Clickbait Detection

SOTA Report & POC Notebook Highlights

2025, NLP Course Project

Jakub Sawicki, Jędrzej Sokołowski, Wiktor Woźniak

What is Clickbait?

- Attention-grabbing headlines, often omitting key facts
- Sparks curiosity and encourages clicks using sensational language
- Difficult to detect automatically: can be subtle and context-dependent
- Exists on a spectrum, not just binary classification

Motivation & Challenges

- Manipulative clickbait undermines trust in news and social media
- Hard to draw the line between catchy and misleading
- Even humans disagree about what counts as clickbait
- Growing sophistication with AI-generated clickbait

Related works, SOTA Approaches

- Early: Hand-crafted features (punctuation, length, headline-article overlap)
- Classical ML: SVM, Random Forests, high accuracy with engineered features
- Deep Learning: CNNs, RNNs using word embeddings (Word2Vec, GloVe)
- Transformers: BERT, RoBERTa, current state-of-the-art for both detection and “spoiling”

Key Open Datasets

- Webis Clickbait Corpus 2017 (Twitter, 38,517 posts, 9,276 clickbaits)
- Wikinews Clickbait Corpus (crowdsourced English news headlines)
- Thai & Chinese Headline Datasets (expands multilingual, cross-domain evaluation)

Project Goals & Methods

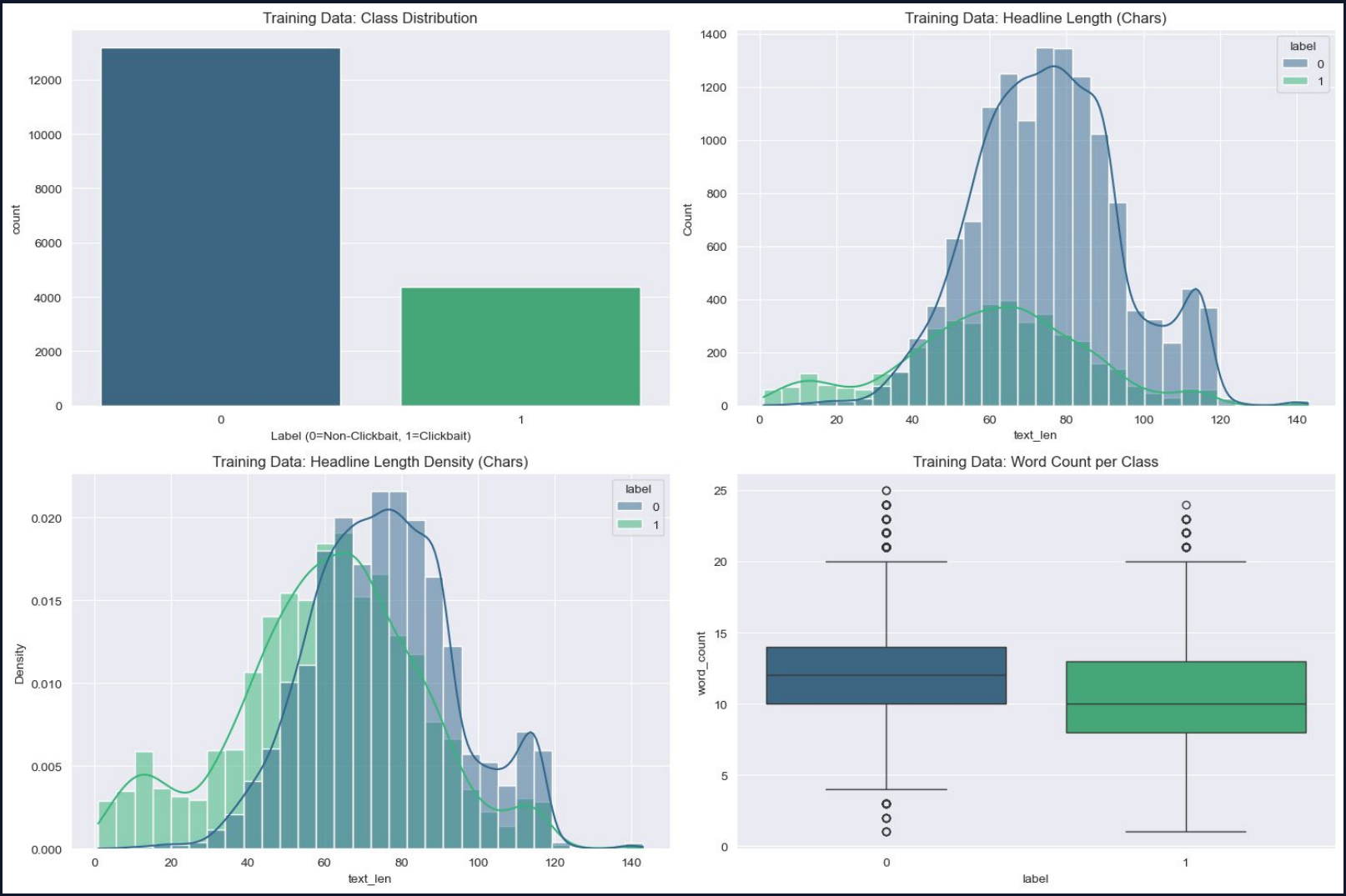
- Build interpretable, effective clickbait detection system using open benchmarks
- Compare classical ML, neural, and transformer models for detection
- Extract & analyze which textual patterns are most predictive
- Evaluate accuracy, precision, recall

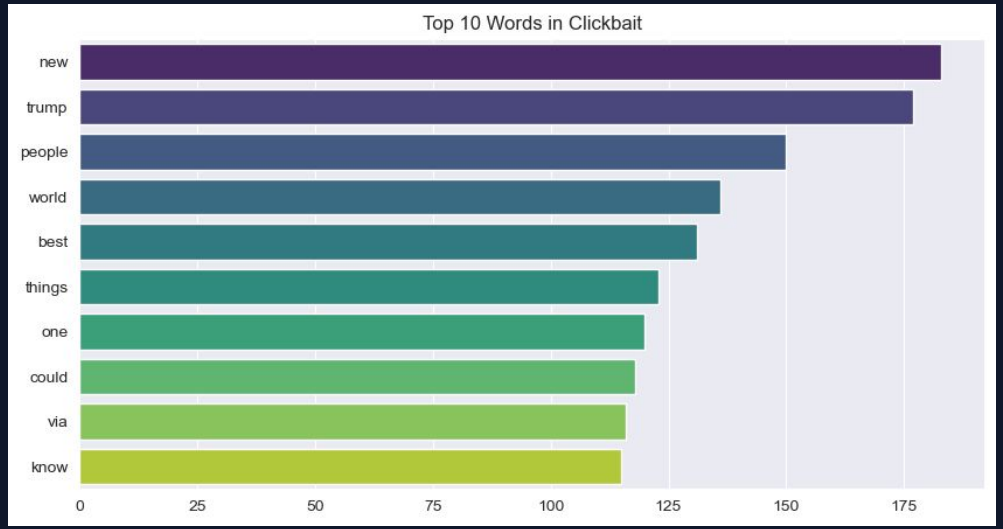
POC: Data Processing

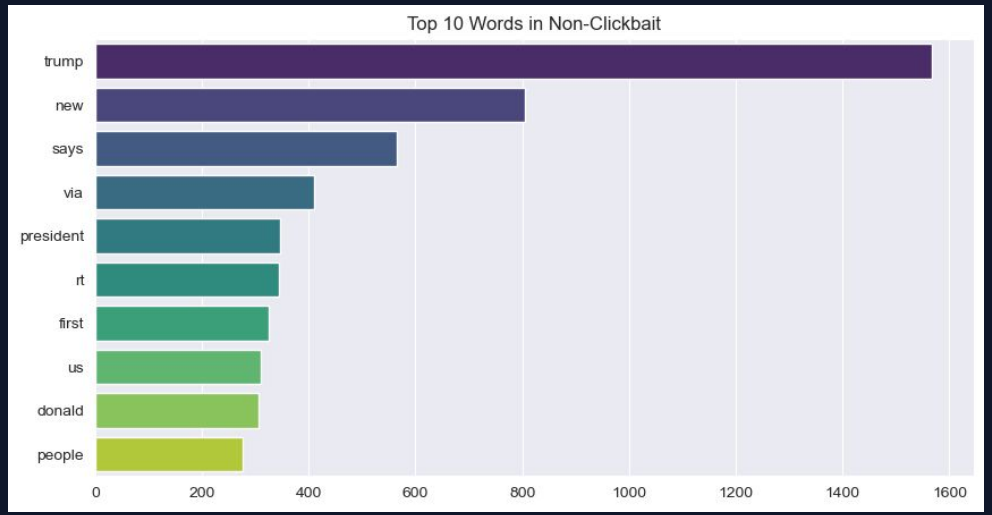
- Remove empty rows

- Add new features:

  Character count
  Word count
  Exclamation marks (!)
  Question marks (?)
  ALL CAPS presence
  Sentiment polarity
  Sentiment subjectivity
  Question word markers


Training Data: Class Distribution
count
Label (0=Non-Clickbait, 1=Clickbait)
Training Data: Headline Length (Chars)
Count
text_len
label
0
1
Training Data: Headline Length Density (Chars)
Density
text_len
label
0
1
Training Data: Word Count per Class
word_count
label


Top 10 Words in Clickbait
new
trump
people
world
best
things
one
could
via
know
0
25
50
75
100
125
150
175


Top 10 Words in Non-Clickbait
trump
new
says
via
president
rt
first
us
donald
people
0
200
400
600
800
1000
1200
1400
1600

POC: Feature & Model Summary

- Classical Baseline: Random Forest (TF-IDF + hand-crafted features)
- Transformer: DistilBERT fine-tuned for headline classification
- Metrics: Precision, recall, F1-score for both models

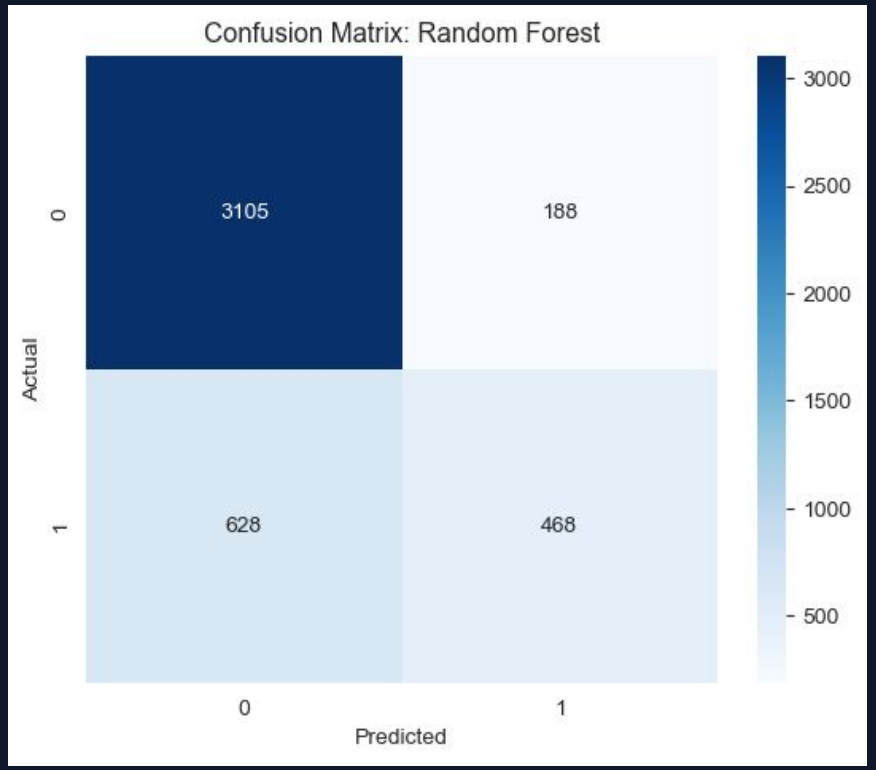
POC: Results Snapshot

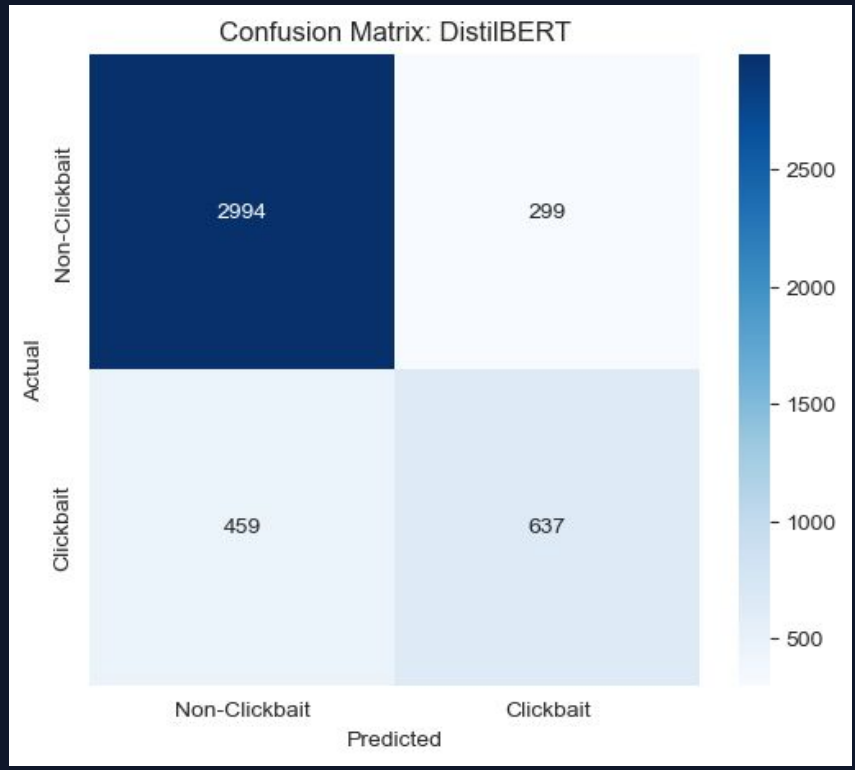
- Classical RF: 81% accuracy, 77% precision, 68% recall, 71% f1-score

  stronger for non-clickbait, recall lower for clickbait class

- DistilBERT: 83% accuracy, 77% precision 75% recall 76% f1-score

  SOTA transformer baseline, expects further improvement with tuning


Confusion Matrix: Random Forest
Actual
0
1
3105
188
628
468
0
1
Predicted
3000
2500
2000
1500
1000
500


Confusion Matrix: DistilBERT
Actual
Non-Clickbait
Clickbait
2994
299
459
637
Non-Clickbait
Clickbait
Predicted
2500
2000
1500
1000
500

Discussion & Future Work

- Expand data sources for better generalization
- Hyperparameter tuning and cross-validation on transformers
- Analyze and explain feature contributions in detail

Questions?