

Analisis Klasifikasi *Correlated Naïve Bayes* Untuk Pemilihan Konsentrasi Program Studi RPL pada Mahasiswa Universitas Dipa Makassar

Jusna¹, Nur Amri Angriadi², Annah³, Nurul Aini⁴, Nurdiansah⁵, Asrul Syam⁶

^{1,2} Jurusan Sistem Informasi Universitas Dipa Makassar
Jln. Perintis Kemerdekaan KM. 9 Makassar

¹ jusna164483@gmail.com, ² nuramriangriadi@gmail.com, ³ annah@undipa.ac.id, ⁴ nurulaini.m11@undipa.ac.id,

⁵ nurdiansah@undipa.ac.id, ⁶ asrulsyam12@undipa.ac.id

Abstrak

Universitas Dipa Makassar merupakan sebuah perguruan tinggi komputer yang menerapkan bahwa mahasiswa pada semester lima diharuskan memilih konsentrasi studi. Dimana terdapat enam jurusan yaitu Sistem Informasi, Teknik Informatika, Rekayasa Perangkat Lunak (RPL), Manajemen Informatika, Bisnis Digital, dan Kewirausahaan. Akan tetapi dari ke enam jurusan tersebut, peneliti berfokus pada jurusan RPL. Mahasiswa pada saat menentukan konsentrasi studi hanya memilih berdasarkan minat tanpa mempertimbangkan aspek pendukung seperti beberapa matakuliah yang perlu diperhatikan dalam memilih konsentrasi tersebut, sehingga nantinya konsentrasi yang telah dipilih oleh mahasiswa tidak sesuai dengan kemampuan dirinya. Untuk itu peneliti melakukan analisis menggunakan metode *Correlated Naïve Bayes*, dimana probabilitas yang dihasilkan oleh *Cloud Software Engineer* sebesar 0,87037037, probabilitas yang dihasilkan oleh *Game Developer* sebesar 0,12962963. Adapun nilai akurasi yang dihasilkan sebesar 70,37%. Pada pengujian *k-fold cross validation* dimana *k* yang digunakan sebanyak 5 menghasilkan akurasi sebesar 51,64%.

Kata kunci: *Correlated Naïve Bayes*, Data mining, Pemilihan Konsentrasi Mata Kuliah.

Abstract

Dipa Makassar University is a computer college that implements that students in the fifth semester must choose a study concentration. There are six majors: Information Systems, Informatics Engineering, Software Engineering, Informatics Management, Digital Business, and Entrepreneurship. However, of the six majors, researchers focused on the Software Engineering department. When determining a study concentration, students only choose based on their interests without considering supporting aspects such as several subjects that need to be considered in selecting the concentration, so that later the concentration chosen by students is not by their abilities. For this reason, researchers conducted an analysis using the *Correlated Naïve Bayes* method, where the probability generated by the *Cloud Software Engineer* was 0.87037037, and the probability generated by the *Game Developer* was 0.12962963. The resulting accuracy value is 70.37%. In the *k-fold cross-validation* test where *k* is used as much as 5, it produces an accuracy of 51.64%

Keywords: *Correlated Naïve Bayes*, Data mining, Choice of subject concentration.

I. PENDAHULUAN

A. Latar Belakang

Universitas Dipa Makassar merupakan sebuah perguruan tinggi komputer yang menerapkan bahwa mahasiswa pada semester lima diharuskan memilih konsentrasi studi. Dimana terdapat enam jurusan yaitu Sistem Informasi, Teknik Informatika, Rekayasa Perangkat Lunak, Manajemen Informatika, Bisnis Digital, dan Kewirausahaan. Penulis berfokus pada jurusan RPL yang baru berdiri pada tahun 2019. Jurusan RPL memiliki dua pilihan konsentrasi yaitu *Game Developer* dan *Cloud Software Engineer*, dimana jurusan tersebut merupakan satu-satunya jurusan yang belum

mengalami perubahan konsentrasi dan terdapat dua angkatan yang telah memilih konsentrasi tersebut yaitu angkatan 2019 dan angkatan 2020 dengan data sebanyak 77 data sehingga hal tersebut menjadi dasar peneliti memilih jurusan RPL sebagai bahan penelitian.

Untuk itu penulis akan menganalisis pemilihan konsentrasi pada jurusan RPL dengan menggunakan metode *Correlated Naïve Bayes*. Adapun teknik yang digunakan untuk mendapatkan informasi dengan cara mencari pola atau aturan tertentu dari sejumlah data yang sangat besar yang disebut dengan istilah data mining dengan tools yaitu *RapidMiner*.

B. Tinjauan Pustaka

1) *Konsentrasi*: Konsentrasi atau peminatan adalah pengkhususan studi yang diambil dalam sebuah jurusan atau program studi. Jadi, yang dipelajari lebih spesifik dan terarah ke suatu bidang. Misalnya, Pada Jurusan RPL yaitu konsentrasi *Game Developer* dan *Cloud Software Engineer* [1].

2) *Data Mining*: Data mining merupakan sebuah hakikat sebagai disiplin ilmu yang tujuannya adalah untuk menemukan, menggali ataupun menambang pengetahuan dari data atau informasi yang kita miliki, hal inilah yang menjadi perhatian utama dari suatu data mining [2].

Secara sederhana, data mining atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya [3].

3) *Algoritma Naïve Bayes*: Metode *Naïve Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (Pattern Recognition). Pendalaman dalam proses klasifikasi dengan menggunakan probabilitas dan biaya yang ditimbulkan dalam keputusan keputusan tersebut [4].

Berikut adalah rumus persamaan teorema Bayes:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \tag{1}$$

Penjelasan dari persamaan (1) sebagai berikut:

- X : Data dengan class yang belum diketahui.
- H : Hipotesis pada data X yang merupakan suatu class khusus.
- P(H|X) : Nilai probabilitas pada hipotesis H berdasarkan kondisi X.
- P(H) : Nilai probabilitas pada hipotesis H.
- P(X|H) : Nilai probabilitas X yang berdasarkan dengan kondisi H.
- P(X) : Nilai probabilitas pada X.

Untuk menjelaskan teorema Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \tag{2}$$

Penjelasan dari persamaan (2) sebagai berikut:

- P : Probabilitas.
- C : Kelas.
- F1...Fn : Karakteristik data yang dibutuhkan.

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan (C,F1, ... , Fn) menggunakan aturan perkalian sebagai berikut [5]:

$$P(c|F) = P(f_1|c).P(f_2|c).P(f_3|c) \dots P(f_n|c).P(c) \tag{3}$$

Persamaan di atas merupakan model dari teorema Naïve Bayes yang selanjutnya akan digunakan dalam proses klasifikasi.

4) *Correlated Naïve Bayes Classifier* : *Correlated-Naïve Bayes Classifier* adalah modifikasi dan pengembangan dari algoritma *Naïve Bayes Classifier*. Metode klasifikasi dengan menggunakan algoritma *Naïve Bayes Classifier* berdasar pada nilai *probability attribute* dari suatu data yang belum diketahui *class*-nya. Sehingga teknik dasar dari klasifikasi algoritma ini berdasar pada *frequency* kemunculan data pada data set. Selain berdasar pada *probability* atau tingkat kemunculan data pada data set, *Correlated-Naïve Bayes Classifier*, juga melibatkan perhitungan terhadap nilai korelasi masing-masing *attribute* terhadap *class* [6].

Formula korelasi :

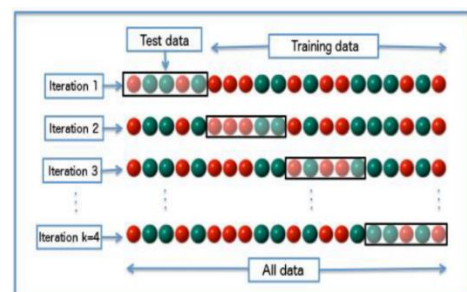
$$r = \frac{n \times (\sum XY) - (\sum X) \times (\sum Y)}{\sqrt{(n \times \sum X^2 - (\sum X)^2)(n \times \sum Y^2 - (\sum Y)^2)}} \tag{4}$$

Penjelasan dari persamaan (4) sebagai berikut:

- r : Korelasi
- Σ : Sigma
- n : Jumlah data
- X : Variabel bebas
- Y : Variabel terikat

5) *K-Fold Cross Validation* : Validasi merupakan proses pengujian performa algoritma. Pada umumnya validasi dilakukan dengan mengulang proses perhitungan sampai beberapa kali. Proses validasi dalam penelitian ini menggunakan *Cross validation*. *Cross Validation* adalah membagi dataset menjadi dua bagian dengan satu bagian dijadikan data *training* dan bagian yang lain dijadikan data *testing* [7].

Langkah pertama dalam melakukan *cross validation* adalah melakukan iterasi sebanyak k, dimana selama iterasi berlangsung data *testing* dan data *training* tidak pernah sama karena data *testing* pada iterasi sebelumnya menjadi data *training* dan pada iterasi selanjutnya diambil beberapa data dari data *training* di iterasi sebelumnya untuk dijadikan sebagai dataset. Gambar 1 merupakan contoh diagram *K-fold cross validation* dimana k=4 [8].



Gambar 1. Contoh Diagram *K-Fold Cross Validation*

No	Stambok	Nama Mahasiswa	Nilai Matakuliah	Pilihan Konsentrasi studi
1	195002	IQBAL SANDY SERU	61	Cloud Software Engineer
2	195003	DHEA ALFIKA RAMADANTY	61	Cloud Software Engineer
3	195004	MUHAMMAD NUR	72	Cloud Software Engineer
4	195005	A. RIFAL FAJAR M	72	Cloud Software Engineer
5	195006	RAYMOND MARTHEN LAISINA	61	Cloud Software Engineer
6	195007	SINTA	72	Cloud Software Engineer
7	195008	NGURAH PUTU SIWANANDA KUSUMA	72	Game Developer
8	195011	BRIYAH HENOKH LEBANG	61	Cloud Software Engineer
9	195012	ANDI MUH BATARA HASMIR	61	Cloud Software Engineer
10	195013	MUH. ARIEF	82	Cloud Software Engineer
11	195014	ANDI MUHAMMAD RAFLI NURYADIN ZAINUDDIN	61	Cloud Software Engineer
12	195015	MUH. ARIEF NOVIAN	61	Cloud Software Engineer
13	195016	DHAFA ALIEFANDI	61	Cloud Software Engineer
14	195017	OEMMOE CHALSUM MIFTAHUL JANNAH	61	Game Developer
15	195018	KHUSNUL HATIMA BASRI	82	Cloud Software Engineer
16	195019	MUHAMMAD YUSRI MASRI	61	Cloud Software Engineer
17	195021	YOHANES BAPTISTA PANDA	45	Cloud Software Engineer
18	195022	NURFADILA SIRAMAN	82	Cloud Software Engineer
19	195026	LULUD ANGGRAINI	82	Cloud Software Engineer
20	195027	FIDYHA FADHILHA	82	Cloud Software Engineer
21	195028	ARDIAN ARISANDY	45	Cloud Software Engineer
22	195029	MARSELINUS TURU PADANG	61	Game Developer
23	195030	MUHAMMAD RIZKY BURNABY	82	Game Developer
24	195032	MUH. RANA ALFATRA	72	Cloud Software Engineer
25	195033	JHORVY	61	Cloud Software Engineer
26	195038	MUHAMMAD FARHAN HAFID	61	Cloud Software Engineer
27	195039	ISWANDI	72	Cloud Software Engineer
28	195040	TRY MUNAZZAR ABDUH	61	Cloud Software Engineer
29	195041	MUH FITRA RAMADHAN	61	Cloud Software Engineer
30	205001	WIDIO PHARDANU NUR	81	Cloud Software Engineer
31	205002	SUNAR AMINULLAH	81	Cloud Software Engineer
32	205003	IRSYAD ALFIANSYAH	81	Cloud Software Engineer
33	205004	YEVGENY RINGGI ALLO	81	Game Developer
34	205005	M. AQSHAL ANUGRAH RUSLI	81	Cloud Software Engineer
35	205007	MUJAHID DZAKY	81	Cloud Software Engineer
36	205009	MUH. RAFU H. UMAR	81	Cloud Software Engineer
37	205010	MICHAEL VINCENT EFREN M	81	Cloud Software Engineer
38	205011	SALMANTO MANDALA PUTRA	62	Cloud Software Engineer
39	205013	BAGOGES RIZKY ADI WIBOWO	81	Cloud Software Engineer
40	205014	ACHMAD FACHRUL AZIS	81	Game Developer
41	205018	DENILS FLORENTINO RADHO	81	Game Developer
42	205018	ANDI MUHAMMAD ASKUR	75	Cloud Software Engineer
43	205021	VIKAR TA'DIAMPANG	62	Cloud Software Engineer
44	205022	ADITYA SAPUTRA	81	Cloud Software Engineer
45	205023	ANDI MUHAMMAD FAHRUL ISMAIL	81	Cloud Software Engineer
46	205025	RAHMAT DWI JAYITA NUR	81	Cloud Software Engineer
47	205027	NANDA PUTRA PRABU KELANA	81	Cloud Software Engineer
48	205028	A. MARILZA YASMIN	81	Cloud Software Engineer
49	205029	WILANDARI RANTETANDUNG	75	Cloud Software Engineer
50	205033	MAKKULAU NASIR	81	Cloud Software Engineer
51	205035	IQBAL	75	Cloud Software Engineer

Tabel 3. Pengklasifikasian Matakuliah ALGO

No	Stambok	Nama Mahasiswa	Nilai Matakuliah	Pilihan Konsentrasi studi
1	195002	IQBAL SANDY SERU	61	Cloud Software Engineer
2	195003	DHEA ALFIKA RAMADANTY	61	Cloud Software Engineer
3	195004	MUHAMMAD NUR	72	Cloud Software Engineer
4	195005	A. RIFAL FAJAR M	72	Cloud Software Engineer
5	195006	RAYMOND MARTHEN LAISINA	61	Cloud Software Engineer
6	195007	SINTA	72	Cloud Software Engineer
7	195008	NGURAH PUTU SIWANANDA KUSUMA	72	Game Developer
8	195011	BRIYAH HENOKH LEBANG	61	Cloud Software Engineer
9	195012	ANDI MUH BATARA HASMIR	61	Cloud Software Engineer
10	195013	MUH. ARIEF	82	Cloud Software Engineer
11	195014	ANDI MUHAMMAD RAFLI NURYADIN ZAINUDDIN	61	Cloud Software Engineer
12	195015	MUH. ARIEF NOVIAN	61	Cloud Software Engineer
13	195016	DHAFA ALIEFANDI	61	Cloud Software Engineer
14	195017	OEMMOE CHALSUM MIFTAHUL JANNAH	61	Game Developer
15	195018	KHUSNUL HATIMA BASRI	82	Cloud Software Engineer
16	195020	MUH. TAUFIK HIDAYAT	82	Cloud Software Engineer
17	195021	RAHMANZAH IBRAHIM	61	Cloud Software Engineer
18	195023	MUHAMMAD YUSRI MASRI	61	Cloud Software Engineer
19	195024	YOHANES BAPTISTA PANDA	45	Cloud Software Engineer
20	195025	NURFADILA SIRAMAN	82	Cloud Software Engineer
21	195026	LULUD ANGGRAINI	82	Cloud Software Engineer
22	195027	FIDYHA FADHILHA	82	Cloud Software Engineer
23	195028	ARDIAN ARISANDY	45	Cloud Software Engineer
24	195029	MARSELINUS TURU PADANG	61	Game Developer
25	195030	MUHAMMAD RIZKY BURNABY	82	Game Developer
26	195032	MUH. RANA ALFATRA	72	Cloud Software Engineer
27	195033	JHORVY	61	Cloud Software Engineer
28	195035	IBRAL MAULANA	61	Cloud Software Engineer
29	195038	MUHAMMAD FARHAN HAFID	61	Cloud Software Engineer
30	195039	ISWANDI	72	Cloud Software Engineer
31	195040	TRY MUNAZZAR ABDUH	61	Cloud Software Engineer
32	195041	MUH FITRA RAMADHAN	61	Cloud Software Engineer
33	205001	WIDIO PHARDANU NUR	81	Cloud Software Engineer
34	205002	SUNAR AMINULLAH	81	Cloud Software Engineer
35	205003	IRSYAD ALFIANSYAH	81	Cloud Software Engineer
36	205004	YEVGENY RINGGI ALLO	81	Game Developer
37	205005	M. AQSHAL ANUGRAH RUSLI	81	Cloud Software Engineer
38	205007	MUJAHID DZAKY	81	Cloud Software Engineer
39	205009	MUH. RAFU H. UMAR	81	Cloud Software Engineer
40	205010	MICHAEL VINCENT EFREN M	81	Cloud Software Engineer
41	205011	SALMANTO MANDALA PUTRA	62	Cloud Software Engineer
42	205013	BAGOGES RIZKY ADI WIBOWO	81	Cloud Software Engineer
43	205014	ACHMAD FACHRUL AZIS	81	Game Developer
44	205018	DENILS FLORENTINO RADHO	81	Game Developer
45	205018	ANDI MUHAMMAD ASKUR	75	Cloud Software Engineer
46	205021	VIKAR TA'DIAMPANG	62	Cloud Software Engineer
47	205022	ADITYA SAPUTRA	81	Cloud Software Engineer
48	205023	ANDI MUHAMMAD FAHRUL ISMAIL	81	Cloud Software Engineer
49	205025	RAHMAT DWI JAYITA NUR	81	Cloud Software Engineer
50	205027	NANDA PUTRA PRABU KELANA	81	Cloud Software Engineer
51	205028	A. MARILZA YASMIN	81	Cloud Software Engineer
52	205029	WILANDARI RANTETANDUNG	75	Cloud Software Engineer
53	205033	MAKKULAU NASIR	81	Cloud Software Engineer
54	205035	IQBAL	75	Cloud Software Engineer

Adapun data *testing* yang digunakan dalam penelitian ini sebanyak 5 mahasiswa diambil secara acak dari data *training*

untuk menguji kebenaran dalam memilih konsentrasi memilih konsentrasi. Data *testing* disajikan pada tabel 4.

Tabel 4. Data Testing

NO	NAMA	MK1	MK2	MK3	MK4	MK5	MK6	MK7	Konsentrasi
1	Ngurah	72	55	58	81	76	88	89	?
2	M.Arif	82	85	91	82	85	84	82	?
3	Khusnul	82	82	95	81	89	85	83	?
4	Yevgeny	81	77	88	90	82	67	82	?
5	Muh.Taufik	82	85	91	82	78	85	81	?

III. HASIL DAN PEMBAHASAN

A. Perhitungan Menggunakan Algoritma Correlated Naive Bayes

Tabel 5. Perhitungan Korelasi ALGO

No	Stambok	Nama Mahasiswa	Algoritma dan Pemrograman	Pilihan Konsentrasi studi	X	Y	XY	X ²	Y ²
1	195002	IQBAL SANDY SERU	61	Cloud Software Engineer	61	2	122	3721	4
2	195003	DHEA ALFIKA RAMADANTY	61	Cloud Software Engineer	61	2	122	3721	4
3	195004	MUHAMMAD NUR	72	Cloud Software Engineer	72	2	144	5184	4
4	195005	A. RIFAL FAJAR M	72	Cloud Software Engineer	72	2	144	5184	4
5	195006	RAYMOND MARTHEN LAISINA	61	Cloud Software Engineer	61	2	122	3721	4
6	195007	SINTA	72	Cloud Software Engineer	72	2	144	5184	4
7	195008	NGURAH PUTU SIWANANDA KUSUMA	72	Game Developer	72	1	72	5184	1
8	195011	BRIYAH HENOKH LEBANG	61	Cloud Software Engineer	61	2	122	3721	4
9	195012	ANDI MUH BATARA HASMIR	61	Cloud Software Engineer	61	2	122	3721	4
10	195013	MUH. ARIEF	82	Cloud Software Engineer	82	2	164	6724	4
11	195014	ANDI MUHAMMAD RAFLI NURYADIN ZAINUDDIN	61	Cloud Software Engineer	61	2	122	3721	4
12	195015	MUH. ARIEF NOVIAN	61	Cloud Software Engineer	61	2	122	3721	4
13	195016	DHAFA ALIEFANDI	61	Cloud Software Engineer	61	2	122	3721	4
14	195017	OEMMOE CHALSUM MIFTAHUL JANNAH	61	Game Developer	61	1	61	3721	1
15	195018	KHUSNUL HATIMA BASRI	82	Cloud Software Engineer	82	2	164	6724	4
16	195019	MUHAMMAD YUSRI MASRI	61	Cloud Software Engineer	61	2	122	3721	4
17	195021	YOHANES BAPTISTA PANDA	45	Cloud Software Engineer	45	2	90	2025	4
18	195022	NURFADILA SIRAMAN	82	Cloud Software Engineer	82	2	164	6724	4
19	195026	LULUD ANGGRAINI	82	Cloud Software Engineer	82	2	164	6724	4
20	195027	FIDYHA FADHILHA	82	Cloud Software Engineer	82	2	164	6724	4
21	195028	ARDIAN ARISANDY	45	Cloud Software Engineer	45	2	90	2025	4
22	195029	MARSELINUS TURU PADANG	61	Game Developer	61	1	61	3721	1
23	195030	MUHAMMAD RIZKY BURNABY	82	Game Developer	82	1	82	6724	1
24	195032	MUH. RANA ALFATRA	72	Cloud Software Engineer	72	2	144	5184	4
25	195033	JHORVY	61	Cloud Software Engineer	61	2	122	3721	4
26	195038	MUHAMMAD FARHAN HAFID	61	Cloud Software Engineer	61	2	122	3721	4
27	195039	ISWANDI	72	Cloud Software Engineer	72	2	144	5184	4
28	195040	TRY MUNAZZAR ABDUH	61	Cloud Software Engineer	61	2	122	3721	4
29	195041	MUH FITRA RAMADHAN	61	Cloud Software Engineer	61	2	122	3721	4
30	205001	WIDIO PHARDANU NUR	81	Cloud Software Engineer	81	2	162	6561	4
31	205002	SUNAR AMINULLAH	81	Cloud Software Engineer	81	2	162	6561	4
32	205003	IRSYAD ALFIANSYAH	81	Cloud Software Engineer	81	2	162	6561	4
33	205004	YEVGENY RINGGI ALLO	81	Game Developer	81	1	81	6561	1
34	205005	M. AQSHAL ANUGRAH RUSLI	81	Cloud Software Engineer	81	2	162	6561	4
35	205007	MUJAHID DZAKY	81	Cloud Software Engineer	81	2	162	6561	4
36	205009	MUH. RAFU H. UMAR	81	Cloud Software Engineer	81	2	162	6561	4
37	205010	MICHAEL VINCENT EFREN M	81	Cloud Software Engineer	81	2	162	6561	4
38	205011	SALMANTO MANDALA PUTRA	62	Cloud Software Engineer	62	2	124	3844	4
39	205013	BAGOGES RIZKY ADI WIBOWO	81	Cloud Software Engineer	81	2	162	6561	4
40	205014	ACHMAD FACHRUL AZIS	81	Game Developer	81	1	81	6561	1
41	205018	DENILS FLORENTINO RADHO	81	Game Developer	81	1	81	6561	1
42	205018	ANDI MUHAMMAD ASKUR	75	Cloud Software Engineer	75	2	150	5625	4
43	205021	VIKAR TA'DIAMPANG	62	Cloud Software Engineer	62	2	124	3844	4
44	205022	ADITYA SAPUTRA	81	Cloud Software Engineer	81	2	162	6561	4
45	205023	ANDI MUHAMMAD FAHRUL ISMAIL	81	Cloud Software Engineer	81	2	162	6561	4
46	205025	RAHMAT DWI JAYITA NUR	81	Cloud Software Engineer	81	2	162	6561	4
47	205027	NANDA PUTRA PRABU KELANA	81	Cloud Software Engineer	81	2	162	6561	4
48	205028	A. MARILZA YASMIN	81	Cloud Software Engineer	81	2	162	6561	4
49	205029	WILANDARI RANTETANDUNG	75	Cloud Software Engineer	75	2	150	5625	4
50	205033	MAKKULAU NASIR	81	Cloud Software Engineer	81	2	162	6561	4
51	205035	IQBAL	75	Cloud Software Engineer	75	2	150	5625	4

$$r = \frac{n \sum XY - (\sum X) (\sum Y)}{\sqrt{(n \sum X^2 - (\sum X)^2) (n \sum Y^2 - (\sum Y)^2)}}$$

$$r = \frac{54 * (7199) - (3859) * (101)}{\sqrt{(54 * 281579 - (3859)^2) (54 * 195 - (101)^2)}}$$

$$r = \frac{388746 - 389759}{\sqrt{(15205266 - 14891881) (1030 - 10201)}}$$

$$r = \frac{-1013}{\sqrt{313385 * 329}}$$

$$r = \frac{-1013}{559,808 * 18,13836}$$

$$r = \frac{-1013}{10154}$$

r Algoritma dan Pemrograman(X1) = -0,00976
 Dari nilai r Algoritma dan Pemrograman(X1) diatas kita jadikan nilai korelasi R-Square dengan Rumus:
 $R\ Square = r^2$
 $R\ Square = -0,00976^2$
 $R\ Square (Algoritma dan Pemrograman) = 0,009953$
 Kita dapatkan nilai Korelasi R Square Algoritma dan Pemrograman (X1) sebesar 0,009953.

Hasil dari perhitungan *Correlated Naive Bayes* untuk 7 matakuliah dengan pemilihan konsentrasi, sehingga dapat dilihat pada tabel 6 akurasi yang dihasilkan oleh variabel matakuliah dalam penelitian ini.

$$58) * R(\text{Game Developer} | X3 = 58) + P(\text{Game Developer} | X4 = 81) * R(\text{Game Developer} | X4 = 81) + P(\text{Game Developer} | X5 = 76) * R(\text{Game Developer} | X5 = 76) + P(\text{Game Developer} | X6 = 88) * R(\text{Game Developer} | X6 = 88) + P(\text{Game Developer} | X7 = 81) * R(\text{Game Developer} | X7 = 81) * P(X | \text{Game Developer}) = (9,611111111 * 0,009953) + (7,925925926 * 0,002844) + (8,166666667 * 0,000153) + (10,37037037 * 0,054278) + (9,314814815 * 0,00216227) + (10,07407407 * 0,031223) + (9,574074074 * 0,00545) * 0,12962963 = 1,023781051$$

$$P(\text{Cloud Software Engineer} | X) = P(\text{Cloud Software Engineer} | X1 = 72) * R(\text{Cloud Software Engineer} | X1 = 72) + P(\text{Cloud Software Engineer} | X2 = 55) * R(\text{Cloud Software Engineer} | X2 = 55) + P(\text{Cloud Software Engineer} | X3 = 58) * R(\text{Cloud Software Engineer} | X3 = 58) + P(\text{Cloud Software Engineer} | X4 = 81) * R(\text{Cloud Software Engineer} | X4 = 81) + P(\text{Cloud Software Engineer} | X5 = 76) * R(\text{Cloud Software Engineer} | X5 = 76) + P(\text{Cloud Software Engineer} | X6 = 88) * R(\text{Cloud Software Engineer} | X6 = 88) + P(\text{Cloud Software Engineer} | X7 = 81) * R(\text{Cloud Software Engineer} | X7 = 81) * P(X | \text{Cloud Software Engineer}) = (61,85185185 * 0,009953) + (50,64814815 * 0,002844) + (58,18518519 * 0,000153) + (57,24074074 * 0,054278) + (49,59259259 * 0,00216227) + (57,55555556 * 0,031223) + (60,59259259 * 0,00545) * 0,87037037 = 6,067181845$$

Menghitung nilai max dari probabilitas akhir dari data testing 1

$$P(\text{Game Developer} | X) = 1,023781051$$

$$P(\text{Cloud Software Engineer} | X) = 6,067181845$$

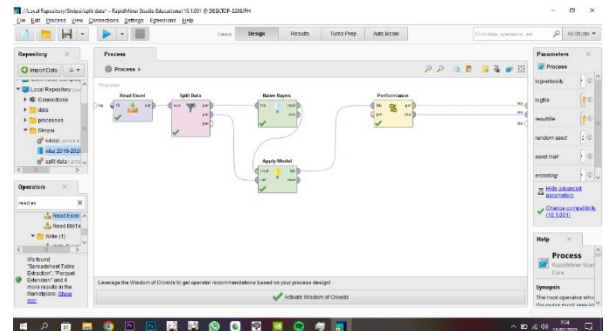
Nilai probabilitas kelas *Cloud Software Engineer* lebih besar dari nilai probabilitas kelas *Game Developer*, sehingga data tersebut diklasifikasikan ke dalam kelas *Cloud Software Engineer*. Untuk itu data testing 1 tidak sesuai dengan data training tersebut.

E. Implementasi dengan RapidMiner

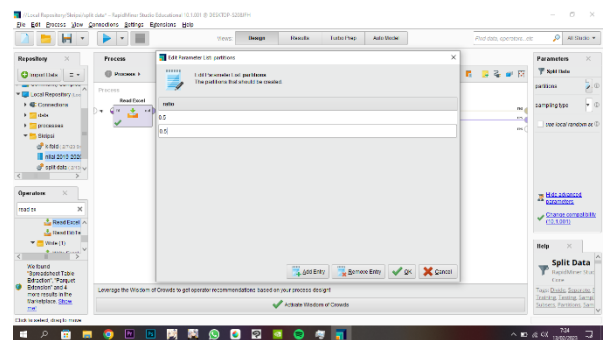
Data yang akan diolah berupa dataset sebanyak 54 data dan menggunakan variabel nilai mata kuliah. Pada variabel pemilihan konsentrasi merupakan variabel penentu atau output dari perhitungan tersebut. Berikut beberapa sampel data yang akan ditampilkan dan selanjutnya akan diuji menggunakan Aplikasi *RapidMiner*. Pada Gambar 3 menunjukkan dataset mahasiswa.

Gambar 3. Dataset Mahasiswa

Pada proses pengujian menggunakan menggunakan operator split data untuk membagi dataset menjadi partisi data *training* dan data *testing* sesuai porsi yang ditentukan. Pada gambar 4 menunjukkan model naïve bayes dan gambar 5 menunjukkan rasio partisi data *training* dan data *testing*.



Gambar 4. Model Naïve Bayes



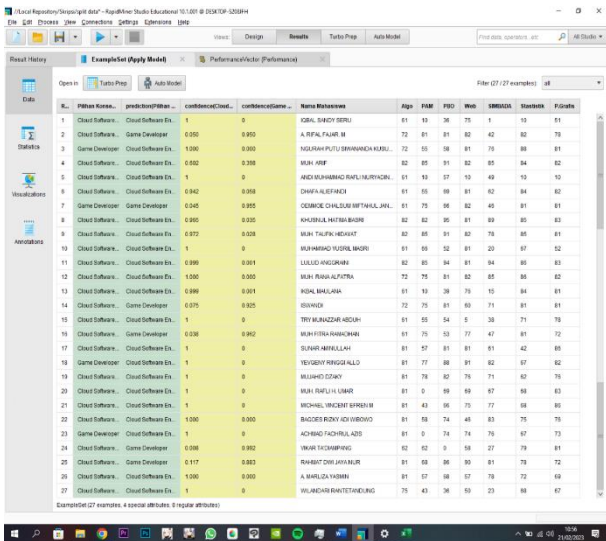
Gambar 5. rasio partisi data training dan data testing

Hasil dari nilai parameter yang diinputkan oleh pengguna (*User*) sistem dengan menggunakan variabel nilai matakuliah dan pilihan konsentrasi. Dengan data *training* sebanyak 50% dari data dan data *testing* sebesar 50% dari data. Berikut hasil dari dataset mahasiswa dengan 2 pilihan klasifikasi *Cloud Software Engineer* dan *Game Developer*. Pada Gambar 6 menunjukkan hasil nilai akurasi sebesar 70,37%

	Real Cloud Software Engineer	Real Game Developer	Class Precision
pred Cloud Software Engineer	5	3	62.5%
pred Game Developer	5	1	16.67%
class total	10.0%	4.0%	

Gambar 6. Hasil Nilai Akurasi

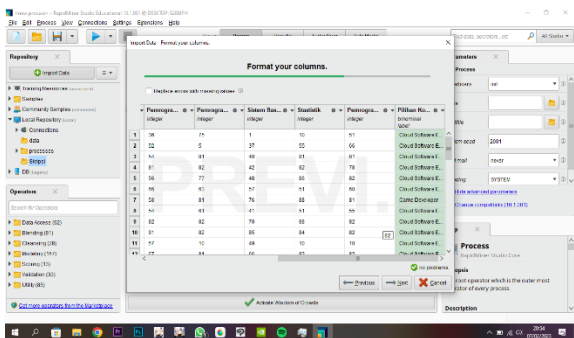
Hasil dari prediksi pemilihan konsentrasi dimana data dibagi menjadi 50% data *training* dan 50% data *testing*. Pada gambar 7 menunjukkan data yang telah diuji.



Gambar 7. Hasil Prediksi Pemilihan Konsentrasi

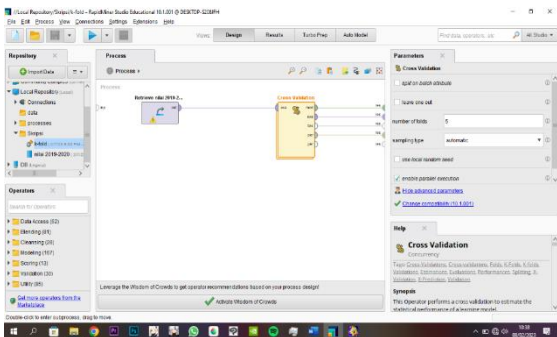
F. Pengujian K-Fold Cross Validaton

Pengujian *K-Fold Cross Validation* yang digunakan untuk mengetahui akurasi yang dihasilkan dari dataset nilai mahasiswa dan pemilihan konsentrasi, adapun k yang digunakan sebanyak k = 5. Pada gambar 8 menunjukkan dataset dari penelitian kami.

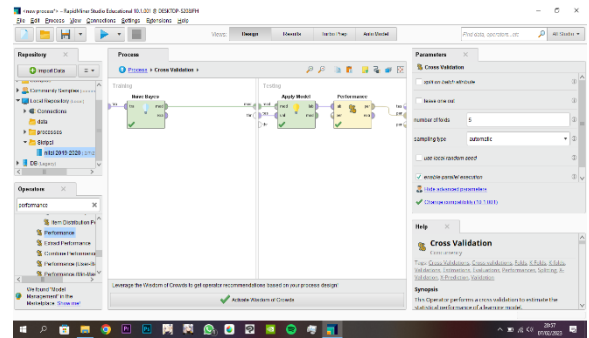


Gambar 8. Dataset Pengujian K-Fold Cross Validation

Pada pengujian *K-Fold Cross Validation* dimana k = 5 dapat dilihat hasil data *training* dari sampel yang peneliti gunakan. Pada gambar 9 dan 10 data *training* dan data *testing* dengan pengujian *K-Fold Cross Validation*.

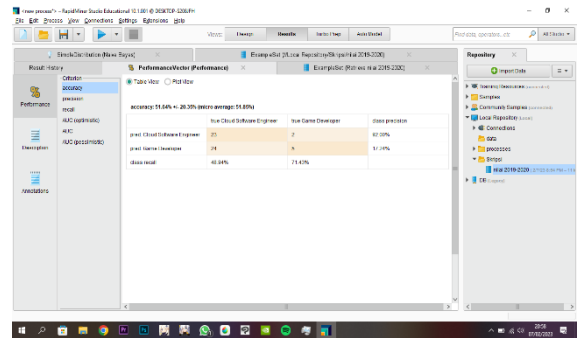


Gambar 9. Uji Validasi K-Fold Cross Validation



Gambar 10. Data Training Dan Data Testing Dengan Pengujian K-Fold

Hasil dari pengujian *K-Fold Cross Validation* dengan dataset yang diolah menghasilkan nilai akurasi sebesar 51,64%. Pada gambar 11 menunjukkan hasil nilai akurasi dengan uji *K-Fold Cross Validation*.



Gambar 11. Hasil Nilai Akurasi Uji K-Fold Cross Validation

Berdasarkan hasil analisis *K-Fold Cross Validation* di atas dengan dataset sebanyak 54 data dan menggunakan k sebanyak 5, menghasilkan tingkat hubungan cukup. Interval koefisien dan tingkat hubungan dapat dilihat pada tabel koefisien korelasi pada tabel 14.

Tabel 14. Koefisien Korelasi

Interval koefisien	Tingkat Hubungan
0 – 0.199	Sangat rendah
0.20 – 0.299	Rendah
0.4 – 0.599	Cukup
0.6 – 0.799	Kuat
0.8 – 1	Sangat kuat

IV. KESIMPULAN

Dalam pengujian *Correlated Naïve Bayes* yang dimana memiliki data *training* dan data *testing*, untuk data *training* berjumlah 54 data dan data *testing* yang kami gunakan sebanyak 5 dari data *training* diambil secara acak. Setelah melakukan pengujian dengan 5 data *testing* untuk mahasiswa Ngurah, M. Arif, Khunsul, Yevgeny dan Muh.Taufik dapat disimpulkan bahwa Hasil akhir nilai probabilitas kelas *Cloud Software Engineer lebih besar* dari nilai probabilitas kelas *Game Developer*, sehingga data tersebut diklasifikasikan ke

dalam kelas *Cloud Software Engineer*. Untuk itu yang sesuai dengan pengujian untuk pemilihan konsentrasinya yaitu mahasiswa M.Arif, Khusnul, dan Muh.Taufik dan yang tidak sesuai dengan pengujian pilihan konsentrasinya yaitu mahasiswa Ngurah dan Yefgeni.

Berdasarkan implementasi *RapidMiner* dengan beberapa simbol yang digunakan didalam *tools*nya tersebut yang dimana fungsi dari *split* data yaitu pembagian antara data *training* dan data *testing*. Untuk itu penulis membagi data tersebut 50% data *training* dan 50% data *testing* dari total data 54. Sehingga menghasilkan akurasi sebesar 70,37%. Hal tersebut dapat dikatakan bahwa 70,37% data yang sesuai dengan pilihan konsentrasi dan 29,63% yang tidak sesuai dengan pilihan konsentrasi

Berdasarkan dari pengujian *K-Fold Cross Validation* untuk pengujian penulis menggunakan simbol *Cross Validation* yang ada di aplikasi *RapidMiner* yang fungsinya membagi data *training* dan data *testing* yang dimana data *training* untuk menjalankan algoritma *Naïve Bayes*nya sedangkan untuk data *testing* menghasilkan atau menampilkan hasil akurasi yang di hasilkan. Untuk itu penulis menggunakan data sebanyak 54 data dengan hasil akurasi sebesar 51,64% dengan koefisien korelasi cukup.

V. SARAN

Adapun saran yang ingin di sampaikan oleh penulis yaitu untuk penelitian selanjutnya yaitu dapat menambahkan data yang akan dijadikan data *training* dalam melakukan metode *Correlated Naïve Bayes* dan pengujian *K-Fold Cross Validation* yang dimana semakin banyak data *training*nya maka hasil penelitian akan semakin baik pula.

REFERENSI

- [1] (2018). Konsentrasi Jurusan. [Online]. Available: <https://campus.quipper.com/kampuspedia/konsentrasi-jurusan>.
- [2] Susanto, S., & Suryani, D. (2010). Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data. Jl.Beo 38-40, Yogyakarta: C.V ANDI OFFSET. 2010
- [3] Pramadhani, A. E., & Setiadi, T. "Penerapan Data Mining untuk Klasifikasi Penyakit ISPA dengan Algoritma Decision Tree". *Jurnal Sarjana Teknik Informatika*, Vol. 2, No. 1, pp. 831–839. 2014.
- [4] Fithri, D. L. "Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naive Bayes". *Simetris : Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, vol. 7, pp. 725. Nov. 2016
- [5] (2021). Contoh implementasi Metode Klasifikasi Naïve Bayes / Naïve Bayes Classifier (NBC) menggunakan PHP dan MySQL untuk memprediksi besarnya penggunaan listrik rumah tangga. [Online]. Available: https://extra.cahyadsn.com/naive_bayes.php
- [6] Bustami, Abdullah, D., & Fadlisyah. (n.d.). *STATISTIKA TERAPANNYA PADA BIDANG INFORMATIKA*. Ruko Jambusari 7A, Yogyakarta 55283: *Graha Ilmu*, 2014.
- [7] Rahmadani, S. "Implementasi Metode K-Nearest Neighbor Dalam Menentukan Kualitas Massa Bantuan". Universitas Islam Riau. 2020. [Online]. Available: <https://repository.uir.ac.id/8689/1/153510492.pdf>
- [8] Septiana, N. "Sistem Klasifikasi Citra Makanan Menggunakan Representasi Anti Textons Dan K-Nearest Neighbour" M.Kom. Thesis, Institut Teknologi Sepuluh Nopember. 2017
- [9] Didin Kusdian, R., Bandung, J. T., Ridwan, A. S., & Syafruddin, A. "Penggunaan Distribusi Normal Dalam Memodelkan Sebaran Persepsi Biaya Perjalanan Dan Transformasi Box-Muller Pada Pengambilan Sampel Acak Model Pemilihan Rute Dan Pembebanan Stokastik". In *Jurnal Transportasi*. Vol. 5, No.2, pp. 129-136, 2005