

Nama : Ferza Reyaldi
NIM : 09021281924060
Mata Kuliah : Pengenalan Pola

UJIAN TENGAH SEMESTER

Rabu, 29 September 2021

1. Kumpulkan 5 research article (jurnal ilmiah) yang membahas mengenai kasus dalam Pengenalan Pola/Machine learning, dan buatlah abstrak nya masing-masing maksimal dalam 300 kata. Resume dalam 1 tabel besar!

Jawab:

Pattern Recognition Algorithm to Identify Detrusor Overactivity on Urodynamics

Detrusor overactivity (DO) kandung kemih adalah temuan pada studi urodinamik (UDS) yang sering berkorelasi dengan gejala saluran kemih bagian bawah dan manajemen dorongan. Namun, interpretasi UDS tetap tidak standar. Kami berusaha mengembangkan model matematika untuk mengidentifikasi DO di UDS dengan andal.

Melalui file arsip UDS antara tahun 2013 dan 2019, penelusuran mentah tekanan vesika, tekanan perut, tekanan detrusor, volume infus, dan semua anotasi selama UDS diperoleh. Pasien berusia kurang dari 1 tahun, studi dengan masalah kalibrasi, atau mereka yang memiliki artifak signifikan dikeluarkan.

Dalam *training set*, lima pola DO yang representatif diidentifikasi. Segmen sinyal kandidat Pdet dicocokkan dengan pola DO yang representatif. Pembelajaran manifold dan algoritma *dynamic time warping* digunakan. *Five-fold cross validation* digunakan untuk mengevaluasi kinerja.

Sebanyak 799 studi UDS dimasukkan. Usia rata-rata adalah 9 tahun (kisaran, 1-33). Terdapat 1.742 kejadian DO yang tidak tumpang tindih dengan artefak beranotasi (batuk, menangis, valsava, gerakan). *Training set* AUC dari *five-fold cross validation* adalah $0,84 \pm 0,01$. *Five-fold cross validation* mengarah ke akurasi keseluruhan 81,35%, dan sensitivitas dan spesifisitas mendeteksi peristiwa DO adalah 76,92% dan 81,41%, masing-masing di *testing set*.

The Application of Pattern Recognition and Machine Learning To Determine Cement Channeling & Bond Quality from Azimuthal Cement Bond Logs

Metode interpretasi log ikatan semen terdiri dari pengenalan pola manusia dan evaluasi kualitas isolasi downhole. Biasanya, juru bahasa log membandingkan data akuisisi dengan klasifikasi kualitas ikatan semen yang telah ditentukan sebelumnya. Makalah ini menguraikan teknik pelengkap evaluasi semen cerdas dan implementasi analisis data evaluasi semen dengan memanfaatkan pencocokan pola otomatis dan *machine learning*. Metode yang diusulkan mampu mendefinisikan kualitas ikatan di beberapa subklasifikasi yang berbeda melalui analisis data citra menggunakan pengenalan pola.

Pustaka tanggapan log nyata digunakan sebagai perbandingan untuk memasukkan data, dan tambahan dapat dilengkapi dengan data sintesis. Menggunakan *machine learning* dan

pengenalan pola berbasis gambar, kualitas ikatan diklasifikasikan ke dalam kategori ringkas untuk menentukan keberadaan saluran. Klasifikasi yang berhasil dari data masukan kemudian dapat ditambahkan ke perpustakaan, sehingga meningkatkan analisis masa depan melalui proses berulang.

Sistem ini menggunakan keluaran dari log evaluasi semen pemindaian ultrasonik azimuth konvensional dan bentuk gelombang CBL 5 kaki untuk menyimpulkan interpretasi ikatan semen. Bentuk gelombang CBL 5 kaki adalah tambahan opsional untuk proses dan meningkatkan interpretasi. Sistem mencari kesamaan antara data akuisisi dengan yang ada di perpustakaan. Kesamaan ini dibandingkan untuk mengevaluasi ikatan.

Proses ini dijelaskan dalam dua bagian: i) pengumpulan gambar dan klasifikasi perpustakaan dan ii) pengenalan dan interpretasi pola. Yang pertama adalah proses menghasilkan pustaka data referensi yang dapat dibaca dari log evaluasi semen historis dan pengukuran laboratorium dan yang terakhir adalah *machine learning* dan metode perbandingan. Sistem ini terbukti sangat mampu dalam identifikasi otomatis penyaluran berbagai ukuran, sesuatu yang akan menjadi tantangan ketika hanya menggunakan representasi kurva skalar dari data azimuth.

Predicting the Tensile Behaviour of Cast Alloys by a Pattern Recognition Analysis on Experimental Data

Kemampuan untuk secara akurat memprediksi sifat mekanik logam sangat penting untuk penggunaan yang benar dalam desain struktur dan komponen. Ini bahkan lebih penting dengan adanya bahan, seperti paduan cor logam, yang sifatnya dapat bervariasi secara signifikan dalam kaitannya dengan elemen penyusunnya, struktur mikro, parameter proses atau perawatan. Studi ini menunjukkan bagaimana pendekatan *machine learning*, berdasarkan analisis pengenalan pola pada data eksperimen, mampu menawarkan prediksi presisi yang dapat diterima sehubungan dengan sifat mekanik utama logam. Sifat metalografi, seperti konten grafit, ferit dan perlit, diekstrapolasi melalui indikator makro dari mikro grafis dengan analisis gambar, digunakan sebagai masukan untuk algoritma *machine learning*, sedangkan sifat mekanik, seperti *yield strength*, *ultimate strength*, *ultimate strain* dan *Young's modulus*, diturunkan sebagai keluaran. Secara khusus, 3 algoritma *machine learning* yang berbeda dilatih mulai dari kumpulan data 20-30 data untuk setiap materi dan hasilnya menawarkan akurasi tinggi, seringkali lebih baik daripada teknik prediksi lainnya.

Predicting The Phase Diagram of Titanium Dioxide with Random Search and Pattern Recognition

Memprediksi stabilitas fase polimorfisme kristal adalah pusat ilmu bahan komputasi dan kimia. Prediksi tersebut membutuhkan pencarian energi potensial minimum dan kemudian melakukan perhitungan energi bebas untuk memperhitungkan efek entropi pada suhu yang terbatas. Sehingga, dikembangkan kerangka kerja yang memfasilitasi prediksi tersebut dengan memanfaatkan semua informasi yang diperoleh dari pencarian acak struktur kristal. Kerangka kerja ini menggabungkan pengelompokan otomatis, klasifikasi dan visualisasi struktur kristal dengan estimasi *machine learning* dari entalpi dan entropi mereka. Kerangka kerja didemonstrasikan pada sistem TiO₂ yang penting secara teknologi, yang memiliki banyak polimorfisme, tanpa bergantung pada pengetahuan sebelumnya tentang fase yang diketahui. Berdasarkan percobaan tersebut, ditemukan sejumlah fase baru dan memprediksi

diagram fase dan *metastability* polimorfisme kristal pada 1600 K, membandingkan hasil dengan perhitungan energi bebas penuh.

Radiomics-based Features for Pattern Recognition of Lung Cancer Histopathology and Metastases

kanker paru-paru adalah penyebab utama kematian terkait kanker di dunia, dan prognosis buruknya sangat bervariasi menurut stadium tumor. Computed tomography (CT) adalah modalitas pencitraan pilihan untuk evaluasi kanker paru-paru, digunakan untuk diagnosis dan stadium klinis. Selain stadium tumor, gambaran lain, seperti subtype histopatologi, juga dapat menambah informasi prognostik. Fitur CT berbasis radiomik digunakan untuk memprediksi histopatologi dan metastasis kanker paru menggunakan model *machine learning*.

kumpulan data gambar lokal dari tumor paru ganas primer yang dikonfirmasi dievaluasi secara retrospektif untuk pengujian dan validasi. Gambar CT yang diperoleh dengan protokol yang sama tersegmentasi secara semi otomatis. Tumor dicirikan oleh gambaran klinis dan atribut komputer dari intensitas, histogram, tekstur, bentuk, dan volume. Tiga pengklasifikasi *machine learning* menggunakan hingga 100 fitur yang dipilih untuk melakukan analisis.

fitur berbasis radiomik menghasilkan area di bawah kurva karakteristik operasi penerima masing-masing 0,89, 0,97, dan 0,92 pada pengujian dan 0,75, 0,71, dan 0,81 pada validasi untuk metastasis kelenjar getah bening, metastasis jauh, dan pengenalan pola histopatologi.

2. Jelaskan peran “Big Data” di era digital saat ini? Komponen apa saja yang mengkategorikan bahwa data tersebut termasuk “Big Data”?

Jawab:

Big Data adalah sekumpulan data yang ukuran sangat besar dan sangat beragam yang tidak dapat diproses menggunakan metode tradisional. Pada era digital Big Data berperan dalam beberapa bidang. Dalam bidang bisnis, Big Data dimanfaatkan dalam meningkatkan *customer experience* dan membantu perusahaan dalam mengambil keputusan berdasarkan data yang telah diproses dan dianalisis.

Data disebut Big Data jika memenuhi kriteria atau komponen berikut yang biasa disebut **3Vs Big Data**. 3 V tersebut adalah *volume*, *velocity*, dan *variety*. Beberapa sumber lain, menyebutkan ada 2 komponen tambahan yang dapat dijadikan kriteria, yaitu *variability* dan *veracity*.

- *Volume*, mengacu pada jumlah data yang dihasilkan melalui dari berbagai sumber,
- *Velocity*, mengacu pada kecepatan data yang dihasilkan, mengacu kecepatan dalam pemrosesan data.
- *Variety*, mengacu pada semua data terstruktur dan tidak terstruktur yang memiliki kemungkinan untuk dihasilkan baik oleh manusia atau oleh mesin.

3. Jelaskan perbedaan antara machine learning dan deep learning? Lalu untuk deep neural networks, apakah termasuk machine learning atau deep learning? Jelaskan alasannya!

Jawab:

Dalam istilah praktis, *deep learning* merupakan bagian dari *machine learning*. Faktanya, *deep learning* dan *machine learning* berfungsi dengan cara yang serupa, namun kemampuannya berbeda. *Deep learning* mengotomasikan sebagian besar bagian ekstraksi fitur dari proses, menghilangkan beberapa intervensi manusia secara manual yang diperlukan. Selain itu juga memungkinkan penggunaan kumpulan data besar, yang membuat *deep learning* disebut *scalable machine learning*. Sedangkan *machine learning* lebih bergantung pada intervensi manusia untuk belajar. Pakar manusia menentukan hierarki fitur untuk memahami perbedaan antara masukan data, biasanya membutuhkan data yang lebih terstruktur untuk dipelajari.

Deep neural networks termasuk dalam *deep learning*. Hal ini dikarenakan *deep neural networks* adalah tulang punggung (*backbone*) dari algoritma *deep learning*. Neural networks yang terdiri dari lebih dari tiga layer—yang mencakup masukan dan keluaran—dapat dianggap sebagai algoritma *deep learning*.

4. Sebutkan 3 contoh terapan masing-masing dari tipe machine learning!

Jawab:

Supervised Learning:

1. Ramalan cuaca.
2. Spam detection pada email.
3. Memprediksi *market value* pemain sepakbola berdasarkan harga pasar pemain (atribut berupa umur, banyak pertandingan, banyak gol, banyak assist dan lain-lain).

Unsupervised Learning:

1. Fraud detection pada transaksi nasabah (*anomaly detection*).
2. Sistem rekomendasi - memberi saran film Netflix yang lebih baik (*clustering*).
3. Segmentasi *customer* dalam penentuan strategi *marketing* suatu bisnis (*clustering*).

Reinforcement Learning:

1. Pengembangan *self-driving car*. Reinforcement learning digunakan dalam melatih mobil berkendara dengan memperhatikan aspek seperti kecepatan, menghindari tabrakan, menentukan *drivable zones*, dan lain-lain.
2. AlphaGo, kecerdasan buatan yang dapat mengalahkan pemain Go terbaik dunia.
3. Horizon, *open-source reinforcement learning platform* dari Facebook untuk mengoptimalkan sistem produksi skala besar. Horizon digunakan untuk mempersonalisasi *suggestions*, memberikan pemberitahuan yang lebih bermakna kepada pengguna, dan mengoptimalkan kualitas streaming video.

5. Jelaskan apa yang dimaksud overfitting, underfitting dan goodfitting dalam model pembelajaran mesin?

Jawab:

Overfitting adalah kondisi dimana kinerja model bernilai baik terhadap *training data* dan *validation data*, namun bernilai buruk saat diterapkan pada *testing data*. Hal tersebut disebabkan model yang dibuat terlalu fleksibel atau terlalu mencocokkan atau menyesuaikan diri dengan *training data*.

Underfitting adalah kebalikan dari overfitting, kondisi dimana model yang dibuat terlalu sederhana, sehingga kinerja model bernilai buruk terhadap *training data*, *validation data* maupun *testing data*.

Goodfitting adalah kondisi dimana kinerja model bernilai baik dan relatif stabil terhadap *training data*, *validation data* dan *testing data*. Model pembelajaran mesin yang dibuat tidak terlalu sederhana dan juga tidak terlalu mencocokkan diri terhadap *training data*.