**Getting Started into Data Science**

Every data scientist has its own journey of getting into Data Science. If you are a beginner looking for "How to start your data science journey?", you are in the right direction. This post will not lay down rules of becoming a Data Scientist, rather will guide you to frame your own Data Science journey and get started with it.

Starting with my journey, as a beginner one of the primary questions I had was, "If there are so many libraries available performing all of the model training , there are so many projects out there doing just the tasks I have thought of, Face recognition, Sentiment analysis, there are so many prepared datasets on so many different forums, where does the role of a Data Scientist come into play?"

Data is much more than the models themselves which are merely the tools to get work done. It is about building intuition from the data, inferring its meaning and transforming data into a story.

I have compiled a list of resources, platforms, data sources and courses which have helped me over the years to step into the field of data science. Feel free to refer to the sources as you create your own roadmap.

**Resources:**

1. https://www.analyticsvidhya.com/blog/?utm_source=feed (blogs and hackathons)

2. https://machinelearningmastery.com/

(Blogs)

3. https://towardsdatascience.com/

4. https://medium.com/

**Courses:**

1. https://padhai.onefourthlabs.in/ (presently offers deep learning and data science course)

2. https://www.coursera.org/learn/machine-learning?utm_source=gg&utm_medium=sem&utm_campaign=07-StanfordML-IN&utm_content=07-StanfordML-IN&campaignid=1950458127&adgroupid=69480953983&device=m&keyword=study%20machine%20learning%20online&matchtype=b&network=g&devicemodel=&adpostion=&creativeid=351281535285&hide_mobile_promo&gclid=CjwKCAiA4o79BRBvEiwAjteoYBOazjWmbe_co_Axu09p2uqtJVne2dSGnHgZ0Cz1isSziwWtDnrt4RoCVpIQAvD_BwE

**Beginners guide**

1. https://www.kaggle.com/general/20856

(Worked on projects, guides, competitions, platform to practice, repository of datasets)

2. Google colab (platform)

3. Machine hack

**Roadmap**

1. Understanding different kinds of datasets (Text based, Time series, Graph data, etc)
2. Data Visualisation tools such as Tableau or using Python libraries
3. Basic pre-processing steps (handling missing values, encoding categorical values, validating features, basic feature transformations)
4. Understanding the use case and how the project is built around various use cases
5. Understanding the classification of machine learning (supervised, unsupervised, Ann, etc)
6. Fine tuning parameters for various models, checking others projects on the same and analysing
7. Generating intuition around why some algorithms work corresponding to a use case

[*Note: Understanding featurisation techniques by reading different research papers based on the domain will surely help work with real projects]*

**Preferred languages:** Python and R ( I chose Python as I was more used to its fundamentals)

**Working with Python Key libraries:**

1. Numpy (mathematical computations, matrix operations )
2. Pandas (built on top of numpy to support working with datasets )
3. Sklearn (ml algorithms)
4. Matplotlib (basic visualisation)
5. Seaborn (alternate visualisation library)
6. Keras, pytorch, tensorflow, caffe, etc (frameworks) _use of libraries should come only when intuition about algorithms and their mathematical concepts are developed (preferred for deep learning tasks)_

Data Science is a journey, it is not an end. As the domain itself keeps progressing each day, we need to put ourselves on a continuous learning journey.

**Wish you good luck for your own Data Science journey!!**

[The above links are used as references and not promotion. They helped me in my Data Science journey and might help you too. Please feel free to share the document with your friends and help them in their journey too]