Dr. Wahyudi,S.T.,M.T.

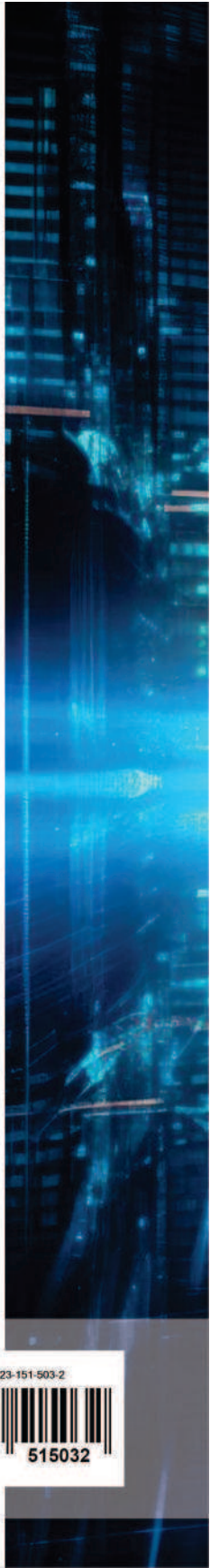
PENGANTAR SAINS DATA

Proses Sains Data dan Pembelajaran Mesin

PENGANTAR SAINS DATA

Proses Sains Data dan Pembelajaran Mesin

Buku Pengantar Sains Data: Proses Sains Data Dan Pembelajaran Mesin terdiri dari tiga bab. Bab pertama yaitu Sains Data Dalam Dunia Big Data menjelaskan tentang defenisi sains data dan big data, proses sains data, dan ekosistem sains data. Bab kedua menjelaskan secara rinci proses sains data. Terdapat 6 langkah proses yang harus dilakukan di dalam sains data yaitu: Menetapkan tujuan penelitian, Mengambil data, Persiapan data, Eksplorasi data, Pemodelan data, Presentasi dan otomatisasi. Bab terakhir menjelaskan tentang pembelajaran mesin. Ilmuwan data sangat bergantung pada teknik-teknik dari statistik dan pembelajaran mesin untuk melakukan pemodelan. Ada 4 tahap proses pemodelan yaitu: Rekayasa fitur, Pelatihan model, Pemilihan dan validasi model, Penilaian model. Bab ini juga menjelaskan 3 teknik pembelajaran mesin yaitu: Pembelajaran Terawasi (supervised Learning), Pembelajaran tanpa pengawasan (Unsupervised learning), dan Pembelajaran semi terawasi (semi-supervised Learning)berada di antara kedua teknik tersebut dan digunakan ketika hanya sebahagian kecil data yang diberi label.



eureka
media aksara
Anggota IKAPI
No. 225/JTE/2021

0858 5343 1992
eurekamediaaksara@gmail.com
Jl. Banjaran RT.20 RW.10
Bojongsari - Purbalingga 53362

ISBN 978-623-151-503-2
9 786231 515032

PENGANTAR SAINS DATA:

Proses Sains Data dan Pembelajaran Mesin

Dr. Wahyudi,S.T.,M.T.

PENERBIT CV.EUREKA MEDIA AKSARA

PENGANTAR SAINS DATA:
Proses Sains Data dan Pembelajaran Mesin

Penulis : **Dr. Wahyudi,S.T.,M.T.**

Desain Sampul : Ardyan Arya Hayuwaskita

Tata Letak : Herlina Sukma

ISBN : 978-623-151-503-2

Diterbitkan oleh : **EUREKA MEDIA AKSARA, SEPTEMBER 2023**
ANGGOTA IKAPI JAWA TENGAH
NO. 225/JTE/2021

Redaksi:

Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari Kabupaten Purbalingga Telp. 0858-5343-1992

Surel : eurekamediaaksara@gmail.com

Cetakan Pertama : 2023

KATA PENGANTAR

Alhamdulillahirobbil'alamin. Penulis bersyukur kehadirat Allah SWT berkat rahmat, karunia dan pertolonganNya, penulis dapat menyelesaikan buku berjudul " **Pengantar Sains Data: proses sains data dan pembelajaran mesin**". Shalawat serta salam semoga senantiasa tercurah atas Nabi Muhammad SAW, para kerabat, serta pengikutnya hingga hari kiamat nanti.

Buku ini hadir untuk menambah literasi tentang teknologi informasi di lingkungan Universitas andalas. Buku ini merupakan seri pertama dari beberapa buku sains data. Buku ini menjelaskan tentang munculnya bidang sains data, hubungan sains data dan big data, proses yang dilakukan di sains data dan pembelajaran mesin yang merupakan bagian paling penting di sains data.

Penulis menyadari bahwa dalam penulisan buku ini masih banyak terdapat kekurangan, untuk itu penulis mengharapkan kritik dan sarannya guna penyempurnaan buku ini di masa mendatang

Padang, Agustus 2023

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1 SAINS DATA DALAM DUNIA BIG DATA

Bab ini mencakup

1. Mendefinisikan Sains data dan big data
2. Mengenali berbagai jenis data
3. Memperoleh wawasan tentang proses sains data
4. Memperkenalkan bidang-bidang sains data dan big data

Big data adalah istilah umum untuk kumpulan data yang begitu besar atau kompleks sehingga sulit untuk diproses menggunakan teknik manajemen data tradisional seperti, misalnya, RDBMS (sistem manajemen basis data relasional). RDBMS yang diadopsi secara luas telah lama dianggap sebagai solusi yang cocok untuk semua, namun tuntutan untuk menangani data besar menunjukkan sebaliknya. Sains data melibatkan penggunaan metode untuk menganalisis data dalam jumlah besar dan mengekstrak pengetahuan yang dikandungnya. Anda dapat menganggap hubungan antara big data dan sains data seperti hubungan antara minyak mentah dan kilang minyak. sains data dan big data berevolusi dari statistik dan manajemen data tradisional, namun kini dianggap sebagai disiplin ilmu yang berbeda.

BAB 2 PROSES SAINS DATA

Tujuan dari bab ini adalah untuk memberikan gambaran umum tentang proses sains data tanpa harus menyelami big data. Anda akan mempelajari cara bekerja dengan kumpulan data besar, data streaming, dan data teks di bab-bab berikutnya.

A. Gambaran Umum Proses Sains Data

Mengikuti pendekatan terstruktur untuk sains data membantu Anda memaksimalkan peluang keberhasilan dalam proyek sains data dengan biaya terendah. Pendekatan ini juga memungkinkan untuk mengerjakan proyek sebagai sebuah tim, dengan setiap anggota tim berfokus pada apa yang mereka lakukan dengan baik. Namun, berhati-hatilah: pendekatan ini mungkin tidak cocok untuk semua jenis proyek atau menjadi satu-satunya cara untuk melakukan sains data yang baik.

Proses sains data pada umumnya terdiri dari enam langkah yang akan Anda lakukan secara berulang, seperti yang ditunjukkan pada gambar 7.

Gambar 7 meringkas proses sains data dan menunjukkan langkah-langkah dan tindakan utama yang

BAB 3 PEMBELAJARAN MESIN

Bab ini mencakup

1. Memahami mengapa ilmuwan sains data menggunakan pembelajaran mesin
2. Mengidentifikasi pustaka Python yang paling penting untuk pembelajaran mesin
3. Mendiskusikan proses pembuatan model
4. Menggunakan teknik-teknik pembelajaran mesin
5. Mendapatkan pengalaman langsung dengan pembelajaran mesin

Tahukah Anda bagaimana komputer belajar untuk melindungi Anda dari orang yang berniat jahat? Komputer menyaring lebih dari 60% email Anda dan dapat belajar untuk melakukan pekerjaan yang lebih baik dalam melindungi Anda dari waktu ke waktu.

Dapatkah Anda secara eksplisit mengajari komputer untuk mengenali orang dalam sebuah gambar? Mungkin saja, tetapi tidak praktis untuk mengkodekan semua cara yang mungkin untuk mengenali seseorang, tetapi Anda akan segera melihat bahwa kemungkinannya hampir tak terbatas. Agar berhasil, Anda harus menambahkan keterampilan baru

DAFTAR PUSTAKA

Apache Spark. (2023). Apache Spark for Python. https://spark.apache.org/

Cielen , Meysman, Ali. (2016). Introducing Data Science. Manning Publications Co. Shelter Island, New York

Cielen. (2016). Overview of the data science process. http://freecontent.manning.com/overview-of-the-data-science-process/

Cyton. (2023). C-Extensions For Python. https://cython.org/

Mike Bostock. (2017). diagram Sankey. https://bost.ocks.org/mike/

Pydata. (2023). NumExpr: Fast numerical expression evaluator for NumPy. https://github.com/pydata/numexpr

Wikipedia. (2023). Time complexity. https://en.wikipedia.org/wiki/Time_complexity