



**Ardian - 2106638173**

**Data Mining**

**Assignment 02 - Studi Kasus Penambangan Data**

1. Bagaimana Anda mengimplementasikan tahapan CRISP-DM jika diminta membuat tool seperti ChatGPT? Jelaskan setiap proses yang perlu dilakukan

**Jawaban:**

- 1. Business Understanding**

Dalam implementasi tahapan CRISP-DM untuk membuat tool seperti ChatGPT, kita akan memulai dengan tahap Business Understanding. Tahap ini fokus pada pemahaman terhadap tujuan bisnis dan masalah yang ingin dipecahkan dengan tool yang akan dibuat. Pada proses ini kita harus mengidentifikasi stakeholder yang terlibat, kemudian melakukan pemahaman tujuan bisnis dari project ini, setelah itu kita melakukan analisis kebutuhan stakeholder, terutama pengguna tool ini, lalu kemudian. Tahap ini akan menjadi dasar yang kuat untuk melanjutkan ke tahapan berikutnya, yaitu Data Understanding.

- 2. Data Understanding**

Setelah tahap Business Understanding, kita tahap atau proses Data Understanding. Pada tahap Data Understanding, fokus utama adalah memahami data yang akan digunakan dalam pengembangan tool seperti ChatGPT. Langkah pertama adalah mengumpulkan data yang relevan dengan tujuan bisnis dan kebutuhan pengguna, dan untuk pembuatan tool seperti ChatGPT tentu data yang digunakan adalah data text. Setelah itu, data tersebut perlu dipahami lebih dalam melalui eksplorasi dan analisis. Ini mencakup pemahaman terhadap struktur data, jenis data yang tersedia, serta kualitas dan kebersihan data. Tahap ini menjadi dasar yang penting untuk langkah selanjutnya dalam proses CRISP-DM ini, yaitu Data Preparation.

- 3. Data Preparation**

Pada tahap Data Preparation, fokus utama adalah mempersiapkan data text agar siap untuk modelling. Pada tahap ini dilakukan pembersihan data text dengan mengidentifikasi dan menangani anomali-anomali pada data. Selanjutnya, data text tersebut diintegrasikan dan disesuaikan dengan diubah formatnya agar sesuai dengan kebutuhan analisis. Data text kemudian di-transmorm lebih lanjut hingga menjadi data yang terstruktur, transformasi ini melibatkan berbagai macam teknik text preprocessing seperti tokenisasi, text representation, dsb. Semua aktivitas pada tahapan ini dilakukan berdasarkan hasil dari tahap Data Understanding.

- 4. Modeling**

Setelah data disiapkan, langkah selanjutnya adalah menggunakan data tersebut untuk membuat model. Tahap awal adalah memilih model yang paling sesuai dengan kebutuhan tool yang akan dibuat. Untuk tool seperti ChatGPT, pilihan yang tepat adalah menggunakan transformer-based model, yang juga merupakan arsitektur yang digunakan oleh ChatGPT itu sendiri. Setelah model dipilih, langkah berikutnya adalah train model menggunakan data yang telah dipersiapkan sebelumnya. Tujuan dari training ini adalah agar model dapat mempelajari pola dan hubungan antar token/term antara satu sama lain. Setelah model berhasil di-train, tahap selanjutnya adalah melakukan fine-tuning dengan menerapkan reinforcement learning. Secara sederhana, pada fine tuning dengan reinforcement learning ini, kita akan mengajari tool kita secara manual dengan cara memberikan pertanyaan atau perintah, yang kemudian jika jawaban yang dihasilkan sesuai dengan harapan, kita memberikan reward, namun, jika jawabannya tidak sesuai, kita beri punishment, dengan begitu, model kita akan semakin pandai dalam memberikan jawaban yang sesuai dengan harapan.

## **5. Evaluation**

Setelah model berhasil dibuat, kita lanjut ke tahap Evaluation, pada tahap ini akan dievaluasi kinerja model yang telah dibuat dalam tahap sebelumnya. Untuk kasus ini cukup sederhana, kita tinggal menguji coba model yang sudah dibuat dan melihat respon jawaban yang diberikan, apakah telah memenuhi standar kualitas yang diharapkan dan sudah sesuai dengan kebutuhan dan harapan pengguna atau tidak.

## **6. Deployment**

Setelah tahap evaluasi selesai dan model terbukti memenuhi standar kualitas yang diharapkan, langkah selanjutnya adalah tahap deployment. Tahap ini bertujuan untuk menerapkan model yang telah dikembangkan ke dalam lingkungan produksi sehingga dapat digunakan oleh pengguna secara langsung. Pada tahap ini, akan dibuat API model yang sudah dibuat, lalu kemudian diintegrasikan ke sebuah platform entah dalam bentuk aplikasi ataupun website, sebagaimana website ChatGPT yaitu chat.openai.com. Selanjutnya, dilakukan uji coba dan validasi untuk memastikan bahwa API model dan platform tool tersebut berfungsi dengan baik dalam lingkungan produksi. Setelah itu, perlu juga dilakukan pemantauan dan pemeliharaan terhadap model yang sudah dideploy untuk memastikan performanya tetap optimal seiring berjalannya waktu. Tidak lupa juga untuk melakukan report terhadap hasil deploy dan pemantuan performa untuk dilakukan review lebih lanjut. Dengan melalui tahap deployment ini dengan cermat, kita dapat memastikan bahwa model ChatGPT yang telah dikembangkan dapat diakses dan digunakan oleh pengguna dengan lancar dan memberikan impact yang nyata sesuai dengan tujuan bisnis yang ditetapkan di tahap Business Understanding.

2. Apa yang dimaksud dengan text mining? Jelaskan apa kaitan antara text mining dengan data mining dan bagaimana pemanfaatan text mining pada ChatGPT?

### **Jawaban:**

Sebagaimana data mining adalah proses ekstraksi informasi berharga dari data, text mining adalah proses ekstraksi informasi yang bermanfaat dan bermakna dari data teks.

Contoh pengaplikasian text mining adalah untuk sentiment analysis, search engine, dan bahkan Large Language Model seperti ChatGPT dan Bard.

Text mining termasuk sebagai data mining, sementara data mining bukan berarti text mining. Text mining adalah data mining dengan text sebagai datanya. Ketika kita mendengar kata data mining, yang muncul di kepala kita umumnya adalah structured data yang berbentuk tabular, padahal, data mining juga melingkup unstructured data, salah satunya adalah text, dan itulah text mining.

Secara teknis, penerapan text mining pada ChatGPT melibatkan serangkaian langkah text processing dan analysis yang terstruktur, mulai dari tokenisasi, berbagai macam preprocessing, teknik-teknik text representation, kemudian dilakukan ekstraksi informasi seperti subject, verb, noun, dsb, hingga pengelompokan topik maupun sentiment analysis, semua dilakukan agar ChatGPT dapat menghasilkan respon yang sesuai dengan keinginan pengguna.

3. Tuliskan hal berikut pada ChatGPT: “Pertanyaan apa yang tidak bisa Anda jawab?” kemudian dari jawaban yang dihasilkan oleh ChatGPT, jelaskan mengapa tool ini tidak dapat menjawab pertanyaan tersebut?

**Jawaban:**

Berikut jawaban dari ChatGPT:

*“Saya tidak bisa memberikan jawaban yang pasti atau spesifik terkait masa depan atau informasi yang tidak tersedia dalam basis data saya. Saya juga tidak bisa memberikan saran medis, hukum, atau keuangan yang dapat diandalkan tanpa bantuan seorang profesional yang sesuai.”*

ChatGPT tentu tidak bisa memberikan jawaban yang pasti atau spesifik terkait masa depan sebagaimana manusia juga tidak bisa. Hal ini dikarenakan ChatGPT menjawab prompt berdasarkan basis data yang ia miliki sebagai mana perkataannya *“atau informasi yang tidak tersedia dalam basis data saya”*, hal ini masuk akal, sebagaimana manusia memiliki memori, AI juga memiliki data. Namun, berbeda dengan memori manusia yang selalu update secara realtime, memori atau data dari ChatGPT ini tidak update secara realtime. ChatGPT ditrain dengan data yang tersebar di internet yang hanya sampai pada January 2022, setidaknya untuk ChatGPT 3.5 yang saya gunakan. Maka dari itu ChatGPT juga tidak bisa menjawab informasi setelah January 2022 seperti pertanyaan “Lagu apa saja yang direlease drake di tahun 2023?” dan tentunya informasi relevan secara realtime seperti pertanyaan “Bagaimana cuaca hari ini di UI”?

4. Dari berbagai manfaat yang diperoleh dari penggunaan ChatGPT, terdapat isu-isu yang perlu menjadi pertimbangan dalam penggunaan tool tersebut, salah satunya adalah isu etika.

- a) Jelaskan isu etika yang mungkin terjadi dari penggunaan ChatGPT di lingkup akademik.

**Jawaban:**

Isu-isu terkait penggunaan ChatGPT di lingkup akademik antara lain adalah ketidakjujuran, ketergantungan yang tidak sehat, dan ketidaksetaraan. Pertama, hal ini merusak prinsip kejujuran dan integritas akademik karena termasuk bentuk kecurangan. Pelajar seharusnya mengerjakan tugas mereka secara mandiri agar pengetahuan dan skill mereka terasah. Tool seperti ChatGPT sebaiknya digunakan

untuk meningkatkan produktivitas dan efisiensi pembelajaran mereka, bukan untuk menggantikan tanggung jawab mereka. Ketika siswa menggunakan ChatGPT untuk menjawab tugas atau ujian tanpa usaha yang sebenarnya, mereka mengabaikan tanggung jawab pribadi mereka sebagai pelajar yaitu adalah belajar untuk memahami materi. Selain itu, penggunaan ChatGPT tentu juga akan menyebabkan ketergantungan, ketergantungan yang berlebihan pada ChatGPT dapat menghambat perkembangan intelektual mereka, karena mereka tidak mengembangkan keterampilan kritis atau pemahaman yang diperlukan, yang ada, malah model ChatGPT yang semakin pintar kedepannya. Di samping itu, penggunaan ChatGPT secara eksklusif sebagai alat belajar juga dapat memperkuat ketidaksetaraan, karena hanya siswa yang memiliki akses ke ChatGPT yang dapat memperoleh keuntungan dari keunggulan. Oleh karena itu, penting bagi lembaga pendidikan untuk menetapkan kebijakan dan pedoman yang jelas terkait penggunaan teknologi ini dalam proses pembelajaran, untuk memastikan kejujuran, integritas, dan keadilan di semua tingkatan Pendidikan

- b) ChatGPT adalah aplikasi yang dibuat oleh OpenAI, sebuah perusahaan yang berbasis di Amerika Serikat. Bagaimana isu budaya dan etika di Indonesia terhadap ChatGPT?

**Jawaban:**

Di Indonesia, penggunaan ChatGPT juga menimbulkan beberapa pertimbangan budaya dan etika. Pertama, soal pekerjaan, beberapa orang khawatir bahwa ChatGPT bisa mengambil pekerjaan manusia, khususnya yang butuh kecerdasan dan kreativitas. Hal ini berdampak pada lapangan kerja. Kedua, soal privasi data, ada kekhawatiran tentang bagaimana data pengguna dijaga. Diperlukan peraturan yang kuat untuk melindungi privasi dan keamanan data kita. Ketiga, dalam hal budaya lokal, Kita harus pastikan penggunaan ChatGPT menghormati dan mendukung keberagaman budaya Indonesia. Untuk mengatasi hal-hal ini, perlu ada regulasi yang tepat dan kesadaran tentang dampak penggunaan ChatGPT dalam pekerjaan, privasi data, dan karya budaya kita. Dengan begitu, penggunaan ChatGPT bisa memberikan manfaat yang positif dan sesuai dengan nilai-nilai budaya kita.

5. Suatu perusahaan menerapkan online technical assessment (contoh: Hackerrank, Codility) dalam proses rekrutmen kandidat karyawan IT. Dengan kehadiran ChatGPT, kandidat dapat menuliskan ulang soal dan ChatGPT dapat memberikan jawaban berupa code secara langsung. Apa yang perlu dilakukan oleh perusahaan dalam memilih platform technical assessment untuk lebih menjaga integritas?

**Jawaban:**

Terdapat beberapa upaya yang bisa dilakukan perusahaan untuk menangani hal ini, seperti peningkatan kompleksitas tes dan penerapan protokol pengawasan yang lebih ketat. Peningkatan kompleksitas tes bertujuan membuat soal menjadi lebih sulit untuk dijawab hanya dengan menyalinnya ke ChatGPT. Tes harus dibuat lebih kompleks hingga mengharuskan calon karyawan benar benar memahami konsep dan ilmu dibaliknya. Perusahaan juga bisa menguji kompleksitas tes mereka pada ChatGPT untuk melihat apakah ChatGPT dapat menjawab menyelesaikan tes tersebut dengan benar atau tidak. Dengan demikian, calon karyawan tidak dapat mengandalkan alat AI untuk memberikan jawaban. Selain itu, penerapan protokol pengawasan yang ketat dapat membantu meminimalkan risiko terjadinya kecurangan, termasuk penggunaan AI. Protokol ini bisa

mencakup penggunaan pengawasan online seperti live video dan live screen recording. Selanjutnya, perusahaan juga bisa melakukan tahapan lebih lanjut atas hasil tes yang telah dilakukan, seperti wawancara untuk mendapatkan penjelasan atas jawaban yang disubmit. Ini memungkinkan perusahaan untuk lebih memahami pemikiran dan pengetahuan calon karyawan serta mendeteksi adanya tanda-tanda kecurangan penggunaan AI. Dengan mengimplementasikan langkah-langkah ini, perusahaan dapat memastikan kejujuran dan keadilan dalam proses rekrutmen karyawan IT.

6. Mengapa terkadang ChatGPT dapat memberikan jawaban yang salah atau tidak sesuai harapan? Bagaimana solusi untuk menghindari hal ini?

**Jawaban:**

Penyebab utama ChatGPT memberikan jawaban yang salah adalah karena terbatasnya informasi/data yang dimilikinya, kurangnya informasi atau konteks yang diberikan, atau pertanyaan/perintah yang terlalu sulit/kompleks. Misalnya, jika ChatGPT tidak memiliki informasi/data tentang topik tertentu, maka jawabannya kemungkinan besar tidak akan akurat. Juga, jika kita tidak memberikan cukup informasi atau konteks dalam pertanyaan kita, jawaban yang diberikan oleh ChatGPT juga kemungkinan besar tidak akan sesuai dengan harapan kita. Terakhir, jika pertanyaannya terlalu rumit, ChatGPT kemungkinan besar tidak bisa menjawabnya dengan benar.

Terdapat beberapa solusi sederhana untuk menghindari hal-hal ini. Yang pertama tentu saja adalah tidak menanyakan hal yang berada di luar kapasitas ChatGPT, karena meskipun ChatGPT sangat canggih, ia tetap memiliki keterbatasan dalam memahami konteks atau topik tertentu. Yang kedua, pastikan untuk memberikan konteks yang cukup dan jelas dalam pertanyaan atau perintah kita kepada ChatGPT. Semakin jelas konteksnya, semakin akurat dan sesuai harapan jawaban yang diberikan oleh ChatGPT. Selain itu, kita juga bisa menjelaskan kepada ChatGPT mengenai bentuk jawaban yang diinginkan. Yang ketiga, jika pertanyaan atau perintah terlalu sulit, cobalah untuk membaginya menjadi bagian-bagian/perintah-perintah kecil. Sebagai contoh, daripada menyuruh ChatGPT untuk membuat website secara lengkap, lebih baik kita menanyakan kepada ChatGPT langkah-langkah apa yang dibutuhkan untuk membuat sebuah website, lalu meminta bantuan ChatGPT untuk setiap langkah tersebut.

## REFERENSI

Semrush. (2022, Januari 5). How Does ChatGPT Work? Retrieved February 10, 2024, from <https://www.semrush.com/blog/how-does-chatgpt-work/#how-was-chatgpt-trained>

Tech Business News. (n.d.). ChatGPT May Lead to the Downfall of Education and Critical Thinking. Retrieved February 10, 2024, from <https://www.techbusinessnews.com.au/blog/chatgpt-may-lead-to-the-downfall-of-eduction-and-critical-thinking/#:~:text=Encourages%20academic%20dishonesty%3A%20Students%20could,thinking%20and%20problem%2Dsolving%20skills.>