

[https://oig.hhs.gov/exclusions/exclusions\\_list.asp](https://oig.hhs.gov/exclusions/exclusions_list.asp)

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier>

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber>

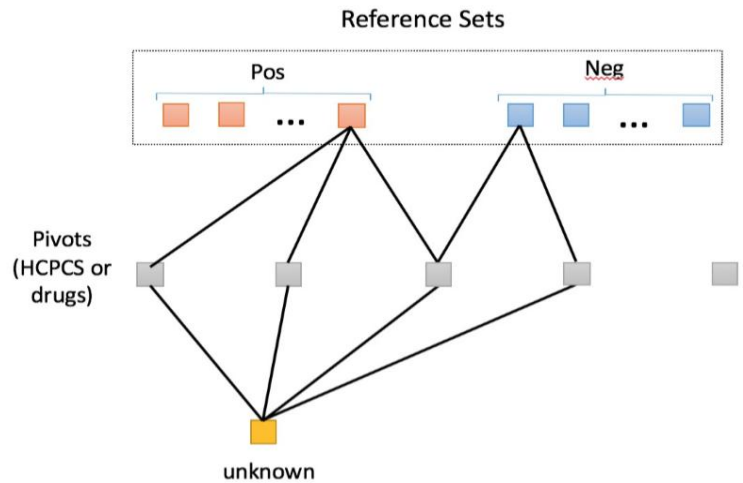
[https://download.cms.gov/nppes/NPI\\_Files.html](https://download.cms.gov/nppes/NPI_Files.html)

## بخش اول:

استفاده از یک دیتابیس مبتنی بر گراف و استخراج 7 ویژگی مبتنی بر گراف از چهار دیتاست فوق

هر ارائه دهنده خدمات پزشکی، هر تجویز دارویی و هر کد درمانی Healthcare Common Procedure Coding System (HCPCS) بعنوان یک نود جدا نمایش داده شد. ارائه دهندگان به هر دارویی که توسط آنها تجویز شده باشد از طریق یک یال PRESCRIBED (تجویز شده) پیوند خورده اند که صفات (ویژگیهای) آن، تعداد بیمارانی که توسط ارائه دهنده نسخه دارویی دریافت کرده اند، کل هزینه تجویزها و سایر داده های Part-D را نمایش میدهد. ارائه دهندگان به هر HCPCS (کد درمان) از طریق یک یال CHARGE\_OF (سرانه خدمات درمان) پیوند خورده اند که ویژگی های آن شامل متوسط (میانگین) پرداخت مدیر (بیمه درمانی سالمندان)، تعداد ذینفعان منحصربفرد (unique / غیر تکراری) و سایر اطلاعات درمانی است که در قالب ستون ها در دیتاست PUF ظاهر میشوند.

ما هر ارائه دهنده ناشناس را با هر عضو از مجموعه های مرجع مثبت (دارای محرومیت) و منفی (فاقد محرومیت) مقایسه میکنیم. شباهت هر جفت از ارائه دهندگان، بنوبه خود، از طریق مقایسه نودهای محوری مجاور هریک تعیین میشود در جاییکه محورها یا مشخص کننده نوع درمان (HCPCS) هستند یا نوع دارو.



برای هر ارائه دهنده (ناشناس یا عضو مجموعه مرجع) ما یک بردار رفتار تعریف میکنیم که دربرگیرنده برداری از وزن یالهای رابط به هر نود محوری مجاور است. مقدار هر عبارت متناظر با یک محور مدنظر در بردار صفر است اگر نود ارائه دهنده هیچ یالی به آن محور نداشته باشد و در غیر اینصورت (بین دو نود یال وجود داشته باشد) برابر ارزشی محاسبه شده به روشی خاص است که به نوع محور وابسته است. در این تحقیق، وزن یک یال (رابط) به یک نود نوع دارو، کل هزینه دارو (TotalDrugCost) است. به معنی کل مقدار فاکتور شده (صورت حساب برای آن صادر شده) توسط این ارائه دهنده برای این داروی خاص (این مقدار یک ستون از دیتاست Part-D است). وزن یک یال (رابط) به یک نود نوع درمان (HCPCS) بصورت حاصلضرب BeneUniqCount و AvgMedicarePayment محاسبه میشود که بترتیب نمایانگر متوسط پرداخت مدیکر (میانگین پرداخت های صورت گرفته برای بهره مندی از خدمات درمانی مدیکر به این ارائه دهنده) و تعداد دینفعان منحصربفرد (تعداد افراد تحت تکفل این نوع از بیمه درمانی که مشتری این ارائه دهنده اند بدون شمارش تکرارها) است. شباهت میان هر جفت ارائه دهنده با توجه به شکلی از رفتار مدنظر (همچون تجویز دارو یا صدور فاکتور (صورت حساب)) میتواند بوسیله اعمال یک معیار شباهت برداری استاندارد بر بردارهای رفتاری ارائه دهندگان برآورد گردد (تخمین زده شود).

معیار شباهت: cosine

برای تبدیل کسینوس وزن دهی شده با معکوس درجه ، بین یک ارائه دهنده و عضو مجموعه های مرجع مثبت یا منفی به ویژگی های قابل استفاده جهت تخمین ( برآورد) ریسک ، ما میانگین K نزدیکترین عضو هر مجموعه را گرفته و سه ویژگی تولید میکنیم :

- مشابهت منفی ( negative-similarity) : میانگین شباهت به K نزدیکترین عضو مجموعه مرجع منفی
- مشابهت مثبت ( positive-similarity) : میانگین شباهت به K نزدیکترین عضو مجموعه مرجع مثبت
- نسبت مشابهت منفی و مثبت ( neg-sim-ratio) : نسبت ویژگی های ۱ و ۲ بهم

تکرار این محاسبات برای هر دو نوع محور ، ۶ ویژگی مقداردهی شده واقعی تولید میکند :

۱. مشابهت منفی HCPCS

۲. مشابهت مثبت HCPCS

۳. نسبت مشابهت HCPCS ( کد درمان )

۴. مشابهت منفی دارو

۵. مشابهت مثبت دارو

۶. نسبت مشابهت دارو

#### **:Provider attributes**

۷. تعداد بیماران جدید به نسبت ذی نفع ( new-patients-per-bene) : نسبت سرویس های ارائه شده به بیماران جدید. (باور کلی اینست که ارائه دهنده ای که تعداد زیادی خدمات به هر بیمار جدید ارائه میدهد، ممکن است سرویس های بیش از ضرورت های درمانی ارائه دهد).

ما بر روی ۱۲۰۰۰ ارائه دهنده محرومیت دار که هر کدام دستکم یک تجویز دارویی ، خدمت درمانی یا آدرس ( یعنی یک لینک CHARGE OF، LOCATED AT یا PRESCRIBED ) آزمایش انجام می دهیم. ارائه

دهندگان فاقد این لینک ها ( پیوندها ) از گراف جدا میشوند و بنابراین در معرض تجزیه تحلیل های گراف قرار ندارند(حذف می شوند).

در انجام اینکار ما محورهای با درجه راس بزرگتر از ۱۰۰۰۰۰ را از محاسبه تشابه بردار رفتار حذف میکنیم ( با این فرض که محورهایی با اینچنین درجه بالا، توانایی تفکیک اندکی داشته و محاسبه را با کندی مواجه میسازند).

نتایج اجرای یک الگوریتم درخت تصمیم با ورودی ویژگی های فوق به صورت زیر است:

Feature set	f-measure	ROC area
HCPCS behavior- similarity features	0.715	0.718
Drug behavior similarity features	0.657	0.666
Provider attributes	0.697	0.697