Introduction to Product List Sorting

Dânia Meira - Head of #datalift @ AI Guild

29 September 2021

About

Dânia Meira

Head of #datalift and
founding member at AI Guild

- Career in data since 2012
- ML models for predictive analytics in marketing and forecasting applications
- Former bootcamp teacher

@MeiraDania

About

The **AI Guild** is the go-to community for data and business professionals advancing AI adoption

- 1000+ members since 2019

- #datacareer: career development by practitioners, for practitioners

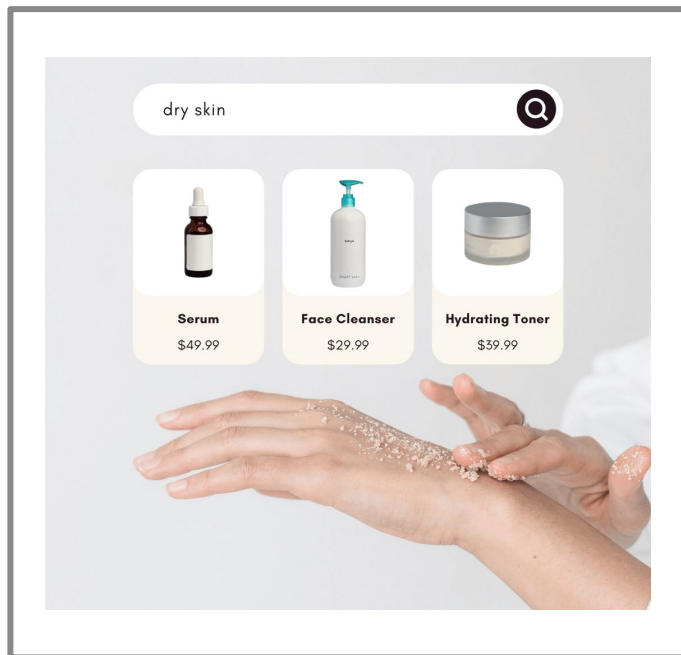- #datalift: deploy data analytics and machine learning

www.theguild.ai

Agenda

o Introduction to Product List Sorting (PLS)

o Using ML to create a PLS

o Measuring the ML model performance

o Conclusion

Introduction to Product List Sorting (PLS)

eCommerce - how do users to find what they need?

Search + Filters
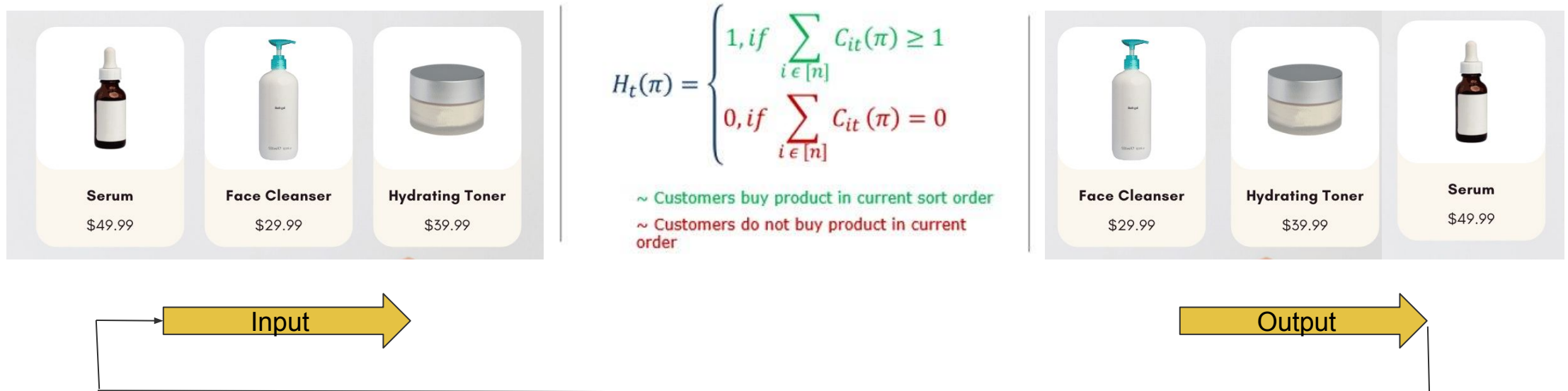
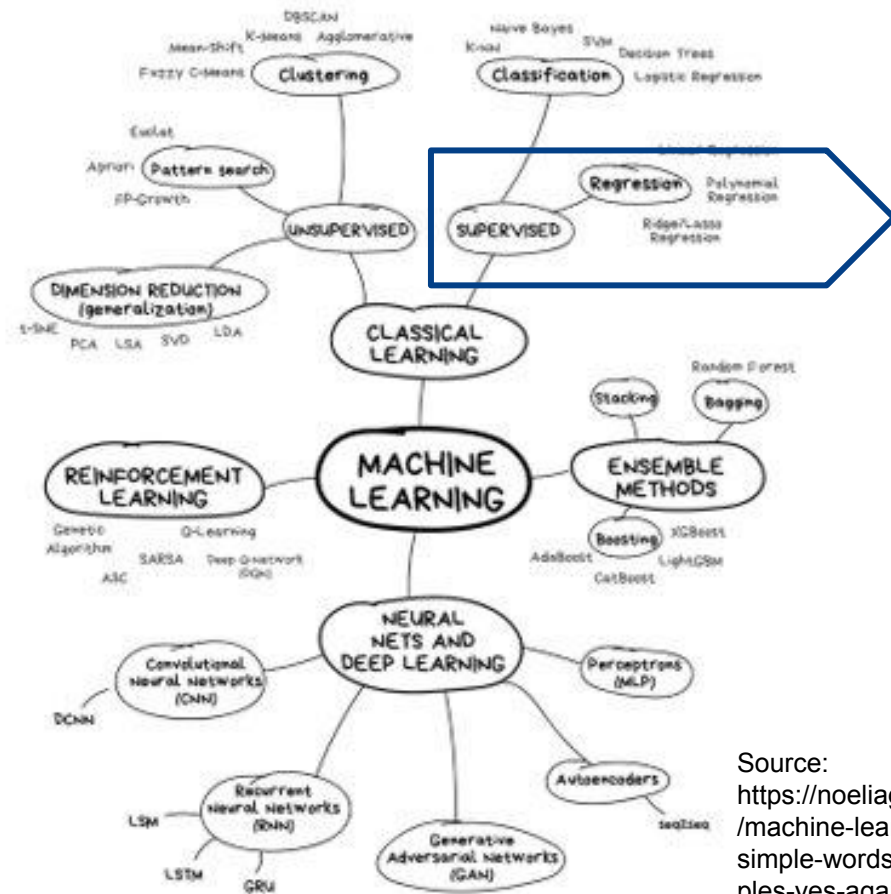**Recommendations**

Categories browsing

Using ML to create a PLS (1/3)

How to use Machine Learning for different contexts?

The statistical algorithm determines the expected purchase probabilities for a product and sorts it according to these probabilities. This probabilities can be used for training a supervised ML algorithm

$$H_t(\pi) = \begin{cases} 1, if \sum_{i \in [n]} C_{it}(\pi) \geq 1 \\ 0, if \sum_{i \in [n]} C_{it}(\pi) = 0 \end{cases}$$

~ Customers buy product in current sort order
~ Customers do not buy product in current order

Input

Output

Using ML to create a PLS (2/3)

How supervised learning works:

Training set

Product	Brand	Price	Conversion Rate
	A	49.99	3.4%
	B	29.99	1.7%

Test set

Product	Brand	Price	Conversion Rate
	C	39.99	?

Source: https://noeliagorod.com/2019/05/21/machine-learning-for-everyone-in-simple-words-with-real-world-examples-yes-again/

ML model

Using ML to create a PLS (3/3)

Linear **Regression Example**

Training set

Product	Brand	Price	Conversion Rate
<image: dropper bottle>	A	49.99	3.4%
<image: pump bottle>	B	29.99	1.7%

Input, or features, or X   Target, or label, or Y

$$cvr = (a_1 * brand) + (a_2 * price)$$

- "Brand" is a category -> needs to be transformed to a numerical
- "Price" can be directly used because it is already numerical

Measuring the ML model performance

You don't want to know if the model predicted the value exactly

You want to know how close the predictions were to the real values, on average.

One common metric used to measure performance of a regression model is Mean Absolute Error (MAE)
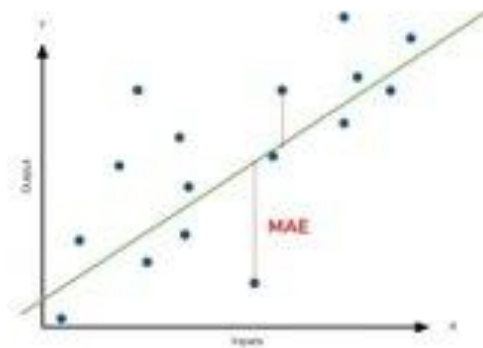
$$\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i|.$$

Conclusion

We learned
… there are many approaches to machine learning

… why it is important to split datasets into training and test sets

… how to analyse data in Python

… how to create a ranking in Python

… how to measure the success of your model

What else you can do
Conversion Rate might be an indicator for the „best" ranking for customers- but is it the best for your company?

Improve your model by using a more company oriented target

Use more features, like customer information

Add a step to cluster customers before you show the Product list

Try different algorithms

Talk to the business

```python
if questions:
    try:
        answer()
    except RuntimeError:
        pass
else:
    print("Thank You.")
```