



Analisis Data Artikel Sistem Pakar Menggunakan Metode *Systematic Review*

Helen Sastypratiwi^{#1}, Rudy Dwi Nyoto^{#2}

[#]Jurusan Informatika, Fakultas Teknik, Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak

¹helensastypratiwi@informatics.untan.ac.id

²rdn@informatika.untan.ac.id

Abstrak—Sistem pakar memiliki berbagai keunggulan dibandingkan kepakaran manusia karena sistem pakar terjangkau, permanen, konsisten, proses yang cepat, dan dapat digandakan. Hal ini menyebabkan sistem pakar berkembang diberbagai bidang. Perlu adanya kajian tentang data implementasi maupun metode yang digunakan dalam sistem pakar. Penelitian ini bertujuan untuk menganalisis data pada artikel sistem pakar dengan menggunakan metode *systematic review* serta menentukan string yang cocok dalam pengumpulan data. Langkah penelitian menggunakan *flowchart* PRISMA dengan penentuan string yang sesuai berdasarkan lingkup sistem pakar dan teknik dalam *data mining*. Pencarian dilakukan pada *database online*. Metode *systematic review* ini, dapat menjadi salah satu alternatif metode dalam penulisan karya ilmiah berdasarkan artikel sebelumnya. Dimana hasil dalam artikel sistem pakar sebagai studi dalam penelitian bahwa bidang yang masih unggul adalah *computer science* serta *prediction* adalah teknik yang paling banyak digunakan.

Kata kunci — Analisis Data, Artikel, *Data Mining*, Sistem Pakar, *String*, *Systematic Review*

I. PENDAHULUAN

Sistem pakar adalah bagian dari Artificial Intelligence (AI), dan ditemukan oleh komunitas AI dipertengahan tahun 1960. Ide dasar dibalik sistem pakar adalah untuk mempermudah para ahli yang memiliki pengetahuan yang spesifik untuk ditransfer ke dalam sebuah komputer. Pengetahuan ini berikutnya disimpan ke dalam komputer dan dapat dipanggil oleh pengguna saat diperlukan. Selanjutnya seperti konsultasi yang terjadi pada manusia, komputer dapat memberikan masukan dan penjelasan [1]. Diperkuat pula bahwa sistem pakar adalah sebuah program komputer atau sebuah perangkat lunak yang memiliki pengetahuan dari seorang pakar dalam menghadapi suatu masalah [2]. Pengetahuan ini digunakan oleh sistem untuk menyelesaikan masalah tersebut sama seperti seorang pakar.

Sistem pakar telah berkembang di berbagai subjek seperti pertanian, ilmu komputer, kimia, kedokteran, geology, *space technology* dan lainnya. Sistem pakar memiliki berbagai keunggulan dibandingkan kepakaran

manusia karena sistem pakar terjangkau, permanen, konsisten, proses yang cepat, dan dapat digandakan. Sedangkan kepakaran manusia mudah rusak, tidak dapat diprediksi, dan mahal serta lambat dalam pemrosesan dan perkembangannya. Namun, metodologi sistem pakar cenderung berkembang menuju kearah *problem-oriented* dan diperlukannya metodologi baru yang memanfaatkan ilmu sosial seperti psikologi, ilmu kognitif dan perilaku manusia yang dapat diimplementasikan oleh sistem pakar sebagai pilihan metode lainnya [3].

Pada saat ini, data telah membanjiri berbagai macam aspek. Data dengan mudah didapat dari manapun. Saat penelitian ini dimulai, era seperti itu disebut dengan era big data. Era big data menjadikan data mudah untuk di akses dalam jumlah yang sangat besar baik data yang terstruktur maupun data yang tidak terstruktur. Secara umum big data dapat diartikan sebagai sebuah kumpulan data yang berukuran sangat besar (*volume*), sangat cepat berubah/bertumbuh (*velocity*), hadir dalam beragam bentuk/format (*variety*), serta memiliki nilai tertentu (*value*), dengan catatan jika berasal dari sumber yang akurat (*veracity*)[4].

Diperjelas dalam penelitian lain, bahwa hal utama yang membedakan big data dengan kumpulan data konvensional terletak pada mekanisme pengelolaannya. Sistem basis data relasional yang saat ini umum digunakan, sudah dirasakan tidak mampu menangani kompleksitas big data secara optimal [5].

Berbagai metode muncul khususnya yang berkaitan dengan pengolahan data dalam berbagai artikel. *Data mining* menjadi pendekatan maupun metode dalam pengolahan data yang banyak digunakan dalam era big data. Sehingga cakupan tidak terbatas mengenai metode yang ada didalam sistem pakar tetapi juga metode yang menggabungkan antara sistem pakar dengan metode-metode yang ada didalam *data mining*.

Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang berisikan dari kecerdasan bautan, mesin pembelajaran, statistic dan sistem basis data [6].

Artikel yang berkaitan dengan sistem pakar dan *data mining* ini tentu memiliki jumlah yang sangat banyak. Sehingga hasil pemikiran atau penelitian ini tidak akan menghasilkan banyak manfaat maupun informasi, bila tidak diolah dengan baik untuk perkembangan ilmu pengetahuan dimasa kini maupun dimasa yang akan datang. Salah satu teknik untuk menggali informasi dari artikel adalah dengan menggunakan metode *systematic review*.

Systematic review dikenal dengan metode yang menggabungkan banyak studi orisinal dengan mengklasifikasikan berdasarkan kriteria dan dilaksanakan secara terstruktur dan terencana sehingga dapat meningkatkan kedalaman dalam *mereview* dan membuat ringkasan dalam *evidence* riset [7]. Metode ini mencari, menilai dan menyusun semua bukti empiris yang relevan untuk memberikan interpretasi lengkap dari hasil penelitian. *Systematic review* juga membantu mengidentifikasi kesenjangan penelitian dalam pemahaman dalam suatu bidang.

Dalam penelitian ini, data yang dikumpulkan akan dianalisis khususnya data artikel berdasarkan judul, metode dan aplikasi dalam sistem pakar serta kesesuaiannya didalam prinsip *data mining* yang sesuai dengan tahapan dalam *systematic review*. Diharapkan metode ini dapat mengidentifikasi gap dalam penelitian sehingga dapat digunakan dalam penelitian selanjutnya dalam mengembangkan ide-ide baru dan meningkatkan keterampilan dengan menggunakan literatur yang ada [8][9].

II. METODOLOGI PENELITIAN

Penelitian ini dibangun terdiri dari empat tahap besar, detail tahapan besar tersebut dijelaskan tersendiri. Tahapan dan detail tahapan akan diikuti dengan metode pelaksanaan. Pada proses pencarian literatur, untuk pemilihan hasil pencarian dilakukan dengan menyaring hasil pencarian berdasarkan kriteria yang telah ditetapkan [10]. Kriteria yang digunakan adalah inklusi dan eksklusi yang kemudian dijadikan dasar pemilihan literatur.

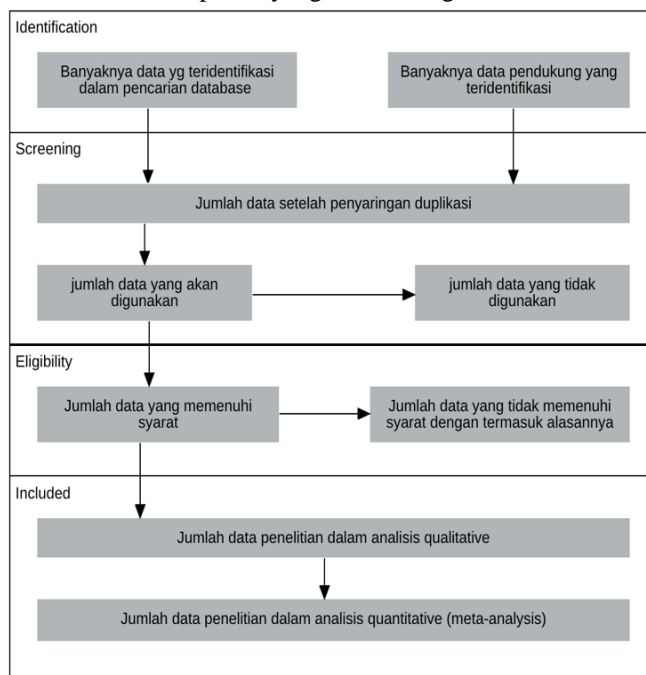
Langkah dalam pencarian dibagi atas beberapa proses yaitu *identification*, *screening*, *eligibility* dan *included*. Langkah ini telah sesuai dengan pedoman dalam PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*). PRISMA adalah serangkaian *evidence-based* minimum berbasis bukti yang bertujuan membantu penulis melaporkan beragam tinjauan sistematis dan meta-analisis yang menilai manfaat. PRISMA berfokus pada cara-cara di mana penulis dapat memastikan pelaporan yang transparan dan lengkap dari jenis penelitian [11].

Rentang khusus pencarian adalah tahun 2015 – 2019 dengan alasan, kepopuleran konsep big data mulai dikenal luas mulai dari tahun 2015. Pencarian ini berbasis pada index dari kata kunci yang digunakan. Secara rinci program penelitian dijelaskan pada gambar 1.

Systematic review dibatasi hanya pada penelitian yang berbentuk artikel. Artikel yang digunakan adalah artikel yang merupakan riset yang telah dikaji dan diterbitkan dalam jurnal berbahasa Inggris. Dalam pengelolaan artikel yang didapat dari *database online*, penulis menggunakan

tools Mendeley. Mendeley adalah program komputer dan web yang dikembangkan Elsevier untuk mengelola dan berbagi makalah penelitian mencari data penelitian, dan bekerja sama secara daring. Mendeley menggabungkan Mendeley Desktop, perangkat lunak manajemen referensi dan PDF, dengan Mendeley Android and iOS dan Mendeley Web, jejaring sosial peneliti.

Pada tahapan *identification*, digunakannya tiga pencarian literatur yang akan dilakukan pada *database online* yang memiliki repositori besar untuk studi akademis yaitu IEEE Xplore, Science Direct dan Emerald. Pada proses kedua dari tahapan *identification*, dilakukan pula penelusuran terhadap data yang mendukung.



Gambar. 1 Alur metode penelitian PRISMA

Tahapan selanjutnya adalah *screening*, pada tahapan ini terbagi atas tiga proses. Pertama, menentukan string (kata kunci) yang akan digunakan dalam penelusuran. Penggunaan string yang digunakan terbagi atas beberapa fase. Pada fase pertama penelusuran menggunakan string “Expert”AND”System”. Setelah didapat hasil, selanjutnya fase penelusuran dengan string terkait *data mining*. String yang digunakan adalah tujuh teknik dari *data mining*. Untuk lebih jelasnya dapat dilihat pada gambar 2.

Adapun teknik dalam *data mining* adalah sebagai berikut [12][13]:

A. Association

Asosiasi saling berhubungan dengan pengenalan pola. Asosiasi lebih spesifik kepada variable-variabel yang saling berhubungan. Dalam kasus ini, data dapat dilihat sesuai dengan spesifikasinya dan yang saling berkorelasi erat. Sebagai contoh, saat berbelanja seorang pelanggan dapat membeli barang tertentu. Selain barang tersebut, pelanggan juga membeli barang lainnya yang terkait dengan barang sebelumnya. Dalam contoh nyata dapat

dilihat pada online store dengan menggunakan kata kunci "orang juga membeli".

B. Classification

Klasifikasi adalah teknik penambangan data yang lebih kompleks yang mengharuskan pengumpulan berbagai atribut secara bersama-sama ke dalam kategori. Kategori-kategori ini digunakan untuk menarik kesimpulan lebih lanjut atau dapat digunakan dalam beberapa fungsi [14].

Misalnya, saat mengevaluasi data tentang latar belakang keuangan pelanggan perorangan dan riwayat pembelian, maka diklasifikasikan sebagai risiko kredit "rendah," "sedang," atau "tinggi". Proses klasifikasi ini dapat digunakan untuk mempelajari pembeli-pembeli lainnya.

C. Clustering

Clustering sangat mirip dengan klasifikasi, tetapi melibatkan pengelompokan potongan data yang berbasis pada similiaritas. Misalnya, Anda dapat memilih untuk mengelompokkan demografi yang berbeda dari audiens Anda ke dalam paket yang berbeda berdasarkan berapa banyak pendapatan yang mereka miliki, atau seberapa sering mereka cenderung berbelanja di toko Anda.

D. Outlier

Dalam banyak kasus, dengan hanya menggunakan pengenalan pola tidak dapat memberi banyak pemahaman yang jelas tentang kumpulan data. Perlu adanya pengidentifikasian anomali, atau outlier dalam data. Misalnya, jika pembeli hampir semuanya laki-laki, tetapi selama satu minggu yang di bulan Juli, ada lonjakan besar pada pembeli perempuan, perlu penyelidikan lonjakan dan melihat apa yang mendorongnya, sehingga dapat meniru atau lebih memahami audiens dalam sebuah proses.

E. Prediction

Prediksi adalah salah satu teknik penambangan data yang paling berharga, karena digunakan untuk memproyeksikan jenis data yang akan Anda lihat di masa mendatang. Dalam banyak kasus, hanya mengenali dan memahami tren historis sudah cukup untuk memetakan prediksi yang agak akurat tentang apa yang akan terjadi di masa depan. Misalnya, Anda dapat meninjau riwayat kredit konsumen dan pembelian sebelumnya untuk memperkirakan apakah mereka akan menjadi risiko kredit di masa depan.

F. Regression

Regresi, digunakan terutama sebagai bentuk perencanaan dan pemodelan, digunakan untuk mengidentifikasi kemungkinan variabel tertentu, mengingat adanya variabel lain. Misalnya, Anda dapat menggunakannya untuk memproyeksikan harga tertentu, berdasarkan faktor-faktor lain seperti ketersediaan, permintaan konsumen, dan persaingan. Lebih khusus lagi, fokus utama regresi adalah untuk membantu mengungkap hubungan yang tepat antara dua (atau lebih) variabel dalam kumpulan data yang diberikan.

Setelah proses penulisan selesai, proses selanjutnya dari *screening* adalah menyaring data ganda atau duplikat. Serta dilakukan filter terhadap judul, abstrak, maupun kata kunci artikel. Eksplorasi serta pemilihan judul, abstrak dan kata kunci pada artikel yang didapatkan dari hasil pencarian berdasarkan kriteria kelayakan yang telah didefinisikan sebelumnya. Rentang waktu berdasarkan tahun 2015 – 2019 termasuk dalam tahapan *screening*. Sehingga pada proses akhir dari tahapan ini didapat jumlah data yang digunakan dan tidak digunakan untuk review selanjutnya.

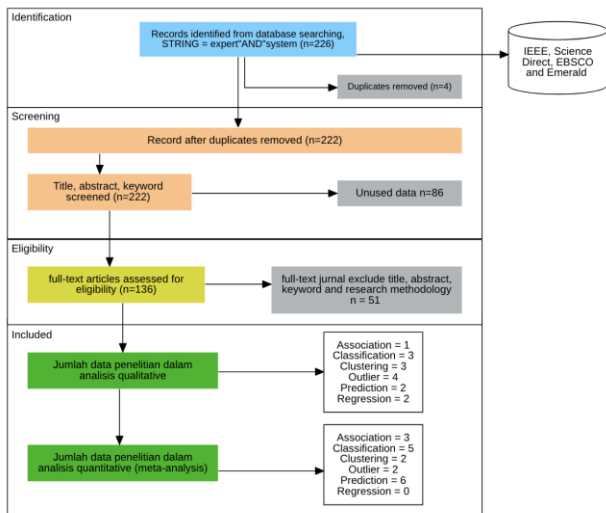
Dalam tahapan *eligibility*, dilakukan dengan cara membaca lengkap atau parsial artikel yang belum tereliminasi pada tahapan sebelumnya untuk menentukan apakah artikel tersebut harus dimasukkan dalam kajian selanjutnya sesuai dengan kriteria kelayakan. Daftar referensi dari artikel terpilih dikaji kembali untuk menemukan studi terkait lainnya. Sehingga diperoleh dua bagian yaitu yang memenuhi syarat dan yang tidak memenuhi syarat. Syarat yang harus terpenuhi dalam tahapan ini adalah kesesuaian metode penelitian dan teknik yang digunakan dalam penelitian.

Tahap terakhir adalah *included*, dimana data dikelompokkan berdasarkan teknik yang digunakan serta bidang dari studi kasus yang digunakan serta pendekatan yang digunakan dan dibedakan berdasarkan dua jenis metode penelitian yaitu kualitatif dan kuantitatif.

Penelitian kuantitatif biasanya melibatkan penyelidikan sistematis dan fenomena yang empiris melalui statistik dan matematika dengan pemrosesan data numerik. Dalam penelitian kuantitatif data biasanya dipilih dan dianalisis dalam bentuk numerik [15]. Sedangkan penelitian kualitatif adalah istilah yang sangat luas yang mencakup metodologi penelitian yang berhubungan dengan fenomena dengan metode menganalisis pengalaman, perilaku dan hubungan tanpa menggunakan statistik dan matematika serta pemrosesan data numerik [16][17]. Pendekatan kualitatif biasanya memberikan jawaban untuk pertanyaan penelitian seperti: (a) apa, (b) bagaimana, (c) kapan dan (d) di mana yang dapat digambarkan sebagai pendekatan penelitian berbasis kata [18].

III. HASIL DAN PEMBAHASAN

Total penelitian yang dapat direview adalah sebanyak 86 artikel dengan menggunakan database IEEE, Science Direct, dan Emerald. Hasil akhir artikel yang digunakan setelah di kategorikan berdasarkan kluster bidang dapat dilihat pada tabel 3 pada tahapan *eligibility*. Gambar 2 dibawah menjelaskan secara rinci semua tahapan dengan mengadopsi *flowchart* PRISMA yang telah didefinisikan pada gambar 1 diatas.



Gambar. 2 PRISMA flowchart artikel sistem pakar

PRISMA membedakan setiap tahapan dalam proses *systematic review* menjadi:

A. Identification

Pada proses identifikasi, Ditentukan *database* yang digunakan adalah IEEE, Science Direct dan Emerald. String yang digunakan dalam pencarian pertama adalah “Sistem”AND”Pakar”. Hasil dari pencarian berdasarkan tiga sumber yang dapat dilihat pada tabel 1 di bawah. Jumlah identifikasi berdasarkan string yang dimasukkan, didapat sebanyak 1991 yang sesuai dengan topik.

TABEL I
JUMLAH DATA TERIDENTIFIKASI

Sumber	URL	Hasil Pencarian
IEEE	ieeexplore.ieee.org	981
Science Direct		983
Emerald		27
Jumlah		1991

B. Screening

Proses *screening* tidak terbatas pada satu string saja, terdiri dari beberapa string yang digambarkan oleh table 2. String ditentukan berdasarkan teknik yang digunakan dalam data mining yang telah dibahas sebelumnya diatas.

TABEL II
STRING PENELITIAN

Kata Kunci	String Penelusuran
Expert System	expert”AND”system
Tracking Patterns	tracking”AND”patterns
Association	association
Classification	classification
Clustering	clustering
Outlier	outlier
Prediction	prediction

C. Egibility

Dari hasil screening diperoleh sebanyak 136 artikel yang sesuai dan 51 artikel yang tidak termasuk untuk tahapan selanjutnya. Sehingga total ada 86 artikel yang digunakan. Pada tabel 3, data dibedakan dalam klaster teknik dalam

data mining. Pada proses ini didapat 33 artikel digunakan. Hal ini tentunya berbeda dengan jumlah sebelumnya yaitu 86 artikel karena proses dilakukan dengan cara membaca lengkap dan atau parsial.

Jumlah sebelumnya didapat berdasarkan judul saja. Setelah dilakukan kajian terhadap metodologi maupun abstrak terhadap 86 artikel, menunjukkan beberapa artikel memiliki lebih dari satu pendekatan yang digunakan dalam tekniknya dan terjadi duplikasi didalam pengklusteran. Sehingga perlu *review* kembali berdasarkan abstrak, *keyword* dan metodologi penelitian. Ditemukan 33 artikel yang sesuai yang dapat dilihat pada tabel 3.

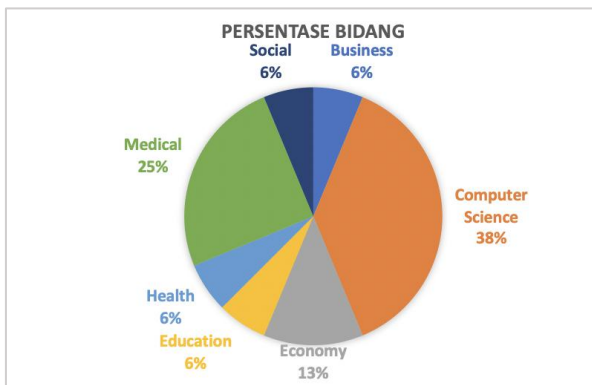
TABEL III
KLUSTER BIDANG PENERAPAN KESELURUHAN

Data mining Task	Applied Approaches	Subject	Authors	
Association Classification	Misdiagnosis Minimization Approach	Health	[19]	
	Relationships in learning systems	Education	[20]	
	Association Rule and Link Prediction	Health	[21]	
	Supervised and unsupervised data mining techniques	Engineering	[22]	
	Bayesian	Economy	[23]	
	Novel semi-supervised learning using apriori algorithm	Technology	[24]	
	Social network analysis	Business & Computer Science	[25]	
	Machine learning; Neural Network	Health	[26]	
	Genetic algorithms	Economy	[27]	
	Classification accuracy	Computer Science	[28]	
Clustering	Machine learning	Social	[29]	
	Machine Learning	Computer Science	[30]	
	Hierarchical clustering, K-Medoids, fuzzy clustering, and Self-Organising Maps (SOM)	Social	[31]	
	Systematic Mapping Study	Computer Science	[32]	
	Fuzzy weighted clustering	Computer Science	[33]	
	MULTIMOORA, K-Means, Cluster analysis for improving Multiple Criteria Decision Analysis	Computer Science	[34]	
	K-Means, SOM	Computer Science	[35]	
	Outlier	Outlier detection	Engineering	[36]
		Fuzzy weighted clustering	Computer Science	[33]
		Rule Based Classification, Complex Event Processing	Computer Science	[37]
Load modelling		Computer Science	[38]	
Data quality assessment		Medical	[39]	
Prediction	Cluster analysis	Medical	[31]	
	Machine Learning	Sport	[40]	
	Matrix factorization	Computer Science	[41]	
	Bayes, Text mining	Business	[23]	
	Machine Learning	Economy	[42]	
	Genetic Algorithm, Neural network	Industry	[43]	
	sliding window-based support vector regression	Geophysics	[44]	
	Regression	Logistic regression; Machine learning; Neural network; Support vector machine	Medical	[26]
Multi-criteria decision analysis; Spatial analysis		Medical	[45]	

D. Included

Dalam penelusuran metode kualitatif dan kuantitatif dilakukan *review* terhadap artikel dengan membaca penuh isi dengan memperhatikan metode pengumpulan data dimana metode yang digunakan dalam penelitian kualitatif adalah observasi, analisis visual, studi pustaka, dan interview. Dapat dilihat pada poin berikut:

1) *Data Kualitatif* : Penelitian kualitatif ini bersifat dinamis, artinya selalu terbuka untuk adanya perubahan, penambahan, dan penggantian selama proses analisisnya. Metode kualitatif merupakan metode yang fokus pada pengamatan yang mendalam. Oleh karenanya, penggunaan metode kualitatif dalam penelitian dapat menghasilkan kajian atas suatu fenomena yang lebih komprehensif. Penelitian kualitatif yang memperhatikan humanisme atau individu manusia dan perilaku manusia merupakan jawaban atas kesadaran bahwa semua akibat dari perbuatan manusia terpengaruh pada aspek-aspek internal individu.



Gambar. 3 Sebaran bidang penggunaan data kualitatif

Pada gambar 3, diagram menunjukkan persentasi sebaran penerapan bidang yang menggunakan sistem pakar dalam implementasinya. *Computer science* menunjukkan persentasi terbesar dibanding enam bidang lainnya dan diikuti oleh *medical*, *economy*. Ini menunjukkan bahwa sistem pakar sudah diterapkan tidak hanya dalam bidang *computer science* saja tetapi dalam bidang lainnya. Terutama *medical* yang dari awal berkembangnya sistem pakar adalah untuk mendiagnosis penyakit atau yang lebih dikenal dengan MYCIN [46].

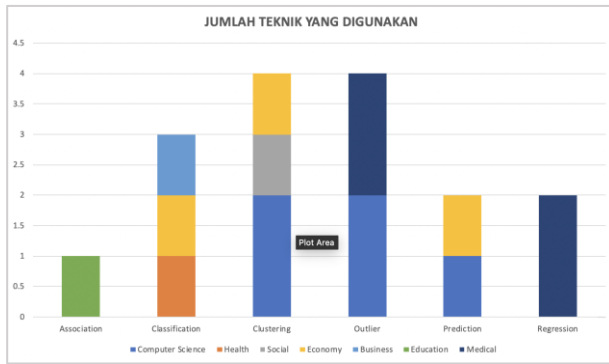
Seiring perkembangan zaman, sistem pakar tidak lagi digunakan dalam dunia medis saja tetapi juga dalam bidang ekonomi. Dibuktikan ekonomi menempati posisi ketiga dari persentasi. *Social*, *business*, *health* dan *education* memiliki persentasi yang sama besar yaitu sebesar 6%. Untuk melihat teknik dan pendekatan yang digunakan dalam masing-masing penelitian, dibuatlah klusterisasi berdasarkan bidang yang dapat dilihat pada tabel 4 dibawah.

TABEL IV
KLUSTER BIDANG PENERAPAN

Subject	Data mining Task	Applied Approaches	Authors
Business & Computer Science	Classification	Social network analysis	[25]
Computer Science	Clustering	Systematic Mapping Study	[32]
	Clustering	K-Means, SOM	[35]
	Outlier	Fuzzy weighted clustering	[33]
	Outlier	Load modelling	[38]
	Prediction	Local search algorithm	[25]
Economy	Classification	Genetic algorithms	[27]
	Prediction	Machine Learning	[42]
Education	Association	Relationships in learning systems	[20]
Health	Classification	Machine learning; Neural Network	[39]
Medical	Outlier	Cluster analysis	[31]
	Outlier	Data quality assessment	[39]
	Regression	Logistic regression; Machine learning; Neural network; Support vector machine	[26]
	Regression	Multi-criteria decision analysis; Spatial analysis	[45]
Social	Clustering	Hierarchical clustering, K-Medoids, fuzzy clustering, and Self-Organising Maps (SOM)	[31]

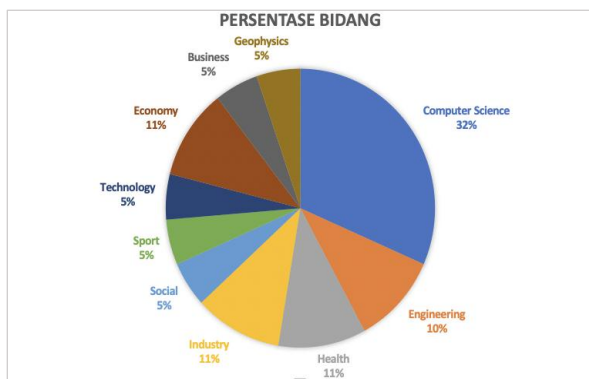
Teknik yang paling banyak digunakan pada penelitian kualitatif adalah *clustering* dan *outlier* dapat dilihat pada gambar 4 dibawah. *Clustering* sendiri merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Sehingga wajar bila tingkat penggunaan teknik ini lebih tinggi baik pada bidang *computer science* maupun pada ilmu sosial itu sendiri dan ekonomi. Dikarenakan menggunakan prinsip yang sederhana yang dapat di jelaskan dalam non-statistik serta sangat fleksibel sehingga adaptasi yang mudah untuk di lakukan.

Senada dengan *outlier*, teknik ini menggunakan banyaknya data mjdcul yang memiliki karakteristik unik yang terlihat sangat jauh berbeda dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi.



Gambar. 4 Teknik dalam penelitian kualitatif

2) *Data Kuantitatif* : Dalam penelitian kuantitatif, data set dikumpulkan, diolah dan dianalisis untuk dicari hubungan antar variabel yang diteliti. Variabel yang digunakan bisa dua atau lebih.



Gambar. 5 Sebaran bidang penggunaan data kuantitatif

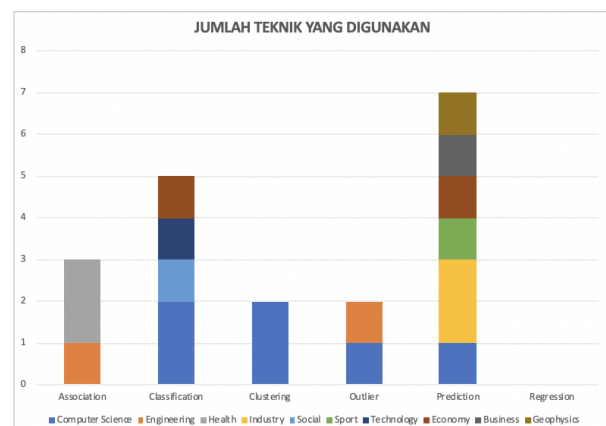
Pada gambar 5, jumlah penelitian yang menggunakan data kuantitatif dalam penelitiannya terbanyak pada bidang *computer science*. Hal ini serupa dengan penelitian yang menggunakan data kualitatif bahwa *computer science* memiliki presentasi terbesar. Selanjutnya adalah ekonomi, lalu disusul berturut-turut oleh kesehatan, industri, *engineering*. Pada bidang *technology*, *sport*, *social* dan *business* besar presentasi adalah sama yaitu 5%. Rincian dari Teknik dan pendekatan yang digunakan dalam memperkuat gambar 5 diperjelas pada tabel 5 yang diklasifikasi berdasarkan bidang.

TABEL V
KLUSTER BIDANG PENERAPAN

Subject	Data mining Task	Applied Approaches	Authors
Computer Science	Classification	Machine Learning	[47]
	Classification	Classification accuracy	[28]
	Clustering	Fuzzy weighted clustering	[33]
	Clustering	MULTIMOORA, K-Means, Cluster analysis for improving Multiple Criteria Decision Analysis	[34]

	Outlier	Rule Based Classification, Complex Event Processing	[37]
	Prediction	Matrix factorization	[41]
Engineering	Association	Supervised and unsupervised data mining techniques	[22]
	Outlier	Outlier detection	[36]
Health	Association	Association Rule and Link Prediction	[48]
	Association	Misdiagnosis Minimization Approach	[19]
Industry	Prediction	Genetic Algorithm, Neural network	[43]
Industry, Economy	Prediction	cluster-based undersampling, genetic algorithm	[27]
Social	Classification	Machine learning	[29]
Sport	Prediction	Machine Learning	[40]
Technology	Classification	Novel semi-supervised learning using apriori algorithm	[24]
Economy	Classification	Bayesian	[23]
Business	Prediction	Bayes, Text mining	[23]
Geophysics	Prediction	sliding window-based support vector regression	[44]

Penelitian kuantitatif menggunakan data numerik dan menekankan proses penelitian pada pengukuran hasil yang objektif menggunakan analisis statistik. Fokus metode kuantitatif adalah mengumpulkan data set dan melakukan generalisasi untuk menjelaskan fenomena khusus yang dialami oleh populasi. Pada penelitian kuantitatif, *prediction* menjadi teknik yang paling banyak digunakan. Karena teknik *prediction* menggunakan data untuk memprediksi data dimasa yang akan datang. Bidang *geophysics* contohnya, dapat dilihat pada tabel 5 menggunakan teknik *prediction* yang tidak ditemukan bidangnya pada penelitian kualitatif. Penelitian tersebut menggunakan teknik *prediction* dalam memprediksi *micrometeorological* data.



Gambar. 6 Teknik dalam penelitian kuantitatif

IV. KESIMPULAN

Metode *systematic review* yang digunakan dalam pencarian artikel sistem pakar yang menggunakan teknik dalam data mining mampu memfilter data sehingga ditemukannya informasi-informasi. *Tools* yang digunakan tentunya menjadi faktor pendukung. Begitu pula *database* yang digunakan mengingat banyak data yang diperlukan guna mencapai tujuan. Digunakannya *systematic review* menjadi metode ini sangat bermanfaat dalam penyaduran dalam bidang sistem pakar terhadap artikel dalam bidang yang sama. Penyaringan data memberikan hasil bahwa bidang *computer science* masih menjadi unggulan dalam penggunaan sistem pakar. Juga pada bidang seperti *medical* dan kesehatan dimana awal dari sistem pakar dikembangkan masih menjadi daya tarik tersendiri. Akan tetapi, bidang sosial saat ini telah banyak pula menggunakan sistem pakar dalam penerapannya. Sehingga perlu kiranya penelitian sistem pakar selanjutnya dapat membantu bidang-bidang yang memang masih minim seperti olahraga maupun sosial itu sendiri.

Didapat pula teknik yang paling banyak digunakan yaitu prediksi, klasifikasi dan *outlier*. Sehingga bila ingin menggunakan teknik yang sama, tidak ada lagi keterbatasan dalam penelusuran yang digunakan sebagai kajian. Hal ini tidak menutup kemungkinan penggunaan teknik *data mining* lainnya guna memperbanyak literatur dalam penelitian ilmiah.

REFERENSI

- [1] E. Turban, T.-P. Liang, and J. E. Aronson, *Decision Support Systems and Intelligent Systems: (International Edition)*. Pearson Prentice Hall, 2005.
- [2] J. Durkin, "Research review: Application of expert systems in the sciences," 1990.
- [3] H. Liao, Z. Xu, X. Zeng, and J. M. Merigó, "Framework of Group Decision Making With Intuitionistic Fuzzy Preference Information," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 1211–1227, 2015.
- [4] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. networks Appl.*, vol. 19, no. 2, pp. 171–209, 2014.
- [5] J. Manyika et al., "Big data: The next frontier for innovation, competition, and productivity," 2011.
- [6] A. C. M. SIGKDD, "Data mining curriculum." ed, 2012.
- [7] I. K. Crombie and H. T. Davies, "What is meta-analysis," *What is*, pp. 1–8, 2009.
- [8] A. H. Eagly and W. Wood, "Using research syntheses to plan future research," 1994.
- [9] I. Chalmers and P. Glasziou, "Avoidable waste in the production and reporting of research evidence," *Lancet*, vol. 374, no. 9683, pp. 86–89, 2009.
- [10] D. Dang and S. Pekkola, "Systematic Literature Review on Enterprise Architecture in the Public Sector," *Electron. J. e-Government*, vol. 15, no. 1, p. 25, 2016.
- [11] V. Welch et al., "Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration," *J. Dev. Eff.*, vol. 8, no. 2, pp. 287–324, 2016.
- [12] L. Alton, "The 7 Most Important Data Mining Techniques," 2017. [Online]. Available: <https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques>. [Accessed: 28-Oct-2019].
- [13] Y. Zhang and Y. Zhao, "Astronomy in the Big Data Era," *Data Sci. J.*, vol. 14, no. 0, p. 11, 2015.
- [14] H. Zaugg, R. E. West, I. Tateishi, and D. L. Randall, "Mendeley: Creating communities of scholarly inquiry through research collaboration," *TechTrends*, vol. 55, no. 1, pp. 32–36, 2011.
- [15] N. Basias and Y. Pollalis, "Quantitative and qualitative research in business & technology: Justifying a suitable research methodology," *Rev. Integr. Bus. Econ. Res.*, vol. 7, pp. 91–105, 2018.
- [16] S. B. Merriam and E. J. Tisdell, *Qualitative research: A guide to design and implementation*. John Wiley & Sons, 2015.
- [17] M. Hennink, I. Hutter, and A. Bailey, *Qualitative research methods*. SAGE Publications Limited, 2020.
- [18] M. B. Miles and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook*. sage, 1994.
- [19] J. A. Rodger, "Discovery of medical Big Data analytics: Improving the prediction of traumatic brain injury survival rates by data mining Patient Informatics Processing Software Hybrid Hadoop Hive," *Informatics Med. Unlocked*, vol. 1, pp. 17–26, 2015.
- [20] L. Cao, "Coupling learning of complex interactions," *Inf. Process. Manag.*, vol. 51, no. 2, pp. 167–186, 2015.
- [21] W. S. Lee, E. J. Han, and S. Y. Sohn, "Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents," *Technol. Forecast. Soc. Change*, vol. 100, pp. 317–329, 2015.
- [22] M. Reder, N. Y. Yürüşen, and J. J. Melero, "Data-driven learning framework for associating weather conditions and wind turbine failures," *Reliab. Eng. Syst. Saf.*, vol. 169, pp. 554–569, 2018.
- [23] K. Coussement, D. F. Benoit, and M. Antioco, "A Bayesian approach for incorporating expert opinions into decision support systems: A case study of online consumer-satisfaction detection," *Decis. Support Syst.*, vol. 79, pp. 24–32, 2015.
- [24] I. Ahmed, R. Ali, D. Guan, Y.-K. Lee, S. Lee, and T. Chung, "Semi-supervised learning using frequent itemset and ensemble learning for SMS classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1065–1073, 2015.
- [25] Y.-N. Park, Y.-S. Lee, J.-J. Kim, and T. S. Lee, "The structure and knowledge flow of building information modeling based on patent citation network analysis," *Autom. Constr.*, vol. 87, pp. 215–224, 2018.
- [26] O. Ben-Assuli, T. Heart, N. Shlomo, and R. Klempfner, "Bringing big data analytics closer to practice: A methodological explanation and demonstration of classification algorithms," *Heal. Policy Technol.*, vol. 8, no. 1, pp. 7–13, 2019.
- [27] H.-J. Kim, N.-O. Jo, and K.-S. Shin, "Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction," *Expert Syst. Appl.*, vol. 59, pp. 226–234, 2016.
- [28] W. Ectors et al., "Optimizing copious activity type classes based on classification accuracy and entropy retention," *Futur. Gener. Comput. Syst.*, 2018.
- [29] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for Big Data professions: A systematic classification of job roles and required skill sets," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 807–817, 2018.
- [30] V. Bolon-Canedo, D. Fernández-Francos, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdiñas, and N. Sánchez-Marroño, "A unified pipeline for online feature selection and classification," *Expert Syst. Appl.*, vol. 55, pp. 532–545, 2016.
- [31] D. Arunachalam and N. Kumar, "Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making," *Expert Syst. Appl.*, vol. 111, pp. 11–34, 2018.
- [32] J. G. Enríquez, F. J. Domínguez-Mayo, M. J. Escalona, M. Ross, and G. Staples, "Entity reconciliation in big data sources: A systematic mapping study," *Expert Syst. Appl.*, vol. 80, pp. 14–27, 2017.
- [33] B. Hosseini and K. Kiani, "FWCMR: A scalable and robust fuzzy weighted clustering based on MapReduce with application to microarray gene expression," *Expert Syst. Appl.*, vol. 91, pp. 198–210, 2018.
- [34] A. Ijadi Maghsoodi, A. Kavian, M. Khalilzadeh, and W. K. M. Brauers, "CLUS-MCDA: A novel framework based on cluster analysis and multiple criteria decision theory in a supplier selection problem," *Comput. Ind. Eng.*, vol. 118, pp. 409–422, 2018.

- [35] W. Qadadeh and S. Abdallah, "Customers Segmentation in the Insurance Company (TIC) Dataset," *Procedia Comput. Sci.*, vol. 144, pp. 277–290, 2018.
- [36] G. Manco *et al.*, "Fault detection and explanation through big data analysis on sensor streams," *Expert Syst. Appl.*, vol. 87, pp. 141–156, 2017.
- [37] N. Mehdiyev, J. Krumeich, D. Enke, D. Werth, and P. Loos, "Determination of Rule Patterns in Complex Event Processing Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 61, pp. 395–401, 2015.
- [38] R. M. Salgado, T. C. Machado, and T. Ohishi, "Intelligent Models to Identification and Treatment of Outliers in Electrical Load Data," *IEEE Lat. Am. Trans.*, vol. 14, no. 10, pp. 4279–4286, 2016.
- [39] V. C. Pezoulas *et al.*, "Medical data quality assessment: On the development of an automated framework for medical data curation," *Comput. Biol. Med.*, vol. 107, pp. 270–283, 2019.
- [40] W. Gu, K. Foster, J. Shang, and L. Wei, "A game-predicting expert system using big data and machine learning," *Expert Syst. Appl.*, vol. 130, pp. 293–305, 2019.
- [41] B. Ait Hammou, A. Ait Lahcen, and S. Mouline, "An Effective Distributed Predictive Model with Matrix Factorization and Random forest for Big Data Recommendation systems," *Expert Syst. Appl.*, 2019.
- [42] B. Weng, L. Lu, X. Wang, F. M. Megahed, and W. Martinez, "Predicting short-term stock prices using ensemble methods and online data sources," *Expert Syst. Appl.*, vol. 112, pp. 258–273, 2018.
- [43] P. Tahmasebi, F. Javadpour, and M. Sahimi, "Data mining and machine learning for identifying sweet spots in shale reservoirs," *Expert Syst. Appl.*, vol. 88, pp. 435–447, 2017.
- [44] Y. Kaneda and H. Mineno, "Sliding window-based support vector regression for predicting micrometeorological data," *Expert Syst. Appl.*, vol. 59, pp. 217–225, 2016.
- [45] D. U. Pfeiffer and K. B. Stevens, "Spatial and temporal epidemiological analysis in the Big Data era," *Prev. Vet. Med.*, vol. 122, no. 1, pp. 213–220, 2015.
- [46] E. Shortliffe, *Computer-based medical consultations: MYCIN*, vol. 2. Elsevier, 2012.
- [47] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data," *Knowledge-Based Syst.*, vol. 86, pp. 33–45, 2015.
- [48] S. I. Park, G. Lee, H. M. Kim, N. Hur, S. Kwon, and J. Kim, "ADT-Based UHDTV Transmission for the Existing ATSC Terrestrial DTV Broadcasting," *IEEE Trans. Broadcast.*, vol. 61, no. 1, pp. 105–110, 2015.