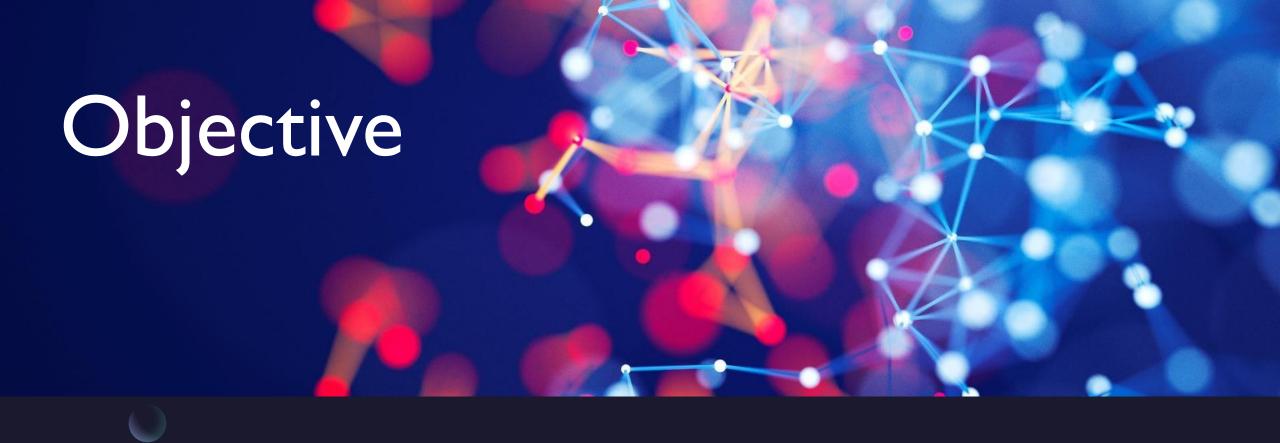


CASE STUDY II

Aires Miguens

Ms. Data Science



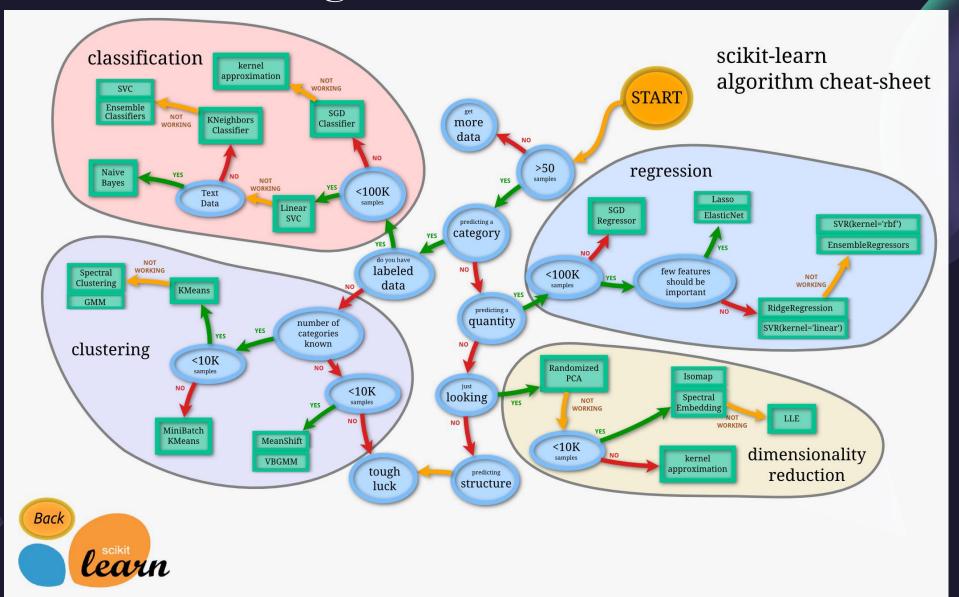
How much data is enough data?

Using Machine Learning algorithm to determine users' location (building and floor number) by analyzing WLAN signal strengths.

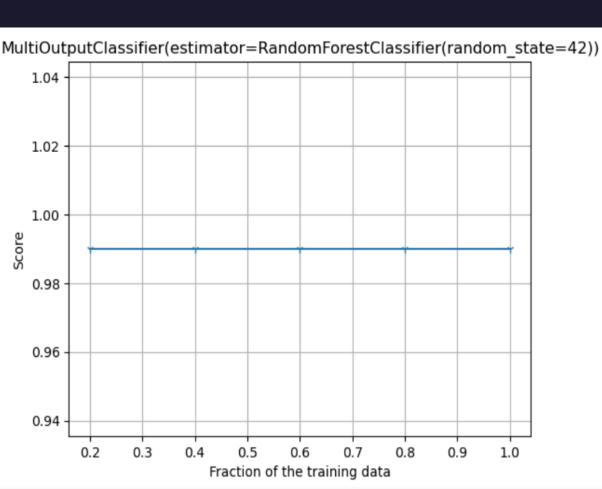


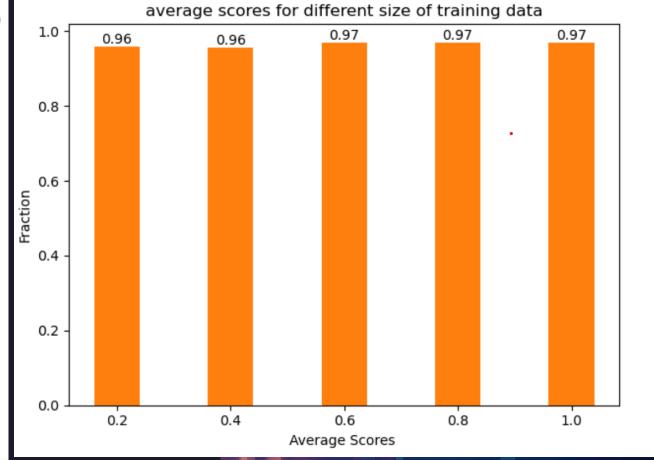
- Classification algorithm selections
- SK-Learn, numpy, matplotlib libraries.
- Pandas for cleaning and storing the data.
- Data analysis

What is the right Estimator?

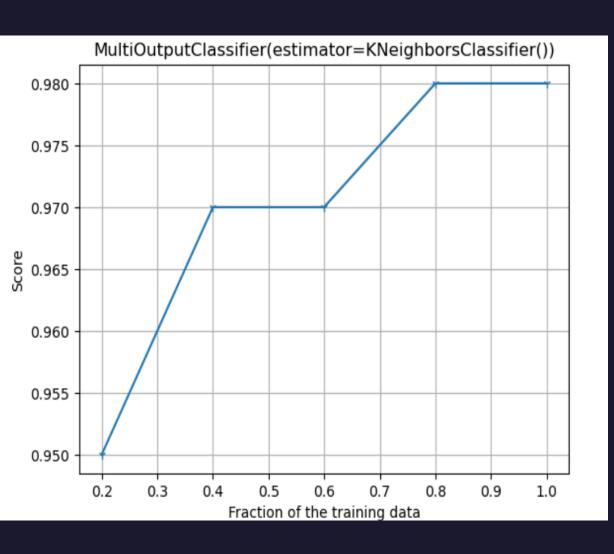


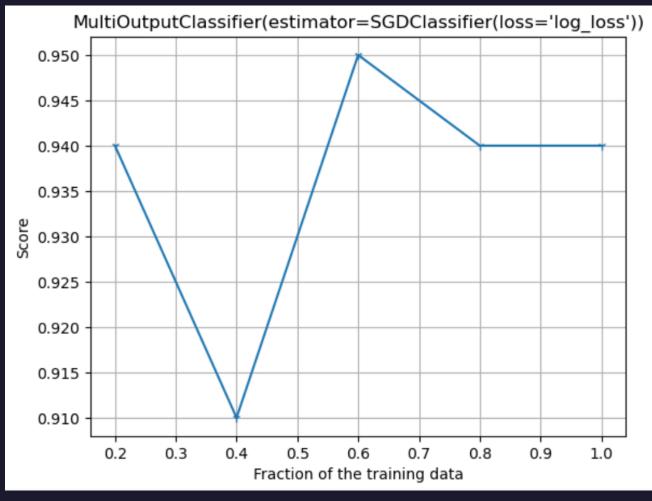
Adding more data vs performance





Adding more data vs performance





Challenges

- Equalizing the testing size to the validation size
- Selecting the appropriate Estimator
- Working on the features of the dataset

Conclusions

More training data does not always imply better score;

It may depend on multiple factors like

- The type of estimator
- The type of data (if it is adding noise or not and how the models fits the data)

It is always good practice to use more than one estimator in order to make a more informed decision.

Thank you

Aires miguens

303-725-8668

amiguns@wpi.edu

