



**PROPOSAL PENELITIAN**

**SEMINAR BIDANG KAJIAN**

**KLASIFIKASI JENIS PERMASALAHAN PADA LAYANAN  
*LIVE CHAT HELPDESK* MENGGUNAKAN *RECURENT*  
*NEURAL NETWORK* (RNN)**

**RHEZA ANDIKA**

**99217015**

**PROGRAM DOKTOR TEKNOLOGI INFORMASI**

**UNIVERSITAS GUNADARMA**

**SEPTEMBER**

**2022**

## DAFTAR ISI

### Bab 1. PENDAHULUAN

1.1. Latar Belakang Masalah

1.2. Batasan Masalah

1.3. Tujuan Penelitian

1.4. Kontribusi Penelitian

### Bab 2. TINJAUAN PUSTAKA

2.1 Media Sosial

2.2 *Text Mining*

2.3 *Pre-Processing*

2.4 *Deep Learning*

2.5 *Recurrent Neural Network (RNN)*

2.6 Klasifikasi

2.7 Penelitian Terkait

### Bab 3. METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

3.2 Rencana Kerja

## DAFTAR PUSTAKA

# 1. PENDAHULUAN

## 1.1. Latar Belakang

Pelanggan atau pemakai suatu produk adalah orang-orang yang berhubungan secara langsung maupun tidak langsung dengan perusahaan-perusahaan bisnis. Pada umumnya pelanggan selalu menginginkan kepuasan terhadap pemenuhan kebutuhannya mulai dari produk yang sesuai serta pelayanan yang baik. Namun memahami keinginan dan kebutuhan setiap pelanggan sangatlah sulit. Adakalanya menemukan pelanggan yang tidak benar-benar puas dengan produk atau layanan yang diberikan dan memiliki masalah-masalah tersendiri. Sangat penting untuk menetapkan proses yang memastikan hubungan baik dengan pelanggan bahkan setelah kesalahan yang terjadi dan mencegah mereka meninggalkan dan berpindah ke produk atau jasa yang lain. Diperlukan manajemen keluhan menggambarkan pengelolaan keluhan pelanggan di sebuah perusahaan. Dimana tujuannya adalah untuk memperkuat loyalitas pelanggan serta jaminan kualitas. Saat ini banyak aplikasi yang dapat menjangkau keluhan pelanggan diantaranya melalui media sosial, *live chat*, WhatsApp, Telegram dan aplikasi *chat* lainnya. Fitur *live chat* digunakan pada sebuah website baik itu website instansi pemerintahan, instansi pendidikan, perusahaan, *e-commerce*, maupun *Marketplace*. Fitur *live chat* dapat memfasilitasi komunikasi antara pelanggan dengan penjual. studi terdahulu telah menunjukkan bahwa penggunaan *live chat* dapat memberikan pengaruh yang positif terhadap keputusan pelanggan.

Klasifikasi teks merupakan salah satu bidang penelitian yang terkenal dalam Pemrosesan Bahasa Alami (NLP) dan terbukti telah menarik perhatian para peneliti dalam banyak aplikasi seperti pencarian *web*, pemfilteran informasi, kategorisasi topik, dan analisis sentimen [2]. Untuk melakukan proses klasifikasi diperlukan sebuah metode yang menunjang proses klasifikasi, salah satunya adalah metode *Recurrent Neural Network* (RNN). Dalam penelitian ini akan dikembangkan sebuah algoritma untuk melakukan klasifikasi permasalahan dari percakapan pada fitur *live chat* kedalam kategori yang diinginkan. Pada penelitian ini terdapat 2 langkah yang akan diproses, langkah pertama klasifikasi untuk menentukan

permasalahan dalam percakapan tersebut mengarah kepada kategori yang mana, kemudian langkah selanjutnya isi percakapan diproses untuk dilakukan pengklasifikasian lain untuk diarahkan kebidang terkait sesuai dengan permasalahannya. Penelitian ini menggunakan metode *Recurrent Neural Network* (RNN) dan diharapkan mesin mampu mengklasifikasikan *toxic comment* pada media sosial Instagram berdasarkan jenis toksisitasnya.

### **1.2. Rumusan Masalah**

1. Bagaimana mengembangkan algoritma ekstraksi permasalahan dalam percakapan pada fitur *live chat*?
2. Bagaimana mengembangkan algoritma klasifikasi permasalahan dalam percakapan pada fitur *live chat*?

### **1.3. Batasan masalah**

Penelitian ini berusaha mengembangkan teknik dan mengatasi kendala yang telah ada dengan difokuskan pada :

1. Bagaimana membuat mekanisme yang dapat membantu mengidentifikasi berdasarkan histori permasalahan dalam percakapan pada fitur *live chat*
2. Mengidentifikasi jenis permasalahan dalam percakapan pada fitur *live chat*.
3. Melakukan tahapan *pre-processing* yang efisien pada data yang akan di analisis.
4. Membuat model RNN untuk melakukan klasifikasi dengan melatih model dengan data pelatihan dan menguji model untuk mengevaluasi tingkat akurasi.

### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini adalah menghasilkan sebuah metode algoritma ekstraksi permasalahan dalam percakapan pada fitur *live chat* dan menghasilkan metode algoritma klasifikasi permasalahan dalam percakapan pada fitur *live chat*.

### **1.5. Manfaat Dan Kontribusi Penelitian**

Dalam bidang teknologi penelitian ini akan menemukan cara baru, menambah, dan mengembangkan dari metode yang sudah ada untuk proses klasifikasi permasalahan dalam percakapan pada fitur *live chat*, nantinya percakapan tersebut harus melewati tahapan *pre-processing* untuk nantinya akan menjadi Dataset. Dalam bidang keilmuan penelitian ini akan memberikan kontribusi berupa pemahaman secara umum mengenai proses dan pengembangan pembentukan klasifikasi dari sebuah komunikasi percakapan.

## 2. TINJAUAN PUSTAKA

### 2.1 *Live Chat*

Menurut situs taptalk.io, Live chat memungkinkan bisnis untuk berinteraksi dengan pengguna ketika saat mengunjungi situs atau media sosial dengan mengirimkan pesan. Penyedia dapat menggunakan otomatisasi sistem untuk memulai percakapan dengan pengguna yang baru pertama kali mengunjungi website atau berinteraksi dengan customer lainnya. Live chat menawarkan pengalaman customer yang dipersonalisasi. Hal ini memungkinkan interaksi brand yang lebih baik serta mengenal pelanggan dengan lebih dekat. Tools ini dapat digunakan oleh tim sales, *marketing*, dan *customer support* untuk menjawab pertanyaan dari *customer*. Di sisi konsumen, live chat digunakan oleh pengunjung untuk mencari bantuan selama mengunjungi website.

### 2.2 *Text Mining*

*Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction*, dan *information retrieval*. *Text mining* sebenarnya tidak jauh berbeda dengan data *mining*, yang membedakan hanyalah sumber data yang digunakan. Pada data *mining* data yang digunakan adalah data yang terstruktur, sedangkan pada *text mining* data yang digunakan adalah data yang tidak terstruktur berupa *teks*. Tujuan dari *text mining* secara keseluruhan adalah pada dasarnya untuk mengubah suatu teks menjadi data yang dapat dianalisis.

### 2.3 *Text Pre-processing*

Tahap *text pre-processing* merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *text pre-processing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data

menjadi bentuk yang lebih mudah diproses oleh system [1]. Tahap *Text Pre-processing* adalah tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen. Proses *pre-processing* ini meliputi : *case folding*, *tokenizing*, *filtering*, dan *stemming*. Dimana penjelasan dari tahap-tahap tersebut adalah sebagai berikut:

a. *Case folding*

*Case folding* merupakan salah satu bentuk *text preprocessing* yang paling sederhana dan efektif meskipun sering diabaikan. Tujuan dari *case folding* untuk mengubah semua huruf dalam dokumen menjadi huruf kecil [8].

b. *Tokenizing*

*Tokenizing* merupakan tahap pemotongan teks input menjadi kata, istilah, *symbol*, tanda baca, atau elemen lain yang memiliki arti yang disebut token. Pada proses, token yang merupakan tanda baca yang dianggap tidak perlu seperti titik (.), koma (,), tanda seru (!), dan lain-lain akan dihapus [8].

c. *Filtering*

*Filtering* merupakan proses yang dilakukan pada tahap ini yaitu menghapus *stop-word*. *Stop-word* adalah kata yang bukan merupakan kata unik dalam suatu artikel atau kata-kata umum yang biasanya selalu ada dalam suatu artikel [8].

d. *Stemming*

*Stemming* digunakan untuk mendapatkan kata dasar dari suatu kata. Hal ini dilakukan untuk menormalisasi kata [8].

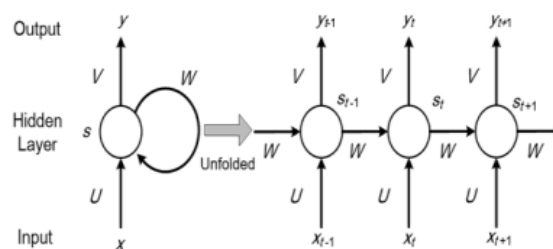
## 2.4 *Deep Learning*

*Deep learning* merupakan salah satu cabang dari *machine learning*. Model *Deep learning* dapat mempelajari komputasinya sendiri dengan menggunakan otaknya sendiri. *Deep learning* dirancang untuk terus menganalisa data seperti otak manusia dalam mengambil keputusan. *Deep learning* memungkinkan sebuah mesin dapat mengklasifikasikan sebuah objek yang dapat berupa gambar, teks, suara, dll. Dalam *deep learning* terdapat sebuah metode yang dikenal dengan nama *Recurrent*

*Neural Network* (RNN), RNN adalah jenis arsitektur jaringan saraf tiruan yang pemrosesannya berulang kali dipanggil untuk memproses *input* data sekuensial dan data diproses melalui banyak lapisan [7]. Pemodelan RNN dapat menyelesaikan berbagai tugas kategorisasi kalimat dan dapat melakukan klasifikasi.

## 2.5 Recurrent Neural Network (RNN)

*Recurrent Neural Network* berfungsi untuk melakukan proses input data secara sekuensial. Dalam tiap pemrosesan, output yang dihasilkan tidak hanya merupakan fungsi dari sampel itu saja, tetapi juga berdasarkan state internal yang merupakan hasil dari pemrosesan sampel-sampel sebelumnya (atau setelahnya, pada bidirectional RNN) [3]. Pemodelan RNN dapat menyelesaikan berbagai tugas kategorisasi kalimat dan dapat melakukan klasifikasi karena kemampuan dalam memprosesnya dipanggil berulang-ulang dengan hasil dapat menangani input dan output variable yang panjangnya bervariasi. Pada intinya RNN adalah jaringan syaraf tiruan yang menggunakan rekurensi dengan memanfaatkan data masa lalu. Karena itu, beberapa studi terbaru mengenai RNN cukup kuat untuk permasalahan klasifikasi. RNN memiliki arsitektur yang dapat digunakan untuk data berbentuk sekuensial seperti yang dapat dilihat Gambar 1 [7].



Gambar 2.1 Recurrent Neural Network

Cara yang dilakukan RNN untuk dapat menyimpan informasi dari masa lalu adalah dengan melakukan looping di dalam arsitekturnya, yang secara otomatis membuat informasi dari masa lalu tetap tersimpan. Pada jaringan RNN



menggunakan fungsi aktivasi sigmoid untuk hidden layer. Fungsi sigmoid memiliki output dengan rentang 0 sampai 1. Fungsi sigmoid dapat dilihat pada Persamaan 1. Dengan turunan fungsi terlihat pada Persamaan 2 [7].

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$f'(x) = f(x)(1 - f(x)) \quad (2)$$

Terdapat inti perulangan pada RNN yang mengambil nilai input  $x$  kemudian dimasukkan ke dalam RNN yang berisi nilai dari hidden layer yang akan diperbarui setiap kali RNN membaca input baru sehingga menghasilkan output pada setiap waktu. Pada komputasi menggunakan RNN terdapat hubungan perulangan dengan fungsi aktivasi yang disimbolkan oleh  $f$  sehingga fungsi tersebut akan bergantung pada bobot  $w$ . Bobot  $w$  akan menerima nilai state baru sebelumnya dari hidden layer dikurangi 1 yang menjadi masukan pada saat keadaan  $x_t$  dan disimpan ke dalam hidden layer berikutnya ( $h_t$ ) atau dengan kata lain pada saat hidden layer diperbarui. Proses ini dilakukan menggunakan Persamaan 3 [7].

$$h_t = f_w(h_{t-1}, x_t) \quad (3)$$

Nilai  $x$  dimasukan ke dalam fungsi aktivasi  $f$  dan bobot  $w$  yang sama pada setiap kali perhitungan. Secara sederhana terdapat matriks bobot  $W_{xh}$  yang dikalikan dengan input  $x_t$  serta matriks bobot lain  $W_{hh}$  yang dikalikan terhadap nilai dari hidden layer sebelumnya atau  $h_{t-1}$ . Kedua matriks tersebut ditambahkan. Jika terdapat data non-linear maka hasil penjumlahan kedua matriks dikalikan dengan  $\tanh$ , seperti pada Persamaan 4 [7].

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \quad (4)$$

Pada arsitektur RNN jika ingin menghasilkan beberapa  $y_t$  di setiap waktu, dikarenakan terdapat matriks bobot lain  $W$  dari hidden layer  $W_h$  sehingga mengubah beberapa nilai  $y$  yang terlihat pada Persamaan 5 [7].

$$y_t = W_{hy}h_t \quad (5)$$

Proses pelatihan pada RNN menggunakan algoritma Backpropagation dengan beberapa putaran. Karena parameter yang dibagikan secara merata pada setiap time step, maka gradient untuk setiap output tergantung tidak hanya pada kalkulasi dari time step saat ini, tetapi juga pada time step sebelumnya. Pada arsitektur RNN terdapat beberapa unit seperti Gate Recurrent Unit (GRU), Backpropagation Through Time (BPTT) dan Long Short-Term Memory (LSTM). Hasil pada setiap data yang telah dilakukan oleh praproses, maka tahapan pemodelan RNN dapat dilakukan. Semua data kalimat yang telah dilakukan proses embedding akan dijadikan sebagai input terhadap neuron. Jumlah neuron dihasilkan dari ukuran window size sebanyak 100 dan jumlah kata unik sebanyak 2696, sehingga menghasilkan jumlah neuron sebanyak 495.204 [7].

## **2.6 Klasifikasi**

Teknik klasifikasi bisa disimpulkan sebagai cara memprediksi suatu data baru sehingga bisa ditentukan pada kategori apakah ia berada, berdasarkan pada data latih, dimana tiap anggota data latih tersebut telah diketahui kategorinya. Kategori ini tentunya bersifat diskrit, dimana urutan tidak mempengaruhi [5], Contohnya seperti: positif, negatif, dan netral; baik dan buruk; dll.

Proses Klasifikasi Dalam teknik klasifikasi ada dua proses utama yaitu proses pembangunan model dan penerapan model [5]. Proses pembangunan model melibatkan tahapan sebagai berikut:

1. Menentukan kategori/kelas/label terlebih dahulu. Misal: positif, negatif, dan netral.
2. Dari sekumpulan data yang diperoleh, tentukan kategori untuk tiap datanya.
3. Sekumpulan data yang telah dikategorisasikan ini disebut dengan data latih yang akan digunakan sebagai model.
4. Model ini bisa digambarkan sebagai aturan klasifikasi, pohon keputusan atau formula matematika.
5. Algoritma berdasarkan model di atas untuk mengklasifikasi disebut dengan classifier (pengklasifikasi).

## 2.8 Penelitian Terkait

Penelitian yang terkait dengan penelitian yang akan dilakukan penulis, mengacu pada penelitian sebelumnya sebagai acuan untuk penelitian selanjutnya. Referensi jurnal-jurnal yang digunakan memiliki kemiripan dari penulisan yang dilakukan penulis, oleh karena itu jurnal-jurnal yang digunakan menjadi pertimbangan bagi penulis dalam melakukan penelitian dan membuat penulisan ini. Berikut adalah jurnal-jurnal terkait yang telah dikaji oleh penulis.

### 1. **Klasifikasi Kategori Pengaduan Masyarakat Melalui Kanal LAPOR! Menggunakan Artificial Neural Network (Mochamad Ihsan Ananto, Wiwiek Setya Winahju dan Kartika Fithriasari, 2019)**

Penelitian yang dilakukan oleh Mochamad Ihsan Ananto dkk ini menggunakan model analisis klasifikasi teks menggunakan *Artificial Neural Network* serta SMOTE untuk mengatasi data *imbalance* dan *Chi-Square* untuk proses seleksi variabel. Penelitian ini dibuat untuk melakukan sebuah pengklasifikasian berdasarkan karakteristik aduan masyarakat yang masuk via LAPOR!, berapa tingkat ketepatan klasifikasi yang didapatkan dari kata kunci pada setiap aduan. Pada penelitian ini terdapat dua variabel yang digunakan setelah tahap *text pre-processing*. Yakni terdiri dari variabel prediktor yaitu bobot dari kata dasar pada setiap aduan dan variabel respon yang telah ditentukan. Data yang diolah terbagi ke dalam enam kategori aduan sehingga sudah dapat diketahui bagaimana karakteristik data awal sebelum proses analisis. Data tersebut memuat isi aduan, kategori, dan status tindak lanjut aduan. Melalui proses seleksi variabel, didapatkan sejumlah 428 term atau kata yang memberikan pengaruh terhadap kategori aduan masyarakat. Ketepatan klasifikasi yang dihasilkan melalui metode *Artificial Neural Network* dengan *feature selection* dan 3 *nodes hidden layer* adalah *precision* 0,794, *sensitivity* 0,818 dan *F1-Score* 0,800. Selain itu didapatkan topik permasalahan yang patut mendapatkan perhatian lebih pada setiap kategori aduan dengan menggunakan *word cloud*.

## **2. Klasifikasi Jenis Laporan Masyarakat dengan K-Nearest Neighbor Algorithm (Heru Pramono Hadi, Titien S. Sukamto, 2020)**

Penelitian ini menggunakan *metode text mining* dengan algoritma *K-Nearest Neighbor* untuk jenis laporan yang disampaikan melalui twitter ke dalam 3 (tiga) kategori, yaitu, pengaduan, permintaan informasi, dan aspirasi. Proses dimulai dengan membagi data laporan yang dikumpulkan pada tahun 2017 ke dalam data latih dan data uji. Kedua, melaksanakan setiap langkah pre-processing, kemudian menghitung bobot dengan TF-IDF dan implementasi algoritma K-Nearest Neighbor (KNN). Terakhir melakukan evaluasi dan uji validasi dengan Confusion Matrix. Penerapan algoritma K-Nearest Neighbor memberikan hasil akurasi yang baik untuk proses klasifikasi data laporan masyarakat ke dalam 3 kategori. Berdasarkan hasil uji evaluasi dan validasi dengan Confusion Matrix, ditemukan bahwa parameter  $k = 11$  memiliki nilai accuracy tertinggi dalam penentuan kelas kategori laporan masyarakat, yaitu 82%.

## **3. Klasifikasi Teks Laporan Masyarakat Pada Situs LAPOR! Menggunakan Recurrent Neural Network (Imam Fahrur Rozi, Vivi Nur Wijyaningrum, Nur Khozin, 2020)**

Penelitian ini dibuat untuk mengklasifikasi data dari Sistem Pengelolaan Pengaduan Pelayanan Publik LAPOR!. menggunakan metode *Long Short-Term Memory Recurrent Neural Network* untuk melakukan proses klasifikasi setiap dokumen laporan masyarakat. Pada pengumpulan data laporan masyarakat dilakukan menggunakan proses scraping pada situs LAPOR! dengan memperhatikan laporan yang sudah terverifikasi dan sudah berstatus selesai, serta dapat mengambil laporan berdasarkan periode yang dibutuhkan. Kemudian dalam proses klasifikasi teks laporan masyarakat dapat dilakukan dengan mengirim sebuah request ke REST API sistem yang telah dibuat pada penelitian. Pada tahap implementasi metode long short-term memory (LSTM) recurrent neural network (RNN) dapat digunakan untuk proses klasifikasi laporan masyarakat berdasarkan instansi yang terkait, dimana pengujian

dilakukan menggunakan k-fold cross-validation sebanyak 10 bagian data, pengujian tersebut menghasilkan rata-rata persentase f-measure sebesar 85,69% untuk dataset seimbang dan 79,44% untuk dataset tidak seimbang, sedangkan dari total keseluruhan pengujian menghasilkan f-measure tertinggi sebesar 88,82%.

#### **4. Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network (Dyan Yuliana, Purwanto, Catur Supriyanto, 2019)**

Penelitian ini dibuat untuk mengklasifikasi pengaduan masyarakat pada situs Lapor Gub! dengan menggunakan Algoritma *Neural Network*. Dataset yang digunakan pada penelitian ini berjumlah 300 data set, terdiri dari 10 kategori dengan masing-masing kategori ada 30 data set, diantaranya Infrastruktur, Sosial Masyarakat, Kependudukan, Energi, Kesehatan, Keuangan dan Aset, Pertanian, Sektor Lain-Lain, Pendidikan, dan Kepegawaian. Pada hasil eksperimen, peneliti memakai jumlah label yang berbeda mulai dari 4 sampai dengan 10 label dalam eksperimen pada *tools rapid miner*. Hasilnya dapat diketahui bahwa semakin banyak label yang digunakan dan semakin tinggi nilai P (seleksi fitur) dalam eksperimen, maka cenderung semakin lama waktu yang dibutuhkan dalam melakukan klasifikasi teks pengaduan masyarakat ini serta nilai akurasi yang dihasilkan pun cenderung semakin kecil. Pada hasil eksperimen dengan 10 label, nilai akurasi tertinggi dihasilkan oleh proses seleksi fitur dengan  $p = 0,08$  yaitu sebesar 43,00% dalam jangka waktu 03 jam 45 menit 14 detik. Dari hasil tersebut, yang paling besar ialah pada prediksi Kesehatan menghasilkan 64.29%.

#### **5. Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes (Dyah Ariyanti, Kurnia Iswardani, 2020)**

Penelitian ini dibuat untuk melakukan *teks mining* dalam mengklasifikasi keluhan masyarakat dengan algoritma *naïve bayes* pada situs Laporo Rek Kota Probolinggo untuk diteruskan ke dalam masing-masing Satuan Kerja Pemerintah Daerah (SKPD) Kota Probolinggo. Data keluhan masyarakat

diambil melalui sms dan telepon kemudian dilakukan proses data training dan data testing, selanjutnya dilakukan proses text mining (praproses). Pada praproses ini kemudian ditentukan kata-kata baku yang termasuk dalam kamus. Kata yang terpilih merupakan kata dasar (tidak berimbuhan) dan bukan termasuk di dalam stopword. Kata yang terpilih ini akan menentukan proses perhitungan Naïve Bayes. Penelitian ini menghasikan tingkat akurasi mencapai 95%, sehingga dapat mengklasifikasikan keluhan masyarakat tiap-tiap instansi di pemerintah Kota Probolinggo

#### **6. Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network (Muhamad Rizal Firmansyah, Ridwan Ilyas, Fatan Kasyidi, 2020)**

Penelitian ini dibuat untuk melakukan pengklasifikasian pada satu kalimat ilmiah tunggal dan tidak terpengaruh oleh kalimat sebelum dan sesudah tetapi hanya berfokus pada satu kalimat ilmiah saja. Menggunakan metode *Recurrent Neural Network* (RNN) yang merupakan arsitektur jaringan saraf tiruan yang telah terbukti berkinerja baik karena pemrosesannya disebut berulang kali untuk memproses *input* data sekuensial. Penelitian ini telah berhasil membuat model komputasi klasifikasi kalimat menggunakan RNN, dengan fitur yang telah diekstraksi menggunakan fungsi Word2Vec untuk menghasilkan satu set vektor. Dalam melakukan proses klasifikasi penelitian ini menggunakan total 2019 data pelatihan kalimat ilmiah yang telah dilabeli menjadi empat kelas, yaitu *Weak*, *Comparison*, *Point*, dan *Neutral*. Penelitian ini telah dibandingkan dengan empat optimasi yaitu Adam, SGD Adadelta, dan Adamax untuk menemukan tingkat pembelajaran terbaik dan cocok untuk klasifikasi kalimat. Hasil tingkat pembelajaran terbaik diperoleh dengan pengoptimalan SGD dengan nilai akurasi 77,48% dan Loss 0,71%.

Tabel 2.1 Ringkasan Penelitian Klasifikasi

No	Peneliti	Metode	Hasil/Ekperimen	Kelebihan	Kekurangan
1.	Mochamad Ihsan Ananto, Wiwiek Setya Winahju dan Kartika Fithriasari (2019)	<ul style="list-style-type: none"> <li>• <i>Artificial Neural Network</i></li> <li>• <i>SMOTE</i></li> <li>• <i>Text Mining</i></li> <li>• <i>Word Cloud</i></li> <li>• <i>Dataset situs LAPOR!</i></li> </ul>	<ul style="list-style-type: none"> <li>• Mengklasifikasikan data aduan berdasarkan seleksi variabel kata yang berpengaruh.</li> <li>• Rata-rata ketepatan klasifikasi yang dihasilkan melalui <i>Artificial Neural Network</i> dengan 428 <i>feature</i> dan <i>nodes hidden layer</i> adalah <i>precision</i> 0,794, <i>sensitivity</i> 0,818, dan <i>F1-score</i> 0,800.</li> </ul>	Penelitian ini berhasil mengklasifikasikan kategori aduan berdasarkan kata kunci aduan	Perlu adanya penambahan kata kunci bahasa sehari-hari maupun tafsiran dari berbagai singkatan yang lebih bervariasi agar kata sebenarnya dapat terhitung dalam frekuensi kemunculan kata.
2.	Heru Pramono Hadi, Titien S. Sukamto (2020)	<ul style="list-style-type: none"> <li>• <i>K-Nearest neighbor</i></li> <li>• <i>Text Mining</i></li> </ul>	<ul style="list-style-type: none"> <li>• Penerapan Algoritma <i>K-Nearest Neighbor</i> untuk jenis laporan yang disampaikan melalui twitter berhasil mengklasifikasi</li> </ul>	Menghasilkan klasifikasi jenis laporan berdasarkan kategori yang telah ditentukan	Diperlukan banyak data aktual pada data uji kelas aspirasi, sehingga memberikan hasil Confusion Matrix yang lebih baik.

No	Peneliti	Metode	Hasil/Ekperimen	Kelebihan	Kekurangan
			<p>kan ke dalam 3 (tiga) kategori, yaitu, pengaduan, permintaan informasi, dan aspirasi.</p> <ul style="list-style-type: none"> <li>• Nilai akurasi tertinggi k=11 sebesar 82%</li> </ul>		
3.	Imam Fahrur Rozi, Vivi Nur Wijayaningrum, Nur Khozin, (2020)	<ul style="list-style-type: none"> <li>• <i>Recurrent Neural Network</i> (RNN)</li> <li>• <i>Dataset</i> situs LAPOR!</li> </ul>	<ul style="list-style-type: none"> <li>• Klasifikasi laporan pengaduan ke instansi terkait</li> <li>• <i>f-measure</i> sebesar 85,69% untuk dataset seimbang dan 79,44% untuk dataset tidak seimbang</li> <li>• <i>f-measure</i> tertinggi sebesar 88,82%</li> </ul>	<p>Penelitian ini menghasilkan klasifikasi laporan pengaduan kepada instansi terkait dengan mengirim sebuah request ke REST API sistem yang telah dibuat pada penelitian kemudian dilakukan implementasi metode <i>long short-term memory</i> (LSTM) <i>recurrent neural network</i> (RNN)</p>	<p>Pada tahap preprocessing perlu menambahkan kamus kata yang berisikan kata singkatan, kata daerah dan kata gaul</p>
4	Dyan Yuliana, Purwanto, Catur	<ul style="list-style-type: none"> <li>• <i>Algoritma Neural Network</i></li> </ul>	<ul style="list-style-type: none"> <li>• Klasifikasi teks pengaduan masyarakat</li> </ul>	<p>Penelitian ini berhasil mengklasifikasikan berdasarkan 10 bidang</p>	<p>Tingkat akurasi masih tergolong rendah dan</p>



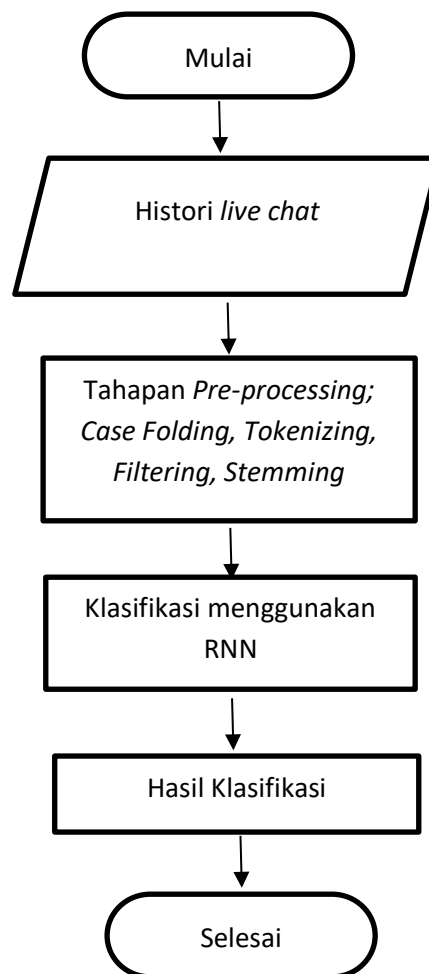
No	Peneliti	Metode	Hasil/Ekperimen	Kelebihan	Kekurangan
	Supriyanto, (2019)	<ul style="list-style-type: none"> <li><i>Rapid miner</i></li> </ul>	<ul style="list-style-type: none"> <li>dengan menggunakan algoritma <i>Neural Network</i></li> <li>Akurasi yang tinggi yaitu sebesar 43,00% dengan jangka waktu 03 jam 45 menit 14 detik dalam mengklasifikasikan dokumen teks berbahasa Indonesia</li> </ul>	pengaduan dengan algoritma <i>Neural Network</i>	waktunya relatif lama
5.	Dyah Ariyanti, Kurnia Iswardani, (2020)	<ul style="list-style-type: none"> <li><i>Teks Mining</i></li> <li>Naïve Bayes</li> <li>Dataset Laporo Rek Kota Probolinggo</li> </ul>	Klasifikasi pengaduan masyarakat ke Satuan Kerja Pemerintah Daerah (SKPD) Kota Probolinggo	Data keluhan masyarakat dipilih menjadi kata yang terpilih dalam menentukan proses perhitungan Naïve Bayes. Penelitian ini menghasilkan tingkat akurasi mencapai 95%	Akurasi data diragukan karena kata yang dimasukan hanya berupa kata dasar
6.	Muhamad Rizal Firmansyah, Ridwan Ilyas,	<ul style="list-style-type: none"> <li><i>Recurrent Neural Network (RNN)</i></li> </ul>	<ul style="list-style-type: none"> <li>Klasifikasi kalimat menggunakan RNN, dengan</li> </ul>	Penelitian ini berhasil membuat klasifikasi untuk kalimat ilmiah yang telah dilabeli	Hanya berfokus pada satu kalimat ilmiah.

No	Peneliti	Metode	Hasil/Ekperimen	Kelebihan	Kekurangan
	Fatan Kasyidi (2020)	<ul style="list-style-type: none"> <li>Algoritma Word2Vec</li> <li>Corpus dataset berjumlah 2019</li> </ul>	<p>fitur Word2Vec untuk menghasilkan satu set vector.</p> <ul style="list-style-type: none"> <li>nilai akurasi 77,48% dan Loss 0,71%.</li> </ul>	<p>menjadi empat kelas, yaitu <i>Weak</i>, <i>Comparison</i>, <i>Point</i>, dan <i>Neutral</i>. Penelitian ini telah melakukan perbandingan dengan empat optimasi yaitu Adam, SGD Adadelta, dan Adamax untuk menemukan tingkat pembelajaran terbaik dan cocok untuk klasifikasi kalimat dan hasil tingkat pembelajaran terbaik diperoleh dengan pengoptimalan SGD dengan nilai akurasi 77,48% dan Loss 0,71%.</p>	

### 3. METODOLOGI PENELITIAN

#### 3.1. Tahapan Penelitian

Tahap penelitian dimulai dengan mendapatkan *corpus* dataset histori *live chat* yang digunakan untuk melatih data. Data pada sistem klasifikasi percakapan akan diproses dengan *pre-processing* yaitu *Case Folding*, *Tokenizing*, *Filtering* dan *Stemming*. Hasil dari tahapan *pre-processing* selanjutnya akan dilakukan pembelajaran mesin menggunakan RNN untuk mendapatkan nilai bobot dan hasil tersebut akan dilakukan klasifikasi menggunakan LSTM. Sistem klasifikasi kalimat ilmiah dapat dilihat pada Gambar 2.



Gambar 3.1 Tahapan Penelitian

### 3.2 RENCANA KERJA

**Tabel 3.1 Rencana Pelaksanaan Penelitian**

	Bulan Ke-																	
Kegiatan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Studi Pustaka																		
Pemilihan Topik																		
Eksplorasi Objek																		
Penyusunan Proposal																		
Pengajuan Proposal																		
Pengerjaan Penelitian																		
Publikasi Ilmiah / Seminar																		
Pengembangan Atas Saran Seminar																		
Penyusunan Desertasi																		
Sidang Desertasi																		

## DAFTAR PUSTAKA

- [1] A. Clark, “Pre-processing very noisy text,” 2003.
- [2] Aggarwal, C. C., and Zhai, C. 2012. A survey of text classification algorithms. In Mining text data. Springer. 163–222.
- [3] D. Yogatama, C. Dyer, W. Ling, and P. Blunsom, “Generative and Discriminative Text Classification with Recurrent Neural Networks,” 2017.
- [4] Dyah Ariyanti, Kurnia Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes" ,2020.
- [5] Dyan Yuliana, Purwanto, Catur Supriyanto, "Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network", 2019
- [6] Heru Pramono Hadi, Titien S. Sukamto, "Klasifikasi Jenis Laporan Masyarakat dengan K-Nearest Neighbor Algorithm", 2020
- [7] J. Han, M. Kamber, “Data mining: concepts and techniques. Second Edition,” San Francisco: Morgan Kaufmann, 2006
- [8] Imam Fahrur Rozi, Vivi Nur Wijayaningrum, Nur Khozin, "Klasifikasi Teks Laporan Masyarakat Pada Situs LAPOR! Menggunakan Recurrent Neural Network", 2020.
- [9] Mochamad Ihsan Ananto, Wiwiek Setya Winahju dan Kartika Fithriasari, “Klasifikasi Kategori Pengaduan Masyarakat Melalui Kanal LAPOR! Menggunakan Artificial Neural Network”, 2019

- [10] M. R. Firmansyah, R. Ilyas, F Kasyidi, “Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network,” Prosiding The 11th Industrial Research Workshop and National Seminar, pp. 488-495, 2020.
- [11] R. Cahyadi, A. Damayanti, D. Aryadani, “*Recurrent Neural Network (RNN) Dengan Long Short Term Memory (LSTM) Untuk Analisis Sentimen Data Instagram*”, Jurnal Informatika dan Komputer (JIKO), Volume 5, Nomor 1, 2020.
- [12] R. Nasrullah, “Media Sosial; Perspektif Komunikasi, Budaya, dan Sosioteknologi”, Bandung : Simbiosis Rekatama Media, 2015.
- [10] R. P. Sidiq, B. A.Dermawan, Y.Umaidah, “Sentimen Analisis Komentar *Toxic* pada Grup Facebook *Game Online* Menggunakan Klasifikasi Naïve Bayes,” Jurnal Informatika Universitas Pamulang, Vol. 5, 2020.
- [13] R. Sharma<sup>1</sup> M.Patel, “Toxic Comment Classification Using Neural Networks and Machine Learning,” International Advanced Research Journal in Science, Engineering and Technology, Vol. 5, Issue 9, 2018.
- [14] S. Mestry, V.Bisht, R.Chauhan, K.Tiwari, H.Singh, “Multi Label Classification Of Toxic Comment Using Fast-Text and CNN,” International Journal of Advances in Computer Science and Cloud Computing, Volume 7, 2019.
- [15] W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang, “Novel architecture for long short-term memory used in question classification,” *Neurocomputing*, vol. 299, pp. 20–31, 2018.
- [16] <https://taptalk.io/blog/memberikan-customer-service-yang-lebih-baik-melalui-live-chat>, diakses pada tanggal 26 September 2022.