

## **Alternatif Topik Tesis: Penerapan Representasi Kontekstual pada Pemrosesan Data Genetik**

### **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.

### **Masalah Khusus**

1. Dibutuhkan suatu representasi semantik dari data genetik yang dapat digunakan pada berbagai tujuan.

### **Referensi**

1. Data genetik dapat dipandang sebagai sebuah teks dengan bahasa tertentu yang berisi informasi tentang makhluk hidup. (Iuchi et. al., 2021).
2. Arsitektur BERT dapat menghasilkan representasi semantik dari sebuah teks yang bersifat independen terhadap *task* pemrosesan bahasa secara khusus (Devlin et. al., 2018).
3. DNABert dikembangkan berdasarkan arsitektur BERT untuk memproses data genetik untuk memecahkan berbagai masalah penelitian seperti identifikasi protein dan klasifikasi gen (Ji et. al., 2021).

### **Hal-Hal Pendukung**

1. *Library* atau perangkat lunak untuk membuat arsitektur *deep learning* dapat diakses dengan mudah dan terdapat banyak tutorial cara penggunaan perangkat lunak tersebut.
2. Data genetik mudah diakses dan gratis.

### **Hipotesis**

1. BERT dapat digunakan untuk menghasilkan model semantik genetik yang dapat digunakan untuk berbagai *task* genetik.

### **Tujuan**

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

### **Manfaat**

1. Penelitian-penelitian *deep learning* yang menggunakan data sama dapat memanfaatkan representasi yang sama sehingga dapat meminimalisasi waktu yang digunakan untuk proses *training* model dari awal.
2. Data genetik yang mirip boleh jadi dapat memanfaatkan representasi yang sama atau memodifikasi representasi yang sudah ada.