

# Information Value

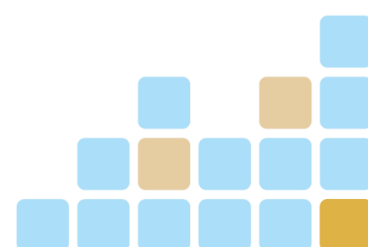
همانطور که در مقاله‌های «[فرآیند توسعه کارت امتیازی](#)» و «[Weight Of Evidence](#)» شرح داده شد؛ کارت امتیازی مجموعه‌ای از متغیرها، مشخصه‌ها و امتیازهای آن‌ها است که با توجه به هدف سازمان متقاضی اعتبارسنجی طراحی می‌شود. انتخاب متغیرها و مشخصه‌ها یکی از گام‌های فرآیند طراحی کارت امتیازی است، به صورتی که متغیرهای در دسترس شناسایی شده و قدرت پیش‌بینی هر کدام به‌طور جداگانه محاسبه می‌شود. سپس متغیرهایی که قدرت پیش‌بینی معنی‌داری دارند در کارت امتیازی به کار گرفته می‌شوند.

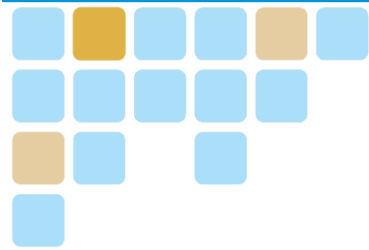
به منظور بررسی قدرت پیش‌بینی هر متغیر از مفهومی به نام ارزش اطلاعاتی یا Information Value استفاده می‌شود. مقدار ارزش اطلاعاتی بر اساس رابطه (۱) بدست می‌آید.

$$IV = \sum_{i=1}^n (Distr\ Good_i - Distr\ Bad_i) * \left[ \ln \left( \frac{Distr\ Good_i}{Distr\ Bad_i} \right) \right] \quad (1)$$

به عبارتی دیگر، برای محاسبه ارزش اطلاعاتی از توزیع خوش رفتارها و بد رفتارها در هر یک از مشخصه‌ها استفاده می‌شود.

در یک نگاه کلی می‌دانیم که قدرت پیش‌بینی متغیرها و در نتیجه سهم امتیازی آن‌ها در کارت امتیازی در چهار دسته طبق‌بندی می‌شود که آن‌ها را در جدول ۱ مشاهده می‌کنید.





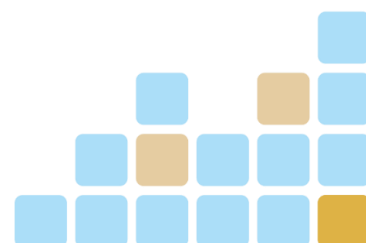
جدول ۱. قدرت پیش‌بینی متغیرها براساس مقدار IV

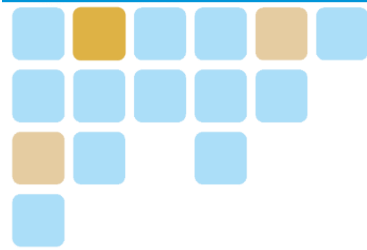
قدرت پیش‌بینی	مقدار IV
صفر	کمتر از ۰,۰۲
ضعیف	۰,۱ < مقدار < ۰,۰۲
متوسط	۰,۳ < مقدار < ۰,۱
قوی	بزرگتر از ۰,۳

باتوجه به این جدول، چنانچه مقدار ارزش اطلاعاتی از ۰,۲ کمتر باشد، متغیر مورد بررسی نسبت احتمال خوش رفتار/ بدرفتار را متمایز نمی‌کند. در بازه مقداری (۰,۱, ۰,۰۲, ۰) متغیر رابطه ضعیفی با نسبت احتمال خوش رفتار/ بدرفتار داشته و برای دو دسته دیگر نیز شرح متغیر به همین ترتیب در جدول ۱ آمده است.

برای مثال متغیر X را در نظر بگیرید که دارای مشخصه‌های A, B, C, D, E, F است. با توجه به خروجی متغیر هدف (خوش رفتار یا بدرفتار) بررسی می‌کنیم که چه تعداد از افراد با خروجی «خوش رفتار» در مشخصه A قرار گرفته‌اند، همان‌طور که در جدول ۲ ذکر شده این تعداد برابر ۱۶۳ نفر است. به همین ترتیب این تعداد برای خروجی «بدرفتار» در این مشخصه و دیگر مشخصه‌ها محاسبه می‌شود. توزیع نسبی خوش رفتارها بر اساس تقسیم تعداد خوش رفتار در آن مشخصه بر تعداد کل خوش رفتارها بدست می‌آید. در مشخصه A توزیع خوش رفتارها از رابطه زیر محاسبه می‌شود:

$$Distr\ Good_A = \frac{163}{1123} = 0.1451$$



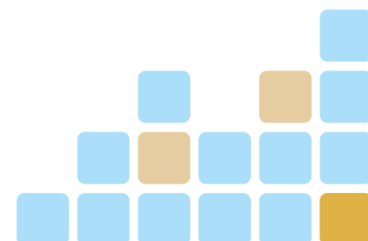


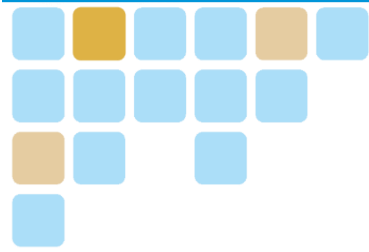
در نهایت بر اساس توزیع نسبی خوش رفتارها و بد رفتارها در هر شش مشخصه F,E,D,C,B,A عدد ارزش اطلاعاتی مطابق رابطه ۱ محاسبه شده که برابر ۰,۱۵۶ است. طبق جدول ۱ این مقدار بیشتر از ۰,۱ بوده و قدرت پیش‌بینی متغیر X متوسط است.

جدول ۲. محاسبه IV

متغیر X	تعداد خوش رفتارها	تعداد بد رفتارها	تعداد کل	توزیع خوش رفتارها	توزیع بد رفتارها	IV	WOE
A	۱۶۳	۱۴۴	۳۰۷	۰,۱۴۵۱	۰,۱۹۸۶	۰,۰۱۶۷۷	-۳۱,۳۷
B	۷۸	۱۵	۹۳	۰,۰۶۹۵	۰,۰۲۰۷	۰,۰۵۹۰۶	۱۲۱,۱۱
C	۳۱۶	۱۲۳	۴۳۹	۰,۲۸۱۴	۰,۱۶۹۷	۰,۰۵۶۵۳	۵۰,۶۰
D	۲۳	۱۲	۳۵	۰,۰۲۰۵	۰,۰۱۶۶	۰,۰۰۰۸۳	۲۱,۳۰
E	۱۹۸	۱۵۲	۳۵۰	۰,۱۷۶۳	۰,۲۰۹۷	۰,۰۰۵۷۷	-۱۷,۳۲
F	۳۴۵	۲۷۹	۶۲۴	۰,۳۰۷۲	۰,۳۸۴۸	۰,۰۱۷۴۸	-۲۲,۵۳
مجموع	۱۱۲۳	۷۲۵	۱۸۴۸	۱	۱	۰,۱۵۶۴۶	

حال فرض کنید که چندین متغیر همانند X با مشخصه‌های مدنظر، کاندیدای حضور در کارت امتیازی هستند. ارزش اطلاعاتی این متغیرها همانند مثال فوق محاسبه می‌شود. هرچه مقدار ارزش اطلاعاتی بیشتر باشد، متغیر موثرتری در پیش‌بینی متغیر هدف می‌باشد. بدین ترتیب متغیرها بر اساس مقدار ارزش اطلاعاتی مرتب شده، سپس متغیرهای معنی‌دار در کارت امتیازی مورد استفاده قرار می‌گیرند.





## منابع

۱. [Louzada, F., Ara, A., & Fernandes, G. B. \(2016\). Classification methods applied to credit scoring: Systematic review and overall comparison. \*Surveys in Operations Research and Management Science\*, 21\(2\), 117-134.](#)
۲. [Siddiqi, N. \(2017\). \*Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring \(2<sup>nd</sup> ed\)\*. John Wiley & Sons, Inc.](#)
۳. [Thomas, L., Crook, J. and Edelman, D. \(2017\). \*Credit scoring and its applications \(Vol. 2\)\*. Siam.](#)

ما را در شبکه‌های اجتماعی با [@icbsco](#) دنبال کنید

