

Seleksi Atribut Indeks Prestasi Mahasiswa Menggunakan Metode Indeks *Jaccard* dan *Rough Sets*

Riswan Efendi^{1,*}, Nur Ain Hera Septia¹, Latifa Nabila¹, Rika Susanti², Ewi Ismaredah²
Reski Mai Candra³, Idria Maita⁴, Misra Hartati⁵, M. Afdal⁶

¹Jurusan Matematika, ²Jurusan Teknik Elektro, ³Jurusan Teknik Informatika, ^{4,6}Jurusan Sistem Informasi,

⁵Jurusan Teknik Industri, Fakultas Sains dan Teknologi,

Universitas Islam Negeri Sultan Syarif Kasim Riau,

Pekanbaru, Indonesia

*riswan.efendi@uin-suska.ac.id

Abstrak

Banyak sekali atribut yang mempengaruhi prestasi akademik mahasiswa di perguruan tinggi. Analisis hubungan antar atribut sering kali dilakukan oleh para peneliti sebelumnya dengan menggunakan pendekatan regresi, dan model-model statistik lainnya. Namun sayangnya, pendekatan statistik ini sangat berkaitan dengan beberapa asumsi klasik, seperti jumlah data, sebaran data, dan sebagainya, sehingga tidak mudah untuk dipenuhi oleh atribut yang tipenya kategori. Melalui makalah ini, penulis ingin mengaplikasikan pendekatan non-statistik untuk menganalisa hubungan antar atribut bersyarat dan atribut keputusan IPK mahasiswa, yaitu metode indeks *Jaccard* dan *rough sets*. Kedua metode ini bekerja atas dasar operasi himpunan bilangan riil, khususnya, operasi irisan dan selisih. Berdasarkan hasil pengolahan data riil diperoleh tiga atribut yang cukup mempengaruhi prestasi akademik mahasiswa Jurusan Matematika, UIN Sultan Syarif Kasim Riau, seperti, lamanya bermain *hand phone* per hari, jumlah unduhan artikel ilmiah per hari dan jumlah halaman Al-Qur'an yang dibaca per hari oleh mahasiswa tersebut. Ketiga atribut ini memberikan pengaruh yang signifikan terhadap IPK mahasiswa.

Kata Kunci: *Indeks Jaccard*, *Rough Sets*, IPK, membaca Al-Quran, unduhan artikel, *hand phone*

1. Pendahuluan

Mahasiswa merupakan aspek penting dalam mengevaluasi keberhasilan atau kegagalan proses pembelajaran di perguruan tinggi. Dalam proses belajar, para mahasiswa harus mampu mengembangkan pola pikir dengan tujuan untuk mempersiapkan diri untuk menjadi sudah berkompetisi di industri publik dan di lingkungan masyarakat nantinya. Kenyataannya tentu akan erat kaitannya dengan indeks prestasi yang mereka raih pada setiap semesternya. Indeks prestasi kumulatif (IPK) mahasiswa menjadi tolok ukur sejauh mana hasil proses pencapaian pembelajaran di setiap semester berhasil dan tuntas dilaksanakan oleh pihak jurusan.

Banyak sekali atribut-atribut yang mempengaruhi IPK mahasiswa di perguruan tinggi, baik internal maupun eksternal. Sebagaimana yang dikaji oleh para peneliti sebelumnya, ada beberapa atribut yang menarik dan unik signifikan terhadap prestasi akademik mahasiswa, seperti, media sosial *Facebook* [6, 7], jumlah bahasa asing yang dikuasai [5], serta makanan cepat saji (*fast food*) [8, 9]. Sehingga dapat disimpulkan bahwa atribut-atribut ini juga berkembang sesuai dengan trend yang terjadi pada era mahasiswa tersebut berada di perguruan tinggi. Disamping itu, berbagai metode statistik juga sering diaplikasikan untuk membentuk hubungan antar atribut tersebut [5-9]. Namun, metode-metode ini tidak semuanya bisa diaplikasikan dengan mudah, khususnya untuk jumlah data yang sedikit dengan tipe kategori.

Melalui makalah ini, tujuan utama adalah untuk memperkenalkan metode non-statistik dalam menginvestigasi atribut-atribut yang mempengaruhi prestasi akademik mahasiswa. Disamping itu, penelitian ini juga dimotivasi oleh kajian sebelumnya tentang aplikasi indeks *Jaccard* dan *rough sets* model terhadap atribut pasien flu [4].

2. Landasan Teori

Pada bagian ini akan dijelaskan tentang metode rough sets dan indeks Jaccard serta formula-formula terkait kedua metode tersebut.

2.1. Sistem Informasi

Sistem informasi dapat disajikan dalam bentuk tabel yang dapat menjelaskan hubungan antar objek dan atributnya. Dalam penelitian ini, penulis mempertimbangkan tiga atribut yang cukup trend dikalangan mahasiswa zaman sekarang (now) yaitu durasi untuk bermain gadget (hand phone) dalam satu hari, jumlah unduhan ilmiah dalam seminggu dan jumlah lembar dari Al-Quran yang dibaca dalam satu hari oleh mahasiswa tersebut.

2.2. Jaccard Indeks

Indeks Jaccard, juga dikenal sebagai persimpangan atau irisan dari koefisien. Persamaan Jaccard diprakarsai oleh Paul Jaccard. Jaccard digunakan untuk membandingkan kesamaan dan keragaman set sampel [2]. Persamaan ini mengukur kesamaan antara kumpulan sampel berhingga, dan didefinisikan sebagai ukuran irisan dibagi dengan ukuran gabungan:

Indeks kesamaan Jaccard diformulasikan sebagai berikut:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{(|A| + |B| - |A \cap B|)}. \quad (1)$$

Jaccard yang mengukur ketidaksamaan antara set sampel, merupakan pelengkap terhadap koefisien Jaccard dan diperoleh dengan mengurangi koefisien Jaccard dari 1, atau, ekuivalen, dengan membagi perbedaan ukuran gabungan dan perpotongan dua set oleh ukuran gabungan:

$$dJ(A, B) = 1 - J(A, B). \quad (2)$$

2.3. Teori Rough Sets

Pawlak (1982) diperkenalkan teori rough set yang umumnya digunakan untuk analisis data kategori. Teori rough set juga dibagi menjadi beberapa komponen, seperti, sistem informasi, hubungan indiscernibility, mengatur perkiraan, pengelompokan kasar, dan lain-lain. Suatu sistem informasi $S = (U, \Omega, Vq, fq)$ terdiri dari himpunan alam semesta yang terbatas, himpunan atribut yang terbatas, dan lainnya [1-3].

2.4. Lower Approximations

Aproksimasi yang lebih rendah (B) adalah kumpulan objek yang dapat diklasifikasikan pasti sebagai anggota dari himpunan A, dengan menggunakan atribut dari himpunan B [1-3].

2.5. Upper Approximations

Aproksimasi atas (B) adalah kumpulan objek yang mungkin dapat diklasifikasikan sebagai anggota himpunan A [1-3].

2.6. Boundary Region

Wilayah batas adalah perbedaan antara aproksimasi atas dan aproksimasi yang lebih rendah. Wilayah batas X ditulis sebagai [1-3]:

$$\underline{B}(X) = \{ \{x \in U \mid [x]_B \subseteq X \} \}, \quad (3)$$

$$\bar{B}(X) = \{ \{x \in U \mid [x]_B \cap X \neq \emptyset \} \}, \quad (4)$$

$$BND(X) = \underline{B}(X) - \bar{B}(X) \quad (5)$$

3. Seleksi Atribut-atribut IPK Menggunakan Indeks Jaccard dan *Rough Sets*

Pada bagian ini akan diimplementasi metode non-statistik untuk menginvestigasi atribut-atribut yang mempengaruhi IPK mahasiswa Jurusan Matematika UIN Sultan Syarif Kasim Riau. Pembahasan secara terperinci akan disajikan pada Sub-bab 3.1-3.3.

3.1. Pemilihan Atribut IPK Menggunakan Indeks Jaccard

Hal yang paling utama pada metode indeks Jaccard adalah menyajikan atribut dan keputusan dalam bentuk tabel hubungan objek dengan atributnya seperti pada Tabel 1.

Tabel 1. Atribut dan keputusan IPK

Kode Mahasiswa	Atribut Kondisi			Keputusan IPK
	HP	Unduh	Quran	
m1	tidak	tidak	tidak	tidak
m2	tidak	tidak	tidak	tidak
m3	ya	ya	ya	ya
m4	tidak	ya	ya	ya
m5	ya	ya	ya	ya
m6	ya	ya	ya	ya
m7	ya	tidak	ya	ya
m8	ya	ya	tidak	ya
m9	ya	ya	ya	ya
m10	tidak	tidak	tidak	tidak
m11	ya	ya	tidak	ya
m12	ya	ya	ya	ya
m13	ya	tidak	tidak	tidak
m14	ya	ya	tidak	ya
m15	ya	ya	ya	ya
m16	tidak	tidak	tidak	tidak
m17	ya	ya	ya	ya
m18	ya	tidak	ya	ya
m19	ya	tidak	tidak	ya
m20	ya	tidak	tidak	ya
m21	ya	ya	ya	ya
m22	ya	ya	ya	ya
m23	tidak	ya	ya	ya
m24	tidak	ya	ya	ya
m25	ya	ya	ya	ya
m26	tidak	ya	tidak	tidak
m27	ya	ya	ya	ya
m28	ya	ya	ya	ya

3.2. Pengumpulan Atribut Menggunakan Jaccard Index

Berdasarkan Tabel 1, maka dikumpulkan atribut bersyarat dan atribut keputusan menggunakan kode mahasiswa seperti yang disajikan di bawah ini:

Langkah 1. Bentuk himpunan-himpunan atribut serta elemennya.

Himpunan Durasi untuk bermain gadget

$$G = \{\text{ya}, \text{tidak}\}$$

$\{\{m3, m5, m6, m7, m8, m9, m11, m12, m13, m14, m15, m18, m19, m20, m21, m22, m25, m27, m28\}, \{m1, m2, m4, m10, m16, m23, m24, m26\}\}$.

Himpunan Jumlah unduhan ilmiah

$$D = \{\text{ya}, \text{tidak}\}$$

$\{\{m3, m4, m5, m6, m9, m11, m12, m14, m15, m17, m21, m22, m23, m24, m25, m26, m27, m28\}, \{m1, m2, m7, m10, m13, m16, m18, m19, m20\}\}$

Himpunan Banyak lembar Al-Quran yang dibaca

$$D = \{\text{ya}, \text{tidak}\}$$

$\{\{m3, m4, m5, m6, m7, m9, m12, m15, m17, m18, m21, m22, m23, m24, m25, m27, m28\}, \{m1, m2, m8, m10, m11, m13, m14, m16, m19, m20, m26\}\}$

Himpunan IPK mahasiswa

$$C = \{\text{ya}, \text{tidak}\}$$

$\{\{m3, m4, m5, m6, m7, m8, m9, m11, m12, m14, m15, m17, m18, m20, m21, m22, m23, m24, m25, m27, m28\}, \{m1, m2, m10, m13, m16, m19, m26\}\}$

Langkah 2. Cari himpunan perpotongan (irisan) antara atribut bersyarat dan atribut keputusan.

Irisan antara himpunan G (ya) dan D (ya)

$$G \cap D = \{m3, m9, m6, m8, m9, m11, m12, m14, m15, m17, m21, m22, m25, m27, m28\}$$

Irisan antara himpunan G (ya), U (ya) dan Q (ya)

$$G \cap D \cap Q = \{m3, m5, m6, m9, m12, m15, m17, m21, m22, m25, m27, m28\}$$

Irisan antara himpunan G (ya), D (ya), Q (ya) dan C (ya)

$$G \cap D \cap Q \cap C = \{m3, m5, m6, m9, m12, m15, m17, m21, m22, m25, m27, 28\}$$

Langkah 3. Hitung indeks kesamaan Jaccard untuk setiap irisan pada Tabel 2.

Tabel 2. Kesamaan dan ketidaksamaan berdasarkan indeks Jaccard

Irisan	Jaccard index
$G \cap D$	$= 15/24$ $= 0.625$ $= 62.5\%$ Jaccard similarity indeks: $J_s(G,D) = 62.5\%$ similar Jaccard dissimilarity indeks: $= 1 - J(G,D),$ $= 1 - 0.625,$ $= 0.375$ $J_d(G,D) = 37.5\%$ dissimilar.

$G \cap D \cap Q$	$= 12/44$ $= 0.273$ $= 27.3\%$ Jaccard similarity index : $J_s(G, D, Q) = 27.3\%$ similar Jaccard dissimilarity index : $= 1 - J(G, D, Q)$ $= 1 - 0.273$ $= 0.727$ $J_d(G, D, Q) = 72.7\%$ dissimilar
$G \cap D \cap Q \cap C$	$= 12/65$ $= 0.185$ $= 18.5\%$ Jaccard similarity index : $J_s(G, D, Q, C) = 18.5\%$ similar Jaccard dissimilarity index : $= 1 - J(G, D, Q, C)$ $= 1 - 0.185$ $= 0.815$ $J_d(G, D, Q, C) = 81.5\%$ dissimilar

Langkah 4. tentukan himpunan mahasiswa yang pasti memiliki $IPK > 3.00$.

Dalam hal ini, kita memperoleh bahwa mahasiswa: m3 m5 m6 m9 m12 m15 m17 m21 m22 m25 m27 m28 adalah kondisi yang baik yang adalah siswa memiliki $IP > 3,00$ berdasarkan kesamaan dan jarak dari $J_s(G, U)$, $J_s(G, U, Q)$ dan $J_s(G, U, Q, C)$.

3.3. Pemilihan Atribut Menggunakan Rough Sets

Rough sets diimplementasikan untuk menentukan ketergantungan antara gejala bersyarat dan atribut keputusan juga menggunakan operasi himpunan dengan mengikuti langkah-langkah berikut:

Langkah 1. Tentukan lower approximation, upper approximation dan daerah batas atau boundary region seperti yang ditunjukkan di bawah ini:

Lower approximations

Para mahasiswa yang pasti memiliki $IPK > 3,00$
 $= \{m3, m4, m5, m6, m7, m8, m9, m11, m12, m14, m15, m17, m18, m21, m22, m23, m24, m25, m27, m28\}$

Para mahasiswa tidak memiliki $IPK > 3.00$
 $= \{m1, m2, m10, m13, m16, m19, m26\}$

Upper approximations

Para mahasiswa yang mungkin memiliki $IPK > 3,00$
 $= \{m3, m4, m5, m6, m7, m8, m9, m11, m12, m14, m15, m17, m18, m21, m22, m23, m24, m25, m27, m28\}$

Para mahasiswa yang mungkin tidak memiliki $IPK > 3,00$
 $= \{m1, m2, m10, m13, m16, m19, m20, m26\}$

-Boundary region untuk pasti memiliki $IPK > 3,00$

$BR = \{m3, m4, m5, m6, m7, m8, m9, m11, m12, m14, m15, m17, m18, m21, m22, m23, m24, m25, m27, m28\} - \{m3, m4, m5, m6, m7, m8, m9, m11, m12, m14, m15, m17, m18, m21, m22, m23, m24, m25, m27, m28\} = \{ \}$

-Boundary region untuk yang mungkin memiliki $IPK > 3.00$

$BR = \{m1, m2, m10, m13, m16, m19, m26\} - \{m1, m2, m10, m13, m16, m19, m20, m26\} = \{m20\}$

Dengan mengikuti semua langkah yang diberikan, dapat disimpulkan dengan menyajikan hasil akhir melalui Tabel 3 di bawah ini:

Tabel 3. Pengelompokkan atribut dan keputusan

Kode	Atribut Kondisi	IPK
	G, U, Q	
m3	ya, ya, ya	3.43
m5	ya, ya, ya	3.33
m6	ya, ya, ya	3.75
m9	ya, ya, ya	3.55
m12	ya, ya, ya	3.58
m15	ya, ya, ya	3.00
m17	ya, ya, ya	3.01
m21	ya, ya, ya	3.41
m22	ya, ya, ya	3.00
m25	ya, ya, ya	3.42
m27	ya, ya, ya	3.13
m28	ya, ya, ya	3.14
m4	tidak, ya, ya	3.38
m23	tidak, ya, ya	3.79
m24	tidak, ya, ya	3.30
m8	ya, ya, tidak	3.21
m11	ya, ya, tidak	3.07
m14	ya, ya, tidak	3.04
m7	ya, ya, tidak	3.70
m18	ya, ya, tidak	3.44
m13	ya, tidak, tidak	2.99
m19	ya, tidak, tidak	2.91
m26	tidak, ya, tidak	2.81
m1	tidak, tidak, tidak	2.83
m2	tidak, tidak, tidak	2.06
m10	tidak, tidak, tidak	2.78
m16	tidak, tidak, tidak	2.82

Berdasarkan Tabel 3, dapat disimpulkan bahwa aturan pendukung keputusan pemilihan atribut data untuk siswa IP mahasiswa dan prediksi nilainya disajikan pada Tabel 4.

Tabel 4. Rule pendukung keputusan IP mahasiswa

Rule	Pernyataan
Rule 1	Jika seorang siswa dengan durasi bermain gadget (G): "ya" dan symptom download ilmiah (U): "ya" dan membaca al-quran (Q): "ya" maka keputusan IPK: "ya" (semua atribut: "ya") $IP > 3.00$

Rule 2	Jika seorang siswa dengan dua gejala: "ya" dan satu gejala: "tidak" maka keputusan IPK: "ya"
Kesimpulan 3	Selain itu, "tidak" IP < 3.00

Berdasarkan aturan-aturan yang sudah diperoleh pada Tabel 4, maka dapat diaplikasikan untuk tujuan prediksi IPK mahasiswa seperti pada Tabel 5.

Tabel 5. Prediksi IP mahasiswa menggunakan aturan-aturan yang disarankan

Kode	Atribut Kondisi			Kesimpulan IPK	Prediksi	
	G	U	Q		Proposed rules	Previous rules
m1	tidak	tidak	tidak	tidak	tidak	tidak
m2	tidak	tidak	tidak	tidak	tidak	tidak
m3	ya	ya	ya	ya	ya	ya
m4	tidak	ya	ya	ya	ya	tidak
m5	ya	ya	ya	ya	ya	ya
m6	ya	ya	ya	ya	ya	ya
m7	ya	tidak	ya	ya	ya	tidak
m8	ya	ya	tidak	ya	ya	tidak
m9	ya	ya	ya	ya	ya	ya
m10	tidak	tidak	tidak	tidak	tidak	tidak
m11	ya	ya	tidak	ya	ya	tidak
m12	ya	ya	yes	ya	ya	ya
m13	ya	tidak	tidak	tidak	tidak	tidak
m14	ya	ya	tidak	ya	ya	tidak
m15	ya	ya	ya	ya	ya	ya
m16	tidak	tidak	tidak	tidak	tidak	tidak
m17	ya	ya	ya	ya	ya	ya
m18	ya	no	ya	ya	ya	tidak
m19	ya	tidak	tidak	tidak	tidak	tidak
m20	ya	tidak	tidak	ya	tidak	tidak
m21	ya	ya	ya	ya	ya	ya
m22	ya	ya	ya	ya	ya	ya
m23	tidak	ya	ya	ya	ya	tidak
m24	tidak	ya	ya	ya	ya	tidak
m25	ya	ya	ya	ya	ya	ya
m26	tidak	ya	tidak	tidak	tidak	tidak
m27	ya	ya	ya	ya	ya	ya
m28	ya	ya	ya	ya	ya	ya

Tabel 5 menunjukkan implementasi aturan yang diusulkan dalam selection atribut untuk IP mahasiswa matematika dan hasil perbandingan dengan aturan yang diusulkan dalam tabel. aturan yang diusulkan mampu memprediksi atribut keputusan dengan tepat, kecuali mahasiswa m20. Sementara, aturan sebelumnya memprediksi pasien m3 m5 m6 m9 m12 m15 m17 m21 m22 m25 m27 m28 s4 m23 s24 m8 m11 m14 m7 m18. Dalam hal ini, penyempurnaan aturan sangat penting untuk dipertimbangkan guna meningkatkan akurasi penentuan.

4. Kesimpulan

Dalam tulisan ilmiah ini, diterapkan indeks Jaccard dan rough sets yang ditetapkan oleh pemilihan atribut untuk menentukan IP mahasiswa Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau. Indeks Jaccard dapat digunakan untuk menentukan kondisi terbaik IP mahasiswa berdasarkan kesamaan dan jarak antara atribut bersyarat dan atribut keputusan. Sementara itu, pendekatan *rough sets* dapat diterapkan untuk menentukan reduksi data dan aturan pendukung keputusan. Jadi, dapat disimpulkan, banyak bacaan Al-

Quran, jumlah unduhan ilmiah dan durasi memainkan gadget sedikit banyak memang berpengaruh pada IP mahasiswa. Mahasiswa yang mendapatkan IP lebih kecil dari 3,00 dapat menggunakan hasil penelitian ini untuk memperbaiki pengetahuan diri mereka sendiri dan juga dapat mempertimbangkan atribut-atribut yang sudah dikaji sebelumnya [5-9] untuk meningkatkan prestasi akademik pada semester berikutnya.

Referensi

- [1] Rissino, S., Torres, G. L. Rough set theory-fundamental concepts, principals, data extraction, and applications. Julio Ponce and Adam Karahoca (Ed), *Data Mining and Knowledge Discovery in Real Life App. Inform.* 2009: 35-58.
- [2] Pawlak Z. Rough Sets. *International Journal of Computer and Information Sciences.* 1982; 11: 341-355.
- [3] Hartama. D dan Hartono. Analisis Kinerja Dosen STMIK IBBI Dengan Menggunakan Metode Rough Sets. *Jurnal Nasional Teknologi Informasi dan Multimedia.* 2016; ISSN : 2302-3805.
- [4] Efendi. R., Samsudin. N. A., Deris. M. M., dan Ting, Y.G. Flu Diagnosis System Using Jaccard Index and Rough Set Approaches. *Journal of Physics: Conference Series.* 2018; 1004(1).
- [5] Martirosyan, N. M., Hwang, E dan R. Wanjohi, R. "Impact of english proficiency on academic performance of International students. *Journal of International Students.* 5; 2015: 60-75.
- [6] Thuseethan dan Kuhanesan. Influence Facebook in academic perfromance of Sri Lankan university students. *Global Journal of Comp. Science and Tech., Network, Web and Security.* 4; 2014: 29-35.
- [7] Junco, R. Student class standing, facebook use, and academic performance. *Journal of Applied Developmental Psychology.* 36; 2015: 18-29.
- [8] F. Kobayashi, "Academic achievement, BMI, and fast food intake of American and Japanese college students," *Nutrition and Food Science*, vol. 39, pp. 555-566, 2009.
- [9] A. Ross, "Nutrition and its effects on academic performance How can our schools improve?," Master Thesis of Art Education, Northen Michigan University, pp. 1-58, 2010.