

Framework Prediksi Serapan Bekerja Alumni Berbasis Pembelajaran Mesin

Devi Fitriyah¹, Yuli Harwani²

¹ Computer Science Department, School of Computer Science,

² Management Department, School of Economics and Business,

Universitas Mercu Buana,

Jakarta, Indonesia 11480

devi.fitriyah@mercubuana.ac.id;

yuli.harwani@mercubuana.ac.id

Abstract – Determining the employability of graduates should become the main concern of a university business in order to be able to measure the curriculum effectiveness. The university must identify the characteristics and processes in this effort to gain a knowledge in this essential information. The characteristics is utilized in the predicting framework, which is a data driven method. A machine learning approach is implemented to discover a distinguished pattern, prediction and also mapping of students type along with the academic characteristics that is quick in the employability. This study answer the urge to have the framework based on machine learning to predict the graduates' employability.

Keywords: prediction framework; employability; graduates; machine learning; clustering; classification.

Abstrak – Menentukan serapan lulusan di dunia industri yang dicetak oleh perguruan tinggi merupakan sebuah usaha yang harus dilakukan dalam rangka untuk dapat melihat ke efektifkan kurikulum akademik yang diberikan pada saat dibangku kuliah. Karakteristik serta proses yang dilakukan untuk mendapatkan prediksi dan pemetaan ini memerlukan analisis data yang kompleks. Sebuah pendekatan pembelajaran mesin dengan framework prediksi dapat diterapkan untuk mendapatkan pola, prediksi dan pemetaan serapan lulusan di dunia kerja. Selain itu tentunya akan didapat sebuah pola jenis mahasiswa dengan karakteristik akademik seperti apa yang cepat diserap di dunia industri. Artikel ini mencoba menjawab untuk mengembangkan sebuah pendekatan prediksi berbasis pembelajaran mesin untuk menentukan serapan lulusan di dunia industri.

Keywords: Framework prediksi; serapan bekerj; alumni; pembelajaran mesin; klastering; klasifikasi.

I. PENDAHULUAN

Dalam perkembangannya perguruan tinggi menciptakan lulusan yang di mana dari tahun ke tahun semakin bertambah jumlahnya. Akan tetapi dengan jumlah lulusan yang kian bertambah tidak diseimbangkan dengan jumlah lapangan kerja, ini menyebabkan serapan di dunia kerja semakin sulit.

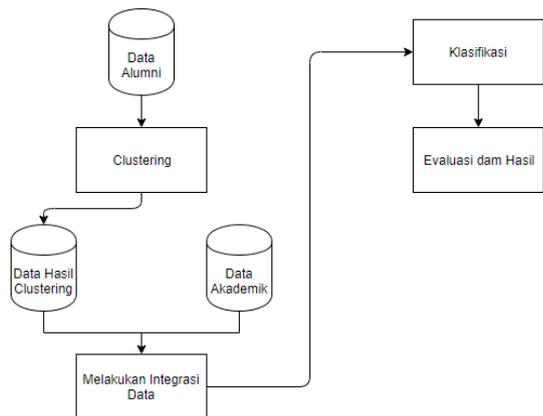
Usaha yang telah dilakukan perguruan tinggi untuk meningkatkan kualitas mahasiswa agar dapat meningkatkan softskill dan dapat memberikan bekal ketika sudah lulus untuk bersaing di dunia kerja. Hal ini merupakan salah satu yang diminta pada kurikulum yang berbasis outcome (Outcome Based Education). Untuk mengetahui kinerja mahasiswa, dengan memprediksi kinerja mahasiswa menggunakan teknik pengelompokan data (Durairaj & Vijitha, 2014). Penambahan data juga dilakukan untuk meningkatkan seleksi personil dan sumber daya manusia di industri teknologi (Chien & Chen, 2008). Studi lain dilakukan untuk memprediksi kegagalan mahasiswa menggunakan teknik penambahan data dengan klasifikasi (Márquez-Vera et al., 2016).

Dalam penyelesaiannya penulis akan membuat sebuah framework berbasis pembelajaran mesin untuk memprediksi serapan bekerja alumni. Diharapkan framework ini dapat digunakan untuk memberikan solusi meningkatkan kualitas mahasiswa.

II. METODOLOGI PENELITIAN

Adapun langkah untuk proses framework berbasis pembelajaran mesin ini ditunjukkan pada Gambar 1 di bawah ini. Langkah klasterisasi digunakan untuk memetakan waktu tunggu kerja alumni yang akan dibagi menjadi 3 klaster. Sedangkan langkah klasifikasi digunakan

untuk memprediksi waktu tunggu kerja alumni berdasarkan karakteristik akademik.



Gambar 1. Framework usulan penelitian

Pada clustering kami mencoba mengategorisasikan waktu tunggu kerja alumni dengan menggunakan Fuzzy C-Means. Setelah dapat hasil clustering kemudian hasil clustering diintegrasikan dengan data akademik yang akan dilanjutkan dengan proses klasifikasi sehingga framework ini dapat memprediksi waktu tunggu kerja alumni berdasarkan karakteristik akademik.

1. Data

Penelitian ini menggunakan 2 data dari sumber data yang berbeda. Data yang dibutuhkan adalah data alumni dan data akademik. Data alumni yang didapat sebanyak 981 record dalam kurun waktu 4 tahun (2013-2016). Sedangkan data akademik, yang didapat sebanyak 3296 record dalam kurun waktu 9 tahun (2006-2014).

2. Clustering

Pada proses clustering ini, data yang diproses hanya data alumni saja untuk mendapatkan sebuah class waktu tunggu kerja alumni. Data yang sudah diproses sebelumnya, akan memasuki tahap implementasi untuk mengetahui hasil algoritma FCM.

Untuk memulai, langkah inialisasi diperlukan untuk membangun matriks prototipe klaster awal dan matriks derajat keanggotaan Fuzzy. FCM dimulai dengan menentukan jumlah klaster. Pada artikel ini, penulis menentukan 3 klaster dan argumen yang tersisa kita biarkan secara default.

```
res.fcm <- fcm(x, centers=3)
v0 <- inaparc::kmpp(x, k=3)$v
print(v0)
```

Dalam hal ini, matriks prototipe dan derajat keanggotaan terbentuk dengan menggunakan inialisasi prototipe yang termasuk ke dalam paket 'inaparc'. Bertujuan agar *alginity* dan *algitu* dapat ditetapkan untuk masing-masing prototipe klaster dan derajat keanggotaan.

```
u0 <- inaparc::imembrand(nrow(x), k=3)$u
res.fcm <- fcm(x, centers=3, memberships=u0)
...
res.fcm <- fcm(x, centers=3,
alginityv="hartiganwong", algitu="imembrand")
```

Argumen *M* merupakan fuzzy eksponen yang terdiri dari beberapa nilai, penulis memilih Argumen *M* ini secara default yaitu $M = 2$ agar mendapat hasil yang lebih maksimal, bisa dilihat dalam baris kode berikut.

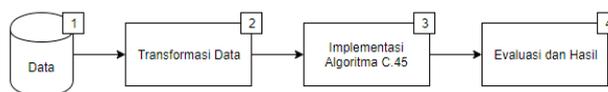
```
res.fcm <- fcm(x, centers=3, m=2)
```

3. Melakukan Integrasi Data

Integrasi data dilakukan pada atribut yang mengidentifikasi entitas-entitas yang ada pada data akademik dan data alumni. Pada artikel ini data alumni dan data akademik diintegrasikan yang akan membentuk suatu data di mana data tersebut akan digunakan dalam proses selanjutnya. Pada kasus ini, atribut HasilCluster dijadikan sebagai label untuk memprediksi. Di mana label itu akan berperan sebagai *supervisor* yang mengawasi proses pembelajaran dalam mencapai tingkat akurasi tertentu.

4. Klasifikasi

Langkah untuk proses klasifikasi ditunjukkan seperti pada Gambar 2 di bawah ini. Penulis melakukan beberapa langkah pada proses klasifikasi ini yaitu data, transformasi data, implementasi algoritma C.45, dan evaluasi dan hasil.



Gambar 2. Langkah proses Klasifikasi Algoritma C.45

a. Data

Data yang digunakan adalah data alumni yang sudah di klaster pada tahap awal dan data akademik yang sudah diintegrasikan. Atribut beserta nilai dari data yang akan digunakan dapat dilihat pada Tabel 1 berikut.

Tabel 1. Data Klasifikasi

Atribut	Deskripsi	Tipe Data	Contoh Data
1	NIM	Int	41506010025
2	Jenis Kelamin	Varchar	Laki-Laki
3	Waktu Dapat Kerja	Varchar	Sesudah Lulus
4	Bulan	Varchar	12 Bulan
5	HasilCluster	Char	Sedang
6	IPK	Char	Cukup
7	Status Kuliah	Varchar	Tidak Tepat Waktu
8	Lama Kuliah	Varchar	8 Tahun
9	Program Kelas	Varchar	REG 1
10	Status SMU	Char	SLTA
11	Program Studi	Char	TI

b. Transformasi Data

Setelah dilakukan integrasi data. Selanjutnya data akan ditransformasikan ke dalam bentuk data yang sesuai dengan kebutuhan algoritma. Pada artikel ini

penulis menyesuaikan nilai dari beberapa atribut agar cocok dengan algoritma yang akan digunakan. JK, WaktuDapatKerja, Bulan, IPK, dan Lama Kuliah. Pada atribut JK ini penulis merubah nilai 1 untuk laki-laki dan 0 untuk perempuan bisa dilihat pada Tabel 2 berikut.

Tabel 2. Transformasi Atribut JK

Nilai	Transformasi
0	Perempuan
1	Laki-Laki

Pada atribut WaktuDapatKerja penulis merubah nilai 0 untuk Tidak Diketahui, 1 untuk Sebelum Lulus, dan 2 untuk Sesudah Lulus. Dapat dilihat pada Tabel 3 berikut.

Tabel 3. Transformasi Atribut WaktuDapatKerja

Nilai	Transformasi
0	Tidak Diketahui
1	Sebelum Lulus
2	Sesudah Lulus

Pada atribut Bulan penulis merubah nilai 0 untuk 0 Bulan, 1 untuk 1 Bulan, 2 untuk 2 Bulan, dan seterusnya. Dapat dilihat pada Tabel 4 berikut.

Tabel 4. Transformasi Atribut Bulan

Nilai	Transformasi
0	0 Bulan
1	1 Bulan
2	2 Bulan
3	3 Bulan
4	4 Bulan
5	5 Bulan
...	...
60	60 Bulan

Pada atribut IPK penulis menyesuaikan dengan ketentuan yang ada di Universitas Mercu Buana yaitu 0 untuk Tidak Lulus, 1 untuk Kurang, 2 untuk Cukup, 2,5 untuk Cukup Baik, 3 untuk Baik, 3,5 untuk Sangat Baik, 4 untuk Istimewa. Contoh atribut IPK yang ditransformasikan dapat dilihat pada Tabel 5 berikut.

Tabel 5. Transformasi Atribut IPK

Nilai	Transformasi
2,01	Cukup
1,63	Kurang
1,43	Kurang
1,93	Kurang
3,37	Baik

Pada Atribut Lama Kuliah penulis merubah nilai 2 untuk 2 Tahun, 2.5 untuk 2.5 Tahun, 3 untuk 3 Tahun, 3.5 untuk 3.5 Tahun, dan seterusnya. Yang

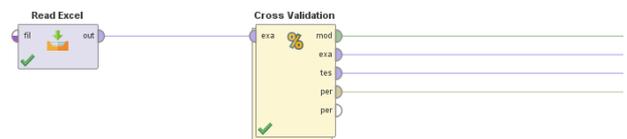
dapat dilihat pada Tabel 6 berikut.

Tabel 6. Transformasi Atribut Lama Kuliah

Nilai	Transformasi
2	2 Tahun
2,5	2,5 Tahun
3	3 Tahun
3,5	3,5 Tahun
...	...
7,5	7,5 Tahun

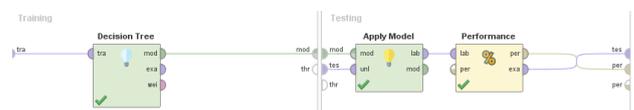
c. Implementasi Algoritma C.45

Pada implementasi algoritma penulis menggunakan algoritma C.45. Penulis menggunakan RapidMiner untuk melakukan uji coba dengan algoritma C.45. Penjelasan dari masing-masing proses yang menggunakan RapidMiner ditunjukkan pada Gambar 3 berikut.



Gambar 3. Operator Pengujian

- Operator Read Excel: Operator ini digunakan untuk membaca ExampleSet dari file Excel yang digunakan (Triyansyah & Fitriana, 2018).
- 5. Operator Cross Validation: operator ini memiliki 2 bagian di dalamnya yaitu Training dan Testing. Prosedur ini secara acak membagi satu set data ke dalam K lipatan terpisah dengan ukuran kira-kira sama, dan setiap lipatan pada gilirannya digunakan untuk menguji model yang diinduksikan dari lipatan K-1 lainnya dengan algoritma klasifikasi klasik. Kinerja algoritma klasifikasi dievaluasi oleh rata-rata akurasi K yang dihasilkan dari validasi *k-fold cross*, dan karenanya tingkat rata-rata diasumsikan titik untuk validasi pada fold (Wong, 2015).



Gambar 4. Operator Cross Validation

Gambar 4 di atas merupakan Operator Cross Validation yang di dalamnya ada beberapa operator lagi yaitu.

- Operator Decision Tree: operator ini menghasilkan model pohon keputusan dengan algoritma C.45 yang bisa digunakan untuk klasifikasi dan regresi.
- Apply Model: operator ini menerapkan model pada ExampleSet. Model pertama kali dilatih pada ExampleSet oleh operator lain. Tujuannya

untuk mendapatkan prediksi pada data yang tidak terlihat atau mengubah data dengan menerapkan model pra pemrosesan data.

- Performance: operator ini digunakan untuk evaluasi kinerja statistik dari klasifikasi. Operator ini juga memberikan daftar kriteria nilai dari klasifikasi.

d. Evaluasi dan Validasi Model

Validasi dilakukan dengan Ten-fold Cross Validation. Ten-fold Cross Validation adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah persentase dari catatan yang diklasifikasikan dengan benar dalam pengujian dataset (Arifin & Firianah, 2018). Validasi silang k-fold digunakan untuk mengevaluasi kinerja algoritma klasifikasi bukan model (Wong & Yang, 2017). Evaluasi yang dilakukan untuk memastikan bahwa hasil pengujian benar benar sesuai dengan pembahasan. Analisa dilakukan dengan melakukan perhitungan kembali hasil validasi dan pengujian (akurasi, presisi, dan recall) secara manual. Apakah perhitungan yang dilakukan akan menghasilkan nilai yang sama atau tidak, dibantu dengan Confusion matrix (Arifin & Fitriana, 2018). Di mana kelas yang diprediksi ditampilkan di bagian atas matriks dan kelas yang diamati disisi kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya sari kelas yang diamati untuk diprediksi (Ardiyansyah & Rahayuningsih, 2018).

III. HASIL DAN PEMBAHASAN

A. Clustering

Eksperimen pengelompokan menggunakan algoritma Fuzzy C-Means telah dilakukan dengan menentukan, sebagai berikut.

1. Tentukan parameter awal
 - a. Jumlah Cluster = 3
 - b. Fuzzy Eksponen = 2 (default)
 - c. Iterasi Maksimum = 1000 (default)
 - d. Kriteria terminasi = 0,00001
2. Buat angka acak U_{ik} , dengan = banyak data dan banyak cluster sebagai elemen matriks partisi awal. Matriks partisi awal yang terbentuk secara acak dapat dilihat pada Tabel 7 di bawah ini.

Tabel 7. Matriks Keanggotaan

Cluster 1	Cluster 2	Cluster 3
0.9951253981	0.0003006984	0.004573904
0.9358419096	0.0030558135	0.061102277
0.9951253981	0.0003006984	0.004573904
0.0140701344	0.0020536752	0.983876190
0.9951253981	0.0003006984	0.004573904

0.9951253981	0.0003006984	0.004573904
...
0.9951253981	0.0003006984	0.004573904
0.1635862307	0.3772940736	0.459119696
0.9951253981	0.0003006984	0.004573904
0.9951253981	0.0003006984	0.004573904
0.8105945795	0.0065915714	0.182813849
0.9507830468	0.0023118936	0.046905060

3. Perhitungan pusat cluster dan peningkatan nilai keanggotaan akan berhenti pada iterasi ke -95. Nilai pada iterasi ini adalah $\Delta = 2669.597411$ yang dapat dilihat pada Gambar 5. Proses perhitungan berhenti pada iterasi ke-95 karena nilai peningkatan fungsi tujuan antara dua iterasi berturut-turut kurang dari jumlah minimum perbaikan yang ditentukan $|P_n - P_{n-1}| < \varepsilon$ atau ketika loop melebihi maksimum iterasi ($n < \max n$).

```
Iteration count = 1, obj. fcn = 5697.612471
Iteration count = 2, obj. fcn = 4624.136838
Iteration count = 3, obj. fcn = 4294.742572
Iteration count = 4, obj. fcn = 3917.251058
Iteration count = 5, obj. fcn = 3553.858263
Iteration count = 6, obj. fcn = 3310.267443
Iteration count = 7, obj. fcn = 3162.965153
Iteration count = 8, obj. fcn = 3047.806670
Iteration count = 9, obj. fcn = 2952.840160
Iteration count = 10, obj. fcn = 2883.080161
...
Iteration count = 90, obj. fcn = 2669.597494
Iteration count = 91, obj. fcn = 2669.597466
Iteration count = 92, obj. fcn = 2669.597446
Iteration count = 93, obj. fcn = 2669.597430
Iteration count = 94, obj. fcn = 2669.597419
Iteration count = 95, obj. fcn = 2669.597411
```

Gambar 5. Hasil Akhir Iterasi

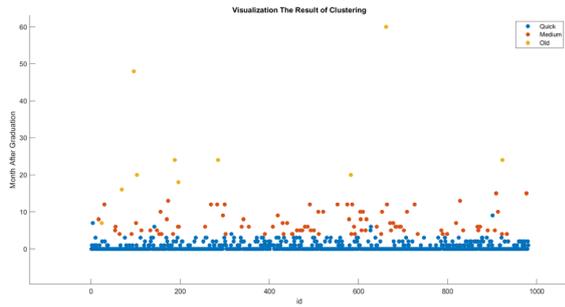
4. Setelah mendapatkan hasil iterasi akhir, kami mendapatkan nilai pusat cluster akhir. Pusat cluster sangat menentukan keanggotaan data yang dimiliki, nilai keanggotaan ditetapkan untuk setiap titik data. Dengan demikian, poin di tepi cluster, dengan nilai keanggotaan yang lebih rendah, mungkin dalam cluster ke level yang lebih rendah daripada poin di pusat cluster. Dalam fuzzy clustering, titik data berpotensi menjadi milik beberapa cluster. Informasi ini dapat dilihat pada Tabel 8 berikut.

Tabel 8. Pusat Klaster Akhir

	Jenis Kelamin	Waktu Dapat Kerja	Bulan Setelah Lulus
Cluster 1	0.7345708	1.086663	0.325424
Cluster 2	0.9483884	1.999757	24.646977
Cluster 3	0.7244739	1.977939	6.242615

Kita dapat melihat visualisasi pengelompokan dari hasil implementasi algoritma FCM. Dari hasil tersebut, ada

3 kluster, yaitu Cepat = 880, Sedang = 91, Lama = 10. Plot dari gugus yang dihasilkan dapat dilihat pada Gambar 6 berikut.

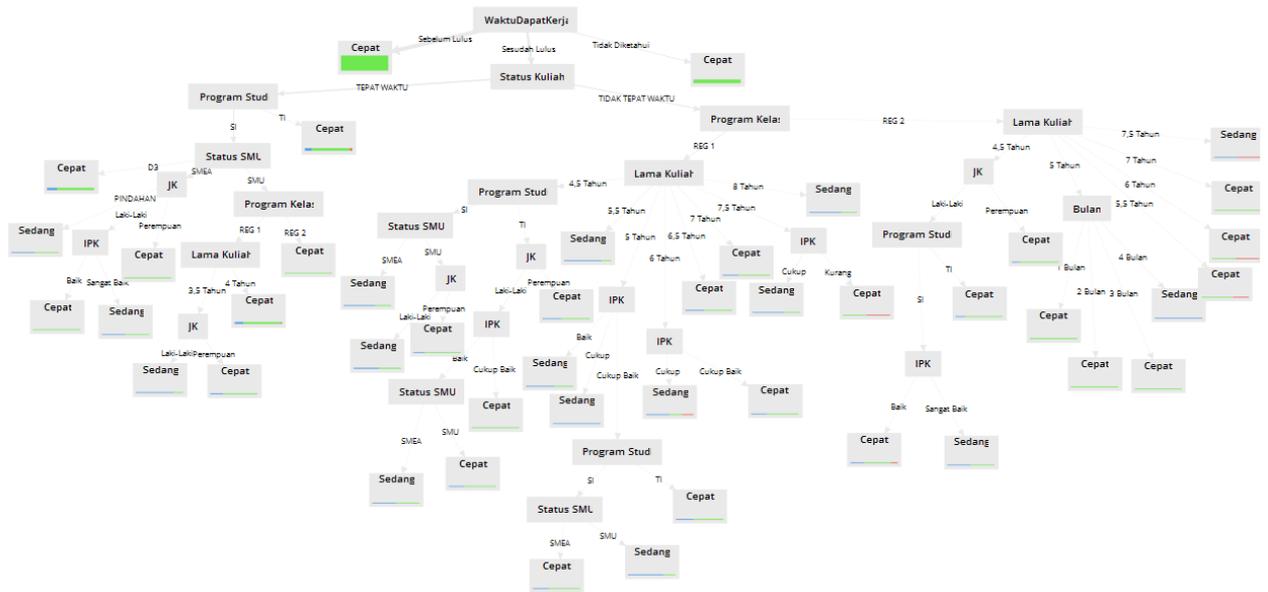


Gambar 6. Visualisasi Hasil Clustering

Untuk memvalidasi hasil kluster penulis menggunakan silhouette index. Dari hasil aklaster ini, silhouette indeks yang diperoleh adalah 0,9168.

B. Klasifikasi

Berdasarkan semua proses yang dilalui pada proses implementasi klasifikasi algoritma C.45 ini maka mohon keputusan akan membentuk seperti yang ditunjukkan pada Gambar 7 berikut ini.



Gambar 7. Pohon Keputusan

Tree

```

WaktuDapatKerja = Sebelum Lulus: Cepat (Sedang=0, Cepat=540, Lama=0)
WaktuDapatKerja = Sesudah Lulus
  Status Kuliah = TEPAT WAKTU
    Program Studi = SI
      Status SMU = DS: Cepat (Sedang=12, Cepat=44, Lama=0)
      Status SMU = PINDAHAN: Sedang (Sedang=1, Cepat=1, Lama=0)
      Status SMU = SMEA
        JK = Laki-Laki
          IPK = Baik: Cepat (Sedang=0, Cepat=3, Lama=0)
          IPK = Sangat Baik: Sedang (Sedang=1, Cepat=1, Lama=0)
          JK = Perempuan: Cepat (Sedang=0, Cepat=5, Lama=0)
        Status SMU = SMU
          Program Kelas = REG 1
            Lama Kuliah = 3,5 Tahun
              JK = Laki-Laki: Sedang (Sedang=4, Cepat=1, Lama=0)
              JK = Perempuan: Cepat (Sedang=3, Cepat=8, Lama=0)
            Lama Kuliah = 4 Tahun: Cepat (Sedang=8, Cepat=34, Lama=1)
          Program Kelas = REG 2: Cepat (Sedang=0, Cepat=2, Lama=0)
          Program Studi = TI: Cepat (Sedang=10, Cepat=55, Lama=3)
      Status Kuliah = TIDAK TEPAT WAKTU
        Program Kelas = REG 1
          Lama Kuliah = 4,5 Tahun
            Program Studi = SI
              Status SMU = SMEA: Sedang (Sedang=4, Cepat=2, Lama=0)
              Status SMU = SMU
                JK = Laki-Laki: Sedang (Sedang=7, Cepat=6, Lama=0)
                JK = Perempuan: Cepat (Sedang=1, Cepat=3, Lama=0)
            Program Studi = TI
              JK = Laki-Laki
                IPK = Baik
                  Status SMU = SMEA: Sedang (Sedang=1, Cepat=1, Lama=0)
                  Status SMU = SMU: Cepat (Sedang=1, Cepat=2, Lama=0)
                IPK = Cukup Baik: Cepat (Sedang=0, Cepat=3, Lama=0)
                JK = Perempuan: Cepat (Sedang=2, Cepat=3, Lama=0)
              Lama Kuliah = 5 Tahun

```

Gambar 8. Deskripsi Pohon Keputusan

Dari pohon keputusan yang terbentuk pada Gambar 7, didapatkan aturan-aturan atau rule model dalam memprediksi alumni dengan waktu tunggu bekerja sebagai berikut:

1. If WaktuDapatKerja = Sebelum Lulus Then Prediksi Cepat
2. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = Tepat Waktu and Program Studi = SI and Status SMU = D3 Then Prediksi Cepat
3. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = Tepat Waktu and Program Studi = SI and Status SMU = Pindahan Then Prediksi Sedang
4. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMEA and JK = Laki-Laki and IPK = Baik Then Prediksi Cepat
5. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMEA and JK = Laki-Laki and IPK = Sangat Baik Then Prediksi Sedang
6. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMEA and JK = Perempuan Then Prediksi Cepat
7. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMU and Program Kelas = REG 1 and Lama Kuliah = 3,5 Tahun and JK = Laki-Laki Then Prediksi Sedang
8. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMU and Program Kelas = REG 1 and Lama Kuliah 3,5 Tahun and JK = Perempuan Then Prediksi Cepat
9. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMU and Program Kelas = REG 1 and Lama Kuliah = 4 Tahun Then Prediksi Cepat
10. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = SI and Status SMU = SMU and Program Kelas = REG 2 Then Prediksi Cepat
11. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TEPAT WAKTU and Program Studi = TI Then Prediksi Cepat
12. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = SI and Status SMU = SMEA Then Prediksi Sedang
13. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program

Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = SI and Status SMU = SMU and JK = Laki-Laki Then Prediksi Sedang

14. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = SI and Status SMU = SMU and JK = Perempuan Then Prediksi Cepat
15. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = TI and JK = Laki-Laki and IPK Baik and Status SMU = SMEA Then Prediksi Sedang
16. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = TI and JK = Laki-Laki and IPK = Baik and Status SMU = SMU Then Prediksi Cepat
17. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = TI and JK = Laki-Laki and IPK = Cukup Baik Then Prediksi Cepat
18. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 4,5 Tahun and Program Studi = TI and JK = Perempuan Then Prediksi Cepat
19. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 5 Tahun and IPK = Baik Then Prediksi Sedang
20. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 5 Tahun and IPK Cukup Then Prediksi Sedang
21. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 5 Tahun and IPK = Cukup Baik and Program Studi = SI and Status SMU = SMEA Then Prediksi Cepat
22. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 5 Tahun and IPK = Cukup Baik and Program Studi = SI and Status SMU = SMU Then Prediksi Sedang
23. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 5 Tahun and IPK = Cukup Baik and Program Studi = TI Then Prediksi Cepat
24. If WaktuDapatKerja = Sesudah Lulus dan Status Kuliah = TIDAK TEPAT WAKTU and Program

- Kelas = REG 1 and Lama Kuliah = 5,5 Tahun
Then Prediksi Sedang
25. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 6 Tahun and IPK = Cukup Then Prediksi Sedang
 26. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 6 Tahun and IPK = Cukup Baik Then Prediksi Cepat
 27. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 6,5 Tahun Then Prediksi Cepat
 28. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 7 Tahun Then Prediksi Cepat
 29. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 7,5 Tahun and IPK = Cukup Then Prediksi Sedang
 30. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 7,5 Tahun and IPK = Kurang Then Prediksi Cepat
 31. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 1 and Lama Kuliah = 8 Tahun Then Prediksi Sedang
 32. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 4,5 Tahun and JK = Laki-Laki and Program Studi = SI and IPK = Baik Then Prediksi Cepat
 33. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 4,5 Tahun and JK = Laki-Laki and Program Studi = SI and IPK = Sangat Baik Then Prediksi Sedang
 34. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 4,5 Tahun and JK = Laki-Laki and Program Studi = TI Then Prediksi Cepat
 35. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 4,5 Tahun and JK = Perempuan Then Prediksi Cepat
 36. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 5 Tahun and Bulan = 1 Bulan Then Prediksi Cepat
 37. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 5 Tahun and Bulan = 2 Bulan Then Prediksi Cepat
 38. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 5 Tahun and Bulan = 3 Bulan Then Prediksi Cepat
 39. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 5 Tahun and Bulan = 4 Bulan Then Prediksi Sedang
 40. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 5,5 Tahun Then Prediksi Cepat
 41. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 6 Tahun Then Prediksi Cepat
 42. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 7 Tahun Then Prediksi Cepat
 43. If WaktuDapatKerja = Sesudah Lulus and Status Kuliah = TIDAK TEPAT WAKTU and Program Kelas = REG 2 and Lama Kuliah = 7,5 Tahun Then Prediksi Sedang
 44. If WaktuDapatKerja = Tidak Diketahui Then Prediksi Cepat

Berdasarkan pengujian dengan metode Ten-fold Cross Validation menghasilkan nilai akurasi sebesar 86,03%. Evaluasi selanjutnya dilakukan secara manual dengan Confusion Matrix dapat dilihat pada Gambar 9 berikut.

accuracy: 86.03% +/- 3.38% (micro average: 86.03%)

	true Sedang	true Cepat	true Lama	class precision
pred. Sedang	22	56	2	27.50%
pred. Cepat	69	822	8	91.43%
pred. Lama	0	2	0	0.00%
class recall	24.18%	93.41%	0.00%	

Gambar 9. Confusion Matrix

IV. KESIMPULAN

Pada artikel ini penulis bertujuan untuk membuat framework berbasis pembelajaran mesin untuk memprediksi serapan lulusan di dunia kerja. Pada eksperimen di atas menunjukkan bahwa framework berbasis pembelajaran mesin yang diusulkan berhasil dibuat dengan menggunakan 2 teknik penerapan data mining yaitu dengan klustering dan klasifikasi. Di mana klustering menggunakan Fuzzy C-Means, sedangkan klasifikasi menggunakan algoritma C.45.

Pada pengelompokan data alumni yang menggunakan FCM ini, kami mengidentifikasi 3 kluster di mana kluster Cepat = 880 (cepat mendapatkan pekerjaan), Sedang = 91 (waktu sedang dalam mendapatkan pekerjaan), dan Lama = 10 (lama dalam mendapatkan pekerjaan). Dan mendapatkan hasil kluster yang baik diuji dengan menggunakan silhouete indeks.

Sedangkan prediksi dengan menggunakan algoritma C.45 dengan data yang sudah diintegrasikan berdasarkan Gambar 10 dapat disimpulkan bahwa perhitungan tingkat akurasi mencapai 86,03%. Di mana karakteristik akademik alumni tidak selalu berpengaruh dalam hal serapan bekerja.

DAFTAR PUSTAKA

- Ardiyansyah, & Rahayuningsih, P. A. (2018). Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. *Journal Khatulistiwa Inform*, VI(1), 20–28.
- Arifin, M. F., & Firianah, D. (2018). Penerapan Algoritma Klasifikasi C4. 5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada. *IncomTech, Jurnal Telekomunikasi Dan Komputer*, 8(2), 87–102.
- Arifin, M. F., & Fitriana, D. (2018). *Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada*.
- Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications*, 34(1), 280–290. <https://doi.org/10.1016/j.eswa.2006.09.003>
- Durairaj, M., & Vijitha, C. (2014). Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), 5987–5991. Retrieved from <https://pdfs.semanticscholar.org/892b/0182c44c34a2ae68daec819eac301c3bd9c.pdf>
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, 33(1), 107–124. <https://doi.org/10.1111/exsy.12135>
- Triyansyah, D., & Fitriana, D. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *IncomTech, Jurnal Telekomunikasi Dan Komputer*, 8(3), 163–182.
- Wong, T. T. (2015). Performance Evaluation of Classification Algorithm by K-Fold and Leave-One-Out Cross Validation. *Pattern Recognition*, 48(9), 2839–2846.
- Wong, T. T., & Yang, N. Y. (2017). Dependency Analysis of Accuracy Estimates in K-Fold Cross. *IEEE Transactions on Knowledge and Data Engineering*, 29(11), 2417–2427.