 ISSN:2302-8149
e-ISSN:2540-9719

Pemetaan Tren *Machine Learning* dalam Penelitian Ilmu Kimia menggunakan *LLM* dengan *Multi-Turn Prompting*

Mapping Machine Learning Trends in Chemistry Research using LLM with Multi-Turn Prompting

[1]Andreo Yudertha*, [2]Riski Dwimalida Putri
[1]Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Sulthan Thaha Saifuddin Jambi
[2]Program Studi Kimia, Fakultas Sains dan Teknologi, UIN Sulthan Thaha Saifuddin Jambi
[1,2]Jl. Jambi - Muara Bulian No.KM. 16, Kec. Jambi Luar Kota, Muaro Jambi, Jambi, Indonesia
*e-mail: andreo@uinjambi.ac.id

(***received*:** 9 January 2025, ***revised*:** 4 February 2025, ***accepted*:** 5 February 2025)

Abstrak

Peninjauan terhadap penelitian di bidang kimia yang melibatkan *machine learning* diperlukan untuk mengidentifikasi perkembangan terkini dan menggali potensi aplikasinya. Artikel-artikel penelitian yang telah terpublikasi memberikan kesempatan untuk menganalisa tren penelitian yang sedang berkembang. Penggunaan teknologi pemrosesan bahasa alami (NLP) tidak hanya mempercepat proses analisis data teks, tetapi juga meningkatkan akurasi dalam memahami isi dan konteks artikel ilmiah. Sebelumnya analisis tren artikel bidang oftalmologi sudah pernah dilakukan dengan *Zero Shot Learning*. Pada penelitian ini dilakukan analisis artikel bidang kimia yang terkait dengan topik *machine learning* dengan teknik *multi-turn prompting*. Tahapan dilakukan dengan mengumpulkan data dengan teknik *scraping* pada abstrak artikel yang memuat kata kunci *machine learning* dan *chemistry*. Data yang diperoleh ditabulasi dan dianalisis menggunakan *Large Language Model* (LLM) dengan menggunakan teknik *Multi-Turn Prompting* dengan memasukkan *prompt* dari hal yang umum, dan menggali informasi yang lebih dalam berdasarkah hasil *prompting* sebelumnya. Selain itu, analisi dilakukan dengan memberikan *prompting* untuk membuat analisis statistik deskriptif. Hasil analisis 200 abstrak artikel didapat tujuh kata kunci utama terkait bidang kimia yang memanfaatkan *machine learning*, yakni *chemical* (138 artikel), *protein* (119 artikel), *drug* (107 artikel), *structure* (100 artikel), *molecular* (96 artikel), *chemistry* (91 artikel), dan *quantum* (84 artikel). Selain itu dari didapatkan adanya tiga topik utama yang dominan dalam kimia yang dikombinasikan dengan *machine learning*, yaitu struktur protein dan molekul , kimia kuantum serta penemuan obat (*drug discovery)*. Terbitan artikel mengenai *machine learning* dalam bidang kimia mulai naik pada pada tahun 2012 dan naik signifikan pada tahun 2019. Berdasarkan hasil yang diperoleh masih banyak peluang pengembangan *machine learning* dalam bidang kimia khusunya bidang kimia kuantum yang baru terpublikasikan pada tahun 2013 dan jumlah artikel yang dipublikasikan tidak terlalu besar ditiap tahunnya, yang menunjukkan bahwa bidang ini masih dalam tahap awal eksplorasi.

Kata kunci: *machine learning, multi-turn prompting,* kimia, *LLM*

Abstract

A review of research in the field of chemistry that incorporates machine learning is essential to identify recent developments and explore its potential applications. Published research articles provide an opportunity to analyze emerging research trends. The use of natural language processing (NLP) technology not only accelerates text data analysis but also enhances accuracy in understanding the content and context of scientific articles. Previously, trend analysis in ophthalmology research had been conducted using Zero-Shot Learning. In this study, an analysis of chemistry-related articles focusing on machine learning was carried out using a multi-turn prompting technique. The process began with data collection through web scraping of abstracts containing the keywords "machine learning" and "chemistry." The retrieved data was then tabulated and analyzed using a Large Language Model (LLM) with a Multi-Turn Prompting approach, where general prompts were initially used, followed by deeper exploration based on previous responses.

 ISSN:2302-8149
e-ISSN:2540-9719

Additionally, statistical descriptive analysis was performed using targeted prompts. Analysis of 200 article abstracts identified seven key terms related to the use of machine learning in chemistry: chemical (138 articles), protein (119 articles), drug (107 articles), structure (100 articles), molecular (96 articles), chemistry (91 articles), and quantum (84 articles). Furthermore, three dominant research topics were found in the intersection of chemistry and machine learning: protein and molecular structure, quantum chemistry, and drug discovery. The number of articles on machine learning in chemistry began to rise in 2012 and saw a significant increase in 2019. The findings suggest that there are still many opportunities for developing machine learning applications in chemistry, particularly in quantum chemistry. This field only began to gain attention in 2013, and the number of published articles remains relatively low each year, indicating that it is still in the early stages of exploration.

Keywords: *machine learning, multi-turn prompting , trend, chemistry, LLM*

1 Pendahuluan

Pemanfaatan *machine* learning mengalami pertumbuhan yang sangat signifikan seiring dengan meningkatnya kekuatan komputasi dan ketersediaan data yang besar (*big data*). Hingga sekarang penerapan *Machine Learning* semakin luas, mencakup berbagai bidang mulai dari kesehatan, keuangan, hingga mobil otonom[1][2][3]. Perkembangan pemanfaatan *Machine Learning* menunjukkan bagaimana teknologi ini telah berevolusi dari konsep teoritis menjadi alat yang esensial dalam dunia modern, dengan dampak yang signifikan pada banyak aspek kehidupan.

Salah satu bidang yang dapat memanfaatkan *machine learning* dengan efektif adalah kimia. Dengan kemampuan *Machine Learning* untuk menganalisis data dalam jumlah yang besar dan kompleks, para peneliti kimia dapat mempercepat penemuan dan pengembangan senyawa baru [4]. *Machine Learning* dapat digunakan untuk memprediksi sifat kimia dan fisik dari molekul-molekul baru, sehingga mempercepat proses pengembangan obat atau material baru [5]. Integrasi *Machine Learning* dalam penelitian kimia tidak hanya meningkatkan efisiensi tetapi juga membuka peluang untuk inovasi yang lebih cepat dan lebih tepat sasaran.

Meskipun penelitian sistematis mengenai penerapan machine learning (ML) telah banyak dilakukan, seperti dalam *software engineering* (SE) dan kesehatan, terdapat kesenjangan yang signifikan dalam konteks ilmu kimia. Studi oleh Nyaga Fred dan I.O. Temkin berhasil mengidentifikasi domain aplikasi ML di SE [6], sementara Katarzyna Kolasa dan rekan-rekannya menyoroti keberhasilan dan tantangan penerapan ML di bidang kesehatan [7], termasuk kebutuhan akan data yang berkualitas dan interpretabilitas model. Namun, dalam ilmu kimia, belum ada tinjauan sistematis yang mendalam yang secara eksplisit memetakan teknik ML terhadap tantangan spesifik, seperti prediksi sifat material, pemodelan reaksi kimia, dan sintesis senyawa baru. Kurangnya upaya sistematis ini menunjukkan perlunya eksplorasi lebih lanjut untuk memahami bagaimana ML dapat digunakan secara optimal dalam mengatasi kompleksitas dan ketidakpastian sistem kimia. Hal ini menciptakan peluang penelitian untuk mengisi celah pengetahuan ini dengan mengadaptasi metodologi yang telah berhasil diterapkan di bidang lain.

Menganalisis tren machine learning (ML) dalam ilmu kimia sangat penting untuk memberikan gambaran yang jelas mengenai bidang-bidang yang memanfaatkan teknologi ini. Dalam artikel *"Machine Learning for Chemistry: Basics and Applications"* oleh Yun-Fei Shi et al. [8], dijelaskan bahwa ML telah merevolusi beberapa aspek penting dalam ilmu kimia, seperti retrosintesis, simulasi atom, dan katalisis heterogen. Dengan memanfaatkan ML, peneliti dapat memprediksi rute sintesis organik yang lebih efisien, mempercepat simulasi atom untuk memahami potensi energi permukaan, dan mengoptimalkan desain katalis untuk reaksi kimia yang lebih efektif. Tren ini sangat krusial karena memberikan wawasan tentang bagaimana ML mampu meningkatkan akurasi dan efisiensi dalam penelitian kimia, serta mempercepat penemuan material dan senyawa baru. Oleh karena itu, menganalisis tren ML dalam ilmu kimia memungkinkan kita untuk memahami penerapan teknologi ini dalam riset dan pengembangan di masa depan, serta potensi revolusi yang bisa dihasilkan dari penerapannya di bidang kimia.

Artikel penelitian yang terpublikasikan menjadi peluang dalam melihat bagaimana tren penelitian berkembang. Dengan menganalisis artikel-artikel tersebut, peneliti dapat mengindentifikasi topik-topik yang menjadi fokus utama, metode atau pendekatan yang paling dominan, serta arah

 ISSN:2302-8149
e-ISSN:2540-9719

inovasi yang berkembang. Pemetaan tren dari publikasi dapat membantu dalam menemukan area yang kurang terjamah atau gap dalam penelitian, dengan demikian artikel penelitian yang terpublikasikan tidak hanya mencerminkan kondisi saat ini dari suatu bidang studi, tetapi juga membantu dalam mendapatkan gap penelitian.

Teks-teks yang terdapat dalam literatur dapat dianalisis secara mendalam menggunakan pendekatan Natural Language Processing (NLP) untuk mempermudah dalam mengekstraksi ide-ide penting yang terkandung di dalamnya [9]. Saat ini berbagai aplikasi berbasis NLP telah tersedia dan dapat dimanfaatkan untuk keperluan tersebut. Aplikasi seperti ChatGPT, Gemini, dan lainnya yang berbasis pada *Large Language Models (LLM)* menawarkan kemampuan analisis teks yang canggih, memungkinkan pengguna untuk mengidentifikasi informasi utama dengan cepat dan efisien [10]. Penggunaan teknologi NLP ini tidak hanya mempercepat proses analisis data teks, tetapi juga meningkatkan akurasi dalam memahami isi dan konteks suatu artikel, sehingga sangat bermanfaat dalam berbagai bidang, termasuk penelitian, pendidikan, dan pengembangan teknologi.

Pemanfaatan analisis teks untuk mendapatkan tren penelitian membutuhkan banyak artikel. Teknik *scraping* merupakan salah satu cara untuk mendapatkan data dari berbagai sumber online yang diekstraksi secara otomatis dalam jumlah besar[11]. Dalam dunia penelitian dan pengembangan teknologi, penggunaan *web scraping* untuk memperoleh data teks merupakan langkah awal yang krusial untuk melakukan analisis lebih lanjut yang berbasis NLP.

Tinjauan terhadap tren penelitian *machine learning* pada ilmu kimia dengan menelaah artikel-artikel yang terpublikasikan dengan memanfaatkan analisis teks penting untuk dilakukan. Hal ini bertujuan untuk memberi gambaran pemanfaatan *machine learning* terhadap bidang kimia, dengan memanfaatkan teknik dan teknologi yang ada saat ini, untuk mempermudah dan meningkatkan efisiensi.

2 Tinjauan Literatur

Penggunaan LLM untuk menganalisis tren artikel telah banyak digunakan, salah satunya adalah pada penelitian yang dilakukan oleh Hina et al. [12] yang mengusulkan metode otomatis untuk klasifikasi artikel ilmiah dengan memanfaatkan *Large Language Models* (LLMs) yang berfokus pada bidang oftalmologi. Model berbasis *Natural Language Processing* (NLP) ini memanfaatkan teknik *zero-shot learning (ZSL)* dan dibandingkan dengan model seperti *Bidirectional and Auto-Regressive Transformers (BART)*, serta *BERT* dan variannya (*distilBERT, SciBERT, PubmedBERT, BioBERT*). Menggunakan dataset *RenD* yang berisi 1000 artikel tentang penyakit mata, yang telah diannotasi menjadi 15 kategori oleh panel ahli, model ini mencapai akurasi rata-rata 0.86 dan skor F1 rata-rata 0.85. Hasil ini menunjukkan bahwa LLM efektif dalam mengkategorikan artikel tanpa intervensi manusia, meningkatkan akurasi dan efisiensi. Selain membantu peneliti dan klinisi dalam mengkategorikan, mengambil literatur yang relevan, dan mengidentifikasi tren ilmiah, framework ini juga dapat diterapkan pada berbagai disiplin ilmu lain untuk mendukung penelitian dan analisis tren.

Selain itu Youngjin Chae et al. [13] menggunakan LLM dalam pengklasifikasian teks. Penulis membandingkan sepuluh model LLM dengan ukuran mulai dari 86 juta hingga 1,7 triliun parameter menggunakan empat pendekatan pelatihan: *zero-shot learning*, *few-shot learning*, *fine-tuning*, dan *instruction-tuning*. Hasil menunjukkan bahwa model terbesar umumnya memberikan kinerja prediksi terbaik. Namun, *fine-tuning* pada model yang lebih kecil menjadi solusi kompetitif karena memiliki akurasi yang relatif tinggi dengan biaya lebih rendah. Untuk tugas prediksi yang kompleks, model berbobot terbuka dengan *instruction-tuning* dapat bersaing dengan model komersial mutakhir. Penelitian ini juga memberikan rekomendasi penggunaan LLM untuk klasifikasi teks dalam penelitian sosiologi serta membahas keterbatasan dan tantangan teknologi ini, seperti biaya pelatihan, kebutuhan sumber daya, dan tantangan dalam interpretasi hasil.

Zihao yi et al [14] melakukan sebuah penelitian yang membahas tentang tinjauan komprehensif penelitian terkait *multi-turn dialogue systems*, dengan fokus pada *multi-turn dialogue systems* yang berbasis pada model bahasa besar (LLMs). Tujuan utama dari penelitian tersebut adalah untuk memberikan ringkasan tentang LLM yang ada dan pendekatan-pendekatan untuk mengadaptasi LLM pada tugas-tugas akhir (*downstream tasks*), selain itu penelitian juga menjelaskan kemajuan terbaru dalam *multi-turn dialogue systems*, yang mencakup sistem dialog domain terbuka (*open-domain*

 ISSN:2302-8149
e-ISSN:2540-9719

dialogue, ODD) berbasis LLM dan sistem dialog berorientasi tugas (*task-oriented dialogue*, TOD), serta dataset dan metrik evaluasi yang digunakan.

Berdarkan penjelasan di atas, meskipun LLM telah terbukti efektif dalam mengklasifikasikan dan menganalisis artikel ilmiah di bidang oftalmologi dan sosiologi. Penerapan metode *multi-turn prompting* dalam konteks kimia menawarkan tantangan yang lebih kompleks, mengingat kekhususan dan kedalaman data yang terlibat, mengingat artikel yang digunakan dalam jumlah yang besar. Untuk menganalisis artikel-artikel yang membahas pemanfaatan ML dalam kimia, LLM dengan *multi-turn prompting* dapat lebih menggali dan menyaring informasi secara bertahap. Oleh karena itu, penting untuk mengembangkan pendekatan *multi-turn prompting* yang lebih terfokus pada kimia, guna memaksimalkan potensi LLM dalam mengidentifikasi tren, tantangan, dan inovasi terbaru dalam pemanfaatan *machine learning* di bidang ini.

3 Metode Penelitian

Penelitian dilakukan dengan mengumpulkan judul artikel dengan kata kunci *machine learning* dan *chemistry*. Kemudian mengumpulkan *abstract* dari kesemua judul artikel yang didapat dengan mengunjungi *url* artikel dan mengekstrak data dengan teknik scraping. Data yang diperoleh kemudian di analisis dengan menggunakan chat-GPT untuk mendapatkan pikiran utama dari penelitian.

3.1. Pengumpulan *url* artikel

Pencarian artikel yang terkait dilakukan dengan menggunakan aplikasi Publish or Perish. Kata kunci yang digunakan adalah *Machine Learning Chemistry*. Hasil yang didapat dan digunakan adalah tahun artikel, judul, serta *url* artikel. *Url* artikel digunakan untuk mendapatkan abstract dari setiap artikel yang didapat yang disimpan dalam format csv.

3.2 *Scraping* data

Scraping terhadap web artikel dilakukan untuk mendapatkan abstract dari seriap artikel. Proses *scraping* dilakukan dengan menggunakan bahasa pemrograman python dengan memanfaatkan PyQt6 untuk meng-*compile* javascript serta Beautifulsoup untuk memparsing html dan mengambil data yang sesuai.

Masukkan data berupa file csv yang berisi *url* artikel yang akan diambil datanya. Terdapat dua tipe halaman artikel yang memiliki template yang berbeda. Kode scraping menyesuaikan tipe url dan template scraping yang sesuai. Hasil dari scraping ini disimpan dalam bentuk file csv. Adapun alur program sebagai berikut

```
Running_scrap
Open csv_file
Get urls
Initialization html
for url in urls
  parse url get domain as url_loc
  if url_loc == url_a
    html = get_html_url_a
  else if url_loc == url_a
    html = get_html_url_b
  end if
  abstract = find_abstract(compile(html))
end for
```

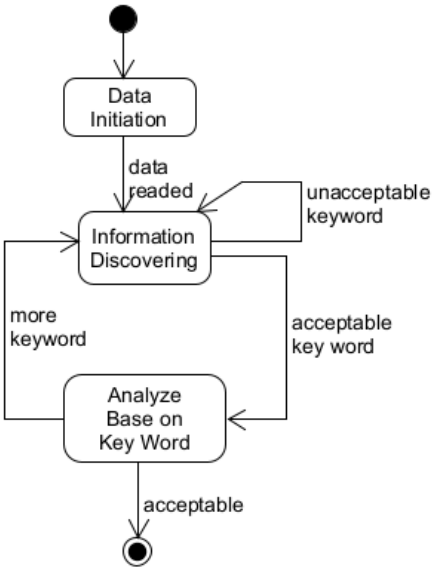
2.3. Analisis Text

Data yang didapat dianalisis secara tekstual dengan menggunakan chat-gpt 4.o. Ada beberapa prompt yang digunakan, yakni dapat dilihat pada Tabel 1.

Tabel 1. Beberapa *prompt* yang digunakan

No	Prompt
1.	Apa yang dapat di analisis dari file tersebut
2.	Berdasarkan data yang sudah diupload topik bidang kimia apa yang paling sering menggunakan machine learning
3.	berdasarkan hasil di atas buat tren tahun penelitian dan jumlah artikel yang mengenai struktur protein dan molekul, kimia kuantum dan penemuan obat
4.	Buatlah grafik tren sitasi per tahun.
5.	Analyze keyword trends by domain
6.	Buatkan grafik Grafik yang menunjukkan peningkatan frekuensi kata kunci terkait *machine learning* seperti "learning", "deep", "neural", "model", dan "AI" yang muncul dari waktu ke waktu
7.	Visualisasi distribusi kata kunci kimia terkait seperti "chemical", "compound", "reaction", dan "synthesis" dibandingkan dengan kata kunci *machine learning*.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1. *Prompting state*

Teknik *prompting* yang digunakan adalah dengan menggunakan pendekatan *Multi-Turn Prompting*, informasi didapatkan berdasarkan data yang ada, kemudian menganalisis berdasarkan informasi tersebut. Adapun state prompting dapat dilihat pada Gambar 1.

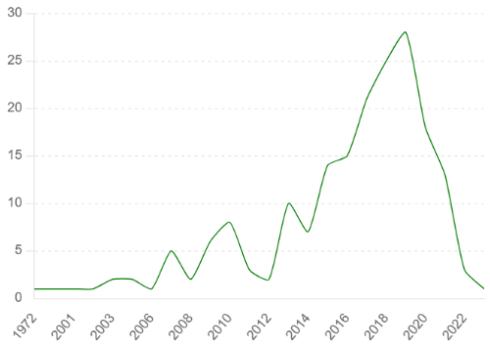
Terdapat tiga *state* dalam menganalisis data artikel yang digunakan, yakni inisiasi data, penggalian informasi dan analisis berdasarkan kata kunci. Pada *state* inisiasi data dilakukan proses menginput data dan memerintahkan aplikasi LLM untuk membacar dan melihat infomrasi apa saja yang ada pada data yang dimasukkan, ada pun contoh *prompt* yang digunakan adalah "Analisis teks pada file berikut, informasi apa saja yang tercantum pada file tersebut".

State penggalian informasi dilakukan untuk mendapatkan kata kunci yang dapat digunakan untuk menganalisis data. Adapun beberapa contoh *prompt* seperti, "Berdasarkan teks tersebut hal apa saja yang dapat di analisis?", "Topik apa saja yang di bahas pada teks tersebut?", "Extrak kata kunci bidang kimia pada data tersebut", dan "Bidang kimia apa saja yang dibahas pada data tersebut".
Pada *state* analisis dilakukan upaya untuk menganalisis berdasarkan kata kun ci yang diharapkan. Adapun beberapa contoh *prompt* yakni "Analisis jumlah kata [*keyword*] pada teks berdasarkan tahun", "Analisis topik [*keyword*] pada teks", dan " Analisis [*keyword*] berdasarkan [*keyword*].

Proses *promping* dilakukan secara manual dan membutuhkan penilaian dari pengguna untuk menentukan apakah hasil *promting* dapat digunakan atau tidak. Begitu juga dengan proses untuk membuat menentukan kalimat pada *prompt* juga dilakukan secara manual sesuai dengan respon dari *prompting* sebelumnya.

4. Hasil dan Pembahasan

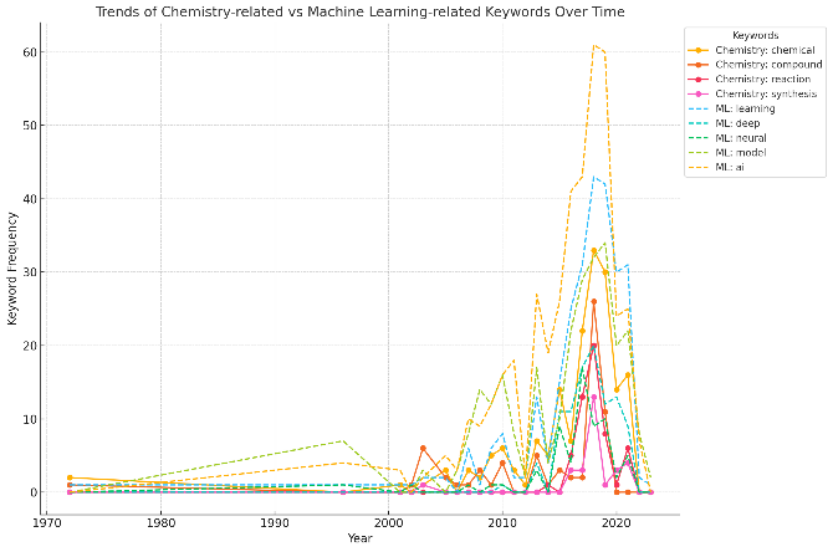
Hasil pengumpulan judul artikel dengan kata kunci *machine learning* dan *chemistry* yang diambil dari basis data Scopus mendapatkan 200 artikel dengan data penting be rupa tahun, judul, h-index dan url artikel. Judul artikel yang didapa dari tahun 1972 sampai 2023. Gambar 2 menunjukkan grafik terbitan artikel yang memuat kata kunci ML dan Kimia pertahun.



Gambar 2. Grafik jumlah artikel ML dalam kimia pertahun

 ISSN:2302-8149
e-ISSN:2540-9719

Analisis mengungkapkan bahwa sekitar 91,5% abstrak yang menyebutkan kata "chemical" juga menyertakan istilah terkait pembelajaran mesin (seperti "machine learning", "deep learning", "neural networks", "AI", atau "model"). Secara khusus, 86 dari 94 abstrak yang menyebutkan "chemical" juga mengandung referensi ke *machine learning*.



Gambar 3. Tren machine learning dan kimia dengan kata kunci terkait dari waktu ke waktu

Gambar 3 membandingkan tren kata kunci yang terkait dengan kimia (seperti *chemical, compound, reaction,* dan *synthesis*) dengan kata kunci terkait dengan *machine learning* (seperti *learning, deep, neuralm model*, dan *AI*). Dari gambar terlihat bagaimana frekuensi istilah tersebut semakin meningkat dari waktu ke waktu.

3.1 Tren Penelitian *Machine Learning* dalam Kimia

Berdasarkan data yang diperoleh, penelitian mengenai kimia yang memanfaatkan *machine learning* dimulai pada tahun 1972 serta mulai naik pada tahun 2006 dan naik pesat pada mulai pada tahun 2010 serta mencapai puncak jumlah penelitian yang paling banyak adalah 2019. Hal ini terkait dengan terjadi covid 19, dimana penggunaan komputer dalam bidang kimia yang semakin besar, karena tuntutan kecepatan dalam mengsimulasikan dengan pendekatan komputerisasi dibandingkan dengan secara manual yang membutuhkan waktu yang lebih lama.

Artikel pertama yang dipublikasikan dari data artikel yang didapat adalah dengan judul *Pattern Recognition.' A Powerful Approach to Interpreting Chemical Data*. Artikel tersebut membahas mengenai penggunaan data kimia melalui pendekatan *pattern recognition*. Penggunaan *pattern recognition* dalam analisis data kimia memungkinkan para ilmuan untuk memproses dan menafsirkan data yang kompleks dengan lebih efektif, serta membuat prediksi yang tidak dapat dilakukan hanya dengan analisis konvensional. Salah satu metode di dalam *pattern recognition* adalah *learning machine* yang telah berhasil diterapkan pada data spektroskopi untuk mendeteksi unit struktural molekul secara langsung [15].

Dari hasil analisis, topik kimia yang paling sering dikaitkan dengan *machine learning* dalam penelitian diantaranya dapaat dilihat pada Tabel 2:

Tabel 2. Jumlah kata topik kimia yang muncul

Kata	Jumlah Muncul (artikel)
chemical	138
protein	119
drug	107
structure	100
molecular	96
chemistry	91
quantum	84

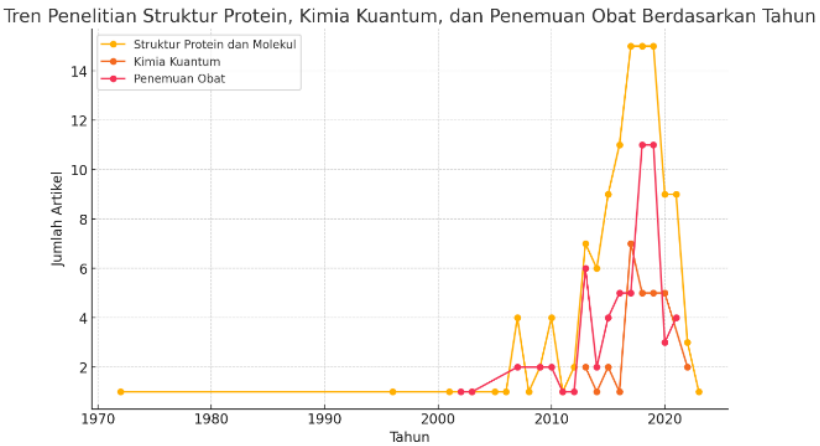
 ISSN:2302-8149
e-ISSN:2540-9719

Dengan mengabaikan kata chemical dan chemistry, terdapat lima kata kunci yang mendominasi penelitian kimia yang melibatkan *machine learning*, yakni protein, *drug*, *structure*, *molecular* dan *quantum*. Topik-topik ini menunjukkan bahwa machine learning sering digunakan dalam penelitian terkait kimia, terutama di bidang struktur protein dan molekul, yang relevan untuk pemodelan protein dan penemuan obat; kimia kuantum, yang merupakan bidang yang berkembang pesat menggunakan machine learning untuk simulasi dan prediksi; serta penemuan obat, di mana machine learning membantu memprediksi sifat kimia dan farmakologi molekul baru

Adapun tren ketiga topik tersebut dapat dilihat pada Gambar 4. Grafik tersebut menunjukkan tren penelitian di bidang struktur protein dan molekul, kimia kuantum, serta penemuan obat berdasarkan jumlah artikel yang diterbitkan per tahun. Terlihat bahwa ketiga topik naik dari tahun ke tahun. Puncak jumlah artikel yang paling banyak untuk ketiga kata kunci tersebut adalah dalam rentan waktu 2019 sampai 2020. Tren tersebut menunjukkan bahwa ketiga bidang tersebut terus berkembang, dengan kontribusi signifikan dari metode *machine learning* dalam penelitian kimia.

Tren struktur protein dan molekul menunjukkan peningkatan yang konsisten, mencerminkan minat yang terus berkembang di bidang ini. Terlihat juga bahwa tren struktur dan molekul lebih mendominasi dari kedua topik yang lain tiap tahunnya. Artikel pertama yang memuat ini adalah dengan judul *Pattern Recognition.' A Powerful Approach to Interpreting Chemical Data* yang diterbitkan pada tahun 1972[15].

Kimia kuantum meskipun fluktuatif, ada peningkatan penelitian dalam beberapa tahun terakhir, sejalan dengan perkembangan teknologi dan metode simulasi. Walaupun topik ini terlihat lebih sedikit dari topik yang lain. Berdasarkan dari artikel yang didapat topik ini pertama kali muncul pada dua artikel pada tahun 2013, yakni pada dengan judul *Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD)* dan *Machine learning of molecular electronic properties in chemical compound space* [16][17].



Gambar 4. Tren penelitian struktur protein, kimia kuantum dan penemuan obat

OQMD adalah basis data besar yang berisi lebih dari 200.000 struktur kristal yang dihitung menggunakan DFT [16]. Basis data ini dirancang untuk mendukung berbagai aplikasi dalam desain material, termasuk pencarian material baru dan pengembangan model yang lebih tinggi melalui penambangan data (data mining). DFT throughput tinggi dan basis data seperti OQMD dapat menjadi alat yang kuat dalam mendesain dan menemukan material baru. Artikel ini juga menyoroti pentingnya penggunaan data mining dan pembelajaran mesin dalam memanfaatkan basis data besar untuk memprediksi dan menguji material baru dengan lebih efisien.

Penelitian kedua ini memperkenalkan model pembelajaran mesin yang dilatih menggunakan hasil perhitungan *ab initio* untuk ribuan molekul organik [17]. Model ini memprediksi berbagai properti elektronik, termasuk energi atomisasi, polarizabilitas, nilai eigen orbital perbatasan (HOMO dan LUMO), potensial ionisasi, afinitas elektron, dan energi eksitasi. Hasil pelatihan menunjukkan bahwa model ini mampu memprediksi properti molekul yang tidak termasuk dalam set pelatihan dengan akurasi yang sebanding, dan dalam beberapa kasus bahkan lebih baik, dibandingkan metode kimia kuantum modern. Akurasi model meningkat seiring dengan bertambahnya jumlah molekul dalam set pelatihan .

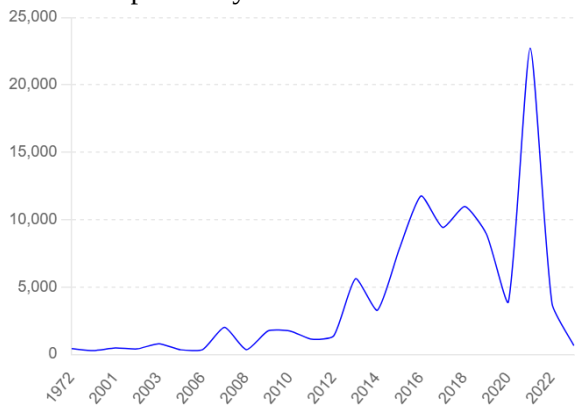
 ISSN:2302-8149
e-ISSN:2540-9719

Topik penemuan obat mengalami peningkatan yang signifikan, terutama dengan penerapan *machine learning* dalam penemuan dan pengembangan farmasi. Topik ini lebih sedikit dari struktur protein dan molekul dikarenakan penelitian mengenai penemuan obat juga berkaitan dengan struktur protein dan molekul. Topik ini awal muncul pada tahun 2002 dengan judul *Prediction of 'drug-likeness'* [18].

Artikel tersebut membahas berbagai metode komputasi yang digunakan untuk memprediksi "*drug-likeness*", yaitu kemampuan suatu molekul untuk berfungsi sebagai obat yang efektif [18]. Artikel tersebut juga mengeksplorasi tantangan yang dihadapi dalam bidang ini dan menyoroti berbagai teknik yang digunakan, mulai dari metode penghitungan sederhana hingga teknik pembelajaran mesin yang lebih canggih. Penulis menyimpulkan bahwa meskipun banyak kemajuan telah dicapai dalam memprediksi "drug-likeness", bidang ini masih dalam tahap perkembangan dan banyak tantangan yang harus diatasi. Mereka juga menekankan pentingnya mengembangkan set data yang lebih besar dan lebih konsisten untuk meningkatkan akurasi prediksi di masa depan.

3.3 Tren Sitasi

Pada data terdapat informasi mengenai jumlah sitasi masing-masing artikel. Berdasarkan data tersebut kita dapat melihat bagaimana tren sitasi untuk penelitian *machine learning* dalam ilmu kimia. Gambar 5 menunjukkan sitasi jumlah sitasi tiap tahunnya.



Gambar 5. Sitasi per tahun

Terlihat bahwa dari tahun 2013 jumlah sitasi mulai naik hinnga puncaknya pada tahun 2021. Hal ini menunjukkan kontribusi penelitian mengenai penggunaan *machine learning* dalam ilmu kimia semakin meningkan tiap tahunnya. Dari data artikel yang paling banyak disitasi adalah dengan judul *Highly accurate protein structure prediction with AlphaFold* oleh J. Jumper sebanyak 16591 sitasi [19].

Artikel tersebut menjelaskan mengenai AlphaFold yang merupakan pendekatan komputasional yang dirancang untuk memprediksi struktur protein dengan akurasi mendekati tingkat eksperimen. Ia juga menekankan pentingnya AlphaFold dalam revolusi bioinformatika struktural. Dengan kemampuannya untuk memprediksi struktur protein dengan tingkat akurasi tinggi [19]. AlphaFold diharapkan dapat mendorong kemajuan signifikan dalam pemahaman kita tentang protein dan aplikasinya dalam biologi dan kedokteran.

Hasil semua data penelitian menunjukkan pergeseran fokus penelitian kimia menuju pendekatan berbasis *machine learning*, yang mampu mempercepat proses penemuan obat, pemodelan struktur protein, dan simulasi kimia kuantum dimana penelitian tersebut terus naik dari tahun ke tahun sampai pada puncaknya pada tahun 2019. Dalam konteks penemuan obat, *machine learning* membantu mengidentifikasi kandidat senyawa potensial secara efisien, sementara dalam pemodelan struktur protein, teknologi seperti AlphaFold telah merevolusi cara para ilmuwan memahami konfigurasi tiga dimensi protein dengan akurasi tinggi. Selain itu, penerapan *machine learning* dalam simulasi kimia kuantum berpotensi menghasilkan prediksi yang lebih presisi mengenai sifat material dan reaksi kimia, yang menjadi landasan penting bagi inovasi di berbagai bidang sains terapan.

Masih banyak peluang untuk pengembangan *machine learning*, khususnya dalam kimia kuantum, yang menawarkan potensi besar untuk mendorong terobosan dalam pemahaman materi pada level molekuler dan atom. Hal ini disebabkan oleh kenyataan bahwa topik ini baru mulai dipublikasikan pada tahun 2013, menjadikannya sebagai bidang yang relatif baru dibandingkan

 ISSN:2302-8149
e-ISSN:2540-9719

dengan aplikasi *machine learning* lainnya. Jumlah artikel yang dipublikasikan tiap tahun masih relatif sedikit, menunjukkan bahwa bidang ini masih dalam tahap awal eksplorasi dan pengembangan. Meskipun demikian, perkembangan yang terjadi dalam beberapa tahun terakhir menunjukkan adanya minat yang semakin besar terhadap penerapan *machine learning* dalam kimia kuantum. Dengan terus meningkatnya pemahaman dan kapabilitas teknologi *machine learning*, khususnya dalam hal pengolahan data dan simulasi kuantum, diharapkan kita dapat menyaksikan pertumbuhan yang lebih pesat dalam penelitian ini, yang pada akhirnya dapat menghasilkan model-model yang lebih akurat untuk prediksi sifat-sifat material, reaksi kimia, dan interaksi molekuler yang sebelumnya sulit dipahami dengan metode konvensional.

Penelitian ini memiliki keterbatasan karena hanya menganalisis 200 artikel yang diperoleh melalui teknik *scraping* data berdasarkan kata kunci tertentu. Meskipun artikel-artikel tersebut mencakup berbagai topik terkait *machine learning* dan kimia, penggunaan kata kunci yang terbatas dapat mengakibatkan ketidaklengkapan dalam mencakup seluruh publikasi yang relevan, sehingga hasil analisis mungkin tidak mencerminkan gambaran yang sepenuhnya representatif dari tren dan perkembangan penelitian di bidang ini. Selain itu, analisis yang dilakukan menggunakan *Large Language Model* (LLM) dengan teknik *multi-turn prompting* memiliki potensi bias yang tidak dapat dihindari. Bias ini bisa muncul tergantung pada formulasi prompt yang digunakan, yang dapat memengaruhi cara model memproses dan menghasilkan informasi. Variasi dalam penyusunan pertanyaan atau instruksi kepada model dapat menghasilkan interpretasi yang berbeda, sehingga hasil yang diperoleh tidak selalu konsisten atau objektif. Oleh karena itu, untuk mendapatkan gambaran yang lebih akurat dan menyeluruh, penelitian lanjutan dengan memperluas jumlah artikel yang dianalisis dan mengoptimalkan teknik *prompting* sangat diperlukan.

Penelitian lanjutan sebaiknya mencakup lebih banyak artikel, dengan memperluas cakupan pencarian untuk mencakup publikasi yang lebih beragam dan relevan. Penggunaan database yang lebih komprehensif dan pengumpulan data dari berbagai sumber, seperti jurnal internasional, repositori penelitian, dan konferensi, dapat membantu memberikan gambaran yang lebih holistik mengenai perkembangan terkini di bidang machine learning dalam kimia. Untuk meningkatkan kualitas data yang dianalisis, penting untuk memperbaiki metode *scraping* dengan memperhitungkan kata kunci yang lebih beragam dan relevan. Penggunaan teknik semantik yang lebih canggih, seperti analisis topik, bisa membantu dalam menangkap artikel yang lebih tepat dan kontekstual, meskipun menggunakan kata kunci tertentu. Selain itu, perlu dilakukan analisis terhadap seluruh isi dari artikel tidak hanya pada abstrak.

4 Kesimpulan

Berdasarkan hasil analisis penelitian bidang kimia terkait dengan *machine learning* saat ini terdapat tiga topik kimia utama yang mendominasi, yakni struktur protein dan molekul, kimia kuantum, dan penemuan obat. Dapat dipastikan bahwa ketiga topik tersebut dapat memanfaatkan *machine learning* dalam proses penelitiannya. Terbitan artikel mengenai *machine learning* dalam bidang kimia mulai naik pada pada tahun 2012 dan naik signifikan pada tahun 2019. Serta puncak sitasi adalah pada tahun 2021. Saat ini artikel yang paling banyak disitasi adalah topik penggunaan AlphaFold yang mampu memprediksi struktur protein dengan akurasi yang tinggi. Berdasarkan hasil yang diperoleh masih banyak peluang pengembangan *machine learning* dalam bidang kimia khusunya bidang kimia kuantum yang baru terpublikasikan pada tahun 2013 dan jumlah artikel yang dipublikasikan tidak terlalu besar ditiap tahunnya. Jumlah artikel yang dipublikasikan tiap tahun masih relatif sedikit, menunjukkan bahwa bidang ini masih dalam tahap awal eksplorasi. Selain itu masih terbuka peluang pemanfaatan *machine learning* terkait topik kimia selain ketiga topik utama tersebut. Hasil penelitian ini dapat dijadikan referensi dan gambaran topik bagi penggiat *machine learning* untuk menggali dan melakukan penelitian bidang kimia. Adapun keterbatasan dalam penelitian ini tidak mencakup seluruh artikel publikasi yang relevan, serta sangat bergantung pada formulasi *prompt* yang digunakan. Pada penelitian selanjutkan disarankan untuk mendapatkan data yang lebih komprehensif dan membandingkan hasil dengan teknik analisis yang berbeda, serta tidak hanya menganalisis pada abstrak pada artikel namun pada seluruh tulisan pada isi artikel.

 ISSN:2302-8149
e-ISSN:2540-9719

Referensi

[1] A. Alanazi, "*Using Machine Learning for Healthcare Challenges and Opportunities*," *Inform Med Unlocked*, vol. 30, p. 100924, Jan. 2022, doi: 10.1016/J.IMU.2022.100924.

[2] P. Vats and K. Samdani, "*Study on Machine Learning Techniques in Financial Markets*," *2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019*, Mar. 2019, doi: 10.1109/ICSCAN.2019.8878741.

[3] A. Soni, D. Dharmacharya, A. Pal, V. Kumar Srivastava, R. N. Shaw, and A. Ghosh, "*Design of a Machine Learning-based Self-Driving Car*," *Studies in Computational Intelligence*, vol. 960, pp. 139–151, 2021, doi: 10.1007/978-981-16-0598-7_11.

[4] X. Wan *et al.*, "*Machine Learning Paves the Way for High Entropy Compounds Exploration: Challenges, Progress, and Outlook*," *Advanced Materials*, p. 2305192, 2023, doi: 10.1002/ADMA.202305192.

[5] S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, and M. J. Ahsan, "*Machine Learning in Drug Discovery: A Review*," *Artificial Intelligence Review 2021 55:3*, vol. 55, no. 3, pp. 1947–1999, Aug. 2021, doi: 10.1007/S10462-021-10058-4.

[6] N. Fred and I. O. Temkin, "*A Systematic Literature Review on the use of Machine Learning in Software Engineering*," Jun. 2024, Accessed: Jan. 28, 2025. [Online]. Available: https://arxiv.org/abs/2406.13877v1

[7] K. Kolasa, B. Admassu, M. Hołownia-Voloskova, K. J. Kędzior, J. E. Poirrier, and S. Perni, "*Systematic Reviews of Machine Learning in Healthcare*: *A Literature Review," Expert Rev Pharmacoecon Outcomes Res*, vol. 24, no. 1, pp. 63–115, Jan. 2024, doi: 10.1080/14737167.2023.2279107.

[8] Y. F. Shi *et al.*, "*Machine Learning for Chemistry: Basics and Applications*," *Engineering*, vol. 27, pp. 70–83, Aug. 2023, doi: 10.1016/J.ENG.2023.04.013.

[9] J. Sawicki, M. Ganzha, and M. Paprzycki, "*The State of the Art of Natural Language Processing—A Systematic Automated Review of Nlp Literature using Nlp Tec*hniques," *Data Intell*, vol. 5, no. 3, pp. 707–749, Aug. 2023, doi: 10.1162/DINT_A_00213.

[10] T. Labruna, J. A. Campos, and G. Azkune, "When to Retrieve: *Teaching Llms to Utilize Information Retrieval Effectively*," Apr. 2024, Accessed: Nov. 07, 2024. [Online]. Available: https://arxiv.org/abs/2404.19705v2

[11] N. R. Ruchitaa Raj, S. Nandhakumar Raj, and M. Vijayalakshmi, "*Web Scrapping Tools and Techniques: A Brief Survey*," *2023 International Conference on Innovative Trends in Information Technology, ICITIIT 2023*, 2023, doi: 10.1109/ICITIIT57246.2023.10068666.

[12] H. Raja *et al.*, "*Using Large Language Models to Automate Category and Trend Analysis of Scientific Articles: an Application in Ophthalmology*," Aug. 2023, Accessed: Jan. 09, 2025. [Online]. Available: https://arxiv.org/abs/2308.16688v1

[13] Y. (YJ) Chae and T. Davidson, "*Large Language Models for Text Classification: from Zero-Shot Learning to Instruction-Tuning*," Aug. 2023, doi: 10.31235/OSF.IO/STHWK.

[14] Z. Yi, J. Ouyang, Y. Liu, T. Liao, Z. Xu, and Y. Shen, "*A Survey on Recent Advances in Llm-based Multi-Turn Dialogue Systems*," Feb. 2024, Accessed: Jan. 09, 2025. [Online]. Available: https://arxiv.org/abs/2402.18013v1

[15] B. R. Kowalski and C. F. Bender, "*Pattern Recognition.1 a Powerful Approach to Interpreting Chemical Data*," *J Am Chem Soc*, vol. 94, no. 16, pp. 5632–5639, Aug. 1972, doi: 10.1021/JA00771A016/ASSET/JA00771A016.FP.PNG_V03.

[16] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "*Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)*," *JOM*, vol. 65, no. 11, pp. 1501–1509, Nov. 2013, doi: 10.1007/S11837-013-0755-4/METRICS.

[17] G. Montavon *et al.*, "*Machine Learning of Molecular Electronic Properties in Chemical Compound Space*," *New J Phys*, vol. 15, May 2013, doi: 10.1088/1367-2630/15/9/095003.

[18] W. P. Walters and M. A. Murcko, "*Prediction of 'Drug-Likeness*,'" *Adv Drug Deliv Rev*, vol. 54, no. 3, pp. 255–271, Mar. 2002, doi: 10.1016/S0169-409X(02)00003-0.

[19] J. Jumper *et al.*, "*Highly Accurate Protein Structure Prediction with Alphafold*," *Nature 2021 596:7873*, vol. 596, no. 7873, pp. 583–589, Jul. 2021, doi: 10.1038/s41586-021-03819-2.

ISSN:2302-8149
e-ISSN:2540-9719