

Real-Time Sentiment Analysis for Political Figures from Youtube

IF5270 – Big Data System

Oleh:

Arief Purnama Muharram (23521013)

Danendra Athallariq Harya P (23522020)



**PROGRAM STUDI MAGISTER INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG**

Daftar Isi

Daftar Isi	1
I. Latar Belakang Persoalan	2
II. Persoalan Project	2
III. Justifikasi Pentingnya Persoalan Diselesaikan	3
IV. Tantangan dari Sudut Pandang: Big Data Processing dan Analytics	3
V. Big Picture dari Sistem yang Dikembangkan	3
VI. Infrastruktur Teknologi Big Data yang Digunakan	4
VII. Rencana Peran Masing-Masing Anggota Kelompok	5
VIII. Usulan Cara/Matrik Evaluasi Kualitas Hasil Pemrosesan/Analisis	5
IX. Referensi	5

I. Latar Belakang Persoalan

Pemilihan Umum (Pemilu) merupakan salah satu perwujudan nyata dari suatu sistem demokrasi. Melalui Pemilu, masyarakat berkesempatan untuk turut serta dalam pemerintahan dengan memilih wakil-wakilnya yang akan mewakili mereka di dalam struktur pemerintahan. [1] Di antara berbagai jenis Pemilu, Pemilihan Umum Presiden (Pilpres) selalu menjadi momen yang sangat dinanti dan dijuluki sebagai "pesta demokrasi" terbesar. Pada tahun 2024, Indonesia akan menggelar Pilpres untuk menentukan sosok calon presiden yang akan memimpin Indonesia selama 5 tahun ke depan.

Menyambut Pilpres 2024, setiap partai politik saling berlomba dalam menentukan calon presiden dan wakil presiden terbaik yang akan diusung. Bagi partai politik, Pilpres tidak hanya sekadar menyangkut posisi Presiden dan Wakil Presiden, tetapi juga menentukan kursi mereka di dalam struktur pemerintahan ke depan. Oleh karena itu, sangat penting bagi partai politik untuk menyusun strategi kampanye politik terbaik demi merebut hati masyarakat. Salah satu upaya yang dapat dilakukan oleh partai politik untuk menilai kualitas kandidat yang diusung yaitu melalui analisis sentimen publik.

Analisis sentimen publik merupakan salah satu alat bagi partai politik untuk memperoleh pemahaman yang lebih baik tentang opini dan pandangan masyarakat terhadap calon politik yang diusung. Dengan analisis sentimen publik, partai politik dapat merancang strategi kampanye yang efektif dan responsif terhadap kebutuhan masyarakat.

Dalam era digital seperti sekarang, media sosial menjadi wadah bagi masyarakat untuk mengekspresikan berbagai hal, termasuk pandangan terhadap calon politik. Ekspresi tersebut dapat berupa dukungan atau penolakan, dan dapat diekspresikan dalam berbagai media seperti tulisan, audio, atau video. Ekspresi yang demikian dapat dijadikan indikator sentimen publik bagi partai politik. [2] [3]

Proyek ini bertujuan untuk mengembangkan sistem analisis sentimen *real-time* terhadap calon politik Pilpres 2024 dari video bernarasi politik di YouTube. Pemilihan topik video di YouTube didasarkan pada kenyataan bahwa *platform* YouTube telah menjadi salah satu sarana ekspresi masyarakat terhadap sosok politik melalui unggahan video yang beragam, mulai dari satir hingga dukungan atau penolakan secara langsung. Dengan sistem ini, diharapkan partai politik dapat memperoleh gambaran dinamika sentimen masyarakat secara *real-time* terhadap calon politik yang ada di *platform* YouTube.

II. Persoalan Project

Bagaimana memanfaatkan data dari berbagai unggahan video di *platform* YouTube untuk menghasilkan analisis sentimen *real-time* terhadap calon politik Pilpres 2024?

III. Justifikasi Pentingnya Persoalan Diselesaikan

Berikut merupakan justifikasi pentingnya persoalan dalam usulan ini diselesaikan.

1. Pilpres 2024 merupakan ajang pesta demokrasi yang penting di mana setiap partai politik perlu memiliki instrumen untuk menilai kualitas calon politiknya di mata publik.
2. Sebagai salah satu *platform* media sosial terbesar, YouTube telah menjadi tempat bagi masyarakat untuk berekspresi dan memberikan respons terhadap calon-calon politiknya. Oleh karena itu, memanfaatkan data dari unggahan video politik di YouTube untuk menganalisis sentimen publik dapat memberikan informasi berharga bagi partai politik dalam merancang strategi kampanye politik yang efektif.

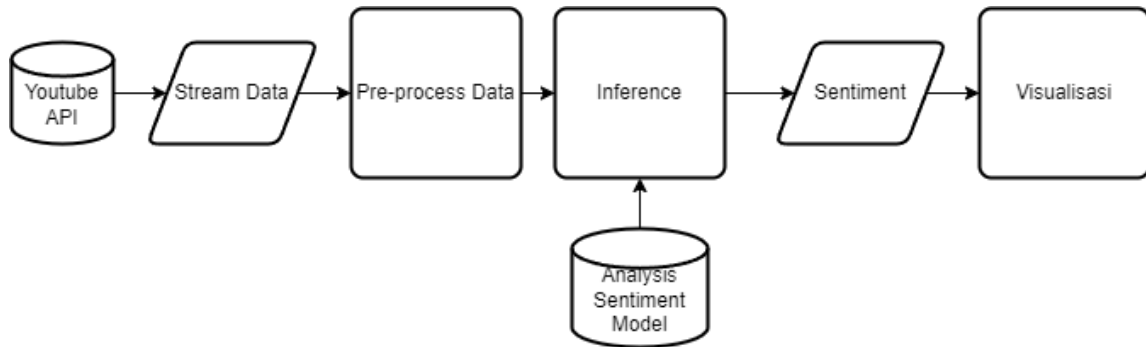
IV. Tantangan dari Sudut Pandang: *Big Data Processing* dan *Analytics*

Berikut merupakan tantangan dalam menyelesaikan persoalan dari sudut pandang *big data processing* dan *analytics*.

1. Bagaimana melakukan *streaming* data secara *real-time* dari *platform* YouTube ke sistem yang dikembangkan?
2. Bagaimana menentukan definisi *real-time* yang tepat (jam, harian, mingguan, dwi mingguan, bulanan)?
3. Bagaimana melakukan penyaringan video yang relevan dengan konteks yang tepat, yaitu "Calon Pasangan Presiden dan Wakil Presiden 2024"?
4. Bagaimana merancang dashboard sederhana untuk memvisualisasikan hasil analisis sentimen yang dihasilkan?

V. *Big Picture* dari Sistem yang Dikembangkan

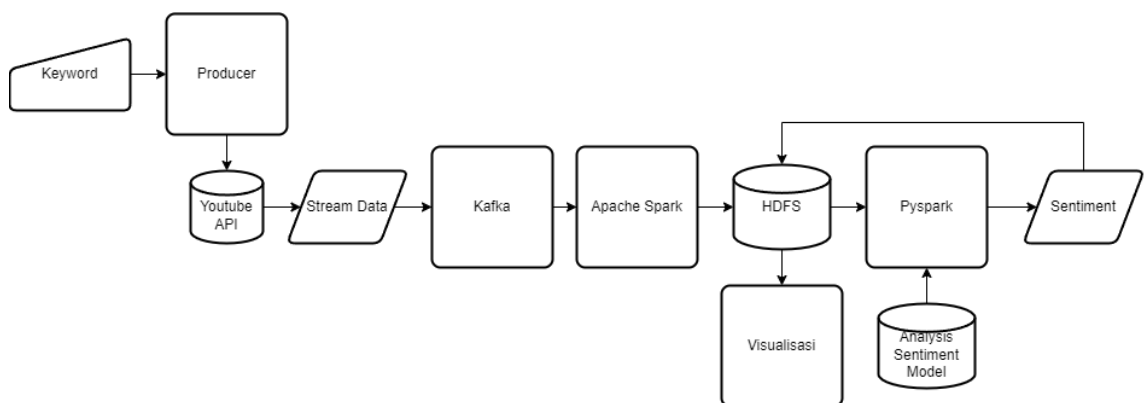
Sistem yang akan dikembangkan akan mengekstraksi sentimen dari data *stream* YouTube yang berkaitan dengan kueri tertentu. Diagram alir dari sistem dapat dilihat pada gambar di bawah ini. Pertama, data *stream* akan diambil dari *database* YouTube menggunakan API yang disediakan berdasarkan kata kunci yang terkait dengan figur politik. Kata kunci ini akan ditentukan terlebih dahulu. *Data stream* yang diperoleh dari YouTube API akan berbentuk JSON. Dari data ini, hanya judul dan deskripsi video YouTube yang akan diekstraksi untuk dilakukan inferensi oleh model analisis sentimen yang telah dilatih sebelumnya. Hasil akhirnya akan berupa persentase sentimen dari semua video YouTube yang terkait dengan tokoh politik tertentu. Persentase tersebut akan divisualisasikan berdasarkan waktu sehingga diperoleh tren sentimen dari masing-masing tokoh politik seiring waktu.



Gambar 1. Big picture dari sistem yang dikembangkan

VI. Infrastruktur Teknologi *Big Data* yang Digunakan

Untuk menjelaskan sistem yang akan digunakan, diagram alir dapat dilihat di bawah ini. *Producer* merupakan program python yang menerima input kata kunci dari pengguna yang terkait dengan tokoh politik tertentu. Kemudian, dengan menggunakan YouTube API, *producer* akan mengambil data dari semua video yang terkait dengan kata kunci tersebut. *Producer* akan mengambil *stream* data setiap satu jam dengan menggunakan *scheduling* yang telah diatur sebelumnya. *Stream* data tersebut akan dimasukkan ke dalam Apache Kafka dan di-*pre-process* dengan menggunakan Apache Spark. Di Apache Spark, *data stream* berformat JSON ini akan diekstrak hanya pada bagian judul dan deskripsi video dan digabungkan dengan waktu data diambil. Data hasil *pre-process* ini kemudian disimpan di HDFS sebagai penyimpanan historis data yang dapat digunakan kembali di waktu lain atau saat didapatkan model analisis sentimen yang lebih baik. Data yang disimpan di HDFS akan ditarik dengan menggunakan PySpark untuk melakukan inferensi model analisis sentimen sehingga akan menghasilkan sentimen dari deskripsi video. Hasil sentimen ini kemudian akan disimpan kembali di HDFS untuk dapat divisualisasikan menjadi trend sentimen tokoh politik terhadap waktu.



Gambar 2. Infrastruktur teknologi *big data* yang digunakan

VII. Rencana Peran Masing-Masing Anggota Kelompok

Nama	Peran
Arief Purnama Muharram	<ul style="list-style-type: none">● Mengambil <i>data stream</i> dari YouTube API dan menghubungkan dengan Apache Kafka.● Visualisasi hasil sentimen dari model.
Danendra Athallariq Harya P	<ul style="list-style-type: none">● <i>Pre-process</i> data dari <i>data stream</i> menggunakan Apache Spark.● Pembuatan model analisis sentimen.

VIII. Usulan Cara/Matrik Evaluasi Kualitas Hasil Pemrosesan/Analisis

Evaluasi akan dilakukan pada tahap pelatihan model analisis sentimen menggunakan matriks akurasi dan F1-score pada data uji.

IX. Referensi

1. Suryanto. 2018. Pengantar Ilmu Politik. Bandung: Pustaka Setia. Halaman 256.
2. M. Arham Aziz, "Analisis Sentimen pada Data Big Political Data untuk Pemilihan Umum 2019 di Indonesia", Jurnal Informatika, vol. 14, no. 1, 2020, pp. 11-20.
3. Nuryanto, B., & Rachmawati, T. (2019). Pengukuran Sentimen Politik Dalam Pemilihan Umum Melalui Media Sosial YouTube. Jurnal Sistem Informasi Bisnis, 3(2), 96-102. <https://doi.org/10.32736/jsib.v3i2.123>