

ISSN : 2620-6897 (Cetak)  
ISSN : 2620-6900 (Online)

Volume 3, Nomor 1, April 2020

# JIRE

JURNAL INFORMATIKA &  
REKAYASA ELEKTRONIKA



Diterbitkan Oleh LPPM STMIK Lombok

Jln. Basuki Rahmat No.105 Praya, Lombok Tengah - NTB  
e-journal.stmiklombok.ac.id/jire - Telp dan Fax (0370) 654310  
email. lppm@stmiklombok.ac.id



## **DEWAN REDAKSI**

### **Jurnal Manager**

**Wire Bagye, S.Kom.,M.Kom** (STMIK Lombok, SINTA ID : 5992010)

### **Reviewer :**

**Resad Setyadi, S.T., S.Si., MMSI., Ph.D (cand)** - Institut Teknologi Telkom Purwokerto  
SCOPUS ID : 57204172534 SINTA ID : 6113570

**Yesaya Tommy Paulus, S.Kom., MT., Ph.D.** - STMIK Dipanegara Makassar  
SCOPUS ID : 57202829909 SINTA ID : 6002004

**Dr. Cucut Susanto, S. Kom. MSi.** - STMIK Dipanegara Makassar  
SINTA ID : 6138863

**Muhamad Malik Mutoffar, ST., MM., CNSS-** Sekolah Tinggi Teknologi Bandung  
SINTA ID : 6013819

**David, M.Cs., M.Kom** - STMIK Pontianak  
SCOPUS ID : 57200208543 SINTA ID : 5977352

**Indo Intan, S.T., M.T.** STMIK - Dipanegara Makassar  
SCOPUS ID : 57200209088 SINTA ID : 6127241

**I Wayan Agus Arimbawa, ST., M.Eng.** - Universitas Mataram  
SINTA ID : 5973017

**Muhammad Fauzi Zulkarnaen, ST., M.Eng.** - STMIK Lombok  
SINTA ID : 6663733

**Yunanri.W, S.T. M. Kom** - Universitas Teknologi Sumbawa (U.T.S)  
SINTA ID : 6723103

**Sitti Aisa, S.Kom., M.T** - STMIK Dipanegara Makassar  
SINTA ID : 6153893

**Sanjaya Pinem, S.Kom, M.Sc** - Universitas Efarina  
SINTA ID : 6689679

**Zamah Sari, S.T., M.T.** - Universitas Muhammadiyah Prof Dr Hamka  
SINTA ID : 6145745

**Fredy Windana, S.Kom., MT** - Sekolah Tinggi Teknologi Stikma Internasional  
SINTA ID : 5974460

**Hijrah Saputra, ST., M.Sc.** - STMIK Lombok  
SINTA ID : 6667974

**Hairul Fahmi, M.Kom.** - STMIK Lombok  
SINTA ID : 5983160

**Sofiansyah Fadli, S.Kom., M.Kom.** - STMIK Lombok  
SINTA ID : 6073057

### **Editor :**

**Wire Bagye, S.Kom., M.Kom** - STMIK Lombok, SINTA ID : 5992010

**Saikin, S.Kom., M.Kom** - STMIK Lombok

**Halena Muna Bekata, M.Pd.** - Universitas Tribuana Kalabahi, SINTA ID : 6168815

### **Desain Grafis & Web Maintenance**

**Jihadul Akbar, S.Kom.** - STMIK Lombok

### **Secretariat**

**Ahmad Susan Pardiansyah, M.Kom** - STMIK Lombok

## DAFTAR ISI

1	KLASIFIKASI ARITMIA DENGAN HEART RATE VARIABILITY ANALISIS MENGGUNAKAN METODE BACKPROPAGATION <i>Wayan Rimba Bazudewa<sup>1</sup>, I Putu Satwika<sup>2</sup>, I Gede Putu Krisna Juliharta<sup>3</sup></i>	1-10
2	IMPLEMENTASI METODE MRP (MATERIAL REQUIREMENT PLANNING) UNTUK MENCAPAI TARGET PRODUKSI PAKAIAN BERBASIS WEB (STUDI KASUS: UD. DARMAWAN DESA SELAGEK) <i>Mohammad Taufan Asri Zaen<sup>1</sup>, Siti Fatmah<sup>2</sup>, Khairul Imtihan<sup>3</sup></i>	11-19
3	DETEKSI KUALITAS BERAS MENGGUNAKAN SEGMENTASI CITRA BERDASARKAN PECAHAN BULIR DAN SEBARAN WARNA <i>Eko Supriyadi<sup>1</sup>, Achmad Basuki<sup>2</sup>, Riyanto Sigit<sup>3</sup></i>	20-29
4	PERMODELAN VISUAL TINGKAT KETAKUTAN PADA SIMULASI EVAKUASI KEBAKARAN 3D MENGGUNAKAN SELF ASSASSEMENT MANIKIN <i>Iqbal Sabilirasyad<sup>1</sup>, Achmad Basuki<sup>2</sup>, Tri Harsono<sup>3</sup></i>	30-39
5	SISTEM KEAMANAN PEMANTAUAN CCTV ONLINE BERBASIS ANDROID PADA RUMAH CANTIK SYIFA MASBAGIK <i>Ahmad Tantoni<sup>1</sup>, Mohammad Taufan Asri Zaen<sup>2</sup></i>	40-47
6	KOMPARASI ALGORITMA MACHINE LEARNING DAN DEEP LEARNING UNTUK NAMED ENTITY RECOGNITION : STUDI KASUS DATA KEBENCANAAN <i>Nuli Giarsyani<sup>1</sup>, Ahmad Fathan Hidayatullah<sup>2</sup>, Ridho Rahmadi<sup>3</sup></i>	48-57
7	SISTEM PENDUKUNG KEPUTUSAN PENENTUAN RESIKO KEMUNGKINAN TERJADI REAKSI DARAH <i>Abd. Halim<sup>1</sup>, Sri Kusumadewi<sup>2</sup>, Linda Rosita<sup>3</sup></i>	58-65
8	MONITORING PENGATUR KECEPATAN KIPAS ANGIN MENGGUNAKAN SISTEM FUZZY BERBASIS WEB DI SMP BAKTI KELUARGA LUBUKLINGGAU <i>Novi Lestari<sup>2</sup>, Nelly Khairani Daulay<sup>1</sup>, Armanto<sup>3</sup></i>	66-76
9	IMPLEMENTASI JARINGAN INTER-VLAN ROUTING BERBASIS MIKROTIK RB260GS DAN MIKROTIK RB1100AHX4 <i>Ahmad Tantoni<sup>1</sup>, Khairul Imtihan<sup>2</sup>, Wire Bagye<sup>3</sup></i>	77-84
10	PERANCANGAN APLIKASI CETAK DOKUMEN ONLINE BERBASIS ANDROID DI BINER JOMBANG <i>Fauzan Adhim<sup>1</sup>, M. Ali Murtadho<sup>2</sup>, Chandra Sukma A<sup>3</sup></i>	85-90

## KOMPARASI ALGORITMA MACHINE LEARNING DAN DEEP LEARNING UNTUK NAMED ENTITY RECOGNITION : STUDI KASUS DATA KEBENCANAAN

Nuli Giarsyani<sup>1</sup>, Ahmad Fathan Hidayatullah<sup>2</sup>, Ridho Rahmadi<sup>3</sup>

<sup>1,2,3</sup>Program Studi Informatika Program Magister, Universitas Islam Indonesia  
Jln. Kaliurang Km 14,5Yogyakarta55584

<sup>1</sup>[nuli.giarsyani@students.uui.ac.id](mailto:nuli.giarsyani@students.uui.ac.id), <sup>2</sup>[fathan@uui.ac.id](mailto:fathan@uui.ac.id), <sup>3</sup>[ridho.rahmadi@uui.ac.id](mailto:ridho.rahmadi@uui.ac.id)

### Abstract

This study aims to classify entity on tweet and analysis the results of two approaches, Machine Learning and Deep Learning. The named entity type included disaster name, location, time, magnitude and others. Deep Learning algoritma used are Long Short-Term Memory, Gated Recurrent Units, and Convolutional Neural Network. Machine Learning algoritma used are Naïve Bayes, Decision Tree, Support Vector Machine and Random Forest. Based on the results of experiments, Deep Learning obtains accuracy that is superior to Machine Learning. It can be seen from the acquisition of the best Deep Learning accuracy value generated from the Gated Recurrent Units and Long Short-Term Memory with a value of 0.999. While the best Machine Learning accuracy generated from the Random Forest with a value of 0.98.

**Keywords :** *named entity recognition, disaster, machine learning, deep learning*

### Abstrak

Penelitian ini bertujuan untuk melakukan Named Entity Recognition guna mengidentifikasi dan mengklasifikasi kata pada tweet yang memuat informasi bencana ke dalam entitas-entitas yang telah ditentukan. Entitas yang diidentifikasi yaitu jenis bencana, lokasi, waktu, magnitude dan others. Adapun algoritma klasifikasi yang digunakan adalah Machine Learning dan Deep Learning. Algoritma Deep Learning yang digunakan yaitu Long Short-Term Memory, Gated Recurrent Units, dan Convolutional Neural Network. Sedangkan algoritma Machine Learning yang digunakan yaitu Naïve Bayes, Decision Tree, Support Vector Machine dan Random Forest. Berdasarkan hasil eksperimen, Deep Learning memperoleh akurasi yang lebih unggul dari Machine Learning. Hal tersebut dilihat dari perolehan nilai accuracy terbaik Deep Learning dihasilkan dari algoritma Gated Recurrent Units dan Long Short-Term Memory dengan nilai 0.999. Sedangkan perolehan accuracy terbaik Machine Learning dihasilkan dari algoritma Random Forest sebesar 0.98.

**Kata kunci :** *named entity recognition, disaster, machine learning, deep learning*

### 1. PENDAHULUAN

Saat ini, media sosial telah menjadi bagian penting dari kehidupan seseorang. Berbagai kalangan, usia, dan hampir semua lapisan masyarakat Indonesia telah memiliki dan menggunakan media sosial sebagai salah satu sarana guna memperoleh dan menyampaikan informasi kepada publik [1]. Pada prakteknya, media sosial sering mengambil peran media tradisional dalam melaporkan peristiwa atau kejadian terkini [2], seperti bencana alam, di

mana laporan atau informasi yang hadir bisa lebih cepat dan berasal langsung dari lokasi bencana [3]. Oleh sebab itu, media sosial banyak dimanfaatkan oleh organisasi atau lembaga yang bergerak pada bidang penanggulangan bencana untuk memberikan informasi terkait bencana.

Salah satu media sosial yang dimanfaatkan dalam melaporkan informasi bencana alam adalah Twitter. Dalam beberapa tahun terakhir, Twitter telah menjadi saluran utama untuk komunikasi selama bencana alam [4]. Hal tersebut dikarenakan Twitter merupakan salah satu

media jejaring sosial dengan pengguna terbanyak diantara beberapa situs jejaring sosial yang ada [5]. Data menunjukkan bahwa Twitter memiliki sekitar 288 juta pengguna aktif, memposting lebih dari 500 juta tweet per hari, dan memiliki jaringan yang paling cepat berkembang [6]. Sedangkan untuk di negara Indonesia sendiri jumlah pengguna Twitter menempati peringkat ke 5 terbesar di dunia.

Indonesia merupakan negara yang rawan akan bencana. Beberapa bencana alam yang sering terjadi di Indonesia mulai dari gempa, tsunami, banjir, tanah longsor, gunung meletus dan masih banyak lagi yang lainnya. Adanya kejadian ini mendorong pengguna Twitter atau lembaga penanggulangan kebencanaan untuk mengunggah informasi tentang kondisi bencana dari tempat terjadinya bencana. Namun tidak semua tweet yang diunggah tersebut memuat informasi tentang bencana. Untuk dapat memanfaatkan informasi tersebut seperti mendapatkan informasi jenis bencana, lokasi bencana, dan waktu kejadian, maka seseorang harus membaca secara keseluruhan isi tweet tersebut. Apabila jumlahnya banyak, maka seseorang akan membutuhkan waktu yang cukup lama untuk mendapatkan informasi bencana pada tweet tersebut. Oleh sebab itu, dibutuhkan *named entity recognition* (NER) yang dapat digunakan untuk mendapatkan informasi penting secara otomatis dari tweet-tweet tersebut dalam waktu yang relatif singkat.

Named entity recognition adalah proses yang mengekstraksi entitas bernama yang dianggap penting di dalam sebuah teks dan menentukan kategorinya ke dalam kategori yang telah terdefinisi [7]. Pada penelitian ini, entitas yang akan diidentifikasi yaitu jenis bencana, lokasi bencana, waktu kejadian, magnitude dan others. NER pada tweet informasi bencana biasanya dibutuhkan oleh pihak tertentu untuk mengidentifikasi bencana, dan memantau keadaan ketika terjadi bencana.

Penelitian ini bertujuan untuk melakukan *named entity recognition* guna mengidentifikasi dan mengklasifikasikan kata pada tweet yang memuat informasi bencana ke dalam entitas-entitas yang telah ditentukan. Selain itu, penelitian ini juga akan memberikan kontribusi terhadap NER pada domain kebencanaan dalam bahasa Indonesia, karena NER pada domain tersebut masih sangat terbatas. NER pada domain kebencanaan ini merupakan langkah awal untuk topik penelitian terkait *information extraction*, *question answering system*, dan sistem monitoring kebencanaan.

Pendekatan yang digunakan dalam penelitian ini adalah *machine learning* dan *deep learning*. Algoritma *machine learning* yang digunakan pada penelitian adalah *naive bayes*, *support vector machines*, *decision tree*, dan *random forest*.

Sedangkan algoritma *deep learning* yang digunakan adalah *long short-term memory* (LSTM), *convolutional neural network* (CNN), dan *gated recurrent units* (GRU). Pendekatan dengan berbagai macam metode ini bertujuan untuk memberikan gambaran secara komprehensif terhadap performa masing-masing metode yang harapannya dapat dijadikan landasan saintifik bagi kasus-kasus yang serupa.

Penulisan makalah ini terdiri dari lima bagian. Bagian pertama merupakan pendahuluan yang memuat latar belakang dari penelitian ini. Bagian kedua membahas tentang penelitian terdahulu yang mendukung penelitian ini. Bagian ketiga membahas tentang metodologi penelitian. Bagian keempat membahas hasil eksperimen yang diperoleh. Bagian kelima adalah bagian terakhir yang membahas kesimpulan dari penelitian.

## 2. TINJAUAN PUSTAKA DAN TEORI

Penelitian mengenai *named entity recognition* pada domain kebencanaan telah dilakukan oleh beberapa peneliti sebelumnya. Dermawan [19] yang membangun model untuk mengklasifikasi apakah suatu tweet termasuk kategori bencana atau tidak. Apabila tweet tersebut masuk dalam kategori bencana, maka akan dilakukan pengenalan entitas seperti lokasi, kondisi, dan kebutuhan masyarakat. Algoritma yang digunakan adalah *Support Vector Machine* (SVM). Metode pembobotan yang digunakan pada penelitian ini untuk klasifikasi tweet ada dua yaitu *tf* dan *tf-idf*. Hasil yang diperoleh berupa *tf-idf* lebih unggul dari *tf* dengan rata-rata akurasi 78%.

Menurut Ashktorab et al. [4] dalam penelitiannya membuat sebuah tool yang bernama *Twitter for Disaster Response* (Tweedr). Tool ini digunakan untuk memberikan informasi yang relevan selama terjadi bencana kepada petugas penanggulangan bencana. Data bencana yang digunakan bersumber dari Twitter. Tujuan dari penelitian ini adalah untuk melakukan ekstraksi informasi terkait bencana. Penelitian ini terdiri dari tiga bagian utama yaitu klasifikasi, ekstraksi dan klustering. Tahap klasifikasi digunakan untuk mengidentifikasi tweet yang melaporkan kerusakan atau korban menggunakan metode *Knearestneighbors*, *decisiontrees*, *naivebayes* dan *logisticregression*. Selanjutnya pada tahap klustering dilakukan filter untuk menggabungkan tweet yang mirip. Pada tahap terakhir yaitu ekstraksi menggunakan *conditional random fields* (CRF), bagian ini mengekstrak token dan frasa yang memberikan laporan informasi spesifik tentang berbagai kerusakan infrastruktur, tipe kerusakan, dan korban jiwa.

Penelitian Dela Cruz et al. [20] mengusulkan sebuah model NER guna membuat model NER yang dapat digunakan untuk mengenali entitas terkait bencana. Data dalam penelitian ini bersumber dari situs berita Pilipino Star NGAYON. Data dibagi menjadi data training dan data testing. Data *training* berjumlah 171 artikel berita tahun 2014 dan data testing berjumlah 74 artikel berita tahun 2015. Entitas yang dikenali adalah jenis bencana, waktu terjadi bencana dan lokasi bencana. Penelitian ini menggunakan deep learning dengan pendekatan *long short-term memory* (LSTM) dan *conditional random fields* (CRF). Nilai akurasi yang diperoleh dari model ini sebesar 99.55 dan *F-measure* sebesar 66.05.

Berdasarkan beberapa penelitian di atas, belum ada penelitian yang secara spesifik melakukan identifikasi entitas bencana berbahasa Indonesia pada Twitter dan belum ada yang melakukan analisa terhadap dua metode yaitu *Machine Learning* dan *Deep Learning*. Oleh karena itu, penelitian ini bertujuan untuk melakukan identifikasi entitas bencana berbahasa Indonesia pada Twitter dan membandingkan akurasi dari kedua metode tersebut. Penelitian ini diharapkan dapat memberikan kontribusi terhadap NER pada domain kebencanaan dalam bahasa Indonesia, karena NER pada domain tersebut masih sangat terbatas.

### A. Machine Learning

Machine learning adalah salah satu cabang dari Artificial Intelligence (AI) yang mengadopsi prinsip dari ilmu komputer dan statistik untuk membuat model yang merefleksikan pola-pola data [8]. Model tersebut dilatih dengan menggunakan berbagai algoritma dalam pendekatan Machine Learning sehingga bisa melakukan klasifikasi terhadap entitas bencana. Algoritma machine learning yang digunakan adalah sebagai berikut :

### B. Naïve Bayes

Naïve Bayes merupakan algoritma yang ditemukan oleh seorang ilmuwan asal Inggris yang bernama Thomas Bayes. Algoritma ini digunakan untuk melakukan klasifikasi dengan menggunakan metode probabilitas dan statistik [9]. *Naïve Bayes* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes [5]. *Naïve Bayes* memiliki asumsi bahwa ada atau tidaknya fitur tertentu dari sebuah kelas tidak ada kaitannya dengan fitur yang ada pada kelas lainnya. Persamaan Teorema Bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X = Bukti.

H = Hipotesis.

$P(H|X)$  = Probabilitas bahwa hipotesis H benar untuk bukti X atau probabilitas posterior H dengan syarat X.

$P(X|H)$  = Probabilitas bahwa bukti X benar untuk hipotesis H atau probabilitas posterior X dengan syarat H.

$P(H)$  = Probabilitas prior hipotesis H.

$P(X)$  = Probabilitas prior bukti X.

### C. Decision Tree

Decision Tree merupakan metode klasifikasi diagram alir yang memiliki bentuk seperti struktur pohon dimana setiap *internal node* dapat menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan *output* dari pengujian tersebut dan *leaf node* menyatakan distribusi kelas [10]. *Decision Tree* terdiri dari tiga bagian sebagai berikut :

1. *Root Node* merupakan *node* yang berada paling atas dari sebuah pohon.
2. *Internal Node* adalah *node* percabangan, *node* ini memiliki satu input dan minimal dua *output*.
3. *Leaf Node* adalah *node* akhir yang memiliki satu *input* dan tidak memiliki *output*. Pada *decision tree*, setiap *leaf node* digunakan untuk menandai label kelas.

Algoritma *Decision Tree* yang akan digunakan dalam penelitian ini adalah Algoritma C4.5. Algoritma ini merupakan pengembangan dari algoritma ID3. Menurut [11] algoritma C4.5 merupakan serangkaian perbaikan atau pengembangan dari algoritma ID3 sehingga bisa menghasilkan sebuah sistem yang dapat berpengaruh untuk *decision tree*. Perbaikan yang dilakukan pada algoritma C4.5 terdiri dari beberapa metode yang digunakan menangani *numeric attributes*, *missing values*, *noisy data*, dan aturan yang menghasilkan *rules* dari *trees*.

### D. Support Vector Machine

Support Vector Machines (SVM) adalah kelas populer dari algoritma *supervised* dari machine learning. SVM berupaya menemukan *hyperplane* pemisah antara dua kelas data berlabel (Stamp, 2017). SVM biasanya digunakan untuk klasifikasi (*Support Vector Classification*) dan regresi (*Support Vector Regression*). Namun, sebagian besar digunakan dalam masalah klasifikasi. Ide dasar metode SVM adalah mengubah fitur input menjadi ruang dimensi yang lebih tinggi untuk memisahkan data secara linear menjadi dua kelas dengan *hyperplane*

dengan memaksimalkan margin dan meminimalkan kesalahan [12].

### E. Random Forest

Metode umum *random forest* pertama kali diusulkan oleh Ho pada tahun 1995. *Random forest* adalah kelas algoritma *machine learning* yang digunakan untuk memecahkan masalah pengenalan pola [14]. Menurut [15] *Random forest* merupakan salah satu metode yang menggunakan metode *bagging*. Metode *bagging* digunakan untuk meningkatkan akurasi dari hasil prediksi pada algoritma klasifikasi. Metode ini yang nantinya akan membangkitkan sejumlah *tree* dari *datasample* dimana pembuatan satu *tree* pada saat *training* tidak bergantung pada *tree* sebelumnya kemudian keputusan diambil berdasarkan voting terbanyak.

Menurut [13] bahwa *random forest* memiliki dua konsep yaitu :

1. Membangun *ensemble*.

*Ensemble* merupakan metode untuk mencari solusi prediksi sehingga bisa memperoleh hasil yang terbaik.

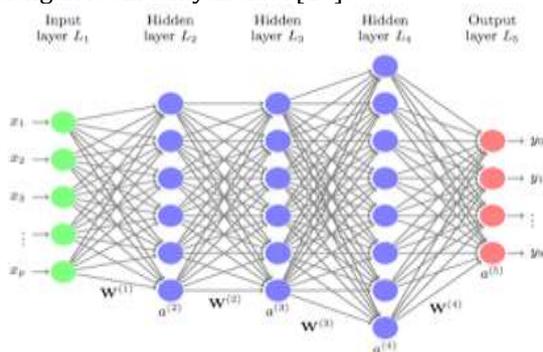
2. Penyeleksian fitur.

Penyeleksian fitur dilakukan secara acak pada setiap *tree* yang telah dibangun.

Hal yang pertama berarti tiap sampel yang diambil dari *data set* untuk *training tree* bisa dipakai lagi untuk *training tree* yang lain, sedangkan hal yang kedua berarti bahwa fitur yang digunakan pada saat *training* untuk tiap *tree* merupakan *subset* dari fitur yang dimiliki oleh *data set*.

### F. Deep Learning

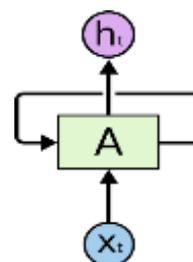
Deep Learning adalah bagian dari *NeuralNetwork* yang memiliki memiliki arsitektur lebih kompleks dan lebih banyak jumlah *layer* yang digunakan, sehingga diharapkan mampu menangani permasalahan yang lebih rumit dengan lebih banyak data [16].



Gambar 1. Deep Learning

### G. Recurrent Neural Network

Recurrent *NeuralNetwork* merupakan bagian dari *NeuralNetwork* yang dapat memproses data yang bersambung (*sequential data*). RNN menerima *data input* dalam bentuk *sequence*, dan pada prosesnya tidak mengharuskan memiliki ukuran yang sama antara *input* dan *output* yang ditangani [16]. RNN dapat menyimpan informasi dari masa lalu dengan cara melakukan *looping* di dalam arsitekturnya. Potongan model RNN dapat dilihat pada gambar berikut.



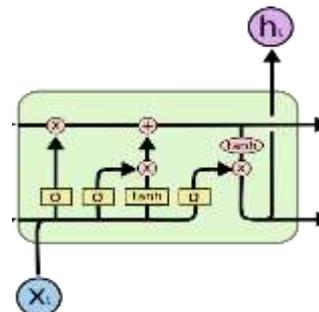
Gambar 2. RNN

*Recurrent neural network* pada bagian *input* digambarkan dengan simbol  $x_t$ , bagian *output* dengan simbol  $h_t$ , dan kotak  $A$  merupakan *hidden layer*.

### H. Long Short-Term Memory

Long Short-Term Memory (LSTM) mulai diperkenalkan pada tahun 1997 oleh Hochreiter and Schmidhuber. LSTM merupakan pengembangan dari *Recurrent Neural Network* (RNN). Alasan dari munculnya model LSTM adalah karena adanya keterbatasan model RNN dalam memproses data yang relatif panjang (*long term dependency*) [16]. Ketika RNN memproses *input sequence* yang panjang, maka semakin banyak *hidden layer* yang terbentuk. *Hidden layer* yang terbentuk dari *input sequence* yang panjang akan menjadi masalah pada RNN, sehingga dapat menyulitkan proses *training*.

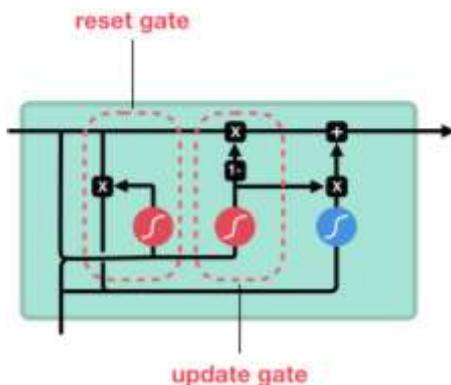
LSTM memiliki 4 komponen mengolah informasi yang masuk. Komponen tersebut adalah *input gate*, *forget gate*, *output gate* dan *cell states*.



Gambar 3. LSTM

### I. Gated Recurrent Units

*Gated Recurrent Unit* hampir mirip dengan LSTM, namun pada GRU hanya memiliki 2 *gate* yaitu *reset gate* dan *update gate*. *Reset gates* sama seperti *forget gate* dan *input gate* pada LSTM, yang akan memilih informasi mana yang harus disimpan atau dibuang. Sedangkan *update gate* digunakan untuk menentukan seberapa banyak data yang perlu dibuang.

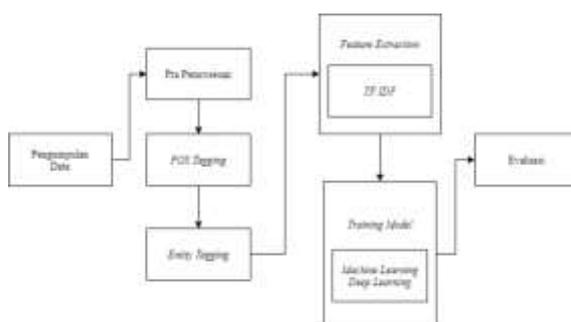


Gambar 4. GRU

### J. Convolutional Neural Network

*Convolutional Neural Networks* memiliki kemampuan baik dalam menyelesaikan masalah visi komputer karena dapat beroperasi secara konvolusional, yaitu melakukan ekstraksi fitur dari *patch* masukan lokal yang memungkinkan modularitas representasi dan efisiensi data [18]. CNN merupakan sebuah konstruk matematika yang disusun oleh 3 tipe *layer* yaitu *Convolution*, *Pooling*, dan *Fully Connected*. *Convolution* dan *Pooling layer* biasa digunakan untuk *feature extraction*. Sedangkan *Fully Connected layer* menempatkan hasil *feature extraction* menjadi hasil akhir keluaran.

## 3. METODOLOGI PENELITIAN



Gambar 5. Tahapan Penelitian

### 3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini bersumber dari Twitter. Data tersebut merupakan data mentah dari tweet-tweet yang berisi informasi tentang bencana di Indonesia. Jumlah data yang dikumpulkan sebanyak 378.602 tweet. Data tersebut diambil dari akun Twitter BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) dan BNPB (Badan Nasional Penanggulangan Bencana) baik yang berada di pemerintah pusat sampai dengan pemerintah daerah.

Data tweet diambil dengan menggunakan library *tweepy* dalam bahasa *python*. Proses pengumpulan data menggunakan *tweepy* dimulai dari melakukan proses autentikasi ke Twitter dengan menggunakan method atau function yang telah disediakan oleh *tweepy*. Proses autentikasi membutuhkan beberapa data diantaranya *consumer key*, *consumer secret*, *access token*, dan *access token secret*. Data untuk proses autentikasi diperoleh dari layanan *Twitter Developer*. Layanan tersebut disediakan oleh Twitter agar pengguna bisa menggunakan fitur-fitur yang ada pada Twitter.

Setelah melalui proses autentikasi, selanjutnya akan dilakukan proses pengambilan data tweet dengan memanfaatkan method atau function pada *library tweepy*. Kemudian data tweet tersebut dikumpulkan menjadi satu kedalam satu dokumen.

Pada penelitian ini, data tweet akan dibagi menjadi dua bagian yaitu data training dan data testing. *Data training* nantinya akan digunakan untuk melatih dengan menggunakan beberapa algoritma klasifikasi dalam *machine learning* maupun *deep learning* sehingga menghasilkan sebuah model. *Data testing* akan digunakan untuk mengetes performa dari model yang telah dibuat.

### 3.2 Pra Pemrosesan

Preprocessing merupakan tahap untuk mempersiapkan tweet yang telah dikumpulkan menjadi dokumen text atau clean data yang siap untuk digunakan pada tahap selanjutnya. Pada tahap preprocessing ini akan dilakukan beberapa proses di antaranya :

1. Menghapus URL.  
Tahap digunakan untuk menghapus URL yang ada pada Tweet.
2. Menghapus *punctuation* atau tanda baca.  
Proses ini dilakukan untuk menghapus tanda baca pada Tweet seperti *hashtag*, *emoticon* dan lain-lain.

### 3. Tokenizing.

*Tokenizing* adalah proses pemotongan Tweet atau text menjadi kata-kata atau yang biasa disebut *token*.

### 4. Stopword Removal.

Proses ini digunakan untuk menghilangkan kata-kata yang tidak penting dalam Tweet

### 3.3 POS Tagging

Untuk mempermudah dalam melakukan *entity labeling*, maka perlu melakukan proses *POS tagging* terlebih dahulu. *POS tagging* digunakan untuk memberikan kelas kata (tag) secara gramatikal ke setiap kata dalam suatu kalimat teks [36]. Kelas kata ini nantinya akan digunakan untuk memudahkan pelabelan entitas.

Proses *POS tagging* yang dilakukan pada penelitian ini menggunakan *tools* yang dikembangkan oleh Wibisono pada tahun 2018. Sedangkan data yang digunakan dalam pembuatan POS tagger tersebut bersumber dari penelitian [37]. *POS Tagging* ini dapat mengkategorikan kata ke dalam beberapa jenis kelas kata yang dapat dilihat pada Tabel 1.

TABEL 1. Kelas Kata

No	POS	Nama
1	OP	Open Parenthesis
2	CP	Close Parenthesis
3	GM	Slash
4	IN	Preposition
5	MD	Modal
6	PRL	Locative Pronouns
7	PRN	Number Pronouns
8	CD	Cardinal Number
9	PRP	Personal Pronouns
10	JJ	Adjective
11	FW	Foreign Words
12	CC	Coor-Conjunction
13	SC	Subor-Conjunction
14	DT	Determiner
15	RB	Adverb
16	NN	Common Noun
17	NNP	Proper Noun
18	NNG	Genitive Noun
19	VB	Verb
20	Z	Z

### 3.4 Entity Tagging

Pada tahap *entity tagging* dilakukan pelabelan entitas pada setiap kata dalam kalimat. Adapun entitas yang akan digunakan dalam penelitian ini adalah "DIS" untuk menandakan jenis bencana, "TIM" untuk waktu bencana, "LOC" sebagai nama tempat atau lokasi terjadinya bencana, "MAG" untuk ukuran kekuatan jika bencana yang terjadi adalah gempa dan "O" untuk kata yang tidak memiliki entitas.

TABEL 2. Entitas Bencana

No	Entitas	Keterangan
1	DIS	Jenis Bencana
2	TIM	Waktu Kejadian
3	LOC	Lokasi Bencana
4	MAG	Kekuatan (jika jenis bencana Gempa)
5	O	Others atau kata yang tidak memiliki entitas

Pada tahap *entity tagging* terdapat dua proses yang dilakukan. Proses pertama, *chunking*. Proses ini melakukan *entity tagging* dengan cara memanfaatkan *part of speech*. Teknik memanfaatkan *part of speech* ini dilakukan dengan membaca pola pada *token* yang sudah dilabeli dengan *part of speech*. Proses *chunking* memanfaatkan method "RegexpParser" yang ada pada *Natural Language Toolkit*.

### 3.5 Feature Extraction

Menurut [23] *feature extraction* atau ekstraksi fitur adalah sebuah proses dimana properti-properti diekstrak dari suatu data. *Feature extraction* merupakan salah satu proses penting yang harus dilakukan sebelum tahap training, karena algoritma *machine learning* yang akan digunakan untuk *training* hanya bisa menerima masukan atau input berupa angka.

*Feature extraction* yang digunakan pada penelitian ini adalah *Term Frequency - Inverse Document Frequency* (TF-IDF). *Tf-idf* merupakan metode untuk menghitung seberapa penting sebuah kata yang berada di dalam dokumen atau korpus. *Tf-idf* terdiri dari *term-frequency* dan *inverse document frequency*. *Term Frequency* adalah frekuensi kemunculan *term* atau kata di dalam dokumen.

Metode TF-IDF dihitung dengan persamaan (1).  $tf_{t,d}$  merupakan kemunculan *term* pada dokumen *d*. Sedangkan  $idf_t$  merupakan nilai yang diperoleh dari persamaan (2). *D* merupakan total keseluruhan dokumen, sedangkan  $df_t$  merupakan banyaknya dokumen yang mengandung *term t*.

$$tf - idf_{t,d} = tf_{t,d} \cdot idf_t \quad (1)$$

$$idf_t = \log \frac{D}{df_t} \quad (2)$$

### 3.6 Training

Sebelum dilakukan proses *training*, data akan dibagi menjadi dua bagian yaitu *training set* dan *test set*. Bagian *training set* adalah data latih yang akan digunakan pada saat *training model*, sedangkan *test set* digunakan untuk menguji

performa dari model yang telah ditraining. Pembagian antara *training set* dan *test set* dengan rasio pembagian 70%-30% dan 80%-20%. Studi empiris menunjukkan bahwa model terbaik diperoleh jika menggunakan 20-30% data untuk *testing* dan 70-80% data untuk *training* [22]. Jumlah tweet yang digunakan pada penelitian ini sebanyak 378.602 tweet. Dari tweet tersebut selanjutnya ditentukan menjadi beberapa sample size. Jumlah sample size dapat dilihat pada Tabel 3.

TABEL 3. Sampel Size

Jumlah Tweet	Sample Size
378.602	189.301 (50%)
	227.161 (60%)
	265.022 (70%)
	302.882 (80%)
	340.742 (90%)

Tahap *training* ini akan menggunakan dua pendekatan yaitu *machine learning* dan *deep learning*. Algoritma yang digunakan pada pendekatan *machine learning* adalah *random forest*, *decision tree*, *naïve bayes* dan *support vector machines*. Sedangkan pada pendekatan *deep learning* menggunakan algoritma *long short-term memory* (LSTM), *convolutional neural network* (CNN) dan *gated recurrent units* (GRU).

Setiap algoritma pada *machine learning* dan *deep learning* akan menggunakan semua *sample size* dari 50% sampai dengan 90%. Berikut adalah skenario dari setiap algoritma *machine learning* dan *deep learning*.

TABEL 4. Skenario Algoritma Machine Learning

Algoritma	Parameter
Naïve Bayes	alpha : 1
Support Vector Machine	penalty : L2
Decision Tree	splitting : criterion gini
Random Forest	-splitting : criterion gini -n estimators : 100

TABEL 5. Skenario Algoritma Deep Learning

Algoritma	Parameter
Long Short-Term Memory	-unit : 50 -dropout : 0.5 -batch size : 32
Convolutional Neural Network	-filter : 50 -kernel size : 1 -batch size : 32
Gated Recurrent Units	-unit : 50 -dropout : 0.5 -batch size : 32

### 3.7 Evaluasi

Evaluasi terhadap model yang telah dibangun dari proses training merupakan hal yang sangat penting. Tujuan evaluasi adalah untuk mengukur apakah model mampu melakukan klasifikasi dengan baik atau tidak. Hasil dari proses evaluasi ini digunakan untuk mengetahui kinerja dari model. Sehingga memperoleh mana model terbaik yang dihasilkan dari setiap algoritma yang telah disebutkan di atas. Metode evaluasi yang digunakan pada penelitian ini adalah perbandingan akurasi. Akurasi sebuah algoritma mengindikasikan persentase dari prediksi yang benar. Formula dari akurasi adalah

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Prediksi} \quad (3)$$

atau lebih detailnya,

$$Akurasi = \frac{PB + NB}{PB + NB + PP + NP} \quad (4)$$

di mana PB = Positif Benar, NB = Negatif Benar, PP = Positif Palsu, dan NP = Negatif Palsu.

Keterangan :

PB: Jumlah data positif yang terklasifikasi dengan benar.

NB : Jumlah data negatif yang terklasifikasi dengan benar.

PP : Jumlah data negatif namun terklasifikasi sebagai data positif.

NP : Jumlah data positif namun terklasifikasi sebagai data negatif.

## 4. HASIL DAN PEMBAHASAN

Hasil penelitian *named entity recognition* ini berfokus pada analisis dari variasi ukuran sampel *data set* terhadap kinerja algoritma *machine learning* dan *deep learning*. Analisis performa kinerja dari algoritma diperoleh berdasarkan akurasi.

### 4.1 Hasil Akurasi

Hasil akurasi dibagi menjadi dua bagian, yaitu akurasi pada rasio 70%-30% dan akurasi pada rasio 80%-20%. Setiap algoritma pada kedua rasio menggunakan *sample size* yang berbeda-beda. *Sample size* yang digunakan mulai dari 50%, 60%, 70%, 80%, dan 90% dari data tweet yang berjumlah 378.602.

1. Rasio 70%-30%

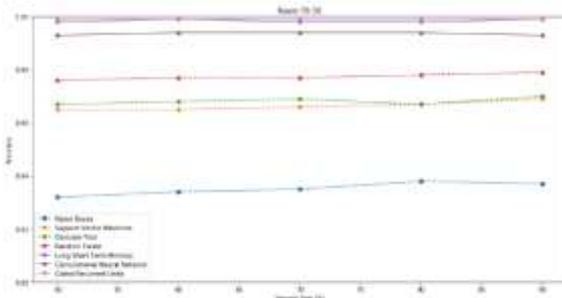
TABEL 6. Akurasi Rasio 70%-30%

Algoritma	Sample Size				
	50	60	70	80	90
NBayes	0.932	0.934	0.935	0.938	0.937
SVM	0.965	0.965	0.966	0.967	0.969
D Tree	0.967	0.968	0.969	0.967	0.97

R Forest	0.976	0.977	0.977	0.978	0.979
LSTM	0.998	<b>0.999</b>	0.998	0.998	<b>0.999</b>
CNN	0.993	0.994	0.994	0.994	0.993
GRU	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>

Berdasarkan dari hasil akurasi yang ditampilkan pada Tabel 6 di atas dapat disimpulkan bahwa pada *sample size* 50% akurasi terbaik dihasilkan oleh algoritma GRU sebesar 0.999, pada *sample size* 60% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU sebesar 0.999, pada *sample size* 70% dan 80% akurasi terbaik dihasilkan oleh algoritma GRU, pada *sample size* 90% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU. Setiap algoritma pada setiap *sample size* memiliki selisih yang berbeda-beda, seperti pada *sample size* 50% rentang perbedaan akurasinya ada diantara 0.001- 0.067, pada *sample size* 60% selisihnya antara 0.005-0.065, pada *sample size* 70% antara 0.001-0.064, pada *sample size* 80% antara 0.001-0.061, pada *sample size* 90% berada diantara 0.006-0.062.

Selain dilihat berdasarkan *sample size*, juga dapat disimpulkan berdasarkan algoritma, seperti pada algoritma *naïve bayes* mendapatkan akurasi terbaik dengan nilai 0.938 pada *sample size* 80%, dan akurasi terburuk dengan nilai 0.932 pada *sample size* 50%. Algoritma SVM memperoleh akurasi terbaik pada *sample size* 90% dengan nilai 0.969. Algoritma *decision tree* memperoleh hasil akurasi terbaik sebesar 0.97 pada *sample size* 90%. Algoritma *random forest* memperoleh hasil akurasi terbaik pada *sample size* 90% dengan nilai 0.979. Algoritma LSTM memperoleh akurasi terbaik dengan nilai 0.999 pada *sample size* 60%, dan 90%. Algoritma CNN memperoleh akurasi terbaik dengan nilai 0.994 pada *sample size* 60%, 70% dan 80%. Algoritma GRU memperoleh akurasi terbaik dengan nilai 0.999 pada semua *sample size*.



Gambar 7. Grafik Akurasi Rasio 70%-30%

Keterangan Gambar 7 :

- Garis lurus menunjukkan algoritma machine learning.
- Garis putus-putus menunjukkan algoritma deep learning.

Pada Gambar 7 dapat dilihat secara keseluruhan bahwa pendekatan *deep learning* menunjukkan nilai akurasi lebih tinggi dibandingkan dengan pendekatan *machine learning*. Pada pendekatan *machine learning*, nilai akurasi tertinggi dihasilkan oleh algoritma random forest dan terburuk dihasilkan oleh algoritma *naïve bayes*. Sedangkan pada pendekatan *deep learning*, nilai akurasi tertinggi dihasilkan oleh algoritma GRU dan nilai akurasi terburuk dihasilkan oleh algoritma CNN.

## 2. Rasio 80%-20%

TABEL 7. Akurasi Rasio 80%-20%

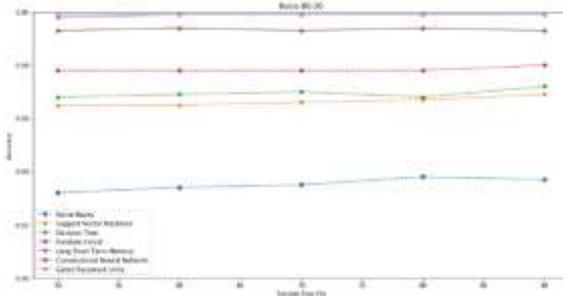
Algoritma	Sample Size				
	50	60	70	80	90
N Bayes	0.932	0.934	0.935	0.938	0.937
SVM	0.965	0.965	0.966	0.967	0.969
D Tree	0.968	0.969	0.97	0.968	0.972
R Forest	0.978	0.978	0.978	0.978	0.98
LSTM	0.998	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>
CNN	0.993	0.994	0.993	0.994	0.993
GRU	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>

Berdasarkan dari hasil akurasi yang ditampilkan pada Tabel 7 di atas dapat disimpulkan bahwa pada *sample size* 50% akurasi terbaik dihasilkan oleh algoritma GRU sebesar 0.999, pada *sample size* 60% sampai dengan 90% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU dengan nilai 0.999.

Setiap *sample size* memiliki selisih yang berbeda-beda, seperti pada *sample size* 50%, rentang perbedaan akurasi setiap algoritmanya ada diantara 0.001-0.067, pada *sample size* 60% selisihnya antara 0.005-0.065, pada *sample size* 70% antara 0.006-0.064, pada *sample size* 80% antara 0.005-0.061, pada *sample size* 90% antara 0.005-0.062.

Selain dilihat berdasarkan *sample size*, juga dapat dilihat berdasarkan algoritma, seperti pada algoritma *naïve bayes* mendapatkan akurasi terbaik dengan nilai 0.938 pada *sample size* 80%, dan akurasi terburuk dengan nilai 0.932 pada *sample size* 50%. Algoritma SVM memperoleh akurasi terbaik pada *sample size* 90% dengan nilai 0.969. Algoritma *decision tree* memperoleh

hasil akurasi terbaik sebesar 0.972 pada *sample size* 90%. Algoritma *random forest* memperoleh hasil akurasi terbaik pada *sample size* 90% dengan nilai 0.98. Algoritma LSTM memperoleh akurasi terbaik dengan nilai 0.999 pada *sample size* 60%, sampai 90%. Algoritma CNN memperoleh akurasi terbaik dengan nilai 0.994 pada *sample size* 60% dan 80%. Algoritma GRU memperoleh akurasi terbaik dengan nilai 0.999 pada semua *sample size*.



Gambar 8. Grafik Akurasi Rasio 80%-20%

Keterangan Gambar 8 :

- Garis lurus merupakan algoritma machine learning.
- Garis putus-putus merupakan algoritma deep learning.

Pada Gambar 8 dapat dilihat secara keseluruhan bahwa pendekatan deep learning menunjukkan nilai akurasi lebih tinggi dibandingkan dengan pendekatan machine learning. Pada pendekatan machine learning, nilai akurasi tertinggi dihasilkan oleh algoritma random forest dan terburuk dihasilkan oleh algoritma naive bayes. Sedangkan pada pendekatan deep learning, nilai akurasi tertinggi dihasilkan oleh algoritma GRU dan nilai akurasi terburuk dihasilkan oleh algoritma CNN

#### 4.2 Perbandingan Hasil Pengujian

Hasil komputasi menggunakan algoritma *machine learning* dan *deep learning* pada kasus data bencana di Indonesia dengan data yang bersumber dari Twitter dengan rasio 70-30 dan 80-20 pada Tabel 6 dan Tabel 7 menghasilkan beberapa perbedaan akurasi, diantara :

1. Hasil akurasi menggunakan algoritma *decision tree* pada rasio 80-20 untuk *sample size* 50%, 60%, 70% dan 80% sedikit lebih baik daripada rasio 70-30, selisih antara keduanya adalah sebesar 0.001. Pada *sample size* 90% selisihnya sebesar 0.002.
2. Hasil akurasi algoritma *random forest* pada rasio 80-20 untuk *sample size* 50% sedikit lebih baik dari rasio 70-30, dengan perbedaan

sebesar 0.002. Sedangkan pada *sample size* 60%, 70%, dan 90% selisihnya sebesar 0.001.

3. Pada algoritma *long short-term memory*, akurasi yang lebih baik dihasilkan oleh rasio 80-20 untuk *sample size* 70% dan 80% dengan selisih sebesar 0.001.
4. Pada algoritma *convolutional neural network*, akurasi yang lebih baik dihasilkan oleh rasio 70-30 pada *sample size* 70% dengan selisih sebesar 0.001.

#### 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka, guna mendapatkan model terbaik, dilakukan proses *training* menggunakan algoritma *machine learning* dan *deep learning*. Algoritma *machine learning* yang digunakan pada penelitian adalah *naive bayes*, *Support vector machines*, *decision tree*, dan *random forest*. Sedangkan algoritma *deep learning* yang digunakan adalah *long short-term memory*, *convolutional neural network*, dan *gated recurrent units*. Berdasarkan eksperimen, metode *deep learning* menghasilkan akurasi yang lebih baik daripada metode *machine learning* dengan akurasi terbaik dihasilkan dari algoritma *deep learning* yaitu *gated recurrent units* dan *long short-term memory* sebesar 0.999. Adapun hasil akurasi terbaik pada metode *machine learning* dihasilkan oleh algoritma *random forest* dengan akurasi 0.98. Kemudian pada proses *training*, diketahui bahwa semakin besar ukuran *sample size*, maka akurasinya semakin tinggi. Namun selisih akurasi antara *sample size* terkecil hingga terbesar tidak terlalu jauh dan pada algoritma *deep learning* menunjukkan bahwa ukuran *sample size* tidak terlalu mempengaruhi akurasi. Akurasi yang dihasilkan oleh algoritma *deep learning* cenderung stabil, baik pada *sample size* terkecil maupun *sample size* terbesar.

#### Daftar Pustaka:

- [1] A. S. Cahyono, "Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia," J. ilmu Sos. dan ilmu Polit., pp. 140-157, 2016.
- [2] J. An, M. Cha, K. Gummadi, and J. Crowcroft, "Media landscape in Twitter: A world of new conventions and political diversity," Assoc. Adv. Artificial Intell., pp. 18-25, 2011.
- [3] H. Februariyanti, E. Zuliarso, D. Bulan, and L. Suryati, Rancang bangun sistem layanan informasi bencana melalui twitter menggunakan basis data xml, vol. 035, no. November. 2013.
- [4] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta, "Tweedr: Mining Twitter to Inform

- Disaster Response,” Proc. 11th Int. ISCRAM Conf., vol. 12, no. 4, pp. 354– 358, 2014.
- [5] A. F. Hidayatullah, “Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter,” TESIS, vol. 2014, no. semnasIF, p. A-1, 2014.
- [6] L. Derczynski et al., “Analysis of named entity recognition and linking for tweets,” Inf. Process.Manag., vol. 51, no. 2, pp. 32–49, 2015.
- [7] Y. Wibisono and M. L. Khodra, “Pengenalan Entitas Bernama Otomatis untuk Bahasa Indonesia dengan Pendekatan Pembelajaran Mesin,” pp. 1–5, 2018.
- [8] Jan Wira Gotama Putra, “Pengenalan Konsep Pembelajaran Mesin dan Deep Learning,” 2017.
- [9] Bustami, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi,” TECHSI J. Penelit. Tek.Inform., vol. 8, no. 1, 2014.
- [10] P. Kasih, “Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara,” vol. 2, pp. 63–69, 2019.
- [11] A. Andriani, “Sistem Prediksi Penyakit Diabetes Berbasis Decision Tree,” Bianglala Inform., vol. 1, no. 1, pp. 1–10, 2013.
- [12] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, “Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines,” Ore Geol. Rev., vol. 71, pp. 804–818, 2015.
- [13] V. Sazonau, “Implementation and Evaluation of a Random Forest Machine Learning Algorithm,” Univ. Manchester, p. 9, 2012.
- [14] E. Scornet, “Random forests and kernel methods,” IEEE Trans. Inf. Theory, vol. 62, no. 3, pp. 1485–1500, 2016.
- [15] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” R News, vol. 2, no. 3, pp. 18–22, 2002.
- [16] W. Widayat, “PENGARUH WORD EMBEDDING DIMENSION REDUCTION TERHADAP KINERJA LSTM UNTUK ANALISIS SENTIMEN,” 2018.
- [17] A. Ahmad, “Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning,” J. Teknol. Indones., no. October, p. 3, 2017.
- [18] R. A. N. Nayoan, “ANALISIS SENTIMEN BERBASIS FITUR PADA ULASAN TEMPAT WISATA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK(CNN),” 2019.
- [19] R. Dermawan, “Klasifikasi Tweet Dan Pengenalan Entitas Bernama Pada Tweet Bencana Dengan Support Vector Machine,” *입법학연구*, vol. 제 13 집 1호, pp. 31–48, 2016
- [20] B. M. Dela Cruz, C. Montalla, A. Manansala, R. Rodriguez, M. Octaviano, and B. S. Fabito, “Named-Entity Recognition for Disaster Related Filipino News Articles,” TENCON 2018 - 2018 IEEE Reg. 10 Conf., no. October, pp. 1633–1636, 2018
- [21] M. N. Saadah, R. W. Atmagi, D. S. Rahayu, and A. Z. Arifin, “Sistem Temu Kembali Dokumen Teks Dengan Pembobotan Tf-Idf Dan LCS,” JUTI J. Ilm. Teknol.Inf., vol. 11, no. 1, p. 19, 2013.
- [22] A. Gholamy, V. Kreinovich, and O. Kosheleva, “Why 70 / 30 or 80 / 20 Relation Between Training and Testing Sets : A Pedagogical Explanation,” pp. 1–6.
- [23] R. De Groot, “Data Mining for Tweet Sentiment Classification,” p. 63, 2012.
- [24] L. Mutawalli, M. T. A. Zaen, and W. Bagye, “Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto),” JIRE (Jurnal Inform. dan Rekayasa Elektron., vol. 2, no. 2, pp. 43–51, 2019.
- [25] Fadli, S., Ashari, M., & Imtihan, K. (2020). SISTEM PENJADWALAN EVENT ORGANIZER DENGAN METODE ROUND ROBIN (RR). *Jurnal Manajemen Informatika dan Sistem Informasi*, 3(2), 100-107.
- [26] Imtihan, K., & Fahmi, H. (2020). ANALISIS DAN PERANCANGAN SISTEM INFORMASI DAERAH RAWAN KECELAKAAN DENGAN MENGGUNAKAN GEOGRAPHIC INFORMATION SYSTEMS (GIS). *Jurnal Manajemen Informatika dan Sistem Informasi*, 3(1), 16-23.