

# Python for Data Science

Ch.0 Introduction

Ahmad Rio Adriansyah

# Instruktur

- Nama : Ahmad Rio Adriansyah
- Email : [ahmad.rio.adriansyah@gmail.com](mailto:ahmad.rio.adriansyah@gmail.com)
- HP : 081573954126
- Panggilan : Rio / Arasy
  - Matematikawan,
  - Programmer,
  - Dosen,
  - Tech. Enthusiast



# Peserta

- Nama ?
- Pekerjaan ?
- Pengalaman dengan bahasa pemrograman ?
- Pengalaman dengan data analisis ?

# Outline

- Introduction
- Review Python (Data Structure)
- Anaconda dan Jupyter Notebook
- Modul Pandas : Series dan Data Frame
- Input Data
- Pengolahan Data
- Visualisasi Data
- Machine Learning

# Tips and Trick

- Ada teori dan praktek.
- Dicoba baris per baris, latihan per latihan.
- Jangan takut salah. Tidak ada kesalahan yang tidak bisa diperbaiki.
- Jangan ragu bertanya. Instrukturnya tidak menggigit orang yang bertanya (mungkin).

*“Belajar programming itu seperti belajar renang. Gak ada yang jadi ahli hanya dengan melihat saja.”*

Ayo kita mulai  
Mulai dengan basmalah

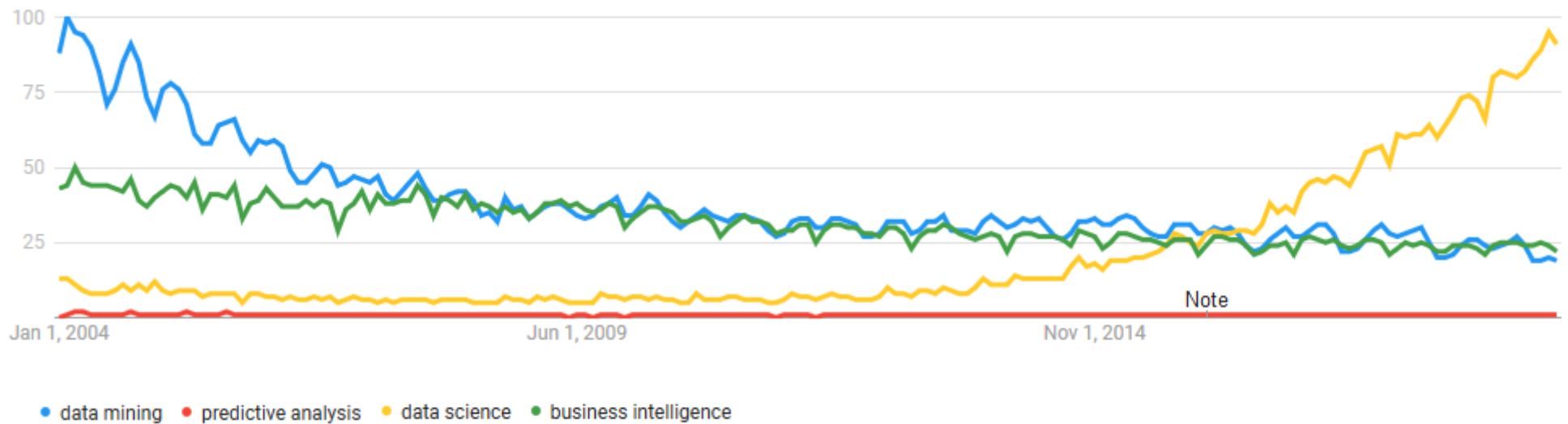
# Apa itu Data Science?

Ilmu yang berkaitan dengan **analisis data** dan **mengekstrak informasi** dari data.

Istilah-istilah lainnya :

- Business Intelligence (BI)
- Business Analytics
- Data Analytics
- Data Mining
- Predictive Analysis

# Google Trends





# Bidang Terapan Data Science



- Shelf-space optimization
- Pricing
- Product Bundling
- Promotions
- Up-sell / Cross-sell



- Risk Modeling
- Fraud Prediction
- Customer Segmentation
- Portfolio Optimization



- Market Basket Analytics
- Marketing Mix Analysis
- Product Pricing
- Sales Forecasting
- A/B Testing



- Clinical Trials of New Drugs
- A/B Testing
- Genetics Analysis
- Epidemic Forecasting & Control



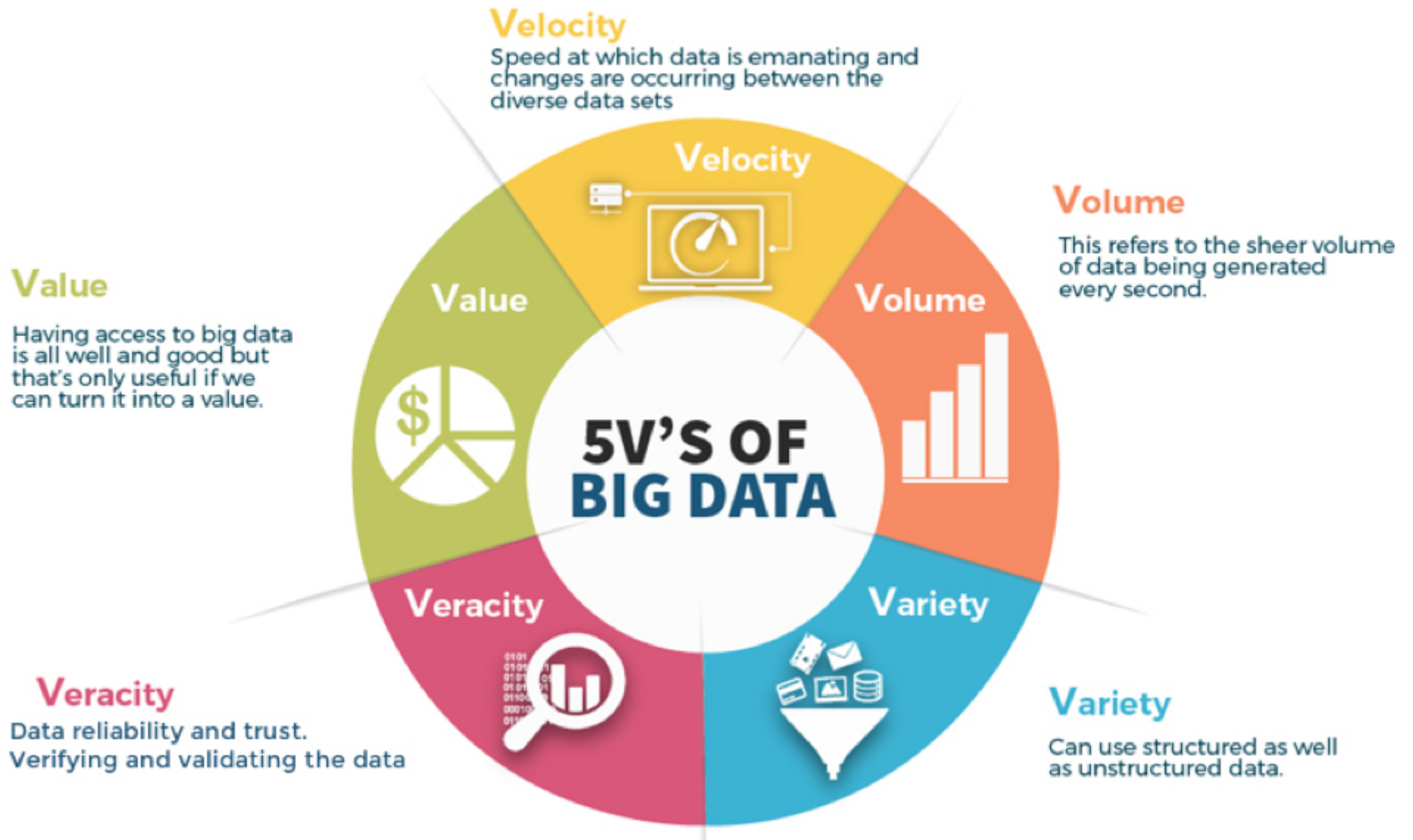
- Customer Acquisition Strategies
- Churn Analytics and Control
- Up-Sell/ Cross-Sell
- Product Bundling



- A/B Testing
- Recommendation Engine
- Media Mix Analytics
- Advertisements Optimization
- Customer Acquisition

Analytics is used by almost all industries and functional areas

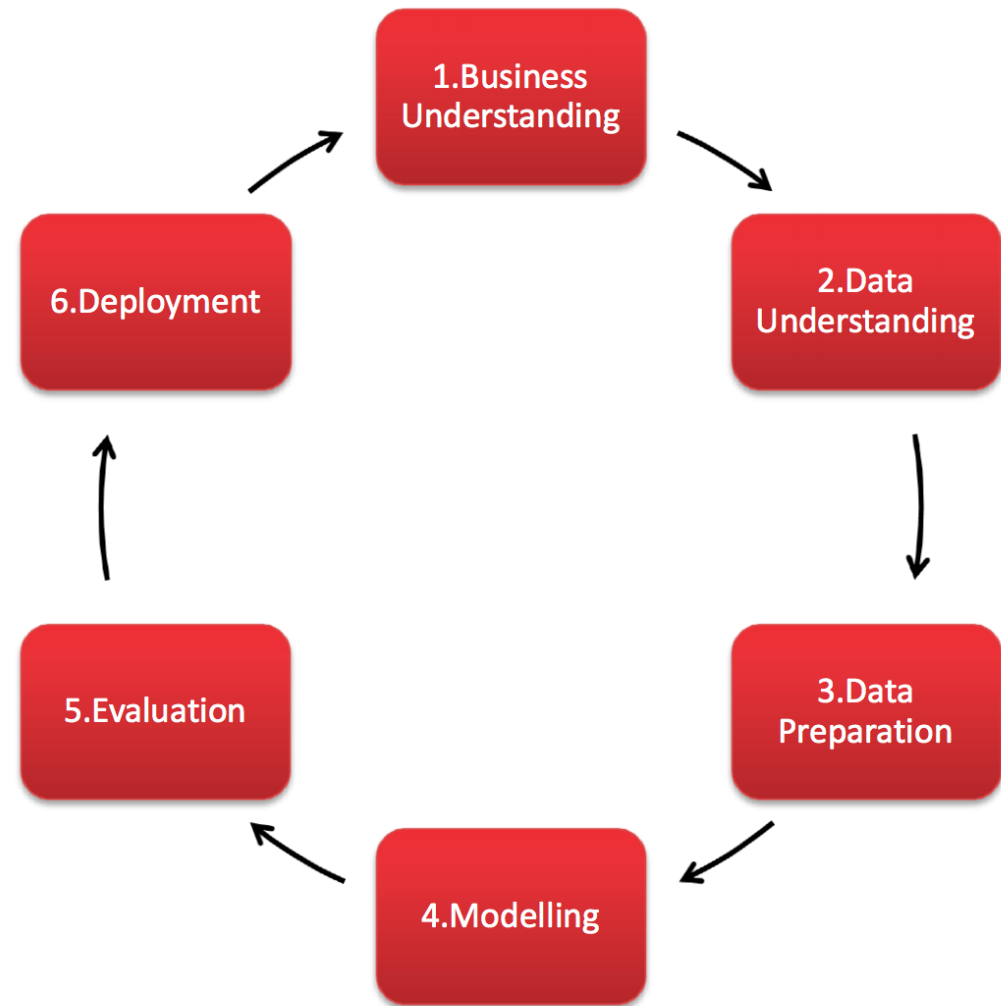
# Big Data



# Pekerjaan Data Science

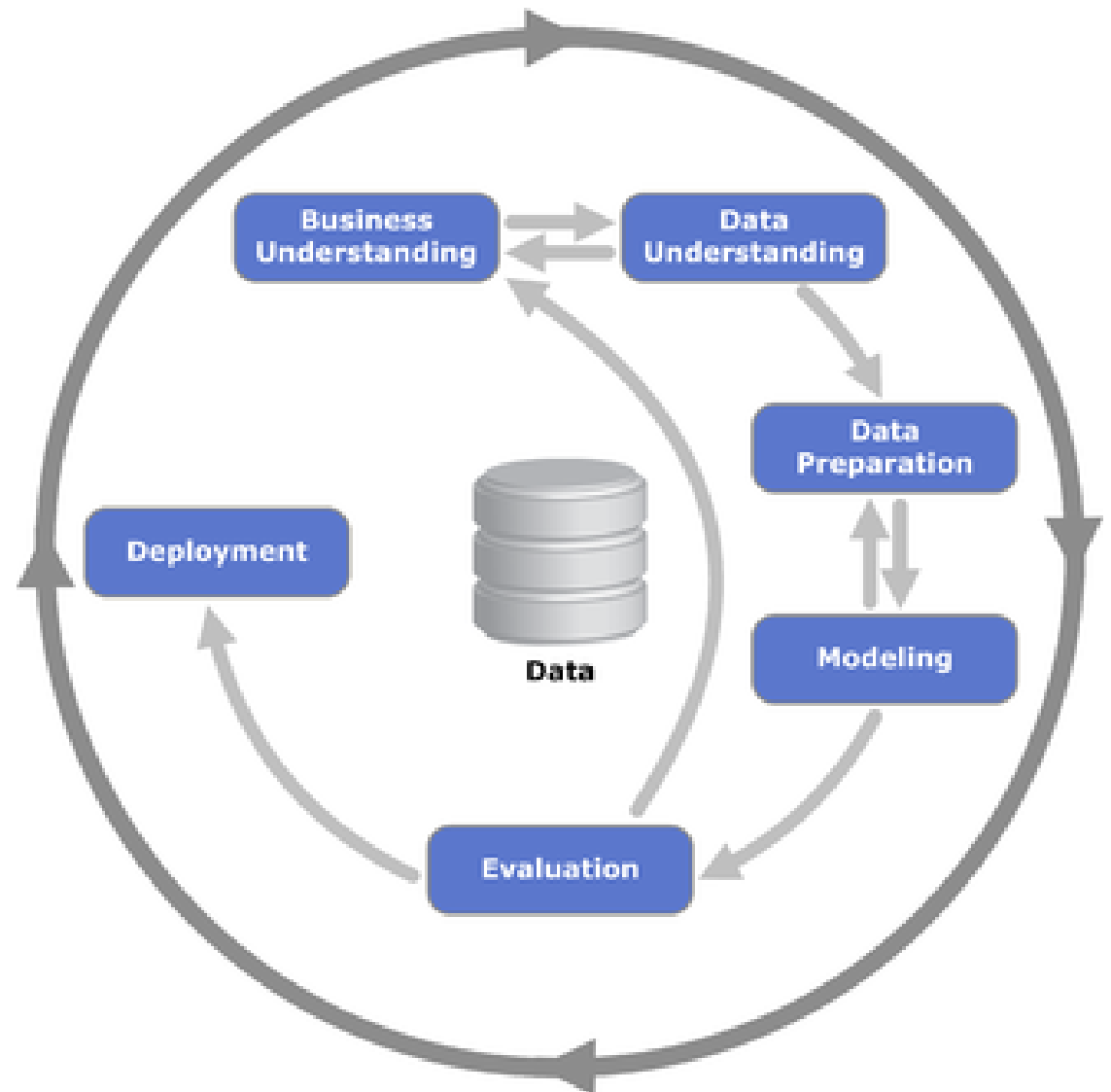
- CRISP-DM

Cross Industry  
Process  
for  
Data Mining



# Pekerjaan Data Science

- Diagram proses hubungan antara fase-fase pada CRISP-DM



# Prerequisite

- Pengetahuan tentang statistik
- Pengetahuan tentang programming
- Pengetahuan tentang area yang dianalisis (domain knowledge)

Bahasa yang populer digunakan

- R
- Python

# R vs Python

- R berasal dari komunitas statistikawan
- Python dari komunitas ilmuwan komputer

Secara kegunaan dan performa kurang lebih sama, pilihlah yang sesuai dengan kebutuhan

# Kenapa Python

- Mudah dipelajari
- Mudah dibaca
- Skalabel
- Memiliki banyak library pendukung
- Mudah diintegrasikan
- Komunitas dan ekosistem yang aktif

# Library Python

Library	Usage
numpy, scipy	Scientific & technical computing
pandas	Data manipulation & aggregation
mlpy, scikit-learn	Machine learning
theano, tensorflow, keras	Deep learning
statsmodels	Statistical analysis
nltk, gensim	Text processing
networkx	Network analysis & visualization
bokeh, matplotlib, seaborn, plotly	Visualization
beautifulsoup, scrapy	Web scraping