

# LECTURE NOTES

## ISYS8036 - Business Intelligent and Analytics

### Topic 9

#### VISUALISASI PERFORMANS MODEL

# LEARNING OUTCOMES

Setelah mempelajari materi ini peserta kuliah diharapkan mampu mengidentifikasi dan memahami:

- Visualization of model performance under various kinds of uncertainty;
- Further consideration of what is desired from data mining results.
- Profit curves;
- Cumulative response curves;
- Lift curves;
- ROC curves.

## OUTLINE MATERI :

1. Klasifikasi dan perbandingan
2. Kurva Profit
3. Grafik dan Kurva ROC
4. Area Under the ROC Curve (AUC)
5. Kurva Respons Cumulative Dan Kurva Lift
6. Kesimpulan

# PENDAHULUAN

Perhitungan profit dengan Persamaan 7-2 memperhitungkan sejumlah kondisi spesifik untuk menghasilkan satu angka, yang menunjukkan keuntungan yang diharapkan dalam skenario itu. Stakeholders di luar tim data sains mungkin tidak memiliki kesabaran untuk memperhatikan detail, dan sering menginginkan tampilan kinerja model di level yang lebih tinggi. Kendati para ilmuwan data yang cukup merasa nyaman dengan persamaan dan perhitungan matematis sering menemukan perkiraan perkiraan yang dihasilkan terkesan miskin dan tidak informatif. Karena itu seringkali dirasa perlu untuk menyajikan visualisasi bukan hanya menyajikan perhitungan perhitungan yang kelihatannya tidak informatif.

## Perankingan dan Klasifikasi

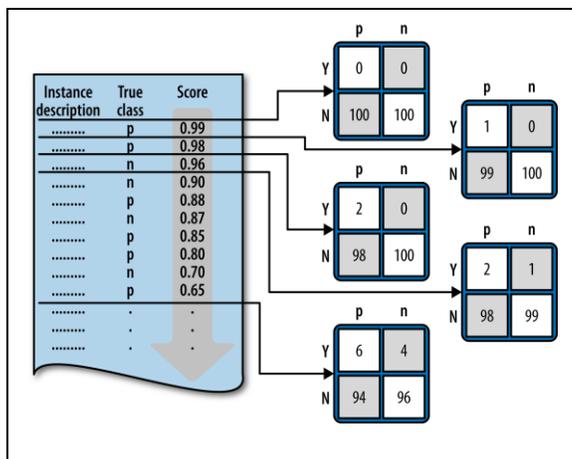
Telah dibahas bagaimana skor yang ditetapkan model dapat digunakan untuk menghitung keputusan untuk setiap kasus individual berdasarkan nilai yang diharapkan. Suatu strategi yang agak berbeda dalam membuat keputusan adalah dengan memberi peringkat (rank) pada satu himpunan kasus dengan menggunakan skor ini, dan kemudian mengambil suatu tindakan pada kasus-kasus yang berada di rank lebih rendah (atas). Daripada memperlakukan kasus secara individual, mungkin lebih baik mengambil  $n$  kasus teratas dari suatu ambang batas (threshold) memutuskan untuk mengambil kasus teratas (atau, dengan kata lain, semua kasus yang mendapat skor di atas ambang batas tertentu). Ada beberapa alasan praktis dalam melakukan ini.

Ketika memiliki anggaran terbatas untuk suatu tindakan, seperti anggaran pemasaran, jelas aksi perlu ditunjukkan pada kandidat yang paling menjanjikan. Jika seseorang akan menargetkan kasus dengan nilai ekspektasi tertinggi menggunakan perhitungan biaya dan manfaat yang konstan untuk setiap kelas, maka melakukan peringkat kasus dengan kemungkinan kelas target sudah cukup.

Ketika bekerja dengan pengklasifikasi yang memberikan skor ke instance, dalam beberapa situasi keputusan pengelompokan harus sangat konservatif, sesuai dengan fakta bahwa classifier harus memiliki kepastian tinggi sebelum mengambil tindakan positif. Ini sesuai dengan

menggunakan ambang tinggi pada skor output. Sebaliknya, dalam beberapa situasi, classifier dapat lebih permisif, yang sesuai untuk menurunkan ambang.

Ini memperkenalkan komplikasi yang kita perlu memperluas kerangka analitis kita untuk menilai dan membandingkan model. Dengan classifier peringkat, classifier ditambah ambang batas menghasilkan confusion matriks tunggal. Setiap kali ambang batas berubah, matriks confusion akan berubah juga karena jumlah positif benar dan kesalahan positif berubah.



Gambar 8.1 Memberlakukan ambang batas pada daftar instans yang diurut berdasarkan skor.

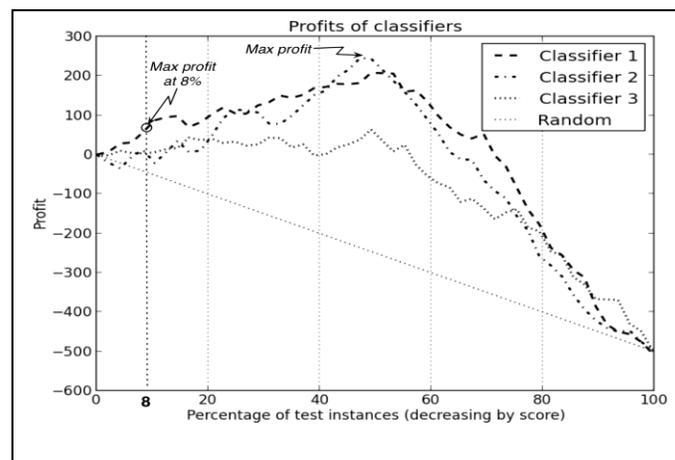
Bagaimana kita membandingkan peringkat yang berbeda? Dan, bagaimana kita memilih ambang yang tepat? Jika kita memiliki perkiraan probabilitas yang akurat dan matriks biaya-manfaat yang ditentukan dengan baik, maka kita sudah menjawab pertanyaan kedua dalam diskusi kita tentang nilai yang diharapkan: kita menentukan ambang batas di mana laba yang kita harapkan berada di atas tingkat yang diinginkan (biasanya nol). Mari jelajahi dan perpanjang ide ini.

## Profit Curves

Telah diketahui cara menghitung ekspektas, dan baru saja diperkenalkan gagasan menggunakan model untuk memberi peringkat pada instance. Kita dapat menggabungkan ide-ide ini untuk membangun berbagai visualisasi kinerja dalam bentuk kurva. Setiap kurva didasarkan pada gagasan memeriksa efek dari thresholding nilai dari classifier pada titik-titik yang berurutan, secara implisit membagi daftar instance menjadi banyak set yang berurutan dari kejadian positif

dan negatif yang diprediksi. Saat ambang diturunkan, akan diperoleh contoh tambahan yang diprediksi sebagai positif daripada negatif. Setiap ambang, yaitu, setiap rangkaian prediksi positif dan negatif, akan memiliki confusion matriks terkait. Sesi sebelumnya menunjukkan bahwa begitu kita memiliki confusion matriks, bersama dengan pengetahuan tentang biaya dan manfaat dari keputusan, kita dapat menghasilkan nilai yang diharapkan sesuai dengan confusion matriks itu.

Lebih khusus lagi, dengan penggolong peringkat, kita dapat menghasilkan daftar contoh dan skor prediksi mereka, yang diberi peringkat dengan menurunkan skor, dan kemudian mengukur laba yang diharapkan yang akan dihasilkan dari memilih setiap titik potong berturut-turut dalam daftar. Secara konseptual, jumlah ini untuk memberi peringkat daftar contoh dengan skor dari tertinggi ke terendah dan menyapu melaluinya, merekam keuntungan yang diharapkan setelah setiap contoh. Pada setiap titik potong kita mencatat persentase daftar yang diprediksi sebagai positif dan perkiraan laba yang sesuai. Menggambarkan Graphic nilai-nilai ini memberi kita kurva keuntungan (profit). Tiga kurva profit ditunjukkan pada Gambar 8-2.



Gambar 8-2. Contoh tiga kurva profit dari tiga klasifier.

## Grafik dan Curva ROC

Kurva profit cocok ketika diketahui secara pasti kondisi di mana suatu classifier akan digunakan. Secara khusus, ada dua kondisi penting yang mendasari perhitungan laba:

1. Class priors; yakni, proporsi antara instans positive and negative dalam populasi target. Proporsi ini sering dikenal dengan base rate.
2. Costs and benefits. Keuntungan yang diharapkan bergantung secara sensitif terhadap perbedaan relatif biaya dan manfaat untuk sel-sel yang berbeda dalam matriks biaya-manfaat.

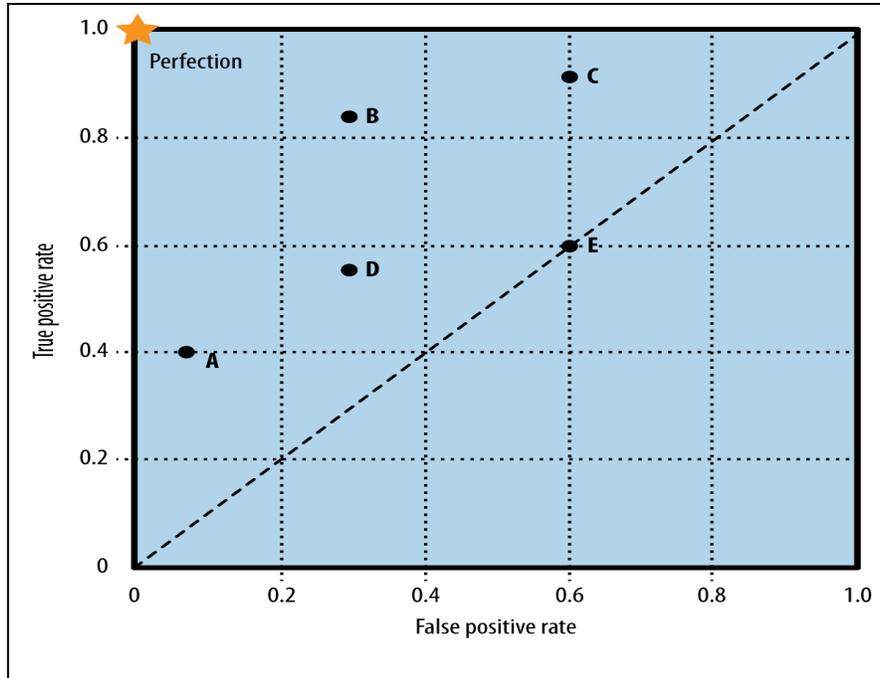
Jika baik nilai class prior dan perkiraan biaya-manfaat diketahui dan stabil, kurva keuntungan dapat menjadi pilihan yang baik untuk memvisualisasikan kinerja model.

Namun, dalam banyak domain, kondisi ini tidak pasti atau tidak stabil. Dalam domain deteksi penipuan, misalnya, jumlah penipuan berubah dari satu tempat ke tempat lain, dan dari satu bulan ke bulan berikutnya. Jumlah penipuan mempengaruhi nilai class prior. Dalam kasus manajemen churn ponsel, kampanye pemasaran dapat memiliki anggaran dan penawaran yang berbeda mungkin memiliki biaya yang berbeda, yang akan mengubah biaya yang diharapkan.

Salah satu pendekatan untuk menangani kondisi yang tidak pasti adalah menghasilkan banyak perhitungan laba yang diharapkan berbeda untuk setiap model. Ini mungkin tidak sangat memuaskan: set model, set prior kelas, dan set biaya keputusan bertambah banyak dalam kompleksitas. Hal ini sering membuat analisis dengan tumpukan grafik keuntungan besar yang sulit dikelola, sulit untuk memahami implikasi, dan sulit untuk dijelaskan kepada pemangku kepentingan.

Pendekatan lain yang dapat mengakomodasi ketidakpastian adalah dengan membangkitkan perhitungan ekspektasi profit yang berbeda untuk setiap model. Pendekatan lain yang juga mungkin adalah dengan menunjukkan seluruh ruang kemungkinan kinerja. Salah satunya dikenal dengan grafik "Receiver Operating Characteristics" (ROC).

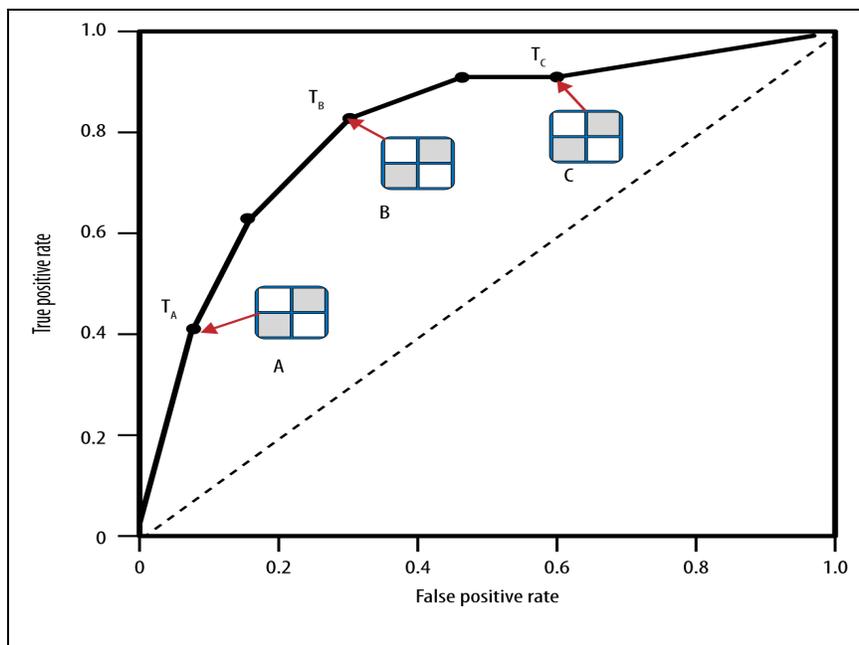
Grafik ROC adalah plot dua dimensi dari classifier dengan false positive rate pada sumbu x terhadap true positive rate pada sumbu y. Dengan demikian, grafik ROC menggambarkan trade-off relatif yang membuat pengelompokan antara manfaat (positif sejati) dan biaya (positif palsu). Gambar 8-3 menunjukkan grafik ROC dengan lima pengklasifikasian berlabel A sampai E.



Gambar 8-3. Ruang ROC dan 5 classifiers (A-E) yang berbeda beserta performance masing masing.

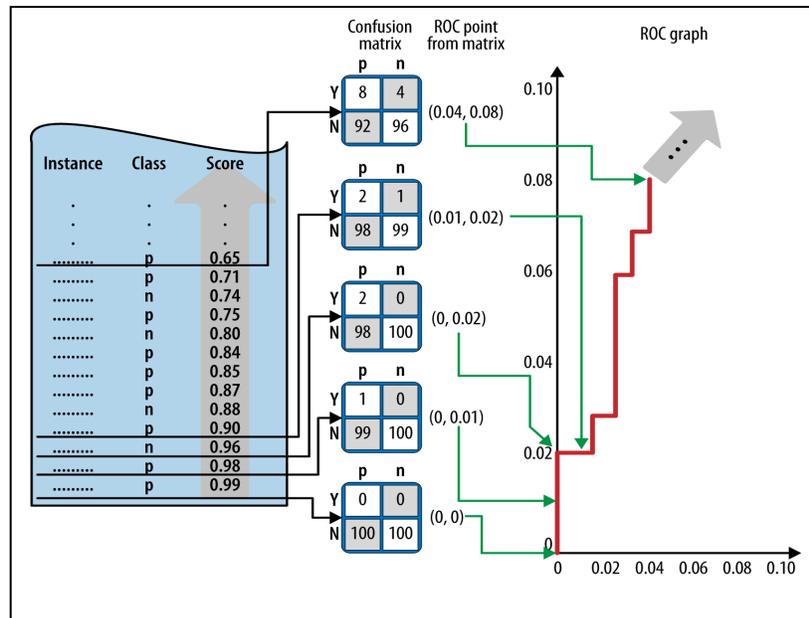
Klasifikasi diskret adalah klasifikasi yang hanya menghasilkan label kelas (sebagai lawan dari peringkat). Seperti yang telah dibahas, masing-masing pengklasifikasian semacam itu menghasilkan confuse matriks, yang dapat direpresentasikan oleh sejumlah nilai statistik tertentu yang menyatakan angka dan skala positif sejati, positif palsu, negatif sejati, dan negatif palsu. Perhatikan bahwa meskipun confusion matriks mengandung empat angka, kita benar-benar hanya membutuhkan dua angka: baik nilai true positive rate atau false negative rate, dan juga false positive rate atau true negative rate. Diberikan satu dari pasangan ini, yang lain dapat diturunkan karena jumlahnya satu. Secara konvensional akan digunakan istilah true positive rate (tp rate) dan false positive rate (fp rate). Setiap classifier diskret menghasilkan pasangan (fp rate, tp rate) yang sesuai dengan satu titik di ruang ROC. Classifier pada Gambar 8-3 semuanya adalah diskrit. Untuk selanjutnya, yang penting diingat adalah, tingkat tp dihitung menggunakan hanya contoh positif yang sebenarnya, dan tingkat fp dihitung hanya menggunakan contoh negatif yang sebenarnya.

Beberapa titik di ruang ROC penting untuk diperhatikan. Titik paling kiri bawah (0, 0) mewakili strategi untuk tidak pernah mengeluarkan klasifikasi positif; pengklasifikasi semacam itu tidak melakukan kesalahan positif palsu tetapi juga tidak memperoleh positif yang sesungguhnya. Strategi yang berlawanan, tanpa syarat mengeluarkan klasifikasi positif, diwakili oleh titik kanan atas (1, 1). Titik (0,1) mewakili klasifikasi sempurna, diwakili oleh bintang. Garis diagonal yang menghubungkan (0, 0) ke (1, 1) mewakili kebijakan menebak kelas. Misalnya, jika penggolong secara acak menebak kelas positif separuh waktu, dapat diharapkan untuk mendapatkan setengah positif dan setengah negatif benar; ini menghasilkan titik (0,5, 0,5) dalam ruang ROC. Jika menebak kelas positif 90% dari waktu, dapat diharapkan untuk mendapatkan 90% dari positif benar tetapi tingkat positif palsu akan meningkat hingga 90% juga, menghasilkan (0,9, 0,9) dalam ruang ROC. Dengan demikian penggolongan acak akan menghasilkan titik ROC yang bergerak bolak-balik pada diagonal berdasarkan frekuensi yang menebak kelas positif. Untuk menjauh dari diagonal ini ke wilayah segitiga atas, penggolong harus memanfaatkan beberapa informasi dalam data. Pada Gambar 8-3, kinerja E pada (0,6, 0,6) hampir acak. E dapat dikatakan menebak kelas positif 60% dari waktu. Perhatikan bahwa tidak ada classifier yang harus berada di segitiga kanan bawah grafik ROC. Ini mewakili kinerja yang lebih buruk daripada dugaan acak.



Gambar 8-4. Setiap titik pada kurva ROC berkaitan dengan satu confusion matrix khusus.

Secara konseptual, dapat dilakukan mengurutkan instance berdasarkan skor dan mengubah ubah ambang batas dari  $-\infty$  ke  $+\infty$  sementara memperhatikan jejak suatu kurva dalam ruang ROC sebagaimana ditunjukkan dalam gambar 8.5.



Gambar 8-5. Sebuah ilustrasi bagaimana kurva ROC dikonstruksi dari data test.

Ketika melewati contoh positif, dilakukan langkah ke atas (meningkatkan positif yang sesungguhnya); setiap kali kita melewati contoh negatif, kita mengambil langkah ke kanan (meningkatkan positif palsu). Jadi “kurva” sebenarnya adalah fungsi langkah untuk satu set tes, tetapi dengan cukup banyak contoh, kelihatannya mulus.

Keuntungan dari grafik ROC adalah bahwa mereka memisahkan kinerja classifier dari kondisi di mana classifier itu akan digunakan. Secara khusus, ROC tidak bergantung pada proporsi kelas serta biaya dan manfaat. Seorang data sains dapat merencanakan kinerja pengklasifikasi pada grafik ROC saat dibuat, mengetahui bahwa posisi dan kinerja relatif dari pengklasifikasi tidak akan berubah. Wilayah (s) pada grafik ROC yang menarik dapat berubah karena biaya, manfaat, dan proporsi kelas berubah, tetapi kurva itu sendiri tidak boleh berubah.

Stein (2005) dan Provost & Fawcett (1997, 2001) menunjukkan bagaimana kondisi operasi dari classifier (prioritas kelas dan biaya kesalahan) dapat dikombinasikan untuk mengidentifikasi wilayah yang menarik pada kurva ROC. Secara singkat, pengetahuan tentang berbagai

kemungkinan prioritas kelas dapat dikombinasikan dengan pengetahuan tentang biaya dan manfaat dari keputusan; bersama-sama ini menggambarkan keluarga garis singgung yang dapat mengidentifikasi penggolong mana yang harus digunakan dalam kondisi tersebut. Stein (2005) menyajikan contoh dari keuangan (pinjaman gagal bayar) dan menunjukkan bagaimana teknik ini dapat digunakan untuk memilih model.

## The Area Under the ROC Curve (AUC)

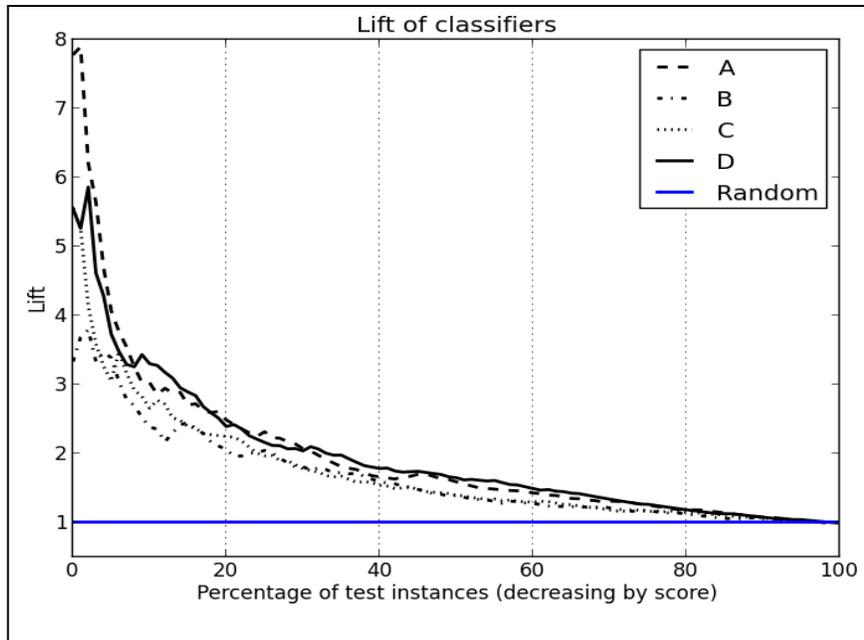
Statistik lainnya yang penting adalah area di bawah kurva ROC (AUC). Seperti namanya, ini hanyalah area di bawah kurva classifier yang dinyatakan sebagai ratio dari satu satuan luas. Nilainya berkisar dari nol hingga satu. Meski demikian kurva ROC memberikan lebih banyak informasi daripada hanya sekedar area. AUC berguna ketika diperlukan cukup satu angka untuk merepresentasikan kinerja.

## Kurva Respons Cumulative Dan Kurva Lift

Kurva ROC adalah alat bantu yang umum digunakan untuk memvisualisasikan kinerja model dalam klasifikasi, estimasi probabilitas kelas, dan scoring. Namun, seperti yang baru saja dialami terlihat bahwa kurva ROC bukanlah visualisasi yang paling intuitif bagi banyak pemangku kepentingan bisnis yang benar-benar harus memahami hasilnya. Penting disadari oleh para ilmuwan data sains untuk menyadari bahwa komunikasi yang jelas dengan para pemangku kepentingan kunci bukan hanya merupakan tujuan utama pekerjaannya, tetapi juga penting untuk melakukan pemodelan yang tepat (selain melakukan pemodelan yang tepat). Oleh karena itu, dapat berguna juga untuk mempertimbangkan kerangka kerja visualisasi yang mungkin tidak memiliki semua sifat kurva ROC yang bagus, tetapi lebih intuitif. (Penting bagi pemangku kepentingan bisnis untuk menyadari bahwa sifat teoritis yang dikorbankan terkadang penting, sehingga mungkin diperlukan dalam keadaan tertentu untuk mengeluarkan visualisasi yang lebih kompleks.)

Salah satu contoh paling umum untuk visualisasi alternatif adalah penggunaan "kurva respons kumulatif," daripada kurva ROC. Keduanya terkait erat, tetapi kurva respon kumulatif lebih intuitif. Kurva respons kumulatif memetakan tingkat klik (tp rate; y axis), yaitu persentase dari yang benar diklasifikasikan dengan tepat, sebagai fungsi dari persentase populasi yang ditargetkan (sumbu x). Jadi, secara konseptual ketika kita memindahkan daftar instance yang diberi peringkat oleh model, kita menargetkan proporsi yang semakin besar dari semua instance. Dalam prosesnya, jika model itu bagus, ketika kita berada di bagian atas daftar, kita akan menargetkan proporsi yang lebih besar dari positif yang sebenarnya daripada yang sebenarnya negatif. Seperti halnya kurva ROC, garis diagonal  $x = y$  mewakili kinerja acak. Dalam hal ini, intuisi jelas: jika Anda menargetkan 20% dari semua instance sepenuhnya secara acak, Anda harus menargetkan 20% dari positif juga. Setiap classifier di atas diagonal memberikan beberapa keuntungan.

Kurva Lift pada dasarnya adalah nilai kurva respon kumulatif pada titik x yang diberikan dibagi dengan nilai garis diagonal ( $y = x$ ) pada titik tersebut. Garis diagonal dari kurva respons kumulatif menjadi garis horizontal pada  $y = 1$  pada kurva lift. Kadang-kadang Anda akan mendengar klaim seperti "model kita memberikan dua kali (atau 2X) lift"; ini berarti bahwa pada ambang pilihan (sering tidak disebutkan), kurva lift menunjukkan bahwa penargetan model dua kali lebih baik daripada acak. Pada kurva respons kumulatif, tp rate yang sesuai untuk model akan dua kali lipat tingkat tp untuk kinerja acak diagonal. (Anda juga dapat menghitung versi pengangkatan dengan memperhatikan beberapa baseline lainnya.) Kurva Lift memplot pengangkatan numerik ini pada sumbu y, terhadap persen populasi yang ditargetkan pada sumbu x (sumbu x yang sama sebagai kurva respons kumulatif).



Gambar 8-7. Empat classifiers (A–D) dari Gambar 8.6 beserta Kurva lift masing masing.

Baik kurva lift dan kurva respon kumulatif harus digunakan dengan hati-hati jika proporsi pasti dari populasi dalam populasi tidak diketahui atau tidak diwakili secara akurat dalam data uji. Tidak seperti kurva ROC, kurva ini mengasumsikan bahwa set tes memiliki target kelas target yang sama persis seperti populasi yang akan diterapkan model. Ini adalah salah satu asumsi penyederhanaan yang disebutkan di awal, yang memungkinkan penggunaan visualisasi yang lebih intuitif.

## KESIMPULAN

Bagian penting dari pekerjaan ilmuwan data adalah menyiapkan evaluasi model yang tepat dan menyampaikan informasi ini kepada para pemangku kepentingan. Melakukan hal ini dengan baik membutuhkan pengalaman, tetapi sangat penting untuk mengurangi kejutan dan mengelola harapan di antara semua yang terkait. Visualisasi hasil merupakan bagian penting dari tugas evaluasi.

Ketika membangun model dari data, menyesuaikan sampel pelatihan dengan berbagai cara mungkin berguna atau bahkan diperlukan; tetapi evaluasi harus menggunakan sampel yang mencerminkan populasi asli dan realistis sehingga hasilnya mencerminkan apa yang sebenarnya akan dicapai. Ketika biaya dan manfaat keputusan dapat ditentukan, ilmuwan data dapat menghitung biaya yang diharapkan per instance untuk setiap model dan hanya memilih model mana yang menghasilkan nilai terbaik. Dalam beberapa kasus, grafik laba dasar dapat berguna untuk membandingkan model minat dalam berbagai kondisi. Grafik ini mungkin mudah dipahami bagi para pemangku kepentingan yang bukan ilmuwan data, karena mereka mengurangi kinerja model ke biaya atau keuntungan "garis dasar" dasar mereka.

Kerugian grafik keuntungan adalah bahwa hal itu mengharuskan kondisi operasi diketahui dan ditentukan dengan tepat. Dengan banyak masalah di dunia nyata, kondisi operasi tidak tepat atau berubah seiring waktu, dan ilmuwan data harus bersaing dengan ketidakpastian. Dalam kasus seperti itu, grafik lain mungkin lebih berguna. Ketika biaya dan manfaat tidak dapat ditentukan dengan keyakinan, tetapi bauran kelas kemungkinan tidak akan berubah, respon kumulatif atau grafik angkat berguna. Keduanya menunjukkan keuntungan relatif dari pengklasifikasi, tidak bergantung pada nilai (moneter atau sebaliknya) dari keuntungan.

Akhirnya, kurva ROC adalah alat visualisasi yang berharga bagi ilmuwan data. Meskipun mereka mengambil beberapa praktik untuk menafsirkan dengan mudah, mereka memisahkan kinerja dari kondisi operasi. Dengan demikian mereka menyampaikan trade-off mendasar yang dibuat masing-masing model.

Banyak pekerjaan dalam komunitas Machine Learning dan Data Mining melibatkan membandingkan pengklasifikasi untuk mendukung berbagai klaim tentang keunggulan algoritma pembelajaran. Akibatnya, banyak yang telah ditulis tentang metodologi perbandingan penggolong. Bagi pembaca yang tertarik, tempat yang baik untuk memulai adalah artikel Thomas Dietterich (1998) “Perkiraan Uji Statistik untuk Membandingkan Algoritma Pengkajian Klasifikasi yang Dibeli,” dan buku Mengevaluasi Algoritma Pembelajaran: Perspektifan Klasifikasi (Japkowicz & Shah, 2011).

## DAFTAR PUSTAKA

1. Foster Provost & Tom Fawcett (2013) Data Science for Business: What you need to know about data mining and data analytic thinking, O'Reilly, ISBN: 978-1-449-36132-7.
2. Sharda, R., Delen, D., Turban, E., (2018). Business intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson.