

PREDIKSI NILAI SISWA DENGAN METODE LEAST SQUARE DAN ANALISIS OUTLIER DENGAN TEKNIK EXACT EXCEPTION

Jessica Belinda
Wimmie Handiwidjojo

Abstrak

Dalam institusi pendidikan, analisis terhadap nilai anak didik dapat membantu guru sebagai bahan pertimbangan dalam membimbing anak didik. Kualitas dari anak didik akan berpengaruh terhadap kualitas institusi pendidikan. Analisis dan pengolahan data nilai jika dilakukan dengan cara manual akan memakan waktu lama. Selain itu, pengolahan data secara manual masih memungkinkan adanya kesalahan.

Berdasarkan permasalahan tersebut, akan dilakukan prediksi nilai dengan analisis outlier untuk menganalisis nilai anak didik. Sistem yang dibangun akan memprediksi nilai siswa. Pengolahan data nilai memanfaatkan database nilai siswa yang kemudian dikenakan analisis outlier. Perhitungan prediksi nilai menggunakan metode Least Square, sementara analisis outlier menggunakan metode Deviasi dan teknik Exact Exception untuk pencarian outlier.

Hasil dari sistem berupa prediksi nilai siswa yang dapat digunakan oleh institusi pendidikan untuk melihat perkembangan belajar anak didik.

Kata Kunci : prediksi, outlier, least square, exact exception.

1. Pendahuluan

1.1. Latar Belakang

Setiap institusi pendidikan perlu melakukan analisis terhadap perkembangan nilai siswa. Hasil analisis tersebut dapat digunakan sebagai bahan pertimbangan dalam membimbing anak didik. Secara tidak langsung, kualitas anak didik yang ada akan mempengaruhi kualitas pendidikan di institusi tersebut. Semakin baik kualitas pendidikan suatu institusi akademik, semakin banyak pula anak-anak didik baru yang berminat untuk mendaftar dan belajar di institusi tersebut.

Dewasa ini, masih banyak institusi pendidikan yang menggunakan cara manual untuk menganalisis nilai anak-anak didik mereka. Analisis manual memerlukan waktu yang lama karena jumlah data yang diperlukan untuk proses analisis cukup besar. Selain itu, proses analisis manual masih memungkinkan adanya kesalahan manusia, baik dalam proses memasukkan data maupun dalam proses analisis.

Berdasarkan permasalahan tersebut, diperlukan suatu sistem untuk membantu proses analisis sehingga proses pengambilan kesimpulan dapat dilakukan dengan tepat. Sistem memanfaatkan data-data nilai untuk diolah menggunakan metode *Least Square* sehingga menghasilkan prediksi nilai akademik siswa. Selain itu, data nilai juga dapat dianalisis dengan pendekatan metode Deviasi dengan teknik *Exact Exception* yang berfungsi untuk mengelompokkan data berdasarkan tingkat kemiripan. Data yang memiliki tingkat kemiripan yang tinggi akan dikelompokkan menjadi satu kelompok data dan data dengan tingkat kemiripan rendah disebut *Outliers*. Dengan metode ini, diharapkan pembuatan prediksi dan analisis pengelompokan data dapat dilakukan dengan baik sehingga dapat mendukung dalam proses pengambilan keputusan.

1.2. Rumusan Masalah

Rumusan masalah yang akan diselesaikan dalam penelitian adalah :

- a. Apakah metode *Least Square* dapat diterapkan untuk membuat prediksi nilai siswa?
- b. Apakah teknik *Exact Exception* dapat diterapkan untuk mengelompokkan data nilai akademik siswa menjadi data *outlier* dan bukan *outlier*?

2. Teori Pendukung

2.1. Web Design

Desain web adalah desain grafis yang bertujuan untuk menyediakan fitur antarmuka yang memiliki nilai estetika kepada konsumen. Menurut Church (2007 : 87), ada beberapa hal yang perlu diperhatikan dalam desain web.

- 1) **Penampilan**, penampilan web harus rapi dan memperhatikan penggunaan ruang kosong secara strategis.
- 2) **Warna**, warna yang dapat digunakan untuk latar adalah warna cerah, sedangkan teks menggunakan warna yang lebih gelap. Penggunaan warna juga perlu sewajarnya saja supaya tidak terjadi *color pollution*.
- 3) **Layout**, *layout* disamping sederhana juga harus konsisten di setiap halaman untuk memudahkan pengguna dalam menggunakan web.
- 4) **Navigasi**, perlu ada baris petunjuk navigasi di setiap halaman untuk memberikan informasi bagi pengguna. Navigasi yang baik adalah navigasi yang memudahkan pengguna untuk kembali ke halaman awal melalui halaman manapun dalam web.
- 5) **Isi**, kalimat-kalimat yang pendek dan jelas justru lebih mudah dipahami. Jenis huruf, warna huruf, ukuran dan pengaturan huruf harus disesuaikan dengan informasi yang ingin disampaikan pada pengguna.
- 6) **Tautan**, harus merujuk ke informasi yang dibutuhkan.
- 7) **Grafik**, penggunaan grafis harus sesuai dengan isi yang ada dalam web.
- 8) **Informasi Dasar**, informasi harus jelas supaya tidak membingungkan pengguna, seperti misalnya: pemberian judul harus jelas pada setiap halaman dalam web, serta informasi lain yang dapat digunakan pengguna ketika pengguna mengakses web.
- 9) **Tata Bahasa dan Ejaan**, penggunaan tata bahasa dan ejaan harus disesuaikan dengan konteks web serta menggunakan gramatika yang benar.

Desain antarmuka terdiri dari beberapa bagian, yaitu *header*, menu navigasi, isi dan *footer*. Website yang dibangun ini akan menggunakan desain *header*, menu navigasi dan *footer* yang sama untuk semua halamannya. Bagian yang berbeda hanya bagian isi halaman. Hal ini ditujukan untuk menjaga konsistensi sebuah web sehingga memudahkan pengguna dalam menggunakan sistem.

Web yang baik adalah web yang mudah dipelajari dan mudah digunakan oleh pengguna. Ketika sistem aplikasi berbasis web dibangun, *developer* harus memperhatikan desain web tersebut sehingga web yang dibangun dapat digunakan oleh pengguna dengan baik. Sebaiknya web menggunakan istilah-istilah yang mudah dan sudah dimengerti oleh pengguna. Selain itu, web juga harus informatif terhadap kesalahan-kesalahan yang dilakukan pengguna. Hal tersebut bertujuan agar pengguna tidak melakukan kesalahan dalam penggunaan web.

2.2. Outlier Analysis

Dalam penerapan *Data Mining* di dunia nyata, kemungkinan suatu data memiliki berbagai macam nilai atribut, di mana nilai atribut tersebut ada yang memiliki tingkat kemiripan yang tinggi dan sebaliknya. Untuk data-data yang memiliki tingkat kemiripan data yang tinggi dapat

dikelompokkan menjadi satu kumpulan data. Di sisi lain, data-data yang memiliki tingkat kemiripan rendah dengan data disekitarnya disebut sebagai data *outlier* (Kantardzic, 2003 : 7).

Salah satu metode untuk menangani deteksi outlier, yaitu: deteksi *Outlier* berbasis Deviasi. Metode ini mengidentifikasi *outlier* dengan menguji karakteristik utama dari setiap objek dalam kelompok. Algoritma yang termasuk dalam metode ini antara lain : teknik *sequential exception*, teknik *exact exception* dan teknik *OLAP data cube* (Jiawei, 2006 : 458).

2.2.1. Teknik *Exact Exception*

Jiawei (2006 : 458) menuliskan bahwa ada beberapa hal yang perlu diperhatikan dalam proses teknik *Exact Exception*, antara lain :

1. Exception Set

Exception Set didefinisikan sebagai subset kecil dari kumpulan data set.

2. Dissimilarity Function

Untuk setiap fungsi *Dissimilarity Function*, dapat mengembalikan nilai yang rendah jika setiap elemen dari I sama satu dengan yang lainnya, dan nilai yang lebih tinggi jika setiap elemen tidak sama. Fungsi ini tidak memerlukan jarak metris antar data. Diberikan sebuah subset yang terdiri dari n anggota $\{x_1, \dots, x_n\}$, maka fungsi pertidaksamaan yang mungkin terjadi adalah variansi dari anggota dalam kelompok tersebut, yaitu :

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad [1]$$

3. Cardinality Function

Merupakan banyaknya objek yang diberikan dalam satu set data. Fungsi $C(I - I_j)$ merupakan jumlah data dalam kelompok data $(I - I_j)$.

4. Smoothing Factor

Faktor penghalusan menunjukkan seberapa besar nilai pertidaksamaan bisa dikurangi dengan cara menghilangkan subset (I_j) untuk setiap elemen dari kelompok data set (I) .

Untuk mengurangi pengaruh yang mungkin terjadi pada hasil dari inputan, proses tersebut dapat dilakukan secara berulang selama beberapa waktu. Subset dengan nilai faktor penghalusan terbesar dari semua iterasi akan menjadi kelompok pengecualian.

2.2.2. Algoritma *Exact Exception*

Menurut Jiawei (2006 : 458 - 459), berikut adalah algoritma dari teknik *Exact Exception* :

- 1) Inputan berupa suatu set data I.
- 2) Pilih data subset I_j ($j = 1, 2, 3, \dots, n$) dari data set I.
- 3) Proses perhitungan :
 - a. Hitung *Dissimilarity Function* untuk data set $I \rightarrow D(I)$
 - b. Cari selisih (data komplemen) dari I dan $I_j \rightarrow (I - I_j)$
 - c. Hitung jumlah data $(I - I_j) \rightarrow C(I - I_j)$
 - d. Hitung *Dissimilarity Function* untuk data $(I - I_j) \rightarrow D(I - I_j)$
 - e. Hitung *Smoothing Factor* data $I_j \rightarrow (SFI_j)$
 - f. Subset dengan nilai *Smoothing Factor* tertinggi akan menjadi outlier.

2.3. Prediksi Nilai

Prediksi numerik adalah proses memperkirakan kelanjutan nilai dari data masukan (Jiawei, 2006: 354). Dalam kehidupan, prediksi numerik sering kali diterapkan pada kasus-kasus

sederhana. Sebagai contoh, membuat prediksi tingkat penjualan produk baru berdasarkan harga, atau membuat prediksi gaji karyawan berdasarkan lama kerja. Pembuatan prediksi nilai dimanfaatkan untuk pendukung pengambilan keputusan.

Metode Kuadrat Terkecil atau *Least Square Method* digunakan untuk menentukan hubungan linear yang terbaik dengan cara meminimalkan jumlah eror kuadrat data yang dihasilkan oleh fungsi matematika (Black, 2004 : 484). Rumus persamaan garis yang digunakan adalah :

$$Y' = a + bt \quad [2]$$

Dengan :

Y' = variabel nilai tren

a = nilai konstanta saat t sama dengan 0

b = kemiringan garis

t = variabel waktu

Untuk mencari nilai kemiringan garis (b) digunakan rumus :

$$b = \frac{n \sum tY - \sum Y \sum t}{n \sum t^2 - \sum t^2} \quad [3]$$

Untuk mencari nilai konstanta (a) digunakan rumus :

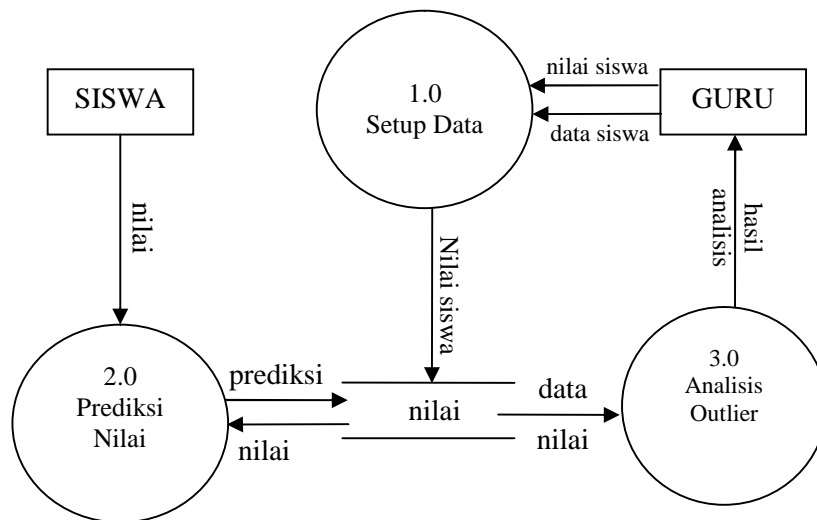
$$a = \frac{\sum Y}{n} - b \frac{\sum t}{n} \quad [4]$$

Langkah selanjutnya adalah menerapkan nilai tersebut pada persamaan garis sehingga dapat diketahui nilai prediksi di periode berikutnya.

3. Rancangan Sistem dan Database

3.1. Rancangan Sistem

DFD adalah suatu diagram yang menggunakan notasi-notasi untuk menggambarkan arus data dari sistem. Perancangan dengan menggunakan DFD akan mempermudah penggambaran aliran data mulai dari *input*, proses yang harus dialui, sampai dengan *output* yang dihasilkan dari sistem. Dalam DFD juga menggambarkan aliran data ke *storage* atau tabel-tabel yang terlibat pada sistem.

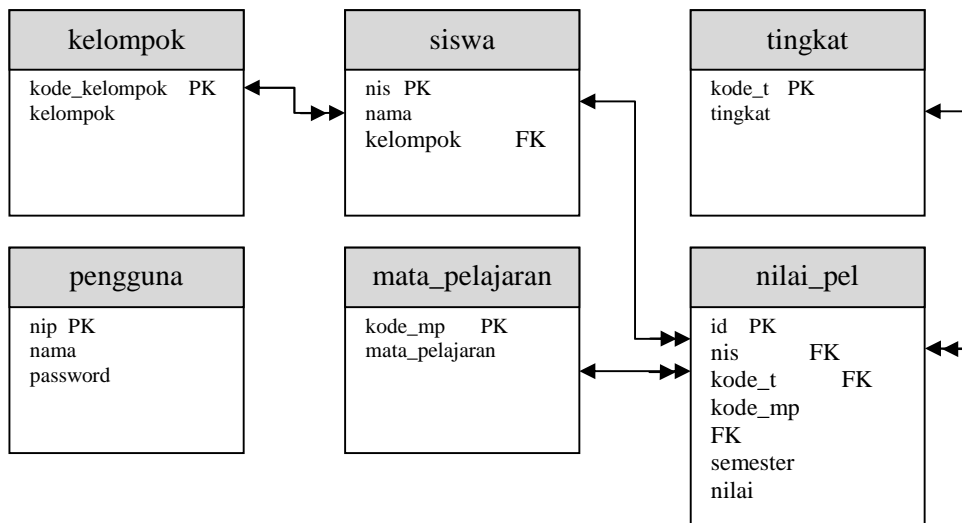


Gambar 1. Diagram Konteks

Dalam diagram konteks di atas menggambarkan bahwa data yang diolah oleh sistem berasal dari data siswa, baik data nilai maupun data pribadi siswa. Sistem akan menghasilkan hasil analisis yang selanjutnya dapat digunakan guru untuk pendukung pengambilan keputusan.

3.2. Rancangan Database

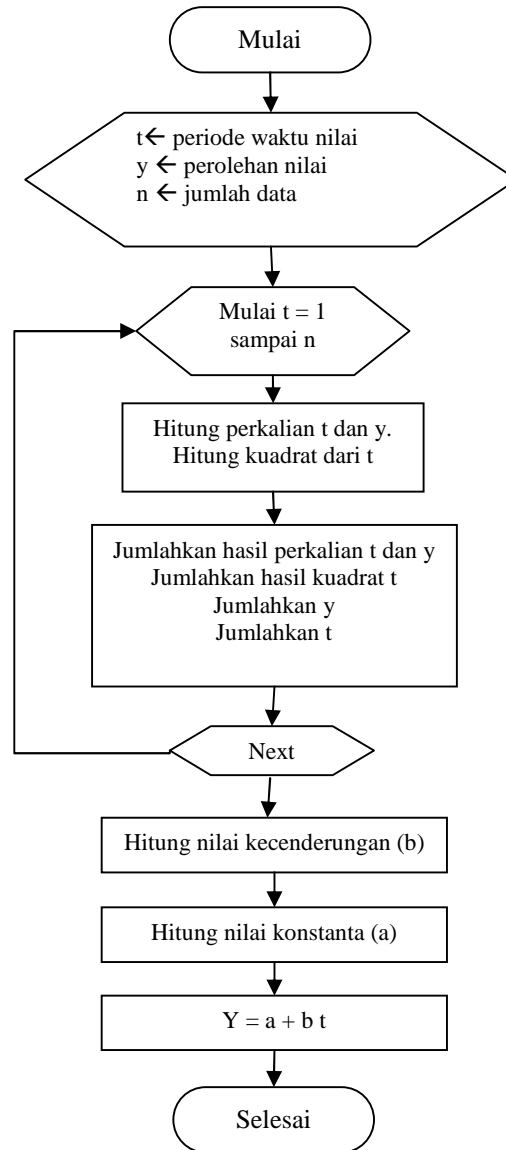
Berikut ini merupakan rancangan database yang akan digunakan pada sistem. Pada Gambar 2 dapat dilihat beberapa tabel yang terlibat. Hubungan antar tabel dibentuk untuk menjaga agar data tetap valid.



Gambar 2. Rancangan Basis Data

3.3. Rancangan Proses

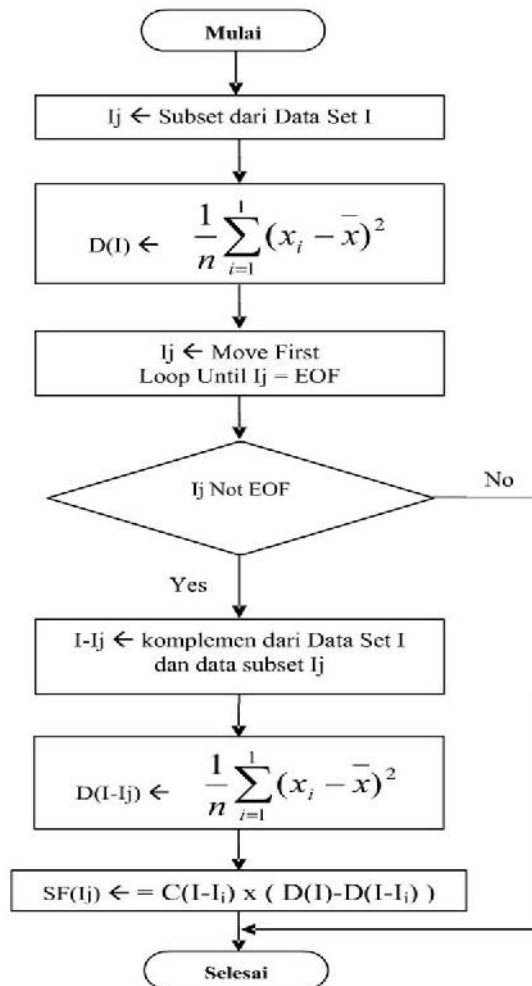
Flowchart menampilkan langkah-langkah penyelesaian masalah yang ada dalam proses. Berikut adalah beberapa *flowchart* yang menggambarkan proses yang ada dalam sistem.



Gambar 3. Flowchart Metode Least Square

Proses perhitungan prediksi nilai menggunakan metode *Least Square* dijabarkan dalam gambar 3. Data yang digunakan dalam perhitungan diambil dari tabel nilai siswa. Pengolahan data dilakukan per siswa dalam kelompok. Dalam setiap kelompok terdapat 5 orang siswa. Masing-masing siswa memiliki 10 nilai untuk tiap mata pelajaran. Tingkatan kelas dijadikan atribut waktu dengan menggunakan angka yang meningkat. Kelas 4 semester 1 diberi atribut 1, kelas 4 semester 2 diberi atribut 2, hingga kelas 8 semester 2 diberi atribut 10. Nilai

siswa pada setiap tingkatan akan dihitung sesuai dengan atribut waktu hingga diperoleh nilai prediksi untuk kelas 9.



Gambar 4. Flowchart Teknik Exact Exception

4. Implementasi dan Analisis Sistem

4.1. Implementasi Sistem

Langkah awal dalam prediksi nilai adalah pengambilan semua data nilai siswa dari tabel nilai_pel. Proses prediksi nilai dilakukan per mata pelajaran. Data nilai setiap siswa kemudian diurutkan berdasarkan tingkat dan semester. Setiap nilai akan diberi bobot dari 1 hingga 10. Selanjutnya, dibuat asumsi bahwa data nilai kelas 9 berbobot 11. Kemudian dilakukan proses perhitungan dengan metode *Least Square*.

Hasil perhitungan berupa prediksi nilai siswa untuk setiap mata pelajaran. Data nilai prediksi akan disimpan di tabel nilai_pel dengan kode tingkat 6 yang berarti kelas 9 dan semester 0. Proses akan dilakukan terus-menerus hingga seluruh siswa dalam kelompok yang dipilih memiliki prediksi nilai.

Daftar Nilai Prediksi Siswa Kelompok C Kelas 9 Semester 1

No	NIS	Nama Siswa	Matematika	Bahasa Indonesia	Bahasa Inggris	Ilmu Pengetahuan Alam
1	19136	Adiga Dwi	85.2	90.5	79.0	87.3
2	19140	Ar Yulian	67	73.3	81.8	86.5
3	19141	Lisa Nuriani	83.5	80.0	86.4	75.5
4	19142	Gloria Fortuna	82.7	71.0	79.5	82.8
5	19143	Ira Tania	80.0	74.0	80.9	85.1

Gambar 5. Hasil Prediksi

Setelah proses prediksi selesai, pengguna dapat melihat hasil prediksi tiap kelompok seperti pada Gambar 5. Untuk dapat diolah dalam proses prediksi, semua siswa dalam kelompok harus memiliki nilai yang lengkap. Jika siswa dalam kelompok belum memiliki nilai lengkap, kelompok tersebut tidak akan muncul dalam pilihan kelompok yang dapat diprediksi.

Proses lain yang dapat diolah dalam sistem adalah analisis *outlier*. Sama halnya dengan proses prediksi nilai, proses analisis *outlier* juga hanya dapat mengolah data kelompok dengan siswa yang telah memiliki nilai yang lengkap. Namun proses analisis *outlier* mengolah data kelompok berdasarkan mata pelajaran yang dipilih. Hasil dari proses ini adalah pengelompokan data nilai tiap kelompok menjadi data *outlier* dan bukan *outlier*.

Langkah awal yang dilakukan dalam analisis *outlier* adalah membentuk subset data. Total subset data untuk setiap kelompok adalah 30 buah. Bila jumlah anggota kelompok bertambah menjadi 6 anak, maka jumlah subset yang harus dihitung dalam setiap kelompok bertambah menjadi 62. Hal ini menjadi salah satu alasan jumlah kelompok terbatas pada 5 orang. Kemudian sistem akan menghitung *dissimilarity function* dari data set (D(I)). Selanjutnya, sistem akan mencari komplemen dari data set dan data subset. Untuk setiap komplemen dilakukan perhitungan *dissimilarity function* (D(I-Ij)) dan *smoothing factor* (SF(Ij)). Data subset dengan nilai *smoothing factor* terbesar merupakan *outlier* dalam kelompok yang dipilih.

Tabel Nilai Siswa

No	Nis	Nama	Nilai	Keterangan
1	19139	Aditya Dwi	89.8	Bukan Outlier
2	19140	Ar Yulian	65.5	Outlier
3	19141	Lisa Nuriani	85.2	Bukan Outlier
4	19142	Gloria Fortuna	93.0	Bukan Outlier
5	19143	Ira Tania	85.1	Bukan Outlier

Gambar 6. Hasil Analisis Outlier

Pada Gambar 6, pengguna dapat melihat hasil pengelompokan *outlier* yang dihasilkan dalam proses perhitungan. Data nilai yang merupakan *outlier* dalam kelompok akan diberikan keterangan **Outlier**.

4.2. Analisis Sistem

Hasil analisis tersebut tidak sepenuhnya menunjukkan bahwa siswa dengan nilai yang merupakan data *outlier* adalah siswa yang memerlukan bimbingan khusus. Ada beberapa kasus data yang menunjuk siswa dengan nilai terbesar dalam kelompok tersebut sebagai *outlier*. Oleh karena itu, penulis melakukan analisis pada beberapa mata pelajaran berbeda dalam kelompok yang sama. Berikut adalah hasil analisis yang dihasilkan dengan urutan mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris dan Ilmu Pengetahuan Alam.

No	NIS	Nama	Nilai	Keterangan
1	19139	Aditya Dwi	89,8	Bukan Outlier
2	19140	Ar Yulian	69,5	Outlier
3	19141	Lisa Noviani	85,2	Bukan Outlier
4	19142	Gloria Fortuna	93,6	Bukan Outlier
5	19143	Ira Tania	85,1	Bukan Outlier

No	NIS	Nama	Nilai	Keterangan
1	19139	Aditya Dwi	82	Outlier
2	19140	Ar Yulian	77,5	Bukan Outlier
3	19141	Lisa Noviani	88,1	Outlier
4	19142	Gloria Fortuna	90,3	Bukan Outlier
5	19143	Ira Tania	79,2	Bukan Outlier

No	NIS	Nama	Nilai	Keterangan
1	19139	Aditya Dwi	77,1	Bukan Outlier
2	19140	Ar Yulian	83,6	Bukan Outlier
3	19141	Lisa Noviani	96,3	Outlier
4	19142	Gloria Fortuna	75,7	Bukan Outlier
5	19143	Ira Tania	70,8	Bukan Outlier

No	NIS	Nama	Nilai	Keterangan
1	19139	Aditya Dwi	76,7	Bukan Outlier
2	19140	Ar Yulian	91,2	Outlier
3	19141	Lisa Noviani	78,3	Bukan Outlier
4	19142	Gloria Fortuna	81,9	Outlier
5	19143	Ira Tania	56,2	Bukan Outlier

Gambar 7. Perbandingan Hasil Analisis Outlier

Pada Gambar 7, terlihat beberapa hasil perhitungan analisis *outlier*. Untuk mata pelajaran matematika, siswa yang merupakan *outlier* kelompok adalah Ari Yulian dengan nilai 65,5. Untuk mata pelajaran Bahasa Indonesia, siswa yang merupakan *outlier* kelompok adalah Aditya Dwi dengan nilai 92 dan Liza Noviani dengan nilai 88,1. Untuk mata pelajaran Bahasa Inggris, siswa yang merupakan *outlier* adalah Liza Noviani dengan nilai 96,3. Sedangkan untuk mata pelajaran Ilmu Pengetahuan Alam, siswa yang merupakan *outlier* adalah Ari Yulian dengan nilai 91,2 dan Gloria Fortuna dengan nilai 91,9.

Dari hasil perbandingan, terlihat bahwa ada 2 macam data *outlier*, yaitu *outlier* yang berada di luar batas atas dan *outlier* yang berada di luar batas bawah. Suatu data disebut *outlier* ketika data tersebut memiliki kemiripan yang rendah bila dibandingkan dengan data lain dalam kelompok. Dalam penggunaan hasil analisis sebagai pendukung pengambilan keputusan bimbingan, pengguna masih harus memperhatikan batas tuntas sekolah. Selain itu, perlu diperhatikan juga bahwa nilai kelas 4,5, dan 6 setiap siswa bias berbeda arti tergantung asal sekolah dasar dari masing-masing siswa dalam kelompok. Nilai 90 untuk mata pelajaran Matematika siswa A di kelas 4 di SD X dapat memiliki arti berbeda dari nilai 90 untuk mata pelajaran yang sama dari siswa B di SD Y.

Siswa dengan nilai *outlier* dalam kelompok yang nilainya ada di bawah batas tuntas merupakan siswa yang memerlukan bimbingan lebih, sedangkan siswa dengan nilai *outlier* yang nilainya diatas batas tuntas tidak. *Outlier* yang dihasilkan dalam perhitungan setiap kelompok biasanya merupakan nilai terendah atau nilai tertinggi yang kemiripannya paling rendah dibanding nilai lain dalam kelompok. Tidak semua *outlier* yang merupakan data nilai terendah dalam kelompok bernilai di bawah batas tuntas.

Siswa dengan nilai *outlier* dalam kelompok yang nilainya ada di bawah batas tuntas merupakan siswa yang memerlukan bimbingan lebih, sedangkan siswa dengan nilai *outlier* yang nilainya diatas batas tuntas tidak. *Outlier* yang dihasilkan dalam perhitungan setiap kelompok biasanya merupakan nilai terendah atau nilai tertinggi yang kemiripannya paling rendah dibanding nilai lain dalam kelompok. Tidak semua *outlier* yang merupakan data nilai terendah dalam kelompok bernilai di bawah batas tuntas.

Sistem selalu menghasilkan *outlier* dalam setiap perhitungan yang dilakukan. Hal ini terjadi karena system mencari subset dengan *smoothing factor* tertinggi. Tidak ada batasan nilai *smoothing factor* untuk penentuan nilai *outlier*. Oleh karena itu, dalam setiap proses perhitungan akan selalu dihasilkan minimal 1 nilai yang menjadi *outlier* dalam kelompok.

5. Kesimpulan

Berdasarkan hasil penelitian dan analisa dapat disimpulkan beberapa hal sebagai berikut :

- Metode *Least Square* dapat digunakan untuk proses perhitungan prediksi nilai siswa.
- Teknik *Exact Exception* dapat diterapkan untuk pengelompokan data nilai akademik siswa menjadi data *outlier* dan bukan *outlier*.
- Outlier* pasti dihasilkan dalam setiap kelompok walau seharusnya belum tentu ada *outlier*.

Daftar Pustaka

- Black, K. (2004). *Business Statistic For Contemporary Decision Making*. United States of America: Leyh Publishing.
- Church, A. P. (2007). *Your Library Goes Virtual*. Worthington, OH : Linworth Books.
- Jaiwei, H. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kauffman.
- Kantardzic, M. (2003). *Concepts, Models, Methods, and Algorithms*. New Jersey: John Wiley & Sons, Inc.
- Montgomery, D. C. & Jennings, C. L. (2008). *Introduction to Time Series Analysis and Forecasting*. United States of America: Willey-Interscience.