**Mengatur Nilai Cutoff**

K-NN memayoritaskan rekaman yang baru untuk membuat keputusan, dimana rekaman diklasifikasikan sebagai mayoritas dari kelas turunan, definisi dari mayoritas yaitu secara langsung ditautkan kedalam nilai Cutoff dan diterapkan ke kelas turunannya. Mari kita pertimbankan sebuah binary dari hasil studi kasus. Untuk sebuah rekaman baru, proporsikan sebuah kelas diantara turunannya untuk perkiraan dari probability untuk memiliki 1 buah kelas, untuk contohnya dengan *K* = 4 kita temukan kelas terdekat ke kelas 1 maka (didapatkan = $60.000 dan ukurannya = 20.000 ft) dan hasil akhir 9, 14, 1 dan 20. Karena dua diantaranya adalah owner dan dua lainnya adalah no owner. Menggunakan mayoritas yang sederhana sama dengan mengatur nilai Cutoff dari 0,5. Dari tabel 7.4 subuah aplikasi bisa melihat secara acak dari sebuah kelas.

Sebagaiman disebutkan dalam bab 5, mengubah nilai cutoff mempengaruhi sebuah matrix (I.e.., tingkat kesalahan) karena dibeberapa kasus memungkinkan memilih sebuah cutoff lain dengan default 0.5 dengan tujuan untuk memaksimalkan akurasi atau untuk menggabungka biaya kesalahan klasifikasi.

***K*-NN Dengan Lebih Dari 2 kelas**

Sebuah *K*-NN tergolong mudah diaplikasikan untuk suatu hasil dengan *M* Classes, dimana *M* > 2. Sebuah aturan akan di klasifikasikan sebagai rekaman baru sebagai anggota dari kelas turunannya. Sebuah Alternatif, ketika ada sebuah kelas tertentu dan untuk mengidentifikasinya ( bersedia mengidentifikasi seabagai catatan kelas ini), untuk mengkalkulasi sebuah proporsi dari *K* kelas turunannya yang termasuk kedalam kelas ini, gunakan sebagai estimasi dari sebuah probability dan akan memiliki sebuah kelas baru, dan kemudian untuk spesifikasi dari nilai Cutoff dan menentukan nilai baru dari sebuah kelas. Untuk lebih jelas tentang mengklasifikasikan sebuah nila Cutoff dan penggunaan single class, lihat di bab 5

**Mengkonversi Categorical Variable ke Binary Dummies**

Tidak biasanya untuk menghitung sebuah jarak lebih dari 2 dan bukan kategori numeric (e.g., cookbooks and maps, in a bookstore). Karena itu sebelum *K*-NN bisa menerapkan, sebuah variabel dan lalu dikonversi ke binaru dummies. Berbeda dengan model regrasi statistic, semua *M* binry sebaiknya dibuat dengan menggunakan *K*-NN, sementara secara matematis ini adalah redundant, Sejak *M*-1 dummies berisi informasi yang sama sebagia *M* dummies, informasi redundant tidak dibuat sebuah masalah multicollinearity untuk model linear. Bahkan di *K*-NN digunakan untuk *M* dummies bisa menghasilkan klasifikasi dari penggunaan *M* dummies, dan menyebabkan ketidakseimbangan untuk kategori berbeda to sebuah model.

***K*-NN For A Numercial Outcome**

Ide dari *K-*NN dapat dengan mudah diperluas untuk memprediksi nilai continuous (dengan tujuan sebagai mutlitple linear model regresi) Langkah peratama untuk menentukan kelas turunan denan komputer yang tidak berubah, Langkah kedua dengan memayoritaskan pilihan dari turunan untuk mementukan kelas, lalu modifikasi untuk mengambil nilai rata-rata dari *K*-nearest turunan untuk menentukan prediksi. Dengan sebuah penurunan dari bertambahnya jarak dari titik dimana prediksis diperlukan, kita bisa membuat sebuah function knn() didalam class ke komputer *K*-NN prediksi numerical untuk mengatur validasi.

Modifikasi lain adalah kesalahan metrik pengguna untuk menenukan nilat terbaik *R* , daripada kesalahan digunakan dalam klasifikasi RMS error or another prediction error metric should be used in prediction (lihat di bab 5)

**Pandora**

Pandora adalah layanan radio music yang memungkinkan pengguna mengubah “statition” lalu memutar musik yang mirip dengan artis yang mereka tentukan, pandora menggunakn sebuah *K*-NN style clustering/classification, proses membatalkan sebuah music dan mencari lagu baru atau menutup lagi seorang artis.

Pandora adalah bagian tim Westergren, yang bekerja sebagai pengasuh ketika dia lulus pada tahun 1980, berasama denan Nolan Gasser dia mempelajari music abab pertengan dia mengembangkan mesin, dari memasukan data tentang karakteristik dari sebuah lag kedalam spreadsheet, hasil pertama mengejutkan lagu beatle cocok dengan lagu Bee Gees, lalu ia membuat sebuah perusahaan, hari-hari awal sangat sulit, westergen mempunyai hutang pribadi lebih dari $300.000, dan berakhir dirumah sakit karena penyakit jantung, seorang investor akhirnya mengivestasikan dananya pada tahun 2004 untuk menyelamatkan perusahaan, dan pada 2013 terdaftar dibursa evek

* Pendora telah menetapkan ratusan variabel untuk megukur sebuah lagu pada sekala 0-5 dan berikut adalah daftarnya
* Acid Rock Qualities
* Accordion Playing
* Acousti-Lectric Sonority
* Acousti-Synthetic Sonority
* Pendora membayar musisi dan menganalisa puluhan ribu lagu, dan memberi peringkat pada masing-masing lagu, setiap lagu kemudian dikasih nilai dari 0 - 5, misalnya untuk Led Zeppelin's Kashmir

Kashmir 4 0 3 3 … (high on acid rock attributes, no accordion, etc.)

Langkah ini merupakan investasi yang mahal yang ada di pandora karena teknik tersebut telah diuji dan terpilih karena datanya akurat dari sebuah lagu. dan memberikan dasar untuk menentukan preferensi yang sangat individual

* Pengguna yang menggunakan dapat meentukan lagu yan dia suka(lagu yang berada didalam padora)
* Pandora kemudian menghitung statistik lagu yan sedang didengarkan yang ada didalam penyimpanan, lagu yang didengarkan dipilih oleh pengguna
* Pengguna kemudian memiliki pilihan lagu "saya menyukai lagu ini", "saya tidak menyukai lagu ini" atau diam saja
* jika lagu yang dipilih suka, lagu aslinya, ditambah lagu baru yang baru yang kemungkinan disukai
* Jika lagu yang dipilih tidak suka, lagu yang tidak disukai disimpan untuk referensi(jika pengguna tidak mengungkapkan pendapat tentang lagu yang dipilih, contoh sederhananya lagu baru tidak digunakan lagi untuk perbandingan)
* Pandora mencari lagu baru di penyimpanannya, dari statistik lagu yang disukai dan yang tidak disukai, tergantung yang dipilih pengguna, lagu baru akan ditambahkan baik yang disukai atau yang tidak disukai

Pandora mengembangkan kemampuan untuk membawakan lagu yang sesuai dengan selera pengguna, seorang pengguna akan membangun banyak stasiun dari kelompok lagu yang berbeda, ini adalah cara untuk membatasi pengguna memilih jenre lagu.

Proses yang dijelaskan diatas lebih sedikit kompleks dari proses biasanya, proses yang dijelaskan dalam bab ini, adalah pundamental proses untuk menklasifikasikan sebuah rekaman, proses dalam pembelajaran ini yaitu memilih seorang pemimpin, dan pemilhannya dibuat oleh manusia.

Baca lebih lanjut, di [www.pandora.com](http://www.pandora.com/), Wikipedia artikel sebuah musik, genoa project, and Joyce John’s article “Pandora and the Music Genome Project,” Scientific Computing, vol. 23, no. 10: 14, p. 40–41, Sep. 2006.