



# Data Science Methodology

Learning Progress Review Week 2

By:  
**MARVEL TEAM**

Fikrie | Natalia | Satria

# 1. Data Analytics Lifecycle

*Memahami data analytics lifecycle*

# Data Analytics Lifecycle

**Data analytics lifecycle** ini mampu mendorong nilai bisnis dan inovasi melalui pengembangan yang berkelanjutan.

Dalam kinerjanya, *data analytics lifecycle* dapat bekerja hanya dalam tiga tahap hingga pada tahap *diagnostic analytics*. Hal ini dimungkinkan karena tidak semua pendekatan pemecahan masalah perlu masuk pada tahap 4 dan 5 (*predictive & prescriptive analytics*) yang memerlukan gambaran bisnis di masa depan.

## Planning Analytics

Apa rencana dalam menjawab masalah bisnis yang kita hadapi?

## Descriptive Analytics

Apa yang sedang terjadi?

## Prescriptive Analytics

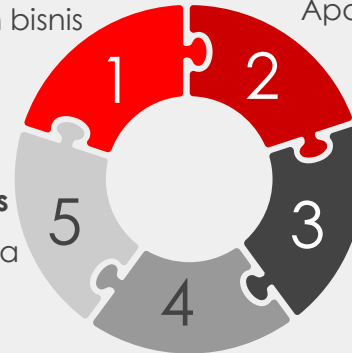
Apa yang seharusnya dilakukan?

## Diagnostic Analytics

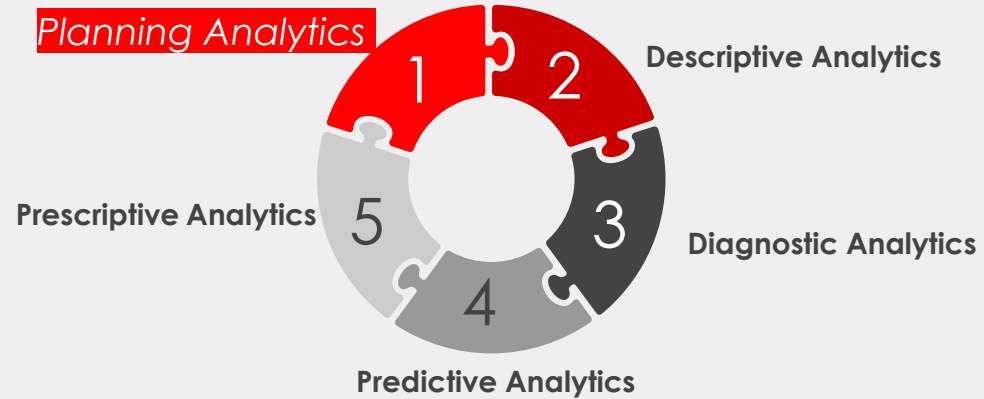
Mengapa itu bisa terjadi?

## Predictive Analytics

Apa yang akan terjadi di masa depan?



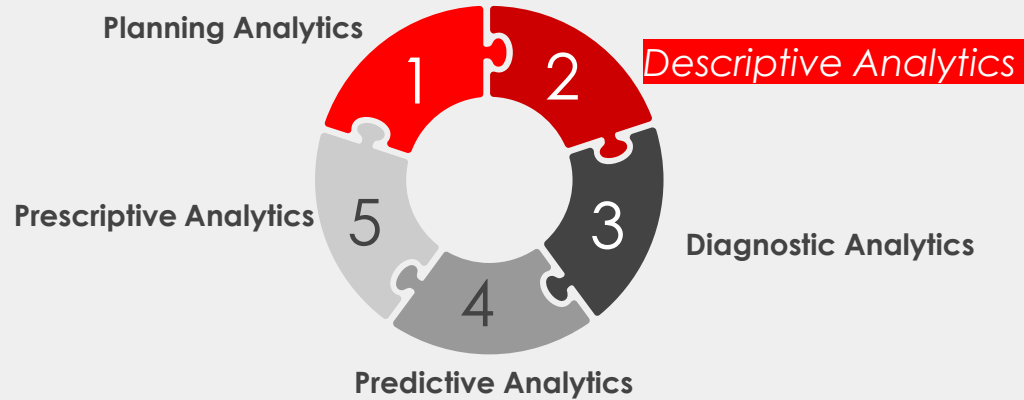
# Data Analytics Lifecycle



**Planning Analytics** berbicara tentang suatu proses di dalam proses *business understanding*. Hal-hal yang perlu dipahami:

- Kinerja bisnis perusahaan
- Identifikasi penyimpangan (deviasi) dari rencana yang ada (plan vs actual)
- Evaluasi skenario yang kemungkinan terjadi ketika menganalisis masalah bisnis yang sedang terjadi
- Memprediksi hasil dan penilaian terhadap resiko serta pertimbangan yang ada

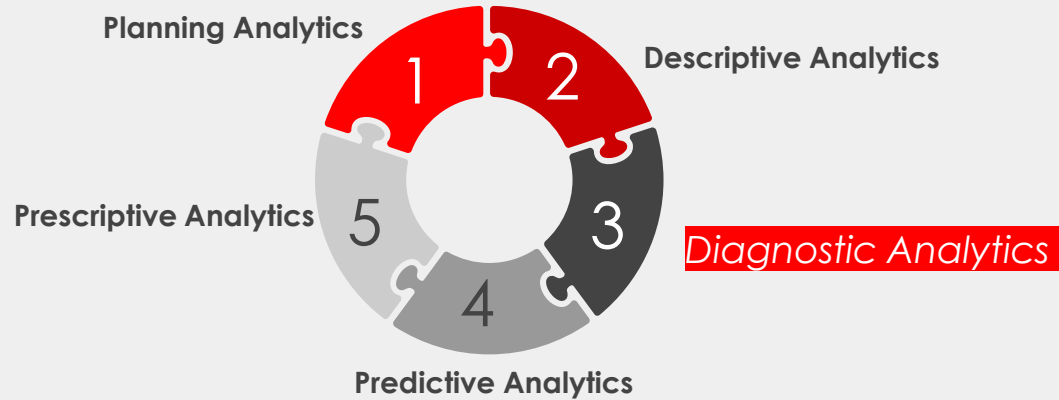
# Data Analytics Lifecycle



**Descriptive Analytics** berkaitan dengan visualisasi data *melalui reports, dashboards, dan scorecards*. Pada tahap ini, hal-hal yang dilakukan :

- Mendefinisikan *business metrics*
- Identifikasi data-data yang dibutuhkan
- *Filter* dan menyiapkan data-data
- Menganalisis data-data
- Mempresentasikan data-data

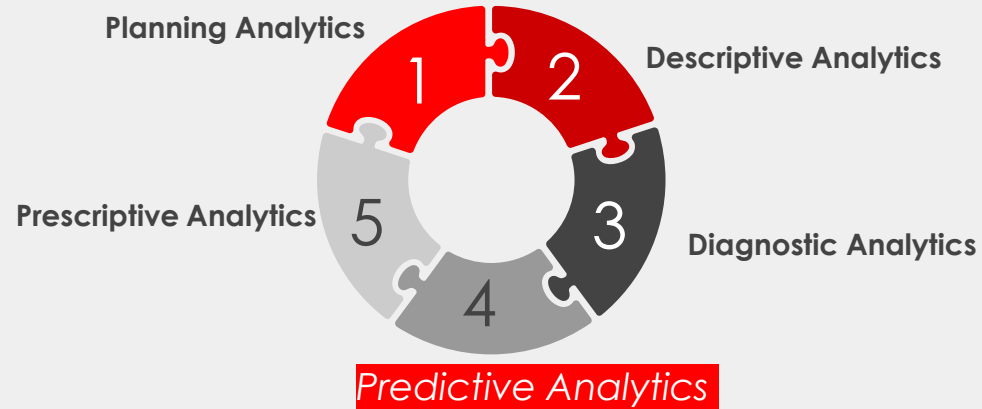
# Data Analytics Lifecycle



**Diagnostic Analytics** membuka secara detail mengapa suatu masalah terjadi. Pada tahap ini, hal-hal yang dilakukan:

- Identifikasi anomali-anomali
- Drilling terhadap analisa data yang ada
- Menentukan hubungan sebab-akibat, contoh : penjualan menurun karena biaya marketing diturunkan dari manajemen

# Data Analytics Lifecycle

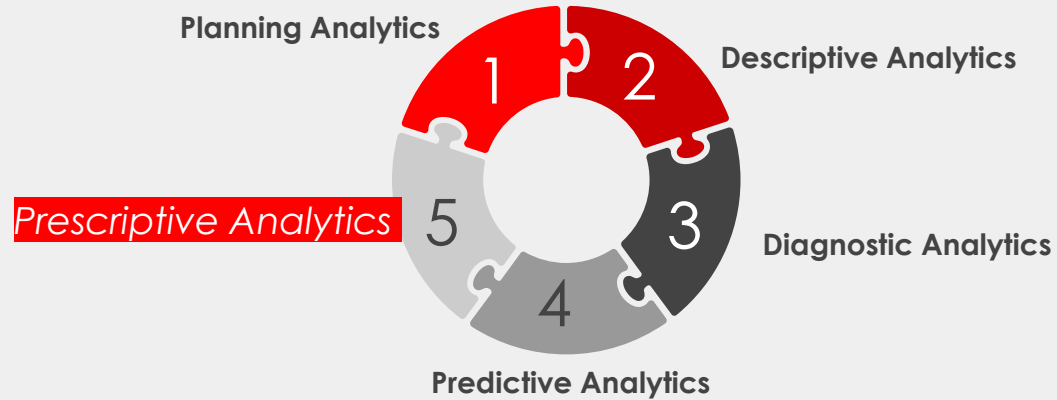


**Predictive Analytics** memprediksi apa yang akan terjadi di masa depan dalam jangka pendek berdasarkan data-data historis. Analisis ini memanfaatkan model *machine learning*.

**Proses ini dilakukan ketika kita membutuhkannya saja.** Tujuan dalam proses ini yaitu:

- Identifikasi *business outcomes*
- Menentukan data yang dibutuhkan untuk data training di dalam model *machine learning*
- Menentukan tipe-tipe analisis yang ada
- Mem-validasi hasil modelling *machine learning*
- Test hasil data prediksinya

# Data Analytics Lifecycle



**Prescriptive Analytics** mencari optimasi solusi dari masalah bisnis yang sedang terjadi. Tujuan tahap ini adalah untuk menghasilkan suatu *decisions making*. **Proses ini mengikutsertakan seluruh peran seperti stakeholder, developer, data engineer dan data scientist. Proses ini juga dilakukan ketika kita membutuhkannya.** Pada tahap ini, hal-hal yang dilakukan adalah :

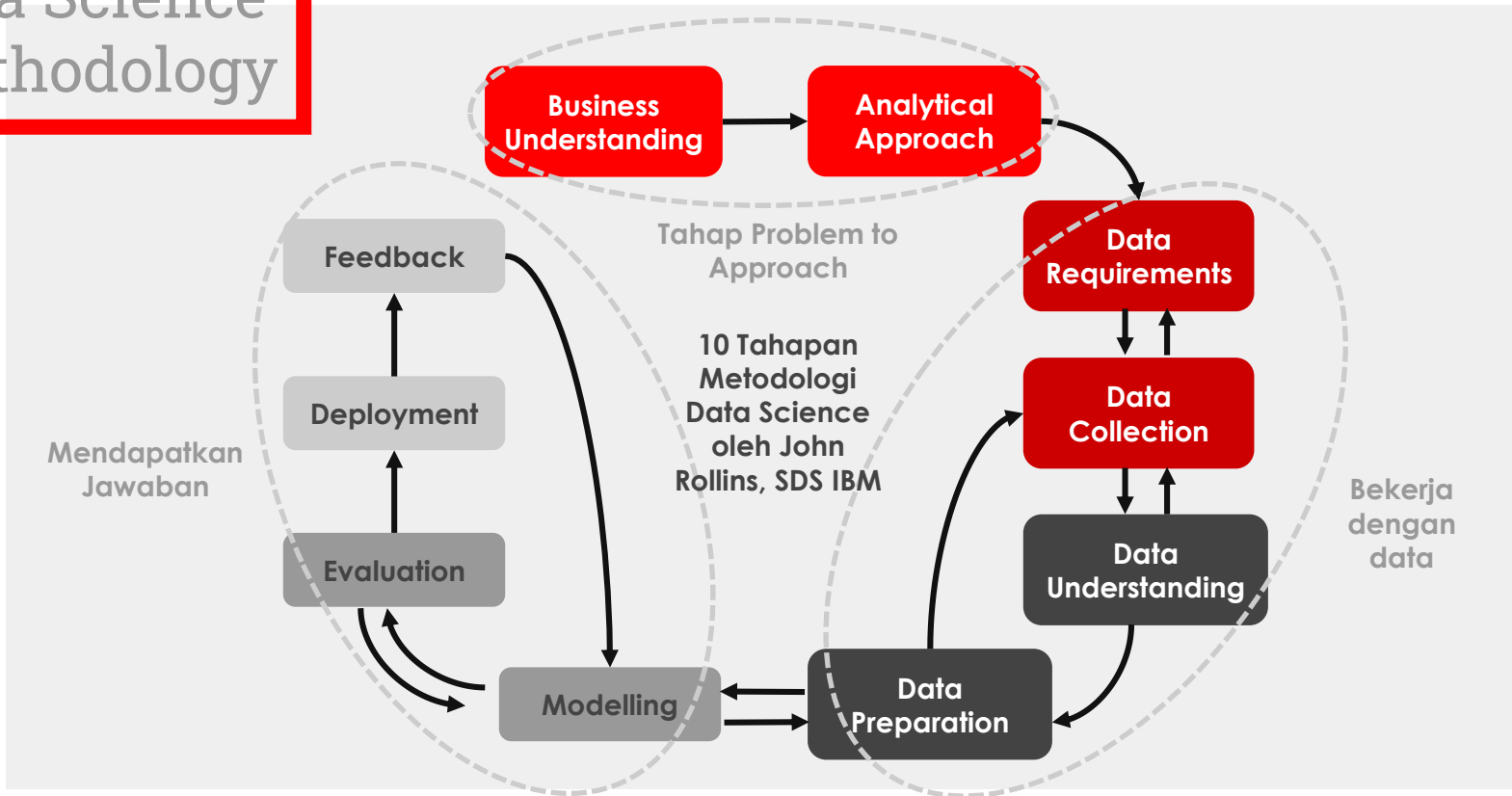
- Input data dari model prediksi
- Mengkombinasikan/membandingkan data-data model dengan aturan tertentu
- Menyediakan suatu solusi optimasi berdasarkan pertimbangan-pertimbangan tertentu
- Mengimplementasikan keputusan yang lebih baik



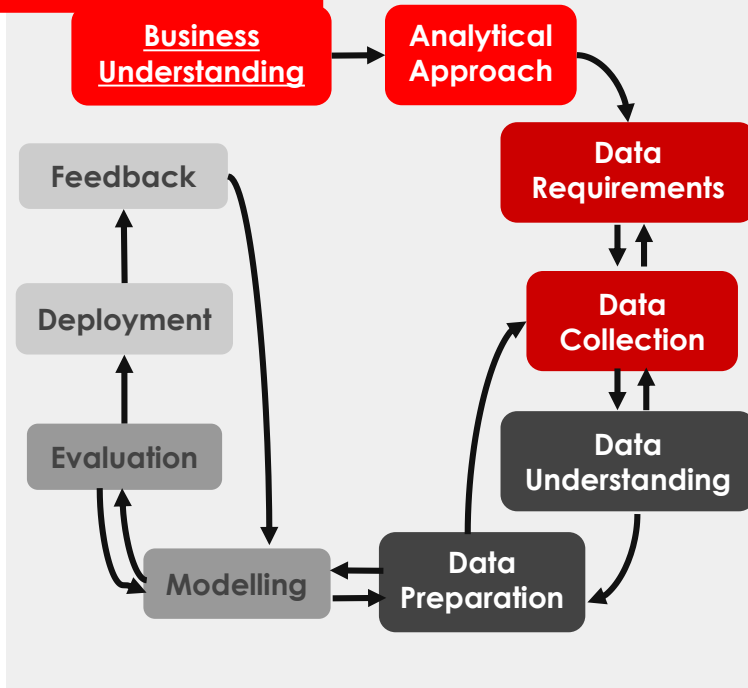
## 2. Data Science Methodology

Untuk memastikan data yang digunakan relevan dalam pemecahan masalah

# Data Science Methodology



# Data Science Methodology



- **Data Science Methodology** adalah langkah-langkah yang digunakan dalam proyek data science agar dapat menghasilkan hasil optimal dalam rangka menjawab pertanyaan suatu masalah yang ingin di selesaikan.
- **Tahapan** dalam metode **Data Science** merupakan proses berulang, yang mana jika dalam suatu tahapan dirasa masih belum sesuai, bisa Kembali ke tahap sebelumnya.

# Data Science Methodology

## *Business Understanding*

Tahapan pertama yang harus dilakukan adalah business understanding. Setiap project harus dimulai dengan pemahaman bisnis yang baik agar permasalahan bisnis yang ditangani dapat diselesaikan secara sempurna. Tahap ini memainkan peranan penting, hal ini terkait dengan definisi yang jelas dan lugas mengenai masalah, tujuan, dan kebutuhan solusi dari perspektif bisnis.

## *Analytical Approach*

Tahap metodologi data science selanjutnya yang dilakukan data scientist adalah menentukan pendekatan analitis untuk menyelesaikan masalah tersebut. Dalam tahap ini, definisi masalah yang jelas dan lugas diaplikasikan dalam bentuk pertanyaan yang berpengaruh pada sejauh mana analisis itu perlu dilakukan. Hanya sekedar ingin mengetahui mengapa suatu permasalahan terjadi atau lebih jauh ingin melihat perkiraan di masa datang apa yang akan terjadi.

# Data Science Methodology

## *Data Requirements*

Tahap ketiga di metodologi data science adalah data requirements. Identifikasi data apa saja yang diperlukan, bagaimana mengumpulkan data tersebut, bagaimana mengolah data tersebut, dan bagaimana menyiapkan data tersebut agar sesuai dengan hasil yang diinginkan.

## *Data Collection*

Setelah menentukan data requirements, data scientist mulai mengidentifikasi dan mengumpulkan data yang relevan dengan domain masalah. Maka saat mengumpulkan data yang relevan, sangat mungkin bagi kita untuk merevisi kembali data requirement dan memutuskan apakah akan mengumpulkan data lebih banyak atau lebih sedikit data.

# Data Science Methodology

## *Data Understanding*

Dalam proses ini, kita mencoba mempelajari data dan melihat kualitas data yang kita miliki, apakah sudah cukup baik untuk proses modeling atau belum. Proses data understanding yang populer adalah dengan menggunakan statistik deskriptif dan teknik visualisasi. Teknik ini membantu data scientist memahami isi data, menilai kualitas data, dan menemukan insight awal dari data tersebut.

## *Data Preparation*

Tahap data preparation terdiri dari semua kegiatan atau proses yang dilakukan untuk membangun dataset yang akan digunakan dalam tahap pemodelan, termasuk membersihkan data, menggabungkan data, dan mengubah data menjadi variabel yang lebih berguna. Sebagian data scientist mengatakan tahap ini bisa menghabiskan sekitar 70% atau bahkan 80% dari keseluruhan proses dalam proyek data science.

# Data Science Methodology

## Modelling

Modeling atau pemodelan adalah tahap dalam metodologi data science dimana data scientist membuat model untuk menjawab permasalahan. Proses pemodelan untuk model prediktif menggunakan data training. Data training adalah kumpulan data dimana hasil labelnya telah diketahui. Data training bertindak sebagai pengukur apakah model tersebut perlu penyesuaian lagi atau tidak.

## Evaluation

Setelah proses pemodelan, tahap selanjutnya adalah mengevaluasi kualitas model dan mengujinya apakah dapat mengatasi permasalahan bisnis dengan tepat. Evaluasi model dapat memiliki dua fase yaitu, fase diagnostic measures dan fase statistical significance testing.

# Data Science Methodology

## Deployment

Setelah model yang dikembangkan dirasa telah menghasilkan output yang memuaskan dan telah mendapat persetujuan dari stakeholder, model tersebut siap di-deploy atau digunakan di lingkungan bisnis.

## Feedback

Setelah proses model deployment, perusahaan akan mendapatkan feedback atau umpan balik tentang kinerja model. Dengan menganalisis feedback memungkinkan data scientist memperbaiki model serta meningkatkan akurasi dan kegunaannya.



Special Thanks to :



Slide template by SlideCarnival