

DAFTAR ISTILAH

Argumen	: Nomina / frasa nominal yang bersama-sama dengan predikat membentuk preposisi.
<i>Constituent</i>	: Sebuah kata hingga kelompok kata yang berfungsi sebagai satu kesatuan unit dalam struktur hirarki kalimat.
<i>Data testing</i>	: Data/vektor yang belum diketahui / dianggap belum diketahui kelasnya untuk selanjutnya diprediksi kelasnya menggunakan model klasifikasi yang sudah dibangun.
<i>Data training</i>	: Data/vektor yang sebelumnya sudah diketahui kelasnya, dan digunakan untuk membangun model klasifikasi.
<i>Dataset</i>	: Kumpulan data.
<i>Feature</i>	: Kuantitas yang digunakan sebagai dasar ukuran dalam menilai suatu objek.
<i>File XML</i>	: 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.
<i>F-score</i>	: Rata-rata harmonik berbobot dari <i>precision</i> dan <i>recall</i> . Bisa juga disebut <i>f-measure</i> .
<i>Hyperplane</i>	: Garis batas yang memisahkan antar dua kelas.
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> .
<i>Parse Tree</i>	: Pohon makna yang diperoleh dari suatu kalimat.
<i>Parsing</i>	: Proses analisis kalimat ke dalam beberapa bagian dan menjelaskan peran sintaksis mereka yang lebih kompleks daripada pembacaan sederhana.
<i>Part of Speech</i>	: Pelabelan kata, frase, hingga klausa sesuai dengan aturan gramatikal. Biasanya juga disingkat menjadi POS.
<i>Precision</i>	: Metrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan (disebut juga <i>positive prediction value</i>).
<i>PropBank</i>	: Korpus semantik yang telah memiliki rincian penjelasan proposisi <i>verbal</i> dan argumen.
<i>Recall</i>	: Metrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan yang terbaca (disebut juga sensitifitas).
Semantik	: Cabang linguistik yang mempelajari arti/makna yang terkandung pada suatu bahasa, kode, atau representasi lainnya.
<i>String</i>	: Tipe data yang digunakan untuk menyimpan barisan karakter.
<i>XML Tag</i>	: Penanda elemen-elemen, dalam hal ini adalah elemen yang ada di dalam <i>file XML</i> .