

**Departemen Statistika**

Fakultas Matematika dan Ilmu Pengetahuan Alam  
Institut Pertanian Bogor



**IPB University**  
— Bogor Indonesia —

# **PEMODELAN KLASIFIKASI**

## **PERTEMUAN #1**

## **PENGANTAR**

**Bagus Sartono**

[bagusco@apps.ipb.ac.id](mailto:bagusco@apps.ipb.ac.id)

**2020**

# DESKRIPSI MATA KULIAH

**Deskripsi:** Mata kuliah ini mendiskusikan beberapa algoritma dalam analisis data dan data mining untuk tujuan klasifikasi, yaitu menentukan kelas atau kelompok dari setiap amatan. Topik yang akan dibahas meliputi pendekatan un-supervised dan supervised, dengan penekanan lebih banyak pada yang kedua. Algoritma un-supervised yang akan dibahas adalah k-means, sedangkan algoritma supervised meliputi k-NN, regresi logistik, pohon klasifikasi, pengenalan support vector machine, dan naïve bayesian classifier. Juga akan didiskusikan pendekatan ensemble yaitu bagging, boosting, dan random forest. Tidak hanya algoritma yang akan dipelajari tetapi juga membahas proses evaluasi dan validasi model. Mata kuliah ini memiliki sks praktikum yang didalamnya akan mendiskusikan penggunaan software R.



# PENGAJAR

Dr. Bagus Sartono

Dr. Anang Kurnia

# BAGUS SARTONO



- Dosen di Departemen Statistika – FMIPA IPB University
- Koordinator Working Group Data Mining – FMIPA IPB University
- Wakil Ketua FORSTAT (Forum Penyelenggara Pendidikan Tinggi Statistika)
- Pendidikan:
  - S1, Statistika – IPB
  - S2, Statistika – IPB
  - PhD, Applied Economics – Univ of Antwerp

# SILABUS

- Pengantar
- Un-supervised Classification: k-means
- Pengantar mengenai supervised classification
- k-NN
- Penilaian kebaikan dugaan klasifikasi
- Analisis Diskriminan
- Regresi Logistik
- Diskretisasi
- Pohon Klasifikasi
- Bagging
- Boosting
- Random Forest dan Rotation Forest
- Pengenalan Support Vector Machine
- Pengenalan Naïve Bayes Classifier

# PENDEKATAN PEMBELAJARAN

- Integrasi antara ceramah teori dan praktek
- Memerlukan (dan menuntut) keaktifan mahasiswa
- Di setiap pertemuan mahasiswa membawa komputer/laptop, dan dosen akan menyiapkan data dan programnya
- Menggunakan R

# BAHAN BACAAN

Trevor Hastie, Robert Tibshirani, Jerome Friedman. 2013. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer.

Daniel T. Larose, Chantal D. Larose. 2015. Data Mining and Predictive Analytics. Wiley.

John D. Kelleher & Brian Mac Namee & Aoife D'Arcy. 2015. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. The MIT Press.

# PENILAIAN

## Komponen

- Keaktifan di Kelas
- Presentasi
- Laporan-laporan Tugas
- Laporan Tugas Akhir
- Kualitas Prediksi Tugas Akhir

## Bobot

15%

20%

25%

20%

20%





**PENGANTAR**

# ANALISIS KLASIFIKASI

Tujuan analisis: menentukan keanggotaan grup/kelompok dari suatu individu

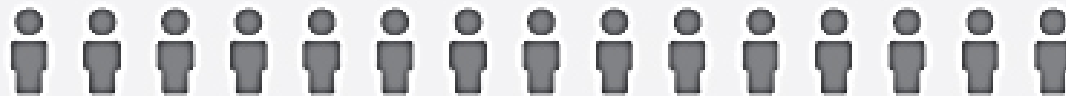
## Tipe metode

- Unsupervised, tidak terdapat informasi mengenai kelompok/grup dari amatan pada data yang digunakan. Analisis dilakukan untuk menentukan keanggotaan grup dari amatan tersebut. Sering juga dikenal sebagai analisis gerombol (clustering, cluster analysis)
- Supervised, data memiliki informasi mengenai kelompok/grup sesungguhnya dari amatan. Analisis dilakukan untuk menentukan pembeda antar grup, dan aturan pembeda tersebut dapat dimanfaatkan untuk menentukan keanggotaan dari amatan lain yang tidak ada dalam data.

# KEGUNAAN ANALISIS KLASIFIKASI

## Before Propensity Modeling

Blanket Marketing Communications efforts to all prospects

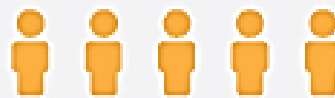


## After Propensity Modeling

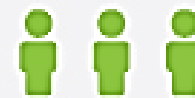
Target prospects likely to respond to Marketing Communications efforts



Low



Medium



High



**GEOGRAPHIC**

**DEMOGRAPHIC**

**MARKET  
SEGMENTATION**

**BEHAVIORAL**

**PSYCHOGRAPHIC**

# CREDIT scoring



- Scoring model to **predict the risk level** of debtors
- Classification model involving predictors: socio-demographical variables, historical payment, other transaction records
- Scores
  - **Good**/Excellent Risk
  - **Bad**/Poor Risk
- **Common algorithms:**
  - Logistic Regression
  - Classification Tree

# PREDICTIVE ANALYTICS IN BUSINESS



- **Propensity model** to predict the likelihood-to-buy of individuals
  - Up-Sell / Cross-Sell campaign
- Selective campaign
  - **High propensity** → give the offering
  - **Low propensity** → no offer
- Common algorithms: Random Forest, Boosted Tree

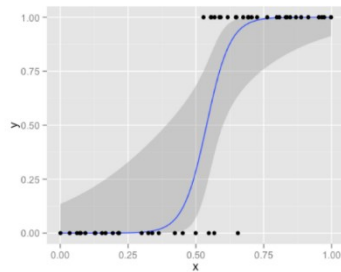
- Identifying the probability of dormant cards to be active
- Recall Campaign to the prospective active card holder
- Common Algorithm:
  - k-Nearest Neighbor

## Debit/Credit Card activation



# COMMON CLASSIFICATION MODEL ALGORITHMS

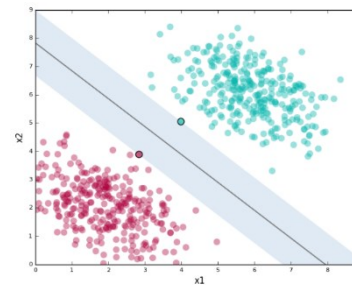
## Logistic Regression



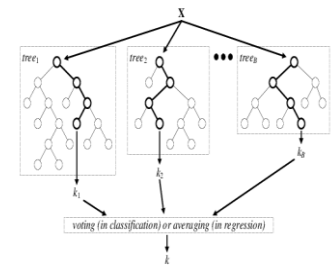
## Classification Tree



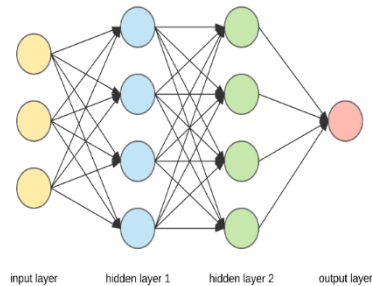
## Support Vector Machine



## Random Forest



## Neural Network



## Bayesian Classifier

THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE

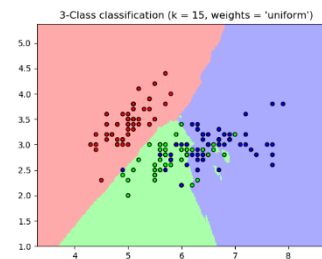
THE PROBABILITY OF "A" BEING TRUE GIVEN THAT "B" IS TRUE

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

THE PROBABILITY OF "A" BEING TRUE

THE PROBABILITY OF "B" BEING TRUE

## k-Nearest Neighbor



## Boosting

