

ANALISA KEMUNGKINAN CALON MAHASISWA BARU MEMILIH PRODI TEKNIK INFORMATIKA MELALUI PENERAPAN DATA MINING DECISION TREE

Oleh Soleh¹⁾, Ollynia Devega²⁾, Pandu Lutfi Asoti³⁾, Katri Olivia⁴⁾

^{1), 2) 3) 4)} Teknik Informatika STMIK Raharja Tangerang
Jl Jendral Sudirman No. 40 Modernland, Tangerang
Email : olehsoleh@gmail.com¹⁾, vegakusdaryanto@gmail.com²⁾, pandu_campus@yahoo.com³⁾,
rae.fahreza@gmail.com⁴⁾

Abstract

Informasi mengenai mahasiswa merupakan hal yang sangat penting bagi sebuah universitas. Dengan mengetahui informasi mengenai data diri, minat atau ketertarikan mahasiswa, sebuah universitas dapat mengambil berbagai kesimpulan yang dapat digunakan sebagai analisa dalam membuat sebuah keputusan. Decision tree merupakan metode yang paling banyak digunakan untuk membuat sebuah prediksi atau kemungkinan dalam mencari sebuah keputusan melalui klasifikasi dari sejumlah data yang besar dalam penerapan data mining. Kelebihan dari metode ini adalah dapat menganalisa data yang besar secara efektif agar mudah dipahami oleh pengguna akhir, serta daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik. Selain itu eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode decision tree maka sample diuji hanya berdasarkan criteria atau kelas tertentu seperti kabupaten, kecamatan, jenis kelamin, fakultas dan prodi. Metode ini dapat diterapkan untuk menganalisa kemungkinan calon mahasiswa baru yang memilih prodi teknik informatika pada studi yang dilakukan di Universitas Islam Syekh Yusuf (UNIS) Tangerang dalam daerah pengambilan keputusan pada proses penerimaan calon mahasiswa tahun ajaran 2013/2014. Untuk membuat sebuah decision tree, dibutuhkan sejumlah sample dari data calon mahasiswa baru dan data mengenai universitas, terutama pada prodi teknik informatika melalui beberapa criteria. Dengan penerapan data mining decision tree, pihak universitas akan mengetahui seberapa besar minat dan ketertarikan calon mahasiswa baru, serta memprediksi seberapa banyak calon mahasiswa yang akan memilih prodi teknik informatika. Prediksi yang dihasilkan dari klasifikasi decision tree akan mempengaruhi pihak universitas dalam mengambil keputusan untuk tindakan selanjutnya.

Key words: decision tree, data mining, prediksi, mahasiswa baru, minat, criteria

1. Pendahuluan

a. Latar Belakang

Mengetahui informasi mengenai calon mahasiswa baru merupakan hal yang sangat penting bagi sebuah perguruan tinggi atau universitas. Karena, dalam pelaksanaannya, mahasiswa adalah *input* dan hal paling utama dalam kegiatan operasional sebuah universitas yang dapat mempengaruhi banyak aspek dalam universitas tersebut.

Pihak universitas terkadang memiliki kesulitan dalam mengetahui berapa jumlah calon mahasiswa baru yang tertarik dengan sebuah jurusan/prodi di tiap tahunnya, terutama pada universitas swasta. Dengan pertumbuhan penduduk yang semakin maju, juga banyaknya perguruan tinggi/universitas yang saling bersaing dalam menyediakan pendidikan yang lebih tinggi membuat suatu universitas sulit dalam menganalisa kemungkinan minat calon mahasiswa baru di universitas tersebut.

Mengetahui ketertarikan calon mahasiswa baru terhadap Prodi Teknik Informatika pada Universitas Islam Syekh Yusuf (UNIS) Tangerang, merupakan hal yang harus dilakukan oleh pihak universitas untuk menganalisa minat dan jumlah mahasiswa baru yang tertarik dengan melihat kriteria-kriteria yang berhubungan dan mempengaruhi ketertarikan untuk memilih prodi tersebut. Kriteria yang digunakan untuk melihat kemungkinan dan ketertarikan calon mahasiswa baru dapat dilihat berdasarkan kriteria seperti: jenis kelamin, fakultas pilihan, prodi, kecamatan dan kabupaten tempat tinggal.

b. Perumusan Masalah

Dengan menggunakan metode *decision tree* dalam penerapan *data mining*, Universitas Islam Syekh Yusuf (UNIS) Tangerang ingin mengetahui berapa banyak kemungkinan calon mahasiswa baru yang berminat dan memilih Prodi Teknik Informatika berdasarkan informasi (kriteria) mengenai calon mahasiswa baru tahun ajaran 2013/2014.

c. Tinjauan Pustaka

Tinjauan pustaka dimaksudkan untuk memperkaya ilmu pengetahuan melalui penelitian sebelumnya yang berkaitan serta mencegah untuk melakukan penelitian ulang mengenai permasalahan yang sedang diteliti. Untuk itu, penulis menemukan beberapa literatur penelitian yang berkaitan yaitu sebagai berikut:

1. Menurut penelitian yang dilakukan oleh Tacbir Hendro P, dkk dalam penelitian yang berjudul “*Penerapan Data Mining Untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru*” ditemukan bahwa pihak perguruan tinggi kesulitan mengetahui jumlah mahasiswa yang akan registrasi dari sejumlah calon mahasiswa yang sudah dinyatakan lolos seleksi penerimaan mahasiswa baru. Dengan tidak diketahuinya mahasiswa yang akan registrasi, maka pihak-pihak perguruan tinggi tidak dapat mengetahui dengan pasti kapan jumlah target penerimaan mahasiswa baru tercapai. Hal ini dikarenakan jumlah mahasiswa yang telah dinyatakan lolos seleksi penerimaan mahasiswa baru, tidak menggambarkan jumlah mahasiswa yang akan melakukan registrasi. Apabila hal tersebut baru diketahui pada akhir proses penerimaan mahasiswa baru akan mengakibatkan target menjadi tidak tercapai. Hal ini juga dapat menjadi gangguan dalam hal finansial, karena biasanya pembiayaan perguruan tinggi swasta sebagian besar dari dana yang didapatkan dari mahasiswa. Dengan diterapkannya *data mining*, data calon mahasiswa 2007-2009 digunakan sebagai *data sample* untuk pembentukan pohon keputusan, dan didapatkan hasil kecocokan prediksi dengan menggunakan 600 data calon mahasiswa baru 2010 yang diujikan terdapat tingkat kecocokan hasil prediksi dengan hasil nyata sebesar 61,89%. Hasil uji tersebut menyatakan sistem cukup layak untuk digunakan dalam prediksi calon mahasiswa baru meskipun sistem belum menghasilkan tingkat akurasi yang maksimal^[1].
2. Menurut penelitian yang dilakukan oleh Yulfita Aini, dkk dalam penelitian yang berjudul “*Analisis Pengaruh Marketing Mix Terhadap Keputusan Mahasiswa Dalam Melanjutkan Studi Pada Perguruan Tinggi (Studi Pada Universitas Pasir Pangaraian – Riau)*” ditemukan bahwa melalui pengujian secara parsial, ditemukan bahwa variabel harga, produk, lokasi, *people*, bukti fisik dan proses memiliki pengaruh yang signifikan terhadap keputusan mahasiswa, sedangkan variabel promosi tidak signifikan. Hal ini dikarenakan media promosi yang digunakan tidak tepat sasaran. Variabel harga merupakan variabel dominan terhadap keputusan mahasiswa. Hal ini dapat dilihat dari hasil koefisien determinasi parsial variabel harga sebesar 56,1%. Hasil penelitian yang dilakukan terhadap 104 responden menunjukkan bahwa berdasarkan pengujian dengan model regresi linier berganda,

marketing mix bersama-sama berpengaruh secara signifikan terhadap keputusan mahasiswa sebesar 83%^[2].

3. Menurut Penelitian yang dilakukan oleh Jaka Adi Swara dalam penelitian yang berjudul “*Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Pemantauan Distribusi Penjualan Sepeda Motor Di PD. Wijaya Abadi Bandung*” mengemukakan bahwa perusahaan ingin adanya pengolahan data lebih lanjut untuk menemukan informasi/pengetahuan baru yang berguna. Informasi yang diinginkan pihak perusahaan adalah bersifat pengklasifikasian keterkaitan suatu perusahaan target pendistribusiannya dengan atribut tipe motor dan kode warna yang ada pada data distribusi penjualan sepeda motor. Untuk mendapatkan informasi tersebut secara lebih cepat dan efisien perlu kiranya suatu bantuan teknologi informasi, dalam hal ini yaitu data mining. Hasil keluaran dari aplikasi ini adalah informasi yang bersifat klasifikasi dalam bentuk *Decision Tree*. Aplikasi ini dapat menguji nilai dari atribut untuk mengetahui pola aturan dari pendistribusian yang ada sehingga perusahaan mempunyai bekal pengetahuan dalam mendistribusikan sepeda motor ke perusahaan target pendistribusiannya^[3].
4. Menurut penelitian yang dilakukan oleh Wahyudin dalam penelitian yang berjudul “*Metode Iterative Dichotomizer 3 (ID3) Untuk Penerimaan Mahasiswa Baru*” dikemukakan bahwa *Iterative Dichotomizer 3 (ID3)* merupakan suatu metode dalam learning yang akan membangun sebuah pohon keputusan yang merupakan suatu pemodelan dalam mencari solusi dari persoalan. Dalam jurnal ini akan dibahas pemakaian pohon keputusan dalam bidang pendidikan, yaitu dalam hal penerimaan mahasiswa baru. Penerimaan mahasiswa baru yang baik mencakup langkah-langkah yang dapat memaksimalkan peluang keberhasilan. Salah satu tahap yang penting adalah tahap pengambilan keputusan. Kegiatan analisis kemahasiswaan diperlukan untuk mendapatkan keputusan yang bersifat menguntungkan demi maju dan berkembangnya suatu universitas dan analisis penerimaan mahasiswa baru tersebut dapat dilakukan melalui berbagai metode, salah satunya dengan *decision tree* menggunakan *ID3 (Iterative Dichotomiser 3)*. Jurnal ini akan memperlihatkan pemakaian pohon keputusan untuk memudahkan pengambilan keputusan penerimaan mahasiswa baru dalam suatu universitas^[4].

d. Landasan Teori

Data Mining

Secara sederhana, *data mining* adalah ekstraksi dari sejumlah data yang besar. *Data mining* sering disebut

knowledge discovery in database (KDD), yang semua prosesnya adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dengan *set data* yang berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [5].

Decision Tree

Decision Tree merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan data mining. Pemilihan metode ini didasarkan pada informasi yang ingin diperoleh yaitu informasi yang bersifat klasifikasi. Keuntungan dalam metode ini adalah efektif dalam menganalisis sejumlah besar atribut dari data yang ada dan mudah dipahami oleh pengguna akhir [3].

Decision Tree adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah *decision tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *decision tree* melakukan strategi pencarian secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai node akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu [4].

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan *atribut* dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon.

Entropy

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Entropi mengukur jumlah dari informasi yang ada pada atribut. Rumus menghitung entropi informasi adalah:

$$Entropy(S) = -p+ \log p+ - p- \log_2 p-$$

Dimana:

S = ruang (data) *sample* yang digunakan untuk *training*.
 P+ = adalah jumlah yang bersolusi positif (mendukung) pada data *sample* untuk kriteria tertentu.

P- = adalah jumlah yang bersolusi negatif (tidak mendukung) pada data *sample* untuk kriteria tertentu.

2. Pembahasan

Untuk menganalisa kemungkinan mahasiswa baru memilih Prodi Teknik Informatika, maka dibutuhkan ruang sampel untuk pembuatan pohon keputusan/*decision tree*. *Sample* data mahasiswa yang

diambil adalah 100 *sample* dari 988 data calon mahasiswa baru.

Tabel 1. Contoh tabel *sample* pembuatan *decision tree*

Kabupaten /Kota	Kecamatan	Jenis Kelamin	Fakultas	Kode Prodi
Kab Tangerang	Pasar Kemis	Perempuan	FKIP	0502
Kota Tangerang	Jatiuwung	Perempuan	F-TEKNIK	0403
Kab Tangerang	Rajeg	Perempuan	FKIP	0501
Kota Tangerang	Karawaci	Perempuan	FKIP	0502
Kota Tangerang	periuk	Perempuan	F-EKONOMI	0601
Kota Tangerang	Karawaci	Laki-Laki	F-EKONOMI	0602
Kota Tangerang	Tangerang	Laki-Laki	FKIP	0501
Kab Tangerang	Rajeg	Perempuan	F-HUKUM	0201
Kota Tangerang	Tangerang	Perempuan	F-EKONOMI	0601

Tabel 2. Tabel data prodi dan jurusan

Kode Prodi	Jurusan	Fakultas
0101	Adm. Negara	FISIP
0103	I-Komunikasi	FISIP
0201	I-Hukum	F-HUKUM
0302	Tarbiyah	F-AGAMA ISLAM
0401	T-Kimia	F-TEKNIK
0402	T-Industri	F-TEKNIK
0403	T-Informatika	F-TEKNIK
0404	T-Sipil	F-TEKNIK
0501	P-Ekonomi	FKIP
0502	P-B. Inggris	FKIP
0601	Manajemen	F-EKONOMI
0602	Akuntansi	F-EKONOMI
0701	M.Si	PASCA
0702	MH	PASCA
0703	MM	PASCA

Proses pembentukan pohon keputusan dalam pengelolaan data digunakan untuk menurunkan atribut-atribut pada data menjadi pohon keputusan sehingga menghasilkan aturan-aturan. Dari keseluruhan data mengenai mahasiswa, diambil beberapa kriteria yang dapat disebut sebagai atribut yang dimasukkan dalam tabel, yaitu Kode_Prodi, Fakultas, Kabupaten/Kota, Kecamatan, dan Jenis_Kelamin.

Untuk penentuan *node* pertama dalam pembentukan pohon keputusan. Maka harus menghitung nilai *entropy* terlebih dahulu. Setelah mendapatkan hasil *entropy* maka barulah selanjutnya dapat menghitung nilai *information gain* untuk mencari *node* yang akan menjadi *node* induk. *Information gain* adalah ukuran efektivitas suatu atribut dalam mengklasifikasikan data.

Tabel 3. Tabel klasifikasi dan perhitungan entropy

Parameter	Nilai	Jumlah	Memilih	tidak	Entropy	gain
keseluruhan					0.2232	
Jns_Kel	P	59	7	52	0.1582	0.0156
	L	41	14	27	0.2788	
Kab/Kota	Kab. Tangerang	35	5	30	0.1781	0.0134
	Kota Tangerang	58	14	44	0.24	
	Tangerang Selatan	5	0	5		
	Luar Daerah	3	2	1	0.2764	
Kec	Wilayah 1	18	1	17	0.0931	0.0295
	Wilayah 2	1	0	1		
	Wilayah 3	11	3	8	0.2544	
	Wilayah 4	5	0	5		
	Wilayah 5	58	14	44	0.24	
	Wilayah 6	4	3	1	0.2442	

- Jika Kab/Kota B, Kec W5, maka Y
- Jika Kab/Kota B, Kec W6, maka Y
- Jika Kab/Kota C, maka N
- Jika Kab/Kota D, Jns_Kel L, Kec W1,W5,W6, maka Y
- Jika Kab/Kota D, Jns_Kel L, Kec W2,W3,W4, maka N
- Jika Kab/Kota D, Jns_Kel P, maka Y

Keterangan:

- A = Kabupaten Tangerang
- B = Kota Tangerang
- C = Tangerang Selatan
- D = Luar Daerah

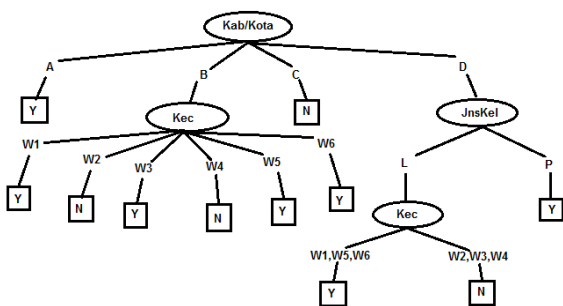
- W1 : Mauk, Rajeg, Pakuhaji, Sepatan, Teluk Naga, Kosambi, Pasar Kemis
- W2 : Kronjo, Balaraja, Cisoka
- W3 : Tigaraksa, Cikupa, Panongan, Curug, legok
- W4 : Cisauk, Serpong, Pondok Aren, Ciputat, Pamulang
- W5 : Neglasari, Periuk, Jatiuwung, Batu Ceper, Cibodas, Karawaci, Cipondoh, Ciledug
- W6 : Kalideres, Serang, Lampung, Karang Tengah, Sindang Jaya

3. Desain Usulan

a. Use Case Diagram

Setelah atribut atau parameter atau kriteria telah diklasifikasikan, maka dapat dicari nilai *entropy* masing-masing atribut, sehingga penentuan *root* atau akar dari pohon keputusan dapat dibuat.

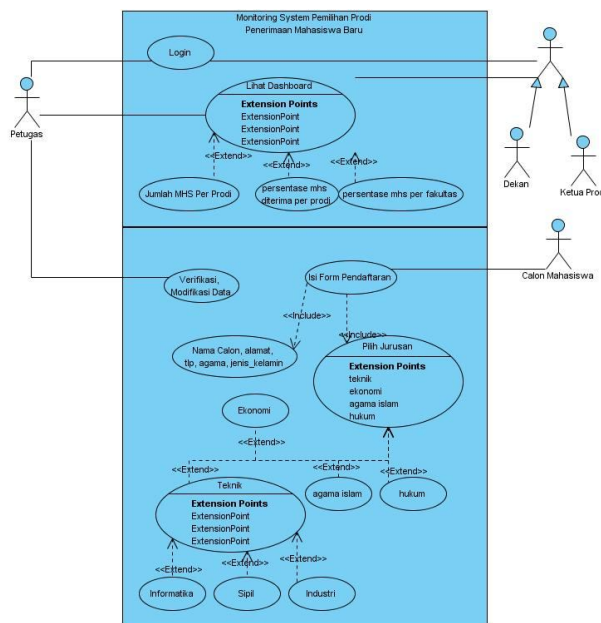
Pohon keputusan yang dihasilkan bersifat *dinamis* tergantung dari jumlah data dan atribut yang digunakan dalam *sample training*. Dari pohon keputusan yang dihasilkan akan terbentuk aturan-aturan dari penurunan data.



Gambar 1. Gambar decision tree kemungkinan calon mahasiswa memilih prodi Teknik Informatika

Daftar Aturan:

- Jika Kab/Kota A, maka Y
- Jika Kab/Kota B, Kec W1, maka Y
- Jika Kab/Kota B, Kec W2, maka N
- Jika Kab/Kota B, Kec W3, maka Y
- Jika Kab/Kota B, Kec W4, maka N



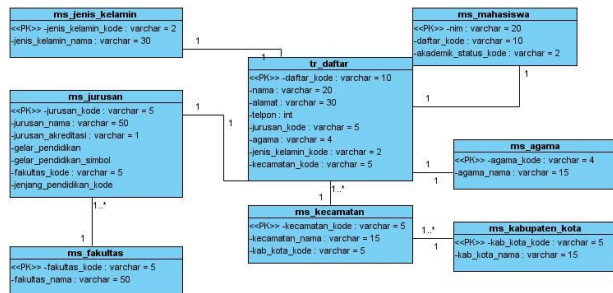
Gambar 2. Use Case Diagram

Berdasarkan Use case diagram diatas menunjukkan actor yang berperan adalah petugas, calon mahasiswa dan staff yang terdiri dari dekan dan ketua prodi. Use case yang ada adalah login, lihat dashboard yang bisa diextend menjadi jumlah mahasiswa per prodi, persentase mahasiswa diterima per prodi, persentase mahasiswa per

fakultas, jumlah mahasiswa per wilayah yaitu kabupaten atau kota.

Untuk actor calon mahasiswa mempunyai kewajiban untuk melakukan input data (isi form pendaftaran) yang didalamnya memuat, profile diri dari calon dan pemilihan fakultas serta prodi yang di inginkan. Untuk fakultas teknik, prodi yang dipilih adalah informatika, sipil dan industry.

b. Class Diagram



Gambar 3 : Class Diagram

Berdasarkan gambar 3 mengenai Class Diagram yang diusulkan terdapat 6 class yaitu tr_daftar, ms_fakultas, ms_kabupaten_kota, ms_jurusan, ms_mahasiswa, ms_fakultas, ms_kecamatan, ms_jenis_kelamin. Class tr_daftar sebagai class transaksi dan selebihnya merupakan class master.

4. Kesimpulan

Dengan menggunakan pohon keputusan, dapat diketahui prediksi kemungkinan calon mahasiswa yang memilih Prodi Teknik Informatika adalah 49,75% dengan menggunakan data *sample* sebanyak 100 responden. Hasil uji tersebut menyatakan sistem cukup layak untuk digunakan dalam prediksi calon mahasiswa baru meskipun sistem belum menghasilkan tingkat akurasi yang maksimal.

Daftar Pustaka

- [1] Pudjiantoro, Tacbir Hendro, dkk. *Penerapan Data Mining Untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru*. Konferensi Nasional Sistem dan Informatika; Bali, November 12, 2011.
- [2] Aini, Yulfitra. *Analisis Pengaruh Marketing Mix Terhadap Keputusan Mahasiswa Dalam Melanjutkan Studi Pada Perguruan Tinggi (Studi Pada Universitas Pasir Pangaraian – Riau)*. Unpublished.
- [3] Swara, Jaka Adi. *Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Pemantauan Distribusi Penjualan Sepeda Motor Di PD. Wijaya Abadi Bandung*. Unpublished.
- [4] Wahyudin. *Metode Iterative Dichotomizer 3 (ID3) Untuk Penerimaan Mahasiswa Baru*. Jurnal Program Studi Ilmu Komputer, Universitas Pendidikan Indonesia, http://file.upi.edu/Direktori/FPMIPA/PRODI_ILMU_KO

MPUTER/WAHYUDIN/metode_ID3_untuk_mhs_baru.p df, diakses tanggal 18 Juli 2013.

- [5] Fayyad, Usama. (1996). *Advance in Knowledge Discovery and Data Mining*. MIT Press.

Biodata Penulis

Oleh Soleh, lulusan S1 dari Universitas Gunadarma untuk Jurusan Sistem Informasi dan melanjutkan ke S2 di Universitas Gunadarma dengan jurusan Sistem Informasi Bisnis. Saat ini merupakan salah satu pengajar di Perguruan Tinggi Raharja, Tangerang.

Ollynia Devega, mahasiswa tingkat akhir Jurusan Sistem Informasi dengan konstrasi Komputer Akuntansi di STMIK Raharja Tangerang yang sedang menempuh skripsi.

Pandu Lutfi Asoti, mahasiswa tingkat akhir Jurusan Sistem informasi STMIK Raharja Tangerang yang sedang menempuh skripsi.

Katri Olivia, mahasiswa tingkat akhir Jurusan Sistem Informasi STMIK Raharja Tangerang yang sedang menempuh skripsi.

