

# Less is More: Human and AI Summarization for Sentiment Analysis

Pablo T. Campos  
Laura Martínez

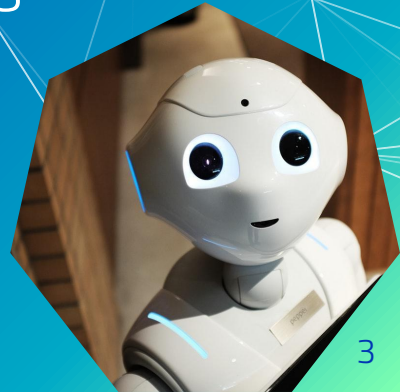
[campos72@purdue.edu](mailto:campos72@purdue.edu)  
[mart2180@purdue.edu](mailto:mart2180@purdue.edu)

# Introduction

- Sentiment Analysis is of great importance for ecommerce. With companies having to analyze great bodies of text, it would be cheaper to analyze summaries of said texts.
- However not all bodies of text come with a summary provided by the user. Could this summary be efficiently produced by a computer rather than by a human?

# Research Question

- Can we perform sentiment analysis on summaries instead of the original texts?
- If so, may AI generated summaries convey more precise information than human summaries about the original texts?



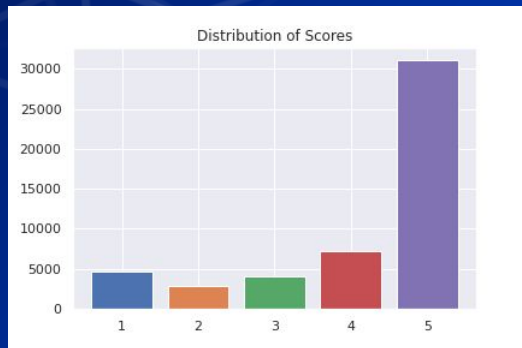
# Modeling

- Fine tune a transformer model to perform abstractive summarization
- Changed transformer model from BERT to T5 for computational reasons
- For sentiment analysis, BoW logistic regression and a LSTM with randomly initialized embeddings
- Used only a sample of the original dataset for computational reasons (10%)

# Experimental Setup: Dataset

## ■ Amazon Fine Food Reviews

Text	Summary	Score	Label
Review left by user	Summary of the review by said user	Score of the product, ranging from 1-5	Scores 1 & 2: 0 (Negative) Scores 3: 1 (Neutral) Scores 4 & 5: 2 (Positive)



# Experimental Setup: Tasks

- Abstractive Summarization

- Fine tune pretrained transformer model on Text with target Summary
- Generate summaries for all instances: Generated Summary

Text	Summary	Generated Summary	Label
Review left by user	Summary of the review by said user	Summary generated by fine tuned transformer model from attribute Text	Scores 1 & 2: 0 (Negative) Scores 3: 1 (Neutral) Scores 4 & 5: 2 (Positive)



# Experimental Setup: Tasks

## ■ Sentiment Analysis

- 3 Logistic regression models trained with BoW features on Text, Summary and Generated Summary to predict Label (Baseline model)
- 3 LSTM model trained on randomly initialized embeddings on Text, Summary and Generated Summary to predict Label

# Experimental Results

## ■ Abstractive Summarization

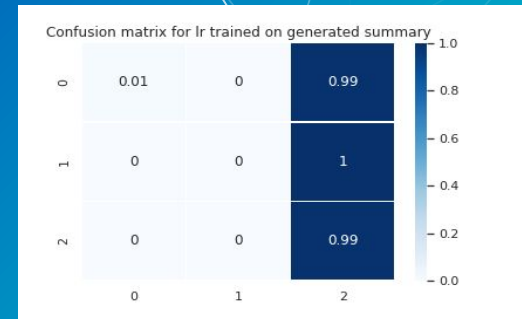
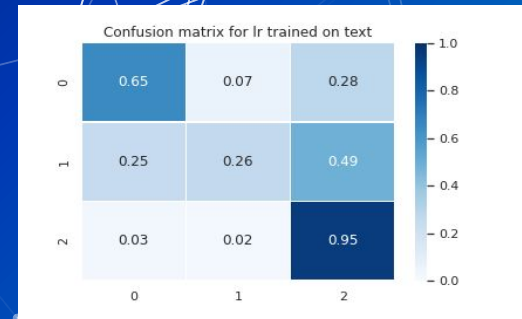
Summary	Generated Summary
Good Quality Dog Food	Great product
Not as Advertised	Jumbo Salted Peanuts - I had this!
"Delight" says it all	The most flavorful treat
Cough Medicine	Best product in the industry
Great taffy	Great taffy





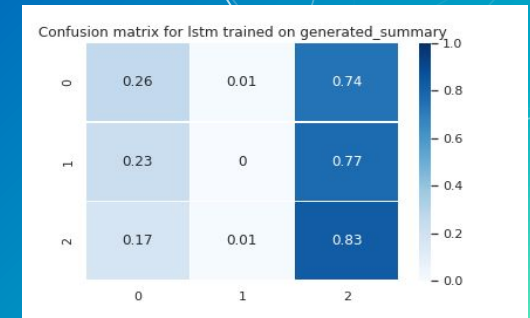
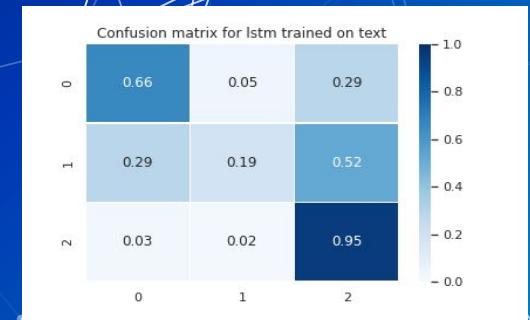
# Experimental Results

- Sentiment Analysis
  - Logistic Regression



# Experimental Results

- Sentiment Analysis
  - LSTM



# Discussion

- Both models trained with either Summary or Generated Summary performed worse than those trained on Text
- LSTM model trained on Generated Summary outperformed its counterpart trained on Summary
- Logistic Regression trained on Generated Summary performed worse than its counterpart trained on Summary

# THANKS!

**Any questions?**

You can find us at

- [campos72@purdue.edu](mailto:campos72@purdue.edu) Pablo T. Campos
- [mart2180@purdue.edu](mailto:mart2180@purdue.edu) Laura Martínez