

GoTo Data Science Conference 2022: Taking Data Science to New Heights with Communities and Industries

Senin, 17 Oktober 2022.

<https://sharingvision.com/2022/goto-data-science-conference-2022-taking-data-science-to-new-heights-with-communities-and-industries/>

Kemudahan berbagai informasi di masa sekarang, membuat *data science* semakin dikenal. *Data science* berfungsi untuk mengolah kumpulan data besar baik yang terstruktur atau tidak menjadi informasi, lalu pengetahuan yang berguna untuk mencapai suatu target. Olahan data menjadi kebutuhan utama untuk menciptakan strategi bisnis.

Berdasarkan kebutuhan tersebut, GoTo pada hari Sabtu, 26 Februari 2022, melaksanakan konferensi pertama mengenai *Data Science* dengan tema “*Taking Data Science to New Heights with Communities and Industries*”. Secara umum konferensi yang berlangsung selama 2 hari ini membahas mengenai :

1. *Tools* Kausalitas dalam *Machine Learning* untuk meningkatkan kualitas prediksi dan putusan
2. Peran komunitas dalam menaikkan level *Data Science*
3. Bagaimana GoTo memanfaatkan *Data Science* dalam berbagai layanannya
4. Kode Generasi pada Pemrograman Kompetitif dengan *AlphaCode*



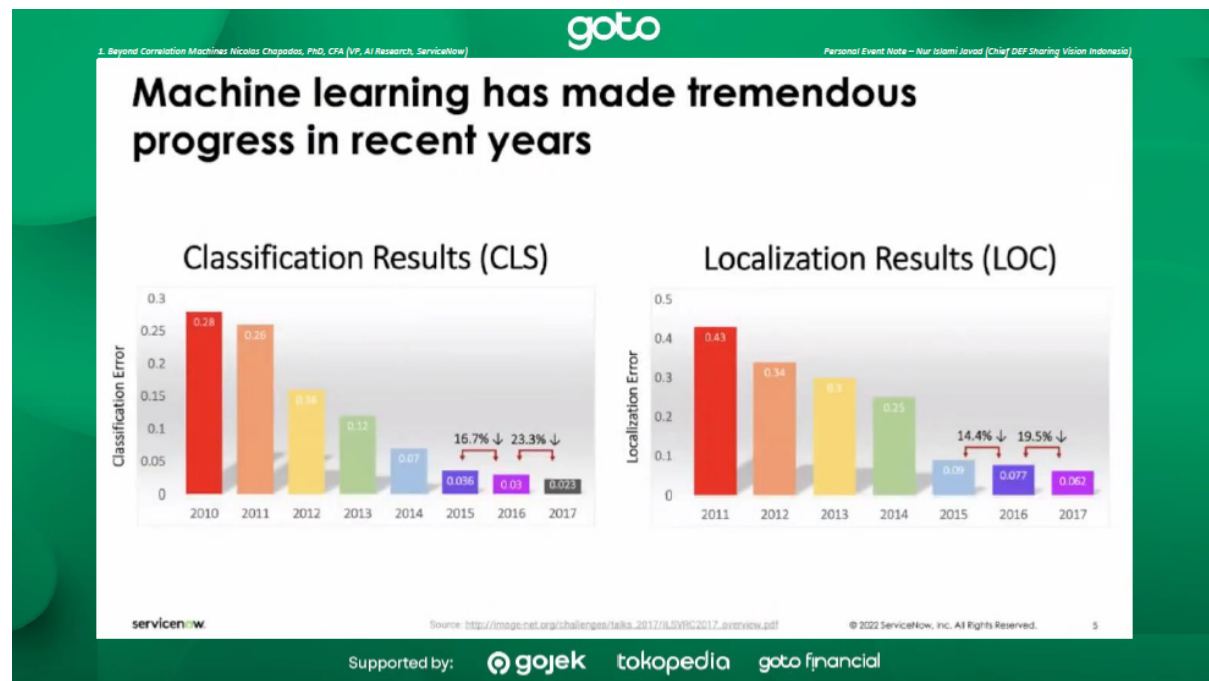
The banner for the GoTo Data Science Conference 2022 features a green background with white and yellow text. On the left, it says 'goto' in white, followed by 'Data Science Conference 2022' in large white letters. Below this, it says 'Take part in shaping the future of Data Science with our leading experts!'. The dates '25th FEB 09:30-11:30 WIB' and '26th FEB 13:30-18:30 WIB' are shown with a large yellow ampersand between them. A yellow button says 'REGISTER NOW !' with the link 'bit.ly/goto-ds-conf22' below it. On the right, it says 'Supported by:' followed by logos for 'gojek', 'tokopedia', and 'goto financial'. Below this are six circular portraits of speakers: Nicolas Chapados (Guest, ServiceNow), Priyanshu Jain (GOJEK), Nitin Handeja (GOJEK), Nando de Freitas (Guest, DeepMind), Gaurav Khanna (Tokopedia), and Hendra Chori (GoTo Financial). At the bottom right, it lists 'Fiona Pan | Syafri Bahar | Dita Amallya | Sengmeng Koo | Quinn Pham', 'Welcome Speech by Andre Soelistyo | Catherine H. S.', and 'CTO Remarks by Severan Rault | Herman Widjaja'.

Causal Inference, Kunci Utama Machine Learning

Pemanfaatan teknologi informasi oleh masyarakat, menyebabkan tersedianya data yang melimpah. Dalam hal mengolah data, yang menjadi pertanyaan adalah “Bagaimana kita dapat memodelkan, memprediksi, dan mendorong perilaku dan interaksi dalam skala besar lautan data?” Menjawab hal tersebut pembicara tamu,

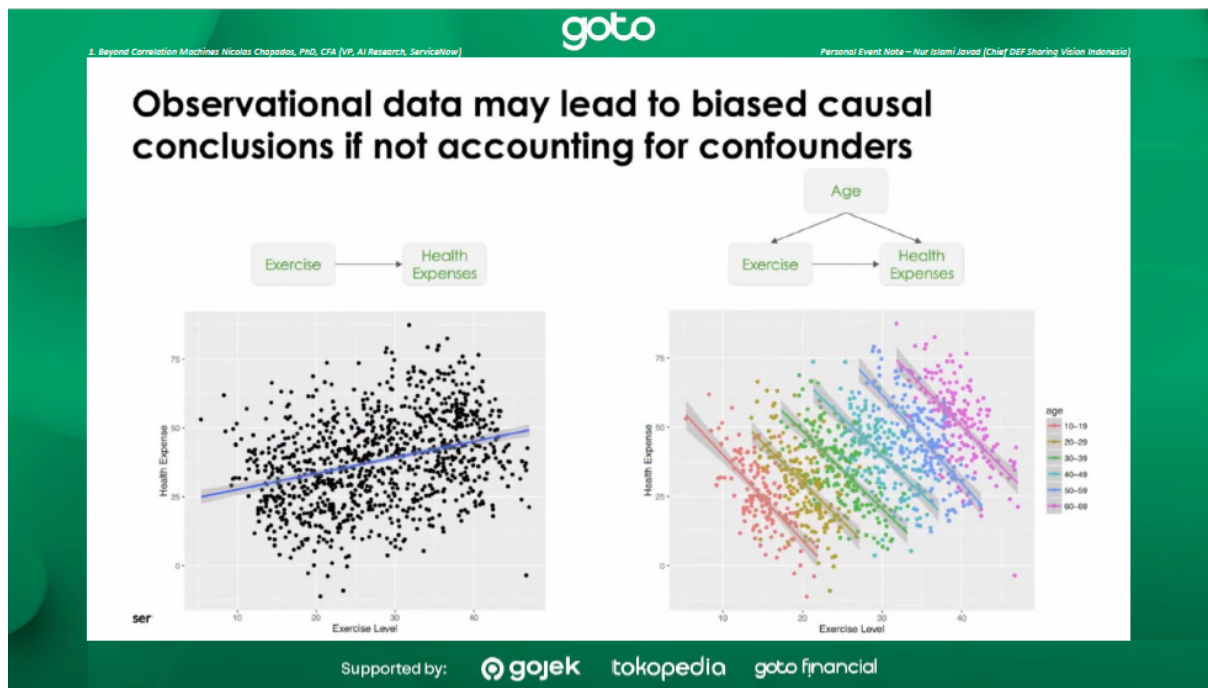
Nicholas Chapados, sebagai VP, AI Research, Service Now, menjelaskan mengenai peran *Machine Learning*.

Machine Learning telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Progres dari arsitektur *deep neural network* dan *transformer* menjadi pendorong utama akselerasi progres tersebut.



Gambar 2. Perkembangan *Machine Learning*

Dalam dunia bisnis, pertanyaan mengenai prediksi dan kemungkinan sering diajukan untuk menentukan strategi bisnis. Jawaban tersebut dapat diberikan dengan mengolah data melalui *Machine Learning*. Namun, yang perlu menjadi catatan dalam penggunaan *Machine Learning* adalah memisahkan antara Kausalitas dan Korelasi. Menerapkan analisis kausalitas dalam data observasi diperlukan untuk menemukan kausalitas (*causal discovery*) agar terhindar dari konsep bias dan kesimpulan yang salah.



Gambar 3. Bias kesimpulan kausalitas

Akselerasi Data Scientist melalui komunitas

Perkembangan dari *Data Science* juga selaras dengan kebutuhan Sumber Daya Manusia yang ahli pada bidang tersebut. Atas dasar tersebut, diperlukan dukungan dari berbagai pihak untuk dapat mengakselerasi *Data Scientist* yang kompeten. Dalam konferensi di hari kedua, beberapa komunitas telah memaparkan upaya yang dilakukan untuk memenuhi kebutuhan yang ada.

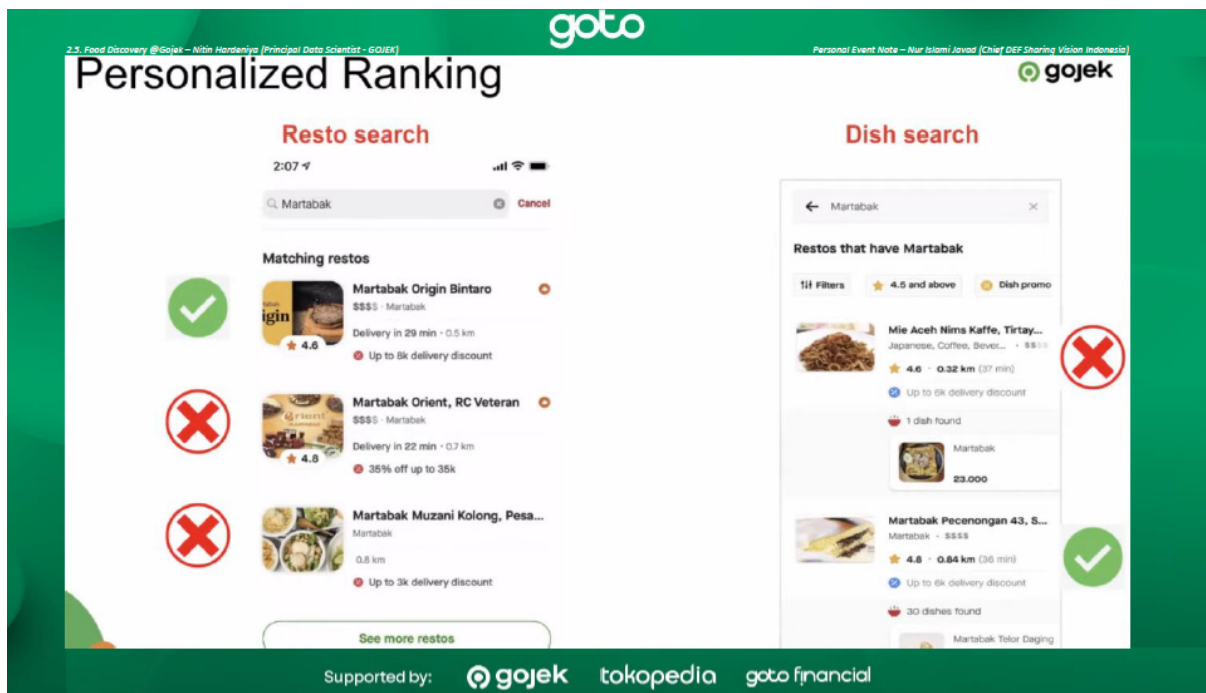


Gambar 4. Skala dampak *She Loves Data*

She Loves Data, yang telah memiliki 17 *chapter* tersebar di berbagai negara, telah memberikan pelatihan kepada 15000 lebih perempuan dalam rangka mempersiapkan perempuan sebagai penggerak revolusi teknologi dan data. Lebih lanjut, *AI Professionals Association* juga turut serta dalam memberikan standar kompetensi dalam berbagai level untuk profesi AI dan ML. Terakhir, Yayasan *Data Science* atau lebih dikenal *Data Science Indonesia* juga memiliki visi untuk menciptakan ekosistem kolaboratif terkait teknologi data untuk kemajuan bersama.

Data Science dan kepuasan pelanggan GoTo

Dalam rangka komitmen memberikan pelayanan terbaik, GoTo juga telah memanfaatkan data science dalam menghilangkan hambatan yang ada. Salah satu hambatan yang dapat diselesaikan oleh Gojek adalah pengalaman *advertising* yang tidak efisien karena harga statis. Lebih lanjut Gojek juga berhasil menyediakan rekomendasi makanan yang paling relevan dengan keinginan pengguna. Hasil pencarian tersebut juga dapat dipilah berdasarkan beberapa kategori dan faktor kemiripan.



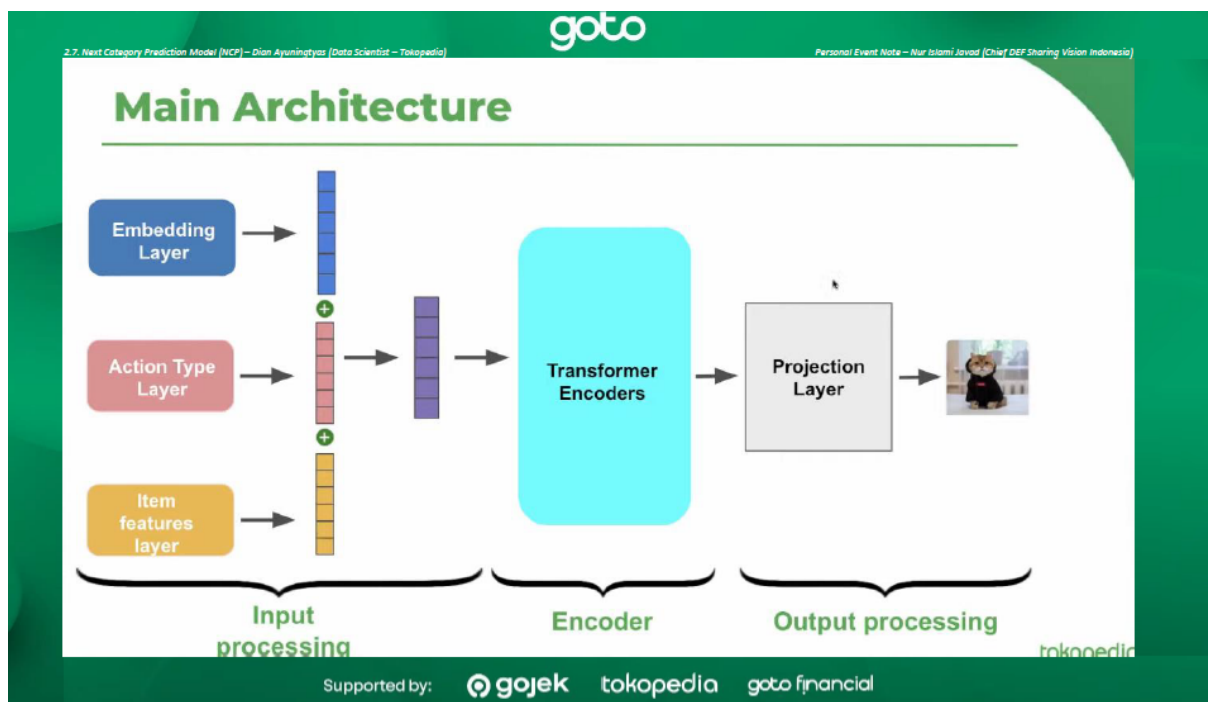
Gambar 5. Personalisasi *Ranking* Gofood pada Gojek

Mirip dengan tujuan Gojek, Tokopedia juga berhasil mempersonalisasi pengalaman pencarian pelanggan dengan pemanfaatan *deep machine learning*. Rekomendasi produk berdasarkan hasil pencarian pelanggan tidak hanya terbatas pada konten dan faktor kemiripan, tetapi juga memberikan rekomendasi lanjutan berdasarkan prediksi kebutuhan setelah pencarian produk awal dilakukan.



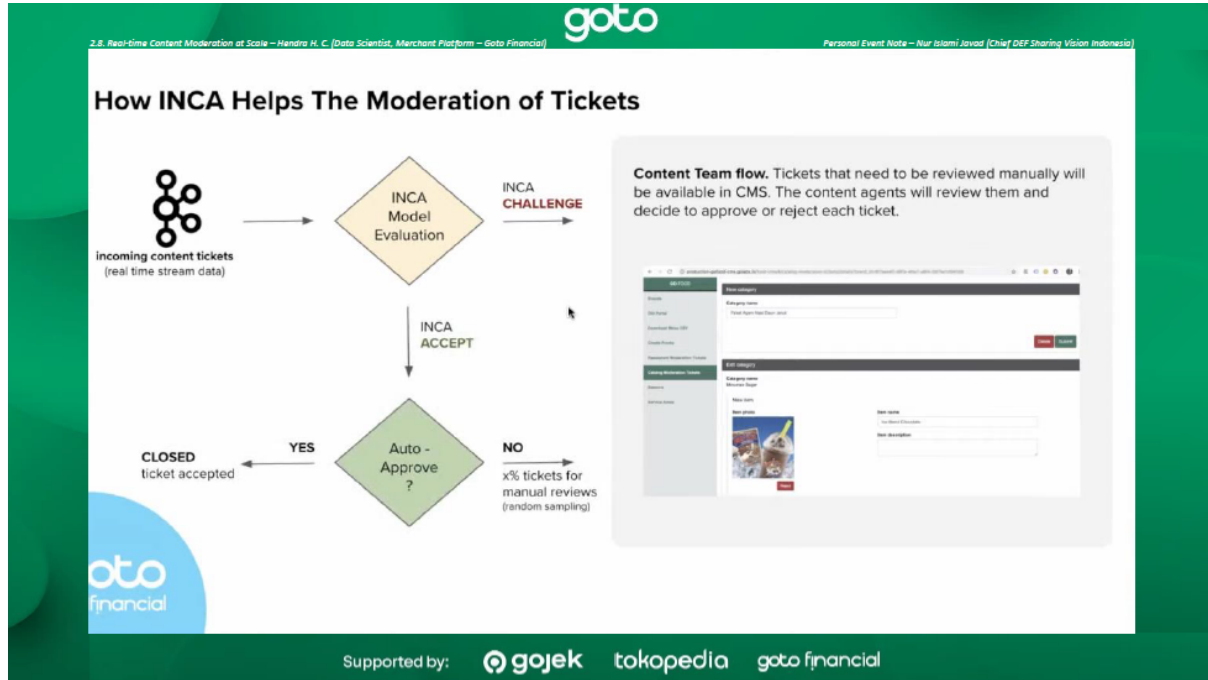
Gambar 6. *Sequential Learning* Tokopedia berdasarkan prediksi kebutuhan

Rekomendasi yang diberikan oleh Tokopedia pun sangat variatif. Dengan *Next Category Prediction Model*, Tokopedia dapat memiliki keragaman prediksi, dimana hanya 20,09 % pengguna mendapat rekomendasi produk yang belum pernah mereka beli sebelumnya dan 10,69 % mendapatkan rekomendasi yang belum pernah mereka hubungi sebelumnya.



Gambar 7. *Main Architecture Next Category Prediction Model* Tokopedia

Hambatan lain baik dari Gojek maupun Tokopedia adalah gambar produk yang di unggah oleh pelaku usaha sering juga tidak sesuai atau pun tidak pantas. Melalui *Inappropriate Content Auditor* (INCA), GoTo financial dapat memfilter 75% konten-konten tersebut secara otomatis dengan 99% presisi dan menghemat 2.600 jam kerja manusia per bulan.

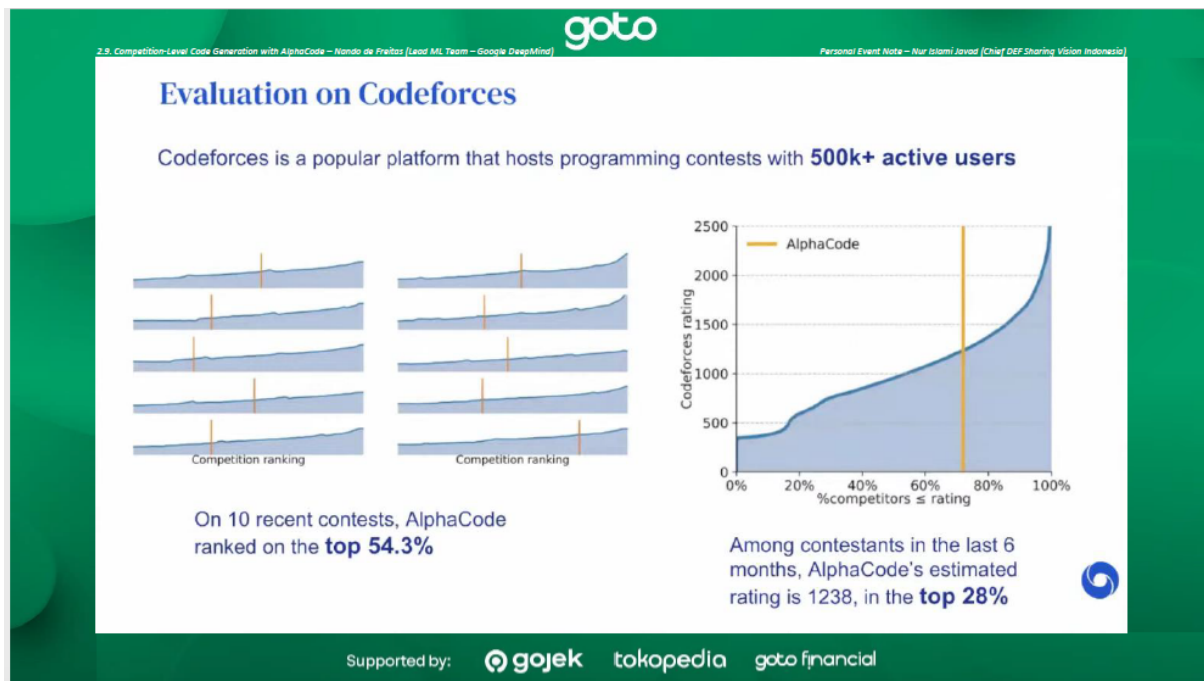


Gambar 8. INCA solusi filter konten dengan cepat dan presisi

Alphacode, solusi pada tingkat kompetitif

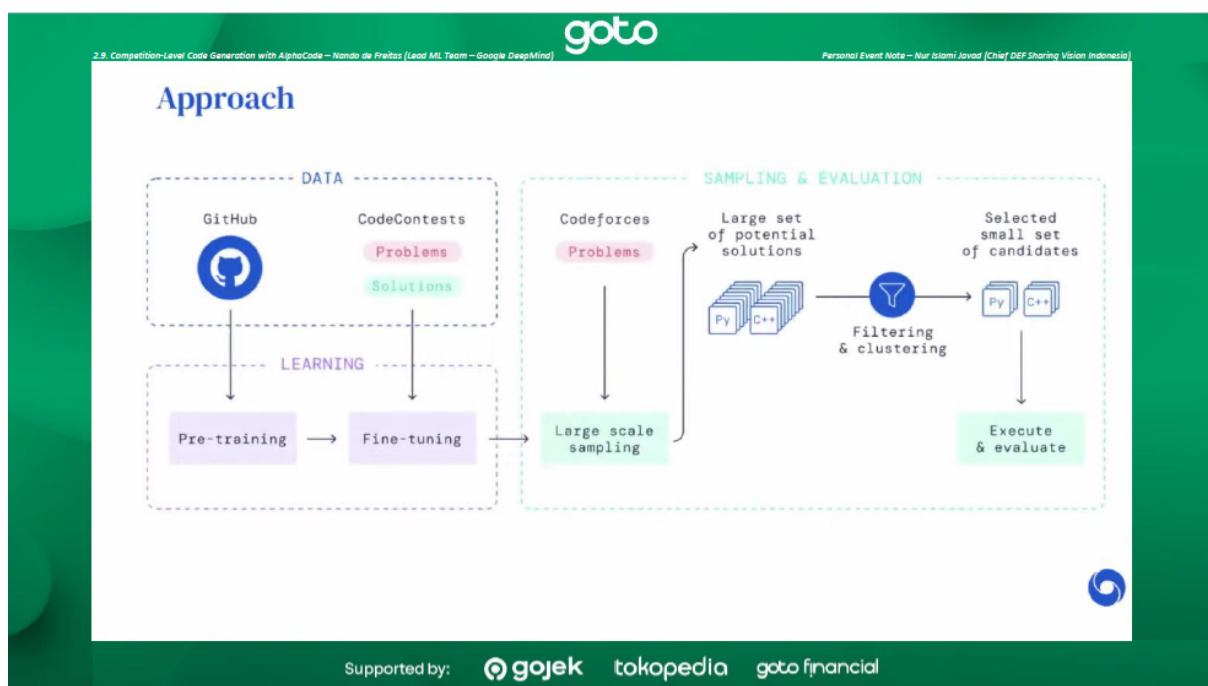
Walaupun *Machine Learning* telah mengalami progres yang signifikan dalam menghasilkan dan mengolah data, namun kemajuan dalam pemecahan masalah belum dapat menyelesaikan masalah sederhana pada matematika simpel dan masalah pada programming. *DeepMind* melalui *AlphaCode* berupaya untuk menyelesaikan masalah tersebut pada programming tingkatan kompetitif.

Nando de Freitas sebagai Scientis Leading the ML Team Goggle DeepMind, menjelaskan bahwa *AlphaCode* menggunakan *transformer-based language models* untuk menghasilkan kode pada skala yang belum pernah terjadi sebelumnya, dan secara cerdas memfilter kode tersebut pada kumpulan kecil program yang menjanjikan. Pada kompetisi *programming* yang diselenggarakan oleh *Codeforces*, *AlphaCode* berhasil menduduki posisi teratas 54% dari peserta.



Gambar 9. Hasil kompetisi *programming Codeforce*

AlphaCode dikembangkan pada programming kompetitif karena pada level ini diuji kemampuan memahami bahasa yang dapat memberikan progres signifikan untuk AI dalam menerjemahkan, meningkatkan *neural architecture*, meningkatkan skala dan mengurangi energi yang kita perlukan dalam program ini.



Gambar 10. Model Pendekatan *AlphaCode*

Pemaparan lebih lanjut pada Konferensi dapat dilihat pada tautan berikut:

Hari ke-1

https://www.youtube.com/watch?v=_x9Ccmh-Gvk&t=3392s

Hari ke-2:

<https://www.youtube.com/watch?v=0sEIXYUQ-wI>