**Pengenalan Kelas**

Kelas ini terdiri dari **enam** modul yang akan fokus membahas dasar-dasar pada ilmu Dasar Data Science. Target kita adalah siswa yang memiliki ketertarikan di dunia data dan ingin menjadi seorang data scientist handal. Latihan menggunakan tools yang mudah dipelajari bagi pemula dan memiliki wawasan yang luas terkait dasar data science hingga memiliki portofolio data scientist di akhir pembelajaran.

Berikut adalah susunan materi yang akan Anda pelajari pada kelas ini.

* **Modul 1: The Power of Data**  
  Modul ini menunjukkan secara luas tentang data mulai dari Berkenalan dengan Data, kemudian menjelaskan terkait data, data, dan keputusan, hingga terdapat latihan terkait data untuk mengasah kemampuan setelah pemaparan materi.
* **Modul 2: Fundamental Data Science**  
  Modul ini mengenalkan ilmu dasar data science mulai dari Say Hi to Data Science yang berisi deskripsi dan fakta menarik tentang data science, kemudian terdapat trend data science di dunia, hubungan antara matematika dan data science, hingga hierarki data science.
* **Modul 3: Menjelajahi Analisis Data**  
  Setelah mempelajari fundamental data science, modul ini berisi penyusun data science yaitu analisis data. Modul ini menunjukkan cara menganalisis hingga memvisualisasikan sebuah data dengan tahapan-tahapan yang runut. Selain itu, terdapat latihan untuk mengasah kemampuan setelah pemaparan materi.
* **Modul 4: Teknologi dan Tools Data Science**  
  Modul ini mengeksplorasi teknologi dan tools yang digunakan oleh data science seperti SQL, NoSQL, Excel/Spreadsheet, SPSS, Tableau, hingga perbandingan antara Python dan R.
* **Modul 5: Machine Learning untuk Data Science**  
  Selain rumpun ilmu dalam dunia data, ilmu data science juga membutuhkan keahlian dalam teknologi di mana meliputi pembelajaran machine learning di dalamnya. Modul ini mengenalkan penggunaan machine learning di dunia data science mulai dari *scope machine learning*, kemudian alasan adanya machine learning di data science, hingga machine learning di mana-mana yang menunjukkan penerapan machine learning di dunia.
* **Modul 6: Mulailah Mencari Peluang: Eksplorasi Hingga Membuat Portofolio**  
  Modul ini menunjukkan peluang data science di dunia, mulai dari peluang karier di dunia data, kemudian tips & triks menjadi data scientist, hingga latihan membuat portofolio data scientist.

**Glosarium**

Berikut adalah glosarium dengan istilah umum yang digunakan pada kelas ini. Anda dapat membaca sekilas materi berikut untuk mengenali istilah-istilah umum yang ada di modul kelas ini. Selain itu, Anda juga dapat mengunjungi kembali halaman ini setiap kali menemukan istilah yang belum dimengerti. Carilah istilah tersebut pada halaman glosarium ini untuk mengidentifikasi makna atau definisinya. Jika masih terdapat kosakata yang tidak Anda pahami dan belum masuk di daftar ini, Anda dapat memberikan saran melalui fitur Laporan Materi.

**B**

**Basis Data**  
Kumpulan data yang diatur dan disimpan dengan cara yang memungkinkan akses dan pengambilan yang mudah.

**Basis Data Grafik**  
Dibuat khusus untuk menyimpan dan menavigasi hubungan.

**Basis Data Terdistribusi**  
Kumpulan data yang digunakan bersama-sama dan memiliki hubungan logis fisik tersebar di jaringan komputer.

**Big Data**  
Aset informasi bervolume tinggi (*high-volume*), berkecepatan tinggi (*high-velocity*), dan/atau memiliki banyak ragam (*high-variety*) yang menuntut bentuk pemrosesan informasi yang hemat biaya dan inovatif yang memungkinkan peningkatan wawasan, pengambilan keputusan, dan otomatisasi proses.

**C**

**CRISP-DM (Cross Industry Standard Process for Data Mining)**  
Proses model yang berfungsi sebagai dasar untuk proses data science.

**D**

**Data**  
Fakta atau pengamatan yang terpisah dan objektif, tidak terorganisir dan tidak diproses, serta tidak menyampaikan makna tertentu. Item data adalah deskripsi dasar dan tercatat dari hal-hal, peristiwa, kegiatan, dan transaksi.

**Dataset**  
Kumpulan data diatur dalam format tertentu.

**Data Analyst**  
Profesi pekerjaan yang bertugas untuk membaca serta menganalisis data yang ada dari perusahaan.

**Data-Driven Decision Making**  
Penggunaan fakta, metrik, dan data untuk memandu keputusan bisnis yang selaras dengan tujuan, sasaran, dan inisiatif perusahaan.

**Data Kategorikal**  
Data yang dapat dikelompokkan dan terbagi berdasarkan karakteristik atau ciri khasnya masing-masing.

**Data Kontinu**  
Data kontinu dapat direpresentasikan dalam berbagai nilai numerik, seperti bilangan desimal, bulat, dan lain-lain

**Data Kualitatif**  
Kumpulan data yang berbentuk kata, skema, dan gambar.

**Data Kuantitatif**  
Kumpulan data yang dinyatakan dalam bentuk angka.

**Data Nominal**  
Jenis pengelompokan data yang tidak memiliki keterkaitan dengan data lainnya dan tidak memiliki arti khusus.

**Data Numerik**  
Data berwujud angka yang bisa didapat dari sebuah pengukuran.

**Data Ordinal**  
Jenis pengelompokkan data yang memiliki urutan, atau harus disusun secara berurutan dengan mekanisme peringkat

**Data Science**  
Wawasan yang sangat berguna untuk kebutuhan bisnis. Ilmu ini merupakan penggabungan prinsip dan praktik dari bidang matematika, statistik, *artificial intelligence,*dan *computer engineering* untuk menganalisis data dalam jumlah besar.

**Data Scientist**  
Praktisi yang memiliki kemampuan dalam mengola dan menganalisis data dalam jumlah besar (big data).

**Data Terstruktur**  
Salah satu jenis data yang disusun dengan rapi dan diatur sedemikian rupa sehingga memiliki format atau bentuk yang tetap.

**Data Warehouse**  
Sebuah sistem atau platform yang dirancang untuk mengumpulkan, menyimpan, menganalisis, dan mengelola data dari berbagai sumber yang berbeda untuk tujuan analisis bisnis dan pengambilan keputusan.

**Data Wrangling**  
Proses atau kumpulan kegiatan yang meliputi mengumpulkan data (*gathering data*), menyiapkan data untuk dianalisis (*assessing data*), dan pembersihan data (*cleaning data*) sebelum data digunakan dalam proses analisis.

**Deep Learning**  
Metode dalam artificial intelligence yang mengajarkan komputer untuk memproses data layaknya otak manusia ketika berpikir.

**Derajat (Degree)**  
Jumlah atribut dalam sebuah relasi. Masih merujuk pada tabel di atas, berarti tabel tersebut memiliki derajat tiga karena terdiri dari tiga atribut.

**E**

**Evolusi**  
Perubahan yang terjadi secara lambat atau waktu yang panjang.

**G**

**GitHub**  
*Software* yang memungkinkan kontrol proyek melalui alat yang dikenal sebagai Git.

**Gradient Descent**  
Digunakan untuk menemukan minimal atau maksimal dari suatu fungsi.

**I**

**Informasi**  
Data yang memberi nilai tambah pada pemahaman suatu subjek. Adapun definisi lainnya bahwa data informasi adalah data yang telah dibentuk menjadi bentuk yang lebih berarti dan berguna bagi manusia.

**K**

**Kebijaksanaan**  
Akumulasi pengetahuan yang memungkinkan Anda memahami cara menerapkan konsep dari satu domain ke situasi atau masalah baru. Adapun pengertian lain yaitu kebijaksanaan adalah kemampuan untuk bertindak kritis atau praktis dalam situasi tertentu.

**Keterampilan Interpersonal**  
kemampuan yang dimiliki seseorang dalam melakukan komunikasi maupun interaksi dengan orang lain.

**L**

**Linear Regression**  
Merupakan algoritma supervised machine learning yang belajar dari kumpulan data berlabel dan memetakkan titik data ke fungsi linier yang paling optimal serta dapat digunakan untuk prediksi pada kumpulan data baru.

**LMS (Learning Management System)**  
Program perangkat lunak berbasis web untuk manajemen, dokumentasi, pemantauan, pelaporan, administrasi, dan distribusi konten pendidikan, program pelatihan, dan lain-lain.

**Logic Regression**  
Berfungsi untuk memprediksi kategori variabel dependen menggunakan seperangkat variabel independen yang diberikan.

**M**

**Machine Learning**  
Ilmu pengembangan algoritma yang memanfaatkan konsep matematis dan statistik dalam menjalankan tugas tertentu tanpa instruksi eksplisit.

**Multivariate calculus**  
M*ultivariate calculu*s adalah bidang kalkulus yang melibatkan banyak variabel.

**N**

**Nilai Korelasi**  
Nilai yang menunjukkan hubungan erat atau tidak antar dua variabel nilai.

**NoSQL**  
Pendekatan untuk desain basis data yang memungkinkan penyimpanan dan kueri data di luar struktur tradisional yang ditemukan dalam basis data relasional.

**O**

**Object-oriented Model**  
Data disimpan dalam bentuk objek.

**On-premise**  
Jenis server berupa *software*yang dijalankan secara internal oleh tim IT perusahaan.

**P**

**Pengetahuan**  
Kombinasi dari data dan informasi yang ditambahkan pendapat ahli, keterampilan, dan pengalaman sehingga menghasilkan aset berharga yang dapat digunakan untuk membantu pengambilan keputusan. Pengetahuan adalah data atau informasi yang telah diatur dan diproses untuk menyampaikan pemahaman, pengalaman, akumulasi pembelajaran, dan keahlian.

**Portofolio**  
Rangkuman atau kumpulan hasil karya Anda baik berupa visual, dokumen, kode, dan sebagainya yang dapat dilihat oleh siapapun.

**Prediksi karier**  
Analisis dan studi yang tepat dari setiap catatan siswa akan membantu memahami kemajuan, kekuatan, kelemahan, minat setiap siswa, dan banyak lagi.

**R**

**Reinforcement Learning**  
Memahami dan menginterpretasikan lingkungannya, mengambil tindakan, dan belajar melalui coba-coba.

**S**

**Sistem Penilaian**  
Kemajuan baru dalam sistem penilaian telah diperkenalkan sebagai hasil dari analisis data siswa yang tepat.

**SPSS (Statistical Package for the Social Sciences)**  
Program komputer yang dipakai untuk analisis statistika. SPSS digunakan untuk pengolahan dan menganalisis data yang memiliki kemampuan analisis statistik serta sistem manajemen data dengan lingkungan grafis.

**Structured Query Language**  
Bahasa pemrograman untuk menyimpan dan memproses informasi dalam database relational.

**Supervised Learning**  
Pendekatan *machine learning*yang menggunakan kumpulan data berlabel (*labeled datasets*).

**T**

**Turunan**  
Laju perubahan fungsi terhadap variabel.

**U**

**Unsupervised Learning**  
Pada prosesnya model akan belajar untuk mengidentifikasi pola tersembunyi yang terdapat dalam data.

**V**

**Variabel**  
Karakteristik, angka, atau kuantitas apa pun yang dapat diukur atau dihitung.

**Variety**  
Variasi tipe dan variasi sifat dari data, apakah data tersebut bersifat terstruktur, semi terstruktur, ataupun tidak terstruktur.

**Velocity**  
Kecepatan dalam men-*generate* data, mengakses data serta memproses data.

**Volume**  
Jumlah data yang dihasilkan dari banyak transaksi serta volume data yang disimpan. Contohnya, seperti penggunaan *history browser*, pencatatan transaksi pada *e-commerce,*data ktp atau data penduduk Indonesia, data pelanggan pada perbankan dan masih banyak lagi.

[1] "What is Data Science?" https://www.ibm.com/topics/data-science [Website]

[2] Maryanto, B (2019). "BIG DATA DAN PEMANFAATANNYA DALAM BERBAGAI SEKTOR". Media Informatika Vol.16 No.2 [Journal]

[3] Nield, Thomas. "Essential Math for Data Science". O'Reilly [E-Book]

[4] Putri, H.N, Saputro D.R (2022). “Clustering Data Campuran Numerik dan Kategorik Menggunakan Algoritme Ensemble Quick RObust Clustering using LinKs (QROCK)”. PRISMA [Journal]

[5] The Council on Quality and Leadership "12 REASONS why data is important" [E-Book]