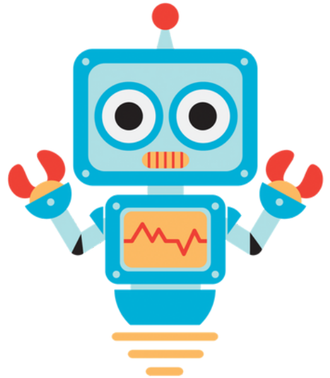
Detecting Student vs. Large Language Model (LLM) Essays

Team Members:

Ashish Agarwal

Project Overview

Classifying whether an essay was written by a student or a Large Language Model (LLM)



OR

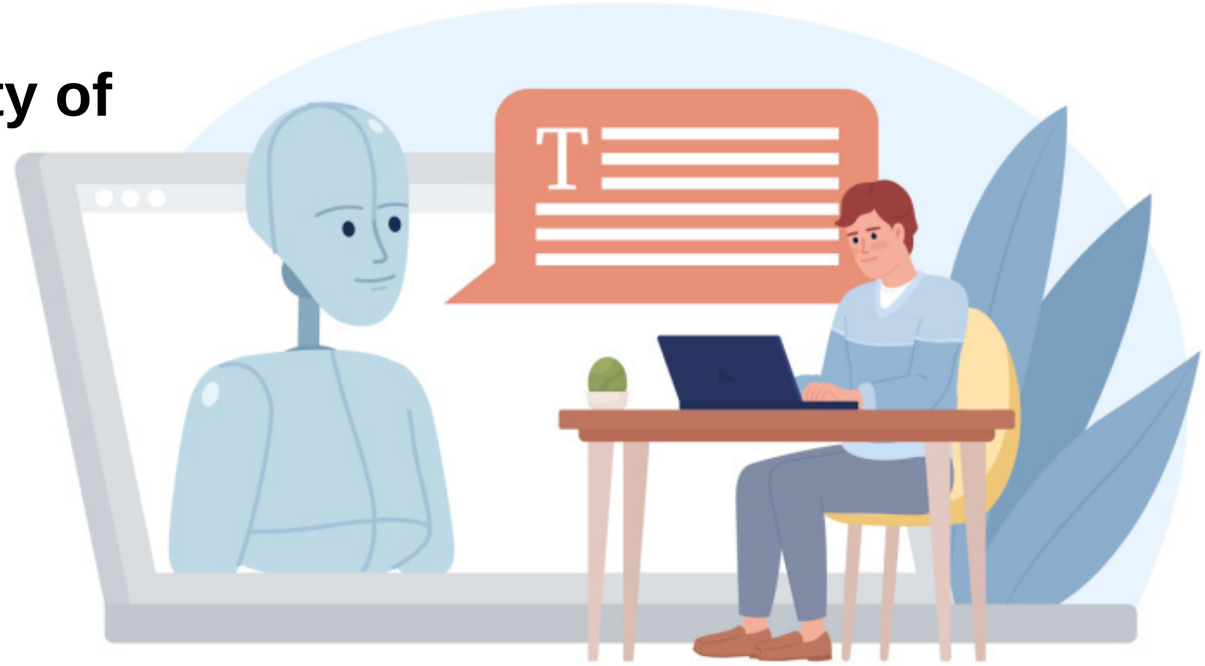
?

Project Objectives

- To build a robust machine learning model for distinguishing between student-written essays and LLM-generated text.
- To identify LLM artifacts that differentiate them from human writing.
- To contribute to addressing concerns about the potential impact of LLMs on education and plagiarism.

Significance

- **Understand the evolving landscape of AI in content creation.**
- **Address concerns about the authenticity of written content.**



Dataset

- **Human generated and LLM generated Texts.**
- **Data based on only two topics or Prompts.**
- **Use of different LLMs for LLM generated texts.** [source]

```
train_essays_df = pd.read_csv('/kaggle/input/llm-detect-ai-generated-text/train_essays.csv')
print(train_essays_df['prompt_id'].value_counts())
print(train_essays_df['generated'].value_counts())
train_essays_df.head()
```

```
prompt_id
0    708
1    670
Name: count, dtype: int64
generated
0    1375
1       3
Name: count, dtype: int64
```

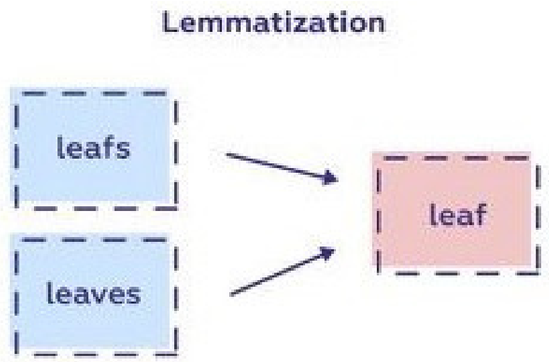
	id	prompt_id	text	generated
0	0059830c	0	Cars. Cars have been around since they became ...	0
1	005db917	0	Transportation is a large necessity in most co...	0
2	008f63e3	0	"America's love affair with it's vehicles seem...	0
3	00940276	0	How often do you ride in a car? Do you drive a...	0
4	00c39458	0	Cars are a wonderful thing. They are perhaps o...	0

```
train_prompts_df = pd.read_csv('/kaggle/input/llm-detect-ai-generated-text/train_prompts.csv')
train_prompts_df.head()
```

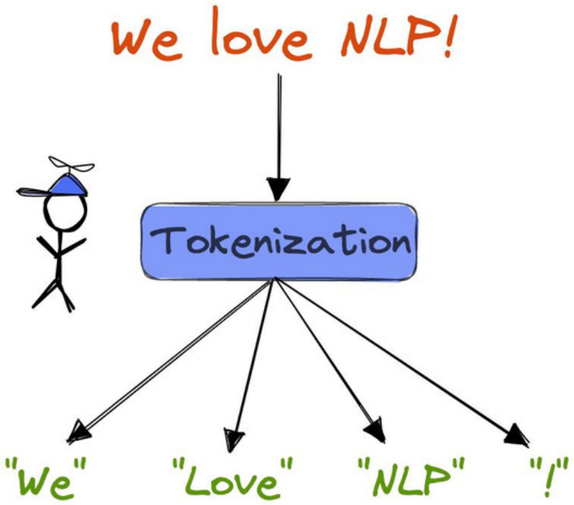
	prompt_id	prompt_name	instructions	source_text
0	0	Car-free cities	Write an explanatory essay to inform fellow ci...	# In German Suburb, Life Goes On Without Cars ...
1	1	Does the electoral college work?	Write a letter to your state senator in which ...	# What Is the Electoral College? by the Office...

NLP Techniques Used

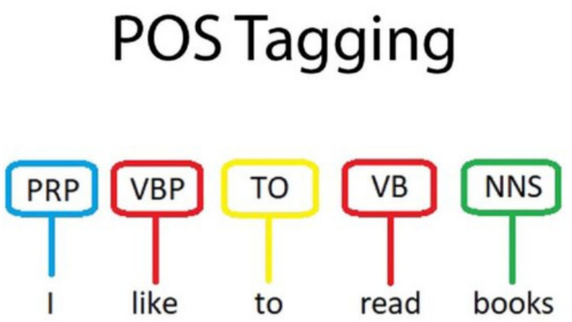
- Tokenization
- Part-of-Speech (POS) Tagging
- Lemmatization
- Stopword Removal
- Spell Checking
- Punctuation Analysis
- Collocation Analysis
- Text Vectorization (TF-IDF)



Source



Source

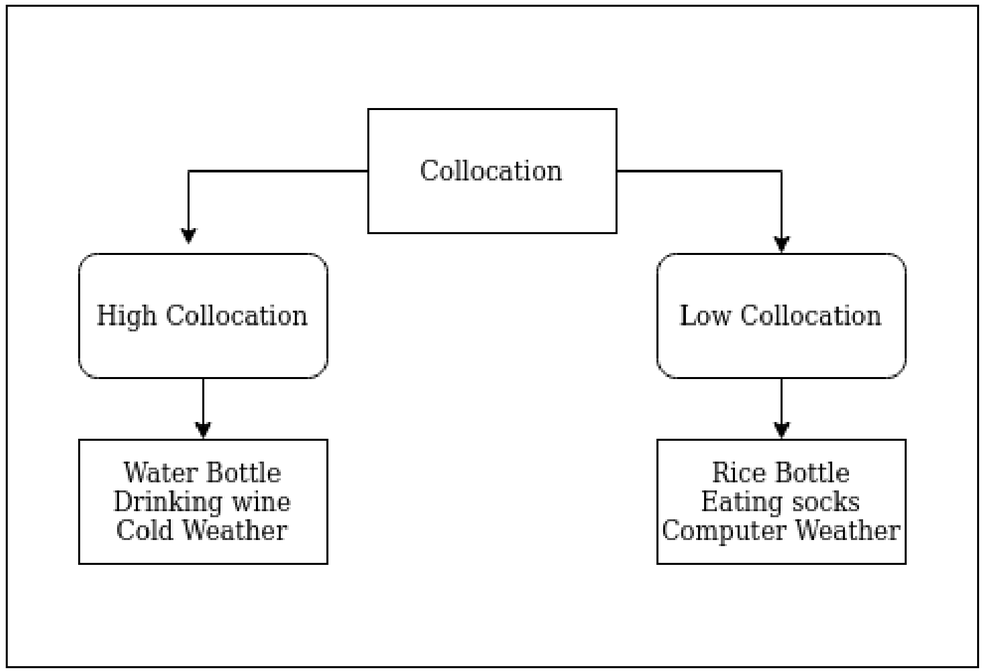


Source

NLP Techniques Used

TF-IDF VECTORIZATION

	text	tf	idf
0	Eddard Stark is a king in the north.	1	3
1	A king but one king : kings are everywhere.	2	3
2	Hodor was different : he was not a king .	1	3
3	But the North could not change without him.	0	3

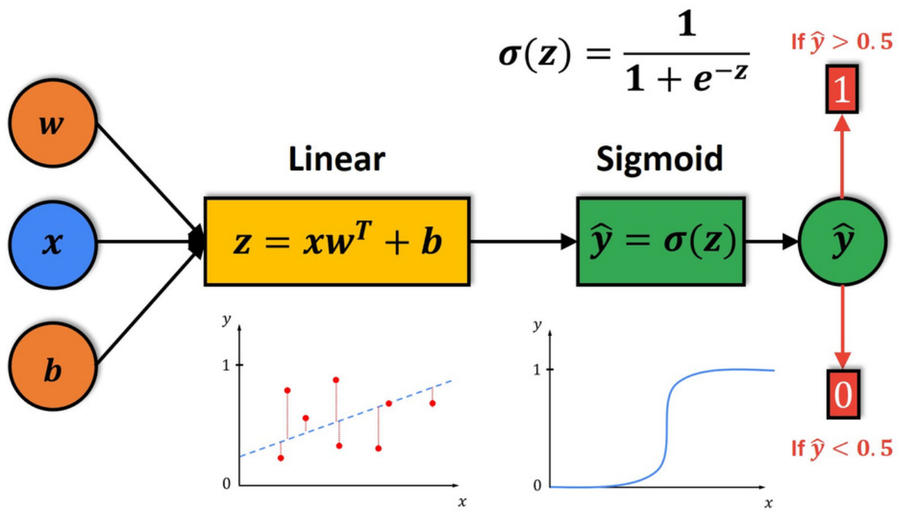


	king	was	the	not	a		he	one	north	kings	is	in	him	everywhere	A	different	could	change	but	are	Stark	North	Hodor	Eddard
0	0.333333	0.0	0.5	0.0	0.5	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
1	0.666667	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
2	0.333333	2.0	0.0	0.5	0.5	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
3	0.000000	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0

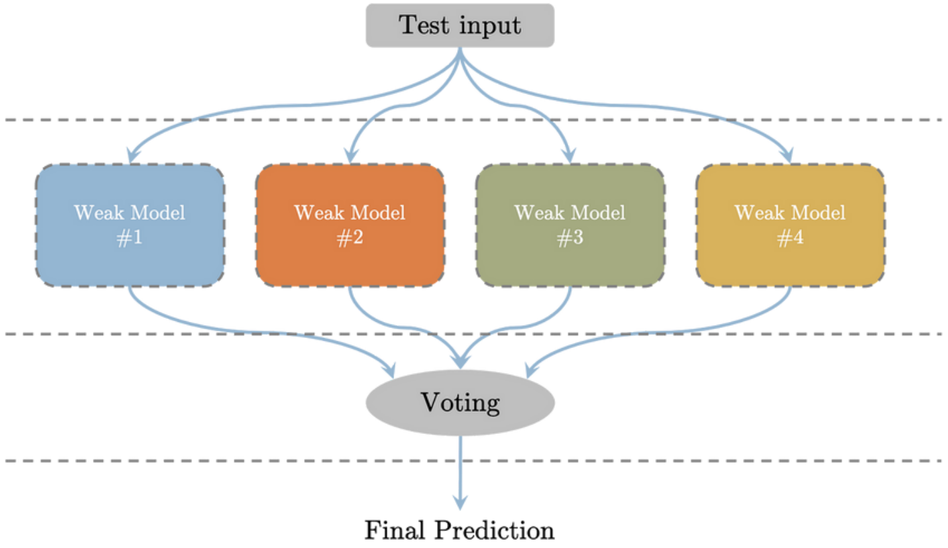
Source

Models Explored

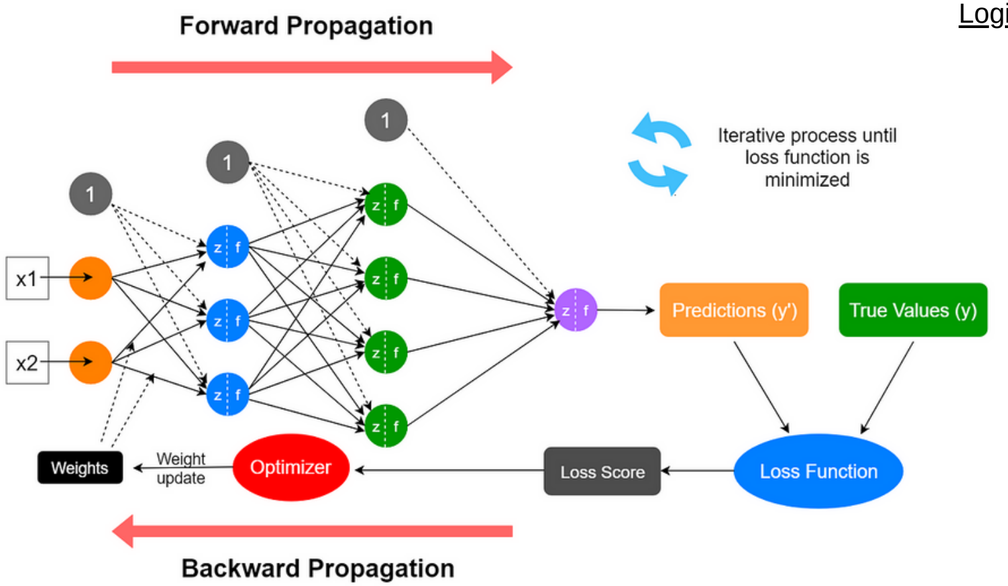
- **Logistic Regression**
- **Ensemble Models**
- **Neural Networks (NN)**



Logistic Regression Source



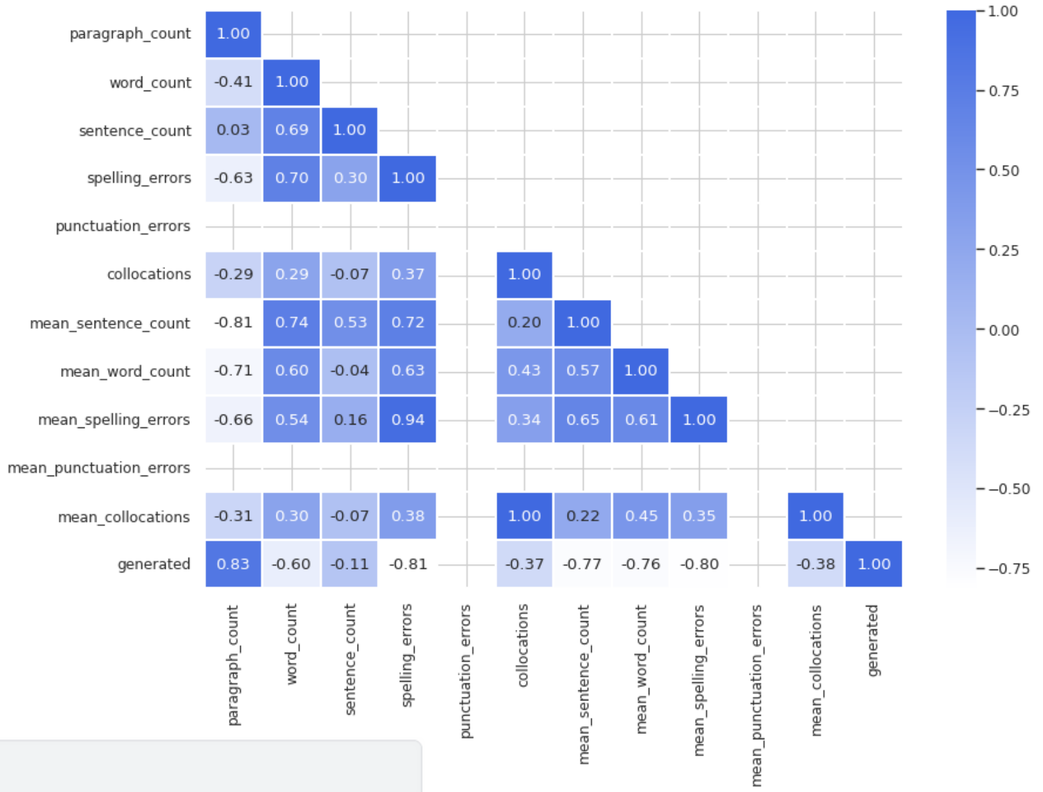
Ensemble Models Source



Neural Networks Source

Feature Development

- **Generating additional features.**
- **Leveraging Language Model Machines for data augmentation.**



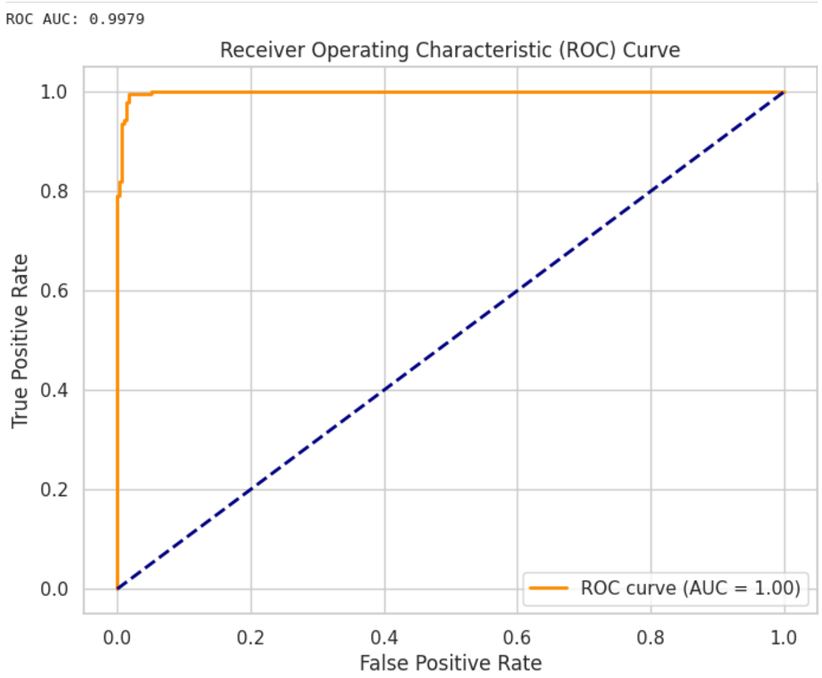

```
X_train_required_df.head()
```

	paragraph_count	word_count	sentence_count	spelling_errors	punctuation_errors	collocations	mean_sentence_count	mean_word_count	mean_spelling_errors	mean_punctuation_errors	mean_collocations	text
2070	5.0	754.0	35.0	52.0	0.0	0.0	7.000	21.151429	0.064626	0.0	0.0	Limiting the usage of cars will bring a lot of...
1213	12.0	556.0	33.0	15.0	0.0	0.0	2.750	14.520833	0.014303	0.0	0.0	## The Advantages of Limiting Car Usage\n\nCar...
424	8.0	450.0	27.0	6.0	0.0	0.0	3.375	18.311012	0.010465	0.0	0.0	In recent years, there has been a growing move...
485	8.0	487.0	32.0	9.0	0.0	0.0	4.000	14.987500	0.017310	0.0	0.0	In the United States, we are a car-centric soc...
999	5.0	634.0	25.0	19.0	0.0	0.0	5.000	25.360000	0.030384	0.0	0.0	There are many fellow citizens in the world th...

Evaluation Methodology

- **Metric:** Accuracy, Precision, Recall, F1 Score for model performance.
- **Cross-Validation:** Testing the model on unseen data for generalizability.
- **Feature Importance:** Analyzing key features contributing to the model's classification.

```
ROC AUC for fold 1: 0.9997
ROC AUC for fold 2: 0.9982
ROC AUC for fold 3: 0.9992
ROC AUC for fold 4: 0.9999
ROC AUC for fold 5: 0.9976
Average ROC AUC: 0.9989
Standard deviation: 0.0009
```




```
Confusion Matrix:
[[275   0]
 [ 95 183]]

Metrics:
Accuracy: 0.8282
Precision: 1.0000
Recall: 0.6583
F1 Score: 0.7939

Classification Report:
              precision    recall  f1-score   support

         0.0       0.74      1.00      0.85       275
         1.0       1.00      0.66      0.79       278

    accuracy                           0.83       553
   macro avg       0.87      0.83      0.82       553
weighted avg       0.87      0.83      0.82       553
```

Challenges Faced

- **Unavailability of large datasets.**
- **Unavailability of variation in dataset.**
- **Handling the dynamic nature of language evolution.**
- **Time to process and train the model.**
- **Computational Resource.**

```
train_essays_df = pd.read_csv('/kaggle/input/llm-detect-ai-generated-text/train_essays.csv')
print(train_essays_df['prompt_id'].value_counts())
print(train_essays_df['generated'].value_counts())
train_essays_df.head()
```

```
prompt_id
0    708
1    670
Name: count, dtype: int64
generated
0    1375
1       3
Name: count, dtype: int64
```

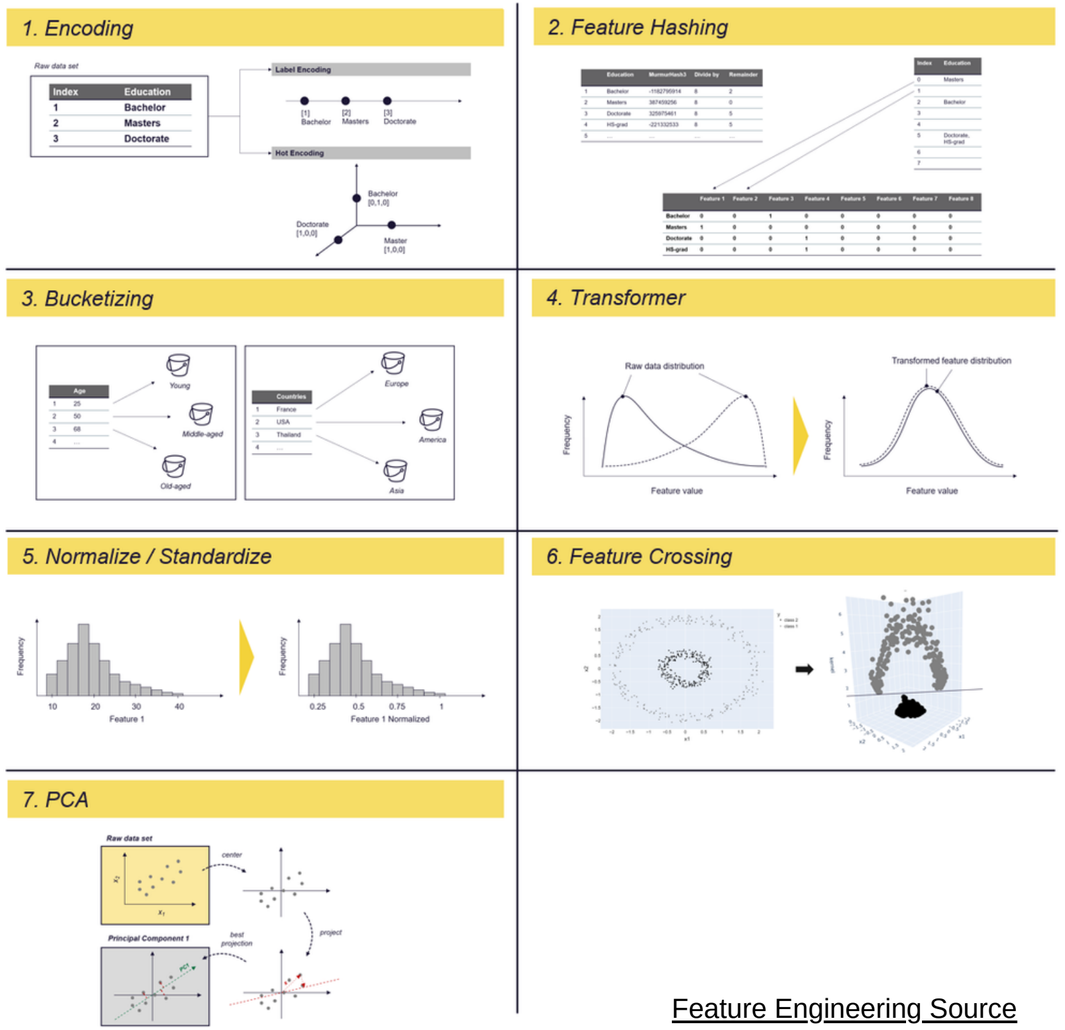
	id	prompt_id	text	generated
0	0059830c	0	Cars. Cars have been around since they became ...	0
1	005db917	0	Transportation is a large necessity in most co...	0
2	008f63e3	0	"America's love affair with it's vehicles seem...	0
3	00940276	0	How often do you ride in a car? Do you drive a...	0
4	00c39458	0	Cars are a wonderful thing. They are perhaps o...	0

Next Steps

- **Experimenting with advanced models.**
- **Incorporating more sophisticated feature engineering.**

Fig. Evolution of NLP Models

Advanced models Source



Feature Engineering Source

Any Questions ?

Thank you