# Machine Learning Jobs for beginners/freshers : My Perspective from various Quora Answers

I will collect some of my answers here regarding fresher Machine Learrning jobs in India as I have some experience in this domain and hire in this domain quite often

### Q. What should I do to get a job in Machine Learning in India ?

It’s a simple rule, work on an interesting project, get some results and put on display (Google, kaggle etc). Most employers can gauge your interest when they look at your profile. An entry level Machine Learning Engineer who can explain the Maths behind the algorithms they apply in some non-trivial problem is good enough for a lot of companies, they are not expecting seasoned veterans.

The following types of projects look very impressive:

1. Taking a research/kaggle competition and working on its dataset. The code to handle such problems is a good proof of ones capability. The other good way is contributions to open source ML. The more you know what works behind your solution, the better.
2. Taking a paper (or taking a set of papers by a research group) and trying to code the algorithm/replicate the results is even better than 1 (at least for us at ParallelDots).
3. An academic/industrial publication is even better.

The following is what is not enough:

1. Forking code for Bicycle challenge or other such competitions (the Titanic one as well) and running it and submitting that as previous work. They are too many tutorials for them and people won't take you seriously.
2. Completing Andrew Ng course.
3. Unless you are applying in a very big enterprise, I will say avoid certification and try working on personal projects. That is way more impressive.
4. Again, Please don’t just fork code on Github and expect employers to believe you have worked. Its easy to see how much you have worked. :(

HTH.

Answer is here: https://www.quora.com/What-should-I-do-next-for-getting-a-job-in-Data-Science-and-Machine-Learning-in-India/answer/Muktabh-Mayank?srid=qie

### Q. Is it not possible for a fresher to work as a Machine Learning Engineer ?

It is not "not possible" for a fresher to join as machine learning engineer if he she is very good at it. Wrong conclusion.

I have seen freshers taking role of machine learning engineer at enterprises and even my startup. The problem is "he she is very good at it" part . Freshers often (not always, I know freshers with CVs as good as a professional ML engineer) over estimate their CVs as they don't know about how real world works , they have very high expectations due to that and real world is unable to meet that.

Answer is here: https://www.quora.com/Why-is-it-almost-impossible-for-a-fresher-to-join-as-a-ML-engineer-even-though-he-is-very-good-at-it/answer/Muktabh-Mayank?srid=qie

### Q. What is the learning path after going through basic Machine Leaning material ?

Does one do projects, go for a summer training in Hadoop in Delhi, read books ?

(Please see link to blog posts below to find all these books)

IMHO, projects are the best way to learn (as you said you want to become a factory data scientist and are not in a mood to pursue a higher degree ). Taking different datasets for different types of problems and trying to come up with a data processing pipeline and then applying Machine Learning algorithms on top of it helped to clear many concepts for me. Its way more fun than all theory. Look at UCI or Kaggle datasets to apply your skills.

1. From where should I do summer training in big data hadoop(Delhi)? (or in any other specific technology)

Dont. Hadoop is just a framework to write programs. If from starting years of your career, you start concentrating on technologies (hadoop / mysql / postgre / java / python / scala and stuff like that), you will end up becoming a <specific technology> professional (say hadoop professional). Technology trends come and go, unless the fundamentals (which here is distributed computing) behind them is clear, one would never be able to adapt to the changing market . As long as you know programming basics well, stuff like Hadoop would be easy to pick up when your job requires it. You can try triggering a mapreduce job on your local system with help of an online tutorial, that is a different thing, but taking a full on course on hadoop is neither required nor recommended.

There are some other points I would like to specify as well:

a) Hadoop is for handling very large amount of data (Terabyte+), which most places do not have. Unlike what the internet hype (and plethora of hadoop training institutes in India want you to believe), most companies still have small data which can fit easily on a postgres cluster if not one server. Data Analysis skill is required almost everywhere, Hadoop at best in few 100 offices in India.

b) Apart from HDFS, most parts of hadoop ecosystem are already being replaced/renovated by the innovation leader companies. Google (the inventor) phased out mapreduce [The Elephant was a Trojan Horse: On the Death of Map-Reduce at Google](http://the-paper-trail.org/blog/the-elephant-was-a-trojan-horse-on-the-death-of-map-reduce-at-google/), projects like mahout already seem to be moving away from hadoop and a lot of companies seem to be moving towards [Apache Spark™ - Lightning-Fast Cluster Computing](https://spark.apache.org/) . Now by the time you become professional, spark might be the talk of the town, in 10 more years even it might be removed.

I would rather advice you to read books like mmds, foundations of data science, ISLR, Deep Learning etc. All of them are freely available on internet. Save your money and time (reading them now than you would read them after you graduate).

Edit : Adding the link to books (will rather link the blog post at ParallelDots we made collecting links to all these books, most of them freely available)

[Free Machine Learning Books](https://blog.paralleldots.com/data-science/24-best-and-free-books-to-understand-machine-learning/)

[Free Data Science Books](https://blog.paralleldots.com/data-science/50-must-read-free-books-for-every-data-science-enthusiast/)

[Free Deep Learning Resources](https://blog.paralleldots.com/data-science/deep-learning/free-resources-deep-learning/)

[Must Follow Blogs](http://blog.paralleldots.com/technology/artificial-intelligence/must-read-blogs-ai-deep-learning-enthusiasts/)

Answer is Here: https://www.quora.com/What-should-be-the-learning-path-after-going-through-basic-machine-learning-material/answer/Muktabh-Mayank?srid=qie

### Q. Do companies hire data scientists with zero internship experience?

Yes. Doing an internship is not necessary to work in Data Science.

But most companies won’t hire Data Scientists with no experience of solving real world problems.

I would say anything out the following gives you equal (or more) weightage than internship at a tech startup.

1. A bronze medal or better at Kaggle,
2. A research project under an academic that resulted in a publication at a decent venue,
3. An internship doing a proper AI project (which is to be deployed within the company), (whose alternatives we are discussing),
4. Contribution to an open source project,
5. Some other way in which you have worked in a real world setting optimizing for different constraints.

The bigger companies working in Data Science, maybe like Fractal and Manthan, might be hiring absolute freshers too, but most companies in this space are small startups who cannot take risk if hiring absolute freshers and training them.

Original Answer Link: https://qr.ae/pNnch5

### Q. Is it true that maths or statistics isn’t required at all in data science while programming skills are required? My friend told me the same thing. He said that's why a majority of data scientists are from BTech backgrounds, not from maths or stats.

Any respectable Engineering program (BTech) has a sizable number of Maths and Statistics courses. At BITS Pilani, I had coursework in Vector Calculus, Statistics, Linear Algebra, Differential Equations, Optimization and Operations Research (each one was a separate 1 semester long course) which every Engineering student had to take. On top of this there were Computer Science specific subjects (Data Mining, Machine Learning) which taught the Maths and Programming part both to everyone. The theory taught in these many courses is frankly enough to read most Data Science papers and understand concepts (You might need to read a couple of tutorials here and there).

So the hypothesis that Engineering students aren’t aware of Maths and Stats is a wrong one.

Now about Data Scientists not needing Maths skill, data scientist as a job description is very broad just like software development.

You can say that Software Engineer doesn’t need to know about Databases, and that will be true for many people: Software Engineers who write Operating Systems, Software Engineers who develop frontend applications, Software Engineers who write compilers and below par Software Engineers who work on web application backends.

There are similarly many job descriptions for Data Scientists. Some of them actually don’t require too much Maths and just programming, but most of them do require some basics of the above mentioned mathematical areas.

Original Answer Here: https://qr.ae/pNnctE

### Q. My answer to “What are the mistakes people make when they start Machine Learning?”

Originally Answered here: https://qr.ae/TiNJou

Some mistakes according to me:

1. Focusing too much on math in the initial stages. To train your first Neural networks, you don’t need to know in detail how backprop works (backprop derivatives of all ops for example), or you don’t need to understand the support vector derivation to train SVMs. Dont start reading complicated mathematical resources in the beginning, it makes learning very slow. Touch on these topics when you are somewhat experienced writing code and training algorithms.
2. Focusing too little on Math is similarly bad too. Not knowing what different parameters to a Conv layer in Keras signify is also sub optimal. A basic book on Machine Learning with theory is best starting point.
3. Practising less. This is what differentiates between good and excellent practitioners. Most excellent practitioners can think of 100s of ideas around a dataset and can iterate quickly on them. That is how accuracy on a dataset will go up. Other practitioners will waste time in thinking about what would be the perfect method and code their “one best” method in hours, which will mostly not work in the end. You cannot think in advance what should be the method for best accuracy, EDA and trail-and-error is the key. Remember during practicing “Getting average accuracy on multiple datasets << multiple round of iteration to get good accuracy on one dataset”.
4. Not focusing on basics numpy and pandas. As I said earlier, you need to iterate over many ideas quickly rather than thinking of “one true idea” that will work. Its grit boring work. To make this quicker, good command on Numpy and Pandas help. Lesser number of Google searhes == more code. Tensorflow/PyTorch have been purposefully written close to Numpy to make sure that Numpy users can iterate quickly.

### Q. What are some signs to recognize inexperienced Machine learning engineers ?

Original Answer Here: https://qr.ae/pNncvm

In my view most inexperienced Machine Learning people (including me say 6–7 years back) focus more on algorithm than data.

Newbies want to try out all the 250 (dummy number) algorithms on the dataset they have got without EDA on the data itself.

Changing the algorithm will generally give what 2–3% (again dummy number for giving an idea about magnitude) gains in accuracy, while, arranging/ balancing/ feature engineering/ augmenting the data can give manifold accuracy gains.

Machine Learning is not yet a cool art where you summon a Charlizard and then it will be burn the competitor to the ground. It requires grit, getting dirty with the data and understanding what is the algorithm learning through many boring iterations.

### Q. I just hired an unqualified machine learning engineer, he know only some basic Tensorflow and have no idea about the maths and build more complex neural networks, should I fire him immediately or give him 2 months to improve?

Originally Answered Here: <https://qr.ae/pNncrn>

Why did you hire an “unqualified” person ? Don’t you have a job description and interview process ? If there is such a huge gap between the hiring and requirement, I think its a problem with the company leadership and process. You should understand who you want to hire for what exact profile before you even put the JD out.

What you should do next is dependent on what is expected of the employee in the long term.

Do you expect him to do something Math(y) ? I don’t think they will be able to learn in 2 months. OTOH, 99.9% of the companies won’t need people who can write “new and complex” neural networks. Its mostly about transforming the data well, using clever training methods and loss functions etc. That said, even to gain enough experience for this, they will need more than 2 months.

The general observation I have is that a lot of companies don’t have clear specs and requirements and they expect their Data Scientists to weave magic. Then they won’t hire senior people who can tell what is possible and what is not and put the entire pressure of expectations on cheap “fresher” employees, who generally have no real world experience (and many times will overestimate their capabilities, thus signing into something they cannot accomplish). This type of org structure is a house of cards. Hope this is not the case with your organization.

If the work is running different open source models and experiments, 2 months is a decent time to learn and apply. If your expectation is in a domain , which is well worked and researched in, 2 months should be a good time for a less experienced person to catch up. If the expected work is this much, the hiring strategy is also not wrong and this is the general capability of talent available in the market. Groom them a bit for 2 months and everything should work out. If the expectation is any more open ended, you have made a wrong hiring decision.

### Q. Will the talent shortage in AI end soon?

Originally Answered Here: <https://qr.ae/TSkAzo>

There is no shortage of AI talent as of 2019 end. Too many people in India know basics about AI and can be trained to work. I am not sure if there is a way to see people applying for AI positions but if there was a way, you could clearly see that way many people apply for all AI/Data Science job (at least here in India). There was a time when just knowing some basics got you a job but no more now.

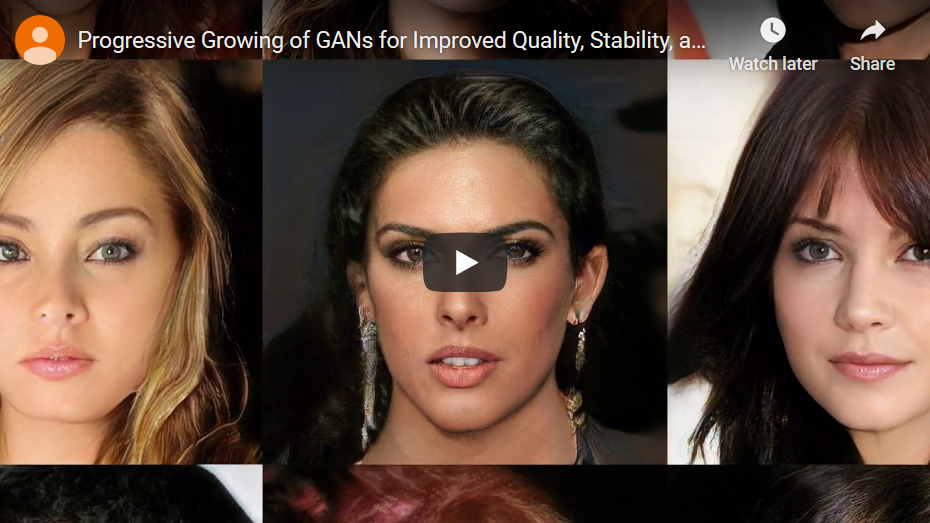
The problem has now shifted to retaining top talent as too many people in the world are applying AI and working abroad is a big deal for most Indians. *So the situation is a lot more like Software Development profiles now for most AI profiles too.*

There is a huge pool of talent and there is arms race for retaining top talent.

### Q. Will data science and machine learning get automated leading to lesser opportunities for data scientists by 2025 ?

This blog post is also an answer by me on Quora [here](https://www.quora.com/Will-data-science-and-machine-learning-get-automated-leading-to-lesser-opportunities-for-data-scientists-as-per-https-www-datasciencecentral-com-profiles-blogs-data-scientists-automated-and-unemployed-by-2025-update/).

Yes. (2025 is not the date I think its going to happen, but its inevitable and will happen in near future). They will be automated to a good extent. So will be Software Developers, designers, manual workers, teachers, linguists, musicians, game developers etc,etc. There are already rudimentary projects like [Turning Design Mockups Into Code With Deep Learning](https://blog.floydhub.com/turning-design-mockups-into-code-with-deep-learning/) which can turn a design mockup into HTML/CSS code, [carpedm20/ENAS-pytorch](https://github.com/carpedm20/ENAS-pytorch) which can design neural networks without a Data Scientist, [Why AutoML Is Set To Become The Future Of Artificial Intelligence](_blank) , system which can generate new characters for games, [Microsoft AI can translate Chinese to English just as accurately as humans](http://www.business-standard.com/article/international/microsoft-ai-can-translate-chinese-to-english-just-as-accurately-as-humans-118031500414_1.html) , [Baidu's Deep Voice can clone speech with less than four seconds of training | Computing](https://www.computing.co.uk/ctg/news/3028065/baidus-deep-voice-can-clone-speech-with-less-than-four-seconds-of-training) and multiple such projects.



Video Link : https://youtu.be/XOxxPcy5Gr4

AI will impact every job profile which exists as of now, Data Scientists no exception, automating some or a lot of work people spend their time on. For a lot of time these systems will become a <Man + Machine> AI systems rather than just a Human working before stuff is totally automated. So its not like everyone becomes redundant day 1, but they will eventually.

Full Automation of any field is going to take way longer than 2025 IMHO. That said, yes a lot less people will be needed for the same task as of today. Then what will people do you ask ? newer more complex tasks.

Automation is not a new phenomenon. Think about the railway breaks a long time back:



Video Link Here: https://youtu.be/EEUkmP2nyxo

So much work was once needed to just stop the train. A lot lesser work is needed today to run/stop the train and not just that, slowly trains are moving towards full automation, but right now they are in a <man + machine> stage. A lot of jobs will stay in this phase for sometime before full automation kicks in. But unlike railways, which take generations to move from one stage of automation to another, AI is causing changes at a very high rate.

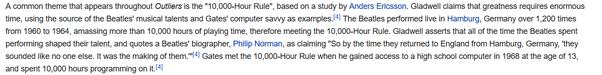
What is the effect on a general Data Scientist (or any white/blue collared worker for that matter):

1. Adapt to AI. Automation has started, but AI aided jobs will stay for a few more years than non-AI jobs. So while no plain X jobs by 2025, X + AI jobs might be around till 2030.
2. Things wont be like earlier generations where one skill learnt gets you a job for entire life. One needs to be open to learning new skills and get started in the middle of life.
3. Average level of education needed will be high. Think of it, 50 years back “High School” was all the education needed for most jobs. Now its somewhere between high school and graduation. Masters and Research might look like the next frontier, but these degrees are too slow and broad. Coursera like courses will become more important in catching new skillsets. You might already see people doing that a lot.
4. With more uncertainty in jobs, millenials probably will want to be a less “spend-y” and more frugal. You can see things happening already [6 reasons why more millennials aren't buying homes](https://www.businessinsider.in/6-reasons-why-more-millennials-arent-buying-homes/articleshow/59005530.cms) and will actually increase. AI is just a trend in a longer cycle of Automation and millenials are at a point in history where education+society was according to old norms but automation has reached a point where jobs have become uncertain. Younger people will be smarter.

### Q. Why are Kaggle Grandmasters in a great demand?

Because they work really really hard on Applied Machine Learning (that’s what kaggle is) and thus have become really good at their job. If there is a hard Machine Learning problem with Billion dollars behind it, with a high probability, Kaggle grandmasters will be able to solve it better than an average joe.

Working consistently on something for years makes one a master of the art. [Outliers (book) - Wikipedia](https://en.wikipedia.org/wiki/Outliers_(book))



People really good at their job (say in top 10%) are really sought after in any field, not just Machine Learning. They drive the innovation, solve open ended problems and hence they get the rewards.

### Q. How do I become a data scientist in 2020 in India by self-teaching?

I originally answered the question on Quora here: <https://qr.ae/TmYc20>

I don’t think the path to being Data Scientist in India is different from being a Data Scientist abroad. It is in fact 99% same if you live in a place with good internet connectivity and understand English.

You have to clarify to yourself the following:

1. 1. As a self-learner, determine if you like learning from books or like learning in virtual classrooms.
2. 2. Either way, you will have to dedicate around 10000 hours of hard work to learning and practical exercises. It looks simple, but most people lose out here. They do not put enough effort.
3. 3. Point 2 above requires self drive. It is not easy. A proxy to that is buying online courses / books. The money you put into them makes you (and probably your parents who drive most people towards learning stuff) have skin-in-the-game and you wont want to lose out. If you arent that wealthy, you will have to push yourself.
4. 4. You have to invest in a good computer. (somewhat 60000 INR cost).
5. 5. You need to have good internet.

That is all. There are enough resources available for free to learn and make yourself a good data scientist.

1. 1. Learn Python programming. This is first and foremost. There are many free books and online courses to learn Python (sometimes Python for Data Science specifically) if you aren’t enrolling in a course. When you are doing a course or reading a book, don’t just read it, force yourself into using python after you are done learning. If book exercises look boring to you, join a startup near you as a free intern and do some coding for them. Forcing yourself to write 1000 lines of code (made up number) is very important.
2. 2. Make yourself comfortable with Python ecosystem : Numpy, Scikit\_learn, Keras, PyTorch, Pandas. There are courses and free books available to do these. Look for their ratings on goodreads (for books) and course stars. Best is to just do the famous courses [Applied Data Science with Python | Coursera](https://www.coursera.org/specializations/data-science-python) , [http://deeplearning.ai](http://deeplearning.ai/) and [Machine Learning | Coursera](https://www.coursera.org/learn/machine-learning) . If you are looking for free books, check out some lists I curated on ParallelDots blog: <https://blog.paralleldots.com/data-science/50-must-read-free-books-for-every-data-science-enthusiast/> , <https://blog.paralleldots.com/data-science/24-best-and-free-books-to-understand-machine-learning/> , <https://blog.paralleldots.com/data-science/deep-learning/free-resources-deep-learning/> , <https://blog.paralleldots.com/data-science/nlp/free-natural-language-processing-resources/> . Force yourself to write more and more code. Do problem sets from courses, exercises from books, work for free in a startup, enter analyticsvidhya competitions or whatever suits you. Unless you practice after you learn, its not going to work.
3. 3. This much will make you employable by many companies. If you still are trying to push to learn more, you will have to start reading research literature. This is not entirely necessary, but if you still want to, you should get yourself associated with some decent university group (or a research startup), where novel problem statements are being solved. This step is really hard without talking to peers, but if you are really self driven, Twitter is a good place to follow people and handles to learn about research while being outside the research world. [Becoming an Independent Researcher and getting published in ICLR with spotlight](https://medium.com/@andreas_madsen/becoming-an-independent-researcher-and-getting-published-in-iclr-with-spotlight-c93ef0b39b8b)

### Q. What signs will tell you that your company is not taking data science seriously?

Originally answered here: https://qr.ae/pN28BW

Note : I am assuming that this question was asked to understand ways to judge a potential employer by an applicant who wants to take a job as Data Scientist. What might be something that is risky. There might be other ways to look at it, but this seems most plausible source of the question to me.

Well, if your company has hired people as Data Scientists, there is definitely a vision to derive value out of data. Funds were allocated, someone actually made effort to hire Data Scientists and (hopefully) a set of problems were defined to solve. If your company is not doing all these 3 (allocation funds, hiring data science team and defining problems to solve), you can safely assume its not going to implement Data Science in near future.

However, something being in the vision is different from getting executed successfully. Sometimes, the company might be looking at Data Science, but there might be some antipatterns in execution that make life of Data Scientists hell and probability of failing high (Similar risks as being a Data Scientist in a company which doesn’t take Data Science seriously). Execution has its own challenges. Some other issues I see :

1. Not hiring Data Scientists at senior positions and allocating Data Scientists to work under senior technology leaders. Although this might work sometimes, a senior traditional technology executive might take decisions that are not really the best to run a Data Science team. So for example a senior backend developer put in charge of Data Science team will often ask Data Scientists working under him to accommodate additional constraints in the data pipelines they implement. For example, they might be (for no reason) asked to work within the box of software architecture which their lead has designed, reducing their efficiency.
2. Sales team not taking feasibility into account is another huge problem. Basically, the Data Science team need not be involved later just in the development of product/service being sold, but also earlier while the sale is being made to make sure something feasible is being sold. They can also suggest applications of data during sales time to make picth better. It is a very common pattern that Sales people of an AI company sell something that is not even feasible.
3. The third is expecting Data Science teams to work while not having/collecting enough data. This often is a result of 1 or 2 where data wasnt collected to optimize some other tech need, or wrong expectations have been set to any client about the performance a low amount of data can deliver. Data Science is not magic, you need enough data to derive insights or train algorithms on it.
4. A spineless leader for Data Science team. Well a spineless leader is always bad as they just pass the pressure downwards. People leave bosses not companies. A lot of the work of senior Data Science folks is expectation management and if they cannot do so, 1,2 and 3 mentioned will haunt the team’s work.