

Nama : Satria Nur Saputro

Kelas: Kangae

Basic Statistics

1. Apa beda statistik dan statistika?
2. Apa yang dimaksud dengan data semi terstruktur dan sebutkan 3 contohnya!
3. Jelaskan yang dimaksud populasi, sensus, sampel dan survey!
4. Sebutkan dan jelaskan pembagian dari statistik

Deep Learning

1. Mengapa Deep Learning menjadi sangat populer hari ini?
2. Apakah harus selalu menggunakan Deep Learning? Jelaskan!
3. Apa yang dimaksud perceptron?
4. Jelaskan apa itu Activation Function dan 3 contohnya!

AI Project Cycle

1. Sebutkan semua tahapan yang ada di project cycle dari AI!
2. Apa yang dimaksud dari problem scoping?
3. Bagaimana kita tahu data apa yang dibutuhkan?
4. Apa yang dimaksud dengan model dan bagaimana mengevaluasinya?

Jawaban

Basic statistics

1. statistik adalah sebuah kumpulan data, angka atau informasi. Sedangkan statistika merupakan ilmu yang mempelajari bagaimana data, angka tersebut dikumpulkan diolah, dan dianalisis untuk menghasilkan sebuah informasi yang bisa digunakan dalam pengambilan keputusan.
2. Data semi-terstruktur merupakan format data terstruktur dan tidak terstruktur. Walaupun belum diklasifikasi oleh repository tertentu (database), namun mengandung informasi yang penting.
Contohnya adalah data dalam bentuk file csv, file xml, dan file json.

3. -Populasi adalah jumlah keseluruhan dari satuan-satuan atau individu-individu yang karakteristiknya dari data yang akan diteliti.
-Sensus adalah cara terstruktur untuk mendapatkan informasi deskriptif tentang jumlah dari sebuah populasi.
-Sampel adalah sebagian dari populasi tersebut yang haruslah dapat mewakili karakteristik dari keseluruhan populasi.
-survei adalah metode pengumpulan data primer dengan memberikan pertanyaan-pertanyaan kepada responden individu atau kelompok untuk mengumpulkan informasi.
4. Berdasarkan pembagiannya:
 - A. Pembagian Statistik berdasarkan bentuk parameternya
 1. Statistik parametrik adalah parameter dari populasinya itu mengikuti suatu distribusi tertentu, seperti halnya distribusi normal serta mempunyai varians yang homogen
 2. Statistik nonparametrik merupakan parameter dari populasinya itu tidak mengikuti suatu distribusi tertentu. mempunyai distribusi yang bebas dari adanya persyaratan serta variansnya tidak perlu homogen.
 - B. Pembagian Statistik berdasarkan ruang lingkup penggunaannya
 1. Statistik Sosial merupakan statistik yang diterapkan didalam ilmu social
 2. Statistik ekonomi merupakan statistik yang diterapkan dalam ilmu ekonomi.
 3. Statistik Pendidikan merupakan statistik yang diterapkan didalam ilmu dalam bidang pendidikan.
 4. Statistik kesehatan merupakan statistik yang diterapkan didalam bidang kesehatan.
 5. Statistik perusahaan merupakan statistic yang diterapkan didalam bidang perusahaan.
 6. Statistik pertanian merupakan statistik yang diterapkan didalam bidang pertanian.
 - C. Pembagian statistik berdasarkan cara pengolahan datanya.
 1. Statistika Deskriptif adalah statistika yang berkenaan dengan metode atau cara mendeskripsikan, menggambarkan, menjabarkan, atau menguraikan data sehingga mudah dipahami mengacu pada bagaimana menata atau mengorganisasi data, menyajikan, dan menganalisis data. Menata , menyajikan, dan menganalisis data dapat dilakukan dengan menentukan nilai rata-rata hitung, median, modus, standar deviasi, dan persen/proporsi.
 2. Statistika Inferensial adalah statistika yang berkenaan dengan cara penarikan kesimpulan berdasarkan data yang diperoleh dari sampel untuk menggambarkan karakteristik atau ciri dari suatu populasi biasanya dilakukan pengujian hipotesis dan pendugaan mengenai karakteristik atau ciri dari suatu populasi, seperti mean dan standar deviasi

Deep Learning

1. Pertumbuhan era data dan AI telah maju sangat pesat hal tersebut mendorong untuk banyak pengembang membuat algoritma untuk mendapatkan insight dari sebuah data yang

kemudian digunakan untuk melakukan problem solving baik untuk komersial maupun non-komersial, pada dasar diatas mengapa deep learning sangat populer karena deep learning merupakan algoritma yang sangat efektif, efisien, dan powerful untuk melatih data yang besar dan kompleks serta memiliki tingkat akurasi yang jauh lebih baik dibanding algoritma machine learning, hal tersebut juga dapat digunakan dalam banyak ragam permasalahan di perusahaan hal ini lah mengapa deep learning menjadi populer pada era saat ini.

2. Tidak, kebutuhan setiap permasalahan, jumlah data, bentuk data, dan juga sumber daya resource merupakan faktor yang mempengaruhi apakah permasalahan harus memakai deep learning atau tidak, hal tersebut menghindari dari sebuah nilai cost, dan ketidak mampuan prediksi jika data terlalu sedikit, karena deep learning merupakan jaringan yang kompleks dan membutuhkan banyak inputan latihan.
3. Perceptron adalah Jaringan Syaraf Tiruan sebuah metode yang mampu melakukan proses perhitungan dengan mengenali variabel-variabel dalam pencocokan pola dan pada akhirnya hasil keluaran dari Jaringan dapat digunakan sebagai bahan pertimbangan pengambilan keputusan.
4. activation function adalah sebuah fungsi di neural network yang berfungsi untuk menentukan apakah neuron tersebut harus “aktif” atau tidak berdasarkan dari weighted sum dari input dengan kata lain membantu jaringan menggunakan informasi penting dan menekan titik data yang tidak relevan.

Contoh : Relu , sigmoid, LeakyRelu,dan Tan

AI Project Cycle

1. Life cycle AI:

a. Gathering data

Merupakan sebuah untuk mengidentifikasi semua data yang diperlukan untuk menyelesaikan masalah hal ini juga bisa disebut dengan Scope dengan melakukan pernyataan apa , dimana , siapa, mengapa masalah harus dipecahkan.

b. Data preparation

melakukan penyiapan suatu data yang dibutuhkan serta melakukan data exploration seperti kualitas data, format, korelasi,dan outlier.

a. Data wrangling

Yaitu melakukan preprocessing data untuk menghilangkan data data yang kosong duplkat , tidak valid, dan berisik sehingga saat model belajar akan memiliki hasil model yang efisien dan akurat.

b. Data analysis

Merupakan tahapan untuk menyiapkan dan memilih serta menganalisis suatu model terhadap permasalahan yang akan dipecahkan, seperti memilih model arsitektur yang cocok untuk klasifikasi , regresi atau clustering. Kemudian nantinya model yang dibuat juga harus dilakukan tahap evaluasi, dengan harapan terhindar dari underfitting dan overfitting.

c. Train model

- Tahapan untuk melatih algoritma yang telah ditentukan dengan data latih yang telah disiapkan sehingga diharapkan mendapatkan model yang akurat.
- d. Test model
Tahapan untuk mencoba model terhadap data-data baru yang belum pernah diuji apakah model dapat membuat prediksi dengan tepat atau tidak.
 - e. Deployment
Tahap terakhir dari life cycle AI setelah semua proses telah dilalui sehingga model dapat diterapkan untuk real-world system untuk pemecahan masalah.
2. Problem Scoping adalah sebuah lingkup masalah yang mengacu pada pemahaman masalah menemukan berbagai faktor yang mempengaruhi masalah, menentukan tujuan atau sasaran proyek terhadap masalah tersebut dalam pemecahannya. Dengan pembahasan seperti apa, siapa, dimana dan mengapa permasalahan harus dipecahkan.
 3. Hal ini tidak lepas dalam peran life cycle AI dan problem scoping yaitu data yang dibutuhkan dalam permasalahan masalah, data yang ditetapkan harus mewakili dari permasalahan apa yang akan dipecahkan, hal ini pada terdapat pada gathering data, data preparation, dan jawaban dari 4W problem scoping
 4. Model adalah sebuah hasil (aturan) dari machine learning atau deep learning yang telah diajarkan serta memiliki arsitektur yang telah ditentukan. Cara untuk mengevaluasi model dengan cara melakukan testing data yang belum diajarkan dan menggunakan confusion metrics untuk evaluasi model untuk setiap permasalahan seperti supervised learning akurasi presisi, dan recall untuk klasifikasi, dan MSE atau RMSE untuk regresi.