# Intro Topic Modelling

## Latar Belakang

Menurut Kementerian Komunikasi dan Informatika RI 30 juta anakanak dan remaja di Indonesia merupakan pengguna internet, dan media sosial menjadi pilihan utama sarana komunikasi bagi mereka, telah tercatat sebesar (73 persen) remaja mengakses internet menggunakan smartphone mereka, dan (4 persen) menggunakan tablet [13]. Selaras dengan perkembangan internet, penggunaan media sosial di kalangan remaja telah menjadi kebiasaan kehidupan sehari-hari mereka. Menurut studi ini ditemukan 98 persen dari anak-anak dan remaja yang disurvei tahu tentang internet dan 79,5 persen diantaranya adalah pengguna aktif media sosial. Faktor yang mendasari para remaja mengakses internet dan media sosial menurut riset Kementerian Komunikasi dan Informatika RI dan UNICEF ada tiga, yakni untuk mencari informasi, terhubung dengan teman (lama ataupun baru) dan untuk hiburan semata [13].

Tingginya rasa ingin tahu remaja terhadap hal baru tentunya mendorong orang tua untuk berperan aktif dalam mengawasi putra-putrinya pada saat mengakses internet, adapun menurut penelitian yang dilakukan dilakukan oleh Kominfo dan UNICEF mengatakan bahwa para orang tua merasa kewalahan dalam memantau atau mendampingi putra-putrinya dalam mengakses internet maupun media sosial, khususnya media sosial yang marak diakses oleh remaja, seperti *Facebook, Twitter, Instagram*. Tingkat penetrasi media sosial di kalangan remaja sendiri, khususnya di wilayah perkotaan, Surabaya, menurut penelitian Pengaruh dan Pola Aktivitas Penggunaan Internet serta Media Sosial pada Siswa SMPN 52 Surabaya didominasi oleh *Facebook* dengan prosentase pengguna (14 persen) dari keseluruhan pengguna, kemudian, disusul dengan *WhatsApp, Twitter, Facebook Messenger, Google+, LinkedIn, Instagram, Skype, Pinterest* dan urutan terakhir ditempati *LINE* dengan presentase (6 persen)[20]. Dilansir dalam survei yang dilakukan oleh Asosiasi Penyelenggaraan Jasa Internet Indonesia [1] menyatakan bahwa pengguna *Instagram* mengalami kenaikan yang signifikan sejak pertama diluncurkan pada Oktober 2010, tercatat ada 500 juta pengguna aktif dunia di tahun 2016, dan di Indonesia tercatat sebanyak 19,9 juta atau sebesar (15 persen) pengguna. Seiring meningkatnya jumlah pengguna media sosial khususnya *instagram*, dilakukan proses percobaan pencarian kata kunci “media sosial remaja” menggukan mesin pencari Google, dan ditemukan sebanyak 785,000 hasil pencarian, yang mayoritas menunjukkan dampakdampak dari penggunaan media sosial di kalangan remaja, mulai dari pergaulan, kenalakan remaja hingga dampak negatif yang tergolong fatal, seperti pertikaian sesama remaja.

Dari fenomena yang terjadi terkait penggunaan media sosial di kalangan remaja dibutuhkan sebuah *platform* yang mampu memberikan informasi visual terhadap aktivitas remaja dalam hal berekspresi di media sosial *Instagram*, dengan cara melakukan analisis topic modelling terhadap perilaku atau kebiasaan remaja ketika *upload* gambar disertai dengan *caption* tertentu, menggunakan metode *Latent Dirichlet Allocation* atau yang akrab disebut dengan *LDA*.

Penelitian ini dikhususkan untuk menganalisa data caption akun Instagram siswa SMP dari 18 sekolah di Surabaya, yang menjadi target untuk kelas mata kuliah Etika Profesi di Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember Surabaya. Setelah data akun dan data caption didapatkan serta dianalisa menggunakan *LDA*, kemudian dilakukan visualisasi terhadap topik atau kategori aktivitas siswa berdasarkan *caption*nya. Melalui penelitian ini diharapkan mampu memberikan informasi kepada para orang tua dan guru untuk lebih peka dan mampu mengarahkan putra-putrinya agar dapat bertanggung jawab dalam hal berekspresi di media sosial, sehingga pelanggaran serta dampak negatif yang dikhawatirkan dari pengaruh media sosial dapat diminimalisir.

## Perumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan permasalahan yang menjadi fokus dan akan diselesaikan dalam Tugas Akhir ini antara lain :

1. Bagaimana cara mengakuisisi data akun dan aktivitas yang dilakukan oleh anak SMP di Surabaya melalui media sosial *instagram*?
2. Bagaimana cara merancang *platform* yang mampu melakukan visualisasi data aktivitas pelajar SMP ke dalam sebuah gambar atau *dashboard*?
3. Bagaimana cara mengelompokkan topik caption atau caption instagram dari akun siswa SMP?

## Batasan Masalah

Dari permasalahan yang disebutkan di atas, batasan masalah dalam tugas akhir ini adalah :

1. Studi kasus yang digunakan pada penelitian ini adalah 18 Sekolah Menegah Pertama di wilayah Surabaya.
2. Pengambilan data dalam tugas akhir ini dilakukan dari *caption account instagram* pelajar SMP di Surabaya

# Tujuan Penelitian

Berdasarkan hasil perumusan masalah dan batasan masalah yang telah disebutkan sebelumnya, maka tujuan yang dicapai dari tugas akhir ini adalah untuk menciptakan sebuah platform yang mampu mengklasifikasi dan memvisualkan perilaku atau aktivitas pelajar SMP di media sosial instagram guna mendorong para orang tua atau guru untuk lebih memperhatikan putra-putrinya dalam hal berekspresi di media sosial. Sehingga dengan adanya platform ini diharapkan para pelajar SMP di Surabaya mampu berekspresi di media sosial dengan lebih bertanggung jawab.

# Daftar Kata Per Topik

|  |  |
| --- | --- |
| 2 Topic | |
| I | II |
| hati | sukses |
| surabaya | selfie |
| sayang | sahabat |
| selamat | kawan |
| senyum | temen |
| photograph | rindu |
| kelas | sehat |
| hbd | usaha |
| tuhan | challenge |
| juang | semangat |
| indonesia | makan |
| kangen | surabaya |
| percaya | nakal |
| lomba | smp |
| family | mannequinchallenge |

|  |  |  |
| --- | --- | --- |
| 3 Topic | | |
| I | II | III |
| hati | surabaya | Selamat |
| sahabat | sayang | Hbd |
| selfie | sukses | Tuhan |
| senyum | sehat | Challenge |
| photograph | family | satriaandthemonster |
| kawan | smile | Indonesia |
| juang | nakal | Makan |
| kangen | drum | Allah |
| kelas | nyata | Pacar |
| percaya | play | Murid |
| rindu | team | Guru |
| sekolah | semangat | mannequinchallenge |
| lomba | video | Sidoarjo |
| keluarga | tua | Jabon |
| usaha | saudara | Kerja |

|  |  |  |  |
| --- | --- | --- | --- |
| 4 Topic | | | |
| I | II | III | IV |
| sayang | surabaya | rindu | selfie |
| hati | sahabat | usaha | kawan |
| percaya | sukses | family | selamat |
| challenge | senyum | dunia | photograph |
| smile | hbd | murid | kelas |
| juang | temen | nyata | makan |
| nikmat | tuhan | kerja | guru |
| accepted | lomba | sehat | keluarga |
| music | kangen | kenang | surabaya |
| muka | smp | nissan | sekolah |
| guitar | mannequinchallenge | surabayavidgram | semangat |
| papa | senang | team | raya |
| doa | satriaandthemonster | jabon | rock |
| play | beach | selamat | nakal |
| candid | iseng | santai | hati |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 Topic | | | | |
| I | II | III | IV | V |
| surabaya | sayang | sukses | keluarga | sahabat |
| selamat | senyum | kawan | percaya | selfie |
| hati | hbd | kelas | surabayavidgram | usaha |
| juang | photograph | kangen | muka | temen |
| indonesia | family | challenge | cover | guru |
| smile | pacar | sekolah | saudara | allah |
| photo | rindu | lomba | kenang | photograph |
| dunia | murid | semangat | hati | iseng |
| nakal | nyata | sehat | laut | santai |
| rock | mannequinchallenge | tuhan | exploresurabaya | tua |
| air | nissan | alhamdulillah | karya | alam |
| music | team | konco | bantu | pahlawan |
| satriaandthemonster | sidoarjo | nikmat | beach | cerita |
| raya | jabon | mama | kapten | drummer |
| manis | play | drum | temen | papa |

Uji Perplexity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STDEV N0\_STEM | | | | |
| Topik 2 | Topik 3 | Topik 4 | Topik 5 | Topik 6 |
| 2,02 | 1,61 | 2,66 | 3,98 | 3,58 |
|  |  |  |  |  |
| STDEV WITH\_STEM | | | | |
| Topik 2 | Topik 3 | Topik 4 | Topik 5 | Topik 6 |
| 1,886 | 2,217 | 3,267 | 1,829 | 4,210 |

|  |  |  |
| --- | --- | --- |
| No. Stemming | | |
| Topic | Perplexity | StDEV |
| 2 | 380,3 | 2,02 |
| 3 | 386,9 | 1,61 |
| 4 | 409,2 | 2,66 |
| 5 | 428,9 | 3,98 |
| 6 | 458 | 3,58 |
|  |  |  |
| With Stemming | | |
| Topic | Perplexity | StDEV |
| 2 | 247,9 | 1,89 |
| 3 | 265 | 2,22 |
| 4 | 275,4 | 3,27 |
| 5 | 294,3 | 1,83 |
| 6 | 312,2 | 4,21 |