



توانمندیهای نرم افزار های متن باز در داده کاوی

محمد باقری، مهیار دریانورد چونچانی، علی ایوبی

باشگاه پژوهشگران جوان و نخبگان، دانشگاه آزاد اسلامی واحد لاهیجان

مسول مکاتبات: محمد باقری

چکیده

حجم زیاد داده هایی که در سازمان به صورت پیوسته در حال افزایش است و هیچ استفاده از این دانش پنهان درون داده ها نمی شود. سازمان ها تصمیم گیری های خود را بدون استفاده از این دانش انجام می دهند و تصمیمات با خطاهایی همراه می شود و کیفیت تصمیم گیری سازمان را بسیار کاهش می دهد. با گسترش سیستم های پایگاهی و حجم بالای داده های ذخیره شده در این سیستم ها، نیاز به ابزاری است تا بتوان داده های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرارداد و این داده ها را به نحو احسن به دانش ارزشمند تبدیل کرد. امروزه سازمانها به اهمیت ارزش نهفته درون اطلاعات پی برده اند که می تواند کیفیت تصمیم گیری درون سازمان را به طور چشمگیری بهبود دهد، ولی سازمانها در دریایی از داده گیر افتاده اند و محتاج دانشند. شناخت اندکی از این قابلیت ها میتواند بسیار سود آور باشد و یا به میزان زیادی از ضررها جلوگیری نماید. در این مقاله چند مورد از بهترین نرم افزار متن باز برای داده کاوی مورد استفاده قرار می گیرد معرفی شده است. هر یک از این نرم افزارها دارای نقاط قوت و ضعف هستند و برای زمینه خاصی مناسب است. بعضی برای داده کاوی ها بسیار پیچیده مناسب هستند و بعضی دیگر برای مقالات دانشجویی با شناخت نیازهای خود و نگاه کردن به توانمندی ها و ضعف نرم افزار می توان بهترین نرم افزار را برای کار به دست آورد.

کلمات کلیدی: داده کاوی، متن باز، نرم افزاری، بهره وری.

مقدمه

داده کاوی عبارت است از شناخت روش هایی برای استفاده از قابلیت های نهفته در کسب و کار، یکی از مؤثرترین راههای موفقیت در بازار رقابتی است ولی پیدا کردن این قابلیت های نهفته بسیار سخت است و داده کاوی این کار را انجام می دهد.

مدیران بسیاری از شرکت ها و ادارات مانند بیمارستان ها، بانکها و مؤسسات مالی، شرکت های بیمه و ... با صرف هزینه های بالا اقدام به ایجاد پایگاههای داده و یا سیستم های مدیریتی همچون مدیریت منابع سازمانی و مدیریت ارتباط با مشتری نموده اند و ولی از گنجینه های پنهان (داده های) خود استفاده نمی کنند. این حجم از داده ها بدون تبدیل شدن به اطلاعات ارزشمند و دانش، در کسب و کار ارزش زیادی نخواهند داشت و هیچ کمکی به غیر از انبار داده نخواهد داشت و درانبوهی از داده ها گیر افتاده اند و محتاج دانش اند. داده کاوی را می توان حاصل سیر تکاملی طبیعی تکنولوژی اطلاعات دانست، که این سیر تکاملی ناشی از یک سیر تکاملی در صنعت پایگاه داده است، نظیر عملیات: جمع آوری داده ها و ایجاد پایگاه داده، مدیریت داده و تحلیل و فهم داده ها. متأسفانه بسیاری از مدیران نمیدانند که این قابلیت های بالقوه وجود دارند و شناخت اندکی از این قابلیت ها میتواند بسیار سود آور باشد و یا به میزان زیادی از ضررها جلوگیری نماید.

دکتر شهرابی در بیان دانش داده کاوی اظهار داشت: (دانش داده کاوی سازمانها را قادر می سازد تا از سرمایه داده هایشان بهره برداری کنند و این ابزار برای پشتیبانی فرایند تصمیم گیری استفاده می شود در واقع داده کاوی با پردازش جامع داده و انجام فرایند تصمیم سازی از طریق استخراج دانش با ارزش از داده، تصمیم گیری را برای مدیران سازمان تسهیل می کند [1].

تعریف داده کاوی

با اهمیت ترین قسمت این موضوع استفاده از تعریف آن است تا درک دستی و یکسانی از این علم داشته باشیم تا بتوانیم درست در این بستر حرکت کنیم. در داده کاوی معمولاً به کشف الگوهای مفید از میان داده ها اشاره می شود. منظور از الگوی مفید: مدلی در داده ها است که ارتباط میان یک زیر مجموعه از داده ها را توصیف می کند، و معتبر، ساده، قابل فهم و جدید است. تعاریف گوناگونی برای داده کاوی ارائه شده اند. برخی از این تعاریف عبارتند از: داده کاوی: عبارت است از فرایند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم گیری در فعالیتهای تجاری مهم.

اصطلاح داده کاوی: به فرایند نیم خودکار تجزیه و تحلیل پایگاه داده های بزرگ به منظور یافتن الگوهای مفید اطلاق می شود

داده ها

ها ی بزرگ.



داده کاوی: یعنی تجزیه و تحلیل مجموعه داده‌های قابل مشاهده برای یافتن روابط مطمئن بین داده‌ها. [4]

عوامل پیدایش داده کاوی

اصلی‌ترین دلیلی که باعث شد داده کاوی کانون توجهات در صنعت اطلاعات قرار بگیرد، در دسترس بودن حجم وسیعی از داده‌ها و نیاز شدید به اینکه از این داده‌ها اطلاعات و دانش سودمند استخراج کنیم. اطلاعات و دانش بدست آمده در کاربردهای وسیعی از مدیریت کسب و کار و کنترل تولید و تحلیل بازار تا طراحی مهندسی و تحقیقات علمی مورد استفاده قرار می‌گیرد. [2]

ابزارهای داده کاوی، داده‌ها را آنالیز می‌کنند و الگوهای داده‌هایی را کشف می‌کنند که می‌توان از آن در کاربردهایی نظیر: تعیین استراتژی برای کسب و کار، پایگاه دانش و تحقیقات علمی و پزشکی، تسخیر نفوذ و ... استفاده کرد. شکاف موجود بین داده‌ها و تحلیلگران داده سبب ایجاد نیاز شدیدی برای ابزارهای داده کاوی شده است تا داده‌های دارای ارزش کم را به دانشی کاربردی ارزشمند تبدیل کنیم. البته اسامی دیگری نیز برای این فرایند پیشنهاد شده که بعضاً بسیار متفاوت با واژه داده کاوی است، نظیر: استخراج دانش از پایگاه داده، استخراج دانش، آنالیز داده/الگو، باستان شناسی داده، و لایروبی داده‌ها. [3]

تاریخچه داده کاوی

برای اولین بار مفهوم داده کاوی در کارگاه در زمینه کشف دانش در پایگاه داده‌ها توسط شیپر مطرح گردید. به دنبال آن در سالهای 1991 تا 1994، کارگاههای کشف دانش در پایگاه داده‌ها مفاهیم جدیدی را در این شاخه از علم ارائه کردند. داده کاوی در ابتدا از حوزه تجارت برخاست اما کاربردهای آن در سایر حوزه‌هایی که به گردآوری حجم وسیعی از داده‌های می‌پردازند که دستخوش تغییرات پویا نیز می‌گردند؛ مفید شناخته شد. بخشهایی مثل بانکداری، تجارت الکترونیک، تجارت سهام، بیمارستان و هتل از این نمونه‌اند. داده کاوی یکی از پیشرفتهای اخیر در راستای فن آوریهای مدیریت داده‌هاست. مجموعه‌ای از فنون است که به شخص امکان میدهد تا وراى داده پردازى معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی و یا پنهان است کمک می‌کند. این تکنولوژی امروزه دارای کاربرد بسیار وسیعی در حوزه‌های مختلف است به گونه‌ای که امروزه حد و مرزی برای کاربرد این دانش در نظر نگرفته و زمینه‌های کاری این دانش را از ذرات کف اقیانوسها تا اعماق فضا می‌دانند. نکته‌ای که باید در این قسمت بیان شود این است که داده کاوی به صورت یک محصول قابل خریداری نمی‌باشد، بلکه یک رشته عمل و فرآیندی است که بایستی به صورت یک پروژه پیاده‌سازی شود [5]

کارکردهای داده کاوی

توصیف و کمک به پیش بینی دو کارکرد اصلی داده کاوی هستند. تحلیل داده مربوط به مشخصه‌های انتخابی متغیرها؛ از گذاشته و حال، و درک الگو مثالی از تحلیل توصیفی است. برآورد ارزش آینده یک متغیر و طرح ریزی کردن روند مثالی از توانایی پیشگوییانه داده کاوی است. توصیفی: فرایند جستجو در یک بانک داده برای یافتن الگوهای پنهان، بدون داشتن یک فرضیه از پیش تعیین شده درباره اینکه این الگو ممکن است چه باشد. مانند تحلیلهایی که برحسب کالاهای خریداری شده صورت می‌گیرد، اینگونه تحلیل‌های سبکی نشانگر مواردیست که مشتری تمایل به خرید آنها دارند. این اطلاعات می‌تواند به بهبود موجودی، استراتژی طراحی، آرایش فروشگاه و تبلیغات منجر گردد. [6]

مدل پیش بینی: فرایندی که الگوهای کشف شده از بانک داده را می‌گیرد و آنها را برای پیش بینی آینده به کار می‌برد. مانند پیش بینی فروش در خرده فروشی، الگوهای کشف شده برای فروش به آنها کمک می‌کند تا تصمیماتی را در رابطه با موجودی اتخاذ کنند. [7]

ارزیابی مدل ها

ارزیابی روش‌های کلاس‌بندی با معیارهای زیر انجام می‌پذیرد:

میزان دقت (Accuracy) در پیشگویی، از جمله معیارهای ارزیابی روش‌های مذکور در کلاس‌بندی است که میزان قابلیت و توانایی یک مدل را در پیشگویی صحیح برچسب یک کلاس، مشخص می‌کند.

سرعت (Speed) و توسعه پذیری از نظر زمانیکه برای ایجاد یک مدل و زمان استفاده از آن مدل لازم است، از معیارهای دیگر ارزیابی روش در کلاس‌بندی



قوی بودن (Robustness) معیار مهمی است که میزان توانایی یک مدل را در برخورد با نویز و مقادیر حذف شده تعیین می‌کند. توسعه‌پذیری (Scalability) معیار دیگری است که از نقطه نظر میزان کارایی در بانک‌های اطلاعاتی بزرگ و نه داده‌های مقیم در حافظه مورد بررسی قرار می‌گیرد.

قابل تفسیر بودن (Interpreted) یعنی میزان و سطح درک ایجاد شده توسط مدل از دیگر مواردی است که می‌بایست در بررسی روش‌های کلاس‌بندی در نظر گرفت. [8]

دلایل استفاده از متن‌باز

به دلیل حجم بزرگ داده‌ها نیاز به سیستم‌هایی است که بتواند این حجم از داده‌ها را پردازش کند و استفاده از متن‌باز یک مزیت بسیار بزرگی برای ایجاد کلاستر کردن و استفاده از پردازش موازی است. دلیل دیگر داده‌ها که در دنیای متن‌باز انواع روش‌های مختلف برای کار است که باید با بررسی بهترین روشها را انتخاب کرد یا بعضی بهینه سازی ها را در روش‌های موجود به وجود آورد تا کارایی بهترین شود و از چهارچوب ساختاری کنار رفت و فراتر از آن فکر کرد که این قسمت هم توسط روش‌های متن‌باز مورد پشتیبانی می‌شود. پروژه‌های داده‌کاوی از 6 مرحله تشکیل شده‌اند که در مرحله آخر بحث توسعه است که باید به یک زبان برنامه‌نویسی در درون برنامه‌ها توسعه داد که توسط یک زبان برنامه‌نویسی پشتیبانی باید شود.

نرم‌افزارها

هدف ما در این مقاله مقایسه نیست بلکه بیان کردن توانمندی‌های هر نرم‌افزار برای کار داده‌کاوی است تا بنا به نیاز خود این یکی از نرم‌افزارهای داده‌کاوی را انتخاب کنند.

مقایسه 4 نرم‌افزار داده‌کاوی که جزو بهترین‌ها این زمینه هستند.

R
Rapid Miner
WEKA

دلیل انتخاب این 3 نرم‌افزار به شرح زیر است:

در آماري که از داده کاوان جهان انجام شده است این نرم‌افزارها از رتبه بالاتری برخوردار بودند.

زبان آماری و داده‌کاوی R

کوچک اما توانا، این توصیف موجزترین تعبیری باشد که می‌توان راجع به نرم‌افزار R بیان کرد. این نرم‌افزار محیط بسیار مناسبی برای محاسبات آماری و ترسیم نمودارها است. این نرم‌افزار در سال‌های اخیر در دنیا شهرت بسزایی یافته و نظر کاربران زیادی را به خود جلب کرده است. ویژگی که باعث شهرتش شده است

رایگان است و حمایت افراد دانشگاهی را داراست

روی اکثر سکوها قابل اجراست

بیش از 5 هزار بسته نسبی دارد که زمینه‌های مختلف آماری و داده‌کاوی را در بر می‌گیرد.

نرم‌افزار R یک زبان برنامه‌نویسی ریاضی شی‌گرای است که بسیار شبیه S-plus است و برای محاسبات آماری طراحی شده است. این پروژه از سال 1995 در گروه آمار دانشگاه Auckland شروع شد.

ویژگی‌های این زبان می‌تواند به موارد زیر اشاره کرد.

رایگان بوده و تحت پلتفرم‌ها برای کلاست رینگ می‌توان استفاده کرد.

قابلیت‌های گرافیکی فراوانی دارد.

یادگیری آسان و توابع پیش‌ساخته آماری فراوانی است

توابع را می‌توان به آسانی به کار برد

توانایی پردازش داده‌های حجیم را دارد



- راهنمای بسیار عالی برای کار دارد.
- با بیش از 5 هزار بسته بسیاری از نیاز برنامه‌نویسان را حل کرده است
- R مجموعه کامل از امکانات نرم‌افزاری برای کار کردن با داده‌ها و محاسبه و رسم نمودار است. از جمله امکانات این مجموعه می‌توان به موارد زیر اشاره کرد:
 - زبان برنامه نویسی ساده و پیشرفته شامل عبارت های شرطی، حلقه و توابع بازگشتی و ...
 - مجموعه قوی از عملگرهای محاسباتی آرایه‌ها و ماتریس‌ها
 - کتابخانه انجام عملیات داده‌کاوی و یادگیری ماشین
 - امکانات گرافیکی برای تجزیه و تحلیل داده‌ها
 - و
- این زبان در پیش‌پردازش داده‌ها و مدل‌سازی آن هیچ محدودیتی ندارد و می‌توان از پردازش موازی برای محاسبات استفاده کرد.[12]

رپدماینر (Rapid Miner)

- این نرم‌افزار یکی از بهترین نرم‌افزارهای متن‌باز داده‌کاوی است. دلیل استفاده از آن هم می‌توان رابط گرافیکی بسیار کاربرپسند این نرم‌افزار نام برد که توانسته است کاربران زیادی را جذب می‌کند. پشتیبانی عالی شرکت RAPID-I از آن بسیار قوی است و آقای پروفیسور نخعیزاده هم در متخصص در زمینه داده‌کاوی کار می‌کند این نرم‌افزار را برای کارهای داده‌کاوی پیشنهاد می‌کنند. استفاده از زبان گرافیکی کار با این نرم‌افزار را، بدون از دست دادن قدرت این زبان، زیاد کرده است.[9]
- از ویژگی‌های این نرم‌افزار می‌توان به موارد زیر اشاره کرد.
 - ارائه گزارش و رونوشت از مراحل اجرای الگوریتم
 - ظاهری مناسب برای کار
 - امکان ورود انواع فایلها برای داده‌کاوی arff, excel and ...
 - امکان تصحیح و خطایابی بسیار دقیق
 - مستندات بسیار زیاد آن
 - قابلیت لوله‌گذاری
 - الگوریتم‌های بسیار زیاد برای کار داده‌کاوی
 - امکان افزودن الگوریتم‌های داده‌کاوی weka به این نرم‌افزار
 - امکان اتصال به نرم‌افزار R برای انجام تحلیل‌ها
 - کیفیت عالی الگوریتم‌ها برای کار
 - امکان استفاده در پلتفرم‌های مختلف
 - افزونه‌های دیگر برای کارهای مثل متن‌کاوی و سری‌های زمانی
 - امکانات گرافیکی بسیار عالی برای کار با این نرم‌افزار
 - امکان اتصال به پایگاه داده‌های مختلف
- امکان برنامه‌نویسی در آن (به دلیل این که 3 ساختار اصلی برنامه‌نویسی ساخت‌یافته در آن پیاده‌سازی شده است)[10]

واسط کاربر رپدماینر

- این نرم‌افزار دارای واسط کاربری بسیار قوی است که از جمله مهم‌ترین نقاط قوت این نرم‌افزار است. ساختار بسیار عالی این نرم‌افزار برای اتصال به انواع پایگاه داده‌ها و پیش‌پردازش قوی و جامع کاربری فراوان این نرم‌افزار می‌توان به جزو ویژگی‌های عالی این نرم‌افزار به شمار می‌رود. ولی در قسمت که بسیار کیفیت مدل‌سازی را کاهش داده است. ولی توسط بسته نرم‌افزار که به آن قسمت استفاده کرد و بسیار قوی شناخته می‌شود.



الگوریتم‌های زیادی را پشتیبانی می‌کنند.

درخت تصمیم

شبکه‌ی عصبی

مدل‌های لجستیک خطی

پوست

روش‌های بیز

Svm

ورودی از انواع داده‌ها

خروجی از انواع داده‌ها

و روش‌های دیگر

واسط کاربری بسیار ساده این نرم‌افزار کار با آن را بسیار ساده کرده است به طوری که در عرض چند ساعت می‌توان کار با آن را یاد گرفت. استفاده از مدل‌های موجود در نرم‌افزار وکا باعث بالا بردن کیفیت مدل‌ها می‌شود.

وکا

WEKA (سرنام Waikato Environment for Knowledge Analysis) نرم‌افزاری آزاد و متن‌باز است که توسط دانشگاه وایکاتو در نیوزیلند طراحی شده است.

میزکار وکا، مجموعه‌ای از الگوریتم‌های روز یادگیری ماشینی و ابزارهای پیش‌پردازش داده‌ها است. این نرم‌افزار به گونه‌ای طراحی شده که می‌توان به سرعت روش‌های موجود را به صورت انعطاف‌پذیری روی مجموعه‌های جدید داده، آزمایش نماید. این نرم‌افزار، پشتیبانی‌های الگوریتمی بسیار ارزشمندی را برای کل فرایند داده‌کاوی‌های فراهم می‌کند. این پشتیبانی‌ها شامل آماده‌سازی داده‌های ورودی، ارزیابی آماری چارچوب‌های یادگیری و نمایش گرافیکی داده‌های ورودی و نتایج یادگیری را در برمی‌گیرند. این نرم‌افزار شامل ابزارهای متنوع پیش‌پردازش داده‌ها است.

جعبه‌ابزار متنوع و جامع، از طریق یک واسط متداول در دسترس است، به نحوی که کاربر می‌تواند روش‌های متفاوت را در آن با یکدیگر مقایسه کند و روش‌هایی را که برای مسایل مد نظر مناسب تر هستند، تشخیص دهد. هر محصول یک وب سایت خوب توسعه یافته دارد.

سادگی یادگیری نرم‌افزار باعث می‌شود تا برای شروع داده‌کاوی معرفی وکا گزینه‌ای مناسب محسوب شود. تقریباً همه شرکت‌های بزرگی که ابزاری برای کار با داده‌ها عرضه می‌کنند، امکانات داده‌کاوی را نیز ارائه می‌کنند. اما در میان نرم‌افزارهای آزاد وکا چیز دیگری است. این نرم‌افزار برای نخستین بار در سال ۱۹۹۷ به شکل مدرنش نوشته شده و تحت مجوز GPL عرضه شد. پلتفرم جاوایی وکا باعث می‌شود تا اجرای آن روی بیشتر سیستم‌عامل‌ها امکان‌پذیر باشد. این سیستم همچنین به یک API عمومی مجهز است که شما می‌توانید با گنجاندن آن در برنامه خودتان از امکانات داده‌کاوی وکا سود

برید. [11]

شروع کار با وکا

آموزش این قسمت از سایت گرفته شده است.

برای اجرای وکا به نصب JRE روی سیستم نیاز است. در صورتی که این موضوع صورت گرفته باشد، مشکل خاص دیگری در اجرای این نرم‌افزار نخواهید داشت.

صفحه آغازین شامل ۴ گزینه است:

1. Explorer: محیطی برای مکاشفه در داده‌ها با استفاده از وکا
2. Experimenter: محیطی برای انجام آزمون و انجام آزمایش‌های آماری میان روش‌های مختلف یادگیری
3. Knowledge Flow: محیطی که تقریباً تمام کارایی اکسپلورر را پشتیبانی می‌کند، اما در عوض از یک رابط drag-and-drop استفاده می‌کنند.
4. SimpleCLI: رابط ساده خط فرمان. که به شما اجازه دسترس مستقیم به دستورات وکا را می‌دهد.

پیط شوید



نتیجه‌گیری

امروزه با گسترش سیستم های پایگاهی و حجم بالای داده‌های ذخیره‌شده در این سیستم‌ها، نیاز به ابزاری است تا بتوان داده‌های ذخیره‌شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربران قرارداد و این داده‌های به نحو احسن به دانش ارزشمند تبدیل کرد. ضعف شدید شرکت‌ها که در حجم عظیمی از داده گیر افتاده و فاقد اطلاعات یا دانش باارزشی که این داده‌ها هستند مشاهده می‌شود. دلیل این امر را می‌توان در کم بودن افراد متخصص و قیمت بالای پروژه‌های داده‌کاوی گفت. دانشی که توسط داده‌کاوی محیا می‌شود می‌تواند مشکلاتی که در سازمان‌ها به‌مرور زمان در لایه‌های شرکت‌ها به وجود آمده است را برطرف کند. در این گشت گزار به دلیل این که خود داده‌کاوی به عنوان ابزاری برای تحلیل داده‌ها استفاده می‌شود باید داده‌ها را برای ابراز تهیه کرد و این قسمت 66 درصد از هزینه داده‌کاوی را شامل می‌شود. کارهایی در زمینه پیش‌پردازش بسیار کارا خواهد بود. پیشنهاد ما که بنا به تجربه ای که در این پروژه به‌دست آورده‌ایم این است که سوخت این موتور قدرتمند باید با دقت تامین شود و از رشته دانشگاهی که باشد مهم نیست به دلیل این که داده‌کاوی توانمندی پردازش تمام این اطلاعات را دارد. پس تحقیقاتی باید در رشته‌های مختلف صورت گیرد تا بتوان این داده‌ها را به نحو احسن به‌دست آورد و در تحلیل از داده‌کاوی استفاده نمود. اگر پروژه‌های بعدی در این راستا صورت گیرد بسیار کارا خواهد بود. در آخر به این جمله مهم از دانشگاه MIT بسنده می‌کنیم که جان کلام را می‌گوید: دانش داده‌کاوی یکی از ده دانش در حال توسعه‌ای است که دهه آینده با انقلاب فناوریانه مواجه خواهد ساخت..

سپاسگزاری

در اینجا بر خود لازم می‌بینیم تا از زحمات بی دریغ جناب آقای دکتر کرد رستمی و جناب آقای دکتر دانشیان در به نتیجه رسیدن تحقیقات کمال تشکر را نماییم واز خداوند منان توفیقات روز افزون ایشان را خواستاریم.

منابع مورد استفاده:

- [1] L. Olson, David— Delen, Dursun— Advanced Data Mining Techniques, 2008.
- [2] NONE YE, "THE HAND BOOK OF DATA MINING", ARIZONA STATE UNIVERSITY, 2663.
- [3] CHRIS RYGIELSKI, "DATA MINING TECHNIQUES FOR CUSTOMER RELATIONSHIP 2005.
- [4] Ian H. Witten, Eibe Frank, Mark A. Hall, Data Mining Third Edition 2006.
- [5] Xindong Wu, Vipin Kumar, "Ten algorithm of data mining", 2669.

[6] دکتر شهرابی، جمال، "داده‌کاوی پیشرفته: مفاهیم و الگوریتم‌ها"، انتشارات: جهاد دانشگاهی

[7] دکتر صنیعی آبادی، محمد "داده‌کاوی کاربردی"، انتشارات: نیاز دانش

[8] مناج محمد باقر "مبانی شبکه های عصبی"، انتشارات: دانشگاه امیرکبیر

[9] سروش علیرضا، "داده‌کاوی و هوشمندی کسب و کار"، انتشارات: ناقوس

[16] علیرضا سمیه، "داده‌کاوی و کشف دانش گام به گام با نرم‌افزار کلمنتاین"، انتشارات: خواجه نصیر الدین طوسی