

چگونه از صفر تبدیل به متخصص داده کاوی بشیم؟

طبق آماری که CNN منتشر کرده، داده کاوی یکی از ده شغل پردرآمد دنیاست. داده کاوی یا **Data Mining** یکی از تخصص‌هایی هست که شرکت‌ها به خصوص شرکت‌های بزرگ نیاز زیادی بهش دارند و علم داده در آینده کاربردهای بسیار زیادی خواهد داشت.

با توجه به این مسئله، سوال‌های زیادی به وجود میاد مثل:

داده کاوی چیه و چطوری کار می‌کنه؟ (**data science** چیست ؟ **data mining** چیست ؟)

هرکسی می‌تونه داده کاوی یاد بگیره؟

من علاقه‌مند به داده کاوی هستم، چطوری اصول داده کاوی رو یاد بگیرم؟

داده کاوی کجاها کاربرد داره؟

داده کاوی یا **data mining** چیه؟

داده کاوی علمیه که به ما کمک می‌کنه بدونیم چطوری میشه داده‌ها رو پیدا و دسته‌بندی کرد، به طوری که در مراحل بعد قابل آنالیز و تحلیل باشند.

در واقع، دیتا ماینینگ به زبان ساده، استخراج اطلاعاتیه که میشه با استفاده از اون‌ها، رفتارها و الگوریتم‌هایی شکل داد تا بتونیم مسائل رو بهتر حل کنیم.

کاربردهای داده کاوی یا دیتا ماینینگ چیه؟

در دنیای امروز رقابت بین شرکت‌های بزرگ بسیار سخت و پیچیده شده. به همین دلیل این شرکت‌ها سعی می‌کنن که با به دست آوردن داده‌ها آینده رو ببینن و با کمک داده کاوی، بر رقبای خودشون غلبه کنن.

پس یک نکته‌ی مهم در تحلیل داده و داده کاوی اینه که خیلی از شرکت‌ها بهش احتیاج دارن تا بتونن پیشرفت کنن.

پس دیتا ماینینگ در خیلی از حوزه ها ورود پیدا کرده مثل:

آموزش

مدیریت ارتباط با مشتریان

بازاریابی

پزشکی

بانکداری

بورس

و.....

به نظر می‌رسد در داده کاوی فرصت‌های شغلی زیادی وجود دارد که کمتر بهش توجه شده.

فواید داده کاوی

دیتا ماینینگ برای گروه‌های مختلف می‌تونه فواید و کارایی متفاوتی داشته باشه.

داده کاوی برای شرکت‌های بزرگ می‌تونه رفتار مشتری رو تحلیل کنه و بزرگتر شدن شرکت رو سرعت ببخسه یا به سازمان‌های آموزشی در بالاتر بردن کیفیت تدریس کمک کنه و خیلی کارهای متفاوت و ضروری دیگه.

نرم افزارهای داده کاوی

برای داده کاوی از نرم افزار های مختلفی میشه استفاده کرد از جمله:

Rapidminer

Oracle Data Mining

SPSS Modeler

Orange

KNIME

PYTHON

R

WEKA

و

کلی شرکت‌ها و بیزینس‌های جدید با استفاده از **Data science** ایجاد شدند.

کلی شرکت هستند که می‌تونن در بورس با **Machine learning** معامله کنند.

Google تونسته با **Data science** همه‌ی رقبا رو تقریباً حذف کنه و کلی مثال دیگه.

خیلیا می‌خوان بدونن دیتا ماینینگ به زبان ساده چه مفهومی داره؟

Data mining یا همون **داده کاوی**، اصطلاح قدیمی در این حوزه هست.

اصطلاح **داده کاوی** در سال ۱۹۹۶ مطرح شد.

اون زمان شرکت‌ها، تازه شروع به وارد کردن دیتاها در دیتابیس‌ها کرده بودند و در **داده کاوی** بحث این بود که از این داده‌ها، علم و دانش به دست بیاریم و داده‌ها رو به **داده‌های فرآوری شده** تبدیل کنیم.

سال ۲۰۰۱ به دلیل اینکه **Computer science** و **Data mining** با هم در روند کار ترکیب می‌شدند، کلمه **Data science** تشکیل شد.

Data science این مسئله رو می‌رسونه که ما با استفاده از **علم کامپیوتری** که داریم، **داده کاوی** رو به شکل جدیدی انجام بدیم.

در اون زمان، حجم داده‌ها شدیداً در حال افزایش بود. برای همین به علم کامپیوتر هم بسیار احتیاج می‌شد و این از اولین جاهایی بود که این دو مقوله باهم ترکیب می‌شدند.

وقتی که قرار بود داده‌ها فرآوری بشوند، به علم کامپیوتر نیازی مبرمی داشتند که تازه حوزه‌ی **Data science** شکل گرفت.

یکی از معروف‌ترین تکنیک‌هایی که توی **Data science** استفاده میشه و گل سرسبد این حوزه هست، **AI** یا دقیق‌ترش **Machine learning** و دقیق‌ترش **Deep learning** هست.

AI مخفف **Artificial Intelligence** هستش یعنی شما یک **هوش مصنوعی** درست کنید که بتونه تصمیم بگیره یا پیش‌بینی کنه.

یک تکنیک دقیق ترش Machine learning همیشه که ما از شبکه‌های عصبی مصنوعی برای این کار استفاده کنیم.

این تکنیک براساس یکسری فعل و انفعالاتی هست که توی مغز اتفاق میوفته و سعی می‌کنه نورون‌های مغزی رو به شکل معادلات ریاضی برداری دربیاره.

باز اگر بخواهیم ریزتر بشیم، به Deep learning می‌رسیم.

در Deep learning به این صورت هست که ما از تعداد زیادی از این نورون‌ها به صورت پشت سرهم استفاده می‌کنیم.

قبلا این امکان وجود نداشت و وقتی از تعداد زیادی از این نورون‌ها استفاده می‌کردیم، شبکه Fail می‌کرد؛ بخاطر اینکه نه دیتا و نه توان محاسباتی زیادی داشتیم.

این مسئله با افزایش دیتاها و پیشرفت تکنولوژی میسر شد و تونستیم از شبکه‌های عصبی عمیق هم استفاده کنیم.

مثلا یکی از کاربردهای این شبکه عصبی عمیق، ایجاد خودران‌های خودرو هست.

این سیستم می‌تونه همه چیز رو تشخیص بده و مشخص کنه که چه مسیری باید طی بشه.

در تشخیص صداهایی مثل Siri یا گوگل، این شبکه‌های عصبی عمیق هستند که صحبت ما رو تبدیل به متن می‌کنند؛ حتی اون متن رو پردازش می‌کنند تا تبدیل به یک عمل بشه.

دیتا ماینینگ در کل در چه حوزه‌هایی به درد می‌خوره؟

در واقع دیتا ماینینگ در جاهایی که خیلی معروف هست مثل تشخیص صدا، خودران‌ها، تشخیص چهره، تشخیص اشیا و ... کاربرد داره.

در واقعیت علم داده برای جاهایی کاربرد داره که ما نمی‌تونیم علم ریاضی سفت و سخت به کار ببریم. مثلا من نمیتونم به کامپیوتر بگم که اگر جایی چیزی با زاویه ۱۵ درجه بالای یک گردی دید، اون ابرو هست. برای این جور موارد قانون سفت و سخت همیشه گذاشت.

در این جاهایی که ما هم نمیدونم باید چطوری عمل کنیم، کار رو به ماشین‌ها می‌سپریم.

برای کامپیوتر مسیری تعیین می‌کنیم که این موارد رو ببینه. در واقع ما با این کار داده‌ها رو وارد سیستم می‌کنیم. در انتها هم خود ماشین تصمیم می‌گیره بر اساس این داده‌ها چطوری در آینده تصمیم گیری کنه.

به همین دلیل هست که هرچه داده‌های ما از داده کاوی انجام شده، بیشتر باشه، نتیجه‌ی بهتری خواهیم گرفت.

علم داده به همین دلیل گستره‌ی فراوانی از کارها رو در بر می‌گیره. مثلا در بازار سرمایه و در بورس به کار میاد که کجا بخرم، کجا بفروشم، کجا داره اتفاقات عجیب و خاصی میوفته و ...

این نکات رو میشه با تحلیل داده‌ها به دست آورد.

توی بحث تشخیص چهره، آدم‌ها رو می‌شناسه و حالات صورتشون رو تشخیص میده. بعدها احتمالا بر اساس حالات چهره، دروغ گفتن یا نگفتن فرد رو می‌تونه تشخیص بده.

در تبدیل صدا به متن و یا پردازش متن از داده کاوی استفاده میشه. مثلا در خواندن خبر افراد ظرفیت محدودی دارن و نمی‌تونن همه‌ی خبرها رو رصد کنن.

در علم داده، سیستمی بر اساس **Machine learning** وجود داره که می‌تونه توی یک ثانیه ۱۰۰ تا خبر بخونه و بگه که احتمالا این خبر به موضوع خاصی مرتبط هست و جذاب‌تر اینکه این خبر، خبر مثبت یا منفی در اون حوزه هست.

تقریبا هر چیزی که ما انسان‌ها به شکل کار تخصصی بهش نگاه می‌کنیم، در آینده می‌تونه تبدیل به کار ماشین بشه.

یکی دیگه از موارد جالب **recommender system** هست مثل ویدیوهایی که در قسمت سمت راست، در **YouTube** وجود داره.

این هم کار متخصص داده کاوی هست که اون ویدیوهای پیشنهادی رو چه چیزهایی بذاره که احتمال اینکه شما روش کلیک کنید خیلی زیاد بشه.

به همین دلیل هست که یوتیوب معروف به اینه که مثل گرداب می‌مونه که وقتی توش میری دیگه هیچ وقت بیرون نمیای

با توجه به این موضوعات، همیشه گفت که داده کاوی در حوزه‌ی خاصی کار می‌کنه بلکه هرجایی که نیاز به تصمیم‌گیری، پیش‌بینی و تحلیل‌های گذشته باشه، این سیستم‌ها کارایی دارن

آیا در ایران کاری به اسم متخصص داده کاوی وجود داره و شرکت‌ها به دنبال فردی با این تخصص هستن؟

می‌تونم بگم که آره شرکت‌ها متخصص داده کاوی می‌خوان ولی یک موضوع مهم اینه که تمام این مثال‌هایی که گفتم بیشتر در حوزه‌ی **Machine learning** بود و این حوزه همه‌ی مباحث علوم داده نیست.

اتفاقا در ایران قسمت کوچکی از کار داده کاوی رو در برمی گیره.

داده کاوی قسمت‌های دیگه‌ای هم دارد. ممکنه با یکسری داده‌ها و تحلیل‌های آماری ساده به اطلاعاتی دست پیدا کرد که اتفاقا تاثیر عمیقی روی سیستم بذارن.

مثلا شرکتی یک کمپین تبلیغاتی راه میندازه. کمپینی که مدیر برای اون چندصد میلیون خرج کرده ولی نمی‌دونه کاربرها در سایت چیکار کردن.

یک متخصص داده کاوی باید دست راست مدیر باشه که مسیر کاربرها رو کاملا تحلیل کنه و میزان سود و یا ضرر شرکت از این کمپین‌ها رو اعلام کنه و برای کمپین‌های آتی بتونه راه‌حل‌های مناسب ارائه بده.

در این حالت یک نفر باید داده‌ها رو جمع آوری و ذخیره کنه. این مورد رو بهش میگن مهندسی داده که بحثش در مورد این هست که چطوری مقدار زیادی از داده رو در زمان کم پردازش کنم.

مثلا سایت Facebook تعداد داده‌های وارد شده بهش زیاد هست و به تعداد زیادی مهندس داده احتیاج داره تا بتونه به فرد به سرعت پاسخ بده. مهندس داده تقریبا قسمت سخت افزاری کار رو انجام میده.

قسمت بعدی، تحلیل رفتاری ساده هست.

توی این قسمت فرد تحلیل می‌کنه که افراد چه کارهایی انجام دادن و شرکت چقدر به اهدافی که مد نظرش بوده، رسیده.

مثلا در بعضی موارد چند سناریو طراحی میشه تا ببینیم کدوم سناریو بهتره و کدوم ارزش بیشتری تولید می‌کنه.

قسمت تحلیل‌های ساده، قسمت اعظم داده کاوی در کل دنیاست که در ایران هم پررنگ‌تره.

اما **Deep learning** قسمتی هست که بیشتر خبرساز میشه و برای مردم جذابیت داره.

در ایران بخش **Deep learning**، کمتر کارایی داره.

اولا چون شرکت‌ها اونقدر دیتا ندارن که بخوان این کار رو انجام بدن.

دوما حتی اگر حجم خوبی از دیتا رو هم داشته باشن، در قسمت‌های پایینی مثل تحلیل ساده، اونقدر مشکل دارن که هنوز از لحاظ اقتصادی بهتره این معضلات رو برطرف کنن و تحلیل ساده انجام بدن.

به این دلایل احتیاج به سیستم تحلیلی عمیق و **Machine learning** در ایران تقریبا کمه.

خبر خوب اینه که شرکت‌های زیادی هستن که برای تحلیل‌های نه چندان سخت، احتیاج به فردی دارند که در حوزه‌ی داده کاوی متخصص باشه.

از توانایی‌هایی که برای متخصص داده کاوی مهم هست، میشه به:

قدرت انجام تحلیل‌های آماری

داشتن تفکر نقادانه

و کمی هم دانستن بیزینس

اشاره کرد که شرکت‌های زیادی هستن که به افرادی با این ویژگی‌ها احتیاج دارن.

به نظرت کسایی که رشته‌ی دانشگاهی شون هیچ ارتباطی به داده کاوی نداره و یا کلا از دیتا ماینینگ هیچی نمی‌دونن، می‌تونن وارد مراحل یادگیری داده کاوی بشن؟

یک رشته به اسم علوم داده وجود داره که دقیقا روی این حوزه، در مقطع ارشد وجود داره.

اگر قرار باشه که من یک درس از دانشگاه رو بگم که به درد کار داده کاوی می‌خوره، آمار و احتمال هستش که اکثر بچه‌های مهندسی می‌گذرونن.

به نظرم خیلی مسئله‌ی مهمیه و خیلی از مواردی که در آمار نیاز داریم رو پوشش میده.

البته خیلی مسائل هست که توی دانشگاه‌ها درس داده نمیشه مثل داشتن تفکر نقادانه. من بعضی وقتا تحلیل‌هایی رو از بچه‌ها می‌بینم که مثلا هر وقت این اتفاق افتاده، در نتیجه اتفاق فلان هم افتاده. بعد نگاه می‌کنم که فرد اصلا توجه نکرده که داده‌های ورودی این قضیه‌ی چه شکلیه یا اصلا خوب فکر نکرده که این مقدار داده اصلا قابل تعمیم هست یا نه.

بچه‌ها توی خیلی از این موارد ابتدایی و ساده ضعف دارن. توی آمار در دانشگاه شاید بهتر باشه در کنار فرمول‌هایی که آموزش داده میشه، تفکر انتقادی هم آموزش داده بشه.

داشتن دانش بیزینسی هم خیلی می‌تونه کمک کنه مخصوصا به کسایی که می‌خوان در قسمت تحلیل‌های نسبتا ساده‌ای که گفتم کار کنن.

البته منظور از تحلیل ساده، این نیست که کار ساده‌ای باشه بلکه پیچیده هم هست ولی بیشتر روی آنالیز داده‌ها متمرکزه.

معمولا افرادی که در این قسمت کار می‌کنن، باید با مدیرهای اجرایی ارتباط نزدیکی داشته باشن تا بتونن در کنار هم کارهاشون رو جلو ببرن.

در این موارد متخصص داده کاوی باید بتونه، رابط مدیر با داده‌های بدست اومده باشه.

پس اگر متخصص داده کاوی یک دید بیزینسی هم داشته باشه، می‌تونه بیشتر از حد تصورات برای سیستم مفید باشه.

معمولا افراد، با مدرک کارشناسی توی داده کاوی کار می‌کنن و طبیعتا کسی از شما انتظار نداره که مدرک خاصی داشته باشین.

رشته‌هایی مثل:

- نرم افزار
- سخت افزار
- IT
- علوم کامپیوتر
- و تقریبا همه‌ی رشته‌های مهندسی

می‌تونن وارد این حوزه مخصوصا توی قسمت آنالیزش بشن.

ولی اگر می‌خوان وارد قسمت Machine learning و Deep learning بشن، معمولا باید ارشد داشته باشن.

شرکت‌های خارجی دوست دارن که حتی شما دکتری این حوزه رو داشته باشین و حتما توش تجربه داشته باشین. حق هم دارن چون این موارد، کارهای بسیار پیچیده‌تریه.

برای یادگیری این حوزه، دوره‌های مختلفی در این حوزه در اینترنت هست. سایت‌هایی مثل Coursera و edX و udacity هستند که دوره‌های این چنینی دارن.

Coursera دوتا از معروف‌ترین دوره‌های این حوزه رو داره.

یکی برای آقای Geoffrey Hinton هست به اسم Neural networks in machine learning که کلاسش رایگانه.

یکی دیگه هم برای آقای Andrew Ng به اسم Deep learning هست.

افرادی هستند که می‌خوان آموزش داده کاوی رو شروع کنن. به نظرت مسیر تبدیل شدن به یک متخصص دیتا ماینینگ چیه؟ چه مهارت‌ها و چه نرم افزارهایی برای داده کاوی باید بلد باشیم.

این مسئله در هر کدام از ۳ بخشی که توضیح دادم متفاوت‌ه.

در حالت اول که بخش مهندسی داده بود، معمولاً باید بچه‌های نرم افزاری و سخت افزاری باشن. به این دلیل که بتونن با سیستم‌های کامپیوتری کارهای پیشرفته‌ای انجام بدن.

این گروه معمولاً لازم دارن که با نرم افزارهای مثل:

- hadoop
- Spark

کار کنن و به دیتابیس‌ها و برنامه نویسی مسلط باشن.

این شد برای مهندس داده که وظیفه‌اش ذخیره سازی حجم زیاد دیتا هست.

قسمت دوم که کارش آنالیز داده‌ها بود، دونستن آمار و احتمالات و کمی برنامه نویسی لازمه.

توی این قسمت نرم افزارهای زیادی استفاده می‌شن ولی به طور معمول:

- پایتون
- R

زبان‌های اصلی در حوزه آنالیز هستن.

R برای تحلیل‌های آماری استفاده میشه و پایتون هم برای برنامه نویسی و هم برای تحلیل‌های آماری کاربرد داره.

قرار نیست که از اول از اینا شروع کنید و احساس کنید که سخته.

من خودم اول از زبان **Woolfram** شروع کردم. یک زبانی هست که شاید خیلی‌ها اسمش حتی به گوششون هم نخورده باشه. توی نرم افزار **Mathematica** کد می‌زدن.

خود نرم افزار **Mathematica** ابزار خوبی برای کارهای آماری و نمایش داده‌ها داشت.

بعدا هم رفتم روی پایتون کار کردم و کار می‌کنم که بتونم باهاش برنامه نویسی هم انجام بدم.

اطلاعات کلی از دیتابیس در این قسمت هم خیلی می‌تونه کمک کنه.

یک مهارت نرمی که خیلی اینجا لازمه تفکر نقادانه یا همون **Critical thinking** هست. بعضی وقتا لازمه که تو سخت گیرانه به دیتاهایی که داری نگاه کنی تا یه وقتی گمراه نشی و مدیران شرکت رو هم گمراه نکنی.

خوب هست که دید کلی در بیزینس و کارهای مدیریتی داشته باشین.

البته یک مسئله‌ی مهم هم این هست که توانایی ارتباطی خوبی باید داشته باشین چون باید یک نفر رو نسبت به موضوعی قانع کنین که با توجه به داده‌ها این کار خوبه یا این کار بده. برخلاف تصور عامه که فکر می‌کنن متخصص داده کاوی یه گوشه‌ای نشسته و فقط با کامپیوتر سر و کار داره، اتفاقا افراد در این حوزه، احتیاج به مهارت ارتباطی بالایی دارن.

اون قسمت **Deep learning** و **Machine learning**، نیاز به اطلاعات ریاضی سفت و سختی داره. معادلات جبری و آماری سنگینی در این قسمت استفاده میشن.

کسی که در این حوزه می‌خواد که کار کنه باید فهم عمیقی از این موضوعات داشته باشه.

خیلی از افرادی که به این حوزه میان رشته‌شون ریاضی یا فیزیک بوده، چون احتیاج به محاسبات سنگینی داره. شاید مجبور باشین یک معادلاتی رو RUN کنین و یک هفته بعد بیاین ببینین که نتیجه‌ی این محاسبات چی شده.

توی این مقام هم لازمه که برنامه نویسی بلد باشین. البته نه در حد زیاد فقط در حدی که ابزارهای **Machine learning** دنیا مثل:

- Tensorflow
- keras
- Open CV

رو بشناسین و بتونی با این‌ها چیزهایی که می‌خواین رو پیاده کنین.

یک فهم کلی از دیتابیس در این قسمت هم لازمه.

توی داده کاوی باید حتما برای جایی کار کنیم یا می‌تونیم به صورت فریلنسر یا شرکت خصوصی کار کنیم؟
و اگر می‌تونیم این کار رو بکنیم چطوری و از کجا باید اقدام کنیم؟

در حوزه‌ی داده کاوی طبیعتا هر سه نوع کار کردن وجود داره ولی در حالت‌های مختلف درجه‌شون فرق می‌کنه.

مثلا در مهندسی داده فریلنسر خیلی زیاده چون خیلی از شرکتها میخوان یک زیرساخت بسازن که بتونن این حجم از پردازش رو مدیریت بکنن.

مهندس داده یکبار اینو میسازه که می تونه به صورت فریلنسر یا شرکت باشه. فرد می تونه شرکتی بزنه که کارش تامین این زیرساختها باشه یا به صورت فریلنسر کار کنه.

در همین حالت هم شرکتها برای تامین زیرساخت و پشتیبانی فردی رو استخدام می کنن.

در بخش آنالیز هم خیلی از شرکتها هستن که نرم افزار تحلیل داده می سازن. من زیاد در مورد فریلنسرهای این حوزه اطلاعی ندارم و ندیدم زیاد کسی در این بخش کار فریلنسری بکنه. توی این قسمت استخدامی خیلی زیاد دیدم.

در قسمت **Deep learning** من فریلنسر کم دیدم چون کاری که قراره صورت بگیره یک کار طولانیه. این کار لزوما جواب نهایی نداره که شرکتی کار رو به فریلنسر بده و رهش کنه. در این قسمت هم بخش استخدامی و هم شرکت زدن مرسومه.

شرکت زدن توی این حوزه جالبه مخصوصا توی آمریکا این اتفاق زیاد میوفته.

مثلا افراد یک شرکت رو تاسیس می کنن و نشون میدن که تواناییهایی دارن. هیچ استفادهای لزوما از شرکت ندارن ولی با اون تواناییها خیلی از مواقع شرکتهای بزرگ این شرکتها رو می خرنند و افراد از این راه کسب درآمد می کنن.

یکی از معروفترین مثالهاش هم شرکت **Deep Mind** هست.

شرکتی بود که در حوزه **Deep learning** کار می کرد و چندسال پیش گوگل این شرکت رو خرید.

در اخبار هم خیلی سر و صدا کرد و گفته شد که این خرید بعد از خرید **اندروید** و **Adwords** بهترین خرید گوگل بوده.

چه آیندهی شغلی رو برای داده کاوی در ایران و جهان متصور هستی؟

داده کاوی در ایران، شاید در حال حاضر یک مقدار ضعیف باشه ولی خیلی داره پیشرفت می کنه.

من حدود یک سال پیش با دکتر اخلاق پور صحبت می کردم. دکتر اخلاق پور، خودشون یکی از متخصصترین افراد ایرانی و حتی جهانی در این حوزه هستن.

ایشون استارت‌آپ در حوزه‌ی هوش مصنوعی در بخش پزشکی در آمریکا ایجاد کرده بودن و استارت‌آپ کاملا موفق‌ی هم بود.

دکتر اخلاق‌پور در حال حاضر یکی از موسس‌های استارت‌آپ استادکار هستن.

من از ایشون پرسیدم که شما با اینکه در اوج تکنولوژی دنیا استارت‌آپ موفق داشتین چرا اینجا اومدین و در حوزه‌ی خدمات که کار غیر تکنولوژیکی هست کار می‌کنین.

اون زمان ایشون گفتند که با خیلی از شرکت‌ها در مورد داده کاوی و علم داده صحبت کردن و به نتیجه نرسیدن.

ولی شرکت‌های بزرگ ما که داده دارن و نمی‌دونستن که باید ازش استفاده کنن، دارن به این سمت گرایش پیدا می‌کنن.

دکتر اخلاق‌پور می‌گفتن که من نمی‌تونستم صبر کنم تا این‌ها به این مرحله برسن. خودم یک بیزینس زدم و از داده کاوی برای کمک به بیزینس‌ام استفاده می‌کنم.

در حال حاضر روز به روز شرکت‌های بیشتری به این سمت میان.

با اینکه کلمه **Data science** اولین بار در سال ۲۰۰۱ مطرح شده، یک سال پیش بود که مجله‌ی کسب و کار هاروارد (HBR) یک مقاله با عنوان «**Data science** جذاب‌ترین شغل قرن ۲۱» چاپ کرد و از ابعاد مختلف مزایای این شغل رو بررسی کرد.

این شغل خیلی در حال رشد و دلایلش هم اینه که حجم داده‌ها روز به روز داره بیشتر میشه. این موضوع موجب شده که تقاضا برای متخصص داده کاوی در حالت‌های مختلف زیاد بشه.

از طرفی هم با بالا رفتن حجم داده‌ها تحلیل‌ها هم دقیق‌تر میشن و شاید به جایی برسیم که تجربه‌ی کامپیوترها اونقدر قوی بشه که خیلی از کارهای تخصصی رو دیگه به انسان نسپرن.

درآمد داده کاوی در ایران چقدره؟ به طور کلی این شغل در کسب و کارها جا افتاده؟

کسای که در داده کاوی کار می‌کنن حقوق خوبی دارن چون کارشون سریع تبدیل به پول یا ارزش برای شرکت میشه و معمولا خوب پول می‌گیرن.

از حدود ماهی ۳,۵ الی ۴ میلیون شروع میشه به بالا.

چون هر شرکتی می‌تونه اعداد مختلفی بده، حقوق‌ها کاملا متفاوته. شرکت‌هایی در ایران هستن که خیلی اعداد بالایی در این حوزه میدن.

ما می‌دونیم که هر کاری سختی‌های خودشو داره، سختی‌هایی که خودت توی دیتا ماینینگ تجربه کردی چیا بوده.

از سختی‌ها و مشکلات داده کاوی به ما بگو.

یک مقدار ناشناخته بودن حوزه داده کاوی مشکلاتی رو به وجود میاره.

من خودم وقتی رفتم سر کار، دوست داشتم که روی **Deep learning** کار کنم ولی وقتی توی شرکت اومدم دیدم که کلی کار آنالیز هست که اول باید این‌ها رو انجام بدم.

البته چیز بدی هم نبود ولی چیزی که من انتظار داشتم نبود.

از طرفی، اگر مدیران شما ندونن که علم داده و داده کاوی چیه، خیلی وقت‌ها شما در شرکت سردرگم توی می‌شید.

خیلی وقت‌ها مدیر نمی‌دونه که از شما چی بخواد و اگر تو هم ندونی که چه چیزی رو باید به مدیر بدی، گیج میشی و شاید احساس کنی که اصلا مفید نیستی.

در مواردی پیش میاد که مدیر از شما انتظارات زیاد داشته باشه. مدیران فکر می‌کنن شما همیشه باید جواب نهایی و درست بدید ولی به این صورت نیست.

خیلی بستگی به مدیر داره ولی من خوش شانس بودم که مدیر خوبی داشتم که این حوزه رو خوب می‌فهمید.

معمولا کسانی که داده کاوی کار می‌کنن، چه چیز هیجان انگیزی به دست میارن؟

یکی از خوبی‌هایی که داره اینه که بعضی وقتا ما آنالیزهایی انجام میدیم و این آنالیزها به سرعت در شرکت خلق ارزش می‌کنه.

به نظرم این خیلی جذابه.

حالا من تبدیل به فردی شدم که داده کاوی بلده. بدون سابقه شغلی چطوری می‌تونم شغل پیدا کنم؟

توی این حوزه همیشه گفت که من فلان مدرک رو دارم و باید یه جورایی با سابقه کاری خودتونو اثبات کنین.

این یعنی اینکه شما توی دور تجربه و کار می‌افتید. پیشنهاد من برای اینجا کارآموزیه.

من خودم اون دوره دوهفته‌ای رو رفته بودم، در اواسط دوران دانشجویی آگهی یک **Summer school** در ارمنستان رو دیدم که شرکت **wolfram** آمریکا این دوره رو برگزار می‌کرد.

<https://www.wolfram.com/?source=nav>

Apply کردم، یک مصاحبه‌ی Skype داشتیم. قبول شدم و برای دو هفته به اونجا رفتم.

یک برنامه‌ی فشرده‌ی دوهفته‌ای که از ۸ صبح تا ۹ شب در حال انجام پروژه‌ها بودیم.

به این صورت بود که همون روز دوم باید پروژه‌ای رو انتخاب می‌کردیم و من یک پروژه حول **Machine learning** انتخاب کردم که در حوزه‌ی **Data science** حساب میشه.

با استادهایی که اونجا بودن، کل این مدت رو کار کردیم و کلی چیز ازشون یاد گرفتیم. در طول این دوره تونستم پروژه‌ای رو برای **تشخیص احساسات از روی چهره** انجام بدم.

در این پروژه، عکس از روی وب کم خونده می‌شد و احساسات شخص رو تشخیص می‌داد.

این در واقع اولین پروژه‌ی عملی بود که در **حوزه داده کاوی** انجام دادم.

این دوره به من این اعتماد به نفس رو داد که می‌تونم توی این حوزه کار عملی انجام بدم.

بعد خودم می‌رفتم توی حوزه‌های مختلف کار انجام می‌دادم.

مثلا رفتم روی آمار بیکاری آمریکا کار کردم. بعد روی بورس یه مقدار تلاش کردم که زیاد نتیجه نداد.

بعد با همین رزومه که چندتا کار انجام داده بودم به شرکت **هاب** (<https://mcihub.ir>) معرفی شدم و اونجا کارم رو به عنوان **Data scientist** شروع کردم.

خودم برای خودم پروژه انجام می‌دادم. اینکه شما خودتون پروژه انجام بدید هم بد نیست ولی به نظرم اپلای کردن برای کارآموزی در یک شرکت راه خوبیه برای اینکه **داده کاوی** رو تجربه کنین.

یک برتری که کارآموزی داره اینه که شما با مسائل واقعی شرکت‌ها دست و پنجه نرم می‌کنین.

مثلا چون من این مرحله رو نرفته بودم، وقتی وارد شرکت شدم تا حدودی جا خوردم چون اون چیزی که از **داده کاوی** در ذهن من بود این نبود.

کارآموزی از لحاظ مهارت‌های ارتباطی هم خیلی کمک می‌کنه و شما در یک کیس واقعی تا حدودی مجبورین با مدیران در ارتباط باشین که قطعاً می‌تونه تجربه‌ی مفیدی باشه.

افراد یا شرکتهای موفق در حوزهی داده کاوی رو معرفی می کنی؟

توی حوزهی داده کاوی، افراد زیادی رو در ایران نمی شناسم فقط دکتر اخلاق پور رو می شناسم و خیلی هم ایشون رو قبول دارم.

در دنیا باز به دلیل اینکه بمبهای خبری در حوزهی **Deep learning** هست، آدمهای معروف این حوزه معمولاً افرادی هستن که در این حوزه کار می کنن.

سلبریتی این حوزه **Andrew Ng** هست که گفتم یک دوره هم در **Coursera** داره.

یک کتاب خوبی هم به اسم **Machine learning yearning** داره و حول این هست که این حوزه رو چطوری جلو ببریم.

دونفر دیگه هم توی این حوزه خیلی قوی هستن.

یکی **Geoffrey Hinton** هست که ایشون هم در **Coursera** دوره ای در این حوزه دارن.

نفر بعدی هم **Yann LeCun** هست که توی این حوزه فرد قدرتمندی محسوب میشه.

از نظر شرکتی هم خیلی از شرکتهای داده کاوی نیاز دارن مخصوصاً شرکتهایی که اطلاعات و دیتاهای زیادی دارن.

ولی طبق معمول بزرگترین شرکت در حوزهی داده کاوی **Google** هست.

هر شغلی روحیات مختلفی رو می طلبه، کار داده کاوی با چه روحیه ای سازگارتره؟

یک روحیه ای که خیلی مهمه اینه که بتونی ابهام رو قبول کنی.

از طرفی یک ذهن پرسشگر هم لازمه که از این دریای دیتایی که وجود داره، پیدا کنه که کدومها می تونن کمکش کنن.

خوبه که یه مقدار سلیقه داشته باشین چون در آخر باید این دیتاها رو در معرض نمایش بذارین. باید تا حد امکان این داده ها رو ساده و قابل فهم به نمایش بذارین و بتونین افراد رو در مورد یک موضوع قانع کنین.

مسیر یادگیری داده کاوی رو خیلی خلاصه بهمون توضیح بده. اینکه از کجا باید شروع به کار در حوزه داده کاوی کنیم و چه مسیری رو طی کنیم؟

واقعیتش اینه که مسیر دقیقی برای این کار وجود نداره ولی میشه تا حدودی برای آشنایی اولیه مسیر مشخص کرد.

اول اینکه شما باید مشخص کنین قراره در کدوم قسمت از داده کاوی کار کنین.

• مهندسی داده

• آنالیز داده

• Machine learning یا همون Deep learning

مسیر هر کدوم تا حدودی متفاوته.

کسی که میخواد مهندس داده بشه، باید درسهای سخت افزاری و نرم افزاری رو خوب بخونه و شاید اگر دانشگاه بره بهتره.

نرم افزارها رو مشخص کردم که چیا هستن ولی دورههای آموزشی برای hadoop و spark رو نمی شناسم ولی اگر بگردین قطعاً پیدا می کنین.

اما در مهندسی داده بیشتر از اینکه مدرک مهم باشه، تجربه ست که اهمیت داره. مثلاً اگر بتونین بگین دو یا سه بار زیرساخت رو ایجاد کردین، قطعاً برای موقعیتهای کاری می تونه خیلی با ارزش باشه. البته این مسئله توی اکثر کارهای کامپیوتری صادقه.

در قسمت آنالیز دادهها همون طوری که گفتم آمار و مقداری برنامه نویسی، تفکر انتقادی و تا حدودی مدیریت یا بیزینس لازمه. این موارد هیچ پیش نیازی نسبت به هم ندارن و میشه اینها رو در کنار هم پیش برد.

آمار و احتمال رو من خودم از دانشگاه یاد گرفتم. نمی دونم دورههایی برای آموزش این مورد هست یا نه.

برای تفکر انتقادی اگر بخوام جایی رو معرفی کنم، میگم که قسمت تفکر انتقادی سایت متمم رو بخونین.

برای برنامه نویسی من خودم پایتون کار می کنم. من هیچ وقت دورهی آموزشی پایتون نگذروندم.

هر موقع که پروژه‌ای برام تعریف میشد، می‌رفتم اون مورد رو یاد می‌گرفتم، مدل ساده شده‌ی طرح رو پیدا می‌کردم و کار می‌کردم تا مسلط می‌شدم. پس میشه بدون هیچ دوره‌ای این موارد رو یاد گرفت.

البته این رو هم بگم که تا حدودی الگوریتم نویسی رو من از قبل بلد بودم و با زبان‌های برنامه نویسی دیگه تا حدودی کار کرده بودم.

باز با این حال با سرچ کردن و گشتن در گوگل، الان به نسبت پایتون رو خوب بلدم.

یکسری ویدیو هم هست که نسیم طالب در یوتیوب خودش منتشر می‌کنه. اسمش رو **Probability Moocs** گذاشته که یه جورایی ترکیب تفکر نقادانه و آماره که ویدیوهای جالبی هم هست. البته چیزی نیست که اگر ببینین کامل این موارد رو یاد بگیرین ولی دید خوبی بهتون میده.

در آخر نکته‌ای اگر در دیتا ماینینگ مونده که نگفتین، خوشحال می‌شیم که بشنویم.

یک نکته‌ای که پیش زمینه‌ی خیلی از مسائل هست، بلد بودن زبان انگلیسیه. باید سعی کنین زبان تون رو خوب کنین تا بتونین اطلاعات بیشتری رو استخراج کنین.

این نکته‌ای بود که البته در خیلی از جاها لازمه ولی میشه گفت در بحث داده کاوی و دیتا ماینینگ به دلیل جدید بودن این حوزه، احتیاج بیشتری به زبان انگلیسی پیدا خواهید کرد.

حالا وقتشه که شروع کنید به یادگیری مهارتهای لازم برای یک متخصص داده.

اگر فکر میکنید یادگیری خودآموز(سلف استادی) مهارتهای مورد نیاز این زمینه شغلی، کار سختیه و تلاشهای گذشته شما برای یادگیری مهارت (مثل زبان انگلیسی یا یک نرم‌افزار) به شکست منجر شده، باید
پیش از هر چیز یادگرفتن رو یاد بگیرید

برای مشاوره شغلی و طراحی رزومه و پیدا کردن موقعیت های

شغلی حوزه داده کاوی و سایر حوزه های کاری مهندسی صنایع با

صنایعشو تماس بگیر

۰۹۳۹۹۶۱۳۷۴۴

واتس اپ ، تماس تلفنی، تلگرام

چنل تلگرام كالج صنایعشو :

<https://t.me/sanayesho>

اینستاگرام كالج صنایعشو:

<https://www.instagram.com/sanaye.sho.clinic>