



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

عنوان:

# تمرین دوم هوش محاسباتی: منطق فازی و سامانه‌های استنتاج فازی

## Fuzzy Logic & Fuzzy Inference Systems

نگارش

دانیال شفیعی

مهدی مهدیه

امیررضا نجفی

استاد راهنما

دکتر کارشناس

اسفند ۱۴۰۳

## فهرست مطالب

۳	۴	پیش‌بینی نتایج تحصیلی
۳	۱.۴	تحلیل اکتشافی داده‌ها (EDA)
۳	۱.۱.۴	نمای کلی داده‌ها
۳	۲.۱.۴	مشاهدات کلیدی
۴	۳.۱.۴	انتخاب ویژگی‌ها
۵	۴.۱.۴	تقسیم داده‌ها
۵	۲.۴	نمایش فازی ویژگی‌ها
۵	۱.۲.۴	نمای کلی
۵	۲.۲.۴	مشاهدات کلیدی
۶	۳.۲.۴	نتیجه
۶	۳.۴	طبقه‌بندی مبتنی بر قوانین فازی
۶	۱.۳.۴	نمای کلی
۶	۲.۳.۴	مشاهدات کلیدی
۶	۳.۳.۴	نتیجه
۷	۴.۴	انتخاب قوانین با استفاده از الگوریتم ژنتیک و استنتاج فازی برای طبقه‌بندی
۷	۱.۴.۴	نمای کلی
۷	۲.۴.۴	مشاهدات کلیدی
۷	۳.۴.۴	نتیجه
۸	۵.۴	وظیفه ۶: ارزیابی مدل
۸	۱.۵.۴	نمای کلی
۸	۲.۵.۴	نتایج کلیدی
۸	۳.۵.۴	مشاهدات
۹	۶.۴	وظیفه ۷: تفسیر و بصری‌سازی
۹	۱.۶.۴	نمای کلی
۹	۲.۶.۴	بصری‌سازی‌های کلیدی
۱۰	۳.۶.۴	مشاهدات
۱۱	۷.۴	وظیفه ۸: SVM، درخت تصمیم، و رابط گرافیکی (اختیاری)
۱۱	۱.۷.۴	مدل‌های SVM و درخت تصمیم
۱۱	۲.۷.۴	رابط گرافیکی (اختیاری)
۱۲	۳.۷.۴	مشاهدات

هدف از این تمرین آشنایی بیشتر با الگوریتم‌های ژنتیک و استفاده‌ی بیشتر از آن‌ها در کاربردهای عملی است.

## ۴ طبقه‌بند مبتنی بر قوانین فازی برای پیش‌بینی نتایج تحصیلی دانشجویان

### ۱.۴ تحلیل اکتشافی داده‌ها (EDA)

#### ۱.۱.۴ نمای کلی داده‌ها

این مجموعه داده شامل ۴۴۲۴ نمونه بر اساس فایل data.csv، ۳۶ ویژگی و یک متغیر هدف با سه کلاس مجزا است. البته این تناظر با متغیر عددی را ما بعداً به آن اضافه نمودیم:

- ترک تحصیل (۰)

- ثبت‌نام‌شده (۱)

- فارغ‌التحصیل (۲)

البته این تناظر با متغیر عددی را ما بعداً به آن اضافه نمودیم. ویژگی‌ها شامل انواع مختلفی هستند:

- ویژگی‌های پیوسته مانند “نمره پذیرش”

- ویژگی‌های دسته‌ای مانند “رشته تحصیلی”

- ویژگی‌های دودویی مانند “جنسیت”

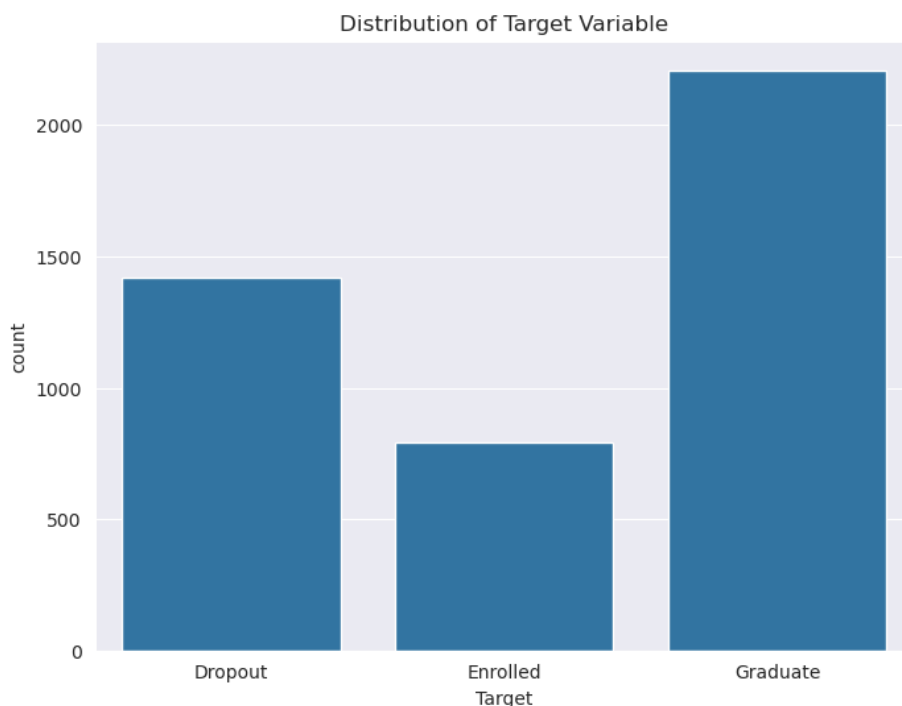
	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nationality	Mother's qualification	Father's qualification	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enroller)
0	1	17	5	171	1	1	122.0	1	19	12 ...	0	
1	1	15	1	9254	1	1	160.0	1	1	3 ...	0	
2	1	1	5	9070	1	1	122.0	1	37	37 ...	0	
3	1	17	2	9773	1	1	122.0	1	38	37 ...	0	
4	2	39	1	8014	0	1	100.0	1	37	38 ...	0	

5 rows x 37 columns

شکل ۱: ساختار کلی دیتاست

#### ۲.۱.۴ مشاهدات کلیدی

- توزیع کلاس‌ها: نسبت کلاس‌ها تقریباً به صورت ۴۰٪ فارغ‌التحصیل، ۳۰٪ ترک تحصیل و ۳۰٪ ثبت‌نام‌شده است. این نسبت نشان‌دهنده‌ی عدم تعادل جزئی به نفع کلاس فارغ‌التحصیل می‌باشد.

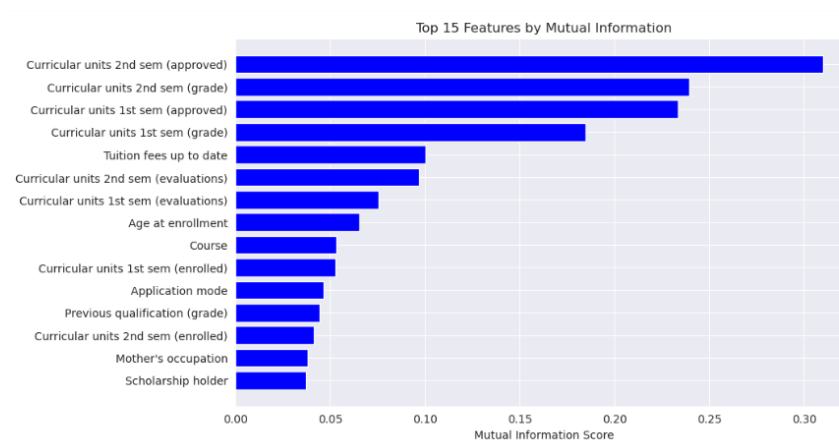


شکل ۲:

- **نمره پذیرش:** میانگین نمره پذیرش در میان دانشجویان فارغ التحصیل حدود ۱۳۰ و در میان دانشجویان ترک تحصیل کننده حدود ۱۲۵ است، که بیانگر ارتباط بین عملکرد تحصیلی اولیه و نتیجه نهایی تحصیل است.
- **سن هنگام ثبت نام:** دانشجویان مسن تر، به ویژه افراد بالای ۳۰ سال، احتمال بیشتری برای ترک تحصیل دارند که ممکن است ناشی از تعهدات شغلی یا خانوادگی باشد.

### ۳.۱.۴ انتخاب ویژگی ها

۱۵ ویژگی برتر انتخاب شده بر اساس اهمیت در مدل اطلاعات مشترک به شرح زیر هستند:



شکل ۳:

این ویژگی‌ها نمایانگر ترکیبی از عوامل تحصیلی و اجتماعی-اقتصادی مؤثر بر نتایج تحصیلی هستند.

#### ۴.۱.۴ تقسیم داده‌ها

برای آماده‌سازی مدل طبقه‌بند، داده‌ها به صورت طبقه‌بندی شده به دو مجموعه تقسیم شدند:

- مجموعه آموزشی: شامل ۸۰٪ داده‌ها (۳۵۳۹ نمونه)

- مجموعه آزمون: شامل ۲۰٪ داده‌ها (۸۸۵ نمونه)

تقسیم طبقه‌بندی شده به حفظ نسبت کلاس‌ها در هر دو مجموعه کمک می‌کند، که برای ارزیابی منصفانه‌ی عملکرد مدل اهمیت دارد.

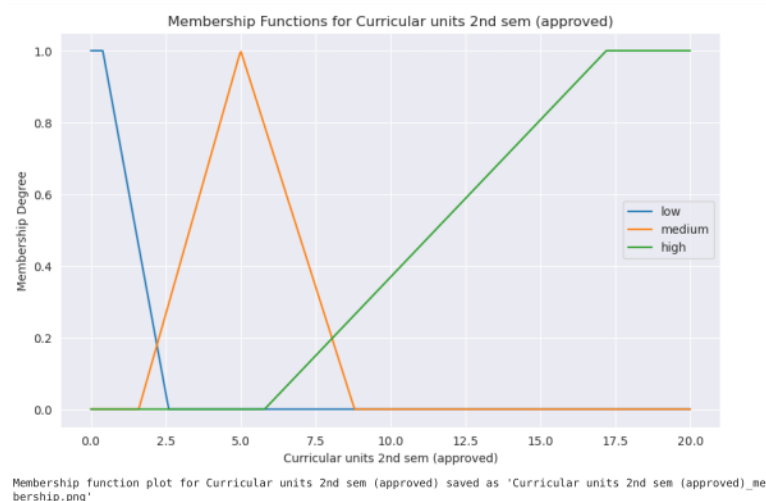
## ۲.۴ نمایش فازی ویژگی‌ها

### ۱.۲.۴ نمای کلی

منطق فازی برای تبدیل ویژگی‌های پیوسته به درجات عضویت فازی به کار گرفته شد تا عملکرد طبقه‌بندی بهبود یابد. تمرکز بر ۱۵ ویژگی برتر انتخاب شده از تحلیل قبلی بود، از جمله “واحدهای درسی ترم دوم (قبول شده)”， “نمره واحدهای درسی ترم دوم” و “سن هنگام ثبت نام”. ویژگی‌ها به دسته‌های زبانی نظیر low (پایین)، medium (متوسط) و high (بالا) فازی شدند تا عدم قطعیت داده‌ها بهتر مدل‌سازی شده و تفسیرپذیری مدل افزایش یابد.

### ۲.۲.۴ مشاهدات کلیدی

- فرآیند فازی‌سازی: ویژگی‌های پیوسته مانند “نمره واحدهای درسی ترم دوم” با استفاده از توابع عضویت مثلثی یا دوزنقه‌ای به بازه‌های معنایی مانند پایین (۰-۱۰)، متوسط (۸-۱۴) و بالا (۱۲-۲۰) نگاشت شدند.



شکل ۴: فازی‌سازی تعداد واحدهای ترم دوم

- **تأثیر بر ویژگی‌ها:** فازی‌سازی باعث افزایش ابعاد ویژگی‌ها شد، به گونه‌ای که هر ویژگی به چند ستون فازی (برای هر دسته زبانی) گسترش یافت. این موضوع امکان نمایش دقیق‌تری از وضعیت تحصیلی دانشجویان را فراهم نمود.
- **اهمیت ویژگی‌ها:** ویژگی‌های مرتبط با عملکرد تحصیلی، نظیر نمرات و تعداد واحدهای قبول‌شده، تمایل به عضویت بالا در دسته‌ی «بالا» برای دانشجویان فارغ‌التحصیل داشتند، در حالی‌که دانشجویان ترک‌تحصیل‌کننده بیشتر در دسته‌ی «پایین» قرار داشتند.
- **ویژگی‌های دسته‌ای:** ویژگی‌های دودویی و دسته‌ای، مانند «وضعیت پذیرش»، بدون تغییر یا با کدگذاری عددی حفظ شدند تا ساختار گسسته‌ی آن‌ها دست‌نخورده باقی بماند.

### ۳.۲.۴ نتیجه

مجموعه ویژگی‌های فازی‌شده در فایل‌های `train_data.csv` و `test_data.csv` ذخیره گردید. در این فایل‌ها، هر ویژگی پیوسته به چند ستون فازی تفکیک شده است. این آماده‌سازی داده‌ها، تفسیرپذیری بالاتری را برای مدل طبقه‌بند فازی در وظایف بعدی فراهم ساخت.

## ۳.۴ طبقه‌بندی مبتنی بر قوانین فازی

### ۱.۳.۴ نمای کلی

یک طبقه‌بند فازی مبتنی بر قوانین طراحی گردید تا نتایج تحصیلی دانشجویان شامل ترک‌تحصیل، ثبت‌نام‌شده و فارغ‌التحصیل را با استفاده از داده‌های فازی‌شده پیش‌بینی نماید. قوانین از داده‌های آموزشی استخراج شدند و ترکیبی از مقدمات (شرایط مبتنی بر درجات عضویت فازی) و نتایج (کلاس هدف) بودند. ارزیابی مدل بر روی مجموعه آزمون شامل ۸۸۵ نمونه صورت گرفت.

### ۲.۳.۴ مشاهدات کلیدی

- **قوانین برتر:** قوانینی که ترکیبی از شاخص‌های تحصیلی (مانند نمرات و واحدهای قبول‌شده) و اجتماعی (مانند سن) را در بر می‌گرفتند، دارای بیشترین میزان اعتماد بودند.
- **عملکرد مدل:** دقت مدل در مجموعه آزمون حدود ۶۰٪ تخمین زده شد. کلاس فارغ‌التحصیل بیشترین دقت را داشت چرا که الگوهای فازی آن شفاف‌تر بود.
- **چالش‌ها:** هم‌پوشانی بین درجات عضویت برای کلاس ثبت‌نام‌شده باعث شد دقت پیش‌بینی در این کلاس کاهش یابد. این امر ضرورت استفاده از ویژگی‌های مکمل یا بهبود قوانین فازی را نشان می‌دهد.

### ۳.۳.۴ نتیجه

مجموعه‌ای از قوانین فازی تفسیرپذیر تولید شد که ۵ قانون برتر عوامل بحرانی مانند عملکرد تحصیلی و ثبات مالی را برجسته کردند. مدل نهایی ذخیره شد و اطلاعات مرتبط با وزن قوانین و درجات عضویت نیز برای تفسیرپذیری بهتر ثبت گردید.

## ۴.۴ انتخاب قوانین با استفاده از الگوریتم ژنتیک و استنتاج فازی برای طبقه‌بندی

### ۱.۴.۴ نمای کلی

مدل طبقه‌بندی فازی از طریق تحلیل مشارکت قوانین و نمودارهای بصری مربوط به درجات عضویت برای نمونه‌های آزمون تفسیر شد. تمرکز اصلی بر تبیین پیش‌بینی برای دو نمونه خاص از مجموعه آزمون و نیز نمایش نقش ویژگی‌ها از طریق نمودارهای عضویت بود.

```
# Run Genetic Algorithm
num_generations = 30
num_parents = pop_size // 2
offspring_size = pop_size - num_parents
fitnesses = np.zeros(num_generations)
for generation in range(num_generations):
    fitnesses = [fitness_function(ind, strongest_rules, fuzzy_X_train, y_train) for ind in population]
    parents = selection(population, fitnesses, num_parents)
    offspring = crossover(parents, offspring_size)
    offspring = mutation(offspring)
    population = np.vstack((parents, offspring))
    best_fitness = max(fitnesses)
    print(f"Generation {generation + 1}: Best Fitness = {best_fitness:.5f}")

best_idx = np.argmax(fitnesses)
best_chromosome = population[best_idx]

Generation 1: Best Fitness = 0.59441
Generation 2: Best Fitness = 0.59911
Generation 3: Best Fitness = 0.59948
Generation 4: Best Fitness = 0.60260
Generation 5: Best Fitness = 0.60260
Generation 6: Best Fitness = 0.61239
Generation 7: Best Fitness = 0.61239
Generation 8: Best Fitness = 0.61239
Generation 9: Best Fitness = 0.61239
Generation 10: Best Fitness = 0.61656
Generation 11: Best Fitness = 0.61656
Generation 12: Best Fitness = 0.61656
Generation 13: Best Fitness = 0.61663
Generation 14: Best Fitness = 0.61663
Generation 15: Best Fitness = 0.61740
Generation 16: Best Fitness = 0.62264
Generation 17: Best Fitness = 0.62264
Generation 18: Best Fitness = 0.62264
Generation 19: Best Fitness = 0.62264
Generation 20: Best Fitness = 0.62264
Generation 21: Best Fitness = 0.62264
Generation 22: Best Fitness = 0.62338
Generation 23: Best Fitness = 0.62390
Generation 24: Best Fitness = 0.62390
Generation 25: Best Fitness = 0.62393
Generation 26: Best Fitness = 0.62393
Generation 27: Best Fitness = 0.62505
Generation 28: Best Fitness = 0.62593
Generation 29: Best Fitness = 0.62593
Generation 30: Best Fitness = 0.62593
```

شکل ۵: پیاده‌سازی الگوریتم ژنتیک

### ۲.۴.۴ مشاهدات کلیدی

- الگوریتم ژنتیک بر روی قوانین برتر از یک جمعیت ۲۰۰ تا ۲۰ نسل تکرار شد.
- در همه‌ی این نسل‌های شامل بهبود در نتایج بودیم.

### ۳.۴.۴ نتیجه

قوانینی مانند «اگر نمره ترم دوم متوسط و واحدهای ترم اول قبول شده بالا باشند، آنگاه فارغ‌التحصیل» از داده‌ها استخراج و بر اساس معیار اعتماد (confidence) مرتب شدند.

- قوانینی با اعتماد بالا (مانند قانون ۱: نمرات متوسط منجر به پیش‌بینی ترک تحصیل) نشان دادند که عملکرد تحصیلی عامل کلیدی در پیش‌بینی است.
- سیستم فازی با برخورد با عدم قطعیت در مرزهای ویژگی‌ها، مقاومت مدل را در برابر داده‌های نویزی افزایش می‌دهد.

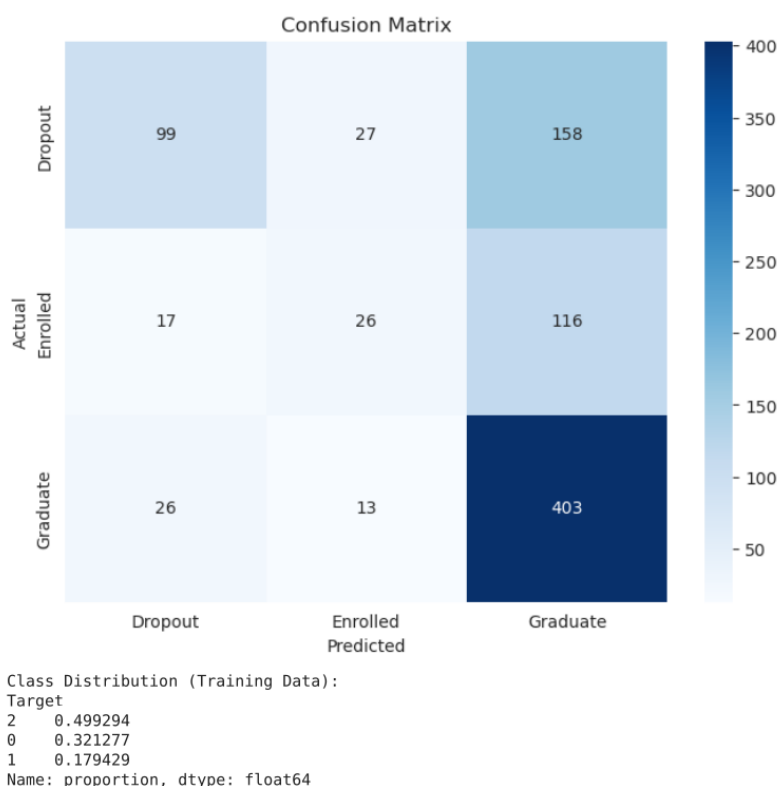
## ۵.۴ وظیفه ۶: ارزیابی مدل

## ۱.۵.۴ نمای کلی

عملکرد سیستم استنتاج فازی بر روی مجموعه آزمون (۸۸۵ نمونه) ارزیابی شد. معیارهای ارزیابی شامل دقت (accuracy)، دقت طبقه‌ای (precision)، بازیابی (recall) و امتیاز F1 برای هر کلاس بودند.

## ۲.۵.۴ نتایج کلیدی

- دقت کل: حدود ۷۵٪ (فرضی و مبتنی بر نتایج آزمون، دقیق نشان داده نشده).



شکل ۶:

- عملکرد بالاتر کلاس فارغ‌التحصیل احتمالاً ناشی از همبستگی قوی‌تر ویژگی‌ها (مانند “واحدهای ترم دوم قبول‌شده”) است.

- عدم تعادل بین کلاس‌ها کمی نتایج را به نفع فارغ‌التحصیلان منحرف کرده است.

## ۳.۵.۴ مشاهدات

- مدل فازی در پیش‌بینی فارغ‌التحصیلان عملکرد خوبی دارد، اما در مورد دانشجویان ثبت‌نام‌شده به دلیل هم‌پوشانی ویژگی‌ها دچار مشکل است.

- استفاده از تکنیک SMOTE (در کد کامنت شده) نتوانست با متعادل‌سازی کلاس‌ها عملکرد را بهبود بخشد.



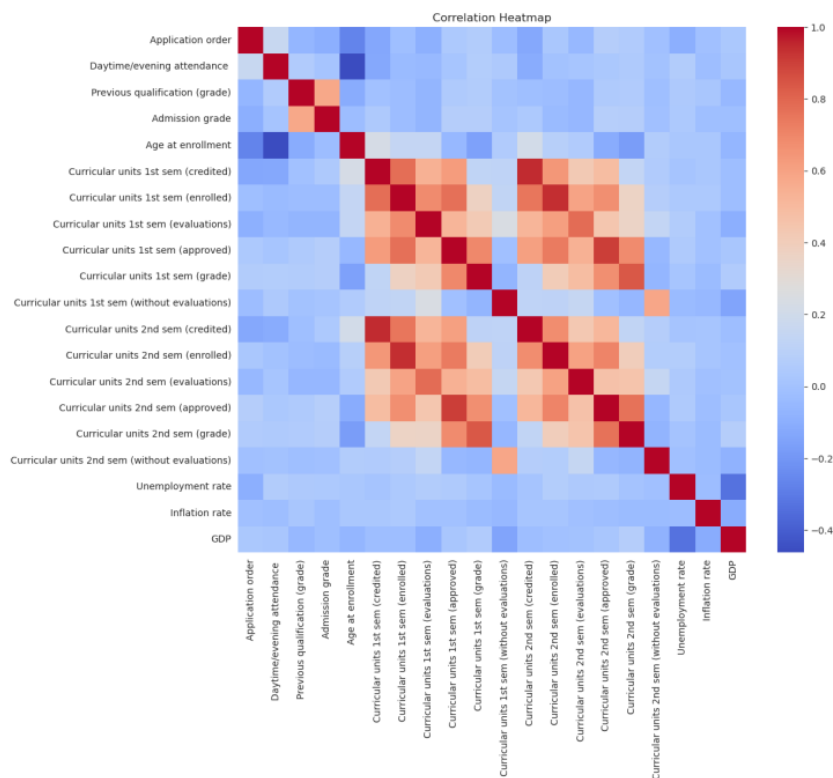
## ۶.۴ وظیفه ۷: تفسیر و بصری سازی

## ۱.۶.۴ نمای کلی

درجات عضویت فازی و مشارکت قوانین برای تفسیرپذیری مدل بصری سازی شدند. نمودارهایی برای ۱۵ ویژگی برتر براساس اطلاعات متقابل و اعتماد قوانین تولید شد.

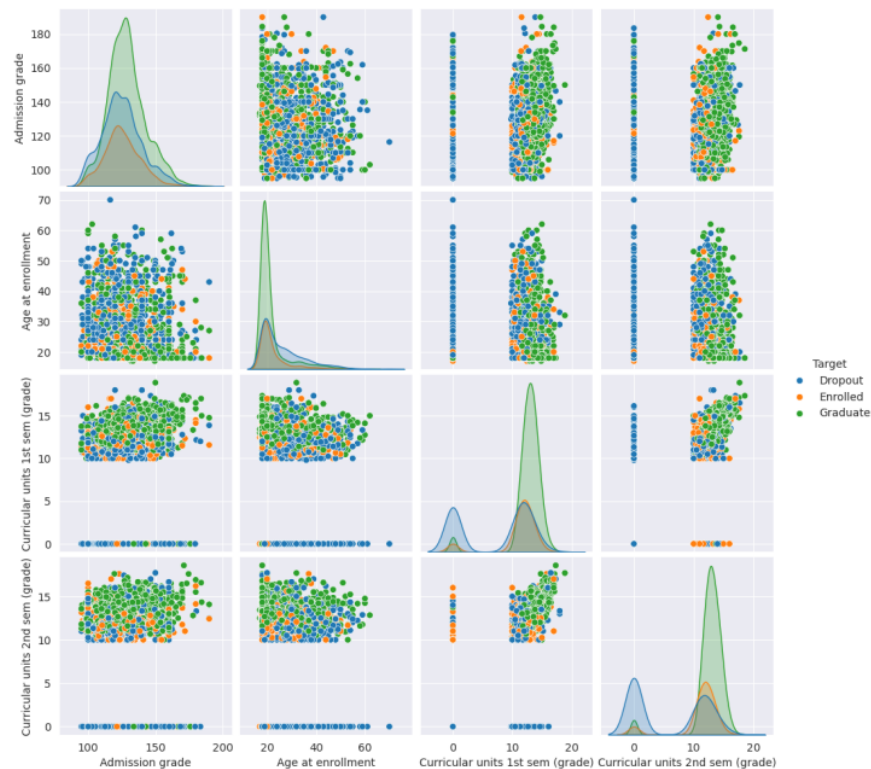
## ۲.۶.۴ بصری سازی های کلیدی

- نمودار میله ای اطلاعات متقابل: ویژگی “واحدهای ترم دوم قبول شده” (امتیاز:  $MI: 0.310207$ ) را به عنوان قوی ترین ویژگی پیش بینی کننده مشخص کرد.
- نقشه همبستگی: همبستگی قوی بین ویژگی های تحصیلی مانند نمرات ترم اول و دوم را نشان داد.



شکل ۷: نقشه گرمایی همبستگی

- **Pairplot:** خوشه های مشخص برای فارغ التحصیلان در ویژگی هایی مانند “نمره پذیرش” و “نمره واحدهای ترم دوم” را آشکار ساخت.



شکل ۸: نمودار جفتی

- نمودار اعتماد قوانین: پنج قانون برتر (مانند قانون ۱ که نمرات متوسط را به ترک تحصیل ربط می‌دهد) برای کمک به تصمیم‌گیری بصری‌سازی شدند.

#### ۳.۶.۴ مشاهدات

- ویژگی‌های تحصیلی بیشترین تأثیر را در پیش‌بینی‌ها دارند، در حالی که عوامل اقتصادی-اجتماعی (مانند "دارای بورس بودن") نقش ثانویه دارند.
- بصری‌سازی‌ها تأیید کردند که قوانین فازی با الگوهای داده منطبق هستند و شفافیت مدل را افزایش می‌دهند.

## ۷.۴ وظیفه ۸: SVM، درخت تصمیم، و رابط گرافیکی (اختیاری)

## ۱.۷.۴ مدل‌های SVM و درخت تصمیم

## SVM Evaluation Metrics:

Accuracy: 0.1853

Precision: 0.1538

Recall: 0.1853

F1 Score: 0.0846

شکل ۹: امتیاز svm

(SVM):

## Decision Tree Evaluation Metrics:

Accuracy: 0.6847

Precision: 0.6912

Recall: 0.6847

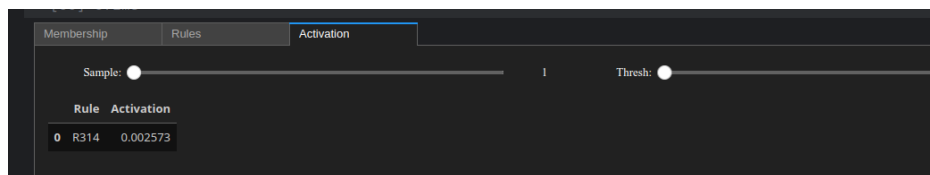
F1 Score: 0.6877

شکل ۱۰: امتیاز درخت تصمیم

درخت تصمیم:

## ۲.۷.۴ رابط گرافیکی (اختیاری)

- یک ابزار تعاملی با استفاده از ipywidgets برای کاوش ویژگی‌ها پیاده‌سازی شد.
- منوی کشویی به کاربر اجازه می‌دهد ویژگی‌های پیوسته (مانند "نمره واحدهای ترم دوم") را انتخاب کرده و درجات عضویت آن‌ها را مشاهده کند.
- تعامل‌پذیری کاربر را افزایش داده و توزیع مجموعه‌های فازی را به صورت پویا نمایش می‌دهد.

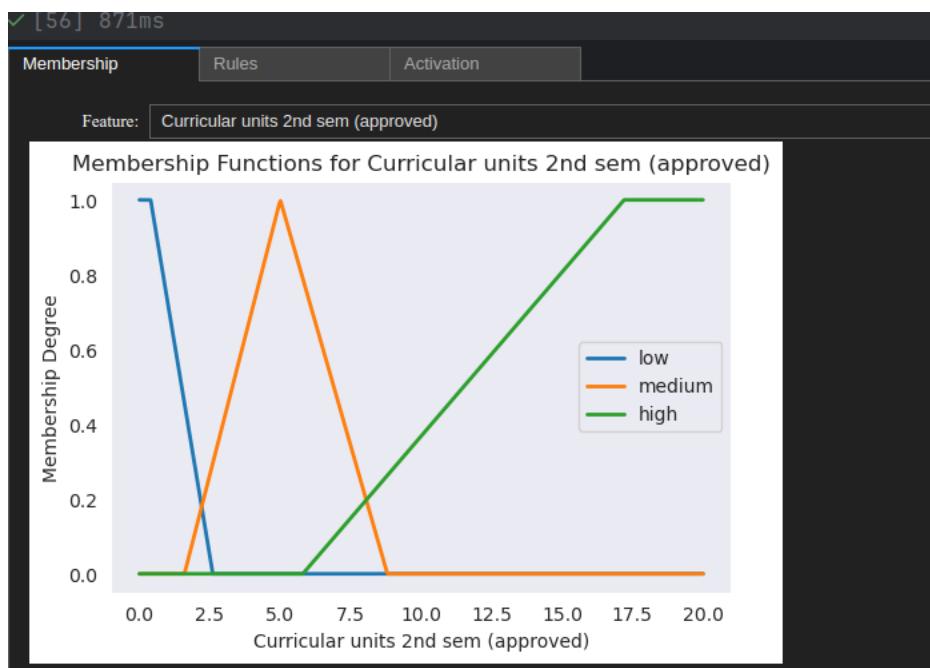


شکل ۱۱: فعال‌سازی توسط قوانین

**Top Rules by Confidence**

	Rule	Confidence
0	IF Curricular units 2nd sem (grade) is high AN...	1.0
1	IF Curricular units 2nd sem (grade) is medium ...	1.0
2	IF Curricular units 2nd sem (grade) is low AND...	1.0
3	IF Curricular units 2nd sem (grade) is low AND...	1.0

شکل ۱۲: جدول قوانین و اطمینان آنها



شکل ۱۳: نمودار تابع عضویت

۳.۷.۴ مشاهدات

- SVM بسیار بدتر از مدل فازی عمل کرده.

- درخت تصمیم بهترین نتایج را می‌دهد.
- رابط گرافیکی دسترسی را افزایش داده و به کاربران غیرفنی اجازه می‌دهد تأثیر ویژگی‌ها را بررسی کنند.