COMP6576 - Natural Language Processing

**Task 1**

# Cause Virtual Run

Cause Virtual Run adalah sebuah platform penyelenggara acara virtual run. Virtual run adalah sebuah acara lari yang bisa diikuti kapanpun, dimanapun, dan oleh siapapun. Dalam virtual run, peserta bisa berlari sambil mengaktifkan aplikasi lari berbasis GPS untuk merekam aktivitas lari mereka dan submit hasilnya ke platform kami.

Kemudian, tim kami akan melakukan validasi hasil lari untuk menentukan apakah hasil yang di-submit valid dan sudah memenuhi kriteria untuk finish. Jika sudah tervalidasi, pada akhir periode acara, medali atau hadiah finisher lainnya akan langsung dikirimkan ke alamat pendaftar.

# The Problem

Saat ini Cause sudah terintegrasi dengan Garmin dan Strava. Artinya, para peserta yang menggunakan jam tangan olahraga Garmin atau aplikasi Strava tidak perlu lagi melakukan submission setiap selesai berlari, karena aktivitas yang dilakukan akan langsung sync secara otomatis ke platform kami.

Namun demikian, tidak bisa dipungkiri bahwa setiap harinya kami masih menerima submission hasil lari berbentuk screenshot dari aplikasi lari berbasis GPS seperti Nike Run Club (NRC), Endomondo, Adidas Running, Mi Fit, dan lain-lain yang tidak memiliki integrasi API.

Hal ini menjadi sebuah kendala bagi Cause karena harus melakukan validasi input dari user, dan juga bagi para pengguna karena harus melakukan input hasil lari secara manual dengan mencantumkan screenshot sebagai bukti.

Jika diperhatikan, hasil lari yang disubmit oleh user tidak selalu seragam, namun tentunya screenshot hasil lari dari setiap running application memiliki karakteristik dan pattern tersendiri. Oleh karena itu, kami percaya bahwa problem ini bisa di-solve menggunakan Optical Character Recognition (OCR) dan Natural Language Processing (NLP).

Salah satu library yang saat ini digunakan untuk eksperimen OCR adalah Google Vision. Mahasiswa/i bisa mencoba implementasi dasar Google Vision di Cause pada link ini [https://cause.id/tests/vision?img=<url\_gambar](https://cause.id/tests/vision?img=%3curl_gambar)>

Sample hasil lari yang di-submit oleh peserta bisa diakses di <https://cause.id/tests/runresultsamples>

# The Task

Sebelum mahasiswa/i memulai menjawab pertanyaan-pertanyaan di bawah ini, mohon untuk mencoba platform Cause baik melalui website ([https://cause.id](https://cause.id/)) ataupun aplikasi Cause Virtual Run (<https://bit.ly/causeid>) agar mendapatkan gambaran proses yang dilalui oleh user saat menggunakan platform Cause.

Berikut adalah beberapa pertanyaan yang kami harap bisa dijawab oleh para mahasiswa/i yang sudah mempelajari sedikit banyak tentang NLP:

* Non-Technical Questions
  1. Apa saja data field yang bisa di-ekstrak dari setiap jenis aplikasi? Field apa yang sama, dan field apa yang hanya ada di aplikasi tertentu?
  2. Bagaimana cara menentukan jenis aplikasi berdasarkan field yang ada di sebuah gambar? Apa faktor unik yang membedakan aplikasi NRC, Endomondo, dan Adidas Running, misalnya?
* Technical Questions

1. Apa algoritma atau library OCR yang paling cocok untuk task ini? Jelaskan selengkapnya bagaimana Cause bisa menggunakan algoritma atau library ini untuk menyelesaikan permasalahan yang diberikan.
2. Apa algoritma atau library NLP yang paling cocok untuk task ini? Jelaskan selengkapnya bagaimana Cause bisa menggunakan algoritma atau library ini untuk menyelesaikan permasalahan yang diberikan.

* Exploratory Questions

1. Selain untuk melakukan otomasi pada user input, kira-kira untuk hal apa lagi Cause bisa memanfaatkan running results ini dan bagaimana teknologi NLP bisa membantu?
2. Selain untuk analisa running results, bagian lain apa dari aplikasi Cause yang bisa dikembangkan dengan teknologi NLP?