

CRISP-DM: Metodologi Proyek *Data Science*

Afika Rianti^{1*}, Nuur Wachid Abdul Majid², Ahmad Fauzi³

^{1,2}*Pendidikan Sistem dan Teknologi Informasi, ³Sistem Telekomunikasi*

Kampus UPI di Purwakarta

^{1*}*afika@upi.edu, ²nuurwachid@upi.edu, ³ahmad.fauzi@upi.edu*

Abstrak— Penelitian ini dilatarbelakangi oleh perkembangan *data science* yang semakin populer. Untuk mendukung perkembangan tersebut, terutama di bidang penelitian, metodologi dibutuhkan supaya proyek yang dibuat dapat menjadi terstruktur dan terorganisir dengan baik. Penelitian ini bertujuan untuk mendiskusikan salah satu metodologi *data science* yang banyak digunakan yaitu *cross industry standard process for data mining (CRISP-DM)*. Jenis penelitian yang digunakan adalah *literature review* dengan teknik pengumpulan data dilakukan menggunakan studi literatur. Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metodologi *CRISP-DM* dapat digunakan dalam proyek bidang *data science* tepatnya dalam area *data mining, artificial intelligence, machine learning, deep learning, big data, data analysis, dan data analytics*. Metodologi ini terdiri dari enam tahapan dimana setiap tahapan tersebut memiliki fase yang berbeda dari satu dan lainnya. Untuk fase-fase tersebut, peneliti dapat mengikuti fase-fase yang telah umum digunakan dalam penelitian lain atau bisa menyesuaikan dengan proyek masing-masing. Kemudian sebagai pertimbangan dalam menggunakan metodologi *CRISP-DM*, peneliti dapat mempertimbangkannya berdasarkan kelebihan dan tantangan dari metodologi ini.

Kata kunci— *Literature review, data science, CRISP-DM.*

Abstract— This research is motivated by the development of data science which is increasingly popular. To support this development, especially in the field of research, methodologies are needed so the projects created can be well structured and organized. This research aims to discuss one of the most widely used data science methodologies, that's *cross industry standard process for data mining (CRISP-DM)*. This research used a type of literature review research with data collection technique used literature study. Based on the research that has been conducted, it can be concluded that *CRISP-DM* methodology can be used in the field of data science projects precisely in the areas of data mining, artificial intelligence, machine learning, deep learning, big data, data analysis, and data analytics. This methodology has six stages that each of these stages has different phases from one to another. For those phases, researchers can follow phases that have been commonly used in other studies or can adjust according to their respective projects. Then as a consideration in using *CRISP-DM* methodology, it can be considered based on the benefits and challenges of it.

Keywords— *Literature review, data science, CRISP-DM.*

I. PENDAHULUAN

Data merupakan salah satu bidang yang sangat populer saat ini. Selain itu, bidang data juga merupakan salah satu bukti perkembangan dari teknologi informasi. Hal tersebut menyebabkan bidang ini banyak dicari di industri. Berdasarkan survei dari *World Economic Forum*, tiga pekerjaan yang permintaannya meningkat di industri adalah: *data analyst and data scientist; artificial intelligence and machine learning specialist; serta big data specialist* [1]. Dapat dilihat bahwa ketiga pekerjaan tersebut masuk dalam bidang data, tepatnya bidang *data science*. Selain dari survei tersebut, terdapat juga survei lain seperti dari Coursera yang menyatakan bahwa *data science* merupakan kursus *online* yang paling banyak diminati di tahun 2022 [2]. Adapun kursus *data science* yang paling populer adalah: *Foundations - Data, Data,*

Everywhere; Machine Learning; serta Ask Questions to Make Data-Driven Decisions. Coursera merupakan satu dari tiga kursus terbaik di dunia yang telah digunakan oleh banyak pengguna dari berbagai negara [3]. Dari pembahasan di atas, dapat disimpulkan bahwa *data science* penting untuk dipelajari.

Data science merupakan bidang yang berhubungan dengan matematika, statistika, *signal processing, computer science* dan pemrograman, *database, model data, machine learning, predictive analytics, visualisasi* dan lain sebagainya [4]. Banyaknya ilmu yang berhubungan dengan *data science* membuat bidang tersebut perlu untuk dipelajari dengan baik. Dalam implementasinya, sebuah metodologi dibutuhkan supaya proyek dapat dibuat dengan lebih terstruktur dan terorganisir. Meskipun demikian, masih ada beberapa

proyek *data science* yang tidak melibatkan metodologi dalam penelitiannya. Hal ini akan berpengaruh terutama dalam penelitian akademik karena proses di dalamnya tidak lengkap. Oleh karena itu, penelitian ini akan berfokus pada salah satu metodologi *data science* yaitu *cross industry standard process for data mining (CRISP-DM)*. Berdasarkan survei, metodologi ini dinyatakan sebagai metodologi yang paling banyak digunakan dalam proyek *data science* dengan persentase sebesar 49% yang kemudian diikuti oleh metodologi *Scrum* dan *Kanban* [5].

CRISP-DM merupakan sebuah metodologi yang menyediakan pendekatan umum untuk membentuk dan merencanakan proyek *data mining* [6]. Metodologi ini dirilis pertama kali sekitar 20 tahun yang lalu atau sekitar tahun 2000-an [7]. Hingga saat ini, metodologi ini menjadi metodologi yang paling populer di bidang *data science*, lebih tepatnya di area *data mining*. Metodologi ini dapat digunakan baik sebagai metodologi proyek maupun sebagai model proses [7]. Sebagai model proses biasanya digunakan ketika praktek mengolah data atau yang bersifat lebih teknis. Sedangkan sebagai metodologi proyek adalah ketika digunakan dalam hal yang lebih formal seperti penelitian dan melibatkan setiap fase di tahapannya. Dalam penelitian ini, peneliti berfokus dalam memperoleh informasi lebih dalam mengenai *CRISP-DM* sebagai metodologi proyek *data science*.

II. METODOLOGI PENELITIAN

A. Desain Penelitian

Jenis penelitian yang digunakan adalah *literature review*. *Literature review* adalah jenis penelitian yang menguji atau meninjau informasi, ide, ataupun penemuan yang terkandung dalam suatu diskusi dengan kritis [8]. Dalam penelitian ini, informasi tentang metodologi *CRISP-DM* akan dibahas sesuai dengan daftar pertanyaan penelitian. Untuk mengumpulkan data penelitian, teknik yang digunakan adalah studi literatur.

B. Pertanyaan dan Tujuan Penelitian

Penelitian ini diharapkan dapat menjawab pertanyaan penelitian dengan data yang mendukung. Daftar pertanyaan penelitian serta tujuan dari masing-masing pertanyaan ditampilkan pada Tabel 1.

Tabel 1. Pertanyaan Penelitian

| No | Pertanyaan | Tujuan |
|----|---|---|
| 1 | Area bidang <i>data science</i> apa saja yang dapat menggunakan <i>CRISP-DM</i> ? | Mengetahui area bidang <i>data science</i> yang dapat menggunakan <i>CRISP-DM</i> |
| 2 | Bagaimana fase dalam setiap tahapan <i>CRISP-DM</i> ? | Mengidentifikasi fase dalam setiap tahapan <i>CRISP-DM</i> |
| 3 | Apa kelebihan dan tantangan dalam menggunakan <i>CRISP-DM</i> ? | Menganalisis kelebihan dan tantangan dalam menggunakan <i>CRISP-DM</i> |

Berdasarkan daftar pertanyaan pada Tabel 1, diketahui bahwa jumlah pertanyaan keseluruhan ada tiga. Ketiga pertanyaan tersebut akan menjadi pokok pembahasan dalam penelitian ini.

C. Pengumpulan Data dan Seleksi

Data yang dikumpulkan diambil dari semua publikasi penelitian yang membahas tentang *CRISP-DM*. Diskusi dalam publikasi tersebut dapat berupa teori, studi literatur, hasil penelitian, dan lain sebagainya. Sumber rujukan yang digunakan berasal dari artikel jurnal, prosiding, buku, skripsi dan sumber lainnya yang dapat digunakan untuk menjawab pertanyaan penelitian.

D. Kriteria Publikasi

Kriteria publikasi yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

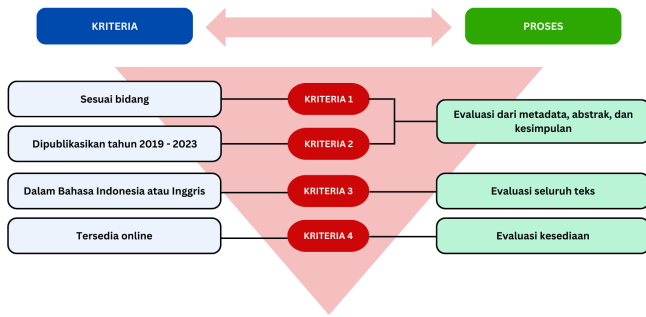
Tabel 2. Kriteria Publikasi

| No | Kriteria |
|----|--|
| 1 | Publikasi sesuai dengan bidang |
| 2 | Publikasi diutamakan diambil dari 5 tahun terakhir |
| 3 | Publikasi dalam Bahasa Indonesia atau Bahasa Inggris |
| 4 | Publikasi dapat diakses secara <i>online</i> |

Berdasarkan kriteria publikasi pada Tabel 2 di nomor 2 tentang tahun, publikasi diutamakan dari lima tahun terakhir. Meskipun demikian, ini tidak berarti bahwa publikasi tidak boleh diambil dari luar rentang waktu tersebut. Publikasi dapat diambil dari luar rentang waktu tersebut apabila benar-benar dibutuhkan.

E. Screening

Tahap *screening* dalam pemilihan publikasi ditampilkan pada Gambar 1.



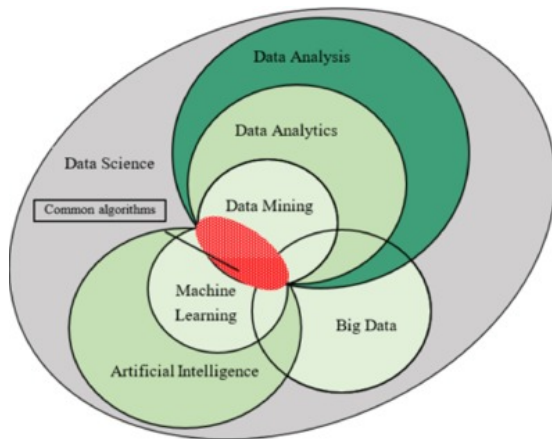
Gambar 1. Tahap *screening*

Berdasarkan informasi pada Gambar 1, *screening* dilakukan dengan melakukan evaluasi terhadap kriteria publikasi satu persatu.

III. HASIL DAN PEMBAHASAN

A. Area Bidang Data Science

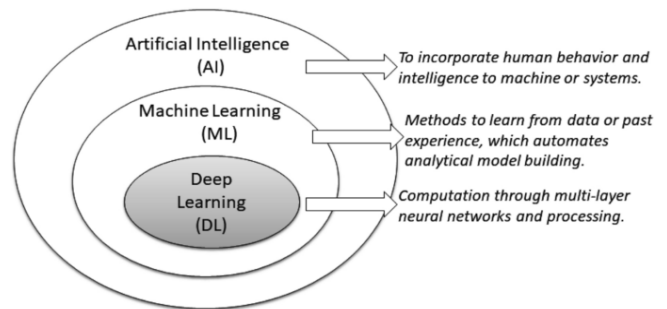
Seperi yang telah dijelaskan sebelumnya di bagian pendahuluan bahwa *CRISP-DM* digunakan untuk proyek *data science*, maka area yang dapat menggunakan metodologi *CRISP-DM* juga merupakan area dalam bidang *data science*. Area yang masuk pada bidang *data science* ditampilkan pada Gambar 2.



Gambar 2. Diagram venn *data science* [9]

Gambar 2 menunjukkan area apa saja yang termasuk dalam bidang *data science*. Terlihat bahwa tiga area utama bidang *data science* adalah *artificial intelligence*, *data analysis*, dan *big data*. Kemudian area-area tersebut

juga diikuti dengan sub area masing-masing. Dimulai dari area *artificial intelligence* yang di dalamnya terdapat sub area *machine learning*. Kemudian dari area *data analysis* yang didalamnya terdapat sub area *data analytics* dan *data mining*. Berdasarkan penjelasan-penjelasan tersebut, dapat dikatakan bahwa area dari bidang *data science* meliputi *artificial intelligence*, *machine learning*, *big data*, *data analysis*, *data analytics*, dan *data mining*. Oleh karena itu, proyek di area bidang *data science* tersebut dapat juga menggunakan *CRISP-DM* baik sebagai metodologi penelitian maupun sebagai model proses. Selain area-area tersebut, ini juga memungkinkan bahwa *CRISP-DM* digunakan dalam area yang lebih spesifik seperti *deep learning*. Hal ini dapat terjadi karena *deep learning* merupakan bagian dari *artificial intelligence* dan *machine learning*. Hubungan dari ketiganya ditampilkan pada Gambar 3.



Gambar 3. Diagram venn *artificial intelligence* [10]

Dari Gambar 3, terlihat bahwa *deep learning* merupakan bagian dari *artificial intelligence* dan *machine learning* dimana area ini lebih kompleks karena ini menggunakan *multi-layer neural networks* dan *processing* [10]. Sebagaimana *deep learning* merupakan bagian dari *artificial intelligence* dan *machine learning*, maka *deep learning* juga merupakan bagian dari *data science* sehingga dapat juga menggunakan metodologi *CRISP-DM*. Berdasarkan pembahasan tersebut, dapat disimpulkan bahwa area dari bidang *data science* yang dapat menggunakan metodologi *CRISP-DM* adalah *artificial intelligence*, *machine learning*, *deep learning*, *big data*, *data analysis*, *data analytics*, dan *data mining*. Contoh dari penelitian yang menggunakan metodologi *CRISP-DM* dalam bidang *data science* ditampilkan pada Tabel 3.

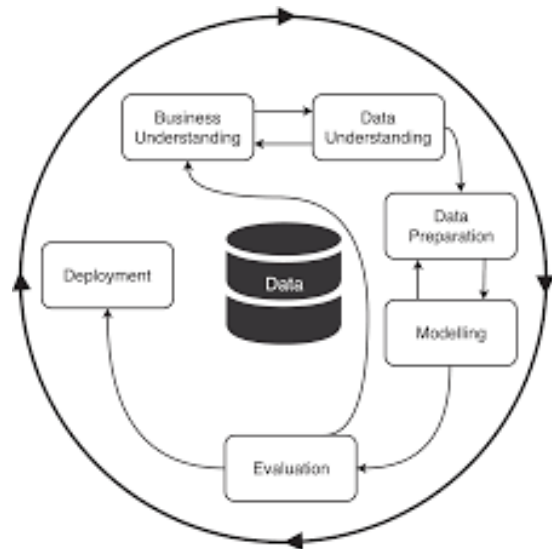
Tabel 3. Penelitian dalam Bidang *Data Science* yang Menggunakan *CRISP-DM*

| Area | Judul Penelitian | Ringkasan |
|--------------------------------|---|--|
| <i>Data mining</i> | Implementasi <i>Association Rule</i> dengan Algoritma <i>Apriori</i> pada Data Peminjaman Buku UPT Perpustakaan Universitas Lampung Menggunakan Metodologi <i>CRISP-DM</i> [11] | Penelitian ini membahas tentang pola <i>Association Rule</i> dengan teknik <i>data mining</i> supaya dapat memberikan rekomendasi pembelian buku menggunakan <i>dataset</i> dari UPT Perpustakaan Universitas Lampung. |
| <i>Artificial intelligence</i> | <i>Artificial Intelligence Model to Predict the Virality of Press Articles</i> [12] | Penelitian ini membahas tentang pemanfaatan model kecerdasan buatan dalam menentukan tingkat viralnya suatu pers artikel dengan tujuan supaya dapat mengingatkan masyarakat agar lebih bijak dalam memperoleh informasi yang viral |
| <i>Machine learning</i> | Pengembangan Model <i>Machine Learning</i> Regresi sebagai <i>Web Service</i> untuk Prediksi Harga Pembelian Mobil dengan Metode <i>CRISP-DM</i> [13] | Penelitian ini membahas tentang pemanfaatan model <i>machine learning</i> untuk melakukan prediksi harga pembelian mobil dengan tujuan supaya dapat meningkatkan penjualan mobil menggunakan informasi dari profil pelanggan |
| <i>Deep learning</i> | <i>Deep Learning for Activity Recognition Using Audio and Video</i> [14] | Penelitian ini membahas tentang <i>neural network</i> yang merupakan jaringan <i>deep learning</i> yang digunakan dalam deteksi audio dan video dengan tujuan untuk mengetahui adanya tindak kekerasan dari kedua media tersebut |
| <i>Big data</i> | <i>Big Data - Knowledge Discovery in Production Industry Data Storages - Implementation of Best Practices</i> [15] | Penelitian ini membahas tentang beberapa pendekatan yang mungkin digunakan dalam menyelesaikan permasalahan <i>big data</i> dimana untuk <i>theoretical framework</i> yang digunakan adalah metodologi <i>CRISP-DM</i> |
| <i>Data analysis</i> | <i>Analysis of Online Consumer Behavior - Design of CRISP-DM Process Model</i> [16] | Penelitian ini membahas tentang pengembangan model <i>CRISP-DM</i> yang disesuaikan dengan tujuan yaitu untuk menganalisis perilaku pelanggan <i>online</i> melalui pendekatan data analisis |
| <i>Data analytics</i> | Penerapan <i>K-Means</i> pada Segmentasi Pasar untuk Riset Pemasaran pada <i>Startup Early Stage</i> dengan Menggunakan <i>CRISP-DM</i> [17] | Penelitian ini membahas tentang segmentasi pasar yang dilakukan dengan memanfaatkan <i>data analytics</i> untuk memberikan solusi dari permasalahan yang dihadapi |

Penelitian-penelitian pada Tabel 3 adalah penelitian pada bidang *data science* yang menggunakan metodologi *CRISP-DM*. Berdasarkan penelusuran lebih lanjut, diketahui kebanyakan penelitian yang menggunakan *CRISP-DM* adalah pada area *data mining*, *artificial intelligence*, *machine learning*, dan *deep learning*. Sementara di area *big data*, *data analysis*, dan *data analytics* tidak terlalu banyak. Hal ini terjadi karena pada area *data mining*, *artificial intelligence*, *machine learning*, dan *deep learning* biasanya melibatkan pemodelan. Sementara pada area lainnya, tidak begitu diperlukan karena beberapa di antaranya hanya fokus pada analisis, bukan pada pemodelan dan evaluasi dari model tersebut. Meskipun demikian, metodologi *CRISP-DM* tetap dapat digunakan di area-area tersebut menyesuaikan dengan proyek *data science* yang dikerjakan. Selain itu, ini juga tidak menutup kemungkinan jika metodologi *CRISP-DM* digunakan di area lain. Tetapi karena *CRISP-DM* merupakan metodologi untuk proyek *data science*, maka *CRISP-DM* akan lebih tepat apabila digunakan di area dalam bidang *data science*.

B. Fase dalam Tahapan *CRISP-DM*

Metodologi *CRISP-DM* memiliki enam tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment* [18]. Skema tahapan dalam metodologi *CRISP-DM* ditampilkan pada Gambar 4.



Gambar 4. Tahapan dalam *CRISP-DM* [18]

Setiap tahapan dalam metodologi *CRISP-DM* memiliki fase yang berbeda-beda. Beberapa penelitian tidak mendeskripsikan fase-fase tersebut secara detail. Meskipun demikian, terdapat beberapa penelitian yang menyebutkan fase-fase tersebut sesuai dengan proses

penelitian masing-masing. Untuk mengetahui fase-fase dalam setiap tahapan, peneliti membandingkan tiga penelitian yang berbeda dimana ketiganya membahas tentang fase-fase dalam metodologi *CRISP-DM*. Perbandingan tersebut ditampilkan pada Tabel 4.

Tabel 4. Perbandingan Fase-fase dalam Tahapan *CRISP-DM*

| Tahapan | Penelitian 1 [19] | Penelitian 2 [20] | Penelitian 3 [21] |
|-------------------------------|--|---|--|
| <i>Business understanding</i> | <ul style="list-style-type: none"> Menentukan tujuan bisnis Mengidentifikasi kondisi terkini Menentukan tujuan <i>data mining</i> Membuat rencana proyek | Menulis deskripsi tekstual Menentukan tujuan <i>data mining</i> Membuat rencana proyek | Menentukan tujuan dan kebutuhan proyek Mendefinisikan masalah dan tujuan <i>data mining</i> Mempresentasikan rencana |
| <i>Data understanding</i> | Mengumpulkan data Mendeskripsikan data Mengeskplor data Memverifikasi kualitas data | Menyebutkan sumber data dan proses hasil Membuat deskripsi data secara terstruktur Mendeskripsikan kebutuhan statistika | Mengumpulkan data Mengidentifikasi kualitas data Menggali informasi Membentuk hipotesis |
| <i>Data preparation</i> | Memilih data Membersihkan data Membentuk data Mengintegrasikan data Memformat data | Mendeskripsikan input dan output data Melakukan transformasi, pemilihan, dan pembersihan data | Membentuk dataset akhir |
| <i>Modelling</i> | Memilih teknik <i>modelling</i> Membuat desain tes Membangun model Mengakses model | Menentukan pendekatan <i>modelling</i> Menyebutkan <i>tool</i> yang digunakan Membangun model untuk tes dan latihan | Memilih dan mencoba bermacam-macam model serta mengatur parameter yang digunakan |
| <i>Evaluation</i> | Mengevaluasi hasil Meninjau proses Menentukan langkah selanjutnya | Menentukan matrik evaluasi Memvisualisasikan hasil | Mengevaluasi kualitas model Memastikan model yang dipilih Menentukan langkah selanjutnya |
| <i>Deployment</i> | Merencanakan <i>deployment</i> Merencanakan <i>monitoring</i> Membuat laporan akhir Meninjau hasil proyek | Jika <i>deployment</i> diikutsertakan, implementasi harus dideskripsikan | Men-deploy model Mengolah hasil Membuat laporan akhir atau mengimplementasikan model |

Berdasarkan perbandingan pada Tabel 4, terlihat bahwa ketiga penelitian mempunyai fase yang berbeda-beda di setiap tahapannya. Meskipun demikian, terlihat juga bahwa fase-fase tersebut masih berhubungan satu sama lain. Dimulai dari tahap *business understanding* dimana fokus pada tahap tersebut adalah memperoleh informasi terkait kondisi saat ini dan menentukan tujuan proyek. Kemudian pada tahap *data understanding* dimana fokusnya adalah mengumpulkan data, melakukan evaluasi terhadap kualitas data, dan menggali informasi. Selanjutnya pada tahap *data preparation* berfokus pada persiapan data hingga membuat dataset akhir yang siap digunakan. Kemudian pada tahap *modelling* berfokus pada pembuatan model serta pada tahap *evaluation* adalah melakukan evaluasi hasil model dan pengambilan keputusan. Terakhir di tahap *deployment*, fokus lebih ke

penyebaran informasi yang dapat dilakukan baik melalui pengimplementasian model maupun pembuatan laporan akhir. Selain kesamaan-kesamaan tersebut, informasi lain yang dapat diperoleh dari perbandingan ketiga penelitian tersebut adalah penelitian pertama hampir mengikutsertakan semua fase yang ada pada dua penelitian lainnya. Hal ini dapat terlihat pada setiap tahapannya dimana penelitian pertama menyebutkan secara detail masing-masing fase yang perlu dilakukan.

Berdasarkan penelusuran lebih lanjut terkait fase dalam penelitian pertama, diketahui juga bahwa fase-fase tersebut sudah cukup sering digunakan. Bahkan beberapa dari penelitian yang menggunakan fase-fase tersebut telah dipublikasikan sejak tahun 2000-an yang berarti tahun tersebut bersamaan dengan waktu pertama kali

metodologi *CRISP-DM* dirilis. Dua dari publikasi penelitian yang dimaksud di antaranya adalah publikasi dengan judul “*CRISP-DM: Towards a Standard Process Model for Data Mining*” [21] dan “*CRISP-DM 1.0: Step-by-step Data Mining Guide*” [22]. Selain dari dua publikasi tersebut, masih ada beberapa penelitian lain yang juga menggunakan fase yang sama. Meskipun demikian, beberapa dari penelitian tersebut tidak menyebutkan referensinya. Hal ini dapat terjadi karena peneliti mengadaptasi dari penelitian sebelumnya. Tetapi ada juga kemungkinan lain seperti peneliti lupa atau dengan sengaja tidak menyebutkan referensi yang digunakan.

C. Kelebihan dan Tantangan *CRISP-DM*

Metodologi *CRISP-DM* banyak digunakan dalam berbagai area di bidang *data science*. Hal ini terjadi karena adanya kelebihan yang dimiliki sehingga peneliti menggunakan metodologi tersebut. Beberapa kelebihan dalam menggunakan metodologi *CRISP-DM* ditampilkan pada Tabel 5.

Tabel 5. Kelebihan *CRISP-DM*

| Kelebihan | Penjelasan |
|---|--|
| <i>De-facto</i> dan dapat dipercaya [7] | Berdasarkan survei dan <i>polling</i> responden, kebanyakan setuju dan menerima metodologi <i>CRISP-DM</i> sebagai standar yang digunakan untuk proyek <i>data mining</i> karena metode ini memiliki rangkaian tahapan yang lengkap dan dapat digunakan dengan baik pada proyek <i>data science</i> sehingga membuat metode ini lebih dipercaya [23] |
| Mudah dan terstruktur [7] | Hal ini dikatakan demikian karena semua tahapannya terdefiniskan dengan jelas, terstruktur dengan baik, dan tidak terlalu kompleks sehingga mudah untuk dimengerti serta metodologi ini dianggap sebagai praktik terbaik termasuk untuk pelajar baru sekalipun |
| Awalan yang benar [23] | Adanya tahap <i>business understanding</i> sangat membantu dalam memahami tujuan bisnis sehingga kebutuhan yang bersifat teknis dapat direncanakan lebih awal melalui identifikasi kebutuhan bisnis |
| Mengurangi biaya dan waktu [24] | Ini dapat mengurangi biaya karena jumlah tahapannya yang tidak banyak serta setiap tahapan mudah untuk diimplementasikan dimana ini juga dapat mengefisienkan waktu |
| <i>Deployment</i> tidak wajib [7] | Karena tahap ini tidak wajib, peneliti dapat memilih apakah hanya akan mengevaluasi model dan membuat laporan akhir saja atau mengimplementasikan model tersebut ke sebuah produk perangkat lunak yang bersifat lebih teknis |

Kelebihan-kelebihan tersebut menjadi alasan mengapa metodologi ini banyak digunakan. Hal ini tidak hanya

berdasarkan pendapat, tetapi beberapa diskusi juga menunjukkan data yang mendukung seperti hasil survei dan *polling* responden. Selain kelebihan, ada juga beberapa tantangan yang menjadi alasan mengapa beberapa peneliti tidak menggunakan metodologi ini atau peneliti merasakan tantangan tersebut saat menggunakan metodologi ini. Beberapa tantangan dalam menggunakan metodologi ini ditampilkan pada Tabel 6.

Tabel 6. Tantangan *CRISP-DM*

| Tantangan | Penjelasan |
|---|---|
| <i>Outdated</i> [23] | Selain metodologi ini telah dirilis sejak 20 tahun lalu, metodologi ini juga tidak mengalami pembaruan sehingga metodologi ini mungkin tidak sesuai dengan kebutuhan proyek <i>data science</i> saat ini |
| Awalan yang lambat [23] | Metodologi ini merupakan turunan dari metode <i>waterfall</i> dimana tahapan lebih banyak berfokus pada tahap di awal yang menyebabkan banyaknya waktu yang diperlukan dalam tahap sebelum <i>modelling</i> |
| Bukan kerangka koordinasi kelompok [23] | Karena metodologi ini mengasumsikan bahwa penggunaannya adalah seorang saja atau kelompok kecil, maka metodologi ini tidak tepat untuk proyek besar tanpa adanya penyesuaian fase di setiap tahapannya |
| Tidak ada kerangka yang jelas untuk berkomunikasi | Dalam tahapan yang ada, tidak dijelaskan bagaimana rangkaian komunikasi dilakukan karena “ <i>loop back</i> ” tidak dijelaskan secara spesifik terutama terkait letak komunikasi dengan <i>stakeholder</i> |
| Tahap <i>deployment</i> tidak ada | Tahapan ini tidak mempunyai acuan yang jelas karena kebanyakan penelitian menyebutkan bahwa tahap ini tidak dibutuhkan |

Tabel 5 dan 6 membahas mengenai kelebihan dan tantangan dari metodologi *CRISP-DM*. Masing-masing tabel memberikan lima kelebihan dan tantangan dalam menggunakan metode ini. Beberapa di antaranya saling terkait satu sama lain, seperti tahap *deployment*, yang memiliki sisi positif dan negatif. Kemudian juga terkait tahapan dimana adanya *business understanding* dapat membantu memahami bisnis, tetapi banyaknya tahapan di awal juga memerlukan waktu lebih. Dari masing-masing kelebihan dan tantangan tersebut, diharapkan dapat menjadi bahan pertimbangan bagi peneliti lain dalam menggunakan metodologi ini. Dengan demikian, ketika peneliti menggunakan metodologi *CRISP-DM*, peneliti telah siap dengan segala konsekuensinya.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metodologi *CRISP-DM* dapat digunakan dalam proyek bidang *data science* tepatnya dalam area *data mining*, *artificial intelligence*, *machine learning*, *deep learning*, *big data*, *data analysis*, dan *data*

analytics. Metodologi ini terdiri dari enam tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. Setiap tahapan tersebut memiliki fase yang berbeda dari satu dan lainnya. Untuk fase-fase tersebut, peneliti dapat mengikuti fase-fase yang telah umum digunakan oleh penelitian lain atau bisa menyesuaikan dengan proyek masing-masing. Kemudian sebagai pertimbangan dalam menggunakan metodologi *CRISP-DM*, peneliti dapat mempertimbangkannya berdasarkan kelebihan dan tantangan dari metodologi ini. Beberapa kelebihan dari metodologi *CRISP-DM* antara lain adalah *de-facto* dan dapat dipercaya, mudah digunakan dan terstruktur, awalan yang benar, mengurangi biaya dan waktu, serta tahap *deployment* tidak wajib. Kemudian untuk tantangan dalam menggunakan metodologi ini antara lain adalah *outdated*, awalan yang lambat, bukan merupakan kerangka koordinasi kelompok, tidak ada kerangka yang jelas untuk berkomunikasi, serta tahap *deployment* tidak ada. Secara keseluruhan, semua pertanyaan dalam penelitian ini telah terjawab. Untuk penelitian selanjutnya, sebaiknya menggunakan referensi yang lebih banyak serta dapat membandingkan dengan metodologi lain.

REFERENSI

- World Economic Forum. (2020, 20 Oktober). *The Future of Jobs Report 2020* [Online]. Available: <https://www.weforum.org/reports/the-future-of-jobs-report-2020/>.
- News18. (2022, 28 Desember). *Data Science Most Preferred Online Course in 2022: Survey* [Online]. Available: <https://www.news18.com/news/education-career/data-science-most-preferred-online-course-in-2022-survey-6605215.html>.
- O. Korableva, T. Durand, O. Kalimullina, and I. Stepanova, "Studying User Satisfaction with the MOOC Platform Interfaces Using the Example of Coursera and Open Education Platforms," in *Proceedings of the 2019 International Conference on Big Data and Education*, 2019. doi: <https://doi.org/10.1145/3322134.3322139>.
- N. P. Volkova, N. O. Rizun, and M. V. Nehrey, "Data Science: Opportunities to Transform Education", *CTE Workshop Proc.*, vol. 6, pp. 48–73, Mar. 2019.
- J. S. Saltz and N. Hotz, "Identifying the Most Common Frameworks Data Science Teams Use to Structure and Coordinate Their Projects," in *2020 IEEE International Conference on Big Data (Big Data)*, 2020. doi: [10.1109/BigData50022.2020.9377813](https://doi.org/10.1109/BigData50022.2020.9377813).
- B. Christoffer and M. Lindovsky, "Machine Learning Project Management - A Study of Project Requirements and Processes in Early Adoption." Master's thesis, Dept. Architecture and Civil Engineering, Univ., Chalmers University of Technology, Gothenburg, Sweden, 2019.
- C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021. doi: <https://doi.org/10.1016/j.procs.2021.01.199>.
- F. Wulandari. (2020, Dec). "Pemanfaatan Lingkungan sebagai Sumber Belajar Anak Sekolah Dasar (Kajian Literatur)". *Journal of Educational Review and Research*, vol. 3(2), pp. 105-110. Available: <https://scholar.archive.org/work/vjcvozxdwfdg7ix4qz2shetmua/access/w/awayback/https://journal.stkipsingkawang.ac.id/index.php/JERR/article/download/2158/pdf>.
- P. M. Seeger, Z. Yahouni, and G. Alpan, "Literature Review on Using Data Mining in Production Planning and Scheduling Within the Context of Cyber Physical Systems," *J. Ind. Inf. Integr.*, vol. 28, no. 100371, 2022. doi: <https://doi.org/10.1016/j.jii.2022.100371>.
- I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 6, pp. 420, 2021. doi: <https://doi.org/10.1007/s42979-021-00815-1>.
- M. Fitriani, "Implementasi Association Rule dengan Algoritma Apriori Pada Data Peminjaman Buku UPT Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM" Skripsi Sarjana, Fakultas Teknik, Universitas Lampung, Lampung, Indoensia, 2019.
- Y. L. Lopez, D. Grimaldi, S. Garcia, J. Ordoez, C. Carrasco-Farre, and A. A. Aristizabal, "Artificial Intelligence Model to Predict the Virality of Press Articles," in *2022 14th International Conference on Machine Learning and Computing (ICMLC)*, 2022. doi: <https://doi.org/10.1145/3529836.3529953>.
- A. M. M. Fattah, A. Voutama, N. Haryana, dan N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p.1669-1678, 2022. doi: <http://dx.doi.org/10.30865/jurikom.v9i5.5021>.
- F. Reynolds, C. Neto, and J. Machado, "Deep Learning for Activity Recognition Using Audio and Video," *Electronics (Basel)*, vol. 11, no. 5, p. 782, 2022. doi: <https://doi.org/10.3390/electronics11050782>.
- J. Abasova, P. Tanuska, and S. Rydzi, "Big Data—Knowledge Discovery in Production Industry Data Storages—Implementation of Best Practices," *Appl. Sci. (Basel)*, vol. 11, no. 16, p. 7648, 2021. doi: <https://doi.org/10.3390/app11167648>.
- E. Exenberger, & J. Bucko, "Analysis of Online Consumer Behavior-design of CRISP-DM Process Model." *AGRIS on-line Papers in Economics and Informatics*, vol. 12, no. 3, pp. 13-22, 2020. doi: <http://dx.doi.org/10.22004/ag.econ.320071>.
- Christian, Yefta, and Katherine Oktaviani Yap Rui Qi. "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol 9, no. 4, pp. 966-973, 2022. doi: <http://dx.doi.org/10.30865/jurikom.v9i4.4486>.
- F. Martinez-Plumed et al., "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, 2021. doi: <https://doi.org/10.1109/TKDE.2019.2962680>.
- A. Zernig, A. Pandeshwar, R. Kern, and M. Rauch, "Machine Learning and Automated Decision Making," in *SemI40 Project Prospective: Industry 4.0 Evolution Revolution, Austria. SemI40 Consortium*, 2019, pp. 58–75. Accessed: Mar. 20, 2023. [Online]. Available: <https://www.researchgate.net/publication/337592264>.
_A_SemI40_Project_Prospective_-_Industry40_from_Evolution_to_Revolution
- V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of Data Mining Methodologies: A Systematic Literature Review," *PeerJ Comput. Sci.*, vol. 6, no. e267, p. e267, 2020. doi: <https://doi.org/10.7717/peerj-cs.267>.
- R. Wirth, & J. HippJ, "CRISP-DM: Towards a Standard Process Model for Data Mining," In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, vol. 1, pp. 29-39, 2000.

22. P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, & R. Wirth, "CRISP-DM 1.0: Step-by-step Data Mining Guide," *SPSS*, 2000.
23. Data Science Process Alliance. (2021). *Evaluation Data Science* [Online]. Available: <https://www.datascience-pm.com/wp-content/uploads/2021/08/CRISP-DM-for-Data-Science.pdf>.
24. W. Y. Ayele, "Adapting CRISP-DM for Idea Mining: A Data Mining Process for Generating Ideas Using a Textual Dataset," *arXiv [cs.IR]*, 2021. doi: <https://doi.org/10.48550/arXiv.2105.00574>.
25. J. S. Saltz, "CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021. doi: <https://doi.org/10.1109/BigData52589.2021.9671634>.