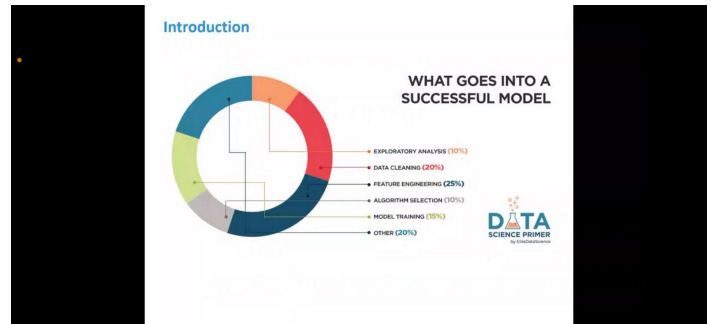


## Lec 4

26 сентября 2024 г. 19:16

1. Feature engineering - synthesizing features on other features of data
2. Some companies don't select algos and develop their own like yandex, nvidia, google
3. Supervised learning - ML on labeled data.
4. Unsupervised learning - ML on unlabeled data. For example, clusterization
5. Semi-supervised. Sometimes model is allowed to label part of data itself, sometimes there is another model that labels data
6. Reinforcement learning. Like a game. Model is acting in a way to maximize reward
7. Clean data is better than cool algo
8. Lack of data can be partially solved by augmentation(same data but with minor changes)
9. Wtf is target variable

The target variable is the variable whose values are modeled and predicted by other variables. A predictor variable is a variable whose values will be used to predict the value of the target variable.



### Machine Learning Tasks

- Supervised Learning
- Unsupervised Learning
- + some special cases
  - ▶ Semi-supervised Learning
  - ▶ Reinforcement Learning
  - ▶ ...
- Major Issues:
  - ▶ Garbage In = Garbage Out
  - ▶ Overfitting (aka Noise Memorization)
    - ▶▶ Model should generalize (in general ☺)

### Additional steps

- Project Scoping
  - ▶ Sometimes, you'll need to roadmap the project and anticipate data needs.
- Data Wrangling
  - ▶ You may also need to restructure your dataset into a format that algorithms can handle.
- Preprocessing
  - ▶ Often, transforming your features first can further improve performance.
- Ensembling
  - ▶ You can squeeze out even more performance by combining multiple models.

### Exploratory Analysis: Basics

- How many observations do I have?
- How many features?
- What are the data types of my features? Are they numeric? Categorical?
- Do I have a target variable?