

# PERTEMUAN 6

## Tools Sains Data

# Tools sains data

Selain kemampuan seperti matematika, statistic dan yang terkait lainnya

# Tools sains data

## Microsoft Excel

Aplikasi ini biasanya digunakan untuk penolahan data numerik, fungsi statistika sampai dengan melakukan visualisasi data

## SAS (Statistical Analysis System)

Aplikasi ini biasanya digunakan khusus untuk pengolahan data statistik

# Tools sains data

## MATLAB

MATLAB dikembangkan oleh The MathWorks yang mendukung data science khususnya statistik hingga pengembangan user interface. Tools ini unggul dalam melakukan analisis data, komputasi aljabar linear atau matriks, dan pembuatan grafik atau plot untuk visualisasi data yang sangat mendukung. MATLAB juga mudah diintegrasikan ke dalam enterprise dan embedded system yang dapat membantu otomatisasi berbagai tugas dalam data science

# Tools Proyek data science

## Python

Python didesain oleh Guido van Rossum<sup>2</sup> pada 1991 sebagai general-purpose language dan bersifat open-source. Python merupakan bahasa pemrograman populer tetapi juga memiliki properti yang sangat baik untuk programmer pemula, sehingga ideal untuk orang yang belum pernah memprogram sebelumnya. Python juga bersifat cross-platform, artinya bisa digunakan pada sistem operasi Windows, Mac OS, dan Linux. Beberapa kelebihan dari properti tersebut adalah kode yang mudah dibaca, pengetikan

# Python

Bahasa pemrograman tingkat tinggi Penulisan kode/sintaks lebih sederhana dan tersedia banyak library

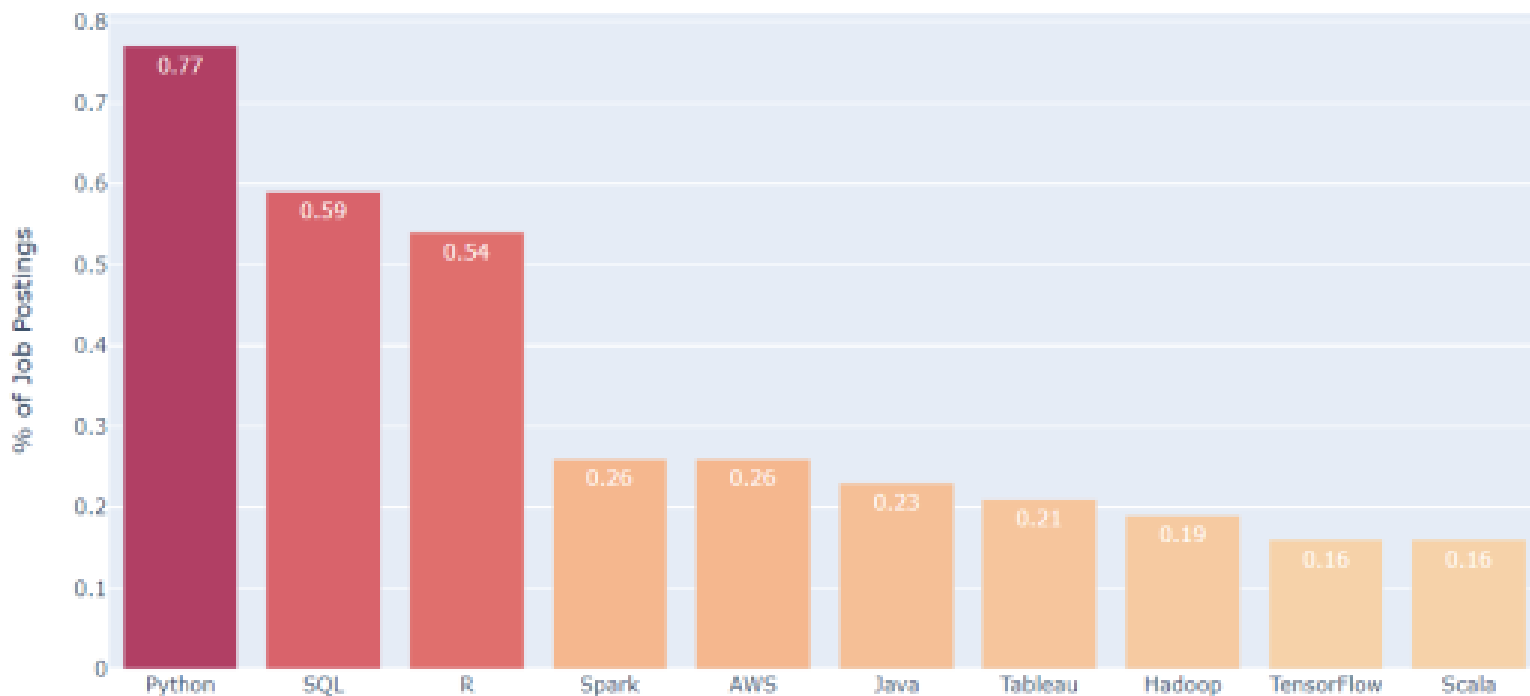
Bersifat open-source dan cross-platform

Python Biasa digunakan oleh:

- Data Analyst
- Data Engineer
- Data Scientist
- Business Intelligence
- ML Engineer

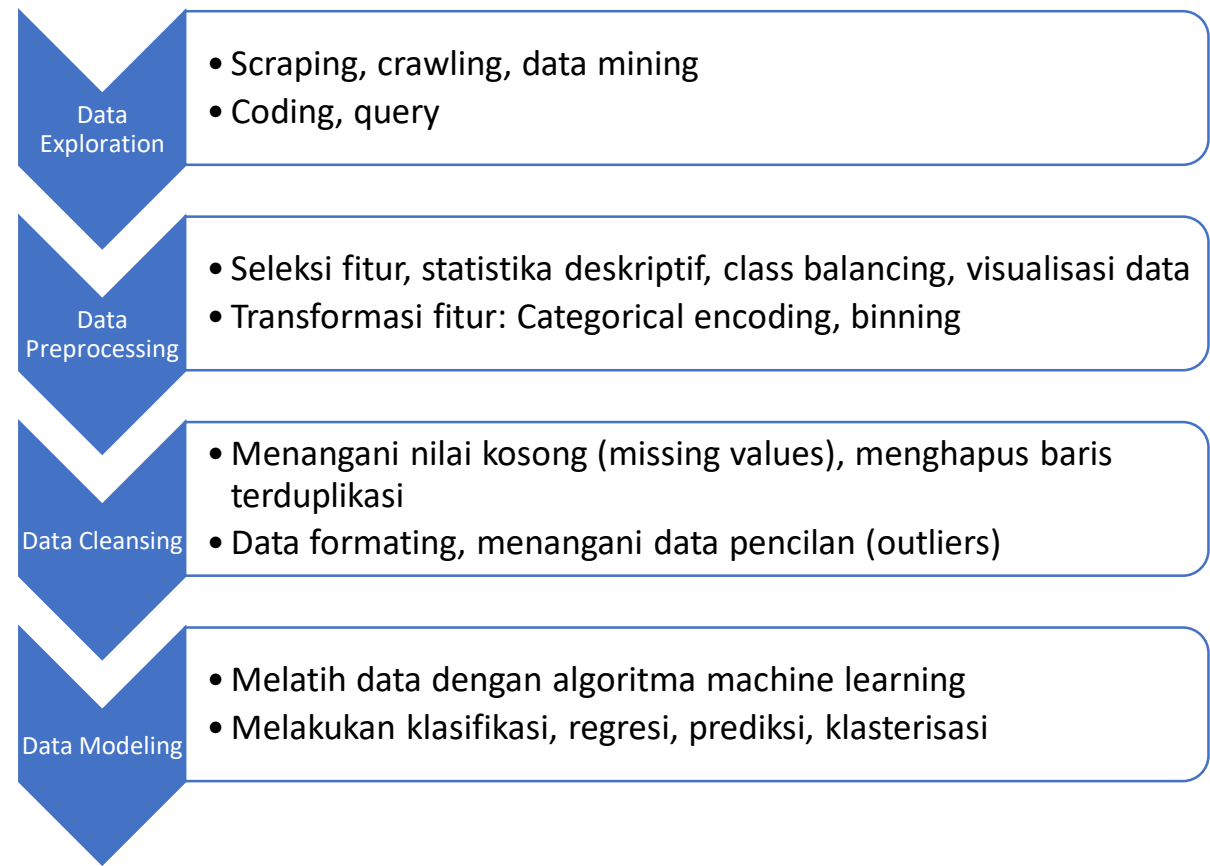
# Python (Lanjutan..)

10 In demand data science skill in 2021



(sumber: <https://towardsdatascience.com>)

# Penerapan Python pada Proyek data Science





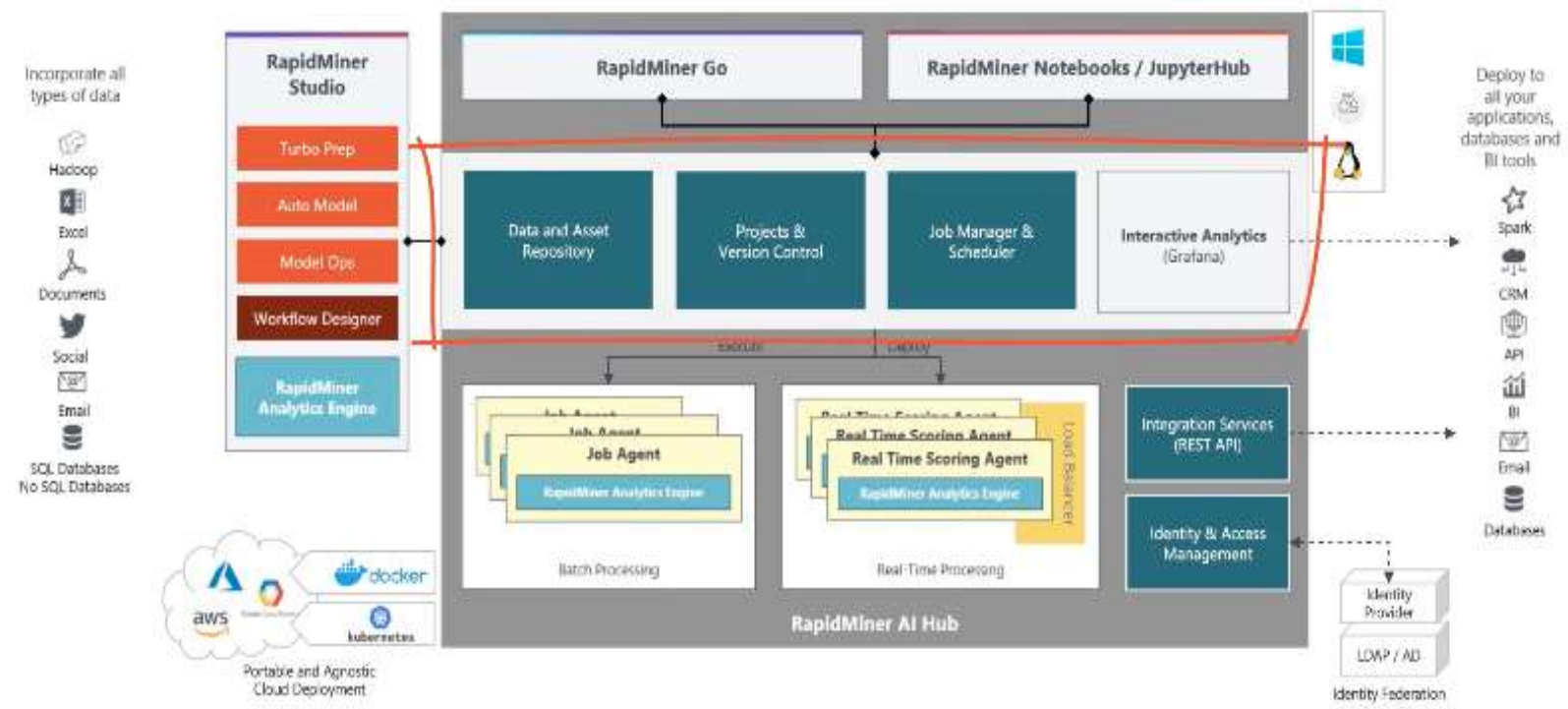
# Rapidminer

Rapidminer merupakan sebuah tools data science yang dikembangkan oleh perusahaan dengan nama yang sama yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analitik prediktif. Ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototipe cepat, dan pengembangan aplikasi dan mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi model, dan pengoptimalan.



<https://rapidminer.com/>

# Penerapan Rapidminer Dalam Datascience



<https://academy.rapidminer.com/learn/video/rapidminer-data-science-platform-overview>

# KNIME

KNIME atau kepanjangan dari Konstanz Information Miner, merupakan sebuah software analisis, pelaporan, dan integrasi data sumber terbuka dan gratis. KNIME mengintegrasikan berbagai komponen untuk pembelajaran mesin dan penambangan data melalui konsep "Building Blocks of Analytics" perpipaan data modular. Antarmuka pengguna grafis dan penggunaan JDBC memungkinkan perakitan node yang memadukan berbagai sumber data, termasuk pra-pemrosesan (ETL: Extraction, Transformation, Loading), untuk pemodelan, analisis data, dan visualisasi tanpa, atau hanya dengan pemrograman minimal



Open for Innovation

# KNIME

<https://www.knime.com/>

# Workflow KNIME Untuk data science

