

Dimentionality Reduction



Big Data
STF1724

Rini Nuraini Sukmana, M.T
0020087901 - 08882024236
rini.nuraini@usbypkp.ac.id



Dimensionality reduction

- Dimensionality reduction atau reduksi dimensi adalah teknik untuk mengurangi dimensi dataset dalam hal ini fitur data.
- Biasanya dataset yang ingin diproses memiliki puluhan bahkan mungkin ratusan fitur atau kolom.
- Dengan reduksi dimensi, kita dapat mengurangi jumlah fitur atau kolom tanpa menghilangkan informasi dari dataset tersebut



Dimensionality reduction

- Reduksi dimensi pada prinsipnya sama dengan ketika kita mengompres file yang berukuran besar menjadi zip file misalnya.
- Kompresi file tidak akan menghilangkan atau mengurangi informasi yang ada di dalam file tersebut, hanya membuatnya lebih sederhana sehingga mengurangi ukuran file yang dapat mempercepat proses transfer file



Dimensionality reduction

- Reduksi dimensi perlu dilakukan karena Banyaknya variabel input dapat menurunkan performa machine learning.
- Dataset yang digunakan pada umumnya direpresentasikan oleh baris dan kolom, sehingga dalam hal ini variabel inputnya adalah kolom atau fiturnya.
- Jumlah fitur yang sangat banyak sering kali dapat mengakibatkan data point merepresentasikan sampel yang tidak representatif.
- Hal ini dapat sangat mempengaruhi performa.



Dimensionality reduction

- Untuk memperjelas data yang didapatkan dan mempermudah peneliti dalam pengumpulan data selanjutnya, maka dilakukan reduksi data.
- Reduksi data adalah proses pemilihan, pemusatan perhatian pada penyederhanaan, pengabstrakan dan transformasi data kasar yang muncul dari catatan-catatan yang muncul di lapangan.
- Reduksi data berlangsung selama proses pengumpulan data masih berlangsung.



Dimensionality reduction

- Tujuan dari reduksi dimensi adalah untuk menghindari kekurangan dari dimensi, membantu menghilangkan fitur yang tidak relevan dan mengurangi noise, mengurangi waktu dan ruang yang dibutuhkan dalam penambang data dan memungkinkan visualisasi yang lebih mudah.
- Ada dua teknik dalam dimensionality reduction ini, yaitu feature selection dan feature extraction.
 1. Feature selection, memilih feature yang berpengaruh dari sekumpulan data asli.
 2. Feature extraction, membentuk feature baru berdasarkan feature yang lama dengan dimensi yang lebih sedikit dibandingkan dengan sebelumnya.



Dimensionality reduction

- Pada tahap ini juga akan berlangsung kegiatan pengkodean, meringkas dan membuat partisi (bagian-bagian).
- Reduksi data juga merupakan suatu bentuk analisis yang menajamkan, menggolongkan, mengarahkan, membuang yang tidak perlu, dan mengorganisasi data dengan cara sedemikian rupa hingga kesimpulan akhirnya dapat ditarik dan diverifikasi.
- Proses transformasi ini berlanjut terus sampai laporan akhir penelitian tersusun lengkap.



ALGORITMA FEATURE SELECTION RANKINGS

- Feature selection atau seleksi fitur adalah salah satu teknik terpenting dan sering digunakan dalam pre-processing.
- Teknik ini mengurangi jumlah fitur yang terlibat untuk menentukan suatu nilai kelas target dengan mengurangi fitur yang tidak relevan dan data berlebih.
- Tujuan utama dari seleksi fitur ialah memilih fitur terbaik dari suatu kumpulan data fitur.
- Metode feature selection dibagi menjadi tiga kelompok: filter, wrapper, dan embedded selector.

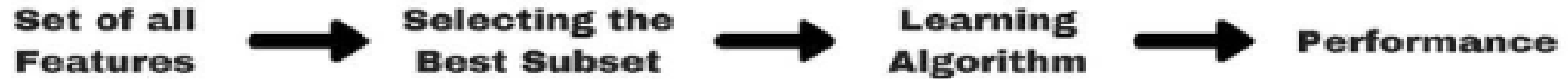


Metode Filter

- Metode filter mengevaluasi setiap fitur secara bebas dari classifier kemudian memberikan peringkat pada fitur setelah mengevaluasi dan mengambil yang unggul.
- Metode filter menerapkan ukuran statistik untuk menetapkan skor untuk setiap fitur.
- Fitur-fitur tersebut diberi peringkat berdasarkan skor dan dipilih untuk disimpan atau dihapus dari dataset.



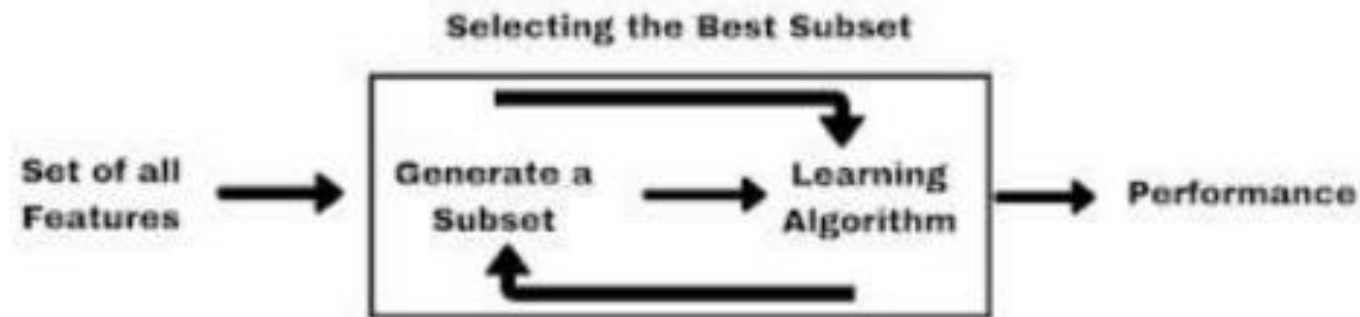
Metode Filter



- Metode filter bergantung pada keunikan umum dari data yang akan dievaluasi dan memilih subset fitur.
- Metode filter menggunakan kriteria penilaian yang mencakup jarak, informasi, ketergantungan, dan konsistensi.
- Metode filter menggunakan kriteria utama teknik pemeringkatan dan menggunakan urutan peringkat untuk pemilihan variabel.
- Alasan untuk menggunakan metode pemeringkatan adalah kesederhanaan, menghasilkan fitur yang sangat baik dan relevan.
- Metode pemeringkatan akan menyaring fitur yang tidak relevan sebelum proses klasifikasi dimulai.
- Metode filter umumnya digunakan sebagai langkah preprocessing data.



Metode Wrapper



- Metode wrapper membutuhkan satu algoritma Machine Learning dan menggunakan kinerjanya sebagai kriteria evaluasi.
- Metode ini mencari fitur yang paling cocok untuk algoritma Machine Learning dan bertujuan untuk meningkatkan kinerja mining.
- Untuk mengevaluasi fitur, akurasi prediktif digunakan untuk tugas klasifikasi.