Sentiment Analysis of Avengers Movie Reviews

Subtitle: Using Python NLP in Google Colab

Engr. Mikko A. De Torres
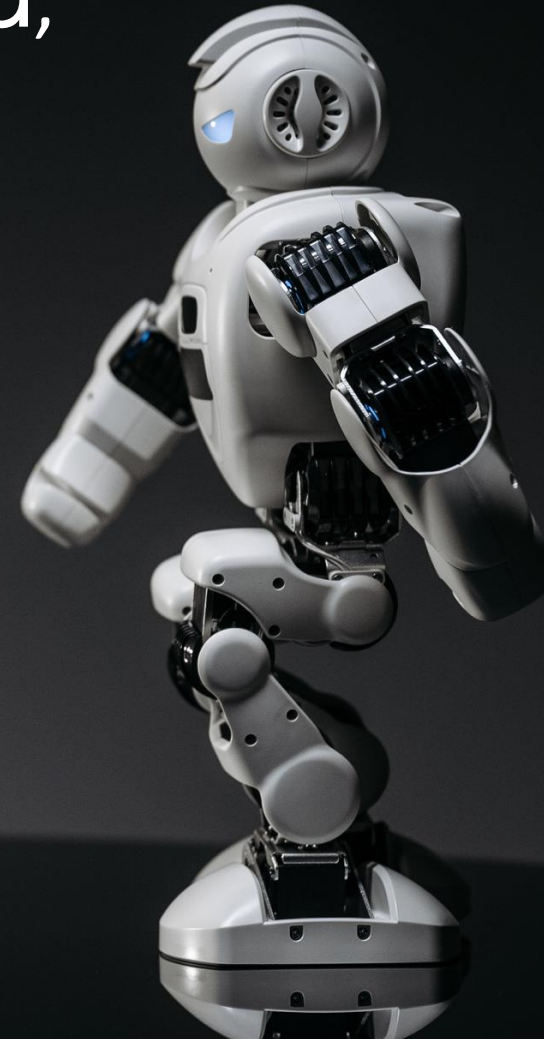**Mechatronics Engineer**
MSAI

Introduction

• Objective: Analyze and compare audience sentiments for Avengers 2012 and 2015

• Method: Natural Language Processing (NLP) with Python (VADER Sentiment Analysis)

Tools & Libraries Used

- Pandas, Matplotlib, WordCloud, NLTK, VADER
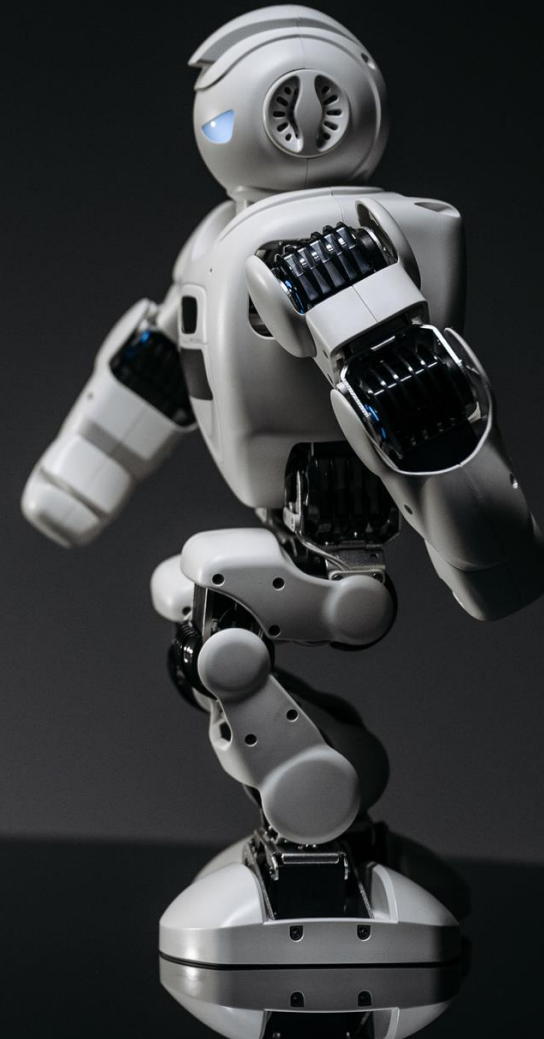
- Google Colab as the development platform

Data Source

• Dataset: **Rotten Tomatoes Extended MCU Movie Reviews**

• Source: Kaggle
https://www.kaggle.com/datasets/bwandow
ando/rotten-tomatoes-extended-mcu-
movie-reviews

• Contents: User reviews and ratings of MCU movies including *The Avengers (2012)* and *Avengers: Age of Ultron (2015)*

Data Preprocessing

- **Convert text to lowercase**
- **Remove stop words and punctuation**
- Tokenization
- Lemmatization

Data Preprocessing

	quote	cleaned_quote
0	An action spectacular with a witty script to m…	action spectacular witty script match big visu…
1	…doesn't have the emotional involvement of t…	doesnt emotional involvement best scifi action…
2	Lovers of the genre will be hard pressed to na…	lovers genre hard pressed favourite superhero …
3	A witty, well-acted sci-fi adventure.	witty wellacted scifi adventure
4	A continued run of quality can't be guaranteed…	continued run quality guaranteed involved aven…

Model & Training

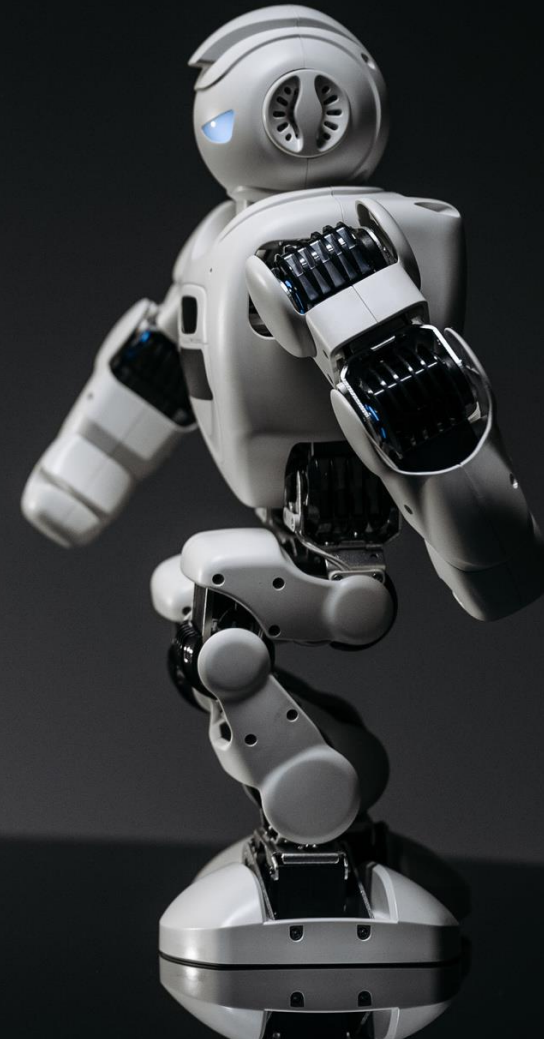**Model Used:**
- Multinomial Naive Bayes
- Train-Test Split: 80–20%
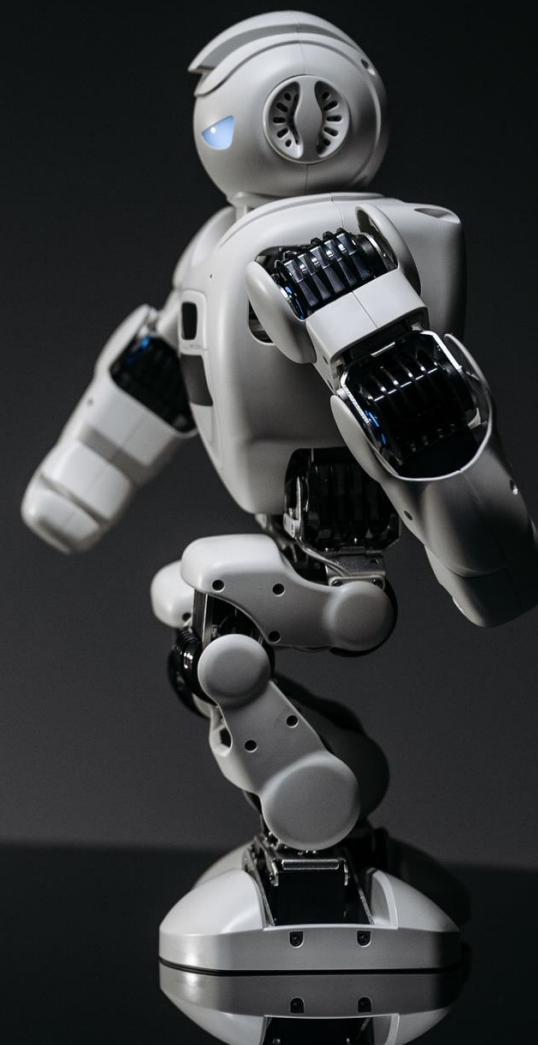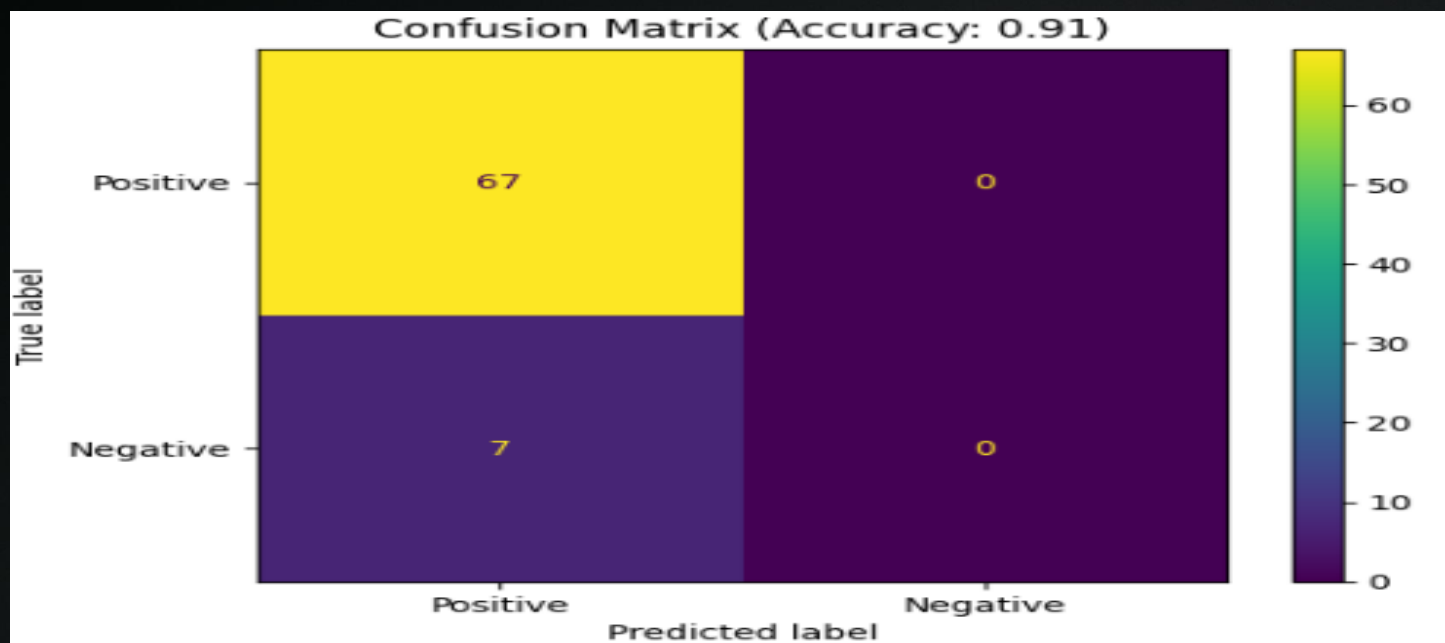
Reason:
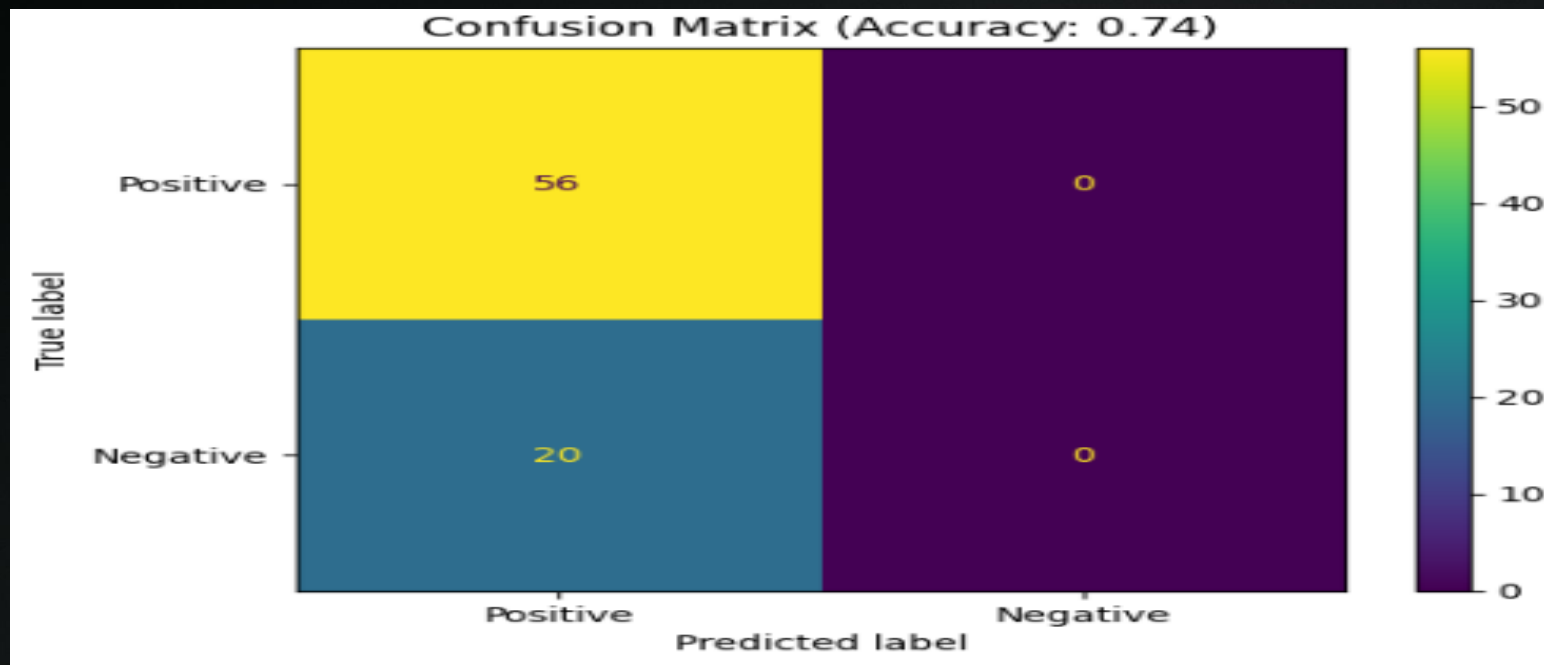Fast and efficient for text classification

Results & Evaluation

Avengers 2012 Review Accuracy: 91%

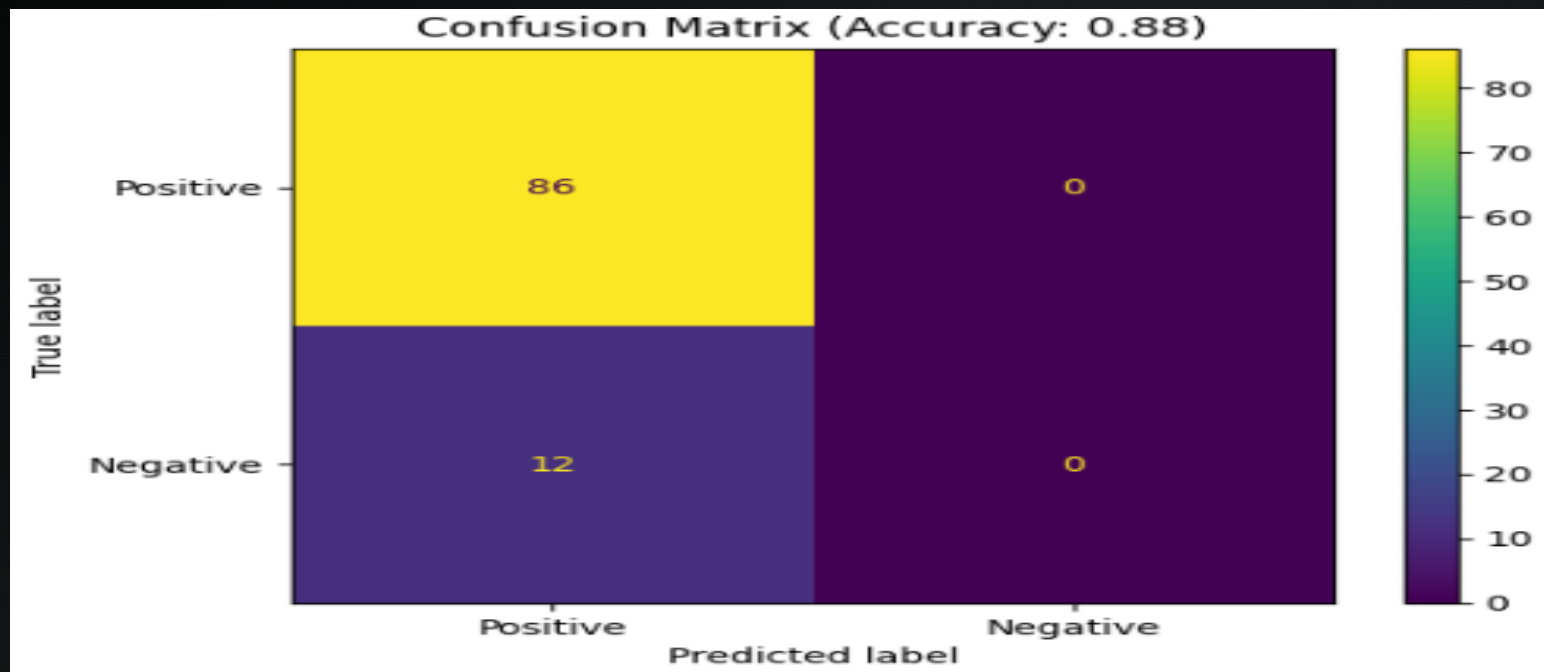Confusion Matrix (Accuracy: 0.91)

True label — Positive: 67, 0; Negative: 7, 0
Predicted label — Positive, Negative

Results & Evaluation

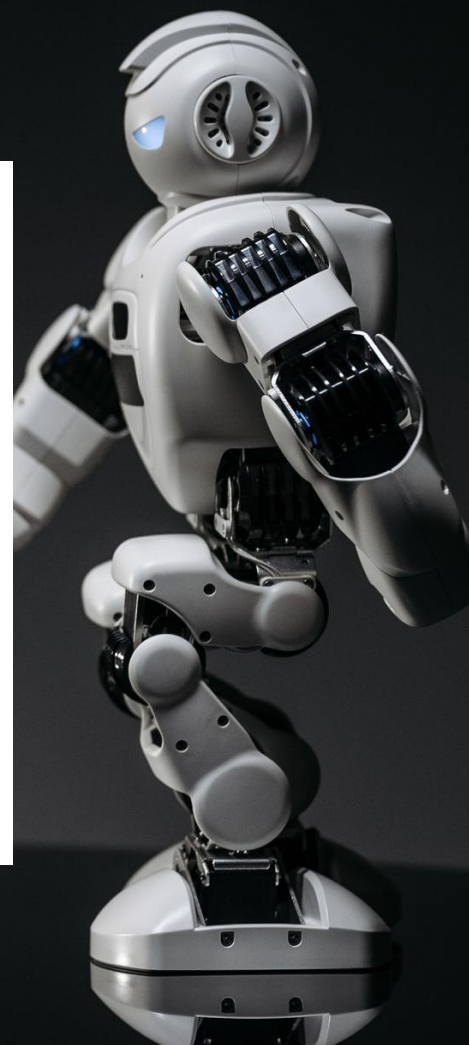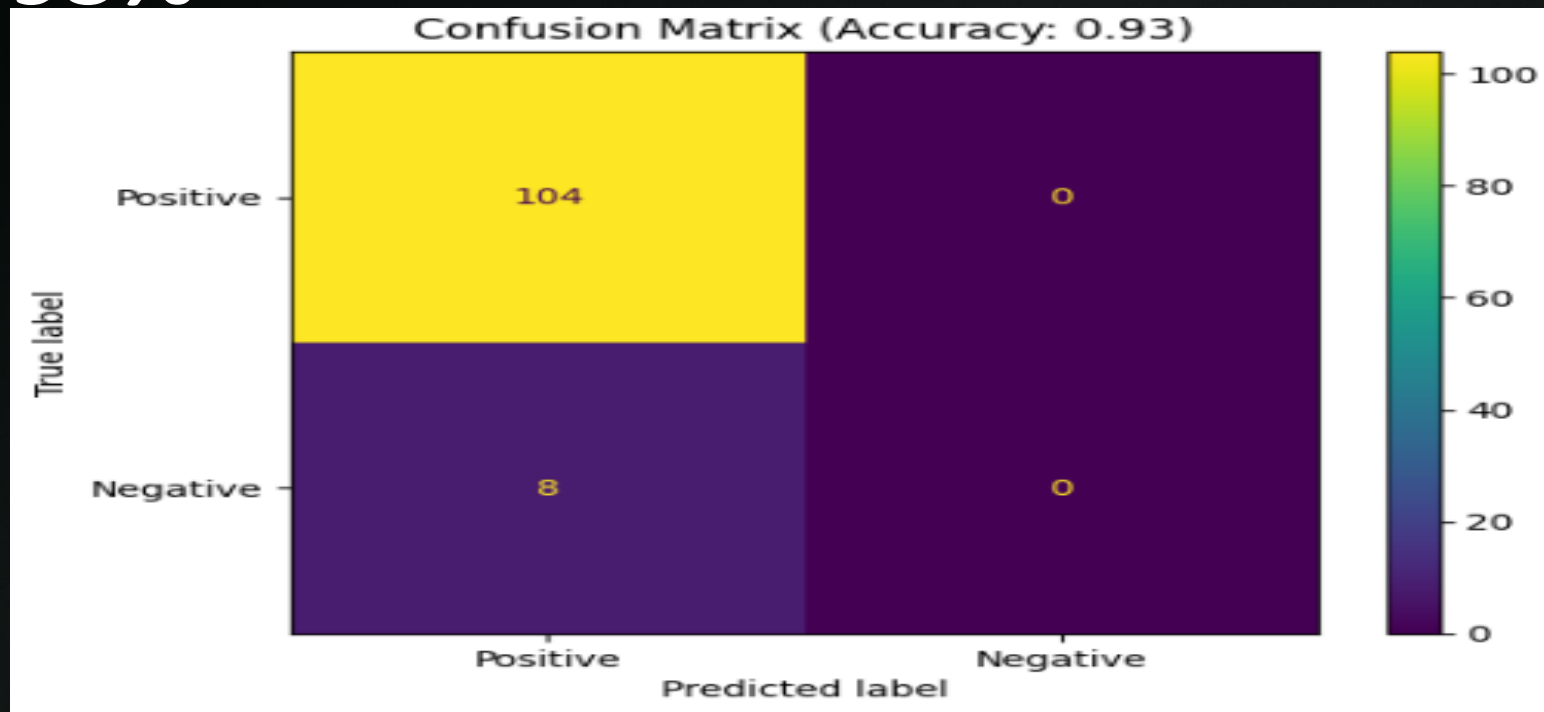## Avengers 2015 Review Accuracy: 74%

Confusion Matrix (Accuracy: 0.74)

Results & Evaluation

Avengers 2018 Review Accuracy: 88%

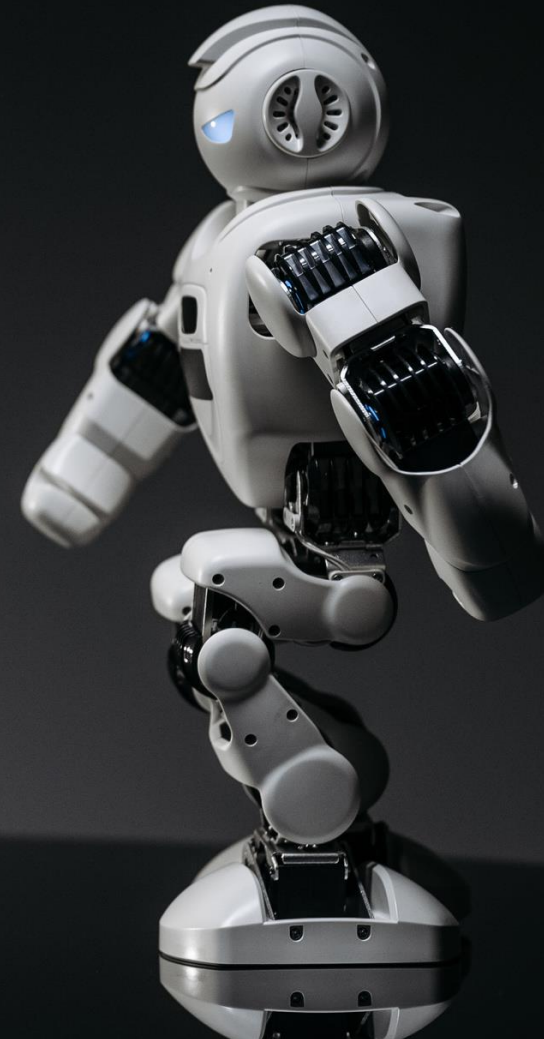# Results & Evaluation

Avengers 2019 Review Accuracy: 93%

Confusion Matrix (Accuracy: 0.93)

Improvements & Future Work

- Use **lemmatization**

- Try **SVM** or **Logistic Regression**

- Explore deep learning: **Word2Vec + LSTM**

- Include emojis/sarcasm handling

Conclusion

- NLP is effective in extracting audience sentiment

- Future: Compare with other Avengers movies, use more advanced models like BERT