**California Housing Price**

**Konteks**

Data ini berasal dari sensus California tahun 1990. Meskipun mungkin tidak relevan untuk memprediksi harga rumah saat ini, dataset ini sangat cocok digunakan sebagai pengenalan untuk mengajarkan dasar-dasar pembelajaran mesin (machine learning). Dataset ini berisi informasi tentang perumahan di distrik-distrik California berdasarkan data sensus tahun 1990. Namun, data ini belum dibersihkan, sehingga diperlukan beberapa langkah praproses.

**Fitur-fitur** Data ini mencakup rumah-rumah di suatu distrik di California, serta beberapa statistik ringkasan dari sensus tahun 1990. Kolom-kolom yang terdapat di dataset ini cukup jelas, dan berikut adalah beberapa di antaranya:

* **longitude:** Garis bujur lokasi
* **latitude:** Garis lintang lokasi
* **Housing median age:** Usia rata-rata rumah
* **total\_rooms:** Total jumlah ruangan di rumah-rumah dalam distrik
* **total\_bedrooms:** Total jumlah kamar tidur
* **population:** Jumlah populasi di distrik
* **households:** Jumlah rumah tangga di distrik
* **median\_income:** Pendapatan rata-rata rumah tangga
* **median house value:** Nilai median harga rumah
* **ocean\_proximity:** Kedekatan rumah dengan laut

**Pengakuan** Data ini pertama kali digunakan dalam makalah berikut: Pace, R. Kelley, dan Ronald Barry. "Sparse spatial autoregressions." *Statistics & Probability Letters* 33.3 (1997): 291-297.