

**USULAN HIBAH PENELITIAN
BELAJAR BERSAMA PEMPROV DKI JAKARTA**



**Deteksi Trending Topic Pada Masa Pandemi Corona Virus 19
(COVID-19)
Pada Tweet Berbahasa Indonesia Di DKI Jakarta
Menggunakan Metode Document Pivot (Doc-p) dan BN-grams
Studi kasus : BNPB**

TIM PENELITI :

Ketua : Dr. Indra, S.Kom, M.T.I (0322038603)
Anggota 1 : Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng
(195809161986011001)
Anggota 2 : Agus Umar Hamdani, M.Kom (0309088302)
Anggota 3 : Dr. M. Syafrullah, M.Kom, M.Sc (0316097401)

**UNIVERSITAS BUDI LUHUR
OKTOBER 2020**

RINGKASAN

Sejak terjadinya pandemi Corona Virus 19 (Covid-19) yang terjadi di Wuhan, China, maka perbincangan yang berkaitan dengan virus tersebut menjadi berita utama di media sosial, televisi, radio dan bahkan antar pribadi di lingkungan keluarga. Masyarakat memberikan beberapa informasi terkait wabah virus Corona seperti: kebutuhan masyarakat yang menjadi korban dan lokasi terdampak Covid-19. Informasi tersebut dapat digunakan sebagai salah satu bahan untuk memetakan peristiwa bencana wabah virus Corona, serta kebutuhan korban bencana yang terjadi di Indonesia. Namun, informasi yang berasal dari media sosial memiliki struktur yang tidak baku serta kredibilitas yang rendah sebagai penyedia informasi. Data media sosial yang tidak terstruktur menyebabkan informasi yang sulit untuk diidentifikasi, terutama yang terkait trending topic bencana wabah virus Corona. Oleh karena itu, pada penelitian ini dilakukan pemodelan deteksi trending topic data tweet menggunakan metode *Document Pivot (Doc-p)* dan *BN-grams*.

Metode Doc-P dan BN-grams diharapkan menghasilkan beberapa informasi trending topic pada masa pandemi Covid-19 sehingga dapat digunakan untuk sumber berita alternatif pada pemerintah, media berita online maupun masyarakat. Garis besar penelitian ini dengan lama penelitian tiga tahun. Tahun pertama dilakukan deteksi trending topic menggunakan BN-grams. Tahun kedua dilakukan deteksi trending topic menggunakan BN-grams dan Doc-p. Tahun ketiga, dilakukan perbaikan kualitas trending topic dan ekstraksi entitas **lokasi, jumlah korban dan kebutuhan terdampak pandemi Covid-19**.

Diharapkan dengan penelitian ini dapat membantu Badan Nasional Penanggulangan Bencana (BNPB) untuk mendapatkan informasi trending topic di masa pandemi virus Corona yang terjadi di Indonesia sedini mungkin, faktual dan actual dengan menggunakan data dari sosial media *Twitter*. Disisi lain, informasi tersebut dapat digunakan sebagai alternatif informasi terkait bencana virus Corona yang digunakan oleh masyarakat.

Deteksi *trending topic* sebagai pondasi utama untuk memonitor dan merangkum informasi dari media sosial terutama twitter untuk menghasilkan topik yang relevan berdasarkan peristiwa yang terjadi secara real dilapangan. Untuk menghasilkan *trending topic* pada penelitian ini diusulkan sebuah metode berbasis *textual content* yaitu metode BN-Grams dan Doc-p. Metode berbasis textual content adalah metode untuk mendeteksi trending topik berdasarkan sekumpulan text dari pesan yang dikumpulkan dari pengguna Twitter. Metode BN-Grams adalah metode yang digunakan untuk mendeteksi trending topik berdasarkan frekuensi kemunculan ngrams dan klasterisasi n-grams pada setiap periode waktu tertentu.

Setiap kluster merupakan representasi dari topik yang dihasilkan. Disisi lain, Doc-p adalah metode deteksi trending topic berdasarkan kesamaan kalimat antar tweet. Diharapkan dengan penggunaan metode BN-Grams dan Doc-p dalam bahasa Indonesia ini dapat meningkatkan akurasi dari proses deteksi *trending topic* di Indonesia terutama terkait Covid-19. Selain itu, *trending topic* yang dihasilkan diharapkan dapat menjadi *Emerging Trend* yang berfungsi sebagai berita yang lebih cepat dihasilkan dibandingkan media lokal dan pelengkap dari berita yang dihasilkan dari media cetak maupun elektronik.

Hasil dari penelitian ini diharapkan dapat menghasilkan publikasi pada jurnal internasional serta prosiding internasional. Publikasi dari hasil akhir dari penelitian ini akan dipublikasikan pada Jurnal of King Saud University Computer and Information Sciences (JKSUCI). Hasil akhir penelitian yang diusulkan juga akan dipublikasikan pada Seminar Internasional bereputasi pada *International Conference on Advanced Computer Science and Information Systems 2021* yang diselenggarakan oleh Universitas Indonesia.

Kata Kunci: Doc-P, BN-gram, Corona Virus 19 (COVID-19)

DAFTAR ISI

RINGKASAN	2
DAFTAR ISI.....	4
BIDANG PENELITIAN	5
LATAR BELAKANG	6
MAKSUD DAN TUJUAN PENELITIAN	8
Maksud Penelitian.....	8
Tujuan Penelitian	8
RUANG LINGKUP	12
JANGKA WAKTU PENELITIAN.....	13
TINJAUAN PUSTAKA	14
METODE, JANGKA WAKTU DAN LOKASI PENELITIAN	16
SASARAN DAN TARGET PENELITIAN	21
LOKASI PENELITIAN.....	22
HASIL YANG DIHARAPKAN	23
DAFTAR PUSTAKA	24
LAMPIRAN 4 BIODATA TIM PENELITI	26
LAMPIRAN 2 BIODATA TIM PENELITI	27

BIDANG PENELITIAN

Bidang Penelitian yang diusulkan adalah **Teknologi**. Pendekatan teknologi Informasi dan komunikasi yang digunakan untuk pemetaan informasi terkait Covid-19 pada Media Sosial terutama Twitter.

LATAR BELAKANG

Pada bulan Januari hingga Mei 2020 Indonesia terkena wabah Covid-19. Status pengguna di Twitter di Indonesia mayoritas berisi tentang informasi atau opini terkait Covid-19. Data twitter (tweet) jumlah terus bertambah dan semakin masif membahas tentang Covid-19. Sekumpulan tweet tersebut memiliki sekumpulan topik yang tersembunyi dan ramai dibahas oleh Netizen. Pada penelitian ini diusulkan deteksi trending topic untuk menjadi pondasi utama sumber informasi bagi pemerintah dan berita online maupun televisi. Informasi yang berasal dari Twitter menurut Aiello dkk.(2013) memiliki ciri: orisinal, tidak diedit dan faktual. Informasi yang dihasilkan Twitter berasal langsung dari masyarakat pengguna Twitter hal ini lebih dikenal dengan Orisinal.

Informasi yang berasal dari Twitter tidak dilakukan pengeditan, tidak seperti di media berita online ada tim editor sebelum ditayangkan beritanya. Informasi dari Twitter merupakan informasi yang menggambarkan realitas peristiwa yang terjadi pada waktu dan tempat tertentu sehingga faktual.

Pada penelitian tersebut dilakukan deteksi trending topic menggunakan metode BN-Grams dan Doc-p. Kedua metode tersebut pertama kali dikenalkan oleh (Aiello et al., 2013) untuk mendeteksi trending topic pada tweet berbahasa Inggris. Pada penelitian (Indra et al., 2018) dikembangkan untuk mendeteksi trending topic terkait peristiwa politik yang terjadi pada tahun 2018 di Indonesia. Pada penelitian Doc-p dan BN-grams digunakan untuk mendeteksi trending topic terkait berbagai peristiwa yang berhubungan dengan Covid-19 di DKI Jakarta.

Pada Penelitian ini, BN-grams menggunakan sistem pengenalan entitas (Named Entity Recognition atau NER) InNer (Indonesian Named Entity Recognition) (Budi et al., 2005) .Jika sebelumnya menggunakan NER berbahasa Inggris. Diharapkan dengan adanya penelitian ini dapat membantu Badan Nasional Penanggulangan Bencana (BNPB) dan Kementerian Kesehatan Republik Indonesia untuk mendapatkan informasi trending topic di masa pandemi Covid-19 sedini

mungkin, faktual dan actual dengan menggunakan data dari sosial media Twitter terutama terkait **lokasi, jumlah korban dan kebutuhan terdampak pandemi Covid-19**. Disisi lain, informasi tersebut dapat digunakan sebagai langkah penyelamatan untuk mengurangi dampak buruk bencana alam yang terjadi.

MAKSUD DAN TUJUAN PENELITIAN

Maksud Penelitian

Penelitian ini bermaksud untuk menyelesaikan permasalahan yang ditemukan yakni Bagaimana mendeteksi trending topic terkait Covid-19 dengan menggunakan metode BN-grams dan Doc-p pada tweet berbahasa Indonesia. Oleh karena itu, tujuan utama dari penelitian ini adalah mendeteksi trending topic Covid-19 pada tweet berbahasa Indonesia menggunakan metode BN-grams dan Doc-p.

Tujuan Penelitian

Tujuan utama dari penelitian ini adalah mendeteksi trending topic Covid-19 pada tweet berbahasa Indonesia menggunakan metode BN-grams dan Doc-p. Urgensi utama pada penelitian ini diharapkan dapat memberikan kontribusi yang nyata pada bidang penelitian ilmu komputer, terutama dalam topik *Topic Detection and Tracking* (TDT) berbahasa Indonesia terutama dibidang pemetaan informasi terkait COvid-19 pada media social (Twitter). Trending topic berbahasa Indonesia ini diharapkan dapat menjadi sebuah emerging trend. Emerging trend adalah informasi berita yang lebih cepat muncul dibandingkan dengan berita lokal. Jika emerging trend menghasilkan sebuah trending topik yang muncul pada hari ini, maka informasi trending topik tersebut dapat menjadi salah satu informasi berita dihari berikutnya. Selain itu, informasi trending topik ini dapat digunakan sebagai pelengkap dari informasi yang belum dipublikasikan di media berita resmi terutama informasi terkait Covid-19.

RUANG LINGKUP

Ruang lingkup pada penelitian ini adalah:

- a. Bidang Penelitian ini adalah Teknologi.
- b. Penelitian ini menggunakan tweet berbahasa Indonesia yang dikumpulkan pada Januari-Oktober tahun 2020.
- c. Data tweet yang digunakan untuk mendeteksi trending topic pada wilayah Pemprov DKI Jakarta.
- d. Kata kunci yang digunakan dalam penarikan data tweet adalah: OTG, ODP, PDP, Covid-19 Jakarta, kasus Suspek, Probable, konfirmasi, kontak erat, pelaku perjalanan, Discarded, selesai isolasi dan kematian.

JANGKA WAKTU PENELITIAN

Jangka Waktu penelitian Covid-19 dengan Pemprov DKI Jakarta dilakukan pada 01 Januari 2020 sampai 30 Juni 2020. Detil kegiatan dapat dilihat pada Tabel 2.

Tabel 2. Jadwal Kegiatan Penelitian

No	Nama Kegiatan	Bulan					
		1	2	3	4	5	6
1	Pengumpulan data	x					
2	Preliminary dataset analysis	x					
3	Preprocessing data	x					
4	Pembuatan aplikasi Deteksi trending topic menggunakan BN-grams dan Doc-p		x				
5	Implementasi dan ujicoba aplikasi deteksi trending topic		x				
6	Perbaikan kualitas trending topic menggunakan NLP			x			
7	Ekstraksi entitas lokasi, jumlah korban dan kebutuhan korban bencana alam			x			
8	Verifikasi dan validasi				x		
9	Pengujian				x		
10	Penulisan paper penelitian	x	x	x	x		
11	Proof read paper					x	
12	Submit paper					x	
13	Publikasi penelitian						x

TINJAUAN PUSTAKA

Penelitian terkait metode deteksi *trending topic* di media sosial diklasifikasikan menjadi dua yaitu metode *trending topic* berdasarkan *textual content* dan *social content*. Penelitian terkait ini merupakan dasar pemikiran utama dalam perancangan penelitian yang akan diusulkan.

2.1. *Trending topic* Berbasis *Textual Content*

Penelitian *trending topic* berbasis *textual content* adalah penelitian *trending topic* dengan mengolah teks atau pesan yang ada dalam tweet atau *post* di media sosial untuk didapatkan sebuah informasi tentang trending atau hot topik (Chen et al., 2012). Beberapa penelitian melakukan pengolahan dengan metode term frequency, burst detection dan LDA.

Penelitian deteksi trending topic dengan menggunakan metode term frequency adalah penelitian yang dilakukan oleh (Benhardus, 2010; Cvijikj & Michahelles, 2011; Phuvipadawat & Murata, 2010). Metode term frequency dalam deteksi trending topik menggunakan TF-IDF dan klasterisasi term untuk menghasilkan sebuah topik pada setiap klaster. Jika setiap klaster memiliki term yang frekuensinya melebihi dari frekuensi term yang lain maka klaster tersebut dapat menghasilkan trending topik.

Metode burst detection adalah salah satu metode untuk mendeteksi trending topik yang berasal dari penelitian yang dilakukan oleh (Guzman & Poblete, 2013; Kleinberg, 2002; Mathioudakis & Koudas, 2010; Matsui et al., 2013; Tamura et al., 2012; Xie et al., 2013). Contoh penerapan metode Burst Detection berdasarkan frekuensi kemunculan keyword NBA pada sejumlah tweet. Tweet yang berisi keyword NBA biasa muncul dalam periode waktu lima menit sekali, tetapi tiba-tiba jumlah tweet yang berisi keyword NBA dalam periode waktu lima menit memiliki jumlah 100. Keyword NBA dapat dikategorikan sebagai trending topik karena frekuensi kemunculannya lebih tinggi dengan jumlah frekuensi meningkat tajam dibandingkan dengan frekuensi kemunculan keyword NBA pada periode waktu seperti biasanya (Mathioudakis & Koudas, 2010).

Metode LDA adalah metode untuk mendeteksi trending topik berbasis probabilistik dengan menggunakan data diskrit pada suatu korpus. Didalam metode LDA, dokumen digambarkan sebagai sekumpulan topik secara acak. Topik dikarakteristikan sebagai distribusi sebuah kata (Blei et al., 2003). Metode LDA dikombinasikan dengan term frequency invers document frequency (TF-IDF) untuk mendeteksi trending topik yang dilakukan oleh (Song & Kim, 2013). Metode LDA dikembangkan menjadi metode TC-LDA. Metode TC-LDA adalah metode untuk mendeteksi trending topik dengan menggunakan distribusi topik dengan urutan pengolahannya berdasarkan hirarki dari dokumen, topik, term inti dan kata (dokumen-topik-term inti-kata) (Ge et al., 2013). Metode TC-LDA dapat diilustrasikan dalam mendeteksi trending topik dengan menggunakan sekumpulan dokumen. Dokumen berisi sekumpulan topik. Topik berisi sekumpulan term. Term merupakan sekumpulan kata. Sedangkan, pada metode LDA dapat diilustrasikan dalam mendeteksi trending topik menggunakan sekumpulan dokumen. Dokumen berisi sekumpulan topik. Topik berisi sekumpulan kata. Metode LDA mendeteksi trending topik dengan menggunakan pola distribusi dokumen-topik-kata (Ge et al., 2013).

2.2. *Trending topic* Berbasis Social Content

Deteksi trending topik berbasis social content adalah deteksi trending topik dengan menggunakan link atau hubungan diantara pengguna twitter. Link atau hubungan diantara pengguna twitter diidentikan dengan jumlah retweet, mention dan follower. Jumlah follower pengguna twitter adalah salah satu komponen yang digunakan dalam mendeteksi trending topik yang dilakukan oleh (Cataldi et al., 2010; Chen et al., 2012).

Jumlah dari retweet dan follower merupakan salah satu komponen penting untuk proses perangkaan setiap klaster. Klaster yang memiliki ranking tertinggi dapat diindikasikan bahwa klaster tersebut berisi sekumpulan kata atau keyword yang merupakan trending topic (Phuvipadawat & Murata, 2010). Beberapa penelitian menggunakan mention untuk mendeteksi trending topic terutama pada penelitian yang dilakukan oleh (Takahashi et al., 2011, 2014). Deteksi trending topik dengan menggunakan mention dilakukan dengan menggunakan probabilitas

mention. Jika probabilitas kemunculan mention lebih tinggi dibandingkan dengan probabilitas mention lain dapat disimpulkan bahwa tweet yang berisi mention tersebut merupakan trending topik (Takahashi et al., 2014).

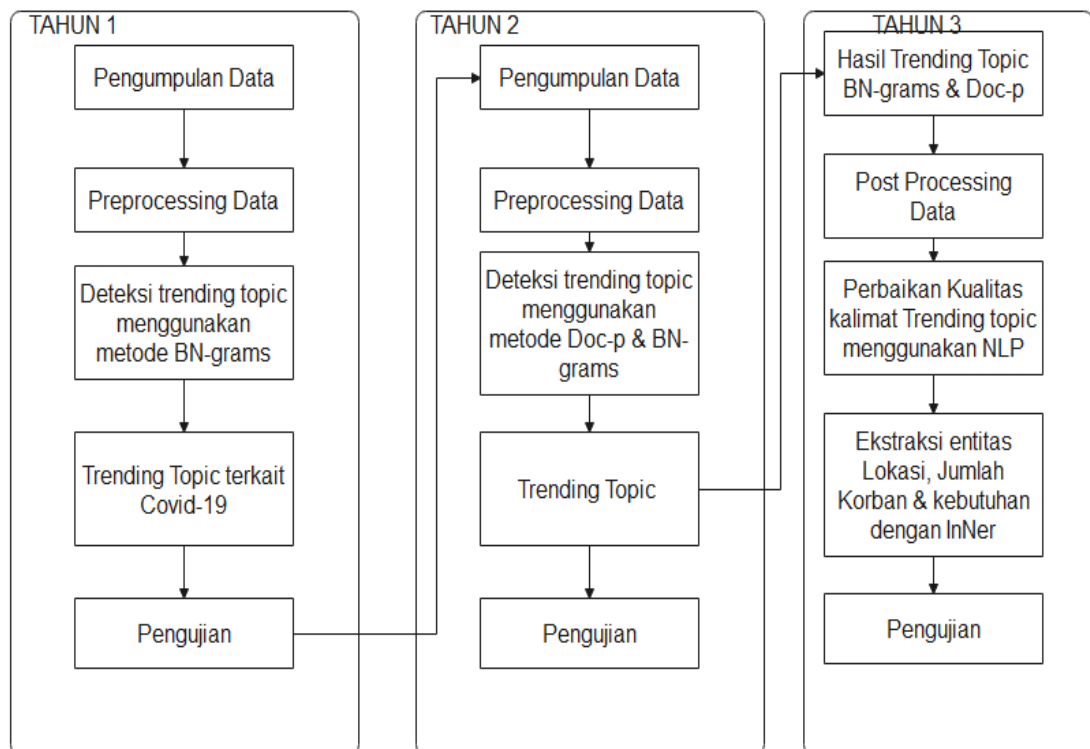
2.3. State of The Art.

Titik tolak penelitian yang dilakukan penulis berdasarkan dari penelitian berbasis *textual content* dengan menggunakan metode BN-Grams dan Doc-p pada Twitter dilakukan oleh (Aiello et al., 2013). Pada penelitian dengan metode BN-Grams dilakukan studi komparasi dari enam metode terkait deteksi *trending topic*. Metode yang paling akurat untuk deteksi *trending topic* pada penelitian Aiello et al. (2013) adalah metode BN-Grams.

Hasil studi komparasi yang dilakukan oleh Aiello et al. (2013) metode BN-Grams dan Doc-p memiliki keakuratan deteksi *trending topic* berbahasa Inggris paling tinggi pada dataset FA Cup tahun 2012. Metode BN-Grams memiliki nilai keakuratan 0.7692 dibandingkan dengan metode LDA dengan nilai 0.693, metode Doc-p dengan nilai 0.7692, metode GFeat-p dengan nilai 0.000, metode FPM dengan nilai 0.3077 dan metode SFPM dengan nilai 0.6154. Oleh karena itu, diharapkan dengan mengusulkan dua metode tersebut mempercepat dan meningkatkan keakuratan dalam deteksi *trending topic* di Twitter

METODE, JANGKA WAKTU DAN LOKASI PENELITIAN

Metode penelitian yang diusulkan dalam penelitian ini menggunakan durasi waktu tiga tahun seperti ditunjukkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

Rincian tahapan, Langkah penelitian, indikator, hasil pencapaian dan lokasi kegiatan dari setiap tahapan penelitian dijabarkan pada Tabel 1. Realisasi dari tahapan penelitian diwujudkan dalam jadwal kegiatan seperti terlihat pada Tabel 2.

Tabel 1. Detil Tahapan Penelitian

Tahun	Tahapan	Langkah	Indikator	Hasil	Lokasi
1	Pengumpulan data	<ul style="list-style-type: none"> - Menentukan kata kunci penarikan data tweet terkait Covid-19 - Penarikan data Twitter dengan Twitter API 	<ul style="list-style-type: none"> - Didefinisikannya kata kunci terkait Covid-19 - Mendapatkan token key dan secret key dari Twitter 	<ul style="list-style-type: none"> - Kata kunci berdasarkan pendapat dari pakar - Token dan secret key sudah didapatkan 	Lab Data Mining Universitas Budi Luhur

		- Mengumpulkan data Twitter dalam database	- Aplikasi dapat mengumpulkan data Twitter dengan langsung tersimpan di database	- Data Twitter tersimpan dalam database MySQL	
	Tahapan	Langkah	Indikator	Hasil	Lokasi
	Preprocessing data	- Tokenisasi - Stemming - Agregasi	- Setiap tweet dipecah menjadi token dan dilakukan penghilangan stop word, tanda baca, mention, karakter berulang dan menghilangkan slangword - Setiap kata dihilangkan awalan dan akhirnya - Menggabungkan beberapa tweet menjadi satu dokumen besar	- Setiap tweet menjadi token/kata - Kata dasar didapatkan - Dataset yang berisi sekumpulan tweet berdasarkan kedekatan waktu dan kesamaan teks	Lab Data Mining Universitas Budi Luhur
	Deteksi trending topic menggunakan BN-grams	- Pembobotan setiap term dengan DFIDFT - Klasterisasi tweet - Topic ranking	- Setiap term atau word memiliki skor DFIDFT - Terbentuk kluster yang berisi sekumpulan term - Pemeringkatan kluster	- Bobot setiap term dengan DFIDFT - Setiap Kluster merepresentasikan satu topik - Kluster dengan ranking pertama menjadi trending topic	Lab Data Mining Universitas Budi Luhur
	Trending topic terkait Covid-19	- Ekstraksi trending topic dari setiap kluster	- Terbentuk beberapa kandidat trending topic	- Top 10 Trending topic	Lab Data Mining Universitas Budi Luhur
	Pengujian	- Mempersiapkan ground truth pengukuran topic recall - Keyword precision - Keyword recall	- Berita populer dari kompas.com dan detik.com dikumpulkan - Nilai topic recall, keyword precision dan keyword recall setiap trending topic	- Daftar ground truth terbentuk - Akurasi, precisi dan recall dari metode BN-grams	Lab Data Mining Universitas Budi Luhur dan BNPB

2	Pengumpulan data	<ul style="list-style-type: none"> - Menentukan kata kunci penarikan data tweet terkait Covid-19 - Penarikan data Twitter dengan Twitter API - Mengumpulkan data Twitter dalam database 	<ul style="list-style-type: none"> - Didefinisikannya kata kunci terkait Covid-19 - Mendapatkan token key dan secret key dari Twitter - Aplikasi dapat mengumpulkan data Twitter dengan langsung tersimpan di database 	<ul style="list-style-type: none"> - Kata kunci berdasarkan pendapat dari pakar - Token dan secret key sudah didapatkan - Data Twitter tersimpan dalam database MySQL 	Lab Data Mining Universitas Budi Luhur
	Tahapan	Langkah	Indikator	Hasil	Lokasi
	Preprocessing data	<ul style="list-style-type: none"> - Tokenisasi - Stemming - Agregasi 	<ul style="list-style-type: none"> - Setiap tweet dipecah menjadi token dan dilakukan penghilangan stop word, tanda baca, mention, karakter berulang dan menghilangkan slangword - Setiap kata dihilangkan awalan dan akhirnya - Menggabungkan beberapa tweet menjadi satu dokumen besar - 	<ul style="list-style-type: none"> - Setiap tweet menjadi token/kata - Kata dasar didapatkan - Dataset yang berisi sekumpulan tweet berdasarkan kedekatan waktu dan kesamaan teks 	Lab Data Mining Universitas Budi Luhur
Deteksi trending topic menggunakan BN-grams & Doc-p	<p>BN-grams:</p> <ul style="list-style-type: none"> - Pembobotan setiap term dengan DFIDFT - Klasterisasi tweet - Topic ranking <p>Doc-p:</p> <ul style="list-style-type: none"> - Pembuatan kamus kata - Pembuatan signature bit array - Collide tweets - Cosine similarity - Pembentukan klaster - Klaster ranking 	<ul style="list-style-type: none"> - Setiap term atau word memiliki skor DFIDFT - Terbentuk klaster yang berisi sekumpulan term - Pemeringkatan klaster BN-grams - Pemeringkatan klaster Doc-p 	<ul style="list-style-type: none"> - Bobot setiap term dengan DFIDFT - Setiap Klaster merepresentasikan satu topik - Klaster dengan ranking pertama pada BN-grams menjadi trending topic - Klaster dengan peringkat pertama pada Doc-p menjadi trending topic 	Lab Data Mining Universitas Budi Luhur	

	Trending topic terkait Covid-19	- Ekstraksi trending topic dari setiap klaster	- Terbentuk beberapa kandidat trending topic	- Top 10 Trending topic	Lab Data Mining Universitas Budi Luhur
	Pengujian	- Mempersiapkan ground truth - pengukuran topic recall - Keyword precision - Keyword recall	- Berita populer dari kompas.com dan detik.com dikumpulkan - Nilai topic recall, keyword precision dan keyword recall setiap trending topic	- Daftar ground truth terbentuk - Akurasi, precisi dan recall dari metode BN-grams	Lab Data Mining Universitas Budi Luhur dan BNPB
Tahun	Tahapan	Langkah	Indikator	Hasil	Lokasi
3	Hasil trending topic BN-grams & Doc-p	- Mengumpulkan hasil trending topic BN-grams dan Doc-p	- List trending topic BN-grams dan Doc-p	- Daftar trending topic yang sudah valid	Lab Data Mining Universitas Budi Luhur
	Post Processing data	- Pengecekan bentuk kalimat trending topic - Pengecekan makna trending topic	- Bentuk kalimat sudah sesuai EYD - Setiap trending topic memiliki makna	- Kalimat trending topic - Makna trending topic	Lab Data Mining Universitas Budi Luhur
	Perbaikan kualitas trending topic menggunakan NLP	- Memperbaiki kalimat trending topic dengan mengecek tweet aslinya - Trending topic hasil modifikasi kalimat	- Kalimat trending topic sudah berpola Subjek , predikat dan objek	- Trending topic sudah berpola SPOK atau EYD - Trending topic dapat dimanfaatkan kalimatnya	Lab Data Mining Universitas Budi Luhur
	Ekstraksi entitas Lokasi, Kebutuhan dan jumlah korban dari data tweet	- Mengecek bentuk kalimat trending topic - Membangun rule ekstraksi entitas - Ekstraksi entitas	- Trending topic minimal berisi SPOK - Rule ekstraksi entitas	- Entitas Lokasi, jumlah korban dan kebutuhan setiap tweet	Lab Data Mining Universitas Budi Luhur
	Pengujian	- Mempersiapkan ground truth - pengukuran topic recall - Keyword precision - Keyword recall	- Berita populer dari kompas.com dan detik.com dikumpulkan - Nilai topic recall, keyword precision dan keyword recall setiap trending topic	- Daftar ground truth terbentuk - Akurasi, precisi dan recall dari metode BN-grams	Lab Data Mining Universitas Budi Luhur

Tabel 2. Jadwal Kegiatan Penelitian

No	Nama Kegiatan	Bulan					
		1	2	3	4	5	6
1	Pengumpulan data	x					
2	Preliminary dataset analysis	x					
3	Preprocessing data	x					
4	Pembuatan aplikasi Deteksi trending topic menggunakan BN-grams dan Doc-p		x				
5	Implementasi dan ujicoba aplikasi deteksi trending topic		x				
6	Perbaikan kualitas trending topic menggunakan NLP			x			
7	Ekstraksi entitas lokasi, jumlah korban dan kebutuhan korban bencana alam			x			
8	Verifikasi dan validasi				x		
9	Pengujian				x		
10	Penulisan paper penelitian	x	x	x	x		
11	Proof read paper					x	
12	Submit paper					x	
13	Publikasi penelitian						x

SASARAN DAN TARGET PENELITIAN

Sasaran dan target penelitian ini dijelaskan dengan detail dibawah ini:

Sasaran Penelitian ini adalah:

1. Membantu Dinas Kesehatan di Pemprov DKI Jakarta dalam memetakan informasi yang valid dan faktual dari media social terutama Twittter terutama terkait Covid-19.
2. Membantu Gugus Tugas Covid-19 dan Dinas Sosial di Pemprov DKI Jakarta dalam memetakan informasi dengan melakukan ekstraksi lokasi, jumlah korban dan kebutuhan terdampak pandemic Covid-19 dari data media social terutama Twitter. Selain itu, untuk melihat monitoring penanganan dan pendistribusian bantuan Covid-19 kepada warga DKI Jakarta sehingga bantuan dapat termonitoring dengan cepat, faktual dan tepat sasaran.

Target Penelitian ini adalah:

1. Terwujudnya model deteksi trending topic dalam bentuk sistem informasi yang berisi dashboard pemetaan informasi dari media social terkait Covid-19 terutama terkait pemetaan OTG, ODP, PDP, Suspek, Probable, konfirmasi, kontak erat, pelaku perjalanan, Discarded, selesai isolasi dan kematian sehingga membantu Dinas Kesehatan Pemprov DKI Jakarta dalam penanganan Covid -19.
2. Terwujudnya model dan ekstraksi informasi dari trending topic yang berisi ekstraksi entitas lokasi, kebutuhan dan jumlah korban terdampak dari Covid-19 yang tersebar di media sosial. Informasi ini menjadi salah satu alternatif dan pelengkap dari Gugus Tugas Covid-19 di Wilayah DKI Jakarta dan Dinas Sosial Pemprov DKI Jakarta dalam penanganan dan monitoring Covid-19.

LOKASI PENELITIAN

Lokasi Penelitian Covid-19 dengan Pemprov DKI Jakarta ini membutuhkan Dinas terkait untuk pengambilan data, konsultasi pakar dan pengujian serta implementasi hasil penelitian pada:

1. Dinas Kesehatan DKI Jakarta, dengan alamat: Jl. Kesehatan No. 10, Kel. Petojo Selatan, Kec. Gambir, Jakarta Pusat Jakarta 10510.
2. Dinas Sosial DKI Jakarta, dengan alamat: Jl. Gunung Sahari II No.6, RT.13/RW.7, Gn. Sahari Sel., Kec. Kemayoran, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10610.
3. Jakarta Development Collaboration Network, dengan alamat: Jl. Medan Merdeka Selatan No. 8-9, Gambir, Jakarta Pusat.
4. Posko Tim Tanggap Covid-19 Dinas Kesehatan DKI Jakarta, dengan alamat: Jl. Kesehatan No. 10, Kel. Petojo Selatan, Kec. Gambir, Jakarta Pusat Jakarta 10510.
5. Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta, dengan alamat: Jl. Kyai H. Zainul Arifin No.71, RT.7/RW.10, Duri Pulo, Kecamatan Gambir, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10140.

HASIL YANG DIHARAPKAN

Hasil yang diharapkan dari penelitian ini adalah:

1. Metode BN-GRAMS dan Doc-p dapat menghasilkan trending topic terkait informasi Covid-19 yang terjadi di Indonesia terutama di wilayah DKI Jakarta
2. Trending topic yang dihasilkan dikelompokkan menjadi 3 kluster yakni kluster pengangguran, kluster pemetaan informasi Covid-19 seperti pemetaan OTG, ODP atau PDP dan kluster pemberlakuan PSBB (Lockdown).
3. Kualitas trending topic yang dihasilkan semakin akurat dengan penerapan Metode BN-grams dan Doc-p.
4. Perbaikan kualitas trending topic dari dua metode usulan semakin mudah dipahami bentuk kalimatnya dan sesuai dengan EYD menggunakan pendekatan Natural Language Processing (NLP).
5. Ekstraksi Lokasi, jumlah korban dan kebutuhan terdampak dari pandemic Covid-19 pada suatu wilayah tertentu berdasarkan informasi trending topic di Twitter terutama di wilayah DKI Jakarta.
6. Adanya Model deteksi trending topic dan ekstraksi lokasi, jumlah korban dan kebutuhan terdampak Covid-19 di Pemprov DKI Jakarta sehingga dapat membantu Dinas Kesehatan, Dinas Sosial dan Gugus Tugas Covid-19 di wilayah DKI Jakarta.

DAFTAR PUSTAKA

- Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., Member, S., IEEE, & Jaimes, A. (2013). Sensing Trending Topics in Twitter. *IEEE Transactions on Multimedia* (2013), 15(6), 1268–1282.
- Benhardus, J. (2010). Streaming trend detection in twitter. In *2010 UCCS Reu For Artificial Intelligence, Natural Language Processing And Information Retrieval Final Report Artificial Intelligence, Natural Language Processing And Information Retrieval Final Report* (Vol. 9, Issue 1). <https://doi.org/10.1504/IJWBC.2013.051298>
- Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022. <http://dl.acm.org/citation.cfm?id=944937>
- Cataldi, M., Torino, U., Caro, L. Di, & Schifanella, C. (2010). Emerging Topic Detection on Twitter based on Temporal and Social Terms Evaluation. *MDMKDD '10 Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 10. <https://doi.org/10.1145/1814245.1814249>
- Chen, J., Yu, J., & Shen, Y. (2012). Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 153–157. <https://doi.org/10.1109/WI-IAT.2012.31>
- Cvijikj, I. P., & Michahelles, F. (2011). Monitoring Trends on Facebook. *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, 895–902. <https://doi.org/10.1109/DASC.2011.150>
- Ge, G., Chen, L., & Du, J. (2013). The research on topic detection of microblog based on TC-LDA. *2013 15th IEEE International Conference on Communication Technology*, 722–727. <https://doi.org/10.1109/ICCT.2013.6820469>
- Guzman, J., & Poblete, B. (2013). On-line relevant anomaly detection in the Twitter stream: an efficient bursty keyword detection model. *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, 31–39. <https://doi.org/http://dx.doi.org/10.1145/2500853.2500860>
- Kleinberg, J. (2002). Bursty and hierarchical structure in streams. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining - KDD '02, 91. <https://doi.org/10.1145/775060.775061>

- Mathioudakis, M., & Koudas, N. (2010). Twittermonitor: trend detection over the twitter stream. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, 1155–1158. <https://doi.org/10.1145/1807167.1807306>
- Matsui, T., Tamura, K., & Kitakami, H. (2013). Location-based burst detection algorithm for georeferenced document streams based on user's moving direction. *2013 IEEE 6th International Workshop on Computational Intelligence and Applications (IWCIA)*, 57–62. <https://doi.org/10.1109/IWCIA.2013.6624784>
- Phuvipadawat, S., & Murata, T. (2010). Breaking News Detection and Tracking in Twitter. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 120–123. <https://doi.org/10.1109/WI-IAT.2010.205>
- Song, M., & Kim, M. C. (2013). RT 2 M : Real-time Twitter Trend Mining System. *2013 International Conference on Social Intelligence and Technology*, 64–71. <https://doi.org/10.1109/SOCIETY.2013.19>
- Takahashi, T., Tomioka, R., & Yamanishi, K. (2011). Discovering Emerging Topics in Social Streams via Link Anomaly Detection. *2011 IEEE 11th International Conference on Data Mining*, 1230–1235. <https://doi.org/10.1109/ICDM.2011.53>
- Takahashi, T., Tomioka, R., & Yamanishi, K. (2014). Discovering Emerging Topics in Social Streams via Link-Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 120–130. <https://doi.org/10.1109/TKDE.2012.239>
- Tamura, S., Tamura, K., Kitakami, H., & Hirahara, K. (2012). Clustering-based burst-detection algorithm for web-image document stream on social media. *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 703–708. <https://doi.org/10.1109/ICSMC.2012.6377809>
- Xie, L., Natsev, A., He, X., Kender, J. R., Hill, M., & Smith, J. R. (2013). Tracking Large-Scale Video Remix in Real-World Events. *IEEE Transactions on Multimedia*, 15(6), 1244–1254. <https://doi.org/10.1109/TMM.2013.2264929>

LAMPIRAN 4 BIODATA TIM PENELITIAN

Tim peneliti adalah usulan penelitian ini terdiri dari 4 anggota. Satu menjadi ketua peneliti dan tiga menjadi anggota peneliti. Satu anggota atas nama Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng dari Institut Teknologi Surabaya sebagai pendamping dan pengarah dalam penelitian ini sekaligus sebagai Profesor Tamu pada Universitas Budi Luhur.

LAMPIRAN 2 BIODATA TIM PENELITIAN

A. Identitas Diri Ketua Tim Peneliti

1. Nama Lengkap : Indra
2. Jenis Kelamin : Laki-laki
3. Jabatan Fungsional : Lektor
4. NIP : 080062
5. Tempat, Tanggal Lahir : Banyumas, 22 Maret 1986
6. E-mail : indra@budiluhur.ac.id
7. Nomor Handphone : 08568287734
8. Alamat : Jl.Kutilang 2 Komp.Pajak

B. Riwayat Pendidikan

	S1	S2	S3
Nama Perguruan Tinggi	Universitas Budi Luhur	Universitas Indonesia	Universitas Gadjah Mada
Bidang Ilmu	Teknik Informatika	MTI	Ilmu Komputer
Tahun Masuk – Lulus	2004 – 2008	2009 – 2011	2013-2018

C. Pengalaman Penelitian (5 Tahun Terakhir)

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber *	Jumlah (Rp.)
1.	2012	Model <i>Extract, Transform dan Load</i> (ETL) Pada <i>Data Warehouse</i> untuk Pelaporan Evaluasi Program Studi Berdasarkan Evaluasi Diri (EPSBED) Menggunakan Model	Universitas Budi Luhur	14.000.000

		<i>Data Warehouse</i> EPSBED: Studi Kasus Universitas Budi Luhur		
2	2013	Model <i>Extract, Transform dan Load</i> (ETL) Pada <i>Data Warehouse</i> untuk Borang Akreditasi Program Studi Tingkat Sarjana: Studi Kasus Universitas Budi Luhur	Universitas Budi Luhur	8.000.000
3	2012	Sistem Pengenalan Wajah Dengan Metode Eigenface Untuk Absensi Pada Pt Florindo Lestari	Mandiri	4.000.000
4	2012	Data Warehouse For Study Program Evaluation Reporting Based on Self Evaluation (EPSBED) using EPSBED Data Warehouse Model: Case Study Budi Luhur University	Mandiri	6.000.000
5	2011	Perancangan Serta Penerapan Absensi Via Bluetooth Pada Fakultas	Mandiri	3.000.000

		Teknologi Informasi Universitas Budi Luhur		
6	2010	<i>Mobile Information</i> pada Rumah Sakit dengan Menggunakan Metode <i>Wireless</i> <i>Messaging Api</i> (WMA)	Mandiri	3.000.000
7	2012	Sistem Informasi Penerimaan Siswa Baru Pada Madrasah Ibtidaiyah Dan Tsanawiyah Nurul Falah	Mandiri	4.000.000
8	2008	Otomasi Sistem Penapakan Tenaga Surya	Mandiri	5.000.000
9	2013	Perancangan serta Penerapan Datawarehouse untuk Borang Akreditasi	Universitas Budi Luhur	12.000.000
10	2013	Perancangan serta Penerapan Perhitungan Suara Pemilihan Kepala Daerah dan Wakil Kepala Daerah (Pilkada) dengan Metode Quick Count Berbasis SMS Studi Kasus: Kecamatan Larangan	Universitas Budi Luhur	8.000.000

10	2014	Pengenalan dan Identifikasi Plat Nomor Kendaraan Berbasis Android dengan Algoritma Static Classifier dan Adaptive Classifier pada Bagian Keamanan Universitas Budi Luhur	Universitas Budi Luhur	8.840.000
11	2014	Perancangan dan Pengolahan Bahasa Alami untuk Info Gempa Bumi Terkini dengan Sumber Data pada Twitter @InfoBMKG	Universitas Budi Luhur	13.750.000
12	2014	Sistem Informasi Penggajian Pada Koperasi Ancol Sayang Lingkungan	Mandiri	7.000.000
13	2015	Perancangan dan Pengolahan Bahasa Alami untuk Informasi Cuaca Penerbangan Studi Kasus: BMKG	Universitas Budi Luhur	11.000.000

D. Identitas Diri Anggota Peneliti 1

1. Nama Lengkap : Prof. Dr. Ir. Mauridhi Hery Purnomo
2. Jenis Kelamin : Laki-laki
3. Jabatan Fungsional : Guru Besar
4. NIDN : 0016095811
5. E-mail : hery@ee.its.ac.id
6. Nomor Handphone : 08113484500

E. Pengalaman Penelitian (5 Tahun Terakhir)

Berada pada Lampiran akhir

F. Identitas Diri Anggota Peneliti 2

1. Nama Lengkap : Agus Umar Hamdani, M.Kom
2. Jenis Kelamin : Laki-laki
3. Jabatan Fungsional : Lektor
4. NIP : 080055
5. Tempat, Tanggal Lahir : Sumberrejo, 09 Agustus 1983
6. E-mail : agus.umarhamdani@budiluhur.ac.id
7. Nomor Handphone : 087809295666
8. Alamat : Jl. Ciledug Indah II D1/9 RT 003 RW 009
Pedurenan Karang Tengah Kota
Tangerang 15157

G. Riwayat Pendidikan

	S1	S2	S3
Nama Perguruan Tinggi	Universitas Budi Luhur	Universitas Budi Luhur	
Bidang Ilmu	Sistem Informasi	Magister Ilmu Komputer	
Tahun Masuk – Lulus	2004 – 2008	2009 - 2011	

H. Pengalaman Penelitian (5 Tahun Terakhir)

Pengalaman Penelitian				
No	Tahun	Judul Penelitian	Pendanaan	
			Sumber*	Jumlah (Rp)
1.	2012	Knowledge Sharing untuk mendukung akreditasi program studi Ilmu Komunikasi Fakultas Ilmu Komunikasi	Universitas Budi Luhur	6.000.000,-
2.	2013	Sistem Informasi Pengadaan Barang Studi Kasus : PT. Sinar Elektronika SEB	Mandiri	4.500.000,-
3.	2014	Pemodelan Sistem Informasi Administrasi Pendistribusian Kartu Asuransi Akda Extra Studi Kasus : PT. Asuransi Bhakti Bhayangkara Jakarta	Mandiri	4.500.000,-
4.	2015	Implementasi Sistem Informasi Polis Asuransi Kendaraan Bermotor Studi Kasus : PT. Asuransi Umum Mega Jakarta Selatan.	Mandiri	5.500.000,-
5.	2016	Sistem Pelanggaran Tata Tertib Kampus	Universitas Budi Luhur	10.784.000,-

		Menggunakan Metode <i>Fuzzy Multiple Attribute Decision Making</i> (FMADM) dan <i>Simple Additive Weighting</i> (SAW) Studi Kasus : Universitas Budi Luhur		
6.	2018	Pengembangan Model Sistem Electronic Customer Relationship Management (E-CRM) Menggunakan Pendekatan Work Centered Analysis dan Framework Dynamic Of CRM Guna Membangun Loyalitas Customer Studi Kasus : PT. XYZ	Mandiri	4.500.000,-
7.	2018	Model E-Commerce Customer Relationship Management Dengan Pendekatan Work Centered Analysis dan Framework Of Dynamic CRM Untuk Meningkatkan Layanan Pelanggan Pada PT. XYZ	Mandiri	4.500.000,-
8.	2019	Model E-Commerce Dengan Metode Web Engineering Method	Mandiri	4.500.000,-

		Untuk Menunjang Pemasaran Produk Pada XYZ Pet Shop		
9.	2019	Perancangan Model Sistem Informasi Manajemen Menggunakan Data Warehouse Untuk Analisis Jasa Pengiriman Barang Studi Kasus : PT. XYZ.	Mandiri	4.500.000,-

I. Identitas Diri Anggota Peneliti 3

1. Nama Lengkap : Dr. M.Syafrullah, M.Kom, M.Sc
2. Jenis Kelamin : Laki-laki
3. Jabatan Fungsional : Lektor
4. NIDN : 0316097401
5. E-mail : mohammad.syafrullah@budiluhur.ac.id

J. Riwayat Pendidikan

	S1	S2	S3
Nama Perguruan Tinggi	Universitas Budi Luhur	Universitas Swiss German	Universiti Teknologi Malaysia
Bidang Ilmu	Teknik Informatika	Magister Ilmu Komputer	Doktor Ilmu Komputer
Tahun Masuk – Lulus	1992 – 1996	2003 - 2005	2010-2015

K. Pengalaman Penelitian (5 Tahun Terakhir)

Pengalaman Penelitian			
No	Tahun	Judul Penelitian	Pendanaan
			Sumber*

1.	2017	Content validity of critical success factors for e-Government implementation in Indonesia		
2.	2018	Analysis of user readiness toward ICT usage at small medium enterprise in south tangerang		
3.	2018	SKF-Based Image Template Matching for Distance Measurement by Using Stereo Vision		
4.	2018	Testing of technology readiness index model based on exploratory factor analysis approach		
5.	2018	Motion Tracker Based Wheeled Mobile Robot System Identification and Controller Design		

**L. Pengalaman Penelitian (5 Tahun Terakhir) Prof.Dr.Ir. Maudhi Hery Purnomo,
M.Eng**