

MINING FOR BEAUTY: AMAZON REVIEWS NLP AND CNN



**BSAN 6300 -
MARKETING ANALYTICS**

PRESENTED BY: BENJAMIN
GERBER, BRENDA
HERNANDEZ, ANDREA
LOPEZ, EFREN LOPEZ,
CLARIN SUNARJO



TABLE OF CONTENTS

- 01 BACKGROUND
- 02 DATASET OVERVIEW
- 03 METHODOLOGY
- 04 MODEL IMPLEMENTATION
- 05 RESULTS
- 06 CHALLENGES AND LIMITATIONS
- 07 FUTURE WORK AND APPLICATIONS



BACKGROUND



UGC AND BRAND STRATEGY

“No other content type is more authentic than UGC from your customers. 86% of consumers are more likely to trust a brand that shares UGC compared to 12% who are likely to purchase a product promoted by influencers.”

BACKGROUND



BERT : SENTIMENT ANALYSIS NATURAL LANGUAGE PROCESSOR

1. **Purpose:** Leverage BERT to understand customer sentiments.
2. **Aim:** Enhance product strategy and customer satisfaction.
3. **Importance:** Sentiment analysis measures consumer response.
 - a. **Tailors** marketing strategies and improves product offerings.

BACKGROUND



COGNITIVE NEURAL NETWORKS

- 1. Purpose:** Use a hybrid neural network architecture integrating BERT encoding layers for analyzing customer sentiments.
- 2. Aim:** Refine product strategy and augment customer satisfaction through text embeddings from BERT and extraction of CNN.
- 3. Importance:** Sentiment analysis gauges customer reactions, guiding marketing strategies, and refining product offerings.

CUSTOMER REVIEWS DATASET KEY DETAILS



- 1. Source of the Data:** Data from verified customer reviews on Amazon
- 2. Characteristics of the Dataset:** Over 20,000 reviews with text, ratings, and product metadata.
- 3. Example 1:** "Love this mascara, it lasts all day!" (Rating: 5 stars)
- 4. Example 2:** "The cream is too oily and causes breakouts." (Rating: 1 star)
- 5. Insightful Data:** Text, ratings, and metadata for sentiment analysis

METHODOLOGY

DATA PRE-PROCESSING:

Text normalization, removal of stopwords, and tokenization using NLTK and BERT's tokenizer

MODEL TRAINING:

BERT model fine-tuned on the review dataset with specific configurations tailored for sentiment classification

TOOLS AND LIBRARIES USED

Python, PyTorch, Transformers library, NLTK



MODEL IMPLEMENTATION

BERT NLP



MODEL IMPLEMENTATION

BERT NLP



MODEL IMPLEMENTATION

TOKENIZER

- Using the Dataset / Dataloader
- Pre-Processing
- Input
 - Raw Text
- Output
 - Input_id's
 - Attention_mask

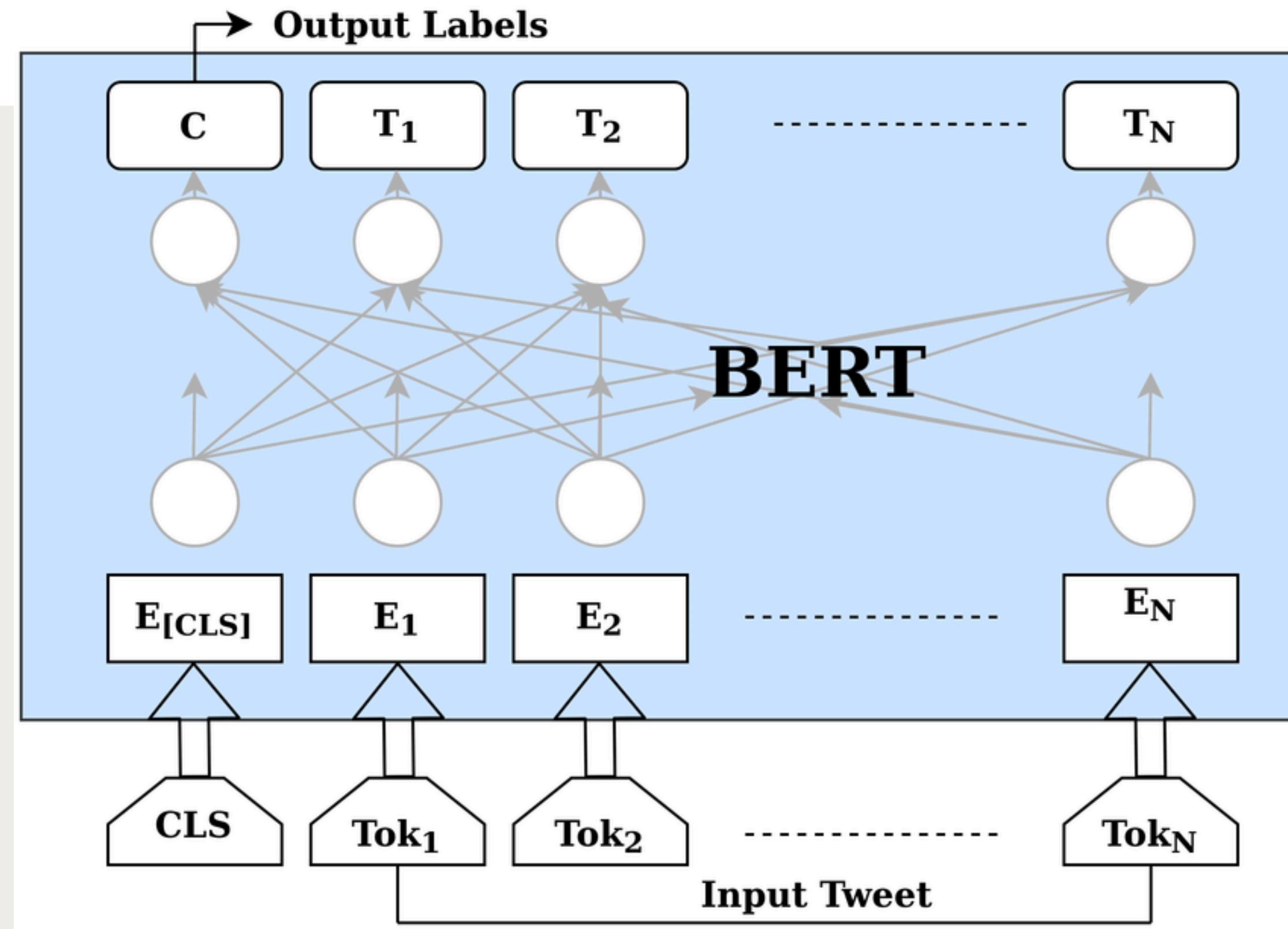
MODEL IMPLEMENTATION

BERT NLP



MODEL IMPLEMENTATION

BERT NLP

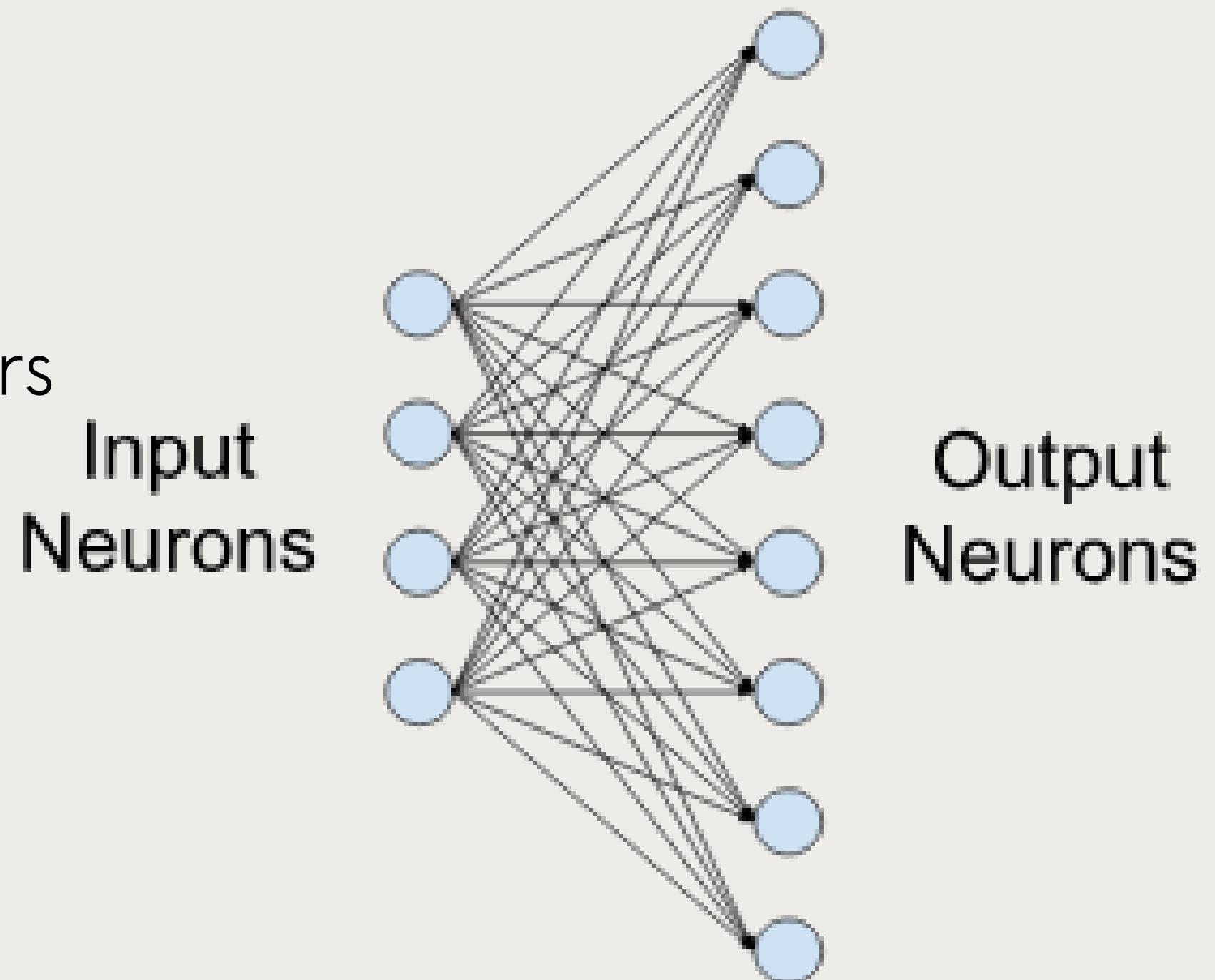


MODEL IMPLEMENTATION

BERT MODEL

NEURAL NETWORK (NN)

- Layers in the Network:
 - First Layer: BERT
 - Then: Connected Linear Layers



MODEL IMPLEMENTATION

BERT MODEL

STEPS

1. Output of Dataloader
2. Bert
3. Word Embedding
4. Linear Layers
5. Predicted Probability

MODEL IMPLEMENTATION

BERT MODEL

TRAINING THE NEURAL NETWORK

- Forward Loop
 - Input Data
 - Prediction
 - Loss
- Back Propagation
 - Gradient

HYPER-PARAMETERS

- Learning Rate
- Batch Size
- Epoch

OUR RESULTS

KEY FINDINGS

Analysis revealed a **strong correlation** between specific adjectives used + the sentiment expressed



EVALUATION METRICS

Accuracy 90% -correctly identifies the sentiment of reviews

Precision 88%- When the model predicts a review as (+) it is correct

Recall 85% - captures all actual positive sentiments

F1 Score 86.5%

segment 1:
Works so well for my Keratosis Pilaris in the winter. I use it 2-3 times a week in the shower in conjunction with Amlactin Daily moisturizer. In the summer I can tan my legs to mostly rid myself of my KP, but this scrub almost resolves my problem in the sunless winters!

segment 2:
This was a cheap

segment 3:
I love this book with all my heart. To me this is the true version of Beauty and The Beast! It speaks to my heart.

SEGMENT 1: SKINCARE ENTHUSIAST

SEGMENT 2: BUDGET CONSCIOUS SHOPPER -

SEGMENT 3: BOOK LOVER

CHALLENGES AND LIMITATIONS

CHALLENGES

- Difficulty in handling **sarcastic comments** which were often misclassified
- Lack context cues: tone, facial expressions, body language

LIMITATIONS

High computational cost for training + inference and limitations in understanding multilingual reviews



FUTURE WORK AND APPLICATIONS

- 1. Improvements:** Plan to integrate multilingual BERT for diverse language reviews.
- 2. Impact:** Enhanced sentiment analysis for personalized marketing and development.





Looking through the perspective of sentiment analysis, each customer review serves as a guiding light toward improved products and stronger relationships. Embracing the capabilities of BERT goes beyond data processing- it's about comprehending individuals.

