

<b>Mata Kuliah</b>	:	<b>ICT Literacy</b>
<b>Bobot Sks</b>	:	<b>2 sks</b>
<b>Dosen Pengembang</b>	:	<b>Cian Ramadhona Hassolthine, S.Kom.,M.Kom</b>
<b>Tutor</b>	:	
<b>Capaian Pembelajaran Mata Kuliah</b>	:	<b>Mahasiswa mampu memahami konsep Data Mining dan Big Data</b>
<b>Kompetensi Akhir Di Setiap Tahap (Sub-Cpmk)</b>		<b>Mahasiswa mampu memahami, Konsep Data Mining, Konsep Big Data dan Big Data Case Study</b>
<b>Minggu Perkuliahan Online Ke-</b>		<b>7</b>

### KONSEP DATA MINING

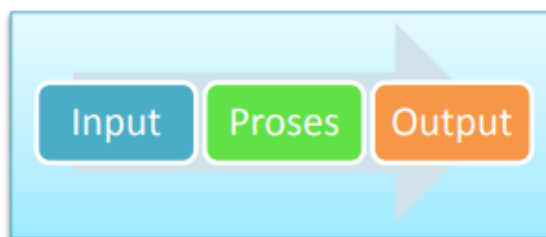
Secara umum Data Mining terbagi atas 2 kata yaitu:

1. **Data** yaitu Kumpulan Fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
2. **Mining** yaitu proses Penambangan Sehingga Data Mining itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah ouput (keluaran) berupa pengetahuan.

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic,database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.

Mengapa kita perlu memahami Data Mining? Karena manusia menghasilkan banyak sekali Data yang sangat besar baik dalam bidang Bisnis, Kedokteran, Cuaca, Olahraga, Politik dan sebagainya. Contohnya dalam Dunia Olahraga kita mengetahui Dari FIFA berapa banyak Lionel Messi Mencetak Gol selama semusim, berapa banyak Lionel Messi memberikan Asisst.

Pada Bidang Bisnis khususnya Saham, kita memperolehnya dari Bursa Efek Jakarta, kapan Harga Saham Naik maupun Turun. Pada Bidang Cuaca kita mengetahui data tentang Curah Hujan, Suhu, Kelembaban dan lain sebagainya. Kita mengetahui bahwa setiap proses terdiri dari 3 (tiga) fase yaitu:



Gambar 1. Siklus Penyelesaian Dari Input Ke Output

Dari gambar di atas bahwa mengetahui suatu hal itu dapat diselesaikan dimulai dengan sebuah Inputan (data) kemudian di Proses sehingga menghasilkan sebuah keluaran. Tentunya di dalam data mining juga mengalami fase tersebut. Yang membedakannya adalah pada Data Mining yang menjadi Input adalah Himpunan Data, Prosesnya adalah Algoritma atau metode dalam Data Mining itu sendiri, dan Keluarannya adalah berupa Pengetahuan dalam Input Proses Output Algoritma Data Mining Dan Pengujiannya bentuk Pola, Decision Tree, Cluster dan lain-lain. Untuk lebih jelas memahaminya berikut ini dapat dijelaskan pada gambar di bawah ini:

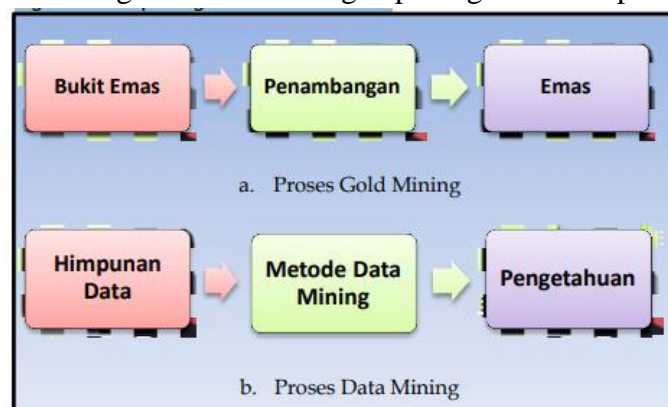


**Gambar 1.2 : Perbedaan Gold Mining dan Data Mining**

Keterangan gambar:

1. Gold Mining. Gambar Tersebut diatas menjelaskan tentang beberapa orang sedang mencari dan melakukan penambangan emas. Jadi dalam hal ini yang menjadi Input adalah Bukit Emas sedangkan proses yang K adalah Penambangan yang dimulai dengan Identifikasi titik pada Bukit Emas tersebut dimana Emas itu berada dan dilanjutkan dengan proses Penggalian sampai mendapatkan Emas sebagai Outputnya.
2. Data Mining: Gambar tersebut di atas menjeleskan sebuah Himpunan Data yang menjadi Input kemudian dilakukan pencarian pengetahuan menggunakan Metode Data Mining sehingga pada akhirnya didapatkan Pengetahuan sebagai Output.

Selain itu proses Gold Mining dan Data Mining dapat digambarkan pada gambar di bawah ini:



**Gambar 1.3 : Proses Gold Mining Dan Data Mining**

Pada proses Knowledge Discovery Database (KDD) terdapat beberapa fase yaitu sebagai berikut:

1. Seleksi Data (Selection) Seleksi (seleksi/pemilihan) data dari merupakan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam Knowledge Discovery Database (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional
2. Pemilihan Data (Preprocessing/Cleaning) Proses Preprocessing mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
3. Transformasi (Transformation) Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining
4. Data Mining Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.
5. Interpretasi/Evaluasi (Interpretation/Evaluation) Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses Data Mining Pola informasi.

## **BIG DATA**

Big data (mahadata) adalah kumpulan data yang super besar dengan perubahan yang begitu cepat. Di dalam mahadata, ada konsep 3V dengan perubahan yang sangat cepat. 3V adalah variety (variasi), volumes (jumlah), velocity (kecepatan). Big data adalah kumpulan data yang sangat besar (mahadata), kompleks, tidak terstruktur dan perubahannya sangat cepat. Contohnya:

- Iklan berdasarkan perilaku keseharian konsumen.
- Algoritma perankingan konten di mesin pencari.
- Smart transportations.
- Smart health care.
- Sistem pemerintahan terintegrasi.

Teknologi manajemen basis data tradisional sudah tidak mampu menangani data yang kompleks dan selalu berubah. Dan semua data itu terakumulasi setiap hari. Dengan begitu, perlu adanya teknik khusus dalam pengambilan data, analisis data hingga visualisasi data. Tujuannya adalah supaya mendapatkan hasil yang akurat. Maka muncul teknologi analisis mahadata. Saat semua serba online, pebisnis online akan mendapatkan data dalam jumlah besar dan perubahannya sangat cepat. Kedepannya, seorang pebisnis yang mahir menganalisis data, dia pasti bisa melihat peluang dengan cepat. Dengan begitu, dia dapat memilih strategi bisnis yang sesuai dengan peluang tersebut.

**Karakteristik Big data****1. Volume**

Sesuai dengan namanya, volume memiliki arti ukuran. Mahadata pasti memiliki sebuah ukuran data yang bukan lagi hitungan Megabytes atau Gigabytes, melainkan sudah Terabytes maupun Petabytes. Saya akan mengambil beberapa contoh ukuran big data di perusahaan.

Saat ini, Facebook mengambil sekitar 500 Terabytes data baru setiap hari dari pengguna. Lalu, sebuah Pesawat Boeing 737 menghasilkan sekitar 240 Terabytes data penerbangan dalam satu penerbangan melintasi Amerika. Bayangkan, seberapa pusing mengurus volume data sebesar itu setiap hari. Hehe.

**2. Velocity**

Pertumbuhan data dan popularitas social media telah mengubah pandangan kita dalam melihat data. Kini, orang-orang atau bahkan kamu seperti berlomba-lomba membalas postingan di social media. Ya, mungkin hanya sekedar untuk update status, atau merespon kejadian terkini.

Platform social media sudah mengaplikasikan big data dengan baik. Proses pengolahan data dari perilaku pengguna juga sangat cepat. Contohnya perusahaan Facebook, dia sudah dapat mengolah, menganalisa dan mengevaluasi data secara real-time.

**3. Variety**

Variety di sini lebih dimaksudkan sebagai berbagai jenis data yang tersedia. Misalnya database relasional, excel, csv akses atau file teks sederhana lainnya. Namun terkadang, kamu juga membutuhkan data yang tidak tersedia dalam format tradisional, seperti: video, sms, pdf.

Seiring berjalannya waktu ada dua V yang muncul namun jarang dijelaskan yaitu Value (nilai) dan Veracity (kejujuran). Value adalah nilai yang ada pada bigdata, sedangkan veracity merupakan definisi yang mengacu pada kualitas data yang didapatkan.

### **Studi Kasus Penerapan Big data**

Setelah kamu paham tentang konsep dasar dan berbagai jenis mahadata, saatnya melihat contoh secara langsung. Pada chapter ini, saya akan membahas contoh penerapan big data di beberapa bidang. Misalnya: pertanian, kehutanan, pemerintahan, kesehatan, media hingga analisis keuangan terkemuka di dunia.

#### **1. Bidang Pertanian dan Kehutanan**

Dalam bidang pertanian dan kehutanan sendiri ada banyak contoh penerapan big data.

Pertama, untuk mempelajari harga produk pertanian di suatu negara.

Kedua, untuk mempelajari rantai ekonomi produk pertanian.

Ketiga, sistem informasi legalitas kayu Perhutani.

Keempat, sistem informasi dini kebakaran hutan

Terakhir, untuk pembuatan sistem informasi cuaca.

#### **2. Bidang Pemerintahan**

Salah satu pemanfaatan big data di bidang pemerintahan yang cukup terkenal yaitu sistem CRVS (Civil Registration and Vital Statistics). Sistem ini dikenalkan oleh WHO sebagai pencatatan sipil yang mencakup data kelahiran, kematian, secara detail termasuk penyebab kematian, riwayat nikah dan perceraian.

#### **3. Bidang Kesehatan**

Salah satu contoh penerapan big data yang jelas terlihat oleh masyarakat luas adalah pusat informasi pandemi (misalnya, pusat informasi COVID-19). Lalu, contoh yang lain adalah pengambilan kebijakan preventif penyakit menular dengan database. Dan yang terakhir adalah sistem informasi rujukan rumah sakit.

#### **4. Bidang Media**

Di bidang media kamu pasti sangat familier dengan mesin pencari Google. Ya, Google telah menerapkan teknologi big data secara menyeluruh di mesin pencari miliknya. Mulai dari pengurutan ranking di mesin pencari, pengumpulan informasi koordinat lokasi di peta digital, hingga berita trending di YouTube dan Google News.