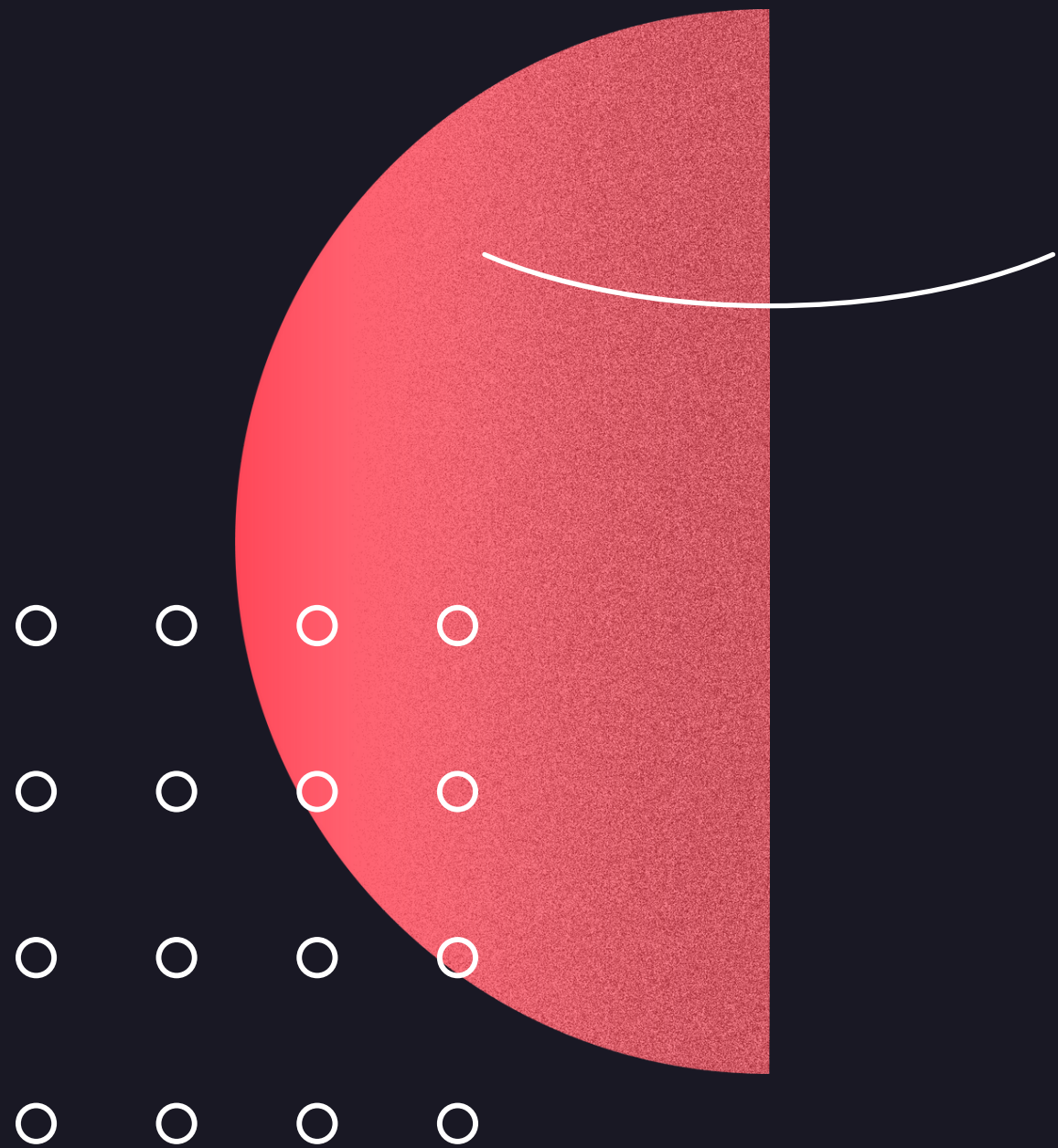


Pengenalan Data Science

Apa itu Data Science?



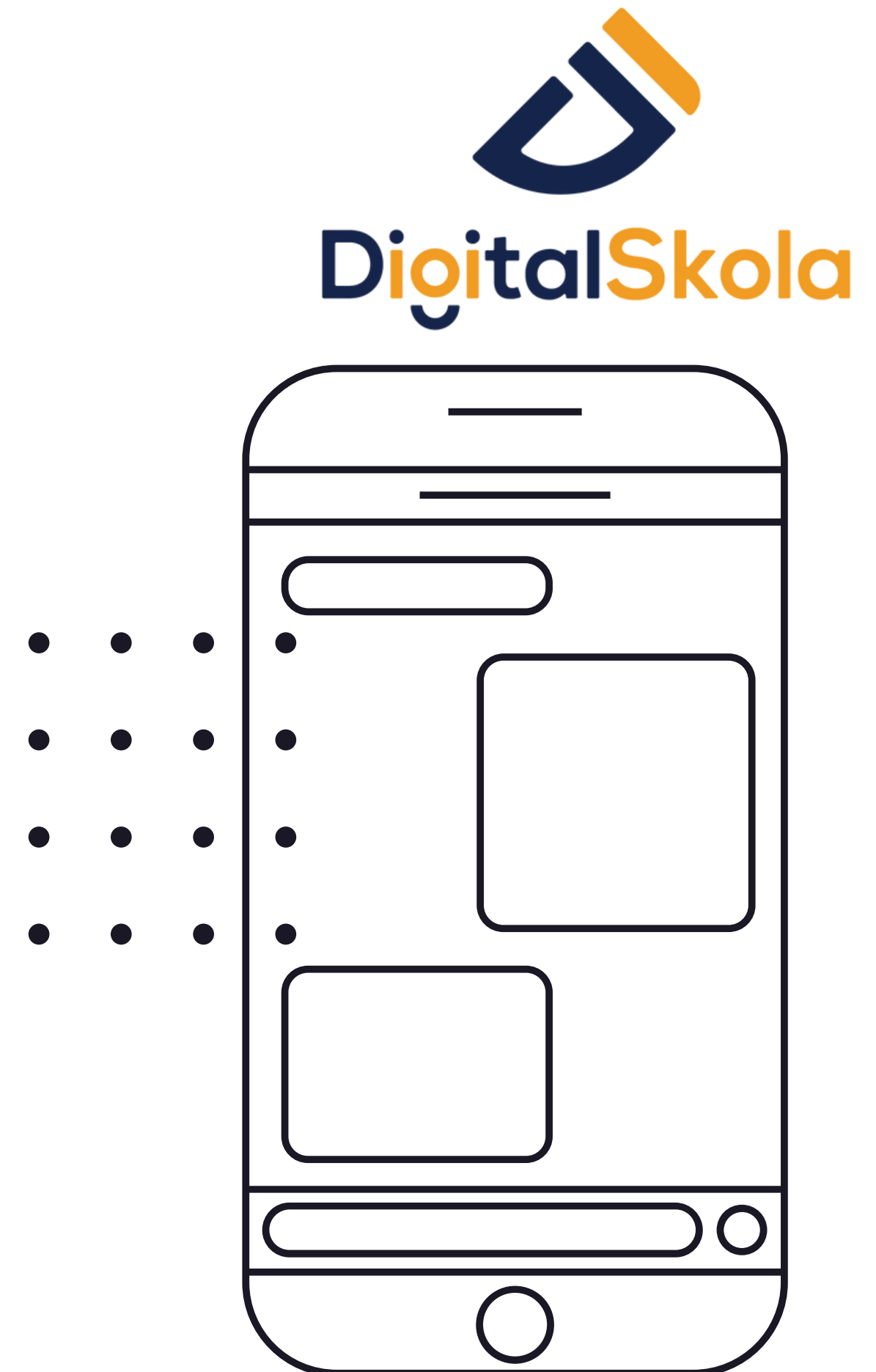
DATA SCIENCE

Terdiri dari 2 kata:

Data : Kumpulan Fakta / Statistik yang dimanfaatkan untuk melakukan analisa

Science : ilmu dan praktikal yang mempelajari sesuatu yang berhubungan dengan kehidupan melalui observasi dan penelitian

Maka dari itu dapat kita simpulkan bahwa data science merupakan ilmu/bidang yang secara khusus mempelajari segala sesuatu tentang data mulai dari mengumpulkan, menyimpan, hingga memproses data tersebut menjadi sesuatu yang bisa disampaikan dan dapat dimanfaatkan dalam proses pengambilan keputusan



BAGIAN DARI DATA SCIENCE

AI (Artificial Inteligence)

Sebuah pengembangan dari ilmu komputer dimana merancang sebuah sistem komputerisasi yang mampu mengerjakan tugas-tugas yang biasanya hanya bisa dilakukan oleh manusia

ML (Machine Learning)

Bagian dari *AI* yang dirancang dengan tujuan agar sistem dapat mempelajari pola dari suatu data tanpa arahan dari penggunaannya

Deep Learning

Bagian dari *ML* yang diciptakan dengan menggunakan beberapa algoritma data yang menyerupai struktur otak manusia

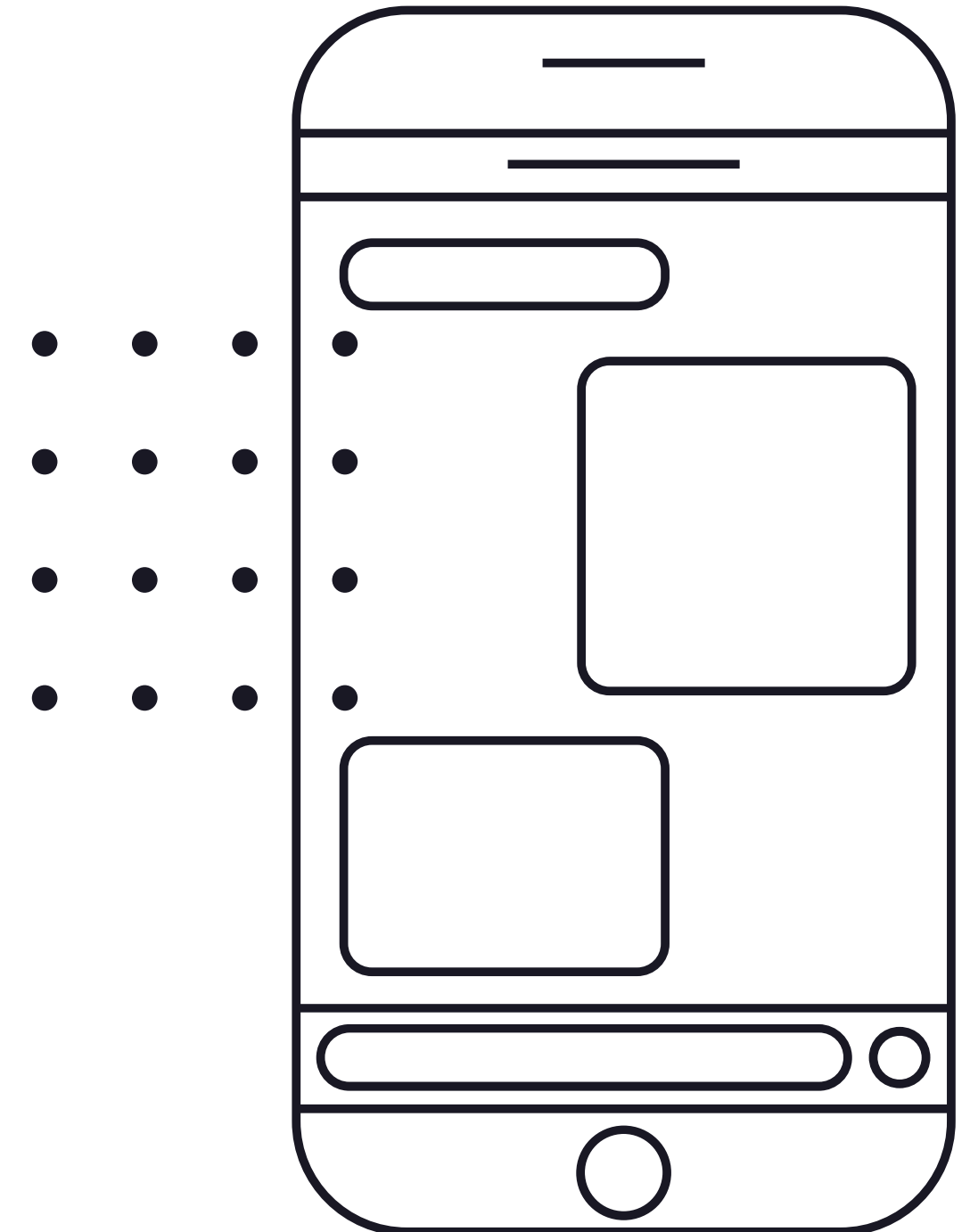
ML (*MACHINE LEARNING*)

Unsupervised Learning

Sistem yang memiliki kemampuan untuk melakukan pengelompokan data berdasarkan sifat dan karakteristik suatu data yang memiliki kemiripan, yang akhirnya terciptalah beberapa kelompok data yang masing masing memiliki data yang identik didalamnya

Supervised Learning

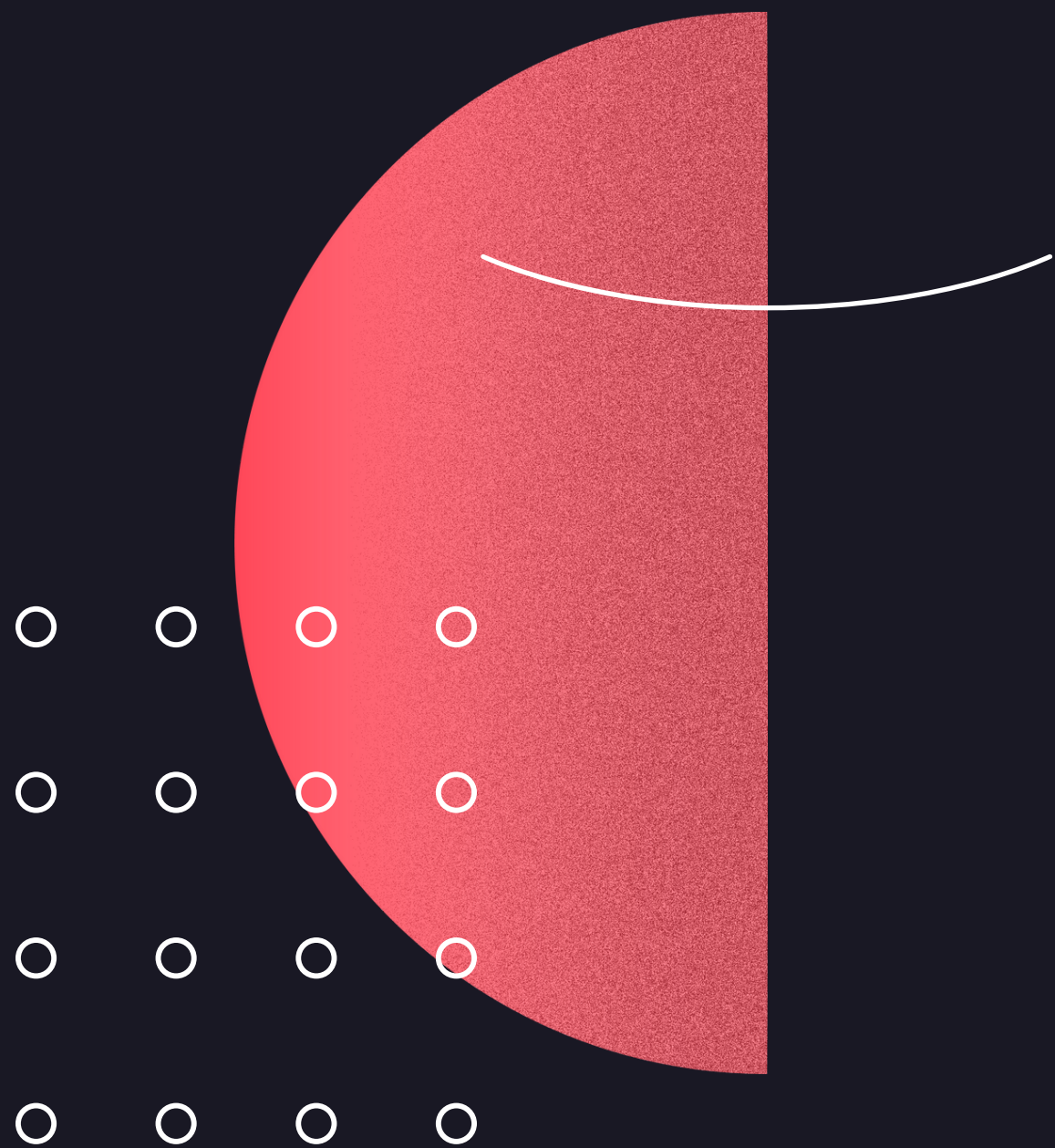
Sistem yang memiliki kemampuan untuk memprediksi dan mengambil keputusan berdasarkan *historikal* atau data yang telah dikumpulkan sebelumnya



PENGUNAAN DATA SCIENCE DALAM BISNIS

1. Membantu dalam pengambilan keputusan perusahaan
2. *fraud detection* (Deteksi data mencurigakan)
3. Memberikan penilaian terhadap suatu data
4. Segmentasi
5. Personalisasi Pelanggan

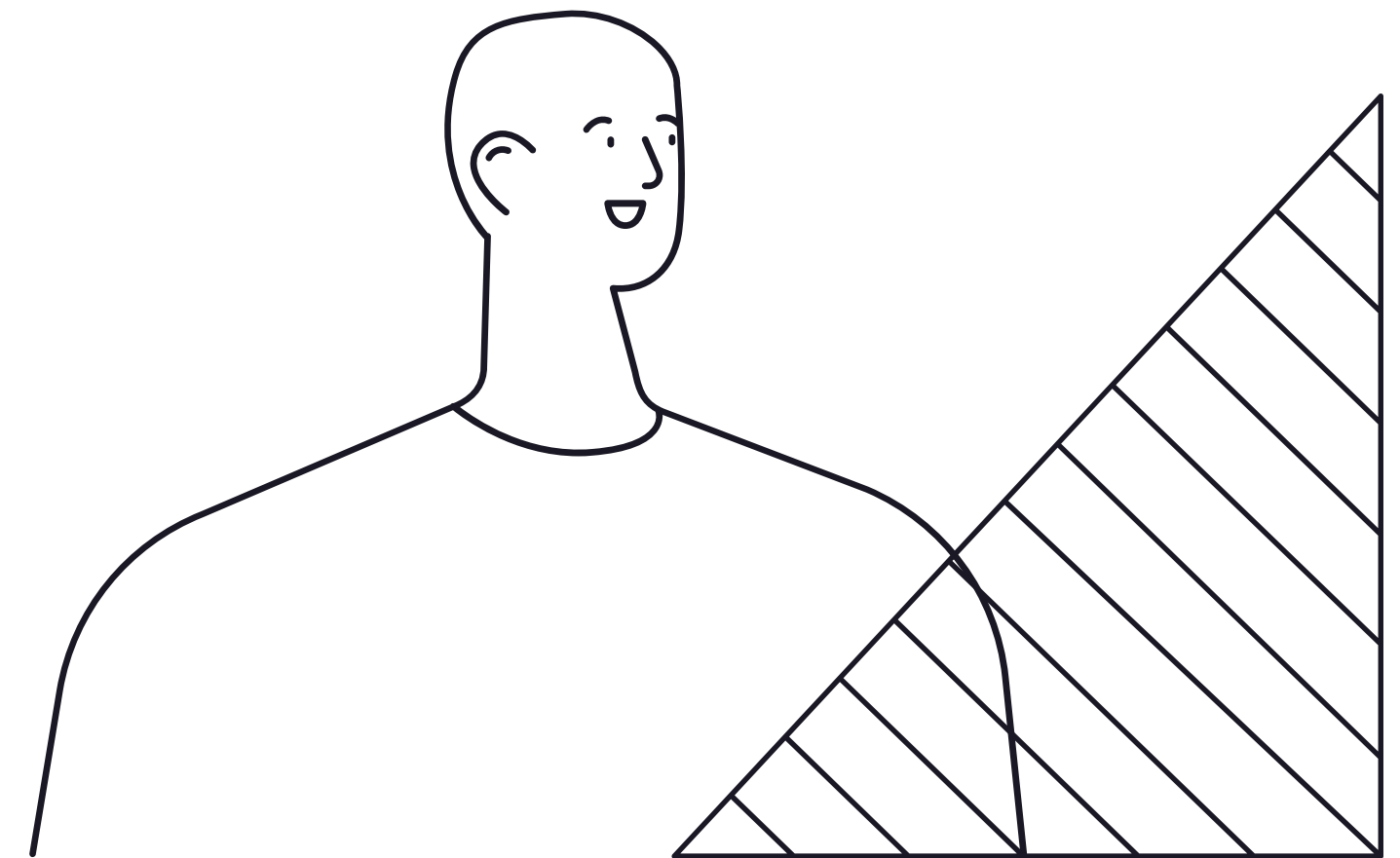
Metodologi Data Science



BUSINESS UNDERSTANDING

Pada Tahap ini seorang Data Scientist harus memiliki pemahaman terhadap bisnis baik secara proses sampai kedalam masalah yang sedang dihadapi, serta mampu menentukan permasalahan dan tujuan akhir yang ingin dicapai oleh pemilik usaha.

Pada tahap ini kita harus sangat aktif menggali informasi yang kita butuhkan serta memahaminya agar dalam proses kedepannya kendala yang akan kita hadpai akan lebih sedikit



ANALYTIC APPROACHES

Setelah memahami bisnis dandan permasalahannya maka kita harus mencari pendekatan *analytics* mana yang paling tepat untuk menyelesaikan permasalahan

- **Descriptive Analytics**
 - Menjelaskan tentang apa yang telah terjadi pada periode sebelumnya, dalam bentuk laporan, presentasi, tabel visual, dll. Dengan tujuan membantu pemilik usaha mengambil keputusan
- **Diagnostic Analytics**
 - Menjelaskan tentang apa yang terjadi pada periode sebelumnya. Hasil dari analisis ini adalah menyimpulkan apa yang menjadi akar permasalahan dari perusahaan.
- **Predictive Analytics**
 - Memprediksi apa yang akan terjadi di masa depan menggunakan data-data yang telah dikumpulkan
- **Prescriptive Analytics**
 - Menyusun Rencana/pola apa yang akan dilakukan untuk mencapai target dimasa depan berdasarkan data-data yang telah dikumpulkan

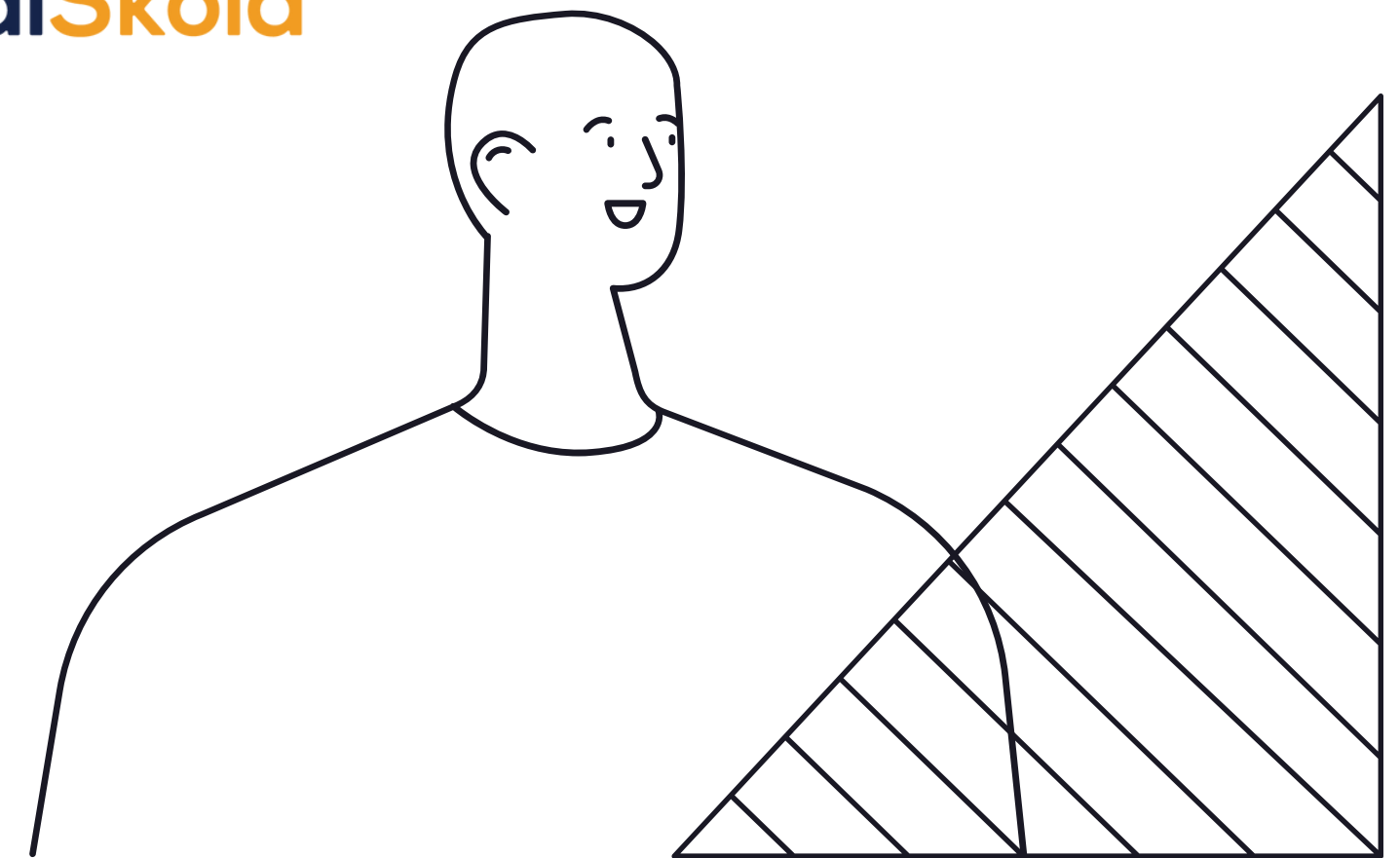
Data Requirements

- Pada tahap ini kita menganalisa dan menentukan data apa yang akan kita perlukan untuk masuk ke tahap berikutnya yaitu data collection.

Data Collection

- Proses Pengumpulan data baik itu data yang terstruktur, tidak terstruktur, maupun data yang semi terstruktur
- Proses pengumpulan data dapat dilakukan dengan mengambil data internal seperti data pelanggan, data penjualan dan lainnya
- Data eksternal juga dapat diperlukan pengumpulan data ini bisa dilakukan dengan melakukan *web scrapping*, kuisisioner, survey dan lain sebagainya

DATA REQUIREMENTS & DATA COLLECTION



DATA UNDERSTANDING & DATA PREPARATION

Data Understanding

- Pada tahapan ini Data Scientist akan mencoba untuk memahami seluruh data yang telah dikumpulkan dan apakah kriteria data yang dikumpulkan sudah sesuai untuk menjadi acuan permasalahan yang dihadapi

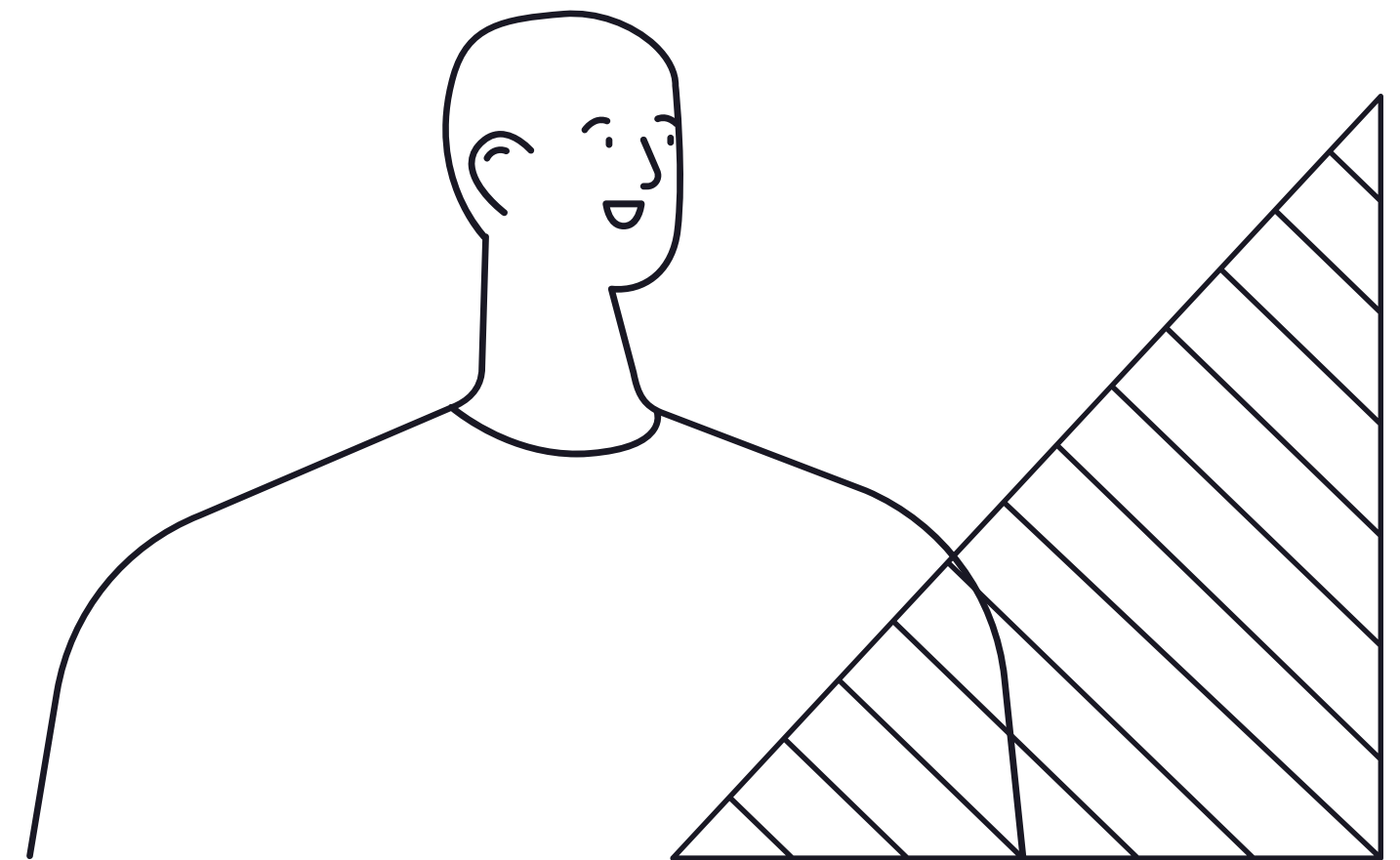
Data Preparation

- Setelah memahami data yang telah dikumpulkan selanjutnya data scientist akan melakukan *data cleansing*, menghapus data yang sama/terduplikasi, data outlier, dll
- Tujuan dari tahap ini adalah memastikan tidak adanya data yang error pada tahap selanjutnya

Setelah data selesai diproses maka tahap selanjutnya adalah mengembangkan model mesin , ada dua model mesin sesuai dengan masalah yang akan diselesaikan

- Descriptive Modelling
 - Proses matematika yang menjelaskan hubungan sebab akibat yang menjadi permasalahan yang dihadapi
- Predictive Modelling
 - Proses yang akan memberikan kemungkinan dan prediksi akan sebuah hasil yang diinginkan berdasarkan data yang telah dikumpulkan

MODELING



DEPLOYMENT & FEEDBACK

Deployment

- Setelah sebuah model telah memenuhi semua kriteria yang diperlukan maka model tersebut akan di luncurkan

Feedback

- Setelah model di luncurkan maka tahap terakhir adalah memastikan model tersebut bekerja secara stabil dan terus bisa diandalkan
- Jika kemudian ada hal yang mempengaruhi kinerja model tersebut maka akan dilakukan penyesuaian lagi



TERIMA KASIH