AN ABSTRACTED PYTHON CLASS FOR TDSP ML PROJECTS:

EXAMPLE USE CASE WITH WINE ID

Kevin Geidel

MSDS 422: Practical Machine Learning

Northwestern University

June 2, 2024

1 Executive summary

At some point ill cite this (”What is the Team Data Science Process?”, 2024) or even this (Hyatt, 2024)!!

2 Research design

-Can we abstract datamining to automate model evaluation and selection? -In a big MS thesis like data sync design where you use ORM youll want a custom pipeline toolkit -Wine defect prediction use case

3 Exploratory Data Analysis (EDA)

4 Datamining data models

5 Data pipeline and ML preprocessing

6 ML engineering, evaluation and deployment

- data transformations!

7 Findings

8 Conclusions

Appendix
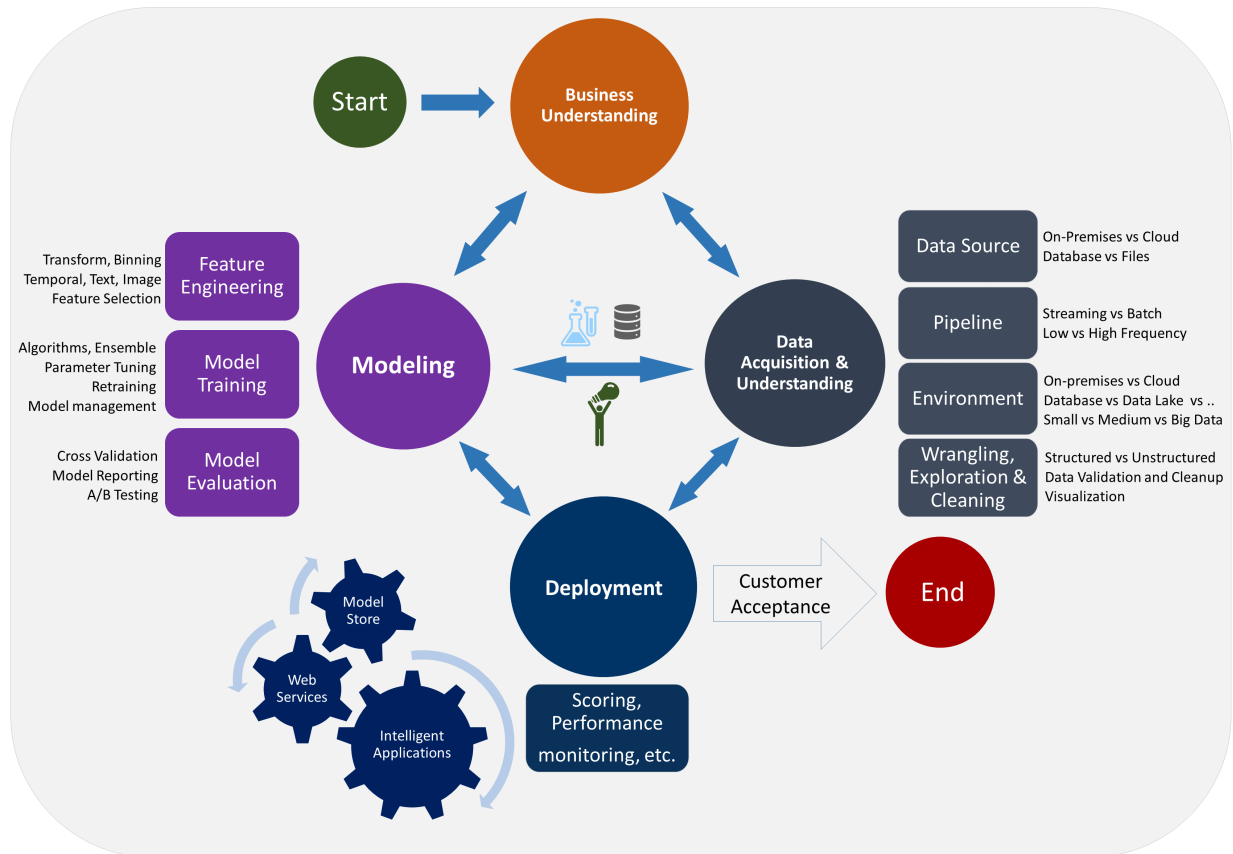
Data Science Lifecycle

Figure 1: (color online) The "data lifecycle" in Team Data Science Process (TDSP)

References

Hyatt, J. (2024). Django + postgres: A data science juggernaut. Master's thesis, Syracuse University.

"What is the Team Data Science Process?" (2024). Azure Architecture Center. `https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview`.