Evaluation of Cyberbullying using Optimized Multi-Stage ML Framework and NLP
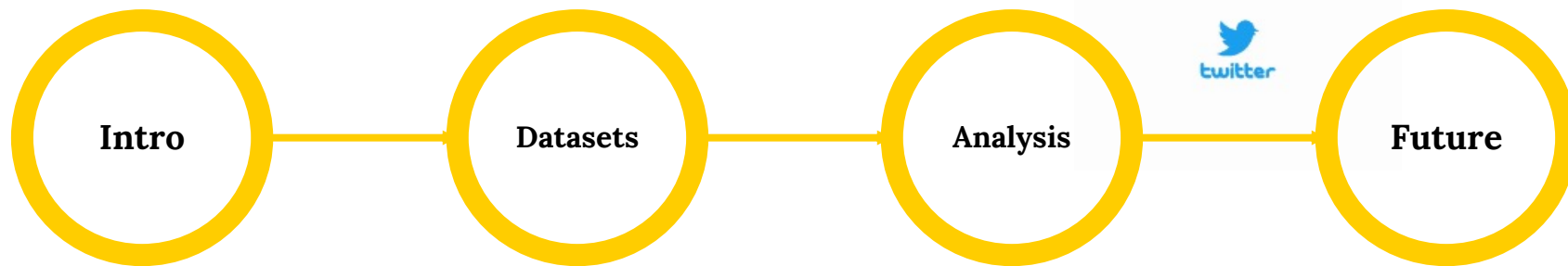
Lida Ketsbaia, Biju Issac and Xiaomin Chen
Computer and Information Sciences, Northumbria University, UK

Northumbria University
NEWCASTLE

Presentation

Intro → **Datasets** → **Analysis** → **Future**

- Optimization: PSO, GA
- NLP: RoBERTa, DstilBERT, XLNET
- Sampling methods

Background

How online hate has changed in the past year

Online Hate

Since the last Annual Progression:

- ◉ Study: 44% of children who had been cyberbullied before lockdown said it happened even more during lockdown.
  - ○ 39% of Cyberbullies said they did it more during lockdown.
- ◉ 84% increase in the number of online disability hate crimes reported to police in Wales last year.
- ◉ 900% increase on online abuse towards Asian people on Twitter since the start of the Coronavirus pandemic.
- ◉ 70% increase in online hate speech amongst children and teens

4

Datasets

Bring the attention of your audience over a key concept using icons or illustrations

Datasets

Dataset 1: based on the article "Automated Hate Speech Detection and the Problem of Offensive Language". The dataset was created using Twitter's API. Tweets were search containing terms from a lexicon resulting in tweets from 33,458 Twitter users. A random sample of 25,000 tweets were chosen and labelled as one of three categories: "hate speech", "offensive", "neither hate-speech or offensive".

Due to the research specifically targeting hate speech this dataset only used the "hate-speech" and "non-hatespeech" data.

○ **Dataset 2:** The researchers created a dataset regarding cyberbullying due to the lack of cyber-bullying datasets within research. The text in the dataset were marked as "cyberbullying" based on whether two of the annotators identified the text and "cyberbullying". The resulting dataset contains 12,772 samples with 86% being labelled as "non-cyberbullying
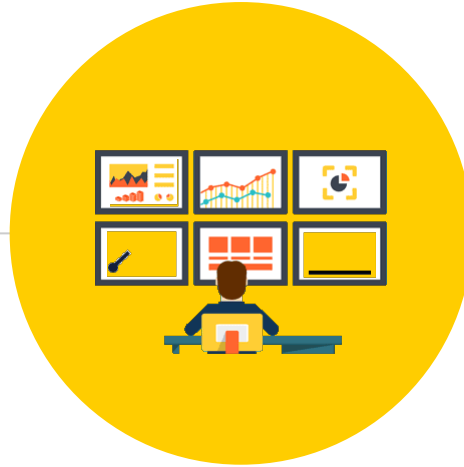
Analysis

Bring the attention of your audience over a key concept using icons or illustrations

Analysis

- **Multi stage framework**
  - Combination of RUS & SMOTE Balancing
  - Implement PCA Feature Selection
  - Employ PSO + Logistic Regression
  - Employ Genetic Algorithm + Logistic Regression

- **NLP Techniques:**
  - XLNET
  - DistilBERT
  - RoBERTa

Machine Learning Framework

○ Multi stage framework

  ○ Combination of RUS & SMOTE Balancing

  ○ Implement PCA Feature Selection

  ○ Employ PSO + Logistic Regression

  ○ Employ Genetic Algorithm + Logistic Regression

Framework

RUS / SMOTE

RUS: Random undersampling removes examples from the majority class.

SMOTE: Synthesize new examples from the minority class

PCA

Dimension reduction tool to reduce a large set of variables to a small set that still contains most of the information in the original set.

PSO / GA

Genetic Algoithm: A method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution.

NLP Techniques

RoBERTa

A robustly optimized method for pretraining natural language processing (NLP) systems that improves on Bidirectional Encoder Representations from Transformers, or BERT,

XLNET

An extension of the Transformer-XL model pre-trained using an autoregressive method to learn bidirectional contexts by maximizing the expected likelihood over all permutations of the input sequence factorization order.

DistilBERT

DistilBERT is a small, fast, cheap and light Transformer model trained by distilling BERT base. It has 40% less parameters than *bert-base-uncased*, runs 60% faster while preserving over 95% of BERT's performances

RESULTS

Classifier	Dataset 1			
	Accuracy	**Precision**	**Recall**	F1
Machine Learning Framework				
LR	85.90%	86.43%	85.30%	86.11%
PCA-GA-LR	88.34%	88.47%	88.34%	87.58%
PCA-PSO-LR	86.98%	87.27%	86.98%	85.90%
NLP				
DistilBERT	86.78%	87.93%	86.77%	85.17%
XLNet	87.78%	87.50%	82.58%	83.71%
RoBERTa	86.43%	87.18%	86.43%	85.36%

Classifier	Dataset 2			
	Accuracy	**Precision**	**Recall**	**F1**
Machine Learning Framework				
LR	85.51%	92.08%	85.51%	88.14%
PCA-GA-LR	86.58%	90.9%	86.58%	88.47%
PCA-PSO-LR	85.64%	92.06%	85.64%	88.21%
NLP				
DistilBERT	94%	92.96%	94%	92.69%
XLNet	94.47%	94.43%	93.48%	93.88%
RoBERTa	93.89%	94.26%	93.89%	91.34%

5 Future Work

Future Work

- Look at other optimization methods such as Ant Colony optimization or Artificial Bee

- Bring together balancing, optimization and NLP methods

- Sarcasm Detection