# Wrangling And Analysing Social Media Data, To Understand How Animals Are Represented In The Australian News Media

Background → Objectives → Methodology → Findings → Challenges


*Prepared by*

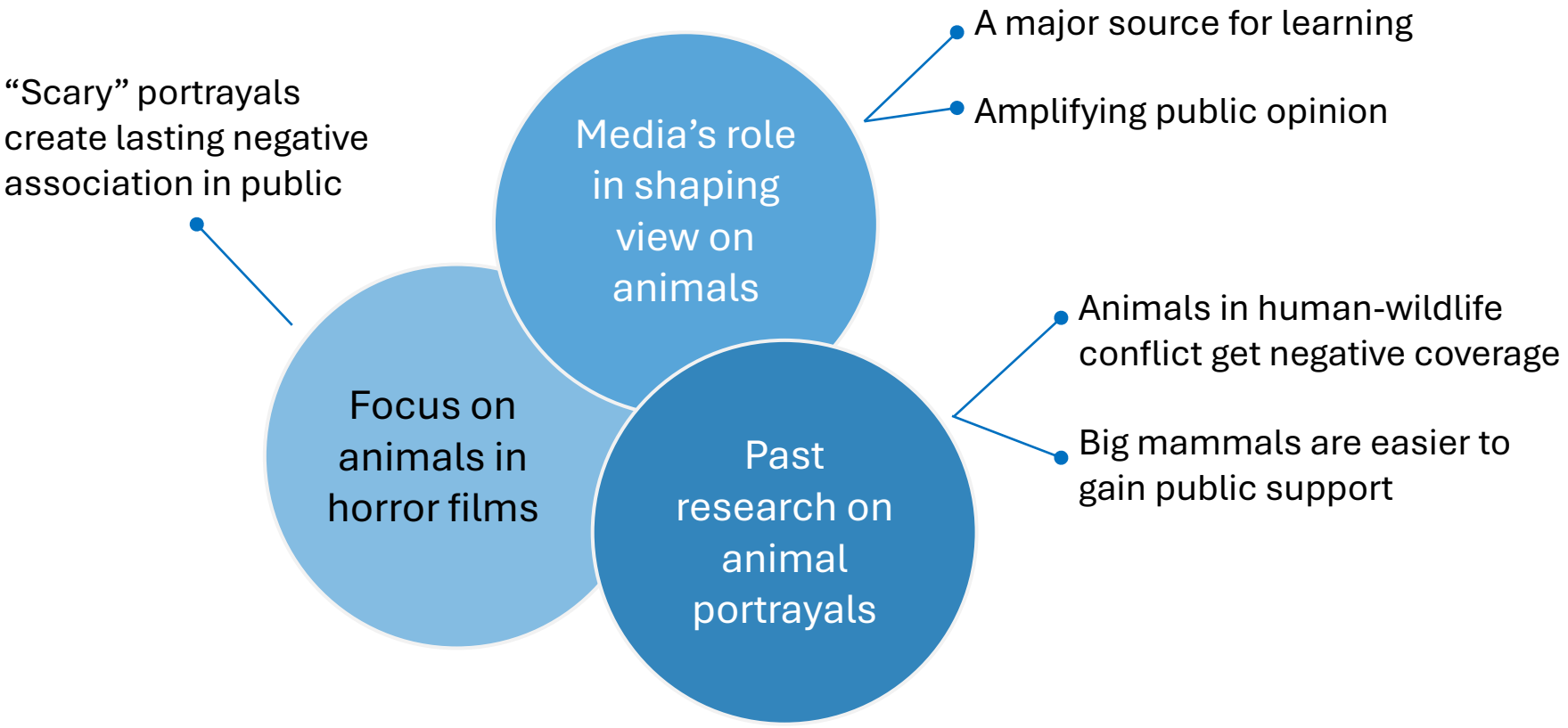
*Thi Xuan Huong Vu*

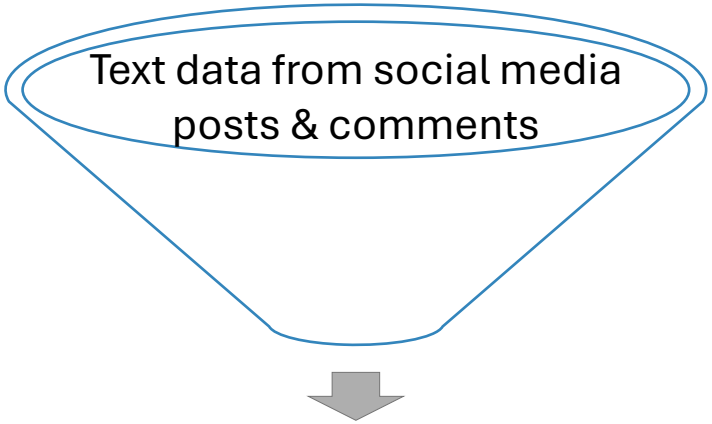
*Akash Suresh*

*Supported by*

*Dr. Brianna Le Busque*

*Dr. Masud Karim*

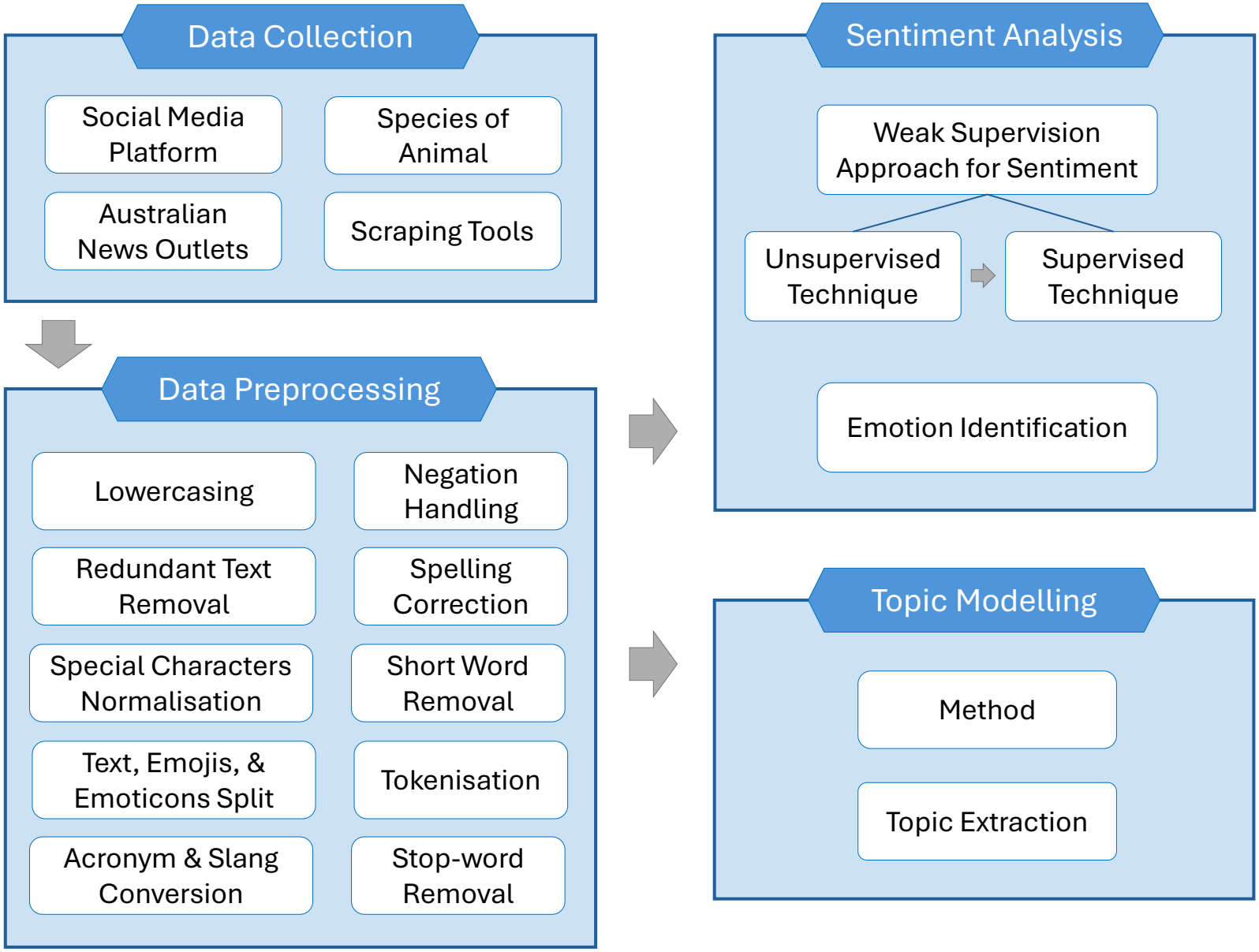

Media's role in shaping view on animals
A major source for learning
Amplifying public opinion
"Scary" portrayals create lasting negative association in public
Focus on animals in horror films
Past research on animal portrayals
Animals in human-wildlife conflict get negative coverage
Big mammals are easier to gain public support




Determine which animals are the most “newsworthy” and how they are portrayed by news media

Reveal public emotional reactions toward the animals

Understand how news media shape public opinion on the animals


Data Collection
Social Media Platform
Species of Animal
Australian News Outlets
Scraping Tools
Data Preprocessing
Lowercasing
Negation Handling
Redundant Text Removal
Spelling Correction
Special Characters Normalisation
Short Word Removal
Text, Emojis, & Emoticons Split
Tokenisation
Acronym & Slang Conversion
Stop-word Removal
Sentiment Analysis
Weak Supervision Approach for Sentiment
Unsupervised Technique
Supervised Technique
Emotion Identification
Topic Modelling
Method
Topic Extraction

## Data Collection

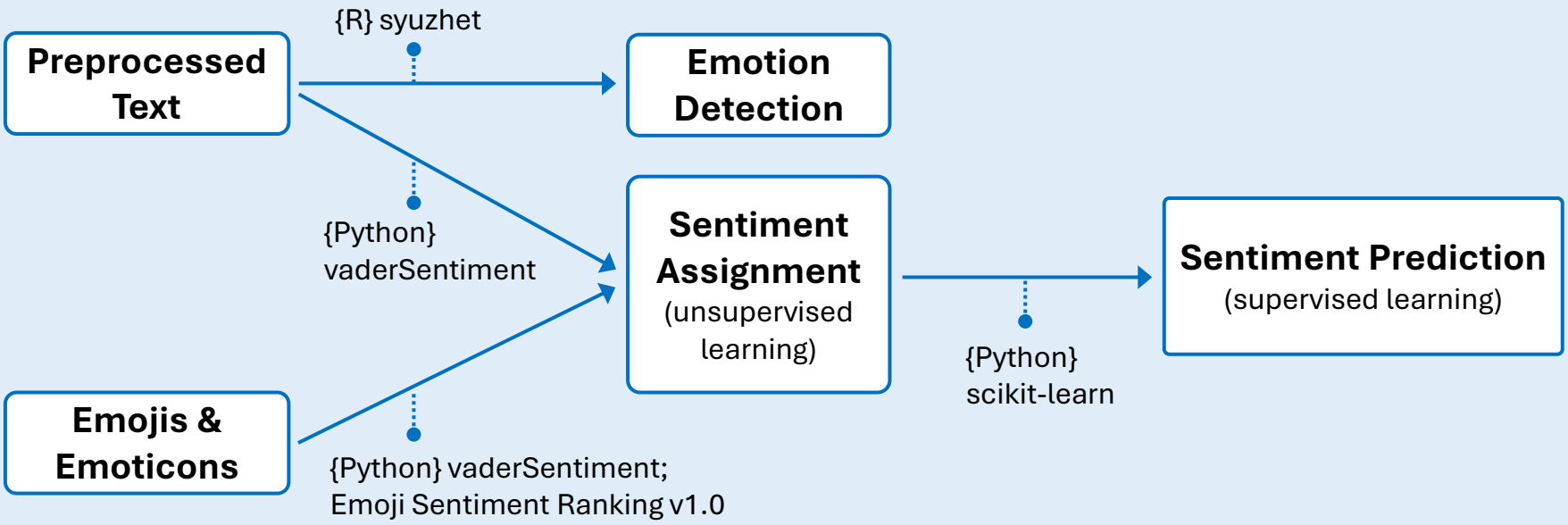
| | |
|---|---|
| Social media platform | Facebook & Instagram |
| Australian news outlet | ABC News, 7News, 9News, 10News First, SBS News |
| Animal species | 76 species: shark, spider, snake, whale, dog, bird, etc. |
| Scraping tool | • ESuit (web scraping service) for Facebook<br>• Web Scraper (web browser extension) for Instagram |

**2,551 posts & 103,538 comments**

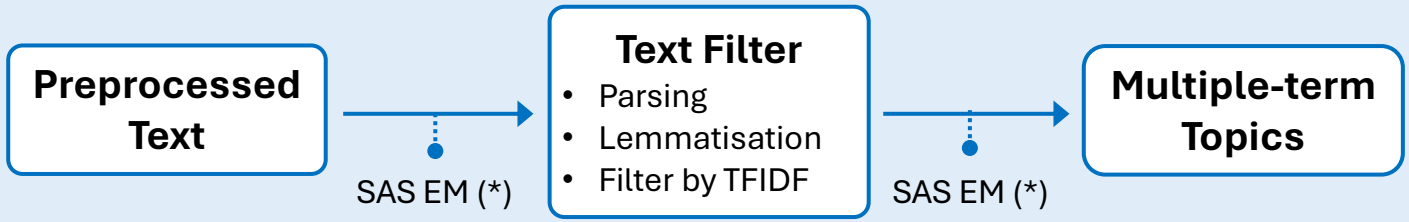
# Data Preprocessing

1. **Lowercasing**
   Regardless of proper names, sentence starters...
2. **Removal of redundant text**
   URLs, mentions, hashtags, extra white space
3. **Normalisation of apostrophes**
   Convert curly apostrophes and backticks to straight ones
4. **Split of text, emojis, emoticons**
   Use Python package 'emot'
5. **Conversion of acronyms & slang**
   Convert microtext to traditional form
   E.g., "idk", "lol" → "I don't know", "laugh out loud"
6. **Negation handling**
   Create a list of negation words & contractions, then normalise them
7. **Spelling correction**
   Use Python packages 'symspellpy', 'textblob', 'multiprocessing'
8. **Removal of short words**
   Words of 1 or 2 letters are removed
9. **Tokenisation**
   Use Python library 'NLTK'
10. **Removal of stop-words**
    Use stop-word list of 'NLTK' package but exclude negation words

## Sentiment Analysis




## Topic Modelling

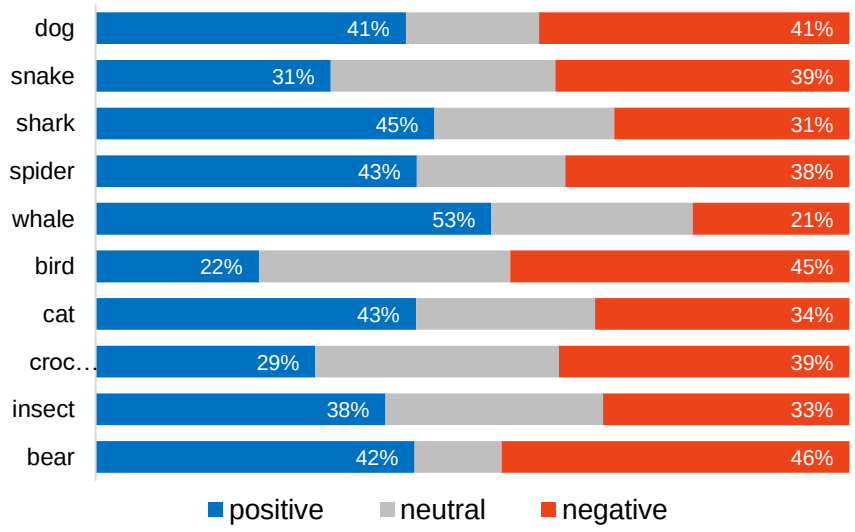



*(*): SAS Enterprise Miner*

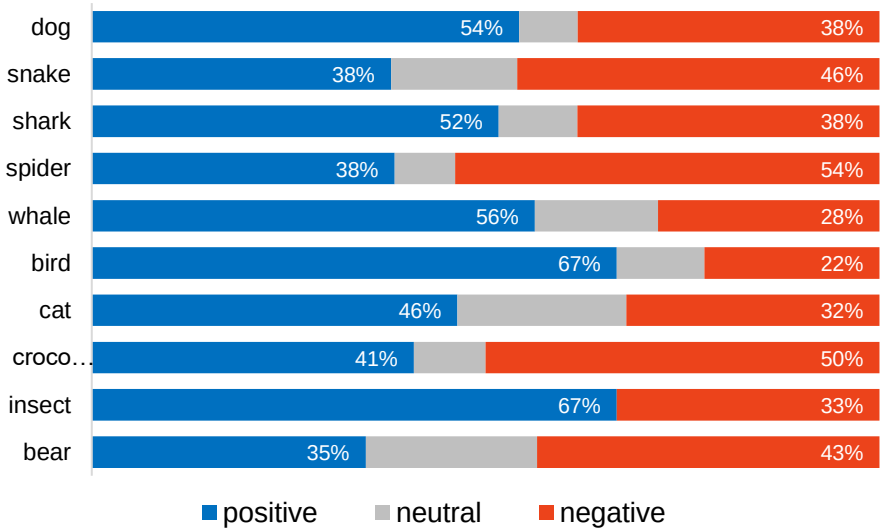
# Sentiment in Posts

| | Facebook | Instagram |
|---|---|---|
| Animals of most **positive** portrayal | Whale, shark, spider | Bird, insect, whale |
| Animals of most **negative** portrayal | Bear, bird, dog | Spider, crocodile, snake |
| **Overall** portrayal | Mixed sentiment across select animals | Emotion-charged portrayals outweigh posts of neutral sentiment |

Sentiment distribution of **FACEBOOK** posts



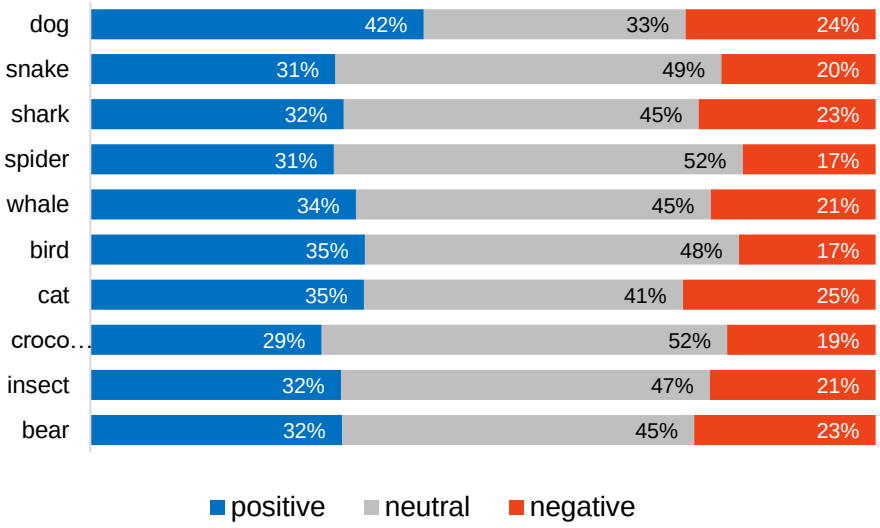

Sentiment distribution of **INSTAGRAM** posts



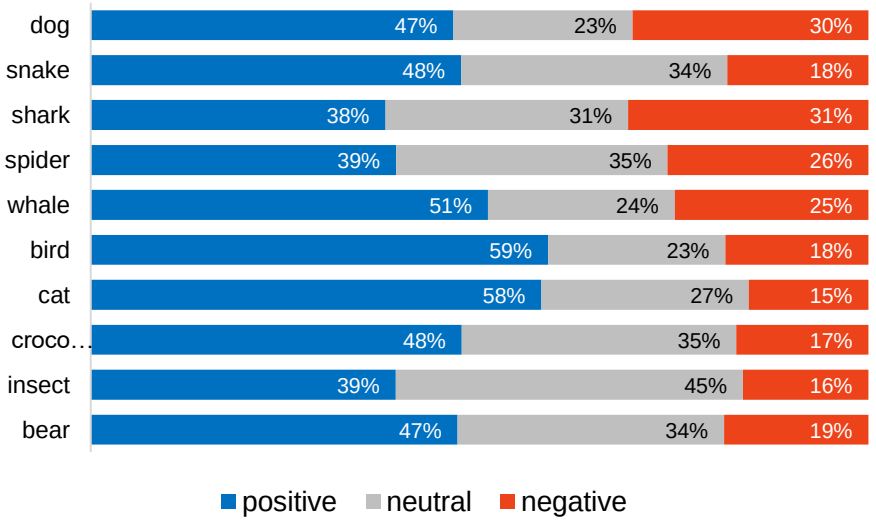
# Sentiment in Comments

| | Facebook | Instagram |
|---|---|---|
| Animals of most **positive** interaction | Dog, bird, cat | Bird, cat, whale |
| Animals of most **negative** interaction | Cat, dog | Shark, dog |
| **Overall** interaction | Positive & neutral comments combined are much larger than negative comments on both platforms | |

Sentiment distribution of **FACEBOOK** comments



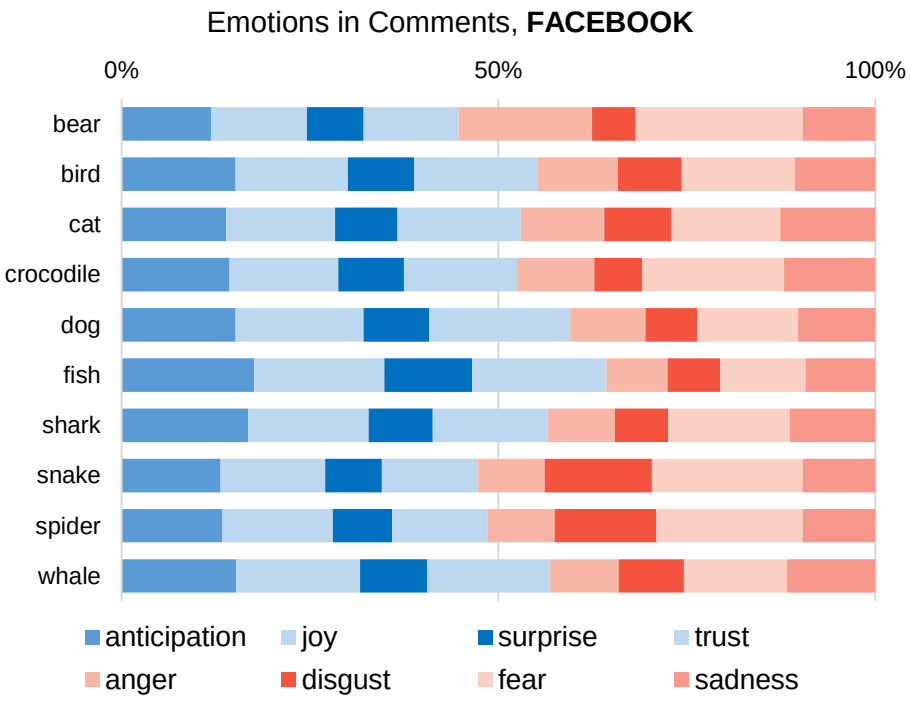

Sentiment distribution of **INSTAGRAM** comments

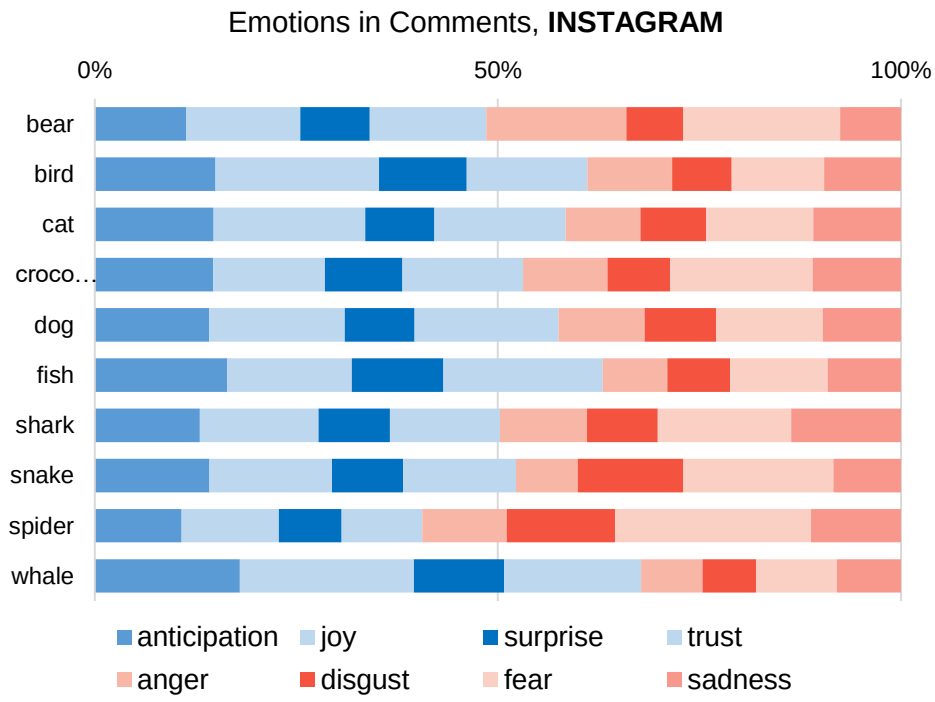


# Emotions in Comments

Generally, in both platforms:

- Positive emotions (anticipation, joy, surprise, trust) appear more frequent.
- Extreme emotion (disgust, in bright orange) is least to be expressed in comments.



# Sentimental impact of posts on comments

| | Facebook | Instagram |
|---|---|---|
| For **positive** posts | People are more inclined to give positive or neutral comments | People are more inclined to give positive comments |
| For **negative** posts | Neutral interaction outweighs emotion-charged comments | Sentiment is relatively balanced among positive, neutral, negative comments |
| **Overall** association | • Public opinion is more likely to agree with positive portrayals by news media<br>• People tend to avoid negative reactions toward animal related posts | |

Sentiment of posts vs. comments, **FACEBOOK**

| Comment Sentiment | Post Sentiment | | | Sum % |
|---|---|---|---|---|
| | positive | neutral | negative | |
| positive | 15% | 8% | 10% | 34% |
| neutral | 19% | 13% | 15% | 47% |
| negative | 6% | 5% | 9% | 20% |
| Sum % | 40% | 26% | 34% | 100% |

Sentiment of posts vs. comments, **INSTAGRAM**

| Comment Sentiment | Post Sentiment | | | Sum % |
|---|---|---|---|---|
| | positive | neutral | negative | |
| positive | 25% | 4% | 18% | 47% |
| neutral | 12% | 2% | 15% | 29% |
| negative | 6% | 2% | 16% | 24% |
| Sum % | 42% | 9% | 49% | 100% |

# Other Findings

## Topics & concerns discussed by the public

- People show common concerns over animal-related dangers & environmental impacts

  e.g., sharks, crocodiles, & snakes are viewed as threats, birds as frustration
- Human-animal interactions, especially unfortunate encounters, form a prominent theme across species

## Performance of sentiment labelling method

- Ground-truth sentiment labels are unavailable → impractical to evaluate accuracy
- Performance of supervised models provides a proxy for **generalisability of labels** obtained from unsupervised method
- Accuracy metric of 3 classifiers (Decision Tree, SVM, Random Forest) > 0.8

→ High generalisability

→ Our method can be applied to unseen data consistently

| Data collection | Data preprocessing | Sentiment labelling | Excluded data | Portrayal differences between platforms | Animal species of interest |
|---|---|---|---|---|---|
| Restrictions on data scraping lead to insufficient data collection | Not all preprocessing jobs work perfectly, especially correction of spelling errors | Ground-truth labels are unavailable for evaluation | Exclusion of images & photos, particularly in data of media posts, limits holistic analysis & outcomes | Different technical challenges & scraping tools hinders firm conclusion on portrayal differences between FB & IG | "Newsworthy" animals are derived from limited sources of data thus may not reflect real world |

Thank you!

# Q & A