

# LECTURE NOTES

**COMP8014**

**Knowledge Data Discovery**

**Anomaly Detection**

## LEARNING OUTCOMES

LO4: Mahasiswa diharapkan mampu untuk menentukan metode data mining yang sesuai dengan sifat kealamiahannya permasalahan yang dihadapi.

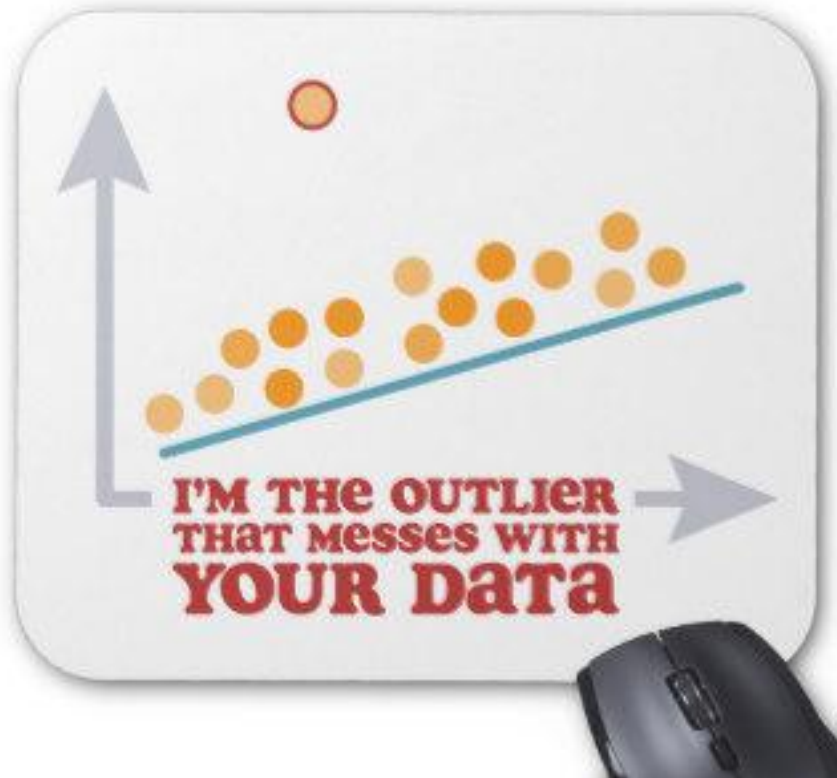
LO5: Mahasiswa diharapkan mampu untuk mengimplementasikan metode data mining.

LO6: Mahasiswa mampu mengevaluasi aplikasi dan trend Data Mining.

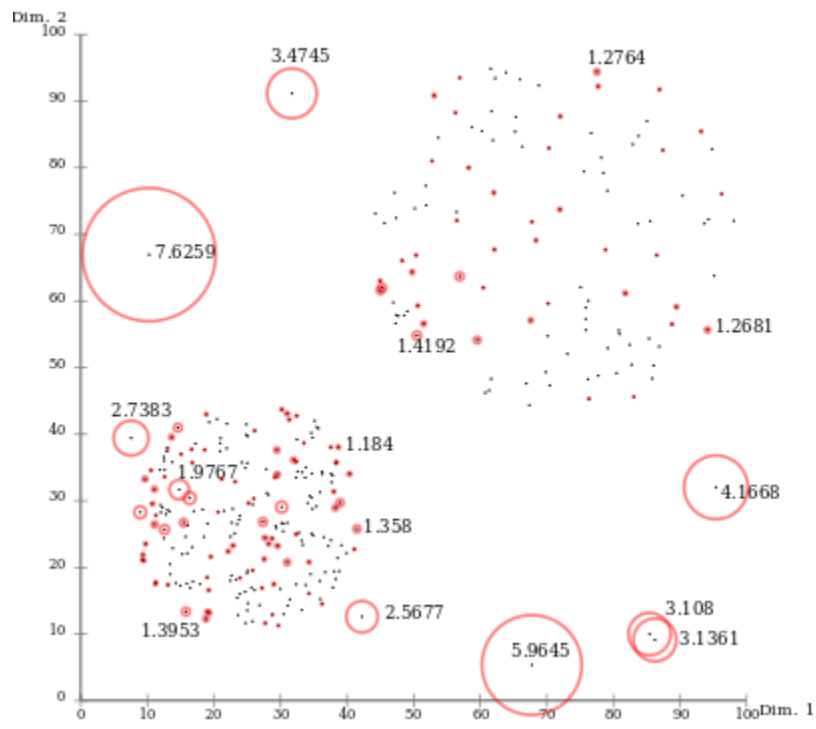
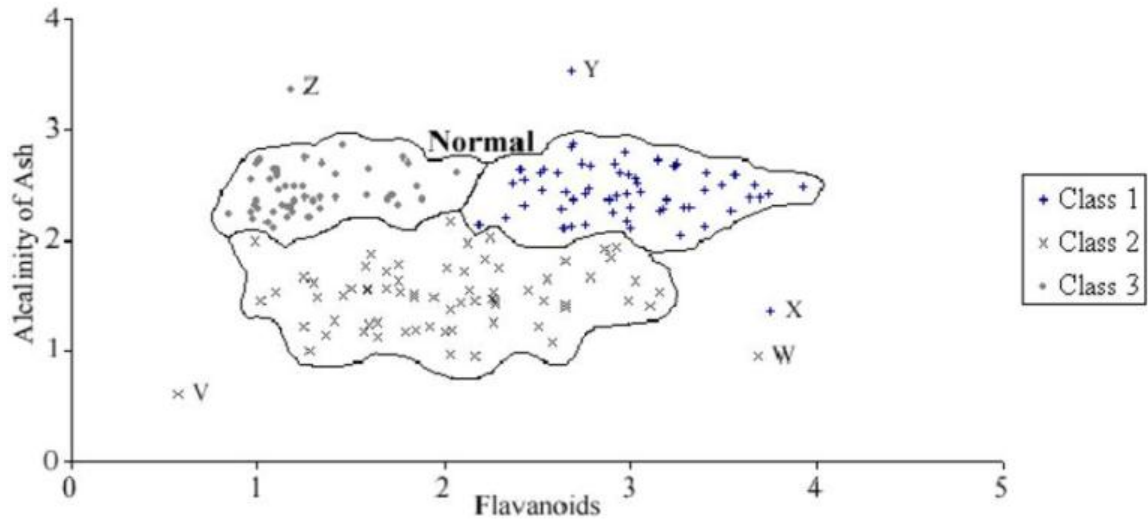
### OUTLINE MATERI :

1. Tinjauan
2. Analisa Outlier
3. Metode Deteksi Outlier
4. Studi Kasus

## ISI MATERI



Graph Showing Outlier Classification - Type 2



Apakah Outlier itu?

- Outlier: Objek data yang menyimpang secara signifikan dari objek normal seolah-olah dihasilkan oleh mekanisme yang berbeda
  - Contoh: Pembelian kartu kredit yang tidak biasa, olahraga: Michael Jordan, Lionel Messi
- Keberadaan outlier menarik karena outlier melanggar mekanisme yang menghasilkan data normal
- Outlier detection vs Novelty Detection: tahap awal, outlier; Namun kemudian bergabung ke dalam model
  - **Outlier detection**; Data pelatihan berisi outlier, dan kita perlu menyesuaikan mode pusat data pelatihan, dengan mengabaikan pengamatan yang menyimpang.
  - **Novelty Detection**: Data pelatihan tidak tercemar oleh outlier, dan kami tertarik untuk mendeteksi anomali dalam pengamatan baru.

#### **Aplikasi dari deteksi outlier:**

- Deteksi penipuan kartu kredit
- Deteksi kecurangan Telecom
- Segmentasi pelanggan
- Analisis medis

#### **Applications Deteksi Outliers?**

- Credit card fraud detection
- Telecom fraud detection
- Customer segmentation
- Medical analysis

### **Metode Outliers Detection?**

- Whether user-labeled examples of outliers can be obtained
  - Supervised method --- Modeling outlier detection as a classification problem
  - semi-supervised method – menggunakan jarak
  - unsupervised method – KNN+Fuzzy Clustering
- Assumptions about normal data and outliers
  - Statistical,
  - proximity-based
  - clustering based methods

## SIMPULAN

1. Jenis outlier
  - outlier global, kontekstual & kolektif
2. Deteksi outlier
  - diawasi, diawasi semi, atau tidak diawasi
3. Pendekatan statistik (atau model)
4. Pendekatan jarak dekat
5. Pendekatan Clustering-base
6. Pendekatan klasifikasi
7. Pertambahan kontekstual dan outlier kolektif
8. Deteksi outlier pada data berdimensi tinggi

## DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, Y. (2006). "Data Mining: Concepts and Technique". Edisi 3. Morgan Kaufman. San Francisco
2. Tan, P.N., Steinbach, M., & Kumar, V. (2006). "Introduction to Data Mining". Addison-Wesley. Michigan
3. Witten, I. H., & Frank, E. (2005). "Data Mining : Practical Machine Learning Tools and Techniques". Second edition. Morgan Kaufmann. San Francisco