
RAHMADYA TRIAS HANDAYANTO
HERLAWATI

# DATA MINING DAN MACHINE LEARNING MENGGUNAKAN MATLAB DAN PYTHON

Konsep Data Mining | Mempersiapkan Matlab | Mempersiapkan Python | Metode Data Mining Berbasis Statistika | Adaptive Neuro-fuzzy Inference System (ANFIS) | Klasterisasi (clustering) | Support Vector Machine (SVM) | Validasi dan Pengujian Model | Data Mining World Wide Web

Penerbit **INFORMATIKA**

# *Data Mining* dan *Machine Learning* Menggunakan Matlab dan Python

Rahmadya Trias Handayanto
Herlawati

Penerbit **INFORMATIKA**

***Data Mining* dan *Machine Learning***
**Menggunakan Matlab dan Python**


| | | |
|---|---|---|
| Penyusun | : | Rahmadya Trias Handayanto<br>Herlawati |
| Penerbit | : | Informatika Bandung |
| Pemasaran | : | **BI-Obses**<br>Pasar Buku Palasari No. 82<br>Bandung 40264<br>Telp.(022)7317812<br>Fax. (022)7317896 |
| Cetakan | : | Oktober 2020 |
| ISBN | : | 978-623-7131-32-8 |

# DAFTAR ISI

# BAB 1

# KONSEP *DATA MINING*

## 1.1 Pendahuluan

Beberapa *literature* yang membahas *data mining* banyak kita jumpai saat ini dan layak untuk dibaca (lihat daftar referensi). Tidak ada buku yang sanggup memberikan informasi lengkap karena Data Mining merupakan bidang yang sangat luas penerapannya. Untuk pembahasan ini beberapa materi diambil dari buku referensi tersebut (kompilasi) sesuai dengan perkembangan terkini yang cenderung mengarah ke sistem *Soft Computing* yang lebih baik dalam mengatasi permasalahan yang ada saat ini.

*Data mining* menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisis terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut (Larose, 2006). Disebutkan bahwa cara yang digunakan adalah cara terkini (Novel) karena perkembangan teknologi yang terus berkembang dan database administrator harus mampu mengikuti perkembangan yang

# BAB 3

# MEMPERSIAPKAN PYTHON

...ir semua bahasa pemrograman dapat digunakan untuk *data* ...ng, salah satunya adalah bahasa pemrograman terkenal python. ...asa ini diluncurkan pertama kali pada tahun 1991 di *Scitching* ...matisch Centrum*,* Belanda. Walaupun menggunakan nama ... satu spesies ular, sesungguhnya perancang Python, Guido van ..., mengambil nama bahasa pemrograman tersebut dari nama ... komedi Inggris terkenal Monty Python. Kemudian dengan ...alannya waktu, Python beralih dari satu laboratorium ke ...ratorium lain dan satu perusahaan ke perusahaan lain.

... tulisan ini dibuat, Python sudah masuk ke versi 3 dengan alat ... terkenalnya, TensorFlow, yang tangguh untuk mengelola data ...ng. Dengan karakteristiknya yang *open source* dengan lisensi ... – *compatible*, bahasa ini sangat diminati oleh pengembang ...ngkat lunak di seluruh dunia. Beberapa literatur mengungkapkan ...rapa kelebihan Python, di antaranya: *Readability,* efisien, ...fungsi, interoperabilitas yang baik, dan dukungan komunitas ... kuat. Oleh karena itu, diharapkan pembaca menginstal terlebih ...lu *software* Python.

```
y_pred=regressor.predict(X_uji)
print(y_pred)
```

```
[655]: #Fitting pada Data Training
       from sklearn.linear_model import LinearRegression
       regressor = LinearRegression()
       regressor.fit(X_train, y_train)
       #Memprediksi hasil dengan Data Tes
       y_pred=regressor.predict(X_uji)
       print(y_pred)

       [3.162 3.376 3.59  3.804 4.018 4.232]
```

**Gambar 4.3 Tampilan hasil fitting pada data training**

amat, Anda berhasil memproyeksikan jumlah penduduk ngan regresi linear untuk enam tahun ke depan. Berikutnya lah proses menyimpan hasilnya ke dalam format CSV yang at dibuka dengan Excel atau Text Editor. Buat sel baru dan ik instruksi berikut untuk menyimpan hasil proyeksi ke file v.

```
#Menyimpan Hasil ke File CSV
tahunuji=np.array(X_uji)
tahun=pd.DataFrame(tahunuji)
prediksi=pd.DataFrame(y_pred)
hasil=pd.concat([tahun, prediksi],axis=1)
np.savetxt('hasil.csv', hasil, delimiter=',')
hasil
```

il perlu dilakukan operasi "concatenation" dalam rangka ggabungan antara tahun uji dengan hasil proyeksi ulasinya.

# BAB 5

# *ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM* (ANFIS)

## 1 Teori Dasar

eda dengan *Fuzzy Inference System* (FIS), atau yang lebih nal dengan fuzzy saja, ANFIS membuat rule berdasarkan data ditraining lewat mekanisme mirip Jaringan Syaraf Tiruan (JST). rule yang bisa dilayani hanyalah yang bertipe **Takagi-Sugeno-g** (TSK), atau dikenal dengan istilah Sugeno saja. Jenis Mamdani dapat diterapkan pada ANFIS. Untuk mengenal lebih dekat an fuzzy TSK berikut ini ringkasannya.

de FSK diprakarsai oleh Takagi, Sugeno, dan Kang pada tahun . Tujuannya adalah untuk memperoleh rule yang berasal dari ngan masukan dan keluaran suatu sistem. Prinsipnya antara

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = f(x, y) \qquad (1)$$

ana A dan B merupakan set Fuzzy sedangkan z merupakan i dalam bentuk *crisp* (bukan *fuzzy*). Jadi perbedaan yang

# BAB 6

# JARINGAN SYARAF TIRUAN (JST)

## 6.1 Teori Dasar

Jaringan Syaraf Tiruan (JST) atau dalam istilah internasionalnya Neural Networks, bermaksud meniru cara kerja otak makhluk hidup. Komponen utama dari JST antara lain neuron dan sinaps. Neuron berisi informasi apakah satu informasi diteruskan atau tidak ke neuron lainnya. Sinaps berisi hubungan antara satu neuron dengan neuron lainnya. Diteruskan atau tidak pada suatu neuron didasarkan pada pengaturan bobot (pengali) dan bias (penambah). Di awal-awal ditemukannya JST, bias dan bobot diatur manual, tetapi sejak ditemukannya prinsip pembelajaran, misalnya propagasi balik (backpropagation) maka bobot dan bias diset berdasarkan pelatihan dari data yang ada. Data yang ada tersebut memiliki label atau target yang harus dicapai oleh model JST. Biasanya ada error antara hasil prediksi model JST dengan label/target. Tetapi jika error cukup kecil, maka model yang diperoleh dari proses pelatihan dapat digunakan untuk memprediksi data lain di luar data pelatihan. Karena adanya prinsip pembelajaran maka JST merupakan salah satu metode dalam mesin pembelajaran (*machine learning*). Juga karena berusaha

# DATA MINING DAN MACHINE LEARNING MENGGUNAKAN MATLAB DAN PYTHON

Buku ini merupakan kelanjutan dari buku sebelumnya tentang soft computing. Banyaknya pembaca yang berminat membuat sistem berbasis mesin pembelajaran (Machine Learning) untuk menggali data (Data Mining) membuat kami melanjutkan membuat buku khusus tentang Data Mining. Tema-tema yang ditulis sebagian besar diambil dari kasus-kasus yang menjadi topik riset mahasiswa-mahasiswa tingkat sarjana dan pascasarjana.

Data mining mengharuskan tersedianya data yang akan digali dan dicari informasi-informasi tersembunyi yang bermanfaat bagi pengambil keptusan. Namun untuk mempermudah proses belajar, dalam buku ini kami hanya menyediakan data-data sederhana yang sengaja dirancang mirip dengan data-data yang jumlahnya banyak yang diperoleh lewat riset di institusi tertentu. Akan tetapi, walaupun sederhana prinsipnya dapat dterapkan langsung dengan data riil.

Buku ini disusun dalam bentuk teori dasar singkat yang dilanjutkan dengan terapannya untuk kasus tertentu dengan bahasa pemrograman Matlab dan Python. Diharapkan pembaca melihat referensi yang kami cantumkan di akhir tulisan jika ingin memperdalam teori dasarnya. Setelah Bab I membahas tinjauan singkat mengenai konsep Data Mining, kami lanjutkan dengan Bab II yang berisi pengenalan terhadap bahasa pemrograman Matlab dilanjutkan dengan bahasa Python di bab III untuk melakukan proses data mining. Kedua bab tersebut menjelaskan teknik-teknik kompulasi, menggunakan Integrated Development Environment (IDE) dan pembuatan Graphic User Interface (GUI) baik dengan bahasa Matlab maupun Python. Bab IV membahas data mining berbasis statistika klasik, antara lain regresi, pohon keputusan dan Naïve Bayes. Bab V dan Bab VI membahas metode-metode yang lebih modern, yaitu ANFIS dan Jaringan Syaraf Tiruan. Berikutnya Bab VII berisi metode-metode dalam klasterisasi (clustering). Bab VIII membahas metode yang saat ini banyak diteliti, yaitu Support Vector Machint (SVM), baik untuk kasus bi-class maupun multi-class. Bab IX disertakan pula teknik-teknik validasi dan pengujian model. Bab X merupakan bab tambahan khusus untuk menggali data dari world wide web yang saat ini sedang ramai diteliti, baik berupa situs web maupun sosial media.

Penerbit **INFORMATIKA**
Pemasaran: **BI-OBSES**
Pasar buku Palasari 82 Bandung 40264
Tel.(022) 7317812 Fax.(022) 7317896
**www.biobses.com**

## BIOGRAFI PENULIS

**Rahmadya Trias Handayanto, S.T., M.Kom., Ph.D.**

Praktisi yang sempat bekerja sebagai staff IT di bank berskala nasional kini menjadi pengajar sejak Tahun 2002 di beberapa Universitas. Penulis sudah tersertifikasi dosen sejak tahun 2011, dan sebagai Asesor Sertifikasi Dosen. Bidang yang digeluti adalah data mining, sistem cerdas, Information Management, Computer Science, Data Spatial, Optimalisasi dan Sistem Informasi Geografis. Saat ini aktif melakukan beberapa penelitian dengan sponsor dari Kementerian Riset dan Teknologi. Penulis menyelesaikan program doctoral dari Asian Institute of Technology (AIT) Thailand. **E-mail:** rahmadya.trias@gmail.com, website: rahmadya.com.

**Herlawati, S.Si., M.M., M.Kom**

Penulis sejak Tahun 1999 menjadi pengajar di beberapa Universitas. Penulis sudah tersertifikasi dosen sejak tahun 2009, dan sebagai Asesor Sertifikasi Dosen. Mata kuliah yang diampu Kalkulus I, Kalkulus II, Aljabar Linier dan Matriks, Statistika, Struktur Diskrit, Data Mining, dan Analisa Perancangan Sistem Informasi. Bidang yang digeluti adalah Data Mining, Computer Science, Data Spatial, Optimaliasasi dan Sistem Informasi Geografis serta Metaheuristics. Penulis aktif melakukan beberapa penelitian dengan sponsor dari Kementerian Riset dan Teknologi. **E-mail:** mrs.herlawati@gmail.com, **website:** herlawati.com.