

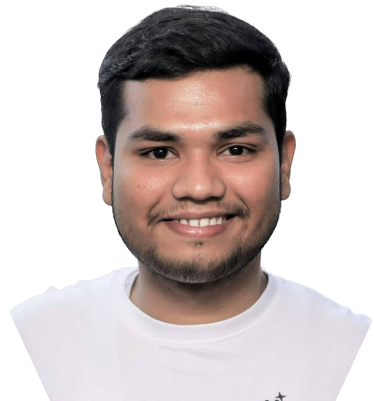
Data Dive

Unveiling the power of numbers

Date : 5th June 2023 | Speaker : Ayon Roy |

Venue : Auditorium @ Dr Akhilesh Das Gupta Institute of Technology & Management, Delhi, India

Visit - AYON-ROY.NETLIFY.APP



Hello Buddy!

I am **Ayon Roy**

Executive Data Scientist @ NielsenIQ

Z by HP Global Data Science Ambassador

Mentored/Judged **100+** Hackathons

Delivered **60+** Technical Talks

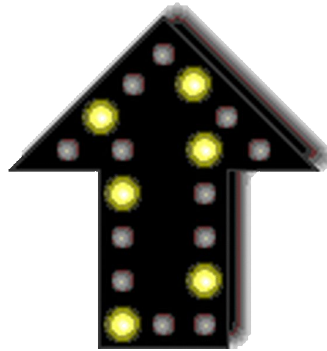
Brought **Kaggle Days Meetup** Community in India for the 1st time

If you haven't heard about me yet, you might have been living under the rocks. Wake up !!

Agenda

- How to start your Data Science journey
- A Primer to Hackathons
- Things to focus on, while making a Data project?
- How & Why organize your Data project?
- Exciting projects to learn from
- A Primer to Competitive Data Science
- A Primer to Data Science Internships

How to start Data Science



Start with Maths for Data Science

But **why should I do Maths**
first for Data Science ?

- Week 1 : Linear Algebra [B] <https://www.khanacademy.org/math/linear-algebra>
- Week 2 : Calculus [B] <https://www.youtube.com/playlist?list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr> or <https://www.mathsisfun.com/calculus/> ; want theoretical notes , find it at <https://the-learning-machine.com/article/machine-learning/calculus> .
- Week 3 : Probability [B] <https://www.edx.org/course/introduction-probability-science-mitx-6-041x-2>
- Week 4 : Statistics [B] <http://alex.smola.org/teaching/cmu2013-10-701/stats.html>
- Algorithms (Only if you want to learn proper software development) [Highly optional]
This is an overview of what the students study as the subject Data Structures & Algorithm . So if you are fluent with this part , you can skip this !! <https://www.edx.org/course/algorithm-design-analysis-pennx-sd3x>

Start with Python
&
try to **implement** those
Mathematical Concepts

**Start exploring Libraries
& then start Data Science /
Machine Learning Courses**

-
- Introduction to python for data science [B] <https://www.datacamp.com/courses/intro-to-python-for-data-science>
 - Want to dive deeper into Data Visualization & Pre-Processing ? Look into Data Visualization & Pre-Processing section in miscellaneous resources . [Highly optional]
 - Want to explore the field of Deep Learning ? See the Deep Learning Section in miscellaneous resources . [Highly optional]
 - Want to explore the field of Natural Language Processing [NLP] ? See the Natural language Processing Section in miscellaneous resources . [Highly optional]
 - See how ML codes are written and made to work at - > <https://github.com/maykulkarni/Machine-Learning-Notebooks> or <https://github.com/GokuMohandas/practicalAI/blob/master/README.md> . [Highly optional]
 - Find useful resources here at <https://github.com/ujjwalkarn/Machine-Learning-Tutorials/blob/master/README.md> . [Highly optional]

Don't rush behind
completing Courses & add
them to Resume

**Understand the concepts
well before starting
Projects**

Now it's time
to start with
Projects

But which projects to
start learning **with ?**

- Beginners Section [B] : Brush your basic concepts and revise them to start doing projects

Titanic Dataset

Iris Dataset

Stock Price Prediction

Stores Sales Forecasting

Housing Price Prediction

Guide for Beginner Projects:

First of all see Below 2 videos to get an idea on how to make projects of Data Science and Machine Learning And then Move to Kaggle for Making your own project.Its is Good if you Make Minimum 2-3 Projects on your own.

- Titanic Survivor : <https://www.youtube.com/watch?v=fS70iptz-XU&t=>
- Credit Card Fraud Detection : <https://www.youtube.com/watch?v=gCWBFyFTxVU>

Intermediate & Advanced Section

- Learn libraries like Opencv , Tensorflow , SkLearn

1) Natural Language Processing : MNIST Handwritten Digit Classification , Twitter Sentiment Analysis

2) Email Spam Classifier

3) Fraud Detection System

4) Computer Vision : Face Recognition , Face Detection

" I am a beginner in AI,ML,Data Science & trying to do projects; but not succeeding as I get stuck more often "

Here is my way ahead if you are facing the same.

" Start Simple Projects & Be Motivated "

We all usually want to do the best projects & showcase them in our [resume](#) & hence sometimes end picking up a complex project at the first go. But do understand that while it's very normal to pick complex projects as a beginner, because we can't analyze the scale of project at first go. Picking up a complex project at a first sight may demotivate you as they have a lot of details, requires lot of studies to progress, thus a beginner ends up leaving the project midway & be traumatized. So start your journey with Simpler & Smaller projects as they require comparatively less details & can be achieved over a short period of time, thus helping us to stay motivated & keep doing projects. And as the learning in AI, [machinelearning](#), [datascience](#) never stops, so as we get motivated with completion of small projects; we learn & practice more while increasing the complexity of our upcoming projects. Still waiting to start ? Start today !! All the best !!

A primer to Hackathon

What do you do in a Hackathon?

- You sit down and identify the problems as a team. Define the problem statement accurately, clearly and discuss its impact on society.
- You learn the different components of the problem, objectively. Ask questions like:
What can be done effectively?
How, why did it happen?
Where do we start?
- You generate different ideas and decide on one final idea to present to experts and SMEs for feedback.
- Based on the feedback, You develop a crude prototype of the idea.
- You iterate, iterate, and then pivot if needed - then iterate, iterate and again iterate!
- You present your last iteration, rough and crude but in the best light possible
- Learn through the process, what worked and what didn't work, what you as a team agreed on and what you as a team didn't agree on

What do you get out of participating in a hackathon?

- You learn to become a problem solver!
- You learn to work with a team of interdisciplinary people.
- You learn to give voice to the problems of society. you live in.
- You learn to become part of the solution making progress and find out the real struggles behind it.
- You or your team may win the hackathon and get recognized as trailblazers in the challenges facing society now, but most importantly you will gain inside knowledge of how technology works, and how it can be leveraged to benefit humanity.



Analytics Project Life Cycle

The 5 Phases



How to organize your Data Project?

Local Project Directory	Github Repository
<ul style="list-style-type: none">▪ Project plans/objectives▪ Project datasets▪ Project codes<ul style="list-style-type: none">○ Jupyter notebook○ R scripts○ Python scripts▪ Output files<ul style="list-style-type: none">○ Visualizations○ Tables○ Other useful outputs▪ Project report	<ul style="list-style-type: none">▪ README file▪ Project datasets▪ Project codes<ul style="list-style-type: none">○ Jupyter notebook○ R scripts○ Python scripts▪ Output files<ul style="list-style-type: none">○ Visualizations○ Tables○ Other useful outputs▪ Project report

<https://gist.github.com/ericmjl/27e50331f24db3e8f957d1fe7bbbe510>

**But why organize
your
Data Project?**

- **Organization increases productivity** as avoid wasting time searching for project files such as datasets, codes, output files, and so on.
- A well-organized project helps you to keep and **maintain a record of your ongoing and completed data science projects.**
- Completed data science projects could be **used for building future models.**
- A well-organized project **can easily be understood by other data science professionals** when shared on platforms such as Github.

Ideas to take reference to build in Hackathon using data

Here are some cool ML hackathon projects as inspiration to get started!

1. A system that detect fires and Smoke, It has two key services: fire tracking and alerting the emergencies

Video : <https://youtu.be/L5eUPXxJVdI>

Github : <https://github.com/Fellah-wassim/IgnitionGuard>

2. Help visually impaired individuals cross the street using machine learning.

Video : https://youtu.be/o_cyugOoDiU

Github : <https://github.com/Szugalew/PedestrianTrafficLightDetectorRaspberryPi>

3. A YouTube sentiment analysis app built using Node.js, Cohere API, and Google Cloud's YouTube Data API

Video : <https://youtu.be/ZrOemkKDYxU>

Github : <https://github.com/Ryan-Diep/ChatRoller>

4. Using Image Analysis to predict what color palettes for clothing would look best on the user.
5. Using neural networks for facial recognition to make paying easier.

What is Competitive Data Science ?

A great opportunity to

- **Sharpen your programming & analytical skills**
- **Enhance domain knowledge**
- **Learn more about practical applications of data science & machine learning algorithms**

by participating in some real world Data Science Competitions hosted by organizations on various platforms.

But why
Competitive Data Science
is gaining traction in 2023?

It's possibly due to the



**BUSINESS
ORGANIZATIONS**

Organizations are having hard time to solve so many data science problems while their data science team is busy with other projects. So hosting a data science competition on certain platform can help & is helping.

Data science competitions help organizations solve complex business problems while enabling data scientists to learn from the experience and win awards.

Organizations need to define the problem, provide data and put a prize on the challenge. Competing data scientists build and present different algorithms to be the winner.

Why should you try
Competitive Data Science
at least once?

To avoid situations like

when you have your first real-world
adult experience after graduating



And to

- Understand how to solve predictive modeling competitions efficiently
- Learn how to preprocess the data and generate new features
- Be taught advanced feature engineering techniques
- Be able to form reliable cross validation methodologies
- Gain experience in analyzing and interpreting the data
- Master the art of combining different machine learning models
- Get exposed to past (winning) solutions

How should you start your **Competitive Data Science** journey?

The only thing you need to know **Before Starting** your CDS journey

“For participating in data science competitions, you only need an urge to constantly learn and improve. Getting a good ranking will follow.”

Initial steps to start your CDS Journey

- Make sure your basics about Python & Mathematical concepts are clear enough.
- Focus on understanding core Data Science & Machine Learning algorithms
- Try to make self projects with the concepts you learned

The next steps

- Try participating in Kudos/Knowledge Competitions (Like Titanic etc.)
- Then try to learn about the approaches from other's notebooks
- Try to apply your learnings from those approaches in Featured/Prized Competitions
- Try exploring variety of techniques you can use to get better results

How to approach a Competitive Data Science Problem?

1. **Start with a very simple baseline model**
2. **Understand the problem and data to create a good validation set**
3. **Try Feature engineering**
4. **Try building a variety of models**
5. **Try stacking or blending of these results using Ensembling**



Time is a very crucial factor in any data science competition.

You should not waste your time writing the same snippets from scratch again and again in multiple competitions. Instead, focus your valuable time on doing something new and better

Where to get involved with Competitive Data Science ?

My personal suggestions

- <https://www.kaggle.com/>
- <https://www.crowdanalytix.com/community>
- <https://zindi.africa/about>
- <https://datahack.analyticsvidhya.com/>
- <https://www.crowdai.org/challenges>
- <https://tianchi.aliyun.com/competition/gameList/activeList>
- <https://www.datasciencechallenge.org/>
- <https://www.drivendata.org/>

Know a few more platforms to kick start your CDS journey [here](#)

Be a part of Communities like

1. Kaggle Days
2. Women in Machine Learning & Data Science
3. GDG, WTM, ODSC
4. Datazoic & a lot more....

Be a part of as many hackathons as you can

who wants to miss networking, free food & swags alongside unlimited learning

How to get involved more with Competitive Data Science ?

1. **Do such courses where the skills learnt in them can be used in Competitions.**
2. **Publish your competition research**, approaches on the forum & do write about the things that you want to share with others via blog etc.
3. **Participate in Discussion forums**, share your knowledge through answering questions & asking genuine questions.
4. **Make notebooks & share them along with great EDA, feature engineering etc** so that others can learn from it.
5. **Try to reproduce interesting kernels.**
6. **Be consistent** in whatever you are trying to share with the CDS community.

Is
Competitive Data Science
everything what the industry
requires?

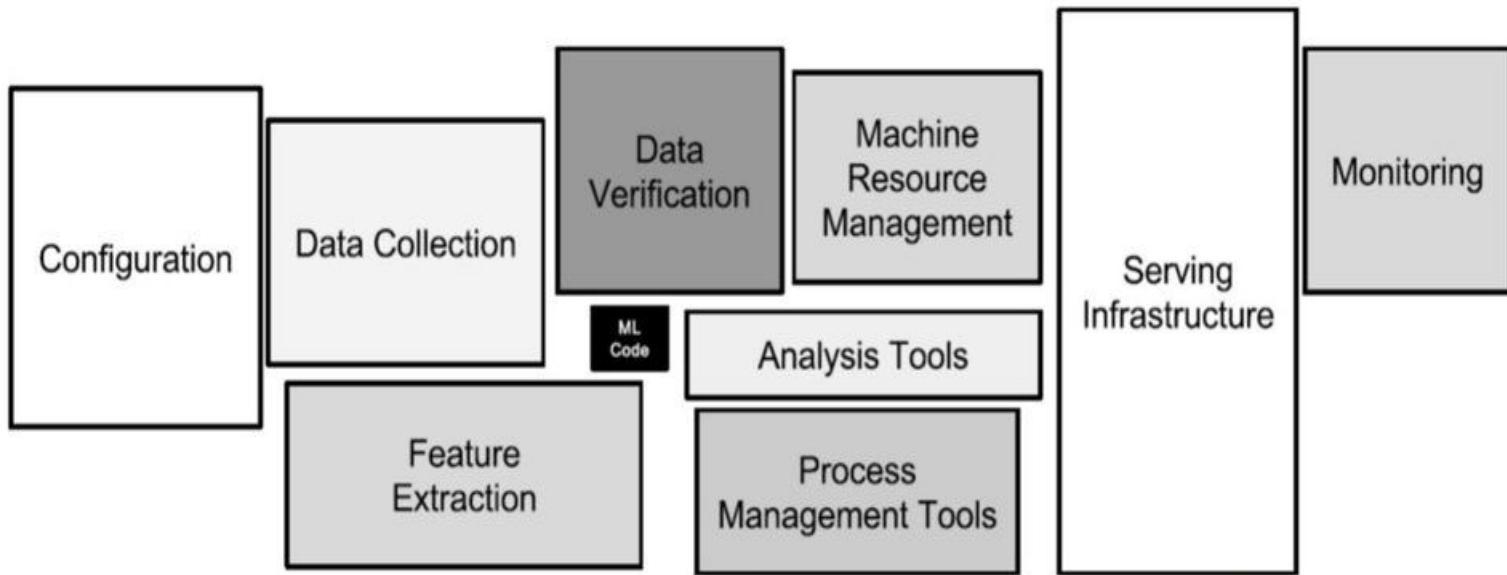


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

View the Google's Research Paper [here](#)

A few important pointers to keep in mind

1. **Focus on understanding what business use case you are trying to solve** before applying Data Science, Machine Learning.
2. **Focus on Communication Skills to convey the result** of your Data Science concepts to the business stakeholders.
3. **Focus on DevOps** to make your models production ready.
4. **Focus on networking & showcasing your work** to the community.

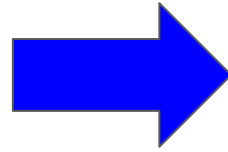
& be prepared to rock the industry

Take your baby steps
with
Internships

The What of Data Science Internships

What are the requirements of a decent Data Science Internship Opportunity ?

Let's divide the variety of Data Science Internships out there in the industry



Requirements for **Entry Level Internships**

- Ideal for Second-Third Year Students of a 4 year Undergraduate course
- Basic Level of Python, OOPs, File Systems
- Good knowledge of Scraping, Numpy, Pandas & Data Visualization libraries
- High level overview of Machine Learning algorithms
- Few basic Data Pre-Processing & Exploratory Data Analysis Projects

Requirements for **Middle Level Internships**

- Ideal for Third-Fourth Year Students of a 4 year Undergraduate course
- Sound knowledge of Data Science concepts & other ML algorithms with good grasp on statistics & concept of maths, SQL
- Good knowledge of Deep Learning concepts, DBMS, API development
- Self projects using ML algorithms

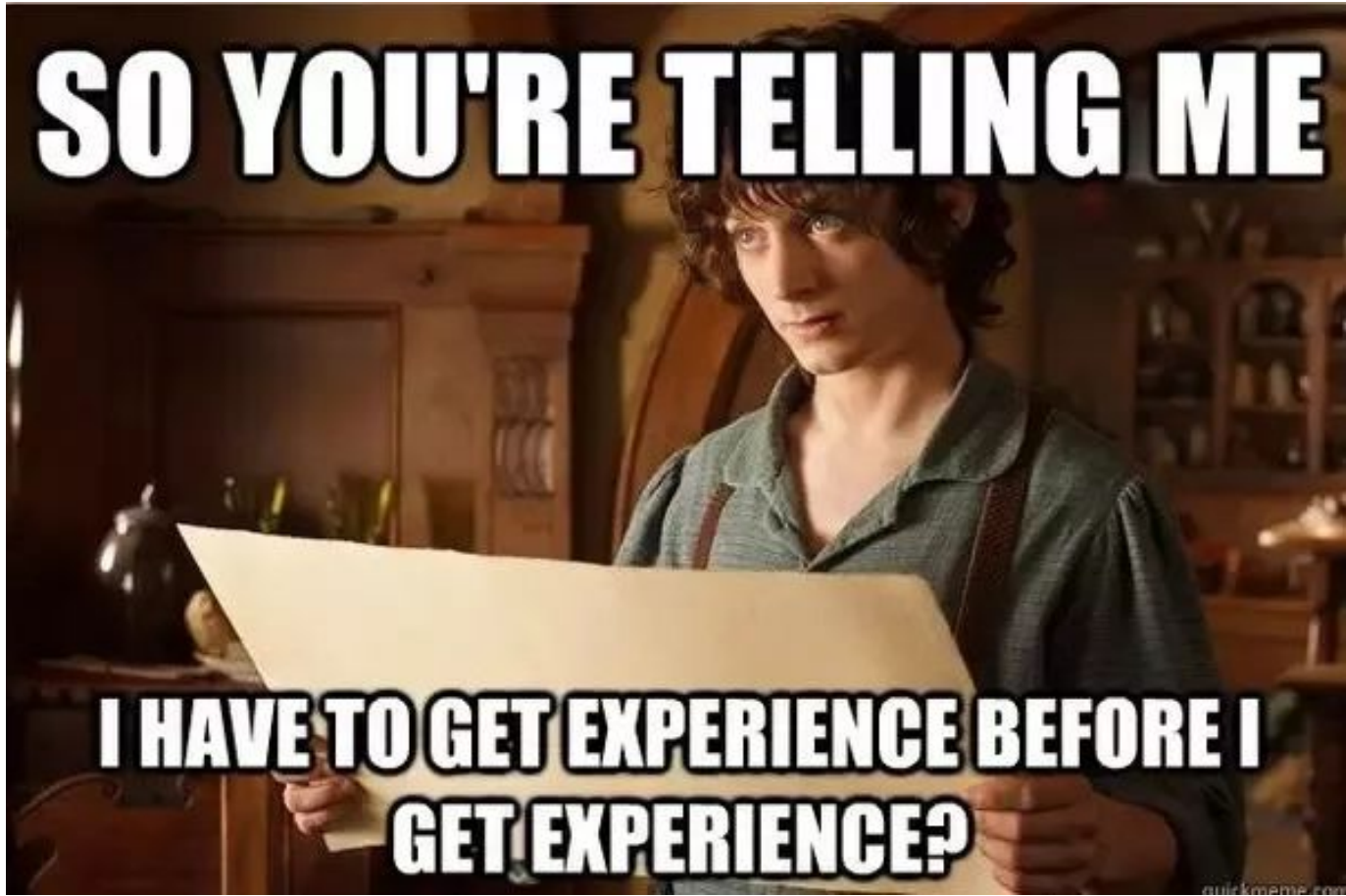
Requirements for **Advanced Level Internships**

- Ideal for Final Year Students of a 4 year Undergraduate course
- Specialised domain of expertise like Computer Vision, Natural Language Processing etc.
- Proficient with whiteboarding of ML algos along with explanation of the basics
- Basic knowledge of Docker, Cloud
- 4-5 very good projects using the complex Deep Learning concepts like LSTMs, Transformers etc.

Best way to get a Pre-Placement Offer before campus hiring starts

The Why of Data Science Internships

SO YOU'RE TELLING ME



**I HAVE TO GET EXPERIENCE BEFORE I
GET EXPERIENCE?**

Why should students do **Data Science Internships** ?

- **To learn dealing with Messy, unstructured, incomplete data** (This is real industry)
- **To experiment & learn new things** as an Intern so that you can save time as a Full Timer excluding the mistakes you did as an intern.
- **To understand how end to end real world Data Science applications works.**
- **To network with people** who look like Future of You.
- **To work under pressure & learn how to deliver** in tough deadlines too.

To make like minded friends & throw a party at Starbucks

The How of Data Science Internships

How to apply for Internships ?

- LinkedIn Jobs [Worked for Me]
- Angel.co [Worked for Me]
- Internshala [Worked for Me]
- Through Organization's Careers Page [Worked for Me]
- Commenting Interested on someone's LinkedIn Posts [Not Worked for Me]
- Career/ Internship Fairs [Never Tried]
- Via Winning Hackathons [Worked for Me]
- Asking for Referrals [Worked for Me]
- Community Events [Have seen it work]

WARNING
for
of Data Science Internships

Pay for Trainings, not Internships



A few useful resources

1. <https://towardsdatascience.com/use-kaggle-to-start-and-guide-your-ml-data-science-journey-f09154baba35>
2. <https://www.coursera.org/learn/competitive-data-science#syllabus>
3. <https://towardsdatascience.com/how-to-successfully-manage-a-data-science-delivery-pipeline-33bdec1a9a27>
4. <http://kaggle.com/learn>

Let me answer your Questions now.

Finally, it's your time to speak.



Danke Schoen

Questions ? Any Feedbacks ? Did you like the talk?
Tell me about it.

If you think I can help you,
connect with me via

Email : ayon-roy@outlook.com

LinkedIn : <https://www.linkedin.com/in/ayon-roy>

Website : AYON-ROY.NETLIFY.APP



Download the slides