

Menggunakan Klasifikasi Model pada data sensus AS(Amerika Serikat)

Masalah umum dalam pembelajaran Machine Learning adalah mengklasifikasikan data menjadi salah satu dari dua jenis, yang dikenal sebagai label. Misalnya, pengecer mungkin ingin memprediksi apakah pelanggan tertentu akan membeli produk baru, berdasarkan informasi lain tentang pelanggan tersebut. Dalam hal ini, kedua label tersebut mungkin "akan membeli" dan "tidak akan membeli". Anda dapat membuat kumpulan data sedemikian rupa sehingga satu kolom mewakili label. Data yang dapat Anda gunakan untuk melatih model regresi logistik biner seperti itu mencakup lokasi pelanggan, pembelian mereka sebelumnya, preferensi yang dilaporkan pelanggan, dan seterusnya.

1. Langkah Pertama: Buat kumpulan data untuk menyimpan model Anda.

Langkah pertama adalah membuat set data BigQuery untuk menyimpan model Anda.

2. Langkah kedua : Periksa data Anda.

Pada langkah ini, periksa kumpulan data dan identifikasi kolom mana yang akan digunakan sebagai data pelatihan untuk model regresi logistik Anda.

3. Langkah ketiga: Pilih data yang akan Anda gunakan.

Langkah selanjutnya adalah menyiapkan data yang Anda gunakan untuk membuat model regresi logistik biner Anda dengan menjalankan kueri terhadap tabel `sensus_dewasa_pendapatan`. Langkah ini mengidentifikasi fitur yang relevan dan menyimpannya dalam tampilan untuk kueri selanjutnya untuk digunakan sebagai data masukan.

4. Langkah empat: Buat model regresi logistik.

Pada langkah ini, gunakan pernyataan `CREATE MODEL` untuk membuat model regresi logistik Anda.

5. Langkah lima: Gunakan fungsi `ML.EVALUATE` untuk mengevaluasi model Anda.

Kemudian, gunakan fungsi `ML.EVALUATE` untuk memberikan statistik tentang performa model.

6. Langkah enam: Gunakan fungsi `ML.PREDICT` untuk memprediksi pendapatan peserta.

Terakhir, Anda menggunakan fungsi `ML.PREDICT` untuk memprediksi braket pendapatan untuk satu set peserta sensus tertentu.

7. Langkah tujuh: Jelaskan hasil prediksi dengan metode AI yang dapat dijelaskan

Untuk memahami mengapa model Anda menghasilkan hasil prediksi ini, Anda dapat menggunakan fungsi `ML.EXPLAIN_PREDICT`.

8. (Opsional) Langkah delapan: Jelaskan model Anda secara global

Untuk mengetahui fitur mana yang paling penting untuk menentukan income bracket secara umum, Anda bisa menggunakan fungsi `ML.GLOBAL_EXPLAIN`. Untuk menggunakan `ML.GLOBAL_EXPLAIN`, model harus dilatih ulang dengan opsi `ENABLE_GLOBAL_EXPLAIN=TRUE`. Jalankan kembali kueri pelatihan dengan opsi ini menggunakan kueri berikut: