

Tinjauan Literatur Sistematis terhadap *Text-to-SQL* Berbasis Model Bahasa Besar untuk Akses Basis Data pada Analitik Olahraga

Roni Habibi^{*1}, Mochammad Rayfan Aqbillah²

^{1,2}Sarjana Terapan Teknik Informatika, Universitas Logistik dan Bisnis Internasional, Indonesia
Email: ¹roni.habibi@ulbi.ac.id, ²mochammadrayfanaqbil@gmail.com

Abstrak

Analitik olahraga berkembang dari analisis statistik deskriptif menuju sistem berbasis kecerdasan buatan yang mendukung pemantauan performa, prediksi hasil, dan pengambilan keputusan strategis. Namun, akses terhadap data olahraga yang tersimpan dalam basis data relasional masih menjadi tantangan bagi pengguna non-teknis karena membutuhkan pemahaman struktur basis data dan *query SQL*. Penelitian ini menyajikan *Systematic Literature Review (SLR)* mengenai *Large Language Models (LLMs)* untuk *Natural Language Database Access* dalam *Sports Analytics Systems*. Kajian ini menggunakan protokol *PRISMA* dengan basis data *Scopus* sebagai sumber utama. Dari 239 artikel awal, diperoleh 65 artikel *Scopus* yang layak dianalisis dan 1 artikel tambahan sebagai pendukung. Hasil kajian menunjukkan adanya peningkatan tren publikasi pada periode 2020-2026, dengan jumlah tertinggi pada tahun 2025. Area studi paling dominan adalah *Sports Analytics*, diikuti *Natural Language Database Access*, *Healthcare/Wearable Analytics*, *Business Analytics*, dan *Urban/Environmental Analytics*. Literatur yang dikaji menunjukkan relevansi *Text-to-SQL*, *Speech-to-SQL*, *LLM*, *wearable sensor*, dan *query generation* dalam mendukung akses serta analisis data olahraga. Meskipun demikian, masih terdapat tantangan berupa generalisasi model, kompleksitas skema basis data, interpretabilitas, integrasi data multimodal, privasi data atlet, dan bias algoritma. Arah pengembangan selanjutnya diarahkan pada arsitektur berbasis *LLM-agent* yang mengintegrasikan *intent routing*, *safety guards*, *query builder*, dan *AI response generation* untuk mendukung akses bahasa alami terhadap data *sports analytics*. Penelitian ini memiliki keterbatasan pada aspek teknis karena proses pencarian literatur hanya dilakukan melalui satu basis data utama, yaitu *Scopus*. Akibatnya, terdapat kemungkinan bahwa sejumlah publikasi yang relevan dan terindeks pada basis data lain belum terakomodasi dalam analisis.

Kata kunci: akses basis data bahasa alami, *large language models*, *sports analytics*, *systematic literature review*, *text-to-SQL*

Systematic Literature Review of Large Language Model-Based Text-to-SQL for Database Access in Sports Analytics

Abstract

Sports analytics has shifted from descriptive statistical analysis toward artificial intelligence-based systems that support performance monitoring, outcome prediction, and strategic decision-making. However, accessing sports data stored in relational databases remains challenging for non-technical users because it requires knowledge of database structures and SQL queries. This study presents Systematic Literature Review (SLR) on Large Language Models (LLMs) for Natural Language Database Access in Sports Analytics Systems. The review follows the PRISMA protocol and uses Scopus as the main data source. From 239 initial articles, 65 Scopus articles were selected for analysis, supported by 1 additional articles from other sources. The results show an increasing publication trend from 2020 to 2026, with the highest number of studies in 2025. Sports Analytics is the dominant study area, followed by Natural Language Database Access, Healthcare/Wearable Analytics, Business Analytics, and Urban/Environmental Analytics. The reviewed literature highlights the relevance of Text-to-SQL, Speech-to-SQL, LLMs, wearable sensors, and query generation in supporting access to sports data analysis. However, several challenges remain, including model generalization, database schema complexity, interpretability, multimodal data integration, athlete data privacy, and algorithmic bias. Future work is directed toward an LLM-agent-based architecture that integrates intent routing, safety guards, query builder, and AI response generation to support natural language access to sports analytics data. This study has a technical limitation in that the literature search was conducted using only one primary database, Scopus. Consequently, some relevant publications indexed in other databases may not have been included in the analysis.

Keywords: *large language models, natural language database access, sports analytics, systematic literature review, text-to-SQL*

1. PENDAHULUAN

Perkembangan analitik olahraga dalam beberapa tahun terakhir menunjukkan pergeseran penting dari penggunaan statistik deskriptif menuju sistem analitik berbasis kecerdasan buatan yang lebih adaptif. Data olahraga saat ini tidak hanya berasal dari catatan pertandingan, tetapi juga dari perangkat *wearable*, sensor biomekanik, sistem pelacakan pemain, video pertandingan, dan basis data historis yang terus bertambah. Kondisi tersebut membuat proses pengambilan keputusan dalam olahraga semakin bergantung pada kemampuan sistem untuk mengolah data yang besar, beragam, dan kontekstual. Perkembangan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa analitik olahraga telah berkembang dari pemodelan statistik dan prediksi performa menuju pendekatan komputasional yang lebih kompleks untuk mendukung evaluasi atlet, strategi tim, dan pengambilan keputusan berbasis data [1], [2].

Di sisi lain, akses terhadap data olahraga yang tersimpan dalam basis data relasional masih menjadi tantangan bagi pengguna non-teknis, seperti pelatih, analis performa, manajer tim, maupun peneliti olahraga. Informasi praktis seperti tren performa atlet, perbandingan hasil pertandingan, catatan waktu, atau pola pencapaian tertentu sering kali membutuhkan pemahaman terhadap struktur basis data dan kemampuan menulis *query SQL*. Keterbatasan tersebut dapat menghambat pemanfaatan data olahraga secara optimal, terutama ketika pengguna membutuhkan informasi secara cepat untuk mendukung analisis dan pengambilan keputusan.

Salah satu pendekatan yang berkembang untuk mengatasi masalah tersebut adalah *Natural Language Interface to Database (NLIDB)* dan *Text-to-SQL*. Pendekatan ini memungkinkan pengguna mengakses basis data melalui pertanyaan bahasa alami yang kemudian diterjemahkan menjadi *query SQL*. Dalam konteks ini, *Large Language Models (LLMs)* memiliki potensi besar karena mampu memahami maksud pertanyaan, menghubungkan pertanyaan dengan skema basis data, serta menghasilkan *query* yang relevan. Penelitian terbaru menunjukkan bahwa *LLM* semakin banyak digunakan dalam tugas *Text-to-SQL* karena mampu meningkatkan pemahaman skema dan membantu pembentukan *query* yang lebih kompleks [3].

Perkembangan *Text-to-SQL* juga menunjukkan pergeseran dari pendekatan berbasis aturan menuju pendekatan berbasis model bahasa yang lebih fleksibel [4]. Namun, sistem berbasis *LLM* masih menghadapi beberapa tantangan, seperti ambiguitas pertanyaan, kompleksitas relasi antartabel, generalisasi terhadap skema basis data baru, dan evaluasi akurasi keluaran *SQL* [5]. Tantangan tersebut menjadi semakin penting dalam konteks analitik olahraga karena data olahraga memiliki karakteristik multidimensi, dinamis, dan spesifik terhadap cabang olahraga tertentu.

Meskipun analitik olahraga dan *Text-to-SQL* sama-sama berkembang pesat, integrasi antara *LLM*, akses basis data berbasis bahasa alami, dan sistem analitik olahraga masih memerlukan pemetaan yang lebih sistematis. Sebagian penelitian berfokus pada peningkatan performa atlet, prediksi pertandingan, *wearable sensor*, atau analisis taktik. Sebagian lainnya membahas *Text-to-SQL*, *NLIDB*, dan peningkatan kinerja *LLM* pada basis data umum. Namun, kajian yang secara khusus merangkum keterkaitan antara *LLM*, akses basis data berbasis bahasa alami, dan sistem analitik olahraga masih terbatas. Padahal, integrasi ketiga area tersebut penting untuk membangun sistem analitik olahraga yang tidak hanya akurat, tetapi juga mudah diakses oleh pengguna yang tidak memiliki latar belakang teknis. Sebagai perbandingan, penelitian tinjauan sebelumnya cenderung mengkaji *LLM* dan *Text-to-SQL* secara terpisah tanpa mengaitkannya dengan domain olahraga [3], [4], [19]. Sementara itu, kajian pada bidang *sports analytics* lebih banyak menyoroti pemanfaatan *AI* dan teknologi sensor untuk analisis performa, namun belum membahas mekanisme akses basis data menggunakan bahasa alami [9], [11], [15], [16]. Berdasarkan hasil penelusuran yang dilakukan, belum ditemukan *SLR* yang mengintegrasikan ketiga area tersebut dalam satu kerangka kajian yang komprehensif. Oleh karena itu, aspek integrasi tersebut menjadi kontribusi kebaruan utama dalam penelitian ini.

Berdasarkan permasalahan tersebut, penelitian ini menyusun *Systematic Literature Review (SLR)* mengenai penggunaan *Large Language Models* untuk *Natural Language Database Access* dalam *Sports Analytics Systems*. Kajian ini menggunakan basis data *Scopus* sebagai sumber utama dengan rentang publikasi tahun 2020-2026. Fokus kajian diarahkan pada tren publikasi, distribusi negara asal penelitian, pendekatan metodologis, area studi, metode analisis, variabel yang dominan, serta hubungan antarkonsep yang muncul dalam literatur. Dengan demikian, artikel ini tidak hanya merangkum perkembangan penelitian, tetapi juga mengidentifikasi *gap* riset dan peluang pengembangan sistem analitik olahraga berbasis *LLM*.

Tujuan dari penelitian ini adalah untuk memetakan perkembangan penelitian mengenai *LLM* dan *Text-to-SQL* pada analitik olahraga, mengidentifikasi pendekatan metodologis dan area studi yang dominan, serta merumuskan *gap* riset dan arah pengembangan masa depan. Hasil kajian ini diharapkan dapat memberikan

gambaran komprehensif mengenai posisi penelitian saat ini serta menjadi dasar bagi pengembangan sistem *sports analytics* yang lebih interaktif, kontekstual, dan mudah digunakan melalui akses basis data berbasis bahasa alami.

Kontribusi utama artikel ini adalah menyajikan pemetaan literatur yang menghubungkan tiga area penting, yaitu *Large Language Models*, *Natural Language Database Access*, dan *Sports Analytics Systems*. Kajian ini diharapkan dapat membantu peneliti dalam memahami arah perkembangan riset, membantu praktisi melihat potensi penerapan *LLM* untuk analitik olahraga, serta memberikan rujukan awal bagi pengembangan sistem yang memungkinkan pengguna mengakses data olahraga melalui pertanyaan bahasa alami. Selain itu, kajian ini juga menyiapkan dasar pembahasan mengenai peluang integrasi *LLM* dengan basis data olahraga, termasuk tantangan pada kualitas data, evaluasi *query*, interpretabilitas model, privasi data atlet, dan kebutuhan validasi pada konteks olahraga nyata.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Systematic Literature Review (SLR)* untuk memetakan perkembangan penelitian mengenai *Large Language Models (LLMs)*, *Natural Language Interface to Database (NLIDB)*, *Text-to-SQL*, dan penerapannya dalam *sports analytics*. Pendekatan *SLR* digunakan karena mampu menyajikan proses identifikasi, seleksi, evaluasi, dan sintesis literatur secara sistematis. Alur seleksi artikel dalam penelitian ini disusun dengan mengacu pada protokol *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*, sehingga proses penyaringan literatur dapat ditelusuri secara transparan.

Sumber data utama dalam penelitian ini adalah basis data *Scopus*. *Scopus* dipilih karena menyediakan artikel ilmiah bereputasi dan memiliki cakupan multidisipliner yang relevan dengan topik analitik olahraga, kecerdasan buatan, pemrosesan bahasa alami, dan sistem basis data. Periode publikasi yang digunakan dibatasi pada tahun 2020-2026 agar kajian berfokus pada perkembangan terbaru, terutama setelah meningkatnya penggunaan *LLM* dalam pemrosesan bahasa alami dan akses basis data.

Strategi pencarian artikel dilakukan dengan menggunakan beberapa kata kunci yang relevan dengan fokus penelitian. Kata kunci yang digunakan meliputi “*Large Language Model Text-to-SQL*”, “*Natural Language Interface to Database*”, “*Text-to-SQL Large Language Models*”, “*Natural Language to SQL*”, “*Sports Analytics*”, dan “*Sports Performance Analytics*”. Penggunaan kata kunci tersebut bertujuan untuk memperoleh artikel yang berkaitan dengan dua area utama, yaitu akses basis data berbasis bahasa alami dan analitik olahraga.

Kriteria inklusi dalam penelitian ini mencakup artikel yang terindeks *Scopus*, diterbitkan pada periode 2020-2026, berada pada kategori jurnal Q1 sampai Q4, memiliki abstrak, tersedia dalam bentuk *full-text*, dan relevan dengan topik *LLM*, *Text-to-SQL*, *natural language database access*, atau *sports analytics*. Sementara itu, kriteria eksklusi mencakup artikel duplikat, artikel di luar rentang tahun penelitian, artikel yang tidak memenuhi kategori Q1-Q4, artikel tanpa abstrak, artikel yang tidak relevan dengan fokus kajian, serta artikel yang tidak dapat diakses secara *full-text*.

Proses seleksi literatur dilakukan secara bertahap. Pencarian awal pada basis data *Scopus* menghasilkan 239 artikel. Pada tahap identifikasi, sebanyak 36 artikel dihapus karena duplikasi, 56 artikel dikeluarkan karena berada di luar periode 2020-2026, 28 artikel tidak memenuhi kriteria kategori Q1-Q4, dan 2 artikel tidak memiliki abstrak. Setelah tahap tersebut, terdapat 117 artikel yang masuk ke proses *screening*. Dari jumlah tersebut, 21 artikel dikeluarkan karena tidak relevan dengan fokus penelitian. Selanjutnya, 96 artikel dicari dalam bentuk *full-text*, tetapi 31 artikel tidak tersedia secara penuh. Dengan demikian, terdapat 65 artikel *Scopus* yang layak dianalisis. Selain itu, dari dua artikel tambahan yang diperoleh melalui penelusuran sumber lain, hanya satu artikel yang memenuhi kriteria kelayakan dan digunakan sebagai literatur pendukung. Dengan demikian, total artikel yang digunakan dalam kajian ini berjumlah 66 artikel.

Proses ekstraksi data terhadap 66 artikel terpilih dilakukan secara manual menggunakan instrumen ekstraksi standar dalam bentuk lembar kerja (*spreadsheet*). Instrumen ini mencatat sembilan elemen utama dari setiap artikel, yaitu judul, penulis dan tahun publikasi, negara asal dataset atau konteks penelitian, kategori kuartil jurnal, area studi, metode penelitian, fokus penelitian, hasil utama, serta gap dan arah penelitian. Kesembilan elemen tersebut sesuai dengan kolom yang disajikan pada Tabel 2. Proses ekstraksi awal dilakukan oleh penulis pertama dan selanjutnya diverifikasi oleh penulis kedua untuk mengurangi potensi bias subjektif serta menjaga konsistensi interpretasi antarartikel. Apabila ditemukan perbedaan interpretasi, keputusan akhir ditetapkan melalui diskusi hingga tercapai kesepakatan bersama. Dengan menggunakan instrumen ini, proses sintesis data menjadi lebih sistematis, transparan, dan dapat direplikasi oleh peneliti lain.

Tabel 1 menunjukkan distribusi kategori jurnal dari 65 artikel utama yang diperoleh melalui basis data *Scopus*. Satu artikel tambahan dari sumber lain tidak dimasukkan dalam tabel ini karena hanya digunakan sebagai pendukung kajian.

Tabel 1. Distribusi Kategori Jurnal Scopus

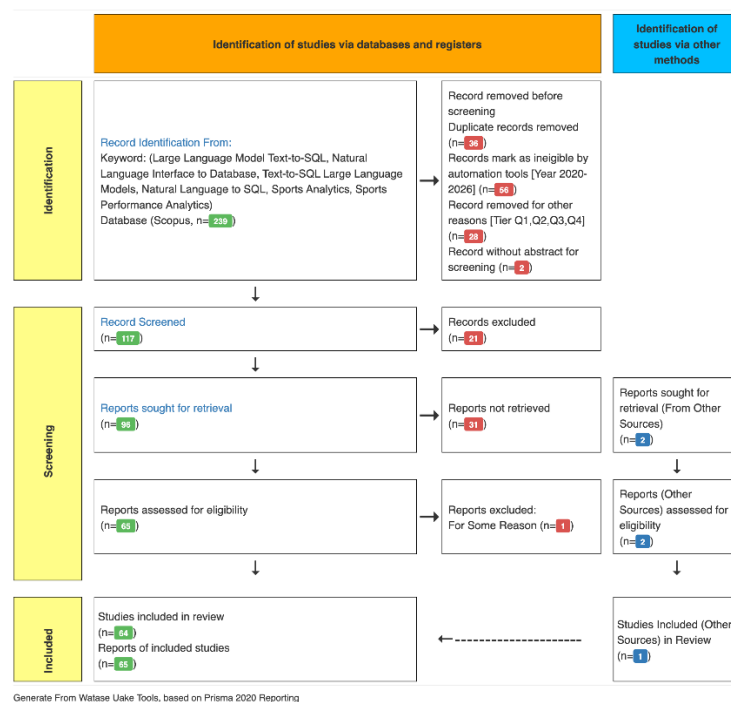
No	Kategori	Jumlah
1	Q1	38
2	Q2	23
3	Q3	4
4	Q4	0
Total		65

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil sintesis literatur mengenai *Large Language Models for Natural Language Database Access in Sports Analytics Systems* berdasarkan artikel yang telah melalui seleksi sistematis. Pembahasan hasil mencakup alur *PRISMA*, tren publikasi, distribusi negara, metode penelitian, area studi, jenis metode analisis, *wordcloud* variabel dan metadata, grafik hipotesis, grafik variabel, serta tabel komparatif artikel. Setelah hasil dipaparkan, bagian pembahasan diarahkan untuk menjawab pertanyaan penelitian mengenai tren, pendekatan dominan, serta *gap* riset dan arah pengembangan masa depan.

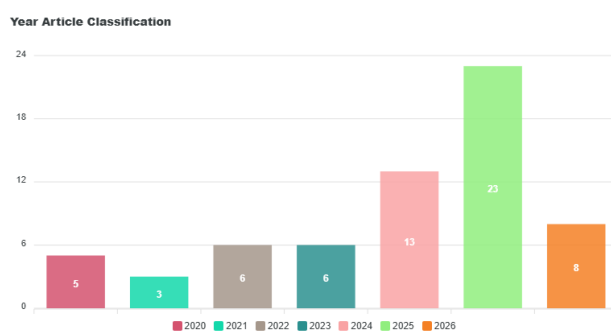
3.1. Hasil

Gambar 1 menunjukkan proses seleksi artikel berdasarkan alur *PRISMA*. Pencarian awal pada *database Scopus* menghasilkan 239 artikel. Pada tahap identifikasi, 36 artikel dihapus karena duplikasi, 56 artikel dikeluarkan karena berada di luar rentang tahun 2020-2026, 28 artikel tidak memenuhi kriteria kategori Q1-Q4, dan 2 artikel tidak memiliki abstrak. Setelah proses tersebut, 117 artikel masuk tahap *screening*. Dari jumlah tersebut, 21 artikel dikeluarkan karena tidak relevan dengan fokus penelitian, sehingga 96 artikel dicari dalam bentuk *full-text*. Sebanyak 31 artikel tidak tersedia *full-text*, sehingga diperoleh 65 artikel *Scopus* yang layak dianalisis. Selain itu, dari dua artikel tambahan yang diidentifikasi melalui penelusuran sumber lain, hanya satu artikel yang memenuhi kriteria kelayakan untuk dimasukkan ke dalam kajian. terdapat 2 artikel tambahan dari sumber lain, sehingga total laporan atau artikel yang digunakan dalam *review* adalah 66. Secara proporsional, hanya sekitar seperempat dari 239 artikel yang diperoleh pada tahap identifikasi awal yang memenuhi seluruh kriteria seleksi. Hal ini mengindikasikan bahwa penerapan kriteria inklusi dan eksklusi yang mencakup penghapusan artikel duplikat, pembatasan periode publikasi, kategori kuartil jurnal, serta ketersediaan artikel *full-text* dilakukan secara ketat untuk memastikan relevansi dan kualitas literatur yang menjadi dasar analisis.



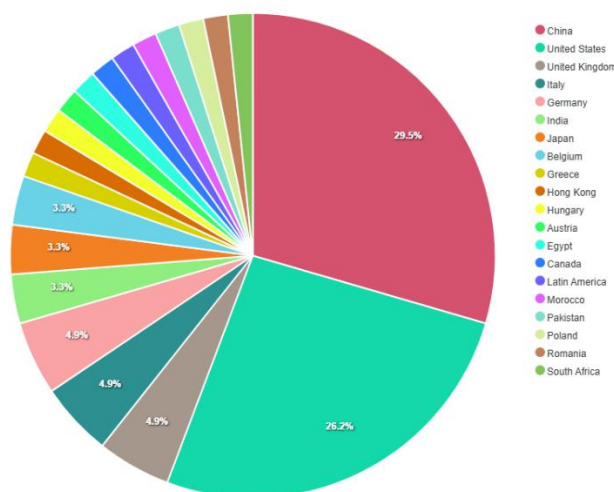
Gambar 1. Diagram Prisma

Gambar 2 memperlihatkan tren publikasi mengenai *LLM* dan *Text-to-SQL* pada analitik olahraga dan bidang terkait pada periode 2020-2026. Jumlah publikasi menunjukkan perkembangan yang meningkat dalam beberapa tahun terakhir. Pada tahun 2020 terdapat 5 artikel, kemudian menurun menjadi 3 artikel pada tahun 2021. Jumlah publikasi kembali meningkat menjadi 6 artikel pada tahun 2022 dan 2023. Peningkatan yang lebih jelas terjadi pada tahun 2024 dengan 13 artikel, lalu mencapai jumlah tertinggi pada tahun 2025 dengan 23 artikel. Sementara itu, tahun 2026 mencatat 8 artikel. Kenaikan publikasi pada periode 2024-2025 menunjukkan meningkatnya perhatian peneliti terhadap integrasi *AI*, *machine learning*, *LLM*, dan *sports analytics*. Pada tahap awal, penelitian terkait topik ini banyak menekankan penggunaan sensor, *big data*, dan manajemen olahraga [6], [7]. Perkembangan berikutnya menunjukkan adanya perluasan fokus ke evaluasi model, pemrosesan bahasa alami, dan akses data berbasis *LLM* [1], [8], [9], [10]. Peningkatan jumlah publikasi pada periode tersebut bertepatan dengan perkembangan pesat model-model *Large Language Models* (LLMs) generasi terbaru yang diperkenalkan pada rentang 2022-2024. Perkembangan ini memperluas peluang penelitian terkait kemampuan pemahaman bahasa alami serta mendorong eksplorasi penerapannya di berbagai domain, termasuk analitik olahraga.



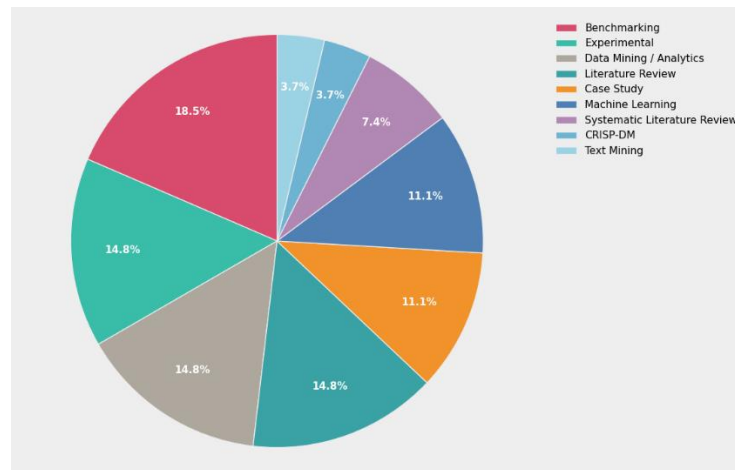
Gambar 2. Tahun Publikasi

Gambar 3 menunjukkan distribusi negara asal dataset atau konteks penelitian. China menjadi negara paling dominan dengan proporsi 29,5%, diikuti United States sebesar 26,2%. United Kingdom, Italy, dan Germany masing-masing memiliki proporsi 4,9%, sedangkan India, Japan, dan Belgium masing-masing sebesar 3,3%. Sisanya tersebar di beberapa negara lain. Dominasi China dan United States dapat dipahami karena kedua negara memiliki ekosistem riset yang kuat dalam *AI*, *data science*, *database system*, dan *sports analytics*. Beberapa studi menunjukkan bahwa pengembangan model analitik dan sistem akses data banyak bertumpu pada ketersediaan *dataset*, infrastruktur komputasi, serta tradisi riset interdisipliner yang kuat [2], [1], [11], [8], [12], [3]. Dominasi penelitian pada kedua negara tersebut mengindikasikan bahwa basis bukti yang tersedia masih terkonsentrasi pada konteks olahraga serta karakteristik basis data yang berkembang di lingkungan penelitian masing-masing. Oleh karena itu, generalisasi temuan ke konteks olahraga, kondisi operasional, maupun infrastruktur data di negara lain perlu dilakukan secara cermat dengan mempertimbangkan perbedaan karakteristik lokal.



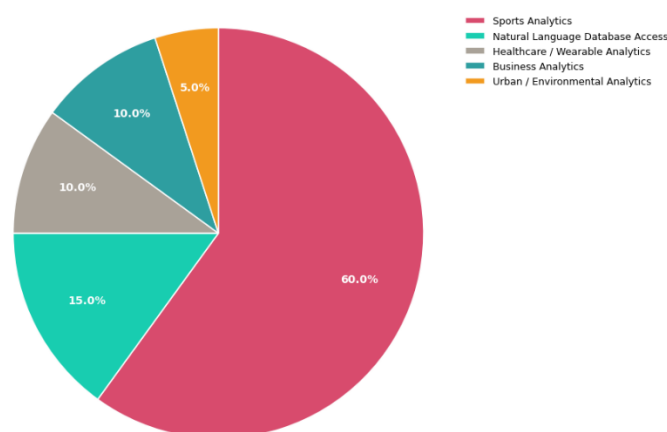
Gambar 3. Negara Asal Dataset

Gambar 4 menampilkan variasi metode penelitian yang digunakan dalam artikel terpilih. Metode *benchmarking* menjadi pendekatan paling banyak digunakan dengan proporsi 18,5%, diikuti *experimental*, *data mining/analytics*, dan *literature review* masing-masing sebesar 14,8%. Selanjutnya, *case study* dan *machine learning* masing-masing mencapai 11,1%, *systematic literature review* sebesar 7,4%, sedangkan *CRISP-DM* dan *text mining* masing-masing sebesar 3,7%. Komposisi ini menunjukkan bahwa penelitian pada bidang *LLM*, *Text-to-SQL*, dan *sports analytics* membutuhkan pengujian performa, perbandingan model, serta evaluasi empiris yang terukur. Hal ini didukung oleh beberapa studi yang memperlihatkan bahwa evaluasi model, pemetaan literatur, analisis graf, dan *benchmark* menjadi bagian penting untuk menilai keandalan pendekatan yang dikembangkan [13], [14], [15], [16].



Gambar 4. Metode

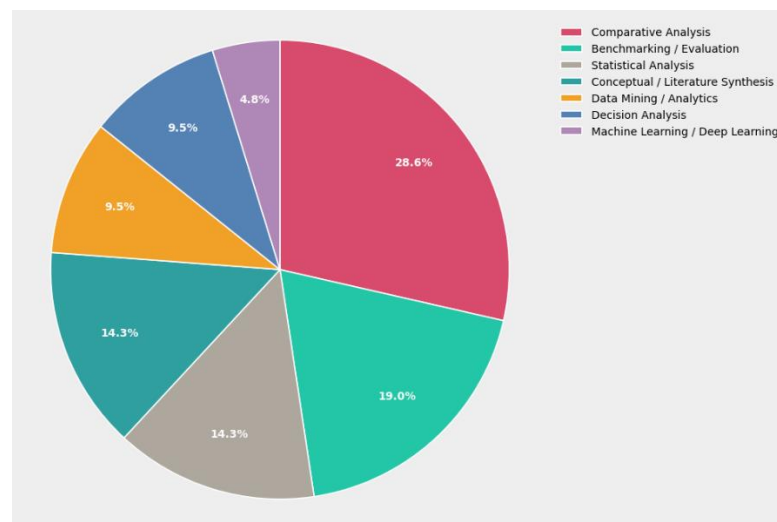
Gambar 5 menunjukkan bahwa area studi terbesar adalah *Sports Analytics* dengan proporsi 60%. Area *Natural Language Database Access* menempati 15%, *Healthcare/Wearable Analytics* sebesar 10%, *Business Analytics* sebesar 10%, dan *Urban/Environmental Analytics* sebesar 5%. Hasil ini menunjukkan bahwa *sports analytics* menjadi domain utama dalam kajian, sedangkan *Natural Language Database Access* berperan sebagai area pendukung untuk memudahkan pengguna non-teknis mengakses data olahraga melalui bahasa alami. Beberapa studi memperlihatkan bahwa *NLIDB*, *Text-to-SQL*, dan *LLM* memiliki relevansi tinggi dalam menjembatani kebutuhan analitik dengan keterbatasan kemampuan teknis pengguna dalam menulis *query* basis data [17], [8], [3], [4], [10].



Gambar 5. Distribusi Area Studi

Gambar 6 memperlihatkan jenis metode analisis yang digunakan dalam literatur. *Comparative analysis* menjadi jenis analisis paling dominan dengan proporsi 28,6%, diikuti *benchmarking/evaluation* sebesar 19,0%, *statistical analysis* sebesar 14,3%, *conceptual/literature synthesis* sebesar 14,3%, *data mining/analytics* sebesar 9,5%, *decision analysis* sebesar 9,5%, dan *machine learning/deep learning* sebesar 4,8%. Dominasi analisis

komparatif menunjukkan bahwa banyak penelitian berupaya membandingkan performa model, metode, atau sistem dalam konteks yang berbeda. Hal ini didukung oleh beberapa studi yang menekankan pentingnya evaluasi yang konsisten, sintesis konseptual, dan perbandingan pendekatan untuk memahami kekuatan serta keterbatasan sistem berbasis *AI* dan *LLM* [13], [18], [14], [19].



Gambar 6. Jenis Metode Analisis

Gambar 7 menunjukkan istilah variabel yang sering muncul, seperti *Text-to-SQL Conversion*, *SQL Query Generation*, *Speech-to-SQL*, *Fine-tuning*, *Machine Learning*, *Athlete Performance Monitoring*, *Wearable Sensor Data*, *Sports Outcome Prediction*, dan *Strategic Decision-Making*. Kemunculan istilah tersebut memperlihatkan dua fokus besar, yaitu akses database berbasis bahasa alami dan penerapan analitik olahraga. Pada sisi akses *database*, beberapa penelitian menekankan pentingnya konversi bahasa alami menjadi *query SQL* yang akurat [20], [3], [21], [13]. Sementara itu, pada sisi *sports analytics*, data sensor dan pemantauan performa atlet menjadi aspek penting untuk mendukung keputusan berbasis data [6], [22].



Gambar 7. Variable Wordcloud

Gambar 8 memperlihatkan istilah metadata dominan, antara lain *Sports Analytics*, *Text-to-SQL*, *Large Language Models*, *Machine Learning*, *Deep Learning*, *Natural Language Processing*, *Natural Language Interface for Databases*, *Data Analytics*, *Wearable Devices*, *IoT*, *Database*, *Query Processing*, dan *Decision-Making*. Istilah-istilah tersebut menunjukkan adanya irisan yang kuat antara *AI*, *database*, dan analitik olahraga. Beberapa studi menunjukkan bahwa *sports analytics* berkembang melalui pemanfaatan data untuk mendukung analisis performa dan pengambilan keputusan [1], sedangkan penelitian lain memperlihatkan bahwa *LLM* dan *Text-to-SQL* semakin penting untuk membuka akses terhadap basis data bagi pengguna non-teknis [3], [4], [10].



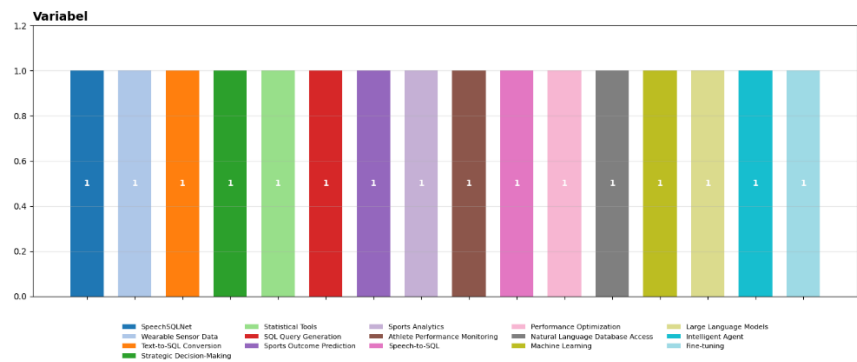
Gambar 8. Metadata Wordcloud

Gambar 9 menggambarkan hubungan antarvariabel utama yang muncul dalam literatur. Hubungan tersebut mencakup *Fine-tuning* terhadap *Text-to-SQL Conversion*, *Machine Learning* terhadap *Sports Outcome Prediction*, *Speech-to-SQL* terhadap *SQL Query Generation*, *Statistical Tools* terhadap *Strategic Decision-Making*, serta *Wearable Sensor Data* terhadap *Athlete Performance Monitoring*. Pola hubungan tersebut menunjukkan dua arah penelitian yang saling melengkapi. Arah pertama berfokus pada pengembangan sistem akses *database* berbasis bahasa alami, terutama melalui pendekatan *Text-to-SQL* dan *Speech-to-SQL* [8], [20]. Arah kedua berfokus pada pemanfaatan data olahraga untuk prediksi, pemantauan, dan pengambilan keputusan berbasis data [2], [23], [24].



Gambar 9. Grafik Hipotesis

Gambar 10 menampilkan variabel seperti *SpeechSQLNet*, *Wearable Sensor Data*, *Text-to-SQL Conversion*, *Strategic Decision-Making*, *Statistical Tools*, *SQL Query Generation*, *Sports Outcome Prediction*, *Sports Analytics*, *Athlete Performance Monitoring*, *Speech-to-SQL*, *Performance Optimization*, *Natural Language Database Access*, *Machine Learning*, *Large Language Models*, *Intelligent Agent*, dan *Fine-tuning*. Tidak terdapat variabel yang benar-benar dominan karena masing-masing muncul dengan frekuensi yang relatif sama. Kondisi ini menunjukkan bahwa bidang *LLM* dan *Text-to-SQL* pada analitik olahraga masih berada pada tahap eksploratif dan multidisipliner. Penelitian tidak hanya bergerak pada satu jalur, tetapi tersebar pada pengembangan model bahasa, akses basis data, pemantauan performa atlet, prediksi olahraga, serta sistem pendukung keputusan.



Gambar 10. Grafik Variable

Secara umum, hasil sintesis menunjukkan bahwa penelitian *LLM* dan *Text-to-SQL* pada analitik olahraga didominasi oleh dua fokus utama. Fokus pertama adalah *sports analytics* untuk evaluasi performa atlet, prediksi hasil pertandingan, pemantauan sensor, dan pengambilan keputusan strategis. Fokus kedua adalah *Natural Language Database Access* untuk *Text-to-SQL*, *Speech-to-SQL*, *SQL Query Generation*, dan pengembangan sistem berbasis *LLM*. Kedua fokus tersebut saling berkaitan karena *sports analytics* membutuhkan akses data yang cepat dan mudah, sedangkan *LLM* dan *Text-to-SQL* dapat menjadi jembatan antara pengguna non-teknis dengan basis data olahraga yang kompleks.

Tabel 2 menunjukkan bahwa artikel *sports analytics* memiliki fokus yang beragam, seperti pemantauan performa atlet, prediksi hasil pertandingan, analisis taktik, dan pengambilan keputusan strategis. Beberapa artikel memperlihatkan bahwa *sports analytics* berkembang dari analisis statistik dan *machine learning* menuju penggunaan sensor, *data tracking*, *visual analytics*, serta model *AI* yang lebih adaptif [6], [2], [1], [15], [20], [9]. Sementara itu, artikel *Natural Language Database Access* menunjukkan perkembangan dari sistem berbasis aturan menuju *LLM*, *Text-to-SQL*, *Speech-to-SQL*, dan *conversational database interface*. Tantangan utama pada area ini meliputi *schema linking*, *query* kompleks, ambiguitas bahasa alami, evaluasi akurasi *SQL*, dan generalisasi lintas domain [25], [8], [3], [4], [10]. Berdasarkan tabel tersebut, *gap* utama yang muncul adalah generalisasi model, interpretabilitas sistem, integrasi data multimodal, privasi data atlet, dan bias algoritma. Pada konteks *sports analytics*, tantangan berkaitan dengan pengolahan data sensor, video, statistik pertandingan, dan data performa atlet. Sementara itu, pada konteks *LLM* dan *Text-to-SQL*, tantangan utama terletak pada akurasi *query*, transparansi sistem, dan kemampuan beradaptasi pada berbagai skema *database* olahraga.

Tabel 2. Komparatif Artikel yang Direview

No	Fokus Penelitian	Area Studi	Metode	Hasil Utama	Gap dan Arah Penelitian
[26]	Analitik GNSS pada olahraga air terbuka dan renang triathlon	Healthcare/Wearable Analytics	SLR PRISMA 2020	GNSS banyak digunakan pada surfing, sailing, dan olahraga air, tetapi pelaporan teknis masih beragam.	Perlu standar pelaporan GNSS dan validasi pada konteks triathlon.
[23]	Evaluasi performa pemain basket secara real-time	Sports Analytics	Model TSF-FUCA dan studi kasus	Model menghasilkan peringkat pemain yang konsisten dan realistis.	Perlu validasi dengan data tracking real-time dan perangkat pendukung keputusan.
[27]	Natural language interface untuk spatial database	Natural Language Database Access	Desain framework dan prototipe	NALSpatial mencapai respons cepat dan keberhasilan translasi tinggi	Masih gagal pada sebagian query spasial kompleks.
[28]	Prediksi dan pemodelan permainan olahraga	Sports Analytics	Pemodelan komputasional TITAN dan	Framework mengungguli pendekatan	Membutuhkan dataset beranotasi dengan kualitas

			<i>PlayNet-Augment</i>	konvensional.	tinggi.
[29]	<i>Natural Language to SQL translation</i>	<i>Natural Language Database Access</i>	<i>Framework G-SQL dan benchmark</i>	<i>G-SQL</i> menunjukkan performa kuat pada <i>IMDB</i> , <i>Yelp</i> , dan <i>MAS</i> .	Masih lemah pada <i>subquery</i> bersarang dan penalaran implisit.
[30]	Pengambilan keputusan olahraga perguruan tinggi	<i>Sports Analytics</i>	Model keputusan komputasional	Model mendukung keputusan olahraga secara sistematis.	Perlu penyederhanaan perhitungan dan validasi empiris.
[16]	Pendekatan graf pada olahraga tim	<i>Sports Analytics</i>	<i>Systematic Mapping Study</i>	Analitik berbasis graf meningkat pada sepak bola, basket, dan <i>baseball</i> .	Perlu mengurangi bias publikasi dan memperluas basis data.
[21]	<i>Cross-domain Text-to-SQL semantic parsing</i>	<i>Natural Language Database Access</i>	Pengembangan model <i>SLENet</i>	<i>SLENet</i> mencapai akurasi <i>exact matching</i> tertinggi.	Masih menghadapi ambiguitas dan skema kompleks.
[31]	<i>Text-to-SQL untuk geospatial database</i>	<i>Natural Language Database Access</i>	Sistem <i>Monkuu</i> dan <i>benchmark</i>	<i>Monkuu</i> meningkatkan akurasi eksekusi pada <i>KaggleDBQA</i> .	Bergantung pada kualitas dokumentasi skema.
[18]	Pengukuran dan evaluasi dalam <i>kinesiology</i>	<i>Healthcare/Wearable Analytics</i>	Artikel konseptual	Evaluasi harus berkembang mengikuti <i>AI</i> dan teknologi baru.	Perlu pengujian empiris terhadap kerangka konseptual.
[32]	<i>Model Text-to-SQL berbasis Gemma</i>	<i>Natural Language Database Access</i>	Pengembangan model dan <i>benchmark</i>	<i>GEMMA-SQL</i> mencapai <i>test-suite accuracy</i> yang kompetitif.	Perlu peningkatan akurasi pada <i>query</i> kompleks.
[3]	<i>Survey LLM untuk Text-to-SQL</i>	<i>Natural Language Database Access</i>	<i>Survey review</i>	<i>LLM</i> meningkatkan kemampuan <i>Text-to-SQL</i> secara signifikan.	Tantangan utama terdapat pada <i>prompt</i> , skema, dan reliabilitas.
[33]	<i>Explainable feature engineering untuk esports</i>	<i>Sports Analytics</i>	Studi kasus komputasional	Fitur temporal lebih prediktif dibandingkan fitur <i>spasial</i> .	Asumsi fisika masih disederhanakan.
[34]	Koneksi <i>LLM</i> dengan <i>database</i> untuk analisis data kompleks	<i>Natural Language Database Access</i>	<i>Case study</i>	<i>LLM</i> dengan koneksi <i>database</i> membantu eksplorasi data.	Terbatas pada pertanyaan kompleks dan validitas jawaban.
[35]	Dampak cedera pada performa pemain NBA	<i>Sports Analytics</i>	Analitik longitudinal	Cedera berhubungan dengan penurunan performa tertentu.	Perlu uji statistik dan variabel kesehatan yang lebih lengkap.
[36]	Evaluasi analitik pendidikan jasmani	<i>Sports Analytics</i>	<i>MCDM</i>	Integrasi <i>AHP</i> dan <i>MEREC</i>	Kompleksitas <i>pairwise comparison</i>

				mendukung proses evaluasi.	perlu dikurangi.
[37]	<i>Refinement benchmark Text-to-SQL</i>	<i>Natural Language Database Access</i>	Evaluasi <i>benchmark</i>	<i>Benchmark</i> yang diperbaiki meningkatkan reliabilitas evaluasi.	Terbatas pada <i>SQLite</i> dan perlu cakupan model yang lebih luas.
[9]	<i>AI dalam analisis performa olahraga</i>	<i>Sports Analytics</i>	<i>SLR dan meta-analysis</i>	<i>AI</i> efektif dalam analisis performa olahraga.	Variasi sampel dan konteks olahraga masih tinggi.
[38]	<i>Natural language access untuk business analytics</i>	<i>Business Analytics</i>	<i>Mixed-method</i> dan eksperimen	<i>GPT-4</i> dan <i>Llama</i> menunjukkan performa terbaik.	<i>Dataset</i> masih kecil dan bersifat <i>domain-specific</i> .
[39]	Optimasi performa olahraga elite	<i>Sports Analytics</i>	Simulasi <i>proof-of-concept</i>	<i>Framework</i> menunjukkan peningkatan performa pada beberapa olahraga.	Data sintetis membatasi generalisasi hasil.
[40]	<i>Pipeline Text-to-SQL CSpider</i>	<i>Natural Language Database Access</i>	Eksperimen <i>machine learning</i>	<i>FGCSQL</i> mencapai hasil kompetitif pada <i>CSpider</i> .	<i>Filtering</i> skema biner berisiko mengabaikan relasi penting.
[14]	<i>Intelligent database interaction</i>	<i>Natural Language Database Access</i>	Model <i>Nabil</i> dan <i>benchmark</i>	<i>Nabil</i> mengungguli <i>baseline</i> pada <i>BIRD</i> .	Dukungan fungsi masih terbatas dan perlu pengendalian halusinasi.
[5]	<i>Evaluasi Text-to-SQL di era LLM</i>	<i>Natural Language Database Access</i>	Pengembangan metrik <i>ETM</i>	<i>ETM</i> mengurangi kesalahan evaluasi dibandingkan <i>EXE</i> dan <i>ESM</i> .	Perlu aturan ekuivalensi tambahan.
[13]	<i>Conversational Text-to-SQL</i>	<i>Natural Language Database Access</i>	<i>Benchmark</i> dan <i>fine-tuning</i>	Model full history lebih baik daripada question rewriting.	Terbatas oleh akses <i>test set</i> publik.
[12]	Prediksi <i>playoff NFL</i>	<i>Sports Analytics</i>	<i>Model-X Knockoff</i>	Mengidentifikasi fitur penting untuk prediksi playoff.	Asumsi distribusi fitur perlu diuji lebih lanjut.
[41]	Prediksi <i>break shot billiards</i>	<i>Sports Analytics</i>	<i>Eksperimental</i>	Metode efektif untuk prediksi break shot.	Kurangnya <i>dataset</i> publik <i>billiards</i> .
[42]	Cedera dan performa pemain NBA	<i>Healthcare/Wearable Analytics</i>	<i>Data mining</i>	Cedera <i>musculoskeletal</i> dominan pada NBA.	Kualitas dan kelengkapan data masih terbatas.
[22]	<i>Sports physiotherapy</i> dan monitoring presisi	<i>Healthcare/Wearable Analytics</i>	<i>Observational cross-sectional</i>	Sensor menunjukkan peningkatan indikator fisiologis.	Sampel perlu diperluas pada profil atlet yang berbeda.
[43]	<i>Dialog natural</i>	<i>Natural Language</i>	<i>Proof-of-</i>	Sistem dapat	Output <i>LLM</i>

	<i>language</i> untuk akses <i>database anatomi</i>	<i>Database Access</i>	<i>concept system</i>	menghubungkan istilah <i>anatomi</i> dengan <i>database</i> .	masih tidak selalu konsisten.
[44]	Prediksi hasil kejuaraan tinju dunia	<i>Sports Analytics</i>	<i>Factorial design methodology</i>	Feature selection paling berpengaruh terhadap prediksi.	Perlu perluasan fitur dan skenario pertandingan.
[45]	<i>Recovery time</i> dan strategi finansial NBA	<i>Sports Analytics</i>	<i>Data mining</i> dan <i>association rule</i>	<i>Recovery time</i> berdampak pada strategi finansial.	Cakupan data dan variabel masih terbatas.
[46]	<i>Text-to-SQL</i> untuk rekam medis elektronik	<i>Healthcare/Wearable Analytics</i>	<i>CRISP-DM</i>	<i>MedT5SQL</i> mencapai akurasi tinggi.	Terbatas oleh sumber daya model dan <i>dataset</i> .
[47]	<i>Precision sport science</i>	<i>Sports Analytics</i>	<i>Literature review</i>	<i>Pervasive computing</i> dan <i>data analytics</i> penting untuk performa atlet.	Transparansi industri dan penggunaan alat analitik masih terbatas.
[15]	<i>AI</i> untuk analisis sepak bola	<i>Sports Analytics</i>	<i>SLR OODA-loop</i>	<i>Review</i> merangkum <i>event data</i> , <i>tracking data</i> , dan analisis sepak bola.	Akses <i>data real-world</i> yang rinci masih sulit.
[48]	<i>Smart sensing</i> pada tenis	<i>Sports Analytics</i>	<i>Eksperimental</i>	<i>SmartDampener</i> mampu mengukur kecepatan dan lokasi bola.	<i>Dataset</i> masih kecil dan perlu diversifikasi.
[8]	<i>Speech-to-SQL</i>	<i>Natural Language Database Access</i>	Model <i>SpeechSQLNet</i> dan <i>benchmark</i>	<i>SpeechSQLNet</i> mengungguli <i>baseline</i> .	<i>Dataset</i> masih <i>single-table</i> dan belum mencakup <i>join</i> kompleks.
[24]	<i>Machine learning</i> pada sepak bola profesional	<i>Sports Analytics</i>	<i>Machine learning</i>	Membahas tantangan metodologis dalam <i>sports analytics</i>.	Kurang <i>ground truth</i> dan data sering <i>noisy</i> .
[20]	Pencegahan cedera olahraga dengan <i>wearable sensor</i>	<i>Healthcare/Wearable Analytics</i>	Model <i>AI</i> berbasis <i>wearable</i>	Model mencapai akurasi tinggi untuk prediksi cedera.	Terbatas pada satu cabang olahraga.
[11]	<i>AI/ML</i> dan sensor dalam <i>sports analytics</i>	<i>Healthcare/Wearable Analytics</i>	<i>Review</i> lanjutan	<i>AI</i> dan <i>ML</i> mendukung analisis data olahraga.	Perlu standardisasi dan integrasi praktis.
[17]	<i>NLIDB</i> untuk <i>relational database</i>	<i>Natural Language Database Access</i>	Pengembangan <i>xDBTagger</i>	Sistem interpretabel dan kompetitif.	Belum optimal untuk <i>nested SQL</i> .
[49]	Peningkatan <i>IRNet</i> untuk <i>NLQ-to-SQL</i>	<i>Natural Language Database Access</i>	<i>Eksperimental</i>	Meningkatkan prediksi <i>query</i> sulit.	Akurasi <i>nested query</i> masih menjadi kendala.
[50]	Cedera NBA dan dampak finansial	<i>Sports Analytics</i>	<i>Data analytics</i> dan <i>text mining</i>	Terdapat perbedaan <i>recovery time</i>	Belum memasukkan tingkat keparahan

				dan beban finansial.	cedera.
[51]	Analitik data pada olahraga tim <i>elite</i>	<i>Sports Analytics</i>	Kualitatif	Analitik meningkatkan pembelajaran pemain dan strategi.	Sampel terbatas pada olahraga tim tertentu.
[1]	Transformasi <i>sports analytics</i>	<i>Sports Analytics</i>	<i>Empirical analysis</i>	<i>Data analytics</i> mengubah praktik olahraga profesional.	Cakupan cabang dan konteks masih dapat diperluas.
[52]	Penyusunan tim basket	<i>Sports Analytics</i>	<i>Mathematical modeling</i>	Metode menghasilkan komposisi tim yang lebih seimbang.	Belum mempertimbangkan perubahan performa pemain.
[53]	Arsitektur <i>big data</i> untuk <i>sports analytics</i>	<i>Sports Analytics</i>	<i>Experimental research</i>	<i>Dockerized big data architecture</i> dapat diterapkan.	Perlu pengembangan pada implementasi dan skalabilitas.
[54]	Pemetaan ruang hijau urban terkait aktivitas	<i>Urban/Environmental Analytics</i>	<i>Geospatial case study</i>	Model mencapai akurasi tinggi untuk pemetaan ruang publik.	Bergantung pada kualitas data <i>OSM</i> .
[55]	Visual <i>analytics</i> pada olahraga raket	<i>Sports Analytics</i>	Visual <i>analytics</i>	Ahli dapat mengidentifikasi perubahan taktik atlet.	Belum mendukung data numerik secara optimal.
[56]	<i>Tac-Trainer</i> untuk olahraga raket	<i>Sports Analytics</i>	<i>Design study</i>	Sistem mendukung pelatihan taktik berbasis data.	Generalisasi pada konteks lain masih perlu diuji.
[2]	Prediksi hasil NFL	<i>Sports Analytics</i>	<i>Machine learning</i>	<i>Random Forest</i> menunjukkan performa terbaik.	Perlu membandingkan strategi <i>imbalance</i> lain selain <i>SMOTE</i> .
[57]	IoT dan <i>big data</i> dalam monitoring olahraga	<i>Healthcare/Wearable Analytics</i>	<i>Experimental design</i>	Sensor mandiri mampu menghasilkan sinyal berbasis pukulan.	Perlu peningkatan perangkat dan validasi penggunaan.
[58]	Analitik pelatihan olahraga <i>elite</i>	<i>Healthcare/Wearable Analytics</i>	<i>Literature review</i>	Penggunaan data masih dominan informatif, belum preskriptif.	Perlu pengembangan <i>advanced</i> dan <i>prescriptive analytics</i> .
[59]	<i>Synthetic soccer event recognition</i>	<i>Sports Analytics</i>	<i>Original software publication</i>	<i>SoccER</i> menghasilkan data pertandingan sintetis lengkap.	<i>Realisme</i> visual masih rendah dibandingkan solusi komersial.
[26]	<i>Natural language query ke relational database</i>	<i>Natural Language Database Access</i>	<i>Systematic Literature Review</i>	Model konseptual <i>NL-to-SQL</i> masih terbatas pada <i>query</i> tertentu.	Perlu peningkatan presisi dan pelatihan model.
[60]	<i>Wearable biosensor</i> untuk	<i>Healthcare/Wearable Analytics</i>	<i>Systematic</i>	Material fleksibel dan <i>mikrofluidik</i>	Perlu standar emas dan validasi jangka

	kesehatan dan olahraga		<i>review</i>	mendukung <i>monitoring real-time</i> .	panjang.
[61]	<i>Transformasi natural language question ke SQL</i>	<i>Natural Language Database Access</i>	<i>Semantic dan pattern matching</i>	Sistem meningkatkan <i>recall</i> , akurasi, dan presisi.	Presisi menurun saat jumlah aturan meningkat.
[7]	<i>Big data dalam sport management</i>	<i>Sports Analytics</i>	<i>Big data analytics</i>	<i>Big data</i> memperluas teori dan praktik manajemen olahraga.	Isu sosial, hukum, dan etika perlu dibahas lebih jauh.
[62]	Penjadwalan olahraga dan <i>operations research</i>	<i>Sports Analytics</i>	<i>Literature review</i>	Teknik optimasi berhasil diterapkan pada penjadwalan.	Kompleksitas jadwal dan kondisi lokal masih menjadi tantangan.
[63]	Akses bahasa alami ke <i>database</i> relasional dan heterogen	<i>Natural Language Database Access</i>	<i>Framework TriSQL dan benchmark</i>	<i>TriSQL</i> mencapai akurasi eksekusi tinggi pada <i>Spider</i> .	Perlu penanganan ambiguitas dan adaptasi ke skema baru.
[10]	<i>Text-to-SQL</i> dunia nyata dengan <i>knowledge graph</i>	<i>Natural Language Database Access</i>	<i>Benchmark eksperimental</i>	Strategi berbasis <i>DANKE</i> meningkatkan <i>schema linking</i> dan kompilasi <i>SQL</i> .	Perlu disambiguasi berbasis <i>user-in-the-loop</i> dan perluasan <i>dataset</i> .
[4]	Sintesis penelitian <i>Text-to-SQL</i>	<i>Natural Language Database Access</i>	<i>Systematic Literature Review</i>	<i>Text-to-SQL</i> berkembang dari pendekatan <i>rule-based</i> menuju <i>LLM</i> .	Perlu sistem yang lebih <i>robust</i> , efisien, dan mampu menangani skema kompleks.
[64]	<i>Natural language query</i> untuk manajemen data <i>IoT</i>	<i>Natural Language Database Access</i>	Desain sistem dan evaluasi komparatif	<i>GPT-o3 mini-high</i> menunjukkan efisiensi valid yang baik.	Perlu validasi pada domain selain <i>precision agriculture</i> .
[19]	<i>Review NLIDB dan Speech-to-NLIDB</i>	<i>Natural Language Database Access</i>	<i>Review</i> komprehensif	Riset <i>ST-NLIDB</i> masih terfragmentasi.	Perlu <i>benchmark</i> dan integrasi antarsistem yang lebih jelas.
[65]	Analisis performa sepak bola	<i>Sports Analytics</i>	Model <i>unbalanced optimal transport</i>	Model meningkatkan rekonstruksi performa dibandingkan pendekatan pembandingan.	Terbatas pada data <i>Serie A</i> dan perlu perluasan konteks liga.

3.2. Pembahasan

RQ1 Bagaimana integrasi *Large Language Models (LLMs)* dengan *sports analytics* dapat mengatasi keterbatasan dalam pengelolaan data olahraga secara *real-time* dan meningkatkan pengambilan keputusan strategis?

Berdasarkan hasil sintesis, integrasi *LLM* dengan *sports analytics* dapat membantu mengatasi keterbatasan akses terhadap data olahraga yang semakin kompleks. Data olahraga tidak hanya berasal dari statistik pertandingan, tetapi juga dari *wearable sensor*, video, *data tracking*, data performa, dan basis data historis. Penelitian sebelumnya menunjukkan bahwa *sports analytics* telah digunakan untuk pemantauan kondisi atlet, evaluasi performa, prediksi hasil pertandingan, dan pengambilan keputusan strategis [1], [2], [6], [7], [20], [23]. Namun, pendekatan tersebut masih sering membutuhkan kemampuan teknis dalam membaca data, mengolah hasil analisis, atau memahami struktur *database*.

Integrasi *LLM* dengan *Natural Language Database Access* dapat memudahkan pengguna non-teknis untuk mengakses data melalui pertanyaan bahasa alami. Pendekatan *Text-to-SQL*, *Speech-to-SQL*, dan *natural language interface* dapat menjadi penghubung antara pengguna dan *database* olahraga [3], [4], [8], [10], [13]. Dengan demikian, pelatih, analis performa, dan manajer tim dapat memperoleh informasi dari data olahraga tanpa harus menulis *query SQL* secara manual.

RQ2 Apa perkembangan tren terkini dalam penggunaan *Text-to-SQL* dan *wearable technologies* untuk analisis performa atlet, serta bagaimana pergeseran paradigma ini memengaruhi pendekatan tradisional dalam *sports analytics*?

Tren penelitian menunjukkan adanya pergeseran dari pendekatan statistik dan *machine learning* tradisional menuju sistem yang lebih adaptif, berbasis *LLM*, dan didukung *wearable technologies*. Pada tahap awal, penelitian banyak berfokus pada *biosensor*, *big data*, dan *wearable sensor* untuk memantau kondisi fisik serta performa atlet [6], [7], [20], [22], [63]. Selanjutnya, penelitian berkembang ke arah *visual analytics*, *predictive analytics*, *graph analytics*, dan *AI-based decision support* untuk menganalisis performa pemain, pola taktik, serta hasil pertandingan [1], [2], [15], [16], [24], [58].

Di sisi lain, *Text-to-SQL* berkembang dari pendekatan berbasis aturan menuju pendekatan berbasis *LLM*. Studi mengenai *Text-to-SQL*, *Speech-to-SQL*, *schema linking*, dan *query generation* menunjukkan bahwa *LLM* dapat meningkatkan kemampuan sistem dalam memahami pertanyaan alami dan menghasilkan *query SQL* [3], [4], [5], [8], [10], [21], [41]. Pergeseran ini membuat *sports analytics* tidak lagi hanya bergantung pada dashboard statis atau laporan manual, tetapi mulai mengarah pada sistem interaktif berbasis bahasa alami.

RQ3 Bagaimana kerangka *TCCM* dapat diterapkan untuk memahami kontribusi teori, konteks, kriteria, dan metodologi dalam integrasi *AI*, *machine learning*, dan *natural language interfaces* di bidang olahraga?

Kerangka *TCCM* dapat digunakan untuk memetakan penelitian berdasarkan aspek theory, context, criteria, dan methodology. Dari sisi theory, *sports analytics* banyak bertumpu pada *data-driven decision-making*, *machine learning*, *statistical analysis*, dan *decision support system* [1], [2], [23], [24]. Sementara itu, *Natural Language Database Access* banyak didasarkan pada *natural language processing*, *semantic parsing*, *schema linking*, dan *Text-to-SQL* [3], [4], [5], [8], [10]. Kedua area tersebut menjadi dasar konseptual bagi pengembangan sistem analitik olahraga yang lebih mudah diakses.

Dari sisi context, penelitian yang dikaji mencakup *Sports Analytics*, *Natural Language Database Access*, *Healthcare/Wearable Analytics*, *Business Analytics*, dan *Urban/Environmental Analytics*, sebagaimana ditunjukkan pada kolom Area Studi di Tabel 2. Dari sisi criteria, ukuran yang sering digunakan meliputi akurasi model dan akurasi *query* [17], [61], [63], efektivitas evaluasi sistem [5], [50], pemantauan performa atlet [1], [65], serta dukungan terhadap pengambilan keputusan strategis [7], [51], [52]. Dari sisi methodology, berdasarkan kolom Metode pada Tabel 2, pendekatan yang banyak digunakan adalah *benchmarking*, *experimental study*, *comparative analysis*, *literature review*, *data mining*, dan *machine learning*. Dengan demikian, *TCCM* membantu menunjukkan bahwa *LLM* dan *Text-to-SQL* pada analitik olahraga tidak hanya berkaitan dengan teknologi, tetapi juga dengan konteks penerapan, indikator evaluasi, dan metode penelitian.

RQ4 Apa agenda riset masa depan yang dapat dikembangkan melalui kolaborasi lintas disiplin antara *sports analytics*, *database management*, dan *AI* untuk menciptakan solusi inovatif dalam pengelolaan data dan strategi olahraga?

Agenda riset masa depan perlu diarahkan pada pengembangan sistem yang menggabungkan *sports analytics*, *database management*, dan *AI* secara lebih terstruktur. Berdasarkan Tabel 2, *gap* utama yang masih muncul meliputi generalisasi model, kompleksitas skema *database*, akurasi *query*, interpretabilitas sistem, integrasi data multimodal, privasi data atlet, dan bias algoritma. Pada *Natural Language Database Access*, sistem *Text-to-SQL* berbasis *LLM* masih menghadapi tantangan pada pertanyaan ambigu, *join* multi-tabel, *nested query*, dan skema *database* yang belum pernah dilihat sebelumnya [3], [4], [8], [10], [21]. Pada *sports analytics*, tantangan muncul dari keragaman data seperti sensor, video, statistik pertandingan, catatan medis, dan data performa atlet [6], [20], [22], [24], [63].

Tabel 3. Perbandingan Penelitian Terdahulu dan Arah Pengembangan Sistem

No	Domain / Fokus	LLM	Text-to-SQL	Routing	Guard	Builder	Response	Sport Context
[8]	<i>Speech-to-SQL</i>	✓	✓					
[3]	<i>Survey LLM untuk Text-to-SQL</i>	✓	✓					
[10]	<i>Text-to-SQL dengan LLM dan knowledge graph</i>	✓	✓			✓		
[38]	<i>Text-to-SQL dengan intelligent agents dan LLM</i>	✓	✓			✓	✓	
[20]	<i>Wearable analytics untuk olahraga</i>							✓
Future Work	Arsitektur <i>chatbot</i> berbasis <i>LLM-agent</i> untuk akses data <i>sports analytics</i> Bahasa alami	✓	✓	✓	✓	✓	✓	✓

Tabel 3 menunjukkan bahwa penelitian terdahulu telah membahas beberapa komponen penting, seperti *LLM*, *Text-to-SQL*, *query generation*, *intelligent agents*, dan *sports analytics*. Namun, komponen tersebut umumnya masih dibahas secara terpisah. Studi pada area *Text-to-SQL* lebih banyak berfokus pada konversi bahasa alami menjadi *SQL*, peningkatan akurasi *query*, *schema linking*, dan evaluasi model [3], [8], [10], [41]. Sementara itu, studi pada area *sports analytics* lebih banyak berfokus pada pemantauan performa, *wearable analytics*, dan pengolahan data olahraga [20]. Dengan demikian, masih terdapat celah pada integrasi komponen *intent routing*, *safety guard*, *query builder*, dan *AI response generation* dalam satu alur sistem yang terpadu untuk konteks *sports analytics*.

Berdasarkan *gap* tersebut, *future work* dari kajian ini diarahkan pada perancangan arsitektur berbasis *LLM-agent* untuk *sports analytics*. Sistem yang dikembangkan diarahkan agar mampu mengelola data seperti data atlet, hasil pertandingan, data kejuaraan, rekor nasional, catatan waktu, kategori lomba, data *club* olahraga dan performa atlet melalui pertanyaan bahasa alami. Dalam arsitektur tersebut, *intent routing* berperan untuk mengenali maksud pertanyaan pengguna, *safety guard* digunakan untuk menyaring *input* dan *output* yang berisiko, *query builder* berfungsi untuk menyusun dan memvalidasi *query*, sedangkan *AI response generation* digunakan untuk menyajikan hasil analisis dalam bentuk jawaban yang mudah dipahami oleh pengguna non-teknis. Dengan arah ini, pengembangan sistem dapat menjadi penghubung antara kebutuhan *sports analytics* dan akses basis data berbasis bahasa alami.

3.3. Ancaman terhadap Validitas

Sebagaimana lazim dalam penelitian *Systematic Literature Review* (SLR), kajian ini memiliki beberapa ancaman terhadap validitas yang perlu diperhatikan. Pertama, proses pencarian literatur hanya dilakukan pada satu basis data utama, yaitu *Scopus*. Meskipun *Scopus* memiliki cakupan multidisipliner yang luas, penggunaan satu basis data berpotensi menyebabkan terlewatkannya artikel relevan yang terindeks secara eksklusif pada basis data lain, seperti *IEEE Xplore* atau *ACM Digital Library*, sehingga pemetaan literatur yang dihasilkan mungkin belum sepenuhnya merepresentasikan keseluruhan penelitian yang tersedia. Kedua, kajian ini berpotensi mengalami bias publikasi (*publication bias*), karena penelitian dengan hasil positif atau signifikan cenderung lebih banyak dipublikasikan dan terindeks dibandingkan penelitian dengan hasil negatif atau tidak signifikan. Ketiga, penerapan kriteria inklusi yang mensyaratkan artikel pada kategori kuartil Q1-Q4 serta ketersediaan naskah full-text berbahasa Inggris berpotensi mengecualikan studi relevan yang dipublikasikan dalam bahasa lain atau pada sumber publikasi di luar kriteria tersebut. Keempat, jumlah artikel pendukung dari sumber *non-Scopus* yang digunakan untuk triangulasi relatif terbatas, yaitu hanya satu artikel, sehingga validasi

silang terhadap temuan utama masih belum optimal. Meskipun demikian, keterbatasan tersebut tidak mengurangi kontribusi kajian ini dalam memetakan perkembangan penelitian pada bidang yang ditinjau, namun perlu dipertimbangkan dalam menginterpretasikan dan menggeneralisasikan temuan yang diperoleh serta menjadi masukan bagi pengembangan metodologi pada penelitian selanjutnya.

4. KESIMPULAN

Berdasarkan hasil Systematic Literature Review yang dilakukan, penelitian mengenai *Large Language Models* untuk *Natural Language Database Access* dalam *Sports Analytics Systems* menunjukkan perkembangan yang semakin kuat pada periode 2020-2026. Dari 65 artikel utama *Scopus* dan 2 artikel pendukung yang dianalisis, terlihat bahwa *sports analytics* menjadi area paling dominan, terutama pada topik pemantauan performa atlet, prediksi hasil pertandingan, *wearable sensor*, *data tracking*, serta pengambilan keputusan strategis. Di sisi lain, *Natural Language Database Access* juga berkembang melalui pendekatan *Text-to-SQL*, *Speech-to-SQL*, *schema linking*, *query generation*, dan penggunaan *LLM* untuk memudahkan akses terhadap basis data.

Hasil kajian menunjukkan bahwa integrasi *LLM* dengan *sports analytics* berpotensi mengatasi keterbatasan akses data olahraga yang selama ini masih banyak bergantung pada kemampuan teknis pengguna. Melalui pendekatan bahasa alami, pengguna non-teknis seperti pelatih, analis performa, manajer tim, dan peneliti olahraga dapat mengakses informasi dari *database* tanpa harus menulis *query SQL* secara manual. Temuan ini menunjukkan bahwa *LLM* dan *Text-to-SQL* pada analitik olahraga dapat menjadi arah penting dalam membangun sistem analitik olahraga yang lebih interaktif, adaptif, dan mudah digunakan.

Dari sisi metodologi, penelitian terdahulu banyak menggunakan *benchmarking*, *experimental study*, *comparative analysis*, *data mining/analytics*, *literature review*, dan *machine learning*. Pola ini menunjukkan bahwa bidang *LLM* dan *sports analytics* masih berfokus pada pengujian performa model, evaluasi sistem, serta perbandingan pendekatan untuk meningkatkan akurasi dan keandalan hasil analisis. Namun, beberapa *gap* masih ditemukan, seperti generalisasi model, kompleksitas skema *database*, interpretabilitas sistem, integrasi data multimodal, privasi data atlet, dan potensi bias algoritma.

Arah penelitian selanjutnya perlu difokuskan pada pengembangan sistem yang mampu menghubungkan *sports analytics*, *database management*, dan *AI* secara lebih terpadu. *Future work* dari kajian ini diarahkan pada perancangan arsitektur berbasis *LLM-agent* untuk *sports analytics*. Sistem tersebut diharapkan mampu mendukung pengelolaan data atlet, hasil pertandingan, data kejuaraan, rekor nasional, catatan waktu, kategori lomba, data *club* olahraga dan performa atlet melalui pertanyaan bahasa alami. Dengan dukungan *intent routing*, *safety guard*, *query builder*, dan *AI response generation*, sistem yang dikembangkan diharapkan dapat menjadi solusi yang lebih aman, terstruktur, dan mudah dipahami dalam mendukung pengambilan keputusan berbasis data olahraga. Secara praktis, federasi maupun klub olahraga yang telah mengelola data dalam sistem basis data terstruktur dapat mempertimbangkan penerapan bertahap mekanisme akses data berbasis bahasa alami pada sistem informasi yang telah tersedia. Pendekatan ini berpotensi mempermudah pelatih, analis performa, dan manajer tim dalam memperoleh informasi terkait pertandingan maupun performa atlet secara lebih cepat dan efisien tanpa bergantung sepenuhnya pada dukungan teknis untuk proses pengambilan data.

DAFTAR PUSTAKA

- [1] B. S. Baumer, G. J. Matthews, and Q. Nguyen, "Big ideas in sports analytics and statistical tools for their investigation," *WIREs Computational Stats*, vol. 15, no. 6, p. e1612, Nov. 2023, doi: 10.1002/wics.1612.
- [2] Y. F. Roumani, "Sports analytics in the NFL: classifying the winner of the superbowl," *Ann Oper Res*, vol. 325, no. 1, pp. 715–730, Jun. 2023, doi: 10.1007/s10479-022-05063-x.
- [3] L. Shi, Z. Tang, N. Zhang, X. Zhang, and Z. Yang, "A Survey on Employing Large Language Models for Text-to-SQL Tasks," *ACM Comput. Surv.*, vol. 58, no. 2, pp. 1–37, Jan. 2026, doi: 10.1145/3737873.
- [4] M. Liu, X. Wang, J. Xu, W. Yi, and O. Wolfson, "A systematic review of natural language interfaces for databases," *Front. Comput. Sci.*, vol. 20, no. 11, p. 2011623, Nov. 2026, doi: 10.1007/s11704-025-50592-w.
- [5] B. G. Ascoli, Y. S. R. Kandikonda, and J. D. Choi, "ETM: Modern Insights into Perspective on Text-to-SQL Evaluation in the Age of Large Language Models," *Future Internet*, vol. 17, no. 8, p. 325, Jul. 2025, doi: 10.3390/fi17080325.
- [6] L. B. Baker *et al.*, "Skin-interfaced microfluidic system with personalized sweating rate and sweat chloride analytics for sports science applications," *Sci. Adv.*, vol. 6, no. 50, p. eabe3929, Dec. 2020, doi: 10.1126/sciadv.abe3929.

-
- [7] N. M. Watanabe, S. Shapiro, and J. Drayer, "Big Data and Analytics in Sport Management," *Journal of Sport Management*, vol. 35, no. 3, pp. 197–202, May 2021, doi: 10.1123/jsm.2021-0067.
 - [8] Y. Song, R. C.-W. Wong, and X. Zhao, "Speech-to-SQL: toward speech-driven SQL query generation from natural language question," *The VLDB Journal*, vol. 33, no. 4, pp. 1179–1201, Jul. 2024, doi: 10.1007/s00778-024-00837-0.
 - [9] P. Pietraszewski *et al.*, "The Role of Artificial Intelligence in Sports Analytics: A Systematic Review and Meta-Analysis of Performance Trends," *Applied Sciences*, vol. 15, no. 13, p. 7254, Jun. 2025, doi: 10.3390/app15137254.
 - [10] E. R. Nascimento *et al.*, "A Text-to-SQL strategy based on large language models and knowledge graphs for real-world databases," *Data & Knowledge Engineering*, vol. 164, p. 102580, Jul. 2026, doi: 10.1016/j.datak.2026.102580.
 - [11] I. Ghosh, S. Ramasamy Ramamurthy, A. Chakma, and N. Roy, "Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective," *WIREs Data Min & Knowl*, vol. 13, no. 5, p. e1496, Sep. 2023, doi: 10.1002/widm.1496.
 - [12] T. Zhao, J. Cabral, and G. Zhu, "A novel explainable artificial intelligence framework using knockoffs techniques with applications to sports analytics," *Ann Oper Res*, Apr. 2025, doi: 10.1007/s10479-025-06575-y.
 - [13] B. G. Ascoli and J. D. Choi, "Advancing Conversational Text-to-SQL: Context Strategies and Model Integration with Large Language Models," *Future Internet*, vol. 17, no. 11, p. 527, Nov. 2025, doi: 10.3390/fi17110527.
 - [14] F. Zhou *et al.*, "Nabil: A Text-to-SQL Model Based on Brain-Inspired Computing Techniques and Large Language Modeling," *Electronics*, vol. 14, no. 19, p. 3910, Sep. 2025, doi: 10.3390/electronics14193910.
 - [15] Z. Pu *et al.*, "Orientation and Decision-Making for Soccer Based on Sports Analytics and AI: A Systematic Review," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 1, pp. 37–57, Jan. 2024, doi: 10.1109/JAS.2023.123807.
 - [16] C. Russo, C. Tommasino, and A. M. Rinaldi, "Leveraging Graphs for Advanced Analytics in Major Team Sports: A Systematic Mapping Study," *ACM Comput. Surv.*, vol. 58, no. 6, pp. 1–35, Apr. 2026, doi: 10.1145/3773024.
 - [17] A. Usta, A. Karakayali, and Ö. Ulusoy, "xDBTagger: explainable natural language interface to databases using keyword mappings and schema graph," *The VLDB Journal*, vol. 33, no. 2, pp. 301–321, Mar. 2024, doi: 10.1007/s00778-023-00809-w.
 - [18] M. Kang and S. Lee, "Is Sport Analytics a Saving Boat for Traditional Measurement and Evaluation in Kinesiology?," *Measurement in Physical Education and Exercise Science*, pp. 1–12, Dec. 2025, doi: 10.1080/1091367X.2025.2600451.
 - [19] S. Acharja and K. Chowdhury, "Natural Language Interfaces for Spatial and Temporal Databases: A Comprehensive Overview of Methods, Taxonomy, and Future Directions," *IEEE Access*, vol. 14, pp. 66971–66992, 2026, doi: 10.1109/ACCESS.2026.3686352.
 - [20] M. Kovoov, M. Durairaj, M. S. Karyakarte, M. Zair Hussain, M. Ashraf, and L. P. Maguluri, "Sensor-enhanced wearables and automated analytics for injury prevention in sports," *Measurement: Sensors*, vol. 32, p. 101054, Apr. 2024, doi: 10.1016/j.measen.2024.101054.
 - [21] Z. Zhan, E. Haihong, and M. Song, "Leveraging Large Language Model for Enhanced Text-to-SQL Parsing," *IEEE Access*, vol. 13, pp. 30497–30504, 2025, doi: 10.1109/ACCESS.2025.3540072.
 - [22] A. Alzahrani and A. Ullah, "Advanced biomechanical analytics: Wearable technologies for precision health monitoring in sports performance," *DIGITAL HEALTH*, vol. 10, p. 20552076241256745, Jan. 2024, doi: 10.1177/20552076241256745.
 - [23] D. Liu and W. Wang, "Advancing sports performance evaluation through T spherical fuzzy FUCA MCDM approach to real time basketball analytics," *Sci Rep*, vol. 16, no. 1, p. 3119, Jan. 2026, doi: 10.1038/s41598-025-32934-7.
 - [24] J. Davis *et al.*, "Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned," *Mach Learn*, vol. 113, no. 9, pp. 6977–7010, Sep. 2024, doi: 10.1007/s10994-024-06585-0.
 - [25] K. Ahkour and M. Machkour, "Towards an interface for translating natural language questions to SQL: a conceptual framework from a systematic review," *IJRIS*, vol. 12, no. 4, p. 264, 2020, doi: 10.1504/IJRIS.2020.111786.

-
- [26] J. Seelhöfer, J. Wick, and M. Witt, "Towards Continuous Swim Leg Analytics in Olympic Triathlon: A Systematic Review of Sensor-Based Assessment Approaches in Open-Water Sports Contexts," *Sensors*, vol. 26, no. 7, p. 2151, Mar. 2026, doi: 10.3390/s26072151.
 - [27] M. Liu, X. Wang, J. Xu, H. Lu, and Y. Tong, "NALSpatial: A Natural Language Interface for Spatial Databases," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 4, pp. 2056–2070, Apr. 2025, doi: 10.1109/TKDE.2025.3525587.
 - [28] S. Zhou and M. Zhu, "Enhancing Athlete Performance Using Deep Learning Techniques in Sports Analytics," *IEEE Access*, vol. 13, pp. 210758–210771, 2025, doi: 10.1109/ACCESS.2025.3631851.
 - [29] H. S. Shalaan, T. H. A. Soliman, and A. M. AbdelAziz, "G-SQL: A Schema-Aware and Rule-Guided Approach for Robust Natural Language to SQL Translation," *IEEE Access*, vol. 13, pp. 158520–158534, 2025, doi: 10.1109/ACCESS.2025.3607879.
 - [30] Y. Gui, "Enhancing Physical Education Analytics: An Analytic Hierarchy Process-Based Approach for College Sport," *IEEE Access*, vol. 13, pp. 40317–40330, 2025, doi: 10.1109/ACCESS.2025.3542557.
 - [31] C. Yu *et al.*, "Monkuu: a LLM-powered natural language interface for geospatial databases with dynamic schema mapping," *International Journal of Geographical Information Science*, vol. 40, no. 2, pp. 588–609, Feb. 2026, doi: 10.1080/13658816.2025.2533322.
 - [32] H. M. Pandey, A. Gupta, S. Sarkar, M. Tomer, S. Johannes, and Y. Gong, "GEMMA-SQL: A Novel Text-to-SQL Model Based on Large Language Models," *Applied Artificial Intelligence*, vol. 39, no. 1, p. 2587371, Dec. 2025, doi: 10.1080/08839514.2025.2587371.
 - [33] A. Augustine, G. P. Redding, and S. L. Moan, "Toward Explainable Data and Sports Analytics: A Case Study on Pass Completion Prediction in American Football," *The American Statistician*, vol. 79, no. 4, pp. 435–448, Oct. 2025, doi: 10.1080/00031305.2025.2541085.
 - [34] Á. Becsei *et al.*, "A case study using sewage metagenomic data for assessment of text-to-SQL capabilities in large language models," *Sci Rep*, vol. 15, no. 1, p. 44557, Nov. 2025, doi: 10.1038/s41598-025-28341-7.
 - [35] V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Sports Analytics for Evaluating Injury Impact on NBA Performance," *Information*, vol. 16, no. 8, p. 699, Aug. 2025, doi: 10.3390/info16080699.
 - [36] H. Gong, "An approach for enhancing physical education analytics for college sports using AHP based MCDM algorithm integrated with MEREC method," *Sci Rep*, vol. 15, no. 1, p. 37034, Oct. 2025, doi: 10.1038/s41598-025-20626-1.
 - [37] R. Zhou and F. Zhang, "Refining Zero-Shot Text-to-SQL Benchmarks via Prompt Strategies with Large Language Models," *Applied Sciences*, vol. 15, no. 10, p. 5306, May 2025, doi: 10.3390/app15105306.
 - [38] S. Ojuri, T. A. Han, R. Chiong, and A. Di Stefano, "Optimizing text-to-SQL conversion techniques through the integration of intelligent agents and large language models," *Information Processing & Management*, vol. 62, no. 5, p. 104136, Sep. 2025, doi: 10.1016/j.ipm.2025.104136.
 - [39] D. C. Mănescu, "Big Data Analytics Framework for Decision-Making in Sports Performance Optimization," *Data*, vol. 10, no. 7, p. 116, Jul. 2025, doi: 10.3390/data10070116.
 - [40] G. Jiang, W. Li, C. Yu, Z. Zhu, and W. Li, "FGCSQL: A Three-Stage Pipeline for Large Language Model-Driven Chinese Text-to-SQL," *Electronics*, vol. 14, no. 6, p. 1214, Mar. 2025, doi: 10.3390/electronics14061214.
 - [41] Q. Zhang, Z. Wang, C. Long, and S.-M. Yiu, "Billiards Sports Analytics: Datasets and Tasks," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 9, pp. 1–27, Nov. 2024, doi: 10.1145/3686804.
 - [42] V. Sarlis and C. Tjortjis, "Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics," *Information*, vol. 15, no. 4, p. 242, Apr. 2024, doi: 10.3390/info15040242.
 - [43] J.-D. Kim and K. Okubo, "Customizing GPT for natural language dialogue interface in database access," *Genom. Inform.*, vol. 22, no. 1, p. 22, Nov. 2024, doi: 10.1186/s44342-024-00020-5.
 - [44] S. K. Thisani, "Developing a High-Performance Sports Results Prediction Artificial Neural Network: Case Study on World Championship Boxing," *International Journal of Computer Science in Sport*, vol. 23, no. 2, pp. 1–21, Jun. 2024, doi: 10.2478/ijcss-2023-0008.
 - [45] V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Leveraging Sports Analytics and Association Rule Mining to Uncover Recovery and Economic Impacts in NBA Basketball," *Data*, vol. 9, no. 7, p. 83, Jun. 2024, doi: 10.3390/data9070083.

-
- [46] A. Marshan, A. N. Almutairi, A. Ioannou, D. Bell, A. Monaghan, and M. Arzoky, "MedT5SQL: a transformers-based large language model for text-to-SQL conversion in the healthcare domain," *Front. Big Data*, vol. 7, p. 1371680, Jun. 2024, doi: 10.3389/fdata.2024.1371680.
 - [47] J. Exel and P. Dabnicki, "Precision Sports Science: What Is Next for Data Analytics for Athlete Performance and Well-Being Optimization?," *Applied Sciences*, vol. 14, no. 8, p. 3361, Apr. 2024, doi: 10.3390/app14083361.
 - [48] R. Liu, T. Lu, S. Yuan, H. Zhou, and M. Gowda, "SmartDampener: An Open Source Platform for Sport Analytics in Tennis," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 3, pp. 1–30, Aug. 2024, doi: 10.1145/3678507.
 - [49] S. Swamidurai, T. S. Murthy, and K. V. Sriharsha, "Translating natural language questions to SQL queries (nested queries)," *Multimed Tools Appl*, vol. 83, no. 15, pp. 45391–45405, Oct. 2023, doi: 10.1007/s11042-023-16987-2.
 - [50] V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Sports Analytics and Text Mining NBA Data to Assess Recovery from Injuries and Their Economic Impact," *Computers*, vol. 12, no. 12, p. 261, Dec. 2023, doi: 10.3390/computers12120261.
 - [51] O. Olaniyan, B. Dehe, D. Bamford, and S. Ward, "Enhancing on-pitch learning capabilities with data analytics and technologies in elite sports," *European Sport Management Quarterly*, vol. 24, no. 6, pp. 1195–1214, Nov. 2024, doi: 10.1080/16184742.2023.2270565.
 - [52] M. Muniz and T. Flamand, "Sports analytics for balanced team-building decisions," *Journal of the Operational Research Society*, vol. 74, no. 8, pp. 1892–1909, Aug. 2023, doi: 10.1080/01605682.2022.2118634.
 - [53] Y. Özgüven, U. Gönener, and S. Eken, "A Dockerized big data architecture for sports analytics," *ComSIS*, vol. 19, no. 2, pp. 957–978, 2022, doi: 10.2298/CSIS220118010O.
 - [54] M. S. Fury, L. S. Oh, and E. M. Berkson, "New Opportunities in Assessing Return to Performance in the Elite Athlete: Unifying Sports Medicine, Data Analytics, and Sports Science," *Arthroscopy, Sports Medicine, and Rehabil.*, vol. 4, no. 5, Oct. 2022, doi: 10.1016/j.asmr.2022.08.001.
 - [55] J. Wu, D. Liu, Z. Guo, Q. Xu, and Y. Wu, "TacticFlow: Visual Analytics of Ever-Changing Tactics in Racket Sports," *IEEE Trans. Visual. Comput. Graphics*, vol. 28, no. 1, pp. 835–845, Jan. 2022, doi: 10.1109/TVCG.2021.3114832.
 - [56] J. Wang *et al.*, "Tac-Trainer: A Visual Analytics System for IoT-based Racket Sports Training," *IEEE Trans. Visual. Comput. Graphics*, pp. 1–11, 2022, doi: 10.1109/TVCG.2022.3209352.
 - [57] X. Ma, X. Liu, X. Li, and Y. Ma, "Light-Weight, Self-Powered Sensor Based on Triboelectric Nanogenerator for Big Data Analytics in Sports," *Electronics*, vol. 10, no. 19, p. 2322, Sep. 2021, doi: 10.3390/electronics10192322.
 - [58] K. C. Houtmeyers, A. Jaspers, and P. Figueiredo, "Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics," *International Journal of Sports Physiology and Performance*, vol. 16, no. 11, pp. 1719–1723, Nov. 2021, doi: 10.1123/ijssp.2020-0958.
 - [59] L. Morra, F. Manigrasso, and F. Lamberti, "SocceER: Computer graphics meets sports analytics for soccer event recognition," *SoftwareX*, vol. 12, p. 100612, Jul. 2020, doi: 10.1016/j.softx.2020.100612.
 - [60] S. Ye, S. Feng, L. Huang, and S. Bian, "Recent Progress in Wearable Biosensors: From Healthcare Monitoring to Sports Analytics," *Biosensors*, vol. 10, no. 12, p. 205, Dec. 2020, doi: 10.3390/bios10120205.
 - [61] A. Solanki and A. Kumar, "A system to transform natural language queries into SQL queries," *Int. j. inf. technol.*, vol. 14, no. 1, pp. 437–446, Feb. 2022, doi: 10.1007/s41870-018-0095-2.
 - [62] G. Durán, "Sports scheduling and other topics in sports analytics: a survey with special reference to Latin America," *TOP*, vol. 29, no. 1, pp. 125–155, Apr. 2021, doi: 10.1007/s11750-020-00576-9.
 - [63] X. Su, Y. Gu, P. Wang, W. Gu, L. Qi, and J. He, "A robust natural language text-to-SQL generation framework with dynamic strategies based on LLMs," *Sci Rep*, vol. 16, no. 1, p. 7892, Feb. 2026, doi: 10.1038/s41598-026-39128-9.
 - [64] M. S. E. D. L. Parte, Y. Wang, J.-F. Martínez-Ortega, and N. L. Martínez, "A Secure and Efficient LLM-Based System for Natural Language Query-to-SQL Translation in IoT Data Management," *IEEE Access*, vol. 14, pp. 28435–28456, 2026, doi: 10.1109/ACCESS.2026.3665980.

- [65] A. Spelta, “An unbalanced optimal transport framework for histogram-valued regression with applications to sports analytics,” *J Big Data*, vol. 13, no. 1, p. 51, Feb. 2026, doi: 10.1186/s40537-026-01388-7.