**چطور Data Science رو شروع کنیم ؟ (Road Map CMD6)**



شغل‌های حوزه تکنولوژی با توجه به تاثیر شگرفی که در زندگی ما گذاشته، برای افراد بسیار زیادی جذاب به نظر می‌رسد، به‌خصوص برای افرادی که قصد ورود به وارد بازار کار دارند، البته Data Science به دلیل ترند بودن خواهان بیشتری دارد. اگر با دیتا و کارکردن با آن لذت می‌برید و به آمار و احتمال علاقه‌مند هستید، Data Science می‌تواند در لیست شغل‌های مورد علاقه شما قرار بگیرد اما بهتر است، ویژگی‌های فردی خودتون رو بهتر بشناسید و کمی با خودتون رو راست باشید و برای انتخاب درست‌تر، قسمت ابتدایی سری اول مجموعه مقالات [RoadMap CMD](https://dataio.ir/%DA%86%D9%87%D8%B7%D9%88%D8%B1%DB%8C-data-analysis-%D8%B1%D9%88-%D8%B4%D8%B1%D9%88%D8%B9-%DA%A9%D9%86%DB%8C%D9%85-pdbouokxzbua" \t "_blank) رو بخونید تا مثل بیشتر افراد شاغل وقتی مشغول کار هستید، عذاب نکشید. ?

این مقاله به کمک: علی شاهد حق قدم (Principal Data Scientist at Freedom Financial Network)، علی موسوی(AI Resident at Google Brain)، کیوان ابراهیمی (Senior Data Scientist at View Inc)، مجید ابوالقاسمی(Data Scientist at Digikala.com) و امیر واهب(Data Scientist & Machine Learning Engineer at SabaIdea) نوشته شده و از زحماتشون متشکرم.

**دیتا ساینس ؟ Data Science ؟ دانشمند داده ؟ ?**

در حال حاضر در ایران بیشتر فعالیت دانشمندان داده (Data Scientists) استخراج و کار با داده‌هایی است که از یک وب سایت یا اپلیکیشن استخراج می‌شود، درحالی که داده می‌تواند از سنسور‌ها، سیستم‌های صنعتی، پژوهش‌های اجتماعی یا … استخراج شود، همچنین می‌تواند عددی یا غیرعددی باشد. شرکتی که کیوان ابراهیمی مشغول به کار هست، بیشتر بروی دیتاهایی که از سیستم‌های صنعتی جمع‌آوری می‌شود، کار می‌کنند.

دیتا ساینس (**Data Science**) باعث بصیرت افزایی از داده‌ها می‌شود ?، یعنی با کشف روابط و معانی بین آن‌ها اطلاعات مفید و قابل فهمی استخراج می‌کند که در نهایت به تحلیل و تصمیم‌گیری درست منجر می‌شود.

دیتا ساینس (**Data Science**)، یک ترکیب چند رشته‌ای است، از جمله رشته‌های آمار و احتمال و آنالیز عددی، که با جمع آوری داده‌ها، پیش پردازش داده‌ها، آماده سازی داده‌‌ها، تحلیل ویژگی‌ها، ماشین لرنینگ و کار با الگوریتم‌ها، منجر به کشف روابط و معانی بین داده‌ها می‌شود.

آنچه از Data Science تصور می‌شود کار با الگوریتم و ماشین لرنینگ است درحالی که بیشتر زمان یک Data Scientist صرف جمع آوری و پردازش داده‌ها می‌شود به همین دلیل مهم است که عاشق دیتا و کار با آن باشید.

**درحال حاضر Data Science مورد تهدید هست یا بیشتر یه فرصت به‌حساب میاد ؟**

باتوجه به وفور داده‌ها، فرصت‌های شغلی بسیاری در این حوزه وجود دارد که ممکن است باعث سر در گمی شود. ممکن است پروژه‌های ترند جذاب به‌نظر برسند اما بعد ورود، از پروژه خسته و دل و دماغ کار کردن نداشته باشید.

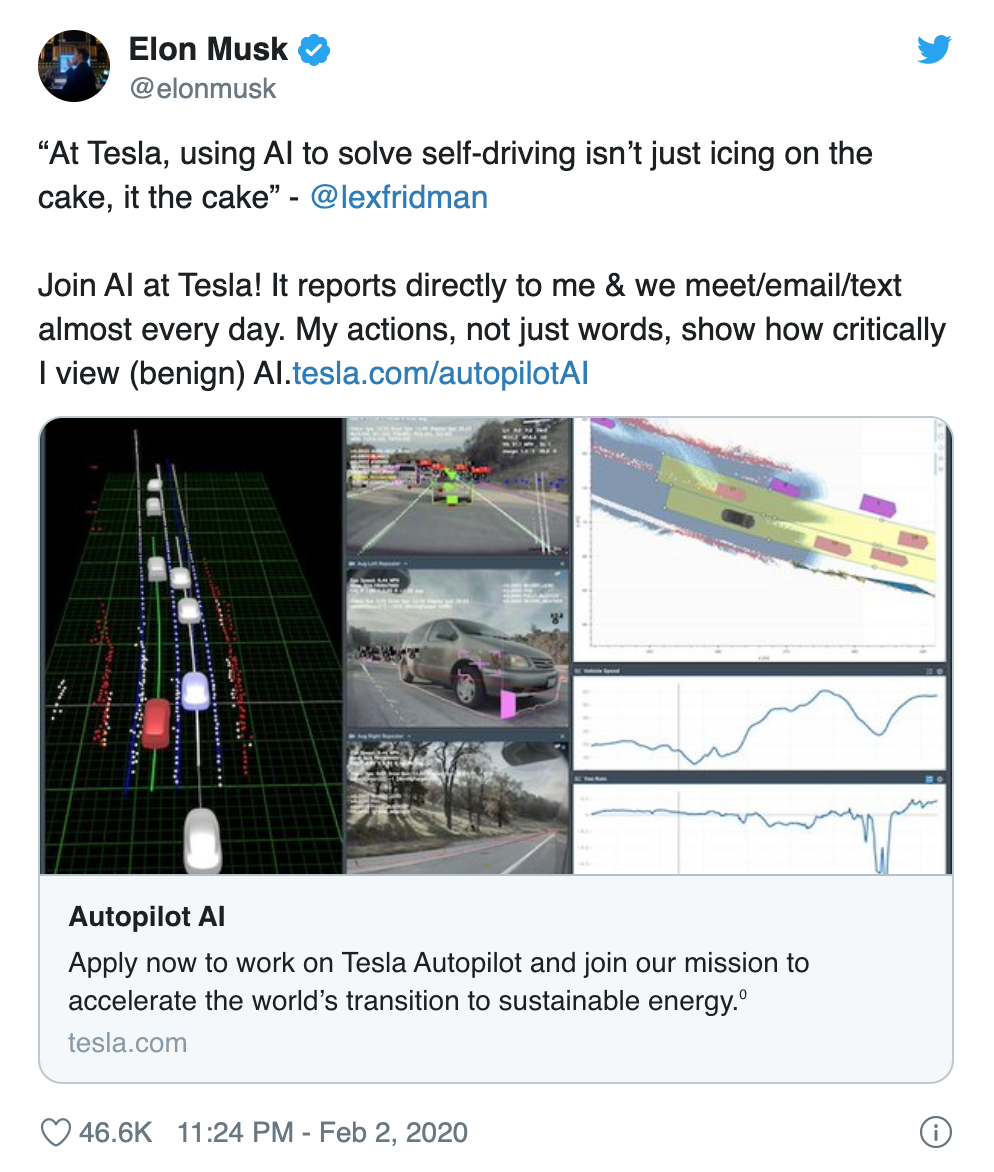
باوجود دوره‌های کوتاه مدتی که وعده متخصص شدن در کوتاه مدت را می‌دهند، انتظار شما را بالا می‌برند درحالی که Data Science از چند رشته تشکیل شده و نیاز به یادگیری پیوسته در طولانی مدت داره و صرفا به دانستن الگوریتم‌های ماشین لرنینگ و قدرت برنامه نویسی ختم نمی‌شود، به همین دلیل نیازمند مطالعه و تلاش مداوم هست.

برای هر کسب و کاری، رقبایی وجود دارد که باعث می‌شود مدیران کسب و کارها تصمیم‌هایی برپایه داده‌های دقیق بگیرند، از این رو نیاز به Data Science در کسب و کارها به شدت احساس می‌شود اما انتظارهای غیر واقعی شرکت‌ها، اعتماد به این حوزه را کم رنگ‌تر می‌کند.

دیتا ساینس (**Data Science**) یکی از موضوعات ترند جامعه تکنولوژی به‌حساب می‌آید و به سرعت درحال بروز رسانی و پیش‌روی است به‌همین دلیل نیازمند مطالعه و یادگیری مداوم است. (از مقالاتی که در کنفرانس‌ها ارائه میشه، غافل نشید)

**دانشگاه پایه‌های اصلی است ؟ ?‍??‍?**

صحبتی که درحال حاضر درباره‌ی دانشگاه و خواندن رشته‌ تحصیلی مرتبط مطرحه، به مهم نبودن مدرک تحصیلی اشاره می‌کند. این موضوع در صحبت‌های بیشتر افرادی که در دنیای تکنولوژی فعال و معروف هستند، وجود دارد و همواره به گفته‌های این افراد استناد می‌شود؛ همچنین در فیلم‌هایی که موضوع استارت‌آپ‌ و تکنولوژی در آن‌ها مطرح است.( مثل سریال سیلیکون ولی )

توییت ایلان ماسک درمورد استخدام در تیم هوش مصنوعی شرکت تسلا

<https://twitter.com/elonmusk/status/1224043531951583233>

نمی‌توان گفت مرتبط بودن رشته تحصیلی مهم نیست اما ضروری هم نیست. درصورتی که رشته تحصیلی مرتبط داشته باشید، پیش‌نیازهایی مانند آمار و احتمال، جبر خطی، تئوری تخمین و اکتشاف (Estimation and detection theory)، مهارت حل مسئله، ساختمان داده و الگوریتم، پروسه‌های اتفاقی و برنامه نویسی رو درصورتی که در دانشگاه به‌خوبی گذرانده باشید مسیر پیش‌رو، مسیر آسان‌تری هست اما درصورت مرتبط نبودن رشته تحصیلی هم می‌توانید با تلاش بیشتر کاستی‌ها را جبران کنید.

**دیتا ساینتیست چه مهارت‌هایی باید داشته باشد ؟**

یکی از مهارت‌هایی که کمتر مورد توجه است، فرایند مطرح کردن سوال و پاسخ دادن به سوال با استفاده از ابزارهاست اما مهم‌تر از این‌ها، توضیح ساده و قابل فهم پاسخ سوال به سایر افراد هست.

از جمله مهارت‌هایی که دیتا ساینتیست باید داشته باشه:

* قدرت برنامه نویسی(آشنایی با یکی از زبان‌های Python و R )
* آشنایی با الگوریتم‌ها و تئوری‌های ماشین لرنینگ
* آشنایی با پایگاه‌های داده‌
* تسلط نسبی بر آمار و احتمال
* آشنایی با ابزارهایی مانند: numpy ,pandas ,tensorflow، keras ,scikit-learn ,matplotlib
* آشنایی با ابزارهای بیگ دیتا مانند spark و hadoop
* مهارت حل مسئله
* مصور سازی داده‌ها

**ویژگی‌های فردی یک دیتا ساینتیست**

به دلیل مهم بودن خروجی داده‌ها و تاثیری که در تصمیم گیری‌های مدیران در کسب و کار دارد، نیازمند دقت فراوان، جزئی‌ نگری و داشتن یک تصویر کلی از مسئله هست. ممکنه برای بدست آوردن پاسخ یک مسئله، مدت زمان زیادی، روی یک مسئله تمرکز کنید به همین دلیل صبر و پشتکار می‌طلبه و حل اون مسئله نیاز به خلاقیت و جسارت دارد؛ همچنین پس از حل مسئله به اولین جواب نیز اکتفا نکنید. شاید هم باید عاشق دیتا باشید که در کنار اون حس کنجکاوی و اکتشاف برانگیخته شود.

**برای شروع چه مهارت‌هایی یاد بگیریم؟**

- برا یادگیری Python می‌تونید در دوره‌های [مقدماتی](https://maktabkhooneh.org/course/%D8%A2%D9%85%D9%88%D8%B2%D8%B4-%D8%A8%D8%B1%D9%86%D8%A7%D9%85%D9%87-%D9%86%D9%88%DB%8C%D8%B3%DB%8C-%D8%A8%D8%A7-%D9%BE%D8%A7%DB%8C%D8%AA%D9%88%D9%86-%D9%85%D9%82%D8%AF%D9%85%D8%A7%D8%AA%DB%8C-mk346/) و [پیشرفته](https://maktabkhooneh.org/course/%D8%A2%D9%85%D9%88%D8%B2%D8%B4-%D8%A8%D8%B1%D9%86%D8%A7%D9%85%D9%87-%D9%86%D9%88%DB%8C%D8%B3%DB%8C-%D8%A8%D8%A7-%D9%BE%D8%A7%DB%8C%D8%AA%D9%88%D9%86-%D9%BE%DB%8C%D8%B4%D8%B1%D9%81%D8%AA%D9%87-mk387/) جادی یا دوره [Coursera](https://www.coursera.org/learn/python-for-applied-data-science-ai" \l "syllabus" \t "_blank), Udemy یا [edx](https://www.edx.org/course/learn-program-using-python-utarlingtonx-cse1309x" \t "_blank) شرکت کنید. همچنین این مقاله از [جادی](https://jadi.net/2017/02/first-steps-python/" \t "_blank)، برای شروع Python هم مطالعه کنید.

دوره‌ها و بوت کمپ‌های متعددی برای یادگیری مهارت‌های دیتا ساینس وجود داره که در بین آن‌ها می‌توان به این بوت کمپ‌ها اشاره کرد.

[Kaggle](https://www.kaggle.com/learn/overview)

[Coursera](https://www.coursera.org/browse/data-science)

[Udemy](https://www.udemy.com/course/the-data-science-course-complete-data-science-bootcamp/)

دوره‌های آنلاین دانشگاه‌های [Stanford](https://www.youtube.com/user/stanfordonline" \t "_blank) و [MIT](https://www.youtube.com/user/MIT" \t "_blank) در یوتیوب هم مفید هستند.

اگر امکان پرداخت هزینه‌ی دوره‌های [Kaggle](https://www.kaggle.com/learn/overview" \t "_blank) براتون وجود داره، حتما شرکت کنید.

**همواره مطالعه کنید و بروز باشید.**

- مقالات [arxiv](https://arxiv.org/" \t "_blank) منبع خوبی هست اما توجه کنید که هنوز این مقالات ریویو نشدن.

- مقالاتی که در کنفرانس‌های معتبر ارائه میشه می‌تونه کمک کنه مثل کنفرانس [ICEFR](http://www.icefr.org/" \t "_blank) و [ICML](https://icml.cc/2015/" \t "_blank).

- مقالاتی که در [Medium](https://medium.com/topic/data-science" \t "_blank) منتشر میشه.

- مقالات [towards](https://towardsdatascience.com/data-science/home" \t "_blank).

- مقالات [analyticsvidhya](https://www.analyticsvidhya.com/" \t "_blank).

- همچنین می‌تونید از سرویس scholar.google.com استفاده کنید تا مقالات بروز براتون ارسال بشه.

در این [کانال](https://www.youtube.com/user/Zan560/featured) یوتیوب هم چندتا از دوره‌های دانشگاه MIT منتشر میشه.

کانال یوتیوب [Deep Mind](https://www.youtube.com/channel/UCP7jMXSY2xbc3KCAE0MHQ-A" \t "_blank) هم که از شرکت‌های زیر مجموعه گوگل هست، ویدئوهای آموزشی منتشر میکنه.

سایت [analyticsvidhya](https://www.analyticsvidhya.com/blog/2019/01/learning-path-data-scientist-machine-learning-2019/) یک مقاله درباره‌ی مسیر یادگیری Data Science در سال ۲۰۱۹ منتشر کرده و درباره‌ی مهارت‌هایی که باید یادبگیرید توضیح داده اما Road map گرافیکی که در کنار این مقاله منتشر شده، مسیر یادگیری رو پله پله ترسیم کرده و خیلی مفید هست.

در این [Repository](https://github.com/academic/awesome-datascience) توضیح کامل و خیلی جامعی در مورد Data Science داده شده، از پادکست، کتاب، اکانت توییتر گرفته تا کانال تلگرام و یوتیوب و مجله، اما بهتره تصمیم نهایی‌تون رو بگیرید و شروع کنید و پله پله پیشرفت کنید. فراموش نکنید که از انجام پروژه غافل نشید و خیلی هم غرق در آموزش و ویدئو دیدن نشید، چون بخشی از یادگیری از انجام پروژه حاصل میشه.

منبع: <https://dataio.ir/data-science-x4zamlhmltae>