

DAFTAR ISTILAH

Argumen	: Nomina / frasa nominal yang bersama-sama dengan predikat membentuk preposisi.
Adaptasi Domain	: Proses learning atau adaptasi dari suatu domain ke domain baru yang berbeda anotasi data.
Constituent	: Sebuah kata hingga kelompok kata yang berfungsi sebagai satu kesatuan unit dalam struktur hirarki kalimat.
Testing Data	: Data/vektor yang belum diketahui / dianggap belum diketahui kelasnya untuk selanjutnya diprediksi kelasnya menggunakan model klasifikasi yang sudah dibangun.
Training Data	: Data/vektor yang sebelumnya sudah diketahui kelasnya, dan digunakan untuk membangun model klasifikasi.
Dataset	: Kumpulan data.
Feature	: Kuantitas yang digunakan sebagai dasar ukuran dalam menilai suatu objek.
File XML	: Bahasa <i>markup</i> untuk keperluan umum yang disarankan oleh W3C untuk membuat dokumen <i>markup</i> guna keperluan pertukaran data antar sistem yang beraneka ragam.
F1 score	: Rata-rata harmonik berbobot dari <i>precision</i> dan <i>recall</i> . Bisa juga disebut <i>f-measure</i> .
Klasifikasi	: Suatu pekerjaan untuk melakukan pembelajaran terhadap fungsi target f yang memetakan setiap vektor (<i>set</i> fitur) x ke dalam satu dari sejumlah label kelas y yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.
Korpus	: Koleksi dokumen atau data berupa teks yang digunakan dalam studi kasus <i>text mining</i> .
Parse Tree	: Pohon makna yang diperoleh dari suatu kalimat.
Parsing	: Proses analisis kalimat ke dalam beberapa bagian dan menjelaskan peran sintaksis mereka yang lebih kompleks daripada pembacaan sederhana.
Part of Speech	: Pelabelan kata, frase, hingga klausa sesuai dengan aturan gramatikal. Biasanya juga disingkat menjadi POS.
PropBank	: Korpus semantik yang telah memiliki rincian penjelasan proposisi <i>verbal</i> dan argumen.
Precision	: Metrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan (disebut juga <i>positive prediction value</i>).
Recall	: Metrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan yang terbaca (disebut juga sensitifitas).
Semantic Role Labeling	: Proses pemberian label peran terhadap argumen yang diambil dari sebuah kalimat berdasarkan predikatnya.
Terjemahan Al Qur'an	: Korpus semantik yang telah memiliki rincian penjelasan proposisi <i>verbal</i> dan argumen yang berasal dari kalimat terjemahan Al Qur'an dalam bahasa Inggris.
Semantik	: Cabang linguistik yang mempelajari arti/makna yang terkandung pada suatu bahasa, kode, atau representasi lainnya.

XML Tag : Penanda elemen-elemen, dalam hal ini adalah elemen yang ada di dalam *file XML*.