# Song Lyric Generator Based on Genre

Vincent Colin Tionando - 2501996171
Steven Andrian Pradita - 2501996764
Caroline Stryadhi - 2502027610
Ammar Yaafi Adhinegoro - 250202791

# Background

Membuat sebuah model yang dapat menghasilkan lirik lagu untuk mempermudah orang dalam membuat lagu

# Related Works

**1.** https://lyrics.mathigatti.com/

**Keywords**

love

Generates song using specified keywords and phrases

**Creativity**

0.7

Should be between 0.1 and 1.0, the higher the value, the more creative/crazy the text.

**Repetition penalty** ✅

Activate it to reduce repetitiveness in the results.

KW: love

You're my one true love, ooh
You know you drive me crazy (drive me crazy)
Drive me crazy, baby
You know you make me so, so lazy

Drive me crazy, baby
You know you make me so, so lazy
Drive me crazy, baby
Drive me crazy, baby (finally kiss me, baby)

I don't wanna be with nobody but you (no, no, no)
I don't wanna be without your touch
(Oh-oh-ohh)
I don't wanna live without your love (fire or water)
If there's nothing else in sight
(Oh-oh-ohh)
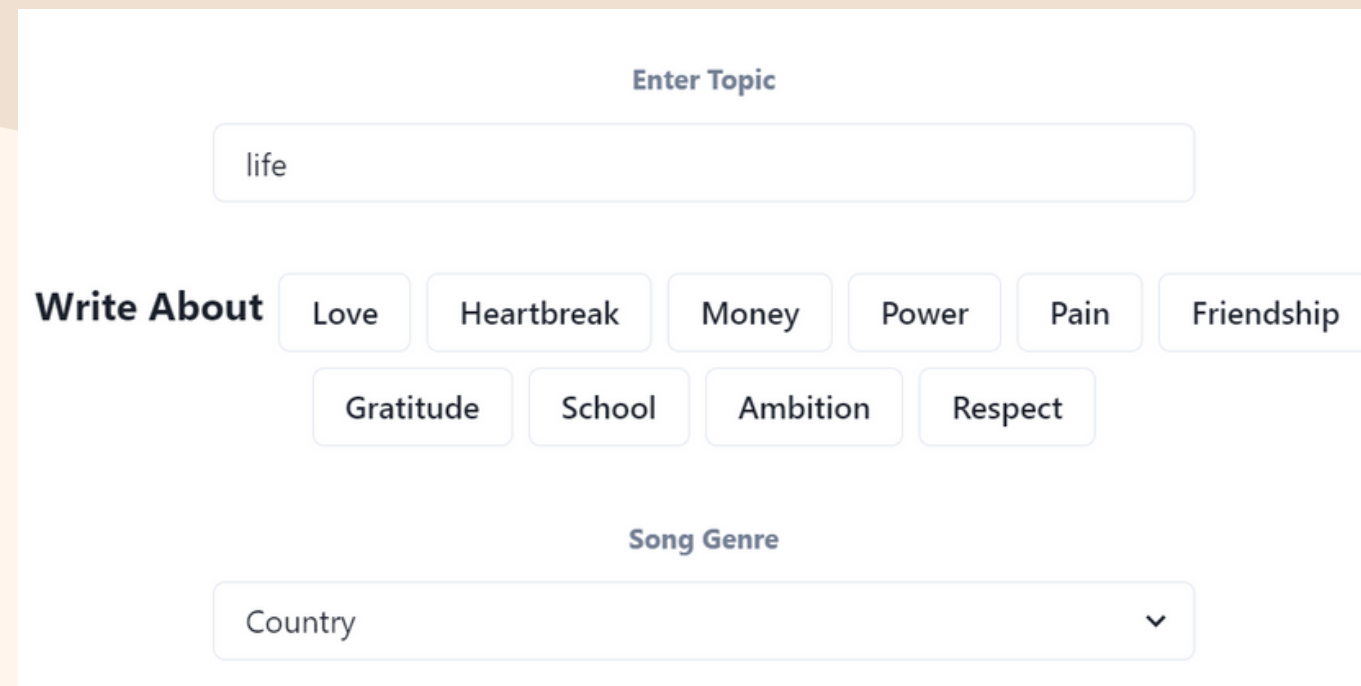I don't wanna be with nobody but you (uh-huh)

You're the sweetest thing on my mind
And I want you to know
Yes, I do and now
When we have an argument
We're not gonna leave this place
I'm sure you will be my one desire
(I don't wanna be)

Drive me crazy, baby
You know you drive me

# Related Works

https://www.freshbots.org/



Verse 1:
Life is a journey,
An adventure to unfold,
With every step and stumble,
We learn to be bold.

Chorus:
Living life to the fullest,
Chasing dreams that we hold,
Taking chances, embracing changes,
Our stories waiting to be told.

Genre Lagu yang tersedia:  Country, Pop, Rap,  Rock

# Related Works

Text Generation

It hasn't rained in weeks, oh, how can I
Doo da doo doo

# Methods

**Preprocessing dataset**
yang berisikan lirik lagu
beserta genre-nya.

->
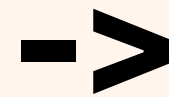
**Model pre-trained**
GPT-Neo 125m

**Generator lirik lagu**
menggunakan model-model
yang telah di fine-tuning

<-

**Fine-tuning**
model pre-trained
dengan menggunakan dataset
yang telah di preprocessing

# Methods

## Dataset - table 1

Kumpulan lirik lagu beserta data genre. Dataset memiliki lebih dari 300,000 sampel lirik yang dikelompokkan kedalam 79 genre lagu yang berbeda.

| A ALink | A SName | A SLink | A Lyric | A language |
|---------|---------|---------|---------|------------|
| /bruce-dickinson/ | Zulu Lulu | /bruce-dickinson/zulu-lulu.html | She fixed me in the corner with that beautiful brown eye She says, I'd like to meet you I was so sur... | en |
| /blink-182/ | Zulu | /blink-182/zulu.html | I'm feeling what I want to feel I'm saying what I've said all along Because I know the things that a... | en |
| /big-sean/ | ZTFO | /big-sean/ztfo.html | Nah man, you gotta be | en |

# Methods

## Dataset - table 2

Kumpulan lirik lagu beserta data genre. Dataset memiliki lebih dari 300,000 sampel lirik yang dikelompokkan kedalam 79 genre lagu yang berbeda.

| A Artist | A Genres | A Songs | A Popularity | A Link |
|---|---|---|---|---|
| Fernandinho | Gospel/Religioso; Pop/Rock | 139 | 9.9 | /fernandinho/ |
| Pharrell Williams | Pop; Hip Hop; Rap | 76 | 9.9 | /pharrell-williams/ |
| SOJA (Soldiers of Jah Army) | Reggae; Surf Music | 111 | 9.9 | /soja-soldiers-of-jah-army/ |
| Vanilda Bordieri | Gospel/Religioso | 244 | 9.8 | /vanilda-bordieri/ |
| Marisa Monte | Trilha Sonora; Romântico; MPB | 160 | 9.7 | /marisa-monte/ |
| Colbie Caillat | Pop; Romântico; Surf Music | 103 | 9.7 | /colbie-caillat/ |
| Florence And | Rock | 137 | 9.6 | /florence-and- |

# Methods

## Preprocessing dataset lirik lagu :

- Menghilangkan sample dengan nilai NaN

- Menghilangkan tanda-tanda kurung '(', '{', '['

- Menghilangkan lirik dengan panjang di atas 200 kata

- Menghilangkan lirik dengan bahasa selain inggris

- Membuat satu dataset untuk setiap genre lagu. Masing-masing dataset yang terbuat berisi lirik dari satu genre.

| | Lyric | Genres |
|---|---|---|
| How many roads must a man walk down\nBefore yo... | | Folk |
| Once upon a time you dressed so fine\nThrew th... | | Folk |
| Mama, take this badge off of me\nI can't use i... | | Folk |
| Pistol shots ring out in the ballroom night\nE... | | Folk |
| Hey! Mr. Tambourine Man, play a song for me,\n... | | Folk |
| Lay, lady, lay, lay across my big brass bed\nL... | | Folk |
| Come you masters of war\nYou that build the bi... | | Folk |
| May God bless and keep you always,\nMay your w... | | Folk |
| Come gather 'round people\nWherever you roam\n... | | Folk |
| "There must be some way out of here," said the... | | Folk |

Dataset genre "Folk" hanya berisikan lirik lagu dari genre folk.

# Methods

## GPT-Neo 125m

GPT-Neo adalah sebuah model pre-trained yang dapat memahami dan menghasilkan teks bahasa manusia.

Dibandingkan dengan model pre-trained lain seperti GPT-2, GPT-Neo lebih ringan dan lebih mudah untuk digunakan atau dilakukan finetuning untuk memenuhi kebutuhan NLP yang lebih spesifik.

# Methods

## Fine-tuning satu model GPT-Neo untuk setiap genre :

Masing-masing model pre-trained GPT-Neo dilatih kembali dengan menggunakan 2000 lirik lagu dari dataset genre yang digunakan.

```python
import torch
from transformers import TrainingArguments, Trainer, GPTNeoForCausalLM, GPT2TokenizerFast
```

```python
tokenizer = GPT2TokenizerFast.from_pretrained('EleutherAI/gpt-neo-125m', pad_token='<|pad|>')
model = GPTNeoForCausalLM.from_pretrained('EleutherAI/gpt-neo-125m')
model.resize_token_embeddings(len(tokenizer))
```

# Methods

## Generator lirik lagu dari model-model yang telah dilakukan fine-tuning :

```python
rock = GPTNeoForCausalLM.from_pretrained('/content/drive/MyDrive/NLP Finetuning/Finetuned Models/Rock Lyrics Generator')
pop = GPTNeoForCausalLM.from_pretrained('/content/drive/MyDrive/NLP Finetuning/Finetuned Models/Pop Lyrics Generator')
indie = GPTNeoForCausalLM.from_pretrained('/content/drive/MyDrive/NLP Finetuning/Finetuned Models/Indie Lyrics Generator')
hip_hop = GPTNeoForCausalLM.from_pretrained('/content/drive/MyDrive/NLP Finetuning/Finetuned Models/Hip Hop Lyrics Generator')
rap = GPTNeoForCausalLM.from_pretrained('/content/drive/MyDrive/NLP Finetuning/Finetuned Models/Rap Lyrics Generator')
```

```python
def generate_lyrics(genre, seed_text, number_of_lyrics):

    model = models.get(genre)

    model.to('cuda')

    generated = tokenizer(seed_text, return_tensors = 'pt').input_ids.cuda()

    output = model.generate(generated, do_sample=True, top_k = 10, max_length = 100, top_p = 0.95, temperature = 2., num_return_sequences = numb

    for i, sample_output in enumerate(output):
        text = tokenizer.decode(sample_output.tolist())
        regex_padding = re.sub('<|pad|>', '', text)
        regex_barra = re.sub('[|+]', '', regex_padding)
        space = re.sub('[ +]', ' ', regex_barra)
        result = re.sub('[\n](2, )', '\n', space)
        print("\n==== Generated lyric {}:\n\n{}".format(i+1, result))
```

# Result

```
generate_lyrics('Indie', 'In the deepest ocean', 3)
```

==== Generated lyric 1:

In the deepest ocean
My deepest love
Found in deepest lovel

I surrender everything
That I have been
I lay it all down
At the edge of my heart

Every tear will be wiped away
Every suffering will be wiped away
All crown my name

 deepest love
Endless striving
The deepest pain
I cannot leave behind
I will reach the ground
 deepest within

I surrender everything
That I have been
I lay it all down

==== Generated lyric 2:

In the deepest ocean
You're not where I want to be
I can't begin to know
but I feel a beginning
of some distant dream

Some in the crowd
noticed on the promises
other in the shadows
You came to me
and I felt Your love start
and I felt Your touch

And I feel Your touch
 touch of warmth and friend
And I feel Your touch
 touch of warm frailty
And I feel Your touch
 touch of far

==== Generated lyric 3:

In the deepest ocean
What remain
What remain
Still ishere
Here on the ocean

We will sing of this and this
We will sing of this and this

And no more questions
aling in us
We will sing of this and this

Love comes again
Just when we need
Two souls come to walk
One will come to know us
The love he needs is more than this

Love comes again
Just when we need
INGS another world

# Result

```
generate_lyrics('Rap', 'This is our', 3)
```

==== Generated lyric 1:

This is our song
And you can keep it jamin
We gonna rock it up
Till it runs off
Its filled with emotion
When your soul starts to rhyme
Hold me back

When your soul starts to rhyme
Hold me back

Im tired of rap realise
Give me a call grab a mic
And it will take it away
Hit the bricks and youll be ready
To get down the phone
Cock back exposed
Im on that

==== Generated lyric 2:

This is our December
This is our time of the year
This is our December
This is why we live today
This is our December
This is why we live today
This is our December
This is why we live today

This is our December
This is our time of the year
This is our December
This is why we live today
This is our December
This is why we live today
This is our December
This is why we live today

==== Generated lyric 3:

This is our year
Im makin shots like never
This is our year
Im makin shots like never
This is our year
Im makin shots like never

This is my shit
Im makin more shots than ever
This is my shit
Im makin more shots than ever

Eat that shit up pretzees
You know that Im aESH
That my beef homie showed no mercy
I eat that shit up pretzees
You know that

# Evaluation

Evaluasi model dilakukan dengan mengukur nilai perplexity dari kelima model yang telah dilatih (model genre rock, pop, indie, hip hop, dan rap).

Pengukuran perplexity dilakukan menggunakan metode *sliding-window strategy* ([ref.](#)).

# Evaluation

| Model Generator Lirik | Nilai Perplexity |
|---|---|
| Rock | 52.25 |
| Pop | 376.93 |
| Indie | 31.70 |
| Hip Hop | 9.73 |
| Rap | 7705.47 |

# Conclusion

- Kelima model memiliki nilai perplexity yang sangat bervariasi, yang mana kami yakini disebabkan oleh karakteristik genre yang berbeda-beda (Seperti halnya pop memiliki lebih banyak variasi tema lirik dibandingkan hip-hop)

- Dari penilaian subjektif, lirik lagu yang dihasilkan sudah cukup bagus untuk dapat memberi ide dan inspirasi dalam menulis lirik lagu

- Model mampu menghasilkan lirik yang berbeda antar genre yang berbeda.

- Performa model masih bisa ditingkatkan melalui pembersihan data yang lebih mendalam, dan penggunaan sistem dengan processing power yang lebih tinggi.