Data Science Class Summary

2190513 Data Science (ICE) (2025/1)

CHULA ENGINEERING
Foundation toward Innovation

COMPUTER

Prof. Peerapon Vateekul, Ph.D.
Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University
Peerapon.v@chula.ac.th
www.cp.eng.chula.ac.th/~peerapon/

Lots of topics: Practical

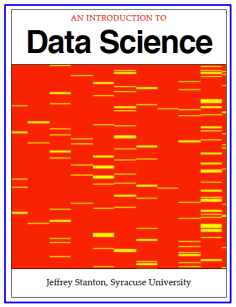
#	Mon (9AM-12PM)	Contents	Instructor	Module
1	4-Aug-25	Intro + Pandas	Aj.Peerapon	Data Sci (1)
	11-Aug-25	No class (Mother's Day)		
2	18-Aug-25	Data preparation	Aj.Peerapon	Data Sci (2)
3	25-Aug-25	ML1: Supervised: Decision Tree	Aj.Peerapon	Data Sci (3)
4	1-Sep-25	ML2: Supervised: Regression	Aj.Peerapon	Data Sci (4)
5	8-Sep-25	ML3: Supervised: Neural Networks & kNN + GridSearch	Aj.Peerapon	Data Sci (5)
6	15-Sep-25	ML4: Unsupervised ML Monitoring Tools (MLflow)	Aj.Peerapon	Data Sci (6)
	24-Sep-25	Midterm Exam Week (22 - 26 Sep) 24 Sep 2025 at 1.00PM - 4.00PM		
7	29-Sep-25	Deep Learning (Graduation Week 29 Sep - 1 Oct) - Online	Aj.Peerapon	Data Sci (7)
8	6-Oct-25	Web Scraping	Aj.Peerapon	Data Eng (1)
	13-Oct-25	No class (H.M. King Bhumibol Adulyadej's Memorial Day)		
9	20-Oct-25	API (FastAPI)	Aj.Peerapon	Data Eng (2)
10	27-Oct-25	Storage (inputs & outputs) + Airflow	Aj.Peerapon	Data Eng (3)
11	3-Nov-25	Streamlit (Web application for AI)	Aj.Peerapon	Data Viz (1)
12	10-Nov-25	BI Dashboard	Aj.Peerapon	Data Viz (2)
13	17-Nov-25	Advanced AI/ML Topics	Aj.Peerapon	Data Sci (8)
	24-Nov-25	Final Exam Week (24 Nov - 8 Dec) 24 Nov 2025 at 8.30AM - 11.30AM		

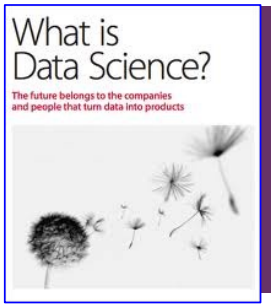
- Finding insights (Pandas)
- AI/ML
  - Data preparation (sklearn)
  - Traditional ML (sklearn)
  - DL (CNN, LSTM, TF) (pytorch)
  - Gen AI (LLM) (pytorch + LangChain)
- Web scraping (BeautifulSoup, Selenium)
- API (FastAPI)
- Visualization (Streamlit)
- Etc., e.g., Text Classification (Huggingface)

14.5. Evaluation

• Attendance	5%
• Assignment	15%
• Project	20%
• Midterm Exam	30% (Lab Test)
• Final Exam	30% (Lab Test)

What is Data Science (DS)? (aka. data analytics)

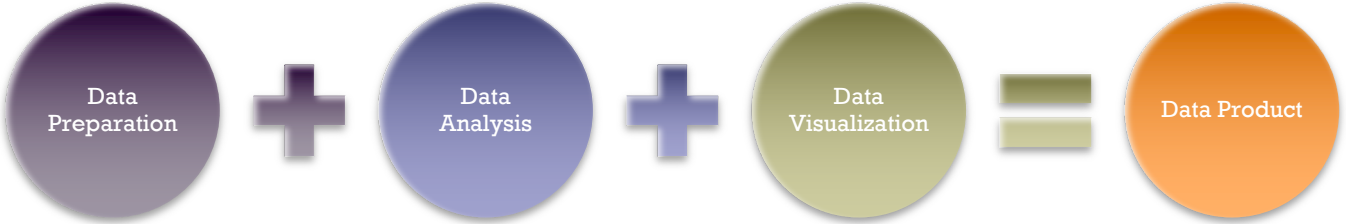




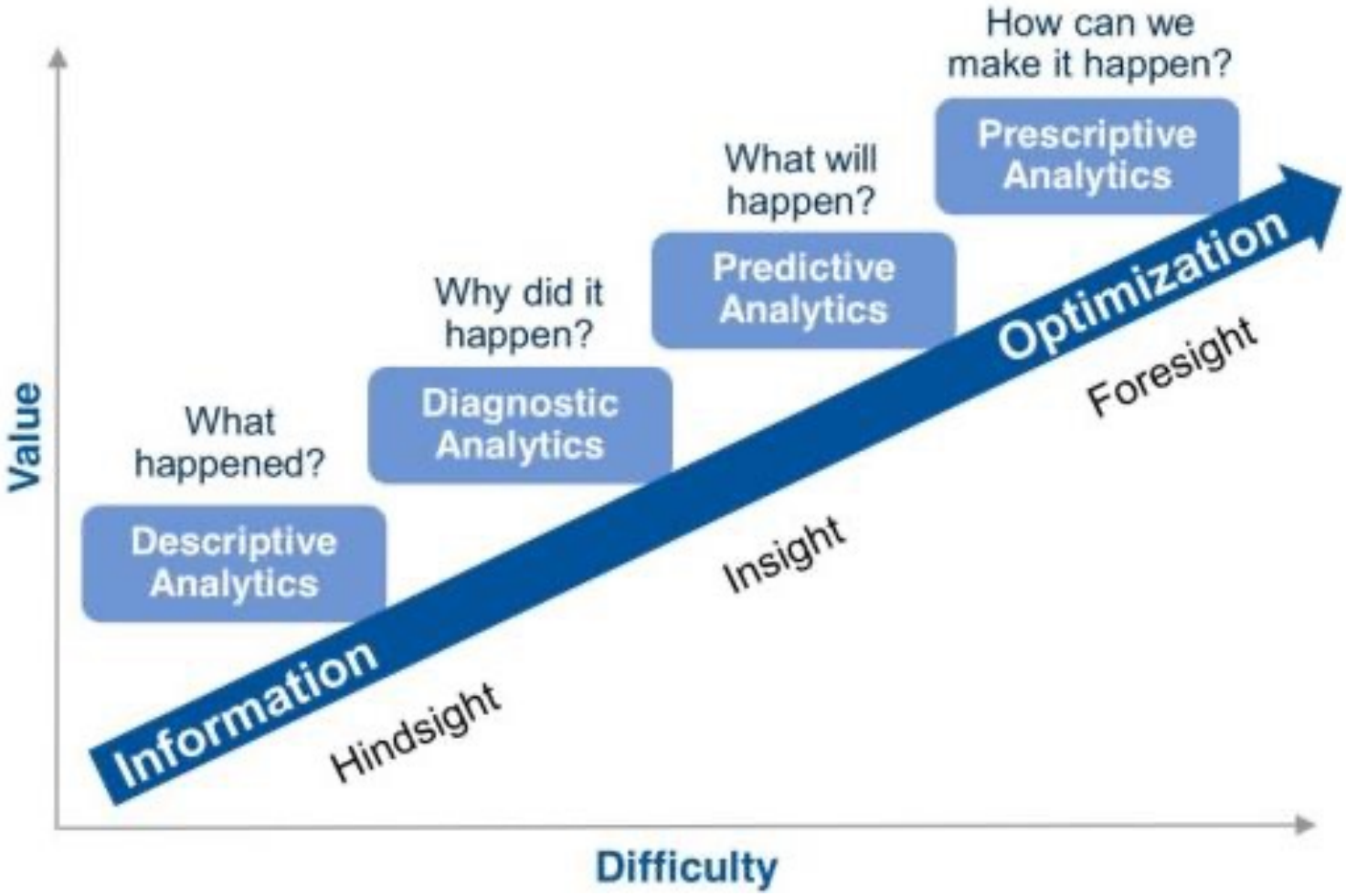
- Data
  - Facts and statistics collected for reference or analysis
- Science
  - A systematic study through observation and experiment
- Data Science
  - The scientific exploration of data to extract meaning or insight,
  - and the construction of software to utilize such insight in a business context.

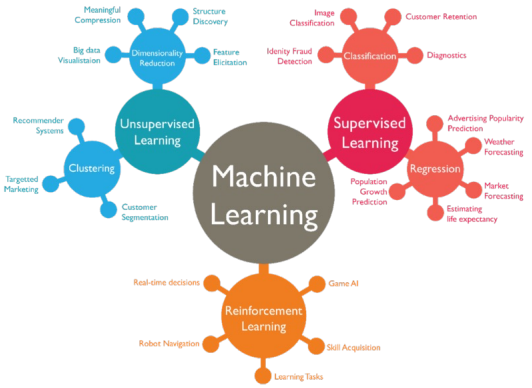
1. Transform data into valuable insights
2. Transform data into data products
3. Transform data into interesting stories

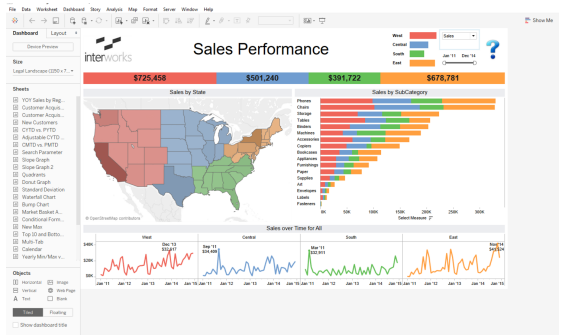
Ta Virot Chiraphadhanakul
Data Scientist, Facebook



Data Analytics (Data Science)







Types of Data Science Projects

DS != AI, but AI can be a tool in DS.

DS emphasizes a practical workflow, which is why we use coding exams to ensure that you can perform real-world industry tasks.

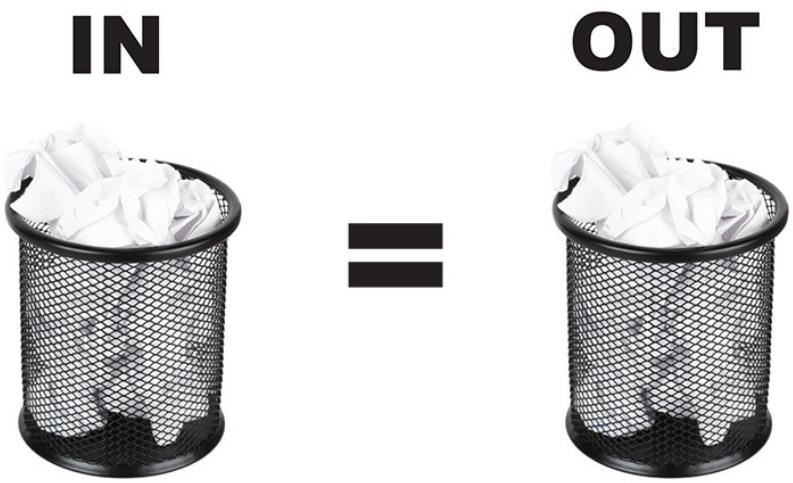
Valuable insights

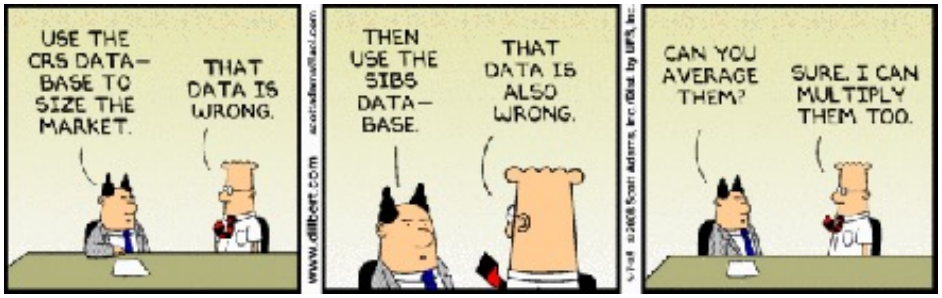
- Data visualization
- Analytical skills & storytelling
  - Infographic

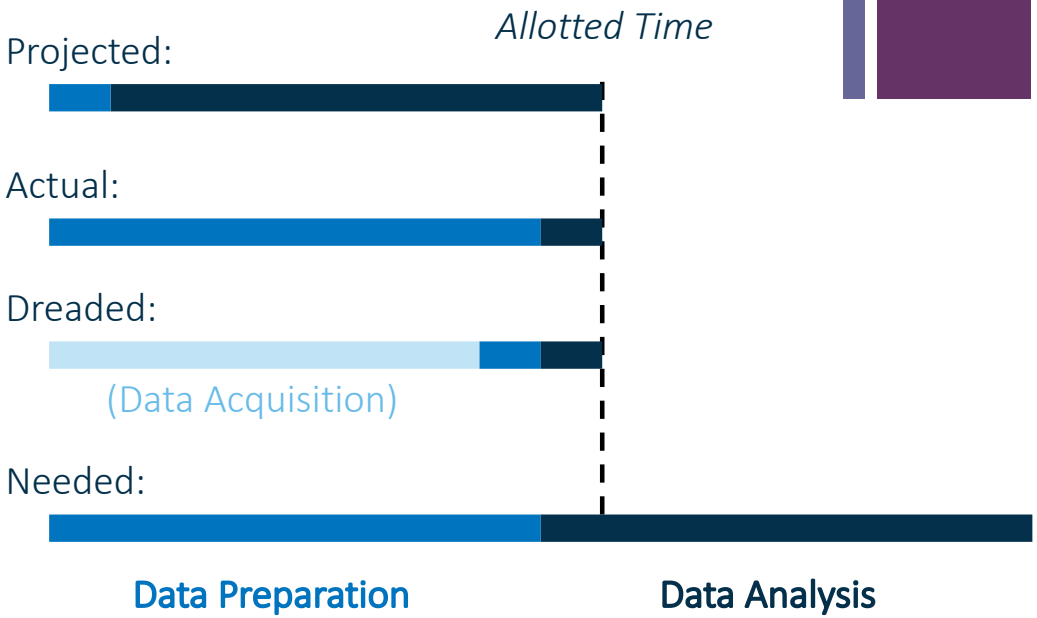
Advanced analytics

- AI/Machine Learning/Deep Learning
  - Prediction, Forecasting, Clustering, etc.

Data is the most important factor. The first thing you should present is not the model, but the data.



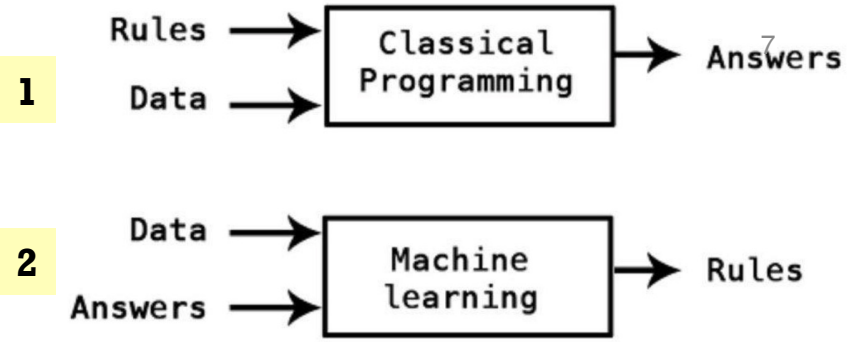


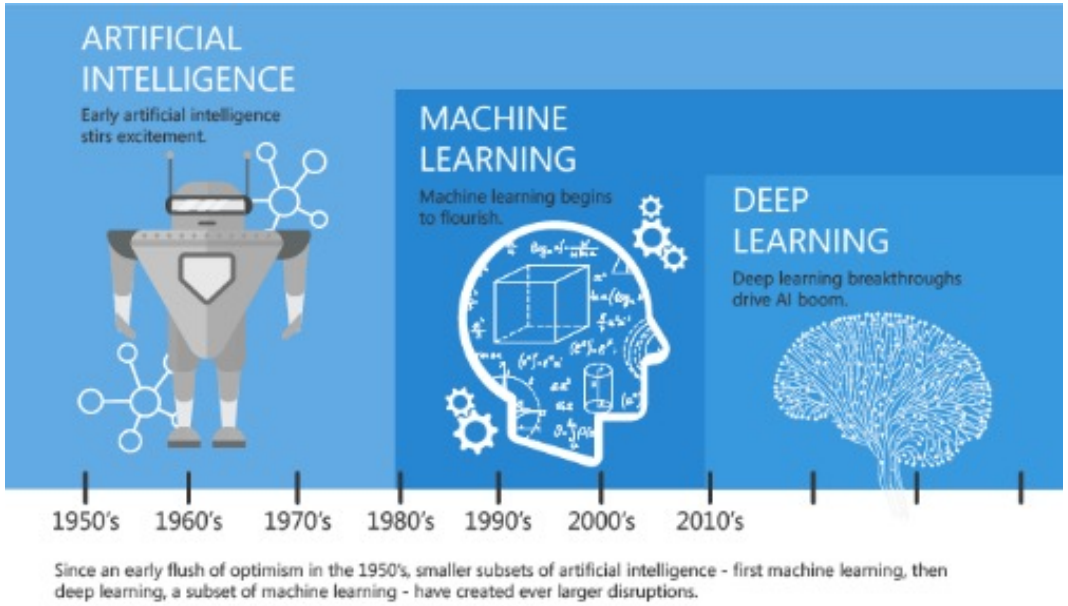


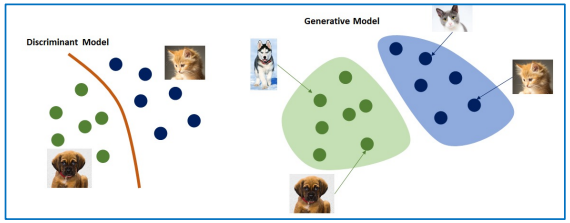
For the AI project, you must choose an appropriate technique for the problem based on the system's conditions and constraints.

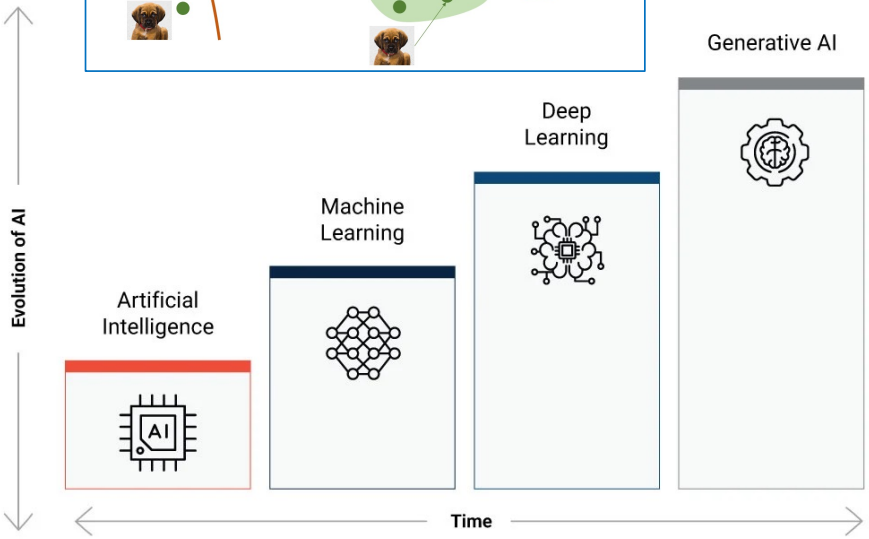
AI = Automation

- 1) Rule-based AI
- 2) Machine Learning (ML)









https://mc.ai/machine-learning-basics-artificial-intelligence-machine-learning-and-deep-learning/

Once the model is done, your job is not done yet!

Do you think that user can use this code to get the prediction result? NO!!!

2_Linear-Regression-v2.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

```
[ ]  1 coeff_df = pd.DataFrame(lm.coef_,lm.feature_names_in_,columns=['Coefficient'])
     2 coeff_df
```

Does this make sense? Probably not because I made up this data. If you want real data to repeat this sort of analysis, check out the boston dataset:

```
from sklearn.datasets import load_boston
boston = load_boston()
print(boston.DESCR)
boston_df = boston.data
```

Predictions from our Model

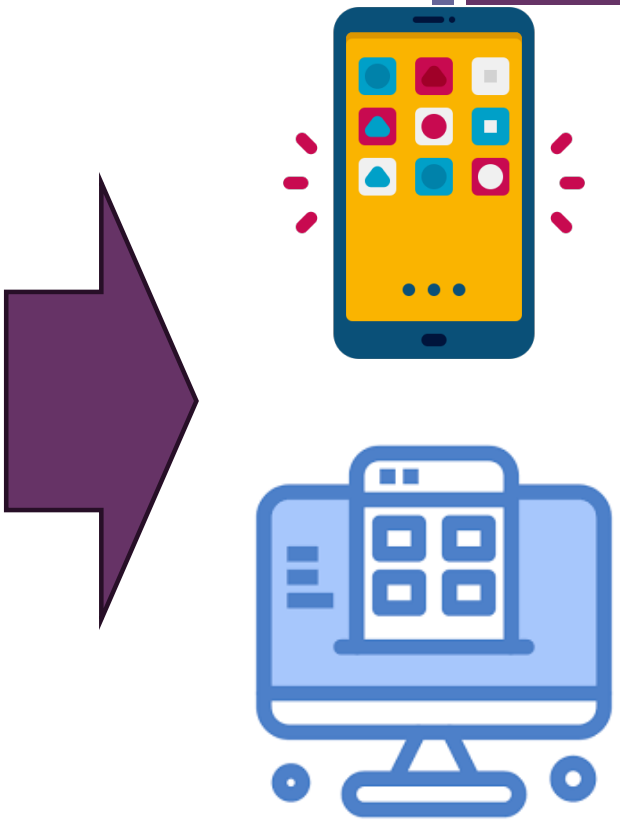
Let's grab predictions off our test set and see how well it did!

```
[ ]  1 predictions = lm.predict(X_test)
```

```
[ ]  1 plt.scatter(y_test,predictions)
```

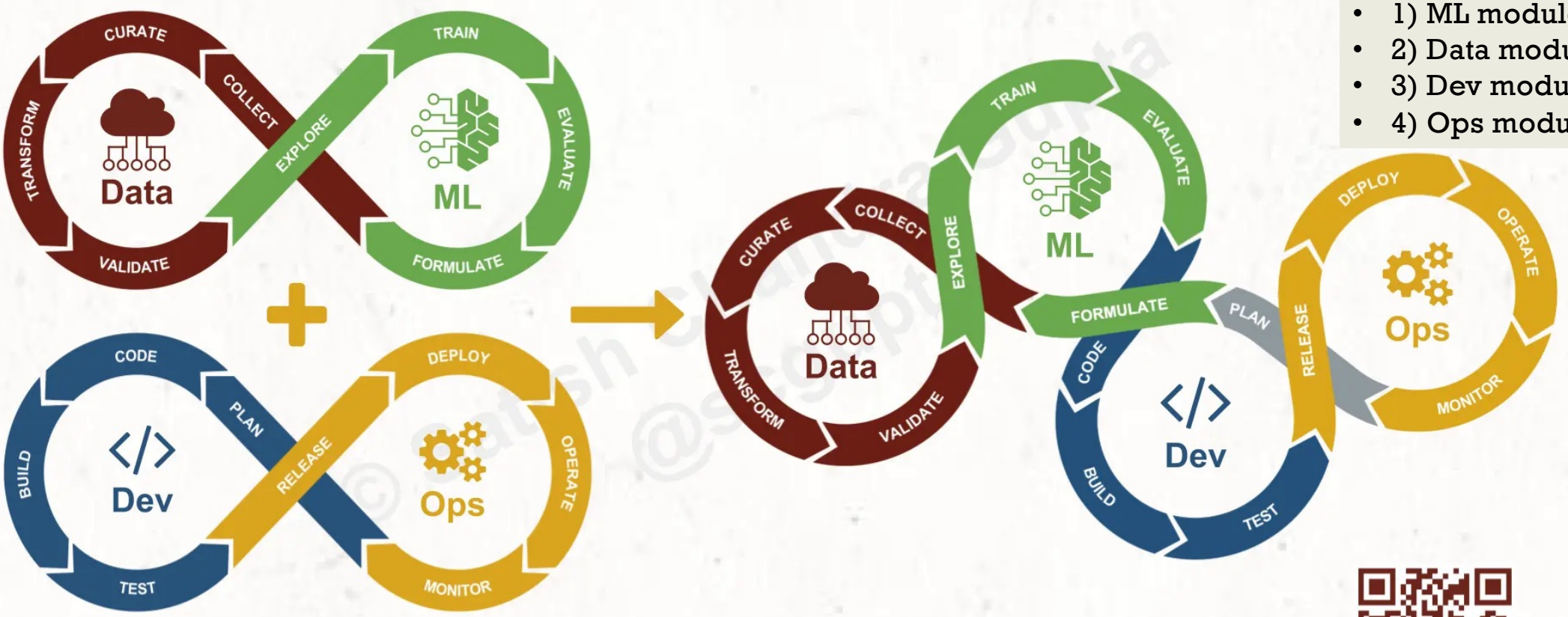
Residual Histogram

```
[ ]  1 sns.distplot((y_test-predictions),bins=50);
```



MLOps = DataML + DevOps

ml4devs.com/mlops-lifecycle



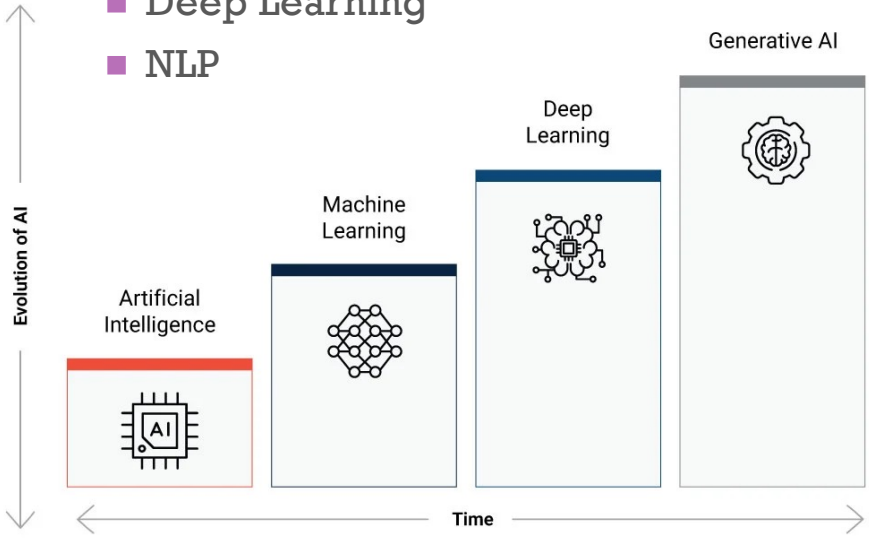
- 1) ML module
- 2) Data module
- 3) Dev module
- 4) Ops module

scgupta.me
twitter.com/scgupta
linkedin.com/in/scgupta

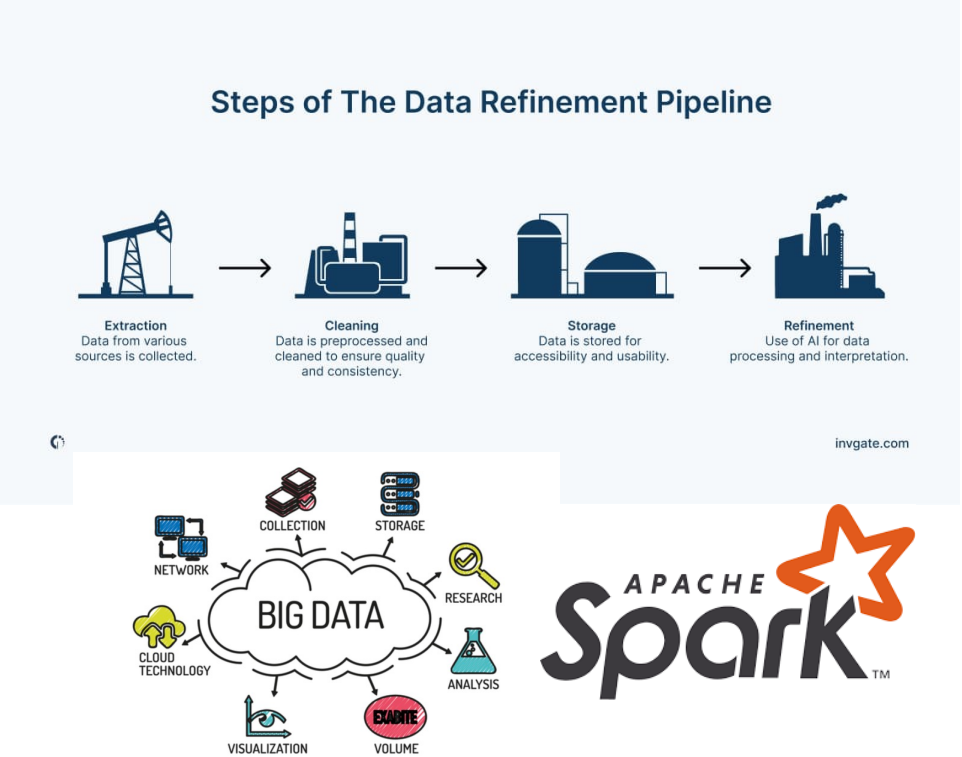
https://www.ml4devs.com/images/illustrations/ml-lifecycle-fusing-model-and-software-development.webp

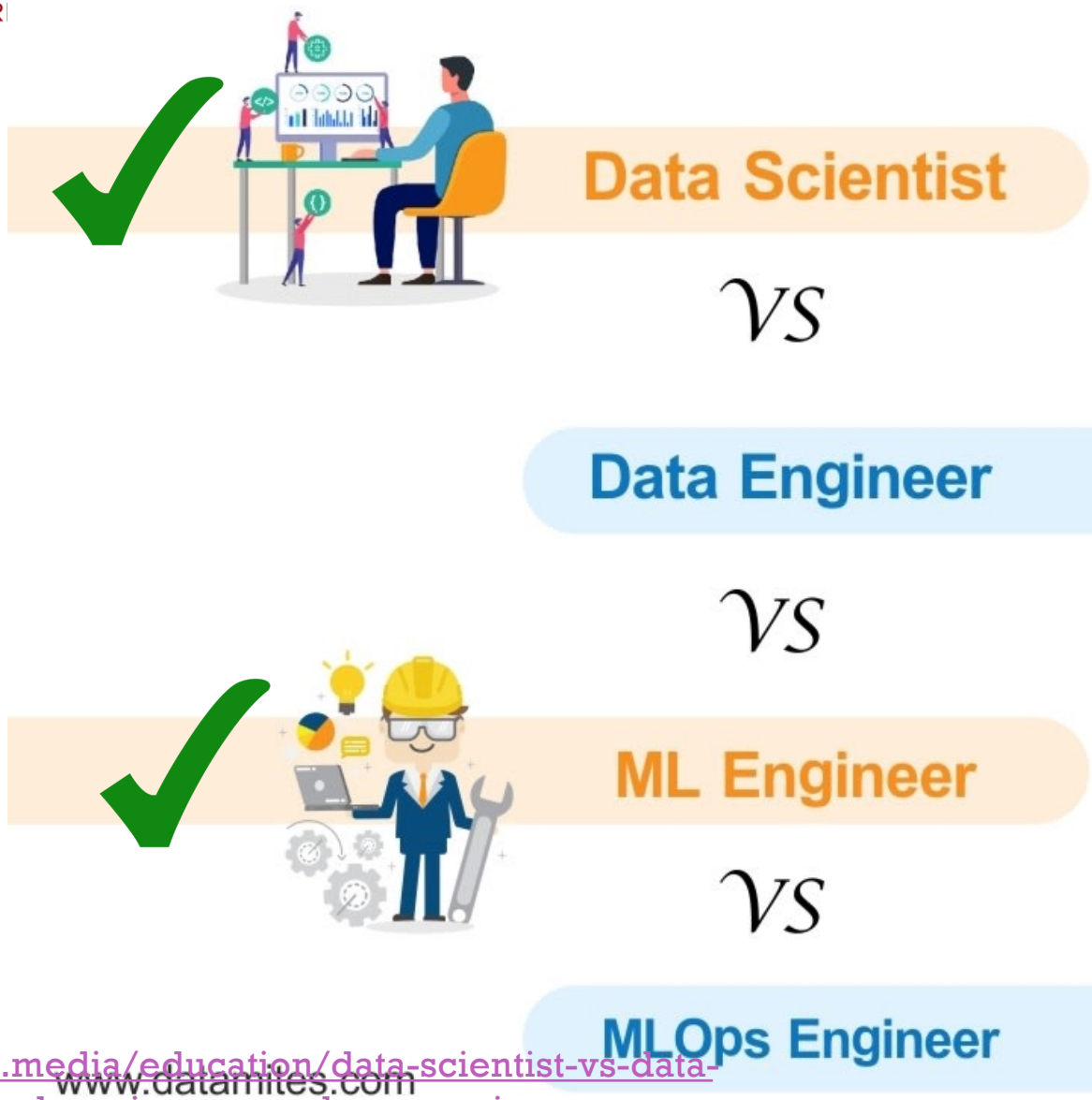
What comes after the Data Science course?

- AI course
  - Algorithms in more details
  - Search & optimization
  - Etc.
- Advanced AI coures
  - Deep Learning
  - NLP



- Data Engineering (aka., Data Warehousing)
  - Very practical





https://vocal.media/education/data-scientist-vs-data-engineer-vs-ml-engineer-vs-ml-ops-engineer

+

Good luck! ☺