“GOODREVIEWS” BOOK REVIEW SUMMARIZATION

Matthew Holm, Kellen Mossner and David Nicolay
26067404, 26024284, 26296918
Data Science 346
Stellenbosch University

INTRODUCTION

Exploring GoodReads

- "GoodReviews " is our deep dive into the realm of GoodReads reviews to piece together what readers truly value about their favourite books.
- Through our analysis of thousands of reviews, we have applied sentiment analysis, crafted review summaries, predicted star ratings based off of reviews, and generated book descriptions solely on what readers think.
- We also apply additional sentiment analysis in the form of Aspect-Based Sentiment Analysis.

Agenda

1
Data Collection & EDA
- How our data was collected
- Data distributions

2
Sentiment Analysis
- Sentiment scores and outliers
- Sarcasm detection

3
Star Classification
- Model architecture (with dropout)
- Prediction

4
Summarization
- Encoder clustering
- BERT model explanation

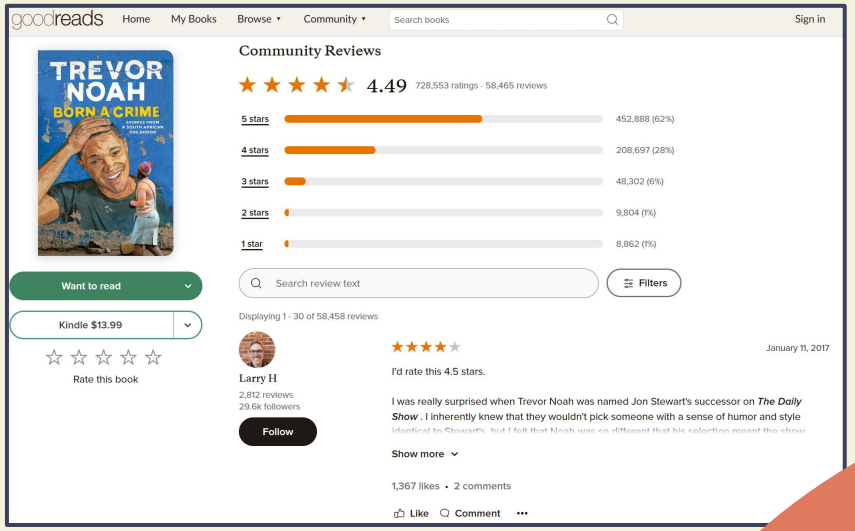
5
Generation
- Process of creating descriptions
- BART model explanation

6
ABSA
- Model hierarchy & explanation
- Final Application

Data Collection

DATA COLLECTION

- Created a web scraper in Python using Selenium in order to scrape javascript content.
- Collected a large corpus of data by scraping the reviews of over 200 books and scraping 120 reviews per book.
- Our total data gathered consists of 29,520 reviews.

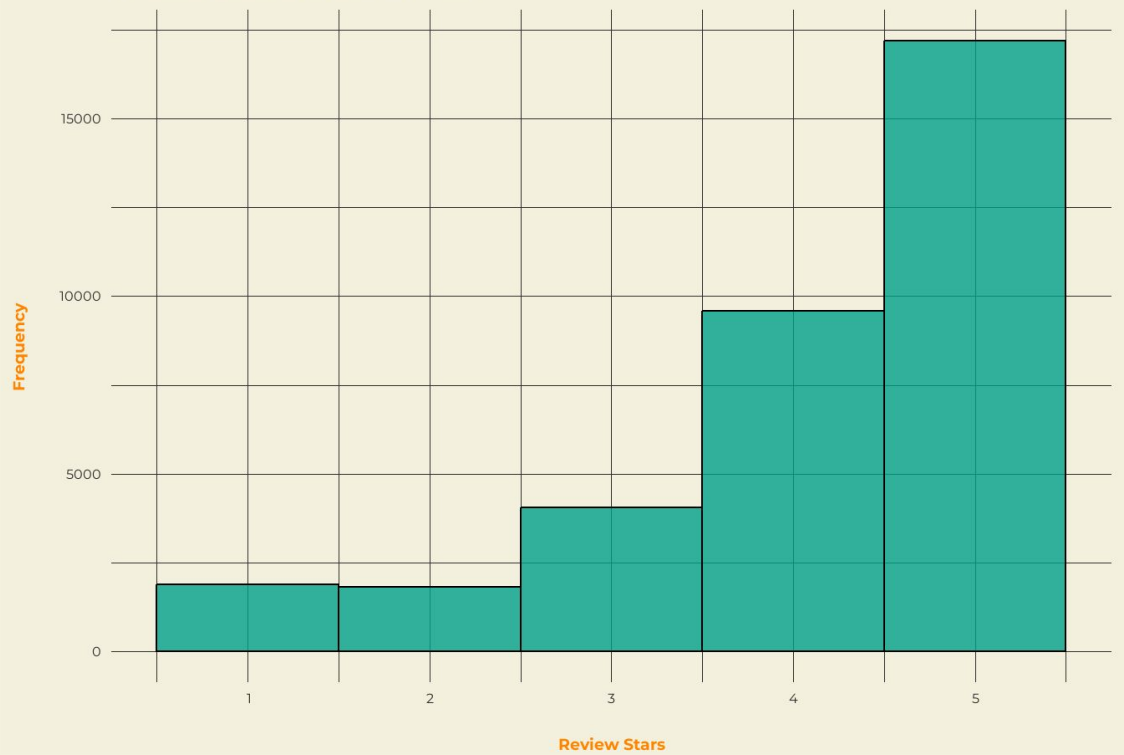


DATA OVERVIEW

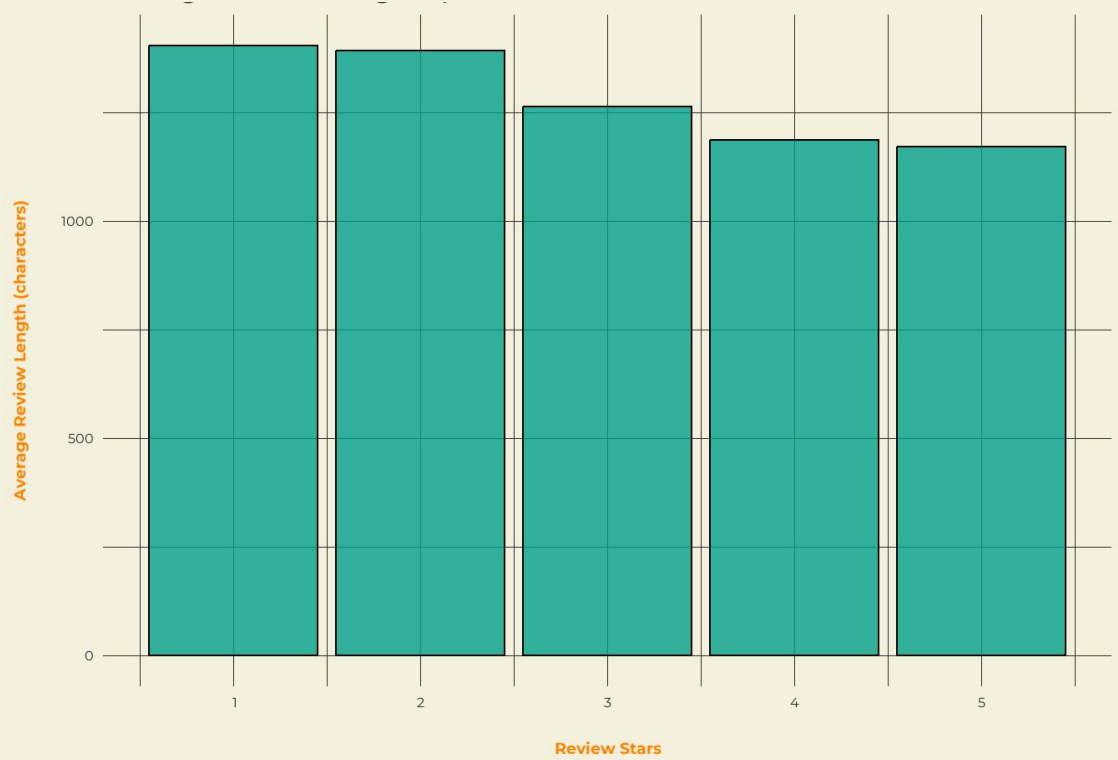
Book Title	Link	Review Text	Review Date	Review Stars	Review Likes	Genres	First Published Date	Author
Ways of Seeing	https://www.goodreads.com/book/show...	This book is based on a television series whic...	September 29, 2014	5	513	Art, Nonfiction, Philosophy, Essays, Art History...	January 1, 1972	John Berger
Cleopatra: A Life	https://www.goodreads.com/book/show...	“Among the most famous women to have lived, Cl...	April 3, 2021	4	201	History, Biography, Nonfiction, Egypt, Historical...	November 1, 2010	Stacy Schiff
Yes Please	https://www.goodreads.com/book/show...	This just may be my favorite autobiography by ...	December 16, 2015	4	20	Nonfiction, Memoir, Humor, Audiobook, Biograph...	October 28, 2014	Amy Poehler

Exploratory Data Analysis

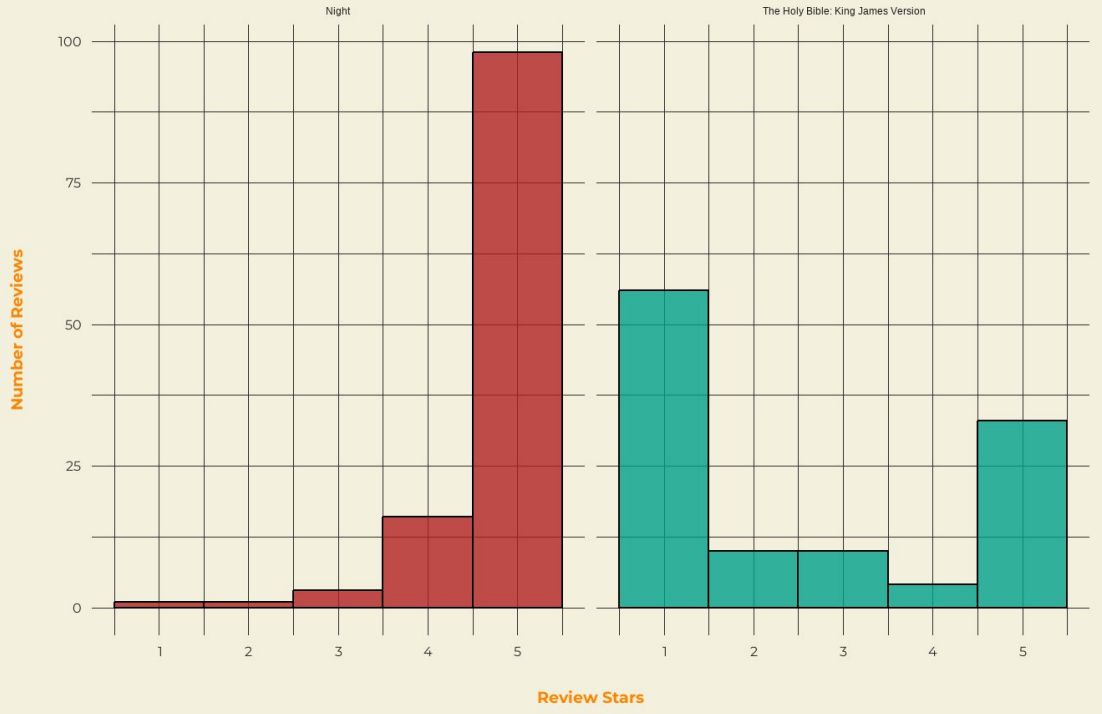
Distribution of Review Stars



Average Review Lengths per Star

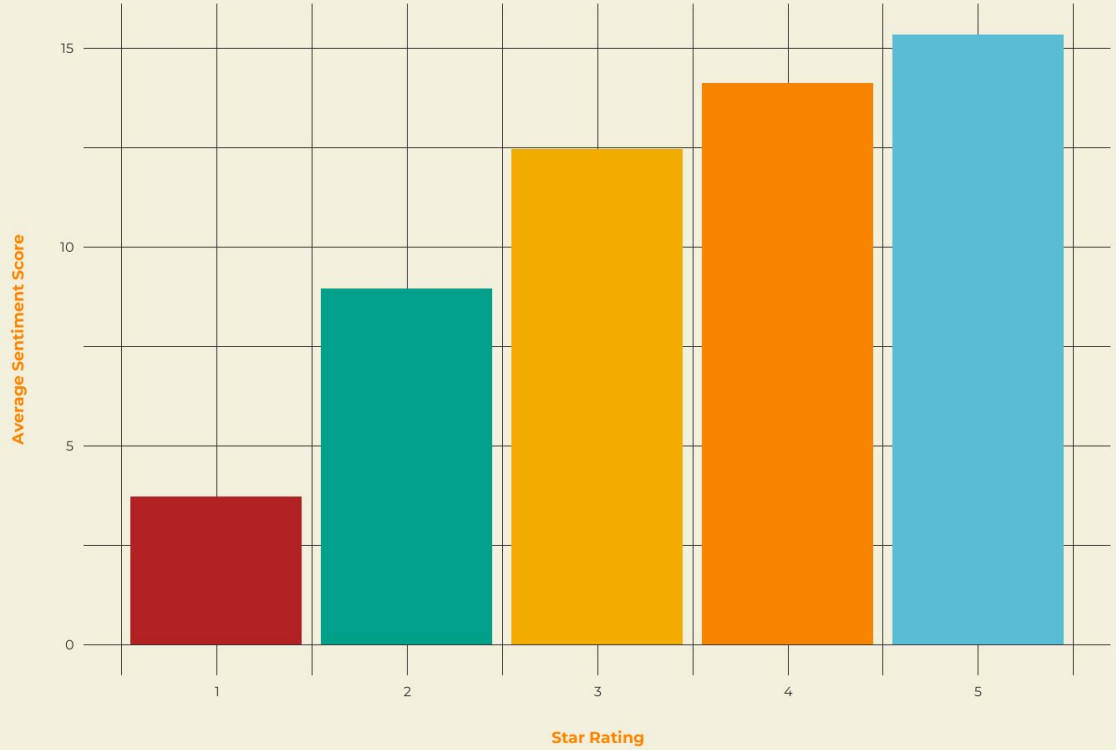


Distribution of Review Stars for Top and Worst Books



Sentiment Analysis

AFINN Sentiment Count in Book Reviews



Lowest Sentiment

Book Title	Sentiment	Average Rating
Night	-367	4.77
Spare	-197	3.40
Bad Blood: Secrets and Lies in a Silicon Valley Startup	-124	4.38
A Grief Observed	-114	4.44
Angela's Ashes	0	4.02
The Lonely City: Adventures in the Art of Being Alone	20	4.025
Unbroken: A World War II Story of Survival Resilience and Redemption	86	4.44
The Autobiography of Malcolm X	123	4.58
Narrative of the Life of Frederick Douglass	130	4.68
Camera Lucida: Reflections on Photography	222	3.90

Highest Sentiment

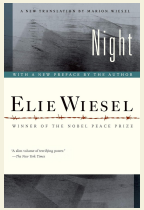
Book Title	Sentiment	Average Rating
Harry Potter and the Sorcerer's Stone	7776	3.87
The Lion, the Witch and the Wardrobe	7584	4.23
Leonardo da Vinci	6840	4.11
Jane Eyre	6498	4.50
The Lightning Thief	6270	4.13
Shoe Dog: A Memoir by the Creator of Nike	6256	4.24
Harry Potter and the Chamber of Secrets	6024	4.59
The Magician's Nephew	5328	4.27
Harry Potter and the Prisoner of Azkaban	5176	4.74
Anne of Green Gables	5000	4.61

Genres for Night

Nonfiction

Classics

Memoir



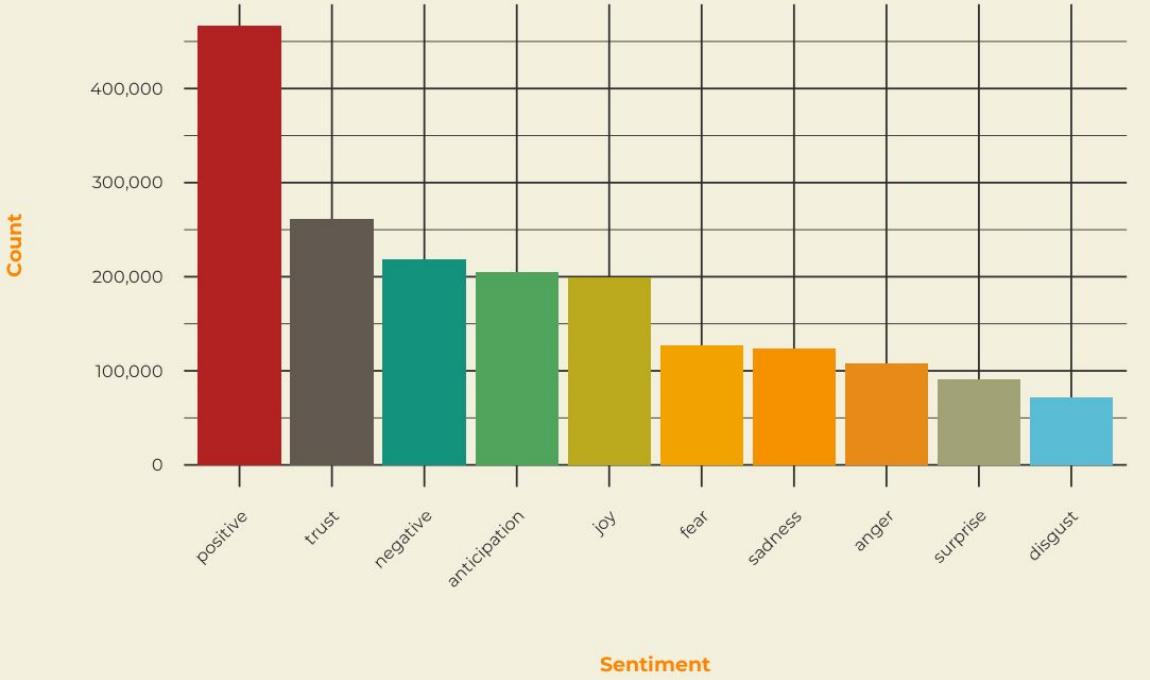
Book Title	Sentiment	Average Rating
Night	-367	4.77

Holocaust

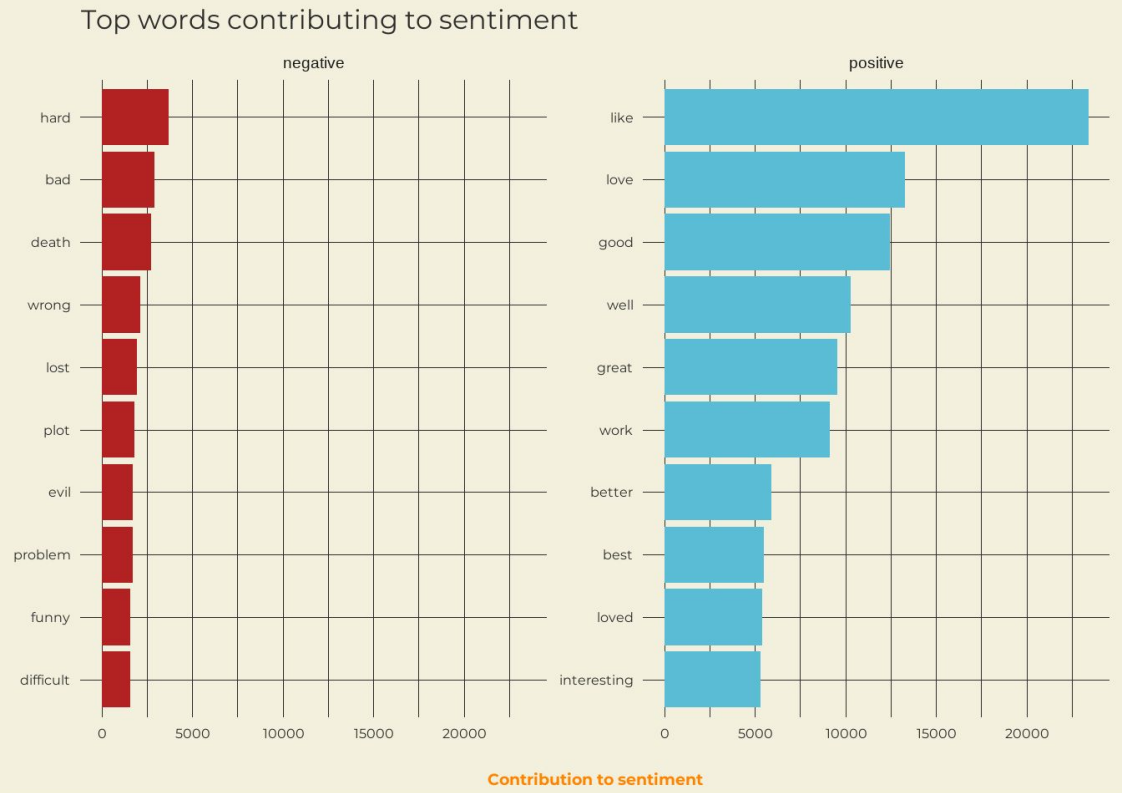
Biography

Historical

Sentiment Count in Book Reviews using the NRC Lexicon



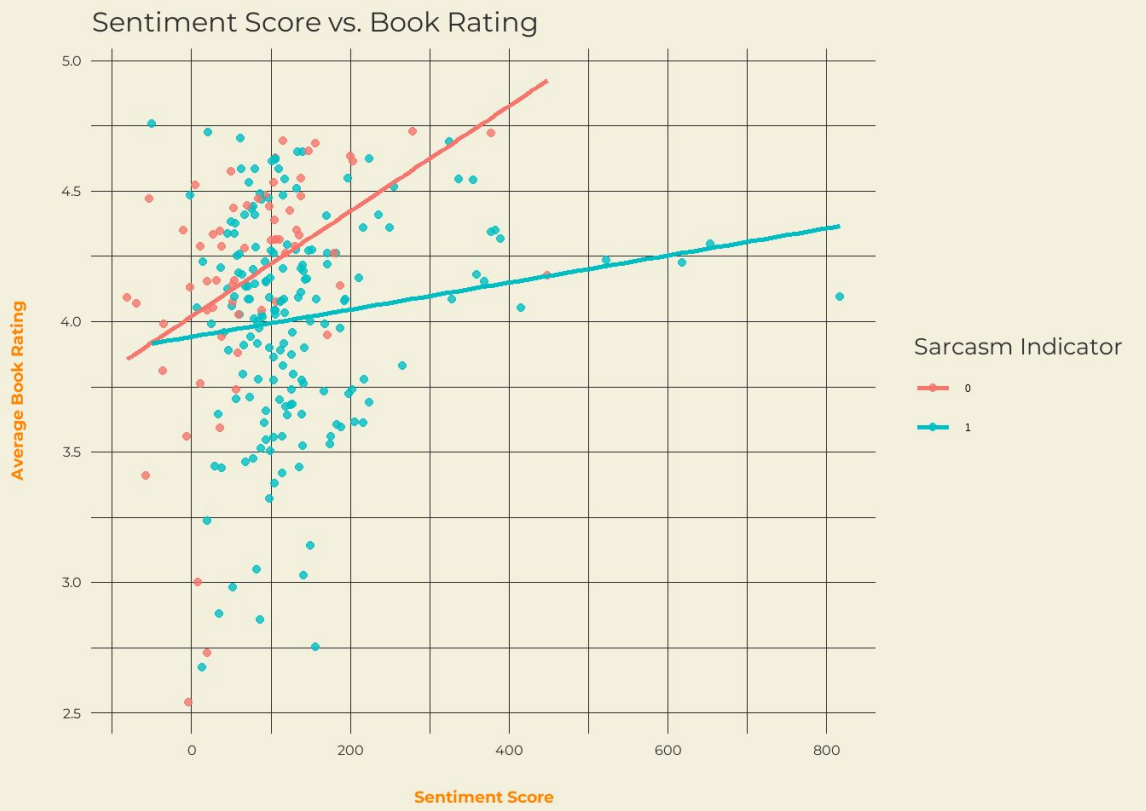
Bing Lexicon Sentiment Analysis



Sarcasm Detection

- Reviews with positive sentiment scores but low review stars could indicate that there is some form of expression where the intended meaning differs from the literal meaning.
- The best way to detect sarcasm in this case would be to examine these reviews ourselves - since sarcasm detection with sentiment alone is quite challenging.

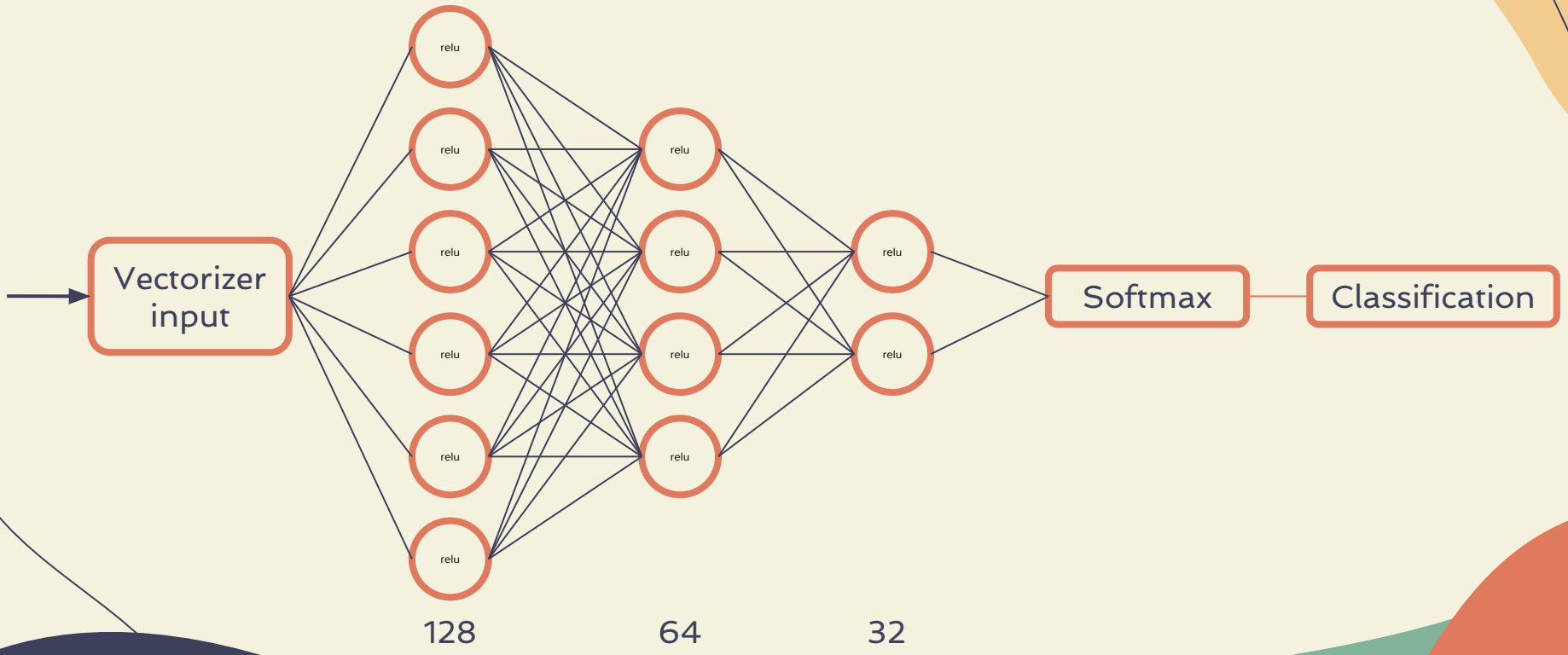
Potential Effects of Sarcasm



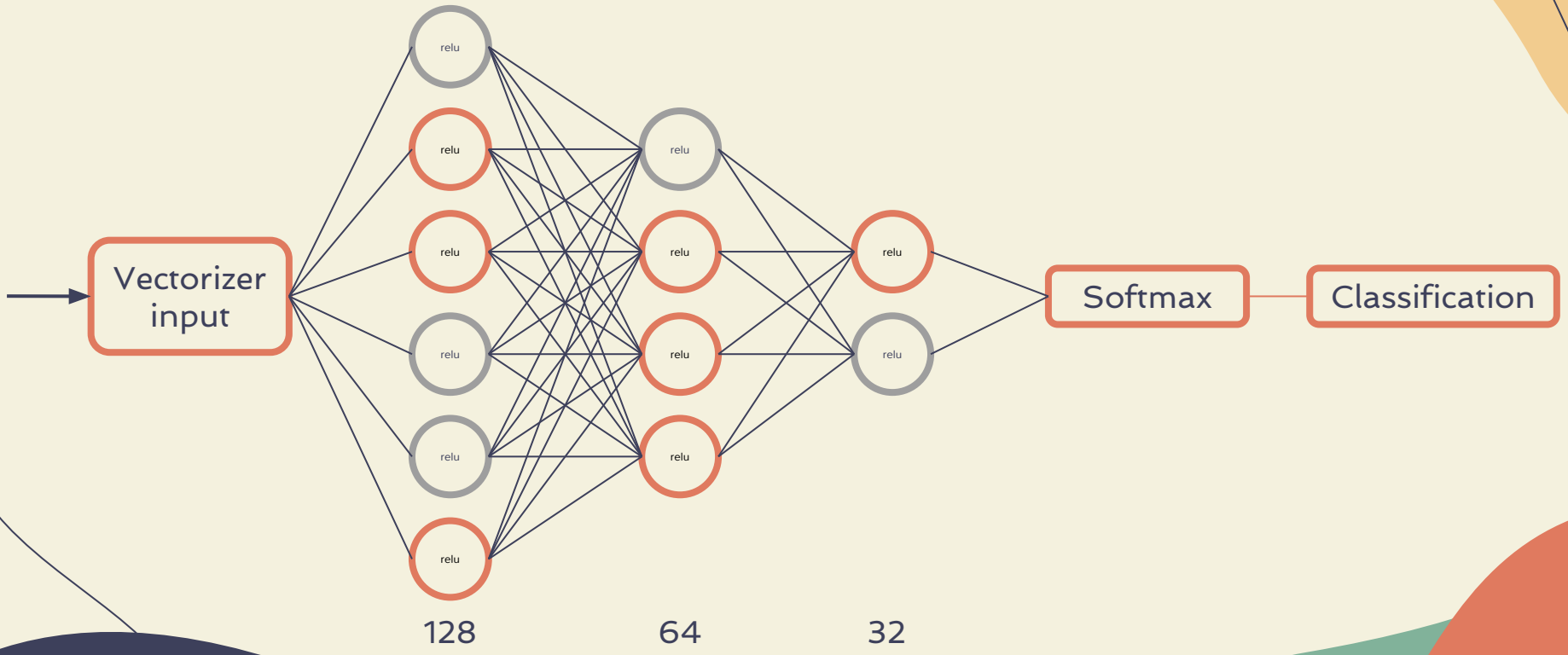
Star Classification

Can a Neural Network be used to accurately predict the number of stars of a review based on the text?

ARCHITECTURE



ARCHITECTURE w/DROPOUT



STAR PREDICTION

TYPE	REVIEW	PREDICTED
BAD	“This book was absolutely terrible! How could you think this was a good idea.”	1
GOOD	“This book was a captivating read from start to finish. The characters felt incredibly real, and the plot twists kept me on the edge of my seat. I couldn't put it down and will definitely be recommending it to everyone!”	5
LONG, MIXED	“I had high hopes for The Infinite Horizon after hearing so much about it. From the beginning, the premise seemed promising, and for the most part, it delivers on its intriguing concept. The plot revolves around a futuristic world where society grapples with the boundaries of artificial intelligence, humanity, and survival—concepts that have always fascinated me. The world-building is impressive, with detailed landscapes and a unique societal structure that keeps you hooked initially. The author has clearly put a lot of thought into constructing the futuristic world, and it shows in the vivid descriptions and creative technologies. However, while the world-building is rich, the characters left much to be desired. The protagonist, Lila, felt underdeveloped. I found myself frustrated at several points because her motivations were either unclear or inconsistent. In the beginning, she starts off as a strong, determined character, but midway through, her actions seem erratic and her growth stagnates. The dialogue, too, felt stilted at times, making it hard to connect with the characters emotionally. There were a few moments where I felt the conversations between key characters were forced, almost like they were inserted to explain plot points rather than feeling organic.On the flip side, I have to give credit where it's due—the pacing of the story is solid for the most part. There are intense moments where you're on the edge of your seat, particularly during the battle scenes. These scenes were written with such vivid detail that I could easily imagine them playing out in a movie. The action sequences are well thought out, and they definitely add excitement to the narrative. That being said, there were also moments where the pacing lagged, especially in the middle sections. Some chapters felt like filler, dragging on with unnecessary exposition and side plots that didn't add much to the overarching story.”	3

Summarization

ENCODER CLUSTERING

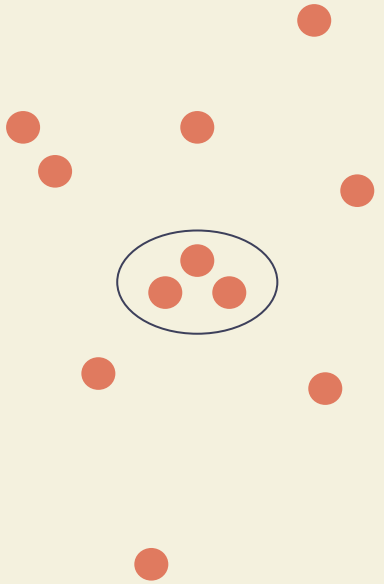
01 **Sentence-BERT**
all-MiniLM-L6-v2

02 **FLATTEN**
Into a single list

03 **ENCODE**
Into embeddings

04 **CLUSTER**
KMeans on sentence embeddings

05 **SELECT**
Closest to cluster center



SENTENCE BERT

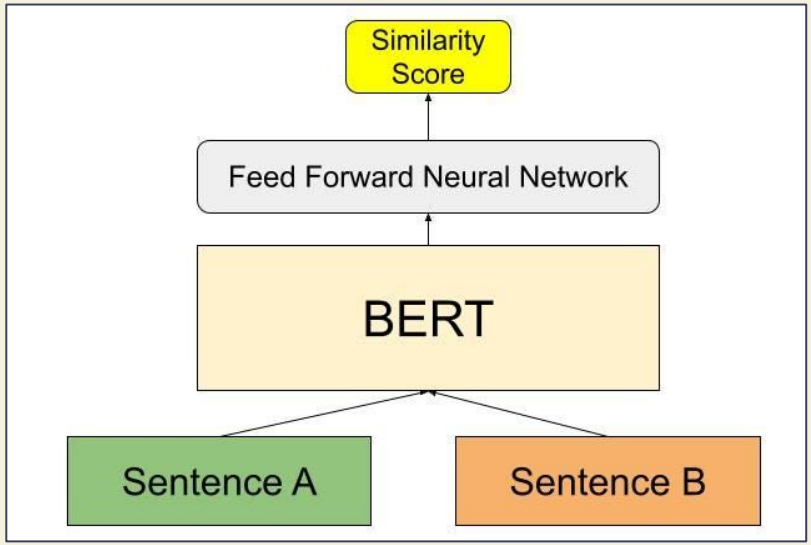


Image Source

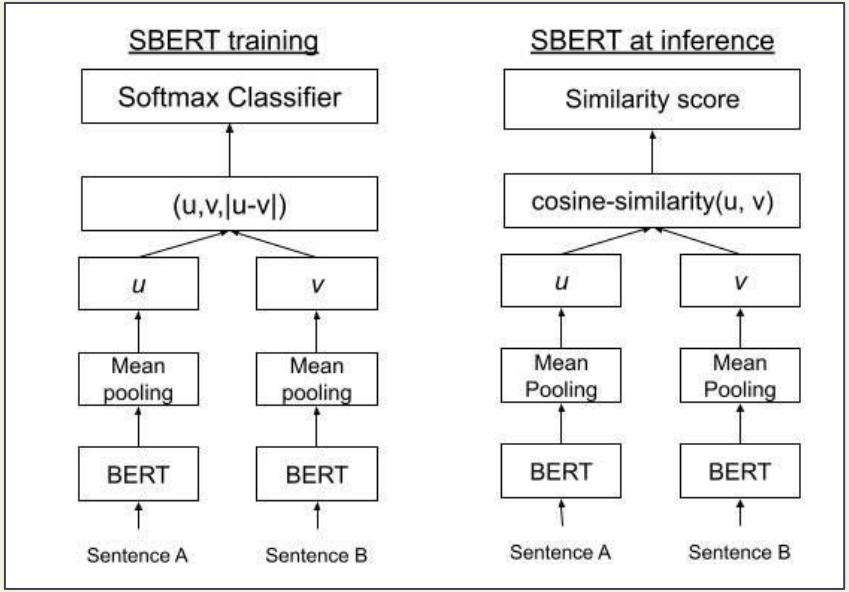


Image Source

Generation

Generation
Book Descriptions

Summarization
Create summaries for each review for a book

Final Summary
Use summarized reviews to form one (longer) final summary

Book Description
Feed final summary into pre-trained BART model with a prompt

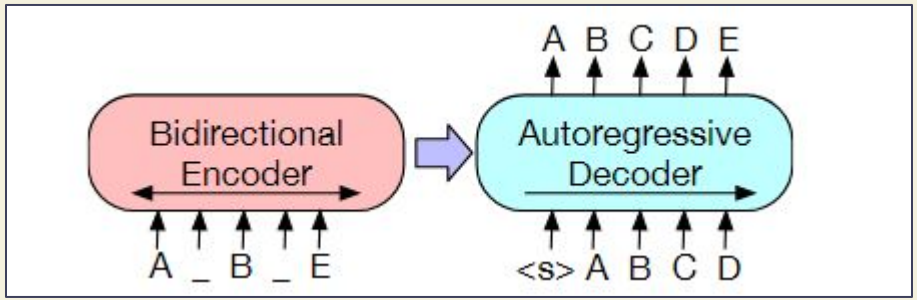
Example: Born a Crime by Trevor Noah

"Trevor Noah's love and respect for his mother & the way she raised him shines through on nearly every page. Eye-opening and perspective changing in a way that's funny and deeply vulnerable, you'll feel educated and entertained at the same time. For an enhanced experience, I highly recommend the audiobook version. Moved out of the house at the age of 17 because of his step-dad and was even jailed for using a fake license plate. Imagine being born from a black mother and a white father in a country where interracial relationships were against the law."

Generation

The BART model

- BART (Bidirectional Auto-Regressive Transformers) is a denoising autoencoder for pretraining seq-to-seq models.
- In less technical terms, BART is trained by first corrupting text with arbitrary noise functions and then learning a model to reconstruct the original uncorrupted text.
- BART is specifically designed for text generation and summarization.



ABSA

Aspect-Based Sentiment Analysis

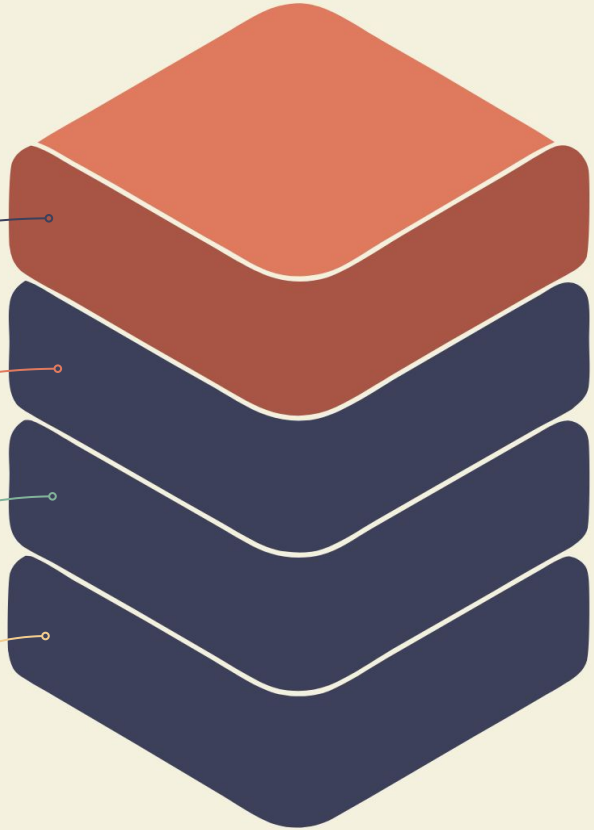
MODEL HIERARCHY

ReviewABSA 4
Our implementation

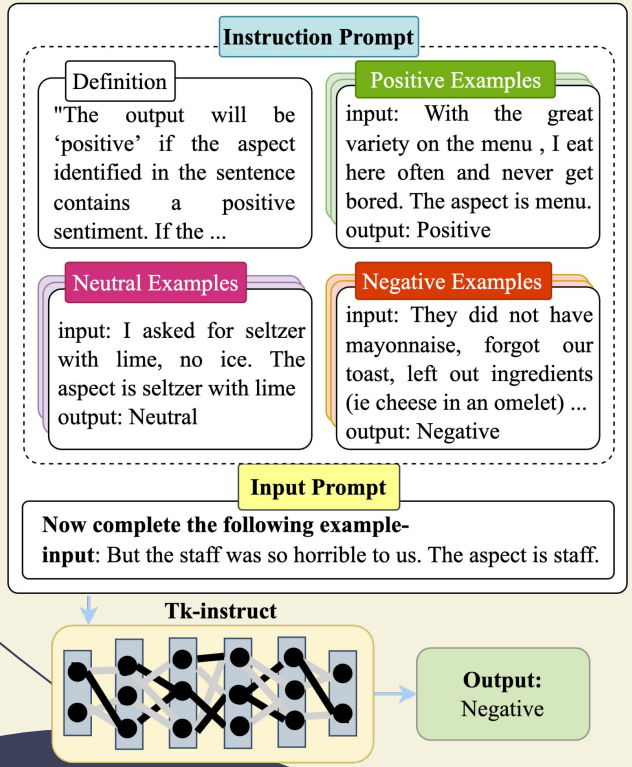
InstructABSA 3
+/- added to training

T*k*-Instruct 2
Instruction tuning

T5 1
Text-To-Text Transformer



INSTRUCT ABSA



sp^1 negative — sp^2 positive

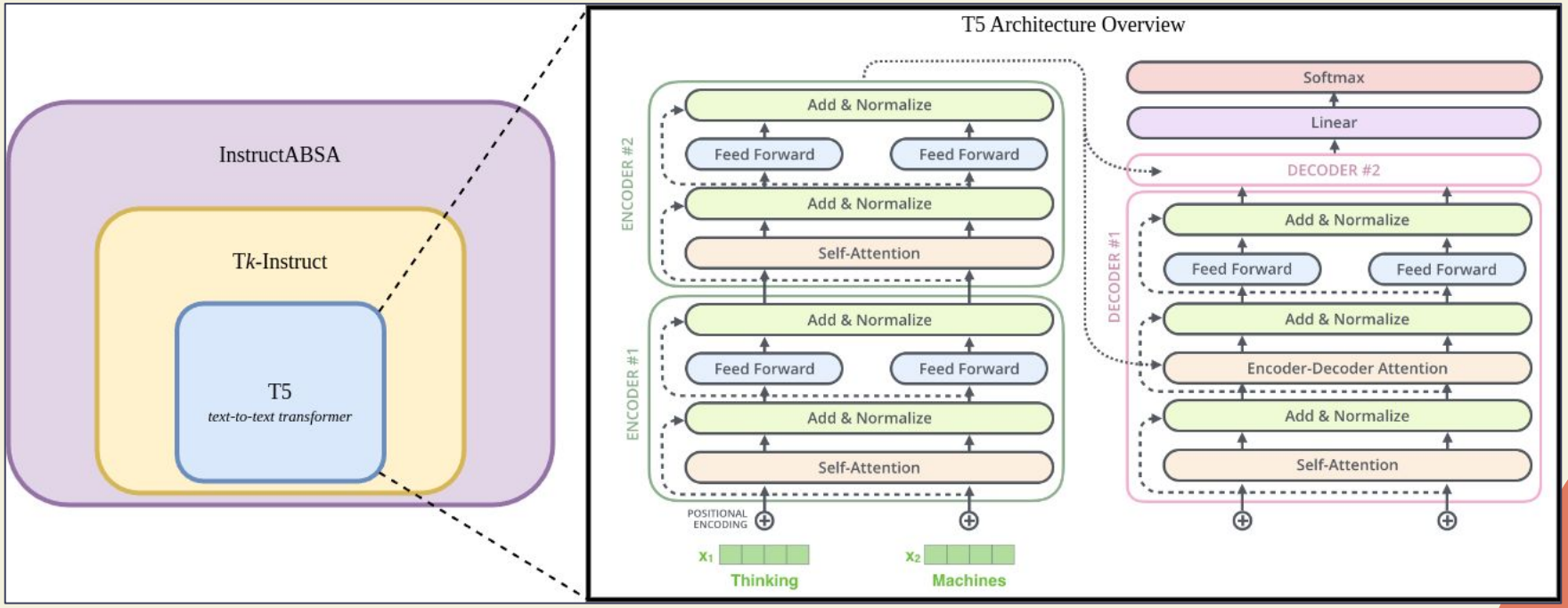
S_i: The *price* was too high, but the *cab* was *amazing*.

a^1 o^1 a^2 o^2

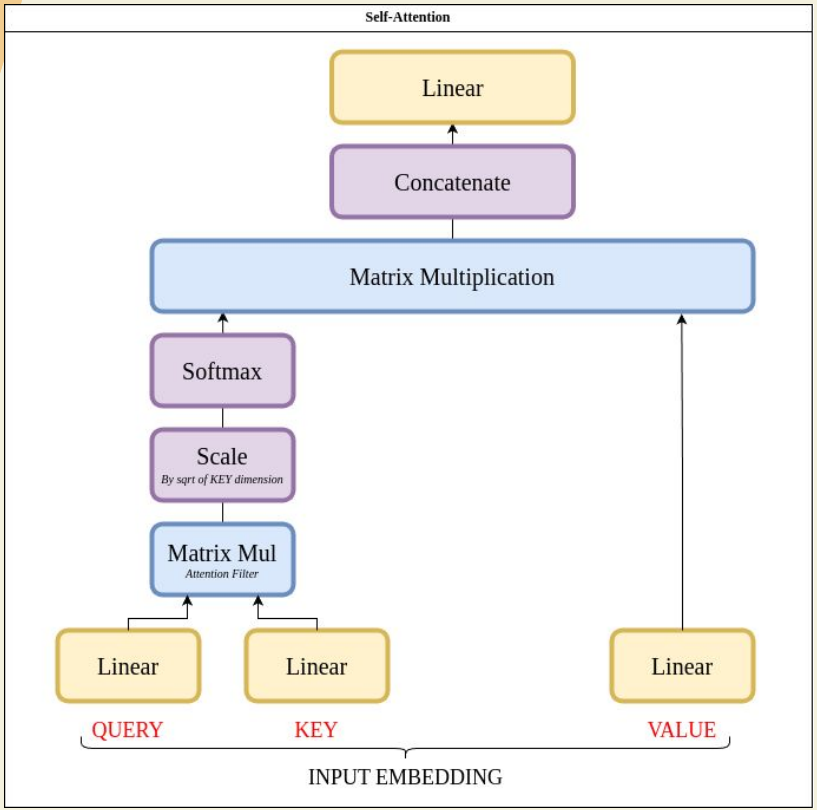
Subtask	Input	Output
Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1$, $S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)	S_i	(a^1, sp^1), (a^2, sp^2)
Aspect Oriented Opinion Extraction (AOOE)	$S_i + a^1$, $S_i + a^2$	o^1, o^2
Aspect Opinion Pair Extraction (AOPE)	S_i	(a^1, o^1), (a^2, o^2)
Aspect Opinion Sentiment Triplet Extraction (AOSTE)	S_i	(a^1, o^1, sp^1), (a^2, o^2, sp^2)

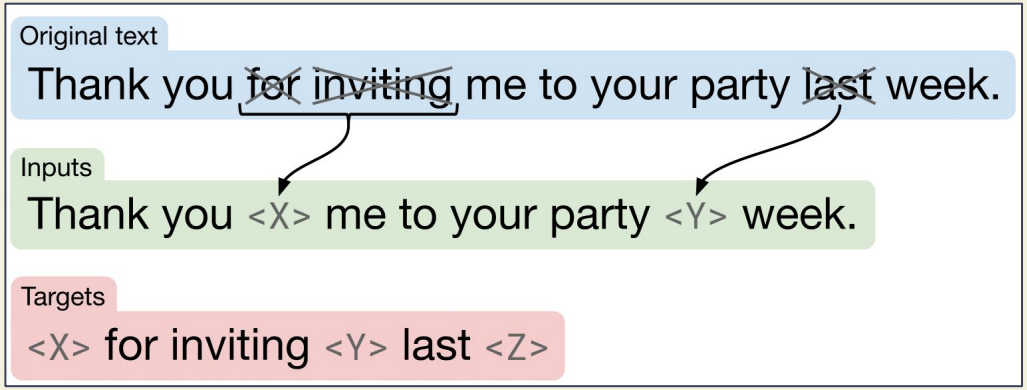
Image Source: InstructABSA

T5 ARCHITECTURE



SELF ATTENTION + TRAINING





$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

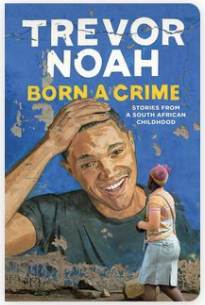
FINAL APPLICATION

ABSA

SUMMARY

Creature Feature From horror to romance, new books with monsters galore

goodreads Home My Books Browse Community Search books Sign in



Want to read

Kindle $13.99

Rate this book

Born a Crime: Stories From a South African Childhood

Trevor Noah

4.49 728,553 ratings · 58,465 reviews

The memoir of one man's coming-of-age, set during the twilight of apartheid and the tumultuous days of freedom that followed.

Trevor Noah's unlikely path from apartheid South Africa to the desk of The Daily Show began with a criminal act: his birth. Trevor was born to a white Swiss father and a black Xhosa mother at a time

Show more

Genres Nonfiction Memoir Biography Audiobook Humor Autobiography Africa ...more

Readers Praise Humor Storytelling Emotion

Readers Dislike Structure Setting

Reviews Summary

Trevor Noah's love and respect for his mother & the way she raised him shines through on nearly every page. Eye-opening and perspective changing in a way that's funny and deeply vulnerable, you'll feel educated and entertained at the same time. For an enhanced experience, I highly recommend the audiobook version. Moved out of the house at the age of 17 because of his step-dad and was even jailed for using a fake license plate. Imagine being born from a black mother and a white father in a country where interracial relationships were against the law.

289 pages, Hardcover

First published November 15, 2016

Book details & editions

30.3k people are currently reading

555k people want to read

About the author

THANK YOU