

# وبینار از علوم اجتماعی تا تحلیل داده

- علی‌رضا چمن‌زار
- مدرسه پردازش و تحلیل دقیقه

## فهرست

2	مقدمه
3	نمونه‌های موفق
3	نمونه‌های علوم اجتماعی محاسباتی
4	نمونه‌ی تحلیل داده در کسب و کار
4	داده Data چیست؟
4	نسبت تحلیل داده و دانشمند داده با داده
5	کاربرد تحلیل داده در کسب و کار
5	بازاریابی و فروش
6	مدیریت عملیات
7	مالی و حسابداری
7	منابع انسانی
7	خدمات مشتری
8	۶ مرحله تجزیه و تحلیل داده
9	۴ نوع تحلیل در تحلیل داده
10	چرخه حیات داده (Data Ecosystem)
10	تصمیم‌گیری مبتنی بر داده و چالش‌های آن
12	چرا پیشینه علوم اجتماعی برای تحلیل داده ارزشمند است؟
14	تفاوت دیتاساینپتیست با دانشمند داده، تحلیلگر داده و مهندس داده

- 15..... نقشه راه رسیدن به شغل تحلیلگر داده
- 15..... مهارت‌های آماری
- 15..... مهارت‌های مورد نیاز برای تحلیلگر داده
- 16..... مهارت‌های آماری و ریاضیاتی مورد نیاز برای دانشمند داده/پژوهشگر علوم اجتماعی محاسباتی
- 17..... مهارت‌های فنی
- 17..... مقدمه
- 18..... ابزارهای محبوب‌تر
- 19..... جدول زبان‌های برنامه‌نویسی و کتابخانه‌ها
- 20..... جدول ابزارهای صفحه‌گسترده
- 21..... جدول ابزارهای بدون کد ((Codeless
- 22..... جدول پلتفرم‌های مصورسازی داده
- 22..... جدول پلتفرم‌های هوش تجاری ((BI
- 23..... جدول ابزارهای تحلیل داده بزرگ
- 24..... جدول ابزارهای یکپارچه‌سازی داده‌ها
- 25..... جدول سیستم‌های مدیریت پایگاه داده
- 26..... مهارت‌های نرم
- 27..... در مسیر علوم اجتماعی محاسباتی ((CSS
- 28..... مهارت‌هایی که به عنوان پژوهشگر علوم اجتماعی محاسباتی باید یاد بگیرید
- 31..... مسیرهای حرفه‌ای تحلیل داده
- 34..... چگونه مهارت‌های خودمون رو بهبود بدیم?
- 35..... آخر قصه

## مقدمه

امروزه داده‌ها به‌طور گسترده‌ای در همه جا وجود دارند و به یکی از مهم‌ترین دارایی‌های سازمان‌ها، حکومت‌ها و البته پژوهشگران تبدیل شده‌اند. از اطلاعات شخصی ما گرفته که البته توی ایران اوپن‌سورس شده تا تراکنش‌های تجاری، میزان داده‌هایی که هر روز تولید می‌شود، خیره‌کننده است. اما دسترسی به داده‌ها به تنهایی کافی نیست؛ بلکه نیازمند تخصصانی هست که توانایی تحلیل و تفسیر داده‌ها را داشته باشند. کسی که با داده‌ها سر و کله بزنه و اونا رو ورز بده تا منجر به تصمیم‌گیری‌های هوشمندانه‌ای مبتنی بر داده بشه یا اگر در علوم اجتماعی محاسباتی استفاده بشه به شناسایی و تحلیل یک مساله اجتماعی برسه.

بگذارید برای روشن‌تر شدن کاربرد تحلیل داده چند تا مثال توی کسب و کار و همچنین کاربرد اون توی علوم اجتماعی محاسباتی بزنم که تصویر ذهنی‌مون از تحلیل داده و تحلیلگر داده روشن‌تر بشه:

## نمونه‌های موفق

### نمونه‌های علوم اجتماعی محاسباتی

پژوهش نظرسنجی‌های موبایلی در کشورهای در حال توسعه که در قالب یک کتاب توسط بانک جهانی در سال ۲۰۱۶ منتشر شده، به بررسی کاربردها و مزایای استفاده از نظرسنجی‌های تلفن همراه برای جمع‌آوری داده‌ها در کشورهای در حال توسعه می‌پردازد. این پژوهش با هدف بهبود جمع‌آوری داده‌ها برای توسعه در کشورهای فقیر انجام شده است. به جای تکیه بر نظرسنجی‌های نمونه‌ای یا سرشماری‌های سنتی، پژوهشگران از داده‌های بزرگ تلفن همراه استفاده کرده‌اند تا تخمین‌های دقیق‌تر و گسترده‌تری از وضعیت فقر در این کشورها ارائه دهند.

1. این تیم با همکاری بزرگترین ارائه‌دهنده تلفن همراه در رواندا، داده‌های تراکنش‌های ناشناس حدود 1.5 میلیون کاربر را بین سال‌های 2005 تا 2009 جمع‌آوری می‌کنند
2. داده‌ها چیا بودن؟ داده‌ها شامل اطلاعاتی درباره زمان شروع، مدت زمان و مکان جغرافیایی تقریبی هر تماس و پیام متنی بودند.
3. مدل‌سازی یادگیری ماشینی (جلوتر در مورد مدل‌سازی و یادگیری صحبت می‌کنیم): پژوهشگران یک مدل یادگیری ماشینی می‌سازند که از رکوردهای تماس برای پیش‌بینی پاسخ‌های نظرسنجی استفاده می‌کند.
4. برای آموزش و ساخت مدل، پژوهشگران با همکاری موسسه فناوری و علم کیگالی یک نمونه تصادفی از حدود هزار کاربر را نظرسنجی کردند.
5. پس از آموزش مدل، از آن برای پیش‌بینی پاسخ‌های نظرسنجی تمامی 1.5 میلیون کاربر استفاده شد.

نتیجه پژوهش چی میشه:

1. تخمین وضعیت مالی: با ترکیب داده‌های مکانی و تخمین‌های وضعیت مالی، پژوهشگران توانستند توزیع جغرافیایی ثروت کاربران را با دقت بالا تخمین بزنند.
2. نتایج این تخمین‌هایی که زدن با نتایج سرشماری‌ها و نظرسنجی‌های دموگرافیک و بهداشتی مقایسه شد و شباهت‌های زیادی بین آنها مشاهده شد، که نشان‌دهنده دقت بالای روش استفاده شده است.
3. اگر چه دقت تخمین‌ها در سطح فردی تا حدودی پرنوسان بود، اما در مجموع توانستند تصویری دقیق از وضعیت فقر در مناطق مختلف رواندا ارائه دهند.

نتیجه نهایی این که همیشه با کلان داده‌ها و هزینه‌ی به مراتب پایین‌تر به نتایج دقیقی رسید و نتیجه موفق این تیم استارتی شد برای پژوهش‌های این چینی در کشورهای در حال توسعه دیگر.

## نمونه‌ی تحلیل داده در کسب و کار

کسب و کارها که بسیار پیش‌تر از دانشمندان علوم اجتماعی با تحلیل داده آشنا بودند و برای نمونه همیشه قصه یوتیوب رو تعریف کرد که با تحلیل داده‌های مخاطبان و ساخت و بهینه‌سازی الگوریتم‌های توصیه ویدیوی شخصی‌سازی شده‌ش، مخاطبش رو به جوری توی سایت نگه میداره که همیشه دومین سایت پربازدید دنیا. این داده‌ها از تاریخچه تماشای کاربر شروع میشه، تا لایک و نظرات و save کردن و ... که البته تا اینجاش حداقلش اینه که داده‌ها ملموس هستند. اما طبق گزارشاتی داده‌های دیگه یوتیوب خیلی دقیق‌تره، در حدی که برای مثال زمانی که در حال اسکرول کردن هستید و یکی دو ثانیه روی ویدیویی مکث می‌کنید، این رو به عنوان علاقه یوزر تعریف می‌کنه و ا به خودت می‌ای اکسپلوررت پیر شده از اون ژانر.

یا قصه فروشگاه اینترنتی **آمازون**؛ آمازون هم مثل شرکت‌های بزرگ دیگه علاقه‌ی زیادی به جمع‌آوری همه‌ی داده‌های کاربراش داره و به قول اون جمله مشهور **Data is information, and information is power** می‌دونه که این داده‌ها چه ارزشی براش دارند. این داده‌ها از اطلاعات شخصی و پرداختی مشتریان فروشگاه تا آی‌پی و مدل و نوع دستگاه‌هایی که کاربر با آن به شبکه آمازون متصل شده‌است را تشکیل می‌دهد و این داده‌ها در مسیر شخصی‌سازی تجربه کاربری، تبلیغات هدفمند و تحلیل بازار و در نهایت تصمیم‌گیری‌های استراتژیک استفاده می‌شود. راه‌اندازی سرویس Amazon Prime و فروش باورنکردنی ناشی از آن از دل تحلیل‌های داده‌های کاربران و علاقه‌ی مشتریان وفادار به تحویل سریع کالا بود.

## داده چیست؟

داده (Data) به مجموعه‌ای از حقایق، اعداد، اندازه‌گیری‌ها، مشاهده‌ها و اطلاعات خام اشاره دارد که می‌تواند به شکل‌های مختلفی مثل اعداد، متون، تصاویر، ویدیوها، صداها و... باشد. این داده‌ها به خودی خود ممکن است معنای خاصی نداشته باشند، اما وقتی که به شکل مناسب جمع‌آوری، سازمان‌دهی، تحلیل و تفسیر شوند، می‌توانند اطلاعات ارزشمندی ارائه دهند که به تصمیم‌گیری‌ها، پیش‌بینی‌ها و حل مسائل کمک کنند.

## نسبت تحلیل داده و دانشمند داده با داده

- **داده** به عنوان ماده خامی است که تحلیل‌گر داده و دانشمند داده با آن کار می‌کنند. داده‌ها پایه و اساس تمامی فعالیت‌های تحلیل و مدل‌سازی هستند.

- **تحلیل داده** فرآیندی است که طی آن داده‌ها تبدیل به اطلاعات می‌شوند. تحلیل‌گر داده با استفاده از تکنیک‌های مختلف سعی می‌کند الگوها، روندها و ارتباطات پنهان در داده‌ها را کشف کند.
  - **تحلیل‌گر/دانشمند داده** کسی است که این فرآیند تحلیل را انجام می‌دهد و نقش اصلی در استخراج اطلاعات و دانش از داده‌ها را دارد. دانشمند داده از ابزارها و تکنیک‌های مختلف برای جمع‌آوری، پردازش، تحلیل و مصورسازی داده‌ها استفاده می‌کند تا نتایج معناداری بدست آورد.
- بنابراین، داده‌ها اساس و نقطه شروع کار تحلیل‌گر داده و دانشمند داده هستند و این افراد با مهارت‌ها و دانش خود داده‌ها را به اطلاعات قابل استفاده تبدیل می‌کنند.

## کاربرد تحلیل داده در کسب و کار

### بازاریابی و فروش

- **تقسیم‌بندی مشتریان:** تقسیم‌بندی مشتریان (Customer Segmentation) فرآیندی است که طی آن مشتریان یک کسب‌وکار بر اساس ویژگی‌های مشترک به گروه‌های مختلف تقسیم می‌شوند. این ویژگی‌ها می‌تواند شامل اطلاعات جمعیت‌شناختی (سن، جنسیت، درآمد)، اطلاعات جغرافیایی (محل سکونت)، رفتار خرید (الگوی خرید، تعداد خریدها)، و روان‌شناختی (علاقه‌مندی‌ها، سبک زندگی) باشد.
- **اهمیت:**
  - **شخصی‌سازی بازاریابی:** با شناخت دقیق هر گروه، می‌توان کمپین‌های بازاریابی هدفمند و شخصی‌سازی شده‌ای برای هر بخش ایجاد کرد.
  - **بهبود ارتباط با مشتری:** تقسیم‌بندی به کسب‌وکارها کمک می‌کند تا ارتباطات بهتری با مشتریان برقرار کنند و پیشنهادات مناسبی برای هر گروه ارائه دهند.
  - **افزایش وفاداری مشتری:** با ارائه خدمات و محصولات منطبق با نیازهای خاص هر گروه، می‌توان رضایت و وفاداری مشتریان را افزایش داد.
  - **بهینه‌سازی منابع:** با تمرکز بر گروه‌های خاص، منابع بازاریابی و تبلیغاتی به طور بهینه‌تری استفاده می‌شود و هزینه‌ها کاهش می‌یابد.

**پیش‌بینی رفتار مشتری:** پیش‌بینی رفتار مشتری (Customer Behavior Prediction) فرآیندی است که با استفاده از تحلیل داده‌های تاریخی و مدل‌های آماری، الگوهای خرید و رفتار آینده مشتریان پیش‌بینی می‌شود. این پیش‌بینی‌ها شامل خریدهای آینده، نیازهای جدید، و واکنش‌های احتمالی مشتریان به کمپین‌ها و محصولات جدید است.

## اهمیت:

- **افزایش فروش:** با پیش‌بینی دقیق نیازها و رفتار مشتریان، می‌توان پیشنهادات شخصی‌سازی شده‌ای ارائه داد که منجر به افزایش فروش می‌شود.
- **کاهش نرخ ریزش مشتری:** شناسایی مشتریانی که احتمالاً از دست می‌روند و ارائه پیشنهادات و مشوق‌های مناسب برای حفظ آن‌ها.
- **بهبود تجربه مشتری:** ارائه تجربه خرید بهتر با پیش‌بینی نیازهای مشتری و آماده‌سازی پاسخ‌های مناسب.
- **توسعه محصولات جدید:** با تحلیل نیازها و ترجیحات مشتریان، کسب‌وکارها می‌توانند محصولات و خدمات جدیدی توسعه دهند که به تقاضای بازار پاسخ دهد.

**تحلیل کارایی کمپین‌های بازاریابی:** تحلیل کارایی کمپین‌های بازاریابی فرآیندی است که در آن اثربخشی و کارایی کمپین‌های تبلیغاتی با استفاده از داده‌ها و شاخص‌های مختلف ارزیابی می‌شود. این شاخص‌ها می‌تواند شامل نرخ بازگشت سرمایه (ROI)، نرخ تبدیل (Conversion Rate)، نرخ باز شدن ایمیل (Open Rate)، و نرخ کلیک (Click-Through Rate) باشد.

## اهمیت:

- **بهینه‌سازی کمپین‌ها:** با تحلیل دقیق عملکرد کمپین‌ها، می‌توان نقاط قوت و ضعف آن‌ها را شناسایی و بهبودهای لازم را اعمال کرد.
- **مدیریت بهتر منابع:** تحلیل عملکرد به کسب‌وکارها کمک می‌کند تا منابع خود را به طور بهینه‌تری تخصیص دهند و هزینه‌های اضافی را کاهش دهند.
- **افزایش اثربخشی تبلیغات:** با ارزیابی دقیق نتایج، می‌توان استراتژی‌های تبلیغاتی موثرتری تدوین کرد که به جذب بیشتر مشتریان منجر شود.
- **آگاهی از رفتار مشتریان:** تحلیل عملکرد کمپین‌ها اطلاعات ارزشمندی درباره واکنش‌ها و ترجیحات مشتریان ارائه می‌دهد که می‌تواند در برنامه‌ریزی‌های آینده مورد استفاده قرار گیرد.

## مدیریت عملیات

- **بهینه‌سازی زنجیره تأمین:** کاهش هزینه‌ها و افزایش کارایی از طریق بهبود فرآیندهای زنجیره تأمین. زنجیره تأمین (Supply Chain) به مجموعه‌ای از فرآیندها و فعالیت‌ها اشاره دارد که از تأمین مواد اولیه تا تولید، انبارداری، حمل و نقل، توزیع و تحویل محصول نهایی به مشتری را شامل می‌شود. مدیریت بهینه زنجیره تأمین با استفاده از تحلیل داده‌ها، به کاهش هزینه‌ها، بهبود کارایی، افزایش رضایت

- مشتری و ایجاد مزیت رقابتی کمک می‌کند. تحلیل داده‌ها می‌تواند در پیش‌بینی تقاضا، بهینه‌سازی موجودی، مدیریت ریسک و بهبود روابط با تأمین‌کنندگان مؤثر باشد.
- **مدیریت موجودی:** پیش‌بینی تقاضا و بهینه‌سازی سطوح موجودی برای جلوگیری از کمبود یا مازاد.

## مالی و حسابداری

- **تشخیص تقلب:** شناسایی الگوهای غیرعادی و تقلب در تراکنش‌های مالی یا برای مثال در حوزه سلامت، یکی از پروژه‌های شرکت دقیقه برای سازمان بیمه سلامت ایران طراحی سامانه‌ای بوده با طراحی الگوریتم‌هایی که می‌تونید توی گزارشی که در سایت هست جزئیاتش رو هم بخونید، «موارد سوءمصرف و تخلف دارویی» رو برای این سازمان شناسایی و کشف می‌کرده
- **مدیریت ریسک:** مدیریت ریسک فرآیندی است که به شناسایی، ارزیابی و کنترل ریسک‌های احتمالی که ممکن است بر اهداف مالی و عملیاتی یک کسب‌وکار تاثیر بگذارند، می‌پردازد. تحلیل داده‌ها در این فرآیند نقش کلیدی دارد؛ با استفاده از داده‌های تاریخی، کسب‌وکارها می‌توانند الگوها و روندهای موجود را شناسایی کرده و از این اطلاعات برای پیش‌بینی ریسک‌های آینده استفاده کنند. با پیش‌بینی دقیق‌تر ریسک‌ها، کسب‌وکارها می‌توانند تصمیمات بهتری در مورد نحوه مدیریت و کاهش ریسک‌ها بگیرند، که منجر به افزایش پایداری و کارایی سازمان می‌شود. اطلاعات حاصل از تحلیل داده‌ها به مدیران کمک می‌کند تا منابع را بهینه‌تر تخصیص دهند و در مواجهه با تغییرات بازار و نوسانات اقتصادی، واکنش‌های سریع و موثرتری داشته باشند.

## منابع انسانی

- **جذب و استخدام:** شناسایی بهترین کاندیداها برای مشاغل با استفاده از تحلیل داده‌های رزومه و مصاحبه.
- **مدیریت عملکرد:** ارزیابی و بهبود عملکرد کارکنان با استفاده از داده‌های عملکردی، به شناسایی نقاط قوت و ضعف، تعیین نیازهای آموزشی، افزایش بهره‌وری و ارائه بازخوردهای سازنده کمک می‌کند، که نتیجه آن تصمیم‌گیری‌های استراتژیک بهتر و افزایش رضایت و انگیزه شغلی کارکنان است.

## خدمات مشتری

- **تحلیل بازخورد مشتری:** جمع‌آوری و تحلیل نظرات مشتریان برای بهبود محصولات و خدمات. با استفاده از تکنیک‌های تحلیل داده می‌توان حجم زیادی از بازخوردهای مشتریان را از منابع مختلف مانند نظرسنجی‌ها، شبکه‌های اجتماعی و تماس‌های مرکز خدمات مشتری جمع‌آوری و تحلیل کرد. این

تحلیل‌ها می‌توانند الگوها و روندهای مهمی را در نظرات مشتریان شناسایی کنند، مشکلات رایج را برجسته کنند و فرصت‌هایی برای بهبود محصولات و خدمات ارائه دهند.

- **پیش‌بینی نیازهای مشتریان:** ارائه خدمات پیشگیرانه بر اساس تحلیل داده‌های مشتریان. مثال ترسناک این پیش‌بینی رو شاید توی اینستاگرام دیده باشید که مثلاً در مورد اون داشتید صحبت می‌کردید یا یه با یه لایک یا Save پست کل اکسپلور خودتون و دوست نزدیکتون پر میشه از همون مورد خاص

## ۶ مرحله تجزیه و تحلیل داده

تجزیه و تحلیل داده‌ها به ما کمک می‌کند تا از داده‌های خام، اطلاعات مفیدی استخراج کنیم و تصمیمات بهتری بگیریم. این مسیر تقریباً همیشه همینه و صرفاً بعضی وقتا ممکنه بین بعضی از مراحل چند بار رفت و برگشت کنید. این مراحل به پژوهشگر کمک می‌کند تا از داده‌ها به نفع تصمیم‌گیری‌های استراتژیک و بهبود عملکرد استفاده کنید و در نهایت به اهداف خود دست یابید.

### ۱. تعریف مسئله

اولین قدم در تجزیه و تحلیل داده‌ها، تعریف دقیق مسئله است. این مرحله شامل شناسایی اهداف تحلیل و پرسش‌های کلیدی است که می‌خواهید به آنها پاسخ دهید. باید مشخص کنید که چه نوع داده‌هایی نیاز دارید و چگونه می‌توانید این داده‌ها را به دست آورید. به عنوان مثال، اگر می‌خواهید بدانید که چه عواملی بر فروش محصولات تأثیرگذار است، باید سوالات مشخصی مانند "آیا قیمت بر فروش تأثیر دارد؟" را مطرح کنید.

### ۲. جمع‌آوری داده‌ها

در این مرحله، داده‌های مورد نیاز را جمع‌آوری می‌کنید. داده‌ها می‌توانند از منابع مختلفی به دست بیایند، مانند پایگاه‌های داده، نظرسنجی‌ها، یا وبسایت‌های مختلف. مهم است که داده‌های جمع‌آوری شده معتبر و مرتبط با مسئله شما باشند. جمع‌آوری داده‌ها ممکن است شامل استخراج اطلاعات از منابع مختلف یا استفاده از ابزارهای مختلف برای جمع‌آوری داده‌های جدید باشد.

### ۳. پاک‌سازی داده‌ها

داده‌های جمع‌آوری شده معمولاً نیاز به پاک‌سازی دارند. این مرحله شامل شناسایی و اصلاح مشکلات موجود در داده‌ها است، مانند داده‌های ناقص، تکراری، یا نادرست. پاک‌سازی داده‌ها به این معناست که باید اطمینان حاصل کنید که داده‌ها برای تحلیل آماده و دقیق هستند. این کار ممکن است شامل حذف خطاها، پر کردن داده‌های گم‌شده، و اصلاح فرمت‌های اشتباه باشد.

### ۴. تحلیل اکتشافی داده‌ها

تحلیل اکتشافی داده‌ها به معنای بررسی اولیه داده‌ها به منظور درک بهتر ساختار و الگوهای آنها است. در این مرحله، با استفاده از ابزارهای مختلف تحلیلی، نمودارها و گراف‌ها، به جستجوی الگوها، روندها و روابط موجود



در داده‌ها می‌پردازید. هدف این است که با تحلیل اکتشافی، نقاط قوت و ضعف داده‌ها را شناسایی کنید و فرضیات اولیه را بررسی کنید.

#### ۵. مدل‌سازی داده‌ها

مدل‌سازی داده‌ها شامل استفاده از روش‌های آماری و در سطوح حرفه‌ای‌تر الگوریتم‌های یادگیری ماشین برای ایجاد مدل‌های تحلیلی است. این مدل‌ها به شما کمک می‌کنند تا پیش‌بینی‌هایی بر اساس داده‌های موجود انجام دهید و الگوهای پنهان را شناسایی کنید. در این مرحله، باید مدل‌های مختلف را تست کنید و بهترین مدل را برای حل مسئله انتخاب کنید.

#### ۶. ارائه نتایج و توصیه‌ها

آخرین مرحله، ارائه نتایج تحلیل و توصیه‌های عملی است. در اینجا، باید نتایج به دست آمده از تحلیل داده‌ها را به شکل واضح و قابل فهم ارائه دهید و توصیه‌های کاربردی برای تصمیم‌گیری‌های آینده ارائه کنید. این مرحله شامل تهیه گزارش‌ها، ارائه‌های تصویری و جلسات مشاوره برای تفسیر نتایج و کمک به اجرای تصمیمات مبتنی بر داده‌ها است.

## ۴ نوع تحلیل در تحلیل داده

**تحلیل توصیفی (Descriptive analytics):** برای شناسایی روندها و خلاصه سازی داده‌ها استفاده می‌شود. به عنوان مثال، یک شرکت تجارت الکترونیک ممکن است از تجزیه و تحلیل توصیفی برای درک اینکه کدام محصولات پرفروش هستند، از کجا مشتریان می‌آیند و چقدر برای هر سفارش هزینه می‌کنند، استفاده کند.

**تحلیل تشخیصی (Diagnostic analytics):** برای توضیح اینکه چرا روندها رخ می‌دهد استفاده می‌شود. به عنوان مثال، شرکت تجارت الکترونیک که در بالا ذکر شد، ممکن است از تجزیه و تحلیل تشخیصی برای درک اینکه چرا برخی از محصولات پرفروش‌تر از سایرین هستند استفاده کند. آنها می‌توانند عواملی مانند قیمت، توضیحات محصول، یا تصاویر را تجزیه و تحلیل کنند تا ببینند چه چیزی باعث فروش بیشتر می‌شود.

**تحلیل پیش‌بینی (Predictive analytics):** برای پیش‌بینی روندهای آینده استفاده می‌شود. به عنوان مثال، شرکت تجارت الکترونیک که در بالا ذکر شد، ممکن است از تجزیه و تحلیل پیش‌بینی برای پیش‌بینی تقاضای آتی برای محصولات خاص استفاده کند. این به آنها کمک می‌کند تا به طور موثر موجودی خود را مدیریت کنند و اطمینان حاصل کنند که محصولات مورد نیاز مشتریان را در انبار دارند.

**تحلیل تجویزی (Prescriptive analytics):** برای ارائه توصیه‌هایی بر اساس آن پیش‌بینی‌ها استفاده می‌شود. به عنوان مثال، شرکت تجارت الکترونیک که در بالا ذکر شد، ممکن است از تجزیه و تحلیل تجویزی برای توصیه اقداماتی برای افزایش فروش محصولات خاص استفاده کند. این ممکن است شامل پیشنهاد تخفیف، اجرای کمپین‌های بازاریابی هدفمند، یا بهبود توضیحات محصول باشد.

## چرخه حیات داده (Data Ecosystem)

چرخه حیات داده، سفری هشت مرحله‌ای است و یک فرآیند مداوم و تکرار شونده است که به سازمان‌ها امکان می‌دهد به طور مستمر از داده‌های خود بیاموزند و بهبود یابند.

1. تولید (Generation): داده‌ها از منابع مختلف ایجاد می‌شوند.
2. جمع‌آوری (Collection): داده‌های تولید شده گردآوری می‌شوند.
3. پردازش (Processing): داده‌های خام پاکسازی و آماده‌سازی می‌شوند.
4. ذخیره‌سازی (Storage): داده‌ها در سیستم‌های مناسب نگهداری می‌شوند.
5. مدیریت (Management): کنترل دسترسی و امنیت داده‌ها تأمین می‌شود.
6. تحلیل (Analysis): بینش‌های ارزشمند از داده‌ها استخراج می‌شوند.
7. تجسم‌سازی (Visualization): نتایج تحلیل به صورت گرافیکی ارائه می‌شوند.
8. تفسیر (Interpretation): بینش‌ها به اقدامات عملی تبدیل می‌شوند.

## تصمیم‌گیری مبتنی بر داده و چالش‌های آن

تصمیم‌گیری مبتنی بر داده یک رویکرد قدرتمند در دنیای مدرن کسب و کار و تکنولوژی است. این روش به سازمان‌ها اجازه می‌دهد تا با استفاده از اطلاعات و تحلیل‌های دقیق، تصمیمات آگاهانه‌تری بگیرند. با این حال، این رویکرد نیز با چالش‌هایی روبرو است که یکی از مهم‌ترین آنها سوگیری مبتنی بر داده است. سوگیری مبتنی بر داده زمانی اتفاق می‌افتد که یک مدل یادگیری ماشین با داده‌هایی آموزش می‌بیند که مغرضانه یا نامتعادل هستند. این امر می‌تواند منجر به نتایج نادرست، تبعیض‌آمیز یا ناعادلانه شود. یا در سطح تحلیل داده‌ی فردی، به دلایل که جلوتر می‌گم، تحلیل دچار سوگیری همیشه

### دلایل ایجاد سوگیری در داده‌ها

1. نمونه‌گیری نامناسب: عدم نمایندگی کافی از جامعه هدف در داده‌های آموزشی.
2. داده‌های ناقص: وجود خلأ در اطلاعات یا عدم پوشش تمام جنبه‌های مورد نیاز.
3. ارتباطات نادرست: برچسب‌گذاری اشتباه داده‌ها یا وجود الگوهای نادرست در مجموعه داده.
4. سوگیری تاریخی: انعکاس تبعیض‌های موجود در جامعه در داده‌های جمع‌آوری شده.
5. سوگیری انتخاب: جمع‌آوری داده از منابعی که به طور ذاتی مغرضانه هستند.

## پیامدهای سوگیری مبتنی بر داده

- تصمیمات نادرست و غیرمنصفانه
- تقویت نابرابری‌های اجتماعی
- کاهش اعتماد به سیستم‌های هوش مصنوعی
- زیان‌های مالی و اعتباری برای سازمان‌ها

## روش‌های کاهش سوگیری مبتنی بر داده

### 1. استفاده از داده‌های متنوع:

- جمع‌آوری داده از منابع متنوع و نماینده جامعه هدف
- همکاری با متخصصان حوزه‌های مختلف برای اطمینان از پوشش کامل

### 2. بررسی و پاکسازی داده‌ها:

- شناسایی و حذف داده‌های پرت یا نامعتبر
- تکمیل داده‌های ناقص با روش‌های علمی و معتبر

### 3. اعتبارسنجی برچسب‌گذاری:

- بازبینی فرآیند برچسب‌گذاری داده‌ها
- استفاده از چندین متخصص برای تأیید صحت برچسب‌ها

### 4. تکنیک‌های پیشرفته کاهش سوگیری:

- نمونه‌گیری بیش از حد (Oversampling) برای گروه‌های کم‌نماینده
- نمونه‌گیری کم (Undersampling) برای گروه‌های بیش‌نماینده
- استفاده از الگوریتم‌های یادگیری عادلانه (Fair Learning Algorithms)

### 5. ارزیابی مستمر مدل:

- تست مدل با داده‌های متنوع و سناریوهای مختلف
- بررسی عملکرد مدل در گروه‌های مختلف جمعیتی

### 6. شفافیت و قابلیت توضیح:

- ایجاد مدل‌هایی که تصمیمات خود را توضیح می‌دهند
- ارائه مستندات کامل در مورد فرآیند توسعه و آموزش مدل

### 7. آموزش و آگاهی‌سازی تیم:

- برگزاری دوره‌های آموزشی در مورد سوگیری داده و اخلاق در هوش مصنوعی
- ایجاد فرهنگ سازمانی حساس به مسائل اخلاقی و عدالت

## نتیجه‌گیری

تصمیم‌گیری مبتنی بر داده، با وجود چالش‌هایی مانند سوگیری، همچنان یک رویکرد ارزشمند است. با آگاهی از این چالش‌ها و اتخاذ استراتژی‌های مناسب برای کاهش سوگیری، می‌توان از مزایای این رویکرد بهره‌مند شد و در عین حال، تصمیمات عادلانه و دقیق‌تری اتخاذ کرد. مسئولیت متخصصان داده و مدیران سازمان‌ها است که اطمینان حاصل کنند سیستم‌های تصمیم‌گیری مبتنی بر داده، منصفانه، شفاف و قابل اعتماد هستند.

## چرا پیشینه علوم اجتماعی برای تحلیل داده ارزشمند است؟

در عصر حاضر که داده‌ها نقشی محوری در تصمیم‌گیری‌های سازمانی و اجتماعی ایفا می‌کنند، ترکیب مهارت‌های تحلیلی با درک عمیق رفتار انسانی از اهمیت فزاینده‌ای برخوردار شده است. تا اینجا کار باید به تصویری از تحلیل داده داشته باشیم. حوزه‌ای که به دلیل کار با ابزارهایی مثل زبان‌های برنامه‌نویسی تا پیش از این مختص بچه‌های کامپیوتر، صنایع، کلا فنی‌ها بود اما نگاه تماماً مهندسی این دوستان نمی‌توانست پاسخگوی همه‌ی تحلیل‌ها، خصوصاً تحلیل‌هایی که یک پای اون به مردم می‌رسد و اونا ذی‌نفعان اون سیاست یا پروژه هستند. خوب بریم که سراغ ویژگی‌ها:

### ۱- مهارت‌های ارتباطی و نوشتاری قوی:

توانایی ارائه یافته‌های پیچیده به شیوه‌ای قابل فهم و متقاعدکننده، یکی از مهارت‌های کلیدی در علم داده است. فارغ‌التحصیلان علوم اجتماعی معمولاً در نگارش گزارش‌های تحقیقاتی و ارائه نتایج به مخاطبان متنوع تبحر دارند. این مهارت می‌تواند در ترجمه یافته‌های فنی تحلیل داده به بینش‌های قابل اجرا برای تصمیم‌گیرندگان سازمانی بسیار ارزشمند باشد.

### ۲- آشنایی با مسائل اجتماعی و فرهنگی:

درک عمیق زمینه‌های اجتماعی و فرهنگی می‌تواند به تفسیر دقیق‌تر داده‌ها کمک کند. فارغ‌التحصیلان علوم اجتماعی می‌توانند نقش مهمی در شناسایی سوگیری‌های احتمالی در مجموعه داده‌ها و تضمین عدالت و انصاف در الگوریتم‌های یادگیری ماشینی ایفا کنند. این مهارت به ویژه در زمینه هوش مصنوعی اخلاقی و مسئولیت‌پذیر حائز اهمیت است.

### ۳- تفکر انتقادی و حل مسئله:

آموزش در علوم اجتماعی بر پرورش مهارت‌های تفکر انتقادی و توانایی حل مسائل پیچیده تأکید دارد. این مهارت‌ها در علم داده، جایی که تحلیلگران اغلب با مسائل چندوجهی و پیچیده روبرو می‌شوند، بسیار ارزشمند است. توانایی بررسی انتقادی فرضیات و تفسیر نتایج در زمینه‌های گسترده‌تر می‌تواند به اتخاذ تصمیمات آگاهانه‌تر منجر شود.

### ۴- تجربه در پژوهش‌های میدانی:

تجربه در انجام پژوهش‌های میدانی می‌تواند در طراحی و اجرای پروژه‌های جمع‌آوری داده در مقیاس بزرگ مفید باشد. فارغ‌التحصیلان علوم اجتماعی معمولاً با چالش‌های جمع‌آوری داده‌های با کیفیت در محیط‌های واقعی آشنا هستند و می‌توانند استراتژی‌های مؤثری برای غلبه بر این چالش‌ها ارائه دهند.

#### **۵- درک عمیق از زمینه اجتماعی و فرهنگی:**

فارغ‌التحصیلان علوم اجتماعی با تکیه بر آموزش‌های گسترده در زمینه‌های روانشناسی، جامعه‌شناسی و انسان‌شناسی، درک عمیقی از پویایی‌های رفتار انسانی دارند. این بینش به آنها امکان می‌دهد تا الگوهای پیچیده داده‌های مرتبط با رفتار انسانی را با دقت بیشتری تفسیر کنند. به عنوان مثال، در تحلیل داده‌های مربوط به رفتار مصرف‌کننده، یک فارغ‌التحصیل علوم اجتماعی می‌تواند عوامل فرهنگی و اجتماعی مؤثر بر الگوهای خرید را شناسایی کند که ممکن است از دید یک تحلیلگر داده با پیشینه صرفاً فنی پنهان بماند.

#### **۶- تسلط بر روش‌های کیفی، کمی و ترکیبی:**

علوم اجتماعی به طور سنتی بر اهمیت روش‌های کیفی و رویکردهای چندرشته‌ای تأکید دارند. این مهارت‌ها می‌توانند به غنی‌سازی تحلیل‌های داده کمک کنند. به عنوان نمونه، ترکیب داده‌های کمی حاصل از تحلیل‌های آماری با بینش‌های کیفی حاصل از مصاحبه‌های عمیق می‌تواند به ارائه تصویری جامع‌تر از پدیده‌های مورد مطالعه منجر شود. این رویکرد ترکیبی به ویژه در مطالعات بازار و تحقیقات مصرف‌کننده ارزشمند است.

#### **۷- تجربه در طراحی و اجرای پژوهش:**

فارغ‌التحصیلان علوم اجتماعی معمولاً تجربه قابل توجهی در طراحی و اجرای پروژه‌های تحقیقاتی دارند. این تجربه شامل مهارت در تدوین سؤالات تحقیق، طراحی متدولوژی مناسب، و اجرای روش‌های جمع‌آوری داده است. در حوزه علم داده، این مهارت‌ها می‌توانند به طراحی مؤثرتر پروژه‌های تحلیل داده و تضمین اعتبار و قابلیت اطمینان نتایج کمک کنند.

#### **۸- مهارت‌های تحلیلی و استفاده از نرم‌افزارهای تحلیلی:**

بسیاری از فارغ‌التحصیلان علوم اجتماعی با نرم‌افزارهای تحلیلی مانند SPSS، NVivo و Atlasit آشنایی دارند. این آشنایی اولیه می‌تواند پایه‌ای برای یادگیری سریع‌تر ابزارهای پیشرفته‌تر تحلیل داده باشد. L مثل شروع یادگیری زبان‌های برنامه‌نویسی

#### **۹- آماده‌سازی و یادگیری مداوم:**

علم داده حوزه‌ای است که به سرعت در حال تکامل است و نیاز به یادگیری مداوم دارد. فارغ‌التحصیلان علوم اجتماعی معمولاً مهارت‌های قوی در یادگیری مستقل و انطباق با رویکردهای جدید دارند. این ویژگی می‌تواند به آنها در کسب سریع مهارت‌های فنی مورد نیاز در علم داده کمک کند.

#### **۱۰. مهارت‌های داستان‌سرایی داده**

یکی از مهارت‌های مهم دانشجویان علوم اجتماعی، توانایی در داستان‌سرایی و تفسیر داده‌ها به گونه‌ای است که برای مخاطبان جذاب و قابل فهم باشد. این مهارت‌ها شامل تحلیل داده‌ها، تفسیر نتایج، و ایجاد یک روایت منطقی و جذاب از داده‌ها است. برای مثال تجربه ارائه‌های کلاسی و کنفرانسی به دانشجویان علوم اجتماعی این امکان را می‌دهد که داستان‌های داده‌ای خود را به شکلی جذاب و متقاعدکننده ارائه دهند.

## نتیجه‌گیری:

در حالی که انتقال از علوم اجتماعی به علم داده مستلزم کسب مهارت‌های فنی جدید است، مجموعه متنوعی از مهارت‌ها و دیدگاه‌هایی که فارغ‌التحصیلان علوم اجتماعی به این حوزه می‌آورند، می‌تواند ارزش قابل توجهی و ارزشمندی برای‌شان ایجاد کند. ترکیب درک عمیق رفتار انسانی با مهارت‌های تحلیلی می‌تواند به تولید بینش‌های عمیق‌تر و کاربردی‌تر از داده‌ها منجر شود. با توجه به اهمیت فزاینده تفسیر اخلاقی و مسئولانه داده‌ها، دیدگاه‌های علوم اجتماعی می‌توانند نقش مهمی در شکل‌دهی آینده علم داده ایفا کنند. همچنین این مهارت‌ها به تحلیلگران داده کمک می‌کند تا داده‌ها را به شکلی روایت کنند که هم برای مخاطبان فنی و هم برای مخاطبان غیر فنی قابل فهم و جذاب باشد.

## تفاوت دیتاساینیست یا دانشمند داده، تحلیلگر داده و مهندس داده

وقتی وارد دنیای داده می‌شوید با مفاهیم زیادی آشنا می‌شوید که می‌تونه گیج‌کننده باشه، البته ترسناک نیست. یکی از این موارد پوزیشن‌هایی هست که به صورت مستقیم داده رو با خودشون یدک می‌کشند. جدول زیر با درصدی از خطا، شغل‌های اصلی داده و وظایف و حوزه‌های کاری هر کدام را بررسی کرده‌است؛

مهندس داده (Data Engineer)	دانشمند داده (Data Scientist)	تحلیلگر داده (Data Analyst)	
طراحی و ساخت زیرساخت‌ها و سیستم‌های مدیریت داده	استفاده از مدل‌های آماری، انجام تحلیل‌های پیشرفته، مدل‌سازی داده‌ها، و پیاده‌سازی الگوریتم‌های یادگیری ماشین	تحلیل داده‌های موجود، استخراج بینش‌ها برای تصمیم‌گیری کسب و کار، تهیه گزارشات و داشبوردهای مختلف	وظایف اصلی
برنامه‌نویسی، مدیریت پایگاه داده، معماری سیستم	آمار، یادگیری ماشین، برنامه‌نویسی، تحلیل داده	تحلیل داده، گزارش‌نویسی، ارتباطات	مهارت‌های کلیدی
Hadoop, Spark, Kafka, Airflow, (Databases (SQL, NoSQL	پایتون، R، TensorFlow، Scikit-learn، Jupyter Notebook	Excel، SQL، Tableau، Power BI	ابزارها

اهداف	استخراج و نمایش بینش‌های تجاری از داده‌ها برای بهبود فرآیندها و تصمیم‌گیری	پیش‌بینی و مدل‌سازی داده‌ها برای ایجاد راه‌حل‌های مبتنی بر داده	ایجاد و مدیریت زیرساخت‌های لازم برای پردازش و ذخیره‌سازی داده‌ها
تحصیلات (معمولا)	لیسانس در رشته‌های مرتبط با آمار، مالی، مدیریت کسب و کار	کارشناسی ارشد یا دکترا در رشته‌های مرتبط با داده‌ها (علوم کامپیوتر، آمار، ریاضیات)	لیسانس یا کارشناسی ارشد در علوم کامپیوتر، مهندسی نرم‌افزار، یا رشته‌های مرتبط
چالش‌ها	تفسیر داده‌های موجود و تبدیل آن‌ها به اطلاعات قابل استفاده	حل مسائل پیچیده و پیاده‌سازی مدل‌های پیش‌بینی با دقت بالا	مدیریت حجم بالای داده‌ها و ایجاد سیستم‌های پایدار و مقیاس‌پذیر
خروجی کار	گزارشات، داشبوردها، تحلیل‌های تجاری	مدل‌های یادگیری ماشین، بینش‌های پیش‌بینی‌کننده	سیستم‌ها و پایگاه‌های داده، لوله‌های داده‌ای (Data Pipelines)

## نقشه راه رسیدن به شغل تحلیلگر داده

### مهارت‌های آماری

تحلیلگران داده، دانشمندان داده و پژوهشگران علوم اجتماعی محاسباتی همگی نیازمند مهارت‌های پیشرفته در ریاضیات و آمار هستند. تحلیلگران داده در ابتدایی‌ترین سطح باید مهارت‌های آمار و ریاضیات پایه را داشته باشند. دانشمندان داده علاوه بر این مهارت‌ها، نیاز به توانایی‌های پیشرفته‌تری در تحلیل داده‌ها و یادگیری ماشین دارند. پژوهشگران علوم اجتماعی محاسباتی باید مهارت‌های تحلیلی و محاسباتی بیشتری داشته باشند تا بتوانند الگوها و ارتباطات پیچیده‌تر اجتماعی را تحلیل کنند.

### مهارت‌های مورد نیاز برای تحلیلگر داده

#### ۱. آمار توصیفی

- کاربرد: خلاصه‌سازی و توصیف ویژگی‌های اصلی یک مجموعه داده.
- نکات عملی: استفاده از میانگین، میانه، مد، انحراف معیار، و مصورسازی داده‌ها با هیستوگرام‌ها و نمودارهای جعبه‌ای برای شناسایی توزیع‌ها و نقاط پرت.

#### ۲. آمار استنباطی

- کاربرد: تعمیم و پیش‌بینی درباره یک جمعیت براساس یک نمونه.
- نکات عملی: استفاده از آزمون‌های  $t$  برای مقایسه میانگین‌ها بین دو گروه و ANOVA برای مقایسه چند گروه. اطمینان از کافی بودن اندازه نمونه برای جلوگیری از خطاهای نوع I و II.

### ۳. تحلیل همبستگی و رگرسیون

- کاربرد: بررسی روابط بین متغیرها و انجام پیش‌بینی‌ها.
- نکات عملی: استفاده از همبستگی پیرسون برای اندازه‌گیری روابط خطی بین متغیرها و رگرسیون خطی برای پیش‌بینی. بررسی ضریب تعیین ( $R^2$ ) برای ارزیابی مناسب بودن مدل.

### ۴. جبر خطی

- کاربرد: مدیریت داده‌ها در قالب ماتریس و انجام تبدیلات خطی.
- نکات عملی: درک عملیات ماتریسی مانند جمع، ضرب و وارون‌سازی و استفاده از ارزش‌های ویژه و بردارهای ویژه برای تکنیک‌های کاهش ابعاد مانند PCA.

## مهارت‌های آماری و ریاضیاتی مورد نیاز برای دانشمند داده/پژوهشگر علوم اجتماعی محاسباتی

### ۱. حساب دیفرانسیل و انتگرال

- کاربرد: تحلیل تغییرات و بهینه‌سازی توابع برای بهبود عملکرد مدل‌ها.
- نکات عملی: استفاده از مشتقات برای درک نرخ تغییر و بهینه‌سازی توابع، و انتگرال‌ها برای محاسبات مساحت زیر منحنی‌ها.

### ۲. نظریه احتمال

- کاربرد: مدل‌سازی و کمی‌سازی عدم قطعیت و رویدادهای تصادفی.
- نکات عملی: استفاده از توزیع‌های احتمالی (مانند نرمال، دوجمله‌ای) برای مدل‌سازی انواع مختلف داده‌ها و محاسبه ارزش مورد انتظار و واریانس برای درک نتایج متوسط و ریسک.

### ۳. تحلیل سری‌های زمانی

- کاربرد: تحلیل و پیش‌بینی داده‌های جمع‌آوری‌شده در طول زمان.
- نکات عملی: استفاده از مدل‌های ARIMA برای پیش‌بینی و مدیریت روندها و فصلی‌ها و تجزیه سری زمانی به اجزای روند، فصلی و باقیمانده.

### ۴. یادگیری ماشین و یادگیری عمیق



- **کاربرد:** ساخت مدل‌های پیشرفته قادر به یادگیری از مجموعه‌های بزرگ داده و الگوهای پیچیده.
- **نکات عملی:** استفاده از شبکه‌های عصبی برای وظایف شامل داده‌های با ابعاد بالا مانند تشخیص تصویر و استفاده از روش‌های ترکیبی (ensemble methods) برای بهبود عملکرد پیش‌بینی.

### ۵. آمار بیزی

- **کاربرد:** ترکیب دانش قبلی با شواهد جدید برای به‌روزرسانی پیش‌بینی‌ها و باورها.
- **نکات عملی:** استفاده از به‌روزرسانی بیزی برای دقیق‌تر کردن پیش‌بینی‌ها با دسترسی به داده‌های بیشتر و انتخاب توزیع‌های پیشین براساس دانش کارشناسی یا داده‌های تاریخی.

### ۶. خوشه‌بندی و بخش‌بندی

- **کاربرد:** شناسایی گروه‌های طبیعی در داده‌ها برای استراتژی‌های هدفمند.
- **نکات عملی:** استفاده از خوشه‌بندی K-means برای بخش‌بندی داده‌ها به گروه‌های مجزا براساس شباهت و ارزیابی کیفیت خوشه‌ها با استفاده از نمره سیلوئت یا مجموع مربع‌های درون خوشه‌ای.

### ۳. بهینه‌سازی

- **کاربرد:** حل مسائل بهینه‌سازی با استفاده از تکنیک‌های مختلف.
- **نکات عملی:** استفاده از الگوریتم‌های بهینه‌سازی مانند نزول گرادیان برای بهینه‌سازی پارامترها و استفاده از تکنیک‌های بهینه‌سازی محدودیت‌دار برای مسائل با محدودیت‌های خاص.

### ۷. تحلیل پیشرفته و داده‌کاوی

- **کاربرد:** استخراج الگوهای پیچیده و اطلاعات مفید از داده‌های بزرگ.
- **نکات عملی:** استفاده از تکنیک‌های داده‌کاوی مانند الگوریتم‌های ژنتیک و شبیه‌سازی تبرید برای یافتن راه‌حل‌های تقریبی در فضاهای جستجوی بزرگ.

## مهارت‌های فنی

### مقدمه

در دنیای تحلیل داده‌ها، ابزارها و نرم‌افزارهای متنوعی وجود دارند که به کاربران کمک می‌کنند تا داده‌های خود را تحلیل، مدل‌سازی و مصورسازی کنند. این ابزارها به دسته‌های مختلفی تقسیم می‌شوند که هر یک کاربردها و

مزایای خاص خود را دارند و البته با پیشرفت‌های امروزی همیشه گفت که ابزارهای صفحه گسترده و Codeless توانسته‌اند بسیاری از حوزه‌هایی که پیش از این در اختیار زبان‌های برنامه‌نویسی بود را تحت پوشش قرار بدهند و این نکته‌ی مثبتی است برای کسانی که همیشه از کدنویسی ترس داشته‌اند.

**زبان‌های برنامه‌نویسی** مانند پایتون، آر و جولیا به تحلیلگران داده و دانشمندان داده امکان می‌دهند تا تحلیل‌های پیچیده‌ای را انجام دهند و مدل‌های یادگیری ماشین را پیاده‌سازی کنند. این زبان‌ها دارای کتابخانه‌های غنی و متنوعی هستند که نیازهای مختلف تحلیل داده را پوشش می‌دهند.

**ابزارهای صفحه‌گسترده** مانند اکسل و گوگل شیتس، برای تحلیل‌های ساده و مصورسازی‌های اولیه مناسب هستند. این ابزارها به خاطر رابط کاربری ساده و قابلیت همکاری آنلاین بسیار محبوب هستند.

**ابزارهای بدون کد (Codeless)** مانند RapidMiner و Alteryx به کاربران غیر فنی این امکان را می‌دهند تا بدون نیاز به دانش برنامه‌نویسی، تحلیل‌های پیشرفته‌ای انجام دهند. این ابزارها به خصوص برای کسب‌وکارهایی که نیاز به تحلیل سریع دارند بسیار مفید هستند.

**پلتفرم‌های مصورسازی داده** مانند Tableau و Power BI ابزارهای قدرتمندی هستند که به کاربران امکان می‌دهند تا داشبوردها و گزارش‌های تعاملی ایجاد کنند. این ابزارها به تحلیلگران کمک می‌کنند تا داده‌ها را به صورت بصری نمایش دهند و از این طریق الگوها و روندها را بهتر درک کنند.

**پلتفرم‌های هوش تجاری** مانند Sisense و Domo ابزارهای جامعی هستند که برای تحلیل و مصورسازی داده‌های کسب‌وکاری طراحی شده‌اند. این پلتفرم‌ها امکان یکپارچه‌سازی داده‌ها از منابع مختلف و ایجاد داشبوردهای مدیریتی را فراهم می‌کنند.

**ابزارهای تحلیل داده بزرگ** مانند Apache Hadoop و Apache Spark برای پردازش و تحلیل داده‌های بزرگ و پیچیده طراحی شده‌اند. این ابزارها به خصوص برای شرکت‌هایی که با حجم زیادی از داده‌ها سر و کار دارند بسیار مهم هستند.

**ابزارهای یکپارچه‌سازی داده‌ها** مانند Talend و Apache NiFi برای اتوماسیون جریان داده‌ها و یکپارچه‌سازی داده‌ها از منابع مختلف استفاده می‌شوند. این ابزارها به تحلیلگران کمک می‌کنند تا داده‌های خام را به اطلاعات قابل استفاده تبدیل کنند.

**سیستم‌های مدیریت پایگاه داده** مانند MySQL و PostgreSQL برای ذخیره‌سازی و مدیریت داده‌ها به کار می‌روند. این سیستم‌ها به تحلیلگران امکان می‌دهند تا داده‌های خود را به صورت ساختاریافته ذخیره و بازیابی کنند.

**ابزارهای محبوب‌تر**

برخی از این ابزارها به دلیل قابلیت‌های گسترده‌تر، جامعه کاربری بزرگ‌تر و منابع آموزشی فراوان، محبوبیت بیشتری در میان تحلیلگران داده و دانشمندان داده دارند. ابزارهایی مانند پایتون، آر، اکسل، Tableau و Apache Spark به دلیل کاربردهای فراوان و کارایی بالایی که ارائه می‌دهند، کاربران بیشتری دارند و در جداول زیر با یک ستاره (\*) مشخص شده‌اند. در لیست زیر به صورت تفصیلی تلاش شده نام ابزارهای زیادی ذکر شود و این لیست بالا و بلند به این منظور نیست که لازم است در تمامی آنها مهارت داشته باشید.

## جدول زبان‌های برنامه‌نویسی و کتابخانه‌ها

زبان‌های برنامه‌نویسی و کتابخانه‌های مختلف ابزارهایی اساسی برای تحلیل و پردازش داده‌ها فراهم می‌کنند. این زبان‌ها به دلیل انعطاف‌پذیری و قابلیت‌های گسترده‌شان در میان تحلیلگران داده و دانشمندان داده بسیار محبوب هستند.

کاربرد	توضیح	ابزار/کتابخانه	زبان برنامه‌نویسی
تحلیلگر داده، دانشمند داده	ساختارها و توابعی برای دستکاری و تحلیل مؤثر داده‌های ساختاریافته.	*pandas	پایتون (Python)
تحلیلگر داده، دانشمند داده	پشتیبانی از عملیات عددی، از جمله پردازش آرایه‌ها و توابع ریاضی.	*numpy	
تحلیلگر داده، دانشمند داده	عملکردهای اضافی برای محاسبات علمی و فنی، از جمله بهینه‌سازی و آزمون‌های آماری.	*scipy	
تحلیلگر داده، دانشمند داده	کتابخانه‌ای پایه برای ایجاد مصورسازی‌های استاتیک، تعاملی و متحرک در پایتون.	*matplotlib	
تحلیلگر داده، دانشمند داده	بر پایه matplotlib، رابط سطح بالایی برای گرافیک‌های آماری جذاب و اطلاع‌رسان.	seaborn	
تحلیلگر داده، دانشمند داده	ایجاد مصورسازی‌های تعاملی و مبتنی بر وب، از جمله نمودارهای پیچیده و داشبوردها.	plotly	
تحلیلگر داده، دانشمند داده	ابزارهایی برای یادگیری ماشین و مدل‌سازی آماری، از جمله طبقه‌بندی، رگرسیون و خوشه‌بندی.	*scikit-learn	

تخلیگر داده، دانشمند داده	کلاس‌ها و توابعی برای برآورد و تفسیر مدل‌های آماری.	statsmodels	
تخلیگر داده، دانشمند داده	سیستمی برای ایجاد گرافیک‌های پیچیده بر اساس "گرامر گرافیک" برای مصورسازی‌های پیچیده.	*ggplot2	آر (R)
تخلیگر داده، دانشمند داده	گرامری برای دستکاری داده‌ها، که کار با داده‌ها را ساده‌تر و شهودی‌تر می‌کند.	*dplyr	
تخلیگر داده، دانشمند داده	کمک به تبدیل داده‌ها به فرمتی مناسب برای تحلیل.	*tidyr	
تخلیگر داده، دانشمند داده	ایجاد برنامه‌های وب تعاملی مستقیماً از R، مناسب برای ساخت داشبوردها و مصورسازی‌ها.	shiny	
تخلیگر داده، دانشمند داده	ساده‌سازی کار با تاریخ‌ها و زمان‌ها در R، با توابعی برای تجزیه و دستکاری اشیاء تاریخ و زمان.	lubridate	
تخلیگر داده، دانشمند داده	ساده‌سازی فرآیند ایجاد مدل‌های پیش‌بینی با ابزارهایی برای آموزش و ارزیابی الگوریتم‌های یادگیری ماشین.	caret	
تخلیگر داده، دانشمند داده	ساختاری انعطاف‌پذیر و کارآمد برای دستکاری داده‌های جدولی.	DataFrames.jl	جولیا (Julia)
تخلیگر داده، دانشمند داده	کتابخانه‌ای چندمنظوره برای تولید مصورسازی‌ها با پشتیبانی از انواع بک‌اندها.	Plots.jl	
تخلیگر داده، دانشمند داده	توابع آماری پایه و ابزارهایی برای تحلیل داده‌ها.	StatsBase.jl	
تخلیگر داده، دانشمند داده	چارچوبی برای یادگیری ماشین در جولیا، با ابزارهایی برای انتخاب و ارزیابی مدل‌ها.	MLJ.jl	

## جدول ابزارهای صفحه‌گسترده

ابزارهای صفحه‌گسترده مانند اکسل و گوگل شیتس، برای تحلیل‌های ساده و مصورسازی‌های اولیه مناسب هستند. این ابزارها به خاطر رابط کاربری ساده و قابلیت همکاری آنلاین بسیار محبوب هستند.

ابزار	توضیح	کاربرد
مایکروسافت اکسل* (Microsoft Excel)	نرم افزار صفحه گسترده ای که امکانات پایه ای برای تحلیل داده ها، مصورسازی و مدل سازی فراهم می کند.	تحلیلگر داده*
گوگل شیت* (Google Sheets)	ابزار صفحه گسترده مبتنی بر وب که امکانات همکاری و ویژگی های پایه ای مصورسازی داده ها را ارائه می دهد.	تحلیلگر داده
لیبر آفیس کالک (LibreOffice Calc)	نرم افزار صفحه گسترده متن باز با قابلیت های مشابه اکسل، از جمله ابزارهای تحلیل و مصورسازی داده ها.	تحلیلگر داده

### جدول ابزارهای بدون کد (Codeless)

ابزارهای بدون کد به کاربران غیر فنی این امکان را می دهند تا بدون نیاز به دانش برنامه نویسی، تحلیل های پیشرفته ای انجام دهند. این ابزارها به خصوص برای کسب و کارهایی که نیاز به تحلیل سریع دارند بسیار مفید هستند.

ابزار	توضیح	کاربرد
RapidMiner*	محیطی یکپارچه برای آماده سازی داده، یادگیری ماشین و پیاده سازی مدل ها بدون نیاز به کدنویسی گسترده.	تحلیلگر داده، دانشمند داده
*Knime	محیط برنامه نویسی بصری برای تحلیل داده ها و یادگیری ماشین با تمرکز بر گردش کارهای مدولار.	تحلیلگر داده، دانشمند داده
Alteryx	تسهیل ترکیب داده ها و تحلیل پیشرفته از طریق رابط کاربری دوستانه با قابلیت کشیدن و رها کردن.	تحلیلگر داده

تحلیلگر داده، دانشمند داده	خودکارسازی فرآیند ساخت و پیاده‌سازی مدل‌های یادگیری ماشین با استفاده از رابط بدون کد.	<b>DataRobot</b>
----------------------------	---	------------------

### جدول پلتفرم‌های مصورسازی داده

پلتفرم‌های مصورسازی داده به کاربران امکان می‌دهند تا داشبوردها و گزارش‌های تعاملی ایجاد کنند. این ابزارها به تحلیلگران کمک می‌کنند تا داده‌ها را به صورت بصری نمایش دهند و از این طریق الگوها و روندها را بهتر درک کنند.

ابزار	توضیح	کاربرد
<b>*Tableau</b>	ابزار قدرتمند مصورسازی داده که داشبوردهای تعاملی و قابل اشتراک‌گذاری را با قابلیت کشیدن و رها کردن ایجاد می‌کند.	تحلیلگر داده
<b>*Power BI</b>	سرویس تجزیه و تحلیل تجاری که امکان ایجاد گزارش‌ها و داشبوردهای تعاملی را فراهم می‌کند.	تحلیلگر داده
<b>*QlikView</b>	ابزار مصورسازی داده که به کاربران امکان می‌دهد داده‌ها را کاوش و تحلیل کنند.	تحلیلگر داده
<b>D3.js</b>	کتابخانه جاوا اسکریپت برای تولید مصورسازی‌های داده تعاملی و مبتنی بر وب.	تحلیلگر داده، دانشمند داده

### جدول پلتفرم‌های هوش تجاری (BI)

پلتفرم‌های هوش تجاری ابزارهای جامعی هستند که برای تحلیل و مصورسازی داده‌های کسب‌وکاری طراحی شده‌اند. این پلتفرم‌ها امکان یکپارچه‌سازی داده‌ها از منابع مختلف و ایجاد داشبوردهای مدیریتی را فراهم می‌کنند.

ابزار	توضیح	کاربرد
Sisense	پلتفرمی برای تجزیه و تحلیل و مصورسازی داده‌ها با تمرکز بر کسب‌وکارهای بزرگ و داده‌های پیچیده.	تحلیلگر داده، دانشمند داده
Domo	پلتفرم هوش تجاری مبتنی بر ابر که به کسب‌وکارها کمک می‌کند داده‌ها را به صورت یکپارچه مدیریت و تحلیل کنند.	تحلیلگر داده، دانشمند داده
*Looker	ابزار تجزیه و تحلیل داده‌ها که امکان ایجاد داشبوردها و گزارش‌های تعاملی را فراهم می‌کند.	تحلیلگر داده
MicroStrategy	پلتفرم هوش تجاری با قابلیت‌های گسترده برای تحلیل داده‌ها و ایجاد داشبوردها و گزارش‌های مدیریتی.	تحلیلگر داده، دانشمند داده

### جدول ابزارهای تحلیل داده بزرگ

ابزارهای تحلیل داده بزرگ برای پردازش و تحلیل داده‌های بزرگ و پیچیده طراحی شده‌اند. این ابزارها به خصوص برای شرکت‌هایی که با حجم زیادی از داده‌ها سر و کار دارند بسیار مهم هستند.

ابزار	توضیح	کاربرد
-------	-------	--------

تحلیلگر داده، دانشمند داده	چارچوبی برای پردازش و ذخیره‌سازی داده‌های بزرگ به صورت توزیع‌شده.	<b>Apache Hadoop</b>
تحلیلگر داده، دانشمند داده	موتور پردازش داده‌های بزرگ که به سرعت بالا و قابلیت پردازش بلادرنگ معروف است.	<b>Apache *Spark</b>
تحلیلگر داده، دانشمند داده	چارچوب پردازش جریان داده‌های بزرگ به صورت بلادرنگ و توزیع‌شده.	<b>Flink</b>
تحلیلگر داده، دانشمند داده	سیستم پردازش جریان داده که امکان انتقال و پردازش داده‌های بلادرنگ را فراهم می‌کند.	<b>Kafka</b>

### جدول ابزارهای یکپارچه‌سازی داده‌ها

ابزارهای یکپارچه‌سازی داده‌ها برای اتوماسیون جریان داده‌ها و یکپارچه‌سازی داده‌ها از منابع مختلف استفاده می‌شوند. این ابزارها به تحلیلگران کمک می‌کنند تا داده‌های خام را به اطلاعات قابل استفاده تبدیل کنند.

کاربرد	توضیح	ابزار
تحلیلگر داده، دانشمند داده	ابزار یکپارچه‌سازی داده‌ها با قابلیت‌های گسترده برای استخراج، تبدیل و بارگذاری (ETL) داده‌ها.	<b>Talend</b>
تحلیلگر داده، دانشمند داده	ابزار یکپارچه‌سازی داده‌ها و مدیریت جریان داده با قابلیت‌های اتوماسیون.	<b>Apache NiFi</b>



داده، تحلیلگر دانشمند داده	پلتفرم یکپارچه‌سازی داده‌ها با تمرکز بر مدیریت داده‌ها و کیفیت داده‌ها.	<b>Informatica</b>
داده، تحلیلگر دانشمند داده	پلتفرم یکپارچه‌سازی داده‌ها با قابلیت اتصال به منابع داده مختلف و مدیریت API‌ها.	<b>MuleSoft</b>

### جدول سیستم‌های مدیریت پایگاه داده

ابزار	توضیح	کاربرد
<b>*MySQL</b>	سیستم مدیریت پایگاه داده رابطه‌ای متن‌باز که برای مدیریت داده‌های ساختاریافته استفاده می‌شود.	تحلیلگر داده
<b>PostgreSQL</b>	سیستم مدیریت پایگاه داده رابطه‌ای متن‌باز با قابلیت‌های پیشرفته و پشتیبانی از انواع داده‌ها.	داده، تحلیلگر دانشمند داده
<b>MongoDB</b>	پایگاه داده NoSQL با ساختار سندی که برای مدیریت داده‌های بزرگ و غیرساختاریافته استفاده می‌شود.	داده، تحلیلگر دانشمند داده
<b>Oracle DB</b>	سیستم مدیریت پایگاه داده رابطه‌ای با قابلیت‌های گسترده برای مدیریت داده‌های بزرگ و پیچیده.	داده، تحلیلگر دانشمند داده

سیستم‌های مدیریت پایگاه داده برای ذخیره‌سازی و مدیریت داده‌ها به کار می‌روند. این سیستم‌ها به تحلیلگران امکان می‌دهند تا داده‌های خود را به صورت ساختاریافته ذخیره و بازیابی کنند. این جداول و ابزارها، به تحلیلگران

و دانشمندان داده کمک می‌کنند تا با استفاده از ابزارهای مناسب، تحلیل‌های دقیق‌تر و مؤثرتری انجام دهند و داده‌ها را به بینش‌های ارزشمندی تبدیل کنند.

## **مهارت‌های نرم**

### **1. ارتباطات مؤثر**

توانایی انتقال مفاهیم پیچیده به زبان ساده و قابل فهم برای مخاطبان مختلف، از جمله مشتریان، همکاران و ذینفعان. ایجاد ارتباط مؤثر برای توضیح داده‌ها و نتایج.

### **2. حل مسئله**

تحلیل و تفسیر داده‌ها برای شناسایی الگوها و روندها. یافتن راه‌حل‌های خلاقانه برای مسائل پیچیده با استفاده از روش‌های تحلیلی و علمی.

### **3. همکاری و تیم‌ورک**

توانایی کار با تیم‌های چند تخصصی، ارائه بازخورد سازنده و اشتراک‌گذاری ایده‌ها و دانش با دیگران. ایجاد هماهنگی و همکاری میان تیم‌ها برای دستیابی به اهداف مشترک.

### **4. هوش هیجانی**

مدیریت و کنترل احساسات خود و دیگران، حفظ نگرش مثبت و مدیریت استرس در مواجهه با چالش‌ها. ایجاد روابط کاری مثبت و موثر.

### **5. انعطاف‌پذیری و تطبیق‌پذیری**

توانایی تطبیق با نیازهای متغیر مشتریان و پروژه‌ها، یادگیری مداوم و پذیرش تکنولوژی‌ها و روش‌های جدید.

### **6. رهبری**

توانایی الهام‌بخشی و انگیزه‌بخشی به تیم، هدایت پروژه‌ها و تصمیم‌گیری مؤثر. نمایش مهارت‌های رهبری در مدیریت پروژه‌ها و تیم‌ها.

### **7. مدیریت زمان**

توانایی اولویت‌بندی وظایف و مدیریت پروژه‌ها به طور مؤثر. کار مستقل و حفظ بهره‌وری در مواجهه با زمان‌بندی‌ها و مهلت‌ها.

### **8. مهارت‌های ارائه**

ایجاد و ارائه گزارش‌ها و ویژوالیزاسیون‌های جذاب و مؤثر، توضیح نتایج به صورت شفاف و قابل فهم برای مخاطبان مختلف.

## 9. کنجاوی

جستجوی عمیق‌تر و طرح سوالات جدید برای حل مشکلات، تحلیل مسائل از زوایای مختلف و یافتن راه‌حل‌های خلاقانه.

## 10. آگاهی فرهنگی

درک و احترام به تفاوت‌های فرهنگی در کار با مشتریان مختلف، ایجاد روابط قوی با مشتریان از زمینه‌های فرهنگی گوناگون.

# در مسیر علوم اجتماعی محاسباتی (CSS)

علوم اجتماعی محاسباتی که استارتش سال ۲۰۰۹ خورده و به عنوان یک رشته معرفی شده، تخصصی میان‌رشته‌ای هست در چند سال گذشته جای پای خودشو محکم کرده. اما فرقی با تحلیلگرهای داده یا دیتاساینس‌ها کجاست؟ مگه پیش از این تحلیلگرها با یه داده‌کاوی از توئیتر یا شبکه‌های اجتماعی دیگه و خوشه‌بندی و ... تحلیل نمی‌کردند؟ تفاوت اساسی و بسیار مهم علوم اجتماعی محاسباتی، نقطه‌ی ورود و اتکای اونه که بر نظریه است و از دریچه نظریه به داده نگاه می‌کنه. برای مثال یکی از کارهایی که برای پاکسازی داده‌ها استفاده میشه، حذف Outlayerها یا همون داده‌های پرت هست. داده‌ی پرت چیه؟ داده‌هایی که یا خیلی بالاتر از حد معمول هستند یا خیلی پایین‌تر از نرمال جامعه و اینجا دقیقا همون جایی هست که علوم اجتماعی محاسباتی حرف برای گفتن داره؛ مطالعه و پرداختن به گروه‌هایی که طرد شده و حاشیه‌ای هستند. البته که این مثال خیلی ساده شده هست و برای این هست که صرفا تصویری توی ذهنتون ساخته بشه.

مهم‌ترین تفاوت علوم اجتماعی سنتی و محاسباتی، اول ماهیت میان‌رشته‌ای این رشته هست که پیش‌زمینه‌هایی در علوم کاربردی، طبیعی، اجتماعی و بیولوژی و ... هست و مورد بعد هم ساختار ویژه‌ای که به طور موثری بر منابع فنی ساخته و پرداخته شده. یا مثلا در علوم اجتماعی سنتی، عموما فرض بر نرمال بودن توزیع نمونه‌های آماری هست در حالی که در مطالعه کلان‌داده‌ها، توزیع‌ها عموما نمایی است. یعنی برخی افراد بیشترین سرمایه‌ها را از شبکه مدنظر در اختیار دارند. مثل توئیتر، لایک و ریتوییت و ... مال یه گروه‌هایی هست که عموم این رو در اختیار ندارند و البته موارد دیگه‌ای که در زمان این وبینار نمی‌گنجه و توی گروه تلگرام می‌تونیم بیشتر در موردش صحبت کنیم.

شاید خنده‌دار به نظر بیاد اما، اولین گام برای ورود به علوم اجتماعی محاسباتی، **آموزش علوم اجتماعی** است. علوم اجتماعی تعداد بی‌شماری از زیر گروه‌های جذاب مانند اقتصاد، جامعه‌شناسی و مردم‌شناسی دارد. مدرک مناسب و تخصص علوم اجتماعی را پیدا کنید، تا جایی که می‌توانید یاد بگیرید.

مهارت‌های فنی و آماری رو که در نقشه راه تحلیلگر داده گفتم و در علوم اجتماعی محاسباتی، همون مهارت‌ها تخصصی‌تر باید آموخته شوند. برای مثال تا پیش از این با زبان برنامه‌نویسی R یا پایتون اگر داده‌ها را پاک می‌کردید و آنها را مصور می‌کردید، در این نقطه باید بتوانید مدل‌سازی کنید و Machin learning بر پروژه‌هایتان پیاده کنید.

علوم اجتماعی محاسباتی به عنوان یک رشته از ابزارهایی مثل مدل‌سازی و شبیه‌سازی، یادگیری ماشین، داده‌کاوی و کار با داده‌های بزرگ، اسکرپینگ وب، تکست‌ماینینگ، رسانه‌های اجتماعی و تحلیل شبکه‌های اجتماعی (Social Network) و البته و دیدگاه‌های جدید برای تجزیه و تحلیل داده‌ها استفاده می‌کند. برای مثال این روزها شاید زیاد تحلیل داده‌های توییتر در موضوعات مختلف دیده‌باشید که به دلیل انتخابات و حاشیه‌های قبل و بعدش خیلی بولد و داغ هم شده.

**یادگیری ماشینی** به شاخه‌ای از هوش مصنوعی اطلاق می‌شود که در آن کامپیوترها از داده‌ها یاد می‌گیرند و بر اساس این یادگیری‌ها، پیش‌بینی‌ها و تصمیم‌گیری‌ها را بهبود می‌بخشند. به عبارت دیگر، الگوریتم‌های یادگیری ماشینی به کامپیوترها این امکان را می‌دهند که الگوها و روابط پیچیده را از داده‌های بزرگ استخراج کنند و بدون برنامه‌ریزی دقیق انسانی، عملکرد خود را ارتقا دهند.

در علوم اجتماعی محاسباتی، یادگیری ماشینی کاربردهای گسترده‌ای دارد. از جمله این کاربردها می‌توان به تحلیل رفتار اجتماعی، پیش‌بینی روندهای جمعیتی، مطالعه شبکه‌های اجتماعی، تحلیل محتوای متنی و حتی بررسی تأثیر سیاست‌های عمومی اشاره کرد. برای مثال، با استفاده از الگوریتم‌های یادگیری ماشینی می‌توان تحلیل‌های دقیقی از محتوای رسانه‌های اجتماعی انجام داد و احساسات عمومی نسبت به یک موضوع خاص را پیش‌بینی کرد. یا مثلاً می‌توان الگوهای مهاجرت را شناسایی و پیش‌بینی کرد که چگونه تغییرات اقتصادی یا سیاسی ممکن است بر رفتارهای جمعیتی تأثیر بگذارد. یادگیری ماشینی با بهره‌گیری از قدرت محاسباتی بالا و داده‌های گسترده، ابزارهای قدرتمندی برای پژوهشگران علوم اجتماعی فراهم می‌آورد تا به تحلیل‌های دقیق‌تر و جامع‌تری دست یابند.

یا مثلاً با تحلیل و پیش‌بینی سری‌های زمانی که یکی از رایج‌ترین تکنیک‌های کمی است و توسط کسب‌وکارها پیش‌تر استفاده می‌شده. امروزه برای شناسایی روندها، پاک کردن داده‌ها و حتی برای پیش‌بینی آینده استفاده می‌شوند. مثلاً توی انگلیس با بررسی آمار جرم و جنایت ضبط‌شده پلیس لندن، برای پیش‌بینی وقوع جرم در محلات مختلف انجام شده

**مهارت‌هایی که به عنوان پژوهشگر علوم اجتماعی محاسباتی باید یاد بگیرید.**

البته این به معنای یاد گرفتن ۱۰۰ درصدی همه‌ی این مهارت‌ها و از این حجم ترسیدن نیست.

- توانایی استفاده از روش‌های آماری برای تجزیه و تحلیل داده‌ها و استخراج نتایج معنی‌دار از آن‌ها.
- 2. **برنامه‌نویسی:**
  - تسلط به زبان‌های برنامه‌نویسی مانند Python و R برای پردازش داده‌ها و اجرای تحلیل‌های پیچیده.
- 3. **تحلیل شبکه‌های اجتماعی:**
  - توانایی تحلیل ساختار و دینامیک شبکه‌های اجتماعی و شناسایی الگوها و ارتباطات میان افراد.
- 4. **جمع‌آوری داده‌ها:**
  - مهارت در استفاده از روش‌های مختلف جمع‌آوری داده‌ها مانند پرسشنامه‌ها، مصاحبه‌ها، و کاوش داده‌های آنلاین.
- 5. **روش‌های کیفی:**
  - توانایی انجام تحلیل‌های کیفی مانند تحلیل محتوا، تحلیل گفتمان و تحلیل تماتیک برای درک عمیق از داده‌های غیرعددی.
- 6. **روش‌های ترکیبی:**
  - توانایی ترکیب روش‌های کمی و کیفی برای ایجاد تصویری جامع و کامل از پدیده‌های اجتماعی.
- 7. **بصری‌سازی داده‌ها:**
  - مهارت در استفاده از ابزارهای بصری‌سازی داده‌ها مانند Tableau و D3.js برای ایجاد نمودارها و گراف‌های قابل فهم.
- 8. **مدل‌سازی اجتماعی:**
  - توانایی ساخت مدل‌های محاسباتی برای شبیه‌سازی و پیش‌بینی رفتارهای اجتماعی.
- 9. **تحلیل محتوای دیجیتال:**
  - مهارت در استخراج و تحلیل داده‌های متنی از منابع دیجیتال مانند رسانه‌های اجتماعی، وبسایت‌ها و اسناد دیجیتال.
- 10. **اخلاق در پژوهش:**
  - درک اصول اخلاقی در پژوهش و رعایت آن‌ها در جمع‌آوری، تحلیل و انتشار داده‌ها.
- 11. **کار با پایگاه داده‌ها:**
  - توانایی مدیریت و کار با پایگاه داده‌های بزرگ و پیچیده مانند SQL و NoSQL برای ذخیره‌سازی و بازیابی داده‌ها.
- 12. **نرم‌افزارهای تحلیل کیفی:**
  - توانایی استفاده از نرم‌افزارهای تحلیل کیفی مانند NVivo و MAXQDA برای مدیریت و تحلیل داده‌های متنی و چندرسانه‌ای.
- 13. **تفسیر نتایج:**

- توانایی تفسیر نتایج به دست آمده از تحلیل داده‌ها و ارائه آن‌ها به صورت گزارشی واضح و قابل فهم.

#### 14. پژوهش‌های بین‌رشته‌ای:

- توانایی کار در پروژه‌های بین‌رشته‌ای و همکاری با متخصصان در حوزه‌های مختلف برای بهره‌گیری از دیدگاه‌های متعدد.

#### مثال پروژه‌هایی که یک پژوهشگر علوم اجتماعی محاسباتی انجام می‌دهد:

۱- تحلیل احساسات و محتوا در شبکه توییتر

۲- داده‌کاوی و بررسی میلیون‌ها مقاله خبری و گزارش‌های نشریات برای مطالعه نگرش‌های سیاسی و تغییرات آن

داشمند داده و پژوهشگر علوم اجتماعی محاسباتی در عمل بسیار شبیه به هم هستند و مرز باریکی بین‌شون هست که در جدول زیر به آن پرداخته شده و البته که این جدول را با نگاه تخصص یا رویکرد اولیه به مسائل ببینید چرا که در عمل، حوزه‌های تحلیل معمولا ابزارهای یکسانی را می‌طلبند.

پژوهشگر علوم اجتماعی محاسباتی	داشمند داده	
جامعه‌شناسی، روانشناسی، علوم سیاسی، انسان‌شناسی	علوم کامپیوتر، ریاضیات، آمار، مهندسی	زمینه تخصص اولیه
استفاده از روش‌های کیفی و کمی برای تفسیر داده‌ها	تمرکز بر روش‌های کمی و تکنیک‌های پیچیده آماری و یادگیری ماشین	رویکرد به تحلیل داده در گام نخست
درک و تبیین پدیده‌های اجتماعی، فرهنگی و سیاسی	بهینه‌سازی فرآیندها، توسعه مدل‌های پیش‌بینی‌کننده، استخراج بینش‌های عملی برای تصمیم‌گیری‌ها	هدف و کاربرد
مصاحبه، پرسشنامه، تحلیل محتوا، روش‌های تاریخی و ...	روش‌های تجربی، آزمون فرضیات، طراحی آزمایش‌ها و مدل‌های ریاضی و آماری	روش‌شناسی پژوهش

دیدگاه نسبت به داده‌ها	داده‌ها را به عنوان منابع خام برای تحلیل و استخراج الگوهای پنهان می‌بیند	داده‌ها را به عنوان منابعی برای درک رفتارها و تعاملات انسانی می‌بیند
تحلیل جوامع انسانی	کمتر با نظریه‌های علوم انسانی درگیر است و بیشتر به تحلیل فنی داده‌ها می‌پردازد.	اساساً نقطه شروع برای او نظریه‌های اجتماعی است و از آن دریچه برای تفسیر داده‌ها استفاده می‌کند

## مسیرهای حرفه‌ای تحلیل داده

مثال‌هایی از شغل‌هایی که از تحلیل داده می‌توان به آن رسید.

### 1. تحلیل داده فروش و بازاریابی (Sales and Marketing Data Analyst):

- مراکز فعالیت: شرکت‌های تجاری، استارت‌آپ‌ها، شرکت‌های تبلیغاتی
- شرح وظیفه: تحلیل داده‌های فروش و بازاریابی برای بهبود استراتژی‌ها، شناسایی الگوهای مصرف‌کننده، ارائه گزارش‌ها و داشبوردهای عملکرد، همکاری با تیم بازاریابی
- ابزارها: Excel، Tableau، Power BI، SQL، Python (pandas, numpy)
- مثال‌هایی از وظایف: تحلیل کمپین‌های بازاریابی، پیش‌بینی تقاضای محصولات، شناسایی فرصت‌های فروش

### 2. تحلیل داده منابع انسانی (HR Data Analyst):

- مراکز فعالیت: شرکت‌های بزرگ، استارت‌آپ‌ها، شرکت‌های مشاوره منابع انسانی
- شرح وظیفه: تحلیل داده‌های منابع انسانی برای بهبود فرآیندها، شناسایی روندها و الگوهای عملکرد کارکنان، ارائه گزارش‌های تحلیلی به مدیریت
- ابزارها: Excel، Power BI، SQL، Python، R
- مثال‌هایی از وظایف: تحلیل داده‌های جذب و استخدام، شناسایی نیازهای آموزشی کارکنان، بررسی نرخ ترک کار

### 3. تحلیل داده زنجیره تامین (Supply Chain Data Analyst):

- مراکز فعالیت: شرکت‌های تولیدی، شرکت‌های لجستیک، خرده‌فروشی‌ها
- شرح وظیفه: تحلیل داده‌های زنجیره تامین برای بهینه‌سازی فرآیندها، شناسایی نقاط ضعف و بهبود، ارائه گزارش‌ها و داشبوردهای عملکرد
- ابزارها: Excel، Tableau، Power BI، SQL، Python (pandas, numpy)

- مثال‌هایی از وظایف: تحلیل جریان کالا، پیش‌بینی تقاضا، بهینه‌سازی موجودی انبار

#### 4. تحلیل داده مالی (Financial Data Analyst):

- مراکز فعالیت: بانک‌ها و مؤسسات مالی، شرکت‌های سرمایه‌گذاری، شرکت‌های بیمه
- شرح وظیفه: تحلیل داده‌های مالی برای ارائه پیشنهادات بهبود، شناسایی روندها و الگوهای مالی، ارائه گزارش‌های مالی به مدیریت
- ابزارها: Excel، R، Python، SAS، SQL
- مثال‌هایی از وظایف: تحلیل داده‌های مالی برای شناسایی نقاط قوت و ضعف، پیش‌بینی روندهای مالی، ارزیابی ریسک‌های مالی

#### 5. تحلیل داده مشتری (Customer Data Analyst):

- مراکز فعالیت: شرکت‌های خدماتی، استارت‌آپ‌ها، شرکت‌های فناوری اطلاعات
- شرح وظیفه: تحلیل داده‌های مشتریان برای بهبود خدمات، شناسایی الگوهای مصرف، ارائه گزارش‌ها و داشبوردهای تحلیلی
- ابزارها: Excel، Tableau، Power BI، SQL، Python (pandas, numpy)
- مثال‌هایی از وظایف: تحلیل نظرسنجی‌های مشتریان، شناسایی نیازهای مشتریان، ارائه پیشنهادات بهبود خدمات

#### 6. تحلیل داده محصولات (Product Data Analyst):

- مراکز فعالیت: شرکت‌های تولیدی، استارت‌آپ‌ها، شرکت‌های فناوری
- شرح وظیفه: تحلیل داده‌های محصولات برای بهبود کیفیت و عملکرد، شناسایی الگوهای مصرف، ارائه گزارش‌ها و داشبوردهای تحلیلی
- ابزارها: Excel، Tableau، Power BI، SQL، Python (pandas, numpy)
- مثال‌هایی از وظایف: تحلیل داده‌های استفاده از محصول، شناسایی نقاط ضعف محصولات، پیش‌بینی تقاضای محصولات

#### 7. تحلیل داده در سازمان‌های دولتی (Government Data Analyst):

- مراکز فعالیت: سازمان‌های دولتی، مراکز آماری دولتی، نهادهای بین‌المللی
- شرح وظیفه: جمع‌آوری و تحلیل داده‌های دولتی، ارائه گزارش‌های آماری برای سیاست‌گذاران، شناسایی الگوها و روندهای اجتماعی و اقتصادی
- ابزارها: Excel، R، Python، SPSS، SAS



- مثال‌هایی از وظایف: تحلیل داده‌های بیکاری، بررسی داده‌های بهداشتی، ارائه گزارش‌های آماری برای برنامه‌ریزی‌های دولتی

#### **8. تحلیل داده‌های بهداشتی (Healthcare Data Analyst):**

- مراکز فعالیت: بیمارستان‌ها، مراکز بهداشتی، شرکت‌های بیمه سلامت  
- شرح وظیفه: تحلیل داده‌های بهداشتی برای بهبود خدمات پزشکی، شناسایی الگوهای بیماری، ارائه گزارش‌های تحلیلی به مدیریت  
- ابزارها: Excel، R، Python، SAS، SQL  
- مثال‌هایی از وظایف: تحلیل داده‌های بیماران، شناسایی روندهای بیماری، ارائه پیشنهادات بهبود خدمات بهداشتی

#### **9. تحلیل داده‌های آموزشی (Educational Data Analyst):**

- مراکز فعالیت: دانشگاه‌ها، مدارس، مؤسسات آموزشی  
- شرح وظیفه: تحلیل داده‌های آموزشی برای بهبود فرآیندهای آموزش، شناسایی الگوهای عملکرد دانش‌آموزان، ارائه گزارش‌های تحلیلی به مدیریت  
- ابزارها: Excel، Tableau، Power BI، SQL، Python (pandas, numpy)  
- مثال‌هایی از وظایف: تحلیل داده‌های نمرات دانش‌آموزان، شناسایی نیازهای آموزشی، ارائه پیشنهادات بهبود فرآیندهای آموزشی

#### **10. پژوهشگر علوم اجتماعی محاسباتی (Computational Social Scientist):**

- مراکز فعالیت: دانشگاه‌ها و مراکز تحقیقاتی، سازمان‌های غیردولتی، مؤسسات آماری  
- شرح وظیفه: تحلیل داده‌های اجتماعی برای شناسایی الگوها و روندها، استفاده از مدل‌های محاسباتی برای تحلیل رفتارهای اجتماعی، ارائه گزارش‌ها و نتایج تحقیقات به سازمان‌های مختلف  
- ابزارها: R، Python، SPSS، Stata  
- مثال‌هایی از وظایف: تحلیل داده‌های نظرسنجی، استفاده از شبکه‌های اجتماعی برای تحلیل روندهای اجتماعی، بررسی تأثیرات سیاست‌های اجتماعی بر جوامع مختلف

## چگونه مهارت‌های خودمون رو بهبود بدیم؟

اول که این مواردی که میگم کلا تجربه شخصی هست و چیزهایی که اینور و اونور خوندم و شما هم باید راه خودتون رو پیدا کنید.

### ۱- اهل چالش باشید:

برای منی که اگر چه سابقه کار فنی و برنامه‌نویسی و موارد این فرمی داشتم اما همچنان تاثیری که از فضای علوم اجتماعی گرفتم، بیشتر از اون سابقه قبلی بود و هر مواجهه‌م با ابزارها و چالش‌های دیتاآنالیز یه شوک سنگین داشتم ((( خصوصا اون بخش‌هایی که به آمار و عدد ربط داره یا حتی مفاهیم عمومی آماری مثل واریانس و رگرسیون و ... اما در نهایت فهمیدم که اینقدرام سخت نیست. سخت هست، اما اینقدری سخت نیست که نشه یاد گرفت! و اینم یادمون باشه که در دنیای امروز و پیشرفت روزانه همه حوزه‌ها، کنجکاو باشید و اخبار دیتا و موضوعات مورد نظرتون رو دنبال کنید و سعی کنید همگام باهاش جلو برید.

برنامه‌نویسی هم اگر چه در علوم اجتماعی محاسباتی و موضوعاتی مثل داده‌کاوی نیازش احساس میشه، اما برای شروع می‌تونید با نرم‌افزارهای Codeless یا همون بدون نیاز به کدنویسی کار کنید یا مثلا برای تحلیل داده‌های توییت‌ر و دیگر شبکه‌های اجتماعی، می‌تونید داده‌ها را از شرکت‌هایی که کارشون جمع‌آوری داده هست، بخرید.

### ۲- دیدگاه‌های مخالف با خودتون رو در نظر بگیرید.

فارغ از داده و تحلیل داده، درگیر شدن به دیدگاه‌های مخالف نگاهتون به شما کمک می‌کنه که دیدگاه خودتون رو بسط بدید و در تحلیل و پژوهش‌هاتون با تعصب‌ها مبارزه کنید چرا که در زمان تحلیل داده‌ها مهم هست که تمام تفاسیر ممکن را در نظر بگیرید و تلاشتون رو بکنید که توی یه بن‌بست فکری نیفتید. اگر چه تعصب، سوپه‌ای سیاسی با خودش حمل می‌کنه اما عمومی صحبت می‌کنم. مثلا ممکنه توی ذهن‌تون این باشه که فلان موضوع بر فلان چیز تاثیر مستقیم می‌گذاره و روند جمع‌آوری داده‌تون رو هم به همین سمت می‌برید اما در واقعیت همبستگی خیلی ضعیفی ممکنه باشه.

### ۳- پروژه‌های انجام شده رو دنبال کنید و بخونید.

برای خودم که این مورد خیلی راه‌گشا بوده همیشه. سایت‌هایی مثل مجله دقیقه که زحمت انتشار مقاله‌ها و گزارش‌های تحلیلی داده‌محور می‌کنند یا در سطوح بالاتر پروژه‌هایی که در ایران یا کشورهای دیگه انجام شده، یه کلاس درس با جزییات عالیه. پس از دستش ندید.

### ۴- کامیونیتی و گروه‌های مشابه خودتون رو پیدا کنید

در مسیر یادگیری داده و تحلیل داده، به کامیونیتی هم‌مسیر می‌تونه نقش به نیروی محرک و حامی براتون داشته باشه. شبکه‌های مجازی مثل لینکدین و توئیتر، فضاهایی مثل مدرسه دقیقه که باعث میشه در فضایی قرار بگیرید که دوستانی با دغدغه شما اونجا هستند یا مثلا نوشتن توی رسانه‌های اختصاصی داده‌محور مثل مجله دقیقه شما را به حلقه‌ای با دغدغه خودتون وصل می‌کنه.

#### ۵- با هوش مصنوعی دوست باشید!

امروزه تحلیل داده خیلی راحت شده و یه روبات هوش مصنوعی می‌تونه کار رو براتون در بیاره اما یادتون باشه که به هوش مصنوعی نه به عنوان راه‌حل، بلکه صرفا به عنوان یک ابزار نگاه کنید. چرا که در بلندمدت و حتی کوتاه‌مدت تبدیل به کسی می‌شوید که نسبت به چند ماه/ سال قبل می‌بینید همه چیز رو فراموش کردید و از قافله عقب افتادید.

#### منابع آموزشی رایگان

۱- [مجله پرانتز - مدرسه پردازش و تحلیل داده دقیقه \(d-learn.ir\)](http://d-learn.ir)

۲- یوتیوب

۳- دوره‌های رایگان کورسرا و یودمی

## آخر قصه

این جزوه و وبینار صرفا بهانه‌ای بود که تصویری از داده و تحلیل داده در کسب و کار و کاربرد آن در علوم اجتماعی محاسباتی برای مخاطبان روشن شود تا تصویری از دنیای پیش روی خود داشته باشند و بتوانند انتخاب صحیحی برای آینده کاری و تخصصی خود داشته باشند.

دلیل عدم ورود به مباحث تخصصی در این جزوه نیز همین موضوع و دایره گسترده مخاطبانی بود که اساسا تصویر درستی از این حوزه نداشتند و نیاز بود که از ابتدایی‌ترین گام‌ها مباحث را شروع کرد و نکات اختصاصی و تخصص‌های نرم افزاری را به دوره‌های دیگر سپرد.

