Requirement Data Science dalam konteks matematika:

* Algoritma, eksponensial, fungsi polinomial, dan bilangan relasional
* Geometri dasar dan teorema, identitas trigonometri
* Bilangan riil dan kompleks
* Deret, penjumlahan, pertidaksamaan
* Grafik dan plotting, koordinat kartesius dan polar, serta irisan kerucut.

**Variabel**

Menurut*Australian Bureau of Statistics*, variabel adalah karakteristik, angka, atau kuantitas apa pun yang dapat diukur atau dihitung. Variabel juga dapat disebut item data. Usia, jenis kelamin, pendapatan dan pengeluaran bisnis, negara kelahiran merupakan contoh variabel dan masih banyak lagi contoh-contoh lainnya.

ada tiga jenis variabel di bidang data science

1. **Numerikal**
2. **Kategorikal**
3. **DateTime**

Kategori variabel ini berkaitan dengan aspek tanggal & waktu. Kategori ini dapat berisi jenis nilai seperti berikut.

Hanya memiliki tanggal.

Hanya memiliki waktu.

Memiliki keduanya.

**Fungsi**

Ekspresi yang mendefinisikan hubungan antara dua atau lebih variabel. Sebagian besar data science difokuskan untuk menemukan hubungan antara variabel dependen dan independen. Sederhananya, variabel independen adalah variabel bebas (x) dan variabel dependen adalah variabel terikat (y) yang dipengaruhi oleh variabel bebas.

**y = 2x + 1**

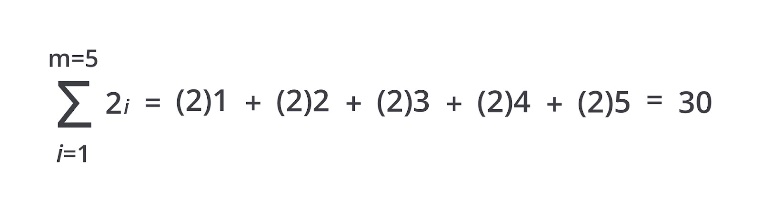
| **Nilai (x)** | **Fungsi (2x + 1)** | **Hasil (y)** | **[x,y]** |
| --- | --- | --- | --- |
| 0 | 2(0) + 1 | y = 1 | [0,1] |
| 1 | 2(1) + 1 | y = 3 | [1,3] |
| 2 | 2(2) + 1 | y = 5 | [2,5] |
| 3 | 2(3) + 1 | y = 7 | [3,7] |

Tabel di atas terbagi menjadi tiga kolom yaitu kolom nilai x yang merupakan variabel bebas (X), kemudian terdapat kolom fungsi yang menjadi soal yang ditanyakan, dan terdapat kolom hasil yang merupakan variabel terikat (Y). variabel Y akan berubah seiring dengan adanya nilai dari variabel X. Sebelum membaca tabel di atas, perlu *disclaimer*bahwa nilai x merupakan sebuah pengandaian, sehingga apabila Anda tidak memasukkan angka dengan nilai tersebut bukan sebuah kesalahan. Tabel di atas dapat dibaca seperti berikut.

* Apabila nilai x = 0 dan dimasukkan ke dalam fungsi, maka nilai y adalah 1.
* Apabila nilai x = 1 dan dimasukkan ke dalam fungsi, maka nilai y adalah 3.
* Apabila nilai x = 2 dan dimasukkan ke dalam fungsi, maka nilai y adalah 5.
* Apabila nilai x = 3 dan dimasukkan ke dalam fungsi, maka nilai y adalah 7.

**Penjumlahan**

Penjumlahan umumnya dinyatakan sebagai sigma (). Nah, agar Anda dapat lebih memahami terkait pembahasan penjumlahan, perhatikan contoh di bawah ini.



Persamaan di atas dapat dibaca sebagai berikut.

i=1 menunjukkan batas bawah (lower limit) atau angka pertama yang akan digunakan; m=5 menunjukkan batas atas (upper limit) atau batas angka paling besar; dan *2i*merupakan fungsi yang akan dihitung.

**Kalkulus**

Sebagian besar bidang data science tentunya memerlukan pemahaman tentang prinsip dasar kalkulus dan pengaruhnya terhadap model *machine learning.*Namun, perlu digaris bawahi bahwa kalkulus pada data science tidak seperti kalkulus pada kelas matematika di sekolah atau perguruan tinggi. Umumnya beberapa konsep kalkulus yang dapat digunakan data scientist adalah *Multivariate calculus*(Kalkulus multivariabel) dan Gradient descent (penurunan Gradien).

1. **Multivariate calculus**

*Multivariate calculus*adalah bidang kalkulus yang melibatkan banyak variabel. Jika output dari fungsi Anda z, bergantung pada satu variabel input (x), maka Anda dapat mendeklarasikannya sebagai berikut.

|  |
| --- |
| z = f(x) |

Selanjutnya, jika output dari fungsi Anda z bergantung pada beberapa input (x dan y), maka Anda dapat mendeklarasikannya sebagai berikut.

|  |
| --- |
| z = f(x,y) |

Variabel-variabel tersebut (x dan y) adalah input dari fungsi, oleh karena itu dapat memengaruhi hasil output.

Berikut merupakan fungsi *Multivariate calculus*pada *machine learning*.

* Dalam algoritma *support vector*(*Support vector algorithm*), *multivariate calculus*digunakan untuk menemukan margin maksimal.
* Masalah pengoptimalan (*optimization problems*) bergantung pada *multivariate calculus.*

1. **Gradient descent**

*Gradient descent* digunakan untuk menemukan minimal atau maksimal dari suatu fungsi. Fungsi tersebut bisa menjadi *cost function*dari algoritma *machine learning.* Prinsip dari *gradient descent*apabila dibuat analoginya adalah Anda seperti melemparkan bola dari permukaan ke atas (kurva naik atau titik tertinggi) maka bola itu akan kembali lagi ke bawah (kurva turun atau titik minimum).

**Turunan**

Dalam matematika, derivatif atau turunan adalah laju perubahan fungsi terhadap variabel. Turunan berasal dari garis kemiringan yang bersinggungan dengan grafik fungsi.

Untuk mengetahui lebih lanjut penerapan fungsi untuk menjadi sebuah grafik, mari kita mulai dengan contoh fungsi sederhana. Seperti sebelumnya, perlu *disclaimer*bahwa nilai x merupakan sebuah pengandaian, sehingga apabila Anda tidak memasukkan angka dengan nilai tersebut bukan sebuah kesalahan.

**f(x) = x2**

Seberapa curamkah grafik yang digambarkan apabila x = 2? Mari hitung terlebih dahulu menggunakan tabel agar Anda lebih paham.

| **Nilai (x)** | **Fungsi = x2** | **Hasil (f(x))** | **[x,y]** |
| --- | --- | --- | --- |
| -2.1 | f(-2.1) = -2.12 | f(x) = 4.41 | [(-2.1),4.41] |
| -2 | f(-2) = -22 | f(x) = 4 | [-2,4] |
| -1 | f(-1) = -12 | 1 | [-1,1] |
| 0 | f(0) = 02 | 0 | [0,0] |
| 1 | f(1) = 12 | 1 | [1,1] |
| 2 | f(2) = 22 | f(x) = 4 | [2,4] |
| 2.1 | f(2.1) = 2.12 | f(x) = 4.41 | [2.1,4.41] |

Pada dasarnya cara membaca tabel di atas sama seperti contoh pada kasus sebelumnya, yaitu

* Apabila nilai x = -1 dan dimasukkan ke dalam fungsi, maka nilai y adalah 1.
* Apabila nilai x = 0 dan dimasukkan ke dalam fungsi, maka nilai y adalah 0.
* Apabila nilai x = 1 dan dimasukkan ke dalam fungsi, maka nilai y adalah 1.
* dst.