**Pengenalan Analisis Data**

Data merupakan sekumpulan fakta yang dapat direpresentasikan dalam berbagai bentuk seperti angka, gambar, video, teks, hasil pengukuran, dll. Sebagai generasi kekinian, kita telah memproduksi banyak data setiap harinya. Selain menghasilkan banyak data, tanpa sadar kita juga sering memanfaatkan data untuk mengambil sebuah keputusan.

Analisis data dapat diartikan sebagai sebuah proses mengumpulkan, mentransformasi, dan menata data untuk menarik kesimpulan, membuat prediksi, serta memberi pertimbangan yang tepat dalam mengambil keputusan. Orang yang melakukan proses analisis data ini sering disebut sebagai *data analyst*.

Untuk mendukung proses analisis data, seorang data analyst membutuhkan beberapa keterampilan seperti berikut.

* **Analytical Skills**  
  Keterampilan ini berhubungan dengan kualitas dan karakteristik yang berhubungan dengan penyelesaian masalah menggunakan fakta. Berikut merupakan beberapa poin penting dalam analytical skill.
  + Memiliki rasa ingin tahu yang tinggi untuk belajar hal baru serta mencari pengalaman dan tantangan yang baru.
  + Memiliki kemampuan untuk memahami konteks atau kondisi di balik suatu kejadian.
  + Memiliki pola pikir teknis sehingga mampu menguraikan suatu hal menjadi langkah-langkah yg lebih kecil dan mengerjakan tiap bagiannya dengan cara yang teratur dan logis.
  + Mampu mengolah atau mengatur informasi secara baik dengan menerapkan *data design*.
  + Mampu menerapkan *data strategy*untuk manajemen orang, proses, dan tools yang digunakan dalam analisis data.
* **Analytical Thinking**  
  Hal ini melibatkan pengidentifikasian dan pendefinisian sebuah masalah, kemudian menyelesaikannya berdasarkan data dengan cara yang terorganisasi dan berurutan. Secara umum, terdapat enam aspek kunci dari analytical thinking antara lain sebagai berikut.
  + Merepresentasikan sebuah informasi dalam bentuk grafik atau visual. Hal ini akan membantu seorang data analyst untuk memahami dan menyampaikan sebuah informasi secara efektif.
  + Selalu menerapkan pola pikir strategis untuk menjaga fokus dengan membuat perencanaan matang.
  + Menerapkan pola pikir yang berorientasi pada masalah untuk mengidentifikasi, mendeskripsikan, dan menyelesaikan sebuah *business problem*.
  + Mengidentifikasi korelasi antara dua data atau lebih.
  + Menerapkan *big-picture thinking* (pemikiran gambaran besar) dalam proses analisis.
  + Mampu menerapkan *detail-oriented thinking* (pemikiran yang berorientasi pada detail) dalam proses analisis.

**Pentingnya Data dalam Dunia Bisnis**

Di zaman yang serba canggih seperti saat ini, terdapat banyak inovasi dalam pemanfaatan teknologi dan informasi terutama pada bidang data. Data merupakan salah satu resource penting bagi perusahaan dari berbagai industri mulai dari *e-commerce*, *entertainment*, *manufacturing*,*healthcare*, *marketing*, *finance*, *tech*, dll.

Berbagai perusahaan tersebut membutuhkan data untuk mengidentifikasi peluang, tren, mengoptimalkan proses yang ada, meningkatkan layanan pelanggan, dan masih banyak lagi penerapannya di industri. Hal inilah yang mengakibatkan data dijuluki sebagai “*The new oil*” karena dianggap sebagai sumber daya paling penting saat ini.

Selain itu, *insigh*t (wawasan) yang diperoleh dari data dapat digunakan sebagai pendukung dalam pengambilan keputusan. Praktek seperti ini mendorong terbentuknya sebuah kultur baru dalam dunia bisnis yang disebut *data-driven decision making*.

Data-driven decision making merupakan sebuah kultur, *best practice*, serta proses dalam penggunaan data dan hasil analisis untuk memandu seluruh pengambilan keputusan bisnis yang tentunya sejalan dengan objektif perusahaan.

Umumnya perusahaan atau organisasi yang menerapkan kultur ini memiliki beberapa karakteristik seperti berikut.

* Memiliki data yang tersentralisasi dan terorganisir dengan baik.
* Memiliki kebijakan dan ketentuan terkait tata kelola data.
* Memiliki tingkat aksesibilitas data yang tinggi.
* Memiliki tool analisis yang mudah digunakan.

Berikut merupakan beberapa manfaat penerapan data-driven decision making.

* Manfaat utama dalam data-driven decision making ialah akuntabilitas dan konsistensi dalam setiap pengambilan keputusan. Hal ini akan mempermudah perusahaan dalam membuat sebuah *benchmark* yang jelas untuk mengevaluasi setiap keputusan dan strategi bisnis yang dibuat sehingga mampu mendorong perkembangan perusahaan secara kontinu.
* Pengambilan keputusan berdasarkan data akan mengarahkan perusahaan untuk berinovasi dan memperoleh peluang bisnis.
* Data-driven decision making mampu meningkatkan efisiensi dengan menyediakan berbagai informasi yang relevan dalam mengambil sebuah keputusan.

Terdapat beberapa poin yang harus diperhatikan ketika ingin menerapkan kultur data-driven decision making yaitu seperti berikut.

* Mengumpulkan data sebanyak mungkin
* Menentukan tujuan yang dapat diukur
* Memastikan Aksesibilitas Data
* Menentukan prioritas

**Mengenal Siklus Analisis Data**

Seperti yang telah kita bahas bersama, analisis data merupakan sebuah proses mengumpulkan, mentransformasi, dan menata data untuk menarik kesimpulan, membuat prediksi, serta memberi pertimbangan yang tepat dalam mengambil keputusan. Sebagai suatu proses, analisis data tentunya memiliki beberapa tahapan seperti berikut.

1. **Mendefinisikan pertanyaan**  
   Pada awal proyek analisis data, kita perlu mendefinisikan berbagai pertanyaan analisis terkait data ataupun permasalahan yang ingin diselesaikan. Pertanyaan tersebut akan membantu kita dalam memahami masalah bisnis yang dihadapi serta tujuan yang ingin dicapai. Bukan hanya itu, dengan adanya pertanyaan tersebut, kita akan merasa terbantu karena bisa lebih fokus pada bagian data yang relevan dengan tujuan.  
     
   Pada proses analisis data, terkadang kita diberikan data terlebih dahulu lalu mulai mendefinisikan pertanyaan analisis berdasarkan data tersebut. Namun, tidak jarang kita juga berangkat dari pertanyaan analisis terlebih dahulu lalu mencari data yang sesuai untuk menjawab pertanyaan tersebut.
2. **Data wrangling**  
   Setelah kita memiliki pertanyaan analisis yang baik, tahap selanjutnya adalah data wrangling. Pada tahap ini, Anda memulai dengan mengumpulkan data yang dibutuhkan untuk menjawab pertanyaan analisis yang telah dibuat sebelumnya. Kemudian Anda perlu menilai kualitas dan struktur dari data tersebut. Hal ini dilakukan untuk mengidentifikasi masalah yang terdapat dalam data serta membuat strategi pembersihan data yang sesuai.  
     
   Proses pembersihan data umumnya dilakukan dengan memodifikasi, mengganti, atau menghilangkan data yang bermasalah sehingga diperoleh dataset yang berkualitas dan terstruktur dengan baik.*FYI*, pembersihan data merupakan salah satu proses yang paling menghabiskan waktu dalam proyek analisis data. Oleh karena itu, kita sangat disarankan untuk menyediakan waktu yang cukup untuk melakukan pembersihan data.
3. **Exploratory Data Analysis**  
   Tahap berikutnya adalah data exploratory data analysis. Pada tahap inilah Anda akan melakukan eksplorasi terhadap data yang telah dibersihkan untuk memperoleh *insight* dan menjawab pertanyaan analisis.  
     
   Proses eksplorasi umumnya dilakukan dengan memanfaatkan berbagai teknik *descriptive statistics* (Penggunaan konsep statistik untuk mendeskripsikan data). Hal ini bertujuan untuk menemukan pola, hubungan, serta membangun intuisi terkait data yang diolah.  
     
   Selain menggunakan descriptive statistic, terkadang kita juga menggunakan teknik machine learning atau *inferential statistic* untuk memprediksi data di masa depan. Namun, kedua teknik tersebut di luar cakupan pembahasan kita pada kelas ini.
4. **Data visualization**  
   Hasil yang diperoleh dari exploratory data analysis perlu divisualisasikan melalui tahap data visualization. Pada tahap ini, kita akan menerapkan berbagai teknik dan jenis visualisasi data yang tepat untuk mengomunikasikan temuan atau insight dari hasil analisis secara efektif.
5. **Draw conclusion & communicate**  
   Pada akhir proyek analisis data, kita perlu membuat conclusion atau kesimpulan dari hasil analisis. Kesimpulan yang dibuat haruslah menjawab semua pertanyaan dan tujuan yang telah didefinisikan di awal.  
     
   Selain membuat kesimpulan, kita juga perlu membagikan temuan atau insight dari hasil analisis. Terdapat banyak cara dalam mengomunikasikan hasil analisis, seperti menggunakan report, slide deck, dan atau presentation.

Di industri, tahapan tersebut tidak selamanya berbentuk liner melainkan ada beberapa tahapan yang berulang. Hal inilah yang membuat tahapan dalam analisis data lebih cocok disebut sebagai sebuah siklus karena terdapat banyak iteratif proses di dalamnya.

**Pengenalan Tool Pengolahan Data**

Sebagai salah satu bahasa pemrograman kekinian, Python dapat digunakan untuk mengerjakan banyak tugas salah satunya adalah mengolah dan memvisualisasikan data.

Untuk mendukung proses pengolahan dan analisis data, Python memiliki banyak *libraries* dan *tools* yang bersifat *open-source* dan dapat digunakan secara bebas. Berikut merupakan tiga libraries yang paling sering digunakan untuk mengolah data.

* **Library NumPy**  
  NumPy (singkatan dari *numerical Python*) merupakan salah satu library Python yang sangat *powerful* untuk membuat dan mengolah *multi-dimensional arrays* (sering juga disebut sebagai matriks atau tensor). NumPy dibuat dengan tujuan untuk mempercepat proses komputasi numerik dalam bahasa pemrograman Python.
* **Library pandas**  
  Selain NumPy, library lain yang sering digunakan adalah pandas. Ia merupakan library Python yang spesifik digunakan untuk memanipulasi dan menganalisis data. Library inilah yang akan kita gunakan sebagai tool utama dalam menganalisis data.
* **Library SciPy**  
  Library penting selanjutnya adalah SciPy (singkatan dari Scientific Python). Ia merupakan library Python yang khusus digunakan untuk kebutuhan komputasi saintifik. Library ini memanfaatkan NumPy sebagai komponen dasar dalam setiap perhitungan komputasi yang akan dilakukan.

**Pengenalan Tool Visualisasi Data**

Pada proses analisis data, selain mengolah data, kita juga perlu merepresentasikan data tersebut ke dalam bentuk visual yang menarik dan mudah dipahami. Nah, pada materi kali ini, kita akan berkenalan dengan berbagai tools visualisasi data yang ada dalam bahasa pemrograman Python.

Sebagai salah satu bahasa pemrograman paling populer di kalangan praktisi data, Python memiliki banyak *libraries* dan *tools* yang bersifat*open-source* untuk mendukung pembuatan visualisasi data yang baik dan menarik. Berikut merupakan dua library Python yang paling sering digunakan oleh praktisi data di industri.

* **Library matplotlib**  
  Matplotlib merupakan sebuah library Python yang *reliable* dan komprehensif untuk mendukung kebutuhan pembuatan visualisasi data. Matplotlib menyediakan banyak *object* dan *method* yang memungkinkan kita membuat visualisasi data statis, *animated*, dan interaktif.
* **Library seaborn**  
  Selain matplotlib, library lain yang sering digunakan untuk visualisasi data adalah seaborn. Seaborn merupakan library Python yang spesifik digunakan untuk membuat visualisasi data yang atraktif dan informatif. Library ini memanfaatkan matplotlib untuk menampilkan grafik visualisasi data.