

Proyek Akhir Semester
CSIE604284 • Analitika Media Sosial – Genap 2021/2022
[Document Version 1.0 – update 2022-05-15, 5.30 PM]

Rangkuman Timeline¹:

1st *checkpoint*: 19 - 27 Mei 2022

2nd *checkpoint*: 10 - 15 Juni 2022

Analytics insight submission: Jumat, 17 Juni 2022, pukul 22.00

Synchronous presentation: Selasa / Rabu, 21 / 22 Juni 2022 (waktu akan diinformasikan)

End-to-End Social Network Analysis

Pada tugas akhir kuliah Analitika Media Sosial ini, peserta kuliah dalam kelompok beranggotakan 4 - 5 orang mengerjakan proyek analitika media sosial *end-to-end* yang mencakup seluruh materi yang dipelajari selama perkuliahan satu semester ini: *social media data collection*, *text mining*, *demography analysis*, dan *social network analytics*. Kompetensi utama yang dievaluasi dalam tugas proyek akhir ini adalah memahami dan menerapkan metodologi analitik yang tepat dan runtut dengan menggunakan data media sosial.

Pengerjaan tugas proyek akhir ini akan didampingi oleh mentor (tim asisten dosen). Terdapat tiga *checkpoint*. Pada dua *checkpoint* pertama, setiap kelompok wajib melakukan sesi mentoring sinkronus dengan asisten dan *submit deliverable* yang ditentukan. Pada *final checkpoint*, setiap kelompok *submit* hasil analisis dan kemudian dijadwalkan sesi sinkronus untuk tanya jawab dengan dosen dan tim asisten.

Tahapan metodologi yang perlu dieksekusi dalam proyek akhir ini adalah sebagai berikut:

1. Pemilihan subtopik dan pertanyaan penelitian
2. Pengumpulan data postingan media sosial (*tweet*)
3. *Text mining* dengan data *tweet*
4. Pengumpulan data pengguna media sosial (*user*)
5. Analisis demografi pengguna
6. Pembangunan data *social network*
7. Analisis *social network*

¹ Detail lihat *section* Timeline & Deadline (halaman 5 deskripsi soal ini)

Tahap 1: Pemilihan Topik

Selama dua tahun terakhir sejak awal 2020, pandemi COVID-19 telah melanda Indonesia. Diskursus topik COVID-19 menjadi salah satu pembicaraan yang mendominasi media sosial di Indonesia dalam kurun waktu 2 tahun tersebut, baik mengenai aspek kesehatan, pendidikan, maupun perekonomian. Masyarakat memiliki pandangan yang beragam terhadap suatu aspek baik itu aspek kebijakan pemerintah maupun aspek ilmiah.

Disediakan 4 subtopik. Kelompok dapat memilih subtopik yang akan dikerjakan (FCFS). Setiap subtopik dikerjakan oleh maksimal 2 kelompok.

No.	Subtopik	Mentor
1	Vaksinasi COVID-19	Douglas
2	Kebijakan pembatasan/ <i>lockdown</i> (PSBB atau PPKM) atau prioritisasi aspek ekonomi atau kesehatan selama pembatasan	Hadi
3	<i>Work From Home</i> (Kegiatan Daring) atau <i>Work From Office</i> (Kegiatan Luring) dalam hal pekerjaan ataupun pembelajaran.	Dean (Raihansyah)
4	Kebijakan mudik atau berkumpul di hari besar / libur panjang.	Dipta

Kelompok boleh bertanya-tanya dulu ke asisten mentor terkait subtopik sebelum memutuskan topik yang akan dikerjakan.

Tahap 2: Tweet Data Collection

Setiap kelompok diminta melakukan pengumpulan data Tweet. Boleh melakukan *crawling* data secara mandiri atau menggunakan data Tweet bahasa Indonesia topik Covid-19 yang tersedia publik, misalnya <https://github.com/douglasraevan/covid19-indonesian-tweet>. Data Tweet ini dikumpulkan dengan *keyword* general Covid-19.

Kelompok Anda perlu melakukan *filtering* data yang relevan sesuai sub topik yang akan dikerjakan. *Filtering* bisa menggunakan *keyword* atau waktu *post*.

Yang dinilai di tahap 2: rancangan dan implementasi data *filtering*

Tahap 3: Text Mining

Kelompok perlu melabeli data Tweet sesuai dengan subtopik yang dipilih, bisa berupa label *sentiment* (positif, negatif, netral), *stance* (pro, kontra, abstrain), atau topik. Karena data Tweet yang dikumpulkan sangat banyak, tentu saja tidak mungkin dilabeli seluruhnya.

Bangunlah model *text classification* yang dilatih dengan menggunakan data berlabel. Labeli minimal 100 Tweet terlebih dahulu.

Eksplorasi model klasifikasi dan akurasi *classifier* tidak menjadi komponen penilaian dalam tugas ini. Jadi, cukup hanya implementasikan satu *classifier* yang dianggap terbaik (termasuk pemilihan fitur) berdasarkan pengalaman tutorial Text Mining atau proyek tengah semester *classification* menggunakan data Tweet.

Yang dinilai di tahap 3:

- pemahaman *text mining task* (ketepatan formulasi *text classification task* yang sesuai dengan subtopik yang dikerjakan)
- ketepatan metodologi dalam mengolah *tweet*, termasuk anotasi dan *model building* (*pre processing* sampai menerapkan *classifier* ke *unlabeled data*)

Tahap 4: User Data Collection

Setiap kelompok diminta melakukan pengumpulan data *user* dari Tweet yang dikumpulkan di tahap 2. Data *user* yang dikumpulkan adalah data *user profile* dan/atau *user tweet*, sesuai kebutuhan model *demography analysis* (tahap 5).

Tahap 5: Demography Analysis

Setiap kelompok diminta memprediksi atribut demografi data *user* yang sudah dikumpulkan di tahap 4.

Eksplorasi model klasifikasi dan akurasi *classifier* tidak menjadi komponen penilaian dalam tugas ini. Jadi, cukup hanya implementasikan satu *classifier* terbaik berdasarkan tugas proyek tengah semester.

Yang dinilai di tahap 4 dan 5: ketepatan metodologi.

Perlu diperhatikan bahwa tahap 4 dan 5 ini memengaruhi tahap 6 dan 7.

Tahap 6: Pembangunan Data Social Network

Setiap kelompok diminta membangun data *social network* dari data *user* yang sudah dikumpulkan di tahap 5 atau data *tweet* dari tahap 3.

Data *social network* adalah graf terdiri dari dua komponen: verteks dan edge. Verteks merupakan pengguna media sosial. Kelompok diminta menentukan definisi *edge* yang sesuai dengan subtopik dan pertanyaan analitik yang akan dijawab pada tahap 7 (contoh edge: *following*, *friends*, *interacts with*, *retweet from*, ...).

Definisikan dengan jelas rancangan Anda apakah *social network* yang dibangun berupa graf *directed* vs *undirected*, *weighted* vs *unweighted*, graf berdasarkan *modality* / *plexity*.

Bisa jadi Anda perlu membangun lebih dari satu *social network*.

Karena jumlah tweet dan user yang akan diperoleh di tahap sebelumnya bisa mencapai ribuan bahkan jutaan, Anda boleh melakukan *network sampling* (lebih banyak verteks lebih baik, sesuaikan dengan *resource* komputasi yang tersedia).

Yang dinilai di tahap 6:

- Pemahaman konsep *social network*, yang tergambar dari pendefinisian graf yang Anda kumpulkan.
- Sampling data *social network*.

Tahap 7: Analytics Finding

Untuk setiap subtopik akan disediakan beberapa pertanyaan analitik. Setiap kelompok diminta menjawab 2 - 3 pertanyaan tersebut. Topik pertanyaan analitik memiliki kisi-kisi sebagai berikut (i) menggabungkan *text mining* dengan analisis demografi, (ii) *centrality* dan/atau *network measurement*, (iii) *homophily*, (iv) *community detection*.

Dalam melakukan analisis, mahasiswa bebas memilih *tool* yang digunakan. Boleh memakai Gephi, boleh menulis code dengan bahasa pemrograman apapun (boleh dari *scratch* atau menggunakan *library*). Perhatikan etika akademis ketika menggunakan / menggubah pekerjaan orang lain, berikan kutipan (*acknowledgement*).

Untuk menjawab pertanyaan ini dengan tepat, tahap 1 - 6 dalam proyek ini harus dieksekusi dengan tepat.

Yang dinilai di tahap 7:

- Pemahaman konsep / materi yang sesuai dengan pertanyaan analitik (misalnya paham *centrality* itu apa, bagaimana cara menghitungnya, dst).
- Penggunaan *tool* yang tepat guna sesuai kebutuhan.
- Analisis dan interpretasi dari ukuran kuantitatif yang dihitung (misalnya jika dalam *social network* ada *node* dengan *centrality* 0,8 artinya apa, dst)

KOMPONEN PENILAIAN

Selain poin-poin penilaian yang sudah disebutkan di tahap 1 - 7, dalam proyek akhir ini juga dinilai

- Penerapan metodologi yang sistematis dari awal sampai akhir.
Ada tahapan yang bisa dikerjakan secara paralel, ada yang harus dilakukan secara sekuensial.
- *Project management*, mencakup *human resource management* (pembagian tugas antar anggota kelompok), *time management* (perencanaan waktu mengerjakan proyek), dan *team management* (bagaimana bekerja dengan orang baru, memotivasi tim, dan kecakapan menyiapkan *backup plan* jika ada anggota yang tidak *perform*)
- *Reporting & presentation skill*.
Selain melaporkan hasil analisis secara tertulis, setiap kelompok akan dijadwalkan sesi tanya jawab sinkronus dengan dosen dan tim asisten.
- Keinginan belajar dan partisipasi.
Hasil akhir bukan 100% pencapaian. Bagaimana setiap individu belajar dari pengerjaan proyek ini akan dinilai.

Di akhir proyek, setiap anggota kelompok akan diminta melakukan *peer review*.

TIMELINE & DEADLINE

1st checkpoint (19 - 27 Mei 2022)

Setiap kelompok memilih subtopik yang akan dikerjakan, menyusun *project plan* sederhana (rencana pembagian tugas antar anggota kelompok dan rencana waktu pengerjaan).

Pada periode ini, diharapkan setiap kelompok sudah menyelesaikan sekurang-kurangnya tahap 1 - 3 dan sudah punya *draft* rancangan pengerjaan tahap 4 - 6 (lebih baik jika sudah mulai dieksekusi). Setiap kelompok perlu melakukan mentoring wajib sekali di periode ini.

Deadline:

- Submit *project plan* (Senin, 23 Mei 2022, pukul 22.00)
- Submit *deliverable* tahap 2 - 3 (Jumat, 27 Mei 2022, pukul 22.00)
 - Tahap 2: filtered data
 - Tahap 3: labeled data yang dianotasi dan hasil prediksi *classifier* (otomatis) serta model *classifier*
 - Max 1 *page report* menjelaskan tahap 2 dan 3

Terkait *draft* rancangan pengerjaan tahap 4 - 6, silakan sepakati mekanisme apa yang perlu dikumpulkan (*customized* antar subtopik / kelompok).

2nd checkpoint (10 - 15 Juni 2022)

Setiap kelompok diharapkan sudah menyelesaikan tahap 4 - 6 dan sudah memiliki gambaran apa yang akan dikerjakan untuk tahap 7. Setiap kelompok perlu melakukan mentoring wajib sekali di periode ini.

Final checkpoint

Setiap kelompok *submit final deliverable* dengan *deadline* **Jumat, 17 Juni 2022, pukul 22.00**

- Tahap 4 & 5: data *user* dengan label atribut demografi
- Tahap 6: data *social network* dalam format *machine readable*
- Tahap 7: Jawaban pertanyaan analitik dan penjelasan langkah (dalam bentuk laporan atau *slide* presentasi)
- Code untuk pengumpulan data dan/atau analisis²

Sesi sinkronus

Setiap kelompok akan dijadwalkan sesi tanya jawab secara sinkronus dengan dosen dan tim dosen terkait proyek yang dikerjakan. Sesi untuk setiap kelompok 15 - 20 menit, dijadwalkan di antara Selasa, 21 atau Rabu, 22 Juni 2022 (waktu akan diinformasikan)

² Silakan sesuaikan jika menggunakan *tool*