**PENERAPAN *DEEP LEARNING* UNTUK KLASIFIKASI BERITA *HOAX* MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK***

**(Studi Kasus: Penyebaran Berita *Hoax* di Media Sosial)**

**TUGAS AKHIR**

Tugas Akhir ini sebagai salah satu syarat untuk memenuhi gelar sarjana

Informatika Universitas Pembangunan Nasional “Veteran” Yogyakarta

****

Disusun oleh:

Yahdi Indrawan

123160039

**PROGRAM STUDI INFORMATIKA**

**JURUSAN INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2020**

**HALAMAN PENGESAHAN PEMBIMBING**

**PENERAPAN *DEEP LEARNING* UNTUK KLASIFIKASI BERITA *HOAX* MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK***

Disusun Oleh:

|  |  |
| --- | --- |
| Yahdi Indrawan  123160039 |  |

Telah diperiksa dan disetujui oleh pembimbing untuk diseminarkan

pada tanggal: ...................................

Menyetujui,

|  |  |
| --- | --- |
| Pembimbing I | Pembimbing II |
|  |  |
| Heru Cahya Rustamaji, S.Si., M.T.  NIK. 2 7106 96 0065 1 | Dr. Awang Hendrianto Pratomo, S.T., M.T.  NIP. 1977 07 25 2005 01 1001 |

# 

# DAFTAR ISI

[HALAMAN JUDUL i](#_Toc47002229)

[HALAMAN PENGESAHAN PEMBIMBING ii](#_Toc47002229)

[DAFTAR ISI iii](#_Toc47078385)

[BAB I 1](#_Toc47078386)

[PENDAHULUAN 1](#_Toc47078387)

[1.1 Latar Belakang Masalah 1](#_Toc47078388)

[1.2 Perumusan Masalah 4](#_Toc47078389)

[1.3 Batasan Masalah 4](#_Toc47078390)

[1.3 Tujuan Penelitian 5](#_Toc47078391)

[1.4 Manfaat Penelitian 5](#_Toc47078392)

[1.5 Metodologi Penelitian 5](#_Toc47078393)

[1.5.1 Metodologi Pengumpulan Data 5](#_Toc47078394)

[1.5.2 Metodologi Pengembangan Sistem 6](#_Toc47078395)

[1.6 Sistematika Penulisan 7](#_Toc47078396)

[BAB II 9](#_Toc47078397)

[TINJAUAN PUSTAKA 9](#_Toc47078398)

[2.1 *Hoax* 9](#_Toc47078399)

[2.2 Media Sosial 10](#_Toc47078400)

[2.3 *Web Scraping* 11](#_Toc47078401)

[2.4 *Text Mining* 12](#_Toc47078402)

[2.5 *Text Preprocessing* 13](#_Toc47078403)

[2.5.1 *Case Folding* 14](#_Toc47078404)

[2.5.2 *Remove Punctuation* 14](#_Toc47078405)

[2.5.3 *Remove Number* 15](#_Toc47078406)

[2.5.4 *Tokenizing* 15](#_Toc47078407)

[2.5.5 *Stopword Removal* 16](#_Toc47078408)

[2.5.6 *Stemming* 16](#_Toc47078409)

[2.6 Klasifikasi 20](#_Toc47078410)

[2.7 *Word Embedding* 21](#_Toc47078411)

[2.8 *Deep Learning* 25](#_Toc47078412)

[2.9 *Convolutional Neural Network* (CNN) 26](#_Toc47078413)

[2.9.1 *Input Layer* 27](#_Toc47078414)

[2.9.2 *Convolution Layer* 27](#_Toc47078415)

[2.9.3 *Activation Layer* 29](#_Toc47078416)

[2.9.4 *Pooling Layer* 30](#_Toc47078417)

[2.9.5 *Fully Connected Layer* 31](#_Toc47078418)

[*2.10* *K-Fold Cross Validation* 31](#_Toc47078419)

[2.11 *Confusion Matrix* 32](#_Toc47078420)

[2.12 Penelitian Terdahulu 34](#_Toc47078421)

[BAB III 35](#_Toc47078422)

[METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM 35](#_Toc47078423)

[3.1 Metodologi Penelitian 35](#_Toc47078424)

[3.2 Pengumpulan Kebutuhan 52](#_Toc47078425)

[3.2.1 Pengumpulan Data 53](#_Toc47078426)

[3.2.1.1 Studi Pustaka 53](#_Toc47078427)

[3.2.1.2 *Web Scraping* 53](#_Toc47078428)

[3.2.2 Analisa Kebutuhan Sistem 54](#_Toc47078429)

[3.2.3.1 Kebutuhan Fungsional 54](#_Toc47078430)

[3.2.3.2 Kebutuhan Non-Fungsional 55](#_Toc47078431)

[3.3 Proses Desain 56](#_Toc47078432)

[3.3.1 Perancangan Sistem 56](#_Toc47078433)

[3.3.1.1 Perancangan Arsitektur 56](#_Toc47078434)

[3.3.1.2 Perancangan Proses 58](#_Toc47078435)

[3.3.1.3 Perancangan Basis Data 77](#_Toc47078436)

[3.3.1.4 Perancangan Antarmuka 83](#_Toc47078437)

[3.3.2 Perancangan Pengujian 103](#_Toc47078438)

[DAFTAR PUSTAKA iv](#_Toc47002229)

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang Masalah

Penyebaran berita *hoax* saat ini sedang marak terjadi, bahkan sudah masuk fase yang cukup mengkhawatirkan. Menteri Komunikasi dan Informatika (Menkominfo), Rudiantara mengatakan tidak kurang dari 900 ribu situs yang melakukan penyebaran berita *hoax* (Fajri, 2019). Penyebaran berita *hoax* ini bertujuan untuk menggiring opini para pembaca sehingga memiliki persepsi yang sama dengan pelaku. Berdasarkan pengamatan terhadap berita hoax yang beredar, ternyata berita hoax memiliki pola-pola tertentu dalam penulisan beritanya yang dilakukan dan disebarkan secara masif (Assidik, 2018). Beberapa contoh dari pola berita *hoax* seperti pesan berantai yang terdapat anjuran untuk menyebarkan, penggunaan tata bahasa yang kurang baik, tidak logis dan tidak konsisten serta bertentangan dengan akal sehat (Nasution, 2017). Apabila penyebaran berita *hoax* dibiarkan akan berdampak buruk bagi masyarakat seperti pengaburan fakta, hilangnya kepercayaan masyarakat hingga dapat mengakibatkan perpecahan antar masyarakat.

Ada beberapa media yang sering digunakan untuk melakukan penyebaran berita *hoax* seperti media cetak, email, situs web dan media sosial. Dari beberapa media penyebaran *hoax* tersebut, menurut hasil survei Masyarakat Telematika Indonesia (Mastel) menunjukkan bahwa media sosial menjadi sumber utama peredaran berita *hoax* (Librianty, 2017). Hal tersebut karena tingginya pengguna media sosial di Indonesia yang mencapai 150 juta orang atau sebesar 56% dari total populasi. Dari data tersebut juga menujukkan bahwa pengguna media sosial meningkat 15% dari tahun sebelumnya (Wearesocial, 2019). Ada beberapa jenis berita *hoax* yang sering diterima masyarakat seperti isu di bidang sosial politik, SARA, kesehatan, IPTEK, bencana alam, dan lain-lain. Banyaknya berita *hoax* yang tersebar membuat sulitnya masyarakat dalam memilah dan membedakan mana berita *hoax* dan tidak. Ada beberapa masyarakat yang kritis terhadap berita yang tersebar, lalu melakukan pengecekan apakah berita tersebut *hoax* atau tidak. Namun, tidak sedikit pula masyarakat yang langsung percaya terhadap berita yang disampaikan. Maka dari itu, perlu dikembangkan sebuah sistem yang dapat digunakan untuk mengklasifikasikan suatu konten berita sehingga dapat diketahui apakah berita tersebut *hoax* atau bukan.

Penelitian yang berkaitan dengan *hoax* pernah dilakukan menggunakan metode *naive bayes* untuk melakukan klasifikasi berita dari *tweet* menjadi tiga kategori yaitu berita palsu (*hoax*), berita asli (*fact*) dan berita yang tidak terindikasi (Tanjung, 2018). Namun, pada penelitian tersebut memiliki kendala karena tingginya kemiripan antar data membuat sulitnya melakukan klasifikasi berita *hoax* sehingga mempengaruhi tingkat error dan akurasinya. Beberapa penelitian lain melakukan klasifikasi artikel *hoax* dengan berbagai metode seperti *support vector machine* dengan pembobotan TF-IDF(Maulina & Sagara, 2018)dan membandingan metode *rocchio* dengan *multinomial naive bayes* (Afriza & Adisantoso, 2018). Kedua penelitian tersebut telah menghasilkan sistem dengan performa yang baik, namun data yang digunakan relatif sedikit dan belum adanya fitur untuk melakukan *upgrade* model menggunakan data yang baru. Selain itu, ada penelitian yang melakukan klasifikasi pengguna twitter untuk melihat potensi penyebaran berita *hoax* berdasarkan isi konten *tweet* dan perilaku pengguna di media sosial. Penelitian tersebut menggunakan metode *backprogation* dengan membandingkan algoritma *gradient descent backprogation* dan *lavenberg-marquad backprogation*, namun memerlukan waktu yang cukup lama untuk melakukan proses *training* (Lhaksmana, Nhita, & Budhiarto, 2017).

Beberapa penelitian tersebut telah menggunakan berbagai metode untuk menyelesaikan permasalahan yang berkaitan tentang berita *hoax* dengan hasil yang baik. Namun, ada beberapa aspek yang perlu ditingkatkan seperti skala data yang digunakan masih relatif rendah, belum adanya fitur untuk *upgrade* model dan tidak dapat menemukan pola berita *hoax* karena hasil klasifikasinya tidak berdasarkan pola melainkan kemiripan data. Padahal dalam beberapa kasus, penambahan atau pengurangan beberapa kata dari berita asli akan merubah maksud dan tujuan dari berita tersebut. Maka dari itu, untuk meminimalisir kendala tersebut dapat dilakukan dengan menggunakan metode *deep learning*. Hal tersebut karena metode *deep learning* sangat efektf dan lebih mudah dalam mengidentifikasi pola-pola dari data yang dimasukkan. Selain itu, kelebihan lain dari *deep learning* mampu menangani skala data yang besar dan mengekstraksi fitur dari data tersebut secara otomatis. Salah satu algoritma yang termasuk metode *deep learning* dan mampu mengatasi permasalahan tersebut adalah *Convolutional Neural Network*.

*Convolutional Neural Network* (CNN) merupakan algoritma yang terdiri dari *neuron* yang memiliki *weight, bias* dan *activation function*. Algoritma ini memiliki layer hirarki/bertumpuk-tumpuk dengan struktuk dan jumlah jaringan syaraf yang sangat banyak sebagai jaringan tersembunyi (*hidden layer*). CNN akan memanfaatkan proses konvolusi untuk mendapatkan informasi yang merepresentasikan dari data yang digunakan dengan menggerakkan sebuah kernel (filter) berukuran tertentu (Lina, 2019). Pada umumnya algoritma CNN digunakan untuk mengolah data yang berbentuk dua dimensi seperti *image classification, object detection* ataupun *face recognation*. Padahal, algoritma ini dapat digunakan dalam lingkup yang lebih luas untuk mengidentifikasi dan mengklasifikasikan data yang berbentuk gambar, video, suara maupun teks. Maka dari itu, algoritma CNN dapat diterapkan pada masalah penelitian ini untuk melakukan klasifikasi berita *hoax*. Beberapa penelitian yang menggunakan algoritma CNN untuk mengatasi data yang berbentuk teks seperti klasifikasi teks medis (Hughes, Li, Kotoulas, & Suzumura, 2017) dan klasifikasi arrtikel berita berbahasa indonesia dengan nilai akurasi sebesar 96,70% dan presisi, recall serta f-measure mencapai 96,60% (Razi, 2017).

Berdasarkan permasalahan yang telah dipaparkan di atas, maka penerapan metode *deep learning* dengan menggunakan *Convolutional Neural Network* menjadi solusi yang ditawarkan pada penelitian ini. Data yang akan digunakan bersumber dari situs turnbackhoax.id yang telah memvalidasi berita *hoax* yang beredar di media sosial. Data tersebut akan dibagi menjadi data latih dan data uji. Data latih digunakan sebagai bahan belajar untuk menghasilkan model yang mampu mengenali pola-pola berita *hoax*, sedangkan data uji digunakan untuk mengukur tingkat akurasi dalam proses klasifikasinya. Selain itu, penelitian ini juga menguji performa dari algoritma *Convolutional Neural Network* dalam menangani data berbentuk teks yang bisanya algoritma ini digunakan untuk pengolahan citra. Dengan menerapkan algoritma *Convolutional Neural Network* diharapkan dapat menghasilkan suatu model yang dapat dimaanfatkan untuk melakukan klasifkasi berita sehingga dapat diketahui apakah berita tersebut termasuk kategori *hoax* atau bukan.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat diambil rumusan masalah sebagai berikut:

1. Bagaimana mengenali pola-pola berita untuk melakukan klasifikasi berita *hoax* menggunakan algoritma *Convolutional Neural Network* ?
2. Bagaimana tingkat performa dari algoritma *Convolutional Neural Network* dalam melakukan klasifikasi berita *hoax* ?

## 1.3 Batasan Masalah

Menghindari terjadinya topik permasalahan yang lebih luas, maka diberikan batasan masalah sebagai berikut:

1. Data penelitian didapatkan dari situs web turnbachoax.id yang telah memvalidasi berita *hoax* yang beredar di media sosial.
2. Data penelitian yang digunakan dari tahun 2015 sampai dengan 2019.
3. Data penelitian yang digunakan akan dikategorikan menjadi dua kelas yaitu *hoax* dan fakta.
4. Data penelitian yang digunakan berbahasa Indonesia.
5. Metode klasifikasi yang digunakan adalah metode *Convolutional Neural Network*.

## 1.3 Tujuan Penelitian

Tujuan pada penelitian ini adalah sebagai berikut:

1. Dapat mengenali pola-pola berita untuk klasifikasi berita *hoax* dengan menggunakan algoritma *Convolutional Neural Network*.
2. Dapat mengetahui tingkat performa algoritma *convolutional neural network* untuk klasifikasi berita *hoax*.

## 1.4 Manfaat Penelitian

Dengan dilakukannya penelitian ini dapat mengetahui performa algoritma *convolutional neural network* dalam menyelesaikan permasalahan tentang klasifikasi berita *hoax*. Selain itu, diharapkan dengan adanya sistem ini dapat memudahkan masyarakat khususnya pengguna media sosial dalam memilah dan menyaring berita yang tersebar. Sehingga dampak dari penyebaran berita *hoax* dapat diminimalisir.

## 1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan pada penelitian ini adalah sebagai berikut:

### 1.5.1 Metodologi Pengumpulan Data

Alur metodologi pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan cara mencari dan menghimpun data ataupun sumber-sumber pustaka yang dapat mendukung penelitian serta memberikan informasi untuk menyelesaikan permasalahan pada penelitian ini. Studi kepustakaan yang digunakan bersumber dari buku, jurnal, artikel dan paper yang berkaitan pada penelitian ini.

1. *Web Scraping*

Data yang digunakan pada penelitian ini bersumber dari situs turnbackhoax.id. Situs tersebut telah memvalidasi kebenaran dari berita-berita yang beredar di media sosial. Pengumpulan data dilakukan dengan menggunakan teknik *web srcaping* dengan ketentuan data dari tahun 2015 sampai dengan data tahun 2019.

### 1.5.2 Metodologi Pengembangan Sistem

Metodologi pengembangan sistem yang digunakan adalah metode *prototype.* Penggunaan metode *prototype* dikarenakan metode ini lebih menekankan pada komunikasi antara pengembang dan pelanggan sehingga kebutuhan pengguna dapat diterjemahkan dalam bentuk model (prototipe) dan hasil yang diharapkan pun dapat tercapai. Ada beberapa tahapan yang dilakukan pada metode *prototype* yaitu (Pressman, 2010):

1. *Communication*

Pada tahapan ini dilakukan proses komunikasi antara pengembang dan pelanggan mengenai tujuan dibuatnya suatu perangkat lunak. Selain itu, akan dilakukan analisis serta identifikasi kebutuhan apa saja yang akan diperlukan selama proses pembuatan perangkat lunak.

1. *Quick Plan dan Modeling Quick Design*

Pada bagian ini akan membuat perencanaan dan pemodelan secara cepat berdasarkan kebutuhan yang telah diidentifikasi pada tahap sebelumnya. Perencanaan yang akan dilakukan dapat berupa pembagian tugas dan rencana kerja. Sedangkan dari sisi pemodelan akan membuat desain model yang merepresentasikan aspek-aspek yang diinginkan dari pelanggan.

1. *Construction of Prototype*

Pada proses *contruction of prototype,* pengembang akan mulai membuat program (perangkat lunak) berdasarkan rencana dan model yang telah dibuat sebelumnya. Pada tahap ini juga termasuk *testing* dari perangkat lunak yang telah berhasil dibuat.

1. *Deployment Delivery & Feedback*

Setelah perangkat lunak berhasil dibuat dan telah melalui *testing,* proses selanjutnya yaitu tahap *deployment delivery & feedback*. Perangkat lunak tadi akan disampaikan kepada pelanggan untuk dilakukan pengecekan. Jika mendapatkan *feedback* baik, maka perangkat lunak akan diterima oleh pelanggan dan pembuatan perangkat lunak telah selesai. Apabila *feedback* yang diberikan kurang baik atau tidak sesuai dengan pelanggan, maka kembali ke tahap *communication* dengan membahas perbaikan yang perlu dilakukan oleh pengembang.

## 1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam menyusun laporan penelitian ini adalah sebagai berikut :

Bab I Pendahuluan

Pada bagian ini membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka

Tinjauan pustaka memuat tentang dasar teori yang digunakan untuk analisis dan perancangan sistem serta implementasi pada penelitian ini. Selain itu juga digunakan sebagai bahan referensi dan pondasi untuk memperkuat argumentasi dalam penelitian ini. Teori-teori yang sesuai dengan penelitian ini antara lain media sosial, *hoax,* klasifikasi, *deep learning* dan *convolutional neural network.*

Bab III Analisis dan Perancangan Sistem

Pada bagian ini akan membahas mengenai analisa dan perancangan sistem hingga implementasi sistem dalam mengklasifikasi berita *hoax* di media sosial.

Bab IV Hasil, Pengujian dan Pembahasan

Pada bab ini akan menyajikan hasil penelitian yang berisi hasil implementasi dari perancangan yang telah dibuat pada bab sebelumnya. Selain itu berisi pengujian terhadap hasil penelitian beserta pembahasannya.

Bab V Kesimpulan dan Saran

Pada bagian ini berisi kesimpulan dari hasil penelitian dan saran yang diajukan oleh penulis untuk pengembangan penelitian selanjutnya.

# BAB II

# TINJAUAN PUSTAKA

## 2.1 *Hoax*

*Hoax* merupakan suatu kepalsuan yang disamarkan seolah-olah itu adalah kebenaran (MacDougall, 1958). *Hoax* dapat diartikan juga sebagai rangkaian informasi yang memang sengaja disesatkan, namun ‘dijual’ sebagai kebenaran (MacDougall, 1958). Beredarnya *hoax* dapat membentuk opini public yang mengarah kepada terjadinya kehebohan di masyarakat (MacDougall, 1958). Ada beberapa jenis informasi yang termasuk *hoax*, antara lain (Vibriza, Rahadi, Marwan, & Ahyad, 2017):

1. *Fake News* (berita palsu)
2. *Clickbait* (tautan jebakan)
3. *Confirmation Bias* (bias informasi)
4. *Misinformation* (informasi yang salah)
5. *Satire* (sindiran)
6. *Post-truth* (pasca kebenaran)
7. *Propaganda* (provokasi)

Dengan mudahnya penyebaran informasi pada saat ini, sangat memudahkan pula berita *hoax* beredar di masyarakat. Sayangnya, masih banyak masyarakat yang percaya dengan berita *hoax* yang beredar. Ada dua faktor yang menyebabkan masyarakat percaya terhadap berita *hoax*. *Pertama,* karena *hoax* yang dilakukan secara berulang-ulang dan dilakukan secara masif dapat dianggap menjadi suatu kebenaran. Selain itu, seseorang akan cenderung percaya terhadap berita *hoax* jika informasinya sesuai dengan opini atau sikap yang dimiliki (Respati, 2017).

Munculnya *hoax* di masyarakat dilatarbelakangi oleh berbagai tujuan, mulai dari main-main, pembentukan opini publik, penipuan hingga hasutan untuk memecah belah masyarakat. Berbagai tujuan tersebut memunculkan dampak negatif dari penyebaran berita *hoax*. Dampak negatif dari berita *hoax* tersebut seperti pengaburan fakta, hilangnya kepercayaan masyarakat hingga dapat mengakibatkan perpecahan antar masyarakat. Hal tersebut akan menjadi semakin buruk jika penyebaran berita *hoax* terus menerus dibiarkan.

Ada berbagai cara yang dapat dilakukan untuk terhindar dari berita *hoax*, seperti membaca seluruh isi berita, menelusuri sumber berita dan membandingkan antar berita. Selain itu, berita *hoax* juga biasanya memiliki pola-pola tertentu. Ada tujuh pola atau ciri-ciri berita *hoax* yang perlu diketahui sebagai antisipasi agar tidak termakan *hoax* yang sedang beredar, antara lain (Nasution, 2017): *Pertama,* terdapat anjuran untuk menyebarkan pesan tersebut ke orang lain. *Kedua*, penggunaan tata bahasa yang kurang sempurna. *Ketiga,* tidak adanya berita lain yang mendukung pesan tersebut. *Keempat,* pesan yang tersebar tidak logis, tidak konsisten dan bertentangan dengan akal sehat. *Kelima,* tidak menyebutkan kenyataan yang dapat dibuktikan dengan fakta-fakta. *Keenam,* pesan berantai atau pesan yang telah di*foward* berkali-kali. *Ketujuh,* pembuat *hoax* akan menghubungkan pesan tersebut dengan sumber resmi yang sebenarnya palsu seperti *hoax* tentang pemenang undian berhadiah dengan menyertakan alamat website yang terlihat seperti asli namun sebenarnya website palsu yang telah dibuat sendiri sebelumnya.

## 2.2 Media Sosial

Menurut *McGraw Hill Dictionary,* media sosial adalah segala bentuk media komunikasi interaktif yang memungkinkan terjadinya interaksi dua arah dan umpan balik. Media sosial dapat didefinisikan sebagai sarana yang digunakan oleh orang-orang untuk berinteraksi satu sama lain dengan cara menciptakan, berbagi serta bertukar informasi dan gagasan dalam sebuah jaringan dan komunikasi virtual. Dengan kata lain, media yang digunakan untuk menjadi sosial (Safko, 2012). Media sosial dapat digunakan untuk berbagi teks, gambar, suara dan video (Kotler & Keller, 2012). Berdasarkan berbagai penjelasan tersebut, dapat disimpulkan bahwa media sosial merupakan media komunikasi yang digunakan untuk melakukan interaksi satu sama lain dengan tujuan untuk berbagi dan bertukar informasi baik berupa teks, gambar, suara maupun video untuk menjadi makhluk sosial.

Media sosial dapat dikenali melalui beberapa karakteristik atau ciri-ciri yang dimiliknya, seperti adanya partisipasi pengguna untuk memberikan umpan balik terhadap konten di media sosial, adanya keterbukaan bagi pengguna untuk memberikan komentar dan sebagainya, adanya interaksi antar pengguna dan semua itu saling terhubung satu sama lain.

Pada saat ini, media sosial memiliki peran besar dalam membangun pola pikir dan perilaku dalam berbagai bidang kehidupan masyarakat. Hal tersebut tidak lepas dari berbagai fungsi yang ditawarkan dalam penggunaan media sosial. Beberapa fungsi dari media sosial, yaitu (Wijayanto, 2012):

## 2.3 *Web Scraping*

*Web scraping* adalah salah satu metode untuk melakukan pengambilan data. Cara kerja dari metode ini dengan mengekstraksi data yang berada di suatu website dan menganalisis data tersebut untuk digunakan bagi kepentingan lain (Turland, 2010). *Web scraping* berfokus pada pengambilan data dan ekstraksinya. *Web scraping* dapat diartikan sebagai bentuk penyalinan, dimana data spesifik dikumpulkan dan disalin dari suatu web dan disimpan di database lokal atau spreadsheet pusat yang dapat digunakan lagi di proses selanjutnya (Pereira & Vanitha, 2015). Pendapat lain menyatakan bahwa *web* scraping adalah sebuah proses yang memanfaatkan dokumen berbentuk website dibangun dengan bahasa *markup* seperti HTML atau XHTML berbentuk *semi-structured* yang berasal dari internet dan kemudian dianalisis untuk mendapatkan suatu informasi yang berguna untuk konteks lain (Nikhit, 2015).

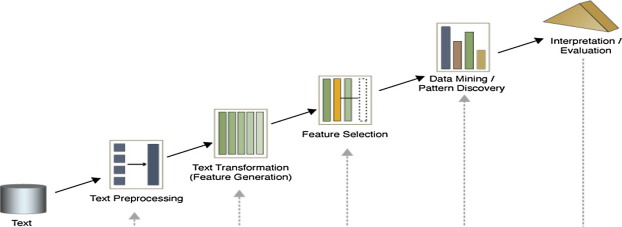
Proses *web scraping* pertama kali dilakukan dengan cara manual yaitu dengan *copy paste* dari website ke tempat penyimpanan lokal. Namun, apabila data yang ingin diambil dalam jumlah banyak, maka cara ini kurang efektif karena memerlukan tenaga untuk menyalin dan juga memakan banyak waktu. Selain dengan cara manual, *web scraping* dapat dilakukan dengan otomatis. Beberapa cara untuk melakukan *scraping* secara otomatis dengan menggunakan *coding,* aplikasi ataupun *extension browser*. Sedangkan pada penelitian ini, metode yang digunakan untuk melakukan *scraping* yaitu dengan *extension browser* dari *google* yang bernama *Data Miner*.

*Data miner* adalah salah satu *extension chrome* yang berfungsi untuk melakukan proses *scraping* data dari suatu website. *Data miner* akan melakukan *scraping* dengan mengekstrak data dari suatu website dan hasilnya disimpan pada *spreadsheet* atau excel serta dapat di*eksport* ke dalam bentuk file XLS, CSV, XLSX atau TSV (.xls, .csv, .xlsx, .tsv). Kelebihan dari *data miner* yaitu bisa digunakan dan dipahami oleh semua orang tanpa harus menguasai struktur HTML dari website, adanya fitur *auto pagination* untuk melanjutkan *scraping* ke halaman berikutnya dari website tersebut dan prosesnya *scraping*nya dilakukan secara otomatis.

## 2.4 *Text Mining*

*Text mining* merupakan proses mencari untuk mencari pola pada kumpulan data-data. Proses pencarian pola pada *text mining* dilakukan oleh komputer untuk mendapatkan sesuatu hal yang baru dalam bentuk informasi atau menemukan kembali informasi yang tersirat secara implisit dengan cara melakukan ekstraksi informasi dari sumber data teks yang berbeda-beda (Feldman & Sanger, 2007). *Text mining* dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata yang terkandung didalamnya serta dapat menemukan kesamaan antar dokumen sehingga dapat mengetahui hubungan antar variabelnya.

Penggunaan *text mining* biasanya untuk menganalisis data teks yang semi-terstruktur ataupun data yang tidak terstruktur. Maka dari itu, *text mining* bertanggung jawab untuk merubah data yang awalnya semi-terstruktur atau tidak terstruktur menjadi data yang terstruktur. Berbagai jenis permasalahan yang dapat menggunakan *text mining* seperti klasifikasi, *clustering, information extraction* dan *information retreival* (Berry & Kogan, 2010)*.* Dari berbagai permasalahan tersebut dapat diaplikasikan pada suatu kasus, seperti klasifikasi artikel berita, analisis sentimen, penyaringan spam email dan lain sebagainya. Beberapa tahapan yang biasanya dilakukan untuk melakukan *text mining,* yaitu *text preprocessing, text transformation, feature selection, pattern discovery* dan *interpretation*. Tahapan *text mining* lebih jelasnya dapat dilihat pada Gambar 2.1.



**Gambar 2.1** Tahapan *Text Mining* (Hashimi, Hafez, & Mathkour, 2015)

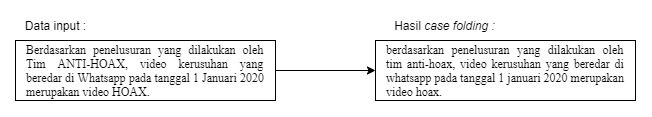
## 2.5 *Text Preprocessing*

*Text preprocessing* merupakan salah satu tahapan yang penting pada proses *text mining*. *Text preprocessing* adalah proses untuk membersihkan dan mempersiapkan data sebelum dilakukannya proses klasifikasi (Haddi, Liu, & Shi, 2013). Penggunaan *preprocessing* dikarenakan data yang digunakan pada proses *mining* tidak selamanya dalam kondisi ideal untuk diproses. Tidak jarang data yang akan digunakan memiliki berbagai permasalahan yang dapat mempengaruhi hasil dari proses *mining* itu sendiri. Permasalahan-permasalahan yang sering muncul pada data yang akan diproses seperti *missing value, data redundant, outliers* ataupun format data yang tidak sesuai dengan sistem. Maka dari itu, proses ini akan mengubah bentuk data yang awalnya semi-terstruktur ataupun tidak terstruktur menjadi data yang terstruktur.

Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa *text preprocessing* bertujuan untuk mendapatkan bentuk data yang siap diolah atau diproses. Tahapan *text preprocessing* yang akan digunakanmeliputi *case folding, remove punctuation, remove number, tokenizing, stopword removal* dan *stemming*.

### 2.5.1 *Case Folding*

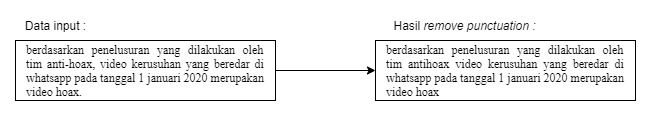
Tahap *case folding* adalah proses merubah semua huruf pada dokumen menjadi huruf kecil (*lowercase*). Huruf yang diterima hanya ‘a’ sampai dengan ‘z’. Perubahan huruf ini bertujuan untuk membentuk suatu standar pada data. Misalnya terdapat data yang berisi “Berita HOAX” akan menjadi “berita hoax”. Contoh lain dari proses *case folding* dapat dilihat pada Gambar 2.2.



**Gambar 2.2** Proses *Case Folding*

### 2.5.2 *Remove Punctuation*

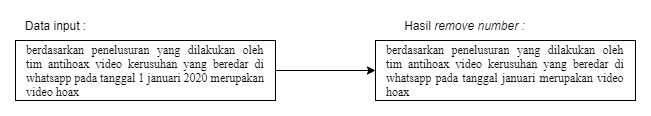
Tahap *remove punctuation* merupakan proses untuk menghapus tanda baca yang terdapat pada suatu data. Pada tahap ini, tanda baca di anggap sebagai *delimiter.* Penghapusan tanda baca pada suatu data akan mengurangi beban pemrosesan. Beberapa contoh tanda baca seperti titik (.), koma (,), tanda tanya (?), tanda seru (!) dan lainnya. Misalnya terdapat data yang berisi “Cek Fakta: Virus Corona tidak bisa masuk Indonesia?” akan berubah menjadi “Cek Fakta Virus Corona tidak bisa masuk Indonesia”. Contoh lain dari proses *remove punctuation* dapat dilihat pada Gambar 2.3.



**Gambar 2.3** Proses *Remove Punctuation*

### 2.5.3 *Remove Number*

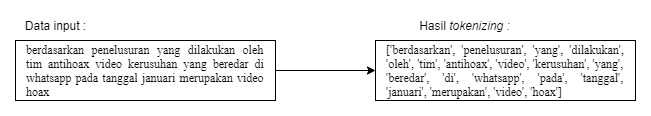
*Remove number* merupakan tahapan yang hampir sama dengan *remove punctuation*. Bedanya pada bagian objek yang akan dihapus. Pada *remove punctuation* menghapus tanda baca, sedangkan *remove number* akan menghapus semua angka pada suatu data. Angka akan dihapus karena dianggap tidak memiliki arti dan termasuk *delimiter*. Misalnya ada data yang berisi “terdapat 2 akun media sosial yang menyebar berita hoax” akan diubah menjadi “terdapat akun media sosial yang menyebar berita hoax”. Contoh lain penggunaan dari *remove number* dapat dilihat pada Gambar 2.4.



**Gambar 2.4** Proses *Remove Number*

### 2.5.4 *Tokenizing*

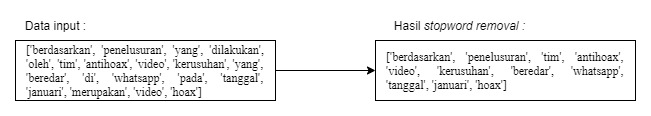
Tahap *tokenizing* atau tokenisasi adalah proses pemisahan teks dari paragraph atau kalimat berdasarkan per kata. Proses *tokenizing* biasanya berpatokan pada karakter *whitespace* seperti spasi, enter ataupun tab sebagai pemisah atau pemotong antar kata. Tiap kata yang akan dihasilkan pada proses ini akan disebut sebagai token*.* Contoh tokenisasi dari kalimat ”komparasi metode deep learninguntuk klasifikasi berita hoax” akan menghasilkan 8 token, yaitu: “komparasi”, “metode”, “deep”, “learning”, “untuk”, “klasifikasi”, “berita”, dan “hoax”. Contoh lain yang dapat memperjelas proses *tokenizing* dapat dilihat pada Gambar 2.5.



**Gambar 2.5** Proses *Tokenizing*

### 2.5.5 *Stopword Removal*

*Stopword removal* adalah proses penghapusan kata *stopword*. *Stopword* adalah kata-kata umum yang sering muncul pada suatu dokumen atau data namun tidak memiliki makna (tidak berkaitan dengan tema dokumen). *Stopword* biasanya berupa kata ganti orang atau kata penghubung. Beberapa contoh dari kata *stopword* seperti ‘aku’, ‘kamu’, ‘yang’, ‘di’, ‘dan’, ‘atau’, ‘tetapi’ dan lain sebagainya. Jadi, *stopword removal* merupakan proses pembuangan kata-kata kurang penting (*stopword*) dari token yang telah dihasilkan pada tahap sebelumnya. Pembuangan *stopword* tidak akan mengubah makna atau isi dari suatu data. Bahkan dengan menghapus *stopword* dapat mengurangi ukuran index, waktu pemprosesan dan *noise* dari suatu data. Contoh penerapan dari *stopword removal* dengan inputan hasil dari proses *tokenizing* dapat dilihat pada Gambar 2.6.



**Gambar 2.6** Proses *Stopword Removal*

### 2.5.6 *Stemming*

*Stemming* adalah suatu proses untuk mentransformasi kata-kata yang terdapat pada suatu dokumen menjadi kata dasar (*root* word). Biasanya proses yang dilakukan dengan menghapus awalan, akhiran maupun sisipan. Hal ini bertujuan untuk mengurangi varian kata dengan makna yang hampir sama pada suatu dokumen. Selain itu, proses *stemming* dapat meningkatkan kinerja pengambilan data karena jumlah indeks dengan istilah yang berbeda menjadi berkurang. Contoh proses transformasi kata pada *stemming* pada kata “menghasilkan” menjadi “hasil”. Terdapat beberapa algoritma yang telah dikembangkan untuk melakukan proses *stemming*. Penggunaan algoritma *stemming* harus disesuaikan dengan bahasa yang digunakan. Algoritma yang digunakan untuk teks dokumen yang menggunakan bahasa Indonesia akan berbeda dengan teks yang berbahasa Inggris. Pada teks berbahasa Indonesia dapat menggunakan beberapa algoritma, salah satunya Algoritma Nazief dan Adriani.

Algoritma Nazief dan Adriani adalah algoritma *stemming* untuk bahasa Indonesia. Algoritma ini memiliki beberapa aturan morfologi untuk merubah suatu kata menjadi kata dasar yaitu dengan menghilangkan *affiks* (awalan, imbuhan dan lain-lain) dan kemudian dicocokkan dengan kamus kata dasar. Kamus kata dasar merupakan hal yang paling utama pada algoritma ini karena semakin banyak kamus kata dasarnya maka semakin tinggi pula akurasinya. Algoritma ini memiliki tahap-tahap sebagai berikut (Nazief & Adriani, 1996):

1. Cari kata yang akan di-*stem* di dalam kamus, jika data tersebut ditemukan maka kata tersebut adalah kata dasar dan algoritma berhenti. Jika tidak ada maka lanjutkan ke langkah-2.

2. Hilangkan *inflectional suffix* (imbuhan infleksional) yaitu (“-lah”, ”-kah”, “-tah”, “-ku”, “-mu”, “-nya”).

3. Hapus *derivation suffix* (imbuhan turunan) yaitu (“-i”, “-an”, atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah-3a.

a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k” maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah-3b.

b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan dan lanjut ke langkah-4.

4. Hapus *derivation prefix* (awalan turunan) yaitu (“be-“, “di-“, “ke-“, “me-“, “pe-“, “se-“, “te-“). Jika pada langkah 3 ada *suffix* yang dihapus maka pergi ke langkah-4a, jika tidak maka pergi ke langkah-4b.

a. Periksa tabel kombinasi awalan-akhiran yang tidak diizinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah-4b. Kombinasi awalan dan akhiran yang tidak diijinkan dapat dilihat pada Tabel 2.1.

**Tabel 2.1** Kombinasi Awalan dan Akhiran yang Tidak Diijinkan (Nazief & Adriani, 1996)

|  |  |
| --- | --- |
| **Awalan** | **Akhiran yang tidak diijinkan** |
| be- | -i |
| di- | -an |
| ke- | -i, -kan |
| me- | -an |
| se- | -i, -kan |

b. Untuk i=1 sampai 3, tentukan tipe awalan kemudian hapus awalan. Jika kata dasar belum ditemukan juga lakukan langkah-5, jika sudah maka algoritma berhenti. Proses ini harus dilihat terlebih dahulu aturan peluruhan kata pada Tabel 2.2.

**Tabel 2.2** Aturan Peluruhan Kata Dasar (Nazief & Adriani, 1996)

|  |  |  |
| --- | --- | --- |
| **Aturan** | **Awalan** | **Peluruhan** |
| 1 | berV. . . | ber-V.. | be-rV.. |
| 2 | Belajar | bel-ajar |
| 3 | berClerC2 | be-ClerC2.. dimana C1!= {‘r’|’l’} |
| 4 | terV. . . | ter-V.. | te-rV.. |
| 5 | terCer. . . | terCer. . . dimana C!=’r’ |
| 6 | teClerC2 | te-CleC2. . . dimana C1!=’r’ |
| 7 | me{I|r|w|y}V. . . | me-{I|r|w|y}V. . . |
| 8 | mem{b|fv}. . . | mem-{b|f|v}. . . |
| 9 | mempe. . . | m-pe. . |
| 10 | mem{r|V|V}. . . | me-m{r|V|V}. . . | me-p{r|V|V}. . . |
| 11 | men{c|d|j|z}. . . | Men-{c|d|j|z}. . . |
| 12 | menV. . . | Me-nV. . . | me-tV. . . |
| 13 | meng{g|h|q|k}. . . | Meng-{g|h|q|k}. . . |
| 14 | mengV. . . | Meng-V. . . | meng-kV. . . |
| 15 | mengeC. . . | Meng-C. . . |
| 16 | menyV. . . | Me-ny. . . | men-sV. . . |
| 17 | memV. . . | Mem-pV. . . |
| 18 | pe{w|y}V. . . | Pe-{w|y}V. . . |
| 19 | perV. . . | Per-V. . . | pe-rV. . . |
| 20 | pem{b|f|v}. . . | Pem-{b|f|v}. . . |
| 21 | pem{rV|V}. . . | Pe-m{rV|V}. . . | pe-p{rV|V}. . . |
| 22 | pen{c|d|j|z}. . . | Pen-{c|d|j|z}. . . |
| 23 | penV. . . | Pe-nV. . . | pe-tV. . . |
| 24 | peng{g|h|q}. . . | Peng-{g|h|q} |
| 25 | pengV. . . | Peng-V | peng-kV |
| 26 | penyV. . . | Pe-nya | peny-sV |
| 27 | pelV. . . | Pe-IV. . .; kecuali untuk kata “pelajar” |
| 28 | peCP. . . | Pe-CP. . . dimana C!={r|w|y|I|m|n} dan P!=’er’ |
| 29 | perCerV. . . | Per-CerV. . . dimana C!={r|w|y|I|m|n} |

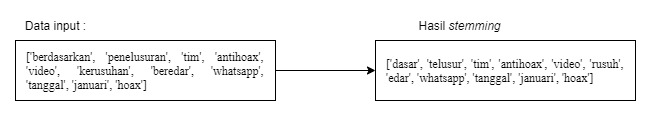
Tipe awalan ditentukan melalui langkah-langkah sebagai berikut:

1. Jika awalannya adalah: “di-”, “ke-“ atau “se-“ maka tipe awalannya secara berturut-turut adalah “di-“, “ke-“ atau “se-”.
2. Jika awalannya adalah: “te-“, “me-“, “be-“ atau “pe-“ maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
3. Jika dua karakter pertama bukan “di-“, “ke-“, “se-“, “te-“, “be-“, “me-“ atau “pe-“ maka berhenti.
4. Jika tipe awalan adalah “none” maka berhenti. Hapus awalan jika ditemukan.

5. Lakukan *recording*. Proses *recording* dilakukan dengan menambah karakter *recording* di awal kata yang dipenggal dengan mengacu pada Tabel 2.5.

6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai kata dasar. Proses selesai.

Sebagai contoh dari proses *stemming,* misalnya terdapat kata “menghasilkan” menjadi “hasil” atau kata “berasal” akan menjadi “asal”. Contoh lain untuk memperjelas proses *stemming* dapat dilihat pada Gambar 2.7.



**Gambar 2.7** Proses *Stemming*

## 2.6 Klasifikasi

Klasifikasi adalah proses untuk menemukan sebuah model atau fungsi yang menggambarkan dan membedakan antar kelas data. Klasifikasi didefinisikan sebagai pengkategorian dokumen secara otomatis ke dalam satu atau lebih kelas yang telah ditentukan berdasarkan isinya (Sebastiani, 2002). Dokumen-dokumen yang memiliki isi yang sama atau relevan akan dikelompokkan pada ketegori yang sama. Dengan kata lain bahwa klasifikasi merupakan pemberian ketegori yang telah didefinisikan kepada dokumen yang belum memiliki kategori (Goller, 2000). Tujuan dari klasifikasi yaitu untuk memperkirakan kelas dari suatu objek yang belum diketahui kelasnya (Gaikwad, Chaugule, & Patil, 2014). Selain itu, akan memudahkan mencari suatu informasi dengan isi yang relevan.

Klasifikasi memiliki dua proses yaitu proses *training* dan *testing*. Proses *training* adalah proses untuk melatih algoritma yang digunakan dengan dokumen yang telah memiliki kategori. Proses *training* akan menghasilkan suatu model yang nantinya dapat digunakan untuk melakukan prediksi kategori dari suatu data. Sedangkan, *proses testing* yaitu proses untuk menguji algoritma dengan data yang belum memiliki kategori untuk memprediksi data tersebut termasuk kategori yang mana. Pada proses *testing* akan diketahui tingkat akurasi dari model yang telah dibangun dari proses *training*.

## 2.7 *Word Embedding*

*Word embedding* adalah proses untuk mengubah suatu kata menjadi sebuah vektor yang terdiri dari kumpulan angka. Jumlah angka yang dihasilkan dari proses *word embedding* disebut dimensi. Penggunaan *word embedding* biasanya untuk algoritma *deep learning,* karenakan algoritma *deep learning* tidak dapat memproses data dalam bentuk *string,* maka dari itu harus dikonversi ke dalam bentuk angka terlebih dahulu. Contoh sederhana untuk merubah teks menjadi vektor angka dengan menggunakan *one-hot encoding*. Misalnya ada kalimat berisi “Indonesia adalah negara berkembang” maka akan dirubah menjadi vektor untuk masing masing kata dimana posisi kata direpresentasikan dengan nilai 1. Vektor yang merepresentasikan kata “Indonesia” adalah [1,0,0,0], kata “adalah” dengan vektor [0,1,0,0], “negara” direpresentasikan dengan [0,0,1,0] dan kata “berkembang” menjadi [0,0,0,1]. Selain menggunakan *one-hot encoding,* ada metode lain yang sering digunakan untuk melakukan *word embedding* yaitu *word2vec*.

*Word2vec* merupakan metode yang digunakan untuk melakukan transformasi dari teks menjadi suatu vektor angka. Proses *word embbeding* dengan metode *word2vec* menggunakan konsep *neural network* yang memetakan kata ke variabel target (Abdullah, 2018). Sama halnya dengan *neural network*, proses yang dilakukan dengan metode *word2vec* ini menggunaka *weight* sebagai representasi vektor kata. *Word2vec* akan mengetahui hubungan semantik/sintatikal antar kata diruang vektor. Pada implementasinya, *word2vec* memiliki dua teknik untuk melakukan proses *word embedding* yaitu *Continous Bag of Words* dan *Skip-Gram Model*. Misalnya ingin membuat model dari suatu data training yang terdiri nilai input berupa kata “king” dan “brave” maka hasilnya adalah “man”. Alur kerja yang dilakukan untuk membuat model tersebut apabila menggunakan *continous bag of words* adalah sebagai berikut:

1. Lakukan proses transformasi pada layer input dan layer output (target) menggunakan *one-hot encoding* sehingga vektor yang dihasilkan berukuran . Dimana nilai adalah jumlah kata dari hasil *tokenizing*. Misalnya pada contoh ini memiliki 6 kata maka matriks input berukuran dimana dan 2 adalah kata “king” dan “brave”. Bentuk matriks input yang dihasilkan dapat dilihat pada Formula 2.1.

(2.1)

Sedangkan matriks dari layer output berukuran seperti pada Formula 2.2.

(2.2)

2. Tahap selanjutnya yaitu melakukan pembobotan. Proses pembobotan dilakukan dua kali yaitu di antara *input layer* dengan *hidden layer* dan di antara *hidden layer* dengan *output layer*. Matriks *input hidden layer* berukuran dan matriks dari *output hidden layer* berukuran . Dimana nilai N merupakan jumlah *neuron*. Misalnya jumlah *neuron*nya adalah 4 maka matriks *input hidden layer* memiliki ukuran seperti pada Formula 2.3.

(2.3)

Sedangkan ukuran matriks dari *output hidden layer* adalah dapat dilihat pada Formula 2.4.

(2.4)

3. Kemudian *input layer* dikalikan dengan *weight* dari *input hidden layer*. Hasil dari proses ini disebut dengan *hidden activation*. Berikut persamaan untuk menghasilkan *hidden activation* pada Formula 2.5. Pada iterasi pertama, bobot akan diberikan secara random.

(2.5)

Setelah itu, cari nilai rata-rata dari vektor *hidden activation*. Hasil rata-rata *hidden activation* pada Formula 2.6.

(2.6)

4. Lakukan proses perkalian antara *hidden activation* dengan *hidden-output weight* untuk memperoleh *output*. Proses perkalian tersebut seperti pada Formula 2.7.

(2.7)

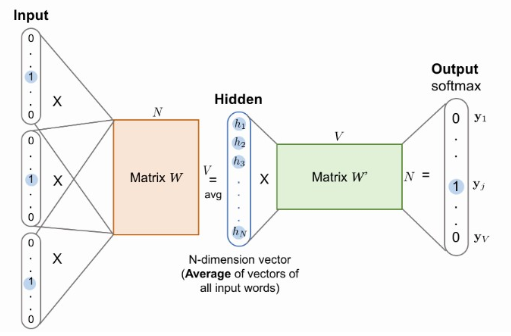
5. Setelah mendapat nilai *output,* maka *output* tersebut ditransformasikan menggunakan fungsi untuk mendapatkan nilai probabilitasnya. Berikut hasil probabilitas dari *output* pada Formula 2.8.

(2.8)

6. Hitung nilai *error* antara *output* dengan target kata. Setelah itu dilakukan *backpropagation* untuk *re-adjust weight*nya untuk memperbaiki nilainya. Cara menghitung yaitu matriks target dikurang matriks *output* seperti pada Formula 2.9.

(2.9)

Berdasarkan alur kerja tersebut, maka dapat diilustrasikan proses dari teknik *continous bag of words* seperti pada Gambar 2.8.



**Gambar 2.8** Ilustrasi *Continous Bag of Words* (Abdullah, 2018)

Sedangkan apabila menggunakan teknik *Skip-Gram Model,* alur kerja yang dilakukan sebagai berikut (Abdullah, 2018):

1. Lakukan transformasi dari nilai input menjadi berukuran , matriks *input hidden weight* berukuran dimana N merupakan jumlah *neuron* dan ukuran matriks dari *hidden output* adalah .

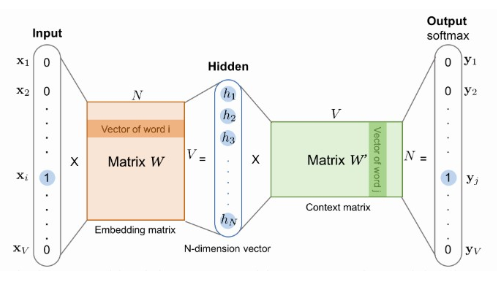
2. C adalah jumlah konteks katanya.

3. Kemudian nilai matriks dari *hidden activation*  dikalikan dengan *weight* yang dilakukan di antara *hidden layer* dan *output layer* untuk menghasilkan *output* dari prediksi.

4. Hasil dari proses sebelumnya berupa nilai *output* ditransformasikan menggunakan fungsi untuk mendapatkan nilai probabilitasnya.

5. Hitung *error* atau selisih antara nilai *output* dengan target. Kemudian dilakukan proses *backpropagation* untuk *re-adjust weight*nya.

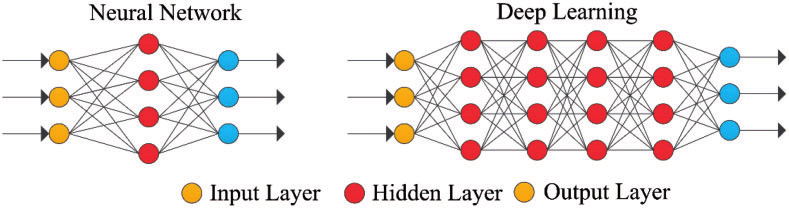
Alur kerja dari *skip-gram model* lebih jelasnya dapat dilihat ilustrasi pada Gambar 2.9.



**Gambar 2.9** Ilustrasi *Skip-gram Model* (Abdullah, 2018)

## 2.8 *Deep Learning*

*Deep learning* adalah salah satu bidang dari pembelajaran mesin (*machine learning*) yang berbasis jaringan syaraf tiruan atau *neural network*. Sama halnya dengan *neural network, deep learning* juga akan mengajarkan kepada komputer untuk melakukan tindakan seperti manusia, dimana tindakan atau keputusan tersebut berdasarkan hasil belajar dari contoh yang diberikan. *Deep learning* merupakan sebuah pendekatan dalam penyelesaian masalah pada sistem pembelajaran komputer yang menggunakan konsep hirarki (Goodfellow, Bengio, & Courville, 2016). Hirarki pada *deep learning* terdiri dari banyaknya struktur dan jumlah jaringan syaraf pada algoritmanya hingga ratusan lapisan yang setiap lapisannya memiliki tanggung jawabnya masing-masing. Secara umum, lapisan dari *deep learning* terdiri dari tiga bagian yatu *input layer, hidden layer* dan *output layer*. Pada bagian *hidden layer* ini lah yang dapat dibuat berlapis-lapis untuk menemukan komposisi algoritma yang tepat agar meminimalisir *error* pada *output* (Primas, 2017). Arsitektur dari *deep learning* dapat diilustrasikan seperti pada Gambar 2.10.

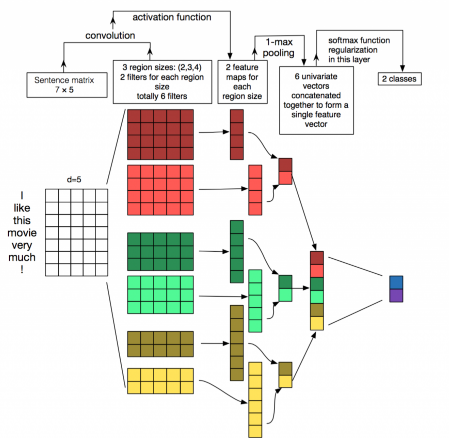


**Gambar 2.10** Diagram *Deep Learning* (Xing & Du, 2018)

*Deep learning* dapat digunakan untuk menangani berbagai permasalahan dengan data berskala besar seperti *computer vision, speech recognition* dan *natural language processing*. Kelebihan dari *deep learning* yaitu memiliki *feature engineering* yang berfungsi untuk mengekstrak pola yang penting dari suatu data sehingga dapat memudahkan model untuk membedakan suatu kelas. Model merupakan hasil dari proses pelatihan pada *deep learning* yang nantinya akan digunakan untuk melakukan prediksi. Algoritma yang digunakan pada *feature engineering* harus disesuaikan dengan kumpulan data dan jenis data yang digunakan. Pemilihan algoritma tersebut berdampak pada baik atau buruknya model untuk melakukan prediksi. Ada beberapa algoritma yang menerapkan konsep dari *deep learning* untuk melakukan *feature engineering* antara lain *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN) dan *Hierarchical Attention Network* (HAN).

## 2.9 *Convolutional Neural Network* (CNN)

*Convolutional neural network* (CNN) adalah salah satu algoritma dari *deep learning*. CNN merupakan jaringan syaraf yang dikhususkan untuk mengolah data yang memiliki grid. Sebenarnya, CNN merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang di desain untuk mengolah data dua dimensi. Namun, ada beberapa penelitian yang menerapkan algoritma CNN pada *Natural Language Processing* (NLP) dengan hasil yang cukup memuaskan. Jadi, selain digunakan untuk mengolah data dua dimensi, CNN juga dapat digunakan untuk mengolah data satu dimensi termasuk melakukan klasifikasi teks (Razi, 2017) . Konsep dari CNN 1D tidak jauh berbeda dari *neural network* biasa dimana terdiri dari *neuron* yang memiliki *weight, bias* dan *activation function*. Secara garis besar, arsitektur dari CNN 1D dibagi menjadi beberapa tahapan yaitu *input layer, convolutional layer, activation layer, pooling layer* dan *fully connected layer.* Arsitektur dari CNN 1D dapat dilihat pada Gambar 2.11:



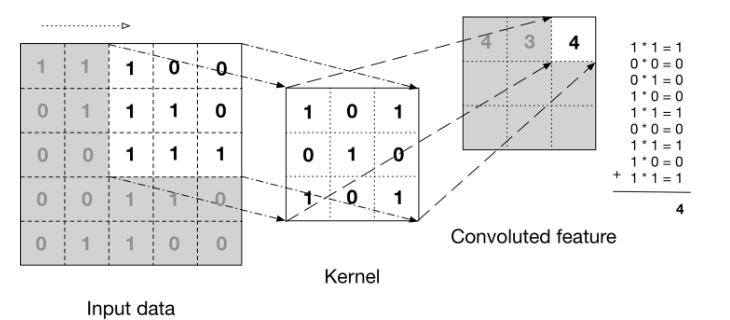
**Gambar 2.11** Arsitektur *Convolutional Neural Network* (Zhang & Wallace, 2015)

### 2.9.1 *Input Layer*

Layer input adalah layer pertama pada algoritma CNN. Layer ini berfungsi untuk menampung data yang menjadi inputan. Data yang ditampung berupa nilai vektor yang merepresentasikan dari masing-masing kata yang merupakan hasil dari proses *word embedding*. Ukuran layer ini sesuai dengan banyaknya kata dikali dengan panjang vektor atau dimensi vektor. Misalnya terdapat data yang terdiri dari 5 kata dan masing-masing kata direpresentasikan dengan 100 dimensi vektor, maka layer input berukuran .

### 2.9.2 *Convolution Layer*

*Convolution layer* adalah proses utama dari CNN. Pada tahap ini, *convolutional layer* akan menggunakan filter atau kernel pada setiap kumpulan vektor yang menjadi masukan. Kernel ini berupa array dua dimensi, bisa berukuran , , atau . Pada proses ini akan menghasilkan *feature map* yang menunjukkan fitur atau ciri dari kumpulan vektor inputan. Hasil *feature map* merupakan jumlah dari proses perkalian antara input data dengan kernel dengan indeks yang sama dan melakukan pergeseran kernel untuk nilai dari indeks berikutnya. Proses perkalian antara input data dengan kernel untuk menghasilkan *feature map* dapat dilihat pada Gambar 2.12.



**Gambar 2.12** *Convolution Layer* pada CNN (Yin, 2018)

Berdasarkan penjelasan tersebut dapat dirumuskan persamaan untuk menghasilkan *feature map* dengan inputan data , kernel dilambangkan dengan dan sebagai *feature map* yang dihasilkan pada Formula 2.10:

(2.10)

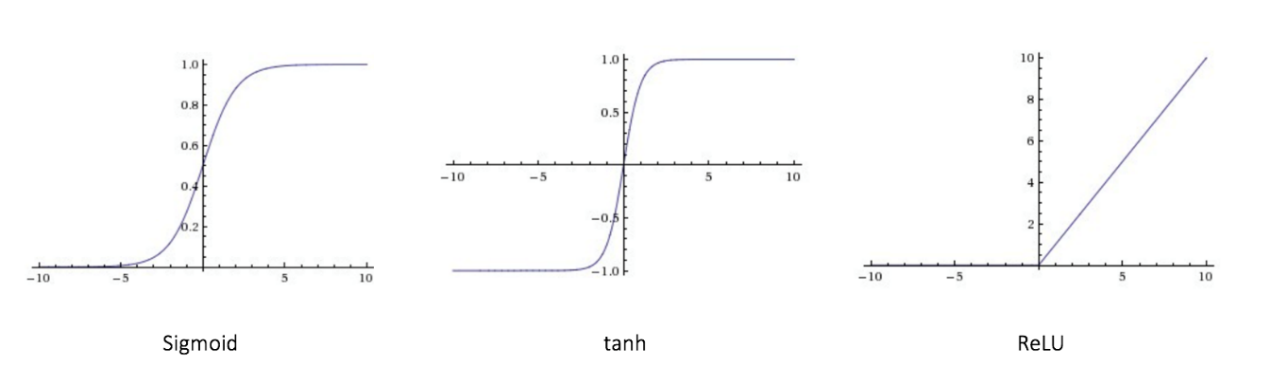
*Convolution layer* memiliki *hyperparameter* yang nilainya mempengaruhi *feature map* yang akan dihasilkan. *Hyperparameter* yang digunakan pada proses konvolusi yaitu *depth, stride* dan *zero-padding*. Penjelasan dari masing-masing *hyperparameter* dapat dilihat pada Tabel 2.1.

**Tabel 2.1** *Hyperparameter* pada *convolutional layer* (Zufar & Setiyono, 2016)

|  |  |  |
| --- | --- | --- |
| **No** | ***Hyperparameter*** | **Keterangan** |
| 1 | *Depth* | Kedalaman *layer* atau jumlah *layer* konvolusi |
| 2 | *Stride* | Jumlah pergeseran *filter* pada proses konvolusi |
| 3 | *Zero-padding* | Jumlah penambahan nilai intensitas nol di daerah sekitar input gambar |

### 2.9.3 *Activation Layer*

Tahap selanjutnya yaitu *activation layer*. Proses pada *activation layer* yaitu dengan memasukkan nilai *feature map* yang telah dihasilkan pada tahap sebelumnya ke dalam fungsi aktivasi. Fungsi aktivasi berfungsi untuk mengubah nilai pada *feature map* pada range tertentu dengan tujuan untuk meneruskan nilai yang menjadi fitur dominan ke layer selanjutnya. Beberapa fungsi aktivasi yang sering digunakan secara umum yaitu sigmoid, tanh dan ReLu: Fungsi aktivasi tersebut memiliki hasil grafik yang berbeda-beda, visualisasi grafik dari masing-masing fungsi aktivasi seperti pada Gambar 2.13



**Gambar 2.13** Contoh Grafik Fungsi Aktivasi (Tarabay, 2019)

Fungsi sigmoid terbentuk dari pembagian antara nilai 1 dengan penjumlahan nilai 1 dan exponensial pangkat minus . Persamaan fungsi sigmoid dapat dibentuk seperti Formula 2.11:

(2.11)

Fungsi selanjutnya yaitu Tanh. Persamaan dari fungsi ini adalah selisih eksponensial dari input dengan eksponensial minus nilai input dibandingkan dengan jumlah dari eksponensial nilai input dan eksponensial minus nilai input. Persamaan tersebut dapat ditulis seperti pada Formula 2.12:

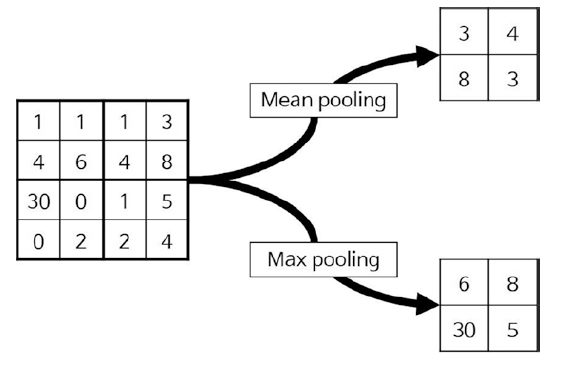
(2.12)

ReLU didefinisikan sebagai fungsi yang 2 kondisi untuk menghasikan nilainya. Jika nilai kurang dari 0 maka fungsi dari tersebut bernilai 0. Sedangkan apabila nilai lebih dari sama dengan 0 maka nilai dari fungsi adalah nilai tersebut. Persamaan dari fungsi ReLu dapat ditulis seperti pada Formula 2.13

(2.13)

### 2.9.4 *Pooling Layer*

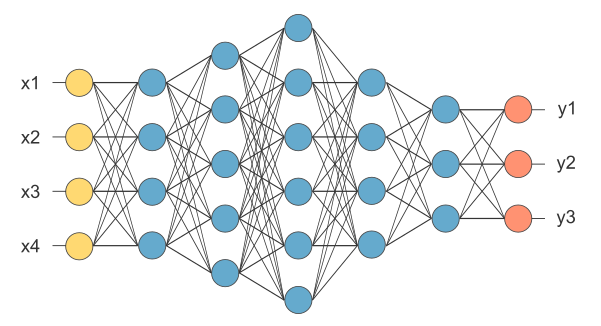
*Pooling layer* dapat disebut juga dengan *subsampling* atau *downsampling*. Hal tersebut karena *pooling layer* akan mengurangi dimensi dari *feature map* yang telah melalui *activation layer*. Pengurangan dimensi pada tahap ini tidak mempengaruhi informasi penting yang ada didalamnya. Proses *pooling* memiliki beberapa macam, seperti *maxpooling*, *meanpooling, sumpooling* dan sebagainya. Proses pada *pooling* dimulai dengan menentukan ukuran *downsampling,* misalnya . Setelah itu akan dilakukan proses *pooling* pada *feature map,* sebagai contoh *feature map* yang digunakan berukuran . Apabila menggunakan *maxpooling*, maka dari ukuran *downsampling* yang diambil nilai tertinggi untuk dijadikan nilai *feature map*. Sedangkan apabila menggunakan *meanpooling* maka nilai rata rata dari *downsampling* yang dijadikan nilai *feature map*. Proses *pooling* diilustrasikan seperti pada Gambar 2.14.



**Gambar 2.14** *Pooling Layer* (Tandungan, 2019)

### 2.9.5 *Fully Connected Layer*

*Fully connected layer* akan menggunakan hasil dari *pooling layer* sebagai inputan. Layer ini memiliki struktur yang sama dengan *neural network* pada umumnya yang terdiri dari *input layer*, *hidden layer* dan *output layer*. Masing-masing layer tersebut terdapat kumpulan *neuron* yang saling terhubungdengan *neuron-neuron* pada layer tetangganya. Berikut bentuk dari *full connected layer* dapat dilihat pada Gambar 2.15:



**Gambar 2.15** *Fully Connected Layer* (Tandungan, 2019)

Layer ini akan menentukan hasil klasifikasi berdasarkan data yang telah dimasukkan. Proses penentuan hasil klasifikasi menggunakan fungsi aktivasi . Fungsi aktivasi yaitu rasio eksponensial dari nilai input dengan jumlah nilai eksponensial. Persamaan dari fungsi aktivasi dapat dirumuskan seperti pada Formula 2.14:

(2.14)

dimana :

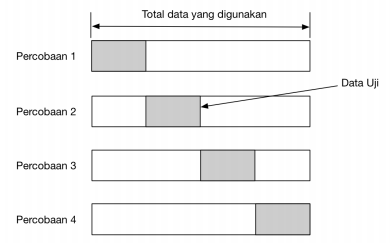
σ = probabilitas input tersebut merupakan class j

z = persamaan linear fungsi pre-aktivasi

j = salah satu class dari K class yang ada

## *K-Fold Cross Validation*

*K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengukur rata-rata tingkat keberhasilan dari suatu model. Hasil tersebut dilakukan dengan melakukan perulangan untuk menguji model dengan data yang acak. Awalnya, data yang dimiliki akan dibagi sejumlah *n-fold* yang diinginkan sehingga membentuk kelompok kelompok secara acak. Lalu akan dilakukan perulangan terhadap data yang telah dibagi menjadi data latih dan data uji secara bergantikan. Jadi setiap kelompok pernah menjadi data latih dan juga data uji. Berikut ilustrasi dari proses *k-fold cross validation* pada Gambar 2.16.



**Gambar 2.16** Ilustrasi *K-Fold Cross Validation*

## 2.11 *Confusion Matrix*

*Confusion matrix* merupakan metode yang sering digunakan untuk mengukur tingkat performa dari suatu model klasifikasi. *Confusion matrix* akan memberikan hasil perbandingan antara hasil klasifikasi yang dilakukan oleh model dengan hasil klasifikasi yang sebenarnya. Terdapat empat istilah yang merepresentasikan hasil dari proses klasifikasi pada *confusion matrix* yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Tabel *confusion matrix* dapat dilihat pada Tabel 3.

**Tabel 3.** *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Actual Values** | |
| Positive | Negative |
| **Predicted Values** | Positive | True Positive (TP) | False Positive (FP) |
| Negative | False Negative (FN) | True Negative (TN) |

Keterangan:

*True Positive* (TP) : Jumlah data positif yang diprediksi benar

*True Negative* (TN) : Jumlah data negatif yang diprediksi benar

*False Positive* (FP) : Jumlah data negatif yang diprediksi salah

*False Negative* (FN) : Jumlah data positif yang diprediksi salah

Berdasarkan tabel *confusion matrix* tersebut dapat menghitung *performance matrix* untuk mengukur model yang telah dibuat. *Performance matrix* yang paling sering digunakan yaitu akurasi, presisi dan recall.

1. **Akurasi**

Akurasi merupakan nilai yang menggambarkan seberapa akurat sebuah model dalam mengklasifikasikan dengan benar. Nilai akurasi ini merepresentasikan nilai kedekatan antara nilai prediksi dan nilai yang sebenarnya. Persamaan untuk menghasilkan nilai akurasi yaitu dengan membandingan nilai prediksi benar dengan keseluruhan data. Rumus menghitung nilai akurasi dapat dilihat pada persamaan 2.

X 100%

1. **Presisi**

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Rumus untuk menghitung nilai presisi dapat dilihat pada persamaan 2.

1. **Recall**

Recall merupakan perbandingan antara prediksi benar positif dengan keseluruhan data yang benar positif. Untuk menghitung nilai recall dapat menggunakan rumus yang dapat dilihat pada persamaan

.

## 2.12 Penelitian Terdahulu

Adapun beberapa penelitian yang telah dilakukan dan dijadikan referensi dalam penelitian ini dapat dilihat pada Tabel 2.

**Tabel 2.** Penelitian Terdauhulu

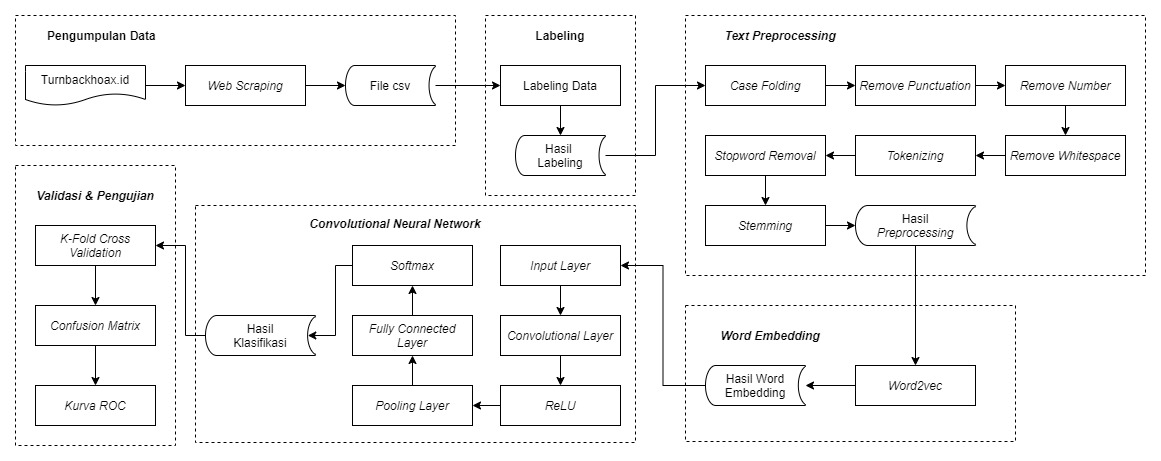
|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Penulis** | **Judul** | **Intisari** |
| 1 | Budi Setiawan Tanjung (2018) | Pendekatan *Text Mining* sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter | Sistem telah berhasil mengklasifikasi berita ke dalam tiga kategori yaitu *hoax, fact* dan *grey*. Namun, tingginya kemiripan antar data membuat sulit menemukan pola berita *hoax* sehingga mempengaruhi tingkat *error* dan akurasinya. |
| 2 | Dina Maulina dan Rofie Sagara (2018) | Klasifikasi Artikel *Hoax* Menggunakan *Support Vector Machine Linear* dengan Pembobotan *Term Frequency-Inverse Document Frequency* | Penelitian ini telah menghasilkan sistem dengan performa yang baik dengan akurasi 95.83% dan memiliki kecepatan training 1.37 detik untuk 240 vektor data. Namun, data yang digunakan relatif sedikit sehingga wawasan yang dihasilkan masih kurang optimal. |
| 3 | Aulia Afriza dan Julio Adisantoso (2018) | Metode Klasifikasi *Rocchio* untuk Analisis *Hoax* | Hasil penelitiannya menunjukan bahwa algoritma *rocchio* memiliki akurasi lebih tinggi dibandingkan multinomial naive bayes. Karena kenadala naive bayes yaitu jika probabilitas kondisionalnya sama dengan 0. |
| 4 | Kemas Muslim Lhaksmana, Fhira Nhita dan Ageng Budhiarto (2017) | Klasifikasi Pengguna Media Sosial Twitter dalam Persebaran *Hoax* Menggunakan Metode *Backpropagation* | Sistem yang dihasilkan memiliki performa yang cukup baik dengan akurasi 72% namun mendapatkan error yang cukup besar karena karakteristik yang hampir sama pada masing-masing kelas. Selain itu membutuhkan waktu yang relatif lama saat proses pelatihan. |
| 5 | Mark Hughes, Irene Li, Spyros Kotoulas dan Toyotaro Suzumura (2017) | *Medical Text Classification using Convolutional Neural Network* | Hasil penelitian menunjukkan bahwa *convolutional neural network* dapat melakukan klasifikasi teks dengan hasil yang cukup meskipun biasanya metode ini digunakan untuk klasifikasi citra. Selain itu, dapat menghasilkan fitur yang lebih optimal dibandingan metode pembelajaran yang dangkal. |
| 6 | Ar Razi (2017) | Klasifikasi Artikel Berita Berbahasa Indonesia Menggunakan *Convolutional Neural Network* | Penelitian yang dihasilkan memiliki akurasi 96,70% dan presisi, recall serta f-measure 96,60%. Namun, dalam penelitian ini hanya menggunakan dataset yang relatif sedikit dan pada bagian *preprocessing* tidak menggunakan *stemming* serta *stopword* sehingga hasilnya kurang optimal. |

# BAB III

# METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM

## 3.1 Metodologi Penelitian

Pada bagian ini akan membahas tentang metodologi penelitian yang digunakan pada penelitian ini. Metodologi penelitian berisi tahapan-tahapan perencanaan yang akan dilakukan selama proses pembuatan sistem pada penelitian ini. Metodologi penelitian bertujuan untuk menjadi bahan acuan dalam meminimalkan risiko kegagalan sehingga mencapai tujuan yang diinginkan. Adapun tahapan metodologi penelitian yang dilakukan dapat dilihat pada Gambar 3.1.



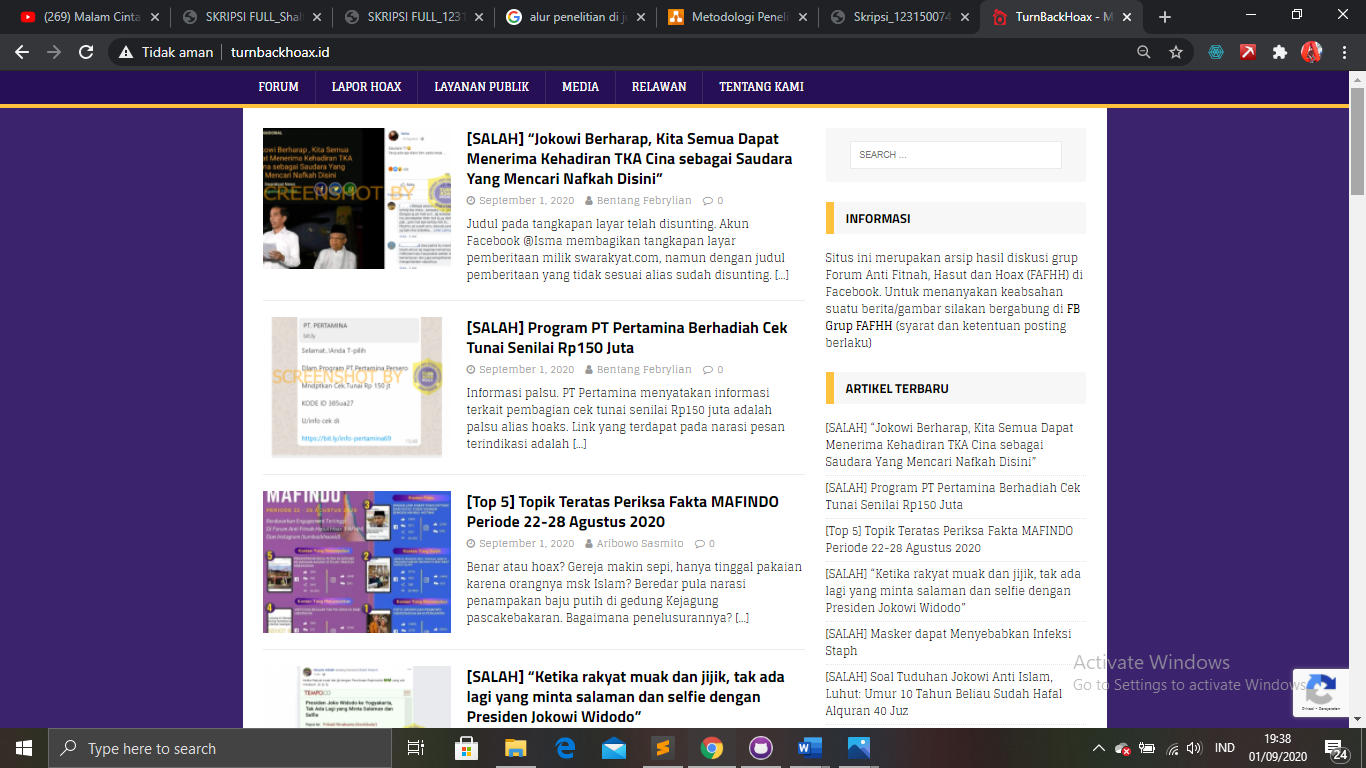
**Gambar 3.1** Metodologi Penelitian

Secara garis besar, terdapat 6 tahapan yang akan dilakukan yaitu pengumpulan data, labeling, *text preprocessing, word embedding* menggunakan *word2vec,* algoritma *convolutional neural network* serta validasi & pengujian.

### 3.1.1 Pengumpulan Data

Data merupakan salah satu hal yang paling penting dalam sebuah penelitian. Tanpa adanya data, suatu penelitian tidak dapat dilakukan. Suatu penelitian memerlukan data untuk diolah sehingga menjadi sebuah informasi yang dapat membantu untuk pengambilan keputusan atau menghasilkan teori. Maka dari itu, data yang diperlukan dalam penelitian harus dikumpulkan terlebih dahulu. Selain itu, data yang digunakan juga harus dapat dipertanggungjawabkan kebenarannya.

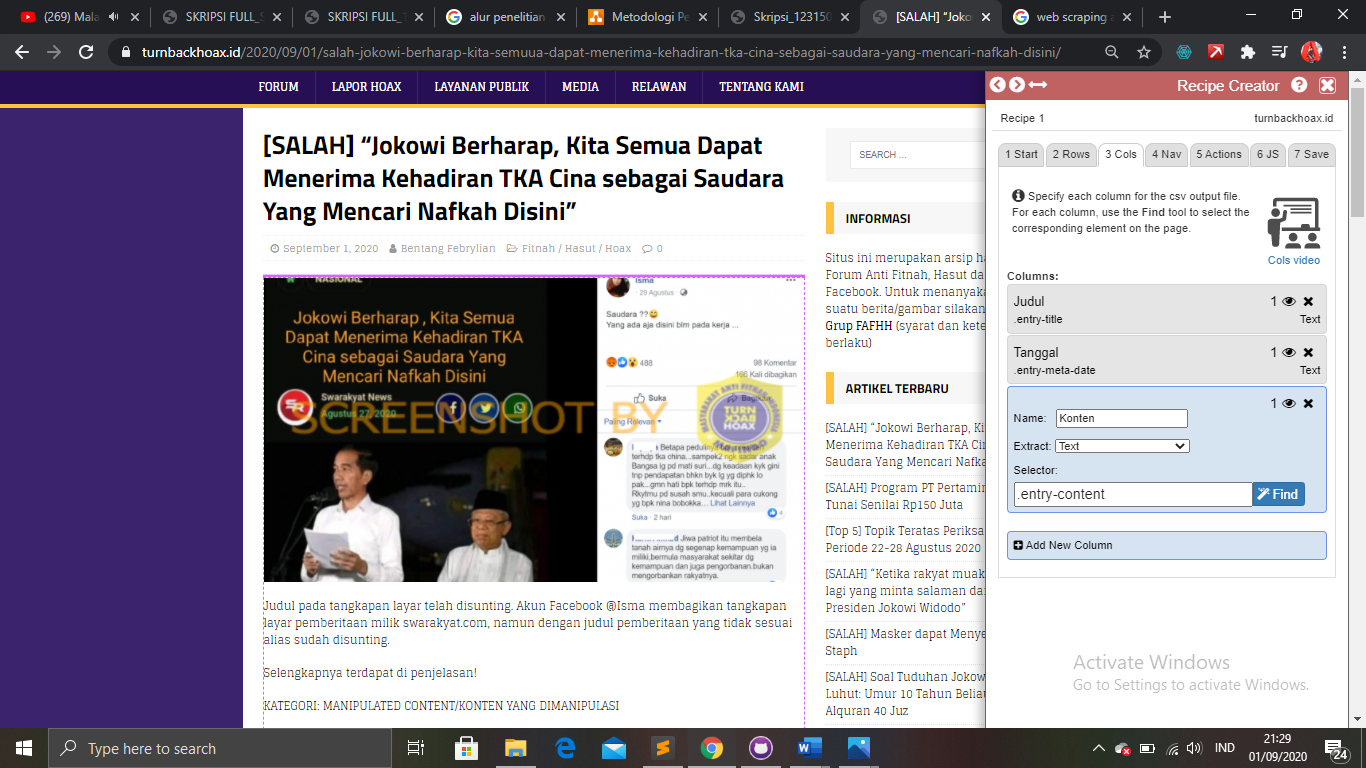
Pada penelitian ini akan menggunakan data berita yang mengandung unsur *hoax* dan berita yang valid kebenarannya. Data yang digunakan adalah data sekunder yang bersumber dari situs turnbackhoax.id. Data yang diambil mulai dari berita tahun 2015 sampai dengan 2019 yang nantinya akan dikategorikan menjadi dua kelas yaitu *hoax* dan fakta.. Alasan situs tersebut menjadi sumber data penelitian karena pada situs tersebut telah memvalidasi kebenaran dari berita-berita yang tersebar di media sosial. Situs tersebut akan melakukan penelusuran berdasarkan aduan dari masyarakat yang menemukan berita *hoax* di media sosial. Berikut tampilan awal situs turnbackhoax.id dapat dilihat pada Gambar 3.2.



**Gambar 3.2** Tampilan Situs Turnbackhoax.id

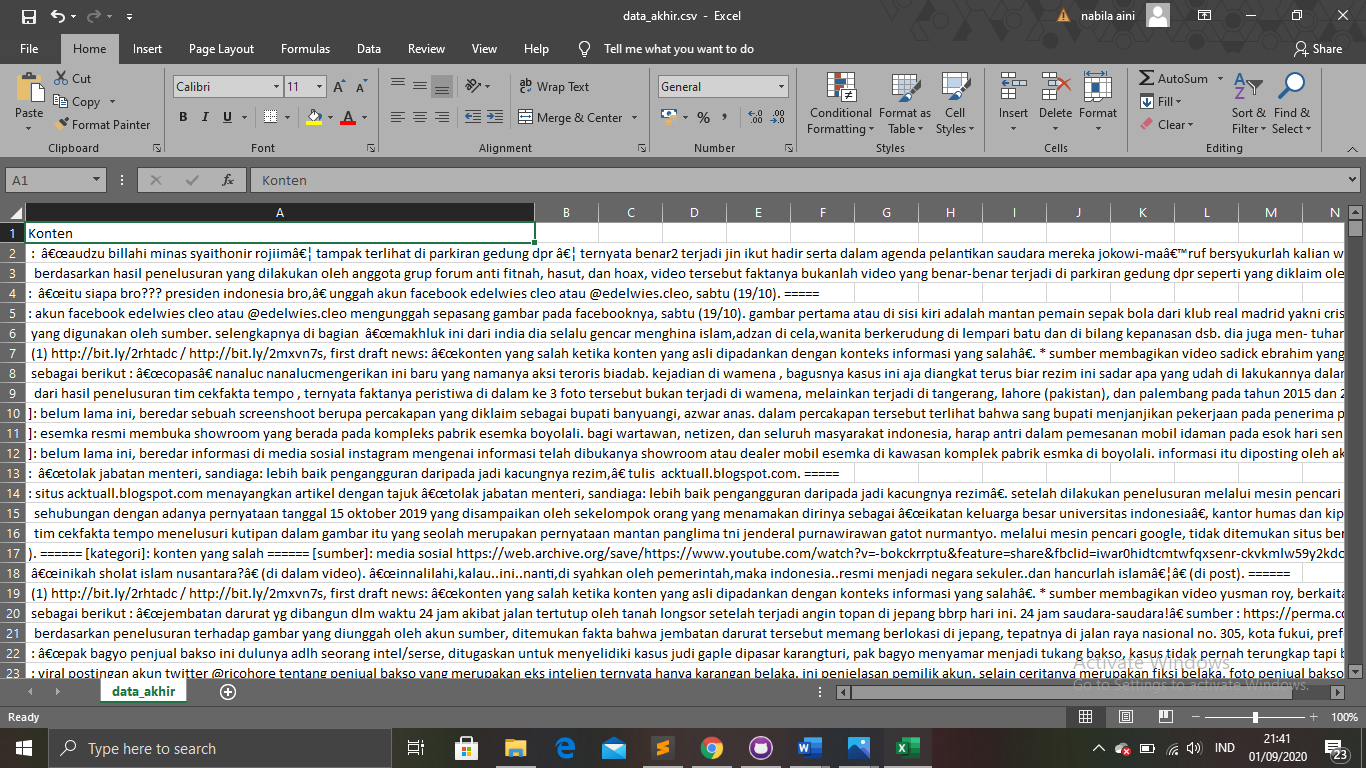
Selanjutnya, akan dilakukan proses pengambilan data dari situs tersebut menggunakan metode *web scraping*. *Web scraping* merupakan teknik pengumpulan data dengan mengekstraksi data dan informasi dari suatu website lalu disimpan ke dalam format tertentu. Dengan menggunakan metode *web scraping* bertujuan untuk memudahkan pengambilan data dari situs tersebut karena dilakukan secara otomatis tanpa harus menyalin data satu per satu. *Web scraping* memiliki beberapa metode yang dapat digunakan, salah satunya menggunakan *extension browser* yang dilakukan pada penelitian ini. *Extension browser* yang digunakan yaitu *Data Miner* dari *browser* Google.

Proses mengambilan data dengan *data miner* cukup mudah karena dapat dilakukan dan dipahami oleh siapapun tanpa harus menguasai pemrograman. Pertama, *install extension data miner* terlebih dahulu di browser google. Kedua, buka situs yang ingin diambil datanya, dan pada penelitian ini pada situs turnbackhoax.id. Setelah itu, buka *data miner* dan pilih bagian website yang memuat data dan informasi yang ingin diambil. Selanjutnya, pilih link *pagination* agar data yang diambil dilakukan secara berulang ke data berikutnya. Proses *web scraping* menggunakan *data miner* dapat dilihat pada Gambar 3.3.



**Gambar 3.3** Proses *Web Scraping* Menggunakan *Data Miner*

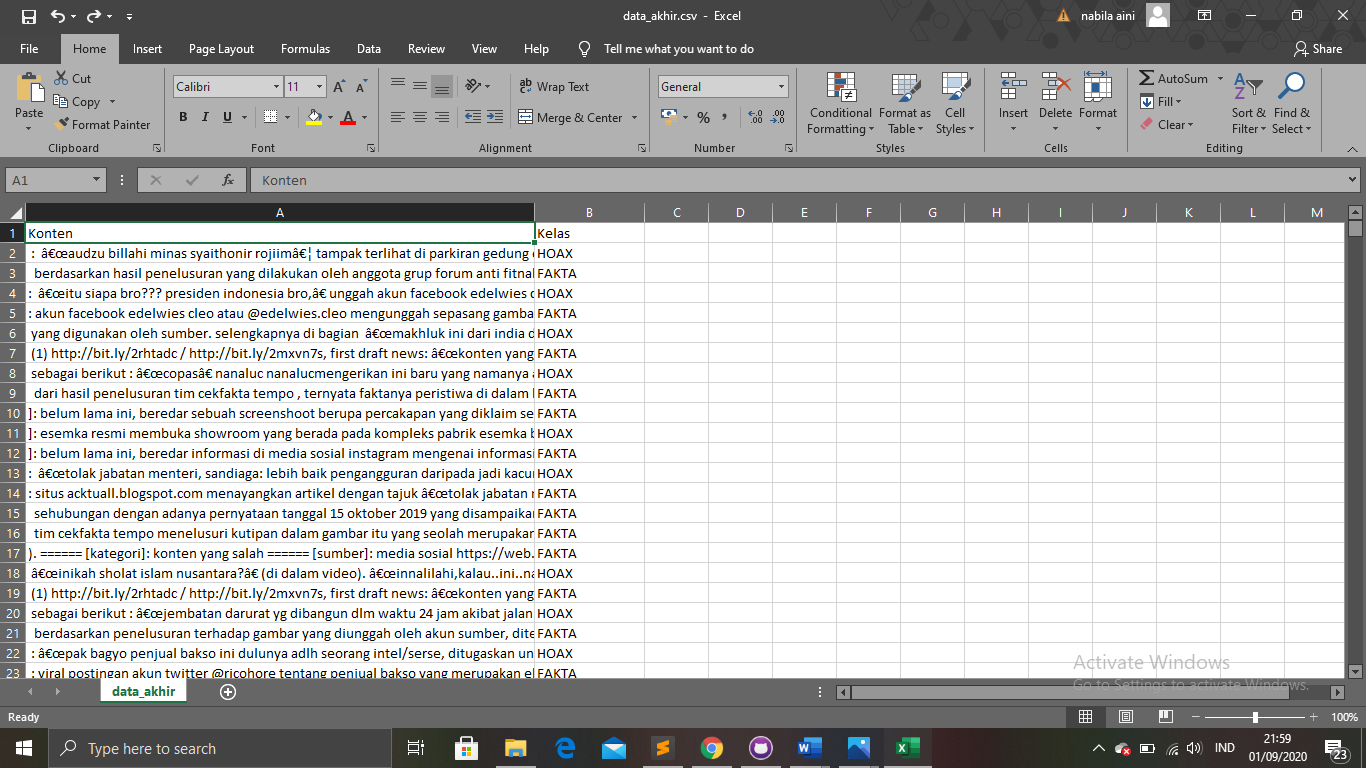
Selanjutnya, hasil data yang telah diambil dapat dilihat apakah sudah sesuai atau belum. Apabila data yang di*scrape* telah sesuai keinginan, maka data tersebut dapat disimpan dalam format csv. Berikut bentuk dataset yang telah diubah ke format csv yang bersumber dari data turnbackhoax.id.



**Gambar 3.4** Dataset Hasil *Web Scraping*

### 3.1.2 *Labeling*

*Labeling* merupakan tahapan untuk memberikan label atau kategori pada setiap data. Semua data akan dikategorikan menjadi 2 kelas yaitu *hoax* atau fakta. Proses pemberian label pada data sesuai dengan judul dari postingan turnbackhoax.id, apabila pada judul tertera tanda “[SALAH]” maka narasi dari data tersebut termasuk kategori *hoax* dan bagian dari penjelasan akan diberi label fakta. Dan apabila ada postingan yang tidak memiliki tanda “[SALAH]” maka akan diberi label fakta. Berikut hasil perubahan dataset setelah dilakukan proses *labeling* dapat dilihat pada Gambar 3.5.



**Gambar 3.5** Dataset Hasil *Labeling*

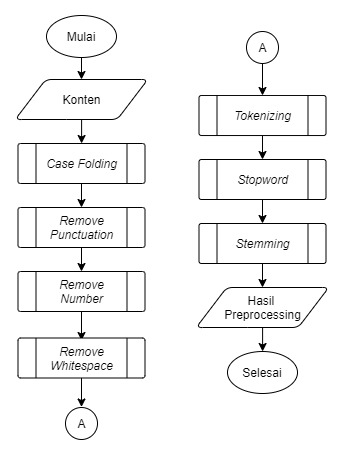
Berdasarkan hasil dari proses *labeling* dapat diketahui rincian dari dataset yang akan digunakan. Data yang didapatkan terdiri dari 1899 berita yang berkategori *hoax* dan 2450 berita yang berkategori fakta. Jadi apabila diakumulasikan maka total dataset yang digunakan pada penelitian ini berjumlah 4349 data. Rincian dari jumlah dataset pada penelitian ini dapat dilihat pada Tabel 3.1

**Tabel 3.1** Rincian Dataset

|  |  |  |
| --- | --- | --- |
| **No** | **Label Klasifikasi** | **Jumlah Data** |
| 1 | Hoax | 1899 |
| 2 | Fakta | 2450 |
| **Total Data** | | **4349** |

### 3.1.3 *Text Preprocessing*

Pada bagianini akan menggambarkan alur proses pada tahap *text preprocessing*. *Text preprocessing* ini bertujuan untuk membersihkan data konten yang menjadi *input* dari sistem. Hal tersebut membuat data yang awalnya semi-terstruktur atau bahkan tidak terstruktur menjadi data yang terstruktur dan siap untuk digunakan. *Text preprocessing* memiliki beberapa tahapan yang menjadi sub proses diantaranya *case folding, remove punctuation, remove number, remove whitespace, tokenizing, stopword* dan *stemming*. *Flowchart* dari *text preprocessing* dapat dilihat pada Gambar 3.6.

****

**Gambar 3.6** *Flowchart Text Preprocessing*

Berikut adalah contoh dari hasil *text preprocessing* terhadap sebuah berita berbahasa Indonesia setelah melalui semua proses *text preprocessing*.

Text awal:

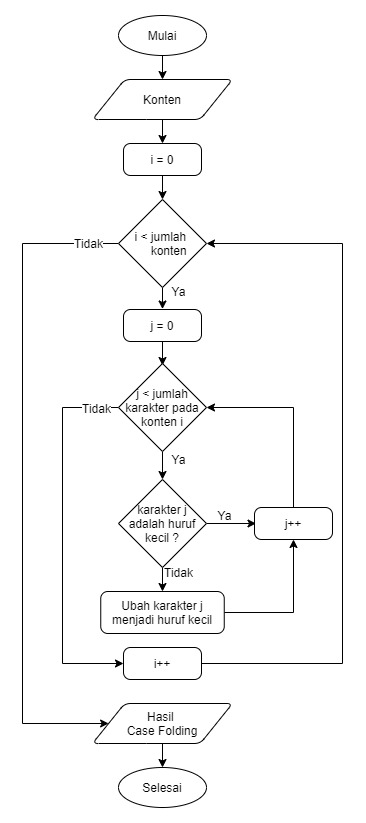
“Terlibat pembuatan virus corona di China, FBI menangkap 10 profesor Harvard.”

Hasil *Preprocessing*:

“['libat', 'buat', 'virus', 'corona', 'china', 'fbi', 'tangkap', 'profesor', 'harvard']”

#### 3.1.3.1 *Case Folding*

*Case folding* yaitu proses merubah semua karakter pada dokumen menjadi huruf kecil (*lowercase*). *Case folding* akan menerima *input* berupa konten dan inisialisasi i yang mewaliki jumlah dan j mewakili jumlah karakter. Prosesnya yaitu melakukan pengecekan ke setiap karakter dari awal hingga akhir. Jika karakter merupakan huruf kecil, maka karakter tersebut disimpan. Tetapi apabila menemukan karakter yang huruf besar maka dirubah terlebih dahulu karakter tersebut menjadi huruf kecil, lalu perubahan tersebut di simpan. Ketika semua karakter telah diperiksa, maka akan mengeluarkan *output* berupa hasil *case folding* yang akan digunakan pada proses berikutnya. *Flowchart* dari proses *case folding* dapat dilihat pada Gambar 3.7.

****

**Gambar 3.7** *Flowchart Case Folding*

Berikut contoh hasil dari proses *case folding* terhadap suatu konten berita yang ingin diperiksa.

Text awal:

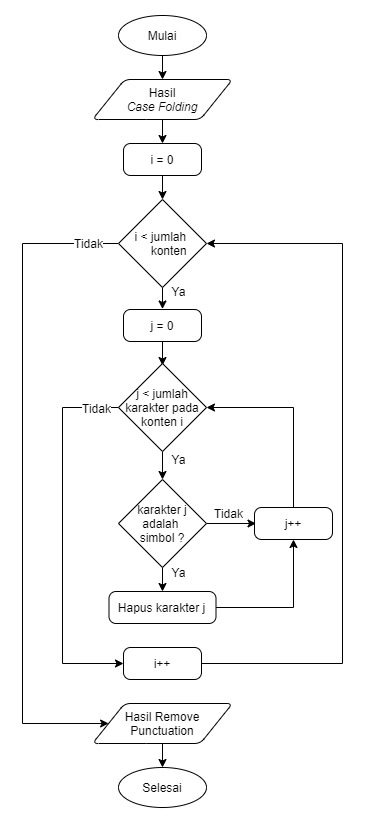
“Terlibat pembuatan virus corona di China, FBI menangkap 10 profesor Harvard.”

Hasil *case folding*:

“terlibat pembuatan virus corona di china, fbi menangkap 10 profesor harvard.”

#### 3.1.3.2 *Remove Punctuation*

*Remove punctuation* adalah proses menghapus tanda baca pada suatu dokumen karena dianggap sebagai *delimiter*. Penghapusan tanda baca ini akan mengurangi beban pemprosesan. Beberapa contoh tanda baca seperti titik (.), koma (,), tanda tanya (?), tanda seru (!) dan lainnya. Proses yang terjadi yaitu melakukan pengecekan jumlah dokumen dan jumlah karakternya. Data yang digunakan pada proses ini merupakan data hasil dari proses *case folding.* Kemudian, setiap karakter dari data tersebut akan dicek apakah karakter merupakan simbol atau tanda baca. Jika karakter merupakan tanda baca maka akan dihapus, tetapi jika bukan tanda baca maka karakter tersebut akan disimpan. Apabila semua karakter telah diperiksa, akan memberikan *output* berupa hasil *remove punctuation*. Alur dari proses *remove punctuation* dapat dilihat lebih detail pada Gambar 3.8.

****

**Gambar 3.8** *Flowchart Remove Punctuation*

Berikut contoh hasil proses *remove punctuation* terhadap data yang merupakan hasil dari *case folding*.

Text *input* (hasil *case folding*):

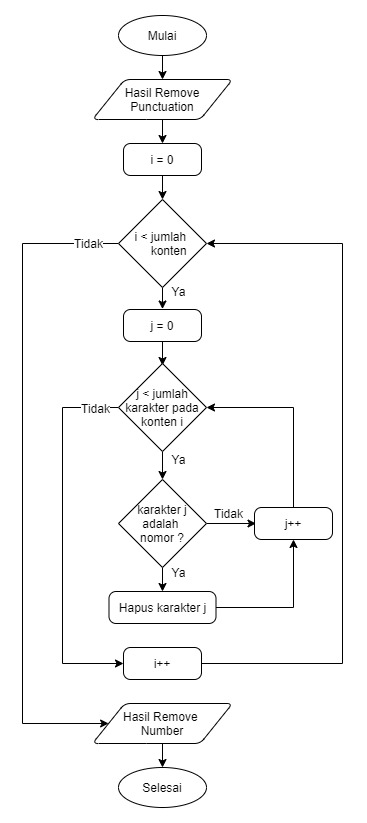
“terlibat pembuatan virus corona di china, fbi menangkap 10 profesor harvard.”

Text hasil *remove punctuation*:

“terlibat pembuatan virus corona di china fbi menangkap 10 profesor harvard”

#### 3.1.3.3 *Remove Number*

*Remove number* tidak jauh berbeda dengan *remove punctuation*, perbedaannya hanya pada karakter yang dihapus. Pada *remove number* ini akan menghapus karakter yang berbentuk angka. Pada proses ini akan menerima *input* berupa hasil dari proses *remove puctuation,* kemudian akan diperiksa per karakternya. Apabila menemukan karakter berbentuk angka, maka karakter tersebut akan dihapus. *Flowchart* dari *remove number* dapat dilihat pada Gambar 3.9.

****

**Gambar 3.9** *Flowchart Remove Number*

Berikut contoh hasil proses *remove number* terhadap data yang merupakan hasil dari *remove punctuation*.

Text *input* (hasil *remove punctuation*):

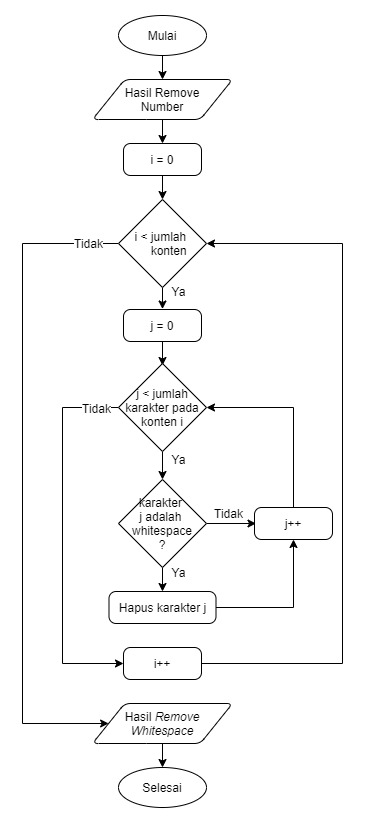
“terlibat pembuatan virus corona di china fbi menangkap 10 profesor harvard”

Text hasil *remove number*:

“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

#### 3.1.3.4 *Remove Whitespace*

*Remove whitespace* merupakan proses penghapusan *whitespace* dari data. Proses ini akan mengecek setiap karakter apakah termasuk *whitespace* atau bukan. Jika karakter merupakan *whitespace* maka karakter tersebut akan dihapus, sebaliknya jika tidak maka karakter akan disimpan. Data yang akan diproses merupakan hasil dari *remove number* dan hasil dari proses ini menghasilkan hasil *remove whitespace*. *Flowchart* dari *remove whitespace* dapat dilihat pada Gambar 3.10.

****

**Gambar 3.10** *Flowchart Remove Whitespace*

Berikut contoh hasil proses *remove whitespace* terhadap data yang merupakan hasil dari *remove number*.

Text *input* (hasil *remove number*):

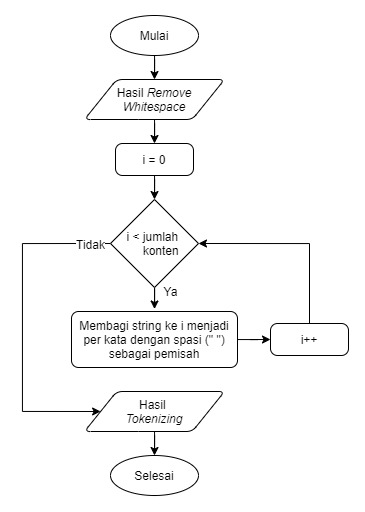
“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

Text hasil *remove whitespace*:

“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

#### 3.1.3.5 *Tokenizing*

Pada bagianini akan menjelaskan proses *tokenizing*. Proses ini kan memotong data menjadi per kata dengan spasi (“ “) sebagai pemisahnya. Data yang akan diproses dari hasil *remove whitespace* dan akan mengeluarkan hasil *tokenizing* dalam bentuk token-token (kata). *Flowchart* dari proses *tokenizing* dapat dilihat pada Gambar 3.11.

****

**Gambar 3.11** *Flowchart Tokenizing*

Berikut contoh hasil proses *tokenizing* terhadap data yang merupakan hasil dari *remove whitespace*.

Text *input* (hasil *remove whitespace*):

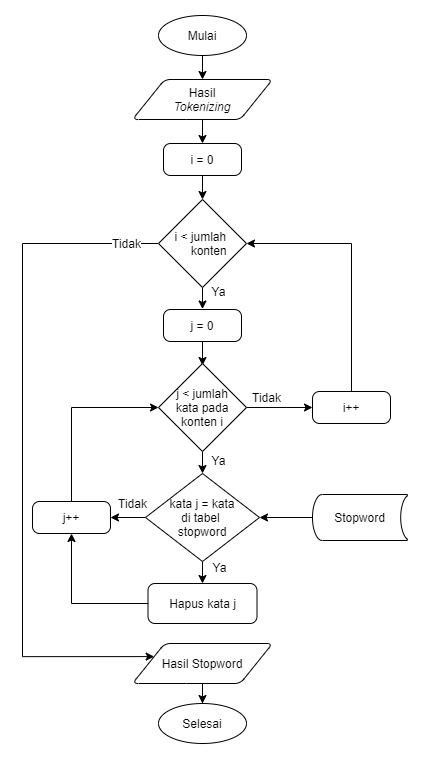
“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

Text hasil *tokenizing*:

“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

#### 3.1.3.6 *Stopword Removal*

*Stopword removal* adalah proses pembuangan kata-kata yang kurang penting (*stopword*) pada data. Contoh kata yang kurang penting seperti kata ganti orang, kata penghubung dan sebagainya. Proses pada tahap ini yaitu melakukan pengecekan per kata dari hasil *tokenizing*. Lalu, setiap kata akan dicek apakah kata tersebut ada di kamus *stopword*. Apabila kata di data ada di kamus *stopword,* maka kata tersebut dihapus. Sebaliknya, jika tidak ada di kamus *stopword* maka akan disimpan. Proses *stopword* dapat dilihat pada Gambar 3.12.

****

**Gambar 3.12** *Flowchart Stopword*

Berikut contoh hasil proses *stopword remova;* terhadap data yang merupakan hasil dari *tokenizing*.

Text *input* (hasil *tokenizing*):

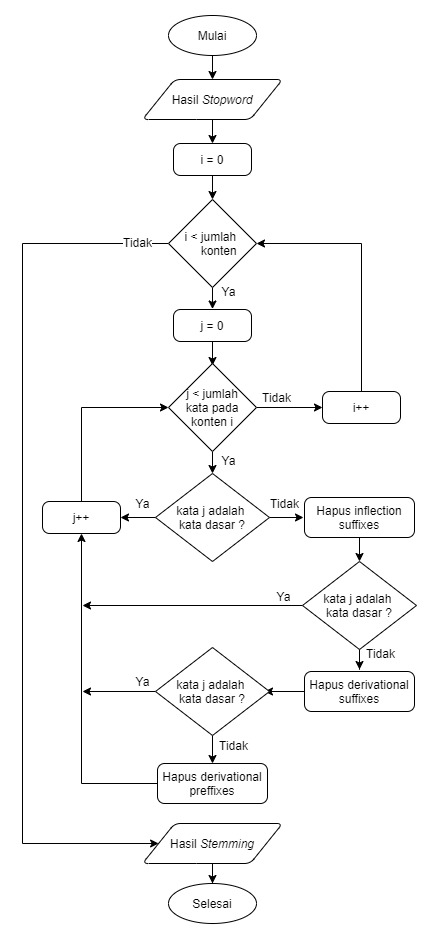
“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

Text hasil *stopword removal*:

“['terlibat', 'pembuatan', 'virus', 'corona', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

#### 3.1.3.7 *Stemming*

Pada bagian ini akan menjelaskan terkait alur dari proses *stemming*. Proses ini akan menggunakan data hasil dari proses *stopword* yang akan diperiksa per kata. *Stemming* akan mencari kata dasar dari setiap kata dengan cara menghapus imbuhan pada kata. Jika kata merupakan kata dasar maka kata tersebut akan disimpan. Apabila kata bukan kata dasar, maka akan di hapus *inflection suffixes, derivational suffixes* dan atau *derivational suffixes* yang ada pada kata sehingga kata tersebut menjadi kata dasar. Setelah itu kata disimpan dan *flowchart* ini akan menghasilkan *output* berupa hasil *stemming* sekaligus menjadi hasil dari proses *text preprocessing*. Berikut *flowchart* dari *stemming* dapat dilihat pada Gambar 3.13.

****

**Gambar 3.13** *Flowchart Stemming*

Berikut contoh hasil proses *stemming* terhadap data yang merupakan hasil dari *stopword removal*.

Text *input* (hasil *stopword removal*):

“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

Text hasil *stemming* dan hasil *preprocessing*:

“['libat', 'buat', 'virus', 'corona', 'china', 'fbi', 'tangkap', 'profesor', 'harvard']”

### 3.1.4 *Word Embedding*

Pada bagian ini akan dijelaskan terkait proses dari *word embedding* menggunakan *word2vec*. Proses ini akan merubah suatu kata menjadi suatu vektor yang berisi kumpulan angka. Kata yang menjadi nilai *input* pada *word embedding* berasal dari hasil proses *stemming*. Vektor yang dihasilkan dari proses ini akan merepresentasikan dari setiap kata yang ada. Vektor yang dihasilkan memiliki panjang tertentu yang dapat diatur, semakin panjang suatu vektor maka akan semakin teliti representasi yang dihasilkan. Panjang vektor pada proses *word embedding* dapat disebut dengan *dimensi*. Pada penelitian ini, dimensi dari vektor yang dihasilkan yaitu 100. Namun, untuk memudahkan dalam memberikan contoh maka akan ditampilkan vektor dari setiap kata dengan jumlah dimensi 5. Berikut *flowchart* dari proses *word embedding* dapat dilihat pada Gambar 3.14 dan hasil dari proses *word embedding* dapat dilihat pada Tabel 3.2.

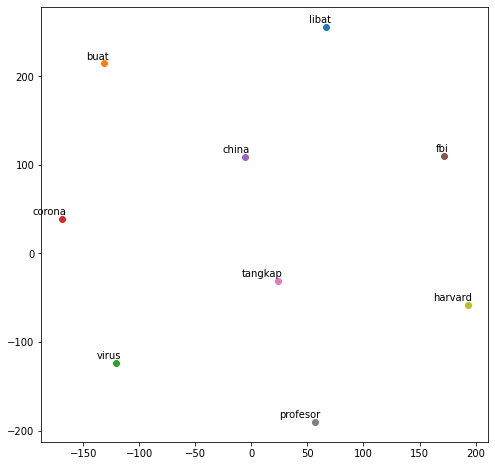


**Gambar 3.14** *Flowchart Word Embedding*

**Tabel 3.2** Hasil *Word Embedding*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Kata** | **Dimensi** | | | | |
| **1** | **2** | **3** | **4** | **5** |
| Libat | 0.07452912 | 0.071270086 | 0.011775045 | -0.0683471 | -0.00932195 |
| Buat | -0.0835347 | 0.0141729405 | -0.028569078 | -0.017134266 | -0.017073445 |
| Virus | -0.045051266 | -0.06300815 | -0.039852243 | 0.06623562 | -0.07374446 |
| Corona | -0.018552037 | -0.05871475 | -0.0063648676 | -0.06412901 | -0.085089624 |
| China | 0.041457042 | -0.015513414 | 0.038284097 | -0.046403337 | 0.074265204 |
| Fbi | -0.061272316 | -0.055307325 | 0.026977174 | 0.089745976 | 0.057193857 |
| tangkap | -0.08523997 | 0.058822744 | 0.007000594 | 0.067885585 | -0.095681004 |
| profesor | 0.08066033 | -0.045340743 | -0.089614525 | -0.0053708707 | -0.028951226 |
| harvard | -0.064775534 | -0.0042433734 | 0.05336014 | 0.084353596 | 0.006926856 |

Berdasarkan vektor-vektor tersebut dapat dibuat sebuah visualisasi yang merepresentasikan kedekatan antar kata ada. Untuk melakukan visualisasi dari data hasil word embedding dapat menggunakan model t-SNE. Model tersebut akan mengkonversi kesamaan antara titik data. Berikut hasil visualisasi dari vektor *word embedding* dapat dilihat pada Gambar 3.15.



**Gambar 3.15** Visualisasi *Word Embedding*

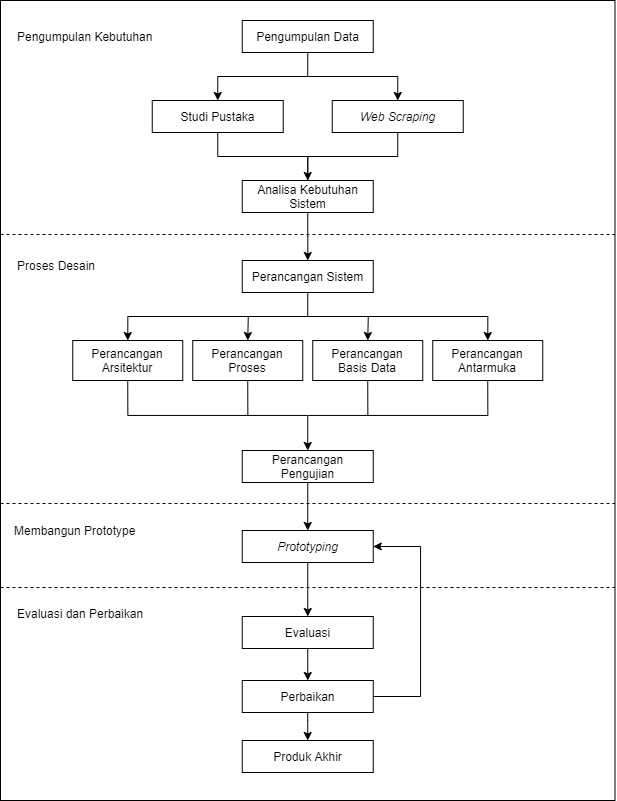
### 3.1.5 *Convolutional Neural Network*

### 3.1.6 Validasi & Pengujian

## 3.2 Metodologi Pengembangan Sistem

## 3.1 Metodologi Penelitian

Pada bagian ini akan membahas tentang metodologi penelitian dan metodologi pengembangan sistem yang digunakan sebagai bahan acuan dalam proses pembuatan penelitian ini. Metodologi penelitian yang digunakan bermodel kualitatif. Penggunaan model kualitatif karena penelitian ini dilakukan berdasarkan data dan memanfaatkan teori yang ada sebagai penguat argumentasi untuk menghasilkan suatu teori. Adapun tahapan metodologi penelitian yang dilakukan dapat dilihat pada Gambar 3.1.



**Gambar 3.1** Metodologi Penelitian

Sedangkan, untuk metode pengembangan sistem yang digunakan pada penelitian ini adalah metode *prototype.* Penggunaan metode *prototype* ini karena pada metode ini lebih menekankan pada komunikasi antara pengembang dan pelanggan. Selanjutnya kebutuhan dari pengguna diterjemahkan dalam bentuk model (*prototype*). Ada beberapa kelebihan dari penggunakan metode *prototype* yaitu terjalin hubungan yang baik antara pengguna dan pengembang, menghemat waktu dan biaya dalam pengembangan sistem, serta pelanggan berperan aktif dalam proses pengembangan sistem sehingga hasil yang didapatkan akan lebih baik karena sesuai dengan keinginan pelanggan. Metode *prototype* terdiri dari empat tahapan yaitu pengumpulan kebutuhan, proses desain yang cepat, membangun prototipe serta evaluasi dan perbaikan (Purnomo, 2017).

Pada tahapan pengumpulan dilakukan pengumpulan data yang bersumber dari studi pustaka dan *web scraping,* dilanjutkan dengan analisa dari data tersebut serta analisa kebutuhan sistem. Kemudian, pada bagian proses desain akan membuat rancangan sistem seperti rancangan arsitektur, proses basis data dan rancangan antarmuka. Selanjutnya membangun *prototype* berdasarkan rancangan yang telah dibuat dengan penulisan program (*coding*). *Prototype* yang dihasilkan akan diuji untuk mengetahui tingkat performa dari algoritma *Convolutional Neural Network* dalam menangai permasalahan klasifikasi berita *hoax,* lalu akan dievaluasi dan dilakukan perbaikan sesuai kebutuhan.

## 3.2 Pengumpulan Kebutuhan

Pada bagian akan dibahas mengenai hal-hal yang dibutuhkan baik berupa data maupun informasi lain untuk dianalisis dan digunakan dalam penelitian ini. Pengumpulan kebutuhan terdiri dari beberapa tahapan, yaitu pengumpulan data, analisa data dan analisa kebutuhan sistem.

### 3.2.1 Pengumpulan Data

Data merupakan salah satu hal yang paling penting dalam sebuah penelitian. Tanpa adanya data, suatu penelitian tidak dapat dilakukan. Penelitian memerlukan data untuk diolah sehingga menjadi sebuah informasi yang dapat membantu untuk pengambilan keputusan atau menghasilkan teori. Maka dari itu, data yang diperlukan dalam penelitian harus dikumpulkan terlebih dahulu. Untuk melakukan proses pengumpulan data, ada beberapa metode yang dapat digunakan. Metode pengumpulan data yang dilakukan pada penelitian ini terdiri dari dua bagian, yaitu studi pustaka dan *web scraping*.

#### 3.2.1.1 Studi Pustaka

Studi pustaka merupakan tahapan yang dilakukan untuk mendapatkan informasi yang berkaitan dengan masalah penelitian tentang berita *hoax,* metode dan teknik yang digunakan pada penelitian sebelumnya serta mengetahui kelebihan ataupun kekurangannya sehingga menjadi bahan evaluasi untuk dikembangkan pada penelitian ini. Selain itu, studi pustaka berfungsi sebagai bahan pembanding dan menjadi bahan referensi untuk memperkuat argumentasi pada penelitian ini. Studi pustaka dilakukan dengan mencari informasi yang bersumber dari buku, jurnal dan sumber lain yang dapat dipertanggung jawabkan.

#### 3.2.1.2 *Web Scraping*

*Web scraping* merupakan metode pengambilan data yang digunakan pada penelitian ini. Data yang digunakan adalah data sekunder yang berasal dari situs turnbackhoax.id. Pada situs tersebut telah memvalidasi kebenaran dari berita-berita yang tersebar di media sosial. Dengan menggunakan metode *web scraping* bertujuan untuk memudahkan pengambilan data dari situs tersebut karena dilakukan secara otomatis tanpa harus menyalin satu per satu. *Web scraping* memiliki beberapa metode yang dapat digunakan, salah satunya menggunakan *extension browser* yang dilakukan pada penelitian ini. *Extension browser* yang digunakan bernama *Data Miner* dari *browser* Google.

Data yang diambil mulai dari berita tahun 2015 sampai dengan 2019 yang dikategorikan menjadi dua yaitu *hoax* dan fakta. Data yang didapatkan terdiri dari 1899 berita yang berkategori *hoax* dan 2450 berita yang berkategori fakta. Jadi apabila diakumulasikan maka total dataset yang digunakan pada penelitan ini berjumlah 4349 data. Rincian dari jumlah dataset pada penelitian ini dapat dilihat pada Tabel 3.

**Tabel 3.** Rincian Dataset

|  |  |  |
| --- | --- | --- |
| **No** | **Label Klasifikasi** | **Jumlah Data** |
| 1 | Hoax | 1899 |
| 2 | Fakta | 2450 |
| **Total Data** | | **4349** |

### 3.2.2 Analisa Kebutuhan Sistem

Pada bagian ini akan dilakukan analisis terhadap kebutuhan-kebutuhan yang perlu disiapkan untuk mengembangkan sistem pada penelitian ini. Analisis ini bertujuan untuk mendefinisikan masalah dan menyiapkan sistem sehingga kendala yang berkaitan dengan sumber daya dapat diminimalisir. Analisis kebutuhan sistem ini terbagi menjadi dua bagian yaitu kebutuhan fungsional dan kebutuhan non-fungsional.

#### 3.2.3.1 Kebutuhan Fungsional

Kebutuhan fungsional adalah kebutuhan-kebutuhan yang berkaitan langsung dengan sistem. Pada kebutuhan fungsional akan memuat gambaran besar dari keseluruhan proses yang dapat dilakukan oleh sistem. Berikut adalah kebutuhan fungsional yang akan dilakukan oleh sistem:

1. Sistem dapat melakukan proses *preprocessing* terhadap dataset.
2. Sistem dapat melakukan proses *word embedding*.
3. Sistem dapat melakukan proses pembuatan model klasifikasi dengan metode *Convolutional Neural Network*.
4. Sistem dapat melakukan proses *training* dan *testing* menggunakan model klasifikasi.
5. Sistem dapat melakukan proses klasifikasi berita.
6. Sistem dapat menampilkan hasil klasifikasi berita *hoax* atau bukan.
7. Sistem dapat melakukan proses *login* dan *logout* untuk *admin*.
8. Sistem dapat memperbaharui model klasifikasi dengan *training* data baru yang dilakukan melalui *admin*.
9. Sistem dapat melakukan pengecekan dan pelaporan berita *hoax* melalui *user*.

#### 3.2.3.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional adalah kebutuhan yang tidak berkaitan langsung dengan fitur yang terdapat pada sistem. Pada kebutuhan non-fungsional terdiri dari kebutuhan perangkat keras (*hardware*) dan perangkat lunak (*software*).

**a. Perangkat Keras (*Hardware*)**

Perangkat keras yang dibutuhkan untuk mengembangkan sistem pada penelitian ini memiliki spesifikasi seperti pada Tabel 3.

**Tabel 3.** Kebutuhan Perangkat Keras

|  |  |  |
| --- | --- | --- |
| **No** | **Perangkat Keras** | **Keterangan** |
| 1 | *Processor* | Intel Core i3-5005U 3th Gen |
| 2 | *RAM* | 4GB DDR4 |
| 3 | *Storage* | Harddisk 500GB |
| 4 | *Graphic* | Intel HD Graphics |
| 5 | Perangkat *Input* dan *Output* | *Keyboard*, *Mouse* dan *Monitor* |
| 6 | Koneksi Internet | Wifi atau Kuota Internet |

**b. Perangkat Lunak (*Software*)**

Adapun perangkat lunak yang dibutuhkan untuk mengembangkan sistem pada penelitian ini dapat dilihat pada Tabel 3.

**Tabel 3.** Kebutuhan Perangkat Lunak

|  |  |  |
| --- | --- | --- |
| **No** | **Perangkat Lunak** | **Keterangan** |
| 1 | Operating System | Windows 10 dengan 64 bit |
| 2 | Bahasa Python | Bahasa pemrograman |
| 3 | Bahasa PHP | Bahasa pemrograman |
| 4 | Sublime Text | Perangkat lunak untuk c*ode editor* |
| 5 | Google Colab | *Code editor* berbasis online |
| 6 | XAMPP | Perangkat lunak sebagai server lokal |
| 7 | MySQL | Tempat penyimpanan database |
| 8 | Google Chrome | *Web browser* |
| 9 | *Draw.io* | Website untuk membuat desain diagram |
| 10 | *CorelDraw* | Perangkat lunak untuk rancangan *interface* |

## 3.3 Proses Desain

Proses desain merupakan tahapan yang membahas mengenai proses pembuatan rancangan sementara untuk disajikan kepada pelanggan berdasarkan hasil komunikasi antara pengembang dan pelanggan. Pembuatan rancangan ini bertujuan untuk memberikan gambaran dan memudahkan proses pembuatan sistem yang akan dibangun. Pada proses desain ini akan dibuat perancangan sistem dan perancangan pengujian.

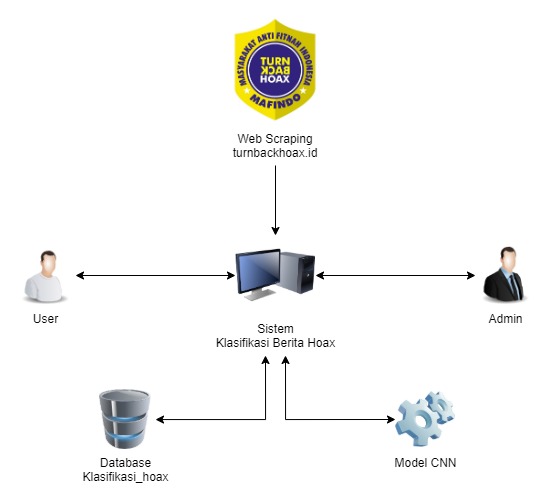
### 3.3.1 Perancangan Sistem

Pada bagian ini akan dibahas tentang rancangan sistem yang akan dibangun. Beberapa perancangan yang akan dibuat untuk membangun sistem klasifikasi berita *hoax* antara lain perancangan arsitektur, perancangan proses, perancangan basis data dan perancangan antarmuka.

#### 3.3.1.1 Perancangan Arsitektur

Arsitektur sistem yang akan dibuat terdiri dari pengguna, database, proses *scraping,* model CNN dan sistem itu sendiri yang akan melakukan klasifikasi berita *hoax*. Terdapat dua aktor pengguna pada sistem ini yaitu *user* dan *admin*. *User* merupakan semua orang secara umum yang menggunakan sistem ini. *User* dapat memberikan *input* kepada sistem berupa konten berita yang ingin diperiksa kebenarannya dan membuat laporan apabila menemukan berita *hoax* di media sosial. Sedangkan *admin* merupakan orang tertentu yang memiliki hak akses khusus dan memiliki hak yang lebih pula di sistem. *Admin* dapat mengolah berbagai data seperti data admin, data cek hoax*,* data lapor hoax, data dataset, data training, data testing dan data pengujian. Selain itu, *admin* juga dapat melakukan pengecekan berita *hoax* dan membuat laporan *hoax*.

Komponen lain yaitu sistem klasifikasi berita hoax. Semua proses akan terjadi di sistem. Sistem mendapatkan data berasal dari proses *scraping* dari web turnbackhoax.id yang setelah itu diolah menggunakan *text preprocessing* sehingga akan menghasil suatu dataset yang siap digunakan. Sistem akan melakukan proses klasifikasi berdasarkan model CNN yang telah dibuat sebelumnya. Model tersebut telah melalui proses training dan testing menggunakan algoritma *Convolutional Neural Network* (CNN). Setelah itu, sistem akan memberikan hasil klasifikasinya kepada *user* atau *admin.* Sistem juga akan memberikan *output* berupa data-data yang berasal dari *database* ataupun menyimpan data ke *database* berdasarkan aksi yang dilakukan oleh *admin*. Ilustrasi dari arsitektur sistem tersebut dapat dilihat pada Gambar 3.2.

****

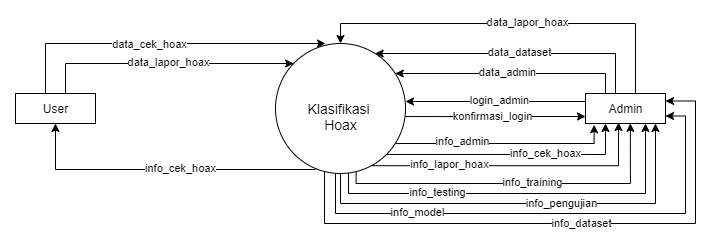
**Gambar 3.2** Arsitektur Sistem

#### 3.3.1.2 Perancangan Proses

Pada tahap perancangan proses ini akan menggambarkan alur dari proses yang ada pada sistem. Perancangan proses ini digambarkan *Data Flow Diagram* (DFD) dari level 0 hingga level 2 serta *flowchart* pada beberapa proses.

* 1. ***Data Flow Diagram Level 0***

*Data Flow Diagram* (DFD) *Level 0* merupakan gambaran besar dari proses beserta alur pergerakan data yang ada pada sistem ini. Pada penelitian ini terdapat dua pengguna yang berinteraksi dengan sistem yaitu *user* dan *admin*. Dimana masing-masing pengguna tersebut memiliki batasan yang berbeda terkait *input* dan *output* pada sistem. Pada bagian *user* dapat memberikan *input* kepada sistem berupa data cek hoax dan data lapor hoax. Selain itu, *user* juga dapat menerima *output* dari sistem berupa informasi cek hoax. Sedangkan dari sisi *admin* mempunyai hak yang lebih besar terhadap sistem. Karena *admin* dapat memberikan *input* kepada sistem berupa login admin, data admin, data dataset dan data lapor hoax. *Admin* juga menerima *output* dari sistem berupa konfirmasi login, informasi admin, informasi cek hoax, informasi lapor hoax, informasi training, informasi testing, informasi pengujian, informasi model dan informasi dataset. Gambaran dari *data flow diagram level 0* dapa dilihat pada Gambar 3.3.

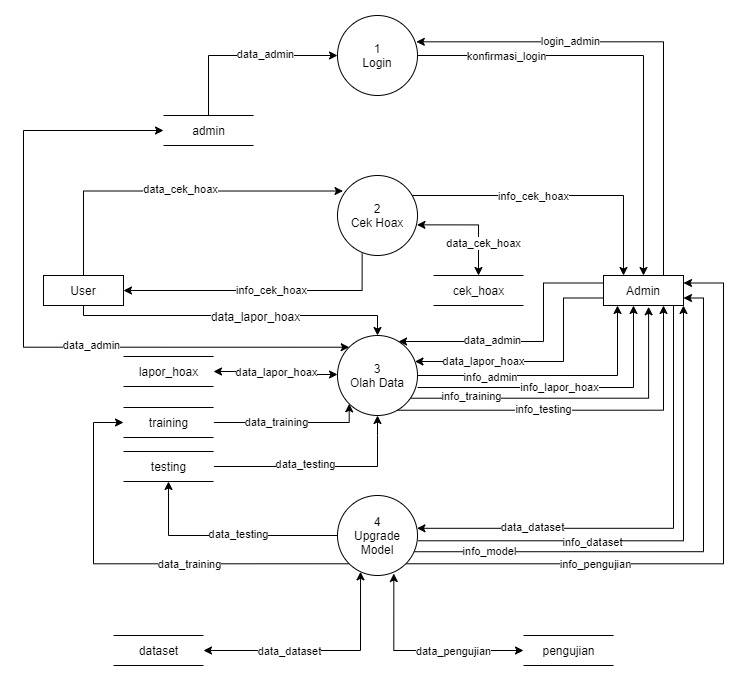
****

**Gambar 3.3** *Data Flow Diagram Level 0*

* 1. ***Data Flow Diagram Level 1***

*Data Flow Diagram* (DFD) *Level 1* merupakan turunan dari DFD level 0 yang akan menjabarkan semua proses yang terdapat pada sistem. DFD level 1 terdiri dari empat proses yaitu proses *login*, proses cek hoax, proses olah data dan proses *upgrade* model. Setiap proses akan menggambarkan aliran data serta interaksi dengan pengguna (*user* dan *admin*) dan atau basis data.

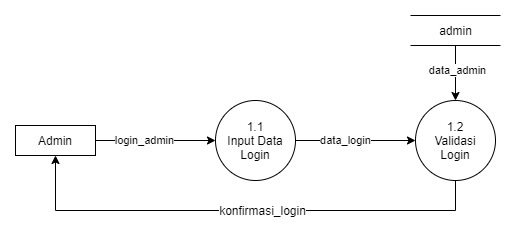
Pada proses *login*, sistem akan menerima *input* dari *admin* berupa login admin dan data admin dari tabel admin. Lalu, proses *login* akan mengeluarkan *output* berupa konfirmasi login kepada admin. Proses kedua pada sistem ini yaitu proses cek hoax. Proses cek hoax dapat menerima *input* dari *user* berupa data cek hoax dan menyimpan data tersebut pada tabel cek hoax. Selain itu, proses cek hoax akan mengeluarkan *output* berupa informasi cek hoax kepada *admin* dan *user*. Selanjutnya proses olah data yang dapat menerima *input* data admin dari *admin* dan data lapor hoax dari *admin* dan *user* lalu disimpan pada tabel lapor hoa. Proses olah data juga akan mengambil data dari tabel training dan tabel testing untuk memberikan *output* berupa informasi training dan testing, serta informasi admin dan lapor hoax kepada *admin*. Terakhir, proses *upgrade* model yang akan menerima *input* data dataset dari *admin* lalu membagi dataset tersebut menjadi data training dan testing untuk diproses dan disimpan di tabel training dan testing. Proses *upgrade* model akan melakukan pengujian yang hasilnya akan disimpan pada tabel pengujian. Proses ini akan mengeluarkan *output* berupa informasi dataset, informasi model dan informasi pengujian. Berikut *data flow diagram level 1* yang dapat dilihat pada Gambar 3.4

****

**Gambar 3.4** *Data Flow Diagram Level 1*

* 1. ***Data Flow Diagram Level 2 Proses* *Login***

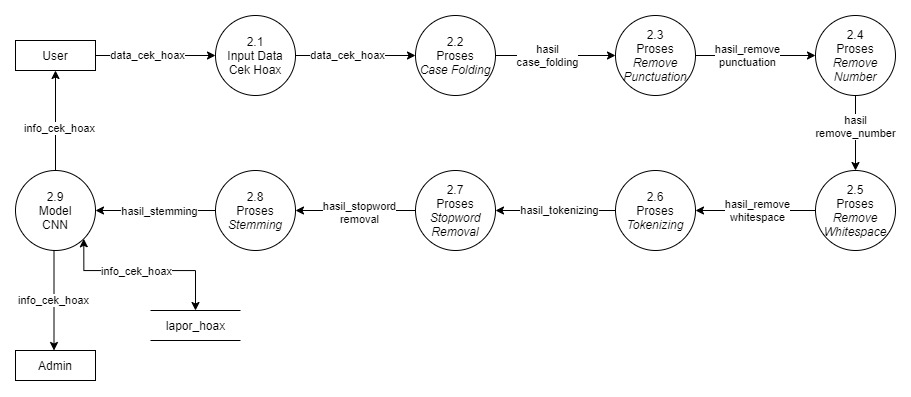
*Data Flow Diagram* (DFD) *Level 2 Proses Login* merupakan penjelasan lebih rinci dari proses login di DFD level 1. Pada DFD level 2 proses login, *admin* akan mengirimkan login admin pada sub proses input data login dan sub proses tersebut akan menghasilkan data login yang akan dikirim ke subproses validasi login. Pada sub proses validasi login akan mengambil data dari tabel admin untuk memvalidasi data login. Hasil validasi tersebut akan dikeluarkan oleh subproses validasi login berupa konfirmasi login kepada *admin*.

****

**Gambar 3.5** *Data Flow Diagram Level 2 Proses* *Login*

* 1. ***Data Flow Diagram Level 2 Proses Cek Hoax***

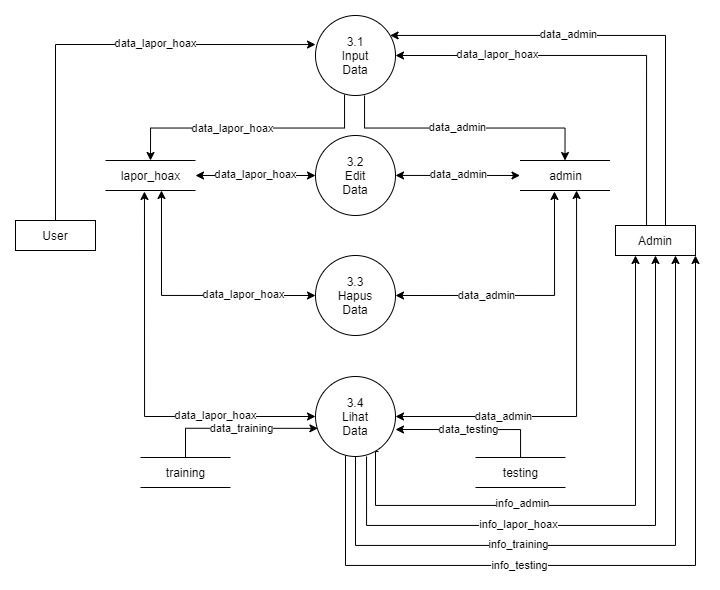
*Data Flow Diagram* (DFD) *Level 2 Proses Cek Hoax* merupakan penjelasan lebih rinci dari proses cek hoax pada DFD level 1. Pada DFD level 2 proses cek hoax akan menerima *input* berupa data cek hoax dan memiliki beberapa sub proses diantaranya input data hoax, *case folding*, *remove punctuation*, *remove number*, *remove whitespace*, *tokenizing*, *stopword*, *stemming* dan model CNN. Pada sub proses model CNN akan menghasilkan informasi cek hoax yang akan dikeluarkan ke *user* dan *admin* serta informasi tersebut di tabel cek hoax. DFD Level 2 Proses Cek Hoax dapat dilihat pada Gambar 3.6.

****

**Gambar 3.6** *Data Flow Diagram Level 2 Proses Cek Hoax*

* 1. ***Data Flow Diagram Level 2 Proses Olah Data***

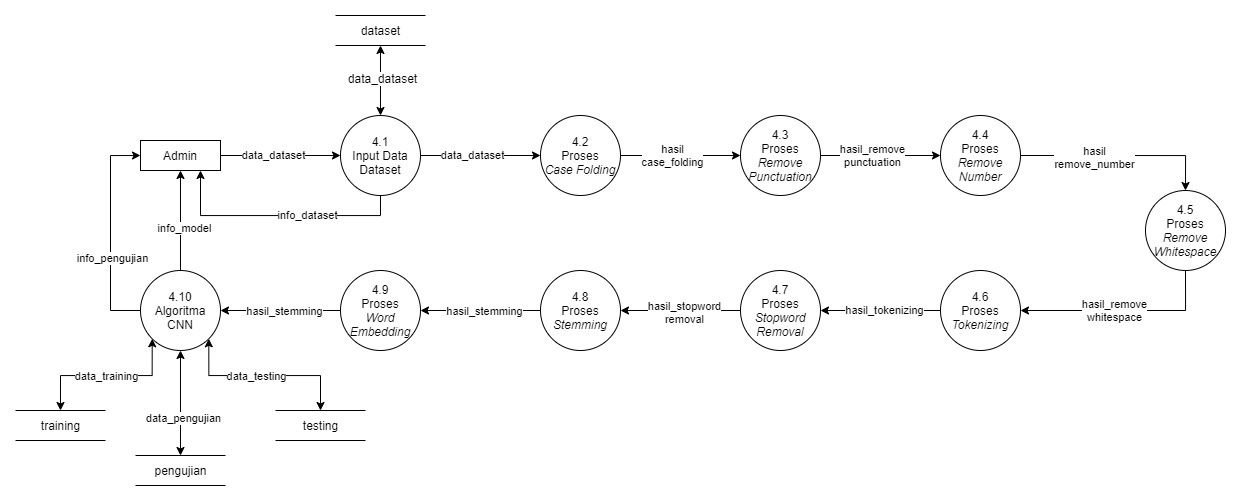
*Data Flow Diagram* (DFD) *Level 2 Proses Olah Data* adalah penjabaran yang lebih rinci dari proses olah data di DFD level 1. Pada DFD level 2 proses olah data terdapat empat sub proses yang memiliki fungsi masing-masing. Pertama, sub proses input data yang akan menerima *input* dari *admin* berupa data admin dan data lapor hoax serta data lapor hoax dari *user*. Kemudian data admin dan data lapor hoax tersebut akan disimpan di tabel admin dan tabel lapor hoax. Kedua, sub proses edit data. Sub proses ini akan mengambil data dari tabel admin dan tabel lapor hoax untuk diubah datanya dan akan disimpan kembali pada tabel masing-masing. Ketiga, sub proses hapus data yang akan mengambil data dan menghapusnya dari tabel admin dan atau tabel lapor hoax. Terakhir, sub proses lihat data. Sub proses ini mengambil data dari tabel admin, tabel lapor hoax, tabel training dan testing yang selanjutnya mengeluarkan *output* berupa informasi admin, informasi lapor hoax, informasi training dan informasi testing. DFD level 2 proses olah data ini dapat dilihat pada Gambar 3.7

****

**Gambar 3.7** *Data Flow Diagram Level 2 Proses Olah Data*

* 1. ***Data Flow Diagram Level 2 Proses Upgrade Model***

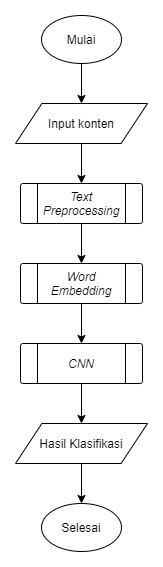
*Data Flow Diagram* (DFD) *Level 2 Proses Upgrade Model* merupakan penjabaran dari proses *upgrade* model pada DFD level 1. Pada bagian ini akan menerima *input* dari *admin* berupa data dataset dan memberikan *output* informasi dataset kepada admin. Data dataset akan diproses pada beberapa sub proses yang terdiri dari *case folding*, *remove punctuation*, *remove number*, *remove whitespace*, *tokenizing*, *stopword*, *stemming*, *word embedding*, dan algoritma CNN. Hasil dari sub proses tersebut berupa data training, data testing dan data pengujian yang akan disimpan di masing-masing tabel serta informasi pengujian dan informasi model yang akan diberikan kepada *admin*. DFD level 2 proses *upgrade* model dapat dilihat pada Gambar 3.8

****

**Gambar 3.8** *Data Flow Diagram Level 2 Proses Upgrade Model*

* 1. ***Flowchart* Klasifikasi Berita *Hoax***

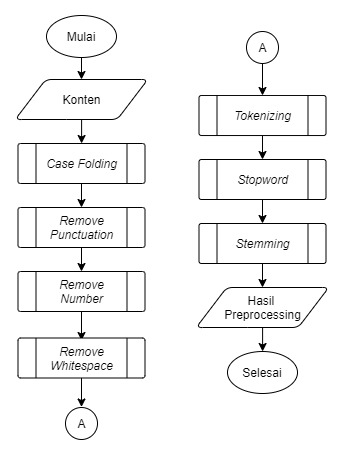
*Flowchart* ini akan menggambarkan secara garis besar alur dari proses klasifikasi berita *hoax.* Proses ini akan dimulai dari menerima *input* konten berita yang ingin diklasifikasi. Kemudian, konten tersebut akan dilakukan proses *text preprocessing* yang terdiri dari beberapa tahapan. Hasil dari *text preprocessing* akan dilakukan proses pembobotan fitur menggunakan metode *word embedding* dengan *word2vec*. Selanjutnya, hasil dari proses *word embedding* tersebut akan diklasifikasi menggunakan Algoritma *Convolutional Neural Network* (CNN) ke dalam dua kelas yaitu *hoax* dan fakta. Hasil dari klasifikasi tersebut yang akan menjadi *output* dari sistem ini. *Flowchart* klasifikasi berita *hoax* dapat dilihat pada Gambar 3.9

****

**Gambar 3.9** *Flowchart* Klasifikasi Berita *Hoax*

* 1. ***Flowchart Text Preprocessing***

*Flowchart* ini akan menggambarkan alur proses pada bagain *text preprocessing*. Pada bagian ini bertujuan untuk membersihkan data konten yang menjadi *input* dari sistem. Hal tersebut membuat data yang awalnya semi-terstruktur atau bahkan tidak terstruktur menjadi data yang terstruktur dan siap untuk digunakan. *Text preprocessing* memiliki beberapa tahapan yang menjadi sub proses diantaranya *case folding, remove punctuation, remove number, remove whitespace, tokenizing, stopword* dan *stemming*. *Flowchart* dari *text preprocessing* dapat dilihat pada Gambar 3.10

****

**Gambar 3.10** *Flowchart Text Preprocessing*

Berikut adalah contoh dari hasil *text preprocessing* terhadap sebuah berita berbahasa Indonesia setelah melalui semua proses *preprocessing*.

Text awal:

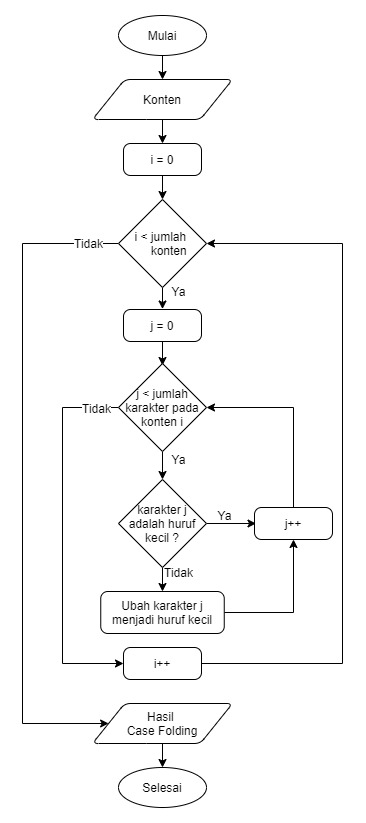
“Terlibat pembuatan virus corona di China, FBI menangkap 10 profesor Harvard.”

Hasil *Preprocessing*:

“['libat', 'buat', 'virus', 'corona', 'china', 'fbi', 'tangkap', 'profesor', 'harvard']”

* 1. ***Flowchart Case Folding***

*Flowchart* ini akan menggambarkan alur dari proses *case folding* yaitu proses merubah semua karakter pada dokumen menjadi huruf kecil (*lowercase*). Flowchart ini akan menerima *input* berupa konten dan inisialisasi i yang mewaliki jumlah dan j mewakili jumlah karakter. Prosesnya yaitu melakukan pengecekan ke setiap karakter dari awal hingga akhir. Jika karakter merupakan huruf kecil, maka karakter tersebut disimpan. Tetapi apabila menemukan karakter yang huruf besar maka dirubah terlebih dahulu karakter tersebut menjadi huruf kecil, lalu perubahan tersebut di simpan. Ketika semua karakter telah diperiksa, maka *flowchart* ini akan mengeluarkan *output* berupa hasil *case folding* yang akan digunakan pada proses berikutnya. *Flowchart* dari proses *case folding* dapat dilihat pada Gambar 3.11

****

**Gambar 3.11** *Flowchart Case Folding*

Berikut contoh hasil dari proses *case folding* terhadap suatu konten berita yang ingin diperiksa.

Text awal:

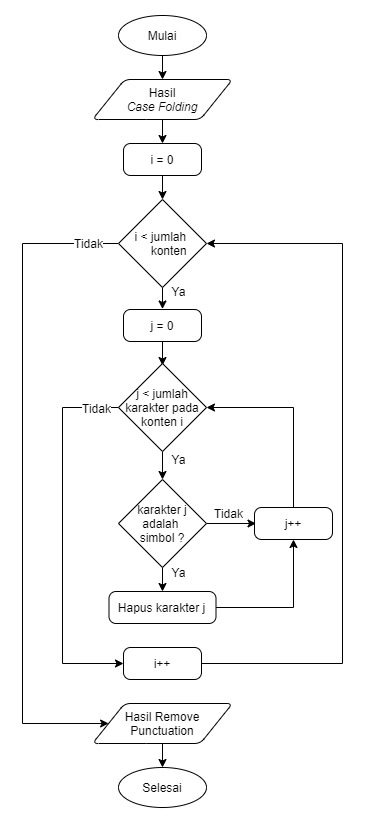
“Terlibat pembuatan virus corona di China, FBI menangkap 10 profesor Harvard.”

Hasil *case folding*:

“terlibat pembuatan virus corona di china, fbi menangkap 10 profesor harvard.”

* 1. ***Flowchart Remove Punctuation***

*Flowchart remove punctuation* yaitu alur dari proses menghapus tanda baca pada suatu dokumen karena dianggap sebagai *delimiter*. Penghapusan tanda baca ini akan mengurangi beban pemprosesan. Beberapa contoh tanda baca seperti titik (.), koma (,), tanda tanya (?), tanda seru (!) dan lainnya. Proses yang terjadi yaitu melakukan pengecekan jumlah dokumen dan jumlah karakternya. Data yang digunakan pada proses ini merupakan data hasil dari proses *case folding.* Kemudian, setiap karakter dari data tersebut akan dicek apakah karakter merupakan simbol atau tanda baca. Jika karakter merupakan tanda baca maka akan dihapus, tetapi jika bukan tanda baca maka karakter tersebut akan disimpan. Apabila semua karakter telah diperiksa, maka *flowchart* ini akan memberikan *output* berupa hasil *remove punctuation*. Alur dari proses *remove punctuation* dapat dilihat lebih detail pada Gambar 3.12.

****

**Gambar 3.12** *Flowchart Remove Punctuation*

Berikut contoh hasil proses *remove punctuation* terhadap data yang merupakan hasil dari *case folding*.

Text *input* (hasil *case folding*):

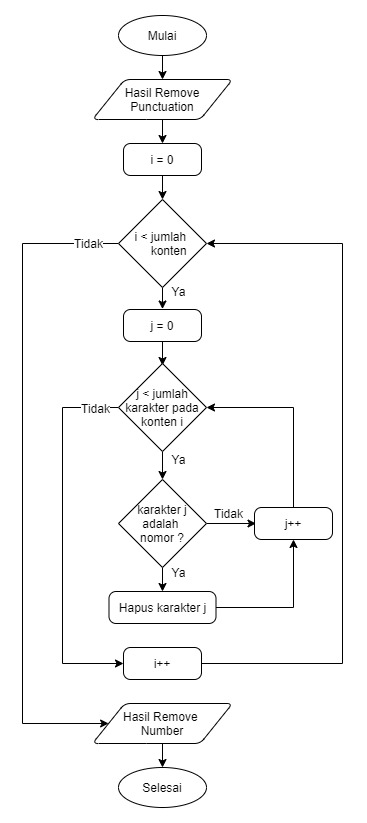
“terlibat pembuatan virus corona di china, fbi menangkap 10 profesor harvard.”

Text hasil *remove punctuation*:

“terlibat pembuatan virus corona di china fbi menangkap 10 profesor harvard”

* 1. ***Flowchart Remove Number***

*Flowchart remove number* tidak jauh berbeda dengan *remove punctuation*, perbedaannya hanya pada karakter yang dihapus. Pada *remove number* ini akan menghapus karakter yang berbentuk angka. Pada proses ini akan menerima *input* berupa hasil dari proses *remove puctuation,* kemudian akan diperiksa per karakternya. Apabila menemukan karakter berbentuk angka, maka karakter tersebut akan dihapus. *Flowchart* dari *remove number* dapat dilihat pada Gambar 3.13

****

**Gambar 3.13** *Flowchart Remove Number*

Berikut contoh hasil proses *remove number* terhadap data yang merupakan hasil dari *remove punctuation*.

Text *input* (hasil *remove punctuation*):

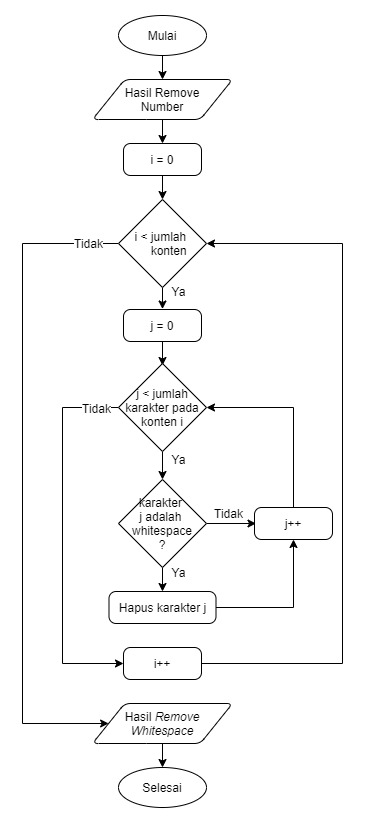
“terlibat pembuatan virus corona di china fbi menangkap 10 profesor harvard”

Text hasil *remove number*:

“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

* 1. ***Flowchart Remove Whitespace***

*Flowchart remove whitespace* akan menggambarkan alur proses dari penghapusan *whitespace* dari data. Proses ini akan mengecek setiap karakter apakah termasuk *whitespace* atau bukan. Jika karakter merupakan *whitespace* maka karakter tersebut akan dihapus, sebaliknya jika tidak maka karakter akan disimpan. Data yang akan diproses merupakan hasil dari *remove number* dan hasil dari proses ini menghasilkan hasil *remove whitespace*. *Flowchart* dari *remove whitespace* dapat dilihat pada Gambar 3.

****

**Gambar 3.** *Flowchart Remove Whitespace*

Berikut contoh hasil proses *remove whitespace* terhadap data yang merupakan hasil dari *remove number*.

Text *input* (hasil *remove number*):

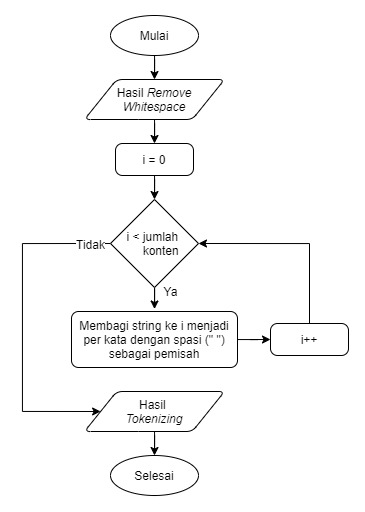
“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

Text hasil *remove whitespace*:

“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

* 1. ***Flowchart Tokenizing***

*Flowchart* ini akan menggambarkan proses *tokenizing*. Proses ini kan memotong data menjadi per kata dengan spasi (“ “) sebagai pemisahnya. Data yang akan diproses dari hasil *remove whitespace* dan akan mengeluarkan hasil *tokenizing*. *Flowchart* dari hasil *tokenizing* dapat dilihat pada Gambar 3.

****

**Gambar 3.** *Flowchart Tokenizing*

Berikut contoh hasil proses *tokenizing* terhadap data yang merupakan hasil dari *remove whitespace*.

Text *input* (hasil *remove whitespace*):

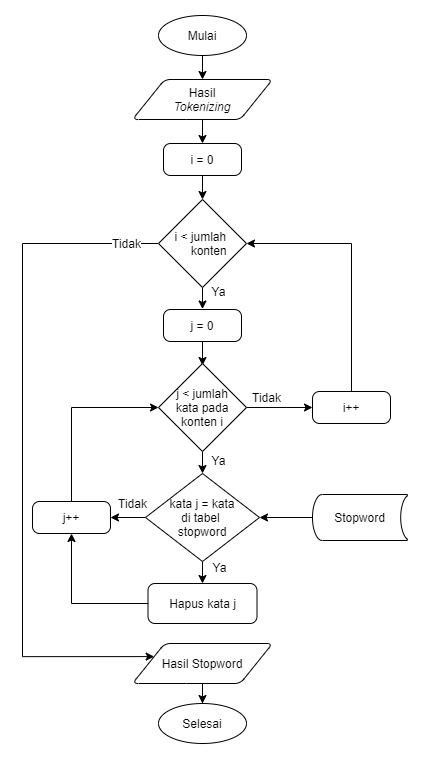
“terlibat pembuatan virus corona di china fbi menangkap profesor harvard”

Text hasil *tokenizing*:

“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

* 1. ***Flowchart Stopword***

*Flowchart stopword* akan menggambarkan proses pembuangan kata-kata yang kurang penting (*stopword*) pada data. Contoh kata yang kurang penting seperti kata ganti orang, kata penghubung dan sebagainya. Proses pada tahap ini yaitu melakukan pengecekan per kata dari hasil *tokenizing*. Lalu, setiap kata akan dicek dicek apakah kata tersebut ada di kamus *stopword*. Apabila kata di data ada di kamus *stopword,* maka kata tersebut dihapus. Sebaliknya, jika tidak ada di kamus *stopword* maka akan disimpan. Proses *stopword* dapat dilihat pada Gambar 3.

****

**Gambar 3.** *Flowchart Stopword*

Berikut contoh hasil proses *stopword* terhadap data yang merupakan hasil dari *tokenizing*.

Text *input* (hasil *tokenizing*):

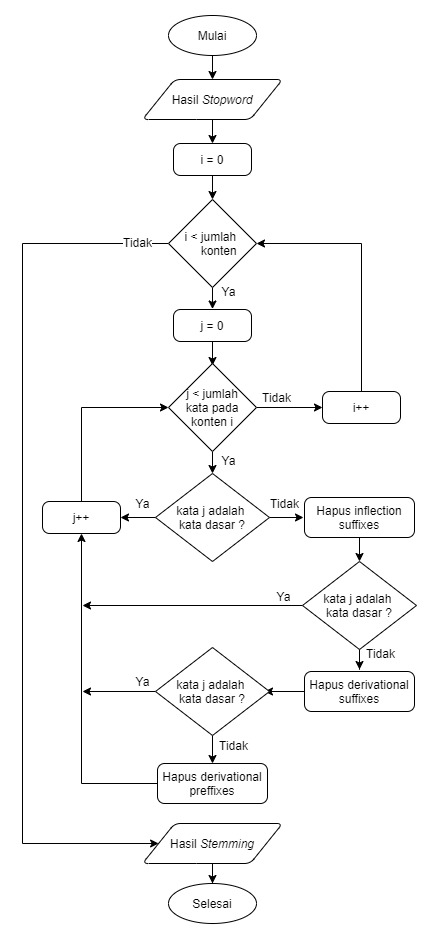
“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

Text hasil *stopword*:

“['terlibat', 'pembuatan', 'virus', 'corona', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

* 1. ***Flowchart Stemming***

Pada bagian ini akan menjelaskan terkait alur dari proses *stemming*. Proses ini akan menggunakan data hasil dari proses *stopword* yang akan diperiksa per kata. *Stemming* akan mencari kata dasar dari setiap kata dengan cara menghapus imbuhan pada kata. Jika kata merupakan kata dasar maka kata tersebut akan disimpan. Apabila kata bukan kata dasar, maka akan di hapus *inflection suffixes, derivational suffixes* dan atau *derivational suffixes* yang ada pada kata sehingga kata tersebut menjadi kata dasar. Setelah itu kata disimpan dan *flowchart* ini akan menghasilkan *output* berupa hasil *stemming* sekaligus menjadi hasil dari proses *text preprocessing*. Berikut *flowchart* dari *stemming* dapat dilihat pada Gambar 3.

****

**Gambar 3.** *Flowchart Stemming*

Berikut contoh hasil proses *stopword* terhadap data yang merupakan hasil dari *tokenizing*.

Text *input* (hasil *stopword*):

“['terlibat', 'pembuatan', 'virus', 'corona', 'di', 'china', 'fbi', 'menangkap', 'profesor', 'harvard']”

Text hasil *stemming* dan hasil *preprocessing*:

“['libat', 'buat', 'virus', 'corona', 'china', 'fbi', 'tangkap', 'profesor', 'harvard']”

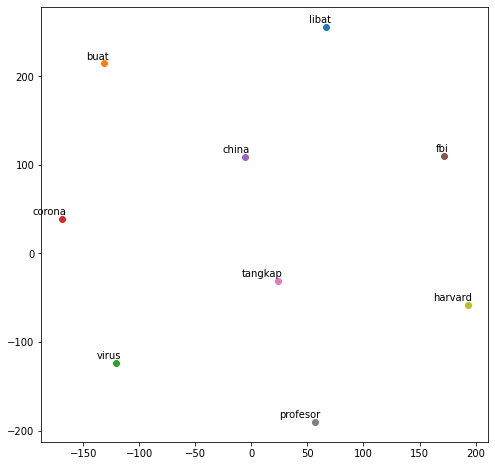
* 1. ***Flowchart Word Embedding***

Pada bagian ini akan dijelaskan terkait proses dari *word embedding*. Proses ini akan merubah suatu kata menjadi suatu vektor yang berisi kumpulan angka. Kata yang menjadi nilai *input* pada *word embedding* berasal dari hasil proses *stemming*. Vektor yang dihasilkan dari proses ini akan merepresentasikan dari setiap kata yang ada. Vektor yang dihasilkan memiliki panjang tertentu yang dapat diatur, semakin panjang suatu vektor maka akan semakin teliti representasi yang dihasilkan. Panjang vektor pada proses *word embedding* dapat disebut dengan *dimensi*. Pada penelitian ini, dimensi dari vektor yang dihasilkan yaitu 100. Namun, untuk memudahkan dalam memberikan contoh maka akan ditampilkan vektor dari setiap kata dengan jumlah dimensi 5. Berikut hasil dari proses *word embedding* dapat dilihat pada Tabel

**Tabel 3.** Hasil *Word Embedding*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Kata** | **Dimensi** | | | | |
| **1** | **2** | **3** | **4** | **5** |
| Libat | 0.07452912 | 0.071270086 | 0.011775045 | -0.0683471 | -0.00932195 |
| Buat | -0.0835347 | 0.0141729405 | -0.028569078 | -0.017134266 | -0.017073445 |
| Virus | -0.045051266 | -0.06300815 | -0.039852243 | 0.06623562 | -0.07374446 |
| Corona | -0.018552037 | -0.05871475 | -0.0063648676 | -0.06412901 | -0.085089624 |
| China | 0.041457042 | -0.015513414 | 0.038284097 | -0.046403337 | 0.074265204 |
| Fbi | -0.061272316 | -0.055307325 | 0.026977174 | 0.089745976 | 0.057193857 |
| tangkap | -0.08523997 | 0.058822744 | 0.007000594 | 0.067885585 | -0.095681004 |
| profesor | 0.08066033 | -0.045340743 | -0.089614525 | -0.0053708707 | -0.028951226 |
| harvard | -0.064775534 | -0.0042433734 | 0.05336014 | 0.084353596 | 0.006926856 |

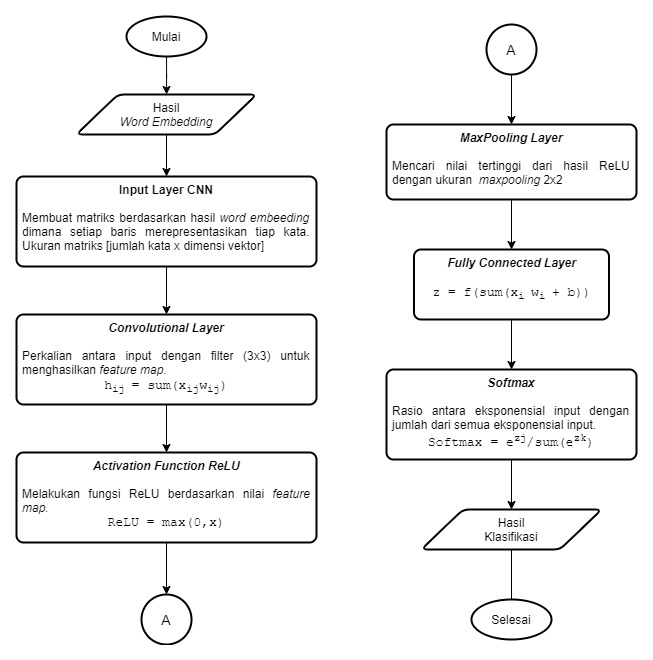
Berdasarkan vektor-vektor tersebut dapat dibuat sebuah visualisasi yang merepresentasikan kedekatan antar kata ada. Untuk melakukan visualisasi dari data hasil word embedding dapat menggunakan model t-SNE. Model tersebut akan mengkonversi kesamaan antara titik data. Berikut hasil visualisasi dari vektor *word embedding* dapat dilihat pada Gambar 3.



**Gambar 3.** Visualisasi *Word Embedding*

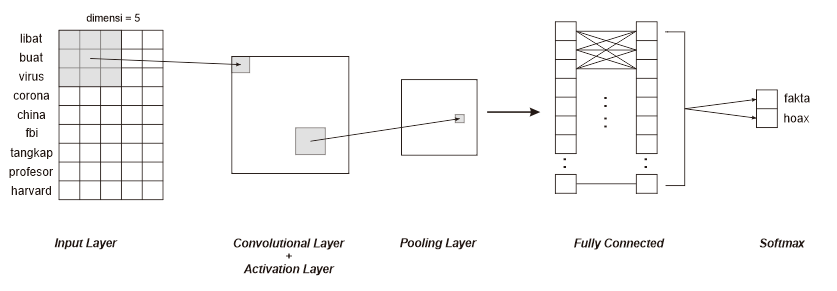
* 1. ***Flowchart CNN***

Pada bagian ini akan menggambarkan proses klasifikasi menggunakan algoritma *convolutional neural network*. Hasil dari proses *word embedding* akan menjadi data *input* yang akan dimasukkan pada matriks *input layer* dengan setiap barisnya merepresentasikan setiap kata. Setelah itu, *input layer* akan dikalikan dengan filter ukuran 3x3 untuk menghasilkan *feature map* yang merupakan proses dari *convolutional layer* dan masing-masing hasil tersebut dilakukan fungsi aktivasi ReLu. Kemudian, akan dilakukan operasi *pooling* dimana pada penelitian ini akan menggunakan *MaxPooling* dengan ukuran 2x2. Hasil dari operasi *pooling* tersebut akan masuk ke bagian *fully connected.* Pada bagian ini merupakan *neural network* yang terdiri dari *input layer, hidden layer* dan *output layer*. Selanjutnya akan dilakukan proses *softmax* untuk menghasilkan probabilitas dari hasil klasifikasi. Berikut *flowchart* CNN dapat dilihat pada Gambar 3.

****

**Gambar 3.** *Flowchart CNN*

Berdasarkan *flowchart* tersebut maka dapat dibuat sebuah arsitektur CNN yang akan digunakan pada penelitian ini untuk memperjelas alur dari proses dari algoritma *convolutional neural network*. Berikut arsitektur CNN dapat dilihat pada Gambar 3.



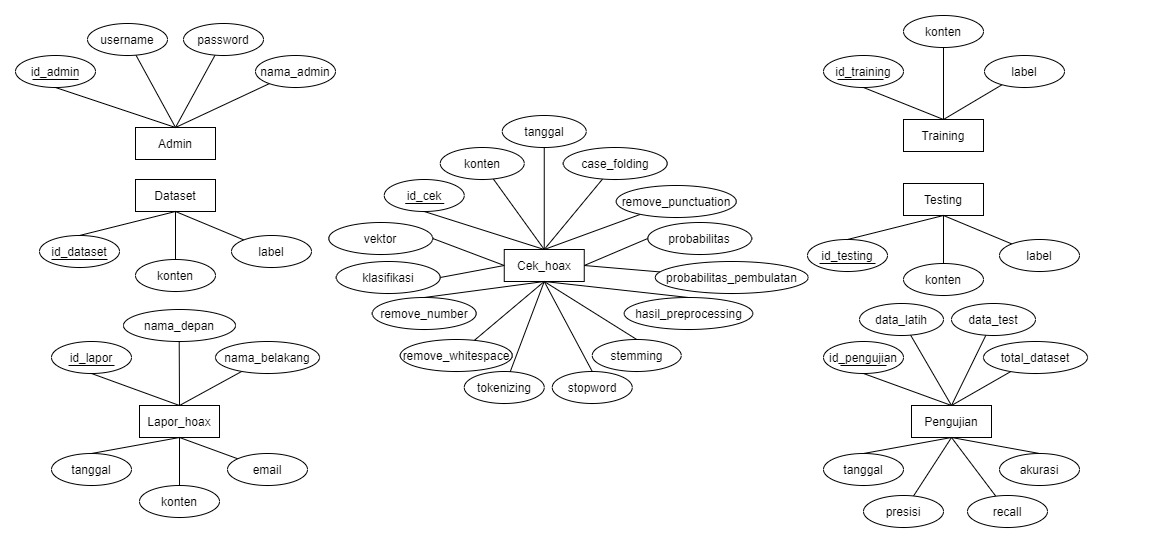
**Gambar 3.** Arsitektur CNN

#### 3.3.1.3 Perancangan Basis Data

Dalam memenuhi kebutuhan sistem, maka diperlukan suatu basis data yang berfungsi sebagai tempat penyimpanan seluruh data yang dibutuhkan. Data-data yang akan disimpan seperti data admin, data cek hoax, dataset, data training dan data testing, data laporan serta data pengujian. Basis data pada sistem ini menggunakan *database* *MySql* dan *Server HTTP* *Apache*. Tetapi sebelum membuat basis data, ada beberapa rancangan yang perlu dipersiapkan terlebih dahulu diantaranya perancangan *Entity Relationship Diagram* (ERD), Relasi Antar Tabel (RAT) dan Struktur Tabel.

* 1. ***Entity Relationship Diagram* (ERD)**

*Entity Relationship Diagram* (ERD) adalah suatu model yang bertujuan untuk menjelaskan hubungan antar data dalam basis data. Proses ini akan memodelkan struktur data dan hubungan antar datanya dengan menggunakan notasi dan simbol. Ada tiga komponen utama pada ERD yaitu entitas(objek), atribut(karakteristik objek), dan relasi (hubungan antar objek). Rancangan ERD pada Penerapan *Deep Learning* Untuk Klasifikasi Berita *Hoax* Menggunakan *Convolutional Neural Network* dapat dilihat pada Gambar 3.

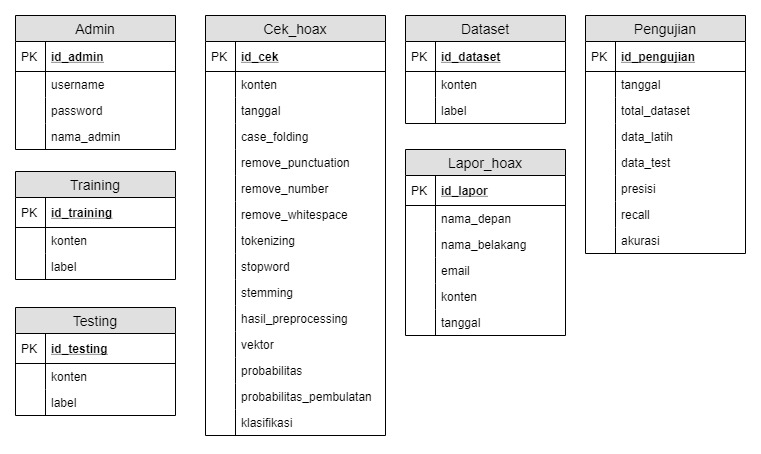


**Gambar 3.** ERD Klasifikasi Berita *Hoax*

ERD dalam sistem ini terdapat tujuh entitas yaitu admin, cek hoax, dataset, lapor hoax, training, testing dan pengujian yang masing-masing entitas akan dijabarkan sebagai berikut:

1. Entitas admin, entitas ini memiliki empat atribut diantaranya id admin, username, password dan nama admin. Id admin merupakan *primary key*.
2. Entitas cek hoax, entitas ini memiliki beberapa atribut yaitu id cek, konten, tanggal, case folding, remove punctuation, remove number, remove whitespace, tokenizing, stopword, stemming, hasil preprocessing, vektor, probabilitas, probabilitas pembulatan dan klasifikasi dimana id cek sebagai *primary key*.
3. Entitas dataset, entitas ini memiliki tiga atribut yaitu id dataset yang merupakan *primary key*, konten dan label.
4. Entitas lapor hoax, entitas ini memiliki beberapa atribut seperti id lapor, nama depan, nama belakang, email, konten dan tanggal. Id lapor merupakan *primary key*.
5. Entitas training, entitas ini terdapat tiga atribut diantaranya id training, konten dan label dimana id training berperan sebagai *primary key*.
6. Entitas testing, entitas ini mempunyai tiga atribut antara lain id testing, konten dan label dimana id testing berperan sebagai *primary key*.
7. Entitas pengujian, entitas ini memiliki beberapa atribut yaitu id pengujian, tanggal, total dataset, data latih, data test, presisi, recall dan akurasi. Id pengujian pada entitas ini adalah *primary key*.
   1. **Relasi Antar Tabel (RAT)**

Relasi Antar Tabel (RAT) akan merepresentasikan hubungan antar tabel yang berfungsi untuk mengatur operasi suatu basis data. Rancangan RAT pada Penerapan *Deep Learning* Untuk Klasifikasi Berita *Hoax* Menggunakan *Convolutional Neural Network* dapat dilihat pada Gambar 3.



**Gambar 3.** RAT Klasifikasi Berita *Hoax*

* 1. **Struktur Tabel**

Basis data dalam penelitian ini terdiri dari beberapa tabel yang masing-masing tabel dapat dilihat sebagai berikut:

1. Tabel Admin

Tabel admin berfungsi untuk menyimpan semua data yang berhubungan dengan admin. Data tersebut digunakan sebagai hak akses untuk masuk ke sistem. Pada tabel admin mempunyai 4 buah atribut yaitu id\_admin, *username*, *password* dan nama\_admin dimana id\_admin sebagai *primary key*. Struktur tabel admin dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Admin

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_admin | Int (11) | *Primary Key, Auto Increment* | Id admin |
| 2 | username | Varchar (50) | *None* | *Username admin* |
| 3 | password | Varchar (50) | *None* | *Password admin* |
| 4 | nama\_admin | Varchar (50) | *None* | Nama admin |

2. Tabel Cek Hoax

Tabel cek hoax berfungsi untuk menyimpan data berupa berita yang ingin dicek atau diprediksi kebenarannya. Selain itu, tabel ini juga berisi hasil dari setiap tahapan dalam proses prediksi. Tabel cek hoax terdiri dari 15 atribut yang salah satunya sebagai *primary key* yaitu id\_cek. Struktur tabel cek hoax dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Cek Hoax

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_cek | Int (11) | *Primary Key, Auto Increment* | Id cek |
| 2 | Konten | Text | *None* | Konten yang ingin dicek |
| 3 | Tanggal | Date | *None* | Tanggal pengecekan |
| 4 | case\_folding | Text | *None* | Hasil case folding |
| 5 | remove\_punctuation | Text | *None* | Hasil remove punctuation |
| 6 | remove\_number | Text | *None* | Hasil remove number |
| 7 | remove\_whitespace | Text | *None* | Hasil remove whitespace |
| 8 | tokenizing | Text | *None* | Hasil tokenizing |
| 9 | stopword | Text | *None* | Hasil stopword |
| 10 | stemming | Text | *None* | Hasil steming |
| 11 | hasil\_preprocessing | Text | *None* | Hasil preprocessing |
| 12 | vektor | Text | *None* | Hasil vektorisasi |
| 13 | probabilitas | Varchar (50) | *None* | Probabilitas prediksi |
| 14 | probabilitas\_pembulatan | Varhcar (50) | *None* | Probabilitas dibulatkan |
| 15 | klasifikasi | Varchar (10) | *None* | Hasil klasifikasi cek hoax |

3. Tabel Dataset

Tabel dataset berfungsi untuk menyimpan data yang akan digunakan sebagai dataset saat admin ingin melakukan proses *upgrade* model prediksi. Tabel dataset terdiri dari tiga atribut yaitu id\_dataset, konten dan label. Struktur dari tabel dataset dapat dilihat lebih jelas pada Tabel 3.

**Tabel 3.** Struktur Tabel Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_dataset | Int (11) | *Primary Key, Auto Increment* | Id dataset |
| 2 | konten | Text | *None* | Konten dataset |
| 3 | label | Varchar (10) | *None* | Label konten dataset |

4. Tabel Lapor Hoax

Tabel lapor hoax merupakan tabel yang berisi kumpulan data dari laporan *user.* Jika *user* menemukan suatu berita *hoax* atau fakta maka *user* dapat melaporkan berita tersebut melalui sistem dan data laporannya akan disimpan pada tabel ini. Atribut dari tabel lapor hoax antara lain id\_lapor, nama\_depan, nama\_belakang, email, konten dan tanggal. Id\_lapor merupakan atribut dari tabel lapor hoax yang bertindak sebagai *primary key.* Untuk lebih jelasnya, tabel lapor hoax dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Lapor Hoax

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_lapor | Int (11) | *Primary Key, Auto Increment* | Id lapor |
| 2 | nama\_depan | Varchar (50) | *None* | Nama depan pelapor |
| 3 | nama\_belakang | Varchar (50) | *None* | Nama belakang pelapor |
| 4 | email | Varchar(100) | *None* | Email pelapor |
| 5 | konten | Text | *None* | Isi konten laporan |
| 6 | tanggal | Date | *None* | Tanggal pelaporan |

5. Tabel Training

Tabel training berisi dataset yang bertindak sebagai data training saat pembuatan model atau saat *upgrade* model prediksi. Maka dari itu, setelah data pada tabel dataset digunakan untuk *upgrade* model maka data pada tabel dataset tersebut yang berperan sebagai data training akan pindah ke tabel training. Tabel training ini memiliki tiga atribut yaitu id\_training, konten dan label. Tabel training dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Training

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_training | Int (11) | *Primary Key, Auto Increment* | Id Training |
| 2 | konten | Text | *None* | Konten training |
| 3 | label | Varchar (10) | *None* | Label konten training |

6. Tabel Testing

Sama halnya dengan tabel training, tabel testing juga berisi data dari tabel dataset yang telah digunakan untuk pembuatan atau *upgrade* model prediksi. Data dari tabel dataset yang akan pindah ke tabel testing hanya data yang berperan sebagai data testing saat proses pembuatan atau *upgrade* model. Tabel testing ini memiliki tiga atribut yaitu id\_testing, konten dan label. Tabel testing dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_testing | Int (11) | *Primary Key, Auto Increment* | Id testing |
| 2 | konten | Text | *None* | Konten testing |
| 3 | label | Varchar (10) | *None* | Label konten testing |

7. Tabel Pengujian

Tabel pengujian berfungsi untuk menyimpan data yang berhubungan dengan hasil dari proses pengujian model prediksi. Tabel pengujian ini terdiri dari delapan atribut yaitu id\_pengujian, tanggal, total\_dataset, data\_latih, data\_test, presisi, recall dan akurasi. Id\_pengujian pada tabel ini sebagai *primary key*. Struktur dari tabel pengujian dapat dilihat pada Tabel 3.

**Tabel 3.** Struktur Tabel Pengujian

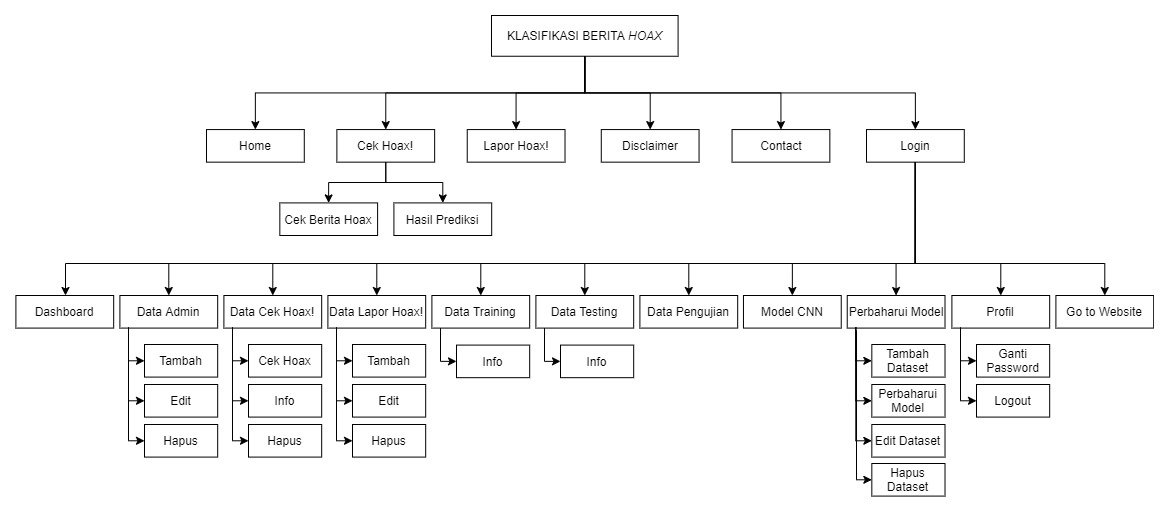
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_pengujian | Int (11) | *Primary Key, Auto Increment* | Id pengujian |
| 2 | tanggal | Date | *None* | Tanggal pengujian |
| 3 | total\_dataset | Int (11) | *None* | Total dataset yang diuji |
| 4 | data\_latih | Int (11) | *None* | Total data latih |
| 5 | data\_test | Int (11) | *None* | Total data test |
| 6 | presisi | Float | *None* | Hasil presisi pengujian |
| 7 | recall | Float | *None* | Hasil recall pengujian |
| 8 | akurasi | Float | *None* | Hasil akurasi pengujian |

#### 3.3.1.4 Perancangan Antarmuka

Pada bagian ini akan memuat hal yang berkaitan dengan perancangan antarmuka sistem. Pada perancangan antarmuka sistem pada penelitian ini meliputi perancangan struktur menu dan *user interface*.

1. **Perancangan Struktur Menu**

Bagian ini akan menggambarkan menu-menu yang terdapat pada sistem Klasifikasi Berita *Hoax* beserta sub-menunya*.* Secara lebih rinci, struktur menu yang dihasilkan dapat dilihat pada Gambar 3.



**Gambar 3.** Struktur Menu

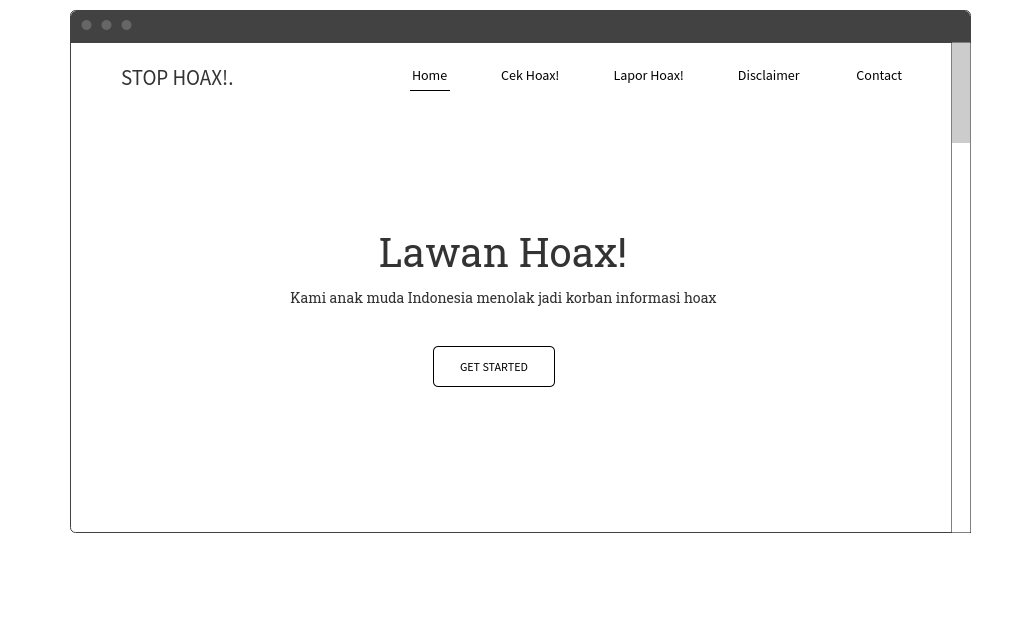
1. **Perancangan *User Interface***

Perancangan *user interface* merupakan model mekanisme komunikasi antara pengguna dengan sistem. Antarmuka atau *user interface* dapat menerima informasi dan memberikan informasi pada pengguna dalam bentuk tampilan yang interaktif. Secara garis besar, *user intercace* pada sistem ini akan dibagi menjadi dua aktor, yaitu *user* dan *admin*.

Berikut merupakan rancangan *user interface* pada aktor *user* yang terdiri dari beberapa bagian, diantaranya adalah sebagai berikut:

1. Rancangan Halaman *Home*

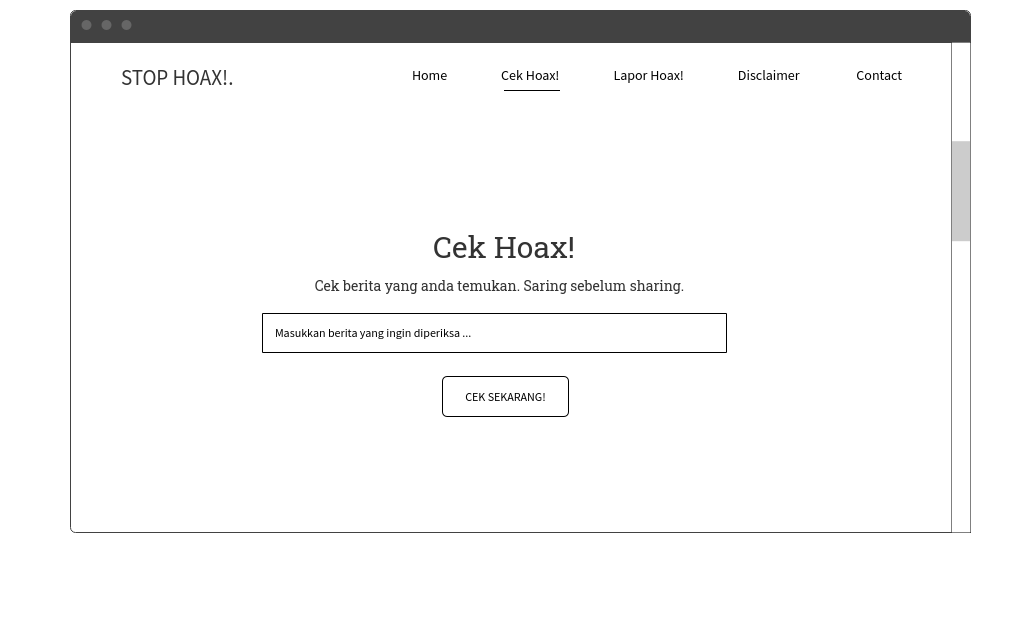
Halaman *home* merupakan halaman yang akan tampil pertama kali saat diakses oleh *user*. Pada halaman ini berisi *header* yang memuat judul dan slogan. Selain itu, terdapat *button Get Started* yang berfungsi untuk mengarahkan *user* ke halaman Cek Hoax!. Rancangan halaman *home* dapat dilihat pada Gambar 3.

­­

**Gambar 3.** Rancangan Halaman *Home*

1. Rancangan Halaman Cek Hoax!

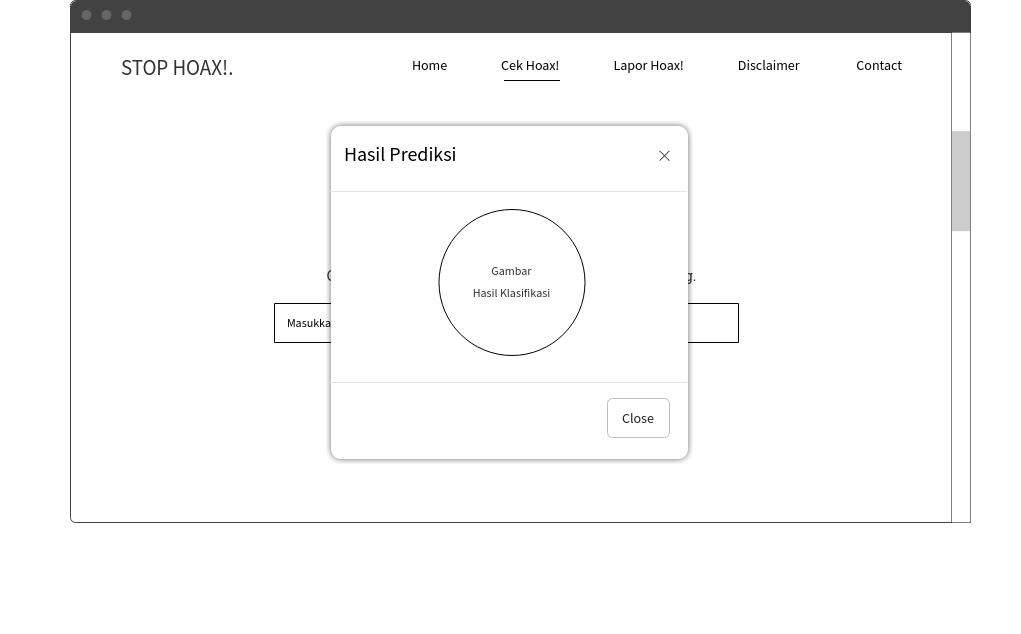
Halaman cek hoax merupakan halaman yang berisi *field* untuk melakukan pengecekan atau prediksi suatu konten berita, apakah berita tersebut *hoax* atau fakta. Pengecekan atau prediksi akan diproses setelah menekan *button* Cek Sekarang. Rancangan dari halaman cek hoax dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman Cek Hoax!

1. Rancangan Halaman Hasil Prediksi

Apabila menekan tombol Cek Sekarang! pada halaman cek hoax maka akan muncul halaman hasil prediksi. Halaman ini akan menampilkan gambar yang menunjukkan bahwa berita yang dicek termasuk berita *hoax* atau fakta. Rancangan halaman hasil prediksi seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Hasil Prediksi

1. Rancangan Halaman Lapor Hoax!

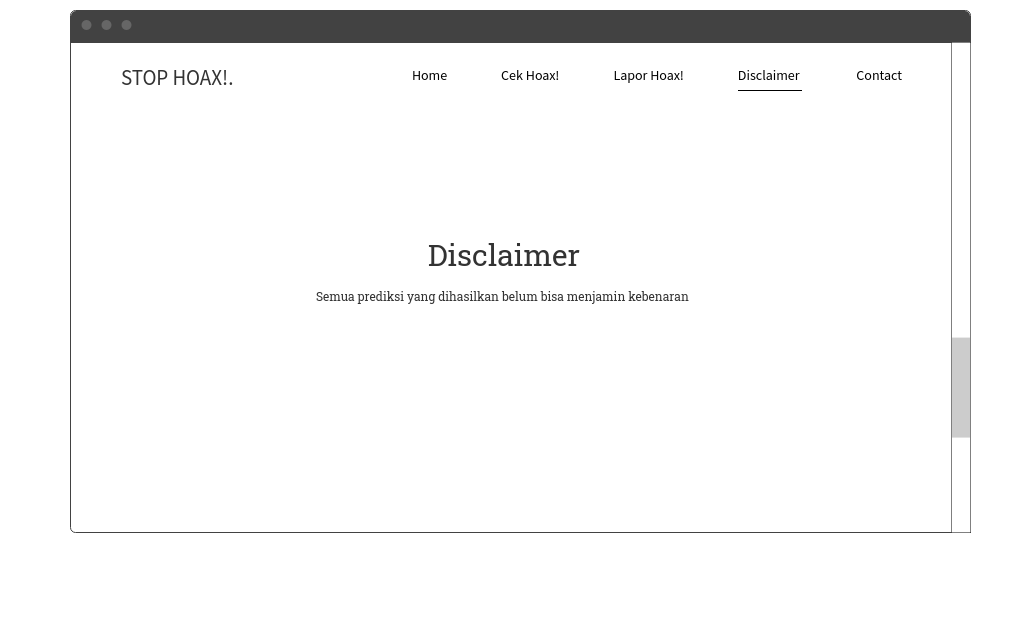
Halaman lapor hoax adalah halaman yang memuat fitur untuk melapor apabila menemukan berita *hoax* atau fakta di media sosial atau sebagainya dengan membuat sebuah form laporan. Pada form laporan ini, *user* mengisi data pribadi seperti nama depan, nama belakang dan email *user* serta konten berita yang ingin dilaporkan. *Laporan* ini akan dikirim ke *admin* untuk ditindaklanjuti. Rancangan dari halaman lapor hoax dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman Lapor Hoax!

1. Rancangan Halaman *Disclaimer*

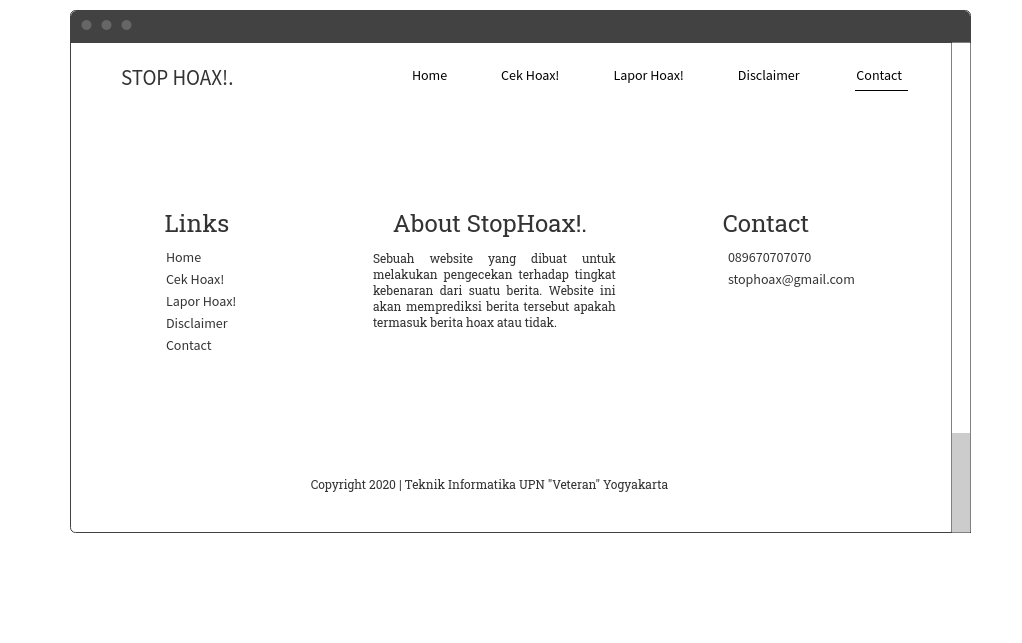
Halaman *disclaimer* ini berisi peringatan berupa batasan dan ruang lingkup bahwa sistem ini hanya menampilkan hasil prediksi dan tidak menjamin kebenaran prediksi tersebut. Rancangan dari halaman *disclaimer* dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman *Disclaimer*

1. Rancangan Halaman *Contact*

Halaman *contact* adalah halaman yang berisi informasi tentang sistem ini seperti link menuju ke menu-menu *user*, penjelasan singkat tentang sistem yang dibuat dan kontak yang dapat dihubungi. Rancangan halaman *contact* dapat dilihat pada Gambar 3.

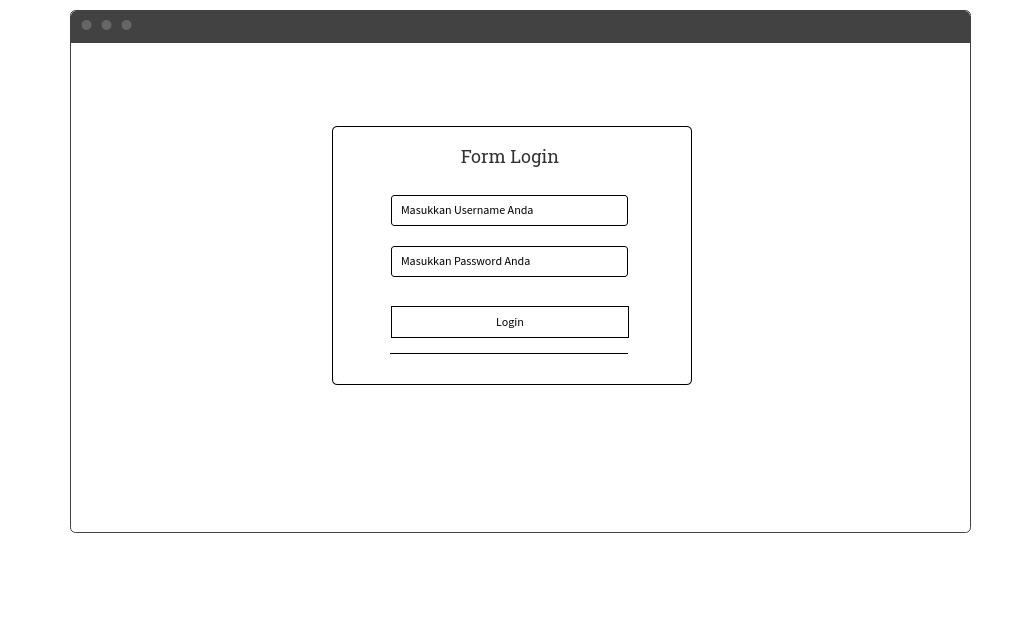


**Gambar 3.** Rancangan Halaman *Contact*

Setelah *user interface* untuk aktor *user,* berikut akan ditampilkan *user interface* pada aktor *admin* yang terdiri dari beberap a bagian, diantaranya yaitu:

1. Rancangan Halaman *Login*

Halaman *login* adalah halaman yang akan tampil pertama kali apabila berperan sebagai aktor *admin* jika ingin masuk ke dalam sistem. Adminmengisi *username* dan *password* yang benar pada *form login*. Rancangan dari halaman *login* dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman *Login*

1. Rancangan Halaman *Dashboard*

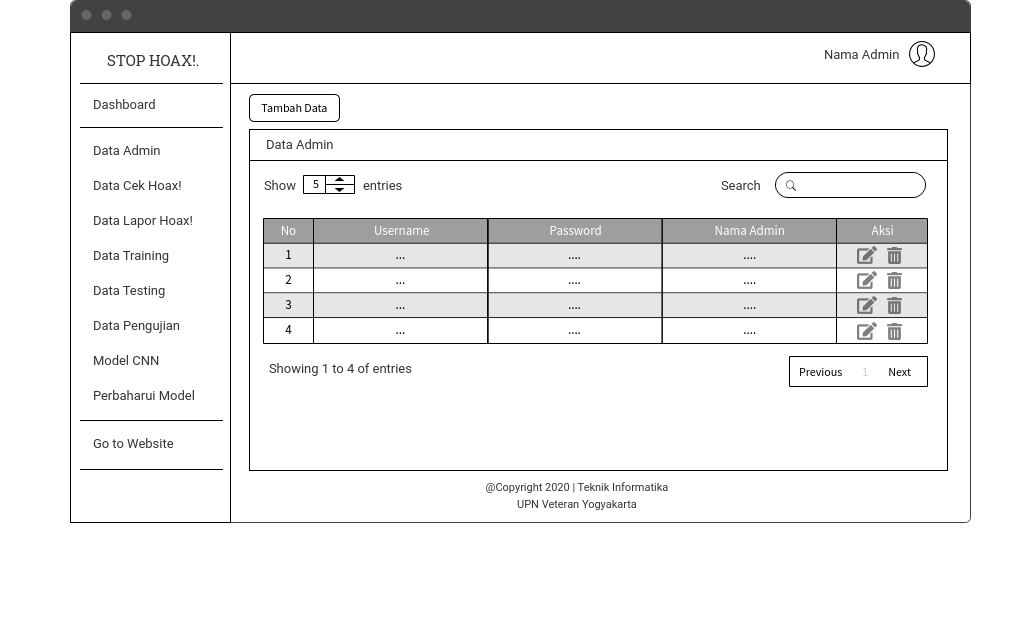
Setelah *admin* berhasil melakukan *login,* maka akan tampil halaman *dashboard*. Halaman *dashboard* berisi jumlah data dari masing-masing tabel pada sistem ini yaitu jumlah data admin, jumlah data cek hoax, lapor hoax, dataset, data training dan data testing serta jumlah data pengujian. Berikut rancangan dari halaman *dashboard* pada Gambar 3.



**Gambar 3.** Rancangan Halaman *Dashboard*

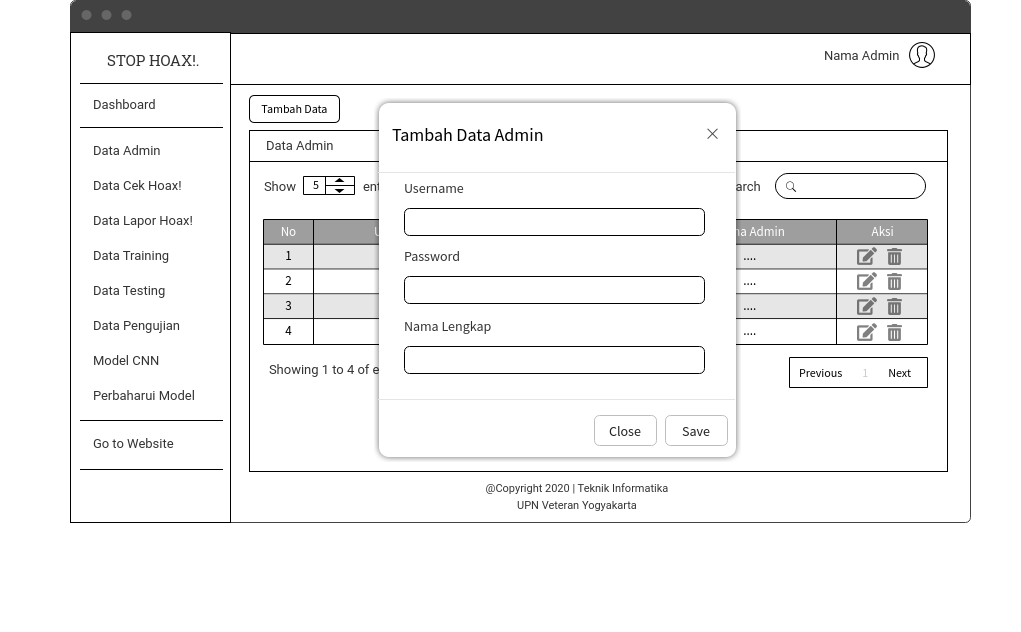
1. Rancangan Halaman Data Admin

Halaman data admin akan menampilkan data-data admin yang divisualisasikan dalam bentuk tabel. Ada beberapa fitur yang tersedia pada halaman ini seperti fitur pencarian dan pengurutan data, navigator tabel dan jumlah baris tabel yang ingin ditampilkan pada setiap halaman tabel. Selain itu, pada halaman ini juga memiliki beberapa fitur untuk mengolah data seperti menambah data admin, mengedit data admin serta untuk menghapus data admin. Rancangan untuk halaman data admin dapat dilihat pada Gambar 3.



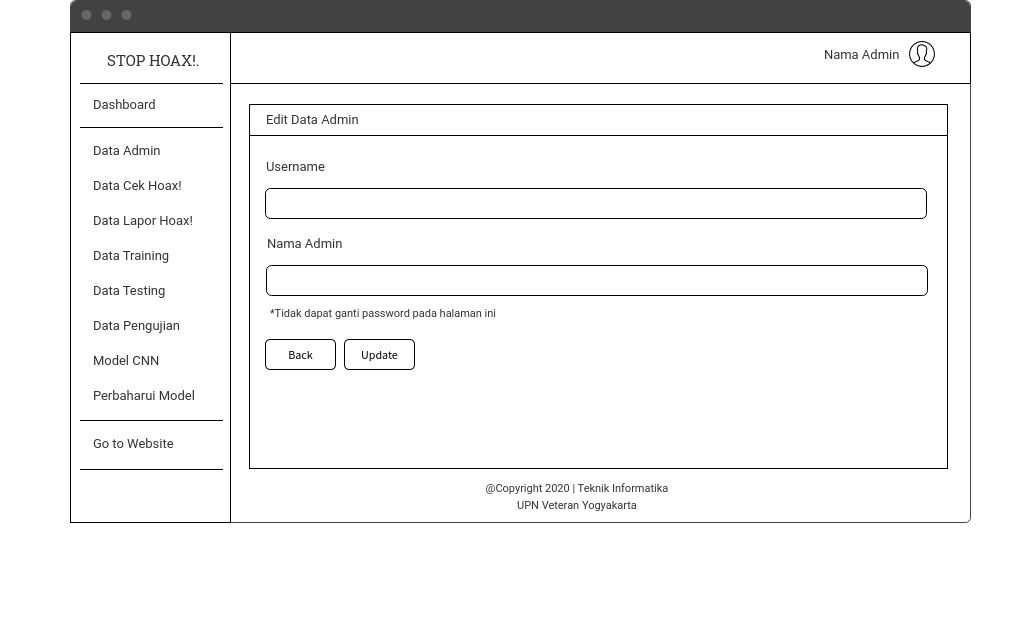
**Gambar 3.** Rancangan Halaman Data Admin

Apabila *admin* menekan tombol tambah data pada halaman data admin maka akan muncul *pop up* yang berisi *form* untuk menambah data admin. Isi dari *form* tersebut terdiri dari *username, password* dan nama lengkap. Jika admin ingin menyimpannya, tekan tombol *save*. Detail dari rancangan halaman tambah data admin ini seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Tambah Data Admin

Fungsi lain yang ada di halaman data admin yaitu edit data. Tombol edit data direpresentasikan dengan gambar pensil pada kolom aksi di tabel. Pada halaman edit data admin, admin merubah sesuai dengan baris data yang dipilih. Data admin yang dapat diubah hanya *username* dan nama lengkap admin, sedangkan untuk *password* hanya dapat diubah oleh admin yang bersangkutan pada menu profil. Apabila sudah mengganti data admin yang dipilih dengan data admin yang baru pada form yang tersedia, selanjutnya tekan tombol *update* untuk menyimpannya. Rancangan halaman pada edit data admin seperti pada Gambar 3.

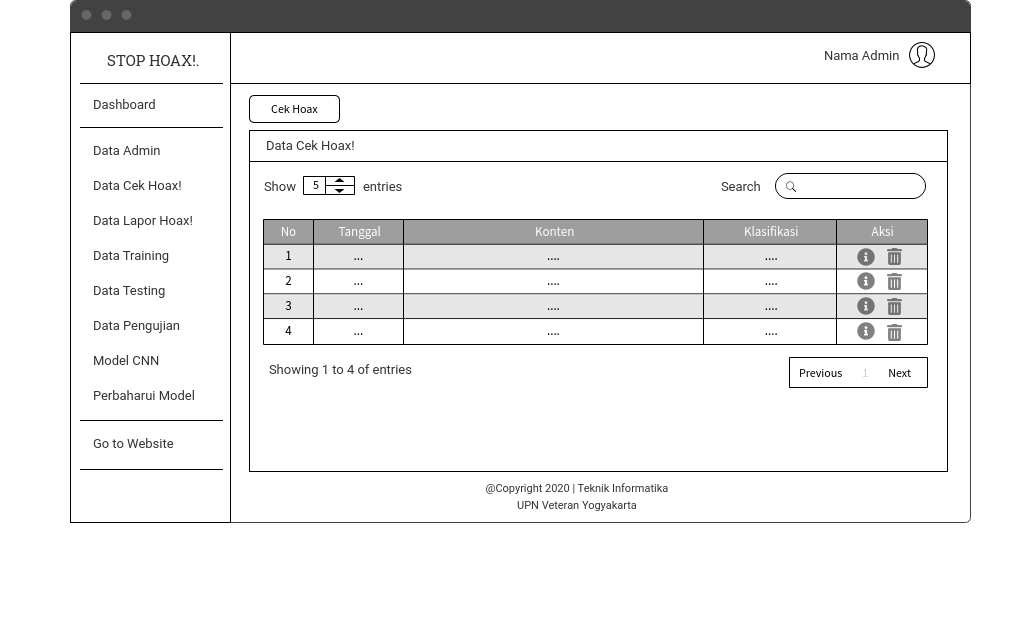


**Gambar 3.** Rancangan Halaman Edit Data Admin

Lalu, tombol yang berfungsi untuk menghapus data pada halaman data admin pada gambar tempat sampah di kolom aksi. Tekan *button* tersebut berdasarkan baris yang dipilih, maka data adminnya akan terhapus.

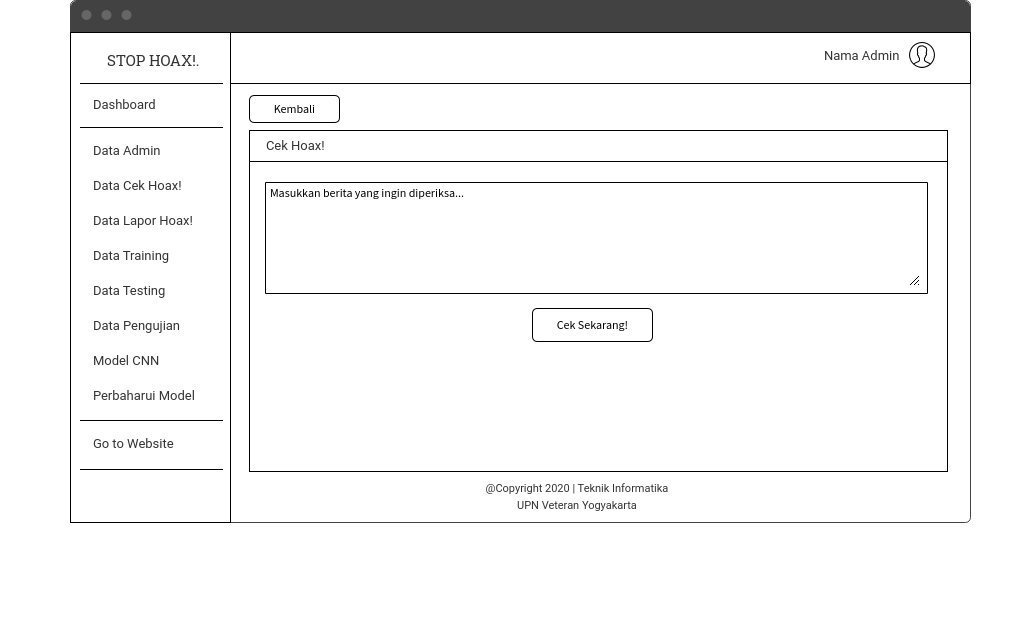
1. Rancangan Halaman Data Cek Hoax!

Halaman data cek hoax adalah halaman pada admin yang akan menampilkan data yang ada pada tabel cek hoax dari database. Data cek hoax yang ditampilkan yaitu no, tanggal, konten, klasifikasi dan kolom aksi untuk proses olah data. Halaman ini juga sama seperti halaman data admin yang mempunyai fitur untuk pencarian dan pengurutan data, fitur cek hoax, lihat data dan hapus data serta navigator tabel. Halaman ini tidak ada proses edit data bertujuan meminimalisir adanya manipulasi data hasil prediksi. Sehingga hasil yang ditampilkan merupakan hasil *real* dari model yang telah dibuat. Berikut rancangan dari halaman data cek hoax yang akan dibuat seperti pada Gambar 3.



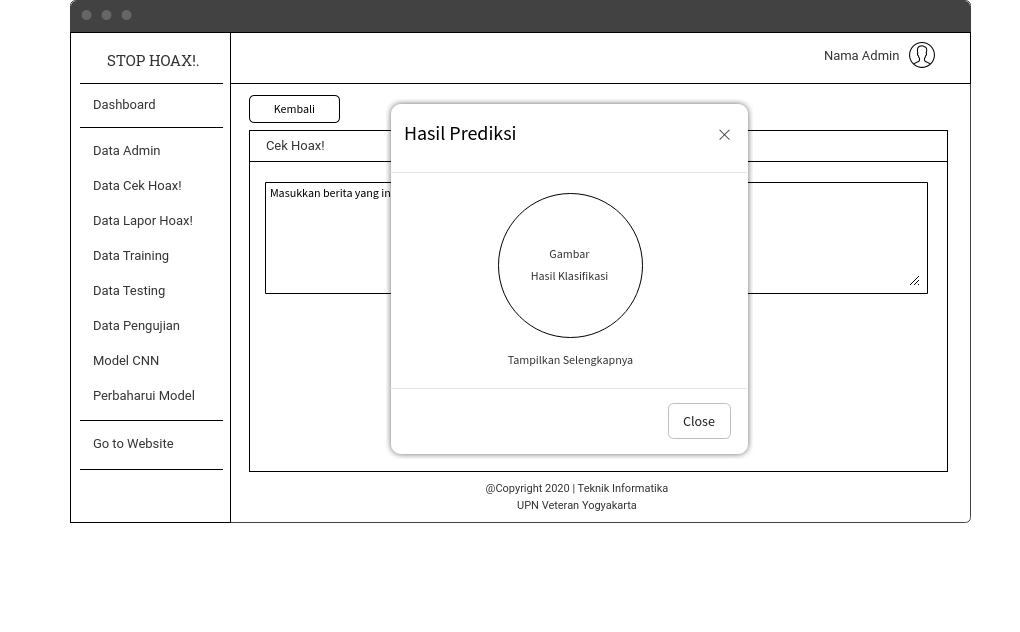
**Gambar 3.** Rancangan Halaman Data Cek Hoax

Selanjutnya yaitu fungsi pada tombol cek hoax pada halaman ini. Tombol tersebut untuk pindah ke halaman cek hoax. Halaman cek hoax berfungsi untuk melakukan proses pengecekan berita hoax seperti pada proses pengecekan hoax di halaman *user*. Dimana admin memasukkan konten berita yang ingin dicek lalu menekan *button* Cek Sekarang untuk melakukan proses pengecekan. Halaman cek hoax pada bagian admin ini akan dibuat seperti rancangan pada Gambar 3.



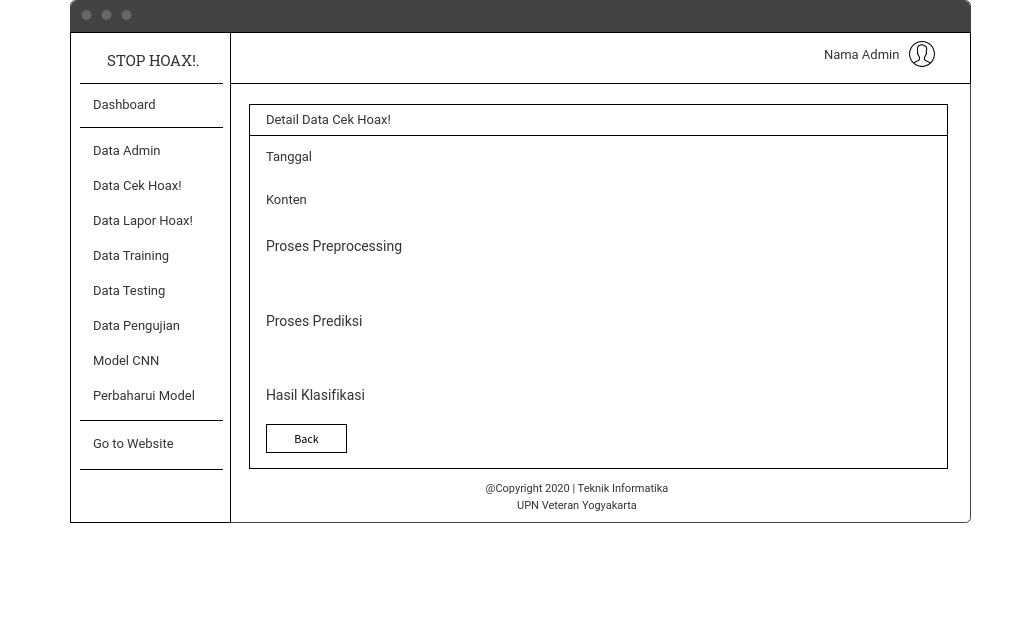
**Gambar 3.** Rancangan Halaman Cek Hoax Admin

Setelah menekan tombol Cek Sekarang, sistem akan melakukan proses klasifikasi dari konten yang dicek. Lalu, akan muncul *pop up* yang akan menampilkan hasil prediksinya dalam bentuk gambar *hoax* atau fakta. Rancangan *pop up* dari hasil prediksi tersebut dapat dilihat pada Gambar 3



**Gambar 3.** Rancangan HalamanHasil Prediksi Admin

Beda dari halaman cek hoax di *user*, pada halaman ini terdapat tombol *tampilkan selengkapnya* yang berfungsi untuk menampilkan hasil dari setiap tahapan proses klasifikasi secara detail. Halaman detail cek hoax tersebut akan seperti pada rancangan di Gambar 3. Halaman tersebut berisi tanggal pengecekan, konten, proses *preprocessing,* proses prediksi dan hasil klasifikasinya.

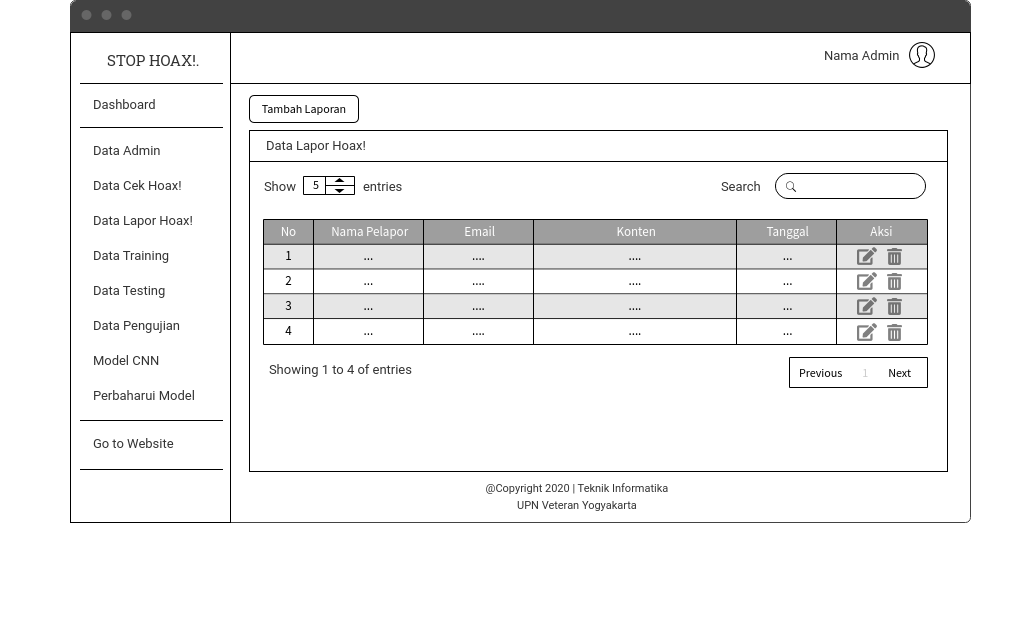


**Gambar 3.** Rancangan Halaman Detail Cek Hoax

Halaman detail cek hoax tersebut juga akan muncul saat menekan tombol yang bergambar i yang ada di kolom aksi pada halaman data cek hoax. Fitur olah data lain yang dapat dilakukan pada halaman data cek hoax yaitu hapus data cek hoax. Cara untuk menghapusnya dengan memilih baris yang dihapus dan menekan tombol yang bergambar tempat sampah di kolom aksi.

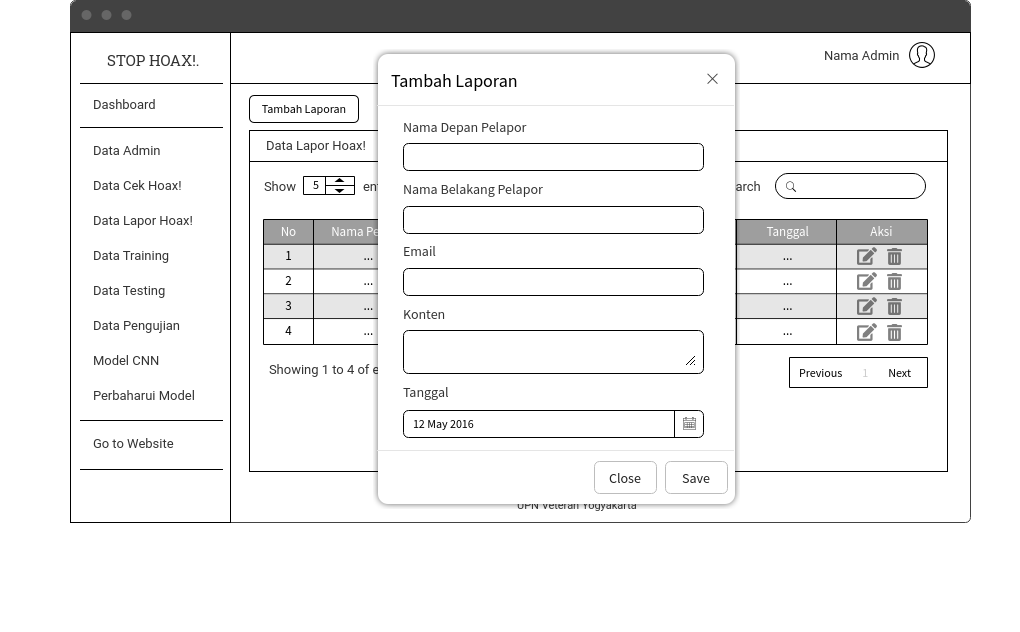
1. Rancangan Halaman Data Lapor Hoax!

Halaman data lapor hoax merupakan halaman yang berisi data-data laporan yang dikirim oleh *user* ataupun yang diinput oleh admin. Data yang ditampilkan yaitu nama pelapor, email, konten yang dilaporkan dan tanggal saat melakukan laporan. Admin dapat mengolah data seperti menambah data laporan, edit laporan dan hapus laporan. Rancangan dari halaman data lapor hoax dapat dilihat pada Gambar 3.



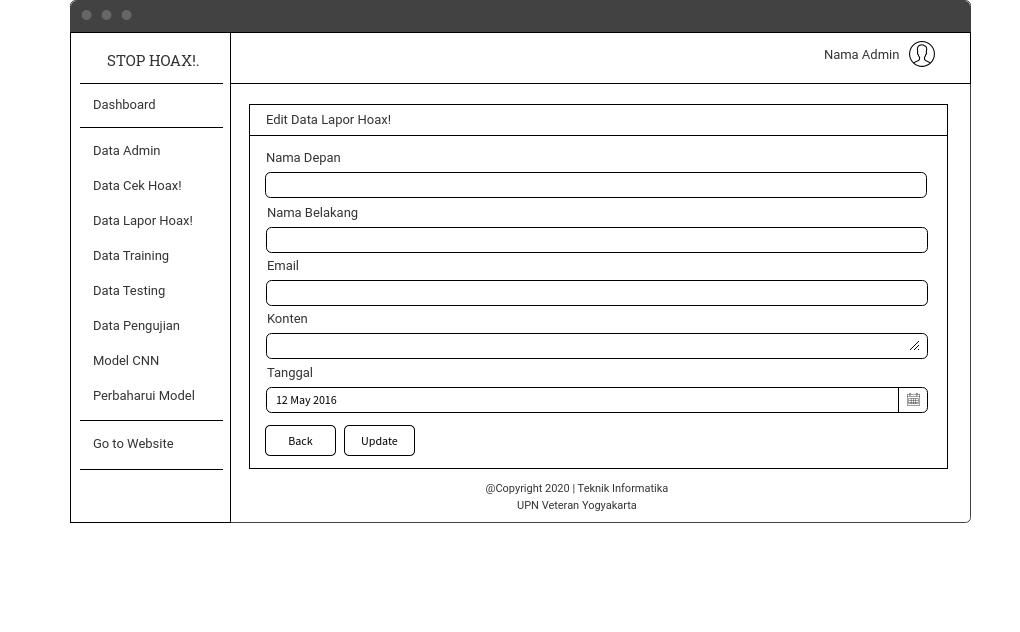
**Gambar 3.** Rancangan Halaman Data Lapor Hoax

Admin dapat menambah data lapor hoax dengan menekan tombol tambah laporan, kemudian mengisi *form* yang terdiri dari nama depan, nama belakang, email, konten dan tanggal pelaporan. Jika sudah, tekan *save* untuk menyimpan data tersebut. Rancangan dari Halaman tambah data lapor hoax dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman Tambah Data Lapor Hoax

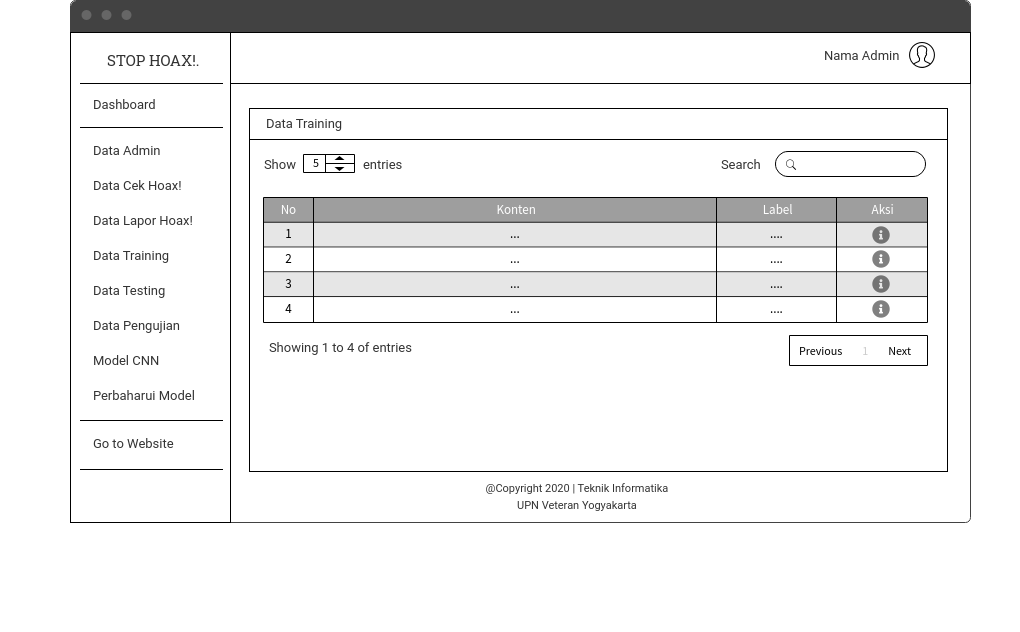
Selain menambah data, admin juga dapat mengubah data lapor hoax. Untuk mengubah data dengan menekan tombol yang bergambar pensil. Pada halaman edit data lapor hoax, akan muncul form yang telah terisi sesuai data yang dipilih. Kemudian admin dapat mengubah dan menyimpannya dengan menekan tombol *update*. Rancangan dari halaman edit data lapor hoax dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman Edit Data Lapor Hoax

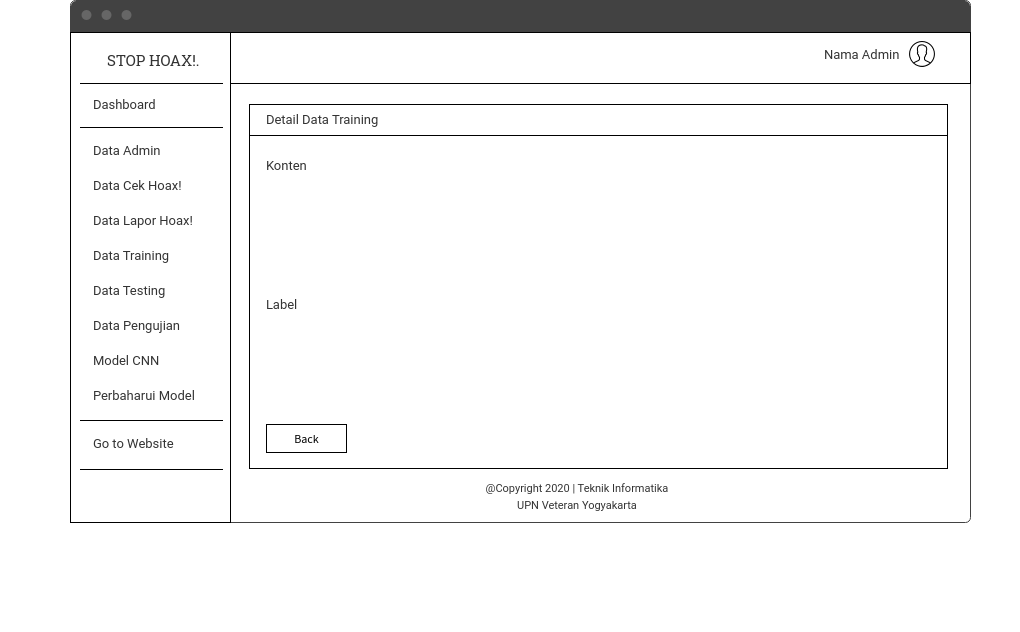
1. Rancangan Halaman Data Training

Halaman data training ini akan menampilkan data training yang berisi konten, label dan aksi untuk melihat data. Pada halaman data training ini tidak ada proses olah data seperti tambah dan hapus, hanya ada tombol bergambar i untuk melihat data secara lengkap. Rancangan dari halaman data training seperti pada Gambar 3.



**Gambar 3.**  Rancangan Halaman Data Training

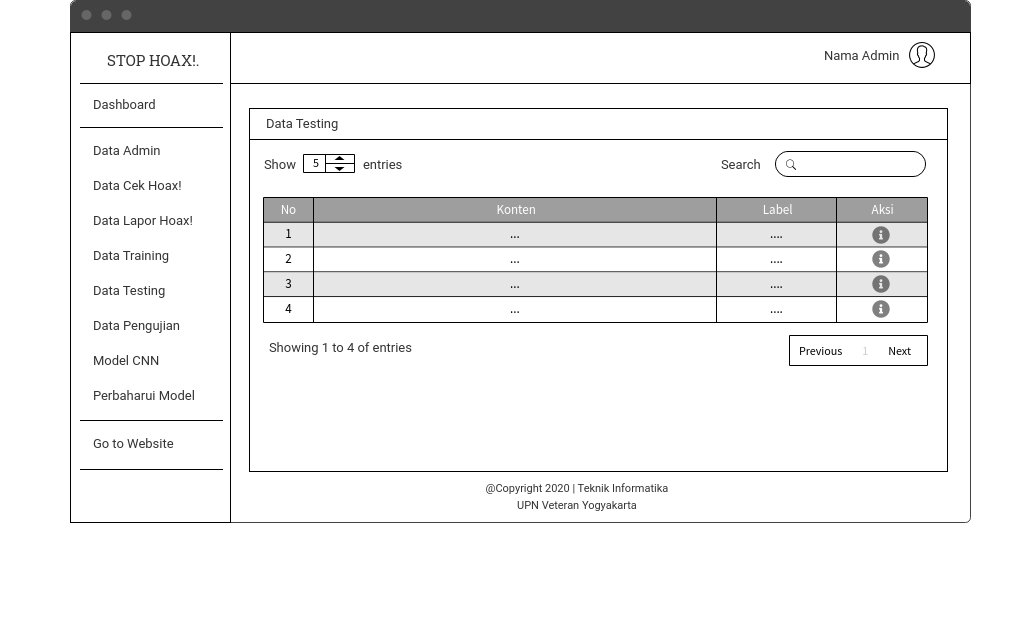
Ketika admin menekan *button* i pada halaman data training, akan muncul halaman detail data training yang menampilkan konten secara lengkap beserta labelnya. Berikut rancangan dari detail data training pada Gambar 3.



**Gambar 3.** Rancangan Halaman Detail Data Training

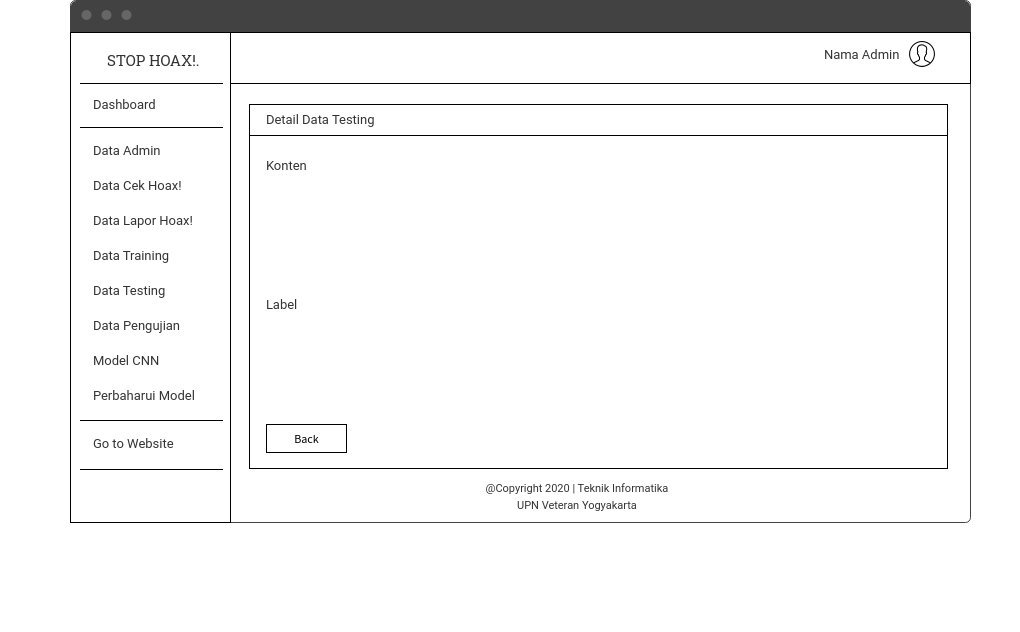
1. Rancangan Halaman Data Testing

Halaman data testing tidak jauh berbeda dengan data training, hanya berbeda di bagian data yang ditampilkan. Halaman data testing ini akan menampilkan tabel training yang berisi data konten dan label serta ada kolom aksi untuk ke halaman detail data testing. Rancangan dari halaman detail data testing dapat dilihat pada Gambar 3.



**Gambar 3.** Rancangan Halaman Data Testing

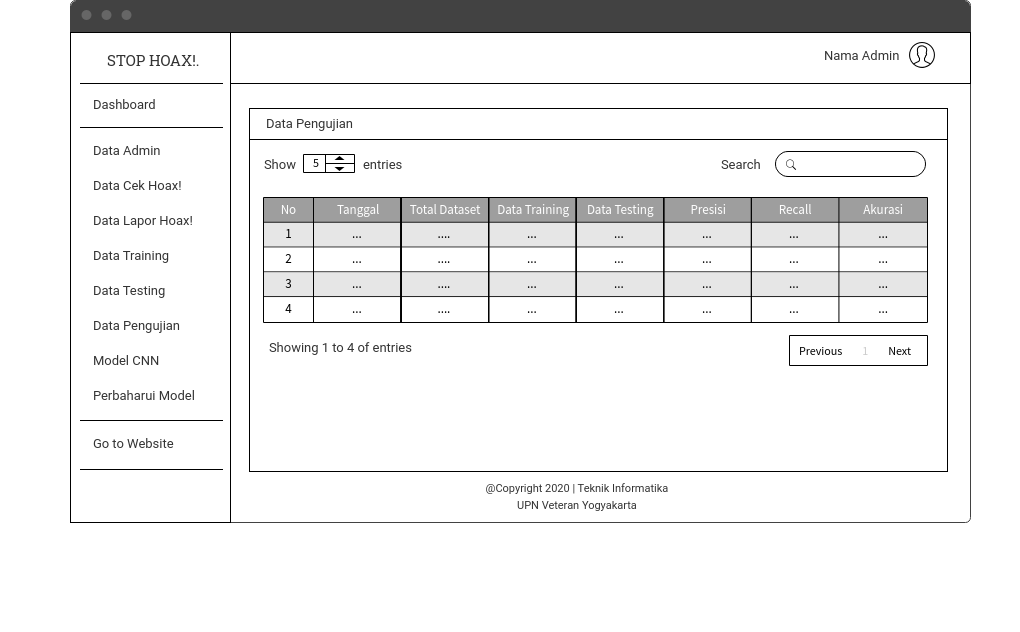
Ketika admin menekan *button* i pada kolom aksi tersebut maka akan ke halaman detail data testing untuk menampilkan data konten secara lengkap beserta labelnya sesuai dengan baris yang dipilih. Berikut rancangan dari halaman detail data testing pada Gambar 3.



**Gambar 3.** Rancangan Halaman Detail Data Testing

1. Rancangan Halaman Data Pengujian

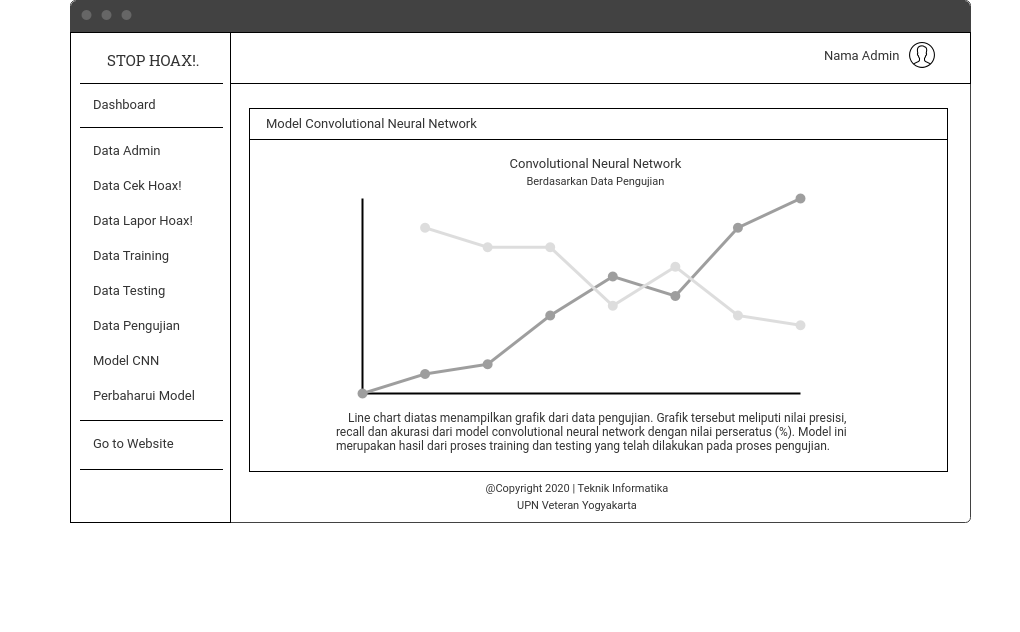
Halaman data pengujian akan berisi data yang memuat tanggal, total dataset, total data traning dan data testing, presisi, recall dan akurasi yang berasal dari tabel pengujian di database. Pada halaman ini juga memiliki fitur yang biasa ada di tabel seperti pencarian dan pengurutan data, navigator tabel serta jumlah data yang ditampilkan per halaman. Rancangan dari halaman data pengujian seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Data Pengujian

1. Rancangan Halaman Model CNN

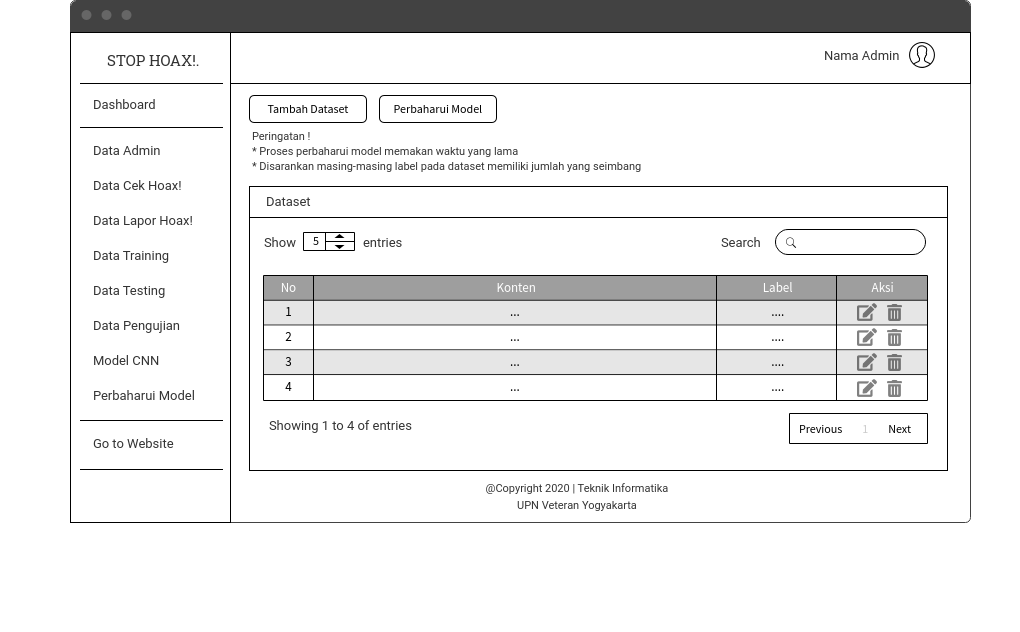
Halaman model CNN yaitu halaman yang akan menampilkan grafik dari prosentase nilai presisi, recall dan akurasi berdasarkan data pengujian. Data tersebut akan ditampilkan menggunakan *line chart.* Rancangan dari halaman model CNN akan dibuat seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Model CNN

1. Rancangan Halaman Perbaharui Model

Halaman perbaharui model merupakan halaman untuk *update* model dengan data baru dengan melalui proses *training* dan *testing* menggunakan model yang sudah dibuat. Dengan adanya perbaharui model ini diharapkan dapat meningkatkan akurasi dari model sehingga hasil prediksinya menjadi lebih akurat. Halaman ini berisi dataset yang akan digunakan dan fitur olah data seperti tambah, edit dan hapus dataset. Rancangan dari halaman perbaharui model seperti pada Gambar 3.



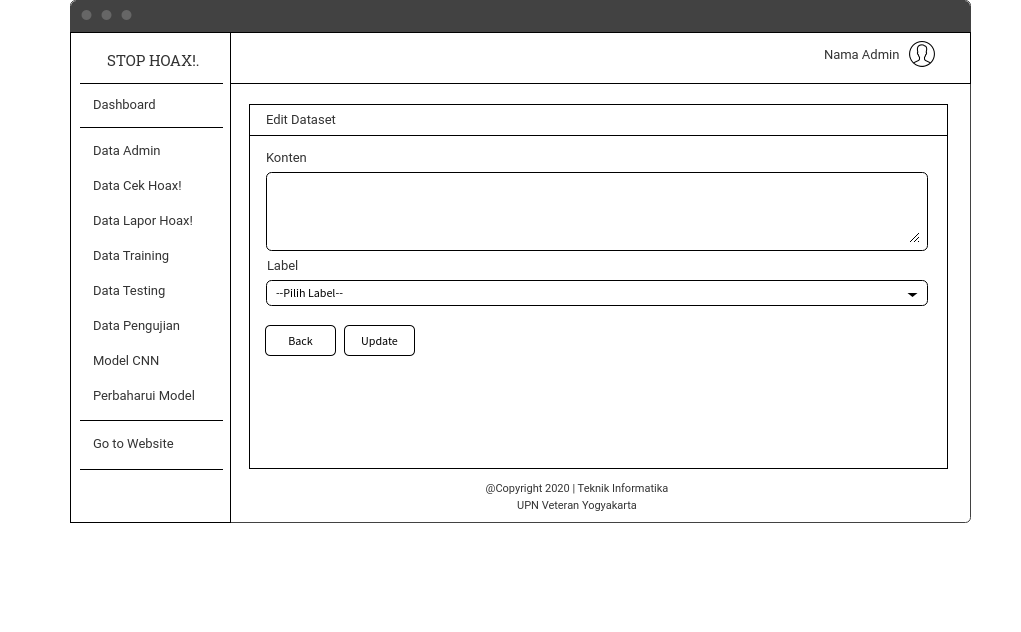
**Gambar 3.** Rancangan Halaman Perbaharui Model

Jika menekan tombol tambah dataset, maka akan muncul *pop up* yang berisi form yang terdiri dari konten dan label. Untuk menyimpannya dengan menekan tombol *save*. Rancangan dari halaman tambah dataset akan seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Tambah Dataset

Dari data yang telah disimpan, admin dapat melakukan edit dataset dengan mengklik *button* yang bergambar pensil di kolom aksi. Maka data yang telah di pilih akan muncul di halaman edit dataset. Terdapat tombol *update* untuk merubah data tersebut. Rancangan dari halaman edit dataset dapat dilihat pada Gambar 3.

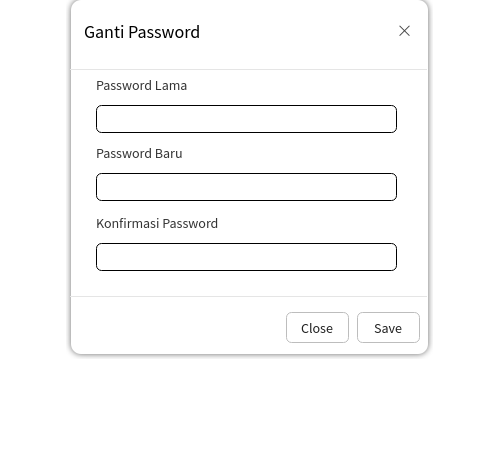


**Gambar 3.** Rancangan Halaman Edit Dataset

Selanjutnya fitur hapus dataset dengan menekan tombol yang bergambar tempat sampah di kolom aksi. Maka data yang dipilih akan terhapus.

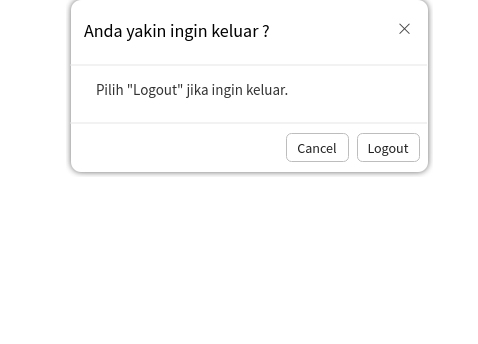
1. Rancangan Halaman Profil

Bagian profil admin terdapat pada bagian sebelah kanan atas di halaman *admin* dengan menekan *icon* akun. Maka akan muncul dua pilihan yaitu ganti password dan *logout*. Untuk mengganti password admin dengan mengisi form yang terdiri dari password lama, password baru dan konfirmasi password. Rancangan dari halaman ganti password akan dibuat seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman Ganti Password

Selain itu, pada menu profil terdapat sub menu *logout*. Menu tersebut berfungsi untuk keluar dari sistem. Rancangan dari tampilan *logout* seperti pada Gambar 3.



**Gambar 3.** Rancangan Halaman *Logout*

### 3.3.2 Perancangan Pengujian

Pada bagian ini akan dibuat suatu rancangan pengujian. Pembuatan dari rancangan pengujian bertujuan untuk mempersiapkan konsep pengujian yang akan dilakukan untuk mengukur kinerja dari suatu sistem yang telah dibangun. Metode pengujian yang akan digunakan pada penelitian ini yaitu *black box testing* serta *k-fold cross validation* dan *confusion matrix*.

Pengujian menggunakan *black box* bertujuan untuk mengukur kinerja dari suatu perangkat lunak yang telah dibangun. Metode pengujian secara *black box* lebih menekankan pada bagian fungsionalitas dari suatu perangkat lunak sehingga dapat memastikan setiap fungsi yang ada berjalan dengan semestinya. Berikut detail rancangan pengujianmenggunakan *black box* dapat dilihat pada Tabel 3.

**Tabel 3.** Rancangan Pengujian *Black Box*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aktor** | **Halaman** | **Detail Pengujian** | **Pengujian** | |
| **Berhasil** | **Gagal** |
| **User** | Home | Menampilkan daftar menu |  |  |
| Menampilkan header |  |  |
| Fungsi tombol *Get Started* |  |  |
| Cek Hoax | Menampilkan daftar menu |  |  |
| Fungsi cek hoax |  |  |
| Menampilkan hasil klasifikasi |  |  |
| Lapor Hoax | Menampilkan daftar menu |  |  |
| Pengajuan laporan hoax |  |  |
| Disclaimer | Menampilkan daftar menu |  |  |
| Menampilkan disclaimer sistem |  |  |
| Contact | Menampilkan daftar menu |  |  |
| Menampilkan informasi sistem |  |  |
| Berpindah ke menu lain via links |  |  |
| **Admin** | Login | Melakukan proses login |  |  |
| Fungsi session |  |  |
| Dashboard | Menampilkan jumlah setiap data |  |  |
| Data Admin | Menampilkan data admin |  |  |
| Menambahkan data admin |  |  |
| Mengedit data admin |  |  |
| Menghapus data admin |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Data Cek Hoax | Menampilkan data cek hoax |  |  |
| Melakukan pengecekan hoax |  |  |
| Menampilkan hasil klasifikasi |  |  |
| Menampilkan detail cek hoax |  |  |
| Menghapus data cek hoax |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Data Lapor Hoax | Menampilkan data laporan hoax |  |  |
| Menambahkan data laporan hoax |  |  |
| Mengedit data laporan hoax |  |  |
| Menghapus data laporan hoax |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Data Training | Menampilkan data training |  |  |
| Menampilkan detail data training |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Data Testing | Menampilkan data testing |  |  |
| Menampilkan detail data testing |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Data Pengujian | Menampilkan data pengujian |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Model CNN | Menampilkan grafik hasil pengujian |  |  |
| Mengolah data grafik |  |  |
| Pembaharui Model | Menampilkan data dataset |  |  |
| Menambahkan data dataset |  |  |
| Mengedit data dataset |  |  |
| Menghapus data dataset |  |  |
| Melakukan proses pembaharui model |  |  |
| Fungsi pencarian dan pengurutan |  |  |
| Profil | Fungsi ganti password |  |  |
| Fungsi logout |  |  |
| Go to Website | Berpindah ke halaman user |  |  |

Selain menggunakan *black box,* penelitian ini juga akan menggunakan metode pengujian *k-fold cross validation* dan *confusion matrix*. *K-fold cross validation* bertujuan untuk melakukan validasi dari suatu model dengan melakukan pengujian sebanyak K, dimana pada penelitian ini menggunakan K=5. Sedangkan, *confusion matrix* untuk mengukur tingkat akurasi dari suatu model. Berikut rancangan pengujian menggunakan *k-fold cross validation* pada Tabel 3. dan rancangan pengujian *confusion matrix* dapat dilihat pada Tabel

**Tabel 3.** Rancangan *K-Fold Cross Validation*

|  |  |  |  |
| --- | --- | --- | --- |
| **Fold** | **Akurasi** | **Presisi** | **Recall** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| **Rata-Rata** |  |  |  |

**Tabel 3.** Rancangan *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
| **Aktual** | **Hasil Prediksi** | | |
| **Kelas** | Fakta | Hoax |
| Fakta |  |  |
| Hoax |  |  |

# BAB IV

# HASIL, PENGUJIAN DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai implementasi dari hasil rancangan yang telah dibuat sebelumnya. Implementasi ini termasuk ke dalam tahap membangun *prototype* yang meliputi pembuatan aplikasi hingga siap dioperasikan pada keadaan sebenarnya. Setelah itu, akan dilakukan pengujian terhadap sistem yang telah dibuat untuk mengetahui kinerjanya baik dari sisi jalannya sistem maupun performa dari algoritma yang digunakan. Metode pengujian pada penelitian ini terdiri dari *black box testing, k-fold cross validation* dan *confusion matrix*. Tahap pengujian ini diharapkan dapat mengetahui kekurangan-kekurangan pada penelitian ini sehingga dapat dilakukan pengembangan selanjutnya.

## 4.1 Hasil Penelitian

### 4.1.1 Hasil Perancangan Klasifikasi Berita *Hoax*

#### Modul *Text Preprocessing*

Pada tahap ini, data konten akan dilakukan proses *text preprocessing* yang bertujuan untuk membersihkan data sehingga siap digunakan. Berikut ini ada modul-modul yang digunakan pada proses *text preprocessing*.

1. **Modul *Case Folding***

Modul *case folding* berfungsi untuk merubah semua karakter pada data konten menjadi huruf kecil (*lowercase*). Apabila pada data konten terdapat huruf kapital maka akan diubah menjadi huruf kecil. Berikut ini adalah modul *case folding*.

|  |
| --- |
| def case\_folding(data):    data = data.lower()    return data  hasil\_case\_folding = []  for i in range(len(df.Konten)):    hasil\_case\_folding.append(case\_folding(df.Konten[i])) |

**Modul Program 4.** *Source Code* Proses *Case Folding*

1. **Modul *Remove Punctuation***

Modul *remove punctuation* merupakan proses untuk menghapus tanda baca pada suatu data karena dianggap sebagai *delimiter*. Data yang digunakan merupakan hasil dari proses *case folding*. Modul *remove punctuation* dapat dilihat pada Modul Program 4.

|  |
| --- |
| def remove\_punctuation(data):  data = data.translate(str.maketrans('','',string.punctuation))  return data  hasil\_remove\_punctuation = []  for i in range(len(hasil\_case\_folding)):  hasil\_remove\_punctuation.append(remove\_punctuation  (hasil\_case\_folding[i])) |

**Modul Program 4.** *Source Code* Proses *Remove Punctuation*

1. **Modul *Remove Number***

Hasil dari proses *remove punctuation* akan dilanjutkan ke proses *remove number*. Dimana pada proses ini akan menghapus karakter yang berupa angka. Modul dari proses *remove number* dapat dilihat pada Modul Program 4.

|  |
| --- |
| def remove\_numbers(data):  data = re.sub(r'\d+','',data)  return data  hasil\_remove\_numbers = []  for i in range(len(hasil\_remove\_punctuation)):  hasil\_remove\_numbers.append(remove\_numbers  (hasil\_remove\_punctuation[i])) |

**Modul Program 4.** *Source Code* Proses *Remove Number*

1. **Modul *Remove Whitespace***

Modul *remove whitespace* berfungsi untuk menghapus karakter *whitespace*. Apabila menemukan karakter *whitespace* akan dihapus sedangkan jika bukan karakter *whitespace* maka karakter tersebut tetap disimpan. Berikut modul *remove whitespcae* pada Modul Program 4.

|  |
| --- |
| def remove\_whitespace(data):  data = data.strip()  return data  hasil\_remove\_whitespace = []  for i in range(len(hasil\_remove\_numbers)):  hasil\_remove\_whitespace.append(remove\_whitespace  (hasil\_remove\_numbers[i])) |

**Modul Program 4.** *Source Code* Proses *Remove Whitespace*

1. **Modul *Tokenizing***

Modul *tokenizing* merupakan tahapan pada *text preprocessing* yang digunakan untuk memotong data menjadi per kata dengan spasi (“ “) sebagai pemisahnya. Data yang digunakan sebagai inputan merupakan hasil dari proses *remove whitespace*. Hasil dari proses *tokenizing* ini akan berbentuk token-token. Modul *tokenizing* dapat dilihat pada Modul Program 4.

|  |
| --- |
| from nltk.tokenize import word\_tokenize  def tokenizing(data):  data = word\_tokenize(data)  return data  hasil\_tokenizing = []  for i in range(len(hasil\_remove\_whitespace)):  hasil\_tokenizing.append(tokenizing(hasil\_remove\_whitespace[i])) |

**Modul Program 4.** *Source Code* Proses *Tokenizing*

1. **Modul *Stopword Removal***

Hasil dari proses *tokenizing* yang berupa per kata (token) akan dilakukan proses *stopword removal*. Proses *stopword removal* yaitu membuang kata-kata yang kurang penting (*stopword*) pada data seperti kata ganti orang, kata penghubung dan sebagainya. Modul *stopword removal* dapat dilihat pada Modul Program 4.

|  |
| --- |
| import nltk  nltk.download('stopwords')  nltk.download('punkt')  from nltk.corpus import stopwords  list\_stopwords = set(stopwords.words('indonesian'))  def stopwords(data):  hasil\_list = []  for i in data:  if i not in list\_stopwords and len(i)>=3 and len(i)<=12 and i not  in hasil\_list:  hasil\_list.append(i)  data = hasil\_list  return data  hasil\_stopword = []  for i in range(len(hasil\_tokenizing)):  hasil\_stopword.append(stopwords(hasil\_tokenizing[i])) |

**Modul Program 4.** *Source Code* Proses *Stopword Removal*

1. **Modul *Stemming***

Modul *stemming* digunakan untuk mencari kata dasar dari setiap kata hasil *stopword removal* dengan menghapus imbuhan pada masing-masing kata. Hasil dari proses *stemming* ini juga merupakan hasil dari proses *text preprocesing.* Berikut modul *stemming* dapat dilihat pada Modul Program 4.

|  |
| --- |
| from Sastrawi.Stemmer.StemmerFactory import StemmerFactory  factory = StemmerFactory()  stemmer = factory.create\_stemmer()  def stemming(data):  data\_akhir = []  for i in data:  kata = stemmer.stem(i)  data\_akhir.append(kata)  return data\_akhir  hasil\_stemming = []  for i in range(len(hasil\_stopword)):  hasil\_stemming.append(stemming(hasil\_stopword[i])) |

**Modul Program 4.** *Source Code* Proses *Stemming*

#### 4.1.1.2 Modul *Word Embedding*

Pada modul *word embedding* ini akan merubah kata dari hasil *preprocessing* menjadi vektor yang berisi kumpulan angka. Setiap kata akan direpresentasikan oleh vektor. Vektor yang dihasilkan pada proses ini memiliki dimensi 100. Maksud dari dimensi ini yaitu panjang vektor atau banyaknya vektor, jadi satu kata akan dibuah menjadi deretan angka sebanyak 100. Penggunaan modul ini karena algoritma hanya bisa memproses data berbentuk angka dan tidak bisa berbentuk string. Modul *word embedding* dapat dilihat pada Modul Program 4.

|  |
| --- |
| from gensim import models  model = Word2Vec(df['Tokens'], size=100, min\_count=1)  model.wv.save\_word2vec\_format('modelakhir.txt', binary=False) |

**Modul Program 4.** *Source Code* Proses *Word Embedding*

#### 4.1.1.3 Modul *Convolutional Neural Network*

**a. Modul *Training***

|  |
| --- |
| from keras.layers import Embedding  from keras.layers import Dense, Input, Flatten  from keras.layers import Conv1D, MaxPooling1D, Embedding, Dropout  from keras.models import Model  from keras.callbacks import ModelCheckpoint  from keras.callbacks import EarlyStopping  embedding\_layer = Embedding(len(train\_word\_index)+1,  EMBEDDING\_DIM,  weights=[train\_embedding\_weights],  input\_length=MAX\_SEQUENCE\_LENGTH,  trainable=True)  sequence\_input = Input(shape=(MAX\_SEQUENCE\_LENGTH,), dtype='int32')  embedded\_sequences = embedding\_layer(sequence\_input)  l\_cov1= Conv1D(128, 5, activation='relu')(embedded\_sequences)  l\_pool1 = MaxPooling1D(5)(l\_cov1)  l\_cov2 = Conv1D(128, 5, activation='relu')(l\_pool1)  l\_pool2 = MaxPooling1D(5)(l\_cov2)  l\_cov3 = Conv1D(128, 5, activation='relu')(l\_pool2)  l\_pool3 = MaxPooling1D(35)(l\_cov3) # global max pooling  l\_flat = Flatten()(l\_pool3)  l\_dense = Dense(128, activation='relu')(l\_flat)  preds = Dense(len(macronum), activation='softmax')(l\_dense)  model = Model(sequence\_input, preds)  model.compile(loss='binary\_crossentropy',  optimizer='adam',  metrics=['acc'])  print("Simplified convolutional neural network")  model.summary()  cp=ModelCheckpoint('model\_cnn.hdf5',monitor='val\_acc',verbose=1,save\_best\_only=True)  num\_epochs = 100  batch\_size = 32  es\_callback =EarlyStopping(monitor='val\_loss', patience=10, verbose=1)  hist = model.fit(x\_train, y\_tr, epochs=num\_epochs,  validation\_split=0.2, shuffle=True ,  batch\_size=batch\_size,callbacks=[es\_callback]) |

**Modul Program 4.** *Source Code* Proses *Training*

**b. Modul *Testing***

|  |
| --- |
| from sklearn.metrics import auc  from sklearn.metrics import plot\_roc\_curve  from sklearn.model\_selection import StratifiedKFold  predictions = model.predict(test\_cnn\_data, batch\_size=1, verbose=1)  labels = [0,1]  prediction\_labels=[]  for p in predictions:  prediction\_labels.append(labels[np.argmax(p)])  akurasi=sum(y[test]==prediction\_labels)/len(prediction\_labels)  y\_test= y[test].values  y\_tes=[]  for i in range(0, len(y\_test)):  y\_tes.append(y\_test[i])  y\_pred=[]  for p in predictions:  y\_pred.append(labels[np.argmax(p)])  cf\_sentimen = pd.DataFrame(data=confusion\_matrix(y\_tes, y\_pred, labels=labels), columns=labels, index=labels)  print(cf\_sentimen)  tps\_sentimen = {}  fps\_sentimen = {}  fns\_sentimen = {}  tns\_sentimen = {}  for label in labels:  tps\_sentimen[label] = cf\_sentimen.loc[label, label]  fps\_sentimen[label] = cf\_sentimen[label].sum() - tps\_sentimen[label]  fns\_sentimen[label] = cf\_sentimen.loc[label].sum() - tps\_sentimen[label]  for label in set(y\_tes):  tns\_sentimen[label] = len(y\_tes) - (tps\_sentimen[label] + fps\_sentimen[label] + fns\_sentimen[label])  accuracy\_global\_new\_sentimen=sum(tps\_sentimen.values())/len(y\_tes)  acc.append(accuracy\_global\_new\_sentimen)  tpfp\_sentimen = [ai + bi for ai, bi in zip(list(tps\_sentimen.values()), list(fps\_sentimen.values()))]  precision=[ai / bi if bi>0 else 0 for ai , bi in zip(list(tps\_sentimen.values()), tpfp\_sentimen)]  precisionSentimen=sum(precision)/2  presisi.append(precisionSentimen)  tpfn\_sentimen = [ai + bi for ai, bi in zip(list(tps\_sentimen.values()), list(fns\_sentimen.values()))]  recall=[ai / bi if bi>0 else 0 for ai, bi in zip(list(tps\_sentimen.values()), tpfn\_sentimen)]  recallSentimen=sum(recall)/2  recal.append(recallSentimen)  fpr, tpr, thresholds = roc\_curve(y\_test, y\_pred)  nilai\_auc = auc(fpr, tpr)  interp\_tpr = interp(mean\_fpr, fpr, tpr)  interp\_tpr[0] = 0.0  tprs.append(interp\_tpr)  aucs.append(nilai\_auc)  ax.plot(mean\_fpr, interp\_tpr, label=r'ROC Fold %d (AUC = %f)' % (fold, nilai\_auc),lw=1, alpha=0.3) |

**Modul Program 4.** *Source Code* Proses *Testing*

**c. Modul Prediksi**

|  |
| --- |
| from keras.preprocessing.sequence import pad\_sequences  from keras.models import load\_model  from keras.models import model\_from\_json  #Load Model  json\_file = open('C:/xampp/htdocs/hoax\_classification/model/model\_cnn.json', 'r')  model\_json = json\_file.read()  model = model\_from\_json(model\_json)  model.load\_weights("C:/xampp/htdocs/hoax\_classification/model/model\_cnn.h5")  with open('C:/xampp/htdocs/hoax\_classification/model/tokenizer.pickle', 'rb') as handle:  tokenizer = pickle.load(handle)  preprocessing = [' '.join(sen) for sen in hasil\_stemming]  sequences = tokenizer.texts\_to\_sequences(hasil\_stemming)  str\_sequences = str(sequences)  word\_index = tokenizer.word\_index  x\_test = pad\_sequences(sequences, maxlen=1000)  prediksi = model.predict(x\_test, batch\_size=1, verbose=0)  probabilitas = "FAKTA : " + str(prediksi[0][0]) + " | HOAX : " + str(prediksi[0][1])  probabilitas\_pembulatan = "FAKTA : " + str(round(prediksi[0][0])) + " | HOAX : " + str(round(prediksi[0][1]))  class\_category = ['Fakta', 'Hoax']  for i in range(prediksi.shape[0]):  klasifikasi = class\_category[prediksi[i].argmax()] |

**Modul Program 4.** *Source Code* Proses Prediksi

### 4.1.2 Hasil Perancangan Aplikasi

## 4.2 Pengujian Sistem

**Daftar Pustaka**

Abdullah, A. H. (2018). Word Embedding. Retrieved February 29, 2020, from https://rpubs.com/ website: https://rpubs.com/ahmadhusain/wordembedding

Afriza, A., & Adisantoso, J. (2018). Metode Klasifikasi Rocchio untuk Analisis Hoax. *Jurnal Ilmu Komputer Dan Agri-Informatika*, *5*(1), 1–10. https://doi.org/10.29244/jika.5.1.1-10

Assidik, G. K. (2018). Kajian Identifikasi dan Upaya Penangkalan Pemberitaan Palsu (Hoax) Pada Pembelajaran Bahasa Indonesia. *Kongres Bahasa Indonesia*.

Berry, M. W., & Kogan, J. (2010). *Text Mining : Applications and Theory*. United Kingdom: WILEY.

Fajri, D. (2019). Menkominfo: Ada 900 Ribu Situs Penyebar Informasi Hoax. Retrieved June 29, 2019, from www.okezone.com website: https://news.okezone.com/read/2019/02/14/337/2018062/menkominfo-ada-900-ribu-situs-penyebar-informasi-hoax

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data* (1st ed.). New York: Cambridge University Press.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text Mining Methods and Techniques. *International Journal of Computer Applications*, *85*(17), 42–45. https://doi.org/10.5120/14937-3507

Goller. (2000). Automatic Document Classification: A Thorough Evaluation of Various Methods. *Proceedings of International Symposium on Information Theory and Its Application*, 145–162.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, (17), 26–32.

Hashimi, H., Hafez, A., & Mathkour, H. (2015). Selection criteria for text mining approaches. *College of Computer and Information Sciences*.

Hughes, M., Li, I., Kotoulas, S., & Suzumura, T. (2017). *Medical Text Classification using Convolutional Neural Networks*.

Kotler, P., & Keller, K. L. (2012). *Marketing Management* (14th ed.). United State of America: Pearson.

Lhaksmana, K. M., Nhita, F., & Budhiarto, A. (2017). Klasifikasi Pengguna Media Sosial Twitter Dalam Persebaran Hoax Menggunakan Metode Backpropagation Classification of Users Social Media Twitter in the Hoax Spread. *E-Proceeding of Engineering*, *4*(2), 3082–3090.

Librianty, A. (2017). Survei: Media Sosial Jadi Sumber Utama Penyebaran Hoax.

Lina, Q. (2019). Apa itu Convolutional Neural Network ? Retrieved June 25, 2020, from https://medium.com/ website: https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4

MacDougall, C. D. (1958). *Hoaxes*. New York: Dover Publication.

Maulina, D., & Sagara, R. (2018). Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency – Inverse Document Frequency. *Jurnal Mantik Penusa*, *2*(1), 35–40.

Nasution, M. A. (2017). HOAX SEBAGAI BENTUK HUDUD MENURUT HUKUM ISLAM. *Yurisprudentia*, *3*(1).

Nazief, B., & Adriani, M. (1996). *Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia*.

Nikhit, K. (2015). A Web Scraping Approach in Node.js. *International Journal of Science, Engineering and Technology Research*, 909–912.

Pereira, R. C., & Vanitha, T. (2015). Web Scraping of Social Networks. *International Journal of Innovative Research in Computer and Communication Engineering*, *3*(7), 237–240.

Pressman, R. S. (2010). *Software Engineering : A Practitioner’s Approach* (7th ed.). McGraw-Hill Education.

Primas, N. (2017). *Deteksi Cacat Permukaan Buah Manggis Berbasis Pengolahan Citra Menggunakan Metode Klasifikasi Support Vector Machine (SVM)*. Universitas Muhammadiyah Yogyakarta.

Purnomo, D. (2017). Model Prototyping Pada Pengembangan Sistem Informasi. *JIMP - Jurnal Informatika Merdeka Pasuruan*, *2*(2), 54–61.

Razi, A. (2017). *Klasifikasi Artikel Berita Berbahasa Indonesia Menggunakan Convolutional Neural Network*. Universitas Gadjah Mada.

Respati, S. (2017). Mengapa Banyak Orang Mudah Percaya Berita “Hoax”? Retrieved February 9, 2020, from www.kompas.com website: https://nasional.kompas.com/read/2017/01/23/18181951/mengapa.banyak.orang.mudah.percaya.berita

Rini, D. P. (2012). *Metodologi Pengembangan Sistem*. Yogyakarta: Graha Ilmu.

Safko, L. (2012). *The Social Media Bible : Tactics, Tools and Strategies for Business Success* (3rd Editio). New York: John Wiley & Sonic Inc.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, *34*(1), 1–47.

Tandungan, S. (2019). Pengenalan Convolutional Neural Network – Part 1. Retrieved February 13, 2020, from http://sofyantandungan.com/ website: http://sofyantandungan.com/pengenalan-convolutional-neural-network-part-1/

Tanjung, B. S. (2018). Pendekatan Text Mining sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter. *Universitas Widya Kartika*, 1–6.

Tarabay, M. R. (2019). PyTorch and deep learning. Retrieved February 13, 2020, from https://rafietarabay.blogspot.com/2019/05/?m=0

Turland, M. (2010). *php|architect’s Guide to Web Scraping with PHP*. Los Angeles.

Vibriza, J., Rahadi, D. R., Marwan, M. R., & Ahyad. (2017). Perilaku pengguna dan informasi. *Jurnal Gunadarma*, *4*(1), 192–208. https://doi.org/https://doi.org/10.22146/jps.v4i2.28586

Wearesocial. (2019). Digital 2019: Indonesia. Retrieved June 29, 2019, from www.datareportal.com website: https://datareportal.com/reports/digital-2019-indonesia/

Wijayanto, F. I. (2012). Social Media: Definisi, Fungsi, Karakteristik. Retrieved February 9, 2020, from www.prezi.com website: https://prezi.com/vddmcub\_-ss\_/social-media-definisi-fungsi-karakteristik/

Xing, W., & Du, D. (2018). Dropout Prediction in MOOCs : Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research*, (March). https://doi.org/10.1177/0735633118757015

Yin, L. (2018). A Summary of Neural Network Layers. Retrieved February 13, 2020, from https://medium.com/ website: https://medium.com/machine-learning-for-li/different-convolutional-layers-43dc146f4d0e

Zhang, Y., & Wallace, B. (2015). *A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification*.

Zufar, M., & Setiyono, B. (2016). Convolutional Neural Network untuk Pengenalan Wajah Secara Real-Time. *Jurnal Sains Dan Seni ITS*, *5*(2), 72–77.

Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *11*(2), 187–198. https://doi.org/10.22146/ijccs.24716