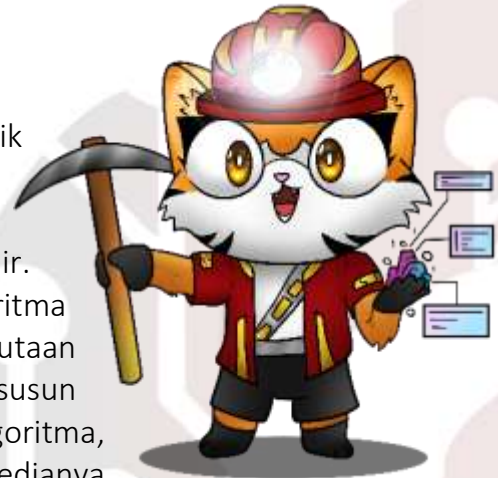


C. PENAMBANGAN DATA

1. Deskripsi

Penambangan Data yang semakin banyak menarik perhatian baik di kalangan industri maupun akademik dunia dewasa ini, terutama dengan penggunaan *deep learning* yang merupakan suatu algoritma mutakhir. Keunggulan *deep learning* dibandingkan dengan algoritma konvensional adalah kemampuan untuk melatih jutaan bahkan milyaran parameter pembelajaran yang tersusun secara hierarkis multi lapis. Selain berkat terobosan algoritma, *deep learning* juga telah dimungkinkan berkat tersedianya *stream data* atau *big data* yang diperlukan untuk melatih parameter dalam jumlah raksasa tersebut.



Dengan adanya konvergensi antara *deep learning* dan *big data* telah terbukti mampu memberikan *insight* atau performa yang sama dengan bahkan melampaui *human level performance*.

Para peserta lomba diharapkan mampu melakukan proses Penambangan Data terhadap *big data* yang tersedia secara publik di web serta sudah terverifikasi validitasnya. Langkah peserta selanjutnya adalah memvisualisasikan hasil proses Penambangan Data tersebut sehingga berguna bagi masyarakat luas.

2. Babak Penyisihan

- a) Pengumpulan makalah Penambangan Data dengan tema “**Penambangan Data untuk Indonesia Maju**”.
- b) Makalah menyajikan beberapa poin penting sebagai berikut:
 - 1) Judul makalah;
 - 2) Latar belakang yang terkait dengan permasalahan seputar *Deep Learning* dan *Big Data* untuk solusi bagi masalah yang ada di masyarakat Indonesia;
 - 3) Tujuan dan manfaat yang diperoleh dari proses Penambangan Data;
 - 4) Batasan yang digunakan;
 - 5) Metode Penambangan Data;
 - 6) Desain dan implementasi Penambangan Data;
 - 7) Analisis;
 - 8) Kesimpulan;
 - 9) Dokumentasi
- c) Data dipastikan bersifat terbuka untuk publik dan boleh digunakan untuk lomba;

- d) Peserta tidak diharuskan menggunakan *Deep Learning* walaupun biasanya *Big Data* seringkali lebih dapat diselesaikan dengan *Deep Learning*;
- e) Peserta boleh tapi tidak diharuskan untuk menggunakan GPU (*Graphical Processing Unit*) untuk melakukan komputasi algoritmanya;
- f) Peserta boleh menggunakan *tools*, *library*, atau *framework* apa saja;
- g) Solusi dan algoritma yang diusulkan belum pernah digunakan atau dipublikasikan sebelumnya, baik untuk lomba maupun publikasi ilmiah;
- h) Jika solusi dan algoritma adalah modifikasi dari apa yang ada sebelumnya, harus dijelaskan modifikasi dan inovasi apa yang dilakukan serta menjelaskan sumber asli maupun sumber inspirasi yang dirujuk.
- i) Pengumpulan makalah untuk babak penyisihan dalam format PDF dengan format penamaan *file* “**GEMASTIK 13 - Penambangan Data - <ID Tim> - <Nama Tim> - <Judul Karya>.pdf**” dan dengan ukuran maksimal *file* 8 MB.
- j) Makalah paling lambat dikumpulkan pada tanggal **12 September 2020**.

3. Babak Final

Pengumuman tim yang lolos ke babak final akan diumumkan pada tanggal **24 September 2020**. Sedangkan untuk pelaksanaan babak final akan dilaksanakan pada tanggal **21 – 24 Oktober 2020**, dengan ketentuan sebagai berikut.

- a) Tim finalis melakukan pendaftaran ulang untuk kepesertaan pada Babak Final.
- b) Tim finalis wajib membuat poster yang dipamerkan pada sosmed dengan ketentuan sebagaimana pada Lampiran.
- c) Peserta akan diberikan dataset disertai deskripsi singkat;
- d) *Dataset* yang diberikan adalah *training data* yang hanya merupakan 50% dari data keseluruhan. Dengan *training data* tersebut, peserta akan diberi waktu **5 jam** untuk membangun model;
- e) Peserta kemudian akan diberikan 50% data uji dan melakukan pengukuran akurasi menggunakan komputer masing-masing di bawah pengawasan dan dicatat oleh dewan juri;
- f) Peserta harus mendokumentasikan pekerjaan Penambangan-Data-nya dalam bentuk *file* **PPT**, dan kemudian mempresentasikan di hadapan juri secara **daring**;
- g) Dalam membangun model, peserta diperbolehkan menggunakan alat bantu Penambangan Data (*tools*, *library*, atau *framework*);
- h) Walaupun kasus yang diberikan adalah *big data problem*, panitia merancang agar problem masih *scalable* untuk diselesaikan tanpa perlu menggunakan GPU.

4. Kriteria Penilaian

Penilaian Babak Penyisihan

- Penilaian utamanya adalah apakah peserta dapat menjadikan *Deep Learning* dan *Big Data* bagi solusi permasalahan yang ada di tanah air atau memberikan manfaat bagi kepentingan masyarakat;
- Laporan yang dikumpulkan tidak menentukan urutan. Laporan adalah syarat mutlak bagi peserta agar dapat membuktikan bahwa peserta mengerjakan penyisihan sendiri. Kegagalan peserta dalam memberikan laporan yang benar akan berakibat terhadap tidak diperhitungkannya peserta tersebut untuk lolos ke final.

Penilaian Babak Final

- Perhitungan skor didasarkan pada tingkat akurasi terhadap data uji yakni 50% data yang tidak digunakan pada saat membangun model dan juga inovasi dan kejelasan dalam presentasi;
- Ranking diurutkan berdasarkan total poin di akhir kompetisi;
- Apabila ada peserta dengan total poin yang sama, maka akan diurutkan berdasarkan waktu terakhir peserta tersebut melakukan *submission flag* ke sistem lomba. Apabila masih sama, maka akan dilihat dari waktu *submission* sebelumnya.

5. Format Penilaian

No	Kriteria	Bobot
1	Babak Penyisihan	
	1. Originalitas	20%
	2. Kebaruan	20%
	3. Manfaat	20%
	4. <i>Clarity</i> dalam tulisan	20%
	5. Kelengkapan Laporan	20%
2	Babak Final	
	1. Nilai dari penyisihan	25%
	2. Skor Akurasi	25%
	3. Inovasi	25%
	4. <i>Clarity</i> dalam menjelaskan	25%

Total Skor (Bobot x Nilai) =