



مجموعه شرکت های مهندسی دانش بنیان رها

## نیازها و پیش نیازهای علم داده و استفاده کاربردی از این چرخه حیاتی

### شرکت رهاکو



## فهرست

- 3 ..... علم داده (Data Science) چیست؟
- 3 ..... علم داده چگونه کار می کند؟
- 4 ..... چرخه حیات علم داده
- 4 ..... چه کسانی بر Data Science نظارت دارند؟
- 5 ..... پیش نیازهای علم داده
- 6 ..... مقایسه علم داده و دانشمندان داده
- 6 ..... علم داده و محاسبات ابری
- 6 نتیجه گیری



علم داده یک حوزه بین رشته‌ای است که از روش‌های علمی، فرآیندها و الگوریتم‌ها برای استخراج دانش و بینش از داده‌های ساختار یافته و بدون ساختار استفاده می‌کند. Data Science به داده کاوی، یادگیری ماشین و کلان داده مرتبط است. امروزه با توجه به حجم انبوه داده‌ای که تولید می‌شود، علم داده بخش ضروری بسیاری از صنایع می‌باشد که یکی از موضوعات داغ در مباحث فناوری اطلاعات بوده که محبوبیت آن در طی سال‌ها افزایش یافته است. امروزه شرکت‌ها شروع به پیاده سازی تکنیک‌های Data Science برای رشد کسب و کار خود و افزایش رضایت مشتری کرده‌اند .

Data Science حوزه مطالعاتی است که تخصص‌هایی از جمله برنامه نویسی، دانش ریاضیات و آمار را برای استخراج بینش معنادار از داده‌ها ترکیب می‌کند. متخصصان علوم داده الگوریتم‌های یادگیری ماشینی را روی اعداد، متن، تصاویر، ویدئو، صدا و موارد دیگر برای تولید سیستم‌های هوش مصنوعی (AI) و انجام کارهایی که معمولاً به هوش انسانی نیاز دارند، اعمال می‌کنند Data Science. به نوبه خود، بینشی ایجاد می‌کند که تحلیلگران و کاربران تجاری می‌توانند آن را به ارزش تجاری ملموس تبدیل کنند.

## علم داده (Data Science) چیست؟

Data Science با حجم وسیعی از داده‌ها با استفاده از ابزارها و تکنیک‌های مدرن برای یافتن الگوهای کاربردی، استخراج اطلاعات معنادار و تصمیم‌گیری‌های تجاری سر و کار دارد. این تکنولوژی از الگوریتم‌های پیچیده یادگیری ماشین برای ساخت مدل‌های پیش بینی شده استفاده می‌کند. داده‌های مورد استفاده برای تجزیه و تحلیل می‌توانند از منابع مختلف و در قالب‌های مختلف ارائه شوند Data Science. یک نیروی محرکه بین تجربیات بسیار تخصصی کاربران است که از طریق شخصی سازی و سفارشی سازی ایجاد می‌شود. این تجزیه و تحلیل می‌تواند برای ایجاد حس دیده شدن و درک مشتریان توسط یک شرکت مورد استفاده قرار گیرد.

## علم داده چگونه کار می‌کند؟

Data Science شامل چندین رشته برای ایجاد نگاهی جامع، کامل و تصفیه شده به داده‌های خام است. دانشمندان داده باید در همه چیز از مهندسی داده، ریاضی، آمار، محاسبات پیشرفته گرفته تا تجسم ماهر باشند که بتوانند به طور موثر توده‌های درهم از اطلاعات را غربال کنند. دانشمندان داده اغلب برای ایجاد مدل‌ها با استفاده از الگوریتم‌ها و تکنیک‌های دیگر، به شدت به هوش مصنوعی، به ویژه زیر شاخه‌های یادگیری ماشینی و یادگیری عمیق، تکیه می‌کنند.



## چرخه حیات علم داده

چرخه حیات شامل پنج مرحله مجزا است که هر کدام وظایف خاص خود را دارند:

1. **ضبط:** این مراحل شامل جمع آوری داده های ساختار یافته و بدون ساختار است. مانند: اکتساب داده، دریافت سیگنال و استخراج داده.
2. **نگهداری:** این مرحله شامل گرفتن داده های خام و قرار دادن آن ها به شکلی است که قابل استفاده باشد مانند: انبار داده، پاکسازی داده، مرحله بندی داده، پردازش داده و معماری داده.
3. **فرآیند:** دانشمندان داده، اطلاعات آماده شده را دریافت می کنند و الگوهای آن را بررسی می کنند تا تعیین کنند که چقدر در تحلیل پیش بینی مفید خواهد بود. داده کاوی، طبقه بندی، مدل سازی داده ها
4. **تجزیه و تحلیل:** این مرحله شامل انجام تحلیل های مختلف بر روی داده ها می باشد. مانند: رگرسیون، متن کاوی، تحلیل کیفی.
5. **ارتباط:** در قسمت نهایی، تحلیلگران تجزیه و تحلیل ها را به شکل های قابل خواندن مانند نمودارها آماده می کنند.

## چه کسانی بر Data Science نظارت دارند؟

### مدیران کسب و کار

مدیران کسب و کار افرادی هستند که وظیفه نظارت بر روش آموزش علوم داده را بر عهده دارند. مسئولیت اصلی آن ها همکاری با تیم Data Science برای مشخص کردن مشکل و ایجاد یک روش تحلیلی است. یک [دانشمند داده](#) ممکن است بر بخش بازاریابی، مالی یا فروش نظارت داشته باشد و به مدیر اجرایی مسئول بخش گزارش دهد. هدف آن ها اطمینان از تکمیل به موقع پروژه ها با همکاری نزدیک با دانشمندان داده و مدیران فناوری اطلاعات است.

### مدیران فناوری اطلاعات

در رتبه دوم مدیران فناوری اطلاعات هستند. اگر عضوی برای مدت طولانی در سازمان بوده باشد، بدون شک مسئولیت های او بیش از سایرین مهم خواهد بود. آن ها در درجه اول مسئول توسعه زیرساخت ها و معماری برای فعال کردن فعالیت های علم داده هستند. تیم های Data Science نظارت می شوند و منابع مورد نیاز تیم تامین می شود تا اطمینان حاصل شود که کارآمد و ایمن عمل می کنند. آن ها همچنین ممکن است مسئول ایجاد و نگهداری محیط های IT برای تیم های تحلیل داده باشند.



## مدیران علوم داده

مدیران علوم داده بخش نهایی کار را تشکیل می دهند. آن ها در درجه اول بر روند کاری همه اعضای تیم Data Science نظارت دارند. که همچنین فعالیت های روزانه تیم علم داده را مدیریت و پیگیری می کنند. مدیران علوم داده تیم سازانی هستند که می توانند برنامه ریزی و نظارت پروژه را با رشد تیم ترکیب کنند.

## پیش نیازهای علم داده

در اینجا برخی از مفاهیم فنی وجود دارد که باید قبل از شروع یادگیری Data Science در مورد آن ها بدانید.

### 1. یادگیری ماشینی

یادگیری ماشینی ستون فقرات Data Science است. دانشمندان داده علاوه بر دانش اولیه آمار، باید درک کاملی از ML داشته باشند.

### 2. مدل سازی

مدل های ریاضی شما را قادر می سازد تا محاسبات و پیش بینی های سریعی را بر اساس آنچه از قبل درباره داده ها می دانید انجام دهید. مدل سازی نیز بخشی از یادگیری ماشینی است و شامل شناسایی این موضوع است که کدام الگوریتم برای حل یک مسئله معین مناسب تر است و چگونه این مدل ها را آموزش دهیم.

### 3. آمار

آمار در هسته Data Science قرار دارد. یک آمار دقیق و درست می تواند به شما کمک کند تا داده بیشتری را استخراج کرده و نتایج معنی داره بیشتری کسب کنید.

### 4. برنامه نویسی

برای اجرای موفقیت آمیز پروژه Data Science به سطحی از برنامه نویسی نیاز است. رایج ترین زبان های برنامه نویسی پایتون و R هستند. R و Python از محبوبیت خاصی برخوردار هستند زیرا یادگیری آن ها آسان است و از چندین کتابخانه برای Data Science و ML پشتیبانی می کنند.

### 5. پایگاه های داده

یک دانشمند داده توانمند باید بداند که پایگاه های داده چگونه کار می کنند، چگونه آن ها را مدیریت کند و چگونه داده ها را از آن ها استخراج کند.



## مقایسه علم داده و دانشمندان داده

Data Science به عنوان یک رشته در نظر گرفته می شود، در حالی که دانشمندان داده متخصصان آن حوزه هستند. دانشمندان داده لزوماً مسائل مستقیم تمام فرآیندهای درگیر در چرخه حیات علم داده نیستند. به عنوان مثال: خطوط لوله داده معمولاً توسط مهندسان داده اداره می شود اما دانشمندان داده ممکن است توصیه هایی در مورد نوع داده ارائه دهند. در حالی که دانشمندان داده می توانند مدل های یادگیری ماشین بسازند، که این تلاش ها در سطح بزرگ تر به مهارت های مهندسی نرم افزار بیشتر و بهینه سازی یک برنامه برای اجرای سریع تر نیاز دارد.

## علم داده و محاسبات ابری

راه حل های ذخیره سازی ابری، مانند دریاچه های داده و دسترسی به زیر ساخت های ذخیره سازی را فراهم می کنند که می توانند حجم زیادی از داده ها را به راحتی دریافت و پردازش کنند. این سیستم ها ذخیره سازی و انعطاف پذیری را برای کاربران نهایی فراهم می کنند. پلتفرم های ابری معمولاً برای اشتراک های مورد نظر کاربر، مدل های قیمت گذاری متفاوتی دارند، تا بتواند به راحتی نیازهای کاربر را برآورده کنند.

از فناوری های open source به طور گسترده در مجموعه ابزارهای Data Science استفاده می شود. چندین ارائه دهنده ابر، از جمله IBM Cloud، کیت های ابزار از پیش بسته بندی شده ای را نیز ارائه می کنند. که دانشمندان داده را قادر می سازد تا مدل هایی را بدون کدنویسی بسازند. و دسترسی بیشتر به نوآوری های فناوری و بینش داده داشته باشند.

## نتیجه گیری

علم داده مفهومی برای یکسان سازی آمارها، تحلیل داده ها، انفورماتیک و روش های مرتبط با آن ها به منظور درک و تجزیه و تحلیل پدیده های واقعی با داده ها است Data Science. از تکنیک ها و نظریه هایی استفاده می کند که در بسیاری از زمینه ها مانند: ریاضیات، آمار، علوم کامپیوتر، علم اطلاعات استخراج شده اند.