Learning Objective

Dalam kursus ini diharapkan:

A. Peserta mampu Melakukan pelabelan data

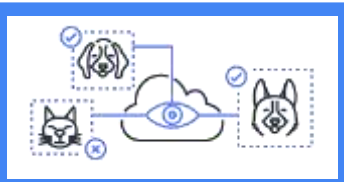
B. Peserta mampu membuat dokumentasi dan laporan pelabelan data

Pengertian *Labelled Data* (data yang dilengkapi label/target)

- Label / target / variable dependent adalah attribute/kolom/field yang menjadi sasaran/target untuk diprediksi. Disebut variable dependent, karena nilai dari attribute ini tergantung dari nilai atribut-atribut yang lain. Label/target biasanya disimbolkan dengan huruf *y,* yang merupakan fungsi dari atribut yang lain (biasanya x). Jadi persamaan y merupakan fungsi dari x, atau y = f(x).

- Seperti namanya, data berlabel (alias data beranotasi) adalah data yang sudah mengandung label yang bermakna, tag, atau kelas. Contoh, misalnya kita membangun sistem pengenalan gambar dan telah mengumpulkan beberapa ribu foto. Penetapan label akan memandu mesin bahwa foto-foto itu berisi *'orang*', *'pohon*', *'mobil*', dan sebagainya.

Contoh Beberapa Label yang umum

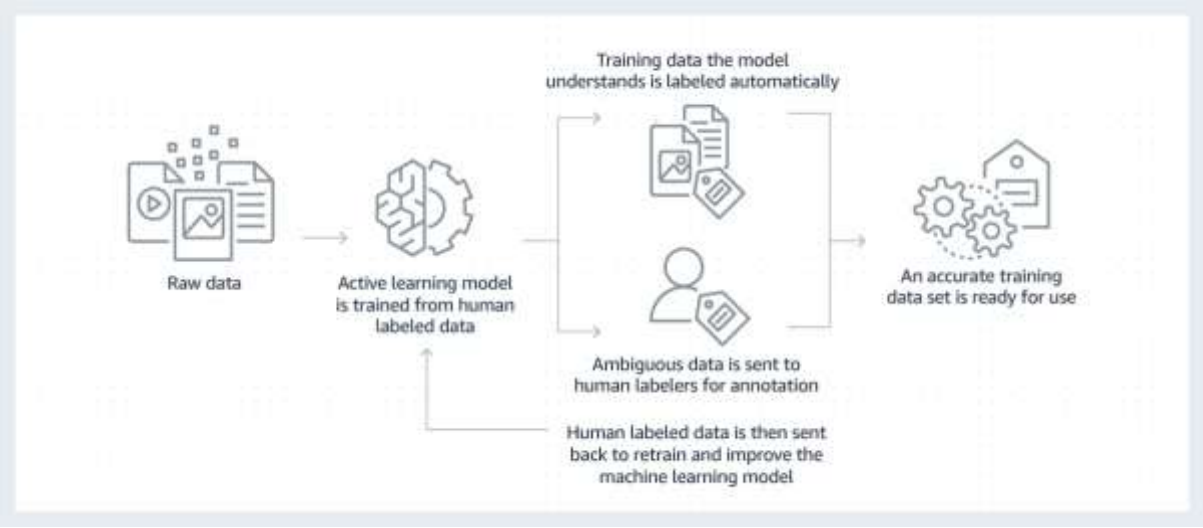
	***Computer Vision:*** • **Label pada gambar, piksel, atau *key point*, batas gambar digital.** • **Klasifikasi: gambar produk vs. gaya hidup; objek wajah vs non wajah, objek hewan vs non hewan**
	***Pemrosesan Bahasa Alami*** • sentimen atau makna uraian teks, • Identifikasi bagian ucapan, • klasifikasikan kata benda • Identifikasi teks, gambar, PDF, atau file lainnya.
	***Pemrosesan Audio*** • Mengubah ucapan ke dalam format terstruktur sehingga dapat digunakan dalam pembelajaran mesin.

Praktik terbaik (best practice) pelabelan data

- **Interface tugas yang intuitif dan efisien.**
- **Konsensus pemberi label**
- **Audit label**
- **Pembelajaran aktif**

Teknik Pelabelan Data Menggunakan Pembelajaran Mesin

Pelabelan dapat dibuat lebih efisien dengan menggunakan model pembelajaran mesin untuk melabeli data secara otomatis



Pendekatan pelabelan data

- **Inhouse Labeling:** Pelabelan dilakukan secara internal pengguna data latih
- **Crowdsourcing** : Penggunaan mekanisme Platform Crowdsourcing
- **Outsourcing ke individu :**Menggunakan pekerja lepas di berbagai situs web rekrutmen, pekerja lepas, dan jejaring sosial
- **Outsourcing ke perusahaan :**Penggunaan jasa perusahaan outsourcing yang mengkhususkan diri dalam persiapan data pelatihan
- **Pelabelan sintetis :**Data sintetis dihasilkan oleh model generatif yang dilatih dan divalidasi pada dataset asli
- **Pemrograman data :** Penulisan fungsi pelabelan — skrip yang secara terprogram melabeli data.

Alat (tools) Pelabelan Data

- **Annotorious :** alat anotasi dan pelabelan gambar web gratis berlisensi MIT

- **LabelMe** : Perangkat lunak online dan bersifat terbuka yang membantu pengguna dalam membangun basis data gambar.

- **Sloth** : Perangkat lunak gratis dengan tingkat fleksibilitas tinggi yang memungkinkan pengguna untuk memberi label file gambar dan video untuk penelitian Computer Vision.

Referensi

- https://www.analyticsvidhya.com/blog/2021/05/detecting-and-treating-outliers-treating-the-odd-one-out/
- https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/#:~:text=In%20machine%20learning%2C%20data%20labeling,model%20can%20learn%20from%20it.
- https://labelyourdata.com/articles/introduction-to-labeled-data-what-why-and-how#what_is_labeled_data
- https://www.altexsoft.com/blog/datascience/how-to-organize-data-labeling-for-machine-learning-approaches-and-tools/

KOMDIGI
Kementerian Komunikasi dan Digital
Republik Indonesia

digitalent

Terima Kasih