# Introduction to Machine Learning

BUSINESS OBJECTIVES, REQUIREMENTS AND DEVELOPMENT PROCESS

STUDENT: PAULO EDUARDO DA SILVA JUNIOR

# Summary

1. Business Objectives and Machine Learning.
2. Requirements for Machine Learning Systems.
3. Iterative Development Process.
4. Defining Machine Learning Problems.
5. Objective Functions.
6. Mind vs. Data in Machine Learning.
7. Conclusion.

# Business Objectives and Machine Learning

1. Impact on business metrics such as revenue and customer retention.
2. Focus on business metrics beyond model accuracy.
3. Validation through A/B testing and impact on key indicators.

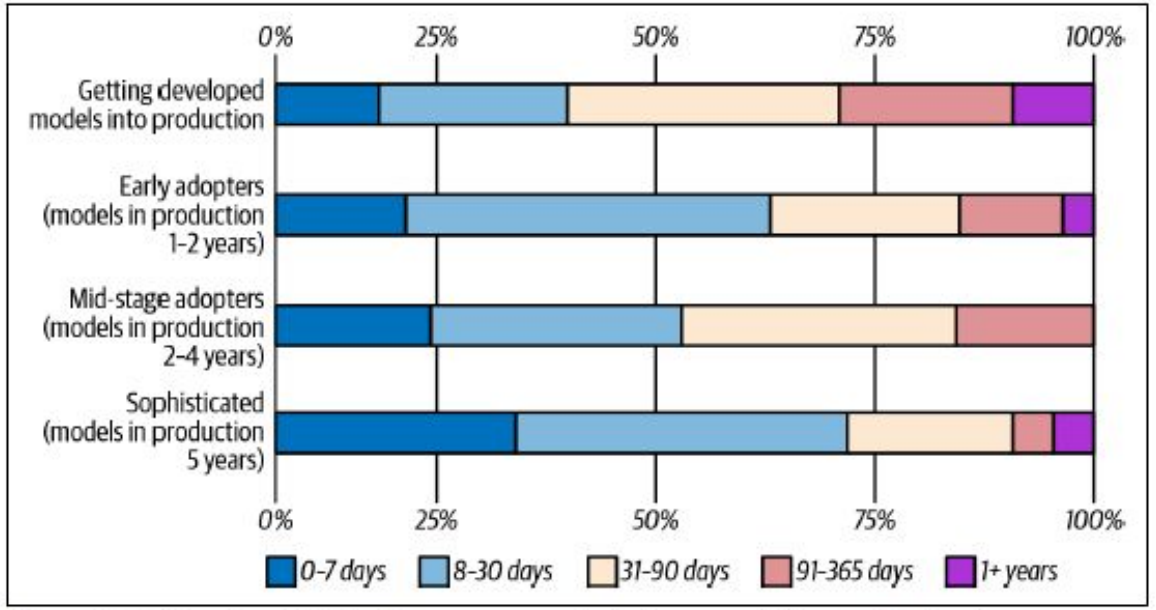


*Figure 2-1. How long it takes for a company to bring a model to production is proportional to how long it has used ML. Source: Adapted from an image by Algorithmia*

# Requirements for Machine Learning Systems

Requirements:

- Reliability
- Scalability
- Maintainability
- Adaptability

# Iterative Development Process

1. Project definition, data engineering, model development.
2. Implementation, ongoing monitoring and adjustments.
3. Coordination and alignment with business objectives.

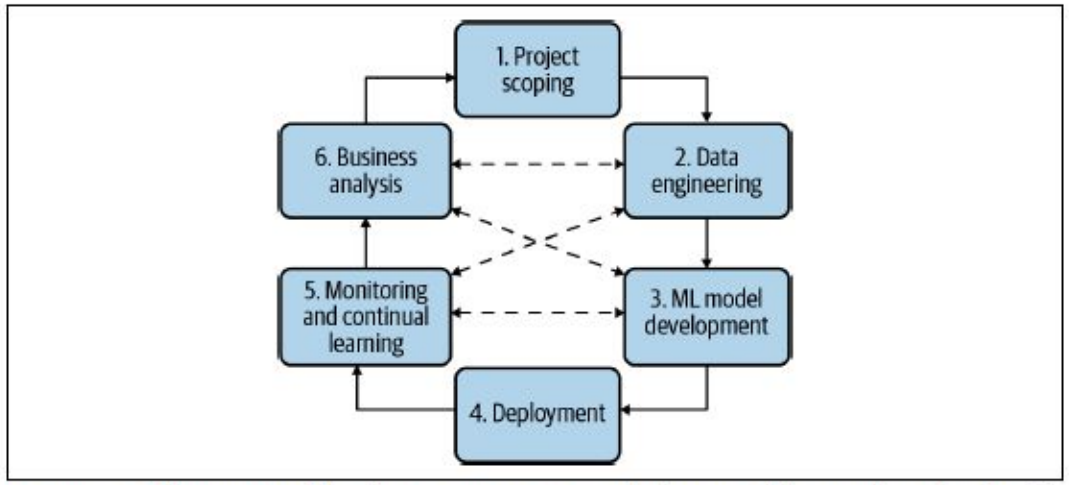


*Figure 2-2. The process of developing an ML system looks more like a cycle with a lot of back and forth between steps*

# Defining Machine Learning Problems

1. Specifying inputs, outputs, and objective function.
2. Difference between classification and regression.
3. Simplifying problems and choosing approaches.

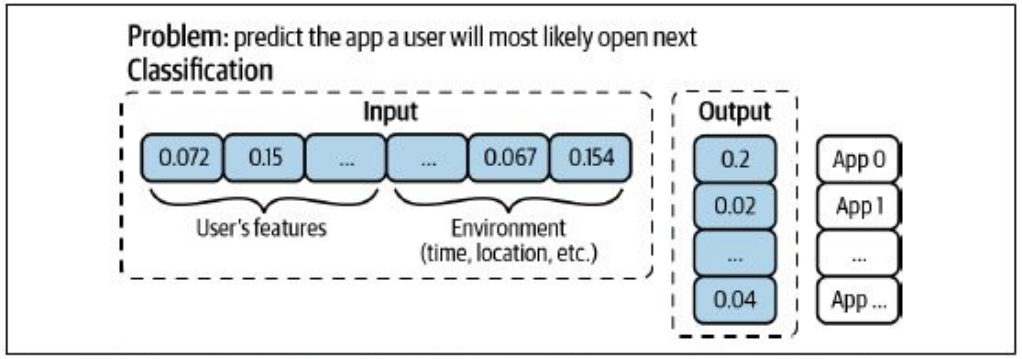


*Figure 2-5. Given the problem of predicting the app a user will most likely open next, you can frame it as a classification problem. The input is the user's features and environment's features. The output is a distribution over all apps on the phone.*

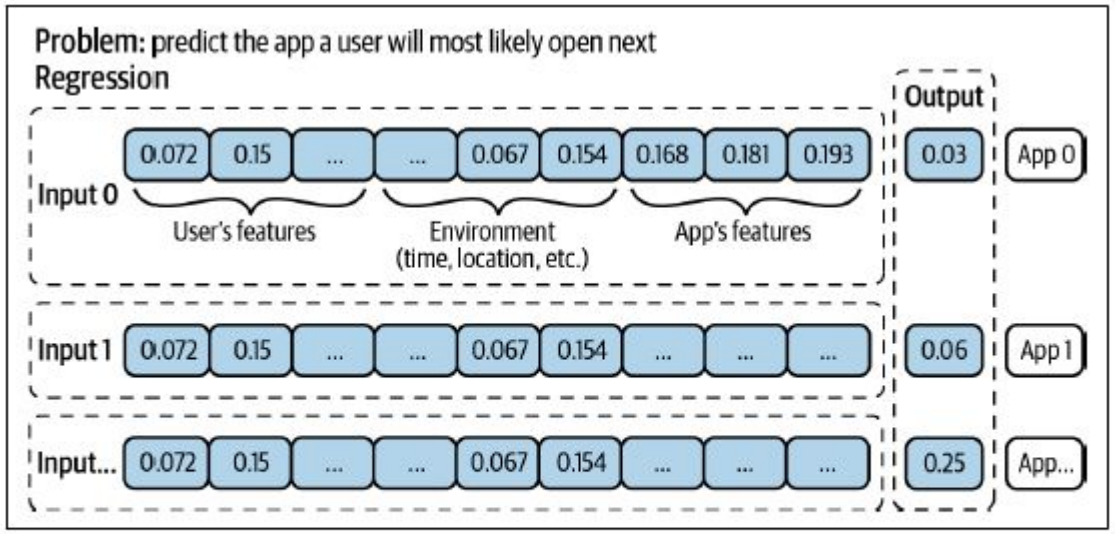


*Figure 2-6. Given the problem of predicting the app a user will most likely open next, you can frame it as a regression problem. The input is the user's features, environment's features, and an app's features. The output is a single value between 0 and 1 denoting how likely the user will be to open the app given the context.*

# Objective Functions

1. Guide learning and minimize errors.
2. Separate multiple objectives into distinct models.
3. Flexibility for specific adjustments.
4. You want to achieve this goal through the following three objectives:
    1. Filter out spam;
    2. Filter out NSFW content;
    3. Filter out misinformation;
    4. Rank posts by quality;
    5. Rank posts by engagement: how likely users will click on it;

```python
import numpy as np

def cross_entropy(p, q):
return -sum([p[i] * np.log(q[i]) for i in range(len(p))])

p = [0, 0, 0, 1]
q = [0.45, 0.2, 0.02, 0.33]
cross_entropy(p, q)
```

*quality_model*
Minimizes *quality_loss* and outputs the predicted quality of each post

*engagement_model*
Minimizes *engagement_loss* and outputs the predicted number of clicks of each post

# Mind vs. Data in Machine Learning

1. Balancing sophisticated algorithms with data volume.
2. Data as a driver of recent advances.
3. Larger models and more data = greater effectiveness.

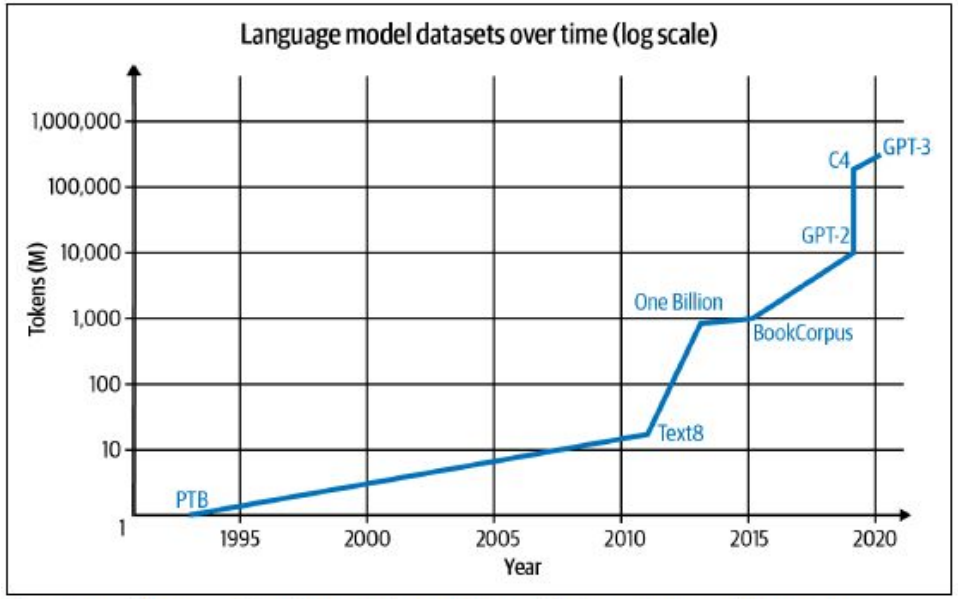


*Figure 2-8. The size of the datasets (log scale) used for language models over time*

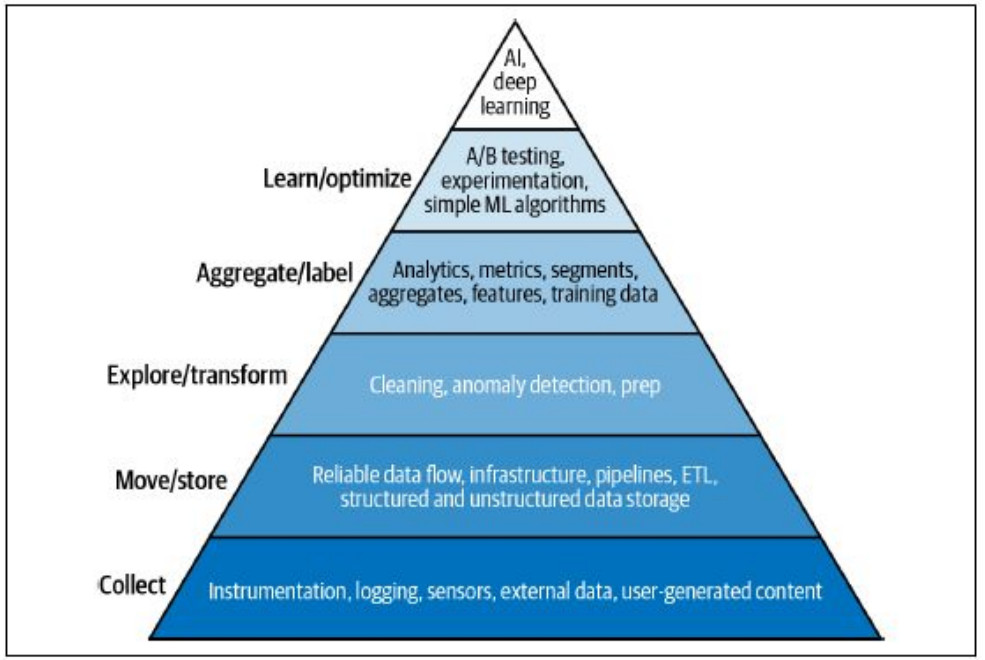


*Figure 2-7. The data science hierarchy of needs. Source: Adapted from an image by Monica Rogati[22]*

# Conclusion

1. Importance of clear business objectives.
2. Essential requirements: reliability, scalability, maintainability, adaptability.
3. Iterative process, adjusting to changes.
4. Next step: data engineering.