

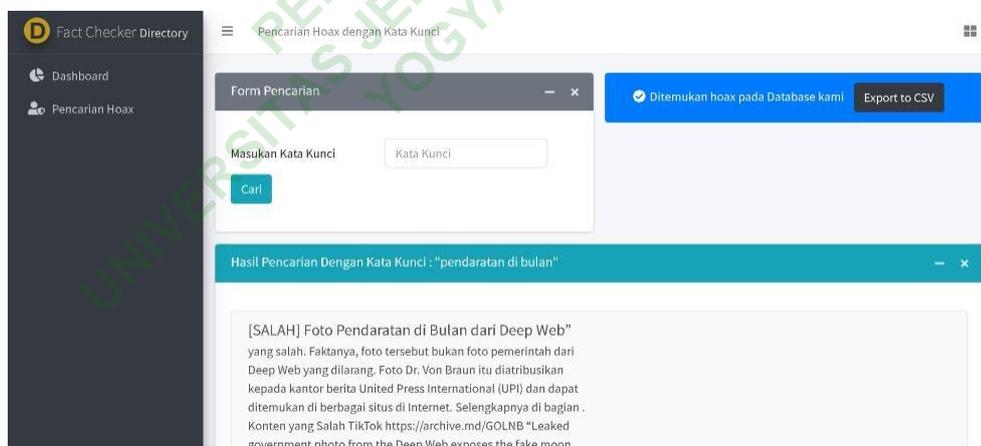
BAB 3

METODE PENELITIAN

Penelitian ini adalah penelitian rancang-bangun. Penelitian berawal dari latar belakang permasalahan yang ada, memetakan proses-proses, mencari penyebab permasalahan, dan terakhir merancang dan mengembangkan suatu sistem yang dapat digunakan untuk mengurangi atau menghilangkan permasalahan yang ada. Berikut ini adalah bahan, alat, dan metode pengembangan sistem serta tahapan penelitian guna merancang sebuah sistem pendukung pengambilan keputusan pemberian beasiswa bagi mahasiswa berprestasi.

3.1 BAHAN DAN ALAT PENELITIAN

Bahan penelitian yang digunakan pada penelitian ini berupa artikel atau berita *hoax* yang beredar pada media *online* yang telah di ekstraksi dan dilabeli menggunakan metode *crawling* yang ada pada platform Fact Checker Directory (Gambar 3.1).



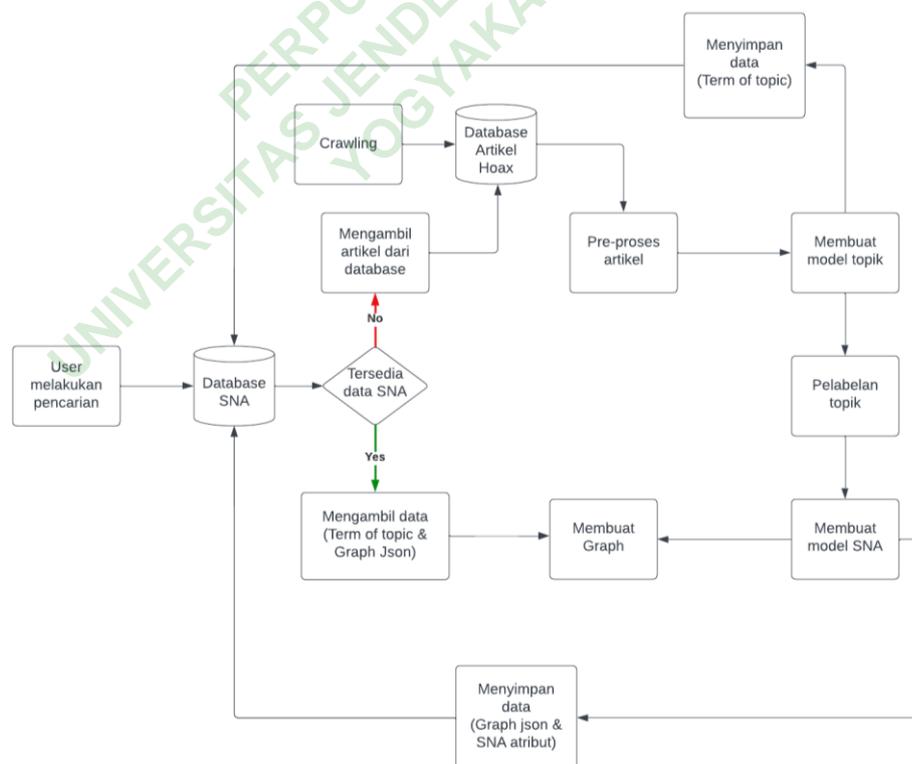
Gambar 3.1 Platform Fact Checker Directory

Pada penelitian ini telah diidentifikasi kebutuhan alat penunjang yang diperlukan untuk melakukan penelitian ini. Berikut rincian kelengkapan alat yang digunakan untuk memenuhi kebutuhan dalam penelitian ini adalah

1. Sistem Operasi : Windows 10.
2. Prosesor : Intel Core i5-8250U CPU @ 1.60GHz 1.80 GHz
3. RAM : 8 Gb
4. IDE : Visual Studio Code
5. *Database manager* : MongoDB

3.2 ARSITEKTUR SISTEM

Pada sistem ini data yang diolah berupa dokumen *text* artikel berita *hoax* yang telah dikumpulkan melalui proses *crawling* pada beberapa *website* dan disimpan ke *database* berupa data json. Data artikel tersebut akan diolah dengan metode *Topic Modeling* untuk menghasilkan data *term of topic* yang kemudian akan data tersebut akan dilabeli sebagai suatu topik lalu dibentuk model SNA dari data tersebut dan divisualisasikan berupa *grafik*. Flowchart pembentukan SNA dapat dilihat pada Gambar 3.2



Gambar 3.2 Flowchart Pembentukan Grafik SNA

Pada Gambar 3.2 dapat dilihat *flowcart* dari proses pembentukan grafik SNA diawali dari pengguna yang memasukan inputan berupa *keyword* untuk melakukan pencarian data SNA. Setelah itu sistem akan melakukan pengecekan apakah *keyword* yang dimasukkan tersedia pada *database* SNA. Jika *keyword* yang dimasukkan ada pada *database* dan cocok, maka sistem akan menampilkan data relasi berupa tabel data *term of topic* dan grafik SNA yang telah dibentuk. Jika *keyword* yang dimasukkan tidak ditemukan didalam *database*, maka sistem akan mengambil data artikel *hoax* yang tersimpan pada *database* artikel *hoax* dan mengolahnya menjadi model topik dan model SNA.

3.3 ARSITEKTUR DATABASE

Pada sistem ini data SNA yang telah diolah disimpan menggunakan MongoDB. Data SNA yang disimpan berupa data *term of topic* yaitu kumpulan kata-kata yang menyusun suatu topik dan data grafik yang tersusun atas data *node* dan data *links*. Tiap data SNA akan mempunyai id dan *keyword* sebagai identifikasi dan pelabelan data relasi.

Struktur data SNA dapat dilihat pada Gambar 3.3, dimana objek memiliki objek id dan *keyword* “Jokowi” sebagai identitas objek, data *term of topic* dengan *key* “data”, dan data grafik dengan *key* “graph”. Data *term of topic* berisikan objek *Topic 1*, *Topic 2*, sampai dengan *Topic n* dimana tiap topik memiliki objek kata 0 sampai dengan kata n. Kemudian terdapat data grafik yang berisikan data *nodes* dan data *links*. Data *nodes* tersusun atas kumpulan objek *node* dimana objek *node* memiliki id sebagai identitas objek, *topic_name*, *degree* sebagai nilai *degree centrality*, *between* sebagai nilai *betweenness centrality*, dan *closeness* sebagai nilai *closeness centrality*. Data *links* tersusun atas kumpulan objek *link* dimana objek *link* memiliki nilai *weight* yang merepresentasikan kekuatan relasi, *source* sebagai titik awal relasi, dan *target* sebagai tujuan relasi.

```

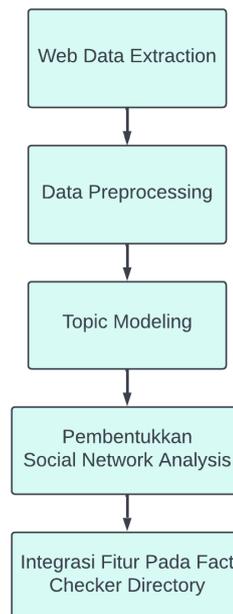
{
  'id': ObjectId('62d01c7b06b076f54f545ee7'),
  'keyword': 'jokowi',
  'data': {
    'Topic 1': {
      '0': 'bloomberg',
      '1': 'pemimpin_dunia',
      ...
    },
    'Topic 2': {
      '0': 'that_saudi',
      '1': 'than_indonesia',
      ...
    },
    ...
  },
  'graph': {
    'directed': False,
    'multigraph': False,
    'graph': {
    },
  },
  'nodes': [
    {
      'id': 'topic 0',
      'topic_name': 'Topic 1',
      'degree': 1,
      'between': 0.17777777777777778,
      'closeness': 0.47407407407407404
    },
    {
      'id': 'topic 1',
      'topic_name': 'Topic 2',
      'degree': 0,
      'between': 0.0,
      'closeness': 0.0
    },
    ...
  ],
  'links': [
    {
      'weight': 1,
      'source': 'topic 0',
      'target': 'topic 0'
    },
    {
      'weight': 1,
      'source': 'topic 2',
      'target': 'topic 5'
    },
    ...
  ]
}

```

Gambar 3.3 Struktur Data SNA

3.4 JALAN PENELITIAN

Penelitian ini bertujuan untuk menelusuri keterkaitan antara topik pada berita *hoax* dan apa pengaruhnya terhadap topik yang saling berkaitan. Ada beberapa tahap yang harus dilakukan untuk melaksanakan penelitian ini mulai dari pengambilan data melalui *Web Data Extraction*, kemudian data yang telah diperoleh akan dilakukan *preprocessing* agar dapat diolah ke tahap selanjutnya, setelah ini lanjut pada tahap pembentukan *Topic Modeling* untuk mengekstraksi topik yang ada, lalu dari topik yang telah didapatkan akan dibentuk SNA yang kemudian model SNA tersebut akan diintegrasikan pada platform Fact Checker Directory untuk menampilkan grafik SNA. Alur jalannya penelitian dapat dilihat pada Gambar 3.4.



Gambar 3.4 Diagram Jalan Penelitian

3.4.1 *Web Data Extraction*

Pada tahap ini dilakukan pengambilan data berupa artikel atau berita *hoax* pada media online dengan metode *crawling* yang telah diterapkan pada Fact Checker Directory. *Website* yang dijadikan target *crawling* adalah *kominfo.go.id* dan *turnbackhoax.id*. Setelah data berita *hoax* telah didapatkan, data tersebut akan disimpan ke *database* dan juga dapat di export dalam bentuk *csv*. *Database* yang digunakan adalah *MongoDb*. *MongoDb* digunakan sebagai *database* karena data yang disimpan berupa *unstructured* data dan tidak *rational model*.

3.4.2 *Data Preprocessing*

Data yang berhasil di ekstraksi selanjutnya akan dilakukan *preprocessing* data, dimana data akan dinormalisasi dalam bentuk dan format datanya sehingga data yang ada dapat diolah pada tahap analisis. Tahap *preprocessing* yang dilakukan berupa *text cleaning*, *tokenizing*, menghilangkan *stopword*, dan penyaringan *bigram* dan *trigram*.

1. *Text Cleaning*

Text cleaning dilakukan untuk membersihkan dokumen dari karakter-karakter yang tidak dibutuhkan pada proses pembuatan model. Contohnya karakter (@) yang tidak memiliki makna dalam suatu kata.

2. *Tokenizing*

Tokenizing adalah proses pemotongan kalimat menjadi beberapa bagian kata yang terpisah dan pada saat yang sama menghilangkan karakter tertentu seperti tanda baca, angka, dan karakter non-abjad. Ini karena karakter ini dianggap sebagai pembatas kata dan tidak memengaruhi pemrosesan kata (Rahutomo et al., 2019).

3. Menghilangkan *Stopword*

Stopword adalah beberapa kata yang umum digunakan dan tidak banyak memengaruhi pemrosesan kata. Sebagai contoh *stopword* pada bahasa Indonesia diantaranya : “yang”, “ini”, “dari”, “ke” (Rahutomo et al., 2019).

4. Penyaringan *Bigram* dan *Trigram*

Pada penelitian ini dilakukan identifikasi frasa yang mengandung 2 kata misalnya “media sosial” dan frasa yang mengandung 3 kata misalnya “Sekolah menengah atas”, oleh karena itu diperlukan penyaringan *bigram* dan *trigram*.

Dapat dilihat pada Gambar 3.5 data mentah yang belum diproses masih mengandung karakter yang tidak diperlukan dan terdapat kata yang tidak memiliki arti serta tiap kata masih menggunakan imbuhan. Oleh karena itu data mentah tersebut perlu dilakukan proses agar data dapat diolah pada tahap selanjutnya.

explanation_detail
Beredar di media sosial WhatsApp, sebuah foto tangkapan layar yang memperlihatkan Presiden Joko Widodo (Jokowi) bersama sejumlah orang tidak menggunakan masker. Pada foto itu terdapat narasi bahwa mereka tidak menggunakan masker dan tidak menjaga jarak, hanya boleh dilakukan saat menghadiri perayaan Imlek. Dari hasil penelusuran, klaim bahwa foto itu memperlihatkan peristiwa saat pandemi Covid-19 adalah salah. Faktanya, foto itu diabadikan sebelum pandemi Covid-19 masuk ke Indonesia. Foto itu memperlihatkan Beredar sebuah unggahan video TikTok yang menyangkan Presiden Jokowi menjanjikan bonus Rp12 M untuk pemain Piala AFF

Gambar 3.5 Data Mentah Artikel Hoax

Hasil olahan data *preprocessing* dapat dilihat pada Gambar 3.6 dimana tiap isi dokumen akan dipecah menjadi kumpulan kata. Dari data tersebut akan dibentuk *corpus* dan *dictionary word* yang nantinya akan digunakan untuk menghitung frekuensi tiap kata dan menentukan seberapa berpengaruh kata tersebut dalam suatu dokumen.

```
[
  [
    'beredar',
    'media',
    'sosial',
    'whatsapp',
    'foto',
    'tangkap',
    'layar',
    'presiden',
    'joko',
    'widodo',
    'jokowi',
    'orang',
    'masker',
    'foto',
    'narasi',
    'masker',
    'menjaga',
    'jarak',
    'menghadiri',
    'perayaan',
    'imlek',
    'hasil',
    'penelusuran',
    'klaim',
    'foto',
    'peristiwa',
    'pandemi',
  ],
  [
    'beredar',
    'unggah',
    'video',
    'tiktok',
    'menayangkan',
    'presiden',
    'jokowi',
    'menjanjikan',
    'bonus',
    'rp',
    'm',
    'pemain',
    'piala',
    'aff',
    'dilansir',
    'insulteng',
    'pikiranrakyat',
    'video',
    'video',
    'piala',
    'aff',
    'piala',
    'aff',
    'beredar_unggah'
  ]
]
```

Gambar 3.6 Data Olahan Preprocessing

3.4.3 Topic Modeling

Setelah data selesai dilakukan *preprocessing*, akan dibentuk pemodelan topik menggunakan metode LDA. Dua masukan utama untuk pemodelan topik menggunakan metode LDA adalah *dictionary* dan *corpus*. Dari data yang telah dilakukan *preprocessing* akan dibentuk *dictionary* berupa pemetaan kata-kata yang ada kedalam id. Kemudian akan dibentuk *corpus* yang berisi frekuensi kemunculan suatu kata. Sebelum dibentuk model LDA, *corpus* akan diolah lagi dengan menghitung frekuensi suatu kata dan memberi bobot relevansi kata tersebut didalam dokumen tertentu.

Setelah semua masukan yang diperlukan telah ada, akan dibentuk model LDA. Supaya model yang diperoleh optimal maka dilakukan evaluasi dengan metode *Coherence Measurement* untuk menghitung skor satu topik dengan mengukur tingkat kesamaan semantik antara kata-kata dengan skor tinggi dalam topik.

Hasil pemodelan topik dapat dilihat pada Tabel 3.1 dimana tiap topik akan tersusun atas beberapa kumpulan kata dengan frekuensi kemunculan kata tertentu. Kemudian hasil pemodelan topik akan disimpan pada *database* data SNA sebagai data *term of topic*.

Tabel 3.1 Hasil Olahan Data Pemodelan Topik

0	1	2	3	4	5	6	7	8	9
bloomberg	that saudi	korban gempa	shooting	silakan	jusuf kalla	johan budi	naufal	pribumi bangkit	fadli zon
pemimpin_dunia	than indonesia	pdi perjuangan	shooting_iklan	shooting	johan_budi	surat	pertamina	johan_budi	utang_pemerintahan
pemimpin	more	megawati soekarnoputri	capres_cawapres	cek	istana_kepresidenan	pasal_danatau	pesan_berantai	istana_kepresidenan	hapuskan
sudut_pandang	china_more	presiden_joko	jusuf_kalla	hasanuddin	salah_satuanya	sim_card	dki jakarta	keterangan_pers	jokowiika
pandang	upset	republik_indonesia	iklan	tulisan	republik_indonesia	pasal	pullman	tantowi	republiknkri
cnr_indonesia	than	sby	surat	shooting_iklan	keterangan_pers	danatau_pasal	guard	undangan	sofan
pemimpin_negara	that	aceh	johan_budi	capres_cawapres	tantowi	danatau	mobil	eriko	sofanperlu
sudut	saudi	beredar luas	tangkapan layar	jusuf_kalla	poin	pelaku	broadcast	pernikahan	terobosan
satusatunya	cuplikan	foto	taqwa	surat	kesehatan	berbahasa_inggris	protokol_pers	johan	berhemat
dinilai	artikel_berjudul	tumpang	zaadit	iklan	negeri	polda_metro	media_sekretariat	ektp	senen
artikelserunya	chavvah	hoaxes	pasal_danatau	johan_budi	undangan	presiden_joko	presiden_joko	arseto	bertambah
klender	disappointedjokowi	gempa	sim_card	gulingkan	selengkapnya_poin	juru_bicara	pesan	poin	menyalin
pemotretan	raditya	berlogokan	mediamedia	perlawanan	selengkapnya	undanganundang_nomor	bey	keputusan	dpr
shalat	photoshop	karib	pasal	jawa_barat	muchtar	bumn	bey_machmudin	dibagikan kali	mengganti nama
positif	disappointed	memotong	antusias	pidato	dimuat situs	umat islam	machmudin	medsos	dibubarkan
donovan	child	acara	settingan	pembaca	pernikahan	johan	deputi_bidang	keluarga	surat
salah_faktanya	stupid	aslinya	danatau_pasal	opini	johan	mengatasnamakan	protokol	selengkapnya	prof
hoax	correspondent	sapaan	danatau	blog	gizi	shalat	selengkapnya_poin	juru_bicara	kritik
rombongan	invest	susilo Bambang	pelaku	pasal_danatau	wakil_presiden	istana_kepresidenan	susilo_bambang	aaach	fadli
kecelakaan	screenshot	yudhoyono_sby	berbahasa_inggris	sim_card	asean	asu	berita_bohong	disarikan	johan_budi

Dapat dilihat pada Gambar 3.7, data *term of topic*. Nama topik dijadikan sebagai *key* dan tiap topik memiliki objek *term of topic* yang terdiri dari index sebagai *key* dan kata sebagai *value*. Dari data *term of topic* tersebut akan dibentuk grafik SNA dengan cara membuat relasi jika topik satu memiliki susunan kata yang sama dengan topik yang lainnya.

```

'data': {
  'Topic 1': {
    '0': 'bloomberg',
    '1': 'pemimpin_dunia',
    ...
  },
  'Topic 2': {
    '0': 'that_saudi',
    '1': 'than_indonesia',
    ...
  },
  'Topic 3': {
    '0': 'korban_gempa',
    '1': 'pdi_perjuangan',
    ...
  },
  'Topic 4': {
    '0': 'shooting',
    '1': 'shooting_iklan',
    ...
  },
  'Topic 5': {
    '0': 'silakan',
    '1': 'shooting',
    ...
  },
  'Topic 6': {
    '0': 'jusuif_kalla',
    '1': 'johan_budi',
    ...
  },
  'Topic 7': {
    '0': 'johan_budi',
    '1': 'surat',
    ...
  },
  'Topic 8': {
    '0': 'naufal',
    '1': 'pertamina',
    ...
  },
  'Topic 9': {
    '0': 'pribumi_bangkit',
    '1': 'johan_budi',
    ...
  },
  'Topic 10': {
    '0': 'fadli_zon',
    '1': 'utang_pemerintahan',
    ...
  }
}

```

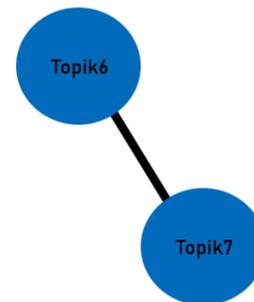
Gambar 3.7 Data Term of topic

3.4.4 Pembentukan Social Network Analysis

Topik yang telah berhasil diekstraksi pada tahap *Topic Modeling* akan dijadikan sebagai komponen utama pembentuk relasi pada model SNA. Relasi pada tiap *node* SNA akan dibentuk berdasarkan tingkat kedekatan topik, dimana topik yang saling tumpang tindih akan saling berelasi.

Pada pengolahan data SNA terdapat tiga tahap pengolahan data. Tahap pertama adalah pembentukan relasi dengan data yang diolah adalah data *term of topic* yang didapatkan dari proses pemodelan topik. Relasi akan dibentuk pada tiap topik yang memiliki susunan kata yang sama, sebagai contoh pada Gambar 3.8 topik 6 dan topik 7 terdapat kata johan, johan_budi, dan istana_kepresidenan pada susunan katanya maka kedua topik tersebut saling berelasi.

5-Topik6	6-Topik7
iusuf_kalla	johan_budi
johan_budi	surat
istana_kepresidenan	pasal_danatau
salah_satunya	sim_card
republik_indonesia	pasal
keterangan_pers	danatau_pasal
tantowi	danatau
poin	pelaku
kesehatan	berbahasa_inggris
negeri	polda_metro
undangan	presiden_joko
selengkapnya_poin	juru_bicara
selengkapnya	undangundang_nomor
muchtar	bumn
dimuat_situs	umat_islam
pernikahan	johan
johan	mengatasnamakan
gizi	shalat
wakil_presiden	istana_kepresidenan



Gambar 3.8 Pembentukan Relasi

3.4.5 Integrasi Fitur Pada Fact Checker Directory

Hasil model dari *Topic Modeling* dan SNA akan ditampilkan pada platform Fact Checker Directory berupa grafik *Social Network*. Dari model tersebut akan dibentuk fitur dimana pengguna dapat mencari hubungan topik *hoax* dengan kata kunci tertentu.