متخصص داده کاوی

چکیده:

داده‌کاوی[1] ، دانشی میان‌رشته‌ای پیرامون استخراج دانش و آگاهی از مجموعه‌ای داده و اطلاعات است که از روش‌ها، فرآیندها، الگوریتم‌ها و سیستم‌های علمی برای استخراج دانش و بینش از داده‌ها در اشکال گوناگون (ساختاریافته و ساختار نیافته) استفاده می‌کند. داده‌کاوی مفهومی برای یکپارچه‌سازی آمار، تحلیل داده، یادگیری ماشین و دیگر مفاهیم مرتبط تحت یک عنوان واحد است. این کار به منظور درک و تحلیل پدیده‌ها با استفاده از داده‌ها انجام می‌شود. داده‌کاوی از ترکیب مباحث مختلفی به وجود آمده و بر مبانی و روش‌های موجود در حوزه‌های مختلف علمی بنا شده است. تعدادی از این حوزه‌ها عبارت‌اند از: ریاضیات، آمار، مهندسی داده، بازشناخت الگو و... . هدف این علم، استخراج مفهوم از داده و تولید محصولات داده محور است . به شاغلین در حوزه ی علم داده، متخصص علم داده (scientist data) یا دانشمند داده می‌گویند. این اصطلاح توسط دی جی پاتیل و جف همربارکر ابداع شده است در صورتی که سال‌ها قبل از آن که آن‌ها استفاده از اصطلاح فوق را به‌طور عمومی مطرح کنند، از آن استفاده‌شده است.

مقدمه:

یکی از حرفه‌هایی که در این مقاله قصد معرفی آن را داریم، متخصص داده‌کاوی است که پیشتر هم راجع به آن صحبت شد و در این مقاله قصد داریم تا حدودی به ماهیت این حرفه، ویژگی‌های متخصص داده‌کاوی، مهارت‌های مورد نیاز متخصص داده‌کاوی، شاخه‌های تخصص داده‌کاوی، و داده‌کاوی در علم اطلاعات و دانش شناسی بپردازیم.

داده‌کاوی:

به مجموعه‌ای از روش‌های قابل اعمال بر پایگاه داده‌های بزرگ و پیچیده به منظور کشف الگوهای پنهان و جالب توجه نهفته در میان داده‌ها، داده‌کاوی گفته می‌شود.

کاربردهای داده‌کاوی:

داده‌کاوی در خیلی از حوزه‌ها ورود پیدا کرده است، مانند:

آموزش

مدیریت ارتباط با مشتریان

تحلیل رفتار مشتری

بازاریابی

پزشکی

بانکداری

بورس

و.....

به نظر می‌رسد در داده‌کاوی فرصت‌های شغلی زیادی وجود دارد که کمتر به آن‌ها توجه شده است.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

۱+۱۰ شغل که به متخصصان علوم داده‌کاوی احتیاج دارند

متخصص متن‌کاوی

متخصص تصویر کاوی (پردازش تصویر)

متخصص ویدیو کاوی (پردازش ویدیو)

متخصص پردازش صوت

متخصص تحلیل داده‌های کسب‌وکار

متخصص مسیریابی هوشمند و تشخیص مسیر

متخصص پردازش داده‌های حجیم و جریان داده‌ای

متخصص تحلیل داده‌های بازی‌ها

داده‌کاوی در علوم اجتماعی و آمارهای دولتی

متخصص داده‌کاوی در بورس و داده‌های اقتصادی

داده‌کاوی در علوم تخصصی و پایه

متخصص علم داده:

به شاغلین در حوزه‌ی علم داده،متخصص علم داده (data scientist) (یادانشمند داده‌می‌گویند. متخصصین علم داده با عمیق شدن در چندین رشته‌ی علمی، مسائل پیچیده‌ی مطرح شده در حوزه‌ی داده را حل می‌کنند. نتایج نظرسنجی‌ها حاکی از این موضوع است که برای متخصص علم داده شدن ۵ تا ۸ سال زمان لازم است. به‌طورکلی انتظار می‌رود که متخصصین علم داده قادر باشند:

در بخش‌هایی از علوم ریاضیات و آمار و علوم کامپیوتر کار کنند.

یک متخصص علم داده می‌بایست در یک یا دو رشته تخصص داشته باشد و در دیگر حوزه‌ها دارای مهارت کافی باشد.

ویژگی‌های متخصص داده‌کاوی:

روحیه‌ی قبول ابهام

ذهن پرسش‌گر

مقداری سلیقه

قدرت انجام تحلیل‌های آماری

داشتن تفکر نقادانه

و کمی هم دانستن تجارت

lib2mag.ir

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

مهارت‌های موردنیاز متخصص داده‌کاوی:

هر کسی که به ساخت موقعیت شغلی قدرتمندتر در این دامنه علاقه‌مند است، باید مهارت‌های کلیدی در سه حوزه تحلیل، برنامه‌نویسی و دانش دامنه را کسب کند.

مهارت برنامه‌نویسی (پایتون[۲]، R[۳]، اسکالا و SAS)

مهارت داشتن در نوشتن کدهای پایگاه داده SQL

توانایی کار با داده‌های ساختار نیافته از منابع گوناگون مانند ویدئو و شبکه‌های اجتماعی

درک توابع تحلیل چندگانه

دانش یادگیری ماشین

توانایی استخراج و تفسیر منابع داده

مدیریت حجم زیاد اطلاعات با سخت‌افزار

محدودیت‌های نرم‌افزاری و پهنای باند

ادغام منابع داده با یک دیگر

تضمین پایداری مجموعه‌های داده

مصورسازی داده برای فهم آن

ساخت مدل‌های ریاضی با استفاده از داده، مانند مدل‌های ریگرسیون و طبقه‌بندی

مقایسه آماری مدل‌های ریاضی گوناگون و انتخاب مدل برتر، فی المثل توسط آزمون B/A

به اشتراک‌گذاری یافته‌ها و دیدگاه‌ها در حوزه داده با متخصصان دیگر یا مخاطب عام

شاخه‌های تخصص داده‌کاوی:

مهندسی داده:

در حالت اول که بخش مهندسی داده است، معمولاً باید بچه‌های نرم‌افزاری و سخت‌افزاری باشند. به این دلیل که بتوانند با سیستم‌های کامپیوتری کارهای پیشرفته‌ای انجام بدهند. این گروه معمولاً لازم دارند که با نرم‌افزارهایی مثل hadoop[۴] Spark[۵] کار کنند و به دیتابیس‌ها و برنامه‌نویسی مسلط باشند. این بخش مختص مهندس داده است که وظیفه آن ذخیره‌سازی حجم زیاد دیتا است. در مهندسی داده وجود افراد فریلنسر[۶] بسیار زیاد است به دلیل اینکه خیلی از شرکت‌ها می‌خواهند یک زیرساخت بسازند که بتوانند این حجم از پردازش را مدیریت بکنند. مهندس داده یک‌بار این را می‌سازد که می‌تواند به صورت فریلنسر یا شرکت باشد. فرد می‌تواند شرکتی تاسیس کند که کار آن تأمین این گونه زیرساخت‌ها باشد و یا به صورت فریلنسر کار کند. در همین حالت هم شرکت‌ها برای تأمین زیرساخت و پشتیبانی، فردی را استخدام می‌کنند.

آنالیز داده‌ها:

در قسمت دوم که کارش آنالیز داده‌ها است،دانستن آمار و احتمالات و کمی برنامه‌نویسی لازم است. در این قسمت نرم‌افزارهای زیادی استفاده می‌شوند ولی به طور معمول زبان‌های اصلی در حوزه‌ی آنالیز (پایتون، R) هستند. R برای

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

تحلیل‌های آماری استفاده می‌شود و پایتون، هم برای برنامه‌نویسی و هم برای تحلیل‌های آماری کاربرد دارد. اطلاعات کلی از دیتابیس در این قسمت هم خیلی می‌تواند کمک کند. یک مهارت که در این مرحله نیاز است تفکر نقّادانه یا همانthinking Critical هست. بعضی وقت‌ها لازم است که سخت گیرانه به دیتاهایی که موجود است، نگاه کرد تا گمراه نشده و مدیران شرکت نیز گمراه نشوند.

خوب است که دید کلی در تجارت و کارهای مدیریتی داشته باشیم. البته یک مسئله‌ی مهم هم این است که توانایی ارتباطی خوبی داشته باشیم چون باید یک نفر را نسبت به موضوعی قانع کنیم که با توجه به داده‌ها این کار خوب است یا این کار بد است. برخلاف تصور عامه که تصور می‌کنند متخصص داده‌کاوی فردی است که تنها با کامپیوتر سروکار دارد، این افراد در این حوزه احتیاج به مهارت ارتباطی بالایی دارند. در بخش آنالیز هم خیلی از شرکت‌ها هستند که نرم‌افزار تحلیل داده می‌سازند. در این بخش کار فریلنسری خیلی کم ولی در عوض آن استخدام بسیار زیاد است.

Deep learning و Machine learning:

این قسمت، نیاز به اطلاعات ریاضی بالا و کاملی دارد (معادلات جبری و آماری سنگینی در این قسمت استفاده می‌شوند). کسی که در این حوزه می‌خواهد کار کند باید فهم عمیقی از این موضوعات داشته باشد. خیلی از افرادی که در این حوزه مشغول به فعالیت هستند به دلیل اینکه نیاز به محاسبات پیچیده ریاضی دارند معمولاً از فارغ‌التحصیلان رشته‌های ریاضی یا فیزیک بوده‌اند. این افراد شاید ناچار باشند یک معادلاتی را RUN کنید و یک هفته بعد بیایند و نتیجه محاسبات را ببینند.

در این مقام هم لازم است که برنامه‌نویسی بلد باشید. البته نه در حد زیاد فقط در حدی که ابزارهای Machine learning دنیا مثل Tensorflow[۷] keras[۸] و Open CV[۹] را بشناسید و بتوانید با این‌ها چیزهایی که می‌خواهید را پیاده کنید. یک فهم کلی از دیتابیس در این قسمت هم لازم است. در قسمتDeep learning فریلنسر کم است چون کاری که قرار است صورت بگیرد یک کار طولانی است. این کار لزوماً جواب نهایی ندارد که شرکتی کار را به فریلنسر بدهد و رهایش کند. در این قسمت، هم بخش استخدامی و هم شرکت زدن مرسوم است.

داده‌کاوی در علم اطلاعات و دانش شناسی:

داده‌کاوی در بخش‌های مختلفی از حوزه علم اطلاعات و دانش‌شناسی از جمله:

جست‌وجوی اطلاعات

فراهم‌آوری

مدیریت و خدمات کتابخانه

جمع‌آوری اطلاعات مراجعان به‌کتابخانه

حفظ وفاداری مراجعان

اخراج مراجعه‌کننده

تحلیل منابع موجود در کتابخانه (موجودی کتابخانه)

بانک اطلاعاتی

گردآوری منابع

استفاده از مجموعه و

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

امانت بین‌کتابخانه‌ای کاربرد دارد.

کاربردهای داده‌کاوی در حوزه علم اطلاعات و دانش‌شناسی بیشتر مختص به بخش‌های امانت، مرجع، و مجموعه‌سازی است. اما می‌تواند در سایر بخش‌ها نیز از داده‌کاوی برای تحلیل داده‌های موجود در راستای دستیابی هر چه بهتر به اهداف رشته و کتابخانه‌ها استفاده کرد.

نتیجه‌گیری:

هنگامی که از متخصص علم داده صحبت می‌کنیم، منظور ما تبدیل شدن به شخصی است که بتواند از میان انبوهی از داده‌ها، الگوها و به دنبال آن، ارزش‌های نهفته در بینَ آن‌ها را کشف کند. کشف این الگوها است که می‌تواند ارزش‌افزوده‌ی یک کسب‌وکار را بالا و بالاتر ببرد. اما برای تبدیل شدن از یک فرد غیرمتخصص به یک متخصص در حوزه‌ی داده کاوی مراحلی لازم است. طبیعی است که مراحلی که در این مقاله نوشته می‌شوند تنها راهکار موجود نیستند.

✓ این متن با نظارت دکتر الهه ابراهیمی مدرس رشته علم اطلاعات و دانش شناسی دانشگاه اصفهان گردآوری شده است.

lib2mag.ir

پانویس ها

[۱] Data mining

[۲] یک ، Python زبان برنامه‌نویسی شیءگرا، تفسیری، سطح بالا ، وهمه منظوره است که فلسفه اصلی طراحی آن «خوانایی بالای کد» است.

[۳] یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای محاسبات آماری و علم داده‌ها است.

[۴] یک چارچوب نرم افزاری منبع باز که پردازش توزیع شده داده های بزرگ را بر روی خوشه هایی از سرورها ممکن می سازد.

[۵] یک چارچوب رایانش توزیع شده متن باز است.

[۶] Freelancer

[۷] یک کتابخانه نرم افزاری متن باز برای یادگیری ماشین در انواع مختلف و وظایف مفهومی و زبان است.

[۸] یک کتابخانه متن باز

[۹] Open computer vision library

منابع:

https://fa.wikipedia.org

https://www.researchgate.net

https://www.sciencedirect.com

تقوی فرد، محمد تقی؛ منصوری، طاها؛ ناصرزاده، سید محمد رضا؛ فراست، علیرضا (۱۳۸۶)، داده کاوی و کاربرد آن در

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

تصمیم گیری ها، دوره ۲۰، شماره ۷۹ (زمستان)، ص ۳ – ۱۴.

رحمانی، مهدی؛ حاجی زین العابدینی، محسن (۱۳۹۴)، کاربردهای داده کاوی در علم اطلاعات و دانش شناسی، شماره ۳ (پاییز)، ص ۲۳ – ۳۲.

میرمحمد صادقی، علیرضا؛ مغان، مهدی؛ علیخانی، سیاوش (۱۳۹۴)، بررسی جایگاه داده کاوی در فرآیند مدیریت دانش و ارائه مدل مفهومی جهت استخراج دانش؛ نشریه مدیریت فردا، ص۶۹.

مشخصات استناددهی به این مقاله

نویسنده(ها):	زهرا میرزایی
عنوان مقاله:	متخصص داده کاوی
عنوان مجله:	کتابدار ۲.۰ –(عنوان لاتین: 2.0 i-Kitbdr)
دوره مجله(Vol):	۶
شماره مجله(Issue):	۶
سال(Year):	۱۳۹۹
شناسه دیجیتال(DOI):	
لینک کوتاه:	http://lib2mag.ir/12211

lib2mag.ir

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.