



**Penguatan Pengambilan Keputusan Berbasis Data
dengan Model Pembelajaran Mesin dan Model Bahasa Besar
di Era Ekonomi Digital Indonesia**

Hendri Murfi

**Pidato pada Upacara Pengukuhan sebagai
Guru Besar Tetap Dalam Bidang *Advanced Data Analytics*
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Indonesia**

Depok, 8 Januari 2025

Daftar Isi

1	Pendahuluan.....	5
1.1	Ilmu Data.....	6
1.2	Pembelajaran Mesin.....	8
2	Analitik Data Aktuaria.....	10
3	Analitik Data Teks.....	15
3.1	Perangkat.....	17
3.2	Kerangka Umum.....	20
3.3	Representasi Teks.....	21
3.4	Model Bahasa Besar.....	24
3.5	<i>Generative AI</i>	28
4	Penutup.....	32
5	Ucapan Terima Kasih.....	33
6	Referensi.....	37
7	Riwayat Hidup.....	40

Assalamu'alaikum Warahmatullahi Wabarakatuh, Selamat pagi dan salam sejahtera untuk kita semua.

Yang terhormat,

- Menteri Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia
- Ketua, Sekretaris dan Anggota Majelis Wali Amanat Universitas Indonesia,
- Rektor, Wakil Rektor, dan Kepala Badan Universitas Indonesia
- Ketua, Sekretaris, dan Anggota Dewan Guru Besar Universitas Indonesia,
- Ketua, Sekretaris, dan Anggota Senat Akademik Universitas Indonesia,
- Dekan dan Pimpinan Sekolah di lingkungan Universitas Indonesia,
- Ketua dan Anggota Tim Penilai Angka Kredit Universitas Indonesia,
- Dekan, Wakil Dekan, dan Pimpinan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia
- Ketua, Sekretaris, dan Anggota Dewan Guru Besar Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia,
- Ketua, Sekretaris, dan Anggota Senat Akademik Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia,
- Guru Besar Tamu, Mitra Undangan
- Teman sejawat Dosen, Keluarga, Kerabat, dan Seluruh Hadirin yang saya muliakan.

Pada kesempatan yang berbahagia ini, perkenankan saya mengajak hadirin sekalian untuk memanjatkan puji dan syukur kepada Allah SWT atas berkat rahmatnya kita semua berkesempatan hadir pada acara Pengukuhan Guru Besar Tetap Universitas Indonesia. Tidak lupa, shalawat dan salam kita sampaikan kepada Rasulullah Muhammad SAW.

Hadirin yang saya muliakan,

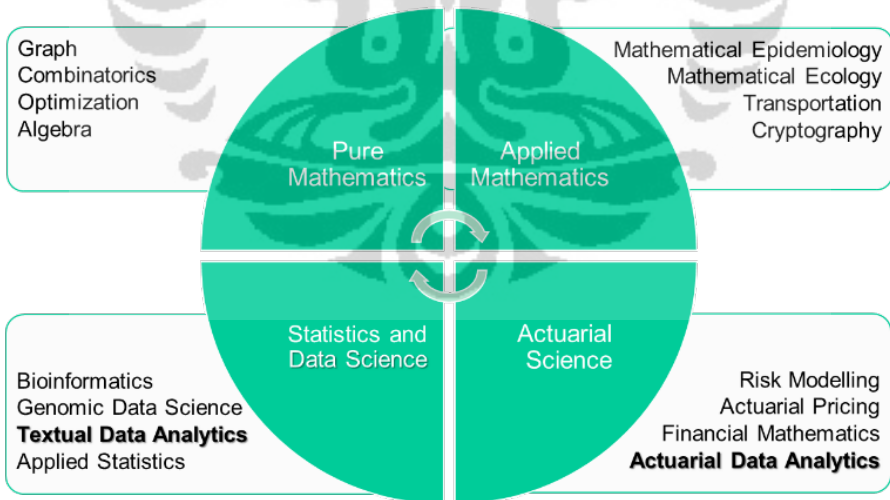
Izinkanlah saya menyampaikan pidato pengukuhan saya sebagai Guru Besar Tetap dalam bidang *Advanced Data Analytics* pada Departemen Matematika FMIPA Universitas Indonesia yang berjudul:

“Penguatan Pengambilan Keputusan Berbasis Data dengan Model Pembelajaran Mesin dan Model Bahasa Besar di Era Ekonomi Digital Indonesia”

1 Pendahuluan

Saat ini, dunia Industri terus mengalami perubahan secara besar-besaran terutama ketergantungan pada perkembangan pesat teknologi. Perubahan ini melalui perpaduan teknologi yang mengurangi sekat antara dunia fisik, digital, dan biologi. Dengan tantangan dan perubahan keadaan industri, sistem pendidikan tinggi perlu adaptif, yaitu sistem pendidikan tinggi yang menghubungkan dunia yang berubah dengan proses pembelajaran dan penelitian.

Universitas Indonesia mengambil peran sebagai *Leading University* dalam menjawab tantangan di atas. Sebagai bagian dari Universitas Indonesia, Departemen Matematika FMIPA Universitas Indonesia mengambil peran yang sama dengan Universitas Indonesia. Apalagi perkembangan bidang Matematika menjadi salah satu bidang pendukung utama dari perkembangan pesat pada teknologi tersebut.

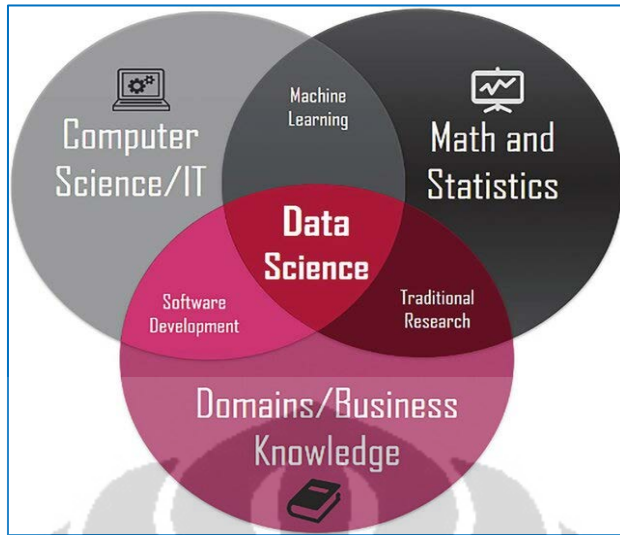


Gambar 1. Bidang Keilmuan Departemen Matematika FMIPA Universitas Indonesia

Saat ini, untuk lebih memenuhi kebutuhan industri, beberapa Bidang Keilmuan pada Departemen Matematika FMIPA Universitas Indonesia telah menjadi Program Studi, yaitu Program Studi Sarjana Statistika dan Program Studi Sarjana Ilmu Aktuaria (Gambar 1). Bahkan, Program Studi Sarjana Ilmu Aktuaria didirikan atas dorongan dari pihak Industri untuk memenuhi kebutuhan profesional Aktuaris di Indonesia. Selain itu, *Data Analysis* dan *Data Scientist* adalah bidang pekerjaan dengan permintaan yang paling meningkat berdasarkan *Future of Jobs Reports 2020* oleh *World Economic Forum*, termasuk juga di Indonesia.

1.1 Ilmu Data

Ilmu Data (*Data Science*) adalah suatu Bidang Keilmuan interdisipliner untuk mengekstraksi wawasan dan pengetahuan dari data yang memadukan berbagai disiplin ilmu seperti matematika, statistika, ilmu komputer, dan pengetahuan bisnis atau domain tertentu (Gambar 2). Pada dasarnya, Ilmu Data memanfaatkan berbagai metode, baik statistik tradisional maupun Pembelajaran Mesin (*Machine Learning*), untuk mengidentifikasi pola, tren, dan wawasan dari data yang dapat berupa data terstruktur atau tabular, seperti angka-angka dalam tabel, atau data tidak terstruktur, seperti teks dari media sosial.



Gambar 2. Bidang Keilmuan Interdisipliner Ilmu Data

Salah satu publikasi yang sering disebutkan sebagai publikasi formal awal terkait dengan istilah Ilmu Data adalah artikel oleh William S. Cleveland pada tahun 2001 yang berjudul "*Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics.*" Artikel ini mengusulkan agar Ilmu Data dipandang sebagai bidang baru yang melampaui statistik tradisional, mencakup pemrosesan data dalam skala besar, visualisasi, dan pemahaman data dari berbagai sumber (Cleveland, 2001).

Popularitas Ilmu Data mulai meningkat pesat pada dekade 2010-an, ketika ledakan data digital dan kemajuan dalam teknologi komputasi membuat analisis data menjadi kebutuhan mendesak di berbagai industri. Istilah ini semakin populer di kalangan praktisi dan akademisi, dan banyak perusahaan teknologi serta lembaga pendidikan mulai membuat posisi dan program khusus untuk Ilmu Data.

Program Studi Sarjana Ilmu Aktuaria, Departemen Matematika FMIPA Universitas Indonesia telah membuka mata kuliah Ilmu Data sebagai mata kuliah pilihan pada tahun 2017, dan kemudian menjadi mata kuliah wajib untuk semua Program Studi Sarjana di Departemen Matematika FMIPA Universitas Indonesia pada tahun 2020.

Theory and Algorithmics-based Methods	
<input type="checkbox"/> Linear Regression	<input type="checkbox"/> Decision Tree
<input type="checkbox"/> Logistic Regression	<input type="checkbox"/> Ensemble Learning (Random Forest, Gradient Boosting)
<input type="checkbox"/> Naïve Bayes	<input type="checkbox"/> Neural Networks
<input type="checkbox"/> Support Vector Machine	<input type="checkbox"/> Deep Learning (Hopfield, Boltzmann, CNN, LSTM, GAN, Transformers)

Gambar 3. Metode-Metode Pembelajaran Mesin

1.2 Pembelajaran Mesin

Pembelajaran Mesin (*Machine Learning*) merupakan salah satu model utama dalam Ilmu Data, yaitu model yang mencoba mempelajari pengetahuan pada data sehingga memiliki kemampuan prediktif untuk data baru (Gambar 3). Dalam sejarahnya, Pembelajaran Mesin mulai berkembang dengan model-model statistik dasar yang diinspirasi oleh teori Probabilitas dan Statistika. Bahkan, model *Decision Tree* dan *Ensemble Learning*, yaitu model populer yang dikembangkan secara algoritmik saat ini, memiliki awal sejarah pengembangan pada bidang Statistika.

Kemajuan besar dan popularitas model Pembelajaran Mesin datang dari *Neural Networks*, yaitu model yang meniru cara kerja otak manusia dalam memproses informasi. Salah satu tokoh berpengaruh di bidang ini adalah John Hopfield, yang dikenal atas kontribusinya dalam mengembangkan model *Hopfield Network*, yaitu model *Neural Network* awal yang dapat menyimpan dan mengingat pola, menyerupai memori asosiatif manusia (Hopfield, 1982). Tokoh lainnya adalah Geoffrey Hinton, yang dianggap sebagai pelopor dalam *Deep Learning* dengan mengembangkan model *Boltzmann Machine* (Hinton & Sejnowski, 1983). Atas kontribusinya, Hinton dan Hopfield dianugerahi Hadiah Nobel pada tahun 2024¹.

Penghargaan Nobel yang diberikan kepada para pionir seperti Hinton dan Hopfield ini, tidak hanya mengakui kerja keras mereka, tetapi juga menunjukkan dampak luas dari model Pembelajaran Mesin terhadap sains, teknologi, dan kehidupan manusia secara keseluruhan.

¹ <https://www.nobelprize.org/prizes/physics/2024/press-release/>

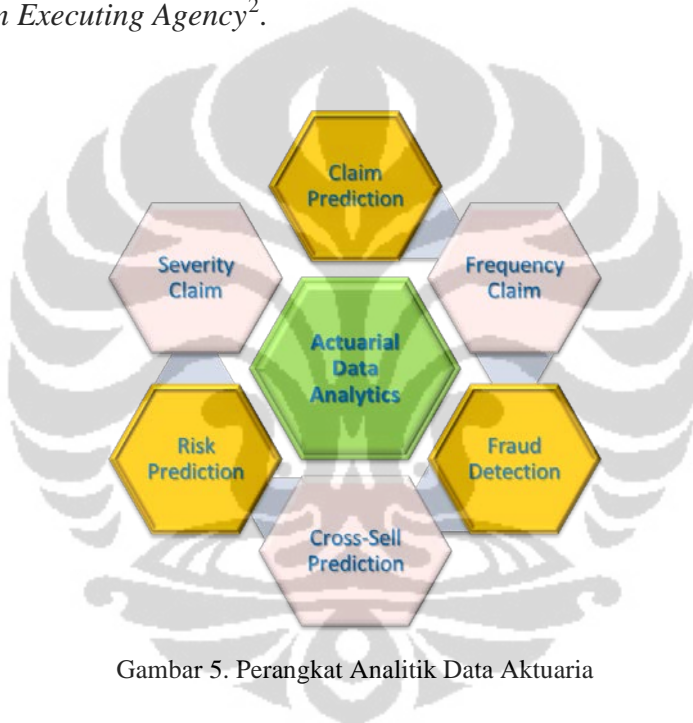
2 Analitik Data Aktuaria

Salah satu fokus penelitian kami adalah Analitik Data Aktuaria (*Actuarial Data Analytics*) yang menerapkan model-model Pembelajaran Mesin pada masalah terkait Ilmu Aktuaria (*Actuarial Science*). Ilmu Aktuaria adalah ilmu yang mengombinasikan matematika, statistika, keuangan, dan manajemen risiko untuk mengukur, memodelkan, dan mengelola risiko. Sehingga, Ilmu Aktuaria memainkan peran penting dalam membantu perusahaan pada berbagai bidang bisnis, seperti asuransi, perbankan, keuangan, dan institusi lainnya. Akan tetapi, serapan tenaga profesional Aktuaris masih didominasi oleh industri asuransi seiring peningkatan penetrasi asuransi di Indonesia. Data dari Otoritas Jasa Keuangan (OJK) menunjukkan bahwa industri asuransi di Indonesia memperlihatkan pertumbuhan yang signifikan dalam beberapa tahun terakhir (Keuangan, 2023).



Gambar 4. Sosialisasi Program Studi Ilmu Aktuaria yang didukung oleh READI *Project*

Ilmu Aktuaria adalah disiplin ilmu yang relatif baru di Indonesia. Program Studi Sarjana Ilmu Aktuaria FMIPA Universitas Indonesia berdiri pada tahun 2017. Sebelumnya, bidang ini hanya merupakan bidang peminatan pada Jurusan Matematika FMIPA Universitas Indonesia. Pendirian Program Studi merupakan dorongan dari industri asuransi yang kemudian membentuk *READI Project* yang didukung oleh Pemerintah Kanada melalui *Global Affairs Canada* dan Universitas Waterloo sebagai *Canadian Executing Agency*².



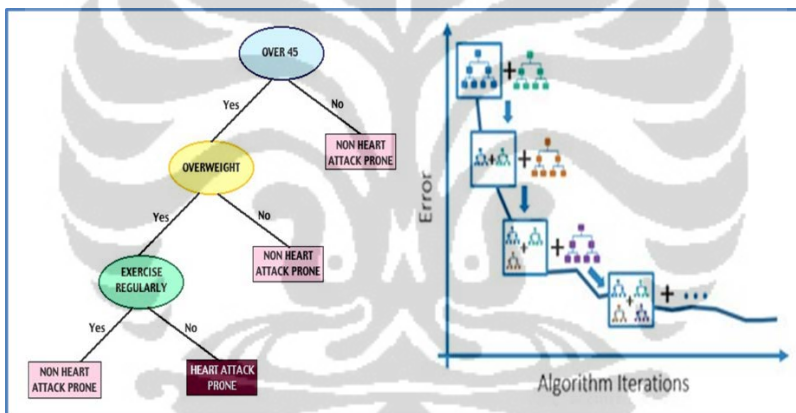
Gambar 5. Perangkat Analitik Data Aktuaria

Peran model Pembelajaran Mesin menjadi sangat relevan dalam Ilmu Aktuaria seperti pada industri asuransi. Model Pembelajaran Mesin memberikan kemampuan bagi para aktuaris untuk menganalisis data secara otomatis dan mendalam, terutama untuk data tabular, seperti data

² <https://uwaterloo.ca/risk-management-economic-sustainability-actuarial-science-development-indonesia/>

risiko dan klaim (Gambar 5), yang sering ditemui dalam industri asuransi. Integrasi Pembelajaran Mesin ke dalam industri asuransi memungkinkan membentuk masa depan analitik data yang lebih adaptif dan responsif terhadap kebutuhan industri dan masyarakat.

Dalam Ilmu Aktuaria, model statistik seperti *Generalized Linear Models* (GLM) terbukti sangat andal dan sudah lama digunakan untuk analisis data asuransi, terutama dalam memprediksi risiko klaim dan penetapan premi. GLM memberikan hasil yang stabil dengan penjelasan yang jelas mengenai variabel risiko, menjadikannya metode standar yang sering digunakan di industri asuransi.



Gambar 6. Model *Decision Tree* (Kiri) dan Model *Gradient Boosting* (Kanan)

Namun, dalam beberapa tahun terakhir, penelitian menunjukkan bahwa model Pembelajaran Mesin mampu memberikan hasil yang setara, bahkan lebih baik, dalam banyak kasus dibandingkan model tradisional seperti GLM. Misalnya, studi kami menunjukkan bahwa model *ensemble* berbasis model *Decision Tree*, seperti model *Gradient Boosting* (Gambar 6), memberikan akurasi yang lebih baik dalam memprediksi klaim dan

risiko (Fauzan & Murfi, 2018; Rusdah & Murfi, 2020). Termasuk juga jika dibandingkan dengan model yang merupakan adaptasi model *deep learning* terkini untuk data tabular (Tjahjono et al., 2024).

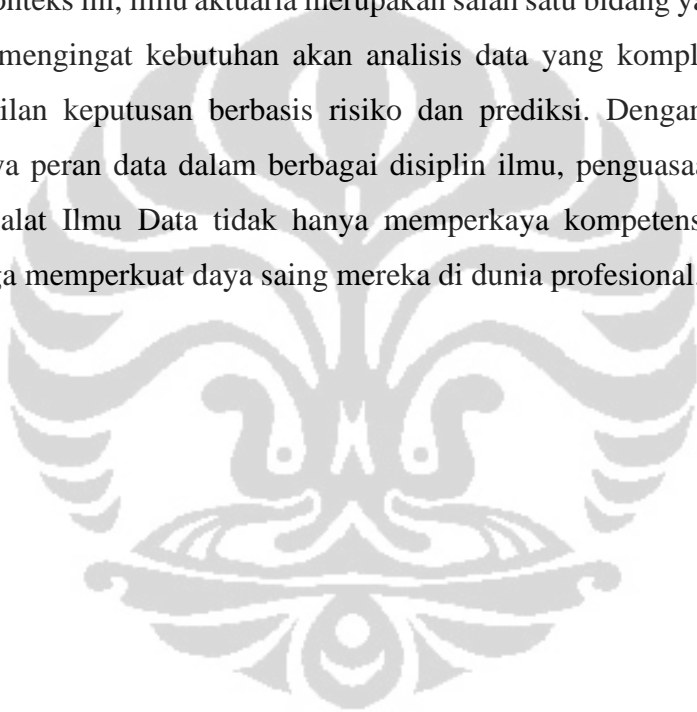
Akan tetapi, untuk penetapan tarif asuransi, model sangat diatur dan harus memenuhi persyaratan khusus sebelum diterapkan dalam praktik, sehingga menimbulkan beberapa tantangan bagi model Pembelajaran Mesin. Salah satu persyaratan tersebut adalah bahwa model penetapan tarif asuransi harus transparan dan mudah dikomunikasikan kepada semua pemangku kepentingan (Kaminski, 2019). Model *ensemble* memberikan kinerja prediktif yang lebih baik tetapi dengan mengorbankan aspek transparansi model. Pada tahun 2021, Henckaerts et. al. berhasil menggunakan perangkat visualisasi untuk menginterpretasikan model yang dihasilkan dan proses pengambilan keputusan yang mendasarinya (*explainability*) (Henckaerts et al., 2021). Temuan ini menunjukkan bahwa model Pembelajaran Mesin dapat menjadi alternatif atau pelengkap bagi model tradisional seperti GLM, terutama dalam lingkungan bisnis yang membutuhkan prediksi risiko yang sangat akurat.

Society of Actuaries (SOA), sebagai salah satu lembaga utama dalam sertifikasi profesi Aktuaris di dunia, menyadari potensi model Pembelajaran Mesin dalam mendukung profesi ini. Pada tahun 2022, SOA menambahkan mata uji untuk pengetahuan mendalam tentang teknik Ilmu Data termasuk model Pembelajaran Mesin yang lebih kompleks. Pada tahun 2024, Persatuan Aktuaris Indonesia (PAI) telah juga mulai memasukan Ilmu Data dan Pembelajaran Mesin pada ujian profesi mereka.

Kondisi ini mencerminkan tuntutan industri yang semakin membutuhkan keterampilan di bidang Pembelajaran Mesin dan

menegaskan pentingnya teknologi ini dalam Ilmu Aktuaria modern. Lembaga profesi tersebut ingin memastikan bahwa para Aktuaris masa depan memiliki keterampilan yang relevan untuk menghadapi tantangan analitik data yang semakin kompleks.

Ilmu Data, seperti halnya Statistika, berpotensi menjadi mata kuliah inti atau setidaknya mata kuliah pilihan yang esensial bagi berbagai Program Studi, terutama yang secara intensif berinteraksi dengan data. Dalam konteks ini, ilmu aktuaria merupakan salah satu bidang yang sangat relevan, mengingat kebutuhan akan analisis data yang kompleks untuk pengambilan keputusan berbasis risiko dan prediksi. Dengan semakin meluasnya peran data dalam berbagai disiplin ilmu, penguasaan konsep dan alat-alat Ilmu Data tidak hanya memperkaya kompetensi lulusan, tetapi juga memperkuat daya saing mereka di dunia profesional.



3 Analitik Data Teks

Fokus penelitian kami lainnya adalah Analitik Data Teks (*Textual Data Analytics*) yang menerapkan model-model Pembelajaran Mesin pada data teks. Selain data tabular, data teks telah menjadi sumber daya penting bagi industri dan bisnis di berbagai sektor. Pertumbuhan data teks dalam industri dan bisnis telah meningkat secara eksponensial seiring dengan transformasi digital dan adopsi teknologi berbasis data. Komunikasi pelanggan melalui email, chat, dan ulasan produk menciptakan aliran data teks yang kaya akan wawasan. Selain itu, dokumen bisnis seperti laporan, kontrak, dan catatan operasional juga berkontribusi pada akumulasi data teks yang besar. Data teks ini dapat dimanfaatkan untuk mengidentifikasi pola, memahami kebutuhan pelanggan, meningkatkan pengalaman pengguna, serta mendukung pengambilan keputusan strategis.

Pertumbuhan data teks telah juga mengalami lonjakan signifikan seiring dengan pesatnya perkembangan media sosial. Platform-platform seperti Facebook, Twitter, Instagram, dan TikTok telah menjadi pusat aktivitas digital yang menghasilkan miliaran konten berbasis teks setiap harinya, mulai dari status, tweet, hingga komentar dan ulasan. Fenomena ini didorong oleh aksesibilitas internet yang semakin luas dan perangkat pintar yang memungkinkan pengguna untuk berinteraksi secara *real-time*. Dengan pengguna aktif yang mencapai miliaran orang, media sosial tidak hanya mencerminkan kehidupan sosial manusia modern tetapi juga menciptakan ekosistem data teks yang melimpah, kompleks, dan beragam.

Seiring dengan pertumbuhan media sosial, *Social Media Monitoring* telah berkembang menjadi lini bisnis di Indonesia. Beberapa contoh perusahaan yang bergerak di bidang ini, antara lain Media Wave

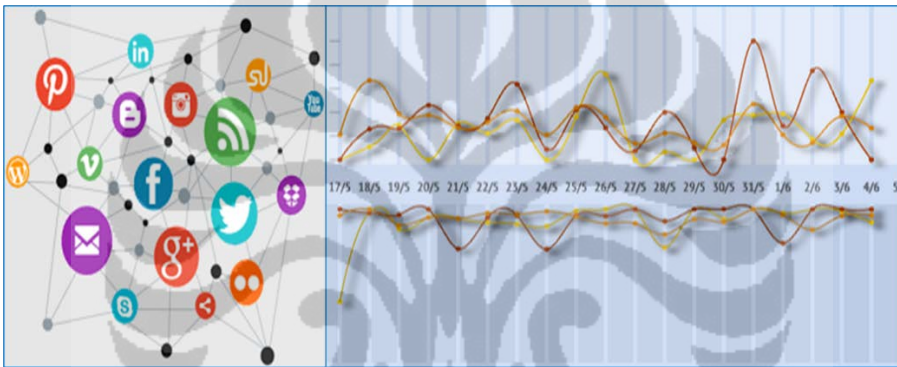
(mediawave.id), Kezee (kazee.id), Indonesia Indikator (indonesiaindicator.com), No Limit (nolimit.id), dll. *Social Media Monitoring* memberikan perusahaan kemampuan untuk melakukan analisis *real-time* terhadap opini konsumen dan merespons kebutuhan pelanggan dengan cepat, mengidentifikasi tren yang sedang berlangsung, atau mengukur efektivitas kampanye pemasaran.



Gambar 7. Perangkat Analitik Data Teks

Pertumbuhan data teks ini memberikan peluang besar bagi model-model Pembelajaran Mesin sebagai perangkat Analitik Data Teks. Beberapa contoh perangkat Analitik Data Teks berbasis model Pembelajaran Mesin adalah Analisis Sentimen (*Sentimen Analysis*) dan Pendeteksian Topik (*Topic Detection*) (Gambar 7). Pada dasarnya, penentuan sentimen dan topik pada teks adalah salah satu kemampuan dari kecerdasan manusia. Akan tetapi, manusia memiliki beberapa kelemahan dan keterbatasan, sehingga membuat model Pembelajaran Mesin mejadi

solusi praktis. Setelah kita melatih model Pembelajaran Mesin untuk kemampuan ini dan jika model tersebut memiliki kemampuan setara manusia atau mencapai level kemampuan yang dapat diterima, maka model Pembelajaran Mesin ini dapat menggantikan peran manusia untuk menentukan sentimen dan topik pada data teks. Model Pembelajaran Mesin memungkinkan penentuan sentimen dan topik pada data teks yang sangat besar, secara terus-menerus, *real time*, dan kemudian menjadi dasar untuk melakukan analitik data teks.

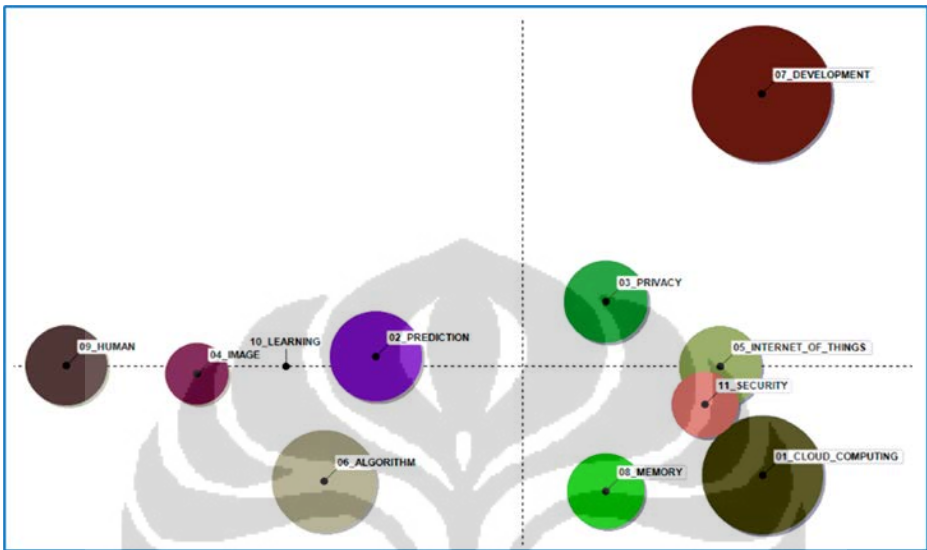


Gambar 8. Pemantauan sentimen pada media *online*

3.1 Perangkat

Salah satu perangkat Analitik Data Teks adalah Analisis Sentimen, yaitu metode yang digunakan untuk menentukan emosi atau sikap pengguna dalam teks, apakah positif, negatif, atau netral. Analisis ini dapat membantu perusahaan memahami bagaimana pelanggan memandang produk atau layanan mereka, dan apakah sentimen terhadap suatu merek cenderung meningkat atau menurun dari waktu ke waktu. Model Pembelajaran Mesin memungkinkan pemantauan sentimen pada media *online* dengan data teks yang sangat besar, secara terus-menerus, *real time*,

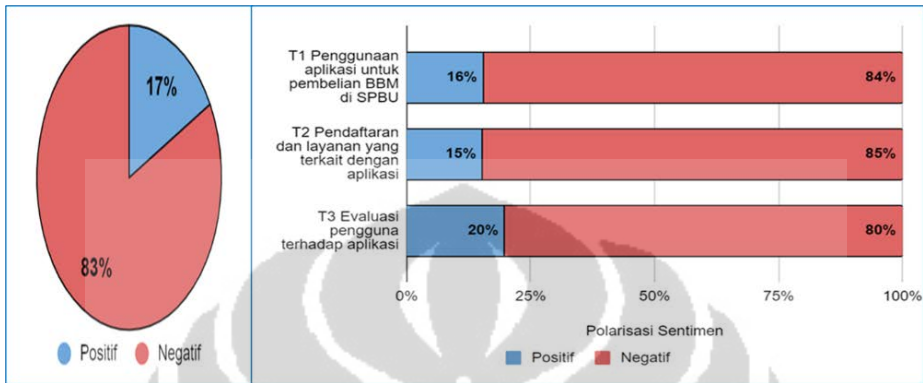
dan kemudian menjadi dasar untuk analisis lebih lanjut (Gambar 8) (Gowandi et al., 2021).



Gambar 9. Topik dari tema *big data* pada data besar dari literatur ilmiah Scopus

Selain Analisis Sentimen, Pendeteksian Topik adalah perangkat lain untuk Analitik Data Teks, yaitu metode untuk mengidentifikasi topik-topik utama dalam dataset besar. Dengan Pendeteksian Topik, perusahaan dapat mengetahui isu-isu atau topik tertentu yang sedang populer di kalangan konsumen, sehingga mereka dapat merespons dengan cepat terhadap isu-isu tersebut. Seperti juga pada Analisis Sentimen, model Pembelajaran Mesin memungkinkan pendeteksian topik pada data teks yang sangat besar, secara terus-menerus, *real time*, dan juga menjadi dasar untuk analisis lebih lanjut. Misalnya, dalam data besar dari literatur ilmiah, Pendeteksian Topik memungkinkan mendeteksi topik terkait dengan suatu tema, misal *big data* (Gambar 9). Topik tersebut dapat digunakan sebagai

referensi bagi pemangku kepentingan untuk mengembangkan penelitian lanjutan, serta untuk mengisi kekurangan dalam tinjauan literatur pada tema tersebut (Parlina et al., 2020, 2021).



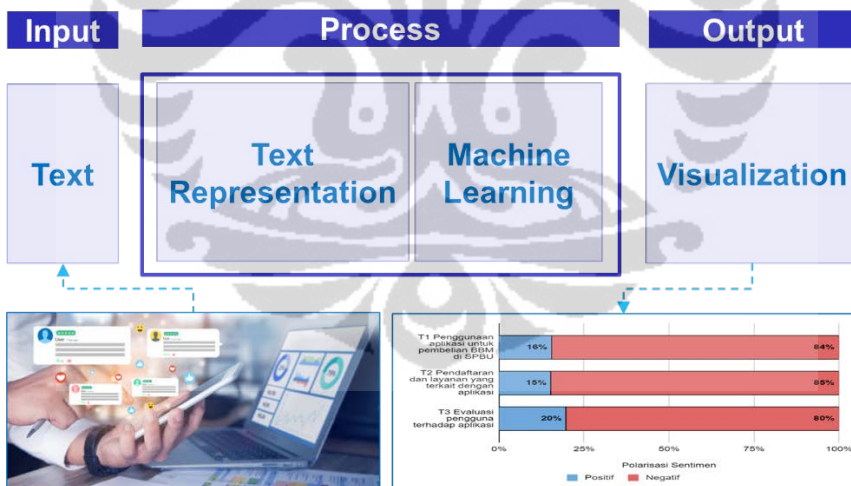
Gambar 10. Analisis Sentimen level dokumen (kiri) dan level topik (kanan)

Salah satu perkembangan terbaru dalam perangkat Analitik Data Teks adalah Analisis Sentimen Level Topik (*Topic-Level Sentiment Analysis*). Berbeda dengan Analisis Sentimen biasa yang hanya menghitung perbandingan jumlah sentimen secara keseluruhan, Analisis Sentimen Level Topik menghitung perbandingan jumlah sentimen untuk setiap topik yang dibicarakan dalam teks tersebut (Gambar 10). Analisis Sentimen Level Topik mengasumsikan bahwa suatu teks merupakan anggota dari satu kategori aspek, yang disebut sebagai topik. Analisis ini menggabungkan metodologi dari Pendeteksian Topik dan Analisis Sentimen. Proses ini melibatkan pendeteksian topik dari teks yang diberikan, menganalisis sentimen sebagai positif atau negatif, dan memberikan wawasan tentang sentimen terhadap topik yang lebih luas dalam teks yang diberikan (Pathak et al., 2021; Wijaya et al., 2024).

Pendekatan ini lebih sesuai pada banyak aplikasi karena memberikan pemahaman yang lebih mendalam mengenai fitur tertentu yang disukai atau tidak disukai oleh pengguna, sehingga lebih informatif untuk pengambilan keputusan.

3.2 Kerangka Umum

Secara umum, kerangka kerja Analitik Data Teks melibatkan beberapa tahap mulai dari input, pemrosesan, hingga visualisasi hasil (Gambar 11). Secara metodologi, tahapan pemrosesan merupakan tahapan inti yang berperan penting untuk mengubah data teks yang tidak terstruktur menjadi informasi yang bisa dianalisis dan menghasilkan wawasan. Tahapan pemrosesan mulai dari representasi data teks hingga penerapan berbagai model Pembelajaran Mesin.



Gambar 11. Kerangka umum Analitik Data Teks

Tahapan representasi teks adalah proses mengkonversi teks menjadi format yang bisa dipahami oleh model Pembelajaran Mesin, misal vektor

atau matriks. Salah satu tantangan terbesar dalam representasi teks adalah adanya aspek semantik yang ditentukan oleh kata-kata yang menyusun teks tersebut. Sementara, terdapat kata-kata yang memiliki ejaan sama tetapi memiliki makna yang berbeda (homonim atau polisemi) dan makna kata yang sesuai ditentukan oleh kata-kata lain di teks. Contoh, kata *mouse* memiliki banyak arti, yaitu hewan pengerat atau perangkat komputer. Arti kata *mouse* yang sesuai akan ditentukan oleh kata-kata lain pada suatu teks (kontektual). Pada teks *cheese-loving mouse* maka *mouse* adalah hewan pengerat. Sementara, pada teks *click on the mouse* maka *mouse* adalah perangkat komputer (Gambar 12).

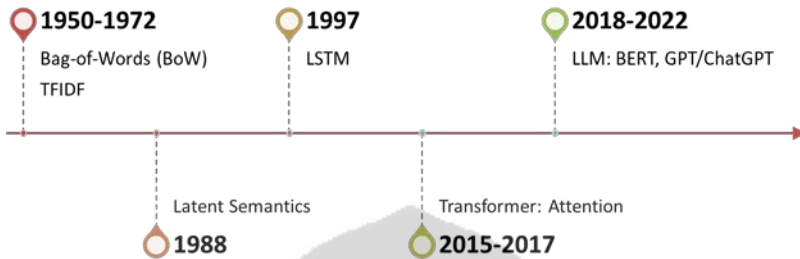


Gambar 12. Kata *mouse* memiliki arti yang berbeda tergantung konteks pada kalimat

3.3 Representasi Teks

Penelitian tentang representasi teks merupakan salah satu penelitian dengan sejarah yang sangat panjang (Gambar 13). Penelitian ini sudah dimulai pada tahun 1950 dimana teks direpresentasikan sebagai frekuensi kata tanpa memperhatikan urutan kata (*Bag-of-Word*). Representasi ini kemudian diperbarui dengan memperhatikan frekuensi kata tidak hanya dalam dokumen tapi juga dalam *corpus* (TFIDF) pada tahun 1972. Metode ini telah digunakan pada banyak masalah analitik data teks. Akan tetapi, metode ini belum memperhatikan aspek semantik pada teks. Untuk mengadopsi aspek semantik ini, penelitian berikutnya menggunakan

asumsi adanya *latent semantics* antar kata dan kemudian merepresentasikan teks berdasarkan *latent semantics* tersebut. Akan tetapi, metode ini belum dapat menangkap dinamika aspek kontekstual pada teks.

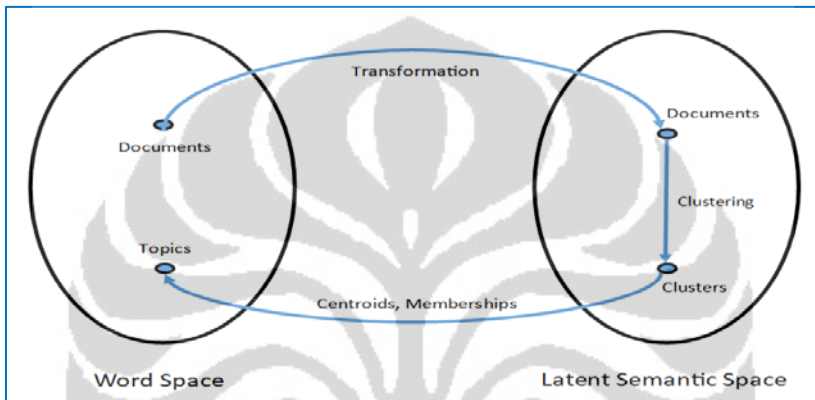


Gambar 13. Peta Jalan Penelitian Representasi Teks

Metode pertama yang dianggap berhasil merealisasi aspek kontekstual ini adalah model *Long Short Term Memory* (LSTM) pada tahun 1997 (Hochreiter & Schmidhuber, 1997). Dengan menggunakan tiga gerbang utama (*input*, *forget*, dan *output*), LSTM mampu mengingat informasi penting dan melupakan yang tidak relevan, sehingga menghasilkan representasi makna yang kaya untuk setiap kata berdasarkan konteksnya. LSTM ditemukan oleh Sepp Hochreiter yang pernah menjadi ketua bidang *Machine Learning* pada *Neural Information Processing Group* TU Berlin – grup riset tempat kami melakukan penelitian program Doktor. Saat itu, model ini dianggap belum layak secara implementasi dan juga arah penelitian lebih ke model-model berbasis teori seperti metode *Large Margin* dan *Kernel*. Sehingga, metode-metode *latent semantics* masih umum dan banyak digunakan untuk representasi data teks.

Kami telah menggunakan beberapa metode matematika untuk merealisasikan konsep *latent semantics* ini terutama pada masalah *Topic*

Detection (Gambar 14), seperti *singular value decomposition* (Muliawati & Murfi, 2017; Murfi, 2018, 2021; Nur'aini et al., 2015), *random projection* (Yusdiansyah et al., 2019), dan juga metode *Neural Networks*, yaitu *Deep Autoencoder* (Murfi et al., 2022). Selain mengakomodasi konsep semantik, metode-metode ini membuat masalah *Topic Detection* menjadi lebih efisien karena bekerja pada ruang dimensi lebih rendah.



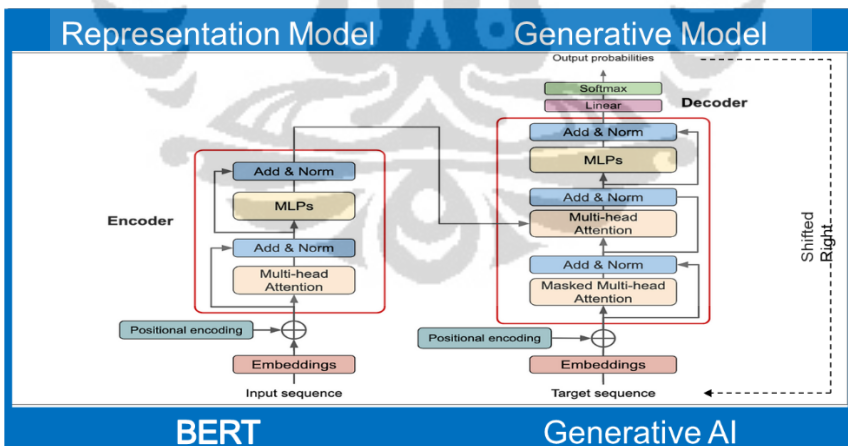
Gambar 14. Metode representasi teks *latent semantics*

Selain itu, metode *clustering* menjadi metode paling relevan sebagai metode *Topic Detection*, karena metode *clustering* memungkinkan untuk memproses data teks dengan representasi positif maupun negatif. Sementara metode-metode standar lain, seperti *Nonnegative Matrix Factorization* dan *Latent Dirichlet Allocation*, hanya sesuai untuk data teks dengan representasi non negatif saja. Sampai saat ini, metode *clustering* dapat dikatakan sebagai metode standar untuk masalah Pendeteksian Topik.

3.4 Model Bahasa Besar

Pada tahun 2010, GPU mulai populer di kalangan peneliti Pembelajaran Mesin karena kemampuan paralelismenya yang sangat baik untuk melatih model *Neural Networks*. Sebagai model berbasis *Neural Networks*, LSTM menjadi layak untuk diimplementasikan dan kemudian menjadi model yang populer. Bahkan model ini diadaptasi dan populer juga untuk banyak masalah lain diluar data teks, yaitu masalah dengan fitur data bersifat sekuensial.

Pada periode yang sama, metode representasi teks lain diajukan, yaitu *Attention* sebagai bagian dari arsitektur *Transformer* [20]. Tidak seperti LSTM yang memproses teks secara berurutan, *Attention* mungkin melihat seluruh teks secara bersamaan dan kemudian menangkap hubungan antar kata. Sehingga, model ini memberikan kinerja yang jauh lebih efisien dari model LSTM.



Gambar 15. Model Representasi Teks (*Encoder*) dan Model Generatif (*Decoder*) pada *Transformer*

Selain model representasi (*representation model*) pada bagian *Encoder*, *Transformer* juga menghasilkan inovasi baru pada Pembelajaran Mesin, yaitu model Pembelajaran Mesin dengan kemampuan generatif (*generative*) atau kemampuan membangkitkan sekumpulan kata (Gambar 15). Pada model generatif ini, kata berikutnya dibangkitkan berdasarkan nilai probabilitas tertinggi secara *autoregressive*. Dengan kata lain, kemampuan generatif ini adalah perluasan dari kemampuan prediktif dimana prediksi suatu kata berikutnya dilakukan secara sekuensial, dan setiap kata yang dihasilkan menjadi input untuk prediksi kata berikutnya.

Pada era "*scaling up*" model, yaitu era yang mengasumsikan bahwa penambahan lebih banyak data dan sumber daya komputasi akan secara konsisten menghasilkan model yang lebih baik, *Transformer* telah menjadi komponen inti dari banyak Model Bahasa Besar (*Large Language Model*), seperti *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al., 2019) untuk representasi teks dan *Generative Pre-trained Transformer* (GPT) (Radford & Narasimhan, 2018) untuk *Generative AI*. Sebagai ilustrasi, BERT adalah model *Neural Networks* yang memiliki 110 juta parameter untuk model BERT_{BASE} dan 340 juta parameter untuk model BERT_{LARGE}. Sementara, GPT memiliki 117 juta parameter untuk model GPT-1, 175 milyar untuk model GPT 3, dan diestimasi mencapai triliunan untuk model GPT 4.

Pretrained BERT telah menjadi model yang populer karena kemampuannya memahami konteks bahasa secara dua arah, sehingga menghasilkan representasi teks yang lebih kaya dan akurat dibandingkan model sebelumnya. Selain itu, model ini juga memberikan kelebihan dalam *transfer learning*, yaitu pengguna hanya menyesuaikan parameter

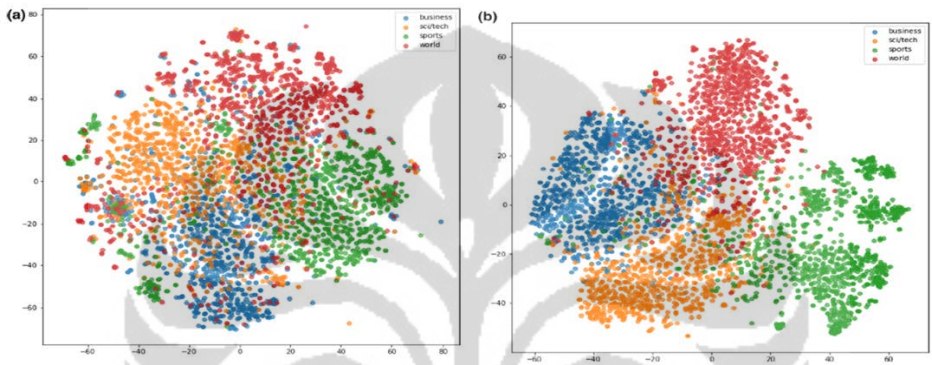
prediktor saja untuk tugas-tugas spesifik. Parameter prediktor ini umumnya jauh lebih sedikit, sehingga mengurangi waktu dan sumber daya pelatihan.

Method	AG news		
	ACC	NMI	ARI
TFIDF+KM	0.3568±0.0059	0.2135±0.0067	0.2121±0.0068
BERT+Max+I+KM	0.3018±0.0131	0.1495±0.0095	0.1479±0.0095
BERT+Max+LN+KM	0.3285±0.0140	0.1797±0.0132	0.1782±0.0132
BERT+Max+N+KM	0.3229±0.0145	0.1739±0.0137	0.1724±0.0137
BERT+Max+MM+KM	0.226±0.0058	0.088±0.0057	0.0864±0.0057
BERT+Mean+I+KM	0.357±0.0079	0.2134±0.0077	0.212±0.0077
BERT+Mean+LN+KM	0.3741±0.0057	0.2302±0.0055	0.2288±0.0055
BERT+Mean+N+KM	0.373±0.0071	0.2286±0.0066	0.2273±0.0066
BERT+Mean+MM+KM	0.1718±0.0066	0.0511±0.0050	0.0493±0.0050
TFIDF+EFCM	0.2482±0.0081	0.1177±0.0018	0.1161±0.0018
BERT+Max+I+EFCM	0.2484±0.0070	0.122±0.0029	0.1204±0.0029
BERT+Max+LN+EFCM	0.2454±0.0069	0.1302±0.0015	0.1287±0.0015
BERT+Max+N+EFCM	0.2374±0.0051	0.1255±0.0014	0.1239±0.0014
BERT+Max+MM+EFCM	0.2043±0.0098	0.0706±0.0044	0.0689±0.0044
BERT+Mean+I+EFCM	0.2486±0.0077	0.1174±0.0016	0.1158±0.0016
BERT+Mean+LN+EFCM	0.2522±0.0037	0.1253±0.0012	0.1237±0.0012
BERT+Mean+N+EFCM	0.2523±0.0032	0.1252±0.0010	0.1236±0.0010
BERT+Mean+MM+EFCM	0.1632±0.0066	0.0415±0.0019	0.0397±0.0019
TFIDF+DEC	0.4024±0.0282	0.2176±0.0154	0.1621±0.0202
BERT+Max+I+DEC	0.1061±0.0143	0.003±0.0074	0.0015±0.0038
BERT+Max+LN+DEC	0.3969±0.0186	0.2301±0.0143	0.1761±0.0133
BERT+Max+N+DEC	0.1±0	0±0	0±0
BERT+Max+MM+DEC	0.1713±0.0708	0.0539±0.0638	0.0312±0.0397
BERT+Mean+I+DEC	0.4661±0.0282	0.286±0.0121	0.2317±0.0193
BERT+Mean+LN+DEC	0.4754±0.0266	0.2907±0.0119	0.2339±0.0172
BERT+Mean+N+DEC	0.427±0.0292	0.2613±0.013	0.1992±0.0172
BERT+Mean+MM+DEC	0.1±0	0.0001±0	0±0
TFIDF+IDEC	0.3975±0.0235	0.2243±0.0109	0.1474±0.0111
BERT+Max+I+IDEC	0.1326±0.0354	0.0225±0.0241	0.0135±0.0158
BERT+Max+LN+IDEC	0.4058±0.0182	0.2394±0.0129	0.1881±0.0131
BERT+Max+N+IDEC	0.1242±0.0342	0.0193±0.0275	0.0097±0.0144
BERT+Max+MM+IDEC	0.1694±0.0511	0.0504±0.0497	0.0278±0.0301
BERT+Mean+I+IDEC	0.477±0.0294	0.2988±0.0126	0.2445±0.0199
BERT+Mean+LN+IDEC	0.487±0.0258	0.3019±0.0118	0.247±0.0167
BERT+Mean+N+IDEC	0.4308±0.0303	0.2687±0.0134	0.2078±0.0170
BERT+Mean+MM+IDEC	0.1015±0.0029	0.0081±0.005	7E-05±0.0004

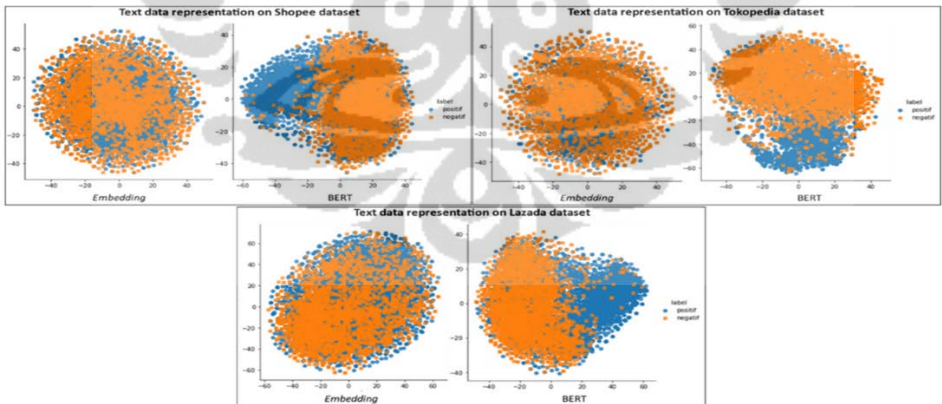
Gambar 16. Evaluasi akurasi metode-metode *Clustering* dengan representasi data teks

TFIDF dan BERT

Simulasi kami menunjukkan bahwa BERT memberikan akurasi yang lebih baik daripada representasi teks standar pada banyak metode *Clustering* yang ada, yaitu 28 dari 36 metrik memberikan hasil yang lebih baik (Gambar 16). Secara visual, representasi yang dihasilkan oleh BERT membuat dokumen yang mirip terletak lebih berdekatan (Gambar 17) (Subakti et al., 2022).



Gambar 17. Visualisasi representasi (a) TFIDF, (b) BERT pada masalah *Text Clustering*



Gambar 18. Visualisasi representasi *fine-tuned embedding* dan BERT untuk masalah *Sentiment Analysis* pada tiga dataset, yaitu Shopee, Tokopedia, dan Lazada

Pada simulasi kami lainnya untuk masalah Analisis Sentimen, BERT juga meningkatkan akurasi model jika dibandingkan dengan model yang menggunakan metode representasi standard *fine-tuned embedding*. Secara visual, representasi yang dihasilkan oleh BERT membuat dokumen dengan sentimen serupa terletak lebih dekat (Gambar 18) (Murfi et al., 2024).

3.5 Generative AI

GPT adalah salah satu *Generative AI* yang telah dilatih dengan data teks yang sangat besar (Radford, 2018). Selanjutnya, GPT dioptimasi untuk dialog seperti dialog manusia dan dikenal sebagai ChatGPT (OpenAI, 2022). Sehingga, ChatGPT dapat dipandang sebagai suatu sistem dengan pengetahuan yang sangat luas, multi bidang ilmu dan multi profesi, serta dapat berkomunikasi dengan pengguna dalam bentuk percakapan bahasa manusia yang dikenal sebagai *prompt*.

Peluncuran ChatGPT pada tahun 2022 telah menyita perhatian dunia karena kemampuannya yang luar biasa dalam memahami dan menghasilkan teks secara alami, menyerupai percakapan manusia. Selain itu, kemampuannya untuk memahami pertanyaan kompleks, memberikan jawaban yang relevan, serta belajar dari konteks percakapan, telah memukau pengguna dari berbagai bidang. Popularitasnya melonjak tidak hanya karena teknologinya yang canggih, tetapi juga karena kemudahannya diakses oleh masyarakat umum, membuka jalan bagi adopsi AI yang lebih luas dalam kehidupan sehari-hari.

Berdasarkan karakteristik di atas, *Generative AI* harusnya memberikan sejumlah manfaat besar bagi institusi Pendidikan maupun Industri dan Bisnis. Bagi Universitas, *Generative AI* dapat menjadi asisten

virtual dan perangkat pendukung pembelajaran, seperti perancangan rencana pembelajaran, penyusunan materi dan konten pembelajaran, serta merancang asesmen atau ujian. Pada bidang penelitian, *Generative AI* dapat menjadi perangkat pendukung penelitian, seperti studi literatur, penulisan dan penyusunan makalah.

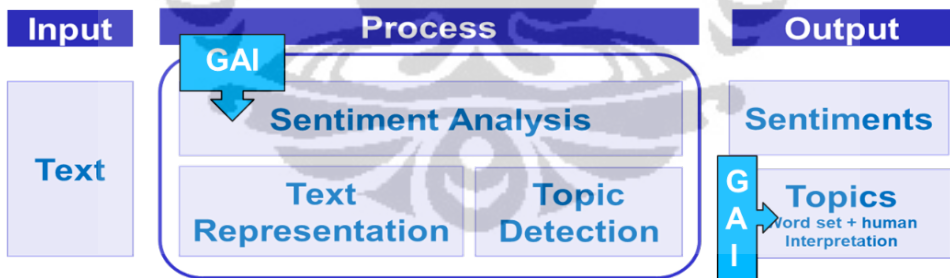
Generative AI juga menawarkan berbagai peluang inovatif dalam sektor industri dan bisnis. Secara umum, industri asuransi menunjukkan minat yang besar terhadap *Generative AI*, khususnya untuk peningkatan layanan pelanggan, efisiensi operasional, dan manajemen risiko. Survei Gallagher Bassett, yang merupakan penyedia layanan klaim dan anak perusahaan Arthur J Gallagher & Co, menyampaikan bahwa mayoritas perusahaan asuransi global aktif mendukung penerapan *Generative AI* dalam proses klaim, *underwriting*, dan layanan pelanggan. Sebanyak 44% perusahaan sedang mengintegrasikan teknologi ini dalam penyelesaian klaim, sementara 42% telah berhasil melakukannya.³

Generative AI telah berkembang pesat dan menunjukkan potensinya tidak hanya dalam menghasilkan konten, seperti teks, tetapi juga dapat diadaptasi dalam menyelesaikan berbagai masalah lain. Dengan memanfaatkan sifat *autoregressive*, model dapat diadaptasi untuk Analisis Sentimen dengan memanfaatkan kemampuannya memprediksi urutan kata secara kontekstual. Sifat *autoregressive* memungkinkan model memproses teks secara sekuensial, di mana setiap kata diprediksi berdasarkan kata-kata sebelumnya, sehingga menangkap hubungan semantik dan emosional secara mendalam. Dalam Analisis Sentimen, pendekatan ini digunakan untuk mengidentifikasi pola bahasa yang

³ <https://mediaasuransinews.co.id/asuransi/chatbots-ai-dan-gen-ai-diprediksi-jadi-primadona-sebagai-alat-proses-klaim-asuransi/>

mengindikasikan emosi positif, negatif, atau netral. Kemampuan *autoregressive* memastikan analisis tidak hanya fokus pada kata-kata individu, tetapi juga mempertimbangkan konteks keseluruhan. Sehingga, *Generative AI* diharapkan menghasilkan penilaian yang lebih presisi terhadap sentimen, baik dalam teks pendek seperti *tweet*, maupun teks panjang seperti ulasan produk atau artikel.

Kami juga telah mengadopsi *Generative AI* untuk beberapa komponen pada perangkat Analisis Sentimen Level Topik, yaitu untuk komponen Analisis Sentimen dan komponen Interpretasi Topik (Gambar 19) (Cahyadi et al., 2024). GPT memungkinkan penentuan sentimen dalam teks tanpa perlu tahapan pelatihan model, sehingga menjadi perangkat yang lebih praktis. Dengan sifat *autoregressive* atau memberikan contoh kalimat positif atau negatif di awal, model dapat memahami perbedaan antara sentimen dan menghasilkan analisis sentimen dengan tingkat akurasi yang setara dengan model yang dilatih dengan data training.



Gambar 19. Adopsi *Generative AI* (GAI) pada perangkat Analisis Sentimen Level Topik

Selain itu, GPT memungkinkan interpretasi topik dilakukan oleh sistem, bukan dilakukan secara manual oleh manusia. Ini berarti bahwa mekanisme ini dapat menghindari aspek subjektivitas dalam penentuan topik. Sehingga, kita dapat memastikan bahwa hasil analisis didasarkan pada data dan algoritma, bukan interpretasi atau bias manusia. Kerangka kerja ini dapat meningkatkan efisiensi proses dan juga konsistensi, keandalan dan kredibilitas hasil.



4 Penutup

Pada era ekonomi digital, penguatan pengambilan keputusan berbasis data menjadi landasan penting dalam menghadapi tantangan dan peluang yang semakin kompleks. Model Pembelajaran Mesin dan Model Bahasa Besar menjadi perangkat penting untuk pengelolaan data besar, terciptanya solusi yang adaptif dan presisi, serta menghasilkan wawasan yang relevan untuk mendukung pengambilan keputusan yang lebih baik. Perkembangan ini diharapkan semakin meningkatkan kapabilitas kita dalam mengambil keputusan berbasis data dalam industri, bisnis, maupun riset akademik.

Kebutuhan ini menjadi peluang bagi Bidang Matematika untuk lebih berperan serta menyiapkan sumber daya manusia Indonesia dalam memenuhi kebutuhan-kebutuhan tersebut. Pada sisi lain, kondisi ini menjadi tantangan bagi Bidang Matematika untuk menyiapkan sumber daya manusia Indonesia yang kompetitif, yang menguasai sains dan teknologi terkini, dan sesuai dengan kebutuhan industri dan bisnis. Penyiapan tersebut mulai dari kurikulum yang adaptif dengan perkembangan keilmuan dan kebutuhan para pemangku kepentingan, maupun metode penyampaian pengetahuan yang mendorong mahasiswa untuk menguasai sains dan teknologi terkini, seperti pembelajaran berbasis proyek maupun pembelajaran berbasis riset.

Sebagai akademisi, kita memiliki tanggung jawab untuk memastikan bahwa perkembangan sains dan teknologi dapat menjadi solusi nyata yang memberikan manfaat bagi para pemangku kepentingan. Kami percaya, dengan semangat inovasi yang berlandaskan pada dasar keilmuan yang kokoh, kita dapat menghadirkan solusi yang relevan dan berkelanjutan bagi tantangan era ekonomi digital.

5 Ucapan Terima Kasih

Hadirin yang saya muliakan,

Sebelum mengakhiri pidato pengukuhan Guru Besar saya ini, perkenankan saya menyampaikan ucapan terima kasih kepada berbagai pihak yang telah memberikan perhatian, dukungan, dan bantuan kepada saya dalam berbagai bentuk. Ucapan terima kasih saya khususkan kepada:

1. Pemerintah Republik Indonesia, melalui Menteri Pendidikan dan Kebudayaan Republik Indonesia Periode 2019-2024, Nadiem Anwar Makarim, B.A., M.B.A., yang telah menetapkan dan mengangkat saya sebagai Guru Besar Tetap Universitas Indonesia.
2. Pemerintah Republik Indonesia, melalui Menteri Pendidikan Tinggi, Sains dan Teknologi, Prof. Dr. Ir. Satrio Soemantri Brodjonegoro.
3. Terima kasih saya sampaikan kepada Rektor Universitas Indonesia Periode 2020-2024, Prof. Ari Kuncoro, SE, MA, PhD, para Wakil Rektor, Sekretaris, para Direktur yang telah mengusulkan pengangkatan saya sebagai Guru Besar Tetap Universitas Indonesia.
4. Terima kasih saya sampaikan kepada Rektor UI: Prof. Dr. Ir. Heri Hermansyah, S.T., M.Eng., IPU, para Wakil Rektor, dan para Kepala Badan serta jajarannya di Pusat Administrasi Universitas Indonesia.
5. Ketua Dewan Guru Besar Universitas Indonesia, Prof. Harkristuti Harkrisnowo, S.H., M.A., Ph.D. beserta seluruh anggota Dewan Guru Besar, Ketua Tim PAK Universitas Indonesia, Prof. Dr. Ir. Gandjar Kiswanto, M.Eng. beserta anggota PAK Universitas Indonesia yang telah menyetujui pengusulan Guru Besar saya, menyetujui

pengukuhan saya pada hari ini dan telah menerima saya dalam lingkungan akademik yang terhormat ini.

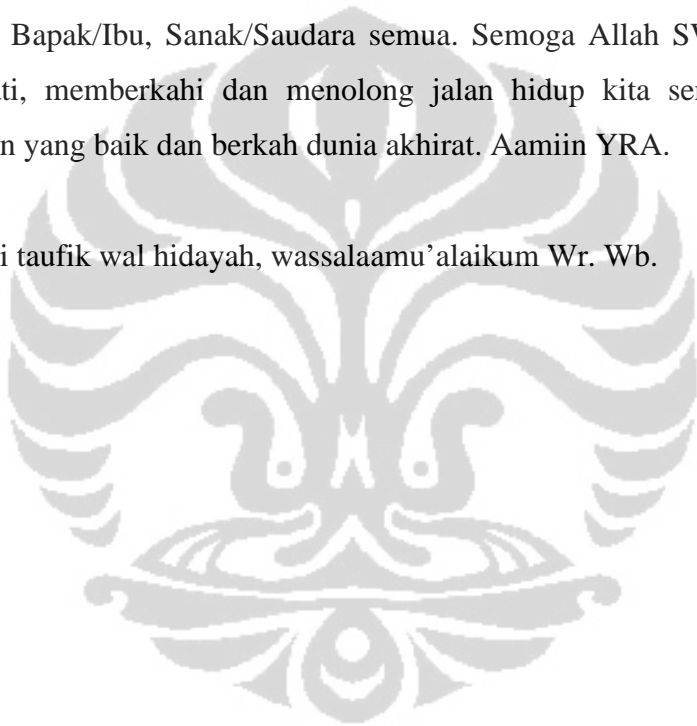
6. Ketua Senat Akademik Universitas Indonesia, Prof. Dr. dr. Budi Wiweko, SpOG, subsp.FER, MPH. Int. Aff. RANZCOG, FICRM. beserta seluruh anggota Senat Akademik Universitas Indonesia atas bantuan dan dukungannya kepada saya sehingga saya dikukuhkan menjadi Guru Besar.
7. Terima kasih saya sampaikan kepada Bapak Dekan FMIPA Universitas Indonesia, Prof. Dede Djuhana, M.Si., Ph.D., Wakil Dekan dan seluruh jajaran Pimpinan Fakultas, yang telah membantu proses pengajuan Guru Besar kami.
8. Kepada Dewan Guru Besar FMIPA Universitas Indonesia yang dipimpin oleh Prof. Dr. Sumi Hudiyono dengan sekretaris Prof. Dr. Jarnuzi Gunlazuardi beserta seluruh anggota, saya mengucapkan terima kasih setinggi-tingginya yang telah mendukung pengusulan saya menjadi guru besar FMIPA Universitas Indonesia.
9. Ucapan terima kasih saya sampaikan terimakasih kepada Ketua Senat Akademik FMIPA Universitas Indonesia, Prof. Dra. Wellyzar Sjamsuridzal, M.Sc., Ph.D., beserta seluruh anggota Senat Akademik FMIPA Universitas Indonesia atas dukungan dan bantuannya.
10. Saya mengucapkan terima kasih kepada semua pihak khususnya Direktur SDM Universitas Indonesia, Prof. Dr. -Ing. Amalia Suzianti, S.T., M.Sc. beserta jajarannya, Dra. Elmida, S., Agus Anang, S.Kom., M.T.I., Arham Akbar, dan juga Manager SDM FMIPA Universitas Indonesia, Fakhur Rizqi, S.Sos., beserta jajarannya, Caya Mikail, atas bantuannya dan dukungannya dalam proses kenaikan pangkat saya.

11. Ungkapan rasa terimakasih kami haturkan kepada Ketua Departemen Matematika Prof. Alhadi Bustamam, S.Si., M.Kom., Ph.D., Dr. Dian Lestari, DEA. (Periode 2017-2021), para Ketua Program Studi Dr. Hengki Tasman, S.Si., M.Si., Dr. Dra. Yekti Widyaningsih, M.Si., Dr. Suci Fratama Sari, S.Si, M.Si., Sarini Abdullah, S.Si., M.Stats., Ph.D., Dr. Fevi Novkaniza, S.Si., M.Si., Prof. Dr. Dra. Titin Siswantining, D.E.A., para Dosen, Tendik, sebagai Keluarga Besar Departemen Matematika FMIPA Universitas Indonesia, serta Drs. Suryadi Slamet, M.Sc., Dr. Dra. Sri Mardiyati, M.Kom, dan semua Dosen Purna Tugas atas dedikasinya pada perkembangan Departemen Matematika.
12. Terima kasih juga untuk Ketua Departemen Teknik Elektro Prof. Dr.Eng. Arief Udhiarto, S.T., M.T., IPU., Dr. Ir. Aries Subiantoro (Periode 2018-2022), dan Prof. Dr. Ir. Kalamullah Ramli, M.Eng yang telah mempercayakan ko-promotor kepada saya, ketua Sidang Promosi Prof. Dr. Ir. Riri Fitri Sari, MSc, serta anggota dewan penguji.
13. Terima kasih banyak untuk semua Pembimbing Tugas Akhir saya Almarhum Prof. Djati Kerami, Prof. Benyamin Kusumoputro, dan Prof. Klaus Obermayer. Terima kasih juga untuk semua Mahasiswa Bimbingan saya atas kerjasama dalam penelitian dan juga publikasi hasil penelitiannya.
14. Saya mengucapkan terima kasih kepada kedua orang tua, atas semua limpahan kasih sayang yang tak berhingga, Almarhum Bapak H. Madani H. Umar dan Ibu Hj. Rusdiana. Serta kedua Mertua, Almarhum Bapak Achyarnis dan Ibu Hj. Sukemi Anisyah. Teriring doa semoga Allah SWT selalu merahmati dan memberkahi mereka semua. Terima kasih buat istriku tercinta Dr. Munaya Fauziah, SKM, M.Kes;

keluarga besar H. Madani H. Umar: Kak Amri Antoni, Kak Desi Wiryani, Kak Nazwar Yunus, Kak Ermaliza, Edi Mulyadi, Sunarsih, Heriman, Septilyta, Hari Ahmadi, Dinda Meta, Heri Yanto, Damyanti; serta keluarga besar Hj. Sukemi Anisyah: Mas John Wisnu Anggono, Mbak Sri Widodo, Mbak Elly, Mas Yogi Sasmito Nugroho, Mbak Enny Wisnu Anggraheni.

Akhirul kalam, hanya Allah SWT yang sanggup membalas semua kebaikan Bapak/Ibu, Sanak/Saudara semua. Semoga Allah SWT selalu merahmati, memberkahi dan menolong jalan hidup kita semua pada kehidupan yang baik dan berkah dunia akhirat. Aamiin YRA.

Wabillahi taufik wal hidayah, wassalaamu'alaikum Wr. Wb.



6 Referensi

- Cahyadi, D. J., Setiawan, T., Murfi, H., Satria, Y., Hariadi, N., Za'in, C., Souza-Daw, T. de, & Haryanto, A. (2024). Neural Networks-based Topic-Level Sentiment Analysis and Its Application for Ride-Hailing User Reviews. *PREPRINT*.
- Cleveland, W. S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review / Revue Internationale de Statistique*, 69(1), 21–26. <https://doi.org/10.2307/1403527>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Fauzan, M. A., & Murfi, H. (2018). The accuracy of XGBoost for insurance claim prediction. *International Journal of Advances in Soft Computing and Its Applications*, 10(2).
- Gowandi, T., Murfi, H., & Nurrohmah, S. (2021). Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis. In A. Mohamed, B. W. Yap, J. M. Zain, & M. W. Berry (Eds.), *Soft Computing in Data Science* (pp. 18–27). Springer Singapore.
- Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2021). Boosting Insights in Insurance Tariff Plans with Tree-Based Machine Learning Methods. *North American Actuarial Journal*, 25(2), 255–285. <https://doi.org/10.1080/10920277.2020.1745656>
- Hinton, G. E., & Sejnowski, T. J. (1983). Optimal Perceptual Inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 448–453.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Kaminski, M. E. (2019). The Right to Explanation, Explained. *Berkeley Technology Law Journal*, 34(1).
- Keuangan, O. J. (2023). *Peta Jalan Pengembangan dan Penguatan*

- Perasuransian Indonesia 2023-2027*. Otoritas Jasa Keuangan.
<https://www.ojk.go.id/id/regulasi/otoritas-jasa-keuangan/rancangan-regulasi/Documents/Draft%20Roadmap%20Pengembangan%20Perasuransian%20Indonesia.pdf>
- Muliawati, T., & Murfi, H. (2017). *Eigenspace-based fuzzy c-means for sensing trending topics in Twitter*. 030140.
<https://doi.org/10.1063/1.4991244>
- Murfi, H. (2018). The Accuracy of Fuzzy C-Means in Lower-Dimensional Space for Topic Detection. In M. Qiu (Ed.), *Smart Computing and Communication* (pp. 321–334). Springer International Publishing.
- Murfi, H. (2021). A scalable eigenspace-based fuzzy c-means for topic detection. *Data Technologies and Applications*, 55(4), 527–541.
<https://doi.org/10.1108/DTA-11-2020-0262>
- Murfi, H., Rosaline, N., & Hariadi, N. (2022). Deep autoencoder-based fuzzy c-means for topic detection. *Array*, 13, 100124.
<https://doi.org/10.1016/j.array.2021.100124>
- Murfi, H., Syamsyuriani, Gowandi, T., Ardaneswari, G., & Nurrohmah, S. (2024). BERT-based combination of convolutional and recurrent neural network for Indonesian sentiment analysis. *Applied Soft Computing*, 151, 111112.
<https://doi.org/https://doi.org/10.1016/j.asoc.2023.111112>
- Nur'aini, K., Najahaty, I., Hidayati, L., Murfi, H., & Nurrohmah, S. (2015). Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter. 2015 *International Conference on Advanced Computer Science and Information Systems (ICACISIS)*, 123–128.
<https://doi.org/10.1109/ICACISIS.2015.7415168>
- OpenAI. (2022). *ChatGPT*. <https://openai.com/chatgpt>
- Parlina, A., Ramli, K., & Murfi, H. (2020). Theme mapping and bibliometrics analysis of one decade of big data research in the scopus database. *Information (Switzerland)*, 11(2).
<https://doi.org/10.3390/info11020069>
- Parlina, A., Ramli, K., & Murfi, H. (2021). Exposing Emerging Trends in Smart Sustainable City Research Using Deep Autoencoders-Based Fuzzy C-Means. *Sustainability*, 13(5), 2876.
<https://doi.org/10.3390/su13052876>
- Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Topic-level sentiment analysis of social media data using deep learning. *Applied Soft*

- Computing*, 108, 107440.
<https://doi.org/https://doi.org/10.1016/j.asoc.2021.107440>
- Radford, A., & Narasimhan, K. (2018). *Improving Language Understanding by Generative Pre-Training*.
- Rusdah, D. A., & Murfi, H. (2020). XGBoost in handling missing values for life insurance risk prediction. *SN Applied Sciences*, 2(8), 1336.
<https://doi.org/10.1007/s42452-020-3128-y>
- Subakti, A., Murfi, H., & Hariadi, N. (2022). The performance of BERT as data representation of text clustering. *Journal of Big Data*, 9(1), 15. <https://doi.org/10.1186/s40537-022-00564-9>
- Tjahjono, S., Murfi, H., & Devila, S. (2024). Claim Severity Prediction of Workers' Compensation Using Tree and Neural Networks-Based Models. *2024 11th IEEE Swiss Conference on Data Science (SDS)*, 225–228. <https://doi.org/10.1109/SDS60720.2024.00039>
- Wijaya, D., Murfi, H., & Ardaneswari, G. (2024). Topic-level Sentiment Analysis for User Reviews in Gasoline Subsidy Application. *The 11th IEEE Swiss Conference on Data Science*.
- Yusdiansyah, M. R., Murfi, H., & Wibowo, A. (2019). Randomspace-Based Fuzzy C-Means for Topic Detection on Indonesia Online News. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11909 LNAI*. https://doi.org/10.1007/978-3-030-33709-4_12

7 Riwayat Hidup

Nama lengkap : Prof. Dr. rer. nat. Hendri Murfi, S.Si., M.Kom.
NIP : 197210231998021002
Pangkat/Golongan : Pembina Tk. I/IVB
Jabatan : Guru Besar
Tempat/Tanggal lahir : Sekayu/23 Oktober 1972
Jenis kelamin : Laki-Laki
Agama : Islam
Nama Ayah : H. Madani H. Umar (alm)
Nama Ibu : Hj. Rusdiana
Nama Istri : Dr. Munaya Fauziah, SKM, MKes

Pendidikan Formal

Tahun	Keterangan
1991 – 1996	Sarjana Sains (S.Si) bidang Komputasi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alama, Universitas Indonesia
2000 – 2002	Magister Komputer (M. Kom) bidang Jaringan Syaraf, Fakultas Ilmu Komputer, Universitas Indonesia.
2005 – 2010	<i>Doktor der Naturwissenschaften</i> (Dr. rer. nat) bidang Pembelajaran Mesin, Fakultas Teknik Elektro dan Ilmu Komputer, Technische Universität Berlin

Jabatan Fungsional

Tahun	Keterangan
1 Februari 1998	Pengajar
1 Januari 2001	Asisten Ahli
1 Mei 2002	Lektor
1 September 2017	Lektor Kepala
1 September 2024	Guru Besar

Jabatan Struktural

Tahun	Keterangan
2010 - 2015	Sekretaris Program Studi Magister Matematika, Departemen Matematika, FMIPA Universitas Indonesia
2017 - 2021	Ketua Program Studi Sarjana Ilmu Aktuaria, Departemen Matematika, FMIPA Universitas Indonesia
2022 - Sekarang	Ketua Program Studi Magister Matematika, Departemen Matematika, FMIPA Universitas Indonesia

Organisasi Akademik

Tahun	Keterangan
2009 - Sekarang	Anggota <i>Association for Computing Machinery</i> (ACM)
2009 - Sekarang	Anggota <i>International Machine Learning Society</i> (IMLS)
2011 - Sekarang	Anggota <i>The Indonesia Mathematical Society</i> (IndoMS)
2022 - Sekarang	Anggota Perkumpulan Akademisi dan Saintis Indonesia (ASASI)
2023 - Sekarang	Peserta Undangan <i>THE's Global Academic Reputation Survey for the Times Higher Education World University Rankings</i>
2023 - Sekarang	Anggota Komite Seleksi Beasiswa <i>Deutscher Akademischer Austauschdienst</i> (DAAD)
2025	Anggota <i>Scientific Program Committee, Swiss Conference on Data Science</i>

Hibah Penelitian

Tahun	Keterangan
2024 – 2025	Analisis dan Pengembangan Metode Pendeteksian Topik Berbasis <i>Fully Deep Learning</i> . Hibah PUTI Q1 Universitas Indonesia
2024 – 2025	Analisis dan Pengembangan Metode Prediksi dan Interpretasi Tarif Asuransi Berbasis <i>Machine Learning</i> . Hibah PUTI Q2 Universitas Indonesia
2023 – 2024	Analisis dan Pengembangan Metode Pendeteksian Topik Berbasis BERT dan <i>Deep Clustering</i> . Hibah PUTI Q1 Universitas Indonesia
2021 - 2022	Pengembangan Metode Pendeteksian Topik Berbasis Fuzzy Clustering dengan Menggunakan <i>Deep Learning</i> dan <i>Lifelong Learning</i> . Hibah PUPT Kemenristekdikti

Karya Ilmiah Pilihan

Tahun	Keterangan
2024	Murfi, H. , Syamsuriani, Gowandi, T., Ardaneswari, G., Nurrohmah, S. (2024). BERT-Based Combination of Convolutional and Recurrent Neural Network for Indonesian Sentiment Analysis. <i>Applied Soft Computing</i> , Vol. 151:111112
2024	Wijaya, D. R. G., Murfi, H. , Ardaneswari, G. (2024) Topic-Level Sentiment Analysis for User Reviews in Gasoline Subsidy Applications. <i>Proceeding of The 11th IEEE Swiss Conference on Data Science</i> , 30-31 Mei 2024, Zurich, Swiss
2024	Tjahjono, S., Murfi, H. , Devila, S. (2024) Claim Severity Prediction of Workers' Compensation Using Tree and Neural Networks-Based Models. <i>Proceeding of The 11th IEEE Swiss Conference on Data Science</i> , 30-31 Mei 2024, Zurich, Swiss
2022	Subakti, A., Murfi, H. & Hariadi, N (2022). The Performance of BERT as Data Representation of Text Clustering. <i>Journal of Big Data</i> , Vol. 9, No. 15

2022	Murfi, H. , Rosaline, N., Hariadi, H. (2022) Deep Autoencoder-based Fuzzy C-Means for Topic Detection. <i>Array</i> , Vol. 13
2021	Parlina, A.; Ramli, K.; Murfi, H. Exposing Emerging Trends in Smart Sustainable City Research Using Deep Autoencoders-Based Fuzzy C-Means. <i>Sustainability</i> 2021, 13, 2876
2021	Gowandi, T., Murfi, H. , Nurrohmah, S. (2021). Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis. <i>Communications in Computer and Information Science</i> , vol 1489. Springer, Singapore
2021	Murfi, H. (2021) A Scalable Eigenspace-Based Fuzzy C-Means for Topic Detection. <i>Data Technologies and Applications</i> , Vol. 55 No. 4, pp. 527-541
2020	Rusdah, D.A., Murfi, H. (2020) XGBoost in Handling Missing Values for Life Insurance Risk Prediction. <i>SN Applied Science</i> , 2(1336)
2020	Parlina, A.; Ramli, K.; Murfi, H. (2020) Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. <i>Information</i> 2020, 11, 69
2019	Yusdiansyah M.R., Murfi H. , Wibowo A. (2019) Randomspace-Based Fuzzy C-Means for Topic Detection on Indonesia Online News. <i>Lecture Notes in Computer Science</i> , vol 11909. Springer, Cham
2019	Murfi, H. , Siagian, F. L., & Satria, Y. (2019) Topic Features for Machine Learning-based Sentiment Analysis in Indonesian Tweets. <i>International Journal of Intelligent Computing and Cybernetics</i> , 12(1), 70-81
2018	Murfi H. (2018) The Accuracy of Fuzzy C-Means in Lower-Dimensional Space for Topic Detection. <i>Lecture Notes in Computer Science</i> , vol 11344. Springer, Cham
2018	Fauzan, M. A., Murfi, H. (2018) The Accuracy of XGBoost for Insurance Claim Prediction. <i>International Journal of Advances in Soft Computing and its Applications</i> , 10(2), 159-171

