

## DAFTAR ISTILAH

| <b>Istilah</b>   | <b>Deskripsi</b>   | <b>Halaman pertama kali digunakan</b> |
|------------------|--|---------------------------------------|
| API              | <i>Application Programming Interface</i> adalah sekumpulan definisi dan protokol yang digunakan untuk membangun dan mengintegrasikan perangkat lunak aplikasi.   | 33                                    |
| <i>Coherence</i> | Ukuran yang digunakan untuk mengevaluasi topik dalam model topik, seperti LDA, untuk memastikan bahwa kata-kata dalam topik memiliki keterkaitan atau kesamaan.  | 16                                    |
| <i>Corpus</i>    | Kumpulan teks atau dokumen yang digunakan dalam analisis teks atau pelatihan model.  | 50                                    |
| CSV              | <i>Comma-Separated Values</i> adalah format file teks yang digunakan untuk menyimpan data tabel, di mana setiap baris data adalah catatan, dan setiap catatan terdiri dari satu atau lebih bidang yang dipisahkan dengan koma. | 27                                    |
| CV               | <i>Coherence Value</i> adalah teknik yang digunakan dalam menghitung <i>coherence score</i> untuk mengevaluasi kualitas model topik.   | 16                                    |
| <i>Dataset</i>   | Kumpulan data yang biasanya dalam bentuk tabel yang digunakan untuk analisis atau pelatihan model <i>machine learning</i> .  | 2                                     |
| <i>Fold</i>      | Salah satu bagian dari data yang dibagi dalam <i>cross-validation</i> , di mana data dibagi menjadi beberapa subset yang digunakan untuk pelatihan dan pengujian mode  | 15                                    |

|                    |  |    |
|--------------------|--|----|
| IDF                | <i>Inverse Document Frequency</i> adalah ukuran seberapa umum atau jarang suatu kata dalam seluruh korpus dokumen.   | 11 |
| KDD                | <i>Knowledge Discovery in Databases</i> adalah proses menemukan pengetahuan yang berguna dari data besar, yang meliputi langkah-langkah seperti seleksi data, pra-pemrosesan, transformasi, penambahan data, dan interpretasi hasil.             | 7  |
| Kernel             | Fungsi yang digunakan dalam algoritma SVM untuk mengubah data ke dalam bentuk yang diperlukan untuk menemukan <i>hyperplane</i> optimal yang memisahkan kelas-kelas data.  | 2  |
| LDA                | <i>Latent Dirichlet Allocation</i> adalah teknik model topik yang digunakan dalam analisis teks untuk menemukan kumpulan topik yang menjelaskan kumpulan dokumen, dengan setiap dokumen dilihat sebagai campuran dari berbagai topik.            | 3  |
| <i>Library</i>     | Kumpulan fungsi atau kelas dalam pemrograman yang dapat digunakan kembali untuk berbagai aplikasi.   | 24 |
| <i>Null</i>        | Istilah yang menunjukkan ketiadaan nilai atau data.  | 9  |
| <i>Overfitting</i> | Situasi dalam machine learning di mana model terlalu kompleks dan menyesuaikan diri dengan terlalu banyak detail atau noise dalam data pelatihan, sehingga kinerjanya sangat baik pada data pelatihan tetapi buruk pada data baru atau data uji. | 47 |
| <i>Subset</i>      | Bagian dari kumpulan data yang lebih besar yang dipilih untuk analisis atau pelatihan model.   | 15 |
| SVM                | <i>Support Vector Machine</i> adalah algoritma <i>machine learning</i> yang digunakan untuk klasifikasi dan regresi, yang bekerja dengan mencari <i>hyperplane</i>   | 2  |

|        |  |    |
|--------|--|----|
|        | optimal yang memisahkan data ke dalam kelas yang berbeda.  |    |
| TF     | <i>Term Frequency</i> adalah jumlah kemunculan suatu kata dalam sebuah dokumen.  | 11 |
| TF-IDF | <i>Term Frequency-Inverse Document Frequency</i> adalah statistik yang digunakan untuk mengevaluasi pentingnya suatu kata dalam dokumen relatif terhadap kumpulan dokumen lain dalam korpus. | 11 |
| URL    | <i>Uniform Resource Locator</i> adalah alamat referensi yang digunakan untuk mengakses sumber daya di internet.  | 25 |