Notes-and-Slides.zip

Open with ZIP Extractor

Chapter 06 - Slides and Notes 3 items

Name	Last modified	File size
.DS_Store	Nov 25, 2021	6 KB
6_2017_02_04101_13_35_PM.ppt	Sep 23, 2021	358 KB
Biases_in_Data_Science.pdf	Nov 22, 2021	258 KB

Ramesh Yadav Jan 13
Get well soon sir.

Add class comment...

Kumar Kisalaya posted a new material: Powerpoint Slides - Unit 5
Jan 3

Chandan Goopta
Nov 25, 2021
Please find notes and slides for your first assessment.

Notes-and-Slides.zip
Compressed Archive

Add class comment...

Chandan Goopta
Nov 23, 2021
Cognitive Biases in Data Science Slides.

ognitivebiasesindatascien...
PDF

Add class comment...

Kumar Kisalaya
Nov 22, 2021
Dear All,

This is a follow-up to our discussion around the use of training data for Unsupervised Learning.

There seems to be some confusion around this, even on the online resource, mostly because of generalized natures of the terms 'data' vs 'labeled data' and 'training' vs 'validation.
In summary, by nature, if we don't use 'labeled data', we are not really 'training' the model/algorithm. Contrast this with the fact that there could be a 'validation' of the output, which can be done either by humans or automated systems.

Please see the references below - I've included the relevant snippets as well, but I would encourage you to go through the whole article:

- *In unsupervised learning, there is no training data set and outcomes are unknown. Essentially the AI goes into the problem blind – with only its faultless logical operations to guide it. Incredible as it seems, unsupervised machine learning is the ability to solve complex problems using just the input data, and the binary on/off logic mechanisms that all computer systems are built on. No reference data at all.*
  Source: https://www.forbes.com/sites/bernardmarr/2017/03/16/supervised-v-unsupervised-machine-learning-whats-the-difference/
- *Unsupervised learning models, in contrast, work on their own to discover the inherent structure of unlabeled data. Note that they still require some human intervention for validating output variables.*
  Source: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

In addition to the above resources, I would encourage you to go through the response at the following link. Even though I don't usually like pointing to forums as reference material, the following link summarizes a lot of this discussion very nicely, please take a look.
https://softwareengineering.stackexchange.com/questions/368671/training-data-in-unsupervised-learning

Add class comment...

Kumar Kisalaya posted a new material: Presentation - Unit 4
Nov 22, 2021 (Edited Jan 3)

Chandan Goopta
Oct 1, 2021
Dear Students,

I am attaching a Google Drive link that contains the following
1. Chapter 01 - Introduction to Data Science
2. Chapter 06 - Ethical Issues in Data Science
3. Free Data Science ebooks
4. Research Papers

Please note that the folders have notes, PDFs, slides, and other related materials that will help you understand the topics in addition to what we discussed during our class.
I will keep adding more resources (eg: Research papers, notes, slides, books, etc) for your reference.

Also, other rough documents that we created during our classes are also in the folder. Use it only for reference.

Let me know if you have any questions.

Thanks,
CG

TU - Masters of Data Scien...
Google Drive Folder

Add class comment...

Kumar Kisalaya
Sep 20, 2021
Vacancy Announcement

Interested candidates can send their resumes to: ***shristi.bajracharya@ranenetwork.com***

Position: **Backend Developer (Python)**

Responsibilities

- Analyze product requirements and design/implement solutions to meet the same
- Maintain, troubleshoot and enhance large scale enterprise applications
- Evaluate application performance and determine any applicable improvements
- Participate in the evolution of our technologies and practices, occasionally pitch new technologies that you believe in, and participate in workshops to learn new skills
- Utilize various open source techgies
- Use various tools to orchestrate solutions
- Build independent scripts, web-based tools, micro-services and solutions
- Write scripts and automation using Perl/Python/Bash etc.
- Configure and manage data sources/solutions like PostgreSQL, Redis, Hadoop, Athena etc.

Desired Skills/Characteristics

- Bachelor's Degree or MS in Engineering or equivalent
- Fluency in programming in core Python on Linux using OOP concepts with focus on data analytics (Pandas, NumPy) and visualization (Bokeh, Matplotlib, Plotly)
- Experience with big data solutions, Hadoop and machine learning is highly desirable
- Experience working with web services using REST APIs
- Adequate knowledge of relational and NoSQL database systems
- Working knowledge of R and Cloud Services (AWS preferred)
- Strong attention to detail while implementing technical specifications
- Ability to work effectively with large code base using version control tools
- Ability to work and thrive in a fast-paced environment, learn rapidly and master diverse technologies and techniques
- Critical thinker, strong analytical, problem-solving and time-management skills
- Excellent verbal and written English communication skills
- Willingness to work in flexible hours overlapping with US office hours, if needed
- Good attitude and strong work ethics

Add class comment...

Kumar Kisalaya posted a new material: Notes/Slides For Unit 2 - Data Validation and Cleanup
Sep 20, 2021

1 class comment

Kumar Kisalaya
Sep 20, 2021
All,

Schedule for the next few classes.
Monday, 20th September: KK
Friday, 24th September: CG
Sunday, 27th September: KK
Monday, 28th September: KK
Friday, 1st October: CG
Sunday, 3rd October: KK
Monday, 4th October: CG

You'll be notified in case there are any changes to the schedule.

Add class comment...