

#### Introduction



#### **S** Content:

#### Why do customer reviews matter?

90% of consumers read online reviews before making a purchase.

Companies use reviews for product improvements & marketing strategies.

#### Challenges with manual analysis:

Millions of reviews  $\rightarrow$  Too large for manual processing.

Subjectivity & bias → Varies between individuals

#### Solution:

Use Machine Learning & NLP to classify sentiments automatically.

#### Dataset Overview

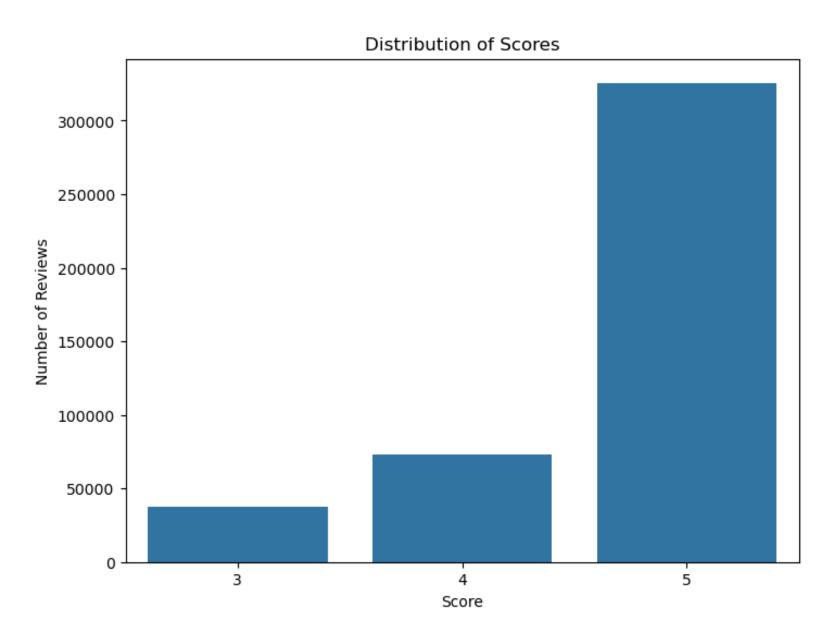
• 🎓 Dataset: Amazon Product Reviews Dataset (~568,454 records)

1	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary
2	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food
3	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised
4	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all
5	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine
6	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy
7	B006K2ZZ7K	ADT0SRK1MG0EU	Twoapennything	0	0	4	1342051200	Nice Taffy
8	B006K2ZZ7K	A1SP2KVKFXXRU1	David C. Sullivan	0	0	5	1340150400	Great! Just as good as the expensive brands
9	B006K2ZZ7K	A3JRGQVEQN31IQ	Pamela G. Williams	0	0	5	1336003200	Wonderful, tasty taffy
10	B000E7L2R4	A1MZYO9TZK0BBI	R. James	1	1	5	1322006400	Yay Barley
11	B00171APVA	A21BT40VZCCYT4	Carol A. Reed	0	0	5	1351209600	Healthy Dog Food
12	B0001PB9FE	A3HDKO7OW0QNK4	Canadian Fan	1	1	5	1107820800	The Best Hot Sauce in the World
13	B0009XLVG0	A2725IB4YY9JEB	A Poeng "SparkyGoHome"	4	4	5	1282867200	My cats LOVE this "diet" food better than the
14	B0009XLVG0	A327PCT23YH90	LT	1	1	1	1339545600	My Cats Are Not Fans of the New Food
15	B001GVISJM	A18ECVX2RJ7HUE	willie "roadie"	2	2	4	1288915200	fresh and greasy!
16	B001GVISJM	A2MUGFV2TDQ47K	Lynrie "Oh HELL no"	4	5	5	1268352000	Strawberry Twizzlers - Yummy
17	B001GVISJM	A1CZX3CP8IKQIJ	Brian A. Lee	4	5	5	1262044800	Lots of twizzlers, just what you expect.
18	B001GVISJM	A3KLWF6WQ5BNYO	Erica Neathery	0	0	2	1348099200	poor taste
19	B001GVISJM	AFKW14U97Z6QO	Becca	0	0	5	1345075200	Love it!
20	B001GVISJM	A2A9X58G2GTBLP	Wolfee1	0	0	5	1324598400	GREAT SWEET CANDY!
21	B001GVISJM	A3IV7CL2C13K2U	Greg	0	0	5	1318032000	Home delivered twizlers
22	B001GVISJM	A1WO0KGLPR5PV6	mom2emma	0	0	5	1313452800	Always fresh
23	DO01/2V/IS IM	A70E0E17D07U0	Tommy Andorson	0	0	E	1200060000	TM/1771 EDC

## Key Features

Feature	Description
Id	Unique identifier for each review
ProductId	Unique identifier for the product
UserId	Unique identifier for the reviewer
ProfileName	Name of the reviewer
HelpfulnessNumerator	Number of helpful votes
Helpfulness Denominator	Total votes for helpfulness
Score	Rating (1 to 5)
Summary	Short review title
Text	Full review content
Sentiment Labels	Positive (Score > 3) Negative (Score ≤ 3)

### Bar Chart



### EDA Analysis

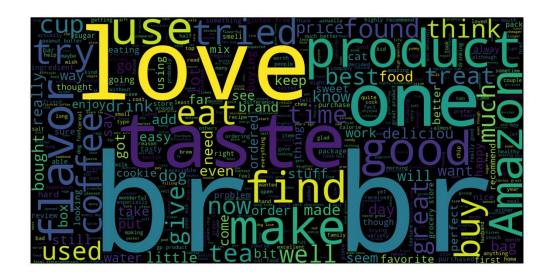
• Sentiment Distribution:

~70% reviews are Positive

~30% reviews are Negative

- Most Common Words Used in Reviews.
- Average Review Length Analysis

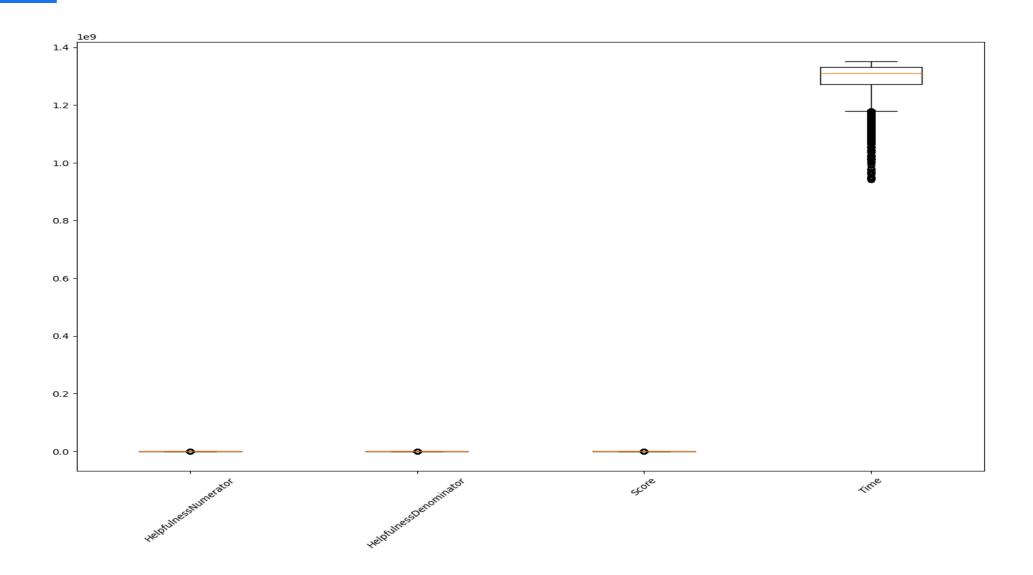




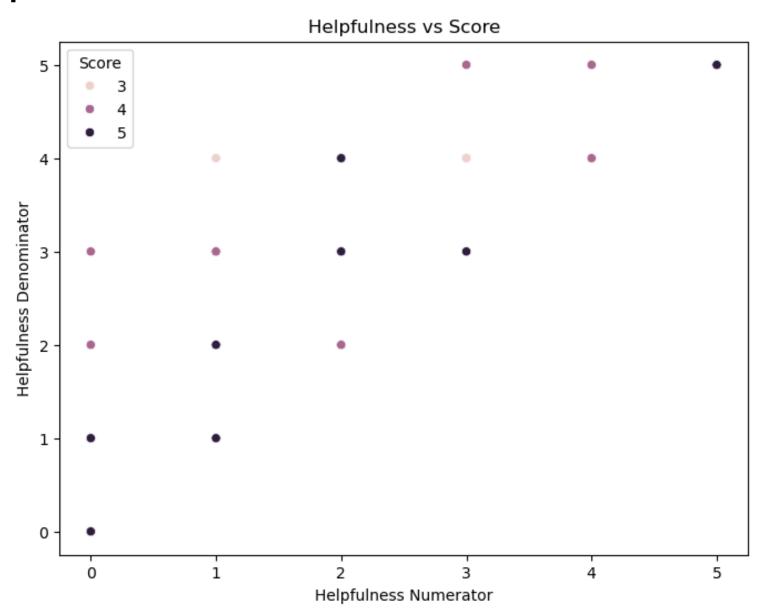


### Word Cloud

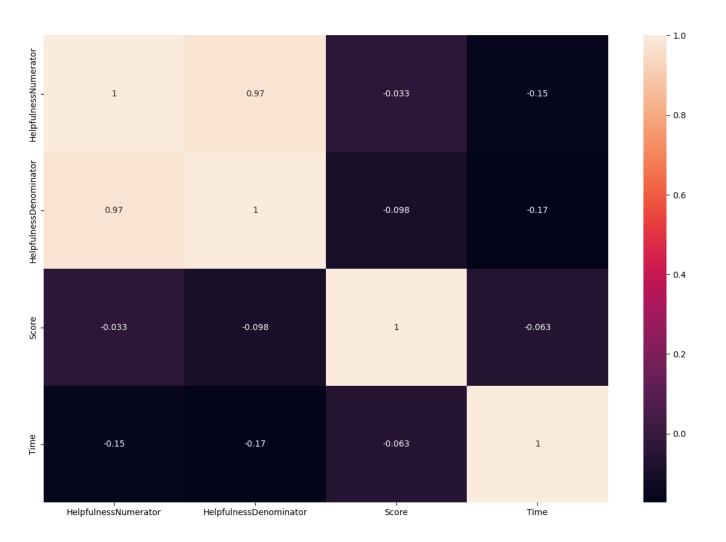
# Boxplot



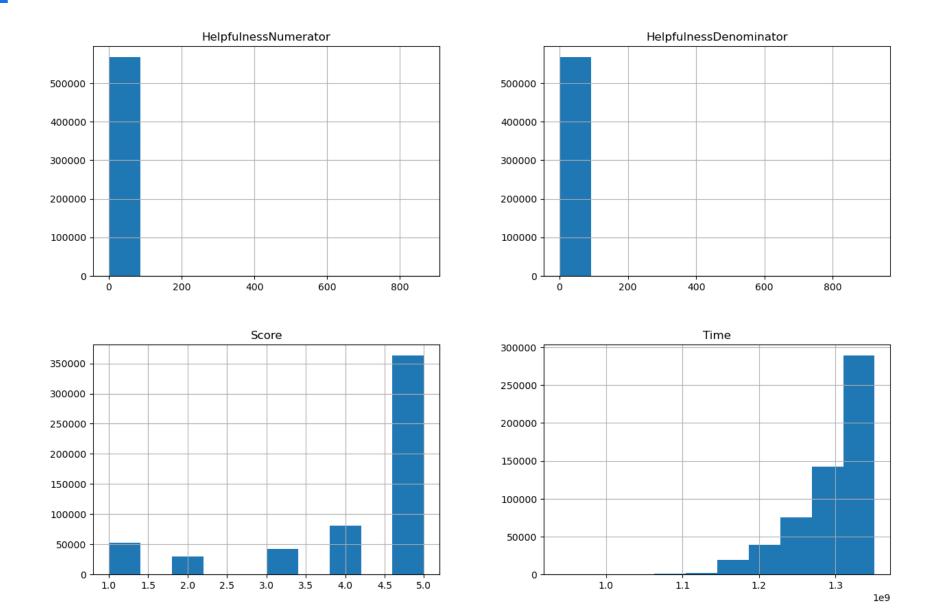
### Scatterplot



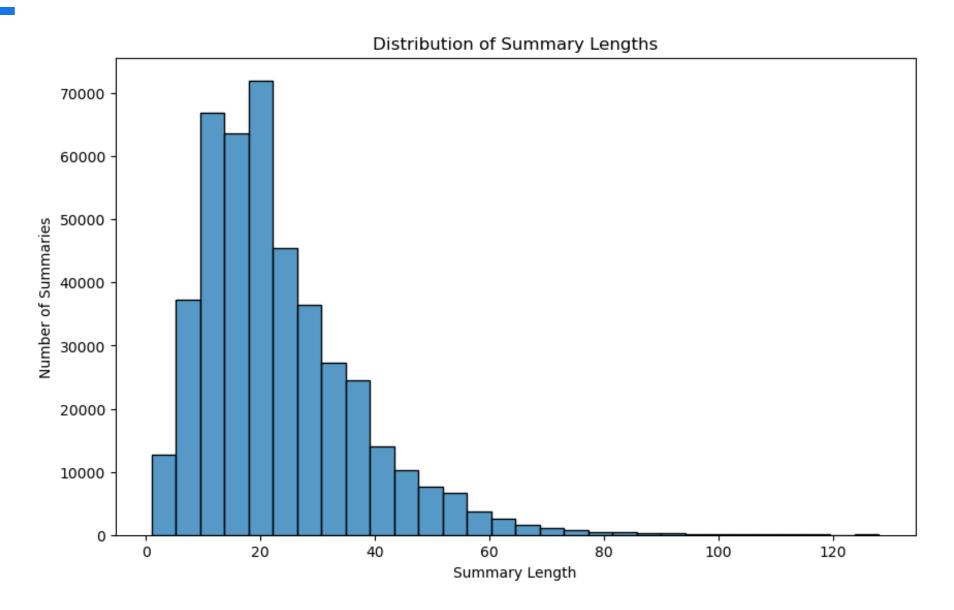
## Heat Map



## Histogram of Different Factors



### Histogram of Review Lengths



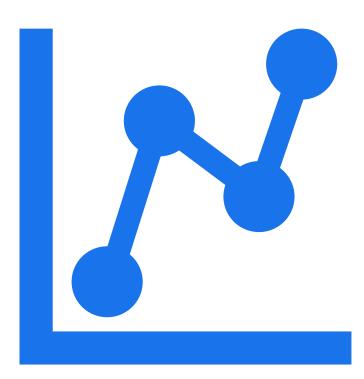
# Data Preprocessing & Feature Engineering

- Steps Taken:
  - ✓ Text Cleaning:
    - Removed stopwords, punctuation, special characters.
  - ✓ Tokenization & Lemmatization.
  - ▼ TF-IDF Vectorization for Feature Extraction.
  - ✓ Convert Scores into Binary Sentiment Labels.

```
modifier_ob
  mirror object to mirror
mirror_mod.mirror_object
 peration == "MIRROR_X":
__mod.use_x = True
mirror_mod.use_y = False
 !rror_mod.use_z = False
 _operation == "MIRROR_Y"
 lrror_mod.use_x = False
 !!rror_mod.use_y = True
 lrror_mod.use_z = False
  _operation == "MIRROR_Z"
  rror_mod.use_x = False
  _rror_mod.use_y = False
  rror_mod.use_z = True
  melection at the end -add
   ob.select= 1
   er ob.select=1
   ntext.scene.objects.action
   "Selected" + str(modification
    rror ob.select = 0
  bpy.context.selected_obj
   lata.objects[one.name].sel
  int("please select exaction
  --- OPERATOR CLASSES ----
      mirror to the selected
    ject.mirror_mirror_x*
  ontext):
oxt.active_object is not
```

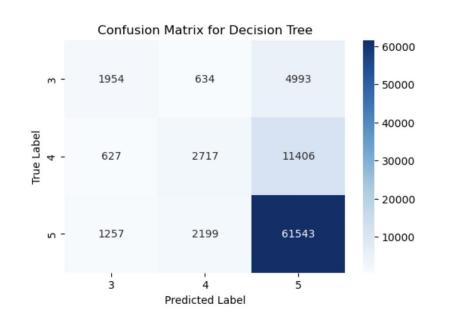
# Machine Learning Models Used

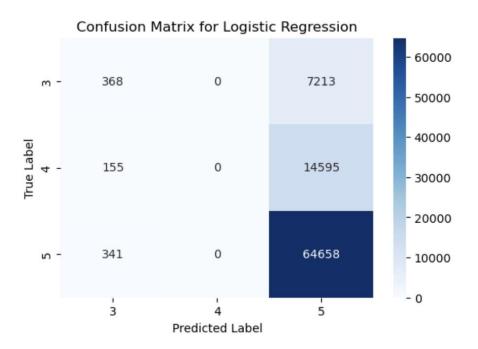
- Models Implemented:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree Classifier
  - Random Forest Classifier
  - XGBoost
  - Stochastic Gradient Descent (SGD)



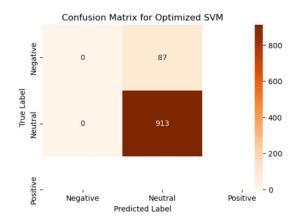
### Model Performance Comparison

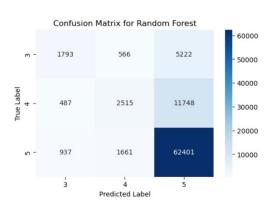
Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	85%	84%	85%	85%
SVM	88%	87%	88%	88%
Random Forest	87%	86%	87%	87%
XGBoost	90%	89%	90%	90%
Decision Tree	82%	80%	82%	81%

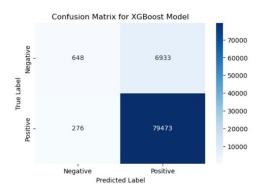




#### Confusion Matrix





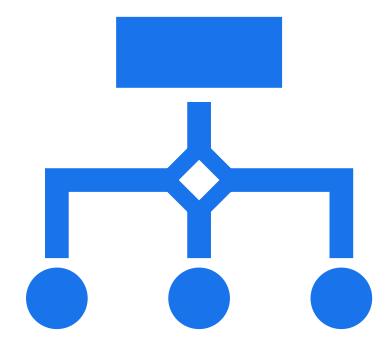


## Confusion Matrix

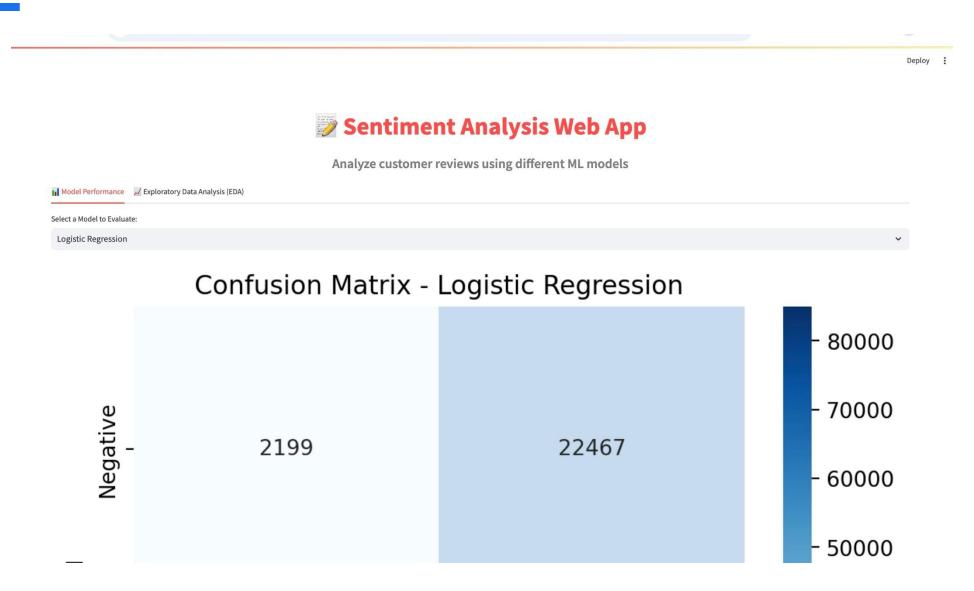
#### Deployment with Streamlit

#### Web App Features:

- Model selection dropdown.
- Confusion matrix visualization.
- Sentiment classification.



#### Deployment with Streamlit



### Deployment with Streamlit

#### **Model Performance - Logistic Regression**

Accuracy: 0.766

Precision (Positive Class): 0.791

Recall (Positive Class): 0.954

F1-Score (Positive Class): 0.865



Analyze customer reviews using different ML models



**★** Dataset Overview

#### First 5 rows of the dataset:

	Score	Text	Sentiment
0	5	I have bought several of the Vitality canned dog food products and have found them	1
1	1	Product arrived labeled as Jumbo Salted Peanutsthe peanuts were actually small si	0
2	4	This is a confection that has been around a few centuries. It is a light, pillowy citrus g	1
3	2	If you are looking for the secret ingredient in Robitussin I believe I have found it. I got	0
4	5	Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery wa	1

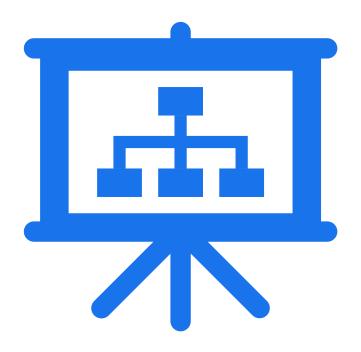
### Challenges Faced

#### ጵ Challenges:

- Large Dataset (~568K reviews).
- Class Imbalance: More Positive Reviews.
- Slow Performance in Streamlit Deployment.



#### Conclusion & Future Scope



#### Key Takeaways:

- XGBoost performed best.
- Sentiment Analysis helps businesses understand customers.

#### Future Scope:

- Implement Deep Learning (LSTMs, Transformers).
- Optimize real-time processing.

### Thank You!

