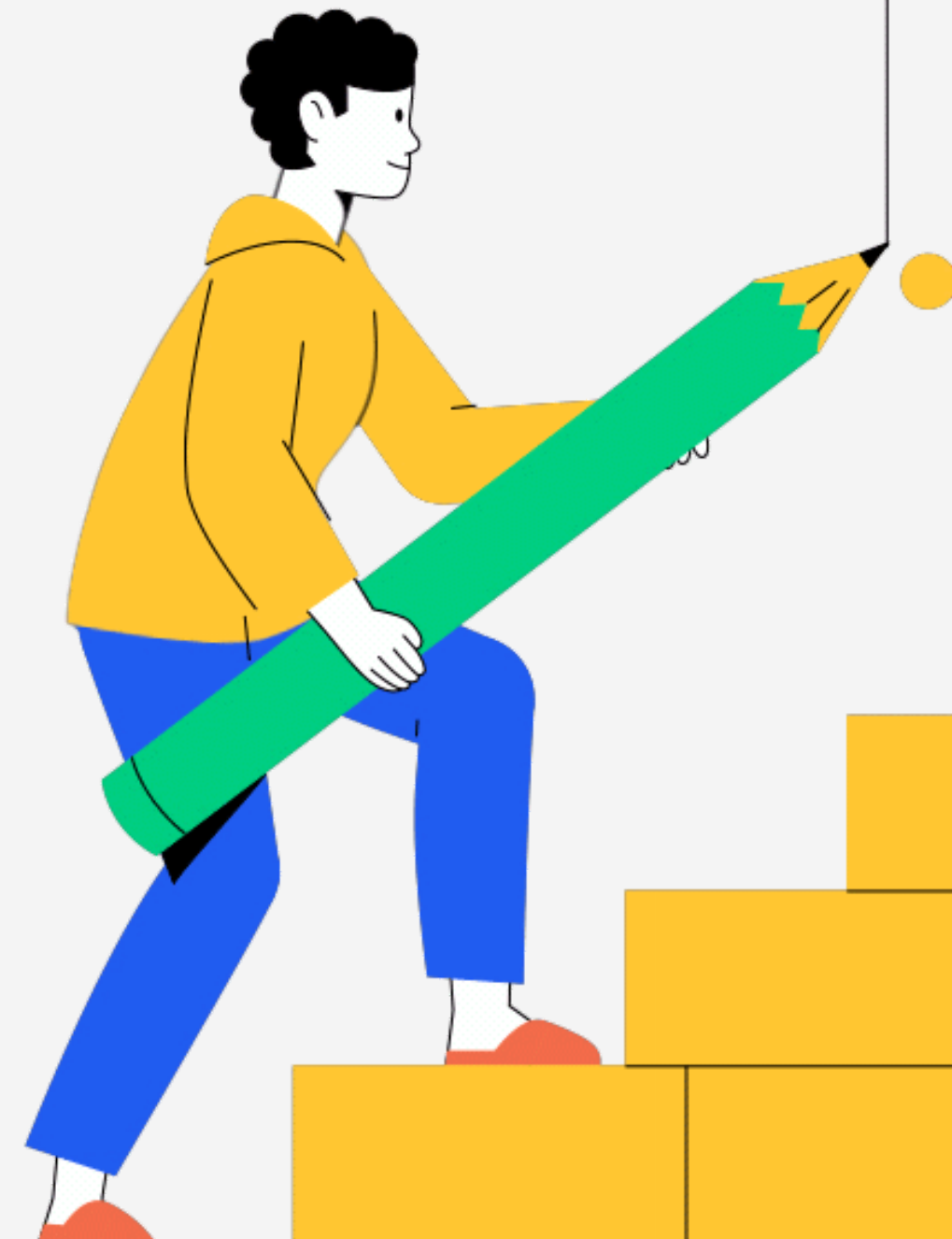


# ESSAY SCORE PREDICTION

Predict scores using natural language processing



HARRY LU



KEVIN DEMSICH



NICHOLAS LEE

December 6, 2022 @ 6:30 P.M. PT

# Project Motivation

We aim to build an ML model that **could learn from current grading** and use it to **score any unseen essay**.



Release the lecturer from repetitive grading tasks to invest in more education / research





# EDA

What is input data and what to predict

	text_id	full_text	cohesion	syntax	vocabulary	phraseology	grammar	conventions
0	0016926B079C	I think that students would benefit from learn...	3.5	3.5	3.0	3.0	4.0	3.0
1	0022683E9EA5	When a problem is a change you have to let it ...	2.5	2.5	3.0	2.0	2.0	2.5
2	00299B378633	Dear, Principal\n\nIf u change the school poli...	3.0	3.5	3.0	3.0	3.0	2.5
3	003885A45F42	The best time in life is when you become yours...	4.5	4.5	4.5	4.5	4.0	5.0
4	0049B1DF5CCC	Small act of kindness can impact in other peop...	2.5	3.0	3.0	3.0	2.5	2.5

## WHAT IS IN INPUT DATA

# Essay

# 430 words

Or, if you may, 2,328 characters per essay

I would pick policy 1 because lunch is free time to eat talk to your friends and to use the restroom. If u need to tell your mom somethin...

Knowing what your future career is in high school is very important. In high school there are many classes that could help you with your...

I think that students should bring phones to school but turn then off and put them in their locker. If the students don't do that job they...

A topic that is debated is if students should do online classes. Some people believe that there should be online classes. They believe that...

Everyone has a talent or a skill. But they are all unique. Skills that I have are: intelligence, talent and athletic. Ways I show intelligence ar...

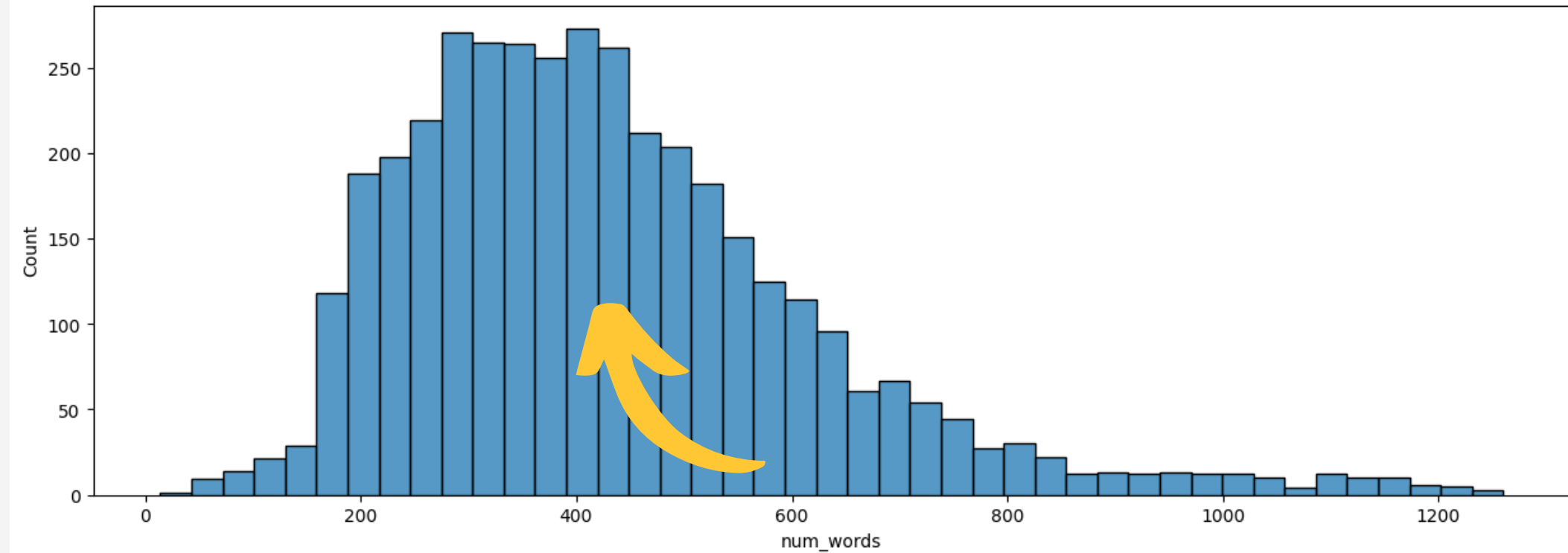
Do you think that should be students finish high school in three years, enter college and get a job early? Some students, parents and...

I want to work at a supreme store for a day in Generic\_City or Generic\_City. And seeing if im a good worker and to see if the boss ...

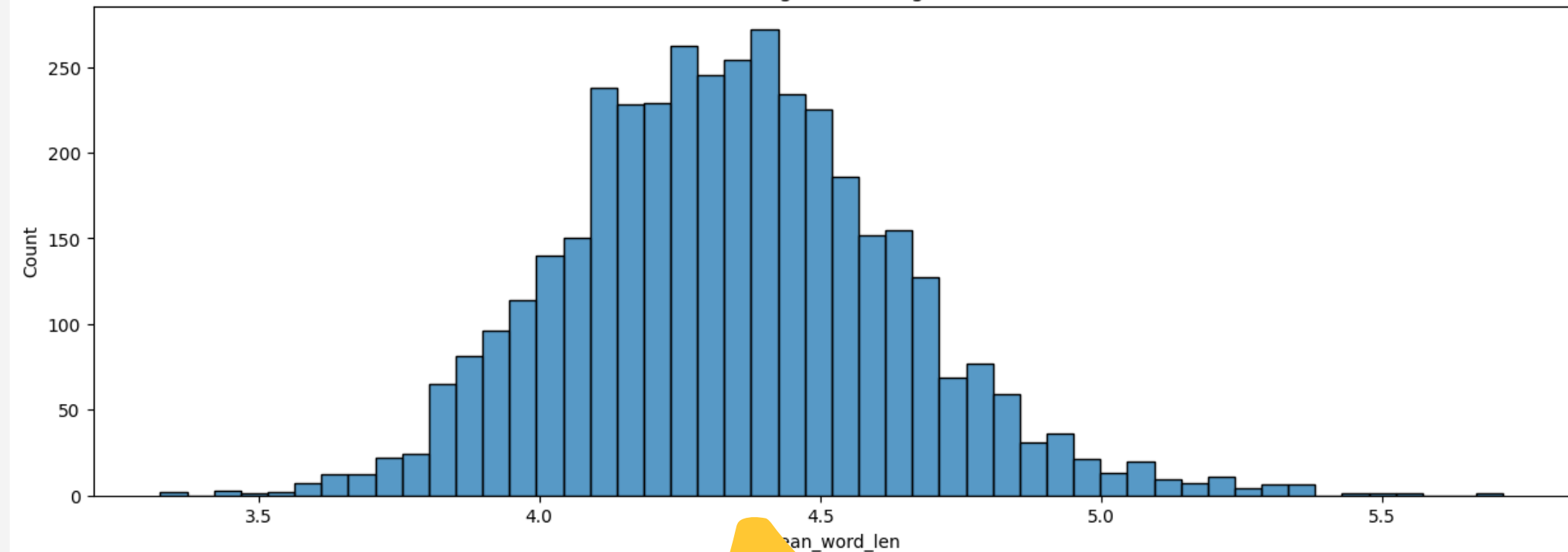
Dear Principal, I was thinking about the policies you were offering about the cell phones allowed in school and I would suggest to choo...

3,910 text records

Number of Words

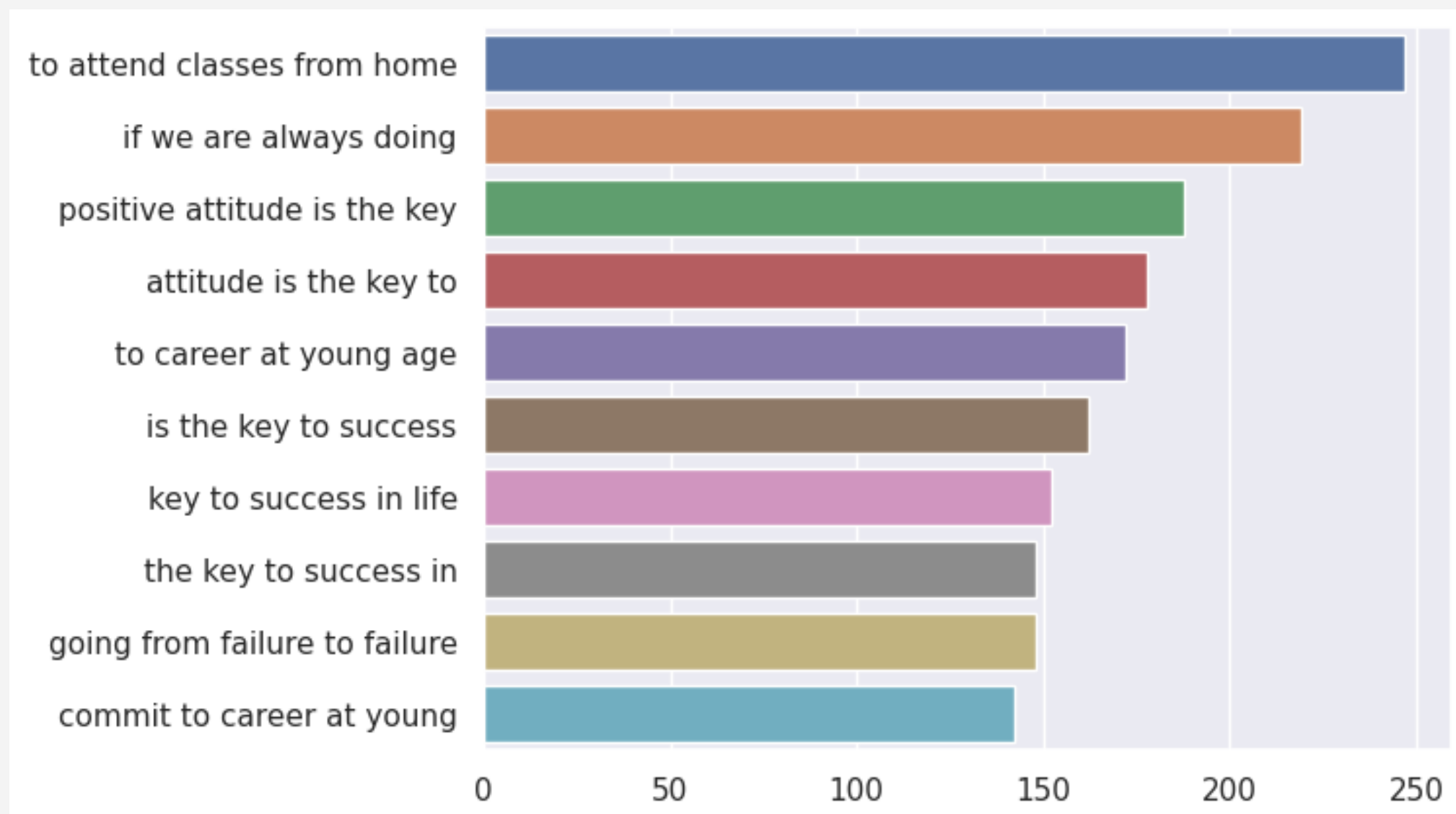
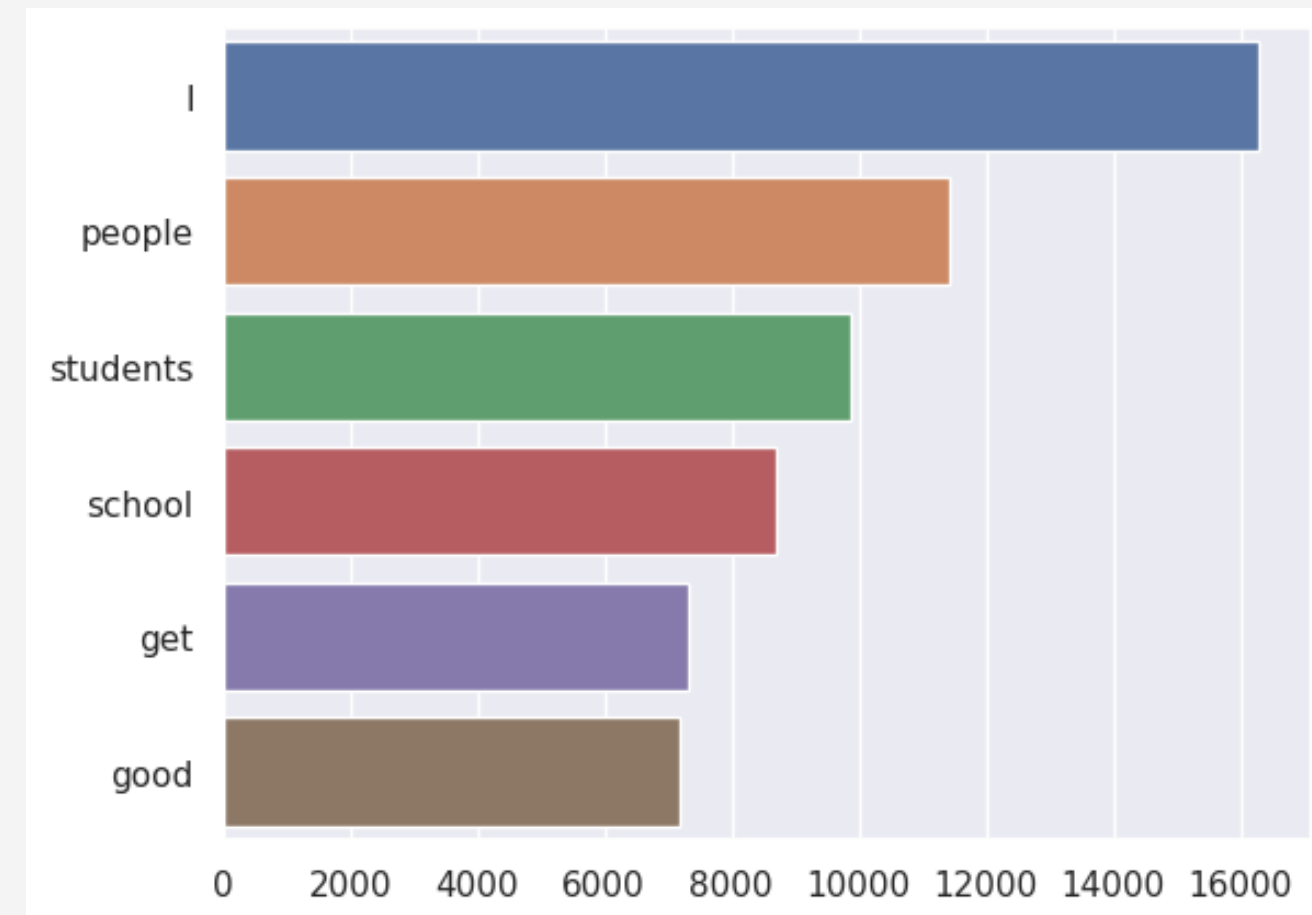
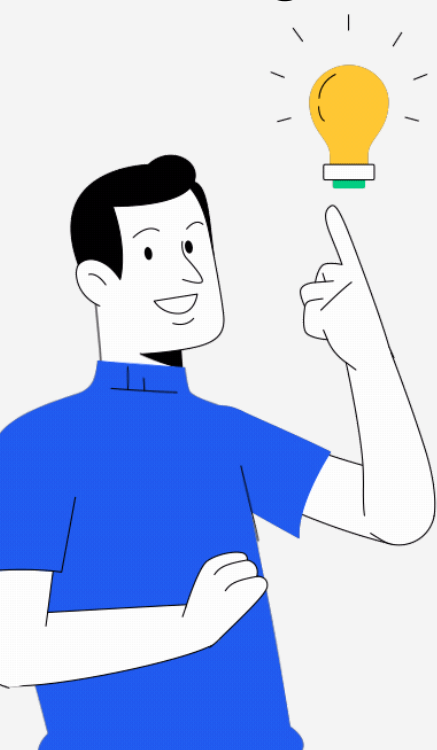


Average Word Length



# Input Characteristics

The popular words and the popular combination of words (ngram=5)



# Bert-base-cased method

Language model which is bidirectionally  
trained (this is also its key technical  
innovation).

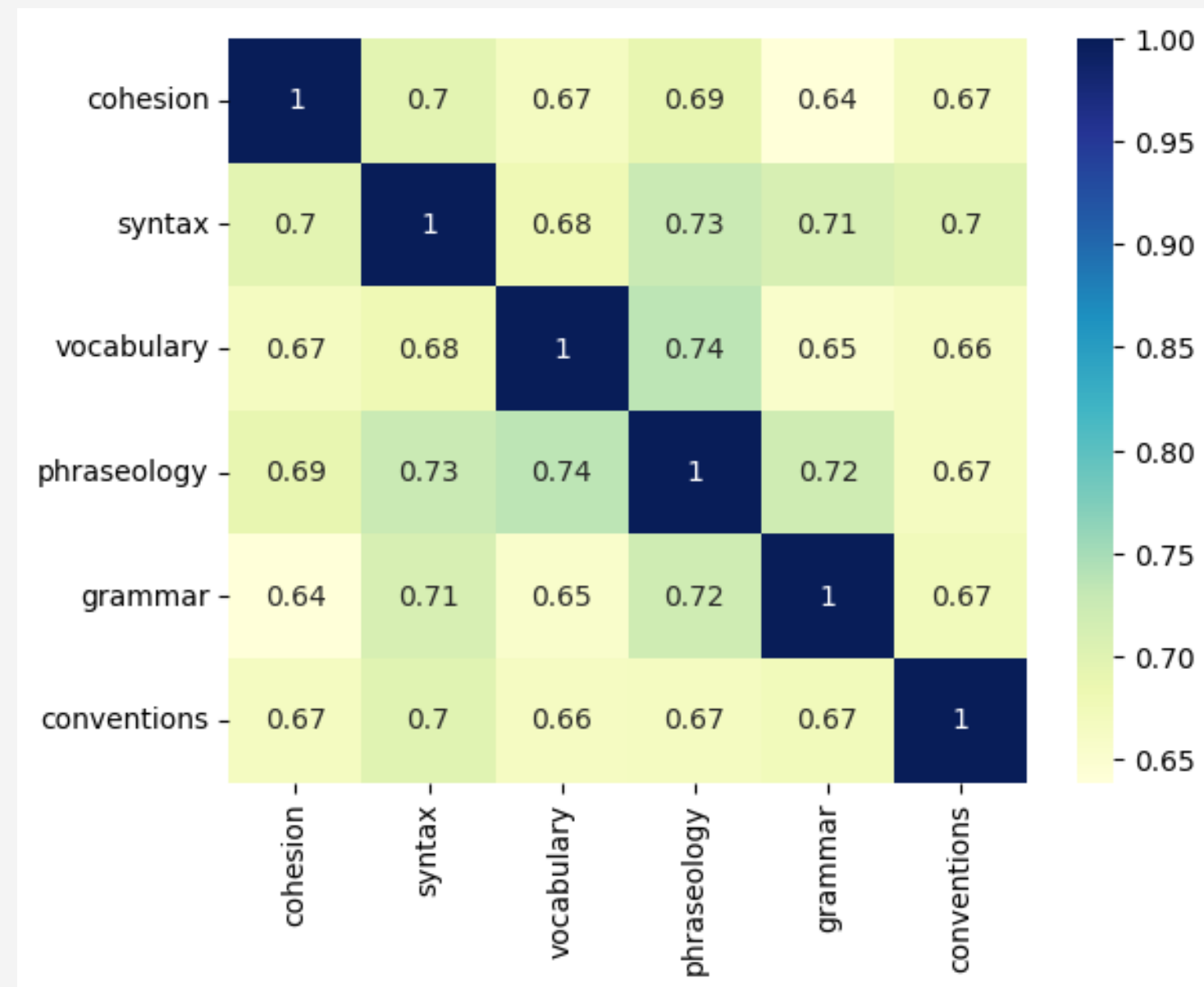
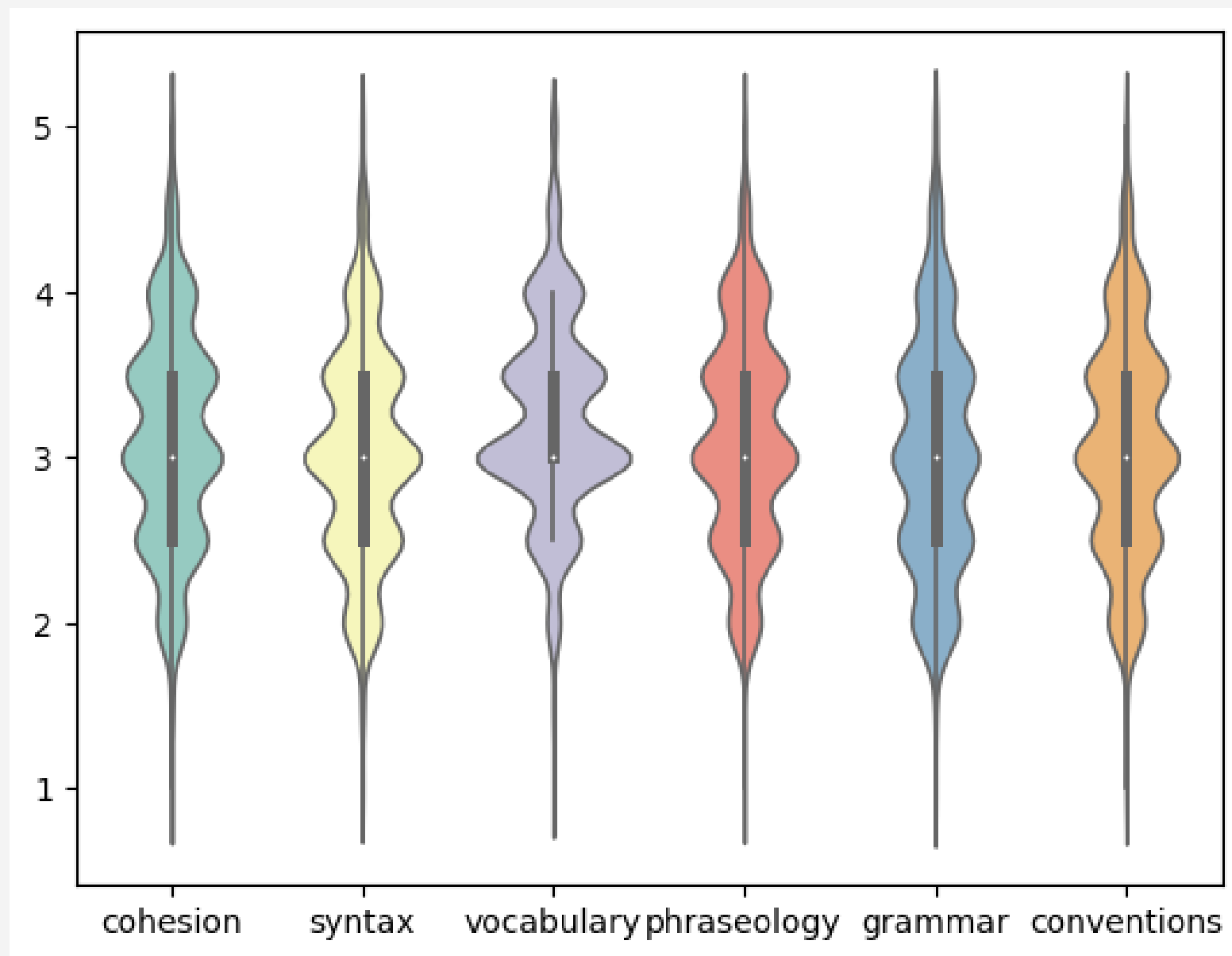
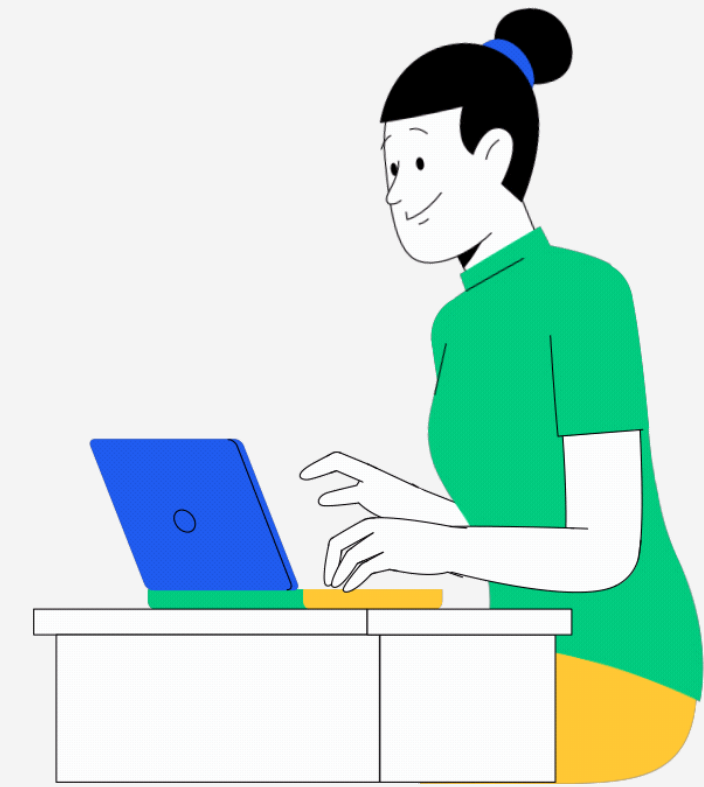
Downstream tasks:  
if you have a dataset of labeled  
sentences, you can train a standard  
classifier using the features produced  
by the BERT model as inputs.



## WHAT IS IN OUTPUT DATA

# Scores

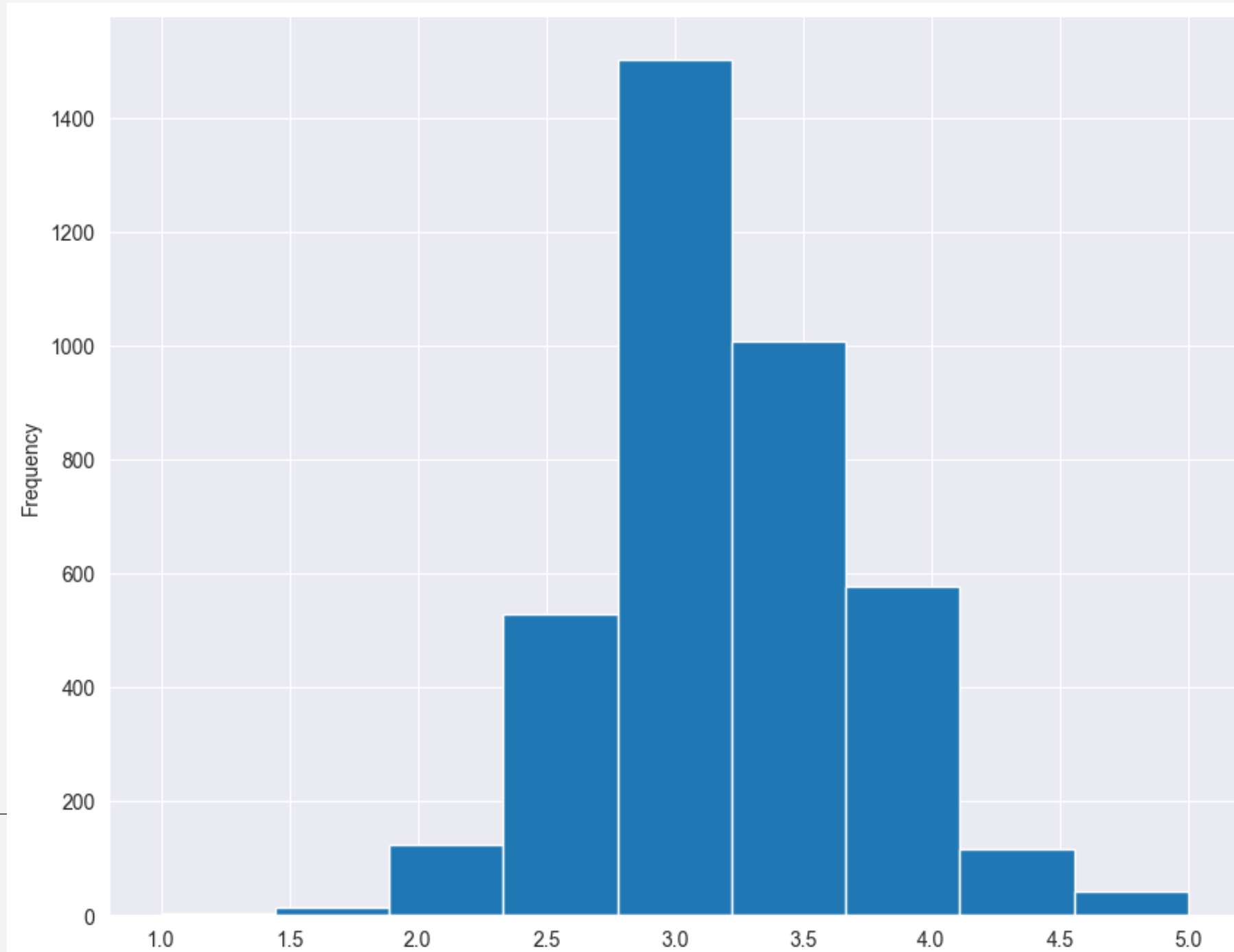
The scores we want to predict centered at 3, kind of **imbalanced** (lack of 1 and 5), and correlated to each other!



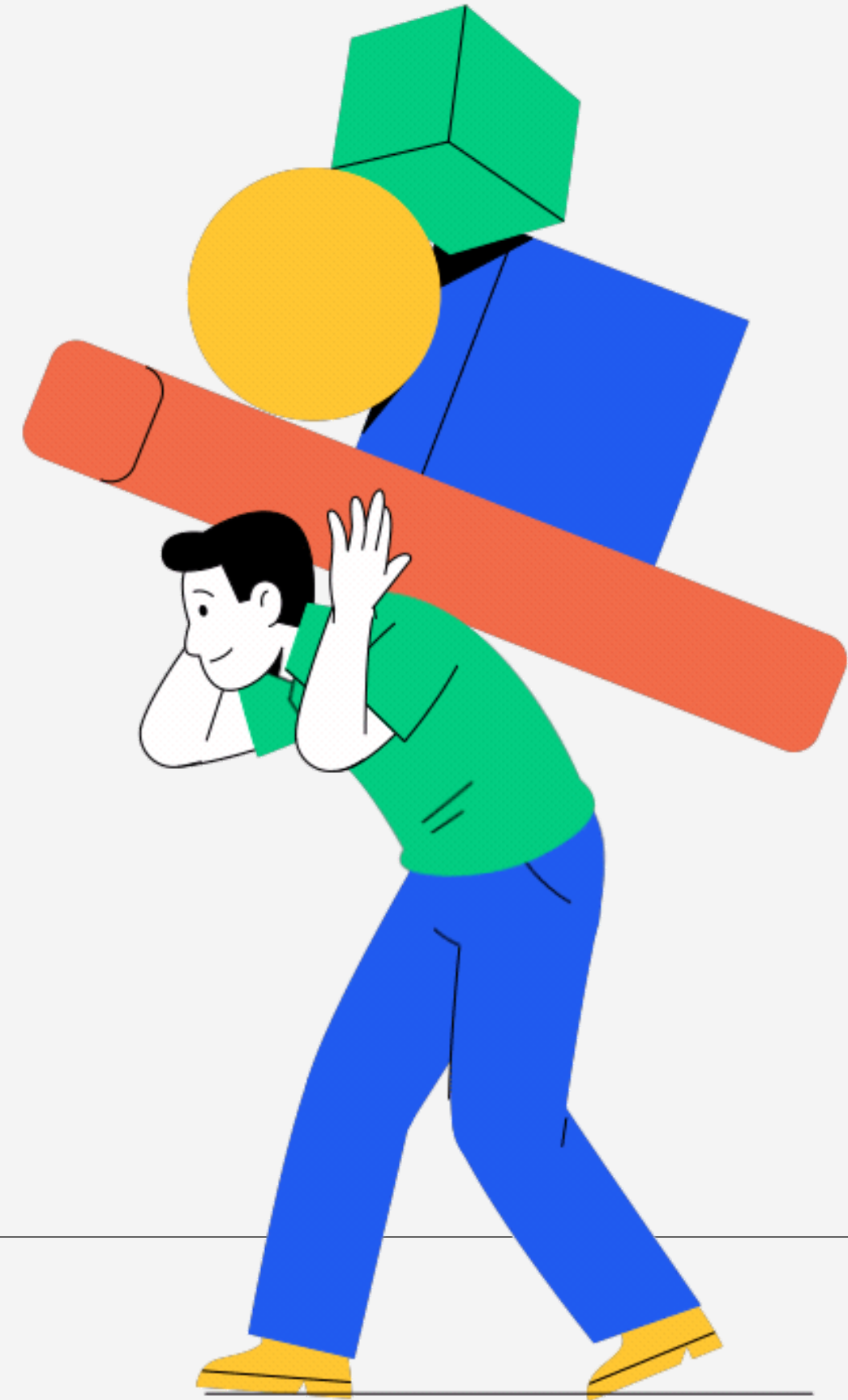


# How to deal with imbalanced data

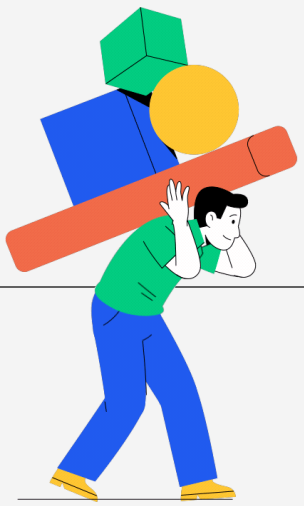
with oversampling



[2, 2.5, 3, 3.5, 4]



[1, 1.5, 4.5, 5]

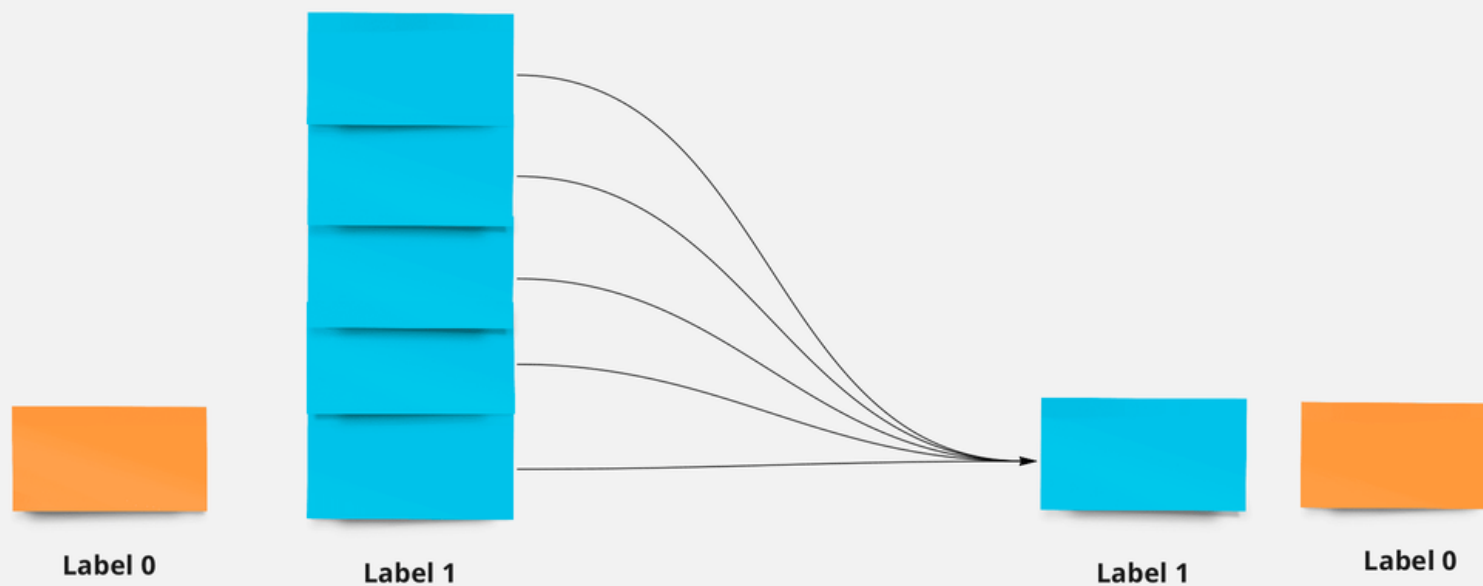




# IMBALANCED DATA

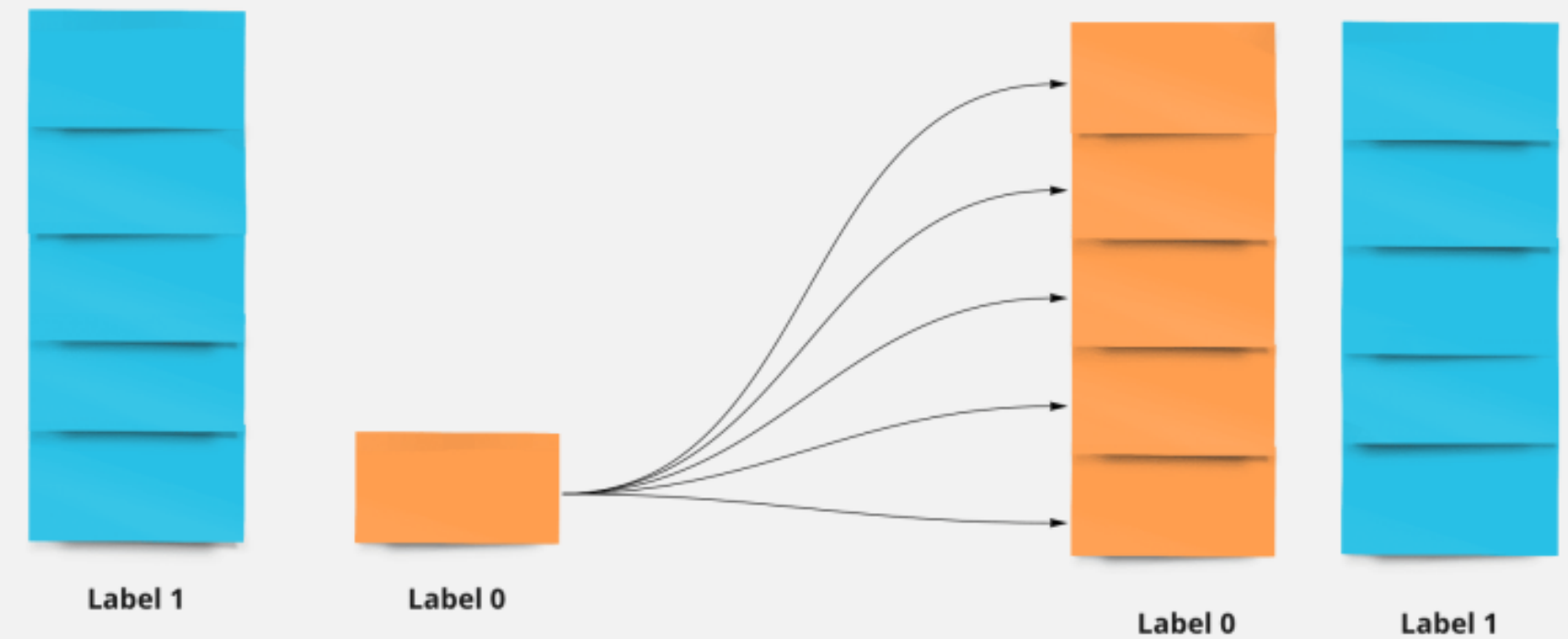
How do we oversample to solve the data imbalance issue

UnderSampling



dataaspirant.com

OverSampling



dataaspirant.com

**USING OVERSAMPLING, WE CAN GENERATE NEW TRAINING EXAMPLES BY AUGMENTING THE DATA**

# Text Preprocessing

Remove  
**Punctuation**

Remove  
**Stop Words**

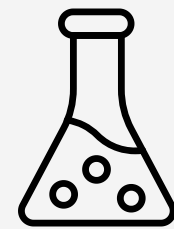
**Truncate** to  
Mean Length

**Pad** Embedding  
Length

**Encode** to  
Numeric Vectors

**Cross Validate**  
small batches

# Experimental Model Building



# Baseline Model

**Just choose 3!**

With all of the effort required to build a model, we decided to use a heuristic baseline model of always predicting 3.

This model produces ~38% accuracy.

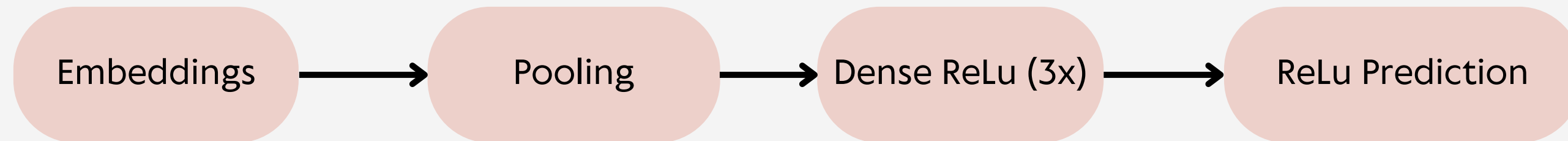
This helps us determine the relative improvement using a trained ML model.



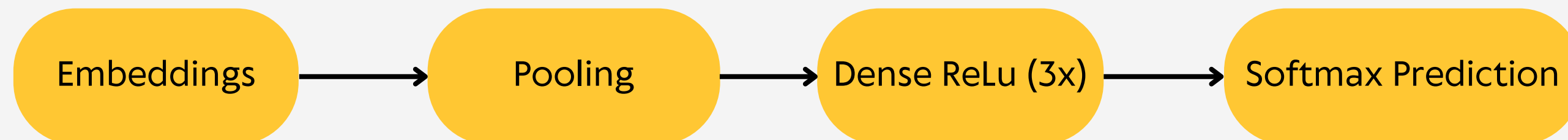


# Building Baselines

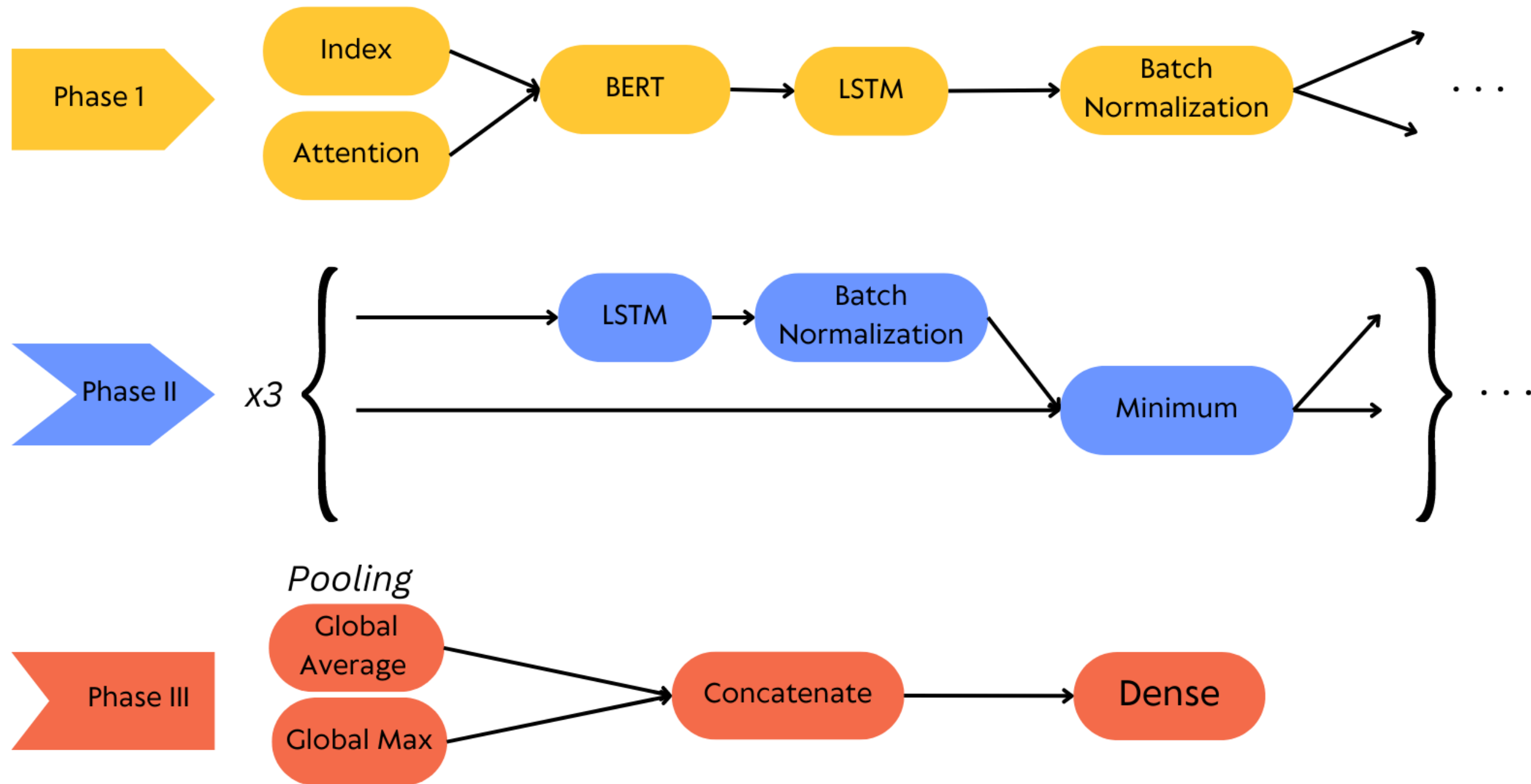
## Regression Model



## Sparse Categorical Model



# Final Model



*LSTM* - Long-Short-term-memory



# Model Evaluation

## BERT v. Sparse

Confusion Matrix for vocabulary



Accuracy: 36.7%

Confusion Matrix for Vocab Sparse Model



Accuracy: 25.7%

# Learning

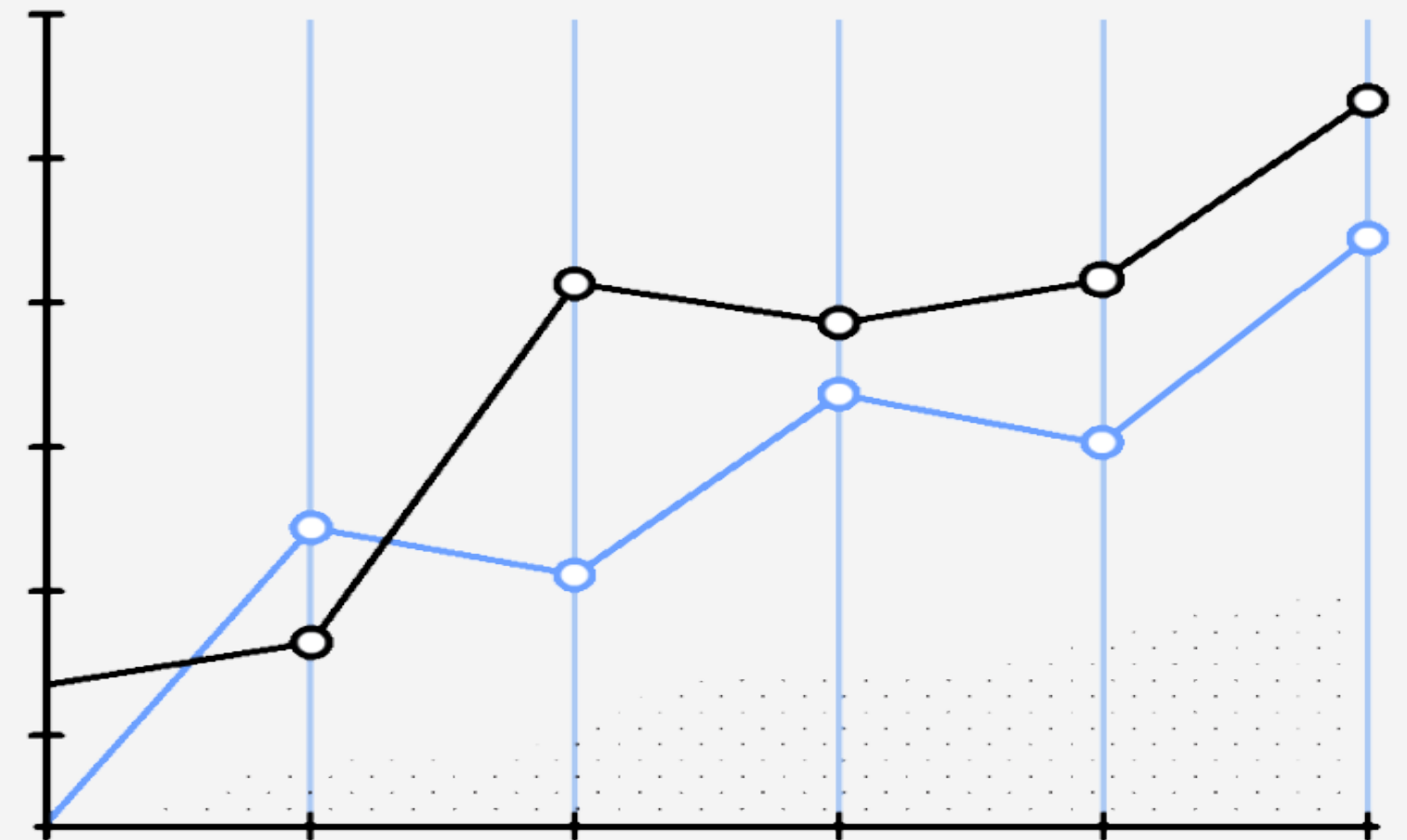
## Regression vs. Classification

### 1. Classification:

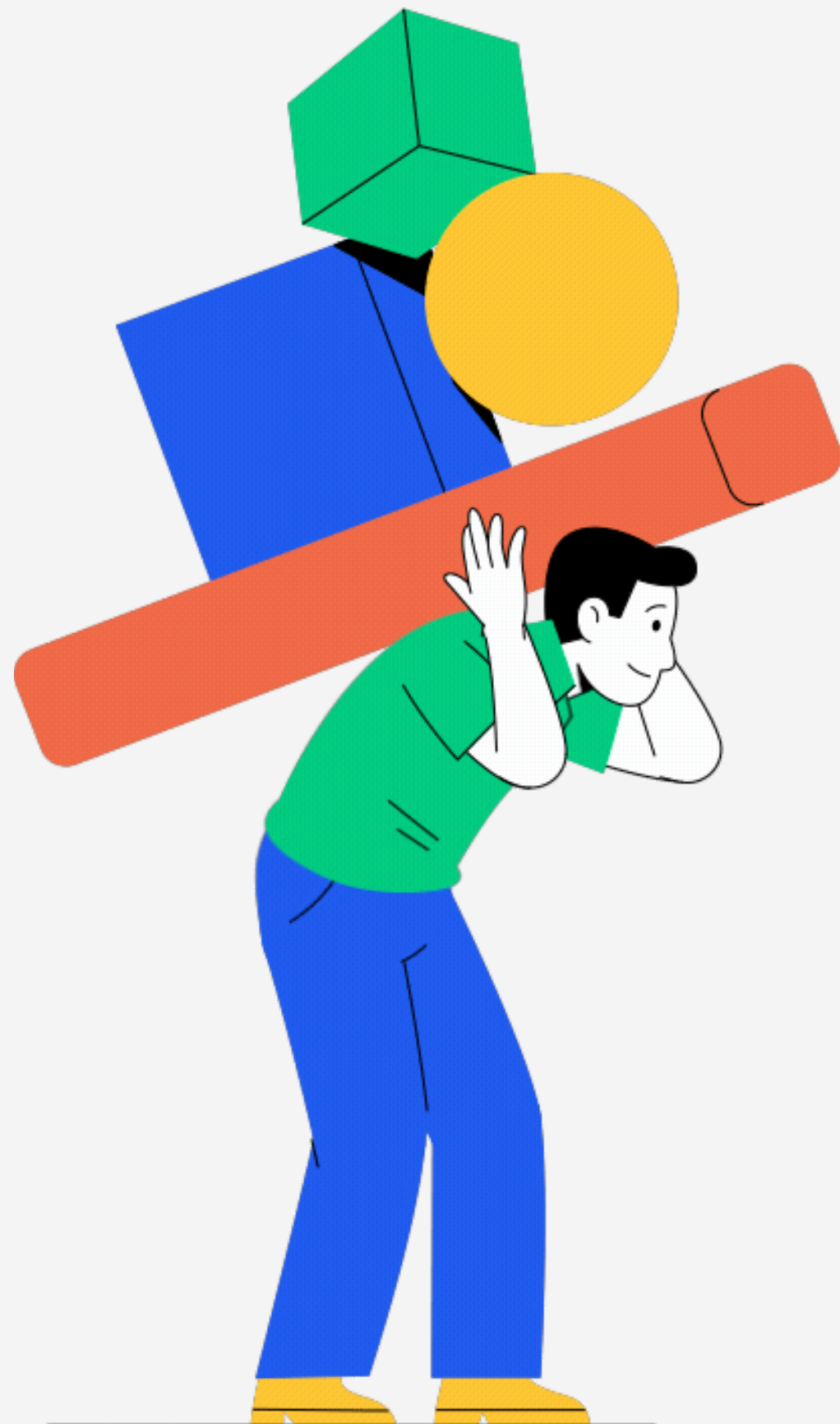
- a. Pros: fixed number of nine outputs.
- b. Cons: punishment for predicting a 1 to a 5 = 4.5 to a 5

### 2. Regression:

- a. Pros: output is a numeric type
  - i. how “off” our predictions are (i.e., MSE)
- b. Cons: The rank-type data is not linear, and the scores are discrete and not continuous.
  - i. The distance between 4.5 and 5 differs from the distance between 1 and 1.5

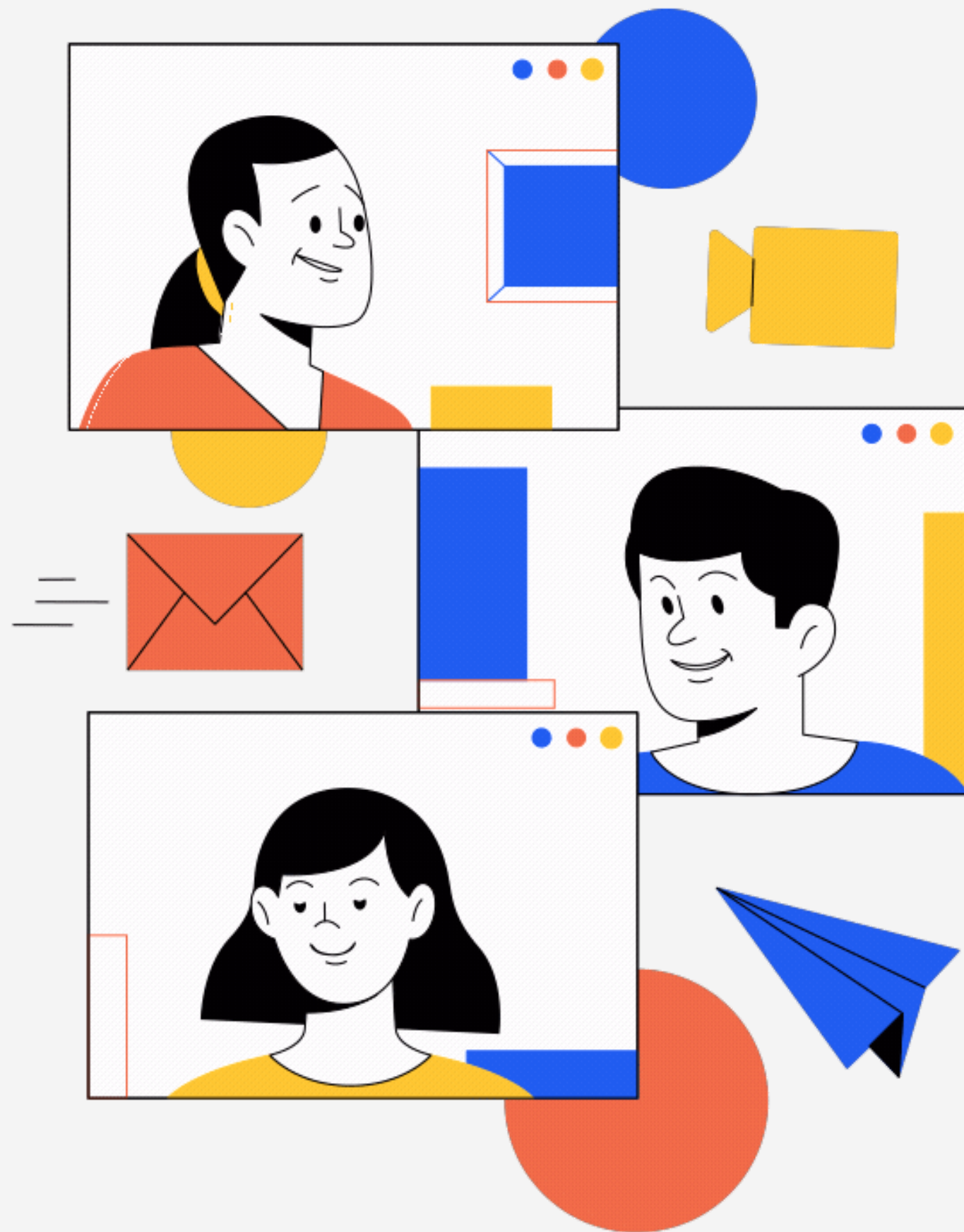






# Conclusions

1. With more iteration, we can use the model to automate English essay grading
2. We plan to further explore the two routes between classification and regression
3. For classification, we want to explore further the imbalance data solutions
4. We want to explore the method to speed up the model training process



# Thank you!

Feel free to slack at **#datasci-207**  
for any questions or clarifications!