

Data Science Use Case in Healthcare Industries

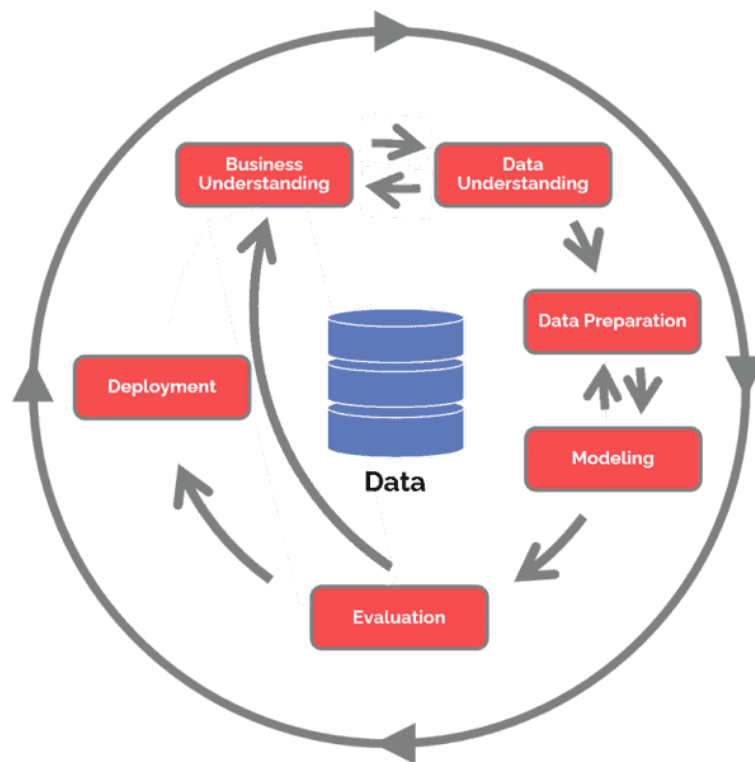
Indonesia merupakan negara dengan jumlah penduduk terbanyak keempat di dunia. Namun, layanan kesehatan yang tersedia tidak sebanding dengan jumlah penduduk. Hal ini terbukti bahwa jumlah dokter di Indonesia menduduki peringkat kedua terendah di Asia tenggara dengan rasio 0,4 dokter per 1.000 penduduk (Jayani, 2020). Nilai ini masih cukup jauh dengan rekomendasi rasio oleh WHO yaitu 1 dokter per 1.000 penduduk. Oleh karena itu, perkembangan industri kesehatan di Indonesia cukup potensial khususnya untuk membantu meringankan beban dan mempermudah pekerjaan dokter sembari menunggu lulusan dokter-dokter baru untuk memenuhi target rasio di masa yang akan datang.

Kita—sebagai calon *engineer*—dapat turut membantu untuk menjawab permasalahan di atas dengan mengembangkan dan memaksimalkan potensi dari sektor industri kesehatan menggunakan bidang kita. Kita akan menggunakan pendekatan menggunakan *data science* untuk mengembangkan layanan kesehatan yang dapat digunakan oleh penduduk Indonesia. *Use case* yang akan kita tinjau menggunakan pendekatan CRISP DM adalah penggunaan *data science* dalam *virtual assistant* di sektor *healthcare*.

Sejak pandemi Covid-19, dimana mobilitas manusia menjadi terbatas, penggunaan telemedicine kian melejit. Hal ini dikarenakan penggunaannya yang cepat dan tidak perlu datang ke dokter secara langsung sehingga dapat mengurangi risiko penyebaran virus. Dalam pengembangannya, dibutuhkan *artificial intelligence* untuk menghasilkan inovasi-inovasi terbaru. Salah satunya adalah penggunaan *virtual assistant* yang saat ini dapat digantikan keberadaannya menggunakan AI.

Intelligent Virtual Assistant (IVA) berbeda dengan *chatbot* pada umumnya. Jika *chatbot* yang sering kita temui merupakan *rule-based artificial intelligent*, maka IVA tidak. IVA lebih terlihat manusiawi dan memiliki empati karena sifatnya yang intuitif, natural, dan *conversational* (Treloar, 2022). Dalam penerapannya, pemanfaatan IVA cukup banyak manfaatnya. Salah satunya adalah IVA mampu mencari database penting dari pasien. Jika suatu saat terdapat satu kasus yang sama, IVA mampu mencari informasi/kasus yang serupa, artikel kesehatan yang relevan, dan beberapa jurnal dari ahlinya (Dreyer, 2013). Oleh karena itu, dibutuhkanlah data (rekam medis, jurnal, artikel, dll) yang sangat banyak untuk membuat sebuah database sehingga dapat digunakan oleh IVA. Untuk mengelola data tersebut hingga akhirnya pengguna dapat mengerti dengan jelas, dibutuhkanlah peran *data scientist* untuk mengolah dan memvisualisasikannya.

Untuk membahas lebih lanjut mengenai IVA, kita akan meninjau IVA menggunakan metode pendekatan CRISP DM (Cross-Industry Standard Process for Data Mining). Secara singkat, berikut adalah gambar diagram dan tahapan dari CRISP DM:



Penjelasan dari masing-masing tahapan CRISP DM adalah sebagai berikut (Mauritsius dan Binsar, 2020)

1. Business understanding & Data understanding
Business understanding merupakan langkah pertama dalam *framework* CRISP DM. Pada tahap ini, kita menganalisis dari segi bisnis: bagaimana kita dapat mendapatkan data dan bagaimana data-data tersebut cocok dengan tujuan bisnis sehingga model terbaik dapat dibangun. Untuk IVA, *business objective* yang ingin didapatkan adalah mampu mendapatkan *customer* dengan memberikan pelayanan yang ramah dan akurat. Selanjutnya, proses berikutnya yang berjalan secara berdampingan adalah *data understanding*. Adapun data yang dibutuhkan adalah rekam medis pasien, jurnal dan artikel yang relevan dengan kondisi pasien, dan frasa-frasa dari percakapan pasien melalui kolom *chat* yang akan digunakan oleh IVA untuk belajar memahami emosi manusia berdasarkan tulisan.
2. Data preparation
Langkah berikutnya adalah *data preparation*, dimana akan dipilih kasus dan parameter yang akan dianalisis (*select data*), ditransformasikan terhadap parameter tertentu (*transformation*), dan dilakukan pembersihan data untuk persiapan tahap *modelling* (*cleaning*). Jika data-data yang diperoleh terdapat banyak kekosongan/kesalahan, maka akan terjadi proses *cleaning*. Pada IVA, semua frasa/kalimat yang masuk melalui *chat* akan dimasukkan ke database. Namun, data tersebut tetap perlu ditinjau kembali apakah frasa/kalimat yang dimasukkan etis atau tidak. Dengan adanya proses *cleansing*, maka frasa yang terekam oleh IVA akan lebih etis sehingga dapat memberikan output yang baik juga.
3. Modelling

Kemudian, dilakukan proses *modelling* yang dapat dibagi menjadi dua yaitu prediktif atau deskriptif. Pada tahapan ini, dilakukan penentuan teknik dan algoritma *data mining* menggunakan metode statistika dan *machine learning*. Untuk dapat memahami bahasa manusia, IVA menggunakan *modelling* tipe *classification* dan *language models* (Vainu, 2021). Model *classification* dibutuhkan untuk mengkategorikan frasa dan kata-kata yang masuk melalui *chat* dari pengguna. Disamping itu, *language models* berguna bagi IVA untuk belajar dan memprediksi kata apa yang mungkin akan muncul berikutnya. Dengan menggunakan *language model*, IVA akan belajar untuk menemukan pola bagaimana menyusun suatu kalimat yang baik.

4. Evaluation

Selanjutnya, dilakukan interpretasi dari hasil pemodelan sebelumnya. Evaluasi akan menilai apakah model yang dipilih telah sesuai dengan tujuan yang diharapkan sesuai pada tahapan *business understanding*

5. Deployment

Terakhir, dilakukan *deployment* atau rencana penggunaan model. Pada proses *deployment*, IVA akan dihadapkan langsung oleh pengguna dan berinteraksi bersama mereka. Dengan berinteraksi terus menerus, IVA akan belajar frasa dan kalimat dari pengguna untuk terus berkembang di kemudian hari. Dengan metode *machine learning*, IVA akan memahami berbagai macam frasa dan bagaimana cara menjawabnya

Github Repository : [angelineshane/zentask1 \(github.com\)](https://github.com/angelineshane/zentask1)

Daftar Referensi

- Dreyer, Jonathan. (2013). *Intelligent virtual assistants*. Diakses pada 17 September 2022, dari <https://healthcare-in-europe.com/en/news/intelligent-virtual-assistants.html>
- Jayani, Dwi Hadya. (2020). *Rasio Dokter Indonesia Terendah Kedua di Asia Tenggara*. Diakses pada 17 September 2022, dari <https://databoks.katadata.co.id/datapublish/2020/04/02/rasio-dokter-indonesia-terendah-kedua-di-asia-tenggara>
- Kosouvorova, Elena. (2022). *Data Science Use Cases Guide*. Diakses pada 17 September 2022, dari <https://www.datacamp.com/blog/data-science-use-cases-guide>
- Mauritsius, T. & Binsar F. (2020). Cross-Industry Standard Process for Data Mining (CRISP-DM). Diakses pada 17 September 2022, dari <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/>
- Treloar, Nate. (2022). *When Automation Meets Empathy: Intelligent Virtual Assistants for Healthcare*
- Vainu, Indrek. (2021). *What is An Intelligent Virtual Assistant?*. Diakses pada 17 September 2022, dari <https://www.alphachat.ai/blog/what-is-an-intelligent-virtual-assistant>