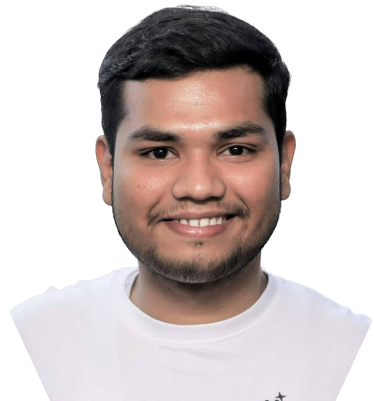


Ways to boost your Data Science Career

Date : 15th June 2023 | Speaker : Ayon Roy |

Venue : Guru Tegh Bahadur Institute of Technology, New 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

- Things to focus on, while making a Data project?
- How & Why organize your Data project?
- A Primer to Competitive Data Science



Analytics Project Life Cycle

The 5 Phases



UNDERSTAND & PLAN

Translate the business question to analytics question

DATA

Work cross functionally to gather and process data

ANALYSIS

Solve problems

PRESENTATION

Present results and tell a story

DOCUMENTATION & REFLECTION

Make your project a happy ending

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.

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.

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 & a lot more....

Be a part of as many hackathons as you can

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

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