**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang Masalah Loh**

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. 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 *naïve bayes* dengan tingkat akurasi 80% untuk mengklasifikasikan 3 kategori yaitu berita palsu (*hoax*), berita asli (*fact*) dan berita yang tidak berindikasi (Tanjung, 2018). Selain itu, ada penelitian lain yang membandingkan antara algoritma *rocchio* dengan *multinomial naïve bayes* dengan akurasi yang didapatkan sebesar 83,501% untuk *rocchio* dan 65,835% untuk *multinomial naïve bayes* (Afriza & Adisantoso, 2018). Algoritma *support vector machine* pun pernah digunakan untuk klasifikasi artikel *hoax* dengan tingkat akurasi yang didapatkan sebesar 95,8333% (Maulina & Sagara, 2018). Ada beberapa metode lain yang pernah digunakan seperti *gradient descent backpropagation* dengan tingkat akurasi 72,04% sedangan *lavenberg-marquad backpropagation* memiliki akurasi 72,19% dalam klasifikasi pengguna media sosial (Lhaksmana, Nhita, & Budhiarto, 2017). Penerapan *deep learning* juga pernah dilakukan dalam klasifikasi sentimen analisis *tweet* dengan menggunakan metode *deep belief network* dengan tingkat akurasi sebesar 93,31% (Zulfa & Winarko, 2017).

Ada tantangan yang harus dihadapi dalam proses klasifikasi berita *hoax* yaitu banyaknya berita *hoax* yang memiliki tingkat kemiripan yang sangat tinggi dengan berita aslinya. Bahkan dalam beberapa kasus, penambahan atau perubahan beberapa kata dari berita asli akan merubah maksud dan tujuan dari berita tersebut. Seperti kendala dari penggunaan metode *naïve bayes*, dimana kemiripan data membuat sulitnya menemukan pola berita *hoax* sehingga mempengaruhi tingkat error dan akurasinya (Tanjung, 2018). Selain itu, banyaknya data yang digunakan membuat waktu yang diperlukan untuk melakukan *preprocessing* dan *training* semakin lama. Untuk meminimalisisr hal tersebut, dapat dilakukan dengan menggunakan metode *deep learning*.

Kelebihan dari penggunaan metode *deep learning* tersebut karena memiliki algoritma dengan struktur dan jumlah jaringan syaraf yang sangat banyak sebagai jaringan tersembunyi (*hidden layer*) yang dapat digunakan untuk berbagai hal, salah satunya klasifikasi teks. Metode *deep learning* yang dapat digunakan untuk klasifikasi teks yaitu *Convolutional Neural Network (CNN), Recurrent Neural Network (RNN),* dan *Hierarchical Attention Network (HAN)*. Ketiga metode tersebutmemiliki layer hirarki banyak/bertumpuk-tumpuk sehingga kendala yang berkaitan dengan kemiripan antar data dapat diminimalisir dan hasil yang didapat menjadi lebih akurat. Kelebihan lain yang dimiliki metode-metode tersebut yaitu menghilangkan ekstraksi fitur secara manual, tidak memerlukan informasi terhadap data yang akan dipelajarinya, dan dapat dikombinasikan dengan *Graphics Processing Unit* (GPU) untuk mempercepat proses *preprocessing* maupun *training*. Algoritma dari *deep learning* dapat melakukan *tuning* (penyetelan) dan pemilihan model yang paling optimal secara mandiri. Selain itu, metode tersebut terbukti memiliki tingkat akurasi yang tinggi dan efisien dalam proses pengklasifikasian. Seperti penelitian yang pernah dilakukan oleh Razi (2017) tentang Klasifikasi Artikel Berita Berbahasa Indonesia Menggunakan *Convolutional Neural Network* dengan nilai akurasi sebesar 96,70% dan presisi, recall serta f-measure mencapai 96,60%. Ketiga metode tersebut juga pernah dibandingkan untuk klasifikasi teks dengan hasil dari CNN, RNN, dan HAN secara berturut turut yaitu 96,12%, 95,76%, 95,6% menggunakan dataset pertama, 87,88%, 82,63%, 87,91% menggunakan dataset kedua, dan dengan dataset ketiga menghasilkan 91,6%, 89,54%, 92,16% (Maheshwari, 2018).

Oleh karena itu, penelitian ini akan mengkomparasikan metode *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax.* Penelitian ini akan menggunakan data yang bersumber dari turnbackhoax.id yang telah memvalidasi berita *hoax* yang beredar di media sosial. Data tersebut akan menjadi data latih dan juga data uji. Penelitian ini diharapkan dapat mengetahui metode *deep learning* mana yang memiliki tingkat performa yang lebih baik. Selain itu, penelitian ini juga diharapkan dapat menghasilkan sebuah sistem yang dapat dimanfaatkan untuk melakukan klasifikasi berita sehingga dapat diketahui apakah berita tersebut termasuk kategori hoax atau bukan.

* 1. **Perumusan Masalah**

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

1. Bagaimana tingkat akurasi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* dalam klasifikasi berita *hoax*?
2. Bagaimana tingkat efisiensi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* dalam klasifikasi berita *hoax*?
   1. **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* (*CNN*)*, Recurrent Neural Network* (*RNN*)dan *Hierarchical Attention Network* (*HAN*).
   1. **Tujuan Penelitian**

Tujuan pada penelitian ini adalah sebagai berikut:

1. Dapat mengetahui tingkat akurasi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax*.
2. Dapat mengetahui tingkat efisiensi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax*.
   1. **Manfaat Penelitian**

Dengan dilakukannya penelitian ini dapat mengetahui algoritma *deep learning* mana yang memiliki tingkat performa yang lebih tinggidalam 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. **Metodologi Penelitian**

Metodologi penelitian yang digunakan pada penelitian ini adalah sebagai berikut:

* + 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. Pengumpulan Data

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. **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. **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 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,* *convolutional neural network, recurrent neural network,* dan *hierarchical attention 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**

* 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)
   1. **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). Beberapa contoh dari media sosial seperti facebook, instagram, twitter dan lain sebagainya.

* 1. ***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). Secara garis besar terdapat dua teknik untuk melakukan *web scraping,* yaitu:

1. *Manual*, yaitu menyalin data dari suatu webseite dengan cara *copy paste*

2. *Otomatis,* yaitu dengan menggunakan koding, aplikasi ataupun *extension browser*

* 1. ***Text Mining***

*Text mining* merupakan variasi dari data *mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar (Feldman & Sanger, 2007). *Text mining* adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen (Han & Kamber, 2006). *Text mining* bertujuan untuk menemukan pola atau kata-kata yang mewakili isi dari dokumen yang selanjutnya akan di analisis.

* 1. ***Text Preprocessing***

*Text preprocessing* adalah proses untuk membersihkan dan mempersiapkan data sebelum dilakukannya proses klasifikasi (Haddi, Liu, & Shi, 2013). Adanya *text preprocessing* bertujuan untuk mendapatkan bentuk data yang siap olah. Ada beberapa tahapan pada *text preprocessing* yaitu :

* + 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’. Karakter selain itu akan dihilangkan dan dianggap *delimiter*.

* + 1. ***Tokenizing***

Tahap *tokenizing* atau tokenisasi adalah proses pemisahan teks dari paragraph atau kalimat menjadi bagian-bagian tertentu yang disebut sebagai token. Pemisah antar token biasanya adalah spasi atau tanda baca. 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”.

* + 1. ***Filtering***

Tahap *filtering* adalah proses pengambilan kata-kata penting (*wordlist*) atau membuang kata kata kurang penting (*stopword*) dari token yang telah dihasilkan pada tahap sebelumnya. Pembuangan *stopword* tidak akan mengubah makna dan isi dari suatu data. Beberapa contoh *stopword* seperti yang, juga, aku, kamu, ini, dan itu.

* + 1. ***Stemming***

Tahap *stemming* adalah proses pemisahan kata yang mengandung awalan, akhiran maupun sisipan sehingga menghasilkan kata dasar. Hal ini bertujuan untuk mengurangi varian kata dengan makna yang hampir sama. Selain itu, proses *stemming* dapat meningkatkan kinerja pengambilan data karena jumlah indeks dengan istilah yang berbeda menjadi berkurang.

* 1. **Klasifikasi Teks**

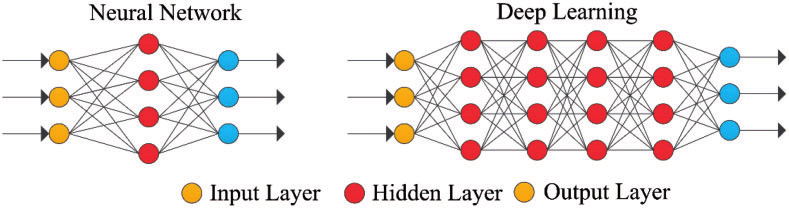
Klasifikasi teks adalah proses untuk menemukan sebuah model atau fungsi yang menggambarkan dan membedakan antar kelas data. Klasifikasi teks didefinisikan sebagai pengkategorian teks secara otomatis ke dalam satu atau lebih kelas yang telah ditentukan berdasarkan isinya (Sebastiani, 2002). Tujuan dari klasifikasi teks yaitu untuk memperkirakan kelas dari suatu objek yang belum diketahui kelasnya (Gaikwad, Chaugule, & Patil, 2014).

* 1. ***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* dikarenakan algoritma *deep learning* tidak dapat memproses data dalam bentuk *string*, maka dari itu harus dikonversi ke dalam bentuk angka terlebih dahulu.

* 1. ***Deep Learning***

*Deep learning* adalah salah satu bidang dari pembelajaran mesin (*machine learning*) yang berbasis jaringan syaraf tiruan. *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.



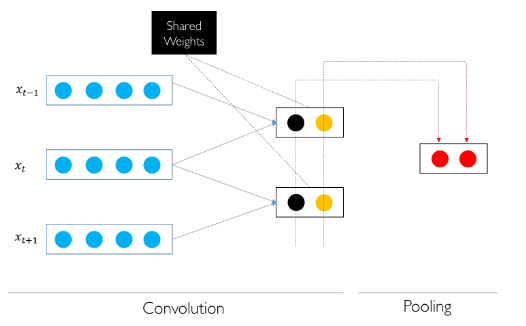
**Gambar 2.1** Diagram *Deep Learning* (Xing & Du, 2018)

* 1. ***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 2 dimensi. Namun, ada beberapa penelitian yang menerapkan algoritma CNN pada *Natural Language Processing* (NLP) dengan hasil yang cukup memuaskan. Konsep dari CNN tidak jauh berbeda dari *neural network* biasa dimana terdiri dari *neuron* yang memiliki *weight, bias* dan *activation function*.

* + 1. ***Feature Extraction***

*Feature extraction* adalah bagian awal dari arsitektur CNN yang tersusun atas beberapa *layer* dan masing-masing *layer* tersusun atas *neuron neuron* yang saling terhubung. Pada bagian ini terdapat 2 jenis *layer* yaitu *convolution layer* dan *pooling layer*.



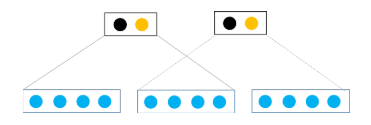
**Gambar 2.2** *Convolution* dan *Pooling* (Putra, 2019)

* + - 1. ***Convolution***

*Convolution layer* adalah *layer* yang terdiri dari *neuron* yang tersusun dan membentuk *filter. Convolution layer* berisi operasi konvolusi dengan operasi kombinasi linear *filter*. Pada *layer* ini memiliki *hyperparameter* dan *parameter*.

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

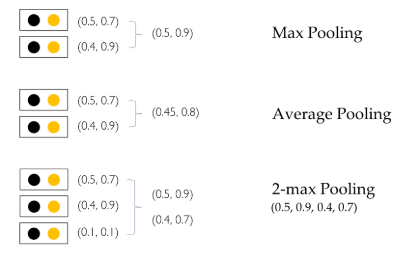
|  |  |  |
| --- | --- | --- |
| **No** | ***Hyperparameter*** | **Keterangan** |
| 1 | *Depth* | Kedalaman *layer* atau jumlah *layer* konvolusi |
| 2 | *Stride* | Jumlah pergeseran *filter* pada proses konvolusi |
| 3 | *Zero-padding* | Jumlah penambahan nilai intensitas nol di daerah sekitar input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

****

**Gambar 2.3** *1D Convolution* (Putra, 2019)

* + - 1. ***Pooling***

*Pooling* adalah sebuah filter yang akan bergeser ke seluruh area *feature map* dengan ukuran dan stride tertentu. Pembuatan model CNN dengan memilih banyak jenis lapisan *pooling* dapat menguntungkan kinerja model (Lee, Gallagher, & Tu, 2016). Ada 3 jenis *pooling* yang sering digunakan yaitu *Max Pooling, Average Pooling* dan kombinasi dari keduanya*.* Sesuai dengan namanya, *Max Pooling* yaitu dengan mengambil nilai terbesar dari *pool,* sedangan *Average Pooling* mengambil nilai rata-rata dari *pool*.Tujuan dari penggunaan *pooling layer* untuk mengurangi dimensi dari *feature map* (*downsampling*), sehingga mempercepat komputasi karena parameter yang harus diupdate semakin sedikit dan mengatasi *overfitting* (Sena, 2017).



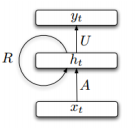
**Gambar 2.4** *Pooling* pada CNN (Putra, 2019)

* + 1. ***Fully Connected Layer***

*Feature map* yang dihasilkan dari *feature learning* masih berbentuk multidimensional array. Selanjutnya, akan dilakukan proses *flatten* atau *reshape feature map* yang bertujuan untuk merubah menjadi sebuah vektor sehingga dapat digunakan sebagai input dari *fully connected layer*.

* 1. ***Recurrent Neural Network* (RNN)**

*Recurrent Neural Network* atau RNN adalah algoritma *neural network* yang memiliki kemampuan *feedback* kembali ke *neuron* itu sendiri ataupun ke *neuron* lainnya sehingga aliran informasi dari *inputan* mempunyai arah jamak. RNN memiliki memori sebagai pengingat histori masa lalu. RNN memiliki minimal satu *feedback loop* di dalam *recurrent*. *Feedback loop* mempengaruhi kemampuan belajar dan kinerja jaringan (Udin et al., 2017).



**Gambar 2.5** *Recurrent Neural Network* sederhana (Mikolov, Joulin, Chopra, Mathieu, & Ranzato, 2014)

**Daftar Pustaka**

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

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

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

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

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.

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

Lee, C. Y., Gallagher, P. W., & Tu, Z. (2016). Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016*, 464–472.

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.

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

Maheshwari, A. (2018). *Report on Text Classifiction using CNN, RNN & HAN*.

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.

Mikolov, T., Joulin, A., Chopra, S., Mathieu, M., & Ranzato, M. (2014). Learning Longer Memory in Recurrent Neural Networks. *International Conference on Learning Representations (ICLR)*, *1*(4), 2.

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

Putra, J. W. G. (2019). *Pengenalan Pembelajaran Mesin dan Deep Learning*. (March 2018).

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.

Sena, S. (2017). *Pengenalan Deep Learning Part 7 : Convolutional Neural Network (CNN)*.

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

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

Udin, M., Kaloko, B. S., Hardianto, T., Elektro, J. T., Teknik, F., Unej, U. J., & Kalimantan, J. (2017). Peramalan Kapasitas Baterai Lead Acid pada Mobil Listrik Berbasis Levenberg Marquardt Neural Network. *BERKALA SAINSTEK*, *2*, 112–117.

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/

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

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