

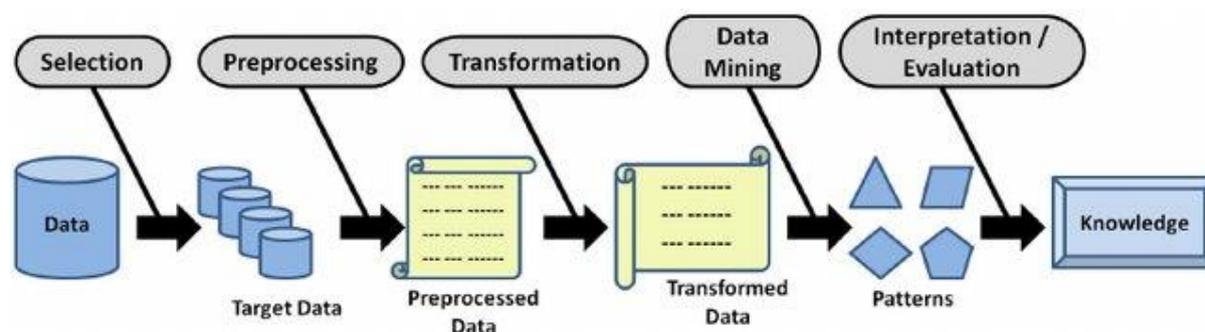
1.2. Framework dalam Data Mining

Dalam dunia data mining yang kompleks, framework berperan sebagai panduan yang terstruktur untuk menjalankan proyek secara efisien dan efektif. Framework ini memberikan kerangka kerja yang jelas, menguraikan tahapan penting, aktivitas kunci, dan pertimbangan yang diperlukan untuk mengekstraksi pengetahuan berharga dari data.

Bagian ini membahas perbandingan framework data mining tradisional dengan salah satu framework yang paling dikenal dan banyak digunakan saat ini, yaitu CRISP-DM (*Cross Industry Standard Process for Data Mining*).

Framework Data Mining Tradisional

Framework data mining tradisional biasanya bersifat lebih linear dan terstruktur, dengan fokus pada tahapan dasar dari proses analisis data. Beberapa literatur di Internet umumnya menggambarkan framework Data Mining tradisional seperti dalam Gambar 1.1.



Gambar 1.1. Gambaran framework DM tradisional

Beberapa framework ini mencakup pendekatan yang lebih teknis dan sering kali digunakan dalam dunia akademis atau penelitian. Framework tradisional ini biasanya terdiri dari tahapan-tahapan berikut:

1. Mengumpulkan dataset atau database yang akan ditambang, atau dikenal sebagai data mentah atau asli yang masih 'kotor'
2. Pemilihan fitur dari dataset/database. Dalam database atau dataset, fitur biasanya diwakili oleh kolom data. Istilah lain dari fitur antara lain: atribut, field, dan variabel. Dimensi dari sebuah data tabular adalah banyaknya kolom fitur, sedangkan ukuran data diwakili oleh banyaknya baris rekaman. Dimensi dari sebuah data tekstual, adalah banyaknya kata di dalam data.
3. Praproses Data: Melakukan pembersihan data, menangani missing values, dan normalisasi dari data yang sudah dipilih fitur-fiturnya
4. Transformasi Data: Mengubah atau mereduksi data hasil pra-proses untuk dapat digunakan oleh algoritma, program atau metode penambangan data
5. Menambang data, dengan sub-tahapan biasanya termasuk:
 1. Memilih teknik atau algoritma DM yang sesuai (seperti clustering, klasifikasi, regresi, atau association rule mining).

2. Menjalankan algoritma DM pada data untuk mendapatkan pola
 3. Eksplorasi pola yaitu langkah lanjutan yang diterapkan pada pola agar lebih mudah diinterpretasikan. Misalnya, membuat struktur pohon (Tree)
 4. Menyimpan model, biasanya pada kegiatan klasifikasi data, untuk digunakan lebih lanjut pada tahap implementasi yang lebih masal.
6. Evaluasi dan Interpretasi terhadap:
1. Model, khususnya pada kualitas akurasi model dan/atau metrik lainnya
 2. Pola hasil penambangan, guna mengekstrak pengetahuan dari Pola, yang berpotensi baru, menarik dan berguna bagi pengambilan Keputusan.

Framework tradisional ini sering kali tidak memperhatikan aspek iteratif dari proses data mining. Prosesnya bersifat linier, dari awal hingga akhir, tanpa penekanan pada evaluasi terus-menerus atau pengulangan siklus untuk penyempurnaan. Keterbatasan lain dari pendekatan tradisional adalah minimnya fokus pada pemahaman bisnis dan penggunaan model di dunia nyata.

CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM adalah framework data mining yang lebih modern dan paling banyak digunakan dalam industri. CRISP-DM menawarkan pendekatan yang lebih iteratif dan berpusat pada bisnis, dengan fase-fase yang saling terkait dan siklus yang fleksibel untuk mengakomodasi perubahan atau kebutuhan yang berkembang selama proyek.

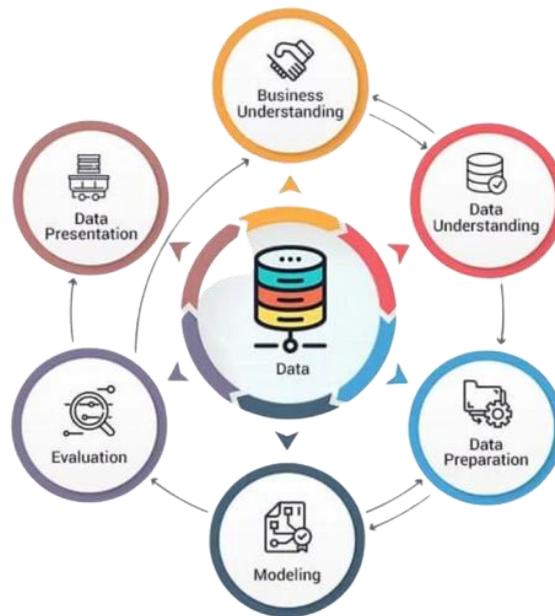
CRISP-DM merupakan model proses standar yang telah menjadi acuan utama dalam dunia data mining sejak dikembangkan pada akhir 1990-an. Model ini dirancang oleh konsorsium yang terdiri dari berbagai perusahaan besar seperti Daimler-Benz, Integral Solutions Ltd., NCR, dan OHRA, dengan masukan dari lebih dari 200 pengguna dan penyedia layanan data mining. Tujuan utama CRISP-DM adalah menyediakan kerangka kerja yang terstruktur dan sistematis untuk menyelesaikan proyek data mining, yang bersifat non-proprietary dan dapat diadaptasi oleh berbagai industri, alat, serta aplikasi.

CRISP-DM terdiri dari enam tahapan utama: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Tahapan ini memastikan bahwa proyek data mining dimulai dengan pemahaman mendalam tentang tujuan bisnis dan kebutuhan organisasi. Setelah itu, data dikumpulkan, diproses, dan dimodelkan menggunakan teknik-teknik data mining yang sesuai. Model yang dihasilkan kemudian dievaluasi untuk memastikan relevansinya terhadap tujuan bisnis, dan jika berhasil, model tersebut diterapkan ke dalam lingkungan bisnis yang nyata.

Salah satu keunggulan utama dari CRISP-DM adalah fleksibilitasnya. Model ini bersifat netral terhadap industri, alat, dan aplikasi, sehingga dapat diterapkan di berbagai sektor seperti pemasaran, keuangan, manufaktur, dan manajemen hubungan pelanggan (CRM). Berbagai organisasi besar, seperti DaimlerChrysler dan OHRA, telah menerapkan CRISP-DM dalam proyek-proyek data mining mereka, menunjukkan keefektifan model ini dalam menghasilkan hasil yang lebih cepat dan lebih baik. CRISP-DM juga telah diadopsi oleh penyedia alat data mining seperti SPSS dan NCR, memperluas pengaruhnya di seluruh industri.

Keberhasilan CRISP-DM dalam memandu proyek data mining berkat pendekatan yang terstruktur namun adaptif, memungkinkan organisasi untuk menyesuaikan setiap tahapan sesuai dengan kebutuhan khusus mereka. Meskipun tidak bersifat kaku, CRISP-DM menawarkan fondasi yang kuat untuk memastikan keberhasilan proyek data mining, baik bagi para pemula maupun praktisi berpengalaman. Model ini telah memberikan dampak signifikan pada standar industri data mining dan sering dijadikan syarat dalam proposal atau tender proyek.

CRISP-DM menggambarkan kegiatan DM sebagai siklus, sebagaimana digambarkan dalam Gambar 1.2.



Gambar 1.2. Siklus DM dalam CRISP-DM

Lebih detail tahapan dalam CRISP-DM adalah sebagai berikut:

1. Business Understanding: Langkah pertama berfokus pada memahami tujuan bisnis yang ingin dicapai dengan data mining dan mengidentifikasi masalah yang akan diselesaikan. Tahapan ini juga menetapkan kriteria keberhasilan proyek dari sudut pandang bisnis.
2. Data Understanding: Pada tahap ini, data dikumpulkan dan dieksplorasi untuk mendapatkan pemahaman awal tentang isinya, termasuk potensi masalah dalam data yang harus diatasi sebelum analisis.
3. Data Preparation: Data kemudian disiapkan melalui proses pembersihan, transformasi, dan format ulang, sehingga siap untuk dianalisis. Ini termasuk menangani missing values, menggabungkan beberapa sumber data, dan memilih atribut yang relevan.
4. Modeling: Berbagai teknik data mining diterapkan pada data yang telah disiapkan. Tahapan ini melibatkan eksperimen dengan berbagai algoritma dan parameter untuk menemukan model terbaik.

5. Evaluation: Setelah model dibuat, langkah berikutnya adalah mengevaluasi kinerjanya dengan tujuan untuk memastikan bahwa model tersebut memenuhi tujuan bisnis yang telah ditentukan pada tahap awal. Jika model tidak memadai, tim mungkin kembali ke tahap sebelumnya.
6. Deployment: Hasil akhir dari proyek data mining diterapkan dalam lingkungan bisnis. Ini bisa berupa pelaporan, implementasi sistem yang otomatis, atau bahkan integrasi model ke dalam sistem operasional. Feedback dari implementasi juga sering digunakan untuk memperbaiki model di masa depan.

Perbandingan Utama Antara Framework Tradisional dan CRISP-DM:

Perbedaan kedua framework dirangkum sebagai berikut:

- Pendekatan Berfokus pada Bisnis: CRISP-DM memulai dengan pemahaman yang kuat tentang tujuan bisnis, sedangkan framework tradisional lebih teknis dan tidak terlalu menekankan aspek bisnis.
- Iteratif vs Linier: CRISP-DM bersifat iteratif, artinya memungkinkan pengulangan tahapan untuk penyempurnaan hasil. Framework tradisional umumnya bersifat linier, di mana proses dimulai dari tahap pertama dan berakhir di tahap terakhir tanpa banyak fleksibilitas untuk kembali ke tahap sebelumnya.
- Fokus Evaluasi Berkelanjutan: CRISP-DM menekankan pada evaluasi terus-menerus di setiap tahap untuk memastikan hasil yang dicapai relevan dan mendukung tujuan bisnis. Framework tradisional lebih fokus pada evaluasi hasil setelah model selesai.
- Keterhubungan Tahapan: Dalam CRISP-DM, tahapan saling terkait, memungkinkan untuk kembali ke tahapan sebelumnya jika diperlukan. Framework tradisional cenderung lebih kaku, dengan tahapan yang kurang fleksibel.
- Dukungan Industri: CRISP-DM dirancang untuk digunakan dalam berbagai industri, membuatnya lebih relevan dan praktis untuk aplikasi bisnis. Sementara framework tradisional lebih banyak diterapkan dalam penelitian atau lingkungan teknis tertentu.

Framework data mining tradisional menawarkan pendekatan yang terstruktur dan teknis, tetapi sering kali tidak cukup fleksibel untuk menangani kebutuhan iteratif dan bisnis dalam proyek data mining modern. Di sisi lain, CRISP-DM adalah framework yang berpusat pada bisnis, fleksibel, dan iteratif, yang membuatnya lebih relevan untuk aplikasi industri saat ini. Dengan pendekatan yang lebih komprehensif, CRISP-DM memungkinkan proyek data mining mencapai tujuan yang lebih sesuai dengan kebutuhan bisnis serta mendukung pengambilan keputusan yang efektif di dunia nyata.

Selain perbedaan, kedua framework ini memiliki beberapa titik persamaan, yaitu:

Baik dalam framework data mining tradisional maupun CRISP-DM, tahapan pra-proses atau persiapan data sering kali menjadi pekerjaan yang paling banyak menyita waktu dan tenaga. Hal ini disebabkan oleh berbagai aktivitas seperti pengumpulan, pembersihan, transformasi, dan integrasi data, yang sangat penting untuk memastikan bahwa data yang digunakan dalam analisis adalah bersih, konsisten, dan siap digunakan. Kualitas data yang buruk atau tidak memadai dapat mengurangi akurasi model dan hasil analisis, sehingga

membuat tahapan ini menjadi sangat kritis. Dalam kedua framework, kesuksesan proyek data mining sangat bergantung pada sejauh mana data dipersiapkan dengan baik.

Dalam Framework Data Mining Tradisional:

Tahapan penyiapan data dilakukan setelah pemilihan data dan sebelum pemilihan algoritma. Aktivitas pada tahap ini meliputi pembersihan data (menghapus data yang tidak valid atau duplikat), transformasi data (misalnya normalisasi), serta pemilihan atribut yang relevan. Karena tahap ini mencakup banyak proses manual seperti deteksi dan penanganan missing values, penggabungan sumber data, dan perubahan format, tidak mengherankan bahwa penyiapan data bisa memakan waktu hingga 70-80% dari keseluruhan waktu proyek.

Dalam CRISP-DM:

Tahap Data Preparation (Persiapan Data) merupakan fase penting setelah tahap Data Understanding. Di sini, data yang telah dipahami dan dieksplorasi lebih lanjut diolah menjadi dataset akhir yang siap digunakan dalam proses pemodelan. Sama seperti framework tradisional, proses persiapan data ini melibatkan pembersihan, transformasi, serta pemilihan data yang sesuai untuk keperluan modeling. Penyiapan data pada CRISP-DM juga dapat memakan banyak waktu karena memerlukan pemrosesan yang cermat untuk memastikan model dapat bekerja dengan baik.

Inti dari Kegiatan Data Mining di Kedua Framework:

- Framework Tradisional: Inti dari kegiatan data mining terletak pada Pemilihan Algoritma dan Eksekusi Algoritma. Ini adalah tahap di mana algoritma data mining seperti clustering, classification, regression, atau association rule diterapkan pada data untuk menemukan pola atau membangun model prediksi. Tahap ini merupakan esensi dari analisis data mining itu sendiri, di mana data yang telah diproses digunakan untuk menghasilkan model yang bermanfaat bagi pengambilan keputusan.
- CRISP-DM: Inti dari kegiatan data mining terletak pada tahap Modeling. Setelah data dipersiapkan, berbagai teknik data mining diterapkan pada dataset untuk membangun model yang sesuai dengan tujuan bisnis. Proses pemodelan di CRISP-DM mencakup eksperimen dengan berbagai algoritma, serta optimasi parameter untuk menghasilkan model terbaik. Tahap ini penting karena hasil akhirnya berupa model yang mampu memprediksi atau mengungkap pola dari data, yang selanjutnya dievaluasi untuk melihat apakah model tersebut memenuhi tujuan bisnis.

Secara keseluruhan, meskipun penyiapan data memakan waktu yang cukup lama, pemodelan dan eksekusi algoritma tetap menjadi pusat utama dari kegiatan data mining di kedua framework ini. Namun, tanpa penyiapan data yang baik, tahap inti tersebut tidak akan menghasilkan model yang akurat atau andal.

Kelebihan dan Keterbatasan CRISP-DM

CRISP-DM memiliki sejumlah keunggulan yang membuatnya populer di dunia industri dan akademis. Pertama, framework ini terstruktur dan mudah diikuti, dengan tahapan yang jelas

mulai dari pemahaman bisnis hingga deployment. Pendekatan ini membantu tim proyek memiliki panduan yang sistematis, sehingga setiap fase dilalui dengan urutan yang logis dan terorganisir. Selain itu, CRISP-DM memastikan semua aspek penting dipertimbangkan, termasuk pemahaman bisnis, pemrosesan data, dan evaluasi model, yang memungkinkan hasil yang lebih relevan dan dapat diterapkan dalam konteks bisnis.

Framework ini juga cocok digunakan untuk proyek dengan kompleksitas rendah hingga menengah. Dengan struktur yang jelas dan tahapan yang bisa diulang, CRISP-DM membantu memastikan bahwa tidak ada detail yang terlewat. Keunggulan lainnya adalah dukungan luas dari alat dan sumber daya, baik dalam bentuk software maupun komunitas pengguna. Banyak alat data mining dan platform analisis seperti KNIME, RapidMiner, dan SAS yang mendukung penerapan CRISP-DM, sehingga pengguna dapat dengan mudah mengintegrasikan proses ini dalam alur kerja mereka.

Namun, CRISP-DM bukan tanpa keterbatasan. Framework ini cenderung kurang fleksibel untuk proyek yang sangat kompleks atau yang menghadapi perubahan cepat. Meskipun bersifat iteratif, iterasi dalam CRISP-DM dapat menjadi lambat dan membutuhkan waktu, terutama ketika proyek membutuhkan banyak pengulangan antara tahapan. Perubahan kebutuhan bisnis di tengah proyek bisa menjadi tantangan dalam framework ini, karena memerlukan adaptasi yang signifikan di setiap fase untuk mengakomodasi perubahan tersebut.

Selain itu, karena iterasi antar tahapan mungkin membutuhkan waktu yang lama, CRISP-DM terkadang kurang efisien dalam proyek dengan perubahan kebutuhan bisnis yang dinamis. Hal ini menjadikan CRISP-DM lebih cocok untuk situasi di mana ruang lingkup dan tujuan proyek sudah jelas dan stabil sejak awal.

Kapan Menggunakan CRISP-DM? CRISP-DM paling efektif digunakan ketika proyek memiliki tujuan dan ruang lingkup yang jelas. Misalnya, jika sebuah organisasi telah menetapkan dengan pasti apa yang ingin dicapai dari proyek data mining, seperti prediksi penjualan atau analisis churn pelanggan, CRISP-DM dapat membantu memandu proyek secara sistematis hingga mencapai hasil yang diinginkan.

Framework ini juga ideal jika kebutuhan bisnis relatif stabil dan tidak banyak berubah selama proyek berlangsung. CRISP-DM menawarkan pendekatan yang sistematis, dan dengan sedikit perubahan, setiap tahap dapat dikelola dengan lancar. Selain itu, framework ini sangat cocok bagi tim proyek yang memiliki pengalaman terbatas dalam data mining. Dengan struktur yang sudah terdefinisi, CRISP-DM memberikan panduan yang jelas dan membantu tim-tim baru untuk mengikuti proses tanpa banyak kesulitan.

Penerapan CRISP-DM

CRISP-DM telah diimplementasikan dan/atau diadaptasi oleh institusi di bidang industri, akademisi dan pemerintahan. Beberapa di antaranya dirangkum dalam uraian berikut ini, meliputi adaptasi CRISP-DM oleh Amazon Web Service (AWS), Akademisi, KNIME, dan Pemerintah Republik Indonesia khususnya di bidang Ketenagakerjaan.

Penerapan CRISP-DM dalam bidang kesehatan dan ilmu hayati berdasarkan paper AWS Whitepaper tentang "Machine Learning Best Practices in Healthcare and Life Sciences", dijelaskan sebagai berikut:

1. **Pemahaman Bisnis:** Pada fase pertama CRISP-DM, paper ini menekankan pentingnya pemahaman yang mendalam tentang tujuan bisnis serta persyaratan regulasi dalam penerapan machine learning (ML) di sektor kesehatan. Dalam industri ini, para ahli dan pemilik bisnis harus mendefinisikan tujuan yang jelas serta memastikan bahwa model ML yang dikembangkan sesuai dengan standar regulasi, seperti Good Machine Learning Practices (GMLP) yang dikeluarkan oleh FDA. Contoh yang disebutkan adalah bagaimana perusahaan farmasi menggunakan AI/ML untuk mendeteksi kejadian buruk (adverse event) dan memastikan model tersebut mematuhi standar pelaporan ketat dari regulator.
2. **Pemahaman dan Persiapan Data:** Fase kedua dari CRISP-DM, yaitu Data Understanding dan Data Preparation, sangat ditekankan dalam konteks sektor kesehatan, terutama terkait keamanan dan kepatuhan. Paper ini menjelaskan bagaimana pengumpulan dan persiapan data harus dilakukan dengan hati-hati untuk menjamin privasi dan memenuhi persyaratan regulasi. AWS menyarankan penggunaan alat-alat seperti Amazon S3 dan AWS Glue untuk mengelola dan memproses data besar dari sektor kesehatan dengan aman. Fase ini sangat penting untuk memastikan bahwa data yang digunakan untuk model ML diolah dengan cara yang tepat, terenkripsi, dan dapat dilacak, yang merupakan hal krusial dalam lingkungan yang diatur ketat seperti layanan kesehatan.
3. **Pemodelan dan Evaluasi:** Fase Modeling dan Evaluation dari CRISP-DM tercermin dalam rekomendasi AWS untuk membangun dan mengevaluasi model ML menggunakan alat seperti Amazon SageMaker. Paper ini membahas bagaimana model dilatih dan divalidasi untuk memastikan bahwa model tersebut memenuhi standar kinerja sekaligus mematuhi persyaratan regulasi seperti Good Clinical and Manufacturing Practices (GxP). Dalam bidang kesehatan, kemampuan untuk menjelaskan dan menginterpretasi model sangat penting karena keputusan yang diambil dari model ML memiliki dampak langsung pada keselamatan pasien.
4. **Deployment dan Monitoring:** Fase Deployment dalam CRISP-DM terkait dengan penerapan model ML dalam produksi. Dalam paper ini, AWS merekomendasikan alat seperti SageMaker Model Monitor untuk memantau kinerja model ML secara berkelanjutan di lingkungan produksi. Pengawasan yang terus-menerus ini sejalan dengan prinsip CRISP-DM untuk memastikan bahwa model yang digunakan tetap akurat dan aman dalam aplikasi dunia nyata, terutama dalam konteks keputusan penting di bidang kesehatan.

Dalam artikel "Big Data–Knowledge Discovery in Production Industry Data Storages–Implementation of Best Practices," framework CRISP-DM diterapkan sebagai dasar untuk melakukan analisis data yang kompleks dalam lingkungan Industri 4.0. Berikut adalah ringkasan penerapannya:

1. **Pemahaman Bisnis (Business Understanding):** Artikel ini menekankan pentingnya memahami proses produksi, sumber data, dan konteks keseluruhan di industri sebelum memulai analisis data. Dalam konteks Industri 4.0, pemahaman ini menjadi dasar untuk melakukan analisis kesalahan dan pencarian pola tersembunyi di data produksi.
2. **Pemahaman Data (Data Understanding):** Dalam industri manufaktur, data yang sangat besar dan heterogen sering ditemukan. CRISP-DM membantu dalam

mengidentifikasi dan memahami sumber-sumber data yang berbeda, termasuk data dari proses produksi, sensor, dan data terkait kualitas produk. Data ini kemudian digunakan untuk menemukan hubungan antarproses dan prediksi kegagalan.

3. **Persiapan Data (Data Preparation):** Proses ini melibatkan penyesuaian struktur data, integrasi berbagai sumber data, dan penghapusan inkonsistensi seperti perbedaan format atau kesalahan pengkodean. Dalam industri, pengolahan data yang tepat sangat penting untuk memastikan keakuratan model prediksi kegagalan.
4. **Pemodelan (Modeling):** Pada tahap ini, CRISP-DM digunakan untuk membuat model prediksi, misalnya, untuk memprediksi kegagalan pada proses produksi berdasarkan data proses. Algoritma machine learning digunakan untuk memodelkan data dan mengidentifikasi pola atau potensi kesalahan.
5. **Evaluasi (Evaluation):** Model yang dibangun kemudian dievaluasi untuk memastikan efektivitasnya dalam lingkungan industri. Model dievaluasi terhadap data historis untuk memprediksi masalah yang mungkin terjadi.
6. **Penerapan (Deployment):** Hasil analisis diimplementasikan dalam sistem monitoring yang memungkinkan deteksi dini terhadap kegagalan atau anomali dalam proses produksi. CRISP-DM membantu dalam pengaturan proses monitoring secara real-time.

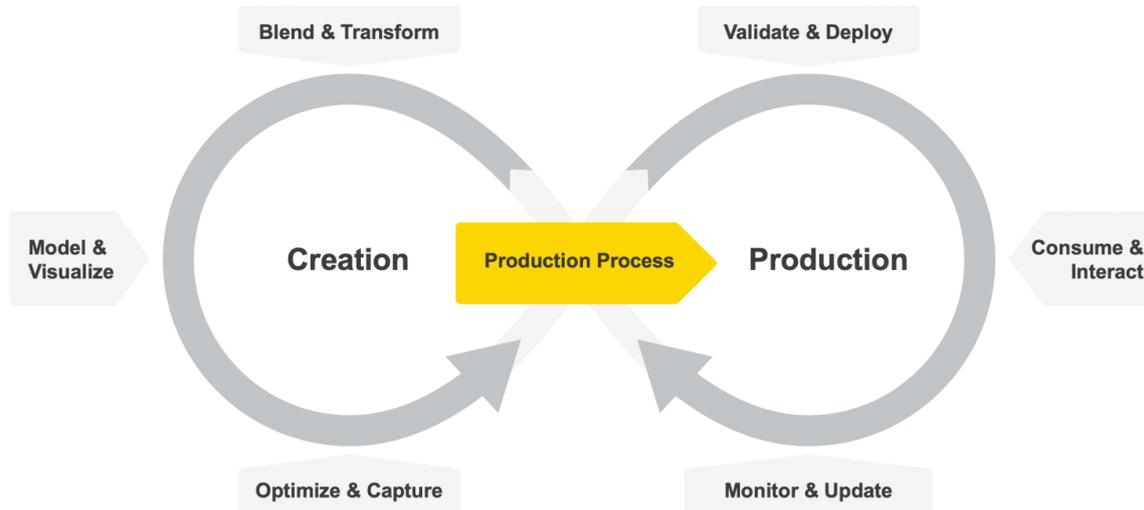
Dalam artikel "A New Standard for the Data Science Process," KNIME mengembangkan pendekatan baru untuk proses data science, yang dibangun di atas framework yang sudah ada seperti CRISP-DM, SEMMA, dan KDD. Artikel ini menyebutkan bahwa meskipun CRISP-DM menyediakan struktur yang baik untuk analisis data, framework ini kurang memperhatikan tahap setelah deployment, yaitu bagaimana model diproduksi dan dimanfaatkan dalam dunia nyata.

KNIME mengadopsi dan memperluas CRISP-DM dengan memperkenalkan siklus hidup data science yang mencakup produksi model, konsumsi oleh pengguna, dan monitoring serta pembaruan model berdasarkan umpan balik dari lingkungan produksi. Artikel ini menekankan pentingnya iterasi dalam siklus hidup data science, di mana model terus dievaluasi, diperbaiki, dan dioptimalkan setelah deployment.

Tahapan dalam pengembangan CRISP-DM di KNIME meliputi:

1. **Blend & Transform:** Menggabungkan dan membersihkan data dari berbagai sumber.
2. **Model & Visualize:** Membangun model dan memvisualisasikan hasilnya.
3. **Optimize & Capture:** Mengoptimalkan model dan mengatur transformasi data yang diperlukan untuk implementasi produksi.
4. **Validate & Deploy:** Memvalidasi model dan melakukan deployment ke lingkungan produksi.
5. **Monitor & Update:** Setelah deployment, model dimonitor dan diperbarui berdasarkan performa dan umpan balik nyata.

Dengan pengembangan ini, KNIME menutup kesenjangan yang ada di CRISP-DM dalam hal deployment dan operasionalisasi model, menjadikannya lebih sesuai untuk lingkungan bisnis modern yang membutuhkan solusi data science yang berkelanjutan dan responsif. Data Science Life Cycle atau siklus hidup data science yang dikembangkan KNIME, diberikan dalam Gambar 1.3.



Gambar 1.3. Framework Data Science dalam KNIME

Framework CRISP-DM secara eksplisit digunakan sebagai pedoman utama dalam SKKNI 2020-299 di bidang Artificial Intelligence sub-bidang Data Science di Indonesia. Adopsi framework ini bertujuan untuk menyediakan panduan yang sistematis dalam proses data science, mulai dari pemahaman kebutuhan bisnis hingga implementasi model di lingkungan produksi.

Tahapan CRISP-DM dalam SKKNI 2020-299:

1. Business Understanding (Pemahaman Bisnis): Ditujukan untuk memahami dan mendefinisikan tujuan proyek berdasarkan kebutuhan organisasi. Pada SKKNI ini, fokus pada menentukan objektif bisnis, mengidentifikasi risiko, dan merumuskan langkah mitigasi risiko sangat penting.
2. Data Understanding (Pemahaman Data): Tahapan ini mencakup pengumpulan, validasi, dan evaluasi data. SKKNI menguraikan bahwa proses menelaah dan memvalidasi data menjadi langkah kunci dalam memastikan kualitas dan kelengkapan data untuk analisis lebih lanjut.
3. Data Preparation (Persiapan Data): Dalam SKKNI, langkah ini mencakup aktivitas pembersihan, konstruksi, dan integrasi data untuk menghasilkan dataset siap pakai yang konsisten dan relevan dengan kebutuhan teknis proyek data science.
4. Modeling (Pemodelan): Tahapan ini berfokus pada pemilihan algoritma dan teknik pemodelan yang sesuai dengan karakteristik data. Pada SKKNI, penekanan pada membangun model yang dapat dioptimasi dan dievaluasi berdasarkan parameter kinerja yang sesuai menjadi hal yang kritis.

5. Evaluation (Evaluasi): SKKNI memberikan perhatian pada evaluasi hasil pemodelan berdasarkan metrik yang telah ditentukan, memastikan model yang dibangun memenuhi tujuan teknis dan bisnis yang diharapkan.
6. Deployment (Penerapan): Tahapan ini mencakup penerapan model ke dalam produksi dan pemantauan kinerja model secara berkelanjutan. SKKNI menyoroti pentingnya langkah ini untuk memastikan model tetap memberikan hasil yang optimal dalam aplikasi nyata.

Framework CRISP-DM yang digunakan dalam SKKNI 2020-299 menunjukkan bagaimana pendekatan standar ini diadaptasi untuk mengembangkan keterampilan tenaga kerja yang relevan dengan industri data science dan artificial intelligence di Indonesia.

Referensi:

1. Shearer, C. (2000). *The CRISP-DM Model: The New Blueprint for Data Mining*. CRISP-DM Consortium. Retrieved from [CRISP-DM Documentation].
2. AWS Whitepaper - Machine Learning Best Practices in Healthcare and Life Sciences: Amazon Web Services. (n.d.). *Machine Learning Best Practices in Healthcare and Life Sciences*. AWS Whitepaper. Retrieved from [AWS website].
3. Big Data–Knowledge Discovery in Production Industry Data Storages–Implementation of Best Practices: Bogdan, W., Wójcik, A., Lysenko, T., & Konik, R. (2019). Big Data–Knowledge Discovery in Production Industry Data Storages–Implementation of Best Practices. *Proceedings of the International Conference on Computer Science and Software Engineering* (pp. 146-160). https://doi.org/10.1007/978-3-030-12385-7_15
4. A New Standard for the Data Science Process _ KNIME: KNIME. (n.d.). *A New Standard for the Data Science Process*. Retrieved from <https://www.knime.com/new-standard-for-data-science-process>
5. Standar Kompetensi Kerja Nasional Indonesia (SKKNI) 2020-299: Kementerian Ketenagakerjaan Republik Indonesia. (2020). *Standar Kompetensi Kerja Nasional Indonesia (SKKNI) Bidang Artificial Intelligence Sub-Bidang Data Science (SKKNI 2020-299)*. Jakarta, Indonesia: Kementerian Ketenagakerjaan.