

IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKAN CUSTOMER PT. KASANA TEKNINDO GEMILANG MENGGUNAKAN ALGORITMA K-MEANS

Teguh Budi Santoso, Raisal Zulkarnain

ABSTRAK

PT. Kasana Teknindo Gemilang merupakan perusahaan yang bergerak dibidang distributor penjualan dan *service* alat-alat berat yang berada di kawasan jakarta barat. Salah satu hal yang penting dalam pengelolaan *customer* adalah bagaimana suatu perusahaan dapat mempertahankan *customer* yang dimilikinya. Usaha untuk mempertahankan *customer* ini menjadi hal yang penting bagi PT. Kasana Teknindo Gemilang mengingat semakin banyaknya perusahaan-perusahaan pesaing membuat PT. Kasana Teknindo Gemilang harus mempunyai strategi baru agar dapat mempertahankan customer potensialnya dengan cara memberikan diskon.

Penerapan data mining dapat membantu untuk menganalisa data yang di peroleh dari transaksi. Pada penelitian ini dilakukan dengan penggunaan algoritma *K-Means*, dengan harapan dapat memberikan solusi untuk penentuan customer potensial berdasarkan pengelompokan. Pada penelitian ini proses algoritma *K-Means* akan diimplementasikan kedalam *software* untuk menghasilkan penentuan penggalian data dalam mencari penentuan customer potensial yang selama ini data sudah ada pada PT. Kasana Teknindo Gemilang.

Berdasarkan proses algoritma k-means pada penelitian ini dengan memiliki himpunan attribut yaitu: quantity, jumlah transaksi dan total harga, dapat ditarik kesimpulan berdasarkan pengelompokan dimana hasil didapat untuk menjadi titik centroid yaitu C1 (PT. Sumiden Serasi Wire Products), C2 (PT. Surya Intan Sentana Jaya), C3 (PT. Ika Maestro Industri). Dalam penelitian ini telah berhasil di buat program K-Means untuk pengelompokan *customer* potensial, informasi yang ditampilkan berupa nilai centroid dari tiap-tiap cluster dan kelompok-kelompok pelanggan potensial

Kata Kunci : Customer potensial , *K-Means*, quantity, jumlah transaksi dan total harga

ABSTRACT

PT. Kasana Teknindo Gemilang is a company engaged in the distributor of heavy equipment sales and service located in the area of West Jakarta. One of the most important things in customer management is how a company can maintain its customer. Attempts to maintain this customer becomes an important thing for PT. Kasana Teknindo Gemilang considering the increasing number of competing companies make PT. Kasana Teknindo Gemilang must have a new strategy in order to maintain its potential customer by giving discount.

Implementation of data mining can help to analyze the data obtained from the transaction. In this research is done by using K-Means algorithm, hopefully can give solution for determination of potential customer based on grouping. In this research the process of K-Means algorithm will be implemented into the software to generate the determination of data mining in searching for the determination of potential customer which has been existing data at PT. Kasana Teknindo Gemilang.

Based on the process of k-means algorithm in this study by having set of attributes are: quantity, total transaction and total price, can be drawn conclusion based on the grouping where the results obtained to become the point of centroid is C1 (Sumiden Serasi Wire Products), C2 (PT. Surya Intan Sentana Jaya), C3 (PT Ika Maestro Industry). In this research has been successfully made K-Means program for the grouping of potential customers, the information displayed in the form of centroid value of each cluster and groups of potential customers. Keywords: K-Means, Clustering, Potential

Keywords: *Potential customers, K-Means, quantity, number of transactions and total price*

PENDAHULUAN

Sebagai perusahaan yang bergerak di bidang distributor penjualan dan service pada PT. Kasana Teknindo Gemilang, sudah seharusnya memiliki suatu strategi dalam pengelolaan manajemen customer. Dikarenakan timbulnya perusahaan-perusahaan pesaing membuat PT. Kasana Teknindo Gemilang, harus memiliki strategi yang dapat menjadi acuan proses kesinambungan antara customer dengan perusahaan, berkaitan dengan data bahwa terdapat 842 pada tahun 2016 transaksi yang mengalami kemunduran, sedangkan kalau dilihat pada tahun 2016 terdapat 895 transaksi di tahun 2015.

Penerapan metode data mining dengan menggunakan algoritma *k-means* dengan proses penentuan berdasarkan titik centroid, untuk mendapatkan hasil penentuan customer potensial yang sudah ada, agar dapat memberikan solusi untuk perusahaan mencari PT apa yang akan diberikan apresiasi dari perusahaan. Data mining merupakan proses penggalian data berdasarkan jumlah data yang besar, proses pengelompokan dengan menggunakan algoritma *k-means* dengan mencari dari jumlah data dengan proses mencari titik centroid untuk pengelompokan data – data.

Adapun berdasarkan penelitian-penelitian sebelumnya yang sudah menerapkan data mining dengan menggunakan algoritma *k-means*, pada penelitian . (Afrisawati, 2013) dimana pada penelitian ini menggunakan algoritma *k-means* untuk mencari customer potensial berdasarkan himpunan attribute yang dimiliki antara lain: data nama pelanggan, jumlah transaksi dan total belanja. Pada penelitian ini telah didapat hasil berupa titik centroid terbesar dalam menentukan customer potensial.

Pada penelitian ini ingin menerapkan algoritma *k-means* berdasarkan himpunan quantity, jumlah transaksi dan total harga, untuk menjadi proses pengelompokan data-data perusahaan dalam hal menentukan siapa saja daftar rekan PT. Kasana Teknindo Gemilang yang saat ini sudah berjalan, untuk diberikan reward, dan menjadikan factor dalam menjalin kerjasama yang baik dalam mempertahankan untuk selalu menjadi customer PT. Kasana Teknindo Gemilang.

TINJAUAN PUSTAKA

Adapun pada penelitian sebelumnya dengan topic untuk menentukan customer potensial menggunakan Data Menggunakan menggunakan model Algoritma K-Means, pada penelitian oleh Afrisawati pada tahun 2013 yang berobjek di CV Cahaya Sejati bergerak dibidang retail. Pada penelitian ini menggunakan algoritma K-Means untuk mencari pelanggan yang potensial berdasarkan himpunan jumlah transaksi dan total

belanja, pada penelitian ini menggunakan *tools tanagra* untuk memproses Algoritma K-Means berdasarkan data nama-nama pelanggan. Pada penelitian ini dapat ditarik kesimpulan bahwa dapat menjadi suatu proses dalam mencari data pelanggan yang potensial. Dalam hal ini, penerapan data mining mampu menganalisa data transaksi yang cukup banyak dengan cara *clustering*. Algoritma K-Means merupakan salah satu metode *non-hierarchical clustering* yang dapat mengelompokkan data ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut. Algoritma K-Means dapat menjadi solusi dalam memberikan informasi untuk bahan pertimbangan pihak perusahaan dalam pemberian diskon dan pelayanan yang baik terhadap *customer*.

Pada tahun 2014 penelitian yang dilakukan oleh Fina Nasari, Surya Darma dalam hal untuk penentuan pengambilan jurusan berdasarkan nilai waktu SMA, pada penelitian ini dapat disimpulkan berdasarkan proses 2 cluster dari hasil cluster dapat diambil bahwa jika asal sekolah adalah SMA maka rata-rata mengambil jurusan Sistem Informasi dan jika asal sekolah SMK rata-rata jurusan yang diambil adalah Teknik Informatika. (Fina Nasari, Surya Darma, 2015)

Pada penelitian yang dilakukan oleh Istiqomah Sumadikarta, Evan Abeiza menyimpulkan bahwa penerapan algoritma K-Means guna mengelompokkan data produk dan data pelanggan untuk mengetahui data yang memiliki potensi dengan melakukan proses jumlah *invoice* terhadap transaksi untuk mengelompokkan data produk sedangkan, total penjualan terhadap pelanggan dan jumlah transaksi untuk mengelompokkan data pelanggan di PT. Mega Arvia Utama. (Istiqomah Sumadikarta, Evan Abeiza, 2014)

LANDASAN TEORI

Pengertian Pelanggan Potensial

Pelanggan potensial adalah pelanggan yang loyal terhadap perusahaan. Pelanggan yang potensial umumnya akan melanjutkan pembelian produk atau jasa tersebut walaupun dihadapkan pada banyak alternatif produk atau jasa yang lebih unggul dipandang dari berbagai sudut atributnya. Menurut Kotler dalam bukunya Zulkarnain, konsumen yang loyal tidak diukur dari berapa banyak dia membeli, tapi dari berapa sering dia melakukan pembelian ulang, termasuk merekomendasikan orang untuk membeli (Zulkarnain, 2012).

Pengertian Data Mining

Data mining adalah proses menemukan pola yang menarik, dan pengetahuan data dari jumlah yang besar. (Witten, Frank, & Hall, 2011)

Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti Database System, Data Warehousing, Statistic, Machine Learning, Information Retrieval, dan Komputasi Tingkat Tinggi. Selain itu data mining didukung oleh ilmu lain seperti Neural Network, Pengenalan Pola, Spatial Data Analysis, Image Database, Signal Processing.

Tahapan Data Mining

Sebagai suatu rangkaian proses data mining dapat dibagi menjadi beberapa tahap yang bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base.

Tahapan data mining ada 6 yaitu :

- a. **Pembersihan Data (*Data Cleaning*)**
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari data base suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki.
- b. **Integrasi Data (*Data Integration*)**
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks.
- c. **Seleksi Data (*Data Selection*)**
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
- d. **Transformasi data (*Data Transformation*)**
Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal.
- e. **Proses mining**
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- f. **Evaluasi pola (*Pattern Evaluation*)**
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model karakter dievaluasi untuk menilai sepeda motor apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

Algoritma K-Means

Web Server adalah suatu program komputer yang mempunyai tanggung jawab atau tugas menerima permintaan HTTP dari komputer Klien, yang dikenal dengan nama *web browser* dan melayani dengan menyediakan respon HTTP berupa konten data, biasanya berupa halaman web yang terdiri dari dokumen HTML dan objek terkait seperti gambar dan lain-lain.

Metodologi Analisis K-Means Clustering

Metodologi dasar analisis Algoritma K-Means Clustering adalah sebagai berikut (Prasetyo, 2012) :

1. Tentukan jumlah kluster (k), tetapkan pusat kluster secara random.
2. Hitung jarak setiap data ke pusat kluster.
3. Kelompokkan data ke dalam kluster yang dengan jarak yang paling pendek.

4. Hitung pusat kluster yang baru.
5. Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

Proses klustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara random, C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat kluster. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat kluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean seperti pada persamaan berikut yaitu:

$$d_{ij} = \sqrt{\sum_{k=1}^m X_{ij} - C_{jk}^2}$$

Dimana :

d_{ij} : Jarak objek antar nilai data dan nilai pusat cluster

m : Jumlah dimensi data

X_{ij} : Nilai data dari dimensi ke- k

X_{jk} : Nilai pusat cluster dari dimensi ke- k

Sedangkan formula membangkitkan *centroid* baru, yaitu :

$$C = \frac{\sum m}{n}$$

Dimana :

C : *centroid* data

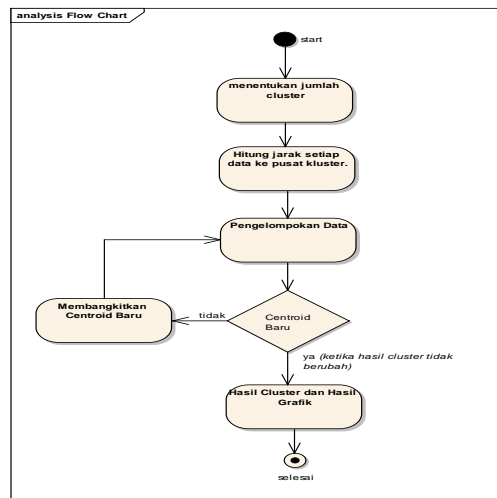
m : anggota data yang termasuk kedalam *centroid* tertentu

n : jumlah data yang menjadi anggota *centroid* tertentu

METODE PENELITIAN

Analisa Sistem Berjalan

Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, dapat dilihat pada gambar 1 Tahapan Proses Algoritma K-Means. Dimana data akan dilakukan untuk proses menentukan jumlah kluster berdasarkan jarak data ke titik pusat *centroid* yang nantinya hasil akan dapat dikelompokkan kedalam titik nilai *centroid* dari masing-masing nilai data berdasarkan dari proses perhitungan menggunakan algoritma k-means, hasil akhir akan dibuat berdasarkan grafik yang membetuk masing-masing klompok data.



Gambar 1 Tahapan Proses Algoritma K-Means

Proses Menentukan Jumlah Cluster dan Titik Centroid

Dalam penelitian ini titik pusat awal random dan didapat titik pusat dari setiap cluster dapat dilihat pada tabel 1. Titik Pusat Centroid Awal.

Tabel 1. Titik Pusat Centroid Awal.

Titik Pusat Awal	Nama Perusahaan	Qty	Jumlah Transaksi	Total Harga
C1	PT. Sumiden Serasi Wire Products	98	4	78.847.835
C2	PT. Surya Intan Sentana Jaya	29	3	25.117.000
C3	PT. Ika Maestro Industri	47	12	6.974.000

Proses Hitung Eucuidean

Berikut ini merupakan proses menghitung untuk mencari titik terdekat yang sebelumnya titik centroidnya yang telah ditentukan berikut ini:

$$d_{ij} = \sqrt{\sum_{k=1}^m X_{ij} - C_{jk}^2}$$

d(C 1, data 1)

$$= \sqrt{(data\ 1\ dimensi\ 1 - C1\ dimensi\ 1)^2 + (data\ 1\ dimensi\ 2 - c1\ dimensi\ 2)^2 + (data\ 1\ dimensi\ 3 - C1\ dimensi\ 3)^2}$$

Berikut ini menghitung data 1 dengan titik centrod C1 :

$$d(c1, data\ 1) = \sqrt{(9 - 98)^2 + (2 - 4)^2 + (1.421.000 - 78.847.835)^2}$$

$$d(c1, data\ 1) = \sqrt{7921 + 4 + 5,994949147781e15}$$

$$d(c1, data\ 1) = \sqrt{5,9949147781e15}$$

$$d(c1, data\ 1) = 77.426.835.000051$$

Dst.....

Kelompok Cluster Terpendek

Dalam penelitian ini kelompokan cluster yang terpendek untuk mencari pusat titik baru dilihat pada tabel 2 Iterasi 1 dibawah ini:

Tabel 2 Iterasi 1

Nama Perusahaan	Cluster 1	Cluster 2	Cluster 3	Terdekat	C1	C2	C3
PT. Mitra Bangun Cemerlang	77426800	23696000	5553000	5553000	0	0	1
PT. Karya Abdi Luhur	8544200	45186600	63329600	8544200	1	0	0
PT. Intan Metalindo	67900800	14170000	3973000	3973000	0	0	1
PT. Nippon Steel and Sumikin Logist	48719000	5011800	23154800	5011800	0	1	0
PT. Pundi Kecana	35689700	18041100	36184100	18041100	0	1	0
PT. Multi Sentana Baja	32834800	20896000	39039000	20896000	0	1	0
PT. Wahana Indah Gemaca	76741800	23011000	4868000	4868000	0	0	1
PT. Abadi Pratama Indonesia	68872800	15142000	3001000	3001000	0	0	1
PT. ILC Logistic Indonesia	60648700	6917900	11225100	6917900	0	1	0
CV. Anak Sejati	60227800	5497000	11646000	6497000	0	1	0
PT. Surya Intan Sentana jaya	53730800	0	18143000	0	0	1	0
PT. Fermentech Indonesia	4321200	49409600	67552600	4321200	1	0	0
Bpk. H. Budi Mulyana	77047800	23317000	5174000	5174000	0	0	1

Dst sampai dengan Iterasi 3.....

Tabel 3 Cluster 1

Nama Perusahaan	C1		
	Qty	Jumlah Transaksi	Total Harga
PT. Karya Abdi Luhur		13	70.303.625
PT. Fermentech Indonesia	53	8	74.526.625
PT. Cerestar Flour Milis	74	10	90.822.510
PT. Sumiden Serasi Wire Products	98	4	78.847.835

Dst sampai dengan Cluster 3.....

Proses Centroid Baru

Berikut ini proses menghitung centroid baru

$$C = \frac{\sum m}{n}$$

Dibawah ini cara menghitung *centroid* C1 baru :

$$C1, D1 = \frac{60 + 53 + 74 + 98}{4} = 71,25$$

$$C1, D2 = \frac{13 + 8 + 10 + 4}{4} = 8,75$$

$$C1, D3 = \frac{70.303.625 + 74.526.625 + 90.822.510 + 78.847.835}{4} = 78.625.100$$

Dst sampai dengan C3.....

Tampilan Layar Tampilan Data Perusahaan

PT. KASANA TEKNIKINDO GEMILANG

Mulai Tur | rasil Zul

Jika kamu tidak melihat logo yang ada di kiri atas, tolong reset theme yang ada di sebelah kanan (hanya untuk theme saja). *

Halaman Utama > Perusahaan > Perusahaan

Perusahaan

Data Perusahaan

Tambah Perusahaan

10 data

Cari

No	Nama Perusahaan	Aksi
1	PT. Mitra Bangun Cemerlang	Edit Hapus
2	PT. Karya Abdi Luhur	Edit Hapus
3	PT. Intan Metalindo	Edit Hapus
4	PT. Nippon Steel and Sumikin Logist	Edit Hapus
5	PT. Pundi Kecana	Edit Hapus

Print

Tampilan data Cluster

PT. KASANA TEKNIKINDO GEMILANG

Mulai Tur | rasil Zul

Jika kamu tidak melihat logo yang ada di kiri atas, tolong reset theme yang ada di sebelah kanan (hanya untuk theme saja). *

Halaman Utama > Cluster > Cluster

Cluster

Data Cluster

Tambah Cluster

10 data

Cari

No	Nama Cluster	Aksi
1	Cluster 1	Edit Hapus
2	Cluster 2	Edit Hapus
3	Cluster 3	Edit Hapus

Menampilkan 1 sampai 3 dari 3 data

Sebelumnya 1 Selanjutnya

Tampilan Nilai Perusahaan

Halaman Utama > Nilai Data > Nilai Perusahaan

Data Nilai Perusahaan

No	Nama Pegawai	Quantity	Transaksi	Total Harga	Aksi
1	PT. Mitra Bangun Cemerlang	9	2	1421000	🔗
2	PT. Karya Abdi Luhur	60	13	70303624	🔗
3	PT. Intan Metalindo	23	2	10946960	🔗
4	PT. Nippon Steel and Sumikin Logist	102	8	30128810	🔗
5	PT. Pundi Kecana	106	28	43158056	🔗
6	PT. Multi Sentana Baja	159	2	46012960	🔗

Tampilan Nilai Cluster

Halaman Utama > Nilai Data > Nilai Cluster

Data Nilai Cluster

Nama Cluster	Quantity	Transaksi	Total Harga
Cluster 1	98	4	78847800
Cluster 2	29	3	25117000
Cluster 3	47	12	6974000

Menampilkan 1 sampai 3 dari 3 data

Tampilan Proses Algorithm K-Means import excel

Halaman Utama > Import Excel

Proses | Import Excel

Import

Import Excel | Proses Temporary Excel | Hasil Proses Data

File Excel

© 2017 K-Means Oleh: Praslyo

Tampilan Hasil Proses

Iterasi	Perusahaan
11	Abadi Pratama Indonesia
11	CV. GUNA DHARMA
11	CV. PRAYOGA WILAJAYA
11	CV. SINAR ASIH
11	HENDRA KURNIAWAN
11	HUSNI MOBAROK
11	LIM KOK MENG

KESIMPULAN DAN SARAN

Kesimpulan

Pada penelitian ini dapat di tarik kesimpulan berdasarkan proses penerapan menggunakan algoritma k-means, didapat proses penentuan titik akhir berdasarkan titik centroid dari berdasarkan proses algoritma k-means pada penelitian ini dengan memiliki himpunan attribut yaitu: quantity, jumlah transaksi dan total harga, dapat ditarik kesimpulan berdasarkan pengelompokan dimana hasil didapat untuk menjadi titik centroid yaitu C1 (PT. Sumiden Serasi Wire Products), C2 (PT. Surya Intan Sentana Jaya), C3 (PT. Ika Maestro Industri). Dalam penelitian ini telah berhasil di buat program K-Means untuk pengelompokan *customer* potensial, informasi yang ditampilkan berupa nilai centroid dari tiap-tiap cluster dan kelompok-kelompok pelanggan potensial

Saran

Adapun pada penelitian ini dapat disarankan untuk penelitian selanjutnya yaitu:

1. Perlunya dilakukan penerapan algoritma optimal seperti (PSO) Particle Swam Optimise.
2. Perlunya dilakukan uji coba dalam menambahkan himpunan atribut, agar dapat lebih maksimal.
3. Lakukan perbandingan pengujian algoritma Klustering data mining.

DAFTAR PUSTAKA

Afrisawati. (2013). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Pelita Informatika Budi Darma, Vol 5 Nomor: 3*, 157.

- Fina Nasari, Surya Darma. (2015). Penerapan Algoritma K-Means Clustering pada Data Penerimaan Mahasiswa Baru. *Seminar Nasional Teknologi Informasi dan Multimedia*, 73-78.
- Istiqomah Sumadikarta, Evan Abeiza. (2014). penerapan algoritma K-means pada data mining untuk memilih produk dan pelanggan potensial (PT. Mega Arvia Utama). *Jurnal Satya Informatika*, Vol. 1 No. 1, 12-22.
- Nugroho, A. (2010). *Rekayasa Perangkat Lunak Berbasis Objek dengan Metode USDP*. Yogyakarta: Andi.
- Nugroho, B. (2008). *Aplikasi Pemrograman Web Dinamis dengan PHP dan MySQL*. Yogyakarta: Gava Media.
- Nugroho, Bunafit. (2013). *Dasar Pemrograman Web PHP - MySQL Dengan Dreamweaver*. jogjakarta: gava media.
- Prasetyo, E. (2012). *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Gresik: Andi.
- Pressman. (2012). *Rekayasa Perangkat Lunak*. Yogyakarta: Andi.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining =: Practical Machine Learning Tools and Teqniques Third*. Elsevier.
- Zulkarnain. (2012). *Ilmu Menjual*. Yogyakarta: Graha Ilmu.