



Analyzing Sentiments of Fine Food Reviews on Amazon

Tarani Rupa Neelapu
COMPSCI 715 Programming for ML



Table of Contents

- Introduction
- Methodology
- Results and Evaluation
 - Conclusion
 - Future Work

Introduction

What?

- Analyze customer sentiments on Amazon fine food reviews using VADER and Logistic Regression

Why?

- To understand customer feedback for product improvements
- Performing sentiment analysis will help product owners understand customer satisfaction and dissatisfaction

How?

- Use **VADER** Sentiment Analysis.
- Use **Logistic Regression** Model.
- Compare the results by both models.

Methodolgy



Step 1: Data
Collection



Step 2: Data pre-
processing



Step 3: Basic
Exploratory Data
Analysis



Step 4: Applying
VADER Sentiment
Aanalysis



Step 5: Applying
Logistic Regression
Model



Step 6: Evaluating
Results

VADER Sentiment Analysis

- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based sentiment analysis tool specifically designed to analyze social media texts.
- It is a pre-trained sentiment analysis model that provides sentiment scores for given text
- VADER has 4 categories positive, negative, neutral and compound
- Compound is a calculation of positive, negative and neutral score ranging between -1 to 1
- VADER is easily available as a part of the NLTK (Natural Language Toolkit) library in python.

Results by VADER Sentiment Analysis

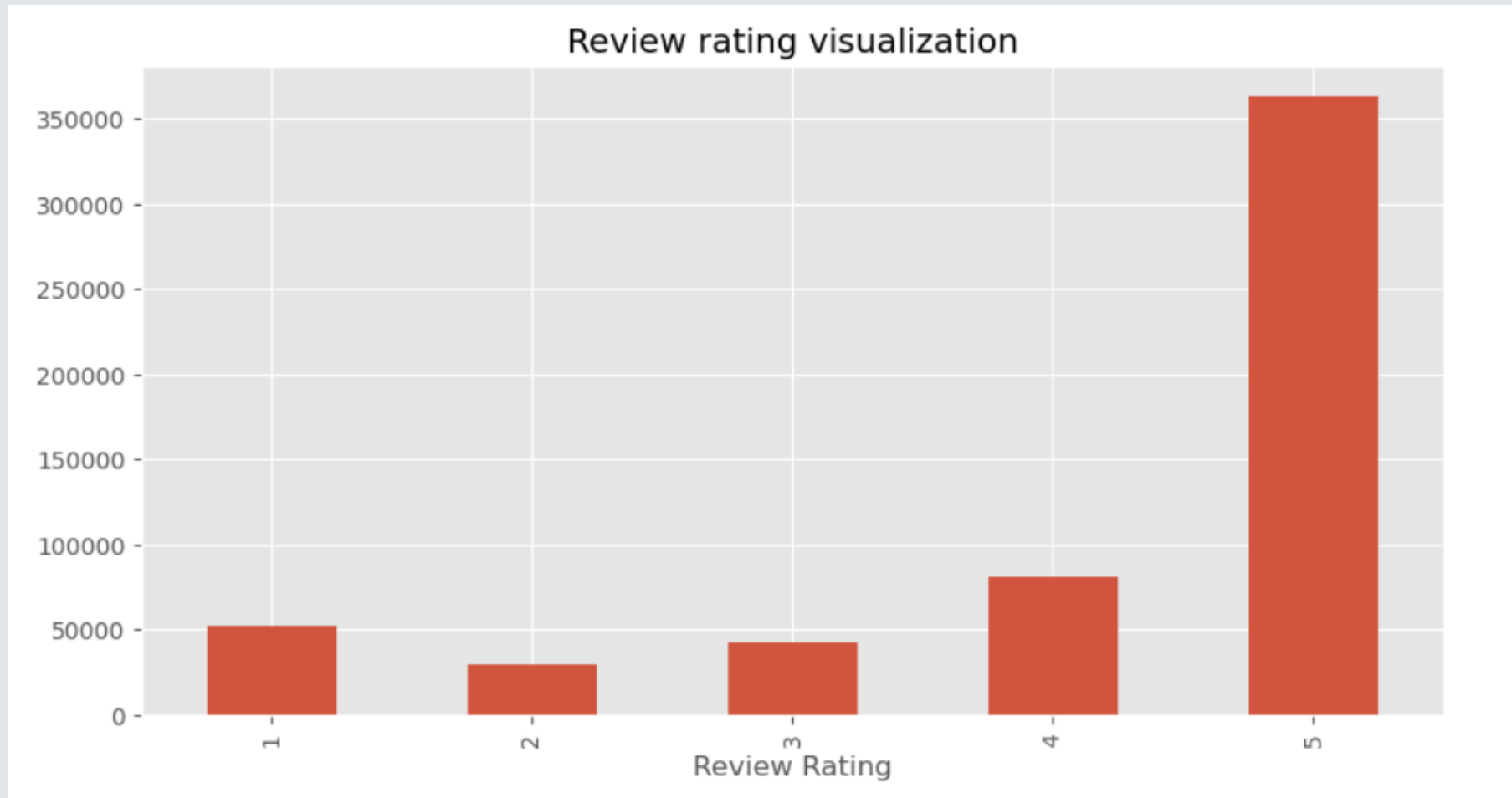


Figure 1: Exploratory Data Analysis

Results by VADER Sentiment Analysis

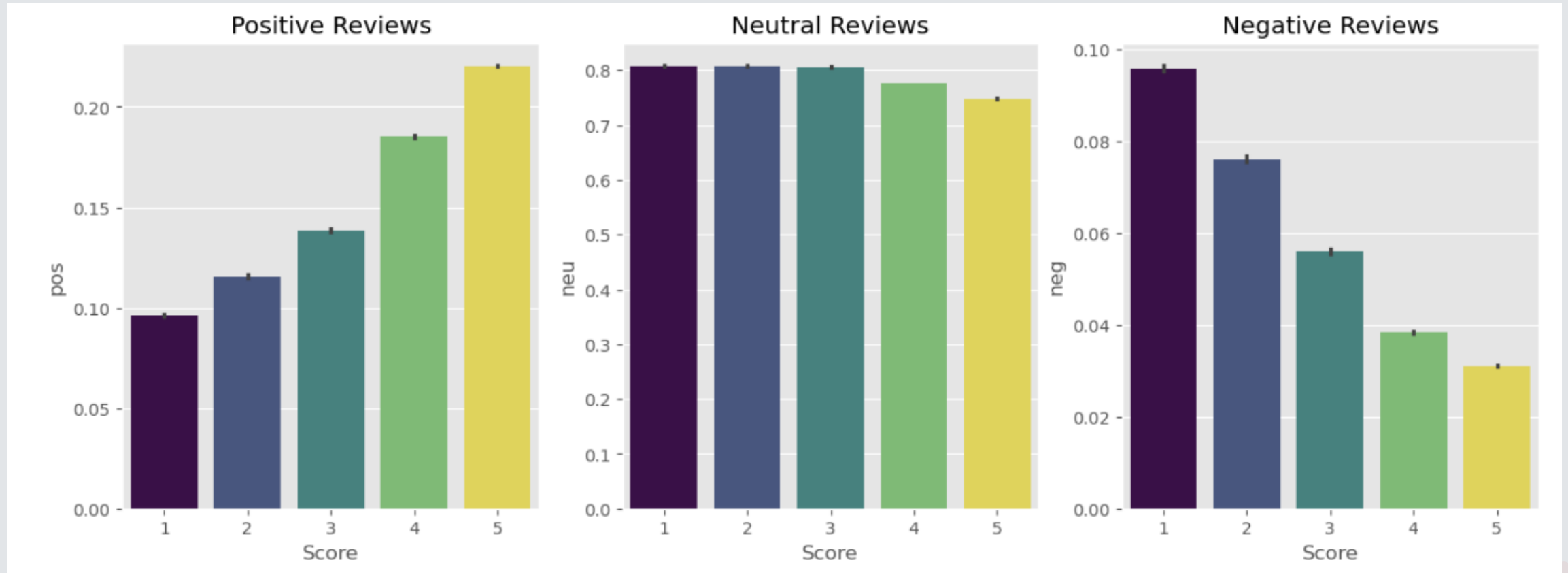


Figure 2: Customer Rating vs VADER Sentiment Score

Results by VADER Sentiment Analysis

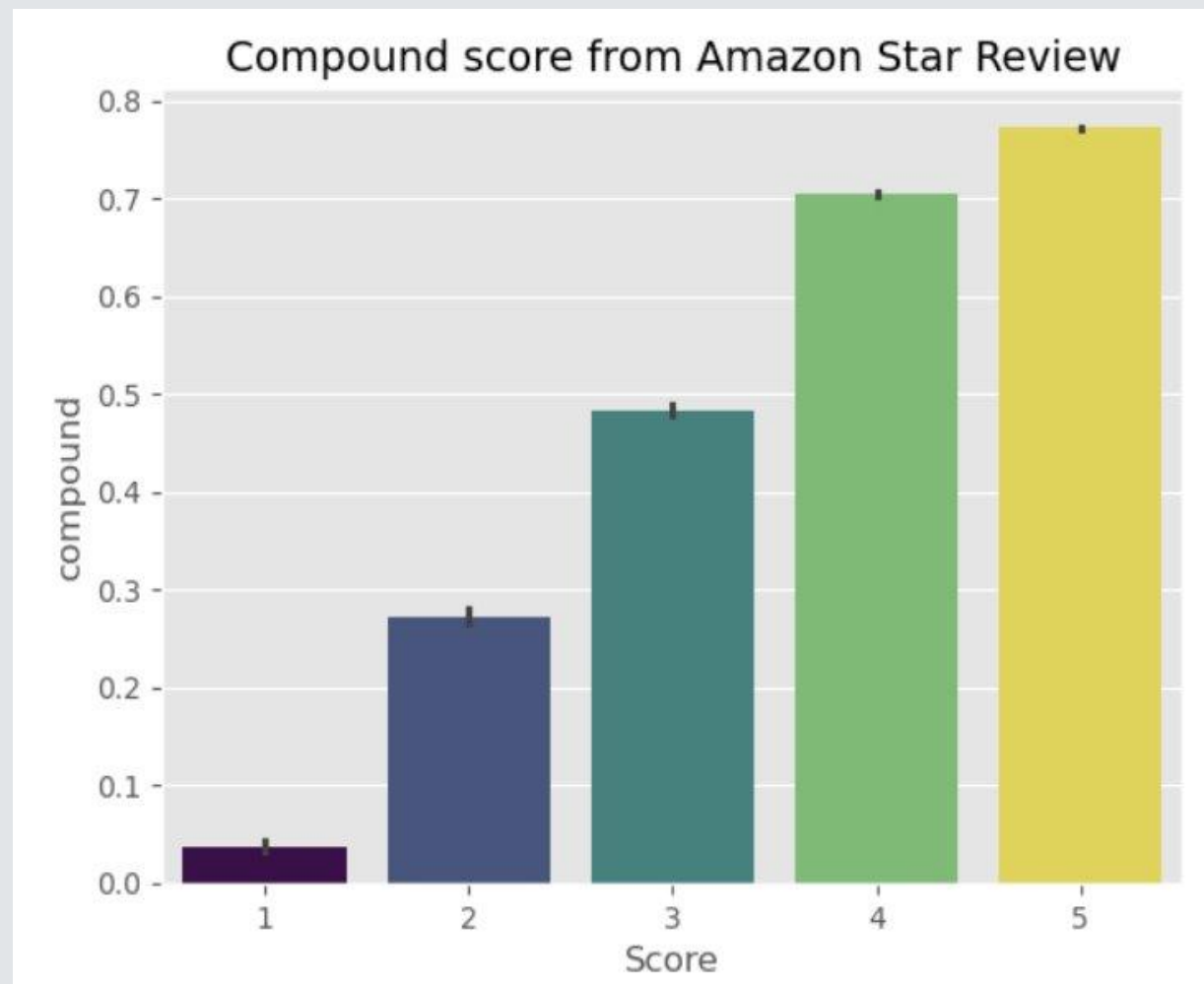


Figure 3: Customer Rating vs VADER Compound score

Results:

VADER Sentiment Analysis Accuracy:

```
Accuracy for VADER model: 0.7970
```

Logistic Regression Results:

```
Classification Report:
```

	precision	recall	f1-score	support
0.0	0.84	0.68	0.75	16407
1.0	0.94	0.98	0.96	88756
accuracy			0.93	105163
macro avg	0.89	0.83	0.86	105163
weighted avg	0.93	0.93	0.93	105163

```
Confusion Matrix:
```

```
[[11232  5175]
```

```
[ 2209 86547]]
```

```
Accuracy Score: 0.9297851906088643
```

Results:

Top 10 Performing Products (Based on Sentiment):

	ProductId	sentiment \
156	B0036VM05I	1.0
35633	B0036UQTQA	1.0
36332	B0036UWPJ0	1.0
105943	B0036VE3V2	1.0
228469	B0036V7K4E	1.0
229039	B0036VFX5M	1.0
309833	B0036UV8RU	1.0
487525	B001XWR35Y	1.0
539149	B0036VFTMO	1.0
544019	B0036UUY9I	1.0

Summary

156	Great Deal
35633	New label, same great stuff
36332	Works really well
105943	My Dog Thinks They Are The Best!
228469	You can't buy cheddar like this at the grocery...
229039	Not for everyone but I LOVE it!
309833	I love the coffee, but:
487525	Best Risotto in the World
539149	Dog + cats LOVE it!
544019	YUMMMMMY

Worst 10 Performing Products (Based on Sentiment):

	ProductId	sentiment \
93743	B00514BIQK	0.0
173990	B005116LQU	0.0
210927	B0025VF8TK	0.0
226725	B0025U00DC	0.0
312697	B004TMAAM8	0.0
369755	B0015DEH8C	0.0
369936	B0015DEH8W	0.0
396416	B0025VP9I0	0.0
456925	B005116KK2	0.0
525741	B005116LMO	0.0

Summary

93743	WARNING to SENSITIVE Celiacs!!
173990	Lowell Foods Dew Sugar
210927	Doesn't look like a salmon, doesn't taste like...
226725	Terrible After Taste
312697	Harm no Charm
369755	Do not order... stale candy!
369936	Sent an open package containing only 28 pieces
396416	item not as shown or name ordered
456925	Worst pork pate ever
525741	Disappointed with quality



Conclusion and Future work

Which model performed better and what could be the reason?

How can this work be used in future for better sentiment analysis?



The slide features a light blue background with decorative elements in the corners. The top-left corner contains several circles in shades of orange, yellow, and green. The top-right corner features a large green circle, a smaller orange circle, and a yellow circle. The bottom-right corner is decorated with a large teal circle, a medium orange circle, a small green circle, and a tiny pink circle. The text "Thank you" is centered in a dark green, sans-serif font.

Thank you