
IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENGUKUR PERFORMA AKADEMIK MAHASISWA TIAP SEMESTER PADA PROGRAM STUDI TEKNIK INFORMATIKA STMIK NURDIN HAMZAH JAMBI

Indri Julia C¹, Lucy Simorangkir², Sukma Puspitorini³

¹²³Program Studi Teknik Informatika, STMIK Nurdin Hamzah, Jambi

E-mail: indrijuliac666@gmail.com

Abstract - The STMIK Informatics Engineering Study Program Nurdin Hamzah Jambi uses Student's Values from Attendance, Tasks, UTS, and UAS values to classify student Academic performance. During this time the technique used to classify the academic performance of students using statistical algorithms but the algorithm does not provide any effective results. Therefore, a combination of Mining Data K-Means Clustering is used to provide more effective results. The purpose of this study is to develop the implementation of the K-Means algorithm to measure the academic performance of students each semester in the STMIK Informatics Engineering Study Program Nurdin Hamzah Jambi. The Data Mining method used is k-means clustering using the Rapidminer application. In this study, the implementation of the k-means clustering algorithm was used to classify student academic performance into 3 clusters. Each cluster has an initial centroid that is randomly determined, the centroid changes if each clustering process transfers data to each cluster and changes the centroid to stop if there is no change in the data. Based on 3 attempts to classify 89 student data from semester 1 to 7 there were 89 students from each semester. The data inputted in the form of Absenteeism, Tasks, UTS and UAS Students of the 2015 Informatics Engineering Study Program. The resulting output is in the form of grouping students categorized as groups that are less satisfying, satisfying and very satisfying. after the rapidminer clustering process with k-means, 22.47% of students obtained an increase in academic performance from semester 1 to semester 7. 3.37% students experienced a decline in academic performance and 74.15% the rest Fluctuated.

Keywords: Clustering, K-Means, Microsoft Excel, Performa Akademik, Rapidminer.

I. PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi saat ini semakin meningkat dan memberi pengaruh besar hampir di setiap *sector* kehidupan dan kenegaraan. Proses globalisasi yang terjadi di setiap negara di dunia saat ini juga mendukung perkembangan dan penggunaan teknologi. Hal ini menjadikan kebutuhan akan informasi yang akurat sangat penting dalam kehidupan sehari-hari. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering kali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan menganalisis, meringkas dan mengekstrak pengetahuan dari data. Metode tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar.

Teknologi informasi mempunyai pengaruh besar dalam berbagai aspek kehidupan karena sudah merupakan bagian dari kehidupan masyarakat. Dunia pendidikan, pemerintahan, bisnis dan usaha, sampai kesehatan dan kebutuhan harian masyarakat selalu membutuhkan teknologi informasi. Kualitas mahasiswa ditentukan oleh keterampilan yang

dimiliki tiap mahasiswa sebagai indikator untuk menyelesaikan perkuliahan. Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi menggunakan Nilai yang di dapat Mahasiswa dari nilai Absensi, Tugas, UTS, dan UAS untuk mengelompokkan performa Akademik mahasiswa. Selama ini teknik yang digunakan untuk mengelompokkan performa akademik mahasiswa menggunakan algoritma statistik tetapi algoritma tersebut tidak banyak memberikan hasil yang efektif. Oleh karena itu, kombinasi Data Mining K-Means Clustering digunakan untuk memberikan hasil yang lebih efektif.

Penelitian ini dilakukan dengan metode K-Means dan menggunakan *software* untuk implementasinya yaitu Rapidminer. K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. RapidMiner adalah salah satu *software* untuk pengolahan *data mining*. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.



Untuk mempermudahnya pihak Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi melihat perkembangan Performa Akademik Mahasiswanya pada tiap semester, maka perlu dibuat suatu implementasi menggunakan metode K-Means, dimana akan mempermudah Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi untuk melihat berapa siswa yang masuk kedalam kelompok Performa Akademik Kurang memuaskan, Memuaskan dan Sangat Memuaskan.

1.2. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah : “Bagaimana Implementasi Algoritma K-Means untuk Mengukur Performa Akademik Mahasiswa Tiap Semester Pada Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi?”.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk Membangun Implementasi Algoritma K-Means untuk Mengukur Performa Akademik Mahasiswa Tiap Semester Pada Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi.

II. TINJAUAN PUSTAKA

2.1. Data Mining

Menurut pramudio *data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak didasari keberadaannya (Kusrini, 2009:3).

Menurut lasore *data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Kusrini, 2009:4).

Data mining merupakan cabang ilmu yang termasuk masih baru tetapi telah menghasilkan keuntungan yang cukup besar saat ini (Prabowo dkk, 2013:3).

Dari definisi–definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.2. Clustering

Clustering merupakan sebuah teknik pengelompokan sejumlah data obyek ke dalam *cluster (group)* sehingga dalam cluster akan berisi data yang semirip mungkin (Nugraha, 2014).

Clustering merupakan upaya untuk mengelompokan record, observasi, atau mengelompokan kedalam kelas yang memiliki kesamaan objek (Kusuma V. M., 2017). Pengklasteran berbeda dengan klasifikasi yang tidak adanya variable target dalam pengklasteran. Pengklasteran tidak digunakan untuk melakukan klasifikasi, mengestimasi, dan memprediksi nilai dari target. Pengklasteran digunakan untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan.

Dengan demikian dapat disimpulkan bahwa *clustering* adalah metode penganalisis data, yang sering dimasukan sebagai salah satu metode *data mining*.

2.3. K-Means

Menurut eko prasetyo (2012:178) K-Means merupakan salah satu metode pengelompokan data nonhirarki (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok .

K-Means merupakan metode analisis kelompok yang mengarah pada pemartisan N objek pengamatan ke dalam K kelompok (*cluster*) di mana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma(Eko Prasetyo, 2012:178).

Menurut salah satu jurnal nasional (Afriswati, 2013), *K-Means* merupakan metode pengelompokan data nonhirarki yang berusaha mempartiksi data kedalam dua bentuk atau lebih kelompok.

Dengan demikian dapat disimpulkan bahwa K-Means adalah pengelompokan data untuk mendapatkan hasil yang efektif dan efisien berdasarkan data yang ada.

III. HASIL DAN PEMBAHASAN

3.1. Kebutuhan Masukan

Adapun kebutuhan input dalam pembuatan klasifikasi k-means untuk mengukur Performa Akademik Mahasiswa ini adalah data NIM, Nama, Nilai Absen, Tugas, UTS, UAS.

3.2. Kebutuhan Proses

Adapun kebutuhan proses dalam pembuatan klasifikasi k-means untuk mengukur performa akademik mahasiswa tiap semester pada program studi Teknik Informatika ini adalah sebagai berikut :

1. Proses olah data Performa Akademik RapidMiner.
Proses ini berfungsi untuk mengolah data Performa Akademik menggunakan Rapidminer.
2. Proses olah data Performa Akademik Microsoft Excel
Proses ini berfungsi untuk mengolah data Performa Akademik menggunakan Microsoft Excel.

3.3. Kebutuhan Keluaran

Adapun kebutuhan *output* dalam pembuatan klasifikasi k-means untuk mengukur Performa Akademik Mahasiswa adalah sebagai berikut :

1. Laporan klasifikasi k-means Performa Akademik dari RapidMiner
Laporan hasil dari program yang dipakai berupa hasil klasifikasi k-means untuk mengukur Performa Akademik Mahasiswa Tiap Semester pada Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi menggunakan RapidMiner.
2. Laporan klasifikasi k-means Performa Akademik dari Microsoft Excel
Laporan hasil dari program yang dipakai berupa hasil klasifikasi k-means untuk mengukur Performa Akademik Mahasiswa Tiap Semester pada Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi menggunakan Microsoft Excel

3.4. Kebutuhan Perangkat Lunak

Untuk mendukung sistem ini dibutuhkan perangkat lunak sebagai berikut :

1. Sistem Operasi : Windows 10
2. Aplikasi : Microsoft Excel
3. Aplikasi : RapidMiner
4. Serta beberapa perangkat lunak pendukung lainnya

3.5. Kebutuhan Perangkat Keras

Perangkat keras (*hardware*) yang digunakan dalam pembuatan klasifikasi k-means untuk mengukur Performa Akademik Mahasiswa Tiap Semester pada Program Studi Teknik Informatika STMIK Nurdin Hamzah Jambi harus disesuaikan dengan besarnya data serta kecepatan dalam mengakses data tersebut meliputi Laptop ASUS, dengan spesifikasi :

- a. Prosesor Intel Dual Core N3060, up to 2.48 GHz
- b. RAM dengan kapasitas sebesar 2 GB
- c. Kapasitas penyimpanan *Harddisk Drive* sebesar 500 GB
- d. Layar berukuran 14 inci dengan resolusi 1366x768 piksel
- e. Printe

3.6. FlowChart

Pada *flowchart* pemasukan data ini dijelaskan tentang bagaimana sebuah data yang telah didapatkan oleh peneliti yaitu data Mahasiswa yang terdiri dari field Nilai Absen, Nilai Tugas, Nilai UTS, dan Nilai UAS. Data yang disiapkan inilah yang akan dimasukkan kedalam program yang telah disiapkan dan akan diolah agar nantinya data siap untuk digunakan kedalam proses selanjutnya. Proses awal yang akan dilakukan adalah normalisasi nilai, pada dasarnya normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tertentu. Normalisasi dilakukan bertujuan untuk mengurangi adanya kesalahan pada proses data mining. Pada nilai Absen, Tugas, UTS dan UAS.



Gambar 1. Flowchart Proses Perhitungan K-Means

3.7. Hasil Implementasi

1. Tampilan Awal

Tampilan awal adalah tampilan dari RapidMiner yang baru dibuka dan untuk memulai proses yang akan dilakukan, pilih New Process untuk memulai proses baru yang akan dilakukan didalam Rapidminer. Adapun Tampilan Awal adalah sebagai berikut :



Gambar 2. Tampilan awal



2. Tampilan Menu Proses

Pada tampilan proses ini terdapat kolom proses yaitu kolom yang digunakan untuk menambah data atau menambah operator kedalam kolom tersebut untuk memulai proses *data mining*.



Gambar 3. Tampilan Proses

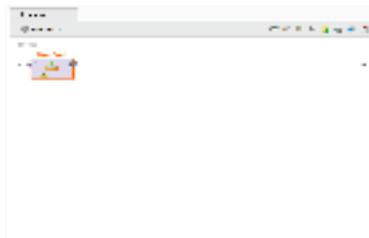
3. Proses Input Data

Pada proses ini dilakukan proses input data dimana data yang akan diolah ke dalam Rapidminer akan diinput terlebih dahulu. Adapun Tampilan Halaman Proses Menghubungkan data adalah sebagai berikut :



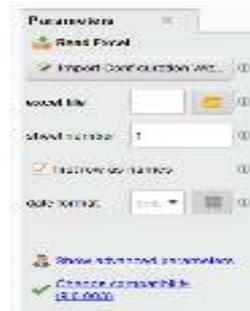
Gambar 4. Operator perintah membaca excel

Pada gambar diatas terdapat folder *data access* yang memiliki folder *files* dan folder *write* yang memiliki perintah menambah data didalam masing-masing folder tersebut, dalam proses kali ini peneliti menarik operator *read excel* yang berfungsi untuk membaca data *excel* untuk dimasukkan kedalam Proses RapidMiner.



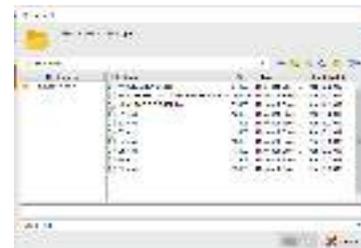
Gambar 5. Process Operator yang digunakan

Gambar diatas menunjukkan operator *read excel* berhasil dimasukkan kedalam proses yang akan dijalankan.



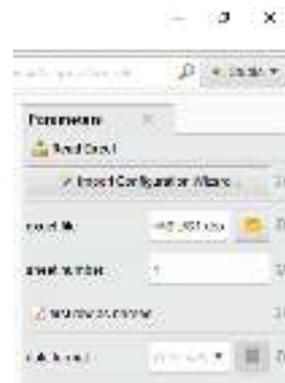
Gambar 6. Parameter Untuk Mengambil DataExcel

Setelah operator *read excel* terdapat dikolom proses untuk memasukkan data *excel* terlebih dahulu kita klik ikon folder yang terdapat diparameter.



Gambar 7. Open File Eexcel

Setelah mengklik ikon folder di parameter akan muncul jendela *open file* dimana peneliti dapat mencari data yang akan di input didalam folder dan memilih data yang akan dipakai.



Gambar 8. Import Configuration

Setelah data diambil peneliti mengklik *import configuration* yang terdapat di parameter untuk melakukan konfigurasi data yang dipakai untuk mengatur format dan pengaturan dari excel diubah kedalam format dan pengaturan RapidMiner.



Gambar 9. Pilih Data yang akan di import

Pada gambar diatas menunjukkan kolom yang terdapat pada data excel peneliti beserta isinya untuk digunakan dalam proses data mining yang dilakukan didalam RapidMiner. Setelah mengikuti langkah dari gambar diatas data excel yang dimiliki sudah terhubung dengan RapidMiner.

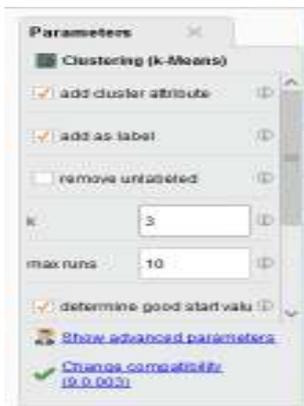
4. Proses Pemilihan Metode

Pada pemilihan metode ini peneliti memilih metode k-means dengan mengambil metodenya pada operator RapidMiner, pada pemilihan operator lalu drag ke dalam proses. Dapat dilihat pada Gambar berikut ini :



Gambar 10. Metode K-Means

Setelah data excel dapat digunakan kali ini peneliti mengambil operator *k-means* dengan cara yang sama dengan pengambilan operator *read excel* sebelumnya dan dimasukkan kedalam kolom proses dan dihubungkan dengan operator *read excel*.



Gambar 11. Parameter Metode K-means

Selanjutnya peneliti mengkonfigurasi dalam parameter proses k-means untuk menambah *cluster attribute* dan *add as label* yang bertujuan untuk menunjukkan hasil *cluster* yang akan didapatkan.

5. Tampilan Hasil Proses

Pada tampilan hasil proses ini menunjukkan hasil olahan data yang telah di import dengan menggunakan metode k-means. Dapat dilihat pada Gambar-gambar dibawah ini :



Gambar 12. Hasil Pembagian Cluster Melalui Description

Pada gambar diatas dapat dilihat bahwa 89 data yang ada terbagi menjadi 3 cluster yang dalam pembagiannya :

- 59 data masuk kedalam cluster_0
- 21 data masuk kedalam cluster_1
- 9 data masuk kedalam cluster_2

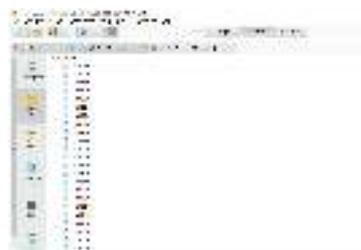
6. Tampilan Folder Cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan data menggunakan metode k-means dalam bentuk folder. Didalam folder terdapat NIM masing masing mahasiswa. Gambar dapat dilihat berikut ini :



Gambar 13. Folder Hasil Pembagian Cluster Melalui Folder View

Pada gambar diatas cluster-cluster yang terbagi 3 tadi dimasukkan kedalam 3 folder yang berbeda.



Gambar 14. Hasil Pembagian Cluster_0 Melalui Folder View

Pada gambar diatas data-data yang termasuk kedalam cluster_0 ada didalam folder cluster 0.



Gambar 15. Hasil Pembagian *Cluster_1* Melalui *Folder View*

Pada gambar diatas data-data yang termasuk kedalam *cluster_1* ada didalam folder *cluster 1*.



Gambar 16. Hasil Pembagian *Cluster_2* Melalui *Folder View*

Pada gambar diatas data-data yang termasuk kedalam *cluster_2* ada didalam folder *cluster 2*.

7. Tampilan Grafik Cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk grafik dengan ukuran yang berbeda-beda dengan menunjukkan besar nilai yang terdapat pada cluster tersebut. Gambar dapat dilihat berikut ini :



Gambar 17. Grafik Pembagian *Cluster* Di Menu *Graph*

Pada gambar grafik di atas menunjukkan dari 89 data yang digunakan *cluster_0* memiliki jumlah data yang lebih besar dibanding *cluster_1* dan *cluster_2*.

8. Tampilan Tabel Centroid

Pada tampilan hasil proses ini menunjukkan centroid yang dimiliki tiap *cluster*. Gambar dapat dilihat berikut ini :

Gambar 18. Tabel *Centroid*

Pada gambar diatas menunjukkan hasil pusat cluster yang didapat dari perhitungan k-means.

9. Tampilan Grafik Batang

Pada tampilan hasil proses ini menunjukkan hasil olahan data dengan menggunakan metode k-means dalam bentuk grafik batang. Gambar dapat dilihat berikut ini :



Gambar 19. Grafik batang *cluster*

Pada grafik di atas menunjukkan jarak cluster yang dimiliki tiap kolom data yang didapat.

10. Tampilan Cluster dan Jarak Cluster

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode k-means dalam bentuk data tiap individu dalam bentuk tabel yang memiliki kelompok *cluster* dan jarak tiap *cluster*. Gambar dapat dilihat berikut ini :

Gambar 20. Hasil Pembagian *Cluster*

Hasil dari pembagian *cluster* yang dilakukan ditambah kedalam kolom bernama label yang menunjukkan pembagian *cluster* tiap data dan ditunjukkan jarak antara tiap data yang digunakan dalam kolomnya masing-masing.

IV. PENUTUP

4.1. Kesimpulan

Setelah peneliti membahas permasalahan yang peneliti angkat dalam laporan ini untuk mengukur performa akademik mahasiswa tiap semester pada program studi TI STMIK Nurdin Hamzah Jambi dan membandingkan hasil yang diperoleh dengan hasil sebelumnya, maka peneliti dapat menarik kesimpulan sebagai berikut:

1. Implementasi klasterisasi k-means menggunakan RapidMiner mempermudah pihak Prodi Teknik Informatika STMIK Nurdin Hamzah Jambi untuk mengolah data Mahasiswa berupa Nilai Absen, Tugas, UTS dan UAS dalam mengetahui meningkat atau menurunnya performa akademik mahasiswa pada tiap semester.
2. Variabel yang digunakan sebagai atribut dalam mengukur performa akademik mahasiswa menggunakan metode k-means *clustering* ini adalah atribut nilai absen, nilai tugas, nilai uts dan nilai uas.
3. Pada penelitian ini dari 89 orang mahasiswa yang diteliti, setelah dilakukan proses *clustering* dengan k-means pada rapidminer diperoleh hasil sebanyak 22,47% mahasiswa mengalami peningkatan performa akademik dari semester 1 sampai dengan semester 2, 7,3,37% mahasiswa mengalami penurunan performa akademik dan 74,15% sisanya Fluktuatif
4. Mahasiswa yang tidak mengalami perubahan cluster dari semester 1 sampai dengan 2 terdapat 15 mahasiswa dan semuanya termasuk kedalam kategori sangat memuaskan.
5. *Output* yang dihasilkan aplikasi Rapidminer adalah pembagian cluster dan Grafik yang dapat dipilih modelnya sesuai dengan yang diinginkan.

4.2. Saran

Dari kesimpulan diatas dan pembahasan bab sebelumnya penulis menyarankan beberapa hal, yaitu :

1. Ketika menggunakan RapidMiner pastikan bahwa pencarian data mining K-Means dibantu dengan pencarian secara manual.
2. Untuk Kedepannya Mahasiswa yang ingin mengembangkan penelitian ini bisa menambahkan antarmuka *interface* untuk proses input data mahasiswa baru.

3. Dibandingkan dengan metode lain misal *Regresi Linear*. Agar lebih bisa mengetahui metode mana yang hasilnya lebih akurat.
4. Bagi para pembaca dan mahasiswa STMIK NH Jambi, laporan skripsi ini dapat menjadi bahan masukan untuk melakukan penelitian lebih lanjut lagi dan menjadi lebih baik lagi di masa mendatang.
5. Penelitian ini bisa lebih dikembangkan lagi di masa mendatang dengan membuat aplikasi untuk mengukur performa akademik mahasiswa menggunakan bahasa pemrograman berbasis web.

DAFTAR REFERENSI

- Afriswati. 2013. "*Jurnal implementasi data mining pemilihan pelanggan potensial menggunakan algoritma K-Means*".
- Eko prasetio. 2012. "*Data Mining Konsep dan aplikasi Menggunakan MATLAB*". Yogyakarta: Andi.
- Feri Sulianta, Dominikus Juju. 2010. "*Data Mining Meramalkan Bisnis Perusahaan*". Jakarta: PT Elex Media Koputindo.
- Han, J., M. Kamber, dan J. pei. 2011. "*Data Mining concept and techniques third edution*". Morgan Kaufmann publishers, san Francisco.
- Kusrini, Emha dkk. 2009. "*Algoritma Data Mining*". Yogyakarta: Andi.
- Rahmad. 2014. "*Jurnal Pengelompokkan Performa Akademik Mahasiswa Berdasarkan Indeks Prestasi Menggunakan K-Means Clustering*".
- Suliyanto. 2012. "*Analisis Statistik Pendekatan Praktis dengan Microsoft Excel*". Yogyakarta: Andi.

IDENTITAS PENULIS

Nama : Indri Julia C
NIM : 1502090
TTL : Nipah Panjang, 08 Juli 1998
Alamat Rumah : Jl. Soemantri Brojonego
Jambi.
Telp. : 081369945656
Email : indrijuliac666@gmail.com