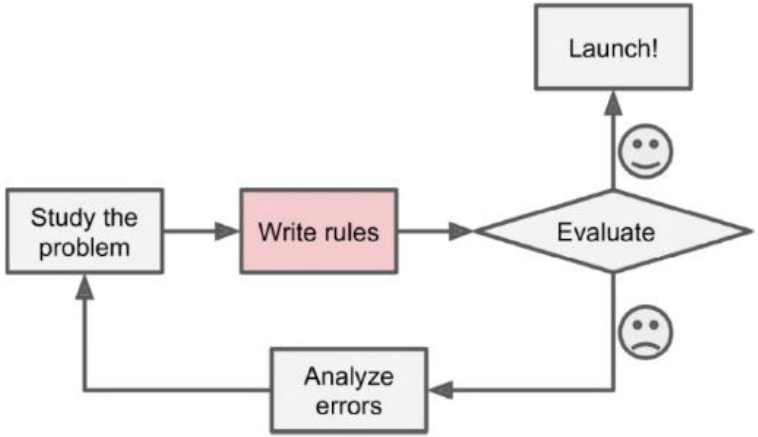
MACHINE LEARNING

CLASS 3

Why Use Machine Learning?

Let's imagine that we have to create a spam filter.

1. Identify patterns: Words like Credit card, free, amazing, 4U.
2. Write detection algorithm
3. Repeat step 1 and 2 until it is good enough

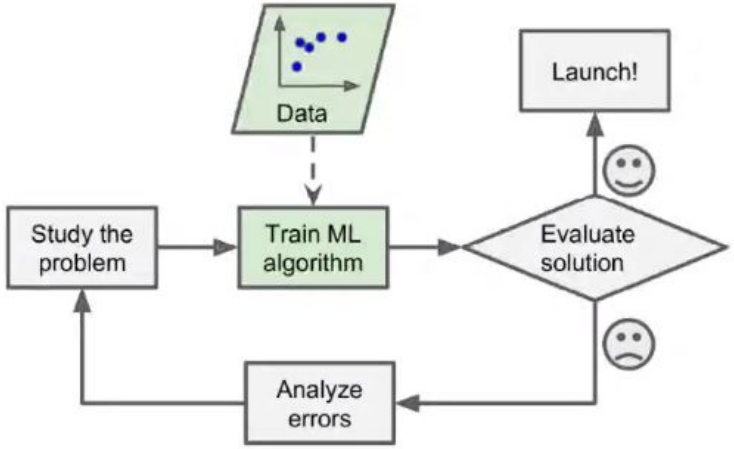


What happen if the spammer detects that emails with “4U” are being blocked and change it to “For U”?

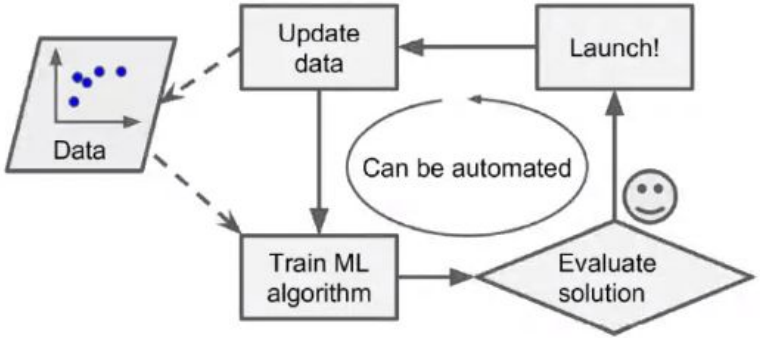
Why Use Machine Learning?

ML automatically learns which words and phrases are good predictors of spam.

The program is much shorter, easier to maintain, and more accurate.

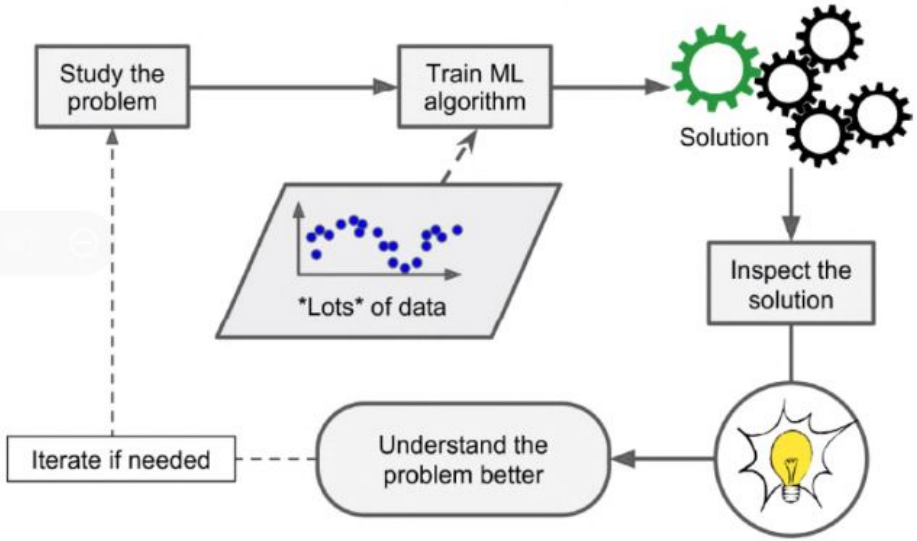


What happen if the spammer detects that emails with “4U” are being blocked and change it to “For U”?



What is ML good at?

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules
- Complex problems for which there is no good solution at all using a traditional approach (Speech Recognition)
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Help humans to learn: Getting insights about complex problems and large amounts of data



Types of Machine Learning

Classification based on:

- Whether or not they are trained with human supervision
- Whether or not they can learn incrementally on the fly
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do

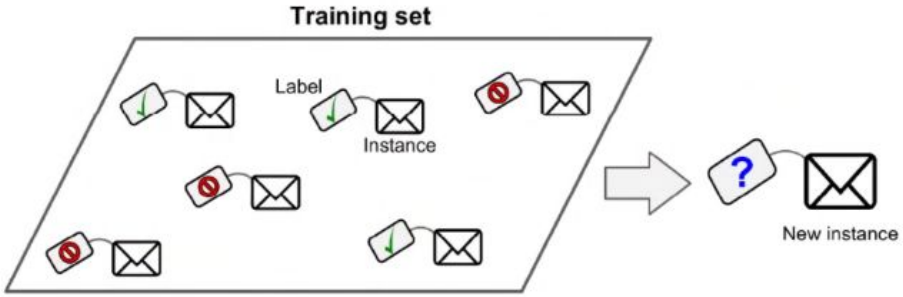
- supervised, unsupervised, semisupervised, and Reinforcement Learning
- online versus batch learning
- instance-based versus model-based learning

Supervised vs Unsupervised Machine Learning

moji

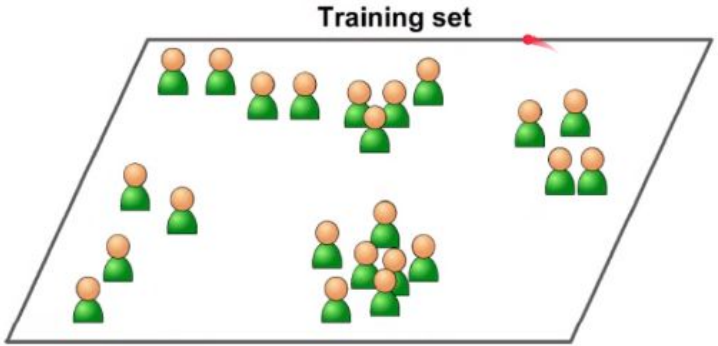
Supervised

The training data you feed to the algorithm includes the desired solutions. Called labels.



Unsupervised

The training data is unlabeled. The system tries to learn without a teacher.



mojix

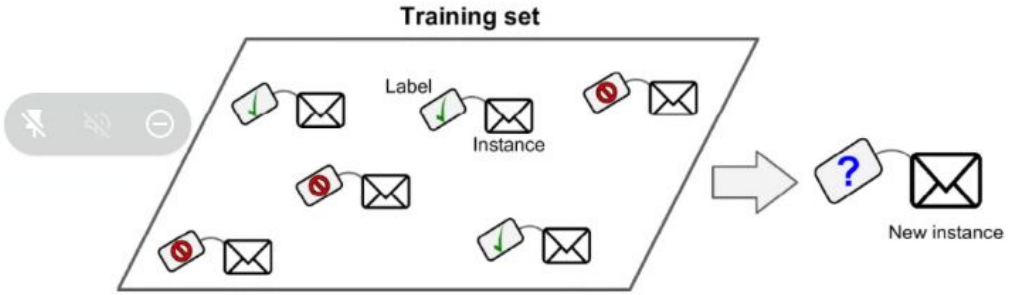
Supervised Learning

How it works?

mojix

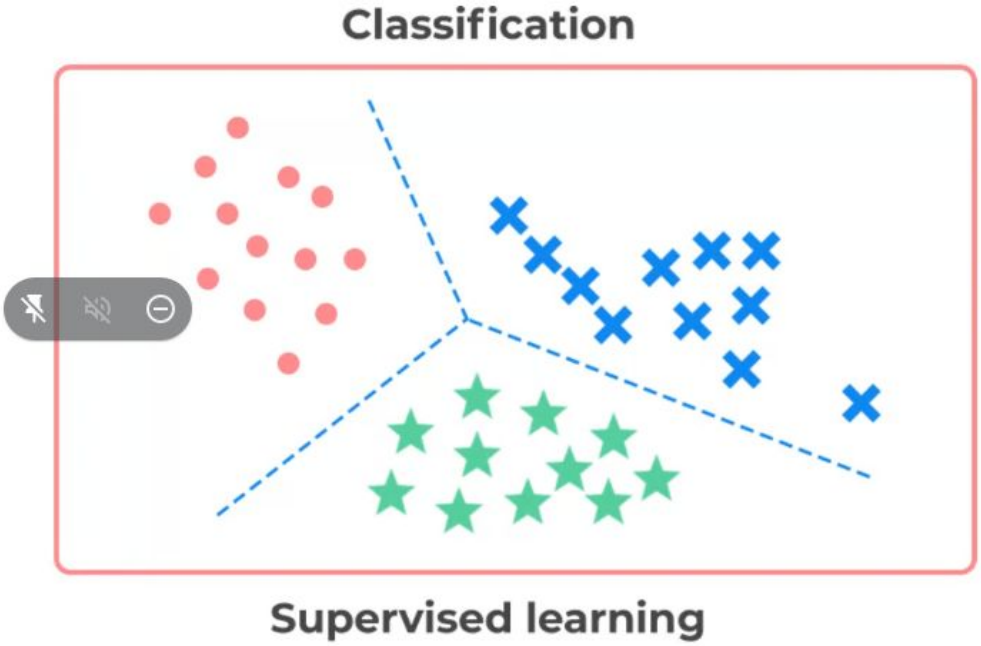
From the training data, the algorithm will learn the characteristics that help to identify a specific label.

The next time that a sample is given to the model. It will identify the label based on what it learnt.



Common models

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks*



Main Challenges of ML

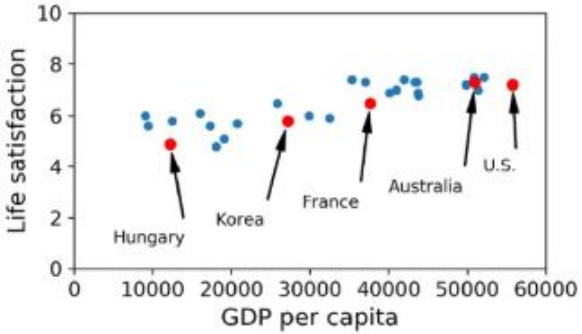
Since your main task is to select a learning algorithm and train it on some data. Two things that can go wrong

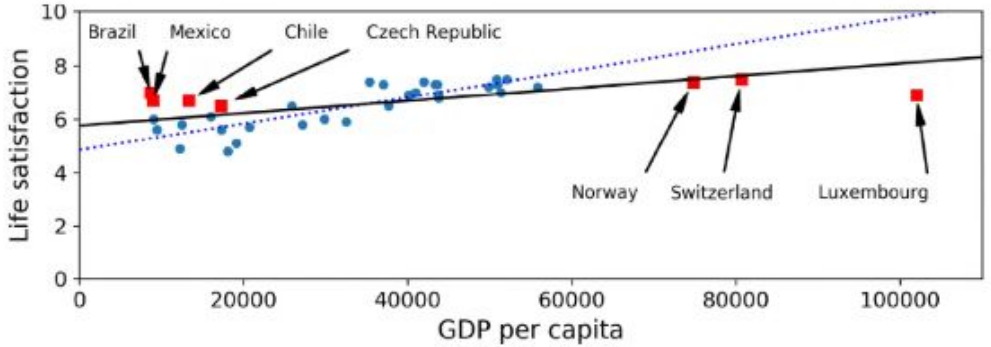
- "bad algorithm"
- "bad data"

Bad Data

mojix

- Insufficient Quantity of Training Data
- Nonrepresentative Training Data
- Poor-Quality Data
- Irrelevant Features
  - Feature selection
  - Feature extraction
  - Creating new features



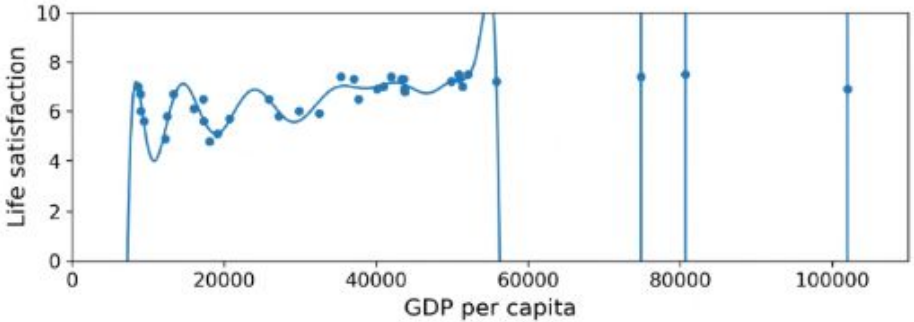


Bad Algorithms

mojix

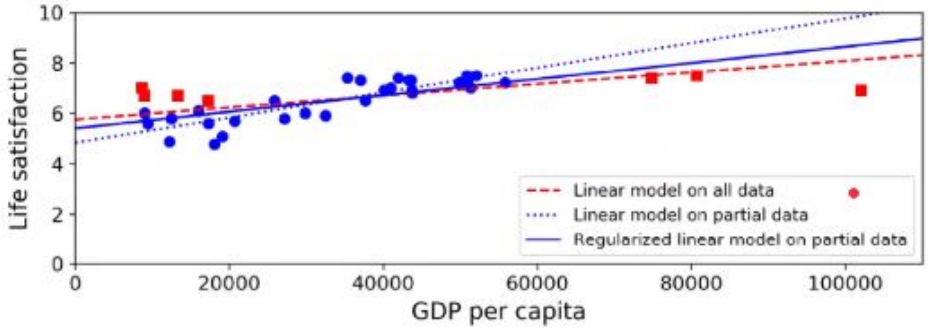
Overfitting the Training Data

Overfitting happens when the model is too complex relative to the amount and noisiness of the training data.



How to fix?

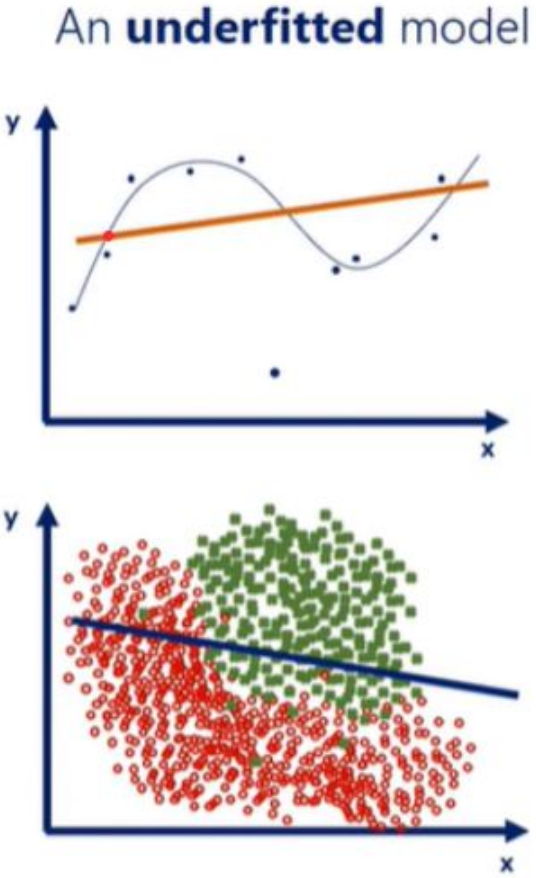
- To simplify the model by selecting one with fewer parameters by reducing the number of attributes in the training data or by constraining the model
- To gather more training data
- To reduce the noise in the training data



Bad Algorithms

Underfitting the Training Data

It occurs when your model is too simple to learn the underlying structure of the data.



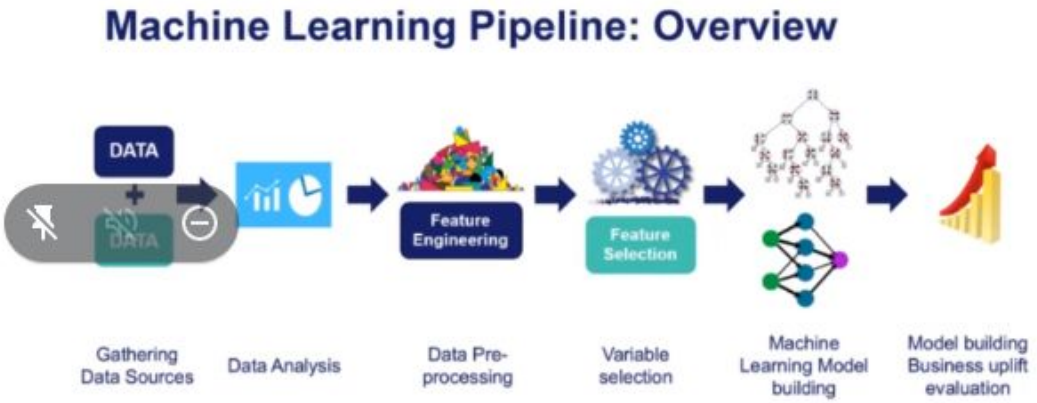
Pipeline

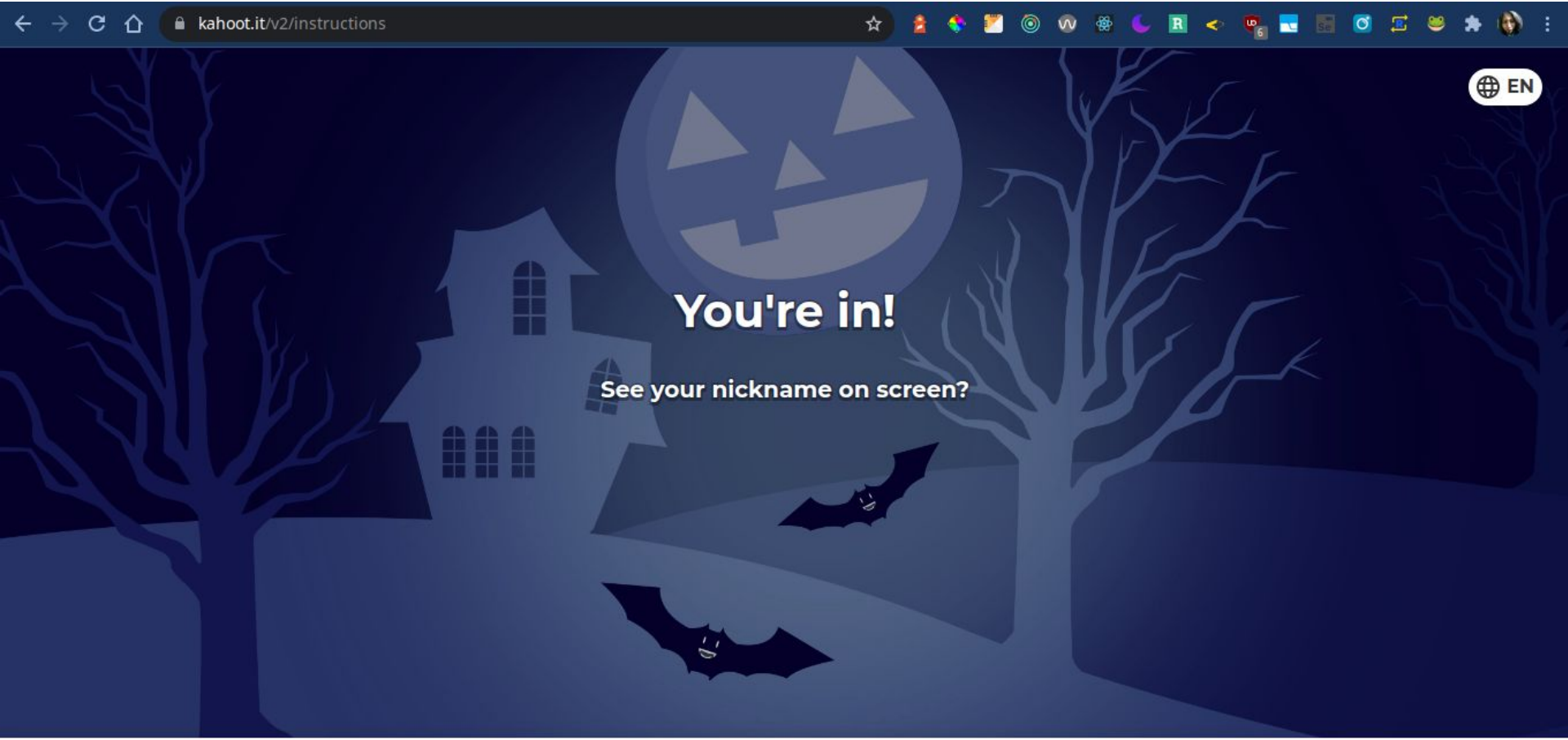
1. Look at the big picture.
2. Get the data.
3. Discover and visualize the data to gain insights.
4. Prepare the data for Machine Learning algorithms.
5. Select a model and train it.
6. Fine-tune your model.
7. Present your solution.
8. Launch, monitor, and maintain your system.

Pipeline

mojix

1. Look at the big picture.
2. Get the data.
3. Discover and visualize the data to gain insights.
4. Prepare the data for Machine Learning algorithms.
5. Select a model and train it.
6. Fine-tune your model.
7. Present your solution.
8. Launch, monitor, and maintain your system.




kahoot.it/v2/instructions
EN
You're in!
See your nickname on screen?
NoblePelican