

Masalah Umum

1. Data sekuens genetik (DNA/RNA) semakin banyak.
2. Berbagai model deep learning digunakan untuk menghasilkan representasi data genetik yang hanya bisa dimanfaatkan untuk tujuan (task) tertentu saja

Referensi

1. Data genetik dapat dipandang sebagai teks yang berisi informasi tentang makhluk hidup. (Luchi et. al., 2021).
2. Arsitektur BERT dapat menghasilkan representasi semantik yang independen terhadap *task* pemrosesan bahasa secara khusus (Devlin et. al., 2018).
3. DNABert (Ji et. al., 2021).

Masalah Khusus

1. Tidak ada representasi generik dari data sekuens genetik yang dapat digunakan untuk berbagai jenis dataset.

Judul

Penerapan Representasi Kontekstual untuk Pemrosesan Data Genetik

Hal-Hal Pendukung

1. Library atau perangkat lunak telah tersedia.
2. Data genetik mudah diakses dan gratis.

Tujuan

1. Menerapkan model BERT untuk menghasilkan representasi data genetik yang dapat digunakan untuk berbagai task.

Manfaat

1. Pemrosesan data genetik dapat semakin efisien
2. Data genetik yang mirip boleh jadi dapat memanfaatkan representasi yang sama atau memodifikasi representasi yang sudah ada.