

تجزیه کلاستر

تألیف: دکتر غلامحسین حسینی

عضو هیئت علمی مؤسسه تحقیقات بیه کشور

تبدیل شدن غیرمجاز هر کتاب به فایل pdf یا powerpoint و انتشار آن در اینترنت؛ قلم نویسنده را برای تولید علم کند می گرداند، چراغ ناشر و کتابفروش بی فروغ می شود و صدها کارگر چاپخانه، لیتوگرافی و صحافی از ادامه کار باز می مانند.
اگر از حق معنوی کپی رایت که یک حق انسانی است بگذریم. کپی و انتشار ناجوانمردانه از کتابها، نمونه سرقت و مال دزدی در روزگار ما نیست؟!
این کار از اهل فرهنگ که مایه سربلندی کشور هستند، انتظار نمی رود.



یادواره کتاب

سرشناسه
 عنوان و نام پدیدآور : تجزیه کلاستر / تألیف غلامحسین حسینی.
 مشخصات نشر : تهران: یادواره کتاب، ۱۳۹۴.
 مشخصات ظاهری : ۱۰۴ ص: مصور، جدول، نمودار.
 شابک : ۹۷۸-۶۰۰-۶۲۳۱-۰۷-۶
 وضعیت فهرست نویسی : فیا
 یادداشت : کتابنامه
 موضوع : تجزیه و تحلیل کلاستر
 موضوع : داده کاوی
 رده بندی کنگره : QA۲۷۸/ت۵۳ ۱۳۹۴
 رده بندی دیویی : ۵۹۱/۵۲
 شماره کتابشناسی ملی : ۳۸۸۱۹۹۸



انتشارات یادواره کتاب

عنوان کتاب : تجزیه کلاستر
 نویسنده : دکتر غلامحسین حسینی
 لیتوگرافی : آیکان
 چاپخانه : دانش پژوه
 چاپ : اول، ۱۳۹۴
 شمارگان : ۱۰۰۰ نسخه
 قیمت : ۹۰۰۰ تومان
 شابک : ۹۷۸-۶۰۰-۶۲۳۱-۰۷-۶

حق چاپ برای انتشارات یادواره کتاب محفوظ است

نشانی: تهران، خیابان انقلاب، خیابان فخر رازی، خیابان وحید نظری، پلاک ۸۲

تلفن: ۶۶۴۸۸۸۱۸-۶۶۴۱۹۹۸۲

www.yadvareketab.ir

فهرست

پیش‌گفتار.....	۹
مقدمه.....	۱۱
تعریف.....	۱۴
الگوریتم.....	۱۷
ضرایب شباهت صفات کمی.....	۲۳
ضریب متوسط فاصله اقلیدسی.....	۲۵
ضریب تفاوت شکل.....	۲۶
ضریب کسینوس.....	۲۶
ضریب همبستگی.....	۲۷
ضریب متریک کانبرا.....	۲۷
ضریب بری-کرتیس.....	۲۷
فاصله مینسکاوسکی.....	۲۸
روش‌های سلسله‌مراتبی.....	۳۳
روش تجمعی.....	۳۵
روش UPGMA.....	۳۵
روش SLINK.....	۳۶

۳۷		CLINK	روش
۳۷		WARD	روش
۳۸		CENTROID	روش
۵۸			روش تقسیمی
۵۸			روش تقسیمی تک واحدی
۵۸			روش تجزیه همبستگی
۶۰			روش تشخیص اثر متقابل خودکار
۶۲			روش های تقسیمی چند واحدی
۶۷			روش های تجزیه ای
۶۹		K	روش میانگین
۷۳			روش دو مرحله ای
۸۳			ضریب شباهت صفات کیفی
۸۷			ضریب جاکارد
۸۷			ضریب انطباق ساده
۸۸			ضریب یول
۸۸			ضریب هامان
۸۸			ضریب سورنسون
۸۹			ضریب راجرز و تانیموتو
۸۹			ضریب سوکال و اسنپس

- ۹۰ ضریب راثو و راسل
- ۹۰ ضریب بارونی - اوربانی و بوسر
- ۹۱ ضریب فاصله دو تایی سوکال
- ۹۱ ضریب اوکیایی کسینوس
- ۹۲ ضریب فی
- ۹۳ ضمائم
- ۹۵ I. سطح زیر منحنی نرمال استاندارد از ۰ تا Z
- ۹۶ II. مقادیر درصدی (t_p) توزیع t- استیودنت با ν درجه آزادی
- ۹۷ III. مقادیر درصدی (x_p^2) توزیع کای اسکوتر با ν درجه آزادی
- ۹۸ IV. مقادیر ۹۵ درصدی توزیع F با درجات آزادی ν_1 و ν_2
- ۹۹ V. مقادیر ۹۹ درصدی توزیع F با درجات آزادی ν_1 و ν_2
- ۱۰۰ VI. مقادیر بحرانی r پیرسون
- ۱۰۱ VII. جدول اعداد تصادفی
- ۱۰۳ منابع
- ۱۰۴ منابع فارسی
- ۱۰۴ منابع انگلیسی

پیش‌گفتار

پژوهش‌گرانی که اقدام به طراحی و اجرای طرح‌های تحقیقاتی در گرایش‌های مختلف علوم می‌کنند قبل از هر کس دیگری نسبت به فرآیند اجراء، کیفیت و کمیت نتایج و فرض‌های صفر و یک آنها اشرافیت و پیش‌بینی‌های لازم را دارند. ولی این پژوهش‌گران برای اقناع مخاطبان از دستاوردها و همچنین انتشار یافته‌های علمی خود، نیازمند دسته‌بندی و ساده‌تر کردن آنها برای ارائه به محققین، نشریات و مجامع علمی هستند. به همین منظور می‌بایست مراحل اجرای پژوهش و بسته‌بندی داده‌های آزمایشی و تجزیه و تحلیل نتایج آنها در چارچوب طرح‌ها و مدل‌های آماری قابل فهم‌تری طبقه‌بندی و ارائه شود. تجزیه کلاستر از جمله روش‌های دسته‌بندی آماری خاصی است که منجر به تجمع و ساده‌تر شدن نتایج تحقیقات و کاهش سودمند داده‌های اولیه آماری، آسان‌سازی ساختار آنها، دسته‌بندی اطلاعات افراد با ضرایب همبستگی بالا در درون گروه و در نتیجه دسته‌بندی آنها در کلاسترهای خاص مفهوم‌دار برای مخاطبان علمی عرصه‌های مختلف علوم پایه، انسانی، کشاورزی، پزشکی، مهندسی و علوم دیگر می‌شود. لذا کتاب حاضر در راستای استفاده عملی و کاربردی از روش تجزیه کلاستر جهت تجزیه و تحلیل داده‌های خام پژوهشی حاصل از طرح‌ها و پروژه‌های تحقیقاتی و تبدیل آنها به دستجات و کلاسترهای مفهوم‌دار به نگارش در آمده است.

غلامحسین حسینی

عضو هیئت علمی مؤسسه تحقیقات پنبه کشور

تجزیه کلاستر نوعی گروه‌بندی است که در آن مجموعه‌ای از افراد با استفاده از روش‌های تعیین شده در گروه یا کلاستر مشابهی قرار می‌گیرند که نسبت به افراد گروه یا کلاسترهای دیگر از درجه مشابهت بیشتری برخوردار هستند. تجزیه کلاستر به خودی خود یک الگوریتم یا روش محاسباتی خاص نیست، بلکه یک مبحث کلی گروه‌بندی است که با استفاده از روش‌های محاسباتی یا الگوریتم‌های مختلف انجام می‌گیرد و به انواع مختلف تجزیه کلاستر طبقه‌بندی می‌گردد. تجزیه کلاستر اولین بار در مطالعات انسان‌شناسی توسط درایور و کروبر^۱ در سال ۱۹۳۲ ابداع شد. در سال‌های ۱۹۳۸ و ۱۹۳۹ به ترتیب توسط زوبین^۲ و روبرت ترایون^۳ توسعه پیدا کرد.

از تجزیه کلاستر می‌توان در مقایسه‌ها و طبقه‌بندی مکانی و زمانی جوامع موجودات و محیط‌های مختلف و همچنین در بررسی‌های زیست‌شناختی، بیوانفورماتیک و اکولوژیکی جانوران و گیاهان استفاده کرد. در سیستماتیک گیاهی نیز برای ایجاد طبقه‌بندی‌های مصنوعی مورد استفاده قرار می‌گیرد که موجودات را در گونه و یا جنس‌های متمایز خاص طبقه‌بندی می‌کند. در همین راستا زیست‌شناسان وقت خودشان را سالیان متمادی صرف ایجاد رده‌بندی جانداران^۴ یا طبقه‌بندی مراتبی آنها^۵ نمودند تا اینکه امروزه شناخت جانداران را در مراتب جاندار، دامنه، سلسله، شاخه (دسته)، رده، راسته، خانواده، جنس و گونه جست‌وجو کنیم. این نوع از طبقه‌بندی جانداران یک نوع طبقه‌بندی توصیفی خوشه‌ای است که امروزه توسط تجزیه کلاستر می‌تواند به شکل و زبان الگوریتم بیان شود.

همچنین از تجزیه کلاستر می‌توان در ساختن گروه‌هایی از ژن‌ها با الگوهای بیان خاص مانند گروه‌هایی از ژن‌های پروتئین‌های خاص از جمله آنزیم‌ها استفاده کرد؛ بدین ترتیب از تجزیه کلاستر برای گروه‌بندی توالی‌های همولوگ‌های فامیل‌های ژنی نیز استفاده می‌گردد. در پزشکی از تجزیه

^۱ - Driver and Kroeber

^۲ - Zubin

^۳ - Robert Tryon

^۴ - Taxonomy

^۵ - Hierarchical Classification



کلاستر برای تمایز بین انواع مختلف بافت‌ها و خون‌ها با استفاده از تصاویر سه بعدی آنها و همچنین طبقه‌بندی آنتی بیوتیک‌ها بر اساس فعالیت آنتی باکتریایی آنها استفاده می‌شود. تجزیه کلاستر در تحقیقات بازرگانی برای تقسیم‌بندی جمعیت مصرف‌کنندگان به طبقات مختلف بر اساس علایق و توانایی خرید آنها به کار برده می‌شود تا در برنامه‌ریزی‌های بعدی از قبیل جایابی فروش، موقعیت محل تولید، توسعه تولید جدید و غیره لحاظ گردد. تجزیه کلاستر در علوم دیگر از جمله علوم اجتماعی، آموزشی، شیمی، هواشناسی و علوم کامپیوتر نیز مورد کاربرد مطالعاتی و تحلیلی دارد.

تعریف

برخی از دانشمندان معتقدند به لحاظ وجود الگوریتم‌های مختلف برای تجزیه کلاستر و استفاده محققین از مدل‌های متفاوت که هر کدام دارای الگوریتم خاصی هستند، نمی‌توان مفهوم کلاستر را به‌طور دقیق تعریف کرد. می‌توان گفت تجزیه کلاستر یک ابزار اکتشافی است که موجبات طبقه‌بندی، شناسایی و تشخیص گروه‌های با مفهوم و ارائه مدل‌های آماری خاص را فراهم می‌کند که در نهایت منجر به کاهش داده‌ها یا آسان‌سازی ساختار آنها، دسته‌بندی اطلاعات و قرار دادن متغیرها با ضرایب همبستگی بالا در گروه‌های مشابه و متمایز می‌گردد.

تجزیه کلاستر روشی است که در آن با داشتن نمونه‌هایی از n فرد با p متغیر، می‌توان افراد مورد مطالعه را طوری گروه‌بندی کرد که افراد مشابه در کلاس‌های مفهوم‌دار و متمایز قرار گیرند. تجزیه کلاستر، داده‌ها را به گروه‌ها یا خوشه‌های متمایز مفهوم‌دار و هم چنین مفید تقسیم می‌کند که ما را در پیدا کردن گروه‌های واقعی یاری می‌کند. از آنجا که هدف از تجزیه کلاستر ایجاد خوشه‌های مفهوم‌دار و با معنی است، می‌بایست ساختار طبیعی داده‌ها حفظ شود تا مجموعه بزرگ داده‌ها به‌صورت دسته‌بندی و مختصر و مفید برای مطالعات بعدی مورد استفاده قرار گیرد.

بنابراین هدف از تجزیه کلاستر پیدا کردن دسته‌های واقعی افراد و کاهش تعداد اولیه داده‌ها است. به بیان دیگر هدف، شناسایی تعداد کمتری از گروه‌هاست به‌طوری‌که افراد دارای شباهت بیشتر با یکدیگر، در یک گروه قرار گیرند. با این تفسیر اگر n فرد داشته باشیم هدف از تجزیه کلاستر تشکیل g گروه بر اساس p صفت است که تعداد آنها کمتر از n باشد.



اگر در تجزیه کلاستر گروه‌های مفهوم‌دار ایجاد شود به این نوع از طبقه‌بندی که با در نظر گرفتن تمامی متغیرها انجام می‌گیرد و طبقات ایجاد شده را موصوف به مفهوم و موضوع خاصی می‌کند طبقه‌بندی افراد^۱ می‌گویند اما اگر هدف کاربر فقط کلاستر بندی داده‌ها باشد به این عمل الگوریتمی، گروه‌بندی^۲ می‌گویند.

www.ketab.ir

^۱ - Classification

^۲ - Clustering