

Summary : Introduction to Data Science

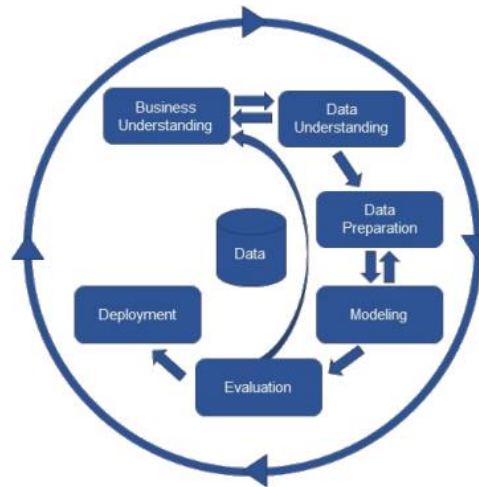
Nama : Diemas Aksya Fachriza

Kelompok : DS-1

A. History & Terminologies

- Decision Support System (1950)
Mendukung proses manajemen, operasi, dan perencanaan pada level organisasi (middle to top management) untuk mendukung penentuan keputusan, koordinasi, pengendalian, analisis, dan visualisasi informasi antar divisi dalam suatu organisasi
- Management Information System (1970)
Sistem informasi yang terintegrasi untuk mendukung penentuan keputusan, koordinasi, pengendalian, analisis, dan visualisasi informasi antar divisi di dalam suatu organisasi
- Business Intelligence (1989)
Terdiri dari strategi dan teknologi yang digunakan oleh enterprise untuk analisis data dari informasi bisnis. Menyediakan hasil analisis berdasarkan data historis dan pandangan terhadap operasi bisnis yang sedang berjalan.
- Data Science (2001)
Ilmu yang menggunakan metode sains, proses, algoritma, dan sistem untuk mendapatkan pengetahuan dan wawasan yang berasal dari data terstruktur dan data tidak terstruktur.

B. CRISP-DM



1. Business Understanding : Melakukan pemahaman bisnis terkait permasalahan yang ingin dipecahkan.
2. Data Understanding : Memahami bagaimana struktur data dapat menyelesaikan permasalahan bisnis.
3. Data preparation : Mempersiapkan data agar dapat diproses dan dianalisis pada tahap selanjutnya. Tahap ini relatif memerlukan waktu 80% dari seluruh tahap CRISP-DM
4. Modelling : Pembangunan model yang dapat memecahkan permasalahan bisnis.
5. Evaluation : Evaluasi terhadap model yang sudah dibangun untuk mengetahui keandalan modelnya.
6. Deployment : Model dipasang pada suatu entitas agar dapat digunakan oleh end-user.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives Background Business Objectives Business Success Criteria Assess Situation Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits Determine Data Mining Goals Data Mining Goals Data Mining Success Criteria Produce Project Plan Project Plan Initial Assessment of Tools and Techniques	Collect Initial Data Initial Data Collection Report Describe Data Data Description Report Explore Data Data Exploration Report Verify Data Quality Data Quality Report	Select Data Rationale for Inclusion/Exclusion Clean Data Data Cleaning Report Construct Data Derived Attributes Generated Records Integrate Data Merged Data Format Data Reformatted Data Dataset Dataset Description	Select Modeling Techniques Modeling Technique Modeling Assumptions Generate Test Design Test Design Build Model Parameter Settings Models Model Descriptions Assess Model Model Assessment Revised Parameter Settings	Evaluate Results Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models Review Process Review of Process Determine Next Steps List of Possible Actions Decision	Plan Deployment Deployment Plan Plan Monitoring and Maintenance Monitoring and Maintenance Plan Produce Final Report Final Report Final Presentation Review Project Experience Documentation

C. Types of Analytics

- **Descriptive**
Menjelaskan data, perhitungan Sum, Count, Average. Laporan berupa tabel, bar chart, pie chart, naratif
- **Diagnostic**
Menjawab mengapa suatu hal dapat terjadi. Melakukan drill down, data discovery, analisis kolerasi, menggabungkan chart.
- **Predictive**
Jenis teknik statistika yang berasal dari data mining, model predictive, dan machine learning yang dapat menganalisis data sekarang dan masa lampau untuk memprediksi kejadian di masa depan atau kejadian yang belum pernah terjadi.
- **Prescriptive**
Menghasilkan jawaban atas pertanyaan kenapa sesuatu dapat terjadi serta memberikan saran pada kondisi yang akan terjadi di masa yang akan datang.

D. Types of Data

- **Categorical**
 - **Binomial** : Dua kategori yang berupa True/False.
 - **Nominal** : Beberapa kategori tetapi tidak memiliki hierarki. Seperti warna, SIM, dan jenis kendaraan.
 - **Ordinal** : Beberapa kategori tetapi memiliki hierarki. Seperti tingkat pendidikan dan jabatan.
- **Numerical**
 - **Discrete** : Angka bulat. Seperti usia dan jumlah barang.
 - **Continuous** : Angka tidak bulat. Seperti suhu, tinggi badan, panjang, dan jumlah uang.

E. Data Science Solution Design

- Top Down

Dimulai dengan menemukan permasalahan awal pada suatu bisnis. Sehingga, dicari solusi dari permasalahan tersebut.

- Bottom Up

Dimulai dengan rasa ingin tahu terhadap sesuatu yang bermula dari data. Sehingga, dapat menemukan wawasan yang menarik berdasarkan hasil analisis.

