

Text Generation Menggunakan LSTM pada Website LinkedIn

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



Oleh :

Muhammad Rizqi Assabil
NIM : 09021281924065

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022

LEMBAR PENGESAHAN SKRIPSI

Text Generation Menggunakan LSTM pada Website LinkedIn

Oleh :

Muhammad Rizqi Assabil
NIM : 09021381924065

Palembang, 20 Januari 2023

Pembimbing II

Pembimbing I

Novi Yusliani, S.Kom., M.T.
NIP : 198211082012122001

Annisa Darmawahyuni M.Kom
NIP : 1671147006900002

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP : 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari **Kamis** tanggal **05 Januari 2023** telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhammad Rizqi Assabil

NIM : 09021281924065

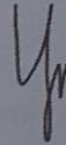
Judul : Text Generation menggunakan LSTM pada *Website* LinkedIn

dan dinyatakan **LULUS**.

1. Ketua Penguji

Yunita, M.Cs

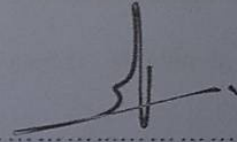
NIP : 198306062015042002



2. Penguji I

Dr. Abdiansah, S.Kom, M.Cs

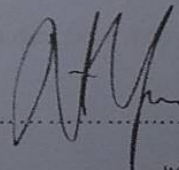
NIP : 198410012009121005



3. Pembimbing I

Novi Yusliani, S.Kom., M.T.

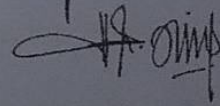
NIP : 198211082012122001



4. Pembimbing II

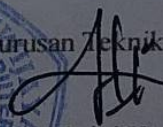
Annisa Darmawahyuni M.Kom

NIP : 1671147006900002



Mengetahui
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom
NIP : 19781222200604200



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Muhammad Rizqi Assabil
NIM : 09021281924065
Program Studi : Teknik Informatika
Judul Skripsi : Text Generation menggunakan LSTM pada *Website*
LinkedIn
Hasil Pengecekan Software (Turnitin) : 11%

Menyatakan bahwa Laporan Penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan penelitian 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, 16 Januari 2023



Muhammad Rizqi Assabil

NIM : 09021281924065

MOTTO

- Semua yang berkilau bukanlah emas
- Skripsi yang baik adalah skripsi yang selesai

Karya tulis ini dipersembahkan kepada:

- Penulis di masa depan
- Orang tua penulis

ABSTRACT

LinkedIn is one of the most popular sites out there to advertise oneself to potential employer. This study aims to create a good enough text generation model that it can generate a text as if it were made by someone who posts on LinkedIn. This study will use a Neural Network layer called Long-Short Term Memory (LSTM) as the main algorithm and the train data consists of actual posts made by users in LinkedIn. LSTM is an algorithm that is created to reduce vanishing and exploding gradient problem in Neural Network. From the result, final accuracy and loss varies. Increasing learning rate from its default value of 0.001, to 0.01, or even 0.1 creates worse model. Meanwhile, increasing dimensions of LSTM will sometimes increases training time or decreases it while not really increasing model performance. In the end, models chosen at the end are models with around 97% of accuracy that has a fairly stable learning graph and predicted output.

Key Word : Artificial Neural Network, Long-Short Term Memory, Natural Language Generation

ABSTRAK

LinkedIn adalah salah satu laman web paling populer untuk mengiklankan diri sendiri ke perusahaan-perusahaan. Penelitian ini ditujukan untuk membuat sebuah model text generation yang cukup baik hingga teks yang dibuat seolah-olah berasal dari sebuah kiriman dari LinkedIn. Penelitian ini akan menggunakan sebuah *layer* jaringan syaraf tiruan bernama Long-Short Term Memory (LSTM) sebagai algoritma utama dengan menggunakan data yang berasal dari kiriman asli oleh pengguna LinkedIn. LSTM adalah algoritma yang akan dibuat untuk menurunkan efek dari *vanishing* dan *exploding gradient problem* pada jaringan syaraf tiruan. Hasil menunjukkan bahwa accuracy dan loss beragam. Menaikkan nilai learning rate dari nilai bawaannya yaitu 0.001 ke nilai 0.01, lalu ke 0.1, menghasilkan model yang lebih buruk. Sementara itu, menaikkan dimensi LSTM terkadang memperlama atau mempercepat waktu latih tanpa menghasilkan perubahan signifikan pada performa model. Di akhir penelitian ini, model yang dipilih adalah model dengan accuracy sekitar 97% dengan perubahan accuracy dan loss yang stabil pada proses latih dan hasil keluaran yang bisa diterima.

Kata Kunci : Jaringan Syaraf Tiruan, Long-Short Term Memory, Natural Language Generation

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT yang memberikan penulis kesempatan untuk menyelesaikan skripsi ini. Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu :

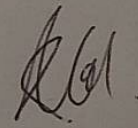
1. Kedua orang tua penulis yang telah memberikan penulis kesempatan untuk menempuh kuliah.
2. Bapak Jaidan Jauhari, M.T selaku kepala Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom. selaku ketua jurusan Teknik Informatika.
4. Ibu Novi Yusliani, S.Kom., M.T dan Ibu Annisa Darmawahyuni M.Kom, kedua pembimbing skripsi penulis yang telah mempermudah proses penulisan skripsi penulis.
5. Seluruh staf tata usaha yang telah membantu dalam menyelesaikan proses administratif.
6. Bapak Samsuryadi, S.Si., M.Kom., Ph.D selaku pembimbing akademik penulis.
7. Teman seperjuangan penulis, Helmi Farizki dan Armenia Yuhafiz, yang telah memberikan penulis motivasi untuk menyelesaikan proposal skripsi dan skripsi.
8. Teman penulis, Ariib, Rezky, Friza, dan seluruh orang pada grup “E.1.3” yang telah menemani penulis selama masa perkuliahan.

9. Nadiem Anwar Makarim, selaku menteri pendidikan, kebudayaan, riset, dan teknologi, yang telah memberikan penulis kesempatan untuk mengikuti program kampus merdeka.
10. Petugas Damri, bus, angkot, dan transportasi lainnya yang telah membawa penulis pulang pergi selama masa perkuliahan.
11. Seluruh pihak lainnya yang telah membantu penulis selama masa perkuliahan, yang namanya tidak bisa ditulis satu-persatu.

Penulis harap hasil dari penelitian ini dapat digunakan sebagai referensi untuk penelitian serupa di masa depan. Penulis harap hasil dan penulisan dari skripsi ini tidak memiliki banyak kesalahan. Meskipun demikian, penulis berharap untuk menerima kritik dan saran yang dapat memperbaiki penulisan penelitian selanjutnya.

Palembang, 16 Januari 2023

Penulis,



Muhammad Rizqi Assabil

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO.....	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xiv
DAFTAR ALGORITMA.....	xv
BAB I PENDAHULUAN	
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah.....	I-4
1.4 Tujuan Penelitian.....	I-5
1.5 Manfaat Penelitian.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-6
1.8 Kesimpulan.....	I-7

BAB II KAJIAN LITERATUR

2.1 Pendahuluan.....	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Natural Language Generation.....	II-1
2.2.2 Neural Network.....	II-3
2.2.3 Long-Short Term Memory.....	II-5
2.2.4 Data <i>Preprocessing</i>	II-9
2.3 Metode Pengembangan <i>Waterfall</i>	II-15
2.4 Penelitian lain yang serupa.....	II-16
2.5 Kesimpulan.....	II-18

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan.....	III-1
3.2 Pengumpulan Data.....	III-1
3.3 Tahapan Penelitian.....	III-2
3.4 Metode Pengembangan Perangkat Lunak.....	III-10
3.4.1 Requirement.....	III-10
3.4.2 Design.....	III-10
3.4.3 Implementation.....	III-11
3.4.4 Verification.....	III-11
3.4.5 Maintenance.....	III-11
3.5 Manajemen Proyek Penelitian.....	III-12
3.5.1 Sumber Daya.....	III-12
3.5.2 Spesifikasi Kebutuhan.....	III-12

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan.....	IV-1
4.2 Metode <i>Waterfall</i>	IV-1
4.2.1 Requirement.....	IV-1
4.2.2 Design.....	IV-3
4.2.3 Implementation.....	IV-7
4.2.4 Verification.....	IV-28
4.2.5 Maintenance.....	IV-33

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan.....	V-1
5.2 Hasil.....	V-1
5.3 Analisis.....	V-30

BAB VI KESIMPULAN DAN SARAN

6.1 Pendahuluan.....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

	Halaman
Gambar II-1. Unit LSTM (Olah, 2015).....	II-6
Gambar II-2. Penjelasan operasi LSTM (Olah, 2015).....	II-7
Gambar II-3. <i>N-Gram Sequence</i> untuk fitur.....	II-12
Gambar II-4. <i>Padded Sequence</i>	II-13
Gambar II-5. <i>One-Hot Encoded</i>	II-14
Gambar II-6. Metode Pengembangan Waterfall (Hughey, 2006).....	II-16
Gambar III-1. Flowchart Penelitian.....	III-3
Gambar III-2. Arsitektur Secara Singkat.....	III-5
Gambar III-3. Arsitektur Neural Network.....	III-6
Gambar III-4. Proses Prediksi.....	III-8
Gambar IV-1. Desain Tampilan.....	IV-4
Gambar IV-2. Use Case Diagram.....	IV-5
Gambar IV-3. <i>Sequence Diagram</i>	IV-6
Gambar IV 4. Diagram Aktivitas.....	IV-6
Gambar IV-5. Detil Berkas.....	IV-9
Gambar IV-6. Hasil Pembuangan.....	IV-10
Gambar IV-7. Melihat Daftar Nama.....	IV-11
Gambar IV-8. Tampilan Depan.....	IV-31
Gambar V-1. Confusion Matrix untuk model Word Based.....	V-27
Gambar V-2. <i>Confusion Matrix</i> untuk model <i>Character Based</i>	V-28

DAFTAR TABEL

	Halaman
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Kebutuhan Non-fungsional.....	IV-3
Tabel IV-3. Use Case Scenario.....	IV-5
Tabel IV-4. Contoh Data.....	IV-13
Tabel IV-5. Contoh Data setelah Dipisah.....	IV-18
Tabel IV-6. <i>Vocabulary</i>	IV-19
Tabel V-1. <i>Hyperparameter</i>	V-2
Tabel V-2. Hasil untuk setiap <i>Hyperparameter</i>	V-4
Tabel V-3. Grafik Accuracy dan Loss.....	V-6
Tabel V-4. Data Prediksi.....	V-17
Tabel V-5. Hasil Kuesioner.....	V-25

DAFTAR ALGORITMA

	Halaman
Algoritma IV-1. Import Pustaka.....	IV-8
Algoritma IV-2. Membaca Berkas.....	IV-9
Algoritma IV-3. Membuang Konten yang Tidak Perlu.....	IV-10
Algoritma IV-4. Membuat Plot.....	IV-11
Algoritma IV-5. Membuang Data yang Kurang.....	IV-12
Algoritma IV-6. Membersihkan Data.....	IV-14
Algoritma IV-7. Lanjutan Pembersihan.....	IV-14
Algoritma IV-8. Filter Jumlah Elemen.....	IV-16
Algoritma IV-9. Filter Elemen Kedua.....	IV-17
Algoritma IV-10. Proses Tokenisasi.....	IV-19
Algoritma IV-11. Menyimpan Vocabulary.....	IV-20
Algoritma IV-12. N-Gram Sequence.....	IV-20
Algoritma IV-13. Padding Sequence.....	IV-21
Algoritma IV-14. Merubah menjadi <i>Feature</i> dan <i>Label</i>	IV-22
Algoritma IV-15. Membuat Model.....	IV-24
Algoritma IV-16. Membuat Callbacks.....	IV-25
Algoritma IV-17: Memisah Data.....	IV-25
Algoritma IV-18: <i>Batching</i>	IV-26
Algoritma IV-19. Inisialisasi proses Training.....	IV-27
Algoritma IV-20. Menjalankan Training.....	IV-27
Algoritma IV-21. Konversi Model ke Tflite.....	IV-28
Algoritma IV-22. Melakukan Proses Prediksi.....	IV-29

BAB I PENDAHULUAN

1.1 Pendahuluan

Bab ini akan menjelaskan mengenai penelitian yang akan dilakukan pada skripsi ini. Adapun yang dibahas adalah latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

1.2 Latar Belakang Masalah

Text Generation adalah sebuah kasus pada kecerdasan buatan dan komputer linguistik di mana sebuah sistem komputer dapat mengkonstruksi sebuah teks yang dapat dimengerti manusia dengan menggunakan informasi atau representasi non-linguistik (Reiter and Dale, 1997). *Text Generation* juga disebut sebagai *Natural Language Generation* (NLG). NLG sendiri sudah ada sejak pertengahan 1960 dengan terciptanya ELIZA, sebuah sistem yang mensimulasi percakapan antar manusia yang dibuat oleh Joseph Weizenbaum, seorang profesor dari MIT (Ireland, 2022).

NLG tidak hanya digunakan untuk menciptakan kalimat yang baku, tetapi juga dalam menciptakan laporan, seperti laporan cuaca (Goldberg et al., 1994), laporan pasien (Portet et al., 2009), dan *image caption* (Hutchison et al., 2010). Tantangan umum NLG dalam menciptakan sebuah kalimat adalah bahasa yang digunakan.

Bahasa yang digunakan oleh manusia sangatlah rumit, contohnya adalah beberapa kata memiliki arti yang berbeda, lalu tiap bahasa juga memiliki kata-kata yang berbeda untuk hal yang sama, tetapi cara penggunaannya dalam sebuah kalimat berbeda-beda.

Teknik yang digunakan untuk menciptakan sebuah sistem NLG sangatlah beragam, tetapi secara umum tekniknya bekerja menggunakan metode statistika dengan pembelajaran mesin yang diberikan data *corpus* sebagai data belajarnya (Perera and Nand, 2017). Beberapa teknik diantaranya adalah *Markov Chain*, *Recursive Neural Network* (RNN), *Long-Short Term Memory* (LSTM), dan yang paling baru ini adalah *Transformer*. Dalam kasus ini, penelitian akan menggunakan algoritma *Long Short Term Memory*. Alasan penggunaan algoritma ini adalah karena *Transformer* masih tergolong teknologi baru, diperkenalkan oleh *Google Brain* pada 2017 (Vaswani et al., 2017). LSTM bisa dibilang adalah algoritma terbaik setelah *Transformer* dalam bidang NLG.

Setelah menghasilkan sebuah keluaran, sistem NLG ini dapat dites performanya untuk mengetahui seberapa bagus keluaran dari sistem tersebut. Ini disebut dengan evaluasi. Beberapa teknik evaluasi di antaranya; (i) Memberikan teks keluaran kepada seseorang, dan memintanya untuk menilai kualitas dari teks tersebut (Dethlefs et al., 2021), (ii) Dalam kasus pembuatan laporan, keluaran dari teks diberikan kepada

seorang ahli, dan ia akan menilai seberapa bergunakah teks tersebut dalam membantu menyelesaikan kepentingan mereka (Portet et al., 2009), (iii) Menggunakan *metric*, dengan membandingkan teks hasil generasi dengan teks yang ditulis oleh manusia (yang dimana memiliki masukan data yang sama), sebuah algoritma bisa digunakan untuk menilai seberapa bagus atau akurat keluaran dari sistem tersebut (Lewis et al., 2019; Novikova et al., 2017).

LinkedIn adalah salah satu laman web paling terkenal yang digunakan oleh masyarakat zaman sekarang. Laman ini memudahkan orang-orang yang sedang mencari kerja, atau kepada orang yang sudah berada di dunia kerja, untuk mendapatkan koneksi yang mungkin akan berguna di masa depan mereka kelak (Hanna, 2022). Umumnya, cara masyarakat menggunakan laman ini adalah dengan membuat sebuah kiriman yang berisi beberapa paragraf yang mendetail kan tentang apa yang mereka lakukan pada kegiatan mereka sehari-hari.

Umumnya kiriman ini dibuat se-profesional mungkin, untuk menarik perhatian orang yang bekerja pada *Human Resource* (HR), sehingga meningkatkan kesempatan penggunaanya untuk mendapatkan kerja. HR sering kali dapat menarik kesimpulan terhadap sifat dan kemampuan seorang-seorang berdasarkan kirimannya pada LinkedIn (Caers and Castelyns, 2011; Van de Ven et al., 2017). Dalam praktis nya, hampir semua kiriman yang dibuat oleh pengguna laman ini bisa dibbilang cukup

“palsu” (Guillory and Hancock, 2012). Mayoritas dari kiriman yang berada di laman ini didesain untuk membuat HR kagum, sehingga seringkali kegiatan yang biasa saja, akan dijelaskan dengan kata-kata sulit dan susah dimengerti. Kiriman yang dibuat juga biasa memiliki konten yang disesuaikan untuk HR, sehingga seseorang dapat memiliki sifat yang berbeda tergantung dari perusahaan yang dituju (Zide et al., 2014). Beberapa orang sadar bahwa kultur ini tidaklah sehat, tetapi jika mereka ingin bersaing di dunia kerja, mereka mau tidak mau harus membuat sebuah kiriman (van Zoonen et al., 2018).

Penelitian ini bertujuan untuk membuat sebuah sistem NLG menggunakan algoritma LSTM untuk menghasilkan beberapa kalimat atau paragraf yang terkesan profesional, sesuai dari masukan oleh pengguna, dan diharapkan memiliki hasil yang cukup bagus hingga memiliki potensial untuk dikirim ke laman web LinkedIn.

1.3 Rumusan Masalah

Berdasarkan kasus yang telah diterangkan pada latar belakang diatas, ditemukan beberapa masalah yang akan dijawab pada penelitian ini.

1. Bagaimana cara membuat sebuah *text generator* yang dapat menghasilkan teks yang mirip buatan manusia?

2. Bagaimana performa algoritma LSTM dalam pembuatan sebuah model *text generation*?

1.4 Tujuan Penelitian

Tujuan penelitian ini di antaranya adalah:

1. Menghasilkan sebuah model pembelajaran mesin untuk membuat sebuah teks.
2. Mengetahui apa yang saja parameter untuk memperbagus prediksi kata dari *text generator*, dan mengaplikasikan pengetahuan tersebut untuk meningkatkan performa.

1.5 Manfaat Penelitian

Manfaat penelitian ini di antaranya adalah:

1. Hasil program dapat dikembangkan agar bisa digunakan oleh masyarakat umum.
2. Program yang sudah jadi akan bersumber terbuka sehingga dapat menjadi rujukan oleh pihak lain untuk kasus mereka masing-masing.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini di antaranya adalah:

1. Karena difokuskan kepada LinkedIn, data bisa didapatkan dengan *scraping* akun LinkedIn secara manual, mengambil dataset berisi kiriman LinkedIn yang sudah tersedia di Internet, atau mengambil dataset NLP lain yang sudah pernah digunakan untuk kasus umum.
2. Hasil *text generation* dalam Bahasa Inggris karena LinkedIn umumnya menggunakan Bahasa Inggris.
3. Pembuatan model akan ditulis dengan bahasa pemrograman Python. Untuk memudahkan penggunaan, HTML akan digunakan sebagai *frontend*.

1.7 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Dalam bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Dalam bab ini akan dijelaskan dasar teori yang digunakan, definisi dari hal teknis yang terkait, seperti jenis pengembangan perangkat lunak, penjelasan algoritma yang digunakan, desain model, dan sebagainya.

BAB III. METODOLOGI PENELITIAN

Dalam bab ini dijelaskan mengenai tahapan yang akan dilaksanakan untuk menyelesaikan penelitian ini. Rencana tahapan pengerjaan akan dijelaskan dengan rinci, dan bab ini juga berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Dalam bab ini akan dijelaskan proses dari pengembangan perangkat lunak, dari pengumpulan data, hingga program berhasil *dideploy*..

BAB V. HASIL DAN ANALISIS PENELITIAN

Dalam bab ini akan dijelaskan tentang apa saja hasil yang didapatkan pada penelitian ini dan analisis mengapa didapatkan hasil tersebut.

BAB VI. KESIMPULAN DAN SARAN

Dalam bab ini akan diberikan kesimpulan tentang hasil penelitian dan saran untuk penelitian selanjutnya.

1.8 Kesimpulan

Penelitian yang akan dilakukan adalah pembuatan sebuah model *text generation* menggunakan algoritma LSTM yang akan digunakan untuk membuat beberapa

paragraf yang terkesan profesional. Hasil ini lalu dapat dijadikan sebuah kiriman untuk laman web LinkedIn. Diharapkan program ini nantinya dapat digunakan kembali untuk keperluan lain.

DAFTAR PUSTAKA

- Bridle, J.S., 1990. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition, in: Soulié, F.F., Héroult, J. (Eds.), *Neurocomputing*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 227–236. https://doi.org/10.1007/978-3-642-76153-9_28
- Caers, R., Castelyns, V., 2011. LinkedIn and Facebook in Belgium: The Influences and Biases of Social Network Sites in Recruitment and Selection Procedures. *Soc. Sci. Comput. Rev.* 29, 437–448. <https://doi.org/10.1177/0894439310386567>
- Choong, A.C.H., Lee, N.K., 2017. Evaluation of convolutionary neural networks modeling of DNA sequences using ordinal versus one-hot encoding method, in: 2017 International Conference on Computer and Drone Applications (IConDA). Presented at the 2017 International Conference on Computer and Drone Applications (IConDA), IEEE, Kuching, pp. 60–65. <https://doi.org/10.1109/ICONDA.2017.8270400>
- Dethlefs, N., Schoene, A., Cuayáhuatl, H., 2021. A divide-and-conquer approach to neural natural language generation from structured data. *Neurocomputing* 433, 300–309. <https://doi.org/10.1016/j.neucom.2020.12.083>
- Dwarampudi, M., Reddy, N.V.S., 2019. Effects of padding on LSTMs and CNNs. <https://doi.org/10.48550/ARXIV.1903.07288>
- Gagniuc, P.A., 2017. *Markov Chains: From Theory to Implementation and Experimentation*. John Wiley & Sons, Inc., Hoboken, NJ, USA. <https://doi.org/10.1002/9781119387596>
- Gerhana, Y.A., Atmadja, A.R., Zulfikar, W.B., Ashanti, N., 2017. The implementation of K-nearest neighbor algorithm in case-based reasoning model for forming automatic answer identity and searching answer similarity of algorithm case, in: 2017 5th International Conference on Cyber and IT Service Management (CITSM). Presented at the 2017 5th International Conference on Cyber and IT Service Management (CITSM), IEEE, Denpasar, Bali, Indonesia, pp. 1–5. <https://doi.org/10.1109/CITSM.2017.8089233>
- Goldberg, E., Driedger, N., Kittredge, R.I., 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert* 9, 45–53. <https://doi.org/10.1109/64.294135>

- Graves, A., 2012. Long Short-Term Memory, in: Supervised Sequence Labelling with Recurrent Neural Networks, Studies in Computational Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 37–45. https://doi.org/10.1007/978-3-642-24797-2_4
- Grefenstette, G., 1999. Tokenization, in: van Halteren, H. (Ed.), Syntactic Wordclass Tagging, Text, Speech and Language Technology. Springer Netherlands, Dordrecht, pp. 117–133. https://doi.org/10.1007/978-94-015-9273-4_9
- Guillory, J., Hancock, J.T., 2012. The Effect of LinkedIn on Deception in Resumes. *Cyberpsychology Behav. Soc. Netw.* 15, 135–140. <https://doi.org/10.1089/cyber.2011.0389>
- Hanin, B., 2018. Which Neural Net Architectures Give Rise To Exploding and Vanishing Gradients? <https://doi.org/10.48550/ARXIV.1801.03744>
- Hanna, K.T., 2022. LinkedIn [WWW Document]. TechTarget. URL <https://www.techtarget.com/whatis/definition/LinkedIn> (accessed 11.17.22).
- Hughey, D., 2006. The Traditional Waterfall Approach [WWW Document]. *Comp. Tradit. Syst. Anal. Des. Agile Methodol.* URL <https://www.umsl.edu/~hugheyd/is6840/waterfall.html> (accessed 11.23.22).
- Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Farhadi, A., Hejrati, M., Sadeghi, M.A., Young, P., Rashtchian, C., Hockenmaier, J., Forsyth, D., 2010. Every Picture Tells a Story: Generating Sentences from Images, in: Daniilidis, K., Maragos, P., Paragios, N. (Eds.), *Computer Vision – ECCV 2010, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 15–29. https://doi.org/10.1007/978-3-642-15561-1_2
- Ireland, C., 2022. Alan Turing at 100 [WWW Document]. *Harv. Gaz.* URL <https://news.harvard.edu/gazette/story/2012/09/alan-turing-at-100/> (accessed 10.4.22).
- Jain, A.K., Jianchang Mao, Mohiuddin, K.M., 1996. Artificial neural networks: a tutorial. *Computer* 29, 31–44. <https://doi.org/10.1109/2.485891>
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. <https://doi.org/10.48550/ARXIV.1910.13461>
- Noam, B., 2022. How to Check the Accuracy of Your Machine Learning Model [WWW Document]. *Check Accuracy Your Mach. Learn. Model.* URL <https://deepchecks.com/how-to-check-the-accuracy-of-your-machine-learning-model/> (accessed 11.23.22).

- Novikova, J., Dušek, O., Curry, A.C., Rieser, V., 2017. Why We Need New Evaluation Metrics for NLG. <https://doi.org/10.48550/ARXIV.1707.06875>
- Olah, C., 2015. Understanding LSTM Networks [WWW Document]. [Htppscolahgithubioposts2015-08-Underst.-LSTMs.](https://colah.github.io/posts/2015-08-Understanding-LSTMs/) URL <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed 11.23.22).
- Pawade, D., Sakhapara, A., Jain, M., Jain, N., Gada, K., 2018. Story Scrambler - Automatic Text Generation Using Word Level RNN-LSTM. *Int. J. Inf. Technol. Comput. Sci.* 10, 44–53. <https://doi.org/10.5815/ijitcs.2018.06.05>
- Perera, R., Nand, P., 2017. Recent Advances in Natural Language Generation: A Survey and Classification of the Empirical Literature. *Comput. Inform.* 36, 1–32. https://doi.org/10.4149/cai_2017_1_1
- Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripatha, S., Freer, Y., Sykes, C., 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artif. Intell.* 173, 789–816. <https://doi.org/10.1016/j.artint.2008.12.002>
- Reiter, E., Dale, R., 1997. Building applied natural language generation systems. *Nat. Lang. Eng.* 3, 57–87. <https://doi.org/10.1017/S1351324997001502>
- Rosenberg, D., 2014. Stop, Words. *Representations* 127, 83–92. <https://doi.org/10.1525/rep.2014.127.1.83>
- Santhanam, S., 2020. Context based Text-generation using LSTM networks. <https://doi.org/10.48550/ARXIV.2005.00048>
- Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Netw.* 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Singh, A., Ramasubramanian, K., Shivam, S., 2019. Natural Language Processing, Understanding, and Generation, in: *Building an Enterprise Chatbot*. Apress, Berkeley, CA, pp. 71–76. https://doi.org/10.1007/978-1-4842-5034-1_5
- van de Ven, N., Bogaert, A., Serlie, A., Brandt, M.J., Denissen, J.J.A., 2017. Personality perception based on LinkedIn profiles. *J. Manag. Psychol.* 32, 418–429. <https://doi.org/10.1108/JMP-07-2016-0220>
- Van Houdt, G., Mosquera, C., Nápoles, G., 2020. A review on the long short-term memory model. *Artif. Intell. Rev.* 53, 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>
- van Zoonen, W., Bartels, J., van Prooijen, A.-M., Schouten, A.P., 2018. Explaining online ambassadorship behaviors on Facebook and LinkedIn. *Comput. Hum. Behav.* 87, 354–362. <https://doi.org/10.1016/j.chb.2018.05.031>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. <https://doi.org/10.48550/ARXIV.1706.03762>

- Wang, S.-C., 2003. Artificial Neural Network, in: *Interdisciplinary Computing in Java Programming*. Springer US, Boston, MA, pp. 81–100. https://doi.org/10.1007/978-1-4615-0377-4_5
- Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. *Proc. IEEE* 78, 1550–1560. <https://doi.org/10.1109/5.58337>
- Zhong Su, Qiang Yang, Ye Lu, Hongjiang Zhang, 2000. WhatNext: a prediction system for Web requests using n-gram sequence models, in: *Proceedings of the First International Conference on Web Information Systems Engineering*. Presented at the WISE 2000: 1st International Conference on Web Information Systems Engineering, IEEE Comput. Soc, Hong Kong, China, pp. 214–221. <https://doi.org/10.1109/WISE.2000.882395>
- Zide, J., Elman, B., Shahani-Denning, C., 2014. LinkedIn and recruitment: how profiles differ across occupations. *Empl. Relat.* 36, 583–604. <https://doi.org/10.1108/ER-07-2013-0086>