

A man wearing a white tank top and a straw hat is standing on the deck of a boat. In the background, the mast and rigging of a sailboat are visible against a hazy sky and water. The overall image has a muted, greyish-blue color palette.

# **Building Machine Learning Powered Applications**

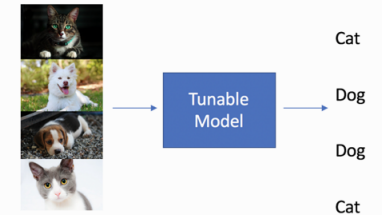


ML particularly useful to build systems for which we are unable to define a heuristic solution.

Algorithmic approach

```
def has_pointy_ears(image):  
    # A manually defined set of filters to detect pointy ears  
    pass  
  
def has_whiskers(image):  
    # A different set of filters to detect whiskers  
    pass  
  
def cat_or_dog(image):  
    if has_pointy_ears(image) and has_whiskers(image):  
        return "Cat"  
    else:  
        return "Dog"
```

Learning approach




*Figure 1-1. From defining procedures to showing examples*



An application that calculates your taxes automatically should rely on guidelines provided by the government. As you may have heard, having errors on your tax return is generally frowned upon. This makes the use of ML for automatically generating tax returns a dubious proposition.

You never want to use ML when you can solve your problem with a manageable set of deterministic rules



A dark, atmospheric image featuring Darth Vader in the foreground, his iconic helmet and mask clearly visible. He is positioned against a blurred background of a city skyline at night, with numerous lights from buildings creating a bokeh effect. The overall color palette is dark, dominated by blues and blacks, with the city lights providing a warm, yellowish glow.

When building products, you should  
start from a concrete business  
problem

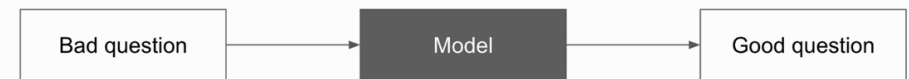
I am your father.



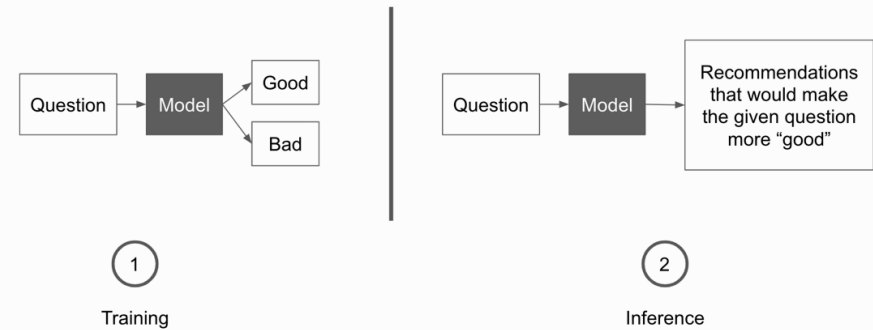
**More projects fail by producing good models that aren't helpful for a product rather than due to modeling difficulties**



# ML based Editor



*Figure 1-10. End-to-end approach*



*Figure 1-11. A middle ground between manual and end-to-end*



**Baseline; designing heuristics based  
on domain knowledge**




**Simple model; classifying text as good or bad, and using the classifier to generate recommendations**





**Complex model; training an end-to-end model that goes from bad text to good text**

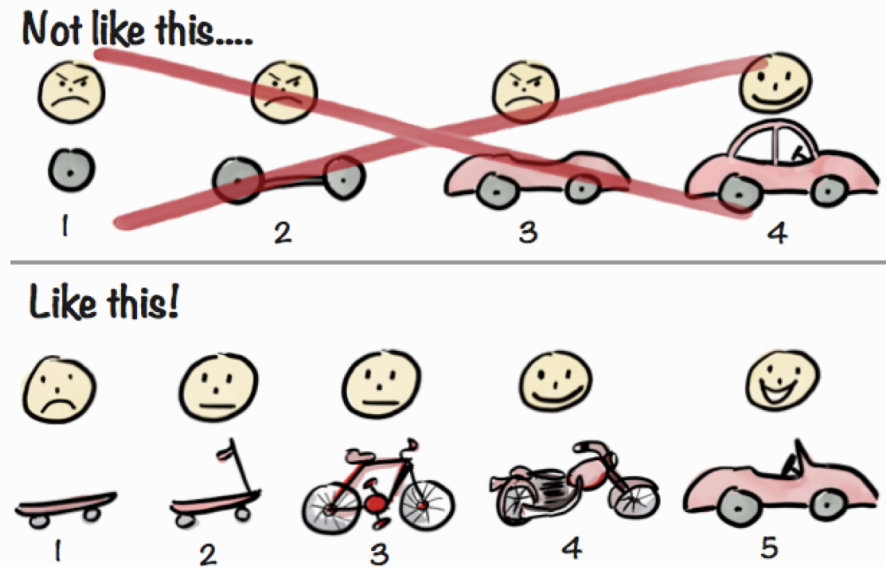




**starting without ML is  
the fastest way to build  
an ML product**



# Build a Working Pipeline



*Figure II-1. The right way to build your first pipeline (reproduced with permission from Henrik Kniberg)*



# Prototype

Flesch readability score

The **Flesch-Kincaid readability tests** are readability tests designed to indicate how difficult a passage in English is to understand.

Count the words, sentences, and syllables and then calculate using a library.

Simple question:

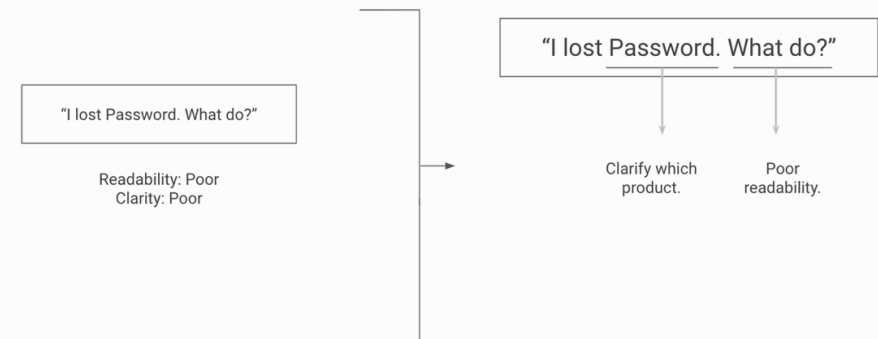
```
$ python ml_editor.py "Is this workflow any good?"  
Adverb usage: 0 told/said, 0 but/and, 0 wh adverbs  
Average word length 3.67, fraction of unique words 1.00  
6 syllables, 5 words, 1 sentences  
6 syllables, 100.26 flesch score: Very easy to read
```

Convolved question:

```
$ python ml_editor.py "Here is a needlessly obscure  
question, that"\  
"does not provide clearly which information it would"\  
"like to acquire, does it?"  
  
Adverb usage: 0 told/said, 0 but/and, 0 wh adverbs  
Average word length 4.86, fraction of unique words 0.90  
30 syllables, 18 words, 1 sentences  
30 syllables, 47.58 flesch score: Difficult to read
```



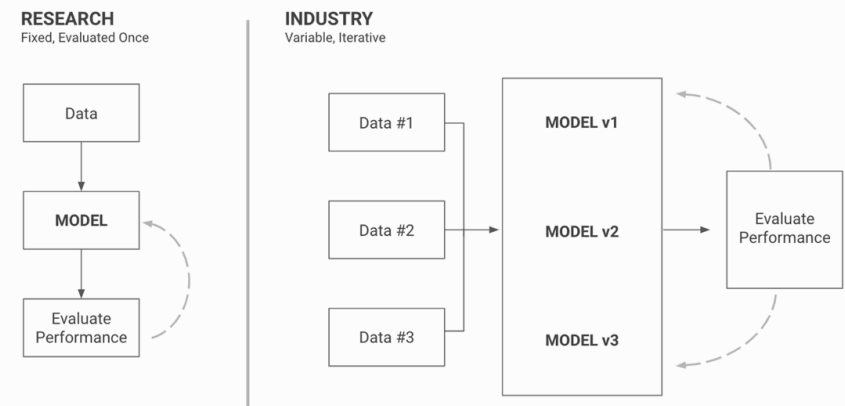
However, you've noticed that returning only a probability of rejection to a user is not the most satisfying of outputs.



*Figure 3-1. More actionable writing suggestions*



# Get a dataset



*Figure 4-1. Datasets are fixed in research, but part of the product in industry*





# Label to Find Data Trends

stackexchange dataset



**Build a heuristic using the data**





```
In [166]: 1 # Most confident correct positive predictions
          2 top_pos[to_display]
```

Out[166]:

|       | predicted_proba | true_label | Title   | body_text   | text_len | action_verb_full | question_mark_full | language_question |
|-------|-----------------|------------|---|---|----------|------------------|--------------------|-------------------|
| Id    |                 |            |   |   |          |                  |                    |                   |
| 38358 | 0.84            | True       | Punctuation when using inline dialogue                      | I am a bit crazy about punctuation and I have a question that I'm struggling to find a consensus... | 277      | False            | True               | False             |
| 7602  | 0.81            | True       | Is it unusual for a flashback to have a very long dialogue? | This flashback is from a short story I'm writing (unedited first draft):\n\nI met Limei last sum... | 870      | True             | True               | False             |

```
In [167]: 1 # Most confident correct negative predictions
          2 top_neg[to_display]
```

Out[167]:

|       | predicted_proba | true_label | Title   | body_text   | text_len | action_verb_full | question_mark_full | language_question |
|-------|-----------------|------------|---|---|----------|------------------|--------------------|-------------------|
| Id    |                 |            |   |   |          |                  |                    |                   |
| 7878  | 0.20            | False      | When quoting a person's informal speech, how much liberty do you have to make changes to what th... | Even during a formal interview for a news article, people speak informally. They say "uhm", they... | 116      | True             | True               | False             |
| 16453 | 0.21            | False      | Printing by the Publisher   | My first book was published through Xlibris. They have reported no sales from numerous authors, ... | 131      | True             | True               | False             |

*Figure 5-10. Top-k most correct*



# **Train and Evaluate Your Model**

avoid using the best model and benchmark them and  
pick one with the best results



True class: High score

Prediction probabilities

|            |                             |
|------------|-----------------------------|
| Low score  | <div><div></div></div> 0.48 |
| High score | <div><div></div></div> 0.52 |

Low score

High score

What  
0.03  
things  
0.03  
up  
0.03  
ideas  
0.03  
feel  
0.03  
useful  
0.02

#### Text with highlighted words

What are the first things I should do with an idea? I have a lot of ideas for things I want to write, but I usually end up getting lost in a stream-of-consciousness and either lose interest in the idea, or suddenly feel overwhelmed by the breadth of the subject I wish to cover. Is there some way of recording my ideas which is less restrictive than outlining, but perhaps more useful than simply writing random sentence fragments on a paper which will be meaningless to me later?

*Figure 5-12. Explaining one particular example*



```
// Following the first three recommendations
>> get_recommendation_and_prediction_from_text(
    """
    I'd like to learn about building machine learning
    products.
    Are there any good product focused resources?
    Would you be able to recommend educational books?
    """
)
```

0.48 score

Increase question length

Increase vocabulary diversity

Increase frequency of adverbs

No need to decrease frequency of question marks

Increase frequency of commas



# Building Machine Learning Powered Applications by Emmanuel Ameisen



A cartoon illustration of Rick Sanchez from the animated series 'Rick and Morty'. He is standing in a desert environment, wearing his signature white lab coat over a teal shirt. His arms are outstretched to the sides. In the background, there is a small, conical tent made of brown and green material on the left, and a wooden fence with a skull hanging from a post on the right. The ground is sandy and brown. The text 'Thank you' and the website 'gauthamsanthosh.com' are overlaid on the center of the image.

**Thank you**  
[gauthamsanthosh.com](http://gauthamsanthosh.com)

