

**TESIS**

**KLASTERING DATA INTERKOMPARASI ANAK TIMBANGAN  
MENGUNAKAN ALGORITMA *SELF ORGANIZING MAPS***



Disusun oleh:

**Nama : Arif Fajar Solikin**  
**NIM : 20.52.1311**  
**Konsentrasi : Business Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM  
YOGYAKARTA  
2022**

**TESIS**

**KLASTERING DATA INTERKOMPARASI ANAK TIMBANGAN  
MENGUNAKAN ALGORITMA *SELF ORGANIZING MAPS***

***WEIGHTS INTERCOMPARATION DATA CLUSTERING USING  
SELF ORGANIZING MAPS ALGORITHM***

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama : Arif Fajar Solikin**  
**NIM : 20.52.1311**  
**Konsentrasi : Business Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM**

**YOGYAKARTA**

**2022**

**HALAMAN PENGESAHAN**

**KLASTERING DATA INTERKOMPARASI ANAK TIMBANGAN  
MENGUNAKAN ALGORITMA *SELF ORGANIZING MAPS***

***WEIGHTS INTERCOMPARATION DATA CLUSTERING USING  
SELF ORGANIZING MAPS ALGORITHM***

Dipersiapkan dan Disusun oleh

**Arif Fajar Sollkin**

**20.52.1311**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, tanggal 5 Januari 2022

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 5 Januari 2022

**Rektor**

**Prof. Dr. M. Suyanto, M.M.**

**NIK. 190302001**

## HALAMAN PERSETUJUAN

### KLASTERING DATA INTERKOMPARASI ANAK TIMBANGAN MENGUNAKAN ALGORITMA *SELF ORGANIZING MAPS*

### *WEIGHTS INTERCOMPARATION DATA CLUSTERING USING SELF ORGANIZING MAPS ALGORITHM*

Dipersiapkan dan Disusun oleh

**Arif Fajar Solikin**

**20.52.1311**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, 5 Januari 2022

**Pembimbing Utama**

**Anggota Tim Penguji**

**Prof. Dr. Kusriani, M. Kom.**  
NIK. 190302106

**Prof. Dr. Ena Utami, S.Si., M. Kom.**  
NIK. 190302037

**Pembimbing Pendamping**

**Alva Hendi Muhammad, S.T., M. Eng, Ph. D**  
NIK. 190302493

**Ferry Wahyu Wibowo, S.Si., M.Cs.**  
NIK. 190302235

**Prof. Dr. Kusriani, M. Kom.**  
NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 5 Januari 2022  
**Direktur Program Pascasarjana**

**Prof. Dr. Kusriani, M. Kom.**  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan di bawah ini,

Nama mahasiswa : Arif Fajar Solikin  
NIM : 20.52.1311  
Konsentrasi : Business Intelligence

Menyatakan bahwa tesis dengan judul berikut:

**"Klastering Data Interkomparasi Dengan Menggunakan Algoritma Self Organizing Maps"**

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M. Kom  
Dosen Pembimbing Pendamping : Ferry Wahyu Wibowo, S.Si., M.Cs.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Bantul, 05 Januari 2022

Yang menyatakan,

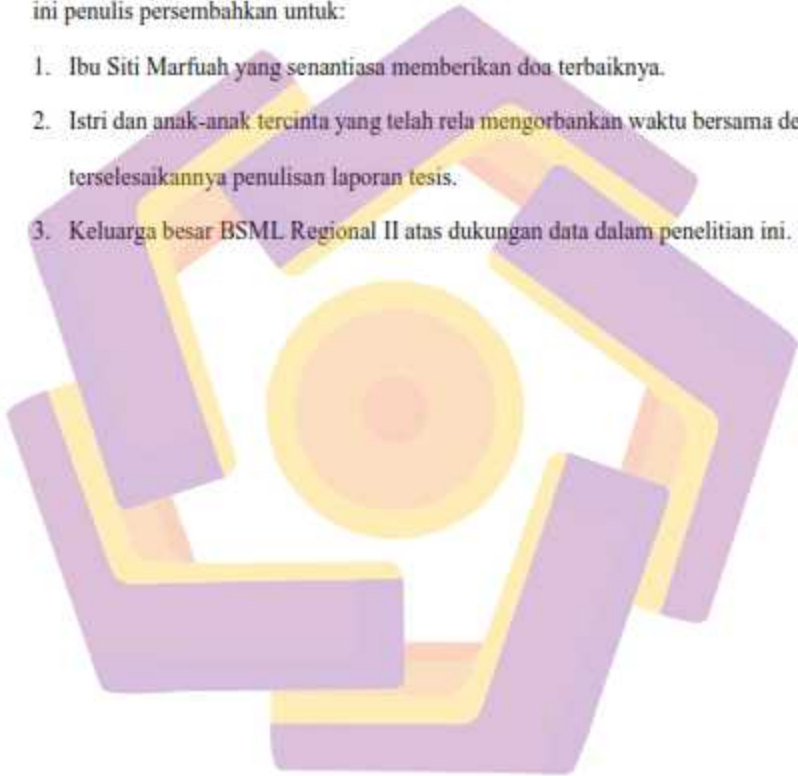


Arif Fajar Solikin

## HALAMAN PERSEMBAHAN

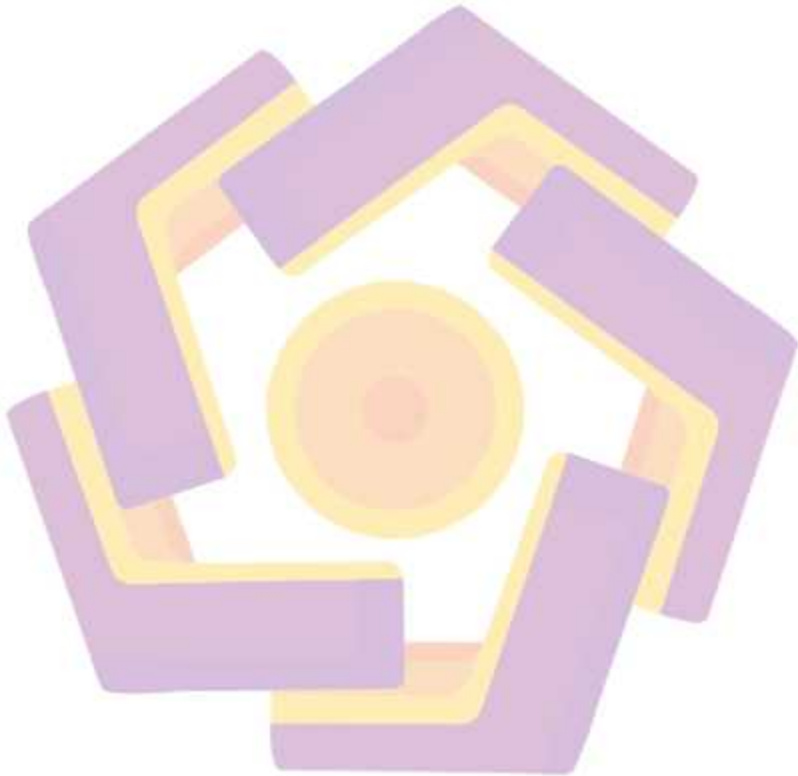
Segala puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan nikmat dan karuniaNya sehingga dapat menyelesaikan tesis ini. Tesis ini penulis persembahkan untuk:

1. Ibu Siti Marfuah yang senantiasa memberikan doa terbaiknya.
2. Istri dan anak-anak tercinta yang telah rela mengorbankan waktu bersama demi terselesaikannya penulisan laporan tesis.
3. Keluarga besar BSML Regional II atas dukungan data dalam penelitian ini.



## **HALAMAN MOTTO**

Fokus pada Solusi, bukan pada masalah. Masalah selalu ada, yang terpenting bagaimana reaksi atau respon kita terhadap masalah tersebut.



## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat, taufik dan hidayahNya kepada penulis, sehingga penulis dapat menyelesaikan penulisan tesis dengan judul: "Klastering Data Interkomparasi Anak Timbangan Dengan Menggunakan Algoritma *Self Organizing Maps*. Sholawat serta salam senantiasa tercurah kepada Baginda Muhammad SAW, Nabi khatimul anbiya yang senantiasa kita nantikan syafa'atnya kelak di yaumul akhir.

Penulis menyampaikan ucapan terimakasih dan penghargaan yang setinggi-tingginya kepada Ibu yang telah memberikan doa terbaiknya dan istri yang senantiasa mendukung dan memberikan semangat untuk menyelesaikan kuliah S2 ini secara tepat waktu. Ucapan terima kasih juga penulis sampaikan kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. sebagai rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M. Kom sebagai Direktur Program Pasca Sarjana Magister Teknik Informatika Universitas AMIKOM Yogyakarta dan sekaligus sebagai dosen pembimbing I yang telah meluangkan waktu dalam membimbing dan memberikan arahan hingga terselesaikannya penulisan tesis ini.
3. Bapak Ferry Wahyu Wibowo, S. Si, M.Cs. sebagai dosen pembimbing II yang telah memberikan bimbingan arahan dan dukungan terhadap penulis.
4. Ibu Prof. Dr. Emma Utami, S.Si., M.Kom dan Bapak Alva Hendi Muhammad, S.T., M.Eng., Ph.D sebagai Dosen Penguji.



5. Seluruh dosen, staff akademik dan rekan-rekan mahasiswa Magister Teknik Informatika yang telah menjalankan sistem perkuliahan di Universitas AMIKOM Yogyakarta secara baik.

Akhirnya dengan kerendahan hati penulis menyadari bahwa penulisan tesis ini masih banyak kekurangan, sehingga kritik dan saran yang bersifat membangun sangat penulis harapkan. Harapannya tesis ini nantinya dapat memberikan manfaat bagi masyarakat, instansi dan bidang pendidikan. Sekian dan penulis sampaikan terima kasih.

Bantul, 05 Januari 2022

Penulis



## DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
INTISARI.....	xiv
<i>ABSTRACT</i> .....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	7
1.5. Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA.....	8
2.1. Tinjauan Pustaka.....	8
2.2. Keaslian Penelitian.....	11

2.3. Landasan Teori.....	14
2.3.1. Interkomparasi.....	14
2.3.2. Anak Timbangan .....	16
2.3.3. Machine Learning.....	19
2.3.4. <i>Self-Organizing Maps</i> .....	21
2.3.5. Evaluasi Cluster .....	27
<b>BAB III METODE PENELITIAN.....</b>	<b>33</b>
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	33
3.2. Metode Pengumpulan Data.....	34
3.3. Metode Analisis Data.....	36
3.4. Alur Penelitian .....	40
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>	<b>43</b>
4.1. Data Penelitian.....	43
4.2. Klastering.....	45
4.3. Evaluasi Hasil Klastering.....	52
4.4. Uji Korelasi Pearson .....	59
4.5. Analisa Data.....	61
<b>BAB V PENUTUP.....</b>	<b>70</b>
5.1. Kesimpulan .....	70
5.2. Saran .....	70
<b>DAFTAR PUSTAKA .....</b>	<b>71</b>
<b>LAMPIRAN.....</b>	<b>74</b>

## DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian.....	11
Tabel 2.1. Lanjutan .....	12
Tabel 2.1. Lanjutan .....	13
Tabel 2.2. Nilai Batas Kesalahan yang Di ijjinkan (BKD) anak timbangan.....	18
Tabel 3.1. Data interkomparasi artefak 1 kg.....	35
Tabel 3.2. Data nilai budget uncertainty artefak 1 kg.....	38
Tabel 3.3. Kriteria korelasi pearson.....	39
Tabel 4.1. Identitas artefak A.....	43
Tabel 4.2. Rerata nilai data interkomparasi .....	44
Tabel 4.3. Hasil percobaan nilai parameter.....	46
Tabel 4.4. Nilai parameter Clustering Dengan Algoritma SOM .....	49
Tabel 4.5. Jumlah kluster dan anggota kluster setiap parameter.....	51
Tabel 4.6. Jumlah kluster berdasar ukuran grid .....	52
Tabel 4.7. Data anggota per cluster parameter kedua .....	53
Tabel 4.8. Data perhitungan nilai jarak cluster ke pusat cluster .....	55
Tabel 4.9. Matriks R.....	57
Tabel 4.10. Nilai DBI hasil klastering .....	58
Tabel 4.11. Analisis hasil pengujian korelasi pearson .....	60

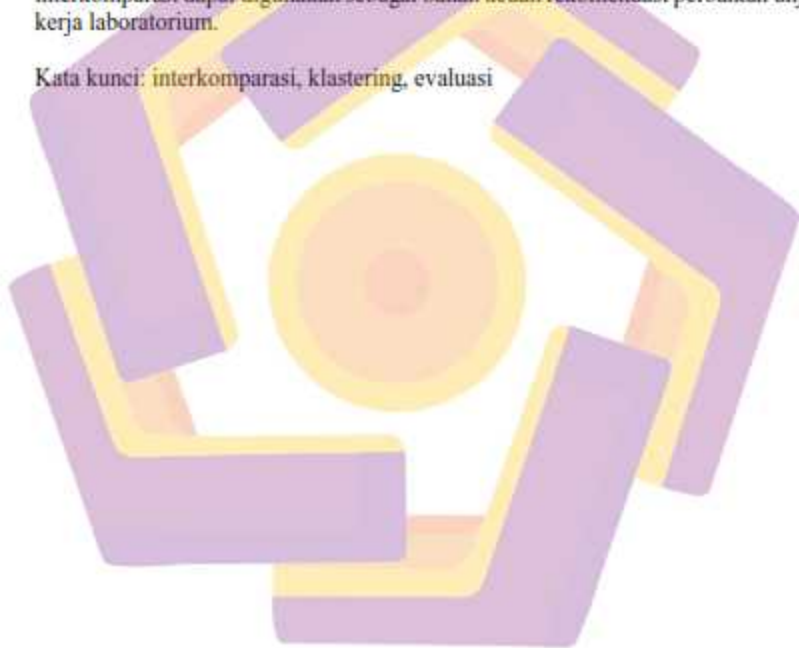
## DAFTAR GAMBAR

Gambar 2.1. Arsitektur SOM .....	23
Gambar 2.2. Rectangular grid .....	24
Gambar 2.3. Hexagonal grid .....	25
Gambar 2.4. Cara kerja SOM.....	25
Gambar 2.5. Representasi cluster .....	28
Gambar 2.6. Datasets DBI .....	30
Gambar 3.1. Clustering dengan KNIME .....	37
Gambar 3.2. Alur Penelitian.....	42
Gambar 4.1. Menu pengaturan parameter pada hub Self Organizing Maps .....	45
Gambar 4.2. Hasil klaster dengan parameter $\alpha = 0.8$ , $E_o$ 6000 dan $E_c$ 2000 .....	47
Gambar 4.3. Hasil klaster dengan parameter $\alpha = 0.8$ , $E_o$ 6000 dan $E_c$ 20000 .....	47
Gambar 4.4. Hasil klaster dengan parameter $\alpha = 0.8$ , $E_o$ 12000 dan $E_c$ 2000 .....	48
Gambar 4.5. Hasil Clustering Parameter 5.....	50
Gambar 4.6. Grafik nilai DBI .....	58
Gambar 4.7. Uji korelasi dengan platform KNIME.....	59
Gambar 4.8. Hasil pengujian korelasi pearson pada cluster0 .....	60
Gambar 4.9. Statistika data parameter ke 4.....	62
Gambar 4.10. Hasil klaster parameter ke-4.....	62
Gambar 4.11. Persebaran data awal .....	64
Gambar 4.12. Persebaran data hasil klastering .....	65
Gambar 4.13. Hasil klastering data nilai sumber ketidakpastian P4.....	66
Gambar 4.14. Statistika hasil klastering sumber ketidakpastian P4 .....	67

## INTISARI

Data interkomparasi adalah data hasil pengujian artefak interkomparasi yang diikuti sebanyak 36 laboratorium metrologi legal. Evaluasi kegiatan interkomparasi selama ini dinyatakan dengan metode Degree of Equivalent atau derajat kesetaraan yang ditunjukkan dengan nilai rasio  $E_0$ . Untuk mengetahui insight data interkomparasi dan menggali informasi tentang data tersebut telah dilakukan klastering data interkomparasi dengan menggunakan algoritma Self Organizing Maps. Algoritma ini memetakan data ke peta 2 dimensi dengan hasil pengujian nilai DBI sebesar 1,803 pada ukuran grid  $2 \times 4$ . Hasil klastering data interkomparasi dapat digunakan sebagai bahan acuan rekomendasi perbaikan unjuk kerja laboratorium.

Kata kunci: interkomparasi, klastering, evaluasi



## **ABSTRACT**

*Intercomparison data is data from the results of intercomparative artifact testing which was followed by 36 legal metrology laboratories. Evaluation of intercomparison activities so far has been stated by the degree of equivalent method that indicated by the ratio value of  $E_n$ . To find insight of intercomparative data and explore information about the data, intercomparative data clustering has been carried out using the self-organizing maps algorithm. This algorithm mapping data into a 2-dimensional map with the results of testing using DBI. The DBI value's is 1.803 on a grid size of  $2 \times 4$ . The results of intercomparative data clustering can be used as reference material for recommendations to improvement laboratory performance.*

*Keyword: intercomparison, clustering, evaluation*



# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Transaksi perdagangan yang dilakukan antara pembeli dan penjual biasanya melibatkan alat ukur dalam penentuan kuantitas barang / jasa. Harga dari barang / jasa yang diserahkan dari penjual ke pembeli dipengaruhi oleh nilai berat atau kuantitas barang yang ditentukan dari alat ukur milik penjual. Alat ukur tersebut disahkan oleh Unit Metrologi Legal (UML) Pemerintah Kabupaten / Kota berdasarkan Undang-Undang No 23 Tahun 2014 tentang Pemerintah Daerah. Untuk menjamin alat ukur yang digunakan dalam transaksi perdagangan memiliki nilai yang sama antara satu daerah dengan daerah lain, setiap unit metrologi legal kabupaten / kota melakukan verifikasi internal standar kerja dengan standar ukuran acuan yang dimiliki oleh masing-masing UML yang tertelusur secara berjenjang ke standar satuan tingkat 4 hingga ke standar satuan tingkat 1.

Selain menjamin ketertelusuran standar di setiap UML, kemampuan atau unjuk kerja dari masing-masing UML juga dapat diketahui melalui kegiatan interkomparasi. Menurut ISO/IEC 17043:2010, interkomparasi atau *interlaboratory comparisons* (ILC) adalah organisasi, kinerja, dan evaluasi pengukuran atau pengujian pada item yang sama atau serupa oleh dua atau lebih laboratorium atau lembaga inspeksi sesuai dengan kondisi yang telah ditentukan. Tujuan dari interkomparasi adalah untuk mengukur kemampuan UPT dan UML Kabupaten / Kota dalam melakukan pengelolaan standar ukuran maupun tera / tera



ulang alat ukur takar timbang dan perlengkapannya (UTTP) dan untuk membangun saling pengakuan yang didasarkan pada keseragaman hasil pengukuran metrologi legal di Indonesia. Kegiatan interkomparasi dipaparkan dalam satu prosedur yang disebut dengan protokol interkomparasi. Protokol interkomparasi mengatur terkait penyelenggara, laboratorium acuan, peserta, desain program (jadwal pelaksanaan), penentuan artefak, prosedur pengujian, pelaporan, dan evaluasi hasil. Pelaporan yang dimaksud adalah pelaporan dari masing-masing peserta interkomparasi berupa laporan pengambilan data dan pengolahan data yang disajikan dalam bentuk cerapan dan surat keterangan hasil pengujian. Lingkup interkomparasi dapat dilakukan untuk satu jenis alat UTTP, seperti anak timbangan. Untuk kegiatan interkomparasi anak timbangan, laporan yang disajikan memuat nilai massa konvensional beserta nilai ketidakpastian pada tingkat kepercayaan 95%. Data ini kemudian dievaluasi menggunakan metode derajat kesetaraan (*Degrees of Equivalent / DoE*) atau yang lebih dikenal sebagai rasio  $E_n$  mengacu pada ISO/IEC 17043:2010. Hasilnya berupa pernyataan kesetaraan laboratorium peserta interkomparasi, laboratorium dinyatakan setara jika nilai  $E_n \leq [1]$ . Sampai saat ini belum pernah dilakukan analisis data interkomparasi melalui *data mining*, analisis data yang dilakukan baru sebatas pada perhitungan nilai  $E_n$  yang akan menentukan tingkat kesetaraan laboratorium peserta interkomparasi. Untuk memberikan evaluasi yang mendalam dan kelengkapan informasi terkait permasalahan-permasalahan yang mungkin terjadi atau dialami oleh laboratorium UML dilakukan analisis *data mining* dengan memanfaatkan *machine learning* salah satunya adalah dengan *clustering*.

Pada penelitian ini, penulis merancang satu penelitian dengan melakukan klastering data hasil interkomparasi menggunakan algoritma *self-organizing maps*. Klastering atau pengelompokan data dilakukan supaya data hasil interkomparasi dari masing-masing laboratorium yang memiliki permasalahan yang sama dapat diidentifikasi. Identifikasi perlu dilakukan oleh penyelenggara interkomparasi agar dapat memberikan bimbingan dan rekomendasi perbaikan pengelolaan laboratorium metrologi legal kepada Unit Metrologi Legal yang menjadi peserta interkomparasi. Penggunaan algoritma *self-organizing maps* didasarkan pada cara kerja algoritma dalam menyederhanakan dimensi data dan hubungan antar *neuron* pada *layer input* dan *output* yang secara terus menerus memperbarui bobotnya sehingga diperoleh kesamaan ciri antar anggota dalam *cluster* yang sama. Penggunaan algoritma *self-organizing maps (SOM)* juga didasarkan pada penelitian sebelumnya yang membandingkan beberapa metode *clustering*, antara lain penelitian (Asriny Novendri Isra et al,2021) yang membandingkan metode *K-Medoids* dengan metode SOM dalam mengelompokkan bencana alam hidrometeorologi di Pulau Jawa. Hasil penelitian menunjukkan bahwa algoritma *self-organizing maps* merupakan metode yang tepat karena memiliki nilai standar deviasi yang lebih kecil dibanding metode *K-Medoids*. Penelitian lain (Marina Marsudi Putri,2015) yang membandingkan pengelompokan kabupaten/kota di Jawa timur berdasarkan indeks kesehatan masyarakat menggunakan metode SOM dan *K-Means* menunjukkan bahwa algoritma *self-organizing maps* menghasilkan jumlah *optimum cluster* dan tingkat *homogenitas internal* dan *heterogenitas eksternal cluster* yang lebih baik dibandingkan metode *K-Means*.

Penerapan algoritma *self-organizing maps* telah dilakukan pada penelitian-penelitian sebelumnya, antara lain: Penelitian yang dilakukan oleh (Ni Putu Nanik, 2018) dengan algoritma *self-organizing maps* dan *k means* dimana  $k=2$  dan bobot awal yang ditentukan adalah sama untuk kedua metode, telah menghasilkan ketepatan klasifikasi yang sama. Proses *clustering* dilakukan dengan menentukan *vector node* yang dipilih untuk selanjutnya ditentukan *best matching unit* (BMU) nya dengan *node* lain. Penelitian lainnya oleh (IY Purbasari et al, 2019) diawali dengan data yang dinormalisasi pada rentang 0 sampai dengan 1. Data yang sudah dinormalisasi kemudian dilakukan *clustering* dengan menetapkan jumlah iterasi dan nilai *learning rate*-nya. Pembuatan aplikasi *clustering* pada data IPM Aceh menggunakan *bubble function* sebagai *neighborhood function* dimana radius *neighborhood* sama dengan nol. Fungsi *bubble* dipilih dengan alasan proses komputasi lebih cepat dan kuantisasi kesalahan lebih kecil (Kania Siti et al, 2019). Penelitian selanjutnya (Firmansyah et al, 2019) adalah pengelompokan wilayah berdasarkan kesejahteraan sosial menggunakan *Silhouette Coefficient* untuk mengukur kualitas hasil *clustering* dengan membatasi nilai rentang lebih dari 1. Penelitian yang dilakukan oleh (Sivakkumaran L., 2020) mencoba menentukan pengelompokan penggunaan BBM berdasar data manuver pengemudi kendaraan dengan menggunakan perluasan algoritma SOM yaitu *super organizing maps*. Dimana data input dibedakan menjadi 4 layer, masing-masing layer kemudian dilakukan proses pengelompokan. Penelitian terakhir oleh (Jin Tian & Mengyi Gu, 2019) menentukan subruang pada *neuron* dengan tujuan mereduksi *noise* pada data. Algoritma SOM menghasilkan *cluster* yang mungkin dari setiap dimensi, kemudian

dilakukan proses penggabungan untuk menggabungkan file kelompok *neuron* dan subruang yang sesuai dengan dimensi terkait. Hasil pengelompokan akhir disimpulkan dari *cluster neuron* dengan mengganti setiap *neuron* dengan poin data di dalamnya.

Penelitian-penelitian tersebut telah menerapkan algoritma *self-organizing maps* untuk pengelompokan data dengan tujuan yang berbeda-beda. Saat ini evaluasi kegiatan interkomparasi didasarkan pada kesetaraan nilai berdasarkan perhitungan nilai rasio  $E_0$ . Pada penelitian yang akan dilakukan, evaluasi kegiatan interkomparasi dilakukan dengan teknik *clustering* menggunakan algoritma *Self Organizing Maps*. Kelompok yang dihasilkan dari proses *clustering* ditentukan oleh berapa banyak anggota *cluster* yang memiliki kedekatan atau kesamaan ciri dengan anggota lainnya. Pengelompokan dilakukan pada 2 jenis data yaitu data asli, dan data sumber ketidakpastian (*repeatability*, *drift standard* dan *standard* yang digunakan, koreksi udara / *bouyancy*, dan resolusi timbangan). Data asli digunakan untuk melihat hasil pengelompokan pada kondisi apa adanya. Data sumber ketidakpastian digunakan untuk melihat kelompok variabel yang berpengaruh dominan pada kegiatan interkomparasi. Proses klastering diharapkan mampu memberikan bahan acuan untuk rekomendasi perbaikan unjuk kerja laboratorium peserta interkomparasi.

## 1.2. Rumusan Masalah

Berdasarkan penjabaran pada latar belakang masalah, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

- a. Dapatkah data interkomparasi dievaluasi dengan teknik *clustering* menggunakan algoritma *Self Organizing Maps* dan seberapa baik hasil *clustering* diuji dengan *Davies Bouldin Index*?
- b. Apakah evaluasi hasil *clustering* dengan algoritma *Self Organizing Maps* dapat dijadikan data atau referensi dalam evaluasi kegiatan interkomparasi?
- c. Variabel apakah yang paling berpengaruh pada kegiatan interkomparasi berdasarkan hasil *clustering* menggunakan algoritma *Self Organizing Maps*?

### 1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Penelitian ini menggunakan algoritma *Self Organizing Maps* dalam meng-*cluster* data interkomparasi anak timbangan.
- b. Data interkomparasi diambil dari kegiatan interkomparasi BSML Regional II tahun 2021 dengan jumlah peserta 36 laboratorium metrologi legal di wilayah BSML Regional II.
- c. Evaluasi hasil *clustering* dilakukan dengan pengujian nilai *DBI*.
- d. Variabel data yang dievaluasi dalam penelitian ini terdiri dari nilai massa konvensional artefak, dan nilai *uncertainty* yang bersumber dari *uncertainty* anak timbangan standar, *drift* standar, kemampuan ulang penimbangan (*repeatability*), *buoyancy udara* dan daya baca timbangan elektronik.
- e. Pengujian terhadap pengaruh setiap variabel menggunakan uji korelasi (Pearson Moment Product).
- f. Implementasi algoritma *Self Organizing Maps* menggunakan platform KNIME.

#### 1.4. Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah sebagai berikut:

- a. Mengetahui performa penerapan algoritma *Self Organizing Maps* dalam evaluasi kegiatan interkomparasi laboratorium unit metrologi legal di wilayah regional II.
- b. Mendapatkan pola evaluasi kegiatan interkomparasi laboratorium metrologi legal berdasarkan hasil *analisis clustering* data interkomparasi anak timbangan menggunakan *algoritma Self Organizing Maps*.

#### 1.5. Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut:

- a. Laboratorium mampu melakukan analisis data interkomparasi dengan teknik *clustering* menggunakan algoritma *Self Organizing Maps*.
- b. Evaluasi interkomparasi laboratorium metrologi legal dengan teknik *clustering* dapat digunakan secara lebih luas tidak terbatas hanya di wilayah Regional II namun juga bisa digunakan secara nasional.
- c. Penyelenggara kegiatan interkomparasi laboratorium dapat lebih efisien dan efektif dalam melakukan evaluasi.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Penelitian ini menggunakan beberapa referensi sumber pustaka yang berasal dari penelitian sebelumnya yang membahas tentang clustering. Pustaka yang relevan dalam penelitian ini ditinjau dari algoritma yang digunakan yaitu algoritma *self-organizing maps*. Penggunaan algoritma *self-organizing maps* telah dilakukan pada pengelompokan pelajar SMU berdasarkan nilai akademik yang diambil dari laporan prestasi belajar (raport) dan nilai ujian nasional 275 pelajar untuk nilai matematika, IPA, Bahasa Inggris, Bahasa Indonesia dan IPS. *Clustering* diawali dengan normalisasi data pada rentang 0 sampai dengan 1 dan menetapkan jumlah iterasi maksimum dan nilai *learning rate* - nya. Penelitian ini menghasilkan 3 kelompok pelajar yang terbagi ke 3 jurusan antara lain IPA, IPS dan bahasa dengan anggota cluster terbesar pada jurusan IPA (IY Purbasari et al, 2020). Penelitian serupa (Rusydi et al, 2018) dilakukan pada pelajar SMP dengan tujuan membantu memberikan rekomendasi pilihan jurusan di SMK berdasar pengelompokan skill bakat dan minat calon siswa dengan algoritma *self-organizing maps*. Pemilihan algoritma SOM didasarkan pada pertimbangan metode *self-organizing maps* tidak memerlukan fungsi objektif seperti *K-Means* dan *Fuzzy CMeans* pada kondisi dimana iterasi telah mencapai tingkat optimal, SOM akan terus berjalan sampai target iterasi yang ditentukan terpenuhi. Kelebihan algoritma SOM dibanding algoritma lain juga telah dikemukakan dari penelitian (Novendri Isra et al, 2021)

yang membandingkan algoritma *K-Medoids* dengan algoritma SOM dalam pengelompokan bencana alam hidrometeorologi di Pulau Jawa. Algoritma SOM memiliki rasio standar deviasi yang lebih baik dibandingkan dengan algoritma *K-Medoids*. Rasio standar deviasi dihitung dari perbandingan standar deviasi dalam kelompok dengan standar deviasi antar kelompok ( $R = \frac{S_w}{S_b}$ ). Perbandingan algoritma SOM dengan algoritma *K-Means* dilakukan oleh (Ruth Ema Febrita et al, 2019) yang menunjukkan algoritma SOM memiliki kemampuan lebih baik dalam membangun validitas *cluster*, dimana data direpresentasikan pada bentuk peta topologi 2 dimensi. Data ditempatkan pada *neuron*, sehingga jarak antar anggota dalam satu *cluster* dan jarak antar *cluster* lebih mudah diukur. Selain perbandingan, terdapat penelitian yang menggabungkan penggunaan dua algoritma atau disebut juga dengan hybrid methods. Seperti yang dilakukan (Ni Putu Nanik Hendayanti et al, 2018) dalam menggabungkan algoritma SOM dengan algoritma *K-Means* untuk mengelompokkan data calon penerima beasiswa STMIK STIKOM Bali. Prinsip kerja dari metode ini adalah menentukan anggota kluster dengan metode SOM dan menghitung nilai pusat dari masing-masing kluster. Nilai pusat yang diperoleh dari masing-masing kluster selanjutnya digunakan untuk menentukan keanggotaan kluster pada algoritma *K-Means*. Kesamaan dari dua algoritma dalam penelitian tersebut terletak pada perhitungan jarak dengan menggunakan *euclidean distance*. Pada SOM, vektor input dihitung jaraknya dengan masing-masing node yang telah ditentukan bobotnya dengan *euclidean distance* dan dipilih jarak minimum sebagai *best matching unit* (BMU). Sementara penggunaan *euclidean distance* pada *K-Means* terletak pada perhitungan jarak antara pusat data dengan anggota kluster.



Penelitian (Sekta Lonir Oscarini Watibhakti et al, 2019) membandingkan penggunaan algoritma SOM dengan algoritma K-Means dalam pengelompokkan hasil tes TOEFL berdasar perbedaan nilai P (performance of global optimum solution) yang merupakan perbandingan nilai variansi centroid dengan nilai variansi anggota kluster.

Evaluasi clustering algoritma SOM dilakukan dengan menggunakan perhitungan *silhouette coefficient* (Dese Narfa Firmansyah et al, 2019), dimana rentang nilai *silhouette coefficient* terletak pada nilai -1 sampai dengan 1. Nilai positif terbagi dalam tiga kategori (*weak, medium, strong*). Nilai negatif menunjukkan banyak data yang klusternya salah tempat. *Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik atau buruknya suatu obyek ditempatkan dalam suatu *cluster* (DA Indah Cahya Dewi et al, 2019). Metode untuk mengevaluasi kluster dapat dilakukan dengan *Davies Bouldin Index* (DBI). *Davies-Bouldin Index* (DBI) merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster* pada suatu metode *clustering* (Bernard Jumadi, 2018). Evaluasi menggunakan *Davies-Bouldin Index* ini memiliki skema evaluasi internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*. Nilai *Davies Bouldin Index* (DBI) yang semakin mendekati nilai 0 menandakan semakin baik *cluster* yang diperoleh. Semakin rendah nilai DBI menunjukkan hasil *cluster* yang optimal (DA Indah Cahya Dewi et al, 2019).

## 2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian  
Klastering Data Interkomparasi Anak Timbangan Menggunakan Algoritma *Self Organizing Maps*

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Ketepatan Klasifikasi Penerima Beasiswa Stikom Bali Dengan <i>Hybrid Self Organizing Maps</i> Dan <i>Algoritma K-Means</i>	Ni Putu Nanik Hendayanti; Gusti Ayu Made Ama Putri, Jurnal Varian Vol.2 No.1 Oktober 2018	Untuk menentukan ketepatan klasifikasi penerima beasiswa dengan menggunakan metode hybrid SOM dan K Means	Ketepatan klasifikasi metode SOM, K Means dan hybrid SOM K Means dibandingkan dengan data riil untuk ketiga metode sama sebesar 54,45%	Nilai pusat yang digunakan untuk ketiga metode adalah sama. Perlu ditentukan nilai pusat yang berbeda-beda untuk setiap metode dengan lebih dari 2 klaster	Pada penelitian ini algoritma yang dilakukan dengan menentukan inisial bobot pada masing-masing node, kemudian vector node terpilih ditentukan jarak terdekatnya sebagai BMU. Terakhir ditentukan node tetangga BMU. Penelitian yang akan dilakukan oleh penulis lebih detail pada jumlah iterasi pada masing-masing fase, nilai learning rate dan status normalisasinya.
2	<i>Using Self-Organizing Map (SOM) for Clustering and Visualization of New Students based on Grades</i>	I Y Purbasari et al, <i>Journal of Physics: Conference Series</i> , IOP Publishing, 2020	Untuk menentukan jurusan pada pelajar SMU dengan pengelompokkan nilai akademik menggunakan metode SOM	Penentuan jurusan siswa dengan metode SOM telah dilakukan dengan hasil terbesar pada jurusan IPA kemudian IPS dan Bahasa	Perlu dicoba pada nilai learning rate lain dan perbaikan pada metode normalisasi data.	Penelitian ini menggunakan matriks 10x10, $\alpha = 0.6$ dan $tm_{max} = 5000$ . Pada penelitian tesis tidak membatasi ukuran layer, nilai $\alpha$ dan jumlah iterasinya. Beberapa skenario dicoba untuk mendapatkan cluster terbaik

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Pengelompokan Wilayah Berdasarkan Kesejahteraan Sosial Menggunakan Algoritme <i>Self-Organizing Maps</i> Dengan Perbaikan <i>Missing Value K-Nearest Neighbors</i>	DN Firman-syah et al, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Unibraw Vol. 3, No. 7, Juli 2019	Mencentukan cluster kesejahteraan sosial kabupaten / kota dengan metode SOM dan KNN untuk data kosong	Terbentuk 2 cluster dengan proporsi 1:37 dari 38 data	Proses interpretasi masih dilakukan secara manual. Untuk otomatisasi interpretasi dapat dicoba dengan decision tree. Hasil clustering tidak konsisten, saran yang diberikan dengan genetic algoritma untuk optimasi bobot awal.	Penelitian ini memperbaiki <i>missing value</i> dengan KNN. Sementara penelitian pada tesis tidak menggunakan KNN. Data penelitian dimungkinkan bernilai nol tetapi bukan <i>missing value</i> .
4	Program Aplikasi Pengelompokan Objek Dengan Metode <i>Self Organizing Map</i> Menggunakan Bahasa R	Siti Kania et all, Jurnal EurekaMatika, Vol.7 No.2, 2019	Membandingkan program aplikasi clustering metode SOM dengan Bahasa R pada data IPM Provinsi Aceh	Program aplikasi Bahasa R menghasilkan jumlah cluster yang sama dengan anggota cluster yang berbeda	Penggunaan <i>neighborhood</i> dan <i>learning function</i> yang berbeda antara dua aplikasi. Sehingga perlu dilakukan dengan <i>function</i> yang sama	Dalam penelitian ini penulis menggunakan <i>Gauss neighborhood function</i> untuk penelitian yang akan dilakukan
5	Analisis Perbandingan Metode <i>Elbow</i> dan <i>Silhouette</i> pada Algoritma <i>Clustering K-Medoids</i> dalam Pengelompokan Produksi Kerajinan Bali	Dewa Ayu Indah Cahya Dewi, Dewa Ayu Kadek Pramita, JURNAL MATRIX, VOL. 9, NO. 3, 2019	menentukan jumlah cluster yang tepat sehingga menghasilkan kualitas cluster yang optimal	clustering k-medoid dengan silhouette menghasilkan kualitas cluster lebih baik karena memiliki nilai DBI lebih rendah daripada clustering k-medoid dengan metode elbow	Metode elbow tepat digunakan untuk menentukan jumlah cluster, tetapi metode silhouette lebih tepat digunakan untuk mengukur kualitas hasil clustering	Pentuan jumlah cluster dan kualitas cluster menggunakan pengujian nilai DBI

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	<i>High Dimensional Data Clustering using Self-Organized Map</i>	Ruth Ema Febrita; Wayan Firdaus Mahmudy; Aji Prasetya Wibawa. Knowledge Engineering and Data Science (KEDS) Vol 2, No 1, Juni 2019	Untuk memberikan rekomendasi harga rumah berdasar pengelompokkan fitur rumah	Algoritma SOM mampu menunjukkan kinerja yang lebih baik dibanding algoritma K-Means berdasar nilai silhouette dan DB Index	Algoritma SOM perlu perbaikan untuk menghasilkan cluster yang lebih baik.	Terdapat perbedaan jenis data dan atribut yang digunakan dalam pengelompokkan

## 2.3. Landasan Teori

### 2.3.1 Interkomparasi

Interkomparasi atau *interlaboratory comparison* adalah organisasi, kinerja dan evaluasi pengukuran atau tes pada dua laboratorium atau lebih sesuai dengan kondisi yang telah ditentukan (ISO 17043, 2010). Interkomparasi atau disebut juga dengan istilah uji profisiensi telah menjadi aspek penting dalam praktik laboratorium di semua bidang pengujian, kalibrasi, dan inspeksi. Sebagian besar skema uji profisiensi memiliki ciri umum berupa perbandingan hasil dari satu laboratorium dengan hasil dari satu atau lebih laboratorium yang berbeda. Sifat pengujian atau pengukuran yang dilakukan dalam interkomparasi dimaksudkan untuk membandingkan kinerja laboratorium. Dapat disimpulkan bahwa tujuan dari interkomparasi adalah untuk mengetahui kemampuan atau unjuk kerja dari suatu laboratorium.

Ada tiga jenis pemeriksaan laboratorium dasar: kuantitatif, kualitatif dan interpretatif (R.M. Voiculescu, 2013).

- a. Hasil pengukuran kuantitatif adalah numerik dan dilaporkan dalam skala interval atau rasio. Pengujian untuk pengukuran kuantitatif dapat bervariasi dalam presisi, kebenaran, kepekaan analitis, dan kekhususan. Dalam skema pengujian profisiensi kuantitatif, hasil numerik biasanya dianalisis secara statistik.
- b. Hasil tes kualitatif bersifat deskriptif dan dilaporkan dalam skala kategorikal atau ordinal, misalnya identitas mikro-organisme, atau dengan identifikasi adanya ukuran tertentu (seperti obat atau penilaian

suatu karakteristik). Penilaian kinerja dengan analisis statistik mungkin tidak sesuai untuk pemeriksaan kualitatif.

- c. Dalam tes interpretatif, uji profisiensi berupa hasil tes seperti pernyataan morfologi deskriptif, atau satu set data (misalnya untuk menentukan garis kalibrasi) atau satu set informasi lainnya misalnya studi kasus tentang fitur interpretatif dari kompetensi peserta.

Hasil pengukuran numerik dari laboratorium diolah secara statistik dengan menggunakan nilai rasio  $E_n$ . Nilai rasio  $E_n$  dihitung dengan rumus (Protokol Interkomparasi, 2021):

$$E_n = \left| \frac{x_i - x_{ref}}{\sqrt{U_i^2 + U_{ref}^2}} \right| \dots \dots \dots (1)$$

Keterangan:

$x_i$  merupakan nilai koreksi anak timbangan pada laboratorium peserta dengan  $i$  merujuk pada urutan no peserta interkomparasi.

$x_{ref}$  merupakan nilai koreksi anak timbangan pada laboratorium referensi.

$U_i$  merupakan ketidakpastian laboratorium peserta pada level kepercayaan 95%.

$U_{ref}$  merupakan ketidakpastian laboratorium referensi pada level kepercayaan 95%.

Hasil perhitungan nilai  $E_n$  berupa pernyataan kesetaraan laboratorium peserta interkomparasi. Laboratorium dinyatakan setara jika nilai  $E_n \leq 1$ .

### 2.3.2 Anak Timbangan

Anak timbangan adalah benda ukur massa yang karakteristik fisik dan sifat kemetrologiannya diatur oleh peraturan. Anak timbangan terdiri dari beberapa kelas yang ditentukan berdasarkan karakter fisik dan sifat kemetrologiannya. Kelas yang dimaksud adalah kelas yang dibuat untuk membatasi nilai massanya pada batas tertentu (SK Dirjen PKTN 123 Tahun 2020). Kelas anak timbangan terdiri dari:

- a. Anak timbangan kelas  $E_1$  adalah anak timbangan yang mampu telusur ke standar massa K46 dan digunakan untuk verifikasi anak timbangan kelas  $E_2$ . Standar massa K46 adalah anak timbangan yang terbuat dari campuran logam platina iridium, berbentuk silinder, memiliki diameter dan tinggi yang sama yaitu 39 mm, memiliki tulisan grafit "46" pada bagian badannya dan mampu telusur ke standar internasional.
- b. Anak timbangan kelas  $E_2$  adalah anak timbangan yang digunakan untuk verifikasi anak timbangan kelas  $F_1$  dan digunakan bersama dengan timbangan kelas akurasi I.
- c. Anak timbangan kelas  $F_1$  adalah anak timbangan yang digunakan untuk verifikasi anak timbangan kelas  $F_2$  dan digunakan bersama dengan timbangan kelas akurasi I atau II.
- d. Anak timbangan kelas  $F_2$  adalah anak timbangan yang digunakan untuk verifikasi anak timbangan kelas  $M_1$  dan digunakan bersama dengan timbangan kelas akurasi II.

- e. Anak timbangan kelas  $M_1$  adalah anak timbangan yang digunakan untuk verifikasi anak timbangan kelas  $M_2$  dan digunakan bersama dengan timbangan kelas akurasi III.
- f. Anak timbangan kelas  $M_2$  adalah anak timbangan yang digunakan untuk verifikasi anak timbangan kelas  $M_3$  dan digunakan bersama dengan timbangan kelas akurasi III.
- g. Anak timbangan kelas  $M_3$  adalah anak timbangan yang digunakan bersama dengan timbangan kelas akurasi III atau IIII.

Anak timbangan mempunyai nilai massa nominal dan nilai massa konvensional. Nilai massa nominal adalah nilai yang berupa bilangan bulat yang digunakan untuk menandakan massa anak timbangan. Nilai massa konvensional adalah nilai massa hasil penimbangan di udara pada kondisi konvensional yaitu suhu  $20^{\circ}\text{C}$ , massa jenis udara sebesar  $1,2 \text{ kg/m}^3$ , dan massa jenis benda yang ditimbang sebesar  $8000 \text{ kg/m}^3$ . Selain itu anak timbangan juga memiliki nilai koreksi dan nilai ketidakpastian. Koreksi adalah nilai yang harus ditambahkan untuk mendapatkan nilai yang benar. Ketidakpastian adalah suatu nilai yang menunjukkan sebaran hasil pengukuran yang dikaitkan dengan besaran yang diukur. Sedangkan timbangan adalah alat ukur yang dipergunakan untuk menentukan massa suatu benda dengan memanfaatkan gravitasi yang bekerja pada benda tersebut.

Berikut tabel batas kesalahan yang diijinkan atau nilai koreksi anak timbangan sesuai dengan SK Dirjen PKTN No 123 Tahun 2020.



Tabel 2.2 Nilai Batas Kesalahan yang Di ijjinkan (BKD) anak timbangan

M.Nominal (g)	E1 (mg)	E2 (mg)	F1 (mg)	F2 (mg)	M1 (mg)	M2 (mg)	M3 (mg)
25000	12.5	37.5	250	375	1250	3750	12500
20000 *	10	30	100	300	1000	3000	10000
20000	10	30	100	300	1000	3000	10000
10000	5	16	50	160	500	1600	5000
5000	2.5	8	25	80	250	800	2500
2000 *	1	3	10	30	100	300	1000
2000	1	3	10	30	100	300	1000
1000	0.5	1.6	5	16	50	160	500
500	0.25	0.8	2.5	8	25	80	250
200 *	0.1	0.3	1	3	10	30	100
200	0.1	0.3	1	3	10	30	100
100	0.05	0.16	0.5	1.6	5	16	50
50	0.03	0.1	0.3	1	3	10	30
20 *	0.025	0.08	0.25	0.8	2.5	8	25
20	0.025	0.08	0.25	0.8	2.5	8	25
10	0.02	0.06	0.2	0.6	2	6	20
5	0.016	0.05	0.16	0.5	1.6	5	16
2 *	0.012	0.04	0.12	0.4	1.2	4	12
2	0.012	0.04	0.12	0.4	1.2	4	12
1	0.01	0.03	0.1	0.3	1	3	10
0.5	0.008	0.025	0.08	0.25	0.8	2.5	0
0.2 *	0.006	0.02	0.06	0.2	0.6	2.0	0
0.2	0.006	0.02	0.06	0.2	0.6	2.0	0
0.1	0.005	0.016	0.05	0.16	0.5	1.6	0
0.05	0.004	0.012	0.04	0.12	0.4	0	0
0.02 *	0.003	0.01	0.03	0.1	0.3	0	0
0.02	0.003	0.01	0.03	0.1	0.3	0	0
0.01	0.003	0.008	0.025	0.08	0.25	0	0
0.005	0.003	0.006	0.02	0.06	0.2	0	0
0.002 *	0.003	0.006	0.02	0.06	0.2	0	0
0.002	0.003	0.006	0.02	0.06	0.2	0	0
0.001	0.003	0.006	0.02	0.06	0.2	0	0

### 2.3.3 *Machine Learning*

Ilmuwan komputer dari Stanford University, Herbert A. Simon dan Pat Langley mendefinisikan *machine learning* sebagai kumpulan metode komputasi yang mampu meningkatkan kinerja melalui pemanfaatan pengetahuan yang berasal dari pengalaman, dengan kata lain tujuan utama dari *machine learning* adalah menaikkan tingkat otomatisasi dari proses rekayasa pengetahuan. Otomatisasi tersebut diharapkan dapat menggantikan kegiatan manusia yang membutuhkan waktu banyak dengan teknik otomatis yang meningkatkan akurasi atau efisiensi melalui pembelajaran dan memanfaatkan pola yang dipelajari dari data *training*.

Di dalam konteks *machine learning*, proses pembelajaran atau *training model* merupakan sebuah proses dimana sebuah algoritma *machine learning* melakukan pembelajaran konsep. Metode pembelajaran atau *training model machine learning* dikelompokkan ke dalam tiga kategori utama sebagai berikut:

#### a. *Supervised learning*

Merupakan metode pembelajaran yang bertujuan untuk memprediksi variabel target sebuah data baru berdasarkan model hasil *training* menggunakan himpunan data berlabel. *Supervised learning* biasanya dipergunakan untuk menyelesaikan masalah komputasi yang dikelompokkan ke dalam dua kategori yaitu masalah klasifikasi dan masalah regresi yang bertujuan untuk memprediksi variabel target berupa data kategorikal.

*b. Unsupervised learning*

Merupakan metode pembelajaran menggunakan sejumlah contoh data tanpa label. *Unsupervised learning* biasanya dipergunakan untuk menyelesaikan masalah komputasi yang dapat dikelompokkan ke dalam dua kategori yaitu:

1. klustering (*clustering*) adalah mengelompokkan data training ke dalam beberapa klaster berdasarkan keiratan atau kesamaan fitur.
2. pengurangan dimensi data (*dimensional reduction*) adalah mengelompokkan variabel input data ke dalam beberapa kelompok variabel yang tidak saling berkorelasi.

*c. Reinforcement learning*

Metode pembelajaran yang didasarkan pada pemberian penghargaan terhadap keputusan yang diinginkan dan atau pemberian pinalti terhadap keputusan yang tidak diinginkan.

Data sebagai input bagi algoritma pembelajaran machine learning harus dapat dibaca oleh komputer, dalam format bilangan, dan berupa tabel. Ketika data akan dipergunakan sebagai data training atau data testing dari proses pembelajaran machine learning, beberapa transformasi data seringkali diperlukan agar model hasil training mencapai kinerja yang baik. Beberapa metode transformasi data terpenting antara lain: diskretisasi (transformasi data dari jenis data bilangan real menjadi data ordinal), menangani nilai NaN (*Not a Number*) dan *rescaling* (transformasi data yang bertujuan agar nilai variabel data memiliki skala atau *range* data yang

sama. Beberapa metode *rescaling* yang umum digunakan di dalam *machine learning* antara lain:

#### 1. Standardisasi

Standardisasi adalah transformasi nilai setiap variabel data sehingga data hasil transformasi memiliki nilai tengah 0 dan deviasi standar 1 menggunakan persamaan

$$x'_i = \frac{x_i - \bar{x}}{s_x} \dots \dots \dots (2)$$

Dimana,  $x'_i$  adalah data hasil normalisasi,  $x_i$  adalah data asli,  $\bar{x}$  adalah data rata-rata, dan  $s_x$  adalah deviasi standar dari data  $x$ .

#### 2. Normalisasi

Normalisasi adalah transformasi nilai setiap variabel data sehingga data hasil transformasi berada pada interval  $[0,1]$  menggunakan persamaan

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \dots \dots \dots (3)$$

Dimana,  $\min(x)$  adalah nilai minimum dari variabel data  $x$  dan  $\max(x)$  adalah nilai maksimum dari variabel  $x$ .

Rescaling data bertujuan untuk meningkatkan kecepatan konvergensi algoritma pembelajaran dan menghindari dominasi variabel dengan rentang nilai besar terhadap variabel dengan rentang nilai kecil di dalam perhitungan fungsi jarak antara dua data. (Heryadi Yaya & Wahyono Teguh, 2020)

#### 2.3.4 Self organizing maps

*Self-Organizing Maps*, disingkat (SOM) atau sering disebut *topology-preserving map* dikenalkan oleh seorang professor asal Finlandia Teuvo Kohonen, yang memperoleh gelar doktor dalam bidang teknik dari *Helsinki University of*

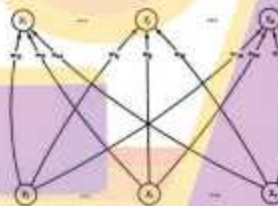
*Technology*. SOM merupakan salah satu teknik dalam *neural network* yang bertujuan mengurangi dimensi data melalui penggunaan *self-organizing neural network* sehingga manusia mampu memahami data berdimensi tinggi yang dipetakan dalam bentuk data berdimensi rendah. Properti utama dari *neural network* adalah kemampuan untuk belajar dari lingkungannya, dan untuk meningkatkan kinerjanya melalui pembelajaran. Terdapat pembelajaran *supervised* dimana supervisor menyediakan data *training* ke *network*, dan pembelajaran *unsupervised* atau pembelajaran tanpa pengawasan dimana pembelajaran diatur sendiri (*self organised*). Selama pembelajaran *neural network* menerima sejumlah pola input yang berbeda dan menemukan sejumlah fitur dari pola-pola tersebut dan belajar bagaimana mengklasifikasikan pola input ke kategori yang sesuai. Metode pembelajaran ini cenderung mengikuti organisasi *neuro biological* pada otak yang bertujuan untuk belajar dengan cepat bahkan jauh lebih cepat daripada jaringan *back propagation* dan dapat digunakan secara real time (Negnevitsky Michael, 2011).

Pada algoritma SOM, vektor bobot untuk setiap unit *cluster* berfungsi sebagai contoh dari input pola yang terkait dengan *cluster* tersebut. Selama proses *self-organizing*, *cluster* satuan yang bobotnya sesuai dengan pola vektor input yang paling dekat dipilih sebagai pemenang. Unit pemenang dan unit tetangganya terus memperbaharui bobotnya, setiap output akan bereaksi terhadap pola input tertentu sehingga hasil SOM akan menunjukkan adanya kesamaan ciri antar anggota dalam *cluster* yang sama (Irwansyah Edy, Faisal M, 2015). Dalam jaringan SOM, *neuron* target tidak diletakkan dalam sebuah baris seperti layaknya model *neural network*

yang lain. *Neuron* target diletakkan dalam dua dimensi yang bentuk /topologinya dapat diatur. Topologi yang berbeda akan menghasilkan *neuron* sekitar *neuron* pemenang yang berbeda sehingga bobot yang dihasilkan juga akan berbeda. Perubahan bobot tidak hanya dilakukan pada bobot garis yang terhubung ke *neuron* pemenang saja, tetapi juga pada bobot garis ke *neuron-neuron* di sekitarnya. *Neuron* di sekitar *neuron* pemenang ditentukan berdasarkan jaraknya dari *neuron* pemenang.

#### 2.3.4.1 Arsitektur dan Topologi SOM

Arsitektur SOM merupakan jaringan yang terdiri dari dua lapisan (layer), yaitu lapisan input dan lapisan output. Setiap neuron dalam lapisan input terhubung dengan setiap neuron pada lapisan output seperti pada gambar 1. Setiap neuron dalam lapisan output merepresentasikan kelas (cluster) dari input yang diberikan.



Gambar 2.1. Arsitektur SOM (Fausett 1993)

Sedangkan untuk topologi, SOM memiliki 3 jenis topologi hubungan ketetanggaan (*neighborhood*) yaitu *linear array*, *rectangular* dan *hexagonal grid*.

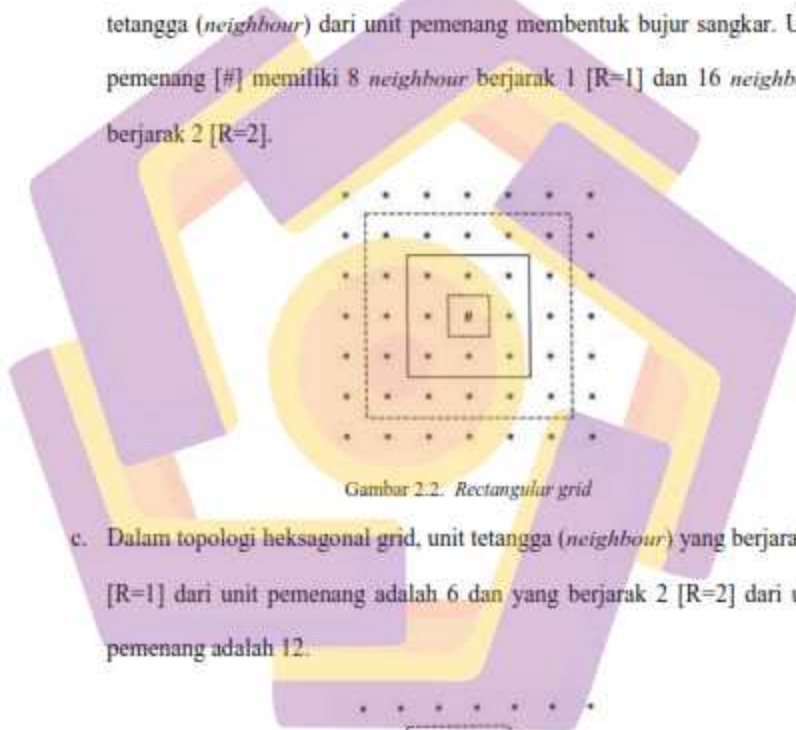
- a. Linear array, topologi ini menunjukkan unit cluster yang tersusun secara linear. Unit cluster yang menjadi pemenang [#] memiliki dua unit tetangga

(neighbour) yang berjarak 1 [R=1], dan mempunyai dua unit tetangga yang berjarak 2 [R=2].

\*\*\* { \* ( [#] \* ) \* } \*\*

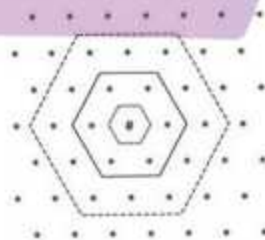
Keterangan: [ ] : R=0 ; ( ) : R=1 ; { } : R=2

- b. *Rectangular grid*, adalah topologi dari *cluster* unit dua dimensi. Unit tetangga (*neighbour*) dari unit pemenang membentuk bujur sangkar. Unit pemenang [#] memiliki 8 *neighbour* berjarak 1 [R=1] dan 16 *neighbour* berjarak 2 [R=2].



Gambar 2.2. *Rectangular grid*

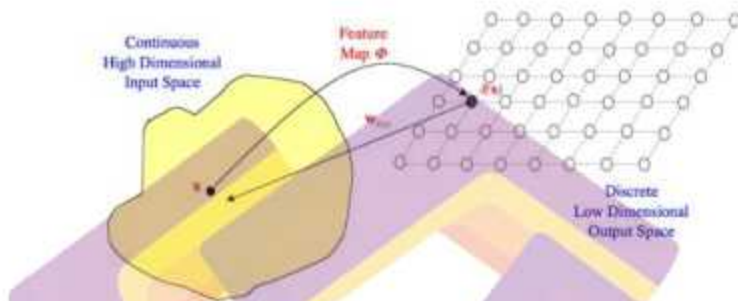
- c. Dalam topologi heksagonal grid, unit tetangga (*neighbour*) yang berjarak 1 [R=1] dari unit pemenang adalah 6 dan yang berjarak 2 [R=2] dari unit pemenang adalah 12.



Gambar 2.3. *Hexagonal grid*

### 2.3.4.2 Cara kerja SOM

Secara umum, cara kerja SOM ditunjukkan oleh Gambar 4 dibawah ini:



Gambar 2.4. Cara kerja SOM

Titik  $(x)$  di ruang masukan dipetakan ke titik  $I(x)$  di ruang keluaran. Setiap titik  $(I)$  di ruang keluaran akan dipetakan ke titik yang sesuai di ruang masukan dengan bobot  $w_i(x)$ .

Menurut Haykin (1999), SOM mengandung tiga komponen penting, yaitu:

- Persaingan: Untuk setiap mode masukan, neuron akan menghitung nilai dari setiap fungsi diskriminan, sehingga memberikan dasar untuk persaingan. Neuron spesifik dengan nilai minimum fungsi diskriminan dinyatakan sebagai pemenang.
- Kerjasama: Neuron pemenang menentukan lokasi spasial dari lingkungan topologi neuron yang tereksitasi, dengan demikian memberikan dasar untuk kerjasama di lingkungan neuron.
- Adaptasi sinaptik: Neuron tereksitasi mengurangi nilai fungsi diskriminan yang terkait dengan mode masukan melalui penyesuaian bobot yang sesuai, sehingga



neuron yang menang akan meningkatkan respons ke aplikasi berikutnya dengan mode masukan yang sama.

Algoritma SOM dapat dijelaskan sebagai berikut, pengelompokan data menggunakan algoritma SOM terdiri dari 4 tahap, yaitu:

- a. Kompetisi: setiap simpul output  $j$ , dihitung nilai  $D(x, w_j)$  yang merupakan fungsi jarak Euclidian antara  $x$  dan  $w_j$ . Fungsi ini didefinisikan sebagai berikut:

$$D(x, w_j) = \sqrt{\sum_{i=1}^n (x_i - w_{ji})^2} \dots\dots\dots (4)$$

dimana  $x$  adalah vector dari node input, sedangkan  $w_m$  adalah vector bobot dari node neuron ke  $m$

- b. Update bobot: setelah mendapat nilai jarak dari tiap-tiap vector input ke vector bobot, nilai jarak yang minimum dipilih sebagai neuron pemenang. Setiap neuron pemenang beserta tetangganya dilakukan proses adaptasi dengan memperbaharui nilai bobot dimana  $h(t)$  adalah fungsi node tetangga (*neighborhood function*) dan  $t$  adalah banyaknya iterasi. Fungsi node tetangga yang digunakan adalah fungsi Gauss (kohonen et al, 2001) dengan formula:

$$h(t) = \alpha(t) e^{-\frac{\|r_i - r_c\|^2}{2\delta^2(t)}} \dots\dots\dots (5)$$

dimana  $\alpha(t)$  adalah nilai laju pembelajaran atau biasa disebut nilai alpha. Laju pembelajaran adalah fungsi penurunan tingkat pembelajaran seiring perubahan waktu (Fausett 1993).

$\|r_i - r_c\|^2$  adalah jarak kuadrat antara neuron ke- $i$  dengan neuron pemenang dalam grid dan  $\delta^2(t)$  adalah lebar tetangga. Nilai laju pembelajaran diperoleh dari

$$\alpha(t) = \alpha_i \left(1 - \frac{t}{t_{max}}\right)$$

dimana  $\alpha_l$  adalah nilai awal laju pembelajaran dan  $t_{max}$  adalah iterasi maksimum.

Perubahan lebar tetangga didapat dari perhitungan berikut ini:

$$\delta(t) = \delta_l \left( \frac{\delta_r}{\delta_l} \right)^{\frac{t}{t_{max}}}$$

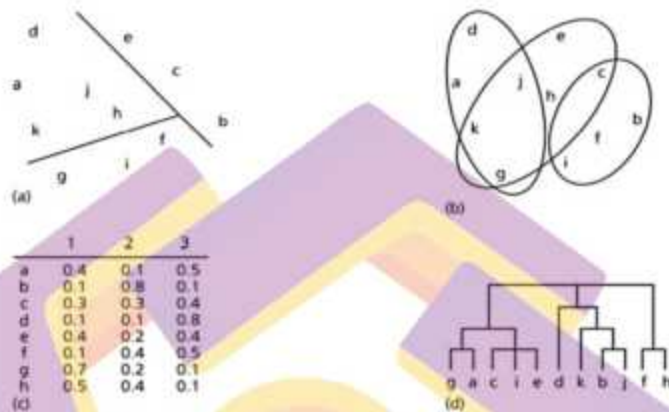
dimana  $\delta(t)$  adalah lebar tetangga yang akan berkurang seiring dengan t langkah pembelajaran.  $\delta_l$  adalah nilai awal lebar tetangga dan  $\delta_r$  nilai akhir lebar tetangga.

- c. Perbaharui  $\alpha$  dan  $\delta$ .  $\alpha$  dan  $\delta$  dilakukan perubahan nilai dengan rumus perubahan laju pembelajaran dan perubahan lebar tetangga seperti yang tersebut diatas.
- d. Hentikan perlakuan: perlakuan dihentikan ketika kriteria pemberhentian dicapai. Kriteria pemberhentian dari proses training dapat ditetapkan berdasarkan banyaknya iterasi, nilai minimum error, nilai alpha dan lebar tetangga.

### 2.3.5 Evaluasi Cluster

*Clustering* adalah proses pengelompokkan sekumpulan objek sedemikian rupa sehingga objek-objek tersebut berada dalam kelompok yang sama (Dony Jollyta et al, 2021). Definisi lain (Manning et al, 2009), data dalam sebuah *cluster* harus semirip mungkin dan data dalam satu *cluster* harus sebeda mungkin dari data dalam *cluster* lainnya. Penyelesaian *cluster* dilakukan dengan berbagai algoritma. Beberapa algoritma pengelompokkan memungkinkan satu *instance* menjadi milik

lebih dari satu *cluster*, sehingga diagram mungkin meletakkan *instance* dalam dua dimensi dan menggambar *subset* yang tumpang tindih mewakili setiap *cluster*.



Gambar 2.5. Representasi *cluster*

(Witten & Frank, 2005)

Gambar 5 menunjukkan beberapa bentuk representasi *cluster* yang sangat bergantung pada variasi data yang dikelompokkan dan algoritma yang digunakan. *Cluster* digunakan di banyak bidang penelitian seperti penggalian data, analisis data statistik, pembelajaran mesin, pola pengenalan, analisis gambar dan pencarian informasi. Umumnya *cluster* menyertakan kelompok dengan jarak kecil antar anggota *cluster*, area padat ruang data, interval atau distribusi statistik tertentu. Oleh karenanya pengelompokkan dapat dirumuskan sebagai multi-tujuan masalah pengoptimalan. Algoritma pengelompokkan yang sesuai dan pengaturan parameter bergantung pada kumpulan data individu dan tujuan penggunaan hasil. Hal ini memperlihatkan adanya kesulitan dalam menggali pengetahuan atau informasi dari hasil *cluster*. Permasalahan yang terjadi adalah belum adanya ketentuan jumlah

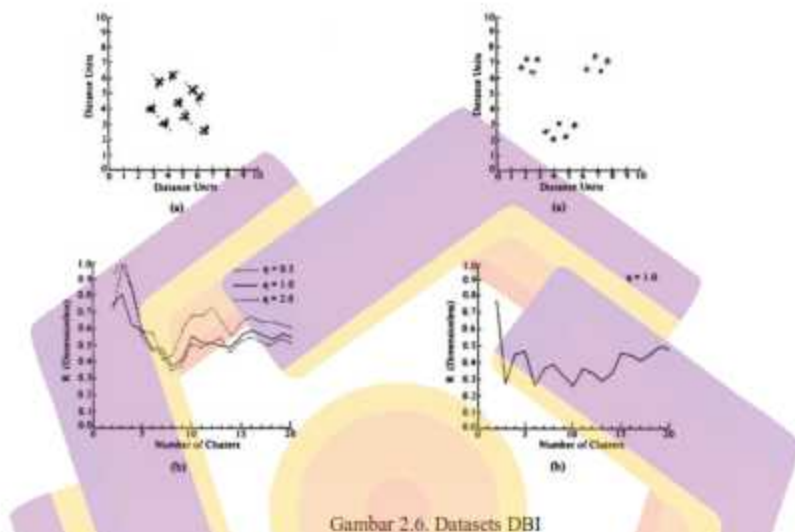
*cluster* yang tepat dari sejumlah  $k$  yang diujikan pada data. Kesulitan dalam menentukan *cluster* terbaik mengarahkan pada penggunaan teknik evaluasi *cluster*. Teknik evaluasi *cluster* digunakan untuk menentukan jumlah *cluster* yang terbaik atau optimal, dimana setiap teknik memiliki cara tersendiri dalam menemukan *cluster* yang optimal. Teknik evaluasi *cluster* tersebut diantaranya adalah metode *Elbow* dengan *Sum of Squares Error* (SSE), *Davies Bouldin Index* (DBI) dan *Silhouette Index* (SI).

#### **2.3.5.1 Davies Bouldin Index**

Davies-Bouldin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. DBI adalah matrik untuk mengevaluasi algoritma pengelompokkan. Matrik ini menunjukkan kesamaan *cluster* yang diasumsikan memiliki kepadatan data yang merupakan fungsi penurunan jarak dari karakteristik vektor *cluster*. Ukuran tersebut dapat digunakan untuk menyimpulkan kesesuaian partisi data dan oleh karena itu dapat digunakan untuk membandingkan kesesuaian relatif dari berbagai divisi data. Pengukuran tidak bergantung pada jumlah *cluster* yang dianalisis maupun metode partisi data dan dapat digunakan untuk memandu algoritma pencarian *cluster* (Deny Jollyta et al, 2021).

DBI merupakan skema evaluasi internal, dimana validasi seberapa baik pengelompokkan dilakukan dengan menggunakan kuantitas dan fitur yang melekat pada dataset. Hal ini menjadi kelemahan DBI dimana nilai terendah yang dihasilkan dari perhitungannya belum tentu *cluster* yang terbaik atau optimal. Untuk mengatasi kelemahan ini, data yang digunakan harus memiliki jarak *cluster* minimum yang

dapat diterima atau deviasi standar minimum yang dapat diterima (Davies & Bouldin, 1979).



Di penjelasan lain DBI merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokkan yang didasarkan pada nilai kohesi dan separasi. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar titik pusat *cluster* dari *cluster*.

Evaluasi dilakukan melalui skema perhitungan jarak *internal cluster*, dimana jarak *internal* terbaik merupakan kuantitas dan kedekatan terhadap titik pusat *cluster* dari *cluster* yang diikuti. Sedangkan jarak antar *cluster* didasarkan pada jarak antar titik pusat *cluster* ke  $i$  terhadap titik pusat *cluster* ke  $j$ . Prinsip yang digunakan pada *Davies-Bouldin Index* ini jarak *inter-cluster* antara *cluster*  $C_i$  dan

$C_j$  dimaksimalkan dan pada waktu yang sama jarak antar titik dalam sebuah *cluster* diminimalkan. Jika jarak *inter-cluster* maksimal, berarti kesamaan karakteristik antar-masing-masing *cluster* sedikit sehingga perbedaan antar *cluster* terlihat lebih jelas. Jika jarak *intra-cluster* minimal berarti masing-masing objek dalam *cluster* tersebut memiliki tingkat kesamaan karakteristik yang tinggi (Wani & Riyaz, 2017). Tahapan dari perhitungan Davies Bouldin Index adalah sebagai berikut:

a. *Sum of Square Within-cluster (SSW)*

Untuk mengetahui kohesi dalam sebuah *cluster* ke- $i$  adalah dengan menghitung nilai dari *Sum of Square Within-cluster (SSW)*. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster* adalah sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \dots\dots\dots (6)$$

dimana  $m$  = jumlah anggota dalam cluster

b. *Sum of Square Between-cluster (SSB)*

Perhitungan *Sum of Square Between-cluster (SSB)* bertujuan untuk mengetahui separasi antar *cluster*. Persamaan yang digunakan untuk menghitung nilai *Sum of Square Between cluster* adalah sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j) \dots\dots\dots (7)$$

c. *Ratio (Rasio)*

Bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$ . Untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*, digunakan persamaan berikut:

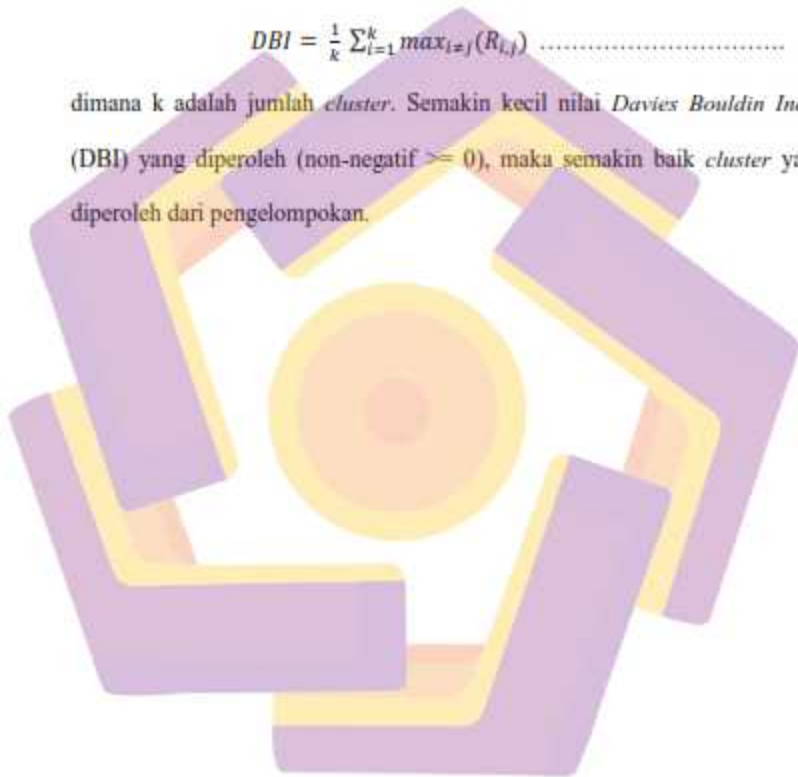
$$R_{l,j} = \frac{SSW_l + SSW_j}{SSB_{l,j}} \dots\dots\dots (8)$$

d. *Davies Bouldin Index*

Nilai rasio yang diperoleh dari persamaan (7) digunakan untuk mencari nilai *Davies-Bouldin Index* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{l=1}^k \max_{l \neq j} (R_{l,j}) \dots\dots\dots (9)$$

dimana k adalah jumlah *cluster*. Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik *cluster* yang diperoleh dari pengelompokan.



## BAB III

### METODE PENELITIAN

#### 3.1. Jenis, Sifat, dan Pendekatan Penelitian

##### 3.1.1 Jenis Penelitian

Jenis-jenis metode penelitian dapat diklasifikasikan berdasarkan, tujuan, dan tingkat kealamiahannya (*natural setting*) obyek yang diteliti. Berdasarkan tujuan, metode penelitian dapat diklasifikasikan menjadi penelitian dasar (*basic research*), penelitian terapan (*applied research*) dan penelitian pengembangan (*research and development*). Selanjutnya berdasarkan tingkat kealamiahannya, metode penelitian dapat dikelompokkan menjadi metode penelitian eksperimen, survey dan naturalistic (Sugiyono, 2013). Penelitian ini termasuk jenis penelitian eksperimen. Penelitian yang dilakukan dengan menerapkan serangkaian tindakan untuk membuktikan suatu konsep. Dapat berupa pengembangan sistem atau penerapan algoritma baru atau penerapan parameter baru terhadap algoritma yang sudah ada dengan tujuan untuk melihat atau membuktikan sebuah konsep. Eksperimen yang dilakukan dengan melakukan pengujian artefak anak timbangan dan mengelompokkan hasil pengujian dengan menggunakan algoritma *Self Organizing Maps*.

##### 3.1.2 Sifat Penelitian

Dilihat dari sifatnya, penelitian ini bersifat deskriptif. Penelitian ini bertujuan untuk mendeskripsikan performa algoritma *Self Organizing Maps* dalam



mengelompokkan data interkomparasi anak timbangan untuk keperluan analisis hasil evaluasi kegiatan interkomparasi.

### 3.1.3 Pendekatan Penelitian

Pendekatan yang digunakan dalam penelitian ini berupa pendekatan kuantitatif, karena data penelitian berupa angka-angka dan dianalisis menggunakan perhitungan statistik.

### 3.2. Metode Pengumpulan Data

Pengumpulan data dapat dilakukan dalam berbagai setting, berbagai sumber, dan berbagai cara. Bila dilihat dari setting-nya, data dapat dikumpulkan pada setting alamiah (*natural setting*), pada laboratorium dengan metode eksperimen, di rumah dengan berbagai responden, pada suatu seminar, diskusi, di jalan dan lain-lain. Bila di lihat dari sumber datanya, maka pengumpulan data dapat menggunakan sumber primer, dan sumber sekunder. Sumber primer adalah sumber data yang langsung memberikan data kepada pengumpul data, dan sumber sekunder merupakan sumber yang tidak langsung memberikan data kepada pengumpul data, misalnya lewat orang lain atau lewat dokumen.

Pada penelitian ini pengumpulan data dilakukan melalui sumber primer pada laboratorium Unit Metrologi Legal Kabupaten /Kota. Data dikumpulkan dari kegiatan interkomparasi laboratorium metrologi legal BSML Regional II Kementerian Perdagangan tahun 2021 yang berlangsung Maret 2021 sampai dengan Oktober 2021 dengan jumlah peserta terdiri dari 36 Unit Metrologi Legal Kabupaten / Kota di wilayah regional II yang meliputi Jawa, Bali dan Nusa Tenggara.

Tabel 3.1. Data interkomparasi artefak 1 kg

No	Nama Laboratorium	Kode	Mc 1kg	Uc 1kg
1	Kab Kediri	A1	999.990	0.0087
2	Kab Demak	A2	1000.0170	0.012
3	Kab Semarang	A3	1000.0025	0.011
4	Kota Magelang	A4	1000.0039	0.0090
5	Kota Cimahi	A5	1000.0394	0.0030
6	Kota Semarang	A6	1000.0013	0.0098
7	Kota Surakarta	A7	1000.035	0.013
8	Kab Madiun	A8	1000.0048	0.0095
9	Kab Sleman	A9	1000.0046	0.0054
10	Kab Jombang	A10	999.9975	0.0064
11	Kota Surabaya	A11	1000.0038	0.010
12	Kab Magetan	A12	1000.010	0.011
13	Kab Pangandaran	A13	1000.0072	0.0095
14	Kab Bantul	A14	999.9997	0.0093
15	Kab Purwakarta	A15	1000	0.014
16	Kab Indramayu	A16	999.9976	0.0089
17	Kota Tasikmalaya	A17	1000.0075	0.0090
18	Kota Pekalongan	A18	999.9999	0.0061
19	Kab Temanggung	B1	1000.0006	0.0087
20	Kab Banyumas	B2	1000.0036	0.0091
21	Kab Kuningan	B3	1000	0.011
22	Kab Pekalongan	B4	999.996	0.010
23	Kab Bojonegoro	B5	1000.0031	0.0033
24	Kab Kudus	B6	1000.0072	0.0093
25	Kab Magelang	B7	1000.0094	0.0080
26	Kab Pati	B8	1000.0006	0.0053
27	Kota Yogyakarta	B9	1000.0037	0.0048
28	Kab Kendal	B10	1000.0039	0.0090
29	Kab Nganjuk	B11	1000.005	0.0041
30	Kab Wonosobo	B12	999.9997	0.0090
31	Kota Kediri	B13	1000.006	0.0087
32	Kab Kulonprogo	B14	1000.0033	0.0065
33	Kab Jepara	B15	1000.0055	0.0089
34	Kab Karawang	B16	1000.0175	0.016
35	Kab Sukoharjo	B17	1000.0026	0.0087
36	Kab Mojokerto	B18	1000.003	0.0071

### 3.3. Metode Analisis Data

Dalam penelitian ini secara umum akan dilakukan dua kali tahapan *clustering*. Tahapan pertama akan melakukan *clustering* pada data massa konvensional (mc) dan ketidakpastian (uc) artefak anak timbangan. Hasil *clustering* tahap pertama dengan menggunakan algoritma *Self Organizing Maps* akan menghasilkan *cluster* laboratorium secara umum. Untuk mengetahui variabel yang memiliki pengaruh terbanyak pada evaluasi hasil interkomparasi, dilakukan *clustering* tahap kedua. *Clustering* tahap kedua dilakukan pada data sumber ketidakpastian masing-masing laboratorium. Sumber ketidakpastian tersebut terdiri dari anak timbangan standar, drift standar, kemampuan ulang penimbangan (*repeatability*), bouyancy udara dan timbangan elektronik yang digunakan. *Cluster* yang terbentuk baik pada tahap pertama, dan tahap kedua kemudian dilakukan analisis lebih lanjut.

Analisa data dilakukan berdasarkan hasil pengelompokkan dengan menggunakan platform analisis KNIME. KNIME atau *Konstanz Information Miner*, adalah platform analisis, pelaporan, dan integrasi data yang bersifat *opensource*. KNIME menyatukan berbagai komponen dalam bidang *machine learning* dan *data mining* dengan konsep alur data yang modular. Proses *clustering* menggunakan platform KNIME tampak seperti pada gambar 3.1 berikut:



Gambar 3.1. Clustering dengan KNIME

Pada gambar 3.1 terdapat beberapa hub dalam tahapan clustering. *Clustering* diawali dengan membaca data pada file dengan format .csv pada hub *csv reader*. Data tersebut kemudian dilakukan *filtering* pada hub *column filter* untuk menentukan atribut/variabel yang menjadi masukan pada proses *clustering*. Selanjutnya pada hub *Self Organizing Maps* dilakukan pengaturan nilai parameter yang terdiri dari nilai *learning rate*, tinggi dan lebar *lattice*, jumlah *epoch* untuk fase *ordering* dan fase *convergence* dan status normalisasi data.

Terdapat dua status normalisasi, *true* jika data dilakukan normalisasi dan *false* jika data tidak dilakukan normalisasi. Normalisasi data dilakukan untuk menskalakan nilai atribut dari data sehingga bisa terbentuk data pada range tertentu di setiap atribut. Normalisasi data dilakukan dengan metode *min-max* dimana data pada atribut yang sama dibagi dengan nilai maksimum pada atribut tersebut. Untuk mengetahui berapa jumlah *cluster* terbentuk dan rincian anggotanya digunakan hub

*weka cluster assigner*. Dari hub ini, data kemudian ditampilkan dalam bentuk image menggunakan hub *data to report*.

Tabel 3.2 Data nilai budget uncertainty artefak 1 kg

Kode	Ur	Ud	Ustd	Udrif	Ub
A1	0.000000	0.004083	0.000800	0.000739	0.001109
A2	0.004082	0.002300	0.010607	0.000000	0.001109
A3	0.002850	0.004080	0.002300	0.000740	0.001110
A4	0.001300	0.004100	0.000800	0.000200	0.001100
A5	0.000200	0.000400	0.000600	0.000700	0.001100
A6	0.000000	0.004080	0.001480	0.002310	0.000069
A7	0.0041	0.0041	0.0023	0.0017	0.0011
A8	0.0019	0.0041	0.0008	0.0007	0.0011
A9	0.0004	0.0004	0.0023	0.0007	0.0011
A10	0.001486	0.000408	0.00115	0.002309	0.001109
A11	0.0019	0.0041	0.0023	0.0007	0.0011
A12	0.0026	0.0041	0.0023	0.0007	0.0011
A13	0.0015	0.0041	0.0014	0.0007	0.0011
A14	0.0016	0.0041	0.0008	0.0007	0.0011
A15	0.005774	0.004083	0.00055	0.000693	0.001109
A16	0.0013	0.0041	0.0006	0.0001	0.0011
A17	0.0013	0.0041	0.0006	0.0007	0.0011
A18	0.0022	0.0004	0.0008	0.0007	0.0011
B1	0.000000	0.004100	0.000800	0.000800	0.001100
B2	0.001900	0.004100	0.000800	0.000200	0.069300
B3	0.003350	0.004082	0.000145	0.000739	0.001109
B4	0.002500	0.004100	0.000600	0.000700	0.001100
B5	0.000342	0.000408	0.008000	0.000739	0.001008
B6	0.001670	0.004080	0.000800	0.000740	0.001110
B7	0.003200	0.000400	0.000800	0.000700	0.001100
B8	0.000195	0.000041	0.002300	0.000700	0.001100
B9	0.000640	0.000400	0.002300	0.000700	0.001100
B10	0.001600	0.004100	0.000800	0.000900	0.001100
B11	0.000200	0.000400	0.000800	0.001800	0.000100
B12	0.001000	0.004100	0.000800	0.000700	0.001100
B13	0.000000	0.004100	0.000600	0.000700	0.001100
B14	0.000590	0.000400	0.001300	0.002800	0.001100
B15	0.001602	0.004083	0.00055	0.000673	0.000111
B16	0.0063	0.0041	0.0006	0.0006	0.0001
B17	0	0.0041	0.0008	0.0007	0.0011
B18	0.0025	0.0004	0.0017	0.0013	0.0011

Pengelompokkan data dilakukan dalam dua tahap, tahap pertama menggunakan data pada tabel 3.1, sedangkan tahap kedua menggunakan data pada tabel 3.2. Untuk mengetahui seberapa baik *cluster* yang dihasilkan, dilakukan evaluasi menggunakan nilai DBI dengan persamaan 8. Nilai DBI yang dicari adalah nilai terkecil mendekati 0 (non-negatif  $\geq 0$ ).

Untuk mengetahui variabel mana yang paling berpengaruh dilakukan analisis korelasi, yaitu metode statistika yang digunakan untuk menentukan kuat lemahnya hubungan antara suatu variabel dengan variabel lain tanpa melihat apakah variabel tertentu tergantung pada variabel lainnya. Analisis korelasi yang digunakan adalah korelasi Pearson (*pearson product moment*). Korelasi Pearson merupakan korelasi sederhana yang hanya melibatkan satu variabel terikat (*dependent*) dan satu variabel bebas (*independent*) dengan menggunakan persamaan sebagai berikut. (Miftahudin et al, 2021)

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \dots\dots\dots (10)$$

Hasil pengujian korelasi mengikuti interpretasi seperti pada tabel 3.3 berikut:

Tabel 3.3 Kriteria korelasi Pearson (Miftahudin et al, 2021)

No.	Nilai r	Interpretasi
1.	0.00 – 0.199	Sangat rendah
2.	0.20 – 0.399	Rendah
3.	0.40 – 0.599	Sedang
4.	0.60 – 0.799	Kuat
5.	0.80 – 1.00	Sangat kuat

Pengujian lanjutan untuk menentukan apakah koefisien korelasi yang didapat bisa digeneralisasi atau mewakili populasi, maka dilakukan uji signifikansi dari uji t dengan mencari nilai  $t_{hitung}$ . Nilai  $t_{hitung}$  dicari dengan menggunakan persamaan 10 berikut ini (Ramli Ayufiah et al, 2021).

$$t_{hitung} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \dots\dots\dots (11)$$

Hasil dari  $t_{hitung}$  kemudian dianalisis, jika  $t_{hitung}$  lebih besar ( $\geq$ ) daripada nilai  $t_{tabel}$  dengan signifikan 5%, maka nilai t signifikan sehingga  $H_1$  diterima dan  $H_0$  ditolak. Hal ini menandakan adanya hubungan antara variabel X dengan variabel Y. Sedangkan jika nilai  $t_{hitung}$  tidak sama dengan atau lebih kecil ( $\leq$ ) daripada nilai  $t_{tabel}$  dengan signifikan 5%, maka nilai t tidak signifikan sehingga  $H_1$  ditolak dan  $H_0$  diterima. Ini menandakan antara variabel X dan Y tidak memiliki hubungan (Ramli Ayufiah et al, 2021).

### 3.4. Alur Penelitian

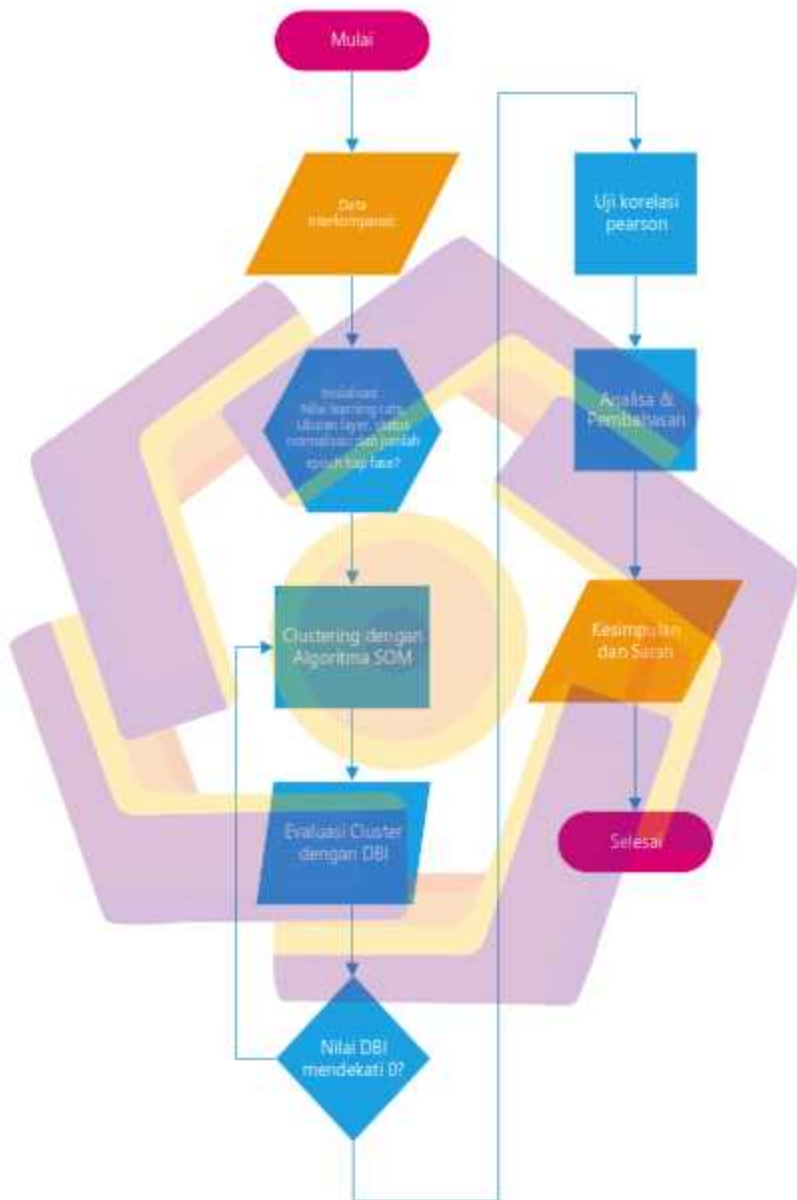
Alur penelitian seperti gambar 3.2 dapat dijelaskan sebagai berikut:

1. Pada tahap awal dilakukan literature review untuk mengumpulkan bahan referensi, tinjauan pustaka bersumber dari jurnal internasional dan jurnal nasional. Sedangkan landasan teori bersumber dari pustaka buku, pustaka makalah, jurnal atau prosiding.
2. Tahap selanjutnya berupa pengumpulan data interkomparasi anak timbangan. Data ini diperoleh dari kegiatan interkomparasi tahun 2021 yang diselenggarakan oleh Balai Standardisasi Metrologi Regional II dengan jumlah peserta sebanyak 36 laboratorium. Dari data yang terkumpul kemudian dilakukan perhitungan nilai rasio  $E_n$  nya. Data Nilai rasio  $E_n$  dan data

uncertainty dari laboratorium peserta interkomparasi akan dilakukan proses pengelompokan data dengan menggunakan algoritma SOM.

3. Sebelum dilakukan proses *clustering* data, pada tahap ini ditentukan terlebih dahulu beberapa setting parameter antara lain ukuran tinggi lebar layer, status normalisasi, nilai *learning rate*  $\alpha(t)$  dan nilai *epoch* untuk fase order dan fase konvergen.
4. Pada tahap ini dilakukan proses *clustering* dengan algoritma SOM pada data nilai massa konvensional (*mc*) beserta nilai ketidakpastian (*uc*) dan pada data nilai sumber ketidakpastian. *Clustering* nilai *mc uc* untuk mendapatkan pengelompokan awal dari 36 peserta interkomparasi.
5. Setelah didapatkan hasil *clustering*, kemudian dilakukan evaluasi seberapa baik hasil *cluster*. Pengujian dilakukan dengan menggunakan kriteria nilai DBI. Untuk menentukan seberapa besar pengaruh variabel terhadap variabel lain dan pengaruh variabel terhadap hasil dilakukan pengujian korelasi.
6. Tahapan terakhir dalam penelitian ini adalah menarik kesimpulan yang diperoleh dari eksperimen dan memberikan saran untuk penelitian selanjutnya.





Gambar 3.2. Alur Penelitian

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Data Penelitian

Penelitian ini menggunakan data interkomparasi anak timbangan yang diselenggarakan oleh Balai Standardisasi Metrologi Legal Regional II. Kegiatan interkomparasi laboratorium Metrologi Legal tahun 2021 diikuti oleh 36 peserta yang merupakan Unit Metrologi Legal di wilayah regional II. Data penelitian dikelompokkan berdasar jenis artefak anak timbangan yang digunakan. Artefak yang digunakan dalam kegiatan interkomparasi berupa artefak anak timbangan kelas M<sub>1</sub> dengan massa nominal 1 kg yang dibedakan identitasnya sebagai artefak A 1 kg dan artefak B 1 kg dengan identitas seperti pada tabel 4.1.

Tabel 4.1 Identitas artefak

Artefak	A	B
Merk	Ankatama	Ankatama
Massa Nominal	1000 g	1000 g
No Seri	ATM.M-1. S.022	ATM.M-1. S.021
Kelas	M <sub>1</sub>	M <sub>1</sub>
Bahan	Stainless steel	Stainless steel
Massa Jenis	7.950 kg/m <sup>3</sup>	7.950/m <sup>3</sup>

Dataset yang akan dikelompokkan berupa data hasil pengujian dari kegiatan interkomparasi seperti pada tabel 4.2 berikut:

Tabel 4.2 Rerata nilai data interkomparasi

No	Kode	Mc 1kg	Uc 1kg
1	A1	999.990	0.0087
2	A2	1000.0170	0.012
3	A3	1000.0025	0.011
4	A4	1000.0039	0.0090
5	A5	1000.0394	0.0030
6	A6	1000.0013	0.0098
7	A7	1000.035	0.013
8	A8	1000.0048	0.0095
9	A9	1000.0046	0.0054
10	A10	999.9975	0.0064
11	A11	1000.0038	0.010
12	A12	1000.010	0.011
13	A13	1000.0072	0.0095
14	A14	999.9997	0.0093
15	A15	1000	0.014
16	A16	999.9976	0.0089
17	A17	1000.0075	0.0090
18	A18	999.9999	0.0061
19	B1	1000.0006	0.0087
20	B2	1000.0036	0.0091
21	B3	1000	0.011
22	B4	999.996	0.010
23	B5	1000.0031	0.0033
24	B6	1000.0072	0.0093
25	B7	1000.0094	0.0080
26	B8	1000.0006	0.0053
27	B9	1000.0037	0.0048
28	B10	1000.0039	0.0090
29	B11	1000.005	0.0041
30	B12	999.9997	0.0090
31	B13	1000.006	0.0087
32	B14	1000.0033	0.0065
33	B15	1000.0055	0.0089
34	B16	1000.0175	0.016
35	B17	1000.0026	0.0087
36	B18	1000.003	0.0071
Rata-rata		1000.005	0.010

## 4.2. Klustering

### 4.2.1 Penentuan nilai parameter

Proses *clustering* dengan KNIME diawali dengan pengaturan nilai *learning rate*, tinggi dan lebar *lattice* serta jumlah *epoch* untuk fase *ordering* dan fase *convergence* pada hub *Self Organizing Maps* seperti tampak pada gambar 4.1 berikut:



Gambar 4.1 Menu pengaturan parameter pada hub *Self Organizing Maps*

Pada gambar 4.1 diatas memperlihatkan menu pengaturan parameter yang digunakan ketika melakukan klustering. Terdapat 7 kolom yang bisa diatur nilai parameternya, diantaranya:

- calStats*, parameter ini untuk menentukan apakah perlu menghitung statistik setiap cluster setelah selesai pelatihan / *training*.
- convergence epoch*, parameter untuk menentukan jumlah *epoch* pada fase konvergensi.

- c. *Height*, parameter untuk menentukan tinggi dari *lattice* / ukuran tinggi grid peta SOM 2 dimensi.
- d. *learning rate*, parameter untuk menentukan jumlah awal bobot yang diperbaharui. Nilai *learning rate* berada pada range nol (0) sampai satu (1).
- e. *normalize attributes*, parameter untuk menentukan status normalisasi data, apakah akan dilakukan normalisasi atau tidak.
- f. *ordering epoch*, parameter untuk menentukan jumlah *epoch* pada fase *ordering*.
- g. *Width*, parameter untuk menentukan lebar dari *lattice* / ukuran lebar grid peta SOM 2 dimensi.

Pada penelitian ini telah dilakukan beberapa percobaan nilai parameter yang meliputi *learning rate*, *ordering epochs*, dan *convergence epochs* pada ukuran grid 2x4 dengan hasil seperti pada tabel 4.3 berikut ini.

Tabel 4.3 Hasil Percobaan nilai parameter

No.	<i>Learning rate</i>	<i>Ordering epochs</i>	<i>Convergenepochs</i>	<i>Time (second)</i>
1.	0.3	6000	2000	60
2.	0.5	6000	2000	46
3.	0.8	6000	2000	34
4.	0.3	6000	20000	190
5.	0.5	6000	20000	184
6.	0.8	6000	20000	178
7.	0.3	12000	20000	204
8.	0.5	12000	20000	197
9.	0.8	12000	20000	190

Pada tabel 4.3 diatas terlihat dari setting nilai paramater *learning rate* ( $\alpha$ ), semakin kecil nilai *learning rate*-nya akan membutuhkan waktu respon yang lebih lama, dan semakin besar nilai *learning rate*-nya akan membutuhkan waktu respon yang lebih cepat. Dari percobaan tersebut nilai *learning rate* 0.8 akan digunakan pada penelitian ini.

Untuk parameter jumlah epoch pada fase order maupun fase konvergen disajikan pada gambar 4.2, 4.3, dan 4.4 berikut.



Gambar 4.2 Hasil kluster dengan parameter  $\alpha=0.8$ ,  $E_0$  6000 dan  $E_c$  2000



Gambar 4.3 Hasil kluster dengan parameter  $\alpha=0.8$   $E_0$ , 6000 dan  $E_c$  20000



Gambar 4.4 Hasil kluster dengan parameter  $\alpha=0.8$ ,  $E_o$  12000 dan  $E_c$  20000

Terlihat gambar 4.2 dengan 4.3 memiliki kesamaan pola kluster terbentuk meskipun berbeda pada jumlah konvergen epoch-nya, sedangkan pada gambar 4.4 ketika nilai epoch pada fase ordering dinaikkan menjadi 12000 terlihat perpindahan ruang *output* pada kluster. Klaster yang semula di *cluster7* berpindah ke *cluster0* berlaku sebaliknya, yang semula di *cluster6* berpindah ke *cluster1* dan berlaku pula sebaliknya. Untuk fase *ordering*, jumlah *epoch* minimal adalah 1000 dan jumlah *epoch* pada fase konvergen sedikitnya 500 kali jumlah *neuron* pada *network* (Umut Asan & Secil Ersan,2012). Pada penelitian ini digunakan parameter jumlah epoch pada fase ordering sebesar 2000 dan pada fase konvergen sebesar 6000. Pemilihan 6000 didasarkan pada kluster yang terbentuk dari data interkomparasi. Pola kluster pada epoch 12000, terjadi pergantian ruang output dengan jumlah anggota kluster tetap pada setiap ruangnya, demikian juga ketika epochnya dibuat 500 kali dari jumlah neuron (500x36) atau sebesar 18.000 pola kluster terbentuk sama dengan jumlah epoch pada fase konvergen sebesar 6000. Adapun untuk parameter normalisasi di set pada nilai *true*, dimana data sebelum di kluster telah dilakukan proses normalisasi data min max pada rentang 0-1.

Tabel 4.4 Nilai Parameter *Clustering* Dengan Algoritma SOM

Parameter ke-	<i>learning rate</i>	tinggi <i>lattice</i>	lebar <i>lattice</i>	<i>Epoch<sub>o</sub></i>	<i>Epoch<sub>c</sub></i>	Normalisasi
1	0,8	2	1	6000	2000	<i>true</i>
2	0,8	2	2	6000	2000	<i>true</i>
3	0,8	2	3	6000	2000	<i>true</i>
4	0,8	2	4	6000	2000	<i>true</i>
5	0,8	3	3	6000	2000	<i>true</i>
6	0,8	3	4	6000	2000	<i>true</i>
7	0,8	4	4	6000	2000	<i>true</i>

Parameter yang digunakan pada tabel 4.4 berbeda pada ukuran *lattice* atau kisi nya. Ukuran *lattice* menentukan ukuran grid (peta 2 dimensi) yang menentukan jumlah *output neuron*. Pada algoritma SOM, masing-masing *input neuron* akan dipasangkan pada *output neuron* sesuai ukuran grid nya. Parameter 1 memiliki 2 *output neuron* yang akan ditempati oleh *input neuron*. Serupa dengan parameter 1, parameter 2,3,4,5,6, dan 7 memiliki jumlah output neuron sesuai ukuran grid yang akan ditempati oleh input neuron. Nilai *learning rate*, jumlah epoch pada fase order dan konvergen dibuat tetap. Data input selanjutnya dilakukan normalisasi sebelum dilakukan pengelompokkan dengan algoritma SOM. Proses normalisasi dilakukan dengan metode *Min-Max Normalization* yang merubah data input pada rentang nilai minimum 0 dan nilai maksimum 1. Pada hub algoritma SOM pengaturan normalisasi terdapat dua pilihan pada kondisi *false*, jika data input sudah dilakukan normalisasi terlebih dahulu, dan *true* jika input data belum dilakukan normalisasi.

Data interkomparasi pada tabel 3.2 kemudian dikelompokkan dengan menggunakan algoritma *self-organizing maps* sesuai dengan nilai parameter pada tabel 4.4. Data input memiliki 2 fitur, fitur x dan fitur y, dimana fitur x adalah nilai massa konvensional artefak dan fitur y adalah nilai ketidakpastian artefak pada level



kepercayaan 95%. Algoritma SOM memiliki keistimewaan pada kemampuannya mengorganisir data secara mandiri. Berapapun data *input* yang diberikan, maka algoritma SOM akan mengelompokkan data – data tersebut ke dalam *neuron output* secara mandiri. Keistimewaan lain dari algoritma SOM terletak pada kemampuannya memvisualisasikan hasil pengelompokkan menjadi data berdimensi rendah (2 dimensi) yang memberikan kemudahan dalam pengamatan distribusi hasil klustering. Dalam penelitian ini data dipetakan secara visual untuk mempermudah analisis data. Hasil pengelompokkan direpresentasikan pada gambar 4.5 berikut ini:



Gambar 4.5 Hasil Clustering Parameter 5

Parameter 5 menghasilkan kluster sebanyak 9, dengan anggota kluster secara berurutan berjumlah (5,3,2,5,8,2,1,4 dan 6) laboratorium. Jumlah kluster terbentuk dipengaruhi oleh ukuran *lattice* atau grid. Ukuran 3x3 akan menghasilkan jumlah kluster maksimal 9 (sembilan).

Serupa dengan hasil kluster pada gambar 4.5, pengelompokkan data

interkomparasi untuk artefak 1 kg dari seluruh parameter ditampilkan pada tabel 4.5. Data pada tabel 4.5 memperlihatkan sebaran jumlah kluster yang berbeda-beda untuk setiap parameter.

Tabel 4.5 Jumlah kluster dan anggota kluster setiap parameter

Parameter	1	2	3	4	5	6	7
Ukuran grid	2x1	2x2	2x3	2x4	3x3	3x4	4x4
C0	26	3	1	4	5	1	2
C1	10	10	12	8	3	4	1
C2		13	3	5	2	4	2
C3		10	9	5	5	2	2
C4			7	3	8	5	2
C5			4	6	2	4	3
C6				4	1	2	4
C7				1	4	2	4
C8					6	2	2
C9						5	2
C10						3	2
C11						2	3
C12							1
C13							1
C14							4
C15							1

Data pada tabel 4.5 memperlihatkan pola sebaran jumlah kluster yang berbeda-beda untuk setiap parameter. Jumlah kluster yang terbentuk pada setiap parameter akan mengikuti ukuran gridnya. Ukuran grid menunjukkan jumlah maksimal ruang yang akan ditempati oleh neuron pemenang. Dari tabel 4.5 dapat diketahui parameter dengan ukuran grid 2x4 akan menghasilkan maksimal 8 ruang yang akan ditempati maksimal 8 neuron pemenang. Pola kluster terbentuk mengikuti ukuran tinggi lebarnya *lattice*. Jumlah maksimal ruang output neuron

dihitung dari 5 (lima) kali akar jumlah baris dataset (Vesanto, 2000). Dataset dalam penelitian ini berjumlah 36 baris, sehingga jumlah ukuran grid maksimum adalah  $5 \times \sqrt{36} = 5 \times 6$ . Dengan 36 data, maka parameter tinggi dan lebar lattice / ukuran grid yang dimasukkan di awal klastering dapat menggunakan kombinasi ukuran 2x1, 2x2, 2x3, 2x4 hingga maksimal ukuran 5x6. Pada penelitian ini dilakukan percobaan klastering dengan beberapa ukuran grid, diperoleh hasil seperti pada tabel 4.6 berikut:

Tabel 4.6 Jumlah klaster berdasar ukuran grid

No.	Ukuran Grid	Jumlah ruang	Jumlah klaster terbentuk
1.	2x1	2	2
2.	2x2	4	4
3.	2x3	6	6
4.	2x4	8	8
5.	3x3	9	9
6.	3x4	12	12
7.	4x4	16	16
8.	4x5	20	19
9.	5x5	25	19
10.	5x6	30	23

#### 4.3. Evaluasi Hasil Klaster

Untuk mendapatkan jumlah klaster optimum atau terbaik dilakukan evaluasi klaster dengan menghitung nilai DBI nya. Nilai DBI diperoleh dari perbandingan nilai kohesi dan nilai separasi dari cluster ke-i dan cluster ke-j. Kohesi menunjukkan kedekatan data terhadap titik pusat klaster dari sebuah klaster yang diikuti. Separasi

menunjukkan jarak antara titik pusat kluster dengan klasternya. Untuk menghitung nilai kohesi cluster ke- $i$  digunakan persamaan 5. Perhitungan nilai DBI dilakukan dengan menggunakan microsoft excel, dan dilakukan melalui 5 tahapan perhitungan yang meliputi:

- Perhitungan nilai titik pusat kluster
- Perhitungan nilai kohesi data pada pusat kluster dengan menggunakan persamaan 6.
- Perhitungan nilai separasi antar kluster dengan menggunakan persamaan 7.
- Perhitungan nilai rasio atau perbandingan kluster dengan menggunakan persamaan 8.
- Perhitungan nilai DBI dengan menggunakan persamaan 9.

Untuk melakukan perhitungan nilai titik pusat kluster, data dipisahkan sesuai dengan outputnya pada cluster 0, cluster 1, cluster 2 dan cluster 3 seperti tampak pada tabel 4.6 berikut, dimana pada kluster ke-0 memiliki 3 anggota, kluster ke-1 memiliki 10 anggota, kluster ke-2 memiliki 13 anggota dan kluster ke-3 memiliki 10 anggota.

Tabel 4.7 Data anggota per *cluster* parameter kedua

Cluster 1			Cluster 0		
data ke	x	y	data ke	x	y
1	0	0.093443	1	0.546559	0.147541
2	0.253036	0.131148	2	0.910931	0.163934
3	0.228745	0.111475	3	0.55668	0.213115
4	0.196356	0.103279	Cluster 2		
5	0.202429	0.180328	data ke	x	y
6	0.153846	0.096721	1	0.281377	0.098361
7	0.214575	0.093443	2	0.299595	0.106557

Tabel 4.7 Lanjutan

data ke	x	y	data ke	x	y
8	0.202429	0.131148	3	0.279352	0.114754
9	0.121457	0.114754	4	0.404858	0.131148
10	0.196356	0.098361	5	0.348178	0.106557
Cluster 3			6	0.354251	0.098361
data ke	x	y	7	0.275304	0.1
1	1	0	8	0.348178	0.103279
2	0.295547	0.039344	9	0.392713	0.081967
3	0.151822	1	10	0.281377	0.098361
4	0.200405	0.05082	11	0.323887	0.093443
5	0.265182	0.004918	12	0.313765	0.096721
6	0.214575	0.037705	13	0.255061	0.093443
7	0.277328	0.029508			
8	0.303644	0.018033			
9	0.269231	0.057377			
10	0.263158	0.067213			

a. Perhitungan nilai titik pusat

Titik pusat klaster pada cluster 0 merupakan nilai rata-rata setiap data pada fitur x dan y.

$$\text{titik pusat } x_0 = \frac{\sum_i^n x}{n} = \frac{0,5465 + 0,9109 + 0,5566}{3}$$

Diperoleh titik pusat  $x_0 = 0,6714$

$$\text{titik pusat } y_0 = \frac{\sum_i^n y}{n} = \frac{0,1475 + 0,1639 + 0,2131}{3}$$

Diperoleh titik pusat  $y_0 = 0,1748$

$$\text{titik pusat } x_1 = \frac{\sum_i^n x}{n} = \frac{0 + 0,253 + 0,228 + \dots + 0,1963}{10}$$

Diperoleh titik pusat  $x_1 = 0,1769$

$$\text{titik pusat } y_1 = \frac{\sum_i^n y}{n} = \frac{0,093 + 0,131 + 0,111 + \dots + 0,098}{10}$$

Diperoleh titik pusat  $y_1 = 0,1154$

$$\text{titik pusat } x_2 = \frac{\sum_i^n x}{n} = \frac{0,281 + 0,299 + 0,279 + \dots + 0,255}{13}$$

Diperoleh titik pusat  $x_2 = 0,3198$

$$\text{titik pusat } y_2 = \frac{\sum_l^n y}{n} = \frac{0,098 + 0,106 + 0,114 + \dots + 0,093}{13}$$

Diperoleh titik pusat  $y_2 = 0,1017$

$$\text{titik pusat } x_3 = \frac{\sum_l^n x}{n} = \frac{1 + 0,295 + 0,151 + \dots + 0,263}{10}$$

Diperoleh titik pusat  $x_3 = 0,3241$

$$\text{titik pusat } y_3 = \frac{\sum_l^n y}{n} = \frac{0 + 0,039 + 1 + \dots + 0,067}{10}$$

Diperoleh titik pusat  $y_3 = 0,1305$

b. Perhitungan nilai kohesi (*sum of square within*).

Digunakan persamaan 5 sebagai berikut:  $SSW_l = \frac{1}{m_l} \sum_{j=1}^{m_l} d(x_j, c_l)$

Untuk menghitung nilai *sum of square within*, terlebih dahulu dihitung jarak antara data ke pusat cluster dengan menggunakan perhitungan jarak euclidean sebagai berikut:

$$d(X_1, C_0) = \sqrt{(0,546559 - 0,6714)^2 + (0,147541 - 0,1748)^2} = 0,1277864$$

$$d(X_2, C_0) = \sqrt{(0,910931 - 0,6714)^2 + (0,163934 - 0,11748)^2} = 0,2397903$$

$$d(X_3, C_0) = \sqrt{(0,55668 - 0,6714)^2 + (0,213115 - 0,11748)^2} = 0,1209195$$

Serupa dengan perhitungan diatas diperoleh nilai jarak data ke masing-masing pusat kluster sebagai berikut:

Tabel 4.8 Data perhitungan nilai jarak cluster ke pusat cluster

Cluster 1				Cluster 0			
data ke	x	y	d	data ke	x	y	d
1	0	0.093443	0.178282	1	0.546559	0.147541	0.127786
2	0.253036	0.131148	0.077723	2	0.910931	0.163934	0.23979
3	0.228745	0.111475	0.051971	3	0.55668	0.213115	0.120919
4	0.196356	0.103279	0.022909	Cluster 2			
5	0.202429	0.180328	0.069749	data ke	x	y	d
6	0.153846	0.096721	0.029695	1	0.281377	0.098361	0.038612

Tabel 4.8 Lanjutan

data ke	x	y	d	data ke	x	y	d
7	0.214575	0.093443	0.043591	2	0.299595	0.106557	0.020802
8	0.202429	0.131148	0.029971	3	0.279352	0.114754	0.042518
9	0.121457	0.114754	0.055469	4	0.404858	0.131148	0.089954
10	0.196356	0.098361	0.025852	5	0.348178	0.106557	0.028742
Cluster 3				6	0.354251	0.098361	0.034581
data ke	x	y	d	7	0.275304	0.1	0.044569
1	1	0	0.688392	8	0.348178	0.103279	0.02838
2	0.295547	0.039344	0.095512	9	0.392713	0.081967	0.075516
3	0.151822	1	0.886409	10	0.281377	0.098361	0.038612
4	0.200405	0.05082	0.147124	11	0.323887	0.093443	0.009255
5	0.265182	0.004918	0.138704	12	0.313765	0.096721	0.007894
6	0.214575	0.037705	0.143537	13	0.255061	0.093443	0.06531
7	0.277328	0.029508	0.111285				
8	0.303644	0.018033	0.114302				
9	0.269231	0.057377	0.091407				
10	0.263158	0.067213	0.087845				

Diperoleh nilai SSW untuk masing-masing kluster sebagai berikut:

$$SSW_0 = \frac{1}{3} \times 0,4884962 = 0,1628321$$

$$SSW_{10} = \frac{1}{10} \times 0,585212 = 0,058521$$

$$SSW_2 = \frac{1}{13} \times 0,524747 = 0,040365$$

$$SSW_3 = \frac{1}{10} \times 0,2,504517 = 0,250452$$

c. Perhitungan nilai separasi (*sum of square between*).

Digunakan persamaan 6 sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j)$$

Kluster  $C_0$  dengan kluster  $C_1$ ,

$$SSB_{0,1} = d(c_0, c_1) = \sqrt{(0,67139 - 0,176923)^2 + (0,174863 - 0,11541)^2}$$

$$SSB_{0,1} = 0,498028$$

Kluster  $C_0$  dengan kluster  $C_2$ ,

$$SSB_{0,2} = d(c_0, c_2) = \sqrt{(0,67139 - 0,319838)^2 + (0,174863 - 0,10176)^2}$$

$$SSB_{0,2} = 0,359071$$

Kluster  $C_0$  dengan kluster  $C_3$ .

$$SSB_{0,3} = d(c_0, c_3) = \sqrt{(0,67139 - 0,324089)^2 + (0,174863 - 0,13049)^2}$$

$$SSB_{0,3} = 0,350124$$

Kluster  $C_1$  dengan kluster  $C_2$ .

$$SSB_{1,2} = d(c_1, c_2) = \sqrt{(0,176923 - 0,319838)^2 + (0,11541 - 0,10176)^2}$$

$$SSB_{1,2} = 0,143565$$

Kluster  $C_1$  dengan kluster  $C_3$ .

$$SSB_{1,3} = d(c_1, c_3) = \sqrt{(0,176923 - 0,324089)^2 + (0,11541 - 0,13049)^2}$$

$$SSB_{1,3} = 0,147937$$

Kluster  $C_2$  dengan kluster  $C_3$ .

$$SSB_{2,3} = d(c_2, c_3) = \sqrt{(0,176923 - 0,324089)^2 + (0,10176 - 0,13049)^2}$$

$$SSB_{2,3} = 0,029039$$

d. Perhitungan nilai rasio / perbandingan antar kluster dengan persamaan 7.

Diperoleh matriks R sebagai berikut:

Tabel 4.9 Matriks R

Matriks R	0	1	2	3
0	0	0,444459	0,565897	1,180393
1	0,444459	0	0,688793	2,088547
2	0,565897	0,688793	0	10,01463
3	1,180393	2,088547	10,01463	0

Dari matriks pada tabel 4.9 selanjutnya dicari nilai Rmax. Nilai Rmax merupakan penjumlahan nilai R terbesar pada masing-masing baris.

$$\text{Nilai Rmax} = 1,1804 + 2,0885 + 10,0146 + 10,0146$$

$$\text{Nilai Rmax} = 23,2981$$

e. Perhitungan nilai Davies Bouldin Index dengan persamaan 8.

Nilai rasio yang diperoleh dari poin d diatas digunakan untuk menghitung nilai



DBI dengan rumus sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

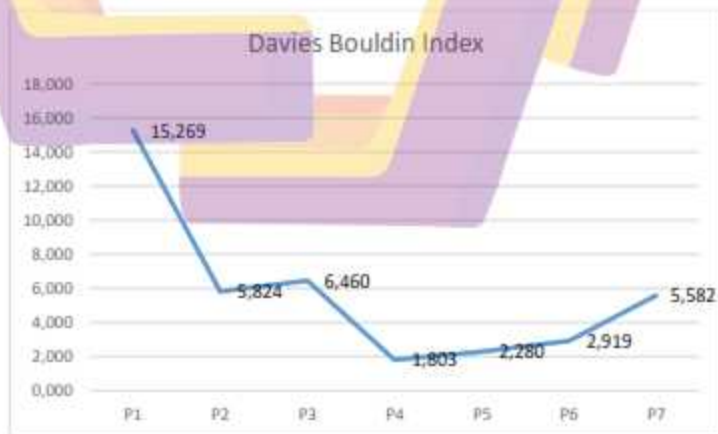
Jumlah kluster terbentuk sebanyak 4, sehingga diperoleh

$$DBI = \frac{1}{4} \times 23,2981 \quad , \quad DBI = 5,824$$

Serupa dengan perhitungan tersebut di atas, nilai DBI untuk masing-masing parameter tercantum pada tabel 4.10 berikut ini.

Tabel 4.10 Nilai DBI hasil klastering

No.	Parameter ke	Ukuran grid	Jumlah kluster	Nilai DBI
1.	1	2 x 1	2	15,269
2.	2	2 x 2	4	5,824
3.	3	2 x 3	6	6,460
4.	4	2 x 4	8	1,803
5.	5	3 x 3	9	2,280
6.	6	3 x 4	10	2,919
7.	7	4 x 4	14	5,582

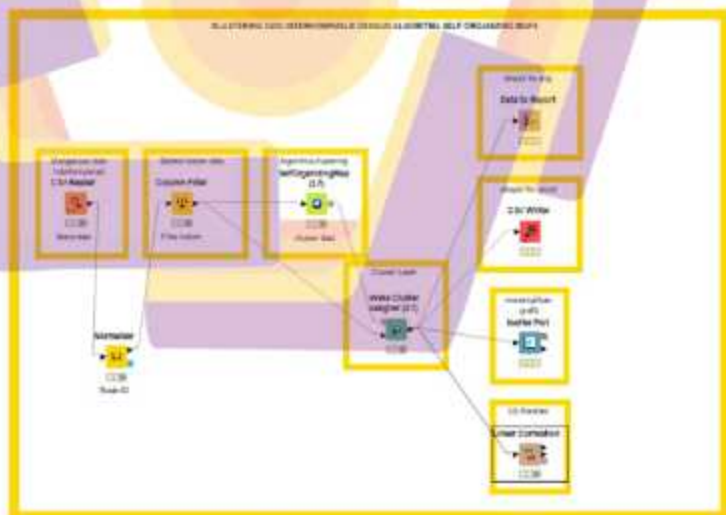


Gambar 4.6 Grafik nilai DBI

Dari nilai DBI pada tabel 4.10 dan grafik pada gambar 4.6 di atas, diperoleh nilai DBI terkecil pada parameter ke-4 sebesar 1,803. Parameter ke-4 adalah parameter clustering dengan pengaturan ukuran grid 2 x 4, jumlah epoch pada fase konvergen sebanyak 2000, dan jumlah epoch pada fase ordering sebanyak 6000. Jumlah kluster yang dihasilkan sebanyak 8 kluster. Nilai DBI terkecil merepresentasikan kluster terbaik dari sejumlah percobaan yang dilakukan dimana homogenitas anggota dalam satu kluster dan heterogenitas antar kluster tinggi.

#### 4.4. Uji Korelasi Pearson

Untuk mengetahui variabel mana yang paling berpengaruh dilakukan analisis korelasi sesuai persamaan 10. Pada penelitian ini, perhitungan nilai korelasi dilakukan dengan menggunakan hub *linear correlation* pada KNIME seperti pada gambar 4.7.



Gambar 4.7 Uji korelasi dengan platform KNIME

Correlation measure - 0.52 - Linear Correlation

File Edit Hilit Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 2kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 3kg	Cluster0	-0.08493551092616...	0.6223521210536...	34
Row2	Uc	Cluster0	0.39106071201437...	0.0183568907277...	34

Gambar 4.8 Hasil pengujian korelasi Pearson pada cluster0

Dari gambar 4.7 dan 4.8 diperoleh nilai korelasi antara mc dengan uc sebesar 0.08334, nilai korelasi antara mc dengan cluster0 sebesar -0.08493, dan nilai korelasi antara uc dengan cluster0 sebesar 0.39106.

Dari persamaan 11 diperoleh nilai  $t_{hitung}$  untuk korelasi mc dengan uc sebesar:  $t_{hitung} = \frac{0.08334\sqrt{36-2}}{\sqrt{1-0.08334^2}} = 0.487648$ . Untuk  $t_{tabel}$  pada  $df = 34$  dan level signifikansi 5% sebesar 1.691. Terlihat bahwa  $t_{hitung} \leq t_{tabel}$ , nilai  $r$  tidak signifikan yang menandakan antara variabel mc dengan uc tidak memiliki hubungan. Serupa dengan perhitungan di atas, perhitungan lengkap disajikan dalam tabel 4.11 berikut.

Tabel 4.11 Analisis hasil pengujian korelasi pearson

No	klm1	klm2	Nilai r	r <sup>2</sup>	Interpretasi	p value	t <sub>hitung</sub>	signifikansi
1	mc	uc	0.0833	0.006946	sangat rendah	0.62892	0.488	tidak signifikan
2	mc	cluster0	-0.084	0.007215	sangat rendah	0.62235	-0.496	tidak signifikan
3	mc	cluster1	-0.132	0.017432	sangat rendah	0.44274	-0.776	tidak signifikan
4	mc	cluster2	-0.380	0.144491	sangat rendah	0.0222	-2.396	signifikan negatif
5	mc	cluster3	-0.084	0.007145	sangat rendah	0.62404	-0.494	tidak signifikan
6	mc	cluster4	0.5816	0.338305	rendah	1.98781	4.169	signifikan positif
7	mc	cluster5	0.0865	0.007498	sangat rendah	0.61555	0.507	tidak signifikan
8	mc	cluster6	-0.169	0.028605	rendah	0.32409	-1.000	tidak signifikan

Tabel 4.11 Lanjutan

No	klm1	klm2	Nilai r	r <sup>2</sup>	Interpretasi	p value	hitng <sup>t</sup>	signifikansi
9	mc	cluster7	0.6230	0.388229	rendah	4.93305	4.645	signifikan positif
10	uc	cluster0	0.3910	0.152928	sangat rendah	0.01835	2.478	signifikan positif
11	uc	cluster1	0.1022	0.010447	sangat rendah	0.55304	0.599	tidak signifikan
12	uc	cluster2	0.0702	0.004935	sangat rendah	0.68391	0.411	tidak signifikan
13	uc	cluster3	-0.599	0.358921	rendah	1.13104	-4.362	signifikan negatif
14	uc	cluster4	0.5428	0.294719	rendah	6.23897	3.769	signifikan positif
15	uc	cluster5	0.0328	0.00108	sangat rendah	0.84912	0.192	tidak signifikan
16	uc	cluster6	-0.278	0.077429	sangat rendah	0.10032	-1.688	tidak signifikan
17	uc	cluster7	-0.348	0.121745	sangat rendah	0.03701	-2.170	signifikan negatif

Pada tabel 4.11 terlihat korelasi positif terbesar terjadi antara mc dengan cluster7 sebesar 0.623 dengan interpretasi rendah yang memiliki signifikansi positif ( $4.169 \geq 1.691$ ), dan korelasi positif terkecil terjadi antara uc dengan cluster5 sebesar 0.0328 dengan interpretasi sangat rendah yang tidak memiliki signifikansi ( $0.192 \leq 1.691$ ). Untuk korelasi negatif terbesar terjadi antara uc dengan cluster3 sebesar -0.599 dengan interpretasi rendah yang memiliki signifikansi negatif ( $4.362 \geq 1.691$ ), dan korelasi negatif terkecil terjadi antara mc dengan cluster0 sebesar -0.084 dengan interpretasi sangat rendah yang tidak memiliki signifikansi ( $0.496 \leq 1.691$ ).

#### 4.5. Analisa Data

Klaster terbaik berupa pengelompokkan dengan parameter keempat selanjutnya dilakukan analisis dengan melihat statistika data pada hub KNIME

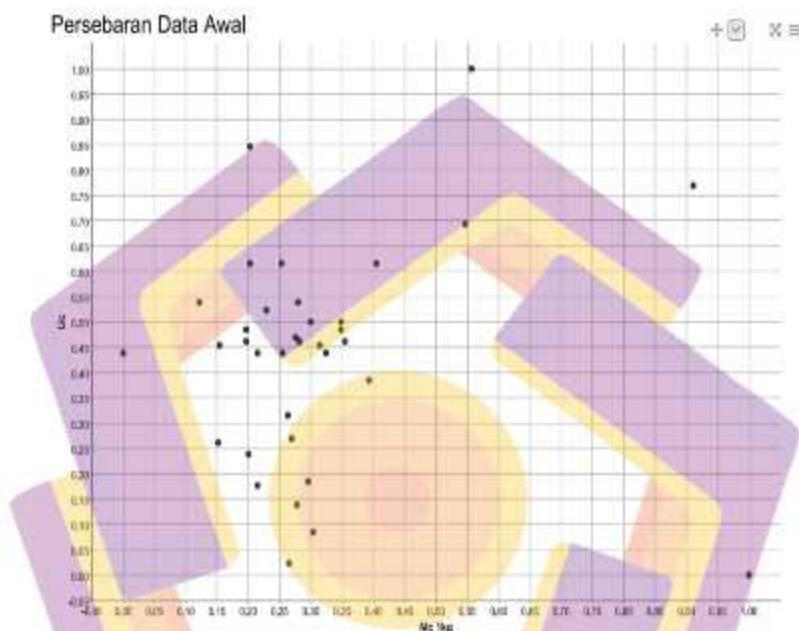


Data interkomparasi dengan nilai massa konvensional dan nilai ketidakpastian pada gambar 4.9 dan 4.10 terbagi menjadi 8 kelompok sesuai dengan kedekatan nilai massa konvensional dan nilai ketidakpastiannya.

- a. Klaster 2 yang beranggotakan 5 laboratorium (14%) memiliki nilai massa konvensional rata-rata di 999,9966 g.
- b. Klaster 0 yang beranggotakan 4 laboratorium (11%) dan klaster 1 yang beranggotakan 8 laboratorium (22%) memiliki nilai rata-rata massa konvensional yang sama (1000,031 g), namun berbeda pada nilai rerata ketidakpastiannya ( $U_c = 0,012$  dengan  $U_c = 0,0092$ ). Rerata yang sama pada nilai massa konvensional ini menunjukkan kedekatan jarak antara klaster 0 dan klaster 1 yang dibuktikan dengan nilai SSB paling kecil diantara klaster lainnya ( $SSB = 0,0414$ ). Nilai rerata ketidakpastian pada klaster 1 ( $U_c=0,0092$ ) memiliki nilai yang lebih kecil dibandingkan pada nilai rerata ketidakpastian pada klaster 0 ( $U_c=0,012$ ).
- c. Klaster 4 ini memiliki nilai ketidakpastian ( $U_c = 0,01$ ) yang lebih besar dibanding klaster lainnya. Anggota klaster 4 terdiri dari laboratorium dengan kode A2, A7 dan B16 memiliki nilai rata-rata massa konvensional sebesar 1000,0232 g.
- d. Anggota klaster 7 memiliki nilai massa konvensional terbesar (1000,039 g) dibandingkan dengan anggota dari klaster lainnya, namun memiliki nilai ketidakpastian yang berdekatan dengan anggota klaster 5.

Pola yang terbentuk dari hasil klastering menunjukkan bahwa pengelompokan laboratorium ditentukan dari fitur x, baru kemudian oleh fitur y.

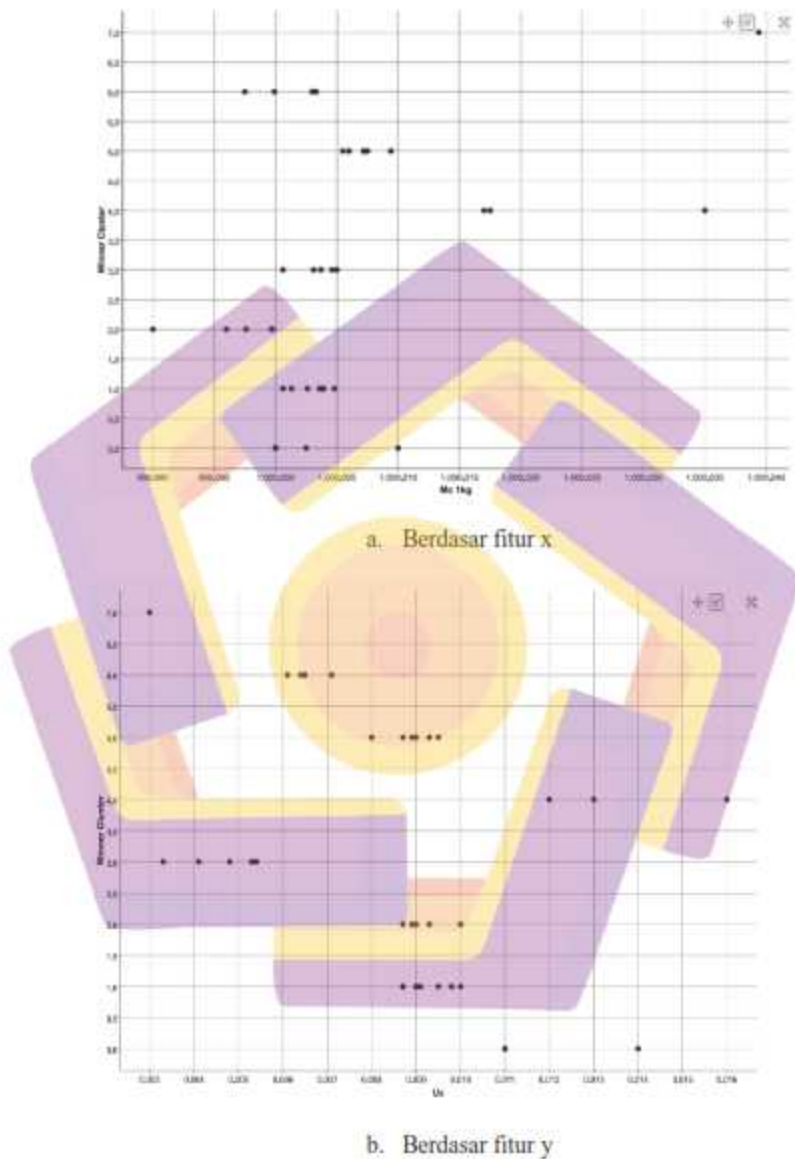
Laboratorium yang memiliki kesamaan pada fitur x dan fitur y mengelompok bersama dan jika terjadi perbedaan pada salah satu fitur maka akan memisahkan diri.



Gambar 4.11 Persebaran data awal

Persebaran awal data interkomparasi seperti terlihat pada gambar 4.11 memperlihatkan posisi awal data dimana terdapat 1 data interkomparasi dengan massa konvensional yang terpisah cukup jauh dari lainnya dan 3 data interkomparasi dengan nilai ketidakpastian yang terpisah cukup jauh dari data lainnya.

Sebaran data awal tersebut setelah dilakukan klustering dengan algoritma menghasilkan sebaran data berdasarkan 2 (dua) fitur awal yang digunakan yaitu fitur x berupa  $Mc$  dan fitur y berupa  $Uc$  seperti terlihat pada gambar 4.12 berikut.



Gambar 4.12 Persebaran data hasil klustering



Pada gambar 4.12 di atas fitur  $x$  dengan nilai rata-rata 1000,005 memiliki sebaran data mayoritas dibawah rata-rata. Pada fitur  $y$  dengan nilai rata-rata 0,010 juga memiliki sebaran data mayoritas dibawah nilai rata-ratanya. Mengacu data statistik pada gambar 4.9, diduga variabel yang paling berpengaruh berdasarkan hasil klustering adalah pada variabe nilai ketidakpastian atau fitur  $y$ .

Untuk mengetahui permasalahan yang mungkin muncul dalam kegiatan interkomparasi dapat dilihat pada sumber ketidakpastian yang terdiri dari:

- Ketidakpastian repeatability timbangan ( $U_r$ );
- Ketidakpastian akibat daya baca / resolusi timbangan ( $U_d$ );
- Ketidakpastian anak timbangan standar yang digunakan ( $U_{std}$ );
- Ketidakpastian *drift* atau pergeseran nilai dari anak timbangan standar ( $U_{drift}$ );
- Ketidakpastian akibat *bouyancy* udara ( $U_b$ ).



Gambar 4.13 Hasil klustering data nilai sumber ketidakpastian P4

Klustering dilakukan dengan menggunakan parameter ke-4 disesuaikan dengan parameter yang digunakan pada klustering data interkomparasi. Hasil pengelompokan data sumber ketidakpastian ditampilkan pada gambar 4.14. Dari gambar 4.11 terdapat 8 kelompok dimana salah satu kelompok memiliki anggota terbanyak yaitu 14 anggota kelompok (38,89%) pada kluster ke-0.

	(14)	(1)	(0)	(6)	(7)	(4)	(3)	(5)
<b>Ur</b>								
value	0,0011	0	0,0015	0,0017	0,0047	0,0025	0,0041	0,0004
min	0	0	NaN	0,0002	0,0019	0,0019	0,0041	0,0002
max	0,0019	0	NaN	0,0032	0,0063	0,0041	0,0041	0,0004
mean	0,0011	0	NaN	0,0017	0,0047	0,0025	0,0041	0,0004
std. dev.	0,0007	NaN	NaN	0,0012	0,0024	0,0008	NaN	0,0002
<b>Un</b>								
value	0,0041	0,0041	0,0015	0,0004	0,0041	0,0041	0,0023	0,0003
min	0,0041	0,0041	NaN	0,0004	0,0041	0,0041	0,0023	0
max	0,0041	0,0041	NaN	0,0004	0,0041	0,0041	0,0023	0,0004
mean	0,0041	0,0041	NaN	0,0004	0,0041	0,0041	0,0023	0,0003
std. dev.	0	NaN	NaN	0	0	0	NaN	0,0003
<b>Uand</b>								
value	0,0008	0,0015	0,0002	0,0011	0,0007	0,0016	0,0101	0,0031
min	0,0005	0,0015	NaN	0,0008	0,0005	0,0001	0,0101	0,0004
max	0,0014	0,0015	NaN	0,0017	0,0008	0,0023	0,0101	0,008
mean	0,0008	0,0015	NaN	0,0011	0,0006	0,0017	0,0101	0,0031
std. dev.	0,0002	NaN	NaN	0,0004	0,0001	0,001	NaN	0,0028
<b>Udrif</b>								
value	0,0006	0,0023	0,0011	0,0016	0,0009	0,0009	0	0,0007
min	0,0001	0,0023	NaN	0,0007	0,0002	0,0007	0	0,0007
max	0,0009	0,0023	NaN	0,0028	0,0007	0,0017	0	0,0007
mean	0,0006	0,0023	NaN	0,0016	0,0005	0,0009	0	0,0007
std. dev.	0,0002	NaN	NaN	0,0005	0,0003	0,0004	NaN	0
<b>U0</b>								
value	0,001	0,0001	0,0015	0,0009	0,0235	0,0011	0,0011	0,0011
min	0,0001	0,0001	NaN	0,0001	0,0001	0,0011	0,0011	0,001
max	0,0011	0,0001	NaN	0,0011	0,0893	0,0011	0,0011	0,0011
mean	0,001	0,0001	NaN	0,0005	0,0235	0,0011	0,0011	0,0011

Gambar 4.14 Statistika Hasil Klustering Sumber Ketidakpastian P4

Berdasarkan data pada gambar 4.14 terdapat 7 kluster dari total 8 jumlah output yang tersedia dengan keterangan sebagai berikut:

- Anggota kluster ke-0 memiliki nilai rata-rata ketidakpastian akibat repeatability ( $U_r$ ) sebesar 0,0011, nilai rata-rata ketidakpastian akibat anak timbangan standar ( $U_{nd}$ ) sebesar 0,0008 dan nilai rata-rata ketidakpastian akibat drift anak

timbangan standar ( $U_{dir}$ ) sebesar 0,0006. Untuk ketidakpastian akibat bouyancy dan ketidakpastian akibat daya baca timbangan memiliki nilai yang sama dengan klaster lainnya

- b. Klaster ke-1 memiliki ciri khusus pada nilai ketidakpastian yang bersumber dari repeatabilitynya dengan nilai 0.
- c. Klaster ke-3 memiliki ciri khusus pada rata-rata nilai ketidakpastian akibat repeatability yang bernilai 0,0017, rata-rata nilai ketidakpastian akibat daya baca timbangan sebesar 0,0004, nilai rata-rata ketidakpastian akibat anak timbangan standar ( $U_{ind}$ ) sebesar 0,0006 dan nilai rata-rata ketidakpastian akibat drift anak timbangan standar ( $U_{dir}$ ) sebesar 0,0005.
- d. Serupa dengan poin 1 sampai dengan 3 di atas, klaster ke-4, klaster ke-5, klaster ke-6 dan klaster ke-7 memiliki rata-rata nilai ketidakpastian yang berbeda pada sumber ketidakpastian akibat repeatability, anak timbangan standar, dan drift anak timbangan standar. Untuk sumber ketidakpastian akibat daya baca timbangan dan nilai bouyancy relatif sama.

Dari keterangan di atas dan berdasar data pada gambar 4.9 dan gambar 4.14, secara umum ada tiga sumber masalah yang berpengaruh pada kegiatan interkomparasi. Pertama pada kondisi peralatan /timbangan elektronik yang digunakan untuk membaca hasil penimbangan, kedua pada kondisi anak timbangan standar yang digunakan dan ketiga pada proses pengujian.

Dari penelitian yang telah dilakukan didapatkan beberapa pengetahuan yang berhubungan dengan klustering diantaranya:

- a. Penentuan parameter klustering dengan algoritma SOM dipengaruhi oleh:

1. Berdasarkan hasil percobaan klustering seperti pada tabel 4.5 dan tabel 4.6, penentuan parameter ukuran grid dipengaruhi oleh 5 kali akar dari jumlah baris data input.
2. Berdasarkan hasil percobaan klustering seperti pada gambar 4.2, 4.3 dan 4.4, jumlah *epoch* baik pada fase *ordering* maupun fase konvergen mempengaruhi penempatan neuron pemenang dan neuron tetangga pada kluster.
  - b. Penginputan data yang salah meskipun hanya satu data, sangat berpengaruh pada hasil klustering dan kualitas hasil klustering.
  - c. Berdasarkan percobaan pada tabel 4.3 semakin besar parameter nilai *learning rate* yang digunakan memerlukan waktu proses yang semakin cepat. Sebaliknya semakin kecil nilai *learning rate*-nya, memerlukan waktu proses yang semakin lambat.
  - d. Berdasarkan perhitungan nilai DBI pada tabel 4.10 dan grafik DBI pada gambar 4.6, jika nilai DBI sudah menunjukkan pola lembah, maka jumlah percobaan dapat dicukupkan karena nilai DBI akan cenderung naik.

## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Data interkomparasi dapat dievaluasi dengan teknik *clustering* menggunakan algoritma *Self Organizing Maps*. Dari tujuh percobaan yang dilakukan didapatkan nilai pengujian terbaik hasil klastering dengan metode *Davies Bouldin Indeks* sebesar 1,803 pada percobaan ke 4 dengan ukuran grid 2 x 4.
2. Evaluasi hasil klastering dengan menggunakan algoritma *Self Organizing Maps* menggambarkan kondisi laboratorium berdasarkan jenis permasalahan yang sama dalam satu kelompok, sehingga dapat dijadikan sebagai bahan acuan dalam evaluasi kegiatan interkomparasi.
3. Berdasarkan analisis korelasi dengan menggunakan metode korelasi pearson, tidak terdapat hubungan antara variabel nilai massa konvensional dengan variabel nilai ketidakpastian. Variabel nilai ketidakpastian memiliki pengaruh pada hasil klastering ketika klaster memiliki persamaan pada nilai rata-rata massa konvensional.

#### **5.2. Saran**

Hasil penelitian ini dapat dijadikan bahan untuk penelitian berikutnya seperti penelitian sistem pendukung keputusan interkomparasi.

## DAFTAR PUSTAKA

### PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Asan U., Ercan S., *An Introduction to Self-Organizing Maps*. In: Kahraman C. (eds) *Computational Intelligence Systems in Industrial Engineering*. Atlantis Computational Intelligence Systems, vol 6. Atlantis Press, Paris, 2012. [https://doi.org/10.2991/978-94-91216-77-0\\_14](https://doi.org/10.2991/978-94-91216-77-0_14)
- Asriny Novendri Isra et all, *Comparison of K-Medoids and Self Organizing Maps Algorithm in Grouping Hydrometeorological Natural Disasters in Java Island*, ICITDA 2020 IOP Conf. Series: Materials Science and Engineering 1077 (2021) 012008 IOP Publishing doi:10.1088/1757-899X/1077/1/01200
- Ayufiah Ramli, Idris Jafar, Sudirman, Hubungan Kebiasaan Belajar dengan Prestasi Belajar Siswa di Kelas V SD Inpres 17 Bajoe, JPPSD, vol 1, 2021
- Cindy Viane Bertan; A. K. T. Dundu; R.J.M. Mandagi, Pengaruh Pendetayagunaan Sumber Daya Manusia (Tenaga Kerja) Terhadap Hasil Pekerjaan (Studi Kasus Perumahan Taman Mapanget Raya (Tamara)), Jurnal Sipil Statik Vol.4 No.1, Januari, 2016
- Cottrell Marie; Olteanu Madalina; Rossi Fabrice; Vialaneix Nathalie, *Self-Organizing Maps, Theory and Applications*, REVISTA INVESTIGACION OPERACIONAL, VOL. 39, NO. 1, 1-22, 2018
- Febrita Ruth Ema; Mahmudy Wayan Firdaus; Wibawa Aji Prasetya, 2019, *High Dimensional Data Clustering using Self-Organized Map*. Knowledge Engineering and Data Science (KEDS) Vol 2, No 1, Juni 2019, pp. 31–40 [https://doi.org/10.17977/um018v2i12019p31\\_40](https://doi.org/10.17977/um018v2i12019p31_40)
- Firmansyah Dese Narfa; Adinugroho Sigit; Rahayudi Bayu, 2019, Pengelompokan Wilayah Berdasarkan Kesejahteraan Sosial Menggunakan Algoritme *Self-Organizing Maps* Dengan Perbaikan *Missing Value K-Nearest Neighbors*, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Universitas Brawijaya Vol. 3, No. 7, Juli 2019, hlm. 7205-7213
- Glozzi Valentina; Plunkett Kim, 2018, *Self-organizing maps and generalization: an algorithmic description of Numerosity and Variability effects*, arXiv:1802.09442v1 [cs.AI] 26 Feb 2018
- Hendayanti Ni Putu Nanik; Putri Gusti Ayu Made Ama; Nurhida Maulida, 2018, Ketepatan Klasifikasi Penerima Beasiswa Stimik Stikom Bali Dengan *Hybrid Self Organizing Maps Dan Algoritma K-Means*, Jurnal Varian Vol.2 No.1 Oktober 2018

- Jollyta Deni; Efendi Syahril; Zarlis Muhammad, Mawengkang Herman, Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index, Prosiding Seminar Nasional Riset Information Science (SENARIS), ISSN: 2686-0260 September 2019 Hal:918-926
- Kania Siti; Rachmatin Dewi; Dahlan Jarnawi Afgani, 2019, Program Aplikasi Pengelompokan Objek Dengan Metode *Self Organizing Map* Menggunakan Bahasa R, Jurnal EurekaMatika, Vol.7 No.2, 2019
- Kapita Syarifuddin N; Abdullah Saiful Do, Pengelompokan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-Som, JIKO (Jurnal Informatika dan Komputer) Vol. 3, No. 1, April 2020, hlm. 56-61 DOI: 10.33387/jiko
- Lakshminarayanan Sivakkumaran, 2020, *Application of Self-Organizing Maps on Time Series Data for identifying interpretable Driving Manoeuvres*, SpringerOpen Jurnal, Germany
- Miftahuddin, Ananda Pratama, Ichsan Setiawan, Analisis Hubungan Antara Kelembaban Relatif Dengan Beberapa Variabel Iklim Dengan Pendekatan Korelasi Pearson Di Samudera Hindia, Jurnal Siger Matematika, Vol. 02, No. 01, Maret 2021
- Purbasari I Y; Puspaningrum E Y; Putra A B S, 2019, *Using Self-Organizing Map (SOM) for Clustering and Visualization of New Students based on Grades*, Journal of Physics: Conference Series, IOP Publishing, 2020
- Riese Felix M; Keller Sina; Hinz Stefan, *Supervised and Semi-Supervised Self-Organizing Maps for Regression and Classification Focusing on Hyperspectral Data*, Remote Sens. 2020, 12, 0007; doi:10.3390/rs12010007, [www.mdpi.com/journal/remotesensing](http://www.mdpi.com/journal/remotesensing)
- Sultan Mahabab Abdula, *A Study of Self-Organizing Maps (SOM) Neural Network Using Matlab*, Research on Humanities and Social Sciences Vol.10, No.6, 2020 [www.iiste.org](http://www.iiste.org)
- TIAN Jin; GU Mengyi, 2019, *Subspace Clustering Based on Self-Organizing Map*, ResearchGate, 2019
- Vesanto J, Alhoniemi E, *Clustering of the Self-Organizing Map*, IEEE Transactions on Neural Networks, Vol. 11, No. 3, May 2000
- Widiyanto Max; Teja Ajie Cipta, Perbandingan Validitas Fuzzy Clustering pada Fuzzy C-Means Dan Particle Swarms Optimazation (PSO) pada Pengelompokan Kelas, JISKa, Vol. 4, No. 1, Mei, 2019, Pp. 22 - 37

Yu Shaoyong et al, *Combination of Self-organizing Map and k-means Methods of Clustering for Online Games Marketing*, Sensors and Materials, Vol. 32, No. 8 (2020) 2697–2707 MYU Tokyo <https://doi.org/10.18494/SAM.2020.2800>

#### **PUSTAKA BUKU**

Heryadi Yaya, Wahyono Teguh, 2020, *Machine Learning (Konsep dan Implementasi)*, Gava Media, Yogyakarta

Irwansyah Edy, Faisal M, 2015, *Advanced Clustering: Teori dan Aplikasi*, Deepublish, Yogyakarta

Ignacio Rojas, Gonzalo Joya, Andreu Catala, 2015, *Advances in Computational Intelligence Part II*, page 550, Springer

Jollyta Deny, Siddik M, Mawengkang Herman, Efendi S, 2021, *Teknik Evaluasi Cluster Solusi Menggunakan Python dan Rapidminer*, Deepublish, Yogyakarta

Negnevitsky Michael, 2011, *Artificial Intelligence: A Guide to Intelligent System, Third edition*, Pearson Education Limited, England

Sugiyono, 2013, *Metode Penelitian Kuantitatif Kualitatif dan R&D*, Alfabeta, Bandung

#### **PUSTAKA LAPORAN PENELITIAN**

Putri Marina Marsudi, 2015, *Pengelompokan Kabupaten/Kota Diprovinsi Jawa Timur Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Metode Kohonen Som Dan K-Means*, Tugas Akhir, Jurusan Statistika FMIPA, ITS, Surabaya





## Lampiran A

Titik Persentase Distribusi t (df = 1 – 40)

Pr	0.25	0.10	0.05	0.025	0.01	0.005	0.001
df	0.50	0.20	0.10	0.050	0.02	0.010	0.002
1	1.0000	3.0778	6.31375	12.70620	31.82052	63.65674	318.30884
2	0.81650	1.88502	2.91999	4.30265	6.96456	9.52484	22.32712
3	0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6	0.71756	1.43976	1.94318	2.44891	3.14267	3.70743	5.20763
7	0.71114	1.41482	1.89458	2.36462	2.99795	3.49940	4.78529
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35538	4.50079
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.29681
10	0.69981	1.37216	1.81246	2.22814	2.76377	3.16927	4.14370
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01226	3.85198
14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	3.78739
15	0.69120	1.34061	1.75305	2.13145	2.60248	2.94671	3.73283
16	0.69013	1.33676	1.74588	2.11991	2.58349	2.92070	3.68615
17	0.68920	1.33338	1.73961	2.10982	2.56805	2.89823	3.64577
18	0.68836	1.33039	1.73406	2.10092	2.55238	2.87944	3.61048
19	0.68760	1.32773	1.72913	2.09302	2.53948	2.86383	3.57940
20	0.68696	1.32534	1.72472	2.08596	2.52798	2.85034	3.55181
21	0.68635	1.32319	1.72074	2.07961	2.51785	2.83136	3.52715
22	0.68581	1.32124	1.71714	2.07387	2.50832	2.81676	3.50499
23	0.68531	1.31946	1.71387	2.06866	2.49907	2.80374	3.48496
24	0.68485	1.31784	1.71088	2.06390	2.49216	2.79694	3.46678
25	0.68443	1.31635	1.70814	2.05954	2.48511	2.78744	3.45019
26	0.68404	1.31497	1.70562	2.05553	2.47863	2.77871	3.43500
27	0.68368	1.31370	1.70329	2.05183	2.47266	2.77068	3.42103
28	0.68335	1.31253	1.70113	2.04841	2.46714	2.76326	3.40816
29	0.68304	1.31143	1.69910	2.04523	2.46202	2.75638	3.39624
30	0.68276	1.31042	1.69726	2.04227	2.45726	2.75000	3.38518
31	0.68249	1.30946	1.69552	2.03951	2.45282	2.74404	3.37490
32	0.68223	1.30867	1.69389	2.03693	2.44868	2.73848	3.36531
33	0.68200	1.30774	1.69236	2.03452	2.44479	2.73326	3.35634
34	0.68177	1.30685	1.69092	2.03224	2.44115	2.72838	3.34793
35	0.68156	1.30621	1.68957	2.03011	2.43772	2.72381	3.34005
36	0.68137	1.30551	1.68830	2.02809	2.43449	2.71948	3.33262
37	0.68116	1.30485	1.68709	2.02619	2.43145	2.71541	3.32563
38	0.68100	1.30423	1.68595	2.02438	2.42857	2.71156	3.31903
39	0.68083	1.30364	1.68488	2.02268	2.42584	2.70791	3.31279
40	0.68067	1.30308	1.68385	2.02108	2.42326	2.70446	3.30688

Catatan: Probabilitas yang lebih kecil yang ditunjukkan pada judul tiap kolom adalah luas daerah dalam satu ujung, sedangkan probabilitas yang lebih besar adalah luas daerah dalam kedua ujung

## Lampiran B

Correlation measure - 0.52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster0	-0.08493551092616...	0.6223521210536...	34
Row2	Uc	Cluster0	0.39106071201437...	0.0183568907277...	34

Correlation measure - 0.52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster1	-0.13202648571986...	0.4427418178050...	34
Row2	Uc	Cluster1	0.10221416060751...	0.5530485432691...	34

Correlation measure - 0.52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster2	-0.3801249681235852	0.0222041375399...	34
Row2	Uc	Cluster2	0.07024985224009...	0.6839144779302...	34

Correlation measure - 0.52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster3	-0.08452588236602...	0.6240379365031...	34
Row2	Uc	Cluster3	-0.5991043210026753	1.1310374650748...	34

Correlation measure - 0:52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster4	0.5816416938133129	1.9878057540023...	34
Row2	Uc	Cluster4	0.5428820543354259	6.2389729240819...	34

Correlation measure - 0:52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster5	0.08659411598524...	0.6155465637838...	34
Row2	Uc	Cluster5	0.0328571087884254	0.8491239359128...	34

Correlation measure - 0:52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster6	-0.16912690598780...	0.3240958001892...	34
Row2	Uc	Cluster6	-0.2782615803414361	0.1003239561014...	34

Correlation measure - 0:52 - Linear Correlation

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 5 Properties Flow Variables

Row ID	S First col...	S Second...	D Correlation value	D p value	I Degree...
Row0	Mc 1kg	Uc	0.08334281632476...	0.6289173568365...	34
Row1	Mc 1kg	Cluster7	0.6230802217625615	4.9330485782927...	34
Row2	Uc	Cluster7	-0.34891780035407...	0.0370079363554...	34

## Lampiran C

Row ID	D Mc 1kg	D Uc	D Cluster0	D Cluster1	I Winner...
A1	999.99	0.009			0
A2	1,000.017	0.012			0
A3	1,000.003	0.011			0
A4	1,000.004	0.009			0
A5	1,000.039	0.003			1
A6	1,000.001	0.01			0
A7	1,000.035	0.013			0
A8	1,000.005	0.009			0
A9	1,000.005	0.005			1
A10	999.997	0.006			1
A11	1,000.004	0.01			0
A12	1,000.01	0.011			0
A13	1,000.007	0.009			0
A14	1,000	0.009			0
A15	1,000	0.014			0
A16	999.996	0.009			0
A17	1,000.008	0.009			0
A18	1,000	0.006			1
B1	1,000.001	0.009			0
B2	1,000.004	0.009			0
B3	1,000	0.011			0
B4	999.996	0.01			0
B5	1,000.003	0.003			1
B6	1,000.007	0.009			0
B7	1,000.009	0.006			0
B8	1,000.001	0.005			1
B9	1,000.004	0.005			1
B10	1,000.004	0.009			0
B11	1,000.005	0.004			1
B12	1,000	0.009			0
B13	1,000.006	0.009			0
B14	1,000.003	0.006			1
B15	1,000.005	0.009			0
B16	1,000.018	0.016			0
B17	1,000.003	0.009			0
B18	1,000.003	0.007			1

Hasil Klastering Parameter 1

Row ID	D	Mt	kg	D	Ut	D	Cluster	D	Cluster	D	Cluster	D	Cluster	I	Waktu
A1	999.99		0.009											1	
A2	1.000.017		0.017											2	
A3	1.000.003		0.011											1	
A4	1.000.004		0.009											2	
A5	1.000.029		0.023											3	
A6	1.000.001		0.01											1	
A7	1.000.025		0.013											0	
A8	1.000.005		0.009											2	
A9	1.000.005		0.005											3	
A10	999.997		0.004											3	
A11	1.000.004		0.01											2	
A12	1.000.01		0.011											2	
A13	1.000.007		0.009											2	
A14	1.000		0.009											3	
A15	1.000		0.024											3	
A16	999.998		0.009											3	
A17	1.000.004		0.004											2	
A18	1.000		0.004											3	
B1	1.000.001		0.006											1	
B2	1.000.004		0.009											2	
B3	1.000		0.011											1	
B4	999.998		0.01											1	
B5	1.000.003		0.003											3	
B6	1.000.007		0.009											2	
B7	1.000.005		0.008											2	
B8	1.000.001		0.005											1	
B9	1.000.004		0.005											1	
B10	1.000.004		0.005											2	
B11	1.000.005		0.004											3	
B12	1.000		0.009											1	
B13	1.000.006		0.009											2	
B14	1.000.003		0.006											3	
B15	1.000.005		0.009											2	
B16	1.000.018		0.014											3	
B17	1.000.003		0.008											2	
B18	1.000.003		0.007											3	

### Hasil Klustering Parameter 2

Row ID	D	Mt	kg	D	Ut	D	Cluster	D	Cluster	D	Cluster	D	Cluster	D	Cluster	I	Waktu
A1	999.99		0.009													1	
A2	1.000.017		0.017													2	
A3	1.000.003		0.011													1	
A4	1.000.004		0.009													2	
A5	1.000.029		0.023													3	
A6	1.000.001		0.01													1	
A7	1.000.025		0.013													2	
A8	1.000.005		0.009													2	
A9	1.000.005		0.005													3	
A10	999.997		0.008													3	
A11	1.000.004		0.01													2	
A12	1.000.01		0.011													2	
A13	1.000.007		0.009													2	
A14	1.000		0.009													3	
A15	1.000		0.024													3	
A16	999.998		0.009													3	
A17	1.000.004		0.004													2	
A18	1.000		0.004													3	
B1	1.000.001		0.006													1	
B2	1.000.004		0.009													2	
B3	1.000		0.011													1	
B4	999.998		0.01													1	
B5	1.000.003		0.003													3	
B6	1.000.007		0.009													2	
B7	1.000.005		0.008													2	
B8	1.000.001		0.005													1	
B9	1.000.004		0.005													1	
B10	1.000.004		0.009													2	
B11	1.000.005		0.004													3	
B12	1.000		0.009													1	
B13	1.000.006		0.009													2	
B14	1.000.018		0.014													3	
B15	1.000.003		0.008													2	
B16	1.000.003		0.007													3	

### Hasil Klustering Parameter 3



Hasil Klastering Parameter 4



Hasil Klastering Parameter 5



Hasil Klustering Parameter 6



Hasil Klustering Parameter 7