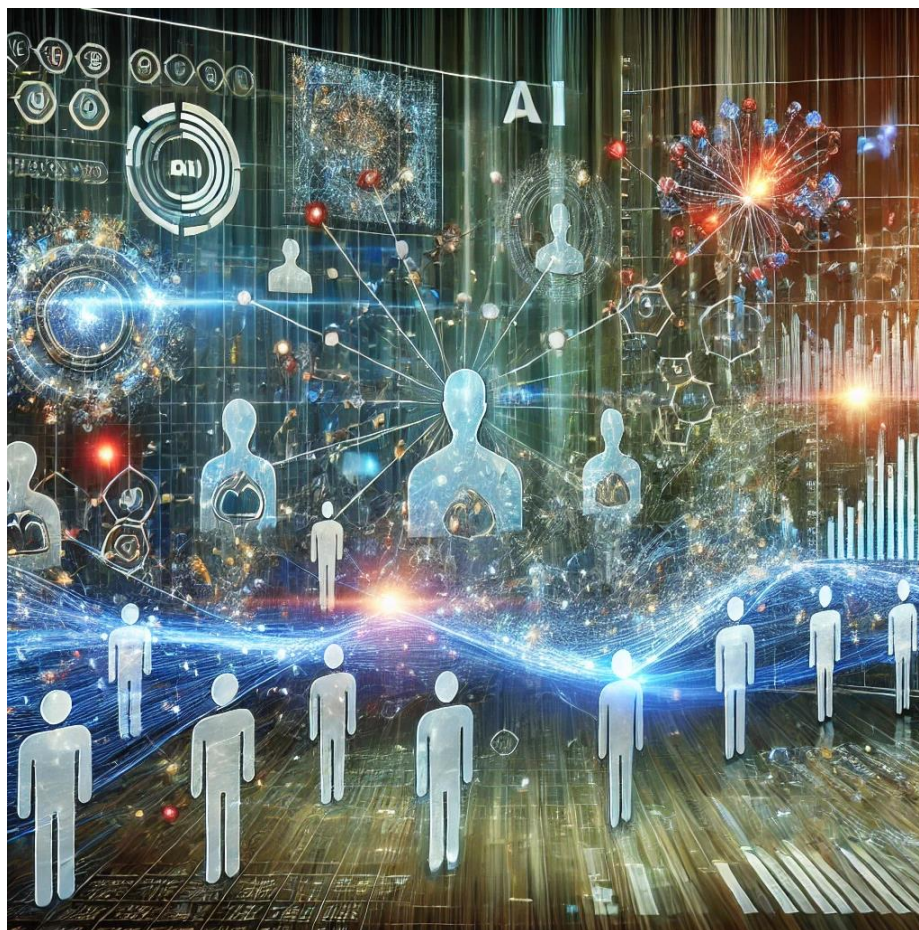


به نام خدا

ارائه الگوریتم پیشرفته و نوین در شناسایی و پیش‌بینی رفتار مشتری:  
تحولی در بهبود تجربه مشتری و ارتقاء درآمدزایی در صنعت پرداخت

تهیه شده توسط تیم هوش مصنوعی شرکت امیدپی



## فهرست مطالب

1.....	مقدمه
1.....	هدف
1.....	معماری پیشنهادی
	انتخاب تکنیک مدل سازی
5.....	ورودی مدل
6.....	الگوریتم انتخابی
7.....	خروجی های مدل
8.....	ارزیابی مدل
9.....	نتیجه گیری

## مقدمه

در دنیای پر رقابت امروز، شناخت دقیق رفتار مشتریان و پیش‌بینی دقیق ریزش آن‌ها برای سازمان‌های مالی و شرکت‌های خدمات پرداخت اهمیت بسیار زیادی دارد. هرگونه ناتوانی در پیش‌بینی این رفتارها می‌تواند به از دست رفتن فرصت‌های درآمدی و افزایش هزینه‌های جذب مشتریان جدید منجر شود. با افزایش پیچیدگی و حجم داده‌های تراکنشی، نیاز به مدل‌های پیشرفته‌تر و دقیق‌تر برای تحلیل این داده‌ها بیش از پیش احساس می‌شود. در این راستا، چارچوبی نوآورانه مبتنی بر گراف پیشنهاد می‌شود که از قدرت شبکه‌های عصبی گراف کانولوشنی برای تحلیل و استخراج ویژگی‌های جدید از روابط بین مشتریان استفاده می‌کند. این مدل نه تنها قادر به پیش‌بینی ریزش مشتریان با دقت بالا است، بلکه توانایی شناسایی رفتارهای نامتعارف را نیز داراست. نتایج بررسی‌ها حاکی از آن است که چارچوب پیشنهادی ما به طور قابل توجهی عملکرد بهتری نسبت به مدل‌های پیشین دارد و می‌تواند دقت و کارایی سازمان‌ها را در مدیریت و حفظ مشتریان به طرز چشمگیری ارتقاء دهد.

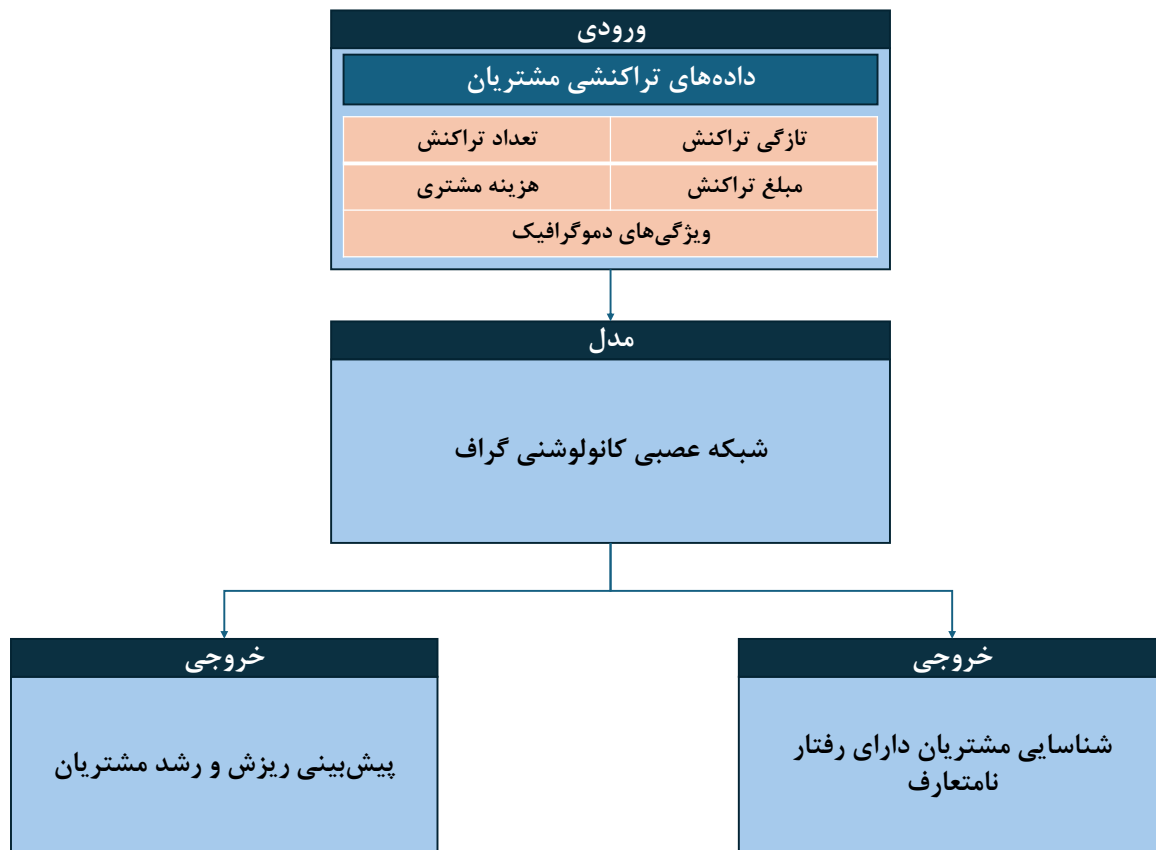
## هدف

هدف از توسعه این الگوریتم، شناسایی دقیق و عمیق مشتریان است. این شناخت به شرکت‌ها کمک می‌کند تا از منابع محدود خود به صورت هوشمندانه و بهینه استفاده کنند و بیشترین بازدهی را به دست آورند. با درک بهتر رفتار مشتریان، می‌توان به سرعت تغییرات در رفتار آن‌ها را شناسایی کرد و در صورتی که مشتری در معرض ریزش باشد، به موقع اقدامات پیشگیرانه را اتخاذ نماید. علاوه بر این، الگوریتم به شرکت‌ها اجازه می‌دهد رفتارهای مشکوک یا نامتعارف مشتریان را شناسایی کرده و با اتخاذ تدابیر مناسب، از مشکلات احتمالی جلوگیری کنند. از دیگر مزایای این الگوریتم، شناسایی مشتریانی است که احتمالاً در آینده نزدیک از خدمات شرکت استفاده نخواهند کرد. این امکان به شرکت اجازه می‌دهد تا با ارائه پیشنهادات ویژه، تخفیفات یا خدمات شخصی‌سازی شده، ریزش مشتریان را کاهش دهد. همچنین، استفاده از این الگوریتم منجر به افزایش رضایت و وفاداری مشتریان می‌شود؛ چرا که با اقدامات به موقع و موثر، مشتریان احساس ارزشمندی بیشتری می‌کنند و ارتباط خود را با شرکت تقویت می‌کنند. در نهایت، این الگوریتم می‌تواند به بهبود عملکرد مالی، کاهش ریسک‌ها، و ارتقاء تجربه مشتریان کمک کند، از آنجا که حفظ مشتریان فعلی هزینه کمتری نسبت به جذب مشتریان جدید دارد. این الگوریتم با پیش‌بینی رفتار مشتریان، به صرفه‌جویی در هزینه‌ها نیز کمک می‌کند.

## معماری پیشنهادی

معماری مدل مفهومی به صورتی است که یک ساختار گرافی برای نمایش ارتباطات مشتریان ایجاد می‌شود. در این گراف، نودها به تعداد مشتریان وجود دارند و به منظور افزودن یال به گراف، ابتدا بر اساس چهار ویژگی تازگی

خرید، تعداد تراکنش، مبلغ خرید و هزینه مشتری، فاصله اقلیدسی بین هر دو مشتری محاسبه می‌شود. سپس با ایجاد یال‌ها بر اساس این فاصله‌ها یک گراف کامل از تمام ارتباطات مشتریان شکل می‌گیرد. از آنجا که فاصله اقلیدسی هر چه کمتر باشد بیانگر شباهت بیشتر بین مشتریان است، مقادیر وزن یال‌ها یک منهای فاصله اقلیدسی در نظر گرفته شده است. این اقدام به معنای این است که هر چه فاصله اقلیدسی کمتر باشد، وزن بیشتری به یال اختصاص می‌یابد و این نشان‌دهنده شباهت بیشتر بین مشتریان است. در ادامه، جهت بهبود کارایی و کاهش ابهامات، برخی از یال‌ها حذف می‌شوند. فرآیند حذف یال‌ها و خلوت‌سازی گراف، به این صورت انجام می‌شود که یال‌های با وزن کمتر حذف می‌شوند تا زمانی که حذف یال بعدی باعث ایجاد گراف ناهمبند نشود، به ترتیب یال‌های با کمترین مقادیر وزن حذف می‌شوند. به این صورت یال‌های با وزن کمتر و غیرقابل اطمینان را که ممکن است نویز به مراحل بعدی الگوریتم وارد کنند حذف شده و تنها اتصالات مهم و ارتباطات موثر در گراف باقی می‌مانند. این عمل به بهبود کارایی الگوریتم یادگیری مدل پیش‌بینی برچسب نهایی کمک می‌کند. ساختار گراف روزانه به دقت می‌تواند تشابه و ارتباطات میان مشتریان را به تصویر کشانده و تحلیل دقیقی از رفتار مشتریان در طول زمان ارائه دهد. این گراف علاوه بر اینکه ارتباطات مشتریان را به صورت واضح نمایش می‌دهد بلکه به پیش‌بینی رفتار مشتریان در روزهای آینده نیز کمک می‌کند. معماری الگوریتم پیش‌بینی رفتار مشتریان طبق تصویر شامل چندین مرحله است. در ادامه فلوچارت مراحل و توضیح مختصر و مفیدی از هر مرحله آورده شده است:



شکل 1. فلوچارت طرح پیشنهادی

## پردازش داده‌های تاریخی مشتریان

- داده‌های تاریخی مربوط به مشتریان پردازش می‌شوند تا ویژگی‌های مختلف برای هر مشتری در هر روز استخراج شود. این ویژگی‌ها می‌توانند شامل اطلاعات تراکنش‌ها، رفتار خرید، تعاملات و سایر مشخصه‌های مرتبط با مشتری باشند.

## ایجاد گراف‌های روابط مشتریان به صورت روزانه

- برای هر روز، گراف‌هایی ایجاد می‌شوند که روابط بین مشتریان را نمایش می‌دهند. این روابط می‌توانند براساس تعاملات، تراکنش‌ها یا دیگر معیارهای ارتباطی باشند. گره‌های گراف‌ها مشتریان و یال‌ها نشان‌دهنده روابط بین آن‌ها هستند.

## تعریف بردار ویژگی برای گره‌های هر گراف

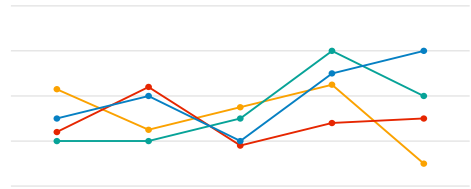
- بردارهای ویژگی برای هر گره (مشتری) در هر گراف تعریف می‌شوند. این بردارها شامل ویژگی‌های استخراج شده در مرحله اول هستند و ممکن است شامل اطلاعاتی مانند تعداد تراکنش‌ها، میزان خرید، زمان تعاملات و غیره باشند.

## ایجاد نمونه‌های آموزشی برای هر روز پیش‌بینی

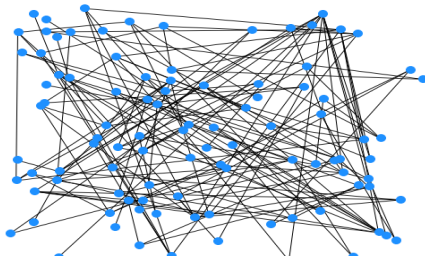
- یک مجموعه آموزشی برای هر روز پیش‌بینی ایجاد می‌شود که شامل گراف‌های چند روز گذشته خواهد بود که این نمونه‌ها برای آموزش مدل استفاده می‌شوند.

## آموزش مدل GCN و پیش‌بینی برچسب‌های نهایی گره‌ها در روز $t$ (time)

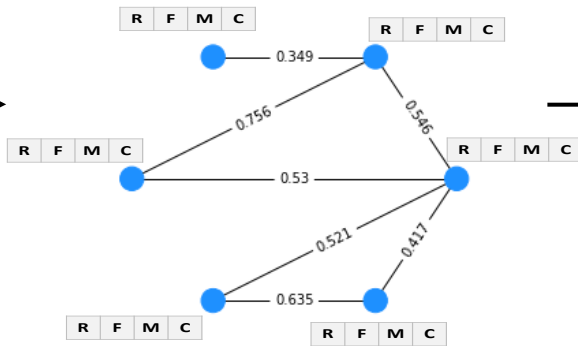
- مدل GCN (Graph Convolutional Network) آموزش داده می‌شود تا برچسب‌های نهایی گره‌ها (مشتریان) در روز مورد نظر را پیش‌بینی کند. این برچسب‌ها نشان می‌دهند که آیا مشتری در معرض ریزش است یا خیر. مشتریانی که پیش‌بینی می‌شود ریزش کنند با رنگ قرمز (Fall prediction) و مشتریانی که احتمال ماندن دارند با رنگ سبز (Rise prediction) نشان داده می‌شوند.



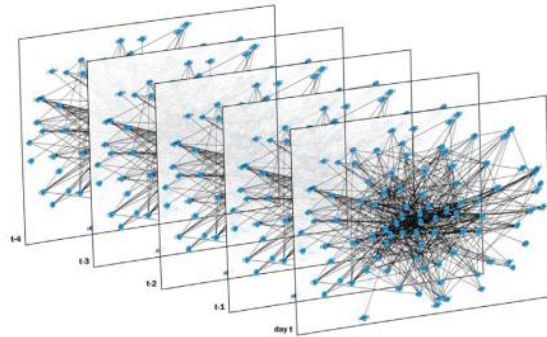
1. پردازش داده‌های تاریخی مشتریان و بدست آوردن مقادیر ویژگی‌ها برای هر مشتری در هرروز



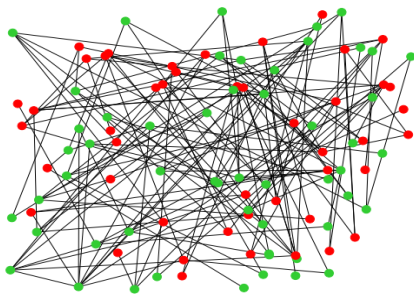
2. ایجاد گراف روابط مشتریان در هر روز



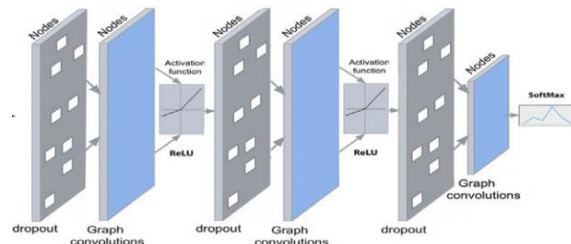
3. تعریف بردار ویژگی برای هر گره‌های هر گراف



4. یک نمونه آموزشی برای هر روز پیش‌بینی ایجاد می‌شود (با استفاده از نمودارهای روز های قبل)



6. گزارش برچسب‌های گره پیش‌بینی‌شده نهایی به عنوان رفتار بعدی مورد انتظار برای مشتری مربوطه



5. آموزش مدل GCN و پیش‌بینی برچسب‌های نهایی گره های روز مشخص

شکل 2. معماری مدل پیشنهادی

## انتخاب تکنیک مدل سازی

در پروژه پیش‌بینی رفتار مشتریان، انتخاب تکنیک مدل سازی مناسب از اهمیت بالایی برخوردار است. برای این منظور، از الگوریتم Graph Convolutional Network (GCN) استفاده شده است که به دلیل توانایی بالای آن در درک و تحلیل داده‌های شبکه‌ای و گراف، انتخاب شده است.

## ورودی مدل

ورودی‌های مدل شامل دو دسته داده اصلی است:

1. **داده‌های تراکنشی:** این دسته شامل اطلاعات مرتبط با تراکنش‌های مشتریان مانند تعداد تراکنش‌ها، مقدار تراکنش‌ها، نوع تراکنش‌ها و زمان‌بندی تراکنش‌ها می‌باشد.
2. **داده‌های دموگرافیک:** این دسته شامل اطلاعات دموگرافیک مشتریان مانند سن، جنسیت، موقعیت جغرافیایی، وضعیت شغلی و درآمد می‌باشد.

این داده‌ها پس از پیش‌پردازش و آماده‌سازی به عنوان ورودی به الگوریتم GCN داده می‌شوند.

## الگوریتم شبکه عصبی کانولوشنی گراف (GCN)

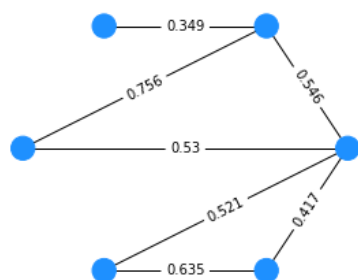
الگوریتم شبکه عصبی کانولوشنی گراف به دلیل توانایی آن در کار با داده‌های ساختار یافته و نیمه‌ساختار یافته انتخاب شده است. قادر است ارتباطات و تعاملات پیچیده میان داده‌ها را در نظر بگیرد و از این طریق به پیش‌بینی دقیق‌تر رفتار مشتریان بپردازد.

## فرآیند مدل‌سازی

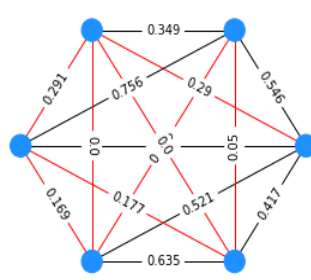
1. **جمع‌آوری داده‌ها:** برای ارزیابی عملکرد مدل ارائه شده، از مجموعه داده‌های مشتریان شرکت خدمات پرداخت الکترونیک استفاده شده است که شامل تازگی تراکنش، تعداد تراکنش، مبلغ تراکنش و هزینه مشتری است. در جدول 1 توضیحات ویژگی‌ها آورده شد است.
2. **پیش‌پردازش داده‌ها:** داده‌های جمع‌آوری شده تمیز، نرمال‌سازی و به فرمتی که برای مدل GCN مناسب باشد تبدیل می‌شوند.
3. **ساختاردهی به داده‌ها:** داده‌های ورودی به صورت گراف مدل‌سازی می‌شوند که در آن گره‌ها نمایانگر مشتریان و یال‌ها نمایانگر ارتباطات تراکنشی میان آن‌ها هستند. مراحل ساخت گراف در شکل 3 آورده شده است.

جدول 1. ویژگی‌هایی که برای نمایش گره‌ها در گراف استفاده شده است.

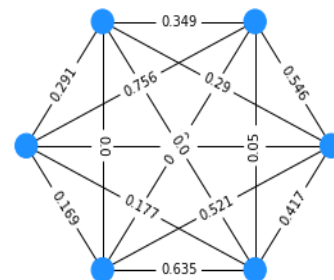
ویژگی	توضیح ویژگی
تازگی تراکنش	تعداد روزهای سپری شده از آخرین تراکنش مشتری، این ویژگی هر چقدر کمتر باشد بهتر است.
تعداد تراکنش	تعداد تراکنش‌هایی که مشتری در بازه یک روزه مورد بررسی داشته است، این ویژگی هر چقدر بیشتر باشد بهتر است.
مبلغ تراکنش	میزان مبلغ که از تراکنش‌های یک مشتری در بازه زمانی یک روزه به دست می‌آید، این ویژگی هر چقدر بیشتر باشد بهتر است.
هزینه مشتری	میزان هزینه‌ای که سازمان برای یک مشتری در بازه زمانی یک ماهه می‌پردازد، این هزینه شامل بازدیدهای دوره‌ای، رول مصرفی و سایر خدماتی که به مشتری ارائه می‌شود است. این ویژگی هر چقدر کمتر باشد بهتر است.
دموگرافیک	این دسته شامل اطلاعات دموگرافیک مشتریان مانند صنف، جنسیت، موقعیت جغرافیایی و وضعیت شغلی می‌باشد.



مرحله 3: یال‌های به رنگ قرمز از کمترین مقدار شروع به حذف شدن می‌کنند تا جایی که گراف ناهمبند نشود.



مرحله 2: یال‌هایی که وزن آنها کمتر از حد آستانه است با رنگ قرمز مشخص شده‌اند.



مرحله 1: مشخص نمودن وزن تمامی یال‌ها بین مشتریان بر اساس فاصله اقلیدسی، وزن بیشتر به معنای شباهت بیشتر بین دو مشتری است.

شکل 3. مراحل ساخت گراف

4. آموزش مدل: مدل GCN با استفاده از داده‌های ورودی آموزش داده می‌شود. در این مرحله، مدل با

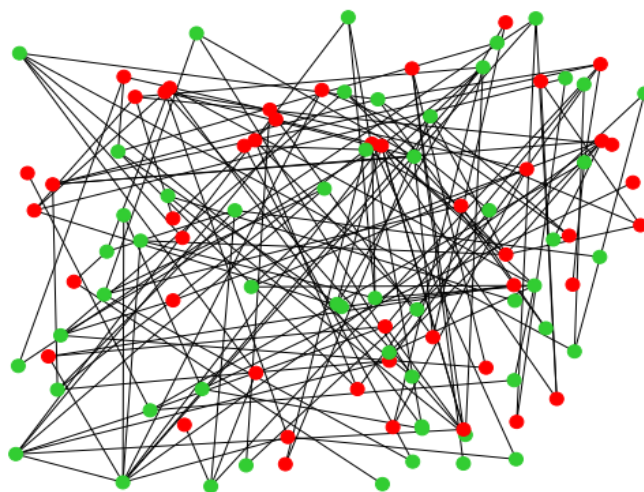
تحلیل ویژگی‌ها و ارتباطات موجود در داده‌ها، یاد می‌گیرد که چگونه مشتریانی که در معرض رفتار هستند را شناسایی کند.

5. پیش‌بینی: پس از آموزش مدل، داده‌های جدید به مدل داده می‌شوند و مدل پیش‌بینی می‌کند که کدام مشتریان احتمال رفتار دارند.



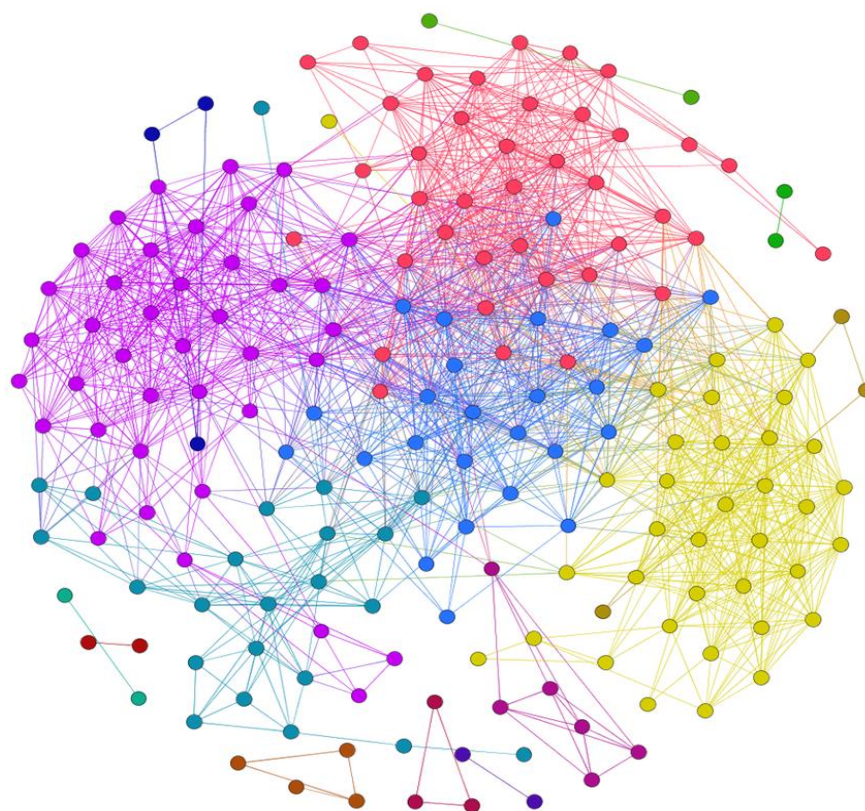
## خروجی‌های مدل

خروجی مدل، پیش‌بینی احتمال رفتار برای هر مشتری و همچنین شناسایی مشتریانی که رفتار نامتعارف دارند، است. این خروجی به تیم‌های بازاریابی و خدمات مشتریان کمک می‌کند تا تدابیر پیشگیرانه‌ای برای حفظ مشتریان اتخاذ کنند. در شکل 4 نودهای سبز رنگ مشتریانی هستند که رفتار نرمال دارند و نودهای قرمز رنگ مشتریان در معرض ریزش هستند.



شکل 4. رفتار بعدی مورد انتظار برای مشتری

در این طرح پیشنهادی، ما از مدل‌سازی روابط بین مشتریان به صورت گراف استفاده کرده‌ایم. هر گره در این گراف نمایانگر یک مشتری و هر لبه نشان‌دهنده تعامل یا ارتباطی است که بین آن‌ها وجود دارد. مشتریانی که تعاملات قوی‌تر و بیشتری با یکدیگر دارند، به صورت خودکار در یک کلاستر قرار می‌گیرند. این کلاستربندی به ما امکان می‌دهد که الگوهای رفتاری مشابه را شناسایی کرده و از طرفی، رفتارهای غیرمتعارف و غیرعادی را به سرعت تشخیص دهیم. برای مثال، اگر یکی از مشتریان در یک کلاستر رفتار غیرمعمولی انجام دهد، این الگوریتم به سرعت آن را شناسایی و اطلاع‌رسانی می‌کند. این قابلیت به شرکت‌ها کمک می‌کند تا به موقع به تغییرات غیرمنتظره در رفتار مشتریان پاسخ دهند و از مشکلات احتمالی جلوگیری کنند. در شکل 5 نمایی از گراف روابط بین مشتریان نشان داده شده است که با کلاستربندی، گروه‌های مشابه را شناسایی کرده و رفتارهای غیرمعمول را به سرعت تشخیص می‌دهد.



شکل 5. کلاسترینگ مشتریان برای تشخیص رفتارهای غیرمعارف

## ارزیابی مدل

در معیارهای ارزیابی پیش‌بینی رفتار مشتریان، از دو معیار دقت و ضریب همبستگی متیوز<sup>1</sup> به عنوان معیارهای ارزیابی پیش‌بینی استفاده می‌شود. با توجه به ماتریس درهم‌ریختگی که شامل تعداد نمونه‌هایی است که به عنوان مثبت درست، مثبت نادرست، منفی درست و منفی نادرست دسته‌بندی شده‌اند. برای محاسبه دقت و ضریب همبستگی متیوز از ماتریس اشتباهات استفاده می‌شود. دقت به صورت تعداد نمونه‌هایی که به درستی دسته‌بندی شده‌اند تقسیم بر تعداد کل نمونه‌ها می‌شود که در فرمول 1 مشاهده می‌شود. ضریب همبستگی متیوز نیز طبق فرمول 2 تعریف می‌شود.

$$ACC = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

$$MCC = \frac{t_p \times t_n - f_p \times f_n}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \quad (2)$$

<sup>1</sup> Matthews Correlation Coefficient

جدول 2. عملکرد مدل معرفی شده در مقایسه با سایر الگوریتم‌ها

مدل	صحت	ضریب همبستگی متیوز
GCN	0.78963	0.16
MLP	0.43141	-0.00126
RNN	0.56940	0.08799
GRU	0.56946	0.07659
CNN	0.51013	0.00276

مدل GCN توانسته است به دقت 78٪ در پیش‌بینی رفتار مشتریان دست یابد. این دقت نشان‌دهنده کارایی مدل در شناسایی مشتریانی است که احتمال ریزش دارند، اگرچه هنوز جای بهبود وجود دارد.

## نتیجه گیری

در این طرح پیشنهادی، یک چارچوب نوآورانه مبتنی بر گراف ارائه شده است که روابط بین مشتریان را به عنوان یک ساختار گراف مدل می‌کند. این ساختار گراف توسط یک شبکه عصبی کانولوشنی گراف تحلیل می‌شود و نمایش‌های جدیدی برای گره‌ها استخراج می‌گردد. این نمایش‌ها به منظور پیش‌بینی ریزش یا عدم ریزش تمامی مشتریان استفاده می‌شود. علاوه بر این، مدل قادر است مشتریانی با رفتارهای نامتعارف را نیز شناسایی کند. بررسی‌ها نشان می‌دهند که مدل پیش‌بینی مبتنی بر گراف به طور قابل توجهی بهتر از الگوریتم‌های پیش‌بینی پیشین عمل می‌کند و در مقایسه با مدل‌های معمول، دقت و همبستگی متیوز را به طور چشمگیری بهبود می‌بخشد. به طور کلی، تحلیل‌های انجام شده روی مدل پیشنهادی نشان‌دهنده ارتقاء قابل توجه عملکرد آن در مقایسه با شبکه‌های عصبی دیگر است. این بهبود عملکرد نه تنها نشان‌دهنده قدرت مدل پیشنهادی در پیش‌بینی رفتار مشتریان است، بلکه اثربخشی بالاتر آن را نسبت به روش‌های مبتنی بر گراف نیز تأیید می‌کند. نتایج نشان می‌دهند که مدل پیشنهادی در پیش‌بینی رفتار مشتریان از میان سایر مدل‌های مقایسه‌ای عملکرد برجسته‌ای دارد و قدرت بالای استفاده از گراف‌ها در بهبود پیش‌بینی رفتار مشتریان را به نمایش می‌گذارد.