

**PENERAPAN *SELF-ATTENTIVE NETWORK* PADA
PENDETEKSIAN *CLICKBAIT* DI JUDUL BERITA *ONLINE*
INDONESIA**

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh :

Tiara Oktavian

NIM: 09021281722059

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN SKRIPSI

**PENERAPAN *SELF-ATTENTIVE NETWORK* PADA
PENDETEKSIAN *CLICKBAIT* DI JUDUL
BERITA *ONLINE* INDONESIA**

Oleh :

Tiara Oktavian

09021281722059

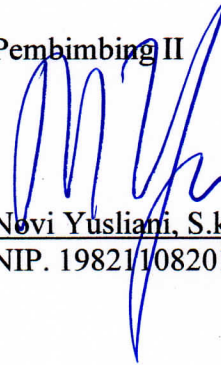
Palembang, Januari 2023

Pembimbing I



Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

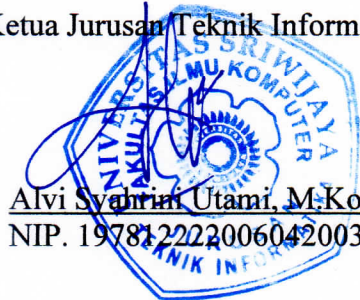
Pembimbing II



Novi Yusliani, S.kom., M.T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

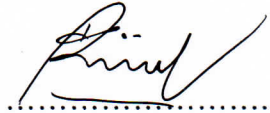
Pada hari Selasa tanggal 13 Desember 2022 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Tiara Oktavian
NIM : 09021281722059
Judul : Penerapan Self-attentive Network Pada Pendeteksian Clickbait di Judul Berita Online Indonesia

dan dinyatakan LULUS.

1. Ketua

Mastura Diana Marieska, M.T.
NIP. 198603212018032001



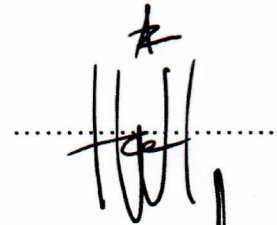
2. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



3. Penguji II

Hadipurnawan Satria, Ph.D
NIP. 198004182020121001



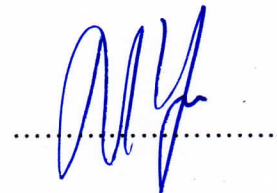
4. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005



5. Pembimbing II

Novi Yusliani, S.kom., M.T.
NIP. 198211082012122001



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Tiara Oktavian
NIM : 09021281722059
Program Studi : Teknik Informatika
Judul : Penerapan Self-attentive Network Pada Pendeteksian
Clickbait di Judul Berita Online Indonesia

Hasil Pengecekan *software ithenticate/Turnitin* : 4%

Menyatakan bahwa laporan tugas akhir merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Januari 2023

Tiara Oktavian

NIM. 09021281722059

MOTTO DAN PERSEMBAHAN

MOTTO :

“Life is a journey to be experienced, not a problem to be solved”

~ Winnie The Pooh ~

“Better an oops than a what if”

~ Unknown ~

Kupersembahkan karya tulis ini kepada :

- Mama dan papa tersayang
- Adik-adikku yang tersayang
- Dosen Pembimbing dan Penguji
- Teman seperjuanganku
- Fakultas Ilmu Komputer
- Universitas Sriwijaya
- Diri sendiri

IMPLEMENTATION OF SELF-ATTENTIVE NETWORK ON CLICKBAIT DETECTION IN INDONESIAN ONLINE NEWS HEADLINE

By:

Tiara Oktavian

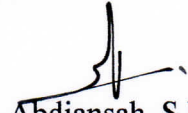
09021281722059

ABSTRACT

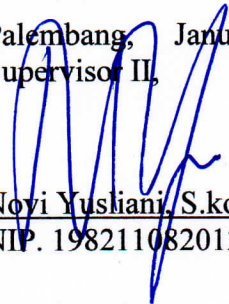
Clickbait is a term that describes a title with the aim of attracting the reader's interest by using flashy and provocative word choices. The problem faced if clickbait detection is done manually is that it takes a long time to be done, because it has to compare the content of the news by reading a whole news for each news title. This study intends to developed software that can detect whether a given news title is clickbait or not by using the self-attentive network method. The data used in this study were 6000 for training data and 2000 for test data. The data will enter the preprocessing stage before going to the self-attentive network. The data were trained using 10-fold Cross Validation. Based on the results of the test data, the model with the best performance is in the 4th fold with precision, recall, and f1-score values of 0.8057, 0.7960, and 0.8008. So that the accuracy value is 0.8020.

Keywords: Clickbait, self-attentive network, news, cross validation

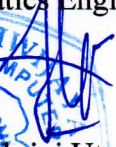
Supervisor I,

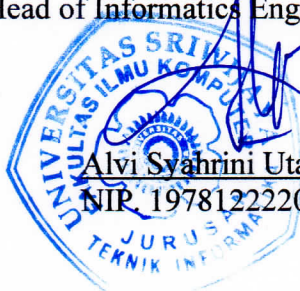

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Palembang, January 2023
Supervisor II,


Noyi Yusliani, S.kom., M.T.
NIP. 198211082012122001

Approve,
Head of Informatics Engineering Department


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



**PENERAPAN *SELF-ATTENTIVE NETWORK* PADA PENDETEKSIAN
CLICKBAIT DI JUDUL BERITA *ONLINE* INDONESIA**

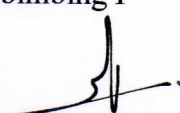
Oleh :
Tiara Oktavian
09021281722059

ABSTRAK

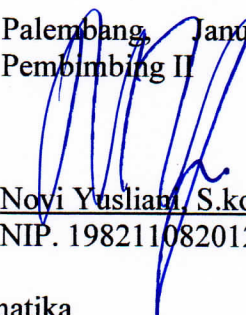
Clickbait adalah istilah yang mendeskripsikan suatu judul dengan tujuan untuk menarik minat si pembaca dengan menggunakan pilihan kata yang mencolok dan terkesan provokatif. Permasalahan yang dihadapi jika pendeteksian *clickbait* dilakukan secara manual adalah memakan waktu yang lama, karena harus membandingkan isi berita dengan membaca berita secara keseluruhan untuk setiap judul berita yang ada. Penelitian ini bertujuan untuk mengembangkan perangkat lunak yang dapat mendeteksi apakah judul berita yang diberikan merupakan *clickbait* atau bukan dengan menggunakan metode *self-attentive network*. Data yang digunakan pada penelitian ini sebanyak 6000 untuk data latih dan 2000 untuk data uji. Data akan memasuki tahap pra-pengolahan sebelum ke *self-attentive network*. Data dilatih menggunakan *10-fold Cross Validation*. Berdasarkan hasil data uji, model dengan performa terbaik ada pada *fold* ke 4 dengan nilai *precision*, *recall*, dan *f1-score* adalah 0,8057, 0,7960, dan 0,8008. Sehingga didapat nilai *accuracy* sebesar 0,8020.

Kata kunci: *Clickbait*, *self-attentive network*, berita, *cross validation*


Pembimbing I


Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Palembang, Januari 2023
Pembimbing II


Novi Yusliani, S.kom., M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya tugas akhir ini dapat selesai dengan baik. Tugas akhir ini dengan judul “Penerapan *Self-attentive Network* Pada Pendeteksian *Clickbait* di Judul Berita Online Indonesia” disusun untuk memenuhi salah satu syarat dalam menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan tugas akhir ini banyak pihak yang telah terlibat dalam memberikan bantuan serta dukungan emosional baik secara langsung maupun tidak langsung. Penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT yang telah memberikan nikmat, kelancaran dan kemudahan dalam menyelesaikan tugas akhir ini.
2. Diriku karena telah memilih untuk tidak menyerah dan terus berjuang.
3. Kedua orang tuaku, mama dan papa yang percaya dan mendukung pilihan anaknya serta selalu mendoakan dan memberikan yang terbaik untuk anaknya.
4. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
5. Ibu Alvi Syahrini Utami, M.Kom. selaku ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Dr. Abdiansah, S.kom., M.Cs. selaku dosen pembimbing I dan Ibu Novi Yusliani, S.kom., M.T. selaku dosen pembimbing II yang

telah membimbing, mengarahkan, dan memberikan kemudahan kepada penulis dalam proses perkuliahan serta pengerjaan tugas akhir.

7. Ibu Alvi Syahrini Utami, M.Kom selaku dosen penguji I, Bapak Hadipurnawan Satria, Ph.D selaku dosen penguji II, serta Ibu Mastura Diana Marieska, M.T. selaku ketua penguji yang telah memberikan saran dan masukan serta ilmu pengetahuan.
8. Bapak Al Farissi, S.Kom., M.Cs. selaku dosen pembimbing akademik dari semester satu hingga semester enam dan Bapak Ali Ibrahim, S.Kom., M.T selaku dosen pembimbing akademik dari semester tujuh hingga kini, yang telah memberikan bimbingan, arahan serta motivasi penulis selama masa perkuliahan.
9. Seluruh Dosen Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penulis selama kegiatan akademik berlangsung.
10. Kak Ricy selaku admin jurusan teknik informatika reguler beserta seluruh staff Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
11. Teman seperjuanganku, Ibal, Nur, Zhicma, dan Zaneva yang telah menemani penulis dari awal perkuliahan hingga saat ini, tempat berbagi informasi hingga motivasi dalam menyelesaikan tugas akhir ini.

12. Teman kelas dan jurusan Teknik Informatika Reguler Angkatan 2017 yang telah berbagi keluh kesah, motivasi, semangat, dan warna selama proses perkuliahan.

Penulis menyadari dalam penyusunan tugas akhir ini masih terdapat banyak kekurangan yang disebabkan atas keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian ke depannya. Akhir kata semoga tugas akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, November 2022

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-4
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 <i>Clickbait</i>	II-1
2.2.2 <i>Text Pre-processing</i>	II-2
2.2.3 <i>Cross Validation</i>	II-6
2.2.4 <i>Word Embedding</i>	II-7
2.2.5 <i>Self-attentive Network</i>	II-9
2.2.6 Pengukuran Kinerja.....	II-14

2.2.7	<i>Rational Unified Process (RUP)</i>	II-16
2.3	Penelitian Lain yang Relevan.....	II-18
2.4	Kesimpulan.....	II-19
BAB III METODOLOGI PENELITIAN.....		III-1
3.1.	Pendahuluan	III-1
3.2.	Pengumpulan Data.....	III-1
3.2.1.	Pelabelan Dataset	III-2
3.3.	Tahapan Penelitian	III-2
3.3.1.	Menentukan Kerangka Kerja	III-3
3.3.2.	Menetapkan Kriteria Pengujian.....	III-5
3.3.3.	Menetapkan Format Data Pengujian	III-5
3.3.4.	Menentukan Alat yang Digunakan dalam Penelitian.....	III-6
3.3.5.	Melakukan Pengujian Penelitian.....	III-7
3.3.6.	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan	III-7
3.4.	Metode Pengembangan Perangkat Lunak	III-7
3.4.1.	Fase Insepsi	III-8
3.4.2.	Fase Elaborasi	III-8
3.4.3.	Fase Konstruksi.....	III-8
3.4.4.	Fase Transisi	III-9
3.5.	Kesimpulan.....	III-9
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-2
4.2.3	Analisis dan Perancangan Perangkat Lunak	IV-3
4.2.4	Desain Perangkat Lunak	IV-17
4.3	Fase Elaborasi.....	IV-23
4.3.1.	Pemodelan Bisnis	IV-23
4.3.2.	Kebutuhan Sistem	IV-24
4.3.3.	Diagram <i>Sequence</i>	IV-24
4.3.4.	Perancangan Antar Muka.....	IV-27

4.4	Fase Konstruksi	IV-29
4.4.1.	Kebutuhan Sistem	IV-29
4.4.2.	Diagram Kelas.....	IV-31
4.4.3.	Implementasi	IV-32
4.5	Fase Transisi.....	IV-38
4.5.1.	Pemodelan Bisnis	IV-38
4.5.2.	Kebutuhan Sistem	IV-38
4.5.3.	Rencana Pengujian	IV-38
4.5.4.	Implementasi	IV-40
4.6	Kesimpulan.....	IV-44
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Percobaan Penelitian	V-1
5.2.1	Analisis Hasil Penelitian	V-4
5.3	Kesimpulan.....	V-24
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1.	Pendahuluan	VI-1
6.2.	Kesimpulan.....	VI-1
6.3.	Saran	VI-2
DAFTAR PUSTAKA		xvii
LAMPIRAN.....		xxi

DAFTAR TABEL

Tabel II- 1. Hasil Case Folding	II-3
Tabel II- 2. Hasil URL Removal.....	II-4
Tabel II- 3. Hasil Special Character Removal	II-5
Tabel II- 4. Hasil Stopword removal.....	II-6
Tabel II- 5. Confusion Matrix	II-15
Tabel III- 1. Confussion Matrix	III-6
Tabel III- 2. Rancangan Tabel Hasil Pengujian	III-6
Tabel IV- 1. Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV- 2. Kebutuhan Non-fungsional Perangkat Lunak	IV-3
Tabel IV- 3. Judul Berita Sebelum Tahapan Pra-pengolahan.....	IV-7
Tabel IV- 4. Hasil Case Folding	IV-8
Tabel IV- 5. Hasil Mention dan Retweet Removal.....	IV-9
Tabel IV- 6. Hasil Hashtag Removal	IV-9
Tabel IV- 7. Hasil URL Removal	IV-10
Tabel IV- 8. Hasil Special Character Removal.....	IV-11
Tabel IV- 9. Hasil Proses Stemming.....	IV-12
Tabel IV- 10. Hasil Proses Stop Word Removal	IV-13
Tabel IV- 11. Hasil Proses Number to Word.....	IV-13
Tabel IV- 12. Hasil setelah diubah kedalam bentuk sequence.....	IV-15
Tabel IV- 13. Hasil padding.....	IV-16
Tabel IV- 14. Definisi Aktor Use Case.....	IV-18
Tabel IV- 15. Definisi Use Case	IV-19
Tabel IV- 16. Skenario Use Case Melihat Dokumentasi Evaluasi Model.....	IV-19
Tabel IV- 17. Skenario Use Case Melakukan Deteksi.....	IV-20
Tabel IV- 18. Tabel Implementasi Kelas	IV-32
Tabel IV- 19. Rencana Pengujian Use Case Melihat Dokumentasi Evaluasi Model	IV-39
Tabel IV- 20. Rencana Pengujian Use Case Input Judul Berita	IV-39
Tabel IV- 21. Rencana Pengujian Use Case Melakukan Deteksi	IV-39
Tabel IV- 22. Pengujian Use Case Melihat Dokumentasi Evaluasi Model	IV-41
Tabel IV- 23. Pengujian Use Case Input Judul Berita	IV-42
Tabel IV- 24. Pengujian Use Case Melakukan Deteksi	IV-43
Tabel V- 1. Penjelasan Arsitektur dan Fitur yang digunakan	V-2
Tabel V- 2. Hasil confusion matrix model fold 1	V-4
Tabel V- 3. Hasil pengujian model fold 1	V-5
Tabel V- 4. Hasil confusion matrix model fold 2	V-6
Tabel V- 5. Hasil pengujian model fold 2.....	V-6
Tabel V- 6. Hasil confusion matrix model fold 3	V-7
Tabel V- 7. Hasil pengujian model fold 3.....	V-8
Tabel V- 8. Hasil confusion matrix model fold 4	V-8
Tabel V- 9. Hasil pengujian model fold 4.....	V-9

Tabel V- 10. Hasil confusion matrix model fold 5	V-10
Tabel V- 11. Hasil pengujian model fold 5.....	V-10
Tabel V- 12. Hasil confusion matrix model fold 6	V-11
Tabel V- 13. Hasil pengujian model fold 6.....	V-12
Tabel V- 14. Hasil confusion matrix model fold 7	V-12
Tabel V- 15. Hasil pengujian model fold 7.....	V-13
Tabel V- 16. Hasil confusion matrix model fold 8	V-14
Tabel V- 17. Hasil pengujian model fold 8.....	V-14
Tabel V- 18. Hasil confusion matrix model fold 9	V-15
Tabel V- 19. Hasil pengujian model fold 9.....	V-16
Tabel V- 20. Hasil confusion matrix model fold 10	V-16
Tabel V- 21. Hasil pengujian model fold 10.....	V-17
Tabel V- 22. Tabel Nilai Akurasi, Precision, Recall dan F1-score.....	V-21
Tabel V- 23. Hasil Prediksi Model fold 4.....	V-22

DAFTAR GAMBAR

Gambar II- 1. Struktur pembagian data 10-fold cross validation.....	II-7
Gambar II- 2. Arsitektur self-attentive network	II-10
Gambar II- 3. Arsitektur Rational Unified Process (RUP)	II-17
Gambar III- 1. Diagram Tahapan Penelitian.....	III-3
Gambar III- 2. Kerangka Kerja	III-3
Gambar III- 3. Flow Chart Pengujian Penelitian	III-7
Gambar IV- 1. Vocabulary index kata	IV-14
Gambar IV- 2. Diagram Use Case	IV-18
Gambar IV- 3. Diagram Aktivitas Melihat Dokumentasi Evaluasi Model.....	IV-22
Gambar IV- 4. Diagram Aktivitas Melakukan Deteksi	IV-23
Gambar IV- 5. Diagram Sequence Deteksi Clickbait	IV-25
Gambar IV- 6. Diagram Sequence Fungsi identify_text.....	IV-26
Gambar IV- 7. Diagram Sequence Melihat Dokumentasi Evaluasi Model	IV-26
Gambar IV- 8. Tampilan halaman utama.....	IV-27
Gambar IV- 9. Tampilan setelah menekan tombol “submit”.....	IV-28
Gambar IV- 10. Tampilan side bar	IV-28
Gambar IV- 11. Tampilan Halaman Dokumentasi Evaluasi Model	IV-29
Gambar IV- 12. Diagram Kelas	IV-31
Gambar IV- 13. Implementasi Halaman utama	IV-33
Gambar IV- 14. Tampilan peringatan ketika masukan melebihi 250 karakter	IV-34
Gambar IV- 15. Tampilan error ketika tidak ada masukan.....	IV-35
Gambar IV- 16. Tampilan error ketika masukan melebihi 250 karakter	IV-35
Gambar IV- 17. Tampilan Hasil Clickbait.....	IV-36
Gambar IV- 18. Tampilan Hasil “non-clickbait”	IV-36
Gambar IV- 19. Tampilan Dokumentasi Hasil Training	IV-37
Gambar IV- 20. Tampilan Dokumentasi Hasil Testing	IV-37
Gambar V-1. Grafik Akurasi Model Saat Training dan Testing.....	V-18
Gambar V-2. Grafik Accuracy Training dan Validation di fold 4.....	V-19
Gambar V-3. Grafik Loss Training dan Validation di Fold 4.....	V-20

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dijelaskan secara umum tentang gambaran penelitian yang dimulai dari latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan serta kesimpulan yang didapat dari bab pendahuluan ini.

1.2 Latar Belakang Masalah

Pada hakikatnya, setiap manusia membutuhkan informasi untuk menjalankan kehidupan sehari-hari. Informasi saat ini bukanlah hal yang sulit didapat karena telah banyaknya platform media penyebaran informasi. Platform media penyebaran informasi ini selalu berkembang dari masa ke masa sejalan dengan perkembangan teknologi yang ada. Segala jenis media cetak baik itu berupa surat kabar, majalah, maupun tabloid telah mengalami penurunan yang tajam sejak tahun 2015, tak lain disebabkan oleh peningkatan pengguna internet di Indonesia. Berdasarkan data yang dikumpulkan oleh *Hootsuite (We Are Social)* pada tahun 2020 pengguna internet di Indonesia mencapai 175,4 juta dan pengguna media sosial aktif mencapai 160 juta. Keadaan ini membuat media digital di Indonesia bertumbuh sangat cepat, salah satunya portal berita *online* (Manan, 2018).

Kesuksesan portal berita *online* dapat dilihat dari tingkat lalu lintas *website*, semakin tinggi dan stabil lalu lintas *website* tersebut, maka keuntungan yang didapat juga akan semakin besar. Keuntungan ini berasal dari biaya sewa klien yang

memasang iklan mereka di *website*. Upaya yang dilakukan untuk meningkatkan lalu lintas di *website* berita tersebut salah satunya yaitu memasang judul berita yang menjebak atau *clickbait*.

Clickbait adalah istilah yang mendeskripsikan suatu judul dengan tujuan untuk menarik minat si pembaca. Terkadang berita yang judulnya mengandung *clickbait* banyak mengecewakan pembaca karena isi berita tersebut tidak sesuai dengan judul berita bahkan ada yang berisi berita fiktif. Penyebaran berita fiktif di Indonesia masih menjadi hal yang harus diberi perhatian lebih. Salah satu media penyebaran berita fiktif ini adalah berita-berita yang memiliki judul *clickbait* (Anand et al., 2017).

Permasalahan yang dihadapi jika pendeteksian *clickbait* dilakukan secara manual adalah memakan waktu yang lama, karena harus membandingkan isi berita dengan membaca isi berita secara keseluruhan untuk setiap judul berita yang ada (Papadopoulou et al., 2017). Oleh karena itu, perlu dikembangkan suatu perangkat lunak yang bisa mendeteksi judul *clickbait* secara otomatis. Beberapa metode yang pernah digunakan untuk pendeteksian *clickbait* dengan parameter yang digunakan dalam pendeteksian berbeda-beda sehingga menghasilkan hasil akurasi yang beragam. Metode tersebut meliputi: *linear regression*, *naive bayes*, dan *random forest* merupakan metode yang digunakan dalam penelitian yang dilakukan oleh Potthast et al., (2016), *linear regression* dan *support vector machine* yang dikombinasikan dengan *Convolutional Neural Networks* (CNNs) dan *word embedding* (GLoVe) yang seluruh metode itu dinamakan *Parallel Neural Networks* dilakukan oleh Adelson et al., (2017). Kemudian ada *Recurrent Neural Network* yang menggunakan arsitektur *Bi-directional Long Short-Term Memory* (LSTM) yang dilakukan oleh Anand et al., (2017).

Sementara itu, penelitian dalam mendeteksi *clickbait* untuk data berbahasa Indonesia masih sedikit yang melaksanakannya. Penelitian yang dilaksanakan oleh Yavi (2018) menggunakan metode *Naïve Bayes* dengan data yang berasal dari situs berita *online*.

Penelitian terdahulu memandang bahwa pendeteksian *clickbait* merupakan masalah pengelompokan biner, label di dataset hanya berisikan *clickbait* atau *non-clickbait*. Namun, Potthast et al., (2018) menemukan bahwa penilaian secara biner ini masih kurang optimal karena setiap orang memiliki pandangan yang berbeda akan seberapa menarik suatu judul berita tersebut. Sehingga Zhou (2017) meneliti pendeteksian *clickbait* ini dalam bentuk regresi menggunakan metode *self-attentive network*. Hasil penelitian ini mendapatkan *Mean Squared Error* (MSE) sebesar 0,033 dengan nilai *F1-Score* sebesar 0,683 dan waktu eksekusi tiga menit dua puluh tujuh detik. Ketiga nilai tersebut mengungguli peserta lain dan menjadi *state-of-the-art* dalam *Clickbait Challenge 2017*.

Berdasarkan penelitian yang dilakukan oleh Zhou (2017) maka penelitian ini akan menerapkan metode yang sama untuk diimplementasikan ke dalam dataset berbahasa Indonesia. Penerapan metode ini bermaksud untuk melihat seberapa efektif jika model tersebut menggunakan dataset berbahasa Indonesia.

1.3 Rumusan Masalah

Berikut merupakan rumusan masalah dalam penelitian ini:

1. Bagaimana mendeteksi *clickbait* pada judul berita *online* di Indonesia menggunakan metode *self-attentive network*?
2. Bagaimana kinerja *Self-attentive Network* dalam pendeteksian *clickbait* pada judul berita?

1.4 Tujuan Penelitian

Berikut merupakan tujuan penelitian ini :

1. Membuat perangkat lunak yang dapat mendeteksi judul berita yang diberikan termasuk *clickbait* atau tidak.
2. Mengetahui kinerja *Self-attentive Network* dalam pendeteksian *clickbait* pada judul berita bahasa Indonesia.

1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah sebagai berikut :

1. Membantu dalam mengurangi penyebaran berita yang judulnya berupa *clickbait*. Sehingga meningkatkan kualitas informasi yang beredar di masyarakat.
2. Dataset yang dikumpulkan dalam penelitian ini dapat menjadi bahan penelitian berikutnya yang berfokus dalam pengelompokan teks ataupun pendeteksian *clickbait*.
3. Dapat menjadi rujukan bagi penelitian lain yang juga menggunakan metode *Self-attentive Network*.

1.6 Batasan Masalah

Penelitian ini memiliki batasan atau ruang lingkup permasalahan yang akan diteliti. Batasan masalah tersebut yaitu :

1. Data yang digunakan hanya berfokus pada judul berita dalam Bahasa Indonesia saja.

2. Data latih yang digunakan merupakan data yang diambil dari beberapa akun resmi portal berita di *Twitter* di antaranya CNN Indonesia (@CNNIndonesia), tvOneNews (@tvOneNews), Liputan6.com (@liputan6dotcom), Detikcom (@detikcom), dan VIVAcoid (@vivacoid).
3. Penelitian ini hanya berfokus pada judul berita saja tanpa mencari keterkaitan antara isi berita dengan judul.
4. Data latih dan data uji yang digunakan merupakan berkas dengan ekstensi csv.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut :

BAB I. PENDAHULUAN

Pendahuluan berisi mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah atau ruang lingkup, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan di bahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi *Clickbait*, *Text Pre-processing*, *Cross Validation*, *Word Embedding*, *Self-Attentive Network* serta beberapa pembahasan mengenai penelitian sebelumnya yang relevan dengan penelitian ini. Penjelasan teori dan pembahasan penelitian ini berguna untuk menganalisis, perancangan dan implementasi tugas akhir.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai proses pengembangan perangkat lunak berdasarkan kerangka kerja yang digunakan. Hal ini meliputi dari analisis kebutuhan perangkat lunak, rancangan perangkat lunak hingga pengujian untuk melihat kinerja perangkat lunak apakah telah sesuai dengan kebutuhan dan rancangan yang telah ditentukan.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dibahas mengenai hasil penelitian berdasarkan perangkat lunak yang telah dibangun dan analisis yang dilakukan akan menjadi dasar dalam menarik kesimpulan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan dibahas mengenai kesimpulan dari penelitian dan saran apa saja yang diharapkan bisa membangun penelitian selanjutnya menjadi lebih baik.

1.8 Kesimpulan

Pada bab ini didapat kesimpulan bahwa penelitian yang akan dilakukan menggunakan data dari akun *Twitter* portal berita *online* Indonesia dan model yang digunakan untuk mendeteksi *clickbait* adalah *self-attentive network*, yang mana arsitekturnya akan dijelaskan lebih lanjut pada bab II.

DAFTAR PUSTAKA

- Adelson, P., Arora, S., & Hara, J. (2017). *Clickbait; Didn't Read: Clickbait Detection using Parallel Neural Networks*.
<https://nlp.stanford.edu/projects/glove/>
- Anand, A., Chakraborty, T., & Park, N. (2017). We used neural networks to detect clickbaits: You won't believe what happened next! *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10193 LNCS, 541–547.
https://doi.org/10.1007/978-3-319-56608-5_46
- Bauskar, S., Badole, V., Jain, P., & Chawla, M. (2019). Natural Language Processing based Hybrid Model for Detecting Fake News Using Content-Based Features and Social Features. *International Journal of Information Engineering and Electronic Business*, 11(4), 1–10.
<https://doi.org/10.5815/ijieeb.2019.04.01>
- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(April), 542–545.
<https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop Clickbait: Detecting and preventing clickbaits in online news media. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 9–16.
<https://doi.org/10.1109/ASONAM.2016.7752207>
- Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment

- analysis. *Procedia Computer Science*, 17, 26–32.
<https://doi.org/10.1016/j.procs.2013.05.005>
- Hestie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2 ed., Vol. 2). Springer International Publishing.
- Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security*, 16(6), 22–32. <https://sites.google.com/site/ijcsis/>
- Kalra, V., & Aggarwal, R. (2018). Importance of Text Data Preprocessing & Implementation in RapidMiner. *Proceedings of the First International Conference on Information Technology and Knowledge Management*, 14, 71–75. <https://doi.org/10.15439/2017km46>
- Manan, A. (2018). Potret Bisnis Media dan Kesejahteraan Pekerjaannya. *Ancaman Baru dari Digital Laporan Tahunan AJI 2018*, 55–68.
- Nurdin, A., Aji, B. A. S., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2Vec , Glove ., *Jurnal TEKNOKOMPAK*, 14(2), 74–79.
- Omidvar, A., Jiang, H., & An, A. (2019). Using neural network for identifying clickbaits in online news media. *Communications in Computer and Information Science*, 898, 220–232. https://doi.org/10.1007/978-3-030-11680-4_22
- Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, I. (2017). A *Two-Level Classification Approach for Detecting Clickbait Posts using Text-*

- Based Features*. <http://arxiv.org/abs/1710.08528>
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <http://www.aclweb.org/anthology/D14-1162>
- Potthast, M., Gollub, T., Komlossy, K., Schuster, S., Wiegmann, M., Garces Fernandez, E. P., Hagen, M., & Stein, B. (2018). Crowdsourcing a large corpus of clickbait on Twitter. *COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings*, 1498–1507.
- Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016). *Clickbait Detection*. 9626, 810–817. https://doi.org/10.1007/978-3-319-30671-1_72
- Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach. Learn. Technol.*, 2.
- Selivanov, D. (2020). *GloVe Word Embeddings*. https://cran.r-project.org/web/packages/text2vec/vignettes/glove.html#glove_algorithm
- Sommerville, I. (2016). Software Processes. In *Software engineering (10th edition)* (10 ed., hal. 46–47). Pearson Education Limited.
- Sravani, L., Reddy, A. S., & Thara, S. (2018). A Comparison Study of Word Embedding for Detecting Named Entities of Code-Mixed Data in Indian Language. *2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018*, 2375–2381. <https://doi.org/10.1109/ICACCI.2018.8554918>
- Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D*, pp, 39–46.

- William, A., & Sari, Y. (2020). CLICK-ID: A novel dataset for Indonesian clickbait headlines. *Data in Brief*, 32, 106231.
<https://doi.org/10.1016/j.dib.2020.106231>
- Yavi, A. F. (2018). Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Zhou, Y. (2017). *Clickbait Detection in Tweets Using Self-attentive Network*.
<http://arxiv.org/abs/1710.05364>