

Rekayasa Perangkat Lunak pada Data Mining Penyakit: Suatu Tinjauan Literatur Sistematis

Ichsan Taufik¹, Cepy Slamet², Agung Wahana³, Barzan Faizin⁴

¹Teknik Informatika, UIN Sunan Gunung Djati Bandung, ichsan@uinsgd.ac.id

² Teknik Informatika, UIN Sunan Gunung Djati Bandung, wahana.agung@uinsgd.ac.id

³ Teknik Informatika, UIN Sunan Gunung Djati Bandung, cepy_lucky@uinsgd.ac.id

⁴ Ilmu Komunikasi, UIN Sunan Gunung Djati Bandung, barzanfaizin@uinsgd.ac.id

Abstrak

Saat ini sedang terjadi wabah penyakit virus corona yang dideteksi berasal dari Wuhan China dan telah menyebar ke seluruh dunia, telah banyak database tentang penyakit Covid-19 yang bisa digunakan untuk melakukan data mining penyakit. Pada artikel ini melakukan tinjauan literatur secara sistematis untuk memberikan gambaran tentang data mining pada penyakit. Artikel yang dipublikasikan pada tahun 2015 sampai dengan 2020 dari tiga database terpilih (IEEE, ACM, Scencedirect). Artikel yang ada dianalisis, dan area yang diteliti tentang rekayasa perangkat lunak untuk data mining penyakit. Metode yang digunakan dalam penelitian ini adalah tinjauan literatur sistematis. Berdasarkan temuan kajian literatur data mining penyakit terdapat banyak ragam penyakit yang diteliti, penyakit yang banyak diteliti yaitu tentang penyakit jantung, serta metode data mining yang banyak digunakan adalah Naive Bayes sedangkan akurasi metode data mining yang paling tinggi yaitu Artificial Neural Networks yang diterapkan pada penyakit Talasemia yaitu sebesar 99,73%, sedangkan negara yang paling banyak melakukan penelitian data mining penyakit yaitu India dan Turki.

Coronavirus disease 2019 (COVID-19) that hits all parts of the world today is allegedly originated from Wuhan China. Thus, there were made many databases about Covid-19 that can conduct disease data mining. This article systematically reveals literature reviews that provide an overview of disease data mining taken from the articles published in three selected databases (IEEE, ACM, Scencedirect) from 2015 to 2020. The scope analysis of this study is focused on software engineering for disease data mining. And the method applied in this study is a systematic literature review. The research findings revealed that heart disease was widely studied among many other types of diseases, Naive Bayes was one of data mining methods mostly applied, and Artificial Neural Networks was seen to have the highest accuracy of data mining methods used to Thalassaemia with 99.73% accuracy. Meanwhile, the countries that extensively conducted data mining researches are India and Turkey.

Kata kunci: data mining, penyakit; tinjauan literatur sistematis;

1 Pendahuluan

Saat ini sedang terjadi wabah virus corona yang dideteksi berasal dari Wuhan China yang mulai menyebar pada akhir bulan 2019 (Bai et al., 2020) yang selanjutnya disebut sebagai Coronavirus Disease 19 yang disingkat Covid-19 dan mulai mewabah ke seluruh dunia dimulai pada bulan Maret 2020, termasuk ke negara Indonesia. Di seluruh dunia Covid-19 telah menginfeksi sekitar 3.349.786 orang dan menyerang sekitar 215 negara di dunia berdasarkan sumber dari World Health Organization (WHO) (WHO, 2020). Sekarang banyak ilmuwan dari seluruh negara mulai meneliti membuat vaksin untuk menyembuhkan Covid-19 serta sudah mulai banyak database tentang Covid-19 salah satunya pada situs www.kaggle.com yang menyediakan data Open Research Dataset (CORD-19) untuk penelitian, selain itu ada juga database dataset lain yang bisa digunakan untuk penelitian data mining yaitu <https://github.com/datasets/covid-19>, <https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data>, <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>.

Data mining adalah “proses pencarian semi otomatis informasi dalam database besar yang menggunakan beberapa teknik dalam pencariannya seperti teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk menggali informasi yang potensial dan berguna” (Turban, Rainer, & Potter, 2005) dalam (Meilani & Asadulloh, 2015). Data mining saat ini banyak digunakan untuk menggali informasi data penyakit berdasarkan data-data yang terdapat dalam database medis yang ada di rumah sakit maupun dataset yang disediakan secara online. Kajian ini bertujuan untuk meninjau artikel-artikel secara sistematis yang terdapat dalam prosiding maupun dalam jurnal-jurnal yang terdapat dalam beberapa database jurnal/prosiding seperti: IEEE, ACM dan Scimedirect..

Penelitian yang akan dilakukan menggunakan metode tinjauan literatur secara sistematis (Kitchenham et al., 2009) (Hidayat, Azzery, & Mahardiko, 2019) yang merupakan cara untuk mengevaluasi dan menginterpretasikan semua penelitian yang tersedia berdasarkan pertanyaan penelitian khusus, area topik atau fenomena. Dengan menggunakan metode ini diharapkan bisa menggali informasi penting secara sistematis dalam suatu database artikel yang tersedia secara online. Diharapkan kajian literatur ini bisa menjadi rujukan awal untuk melakukan penelitian lebih lanjut mengenai data mining Covid-19, khususnya di Indonesia.

Kajian ini bertujuan untuk memberikan tinjauan tentang rekayasa perangkat lunak pada data mining penyakit yang banyak tersedia di database jurnal serta untuk menggali pertanyaan, topik atau fenomena yang terdapat dalam data mining penyakit untuk memperkaya pengetahuan pada mata kuliah rekayasa perangkat lunak atau data mining. Kajian literatur ini diharapkan bisa memberikan tinjauan tentang metode-metode dan penyakit yang sering muncul pada artikel-artikel yang terdapat dalam database jurnal mulai tahun 2015 sampai dengan 2020.

2 Metodologi

Metode yang digunakan pada penelitian ini adalah menggunakan tinjauan literatur secara sistematis untuk menggali informasi yang terdapat pada artikel dalam tiga database jurnal, yaitu: IEEE, ACM, Scimedirect.

Metode tinjauan literatur secara sistematis merupakan metode yang banyak digunakan untuk memperoleh wawasan dalam meningkatkan pengetahuan pada suatu area spesifik (Gokce & Dalveren, 2019). Oleh karena itu, metode ini digunakan untuk lebih memahami literatur tentang data mining penyakit. Pertanyaan penelitian dan kata kunci didefinisikan untuk mengidentifikasi penelitian serupa dengan memeriksa literatur yang ada kemudian menghasilkan temuan dari dokumen yang dicari. Tujuan utama dari penelitian ini adalah untuk lebih memahami area rekayasa perangkat lunak dalam penggunaan metode data mining serta penyakit apa saja yang sering ditemukan dalam penelitian. Lingkup dari artikel ini terdiri dari metoda data mining yang sering digunakan, kinerja metode data mining yang digunakan, penyakit yang sering dijadikan bahan penelitian. Artikel yang dikumpulkan berupa prosiding dan jurnal internasional yang dipublikasikan pada basis data IEEE, ACM dan ScienceDirect mulai tahun 2015 sampai dengan 2020. Kita mencari artikel menggunakan kata kunci “data mining” dan “disease” pada judul dari artikel. Kami mengecualikan bab buku, majalah, tinjauan artikel, artikel yang tidak berbahasa Inggris. Hal pertama yang dilakukan adalah setiap artikel yang dicari dilihat dari judul dan berdasarkan pada kriteria inklusi (Tabel 1) dan eksklusi (Tabel 2)

Tabel 1 Kriteria Inklusi

No.	Kriteria Inklusi
1.	Meneliti di bidang data mining
2.	Meneliti di bidang metode yang digunakan pada data mining
3.	Meneliti di area penyakit yang sering dilakukan penelitian

Tabel 2 Kriteria Eksklusi

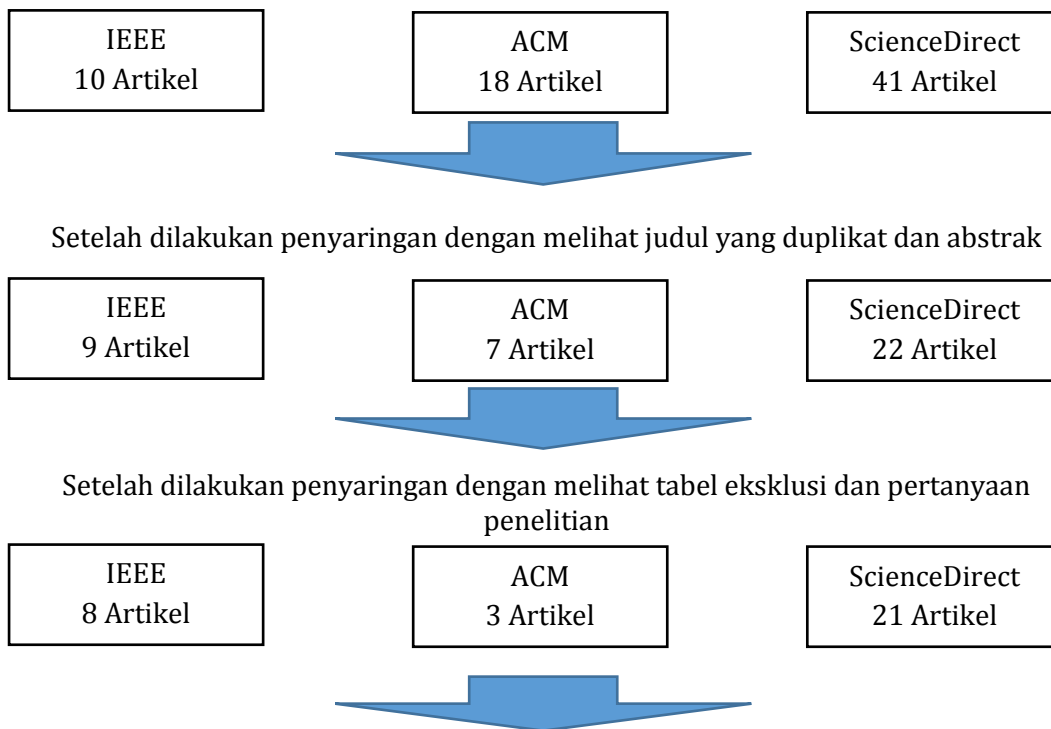
No.	Kriteria Eksklusi
1.	Meneliti yang bukan pada bidang data mining penyakit
2.	Meneliti yang tidak ditulis dalam bahasa Inggris
3.	Artikel yang tidak bisa dilihat secara full text
4.	Artikel yang berupa bab buku, majalah, tinjauan

Hal yang kedua dilakukan adalah memeriksa dengan membaca artikel *full-text* secara kritis. Pada awalnya terdapat 69 artikel. Setelah melakukan penilaian awal dan menghilangkan yang duplikat, artikel berkurang menjadi 38. Kami tidak memasukan artikel yang tidak menggambarkan pertanyaan penelitian yang bisa dilihat pada Tabel 3 dan pada akhirnya didapatkan sejumlah 32 artikel untuk ditinjau.

Tabel 3 Pertanyaan Penelitian

No.	Pertanyaan Penelitian
1.	Penyakit apa saja yang sering diterapkan pada kajian data mining penyakit?
2.	Metode data mining apa yang sering digunakan?
3.	Berapa akurasi kinerja metode data mining yang diterapkan?
4.	Negara mana saja yang sering melakukan kajian data mining penyakit

Sampel yang berhasil didapatkan dari hasil pencarian dan mendapatkan *full-text* bisa dilihat pada Gambar 1 :

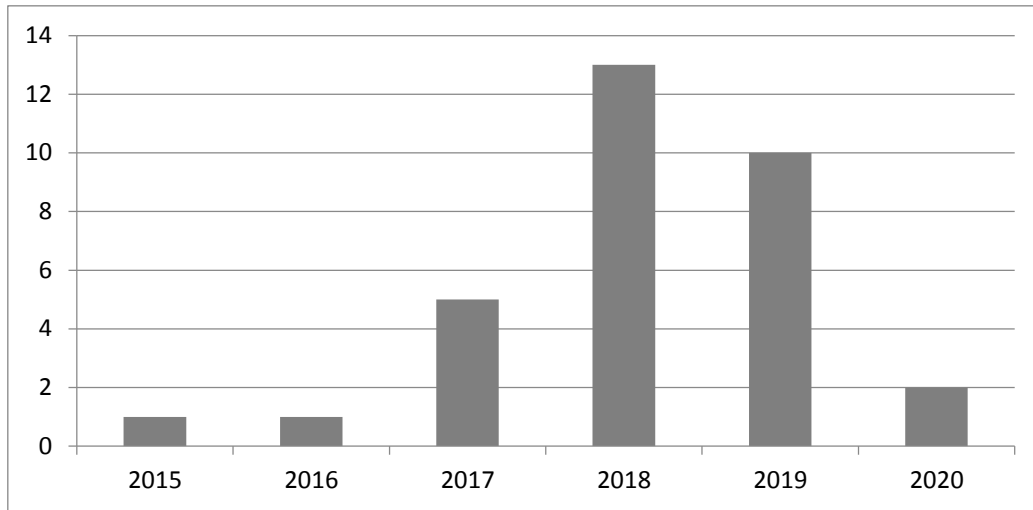


Terdapat 32 artikel yang bisa diproses untuk dianalisis dan diresumekan

Gambar 1 Sampel yang didapatkan

3 Hasil dan Pembahasan

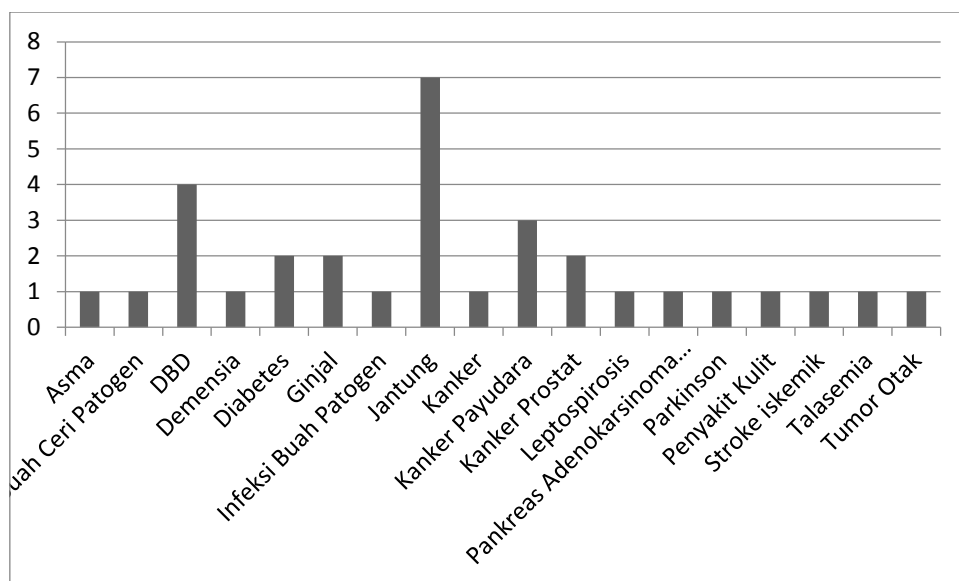
Sebagaimana yang kita lihat pada Gambar 2 penelitian tentang data mining pada penyakit banyak dilakukan pada tahun 2018, tetapi tidak menutup kemungkinan pada tahun 2020 akan terjadi peningkatan penelitian tentang penyakit dikarenakan pada tahun 2020 telah terjadi pandemi global Covid-19. Pada saat artikel ini ditulis kami belum mendapatkan banyak publikasi tentang data mining penyakit Covid-19 khususnya pada database jurnal (IEEE, ACM, ScienceDirect) disebabkan mungkin masih dalam proses publikasi ke dalam database jurnal tersebut.



Gambar 2 Jumlah Publikasi Berdasarkan Tahun

3.1 Penyakit apa saja yang sering diterapkan pada kajian data mining penyakit?

Berdasarkan hasil temuan kajian literatur pada Gambar 3 penyakit yang banyak diteliti oleh para peneliti data mining tentang penyakit adalah penyakit jantung disebabkan penyakit jantung merupakan penyebab kematian pertama tertinggi di dunia sekitar 15,2 juta kematian pada tahun 2016 (WHO, 2018) sehingga banyak tersedia database penyakit jantung di situs tentang dataset penyakit, selanjutnya disusul oleh penyakit demam berdarah yang menempati urutan kedua, kemudian kanker payudara juga cukup banyak diteliti dalam data mining penyakit disusul oleh diabetes dan ginjal.



Gambar 3 Daftar Penyakit yang diteliti

3.2 Metode data mining apa yang sering digunakan?

Menurut hasil kajian terhadap 32 artikel tentang data mining penyakit yang bisa dilihat pada Tabel 4, metode yang paling banyak digunakan adalah Naive Bayes dan disusul oleh Support Vector Machine (SVM), k-Nearest Neighbour (k-NN) (Syahid, Nursantika, Informatika, & Sains, 2016) (Maylawati, Andrian, Sunarto, Wildanuddin, & Wahana, 2019), Random Forest, dan selanjutnya ada Decision Tree (Setiawati & Taufik, 2016) dan pengembangannya yaitu C4.5 dan J48 yang juga merupakan algoritma C4.5 yang diimplementasikan di WEKA serta cukup banyak juga yang menggunakan metode Artificial Neural Network (ANN) (Ramdhani, Maylawati, Zulfikar, & Taufik, 2017), K-Means (Slamet, Rahman, Ramdhani, & Darmalaksana, 2016).

Tabel 4 Daftar Metode Data Mining

Metode	Jumlah
A Hybrid Feature Extraction	1
Artificial Neural Network (ANN)	3
Bayesian Classifier	1
Bernoulli Naive Bayesian	1
C4.5	4
C5.0	1
C5.0 Boosted	1
Classification and Regression Tree (CART)	1
Compact Classification Trees	1
Convolutional Neural Network (CNN)	1
Decision Tree	4
Extra Tree Classifier	1
Gaussian Naive Bayes	1
Gradient Boosting	1
ID3	1
J48	4
JRip	1
K-Means	2
K-Nearest Neighbour	5
Least Angle Regression (LARS)	1
Linear Discriminant Analysis	2
Logistic Regression	4
Modified Apriori	1
Multilayer Perceptron (MLP)	1
Naive Bayes	7
Neural Networks	2
Passive Aggressive Classifier	1
Penalized Logistic Regression	1
Principal Component Analysis	1
Pseudo Linear Discriminant Analysis	1
Quadratic Discriminant Analysis	1
Radial Basis Function (RBF)	1
Radiomics-Driven Conditional Random Field (RD-CRF)	1
Radius Neighbour Classifier	1
Random Forest	5
Stochastic Gradient Boosting	1
Support Vector Machine (SVM)	6
Vote	1
Weighted Classification Based on Association	1

3.3 Berapa akurasi kinerja metode data mining yang diterapkan?

Pada Tabel 5 bisa dilihat daftar akurasi metode data mining penyakit yang digunakan memiliki beragam persentase akurasi setiap metode yang berbeda-beda tergantung kasus/penyakit yang ditelitinya. Metode yang sama bisa memiliki nilai akurasi yang berbeda terhadap penerapan kasusnya. Metode yang paling tinggi nilai akurasinya adalah ANN yang diterapkan pada penyakit Talasemia yaitu sebesar 99,73% disusul oleh metode Decision Tree yaitu 99,69% pada penyakit yang sama, setelah itu nilai akurasi yang merupakan ketiga terbesar yaitu metode Radius Neighbour Classifiers sebesar 99,68% yang diterapkan pada penyakit kulit. Sedangkan metode yang memiliki akurasi dengan nilai terkecil yaitu metode SVM dengan nilai 49,22% yang diterapkan pada penyakit Demam Berdarah di Thailand diikuti oleh metode Multilayer Perceptron dengan nilai 51,5% yang diterapkan pada penyakit ginjal dan yang terkecil ketiga yaitu metode ANN dengan nilai 59,16% yang diterapkan pada penyakit yang sama dengan metode SVM dengan nilai yang terkecil.

Tabel 5 Akurasi Metode Data Mining

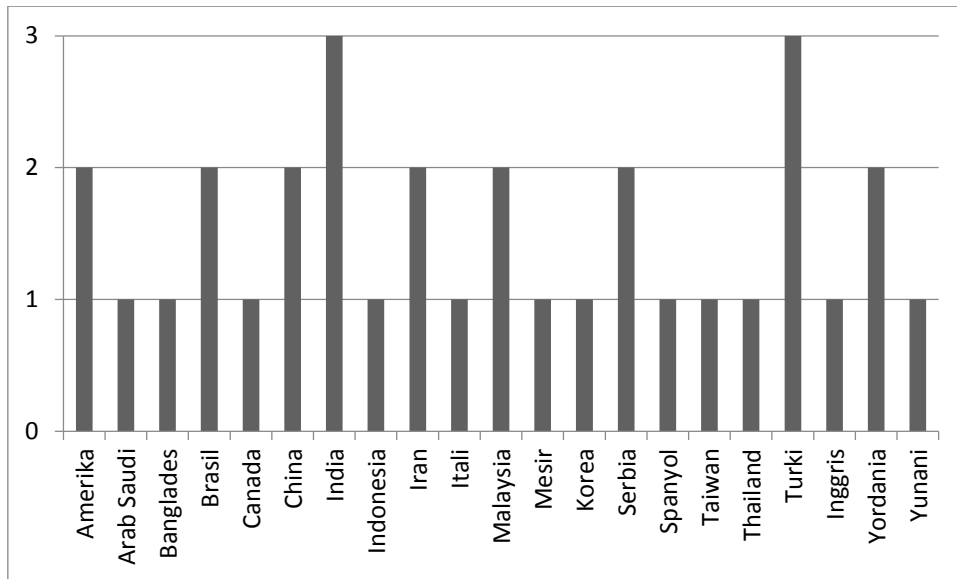
No	Judul	Metode yang digunakan	Akurasi Metode
1	Prostate cancer detection via a quantitative radiomics-driven conditional random field framework (Chung, Member, & Khalvati, 2015)	Radiomics-Driven Conditional Random Field (RD-CRF)	92%
2	Different medical data mining approaches based prediction of ischemic stroke (Arslan, Colak, & Sarihan, 2016)	Support Vector Machines, Stochastic Gradient Boosting, Penalized Logistic Regression	SVM=97,8%, SGB=97,37%, PLR=89,47%
3	Efficient Heart Disease Prediction System using K-Nearest Neighbor Classification Technique (Khateeb & Usman, 2017)	K-Nearest Neighbor	80%
4	A Simple Acute Myocardial Infarction (Heart Attack) Prediction System Using Clinical Data and Data Mining Techniques (Ahmed, 2017)	C4.5 dan Random Forest	96%
5	hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm (Tayefi et al., 2017)	Decision Tree	94%
6	Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data (Varlamis et al., 2017)	C4.5, Random Forests, Logistic Regression, SVM, Naive Bayes	C4.5=80,31%, RF=80,69%, LR=82,95%, SVM=81,08%, NB=78,66%
7	Prediction of leptospirosis cases using classification algorithms (Ruy, Nery, Claro, & Lindow, 2017)	JRip dan J48	JRip = 84%; J48 =82%
8	Early cherry fruit pathogen disease detection based on data mining prediction (Ilic, Ilic, Jovic, & Panic, 2018)	Linear Discriminant Analysis, Quadratic Discriminant Analysis, Pseudo Linear Discriminant Analysis, Compact Classification Trees	LDA=80%, QDA= 87%, PLDA=82%, CCT=93,6%
9	Open Data Mining for Taiwan's Dengue Epidemic (C. H. Wu, Kao, Shih, & Kan, 2018)	ID3	96% not google trend, 94% google trend
10	Prediction of Dengue Disease Through Data Mining by using modified Apriori Algorithm (Hannan, 2018)	Modified Apriori	75%
11	Forecasting Dengue Fever using Classification Techniques in Data Mining (Anitha & Wise, 2018)	J48	86,13%

12	Comparision using Data Mining Algorithm Techniques for Predicting of Dengue fever Data in Northeastern of Thailand (Jongmuenwai, Lowanichchai, & Jabjone, 2018)	ANN; Naïve Bayes, DT, SVM	ANN= 59,16% Naive Bayes = 70,6% DT = 70,08% SVM = 49,22%
13	A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia (Moreira & Namen, 2018)	J48	99,34%
14	Type 2 diabetes mellitus prediction model based on data mining (H. Wu, Yang, Huang, He, & Wang, 2018)	K-means dan Logistic Regression	93,50%
15	Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History (Akben, 2018)	K-means, k-NN, SVM, Naive Bayes	K-means=90,25%, k-NN=96%, SVM=95,50%, Naive Bayes=95,50%
16	Data mining based tool for early prediction of possible fruit pathogen infection (Predic et al., 2018)	J48	89%
17	Soft set and Fuzzy Rules enabled SVM Approach for Heart Attack Risk Classification among Adolescents (Malode, Bhargavi, Gunasheela, Kavana, & Sushmitha, 2018)	Soft set and Fuzzy Rules enabled SVM	90%
18	WCBA: Weighted classification based on association rules algorithm for breast cancer disease (Alwidian, Hammo, & Obeid, 2018)	Weighted Classification Based on Association	WCBA diagnosis Supp=0,1, Conf. 0,5;Supp=0,2, Conf.=0,5; Supp.=0,3, Conf.=0,5=97,4%; 97,4%; 96,8%
19	A data mining framework based on boundary-points for gene selection from DNA-microarrays: Pancreatic Ductal Adenocarcinoma as a case study (Ramos, Castellanos-Garzón, de Paz, & Corchado, 2018)	k-Nearest Neighbour	91,67%
20	Identifying β -thalassemia carriers using a data mining approach: The case of the Gaza Strip, Palestine (AlAgha, Faris, Hammo, & Al-Zoubi, 2018)	k-Nearest Neighbour, Artificial Neural Networks, Naive Bayes, Decision Tree	k-NN=99,59%, ANN=99,73%, NB=99,28%, DT=99,69%
21	A data-mining algorithm to assess key factors in asthma diagnosis (Mozaffarinya et al., 2019)	Decision Tree	98,68%
22	Analysis and Prediction of Diabetes Complication Disease using Data Mining Algorithm (Fiarni, Sipayung, & Maemunah, 2019)	Naive Bayes dan C4.5	68%
23	Prediction of kidney disease stages using data mining algorithms (Rady & Anwar, 2019)	Probabalistic Neural Network (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM), Radial Basis Function (RBF)	PNN = 96,7 %, SVM=60,7%, RBF=87%, MLP=51,5%
24	The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining (Emre, Erol, Ayhan, Özkan, & Erol, 2019)	Naive Bayes, CART, C4.5, C5.0, C5.0 boosted, Random Forest	NB=74%, CART=87%, C4.5=77%, C5.0=77%, C5.0 b=85%, RF=79%
25	Heart Disease Prediction Using Data Mining Techniques (C. M. Wu, n.d.)	Principal Component Analysis (PCA)	69%

26	Identification of significant features and data mining techniques in predicting heart disease (Amin, Chiam, & Varathan, 2019)	Support Vector Machines, Vote, Naive Bayes, Logistic Regression, Neural Network, k-NN, Decision Tree	SVM=86,87%, Vote=86,20%, NB=85,86%, LR=85,86%, NN=84,85%, k- NN=82,49%, DT=82,49%
27	A Deep Learning Approach for Targeted Contrast-Enhanced Ultrasound Based Prostate Cancer Detection (Feng et al., 2018)	Convolutional Neural Network	90%
28	Using gait analysis' parameters to classify Parkinsonism: A data mining approach (Ricciardi et al., 2019)	Random Forests, Gradient Boosted Trees	RF=86,4%, GBT=84%
29	Comparison of skin disease prediction by feature selection using ensemble data mining techniques (Verma, Pal, & Kumar, 2019)	Passive Aggressive Classifier, Linear Discriminant Analysis, Radius Neighbour Classifiers, Bernoulli Naive Bayesian, Gaussian Naive Bayes, Extra Tree Classifier	PAC=97,38%, LDA=98,98%, RNC=99,68%, BNB=95,36%, NB=98,66%, ETC=97,65
30	A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine for Brain Tumor Classification (Gumaei, Hassan, Member, & Hassan, 2019)	A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine	92% s.d. 95%
31	A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival (Simsek, Kursuncu, Kibis, AnisAbdellatif, & Dag, 2020)	Artificial Neural Network, Logistic Regression	ANN=74,2%, LR=73,8%
32	Characterizing basal-like triple negative breast cancer using gene expression analysis: A data mining approach (Hassan Zadeh, Alsabi, Ramirez-Vick, & Nosoudi, 2020)	Decision Tree, LARS, Neural Networks, Logistics Regression, SVM, Gradient Boosting, Random Forest, Bayesian Classifier	DT= 91,38%, LARS=97,70%, NN=98,28%, LR=97,13%, SVM=97,13%, GB=98,28%, RF=97,70%, BC=97,70%

3.4 Negara mana saja yang sering melakukan kajian data mining penyakit?

Berdasarkan temuan kajian literatur pada Gambar 4 tentang negara yang sering melakukan kajian data mining yaitu India dan Turki disusul oleh negara Amerika, Brasil, China, Iran, Malaysia, Serbia dan Yordania. Hasil ini diambil berdasarkan asal negara dari penulis yang melakukan penelitian tentang data mining penyakit.



Gambar 4 Negara yang melakukan Penelitian Data Mining

4 Simpulan

Pada kajian literatur ini, sebanyak 32 artikel yang berkaitan dengan data mining penyakit telah dilakukan proses analisis menggunakan metode kajian literatur sistematis. Berdasarkan temuan pada kajian literatur ini penyakit yang banyak diteliti pada data mining penyakit yaitu penyakit jantung disebabkan menurut WHO, penyakit jantung merupakan penyebab pertama kematian tertinggi pada tahun 2016 sehingga banyak tersedia database dataset yang bisa digunakan untuk penelitian data mining. Metode yang sering digunakan pada kajian data mining penyakit adalah Naive Bayes karena metode ini merupakan yang populer dalam aplikasi pembelajaran mesin dan kesederhanaan metodenya dalam mengizinkan setiap atribut untuk berkontribusi pada keputusan akhir dan independen dari atribut-atribut lainnya (Xhemali, Hinde, & Stone, 2009). Temuan selanjutnya pada kajian ini adalah bahwa metode yang paling tinggi akurasi adalah Artificial Neural Networks yang diterapkan pada penyakit Talasemia yaitu sebesar 99,73% tetapi mendapatkan nilai akurasi yang kecil jika diterapkan pada penelitian penyakit demam berdarah yaitu 59,16% ini membuktikan bahwa metode data mining tidak mutlak akurasi lebih baik dibandingkan dengan metode yang lain tergantung dari atribut/kriteria yang diteliti serta jenis penyakit yang diteliti. Selanjutnya negara yang paling banyak melakukan penelitian tentang data mining penyakit yaitu India dan Turki tapi tidak menutup kemungkinan negara-negara lain seperti Amerika, Brasil dan lain-lain banyak melakukan penelitian dikarenakan publikasi yang didapatkan pada kajian ini masih terbatas tidak semua artikel jurnal/prosiding tidak menyediakan *full-text* serta terdapat jawaban atas pertanyaan penelitian. Untuk kajian literatur data mining penyakit selanjutnya diharapkan bisa ditambahkan database jurnal selain yang digunakan pada kajian ini dan bisa menambahkan pertanyaan penelitian tentang metode pengembangan perangkat lunak yang sering digunakan pada penelitian data mining penyakit serta bisa ditambahkan sumber dari literatur yang digunakan tidak sebatas jurnal/prosiding tetapi seperti bab buku, majalah, atau kajian artikel. Mudah-mudahan kajian literatur ini bisa menjadi referensi/acuan dasar untuk melakukan penelitian data mining tentang Covid-19 yang sudah mewabah ke lebih dari 215 negara di dunia.

Referensi

- Ahmed, F. (2017). A Simple Acute Myocardial Infarction (Heart Attack) Prediction System Using Clinical Data and Data Mining Techniques, 22–24.
- Akben, S. B. (2018). Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History. *Irbm*, 39(5), 353–358. <http://doi.org/10.1016/j.irbm.2018.09.004>
- AlAgha, A. S., Faris, H., Hammo, B. H., & Al-Zoubi, A. M. (2018). Identifying β -thalassemia carriers using a data mining approach: The case of the Gaza Strip, Palestine. *Artificial Intelligence in Medicine*, 88(July 2017), 70–83. <http://doi.org/10.1016/j.artmed.2018.04.009>
- Alwidian, J., Hammo, B. H., & Obeid, N. (2018). WCBA: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing Journal*, 62, 536–549. <http://doi.org/10.1016/j.asoc.2017.11.013>
- Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, 82–93. <http://doi.org/10.1016/j.tele.2018.11.007>
- Anitha, A., & Wise, D. C. J. W. (2018). Forecasting Dengue Fever using Classification Techniques in Data Mining. *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (Icssit), 398–401. <http://doi.org/10.1109/ICSSIT.2018.8748864>
- Arslan, A. K., Colak, C., & Sarihan, M. E. (2016). Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine*, 130, 87–92. <http://doi.org/10.1016/j.cmpb.2016.03.022>
- Bai, L., Yang, D., Wang, X., Tong, L., Zhu, X., Zhong, N., ... Tan, F. (2020). Clinical eHealth Chinese experts ' consensus on the Internet of Things-aided diagnosis and treatment of coronavirus disease 2019 (COVID-19), 3, 7–15. <http://doi.org/10.1016/j.ceh.2020.03.001>
- Chung, A. G., Member, S., & Khalvati, F. (2015). Prostate Cancer Detection via a Quantitative Radiomics-Driven Conditional Random Field Framework, 2531–2541.
- Emre, İ. E., Erol, N., Ayhan, Y. İ., Özkan, Y., & Erol, Ç. (2019). The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining. *International Journal of Medical Informatics*, 123(December 2018), 68–75. <http://doi.org/10.1016/j.ijmedinf.2018.12.009>
- Feng, Y., Yang, F., Member, X. Z., Guo, Y., Tang, F., Ren, F., ... Member, S. J. (2018). A Deep Learning Approach for Targeted Contrast-Enhanced Ultrasound Based Prostate Cancer Detection, (c), 1–8. <http://doi.org/10.1109/TCBB.2018.2835444>
- Fiarni, C., Sipayung, E. M., & Maemunah, S. (2019). Analysis and prediction of diabetes complication disease using data mining algorithm. *Procedia Computer Science*, 161, 449–457. <http://doi.org/10.1016/j.procs.2019.11.144>
- Gokce, G., & Dalveren, M. (2019). Software Engineering in Medical Informatics : A Systematic Literature Review, 1–6.
- Gumaei, A., Hassan, M. M., Member, S., & Hassan, R. (2019). A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine for Brain Tumor Classification. *IEEE Access*, PP(c), 1. <http://doi.org/10.1109/ACCESS.2019.2904145>
- Hannan, A. (2018). PREDICTION OF DENGUE DISEASE THROUGH DATA MINING BY USING MODIFIED APRIORI ALGORITHM.
- Hassan Zadeh, A., Alsabi, Q., Ramirez-Vick, J. E., & Nosoudi, N. (2020). Characterizing basal-like triple negative breast cancer using gene expression analysis: A data mining approach. *Expert Systems with Applications*, 148, 113253. <http://doi.org/10.1016/j.eswa.2020.113253>
- Hidayat, T., Azzery, Y., & Mahardiko, R. (2019). Load Balancing Network by using Round Robin Algorithm : A Systematic Literature Review, 4(2), 85–89. <http://doi.org/10.15575/join.v4i2.446>

- Ilic, M., Ilic, S., Jovic, S., & Panic, S. (2018). Early cherry fruit pathogen disease detection based on data mining prediction. *Computers and Electronics in Agriculture*, 150(May), 418–425. <http://doi.org/10.1016/j.compag.2018.05.008>
- Jongmuenwai, B., Lowanichchai, S., & Jabjone, S. (2018). Comparison using Data Mining Algorithm Techniques for Predicting of Dengue fever Data in Northeastern of Thailand *. *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 532–535.
- Khateeb, N., & Usman, M. (2017). Efficient Heart Disease Prediction System using K-Nearest Neighbor Classification Technique, 21–26.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. <http://doi.org/10.1016/j.infsof.2008.09.009>
- Malode, C. M. C., Bhargavi, K., Gunasheela, B. G., Kavana, G., & Sushmitha, R. (2018). Soft set and Fuzzy Rules enabled SVM Approach for Heart Attack Risk Classification among Adolescents. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–6.
- Maylawati, D. S., Andrian, R., Sunarto, S., Wildanuddin, M., & Wahana, A. (2019). Hybrid principal component analysis and K-nearest neighbour to detect the catfish disease Hybrid principal component analysis and K-nearest neighbour to detect the catfish disease. *4th Annual Applied Science and Engineering Conference*. <http://doi.org/10.1088/1742-6596/1402/7/077008>
- Meilani, B. D., & Asadulloh, M. (2015). DATA MINING UNTUK MENGGALI POLA MAHASISWA BARU (STUDI KASUS: INSTITUT TEKNOLOGI ADHI TAMA SURABAYA), 269–276.
- Moreira, L. B., & Namen, A. A. (2018). A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Computer Methods and Programs in Biomedicine*, 165, 139–149. <http://doi.org/10.1016/j.cmpb.2018.08.016>
- Mozaffarinya, M., Reza Shahriyari, A., Karim Bahadori, M., Ghazvini, A., Shamsadin Athari, S., & Vahedi, G. (2019). A data-mining algorithm to assess key factors in asthma diagnosis. *Revue Francaise d'Allergologie*, 59(7), 487–492. <http://doi.org/10.1016/j.reval.2019.01.013>
- Predic, B., Ilic, M., Spalevic, P., Trajkovic, S., Jovic, S., & Stanic, A. (2018). Data mining based tool for early prediction of possible fruit pathogen infection. *Computers and Electronics in Agriculture*, 154(August), 314–319. <http://doi.org/10.1016/j.compag.2018.09.023>
- Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked*, 15(December 2018), 100178. <http://doi.org/10.1016/j.imu.2019.100178>
- Ramdhani, M. A., Maylawati, D. S., Zulfikar, W. B., & Taufik, I. (2017). An Expert System for Predicting the Early Pregnancy with Disorders Using Artificial Neural Network. Bali.
- Ramos, J., Castellanos-Garzón, J. A., de Paz, J. F., & Corchado, J. M. (2018). A data mining framework based on boundary-points for gene selection from DNA-microarrays: Pancreatic Ductal Adenocarcinoma as a case study. *Engineering Applications of Artificial Intelligence*, 70(January), 92–108. <http://doi.org/10.1016/j.engappai.2018.01.007>
- Ricciardi, C., Amboni, M., De Santis, C., Improta, G., Volpe, G., Iuppariello, L., ... Cesarelli, M. (2019). Using gait analysis' parameters to classify Parkinsonism: A data mining approach. *Computer Methods and Programs in Biomedicine*, 180, 105033. <http://doi.org/10.1016/j.cmpb.2019.105033>
- Ruy, N., Nery, R., Claro, D. B., & Lindow, J. C. (2017). Prediction of leptospirosis cases using classification algorithms, 11, 93–99. <http://doi.org/10.1049/iet-sen.2016.0193>
- Setiawati, D., & Taufik, I. (2016). Jumadi, and WZ Budiawan, “Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile,” J.

Online Inform, 1(1), 24–27.

- Simsek, S., Kursuncu, U., Kibis, E., AnisAbdellatif, M., & Dag, A. (2020). A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival. *Expert Systems with Applications*, 139. <http://doi.org/10.1016/j.eswa.2019.112863>
- Slamet, C., Rahman, A., Ramdhani, M. A., & Darmalaksana, W. (2016). Clustering the Verses of the Holy Qur'an Using K-Means Algorithm. *Asian Journal of Information Technology*, 15(24), 5159–5162. <http://doi.org/10.3923/ajit.2016.5159.5162>
- Syahid, D., Nursantika, D., Informatika, J. T., & Sains, F. (2016). DAUN PHILODENDRON MENGGUNAKAN METODE K-NEAREST NEIGHBOOR (KNN) BERDASARKAN NILAI HUE , SATURATION , VALUE (HSV), 1(1), 20–23.
- Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily, H., ... Ghayour-Mobarhan, M. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer Methods and Programs in Biomedicine*, 141, 105–109. <http://doi.org/10.1016/j.cmpb.2017.02.001>
- Turban, E., Rainer, R. K., & Potter, R. E. (2005). *Introduction to Information Technology* (3rd ed.). USA: John Wiley & Sons.
- Varlamis, I., Apostolakis, I., Sifaki-Pistolla, D., Dey, N., Georgoulas, V., & Lionis, C. (2017). Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry: The case of the island of Crete, Greece. *Computer Methods and Programs in Biomedicine*, 145, 73–83. <http://doi.org/10.1016/j.cmpb.2017.04.011>
- Verma, A. K., Pal, S., & Kumar, S. (2019). Comparison of skin disease prediction by feature selection using ensemble data mining techniques. *Informatics in Medicine Unlocked*, 16(April), 100202. <http://doi.org/10.1016/j.imu.2019.100202>
- WHO. (2018). The top 10 causes of death. Retrieved May 3, 2020, from <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- WHO. (2020). Coronavirus disease (COVID-19) Pandemic. Retrieved May 3, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- Wu, C. H., Kao, S. C., Shih, C. H., & Kan, M. H. (2018). Open data mining for Taiwan's dengue epidemic. *Acta Tropica*, 183, 1–7. <http://doi.org/10.1016/j.actatropica.2018.03.017>
- Wu, C. M. (n.d.). Heart Disease Prediction Using Data Mining Techniques, 1–5.
- Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10(August 2017), 100–107. <http://doi.org/10.1016/j.imu.2017.12.006>
- Xhemali, D., Hinde, C. J., & Stone, R. G. (2009). Naïve Bayes vs . Decision Trees vs . Neural Networks in the Classification of Training Web Pages, 4(1), 16–23.

Biografi Penulis

	<p>Ichsan Taufik, is a lecturer in Department Informatics at UIN Sunan Gunung Djati Bandung. Ichsan has graduated from Informatics Departement at IAIN Sunan Gunung Djati. He received his M.T. Degree from Institut Teknologi Bandung. Research areas: Information System, Algorithm dan Programming.</p>
	<p>Cepcy Slamet, is a lecturer in Department Informatics at UIN Sunan Gunung Djati Bandung. Cepcy has graduated from Informatics Departement at STT Garut. He received his M.Kom. Degree from STMIK LIKMI Bandung. He currently a Ph.D. student of Asia E University, Malaysia. Research areas: Strategic Management Information System (SMIT).</p>
	<p>Agung Wahana, is a lecturer in Department Informatics at UIN Sunan Gunung Djati Bandung. Agung has graduated from Accounting Departement at STIE Pasundan. He received his M.T. Degree from Institut Teknologi Bandung. Research areas:Text Mining.</p>
	<p>Barzan Faizin, is a lecturer of Communication Science Department at UIN Sunan Gunung Djati Bandung. Barzan has graduated from English Education Department of UIN Bandung. He received his M.Pd. Degree from Indraprasta University Jakarta. He is currently a Ph.D. student of Indonesia University of Education, Bandung. Research areas: Media Discourse.</p>