

● سوال 1: تاریخچه داده کاوی

داده کاوی فرآیندی است که در آغاز دهه 90 مطرح شد و با نگرشی نو، به مسئله استخراج اطلاعات از پایگاه داده ها میپردازد. از سال 1995 داده کاوی به صورت جدی وارد مباحث آمار شد و در سال 1996 ، اولین شماره مجله کشف دانش و معرفت از پایگاه داده ها منتشر شد . محققانی نظیر براچمن و آناند (1996) کلیه مراحل واقع گرایانه و رو به جلو کشف دانش از پایگاه داده ها را تشخیص دادند. در حال حاضر، داده کاوی مهمترین فناوری جهت بهره برداری موثر از داده های حجیم است و اهمیت آن رو به فزونی است . به طوریکه تخمین زده شده است که مقدار داده ها در جهان هر 20 ماه به حدود دو برابر می رسد. در یک تحقیق که بر روی گروه های تجاری بسیار بزرگ در جمع آوری داده ها صورت گرفت مشخص گردید که 19 درصد از این گروه ها دارای پایگاه داده هایی با سطح بیشتر از 50 گیگا بایت میباشند و 59 درصد از آنها انتظار دارند که در آینده ای نزدیک در چنین سطحی قرار گیرند .

درصنایعی مانند کارت های اعتباری و ارتباطات و فرشگاه های زنجیره ای و خریدهای الکترونیکی و اسکنرهای بارکد خوان هر روزه داده های زیادی تولید و ذخیره می شوند . افزایش سرعت کامپیوترها باعث به وجود آمدن الگوریتم هایی شده است که قدرت تجزیه و تحلیل بسیار بالایی دارند بدون اینکه محدودیتی در زمینه ظرفیت و سرعت کامپیوترها داشته باشند. در سال 1989 و 1991 کارگاه های کشف دانش و معرفت از پایگاه داده ها توسط پیاتنسکی و همکارانش برگزار شد. در فواصل سالهای 1991 تا 1994 کارگاه های کشف دانش و معرفت از پایگاه داده ها توسط فییاد و پیاتنسکی و دیگران برگزار شد. به طور رسمی اصطلاح داده کاوی برای اولین بار توسط فییاد در اولین کنفرانس بین المللی ”کشف معرفت و داده کاوی” در سال 1995 مطرح شد. امروزه کنفرانسهای مختلفی در این زمینه در سراسر دنیا برگزار میشود. افزایش داده های بسیار باعث پیدایش فرصتهای تازه برای کار در علوم مهندسی و کسب و کار شده است. زمینه داده کاوی و کشف دانش از پایگاه داده ها به عنوان یک رشته علمی جدید در مهندسی و علوم کامپیوتر ظهور کرده است. مهندسی صنایع با حوزه های گوناگون و در بر داشتن فرصتهای بینظیر اکنون برای کاربرد داده کاوی و کشف دانش از پایگاه داده ها و برای توسعه مفاهیم و روشهای تازه در این زمینه آماده است. فرآیندهای صنعتی زیادی اکنون برای مطمئن شدن از کیفیت سفارشات محصول و کاهش هزینه های محصول به طور خودکار و کامپیوتری شده اند.

منابع:

<http://azsoft.ir/history-data-mining/>

<https://tanoco.ir/datamining/%D9%85%D8%B1%D9%88%D8%B1%DB%8C-%D8%A8%D8%B1->

[D8%AA%D8%A7%D8%B1%DB%8C%D8%AE%DA%86%D9%87-](https://tanoco.ir/datamining/%D9%85%D8%B1%D9%88%D8%B1%DB%8C-%D8%A8%D8%B1-%D8%AA%D8%A7%D8%B1%DB%8C%D8%AE%DA%86%D9%87-%D8%AF%D8%A7%D8%AF%D9%87-%DA%A9%D8%A7%D9%88%DB%8C/)

[D8%AF%D8%A7%D8%AF%D9%87-%DA%A9%D8%A7%D9%88%DB%8C/](https://tanoco.ir/datamining/%D9%85%D8%B1%D9%88%D8%B1%DB%8C-%D8%A8%D8%B1-%D8%AA%D8%A7%D8%B1%DB%8C%D8%AE%DA%86%D9%87-%D8%AF%D8%A7%D8%AF%D9%87-%DA%A9%D8%A7%D9%88%DB%8C/)

● سوال 2: تفاوت آمار و داده کاوی

علم داده کاوی با آنالیز های متداول آماری بسیار متفاوت است ، ما در این مقاله تفاوت داده کاوی با آمار را مفصل مورد بررسی قرار داده ایم . و در زیر نمونه ای از تفاوت ها را در انجام پروژه داده کاوی و آمار به تفصیل توضیح خواهیم داد.

در داده کاوی: ما به فرضیه احتیاجی نداریم علم داده کاوی الگوریتم هایی دارد که به طور اتوماتیک روابط را ایجاد می کند.

در داده کاوی: علم داده کاوی به داده های صحیح و درست نیازی ندارد.

در داده کاوی: نتایجی که از داده کاوی به دست می آید پیچیده است و برگردان آن به صورت مطلبی قابل فهم برای مدیران نیازمند متخصص است.

در آنالیز آماری: در تمامی موارد آمار شناسان با یک فرضیه شروع می کنند.

در آنالیز آماری: آنها برای آنالیز آماری از داده های عددی استفاده می کنند آنها رابطه های ایجاد می کنند که در نهایت به فرضیه مرتبط باشد.

در آنالیز آماری: آمار شناسان داده های نابجا و نادرست را در طول آنالیز مشخص کرده و در نهایت نتایج کار خود را تفسیر و برای مدیران بیان می کنند.

یک داده کاو پس از طی مراحلی که در داده کاوی وجود دارد به جمع آوری داده ها می پردازد و سپس داده ها را یکپارچه سازی می کند و در نهایت عملیات داده کاوی را انجام میدهد علم داده کاوی در میان داده ها تمامی الگوهای غیر عادی را که از حالت عادی و نرمال آن انحراف دارد و ممکن است منجر به کلاهبرداری شود را شناسایی کرده و و آن را به صورت نتیجه تحلیل ارائه میکند.

داده کاو نتایجی را که از داده کاوی به دست آورده را بررسی کرده و با توجه به آن مراحل بعدی که باید راجع به آنها تحقیق کند مشخص شده و در نهایت مدلهایی به دست می آید که می تواند امکان کلاهبرداری مشتریان را پیش بینی کند.

همانطور که گفتیم در آمار ما نخست به یک فرضیه نیاز داریم آمارشناس نخست متوجه الگوی رفتاری

می شود که سبب کلاهبرداری شده ، بر اساس این فرضیه یک سوال مطرح می کند تا این موضوع را مورد بررسی قرار دهد اگر نتیجه حاصله مناسب نبود آمارشناس فرضیه را اصلاح می کند یا یک فرضیه دیگری انتخاب می کند و همانطور که کاملاً مشخص است این روش نه تنها وقت گیر است بلکه بسیار به قدرت تجزیه و تحلیل یک آمارشناس بستگی دارد و بدتر از آن که الگوهایی که آمارشناس نتوانست پیش بینی کند و آنها را دریابد منجر به عدم کشف در الگوهای کلاهبرداری خواهد شد.

منابع:

<https://dadehkavy.com/difference-datamining-statistics/>

● سوال 3: تفاوت یادگیری ماشین و داده کاوی

عملکرد داده کاوی به شدت محدود به جمع آوری اطلاعات از منابع مختلف است. خود فناوری تصمیم نمی گیرد. بدون دخالت شما قادر به انجام هیچ عملی نیست. هدف آن یافتن راه های مفید برای استفاده از داده های یافت شده است. بین مجموعه هایی برای بازیابی داده ها و محاسبه داده ها از آنها، آنچه در بخش ها قرار می گیرد و آنچه که انجام نمی شود، تمایز قائل می شود. از این فناوری می توان برای تمایز داده ها و به دست آوردن مجموعه داده ها برای پارتیشن ها استفاده کرد. یادگیری ماشین با آرایه های داده تولید شده توسط فناوری داده کاوی کار می کند. فناوری هوش مصنوعی با استفاده از الگوریتم های از پیش مدل سازی شده برای اقدامات، از داده ها برای تصمیم گیری و پیگیری استفاده می کند. این فناوری بدون پشتیبان گیری دائمی از اطلاعات به روز وجود ندارد. در نهایت، ما اکوسیستم خود را از تصمیم گیری آگاهانه دریافت می کنیم. هر دو فناوری مکمل یکدیگر هستند. استفاده از آنها به تنهایی برای محدود کردن پتانسیل آنها است. به عبارتی دیگر یادگیری ماشین در واقع هدف ساخت ماشینی را دنبال می کند که توانایی یادگیری داشته باشد و بتواند خود را با اطلاعات جدید وفق دهد. اما داده کاوی مربوط به علمی برای استخراج دانش از مجموعه پایگاه داده هاست. داده کاوی بر روی مسائل صنعتی و راه حل های کاربردی تمرکز دارد که علاوه بر نگرانی در مورد اندازه داده ها، از نظر داده های بزرگ یا کوچک، بر روی سرعت پردازش داده ها نیز تمرکز می کند.

منابع:

<https://dataak.com/blog/what-is-the-difference-between-artificial-intelligence-machine-learning-data-mining/#:~:text=%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C%20%D9%85%D8%A7%D8%B4%DB%8C%D9%86%20%D8%AF%D8%B1%20%D9%88%D8%A7%D9%82%D8%B9%20%D9%87%D8%AF%D9%81,%D8%AF%D8%A7%D9%86%D8%B4%20%D8%A7%D8%B2%20%D9%85%D8%AC%D9%85%D9%88%D8%B9%D9%87%20%D9%BE%D8%A7%DB%8C%DA%AF%D8%A7%D9%87%20%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7%D8%B3%D8%AA.>
<https://lancersara.com/differences-between-data-mining-and-machine-learning/>

● سوال 4: تفاوت داده کاوی و علم داده

تفاوت علم داده و داده کاوی در ابعاد مختلفی از ماهیت گرفته تا اهداف قابل بررسی است. داده کاوی فرآیند استخراج اطلاعات مفید، الگوها و روندها مانند تجزیه و تحلیل نمونه‌ای از داده‌های خام موجود در دیتابیس‌های عظیم و ارائه اطلاعات مرتبط و قابل استفاده است که می‌تواند برای حل مشکلات تجاری مورد استفاده قرار بگیرد. در مقابل علم داده به عنوان فرآیند به دست آوردن بینش ارزشمند از داده‌های ساختاریافته و بدون ساختار با استفاده از ابزارها و تکنیک‌های مختلف تعریف می‌شود. این دو اصطلاح برای افرادی که درک صحیحی از آن ندارند، مشابه یکدیگر به نظر می‌رسد، اما حوزه‌های کاربرد این دو مفهوم با یکدیگر بسیار متفاوت است.

مبنای مقایسه	داده کاوی	علم داده
تعریف	فرآیند استخراج اطلاعات مفید، الگوها و روندهای پنهان از دیتابیس عظیم است.	به فرآیند به دست آوردن بینش ارزشمند از داده‌های ساختاریافته و بدون ساختار، جمع‌آوری دیتاها، تجزیه و تحلیل و ترسیم بینش از آن با استفاده از ابزارها و روش‌های مختلف اطلاق می‌شود.
گستره	عمدتاً برای اهداف تجاری استفاده می‌شود.	عمدتاً برای اهداف علمی استفاده می‌شود.
ارتباط	با فرآیندها درگیر است.	بر علم دیتاها تأکید می‌کند.
رویکرد	یک تکنیک است.	یک رشته، محدوده و حوزه است.
تمرکز	بر روی فرآیند کاری متمرکز است.	بر روی مطالعه علمی متمرکز است.
هدف	از داده‌ها بهتر و راحت‌تر استفاده شود و اطلاعات حیاتی و ارزشمند از آن‌ها حاصل شود.	به جمع‌آوری، پردازش، پیش‌بینی‌های دقیق، تجزیه و تحلیل، تصمیم‌گیری آگاهانه و استفاده از داده‌ها در عملیات‌های مختلف می‌پردازد و مفهومی است که برای ساخت محصولات Data محور سازمان‌ها استفاده می‌شود.
خروجی	الگوها به عنوان خروجی در نظر گرفته می‌شوند.	خروجی‌ها انواع متنوعی دارند.
مقصود	به یافتن روندهایی که پیش از این ناشناخته بودند، می‌پردازد.	به تجزیه و تحلیل اجتماعی، ساخت مدل‌های پیش‌بینی، کشف حقایق ناشناخته و سایر موارد می‌پردازد.
دیدگاه حرفه‌ای	فردی با دانش ناوبری در میان داده‌ها و درک آماری، قادر به انجام این تکنیک است.	یک فرد برای تبدیل شدن به یک دانشمند دیتا باید یادگیری ماشین، برنامه‌نویسی و تکنیک‌های اطلاعات گرافیکی را درک کند و دانش لازم برای این حوزه را کسب کرده باشد.
وسعت	این تکنیک زیرمجموعه‌ای از علم دیتا ساینس و بخشی از راستای آن محسوب می‌شود.	چند رشته‌ای بوده و شامل تجسم داده‌ها، علوم اجتماعی محاسباتی، آمار، دیتا ماینینگ، پردازش زبان طبیعی و غیره است.
نوع داده	عمدتاً ساختاریافته است.	کلیه اشکال دیتا از جمله ساختاریافته، نیمه ساختاریافته و بدون ساختار را دربر می‌گیرد.
سایر عناوین فرعی	۱ باستان‌شناسی داده‌ها ۲ برداشت اطلاعات ۳ کشف اطلاعات ۴ استخراج دانش	علم مبتنی بر داده
پیش‌زمینه	تکنیکی است که شامل بخشی از کشف دانش در فرآیندهای پایگاه داده (KDD) تلقی می‌شود.	یک رشته تحصیلی همانند علوم کامپیوتر، آمار کاربردی یا ریاضیات کاربردی است.

منابع:

<https://nokarto.com/%D8%AA%D9%81%D8%A7%D9%88%D8%AA-%D8%B9%D9%84%D9%85-%D8%AF%D8%A7%D8%AF%D9%87-%D9%88-%D8%AF%D8%A7%D8%AF%D9%87-%DA%A9%D8%A7%D9%88%DB%8C/>