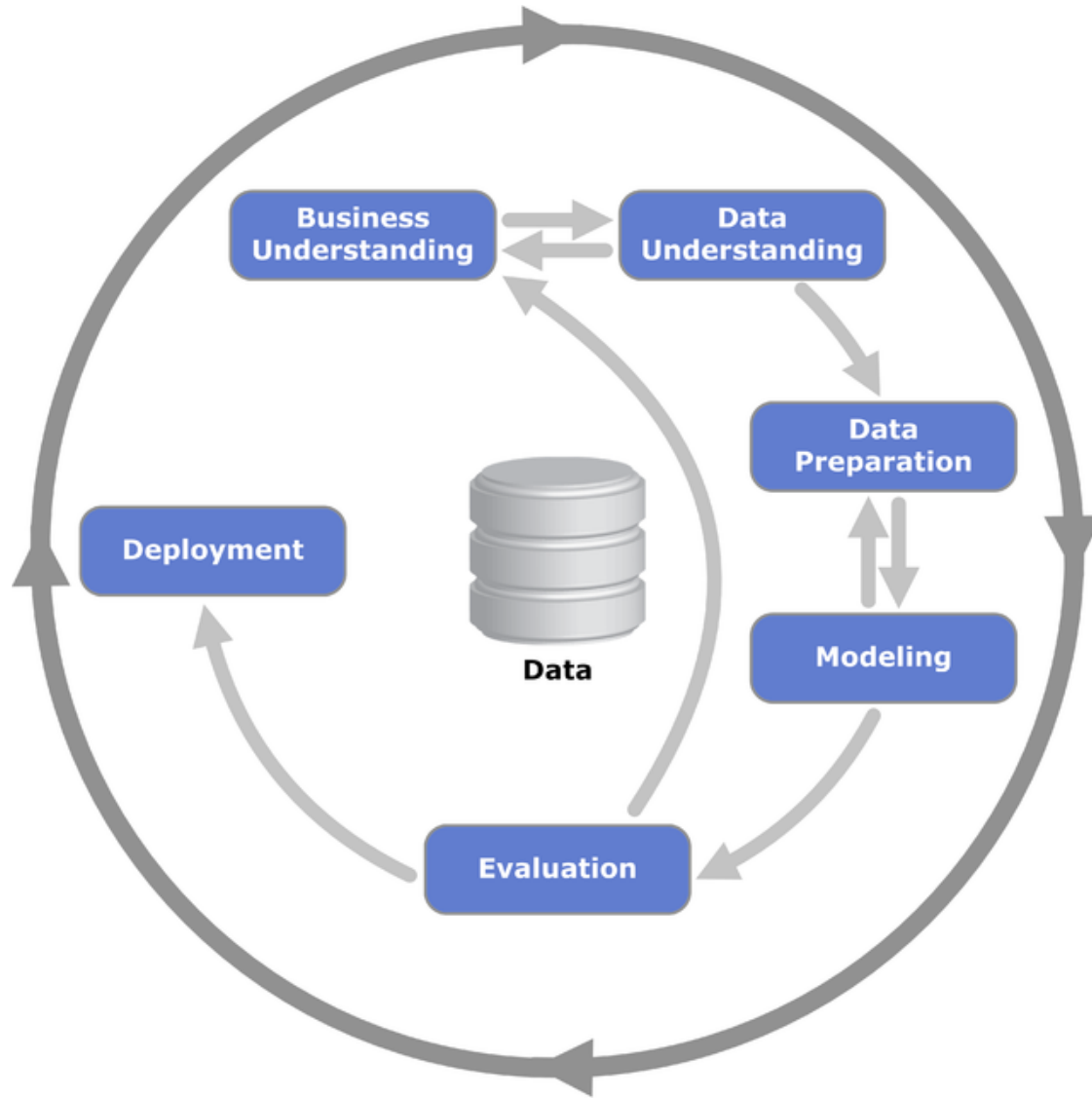
Save

Referensi

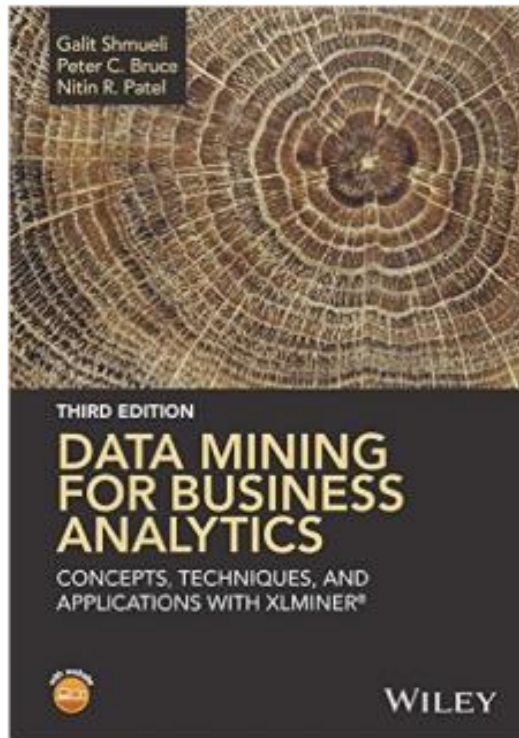


CRISP-DM

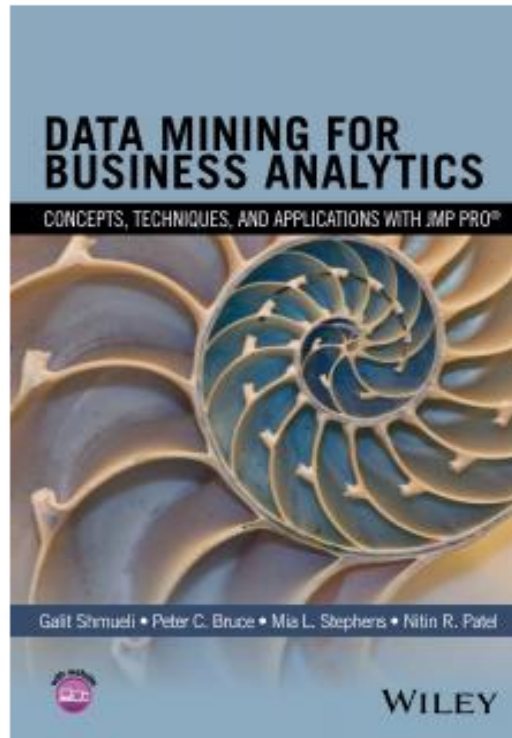
Standar Proses Datamining



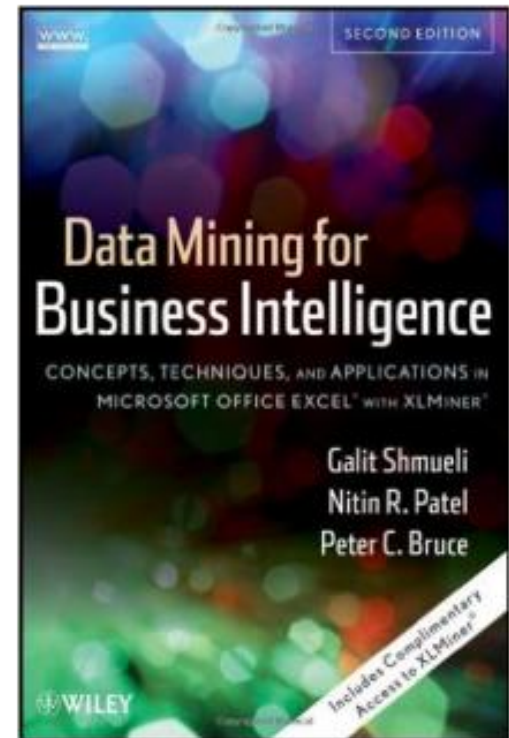
3rd Edition (2016)

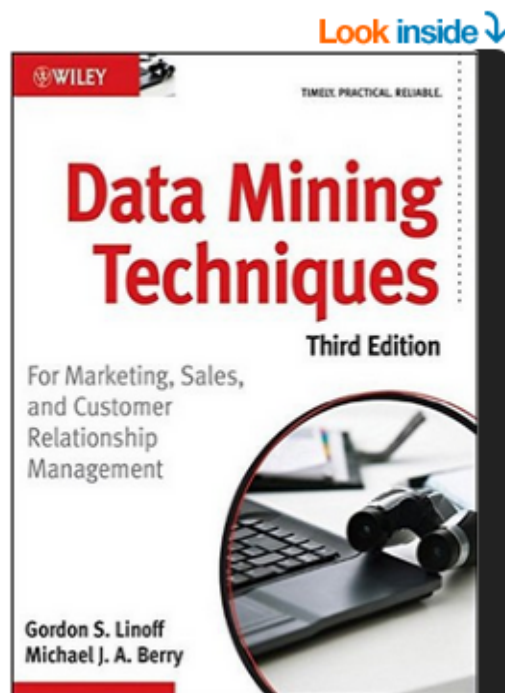


JMP PRO (2016)



2nd Edition (2010)





Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management Paperback – April 12, 2011

by [Gordon S. Linoff](#) (Author), [Michael J. A. Berry](#) (Author)



[27 customer reviews](#)

[See all 4 formats and editions](#)

Kindle
\$26.39

[Read with Our Free App](#)

Paperback
\$36.75

[36 Used](#) from [\\$14.19](#)

[47 New](#) from [\\$14.46](#)

The leading introductory book on data mining, fully updated and revised!

When Berry and Linoff wrote the first edition of *Data Mining Techniques* in the late 1990s, data mining was just starting to move out of the lab and into the office and has since grown to become an indispensable tool of modern business. This new edition—more than 50% new and revised—is a significant update from the previous one, and shows you how to harness the newest data mining

Memahami business (Business Understanding)

- Pemahaman tujuan proyek dan kebutuhan-kebutuhan yang diinginkan untuk bisnis
- Mendefinisikan kebutuhan-kebutuhan bisnis itu dalam data mining untuk mencapai tujuan bisnis

- Mengumpulkan data awal dan dilanjutkan dengan dengan kegiatan-kegiatan untuk mendapatkan data yang lazim serta
- Identifikasi data yang berkualitas,
- Pemahaman data sangat diperlukan untuk mendeteksi bagian yang menarik dari data sehingga dapat membangun hipotesa terhadap informasi yang tersembunyi

[Datasets Home](#) » **Datasets**

Latest News, Stories

[Evolution of IoT Edge Analytics: Strategies of Learning](#)

[Mining Meetings in Analytics, Big Data Mining,...](#)

[Human Vector: Incorporate Speaker Recognition into Machine Learning](#)

[eOne: Enterprise Information Security](#)

[Finance Analytics Europe, 5-6 October, London, UK](#)

[News & Stories](#) | [Top Stories](#)

Datasets for Data Mining and Data Science



45

Share

1.4K



Tweet

See also

- [Government, State, City, Local, public data sites and portals](#)
- [Data APIs, Hubs, Marketplaces, Platforms, and Search Engines](#)
- [Data Mining and Data Science Competitions](#)

Data repositories

More Recent Stories

[Wise Athena: Senior Data Scientist \(Madrid\)](#)

[7 Big Data Steps in Health Science](#)

[Data Science vs Crime: Detecting Pickpocket Suspects from Trajectory Data](#)

[HPE Haven OnDemand: Power Connectors for the Cloud and Edge](#)

[More answers, less theory from Big Data LDN, Nov 3-4](#)

[Top tweets, Aug 24-30: #DataScience sexiest job of the 21st century](#)

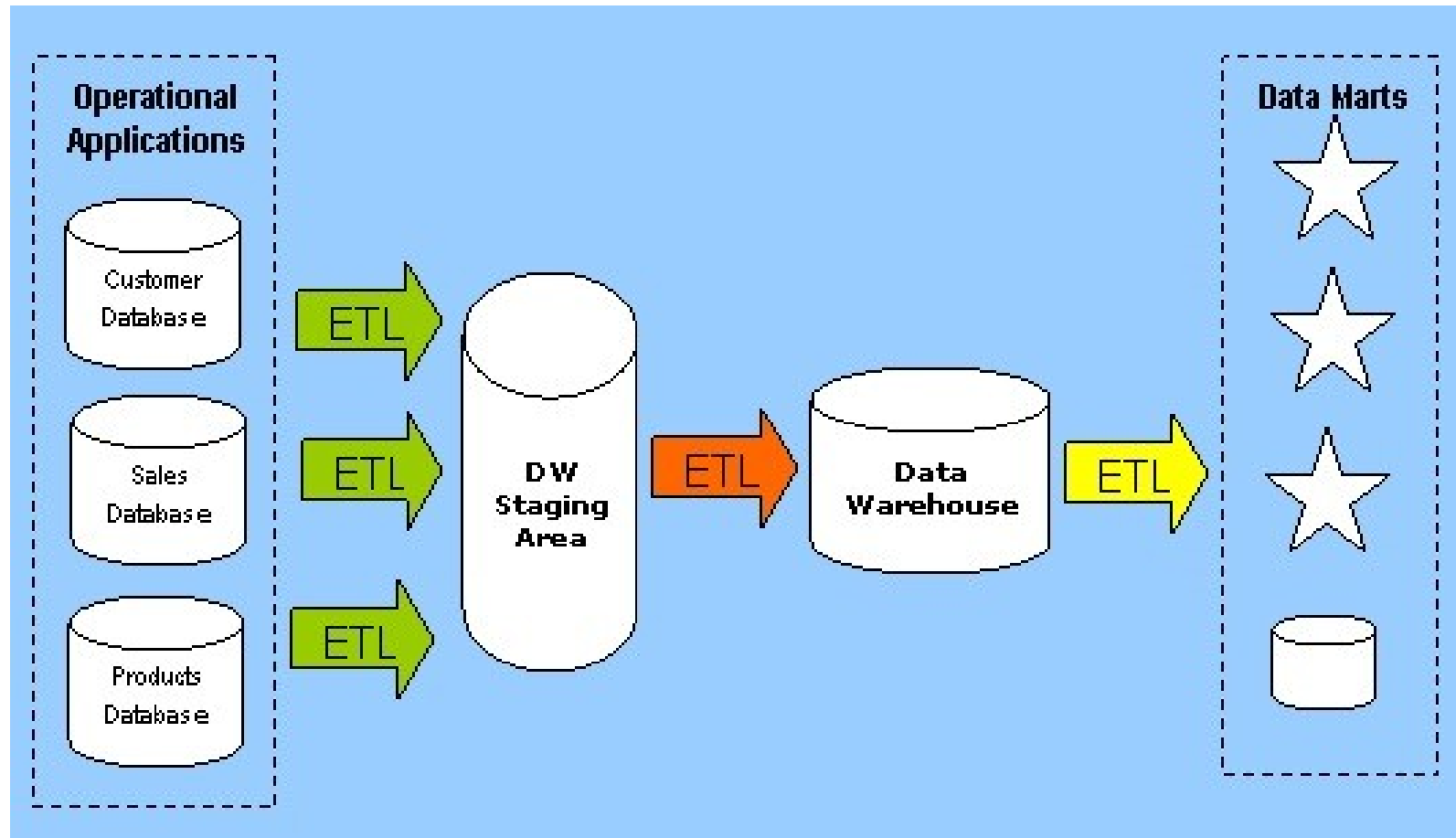
[Learning from Imbalanced Classification](#)

[Booking: Data Scientist – Machine Learning](#)

Mengumpulkan data awal

- Mendaftar data yang ada
- Membutuhkan tools untuk mengumpulkan data untuk integrasi data dari berbagai sumber
 - **SQL POWER ARCHITECT**

Proses integrasi Data



Output Pengumpulan data

- Daftar data yang di hasilkan dan dimana data tersebut berada
- Cara mendapatkan
- Permasalahan dari data

Mendeskripsikan data

- Mengamati secara kasar dan yang tampak dari data yang diperoleh dan mendokumentasikan deskripsi data tersebut.

Output :

- format dari data,
- jumlah data,
 - jumlah record dan field dari masing-masing tabel, identitas dari field-field (atribut-atribut) dan karakteristik yang tampak dari data yang sudah dikumpulkan

-

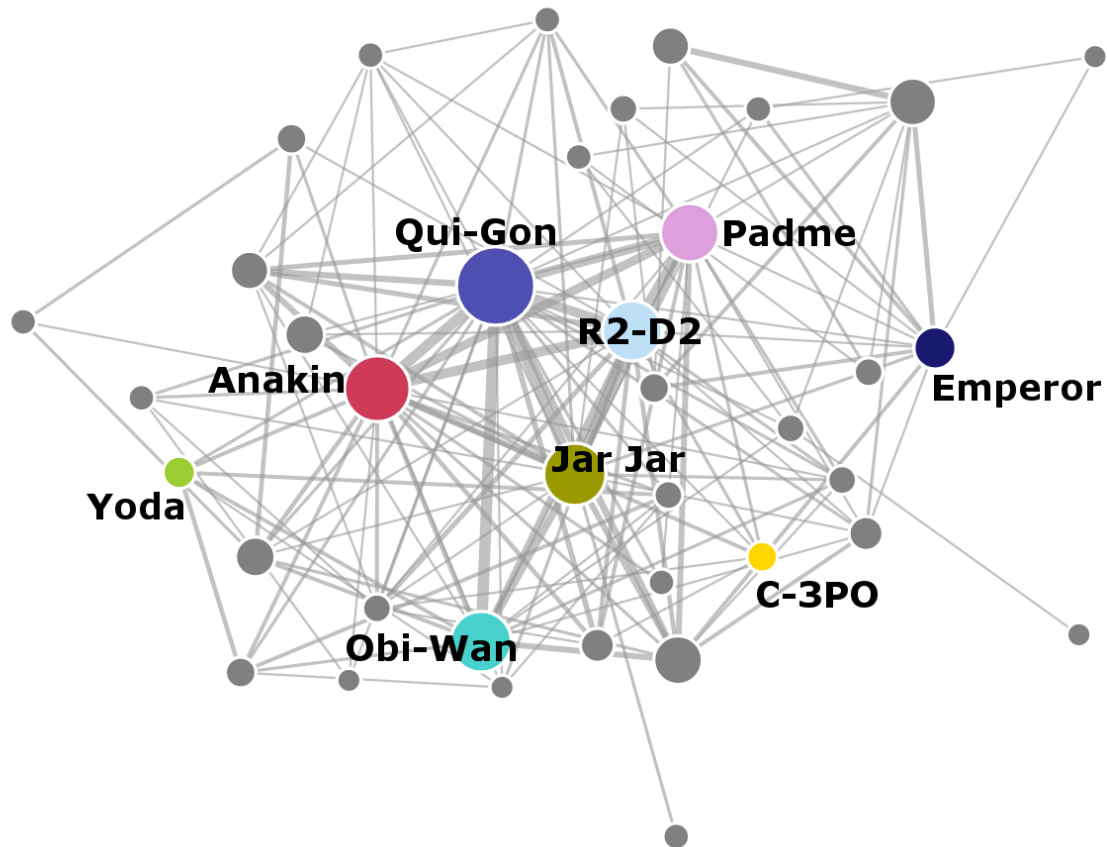
Types of Data Sets

- Record
 - Relational records
 - Data matrix, e.g., numerical matrix
 - Document data: text documents: term-frequency vector
 - Transaction data
- Graph and network
 - World Wide Web
 - Social or information networks
- Ordered
 - Video data: sequence of images
 - Temporal data: time-series
- Spatial, image and multimedia:
 - Spatial data: maps
 - Image data:
 - Video data:

	beer	bread	coke	diaper	milk	soda	sour	sweet	tea	water
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

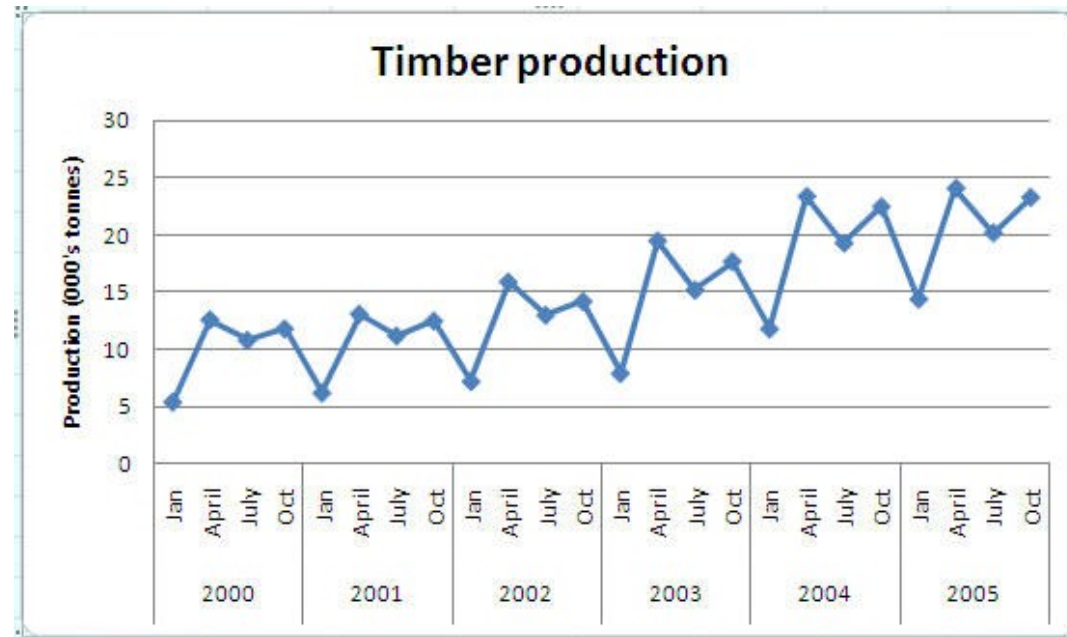
<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Type data graph

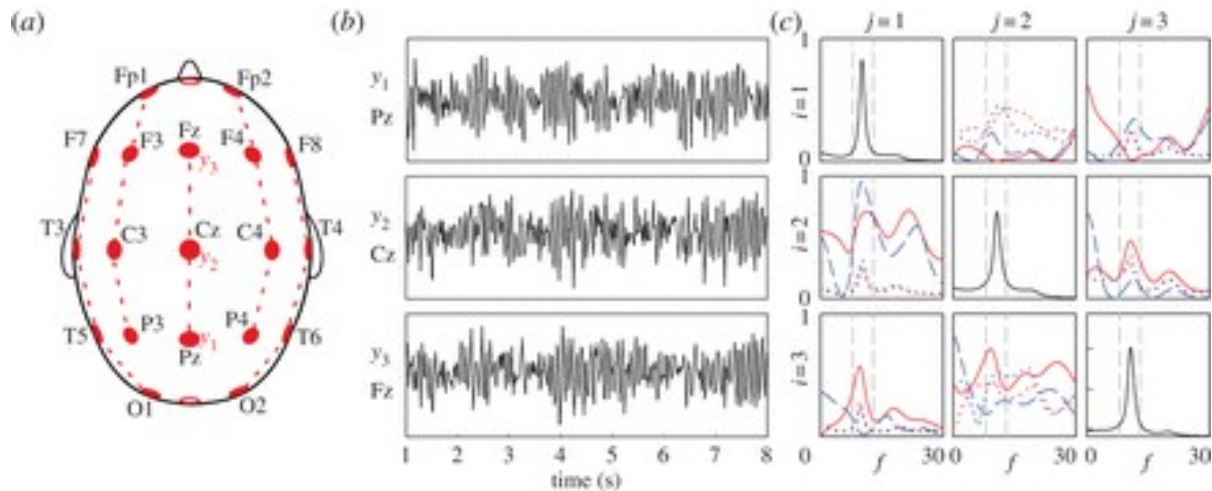


Time series data

Data
produk
si

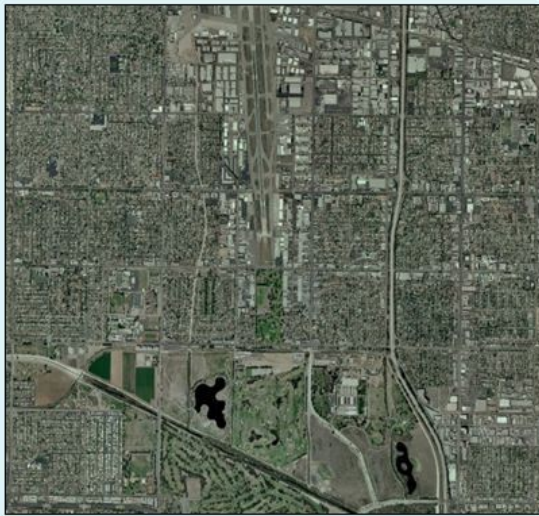


EEG

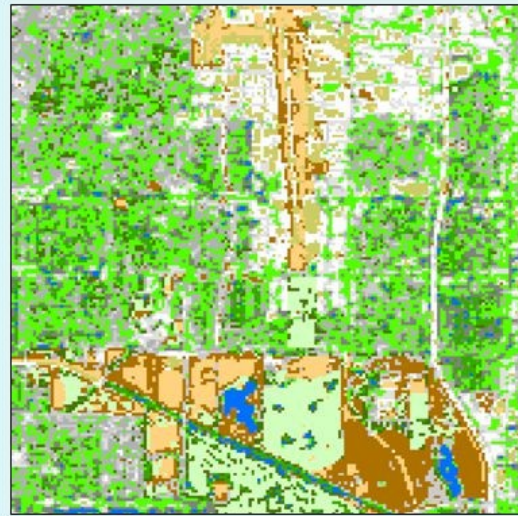


data citra (clustering)

Unsupervised Classification



Natural Color Composite of San Fernando Valley, CA



Data clustered by software and colored to match Land Use types (i.e. blue = water, green = vegetation, etc.)

Data Objects

- **Data object** menyatakan suatu entitas
- Contoh:
 - Database penjualan: customers, barang-barang yang dijual, penjualan
 - Database medis: pasien, perawatan
 - Database universitas: mahasiswa, professor, perkuliahan
- Data objects dijelaskan dengan **attribut-atribut**.
- Baris-baris Database -> data objects;
- Kolom-kolom -> attribut-atribut.

Atribut

- **Atribut (dimensi, fitur, variabel):**
menyatakan karakteristik atau fitur dari data objek
 - *Misal., ID_pelanggan, nama, alamat*
- Tipe-tipe:
 - Nominal
 - Ordinal
 - Biner
 - Numerik:
 - Interval-scaled
 - Ratio-scaled

Attribute Types

- **Nominal:** kategori, keadaan, atau “nama suatu hal”
 - *Warna rambut*
 - Status , kode pos, dll, NRP dll
- **Binary :** Atribut Nominal dengan hanya 2 keadaan (0 dan 1)
 - Symmetric binary: keduanya sama penting
 - Misal: jenis kelamin,
 - Asymmetric binary: keduanya tidak sama penting.
 - Misal : medical test (positive atau negative)
 - Dinyatakan dengan 1 untuk menyatakan hal yang lebih penting (positif HIV)
- **Ordinal**
 - Memiliki arti secara berurutan, (ranking) tetapi tidak dinyatakan dengan besaran angka atau nilai.
 - *Size = {small, medium, large}*, kelas, pangkat

Atribut Numerik

- Kuantitas (integer atau nilai real)
- **Interval**
 - Diukur pada skala dengan unit satuan yang sama
 - Nilai memiliki urutan
 - *tanggal kalender*
 - No true zero-point
- **Ratio**
 - Inherent **zero-point**
 - Contoh: Panjang, berat badan, dll
 - Bisa mengatakan perkalian dari nilai objek data yang lain
 - Misal : panjang jalan A adalah 2 kali dari panjang jalan B

Atribut Discrete dan kontinu

- **Atribut Diskrit**

- Terhingga, dapat dihitung walaupun itu tak terhingga
 - Kode pos, kata dalam sekumpulan dokumen
- Kadang dinyatakan dengan variabel integer
- Catatan Atribut Binary: kasus khusus atribut diskrit

- **Atribut Kontinu**

- Memiliki nilai real
 - E.g., temperature, tinggi, berat
- Atribut kontinu dinyatakan dengan floating-point variables

Explorasi data (menyelidiki data)

- analisa statistik sederhana, hubungan antara atribut
- Output :

Ploting data

hasil ekplorasi berupa grafik

Verifikasi kualitas data

- **Apakah ada missing value (isian dari kolom-kolom apakah lengkap**
- **Melihat kelengkapan data secara keseluruhan**

Persiapan data

- mempersiapkan data mencakup semua aktifitas untuk membangun dataset akhir(data yang siap untuk dijadikan input bagi model data mining
- Tugasnya adalah memilih table, record dan atribut juga tranformasi dan membersihkan data

Preprocessing data

- Tahapan adalah memilih atribut dan record
- Transformasi data
- Membersihkan data

Data Science Training



New York City

March 18, 19, 20, 2015 - attend one, two, or all three days

Data Science Training: New York City, March 18, 19, 20 - attend 1, 2, or 3 days

Subscribe to [our newsletter on Analytics, Big Data, Data Mining](#) | [Follow @kdnuggets](#) | [voted Best Big Data Twitter](#) | [f](#) [in](#) [s](#) [Contact](#)

[KDNuggets Home](#) » [News](#) » [2014](#) » [2014 Oct News, Features](#) » [CRISP-DM, still the top methodology for analytics, data mining, or data science projects \(14:n28 \)](#)

Latest News

- Top stories for Feb 15-21: 10 things statistics taught us about big data analysis; History of Data Science in 5 strands
- Prismatic Interest Graph [API]: Organize and Recommend Content
- Top KDNuggets tweets, Feb 18-19: New Face Detection Algorithm to revolutionize search; How to transition from Excel to R
- Megaputer: Data Analysis Consultant
- Syracuse University: Interdisciplinary Faculty

CRISP-DM, still the top methodology for analytics, data mining, or data science projects

[f](#) [in](#) [g+](#) [2](#) [Share](#) [25](#) [Tweet](#) [48](#)

[◀ Previous post](#)


[Next post ▶](#)

CRISP-DM remains the most popular methodology for analytics, data mining, and data science projects, with 43% share in latest KDNuggets Poll, but a replacement for unmaintained CRISP-DM is long overdue.

[comments](#)



Tools

← → ↻  www.cs.waikato.ac.nz/ml/weka/

Project

Software

Book

Publications

People

Related

Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

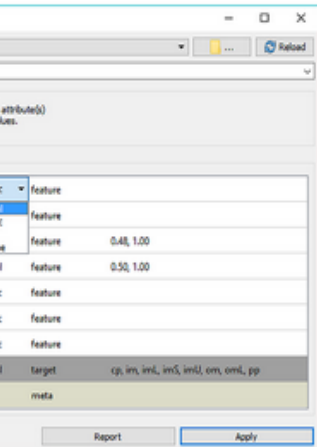
Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

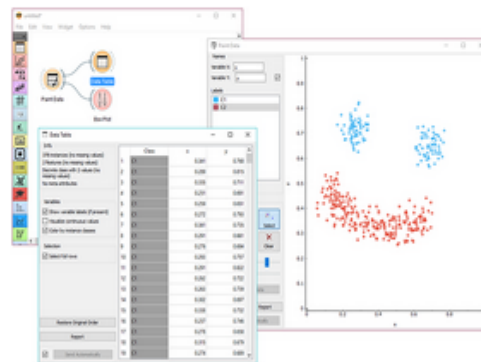
Yes, it is possible to apply Weka to **big data**!

Data Mining with Weka is a 5 week MOOC, which was held first in late 2013. Check out the **MOOC site** for video lectures and details on how to enrol into this course and a new, advanced Weka course.

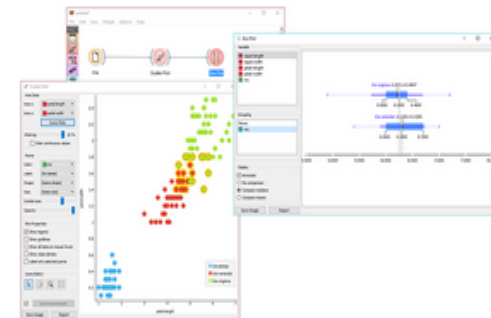
Screenshots



your data in the File widget.



Paint a two-dimensional data set.



Data selection in Scatter Plot is visualised in a Box Plot.

- Memahami Data
 - Objek Data dan Type atribut
 - Statistik deskriptif dari data
 - Visualisasi data
 - Mengukur Data Similarity dan Dissimilarity

- Pre-proses data
 - Pengantar preproses data
 - Membersihkan data
 - Reduksi data
 - Transformasi data dan diskritisasi data
- Association Rule
 - Apriori Algorithm

- Klasifikasi
 - Konsep dasar
 - Pohon Keputusan
 - Naive Bayes
 - Bayesian Network
 - Backpropagation
 - EM
 - Evaluasi model klasifikasi

- Analisa Kluster
 - Konsep dasar
 - Metode Partisi
 - Metode Hirarki

- Outlier Detection
 - Pendekatan Statistik

- Tugas 40 %
- UTS 30 %
- UAS 30 %