**Tugas Sequential Labeling**

Tugas ini merupakan tugas yang bertujuan untuk mendapatkan pemahaman yang komprehensif dalam menyelesaikan task NLP berbasis sequential labeling.

Ketentuan Kelompok:

1. Bentuklah kelompok yang terdiri dari 3 orang
2. Kelompok disarankan heterogen. Jika ada mahasiswa fast track diusahakan tidak satu kelompok. Begitu pula dengan mahasiswa S2 dan S3. Sebaiknya bercampur kelompoknya.
3. Silahkan mengisi daftar anggota kelompok pada tautan berikut:   
   [Daftar Kelompok Sequential Labeling](https://docs.google.com/spreadsheets/d/1HrxBzBCG27gIfVFJlNs6jIAw_ZEcXLhRI6f6CXWi1Qg/edit?usp=sharing)

Ketentuan Tugas:

1. Selesaikanlah beberapa task NLP yang berbasis sequential labeling yaitu NER, POS Tag, dan Keyword Extraction
2. Dataset tersedia pada url berikut:
   1. NER <https://github.com/indobenchmark/indonlu/tree/master/dataset/nerp_ner-prosa>
   2. POS Tag  
      <https://github.com/indobenchmark/indonlu/tree/master/dataset/posp_pos-prosa>
   3. Keyword Extraction

<https://github.com/indobenchmark/indonlu/tree/master/dataset/keps_keyword-extraction-prosa>

1. Masing masing anggota kelompok mengerjakan satu task diatas. Sehingga 3 orang akan mengerjakan task yang berbeda beda.
2. Task yang dikerjakan oleh setiap orang harus diselesaikan dengan menggunakan 3 pendekatan
   1. Traditional ML  
      Misalnya: melakukan ekstraksi fitur menggunakan tf-idf, kemudian menggunakan algoritma decision tree learning untuk modeling
   2. Non Contextual Embeddings  
      Misalnya: menggunakan glove / word2vec / fasttext untuk ekstraksi fitur, kemudian menggunakan RNN, CNN, dan atau FCN untuk modeling
   3. Contextual Embeddings  
      Misalnya: fine-tuning BERT, GPT-3, dll.

Untuk fine tuning BERT, bisa menggunakan trainer class dari transformer di huggingface (https://huggingface.co/docs/transformers/training)

1. Setiap task yang dikerjakan wajib submit pada laman kompetisi indoNLU <https://competitions.codalab.org/competitions/26537#learn_the_details-how-to-submit> . Harap diperhatikan maksimum submisi yang dapat dilakukan dalam satu hari adalah 5 buah submisi.
2. Segera lakukan permintaan untuk join kompetisi tersebut agar admin lebih mudah untuk memberikan persetujuan permintaan bergabung kompetisi
3. Gunakan nama team yang sama dengan nama kelompok tugas ini dan gunakan username yang representatif sesuai nama anggota tim
4. Pengumpulan tugas dilakukan melalui edunex. Paling lambat pada hari Senin, 28 Maret 2021 pukul 09.00 WIB.Deliverables yang dikumpulkan:
   1. Kode program
   2. Laporan singkat yang menjelaskan teknik yang digunakan dan screenshot pada leaderboard pada kompetisi codalab
5. Jika terdapat pertanyaan dapat disampaikan pada grup whatsapp IF6082 NLP Genap 2021 dengan tag asisten (Fajar)