



Twitter sentiment analysis

Sentiment analysis of tweets regarding AI

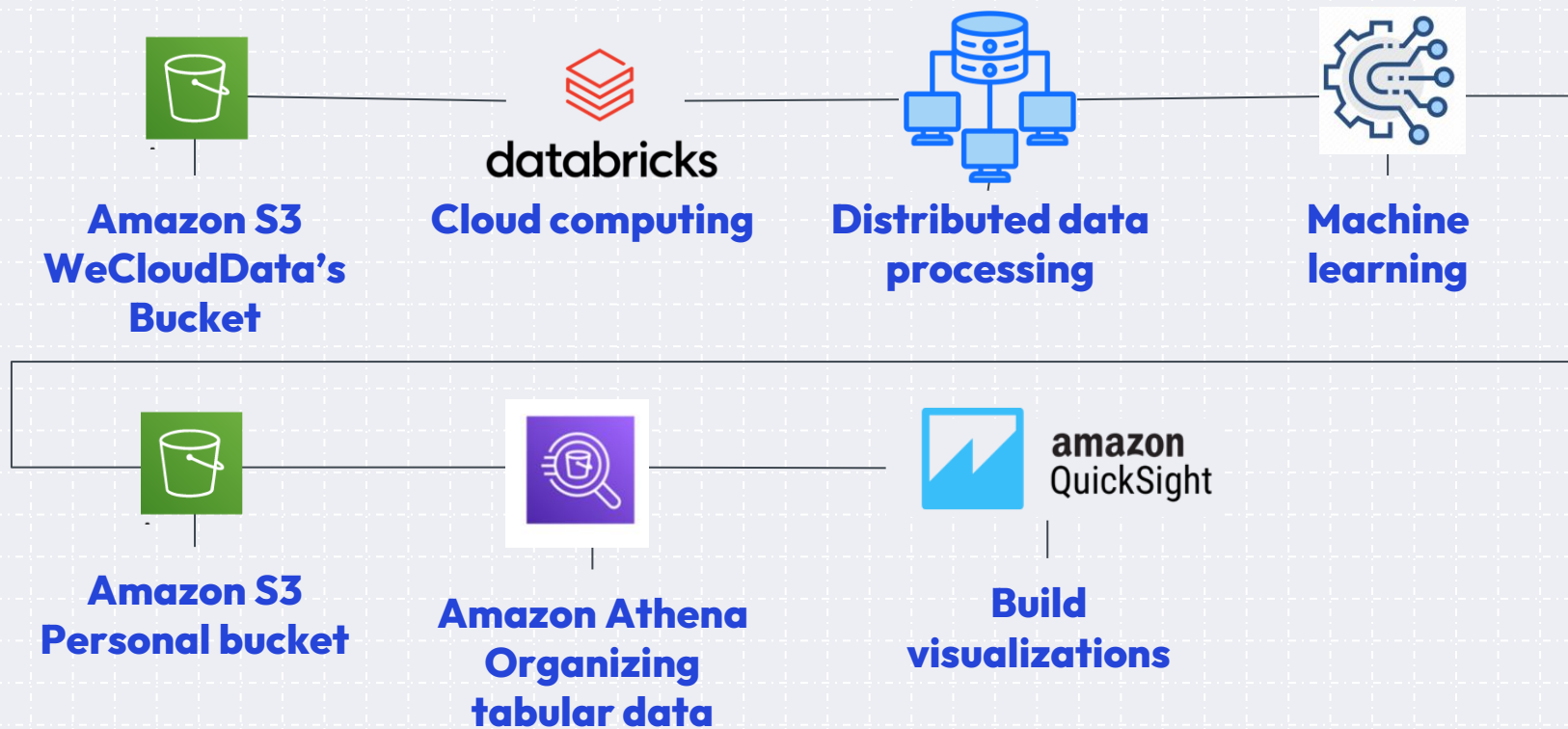


Introduction

- This project's objective was to build a sentiment analysis model using tweets retrieved from the web and present an analysis of the resulting dataset and model.
- The theme of the tweets used was 'AI', which refers to Artificial Intelligence.
- The data used was retrieved from one of weclouddata's public folders available through Amazon Simple Storage Service (AWS S3).
- The date of the tweets analyzed range from December 08 to 09, 2022.



Workflow



Analysis

Total tweets

10,491

- The dataset had a total of 10,491 tweets.

Unique tweets

4,840

46.13%

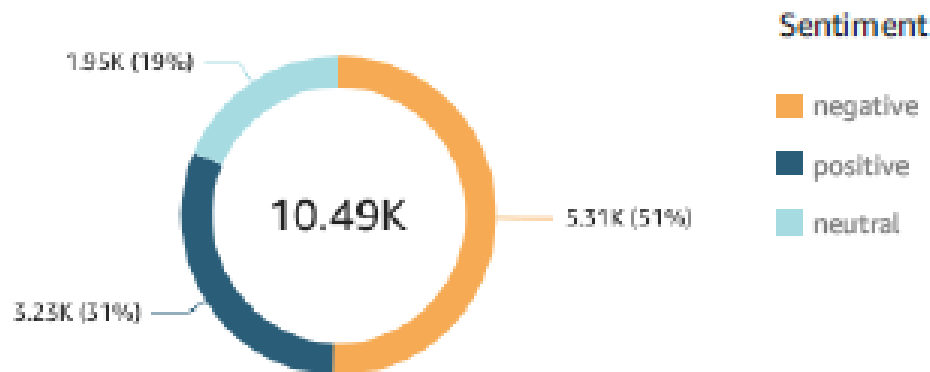
- 4,840 of these (46,13%) were unique.

Analysis

- **Positive:** 31% of the data (~ 3230 tweets).
- **Neutral:** 19% of the data (~ 1950 tweets).
- **Negative:** 51% of the data (~ 5310 tweets).

Sentiment count

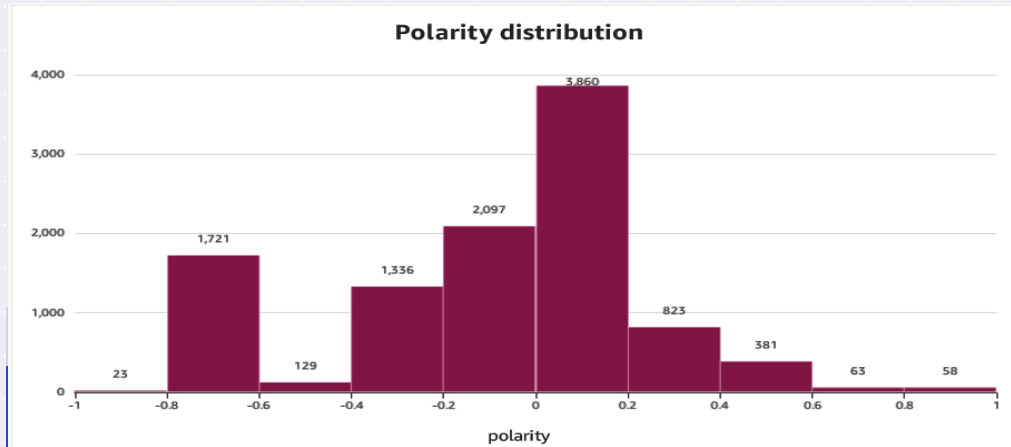
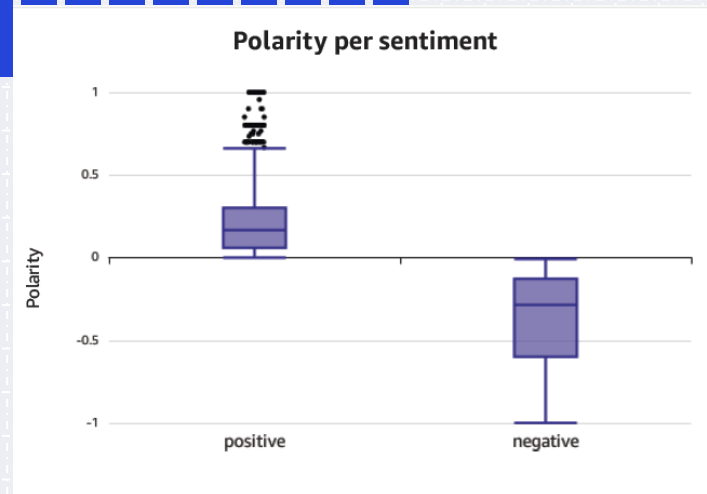
Original dataset



- 4,840 of these (46,13%) were unique.

Analysis

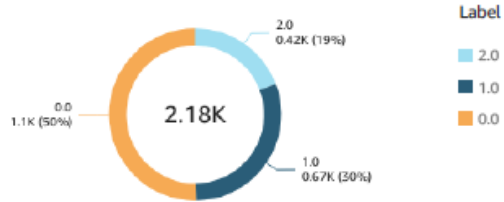
- Median absolute value for positive polarity was lower than the one for negative polarity.
- Extreme positive manifestation was way rarer than negative ones.



Analysis

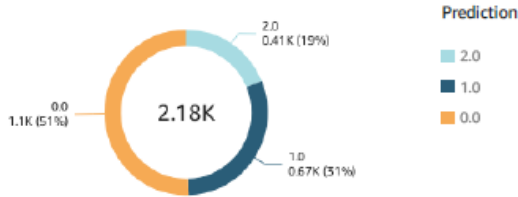
Label count

Test dataset



Predictions count

Test dataset



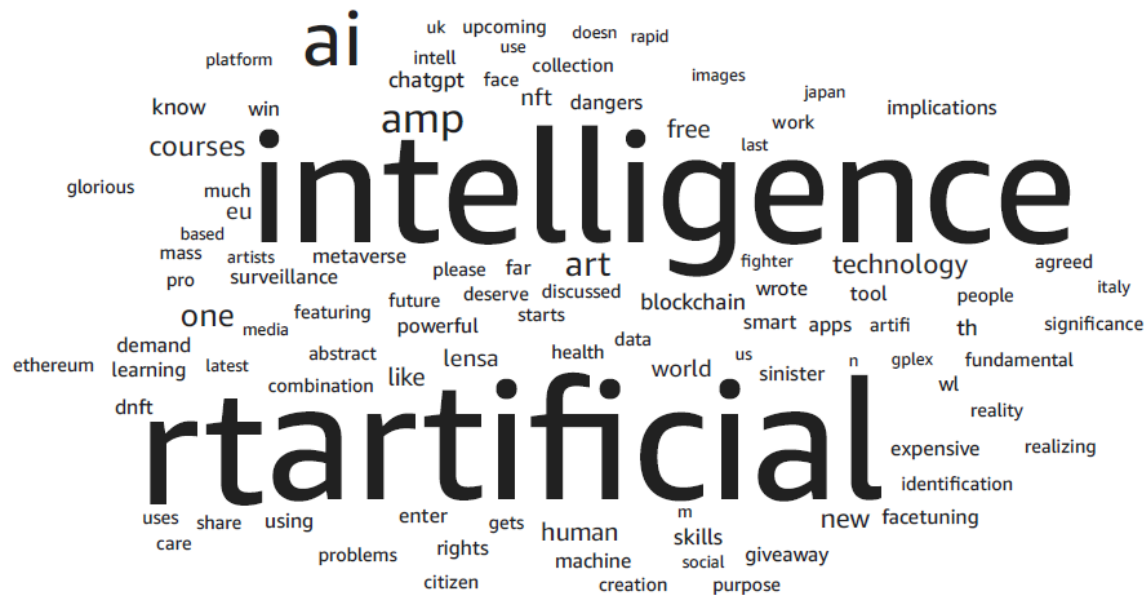
Three models were evaluated on a validation dataset:

- Logistic Regression (1 ngram tfidf)
- Decision Tree (1 ngram tf_idf)
- Random Forest (1 ngram tf_idf)

The Logistic Regression performed best on the validation set and was then scored on the test set.

LR Model scores on test data

Accuracy	Weighted precision	Weighted recall	F1
0.9047	0.9045	0.9047	0.9046

[illegible]

Challenges

- Adapting code to pyspark context.
- Managing AWS functionalities.

Conclusions

Best Model

Logistic Regression was the best model when predicting tweets' sentiments (accuracy= 90% f1 = 90%).

Sentiment frequency

Sentiments represented in the dataset were mostly negative (~51% negative, ~19% neutral, 31% positive).

Polarity

Negative manifestations of sentiments towards AI were more polarized than the expression of positive sentiments.



The slide features a light gray background with a white dashed grid. On the left and right sides, there are decorative elements consisting of horizontal bars of varying lengths, colored in blue and light gray, creating a modern, abstract border.

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**