

PENERAPAN DATA MINING UNTUK ESTIMASI NILAI MATEMATIKA DENGAN MENGGUNAKAN ALGORITMA REGRESI LINEAR PADA SMA KESATRIAN 1 SEMARANG

Cindy Eka Febriana¹, Usman Sudibyo²

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Jl. Nakula I No. 5-11, Semarang

E-mail: 111201206792@mhs.dinus.ac.id¹, usmansudibyo@yahoo.co.id²

Abstrak

Sekolah Menengah Atas (SMA), adalah jenjang pendidikan menengah di Indonesia setelah lulus Sekolah Menengah Pertama (SMP) atau sederajat. SMA ditempuh dalam kurun waktu 3 tahun, mulai dari kelas X sampai kelas XII. Dalam kurun waktu 3 tahun tersebut, setiap jurusan pada SMA terdapat materi pelajaran yang sama salah satunya yaitu pelajaran matematika. Matematika merupakan materi pelajaran yang dipadatkan atau ditambahkan karena disesuaikan dengan materi pelajaran Standar Internasional sehingga pendidikan di dalam negeri dan pendidikan di luar negeri dapat seimbang. Dengan memanfaatkan data mining, penulis melakukan penelitian tentang estimasi nilai mid dengan variabel bebas nilai UHT 1 dan UHT 2 dengan menggunakan algoritma regresi linear berganda. Yang bertujuan untuk mengetahui hubungan antara variabel nilai UHT 1 dan nilai UHT 2 terhadap nilai mid pada masing-masing jurusan (IPA, IPS, dan Bahasa) dengan menggunakan algoritma regresi linear berganda. Hasil dari penelitian ini adalah mendapatkan sebuah model regresi linear berganda untuk setiap jurusan. IPA: $Y = 11.8430 + 0.3692 X_1 + 0.3812 X_2$, IPS: $Y = 4.2233 + 0.4883 X_1 + 0.3801 X_2$, Bahasa: $Y = 7.8158 + 0.3377 X_1 + 0.4971 X_2$. Dengan masing-masing RMSE yang diperoleh 15.5483, 13.9542, dan 12.7714.

Kata Kunci: matematika, nilai, estimasi, algoritma regresi linear berganda, RMSE

Abstract

Senior High School (SMA) is an education level in Indonesia after graduating from Junior High School (SMP) or equivalent. The period of study in SMA is taken within 3 years, starting from class X to XII [1]. Within 3 years of study, each different majors in SMA have several same subjects mathematics. Mathematics is a subject matter that is compacted or added based on International Standard so that the education level can be balanced. By utilizing data mining, the authors conducted a research about mid value estimation by independent variable value of UHT 1 and UHT 2 using multiple linear regression algorithm. Which aims to determine the relationship between UHT 1 and UHT 2 variables towards mid value in each majors (IPA, IPS, and Bahasa) using multiple linear regression algorithm. The result of this study is to get the multiple linear regression model for each majors. IPA: $Y = 11.8430 + 0.3692 X_1 + 0.3812 X_2$, IPS: $Y = 4.2233 + 0.4883 X_1 + 0.3801 X_2$, Bahasa: $Y = 7.8158 + 0.3377 X_1 + 0.4971 X_2$. With each RMSE obtained 15.5483, 13.9542, and 12.7714.

Keywords: mathematics, value, estimation, multiple linear regression algorithm, RMSE

1. PENDAHULUAN

Sekolah Menengah Atas (SMA), adalah jenjang pendidikan menengah di Indonesia setelah lulus Sekolah Menengah Pertama (SMP) atau sederajat. SMA ditempuh dalam kurun waktu 3 tahun, mulai dari kelas X sampai kelas XII[1]. Dalam kurun waktu 3 tahun tersebut, setiap jurusan pada SMA terdapat materi pelajaran yang sama salah satunya yaitu pelajaran matematika. Matematika merupakan materi pelajaran yang dipadatkan atau ditambahkan karena disesuaikan dengan materi pelajaran Standar Internasional sehingga pendidikan di dalam negeri dan pendidikan di luar negeri dapat seimbang.

Dalam sebuah penelitian yang berhubungan dengan data, dibutuhkan sebuah metode yang dapat membantu dalam proses pelaksanaannya. Data mining merupakan disiplin ilmu yang mempelajari tentang metode untuk mengekstrak pengetahuan atau pola dari suatu data, sehingga data mining juga sering disebut *knowledge discovery in database*[2]. Data mining dapat digunakan untuk pengelompokan data, prediksi, estimasi, dan menentukan kaidah asosiasi dalam suatu data yang ada. Perlunya data mining karena adanya sejumlah data besar yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan untuk mengetahui suatu pola dalam suatu data yang banyak, terlebih lagi besarnya kebutuhan untuk mengubah data tersebut menjadi informasi yang berguna. Metode data mining bermacam-macam disesuaikan dengan kebutuhan yang ada[3].

Salah satu penelitian yang menggunakan data mining adalah estimasi dengan menggunakan model regresi linear. Model regresi adalah istilah lain untuk jenis model linear[2].

Analisis regresi adalah metodologi populer yang mengeksplorasi hubungan antara variabel respon dan satu atau lebih variabel penjelas. Metodologi ini telah banyak diterapkan di banyak bidang seperti ekonomi, teknik, biologi, bisnis, dan sebagainya[4]. Selain bidang tersebut, model regresi juga dapat digunakan pada bidang pendidikan, contohnya untuk menghitung estimasi dalam sebuah perhitungan khususnya untuk estimasi nilai siswa. Dalam sejumlah data hubungan sebenarnya jarang dapat diketahui akan tetapi hubungan tersebut dapat diestimasi berdasarkan data pengamatan. Metode yang populer digunakan adalah estimasi kuadrat terkecil. Hal ini didasarkan pada kenyataan bahwa ia meminimumkan jumlah kuadrat perbedaan nilai yang diharapkan dengan nilai observasi[5].

Dalam penelitian ini, penulis menggunakan pendekatan data mining dengan penerapan model regresi linear berganda akan dilakukan untuk estimasi nilai siswa untuk mengetahui pengaruh nilai ulangan harian terhadap nilai mid dengan menggunakan algoritma regresi linear. Penelitian ini menggunakan RMSE (*Root Mean Square Error*) untuk menghitung standar error hasil estimasi. Diharapkan dengan perhitungan estimasi tersebut dapat diperkirakan hasil nilai yang kemungkinan didapat pada saat ujian mid sehingga pihak SMA dapat meningkatkan kualitas belajar siswanya apabila nilai kurang memenuhi syarat ketuntasan. Berdasarkan penelitian-penelitian sebelumnya, metode atau model yang baik digunakan untuk menghitung estimasi adalah model regresi linear. Dengan menggunakan Matlab berguna untuk mengetahui hasil dari kinerja algoritma yang digunakan.

2. LANDASAN TEORI

2.1 Data

Data merupakan keterangan atau fakta yang dikumpulkan dari suatu populasi

yang akan digunakan untuk menerangkan ciri-ciri populasi yang bersangkutan. Data harus memenuhi kriteria sebagai berikut[14]:

- Objektif, ialah data yang sama dengan keadaan sebenarnya.
- Mewakili populasi
- Galat baku (standard error) kecil
- Tepat waktu
- Relevan

2.2 Data Mining

Data mining adalah proses menemukan pola yang berguna dan tren di set data yang besar[9]. Sejak tahun 1990-an, gagasan data mining, biasanya dilihat sebagai proses “pertambangan/mining” data, telah muncul di banyak lingkungan, dari bidang akademik, bisnis atau kegiatan medis, khususnya. Data mining masih diperdebatkan oleh beberapa bidang ilmiah, termasuk Daryl Pregibons yang menyatakan bahwa “data mining adalah campuran statistic, *Artificial Intelligence* (kecerdasan buatan), dan penelitian database”[10]. Terdapat dua jenis tujuan data mining, antara lain:

1. Prediktif (misalnya, klasifikasi, regresi, bias/deteksi anomaly, dll), menggunakan beberapa variabel yang ada untuk memprediksi nilai masa depan (belum diketahui) dari variabel lain.
2. Deskriptif (misalnya, *clustering*, *association rule*, pola sekuensial, dll), identifikasi pola yang menggambarkan data dan dapat dengan mudah dipahami oleh pengguna[10].

2.3 Model Regresi

Regresi linear adalah teknik alami untuk mempertimbangkan atau menyelesaikannya ketika hasil atau kelas adalah numerik dan semua atribut adalah numerik. Ini adalah metode pokok dalam statistik. Berikut adalah bobot yang telah ditentukan dalam

proses penyelesaian dengan menggunakan model regresi linear berganda[2]:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (1)$$

Dimana:

Y = kelas (variabel dependen/tidak bebas)

a = konstanta

b_1, b_2, \dots, b_k = bobot atau koefisien regresi

X_1, X_2, \dots, X_k = nilai atribut (variabel independen/bebas)

Model atau metode regresi linear merupakan metode yang cukup populer dan biasanya digunakan untuk menemukan persamaan dari sebuah data yang dimana data tersebut saling berhubungan antara variabel satu dengan variabel yang lain dalam satu database yang cukup besar. Informasi yang dihasilkan dari data mining dengan metode regresi linear bisa dijadikan sebagai dasar untuk pengambilan keputusan[6].

2.4 Root Mean Square Error (RMSE)

Dalam regresi linear berganda, nilai a, b_1, b_2, \dots, b_k dapat dihitung menggunakan Metode Kuadrat Terkecil (MKT), salah satunya yaitu *Root Mean Square Error (RMSE)*. RMSE adalah ukuran yang sering digunakan dari perbedaan antara nilai-nilai prediksi oleh model atau estimator. RMSE berfungsi untuk menghitung dan menjadi ukuran besaran kesalahan dalam prediksi[6].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Keterangan:

Y_i = data awal (data sebenarnya)

\hat{Y}_i = data akhir (data hasil estimasi)

n = jumlah data

Keakuratan pengukuran estimasi ditunjukkan dengan adanya hasil RMSE yang kecil (mendekati nol). RMSE yang lebih kecil dikatakan lebih akurat dibandingkan RMSE yang lebih besar[15]. Standar error yang baik adalah kisaran 0.0 – 1.0, lebih dari 1.0 sudah tergolong besar[16].

2.5 Estimasi

Estimasi adalah suatu metode dimana kita dapat memperkirakan nilai populasi (suatu objek yang diteliti) dengan memakai nilai sampel (contoh objek yang diambil untuk dijadikan penelitian)[13].

Salah satu contoh tugas estimasi dalam penelitian adalah memperkirakan nilai rata-rata (IPK) dari seorang mahasiswa pascasarjana berdasarkan IPK mahasiswa sarjana[9].

3. HASIL DAN PEMBAHASAN

Data nilai matematika yang digunakan sebagai penelitian berjumlah 974 data dari semua jurusan, yang terdiri dari jurusan IPA (541 data), IPS (358 data), dan Bahasa (75 data).

Langkah-langkah perhitungan yang dilakukan dalam penelitian ini, sebagai berikut:

1. Data yang akan diolah, diubah menurut variabelnya yaitu nilai UHT 1 menjadi X_1 , nilai UHT 2 menjadi X_2 , dan nilai mid menjadi Y .
2. Menghitung jumlah dari perhitungan X_1 , X_2 , Y , X_1Y , X_2Y , X_1X_2 , X_1^2 , X_2^2 , dan Y^2 .
3. Didapatkan hasil pada nomor 2, kemudian dilakukan perhitungan regresi linear berganda dengan perhitungan sebagai berikut:

$$\sum x_1^2 = \sum X_1^2 - ((\sum X_1)^2 / n) = \dots \quad (1)$$

$$\sum x_2^2 = \sum X_2^2 - ((\sum X_2)^2 / n) = \dots \quad (2)$$

$$\sum y^2 = \sum Y^2 - ((\sum Y)^2 / n) = \dots \quad (3)$$

$$\sum x_1 y = \sum X_1 Y - ((\sum X_1 \cdot \sum Y) / n) = \dots \quad (4)$$

$$\sum x_2 y = \sum X_2 Y - ((\sum X_2 \cdot \sum Y) / n) = \dots \quad (5)$$

$$\sum x_1 x_2 = \sum X_1 X_2 - ((\sum X_1 \cdot \sum X_2) / n) = \dots \quad (6)$$

4. Setelah mendapatkan hasil dari perhitungan menggunakan persamaan di atas, kemudian melakukan perhitungan untuk mencari nilai a , b_1 , dan b_2 dengan perhitungan berikut:

$$b_1 = \frac{[(\sum x_2^2 \cdot \sum x_1 y) - (\sum x_2 y \cdot \sum x_1 x_2)]}{[(\sum x_1^2 \cdot \sum x_2^2) - (\sum x_1 x_2)^2]} \quad (1)$$

$$b_2 = \frac{[(\sum x_1^2 \cdot \sum x_2 y) - (\sum x_1 y \cdot \sum x_1 x_2)]}{[(\sum x_1^2 \cdot \sum x_2^2) - (\sum x_1 x_2)^2]} \quad (2)$$

$$a = \frac{(\sum Y) - (b_1 \cdot \sum X_1) - (b_2 \cdot \sum X_2)}{n} \quad (3)$$

5. Dari perhitungan nomer 4, didapatkan persamaan regresi linear $Y = a + b_1 X_1 + b_2 X_2$
6. Dengan menggunakan persamaan regresi linear yang telah didapat, maka dilakukan perhitungan estimasi yang terdiri dari a , b_1 , b_2 , $b_1 X_1$, $b_2 X_2$, dan hasil estimasi.
7. Hasil estimasi diperoleh, kemudian melakukan perhitungan RMSE untuk mengetahui berapa besar kesalahan (*error*) yang didapat dengan menggunakan algoritma regresi linear berganda yang terdiri dari tabel Y_i , \hat{Y}_i , $Y_i - \hat{Y}_i$, dan $(Y_i - \hat{Y}_i)^2$.

Keterangan:

Y_i = data awal (data sebenarnya)

\hat{Y}_i = data akhir (data hasil estimasi)

n = jumlah data

```

1 load ipa.txt
2 X1 = ipa(:,1);
3 X2 = ipa(:,2);
4 Y = ipa(:,3);
5 n = length(X1);
6 EK1 = sum(X1);
7 EK2 = sum(X2);
8 EY = sum(Y);
9 X1Y = dot(X1,Y);
10 X2Y = dot(X2,Y);
11 X1X2 = dot(X1,X2);
12 EK1Y = sum(X1Y);
13 EK2Y = sum(X2Y);
14 EK1X2 = sum(X1X2);
15 X1_2 = X1.^2;
16 X2_2 = X2.^2;
17 Y_2 = Y.^2;
18 EK1_2 = sum(X1_2);
19 EK2_2 = sum(X2_2);
20 EY_2 = sum(Y_2);
21 Ek1_2 = EK1_2 - ((EK1.^2)/n);
22 Ek2_2 = EK2_2 - ((EK2.^2)/n);
23 Ey_2 = EY_2 - ((EY.^2)/n);
24 Ek1y = EK1Y - ((dot(EK1,EY))/n);
25 Ek2y = EK2Y - ((dot(EK2,EY))/n);
26 Ek1x2 = EK1X2 - ((dot(EK1,EK2))/n);
27 b1 = (dot(Ek1_2,Ek1y) - dot(Ek2y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
28 b2 = (dot(Ek2_2,Ek2y) - dot(Ek1y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
29 a = ((EY) - dot(b1,EK1) - dot(b2,EK2)) / n;
30 UHT1 = input('Masukkan Nilai UHT 1 = ');
31 UHT2 = input('Masukkan Nilai UHT 2 = ');
32 H = a + b1*UHT1 + b2*UHT2;

```

Gambar 1. Implementasi algoritma pada jurusan IPA

```

1 load ipa.txt
2 X1 = ipa(:,1);
3 X2 = ipa(:,2);
4 Y = ipa(:,3);
5 n = length(X1);
6 EK1 = sum(X1);
7 EK2 = sum(X2);
8 EY = sum(Y);
9 X1Y = dot(X1,Y);
10 X2Y = dot(X2,Y);
11 X1X2 = dot(X1,X2);
12 EK1Y = sum(X1Y);
13 EK2Y = sum(X2Y);
14 EK1X2 = sum(X1X2);
15 X1_2 = X1.^2;
16 X2_2 = X2.^2;
17 Y_2 = Y.^2;
18 EK1_2 = sum(X1_2);
19 EK2_2 = sum(X2_2);
20 EY_2 = sum(Y_2);
21 Ek1_2 = EK1_2 - ((EK1.^2)/n);
22 Ek2_2 = EK2_2 - ((EK2.^2)/n);
23 Ey_2 = EY_2 - ((EY.^2)/n);
24 Ek1y = EK1Y - ((dot(EK1,EY))/n);
25 Ek2y = EK2Y - ((dot(EK2,EY))/n);
26 Ek1x2 = EK1X2 - ((dot(EK1,EK2))/n);
27 b1 = (dot(Ek1_2,Ek1y) - dot(Ek2y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
28 b2 = (dot(Ek2_2,Ek2y) - dot(Ek1y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
29 a = ((EY) - dot(b1,EK1) - dot(b2,EK2)) / n;
30 UHT1 = input('Masukkan Nilai UHT 1 = ');
31 UHT2 = input('Masukkan Nilai UHT 2 = ');
32 H = a + b1*UHT1 + b2*UHT2;

```

Gambar 2. Implementasi algoritma pada jurusan IPS

```

1 load bahasa.txt
2 X1 = bahasa(:,1);
3 X2 = bahasa(:,2);
4 Y = bahasa(:,3);
5 n = length(X1);
6 EK1 = sum(X1);
7 EK2 = sum(X2);
8 EY = sum(Y);
9 X1Y = dot(X1,Y);
10 X2Y = dot(X2,Y);
11 X1X2 = dot(X1,X2);
12 EK1Y = sum(X1Y);
13 EK2Y = sum(X2Y);
14 EK1X2 = sum(X1X2);
15 X1_2 = X1.^2;
16 X2_2 = X2.^2;
17 Y_2 = Y.^2;
18 EK1_2 = sum(X1_2);
19 EK2_2 = sum(X2_2);
20 EY_2 = sum(Y_2);
21 Ek1_2 = EK1_2 - ((EK1.^2)/n);
22 Ek2_2 = EK2_2 - ((EK2.^2)/n);
23 Ey_2 = EY_2 - ((EY.^2)/n);
24 Ek1y = EK1Y - ((dot(EK1,EY))/n);
25 Ek2y = EK2Y - ((dot(EK2,EY))/n);
26 Ek1x2 = EK1X2 - ((dot(EK1,EK2))/n);
27 b1 = (dot(Ek1_2,Ek1y) - dot(Ek2y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
28 b2 = (dot(Ek2_2,Ek2y) - dot(Ek1y,Ek1x2)) / (dot(Ek1_2,Ek2_2) - (Ek1x2.^2));
29 a = ((EY) - dot(b1,EK1) - dot(b2,EK2)) / n;
30 UHT1 = input('Masukkan Nilai UHT 1 = ');
31 UHT2 = input('Masukkan Nilai UHT 2 = ');
32 H = a + b1*UHT1 + b2*UHT2;

```

Gambar 3. Implementasi algoritma jurusan Bahasa

```

function rmse (data,estimate)
2 sqrt(sum((data(:) - estimate(:)).^2) / numel(data))

```

Gambar 4. Perhitungan RMSE

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Kesimpulan dari hasil penelitian yang telah dilakukan oleh peneliti adalah sebagai berikut:

1. Didapatkan model regresi linear berganda untuk masing-masing jurusan, yaitu:

$$\text{IPA: } Y = 11.8430 + 0.3692 X_1 + 0.3812 X_2$$

$$\text{IPS: } Y = 4.2233 + 0.4883 X_1 + 0.3801 X_2$$

$$\text{Bhs: } Y = 7.8158 + 0.3377 X_1 + 0.4971 X_2$$

2. *Root Means Square Error* yang didapatkan tergolong besar, yaitu:

$$\text{IPA} = 15.5483$$

$$\text{IPS} = 13.9542$$

$$\text{Bahasa} = 12.7714$$

3. Berdasarkan model regresi linear yang telah didapat, maka dapat disimpulkan bahwa tidak ada hubungan antara nilai UHT 1 dan UHT 2 terhadap nilai mid. Karena dengan menggunakan model regresi linear berganda tersebut dapat diketahui bahwa nilai error yang didapatkan tergolong besar melebihi kisaran 0.0 – 1.0.

4.2 Saran

Agar penelitian ini terus berkembang, peneliti memberikan saran untuk penelitian selanjutnya, yaitu:

1. Penelitian ini dikembangkan dengan menambah jumlah atribut, tidak hanya nilai UHT 1, nilai UHT 2, dan nilai mid.
2. Periode pengambilan data dapat diambil dari beberapa semester.
3. Dikembangkan dengan metode data mining estimasi menggunakan algoritma estimasi yang lain.
4. Menggunakan model regresi lain yang bukan linear.

DAFTAR PUSTAKA

[1] Holisatul Munawaroh, Bain Khusnul, and Yeni Kustiyahningsih, "Perbandingan

- Algoritma ID3 dan C5.0 Dalam Identifikasi Jurusan Siswa SMA," *Jurnal Sarjana Teknik Informatika*, vol.1, hlm 1-12, 2013.
- [2] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques Third Edition*. Elsevier. 2011.
- [3] Obbie Kristanto, "Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Jurusan Siswa SMAN 6 Semarang," 2014.
- [4] Silian Shen and Jianling Cui, "Estimation and Inference of the Fuzzy Linear Regression Model with L Fuzzy Observations," *Fifth International Joint Conference on Computational Science and Optimization*, 2012.
- [5] Yuliana, "Penerapan Model Regresi Linear Robust Dengan Estimasi M Pada Data Nilai Kalkulus II Mahasiswa Universitas Widya Dharma Klaten," 2014.
- [6] Ali Fikri, "Penerapan Data Mining Untuk Mengetahui Tingkat Keakuratan Beton Yang Akan Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression," 2013.
- [7] Eggy Inaidi Andana Warih, "Penerapan Data Mining Untuk Menentukan Estimasi Produktivitas Tanaman Tebu Dengan Menggunakan Algoritma Linear Regresi Berganda di Kabupaten Rembang," 2015.
- [8] Yuniarsi Rahayu, "Penerapan Metode Numerik Pada Peramalan Untuk Menghitung Koefisien-koefisien Pada Garis Regresi Linear Berganda," *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*, 2011.
- [9] Daniel T. Larose and Chantal D. Larose. *Discovering Knowledge In Data: An Introduction to Data Mining Second Edition*. Wiley. 2014.
- [10] Florin Gorunescu. *Data Mining: Concepts, Models and Techniques*. Verlag Berlin Heidelberg: Springer. 2011.
- [11] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques Third Edition*. Elsevier. 2012.
- [12] Kusriani and E.T. Luthfi. *Algoritma Data Mining*. Yogyakarta. Andi Offset. 2009.
- [13] E. Prasetyo. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta. Andi Offset. 2014.
- [14] R. Lungan. *Aplikasi Statistika dan Hitung Peluang*. Yogyakarta. Graha Ilmu. 2006.
- [15] Catharina Sri Wahyu Widayati, "Komparasi Beberapa Metode Estimasi Kesalahan Pengukuran," *Jurnal Penelitian dan Evaluasi Pendidikan*. 2009.
- [16] Vijay Kotu and Bala Deshpande, Ph.D. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Elsevier. 2015.
- [17] M. Gugun, "Modul Materi Perkuliahan Ilmu Komputer," Available: <https://gugunawan.wordpress.com/category/matematika-dan-algoritma/page/3/>. [Accessed 3 Januari 2016].