

KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE *K-MEANS CLUSTERING* (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA)

Sabrina Aulia Rahmah¹⁾

¹⁾Teknologi Informasi / ¹⁾Universitas Dharmawangsa

¹⁾Jl.K.L. Yos Sudarso No. 224 Glugur Kota Medan

Email: ¹⁾sabrinaaulia@dharmawangsa.ac.id

Abstrak— *Data Mining* adalah suatu proses pencarian atau penggalian suatu informasi yang dilakukan di dalam sebuah data yang besar. *Clustering* (pengelompokan) merupakan salah satu metode yang dilakukan di dalam proses *Data Mining*. Pengelompokan yang terjadi adalah mengelompokkan penjualan pestisida tanaman padi. Atribut yang digunakan dalam melakukan pengelompokan ini adalah total penjualan selama satu musim dan harga pestisida. Pengelompokan ini dilakukan agar pihak toko dapat mengetahui penjualan jenis pestisida apa saja yang laku dan tidak laku selama satu musim agar tidak terjadi penumpukan barang digudang. Di dalam *clustering* terdapat beberapa metode salah satunya adalah metode *K-means*. *K-means* adalah salah satu proses metode pengelompokan yang non hirarki yang cara bekerjanya membagi data ke dalam bentuk satu atau lebih *cluster*. *Cluster* yang dibuat ada 3 macam *cluster* yaitu, Sangat laku, laku dan tidak laku. Hasil dari *cluster* tersebut 53 *items* tergolong ke dalam kelompok penjualan pestisida yang sangat laku, 21 *items* tergolong ke dalam penjualan pestisida yang laku, dan 126 *items* digolongkan ke dalam penjualan pestisida yang tidak laku. Aplikasi yang digunakan adalah *software Rapid Miner 7.5*.

Kata Kunci— *Data Mining, Clustering, Pestisida, Algoritma K-Means, Rapid Miner 7.5*

Abstract— *Data Mining* is a process of finding or extracting information that is carried out in a large data. *Clustering* is one of the methods carried out in the *Data Mining* process. Groupings that occur are grouping sales of rice pesticides. The attributes used in carrying out this grouping are total sales during one season and the price of pesticides. This grouping is done so that the store can find out what kind of pesticide sales are selling and not selling for one season so there is no accumulation of goods in the warehouse. In *clustering*, there are several methods, one of which is the *K-means* method. *K-means* is a non-hierarchical grouping process that works by dividing data into one or more clusters. Clusters made there are 3 types of clusters, namely, Very salable, salable and unsold. The results of the cluster 53 items belong to the selling pesticides group, 21 items belong to the sale of pesticides that sell, and 126 items are classified as selling pesticides that do not sell. The application used is the *Rapid Miner 7.5* software.

Keywords— *Data Mining, Clustering, Pesticides, K-Means Algorithm, Rapid Miner 7.5*

I. PENDAHULUAN

1.1 Latar Belakang

Toko Juanda Tani merupakan salah satu toko yang terletak di Kecamatan Hutabayu Raja di Kabupaten Simalungun, Sumatera Utara. Toko ini menjual berbagai macam jenis pestisida pada tanaman padi. Namun di dalam melakukan proses pencatatan data-data, toko masih menggunakan sistem secara manual sehingga toko tidak dapat mengetahui dan tidak dapat mengelompokkan jenis pestisida apa saja yang paling tinggi penjualannya. Sehingga kesulitan yang dialami oleh toko adalah seringkali terjadi kekurangan stok barang yang laku dikarenakan penjualan yang cukup tinggi. Data penjualan yang terjadi pada toko Juanda Tani setiap bulannya juga berbeda-beda sehingga perlu adanya pengelompokan (klasterisasi/ *clustering*).

Salah satu metode yang terjadi di dalam *Data Mining* adalah *clustering*. *Clustering* merupakan salah satu metode analisa data yang dapat digunakan dalam memecahkan masalah dalam suatu pengelompokan data. Salah satu metode yang ada di dalam metode *clustering* adalah metode *K-means*. Metode *K-means* merupakan suatu metode yang dapat melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu yang relatif cepat dan efisien.

Ramadhani *et al* (2014), menyatakan bahwa proses *Data Mining* dengan menggunakan metode *K-Means* dapat mengelompokkan perubahan kondisi cuaca yang terjadi di kota Pelaihari dengan menghasilkan 4 pusat *cluster*, seperti meningkatnya kecepatan angin, meningkatnya kecepatan temperatur,

meningkatnya penguapan dan menurunnya kelembaban udara.

1.2 Perumusan Masalah

Berdasarkan dari uraian latar belakang masalah diatas dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana proses klusterisasi dengan metode *K-means* dapat diterapkan dalam mengelompokkan penjualan barang berdasarkan tingkat penjualannya?
2. Bagaimana mengelompokkan data-data penjualan untuk mengetahui tindakan apa yang akan diambil?

1.3 Tujuan Penelitian

Tujuan dalam penelitian ini adalah:

1. Untuk mengetahui skema penjualan pestisida yang terjadi setiap bulannya.
2. Mengelompokkan data penjualan produk yang laku dijual dan yang tidak laku dijual di toko.
3. Menerapkan *Data Mining* dalam mengelompokkan pola penjualan menggunakan algoritma *K-means* pada toko racun.
4. Mengaplikasikan *Data Mining* dalam pola penjualan dengan memanfaatkan *software RapidMiner 7.5*.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Mempercepat waktu yang dipakai dalam mengelompokkan suatu produk penjualan.
2. Memberikan pengetahuan dan mempermudah memberikan informasi kepada pihak toko untuk mengetahui produk mana saja yang laku dijual dan tidak laku dijual
3. Memberikan informasi agar pihak toko tidak salah lagi dalam memesan produk ke pabrik sehingga tidak terjadi penumpukan barang yang ada di gudang.
4. Membantu pengelola toko dalam membuat laporan penjualan setiap bulannya.

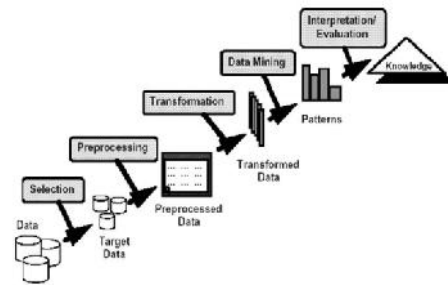
II. LANDASAN TEORI

2.1 Knowledge Discovery in Database (KDD)

Data Mining merupakan proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, seperti data *warehouse* dan tempat penyimpanan data lainnya. *Data mining* juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam database yang dikenal dengan istilah *Knowledge Discovery in Database* (KDD) (Ramadhani *et al*, 2014). Salah satu tahapan dalam keseluruhan proses KDD adalah *Data Mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Nasari dan Darma, 2015) :

1. *Data Selection*
2. *Pre-processing / Cleaning*

3. *Transformation*
4. *Data Mining*
5. *Interpretation / Evaluation*



Gbr 1. Knowledge Discovery in Database (KDD)

2.2 Data Mining

Data Mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. *Data Mining* biasa juga disebut dengan “Data atau *knowledge discovery*” atau menemukan pola tersembunyi pada data. *Data Mining* adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna (Arta *et al*, 2016).

2.3 Pengertian Clustering

Menurut Metisen dan Sari (2015) *Clustering* atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang *similar* satu sama lain dalam *cluster* yang sama dan *disimilar* terhadap objek-objek yang berbeda *cluster*. Objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya.

2.4 Partitioning Clustering

Partitioning Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Konsep dasar dari *Partitioning Clustering* adalah membagi n jumlah *cluster* ke dalam k *cluster*. Metode ini merupakan metode pengelompokan yang bertujuan mengelompokkan objek sehingga jarak antar tiap objek ke pusat kelompok di dalam satu kelompok adalah minimum.

2.5 K-Means Clustering

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang

berbeda dikelompokkan ke dalam kelompok yang lainnya.

Adapun langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut :

1. Menentukan jumlah *cluster k* yang akan dibentuk.
2. Inisialisasi *k* pusat *cluster* dapat dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberikan nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

D (i, j) = Jarak data ke *i* ke pusat cluster *j*

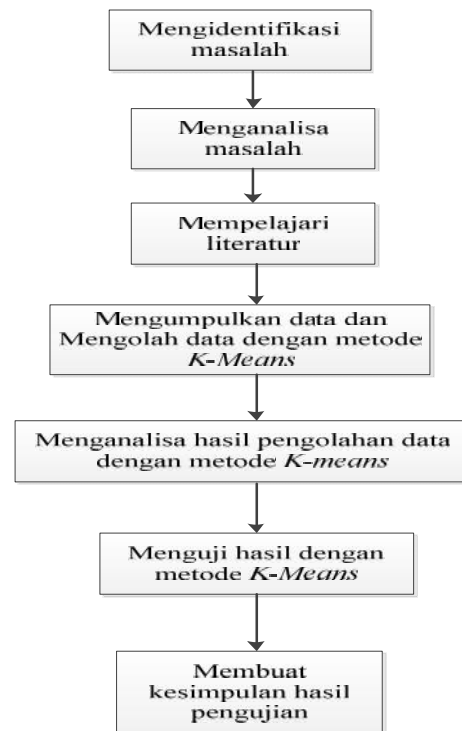
X_{ki} = Data ke *i* pada atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut ke *k*

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

III. METODOLOGI PENELITIAN

Pada bab ini akan diuraikan metodologi dan kerangka kerja penelitian yang digunakan dalam penyelesaian penelitian ini. Kerangka kerja ini merupakan tahapan – tahapan yang akan dilakukan dalam rangka penyelesaian masalah yang akan dibahas. Adapun kerangka kerja dari penelitian ini dapat dilihat pada gbr 2.



Gbr.2 Metodologi Penelitian

IV. ANALISA DAN PEMBAHASAN

Pada analisa sistem yang berjalan di toko Juanda Tani masih menggunakan sistem pencatatan secara manual. Setiap pembeli yang berbelanja akan dicatat didalam buku penjualan. Data yang diperoleh dari toko Juanda Tani Kecamatan Hutabayu Raja merupakan data penjualan selama satu musim yaitu data selama 4 bulan penjualan dari bulan Maret- Juni 2016.

Data yang digunakan dalam hal ini terbagi atas dua jenis, yaitu data sampel dan data latih. Data sampel yang digunakan sebanyak 15 data dan data latih yang digunakan sebanyak 200 data (keseluruhan jumlah data pestisida). Data dapat dilihat pada Gbr 3.

Tabel 1. Data Sampel

| No | Nama Pestisida | Total Penjualan | Harga |
|----|----------------|-----------------|-------|
| 1 | Combiotik | 63 | 85000 |
| 2 | Antracol | 10 | 63000 |
| 3 | Bio ziel | 14 | 35000 |
| 4 | Super K | 2 | 28500 |
| 5 | Besnoid B | 60 | 60000 |
| 6 | Bestok B | 21 | 35000 |
| 7 | Marsal | 106 | 38000 |
| 8 | Nazole B | 7 | 18000 |
| 9 | Bless | 1 | 16000 |
| 10 | Bibit Cikerang | 1 | 16500 |
| 11 | Za | 209 | 85000 |

| | | | |
|----|------------|-----|--------|
| 12 | Centafur | 31 | 23000 |
| 13 | Phonska | 163 | 130000 |
| 14 | Balistik B | 50 | 120000 |
| 15 | Dimpo K | 21 | 40000 |

Pusat awal *Cluster* atau *centroid* didapatkan secara random (acak), untuk penentuan awal *Cluster* ini dapat diasumsikan sebagai berikut:

Pusat *Cluster* 1, diambil dari data ke- 11 : (209, 85000)

Pusat *Cluster* 2, diambil dari data ke- 1 : (63, 85000)

Pusat *Cluster* 3, diambil dari data ke- 3 : (14, 35000)

Tahapan berikutnya adalah menghitung jarak antara data dengan pusat awal *cluster* menggunakan persamaan *euclidean distance*.

Setelah nilai pusat *cluster* didapat selanjutnya dilihat nilai dengan jarak terpendek.

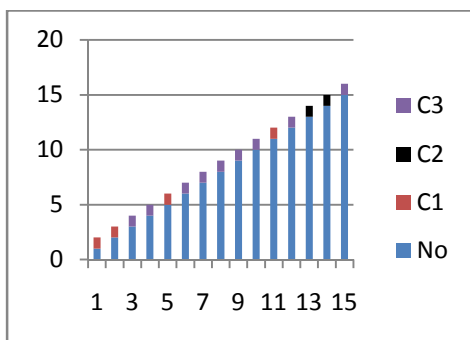
Proses iterasi terus dilakukan hingga nilai *centroid* dan posisi atau letak *cluster* tidak berpindah lagi. Pada penelitian ini iterasi dilakukan sampai iterasi 3.

Setelah nilai *centroid* diperoleh selanjutnya akan diperoleh letak atau posisi *cluster*. Berikut adalah posisi *cluster* dari *centroid* iterasi ke-3, dapat dilihat pada tabel 4.2 berikut ini :

Tabel 2. Hasil dari Iterasi 1

| No | C1 | C2 | C3 |
|----|----|----|----|
| 1 | | 1 | |
| 2 | | 1 | |
| 3 | | | 1 |
| 4 | | | 1 |
| 5 | | | 1 |
| 6 | | | 1 |
| 7 | | | 1 |
| 8 | | | 1 |
| 9 | | | 1 |
| 10 | | | 1 |
| 11 | 1 | | |
| 12 | | | 1 |
| 13 | | 1 | |
| 14 | | 1 | |
| 15 | | | 1 |

Hasil pengolahan data secara manual dengan sampel 15 items pestisida dapat dilihat pada gbr 4 berikut ini :



Gbr 4. Grafik dari Hasil Clustering

Kesimpulan dari hasil perhitungan dan tabel di atas adalah **Y1, Y2, Y5, Y11** adalah anggota *Cluster* 1, **Y13, Y14** adalah anggota *Cluster* 2, dan **Y3, Y4, Y6, Y7, Y8, Y9, Y10, Y12, Y15** adalah anggota *Cluster* 3. Dikarenakan hasil dari iterasi 3 dan iterasi 2 dengan posisi *Cluster* pada iterasi 2 sama/ tidak berubah maka proses pencarian dan perhitungan iterasi dihentikan.

V. IMPLEMENTASI DAN HASIL

Implementasi menggunakan *software Rapid Miner* dengan data sampel sebanyak 200 items pestisida. Berikut adalah pengolahan data menggunakan *K-Means* untuk sampel data dapat dilihat pada Gbr 3. *Example Set (read Excel)* yang Berisi Data berikut ini :

| Item No | Nama | Harga |
|---------|------------|--------|
| 1 | Centafur | 23000 |
| 2 | Phonska | 130000 |
| 3 | Balistik B | 120000 |
| 4 | Dimpo K | 40000 |

Gbr 3. Example Set (read Excel) yang Berisi Data

Berdasarkan hasil pengujian *Data Mining* yang menggunakan algoritma *K-means Clustering* yang telah dilakukan secara perhitungan manual dan pengujian dengan aplikasi *software RapidMiner 7.5*, maka dapat ditarik kesimpulan nilai *centroid* yang diperoleh berbeda dikarenakan jumlah data yang tidak sama sedangkan nilai yang ada didalam *cluster* yang dihasilkan adalah sama. Nilai *centroid* akhir dapat dilihat pada tabel 5.1 berikut ini.

Tabel 3. Nilai Centroid Akhir Data Keseluruhan

| Attribute | Cluster 2 | Cluster 1 | Cluster 0 |
|-----------|-----------|------------|-----------|
| Total | | | |
| Penjualan | 20.075 | 32.905 | 10.960 |
| Harga | 60594.340 | 113142.857 | 27107.143 |

Dari tabel di atas dapat dilihat bahwa atribut total penjualan memiliki nilai tertinggi pada *cluster* 2 dan untuk atribut harga pada *cluster* 2.

VI. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan oleh penulis pada Toko Juanda Tani Kecamatan Hutabayu Raja dalam mengelompokkan jenis-jenis pestisida dan pengujian data dengan memanfaatkan *software RapidMiner 7.5*, dapat ditarik kesimpulan bahwa:

1. Metode *K-Means clustering* dapat diterapkan pada penjualan Pestisida di Toko Jaunda Tani, sehingga metode ini sangat membantu dalam mengelompokan pola penjualan selama satu musim.
2. Menentukan titik pusat (*centroid*) pada tahap awal metode *K-Means* sangat berpengaruh pada hasil *cluster* seperti pada hasil pengujian yang dilakukan dengan menggunakan 15 *dataset* dengan 200 *dataset* dan *centroid* yang dihasilkan berbeda dan menghasilkan hasil *cluster* yang berbeda pula.
3. Hasil akhir dari 200 *items* penjualan pada Bulan Maret- Juni 2016 adalah Penjualan Pestisida Sangat Laku sebanyak 53 *items*, Penjualan Laku sebanyak 21 *items* dan Penjualan Tidak laku sebanyak 126 *items*.
4. Pengujian dengan *RapidMiner 7.5* sangat efektif dan akurat karena nilai *centroid* yang didapat dengan perhitungan yang dilakukan secara manual dan aplikasi adalah sama.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Risanuri Hidayat yang telah meluangkan waktu untuk membuat template ini.

REFERENSI

- [1] Agustin F.E.M., Fitria A., dan S.A.H., (2015). “Implementasi Algoritma *K-Means* Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta)”. *Jurnal Teknik Informatika* Vol. 8 No. 1 April 2015, (Page 73-78).
- [2] Arta I.K.J., Indrawan G., dan Dantes. R., (2016). “Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode *Technique For Others Reference By Similarity To Ideal Solution*”. *Jurnal Sains dan Teknologi*, Vol. 5, No. 2, Oktober 2016. ISSN: 2303-3142 (page 747-760).
- [3] Dalhatu K., Sim H.T.A., (2016). “Density base *K-Mean’s Cluster Centroid Initialization Algorithm*”. *International Journal of Computer Applications* (0975 – 8887) Volume 137 – No.11, March 2016, (pages 48-51).
- [4] Handoko K., (2016). “Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran pada Instansi Perguruan Tinggi Menggunakan Metode *K-Means Clustering (Studi Kasus Di Program Studi TKJ Akademi Komunitas Solok Selatan)*”. *TEKNOSI*, Vol. 02, No. 03, Desember 2016. ISSN: 2476- 8812, (page 31-40)
- [5] Kamagi H.D., Hansun S., (2014). “Implementasi Data Mining dengan Algoritma *C4.5* untuk Memprediksi Tingkat kelulusan Mahasiswa”. *ULTIMATICS*, Vol. VI, No. 1, Juni 2014. ISSN: 2085-4552, (page 15-20).
- [6] Metisen M.B., Sari L.H., (2015). “Analisis *Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila*”. *Jurnal Media Infotama* Vol. 11 No. 2, September 2015. ISSN: 1858 – 2680, (page 110-118).
- [7] Nasari F., Darma S., (2015). “Penerapan *K-means Clustering pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)*”. *Seminar Nasional Teknologi Informasi dan Multimedia 2015*. ISSN : 2302-3805, (page 73-78).
- [8] Ramadhani A., Farmadi A., Budiman I., (2014). “*Clustering Data Cuaca Untuk Pengenalan Pola Perioditas Iklim Wilayah Pelaihari Dengan Metode Fuzzy C-Means*”. *Jurnal Teknologi & Informasi* Vol. 3 No. 1, Juni 2014. ISSN: 2087-6920, (page 57-64).
- [9] Surmayanti., Marfalino H., Rahmi A., (2015). “Penerapan *Analysis Clustering Pada Penjualan Komputer Dengan Perancangan Aplikasi Data Mining Menggunakan Algoritma K-Means (Study Kasus Toko Tri Buana Komputer Kota Solok)*”.
- [10] Prosiding Seminar Ilmiah Nasional Teknologi Komputer (SENATKOM 2015), Vol. 1, Oktober 2015. ISSN : 2460– 4690, (page 50-59).
- [11] Virmani D., Taneja S., Malhotra G., (2015). “*Normalization Based K-means Clustering Algorithm*”. *International Journal of Advanced Engineering Research and Science (IJAERS)* ,Vol. 2, Issue. 2, February 2015. ISSN: 2349-6495, (page 36-40).
- [12] Yuantari MG. C., Widiarnoko B., Sunoko H.R., (2013). “*Tingkat Pengetahuan Petani dalam Menggunakan Pestisida (Studi kasus di Desa Curut Kecamatan Penawangan Kabupaten Grobongan)*”. *Prosiding Seminar Nasional Pengelolaan Sumberdaya Alam dan Lingkungan 2013*. ISBN 978-602-17001-1-2, (page 142-148).