



## Menginterpretasikan Model Machine Learning

Model *machine learning* bisa memberikan akurasi prediksi lebih baik dari model-model sederhana seperti regresi linear. Namun, interpretasi dari model *machine learning* tidaklah mudah. Berikut tiga hal sederhana yang bisa membantu kita menginterpretasikannya.

PENULIS:



**BAGUS SARTONO**  
Dosen di Departemen Statistika  
Institut Pertanian Bogor dan  
Pengurus Ikatan Statistisi  
Indonesia



**YUNANTO PUTRANTO**  
Business Analyst di Hassad Qura

**TULISAN INI** kami susun untuk menanggapi beberapa kolega yang merespons tulisan di *InfoKomputer* bulan Oktober 2018 lalu dengan judul "Persimpangan antara Interpretasi dan Ketepatan Prediksi". Di antaranya menanyakan bagaimana persisnya kompromi pilihan antara penggunaan model sederhana seperti regresi linear atau regresi logistik dengan implementasi model

*machine learning* yang rumit dan cenderung berupa kotak hitam.

Mengulas kembali yang telah dipaparkan sebelumnya, model-model statistika sederhana menarik digunakan karena menawarkan interpretasi pengaruh dari variabel prediktor dengan sangat mudah, termasuk menentukan variabel mana yang memiliki pengaruh lebih besar dibandingkan yang lain. Namun, banyak kajian memberikan hasil bahwa model-model seperti

regresi linear dan regresi logistik, kalah dalam hal akurasi prediksi dibanding model-model rumit seperti *machine learning*.

Pendekatan seperti *random forest*, *boosting*, dan *support vector machine* memberikan keleluasaan penanganan terhadap pola nonlinear dalam data sehingga akurasi sangat memuaskan. Tentu saja, ini kemudian dibayar dengan kerumitan bentuk model yang pada akhirnya banyak orang memandangnya sebagai kotak hitam. Pengguna menerima saja hasil prediksinya tanpa perlu menjelaskan yang terjadi di dalamnya.

### Tuntutan Interpretasi Masih Besar

Kondisi ini tidak cukup memuaskan karena tuntutan terhadap interpretasi dan penjelasan model masih sangat besar di banyak industri. Pada bidang manajemen risiko perbankan misalnya, regulator sangat memerlukan interpretasi dan penjelasan sebelum suatu model *scoring* maupun rating bagi debitur diterapkan bank. Demikian pula halnya dengan kebutuhan perencanaan strategis yang memerlukan penjelasan mengenai bagaimana variabel prediktor bekerja, seperti arah hubungan antarvariabel.

Lalu apa yang bisa dikerjakan? Berikut ini tiga hal sederhana yang bisa membantu kita menginterpretasikan model *machine learning* yang digunakan.

### 1 Mengenali Tingkat Kepentingan Variabel Prediktor

Pertanyaan umum pertama, dari sekian banyak variabel prediktor yang terlibat, mana





sajakah yang penting dalam pemodelan? Ada dua cara umum yang dapat dikerjakan. Pertama, membandingkan tingkat akurasi antara model yang melibatkan seluruh variabel prediktor (*full model*) dengan model yang tidak menyertakan variabel prediktor yang sedang dievaluasi (*reduced model*). Makin tinggi perbedaan akurasi antara *full model* dan *reduced model*, mengindikasikan makin pentingnya variabel prediktor yang tidak disertakan tersebut.

Kedua, melakukan permutasi nilai-nilai dari variabel prediktor yang sedang dievaluasi. Nilai dari variabel prediktor itu diatur sedemikian rupa sehingga nilai dari suatu amatan diberikan ke amatan lain. Akurasi dari model yang variabelnya mengalami permutasi dibandingkan terhadap akurasi dari model dengan data aslinya. Makin besar perbedaannya, makin penting variabel prediktor tersebut. Jika proses tersebut dilakukan untuk setiap variabel prediktor, dapat diperoleh berapa penurunan akurasi masing-masing. Hal ini kemudian bisa digunakan untuk mengetahui tingkat kepentingan variabel prediktor dan menentukan mana yang paling penting. Penurunan tingkat akurasi dapat juga diganti dengan penurunan nilai Gini.

## 2 Mengenali Pola Umum Hubungan antara Variabel Prediktor dengan Variabel Target

Mengenali pola umum hubungan antara variabel prediktor dengan variabel target, dapat diwakili dengan hubungan antara variabel prediktor dengan skor hasil prediksi.

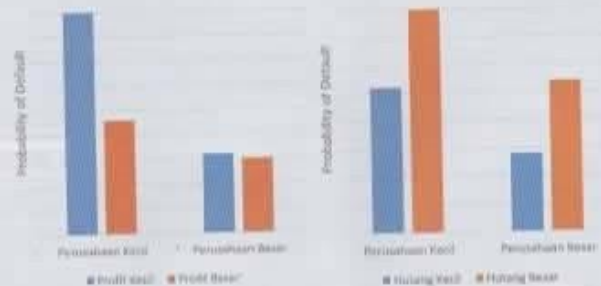
Langkah pertama, gunakan model untuk menghasilkan skor

prediksi untuk setiap amatan pada *data training*. Berikutnya, lakukan *discretization* terhadap variabel prediktor yang akan dilihat hubungannya. Terakhir, hitung nilai rata-rata skor prediksi untuk masing-masing *bin* hasil *discretization*. Pola menurun, menaik, atau kurva tertentu umumnya akan mudah dikenali.

Proses ini barangkali tidak perlu dilakukan pada seluruh variabel prediktor. Yang utama dikerjakan adalah untuk variabel prediktor penting yang telah dikenali sebelumnya. Pada variabel prediktor yang tingkat kepentingannya rendah, pola hubungan akan tidak terlihat dan rata-rata skor prediksi bernilai mirip/sama untuk berbagai *bin* variabel prediktor.

## 3 Mengidentifikasi Keberadaan Interaksi

Interaksi adalah suatu kondisi adanya pengaruh suatu variabel prediktor terhadap variabel prediktor lain. Gambar 1a memperlihatkan untuk perusahaan kecil, keuntungan yang meningkat diikuti dengan menurunnya *probability of default* (PD). Namun tidak demikian pada perusahaan besar, keuntungan yang meningkat cenderung tidak mengubah PD-nya. Ini berarti, pengaruh keuntungan terhadap PD tergantung pada ukuran perusahaan. Jadi, ada interaksi antara keuntungan dengan ukuran perusahaan. Lain halnya dengan pengaruh hutang terhadap PD yang digambarkan pada Gambar 1b. Hutang yang makin besar mengakibatkan PD yang makin besar pula, baik pada perusahaan kecil maupun besar. Artinya tidak ada interaksi antara hutang dengan ukuran perusahaan.



**Gambar 1a.** Ilustrasi keberadaan interaksi antara keuntungan dan ukuran perusahaan.

**Gambar 1b.** Ilustrasi ketiadaan interaksi antara hutang dan ukuran perusahaan.



**Gambar 2a.** Ilustrasi keberadaan interaksi antara keuntungan dan ukuran perusahaan pada *tree*.

**Gambar 2b.** Ilustrasi ketiadaan interaksi antara hutang dan ukuran perusahaan pada *tree*.

Jika model *machine learning* sudah diperoleh, interaksi dapat diidentifikasi melalui dua langkah berikut. Pertama, lakukan prediksi terhadap *data training* menggunakan model yang telah diperoleh sehingga untuk setiap amatan dihasilkan skor tertentu. Kedua, terapkan teknik *decision tree* dengan respons berupa skor prediksi. Interaksi dapat diidentifikasi dari *tree* jika variabel pemisah di bagian kiri berbeda dengan variabel pemisah di sebelah kanan. Gambar 2a dan Gambar 2b adalah representasi sederhana dari kondisi keberadaan dan ketiadaan interaksi yang serupa dengan Gambar 1a dan Gambar 1b.