SemEval-2022 Task 5: Multimedia Automatic Misogyny Identification

Overview

Dataset

Memes were downloaded, by site scraping and manual download from:

Social Media Platforms

e.g. Twitter and Reddit;

Websites

e.g. 9GaG, Knowyourmeme and Imgur;

The procedure for collecting relevant consisted of 4 main activities, performed to collect a proper number of misogynous memes,

Searching for threads dedicated to memes with **women as the subject**;

Searching for **threads or conversations** dedicated to or written by persons who identify as **anti-women or antifeminist**;

exploring **discussions** in recent events involving **famous wome**n (such as Michelle Obama);

searching by **keywords and/or hashtags** such as #girl, #girlfriend, #women, #feminist

Dataset creation - Annotations

The final (duplicates have been removed) benchmark dataset released for the MAMI challenge is composed of **10k memes** for training and **1k** for testing (balanced between classes).

The dataset has been labelled using **crowd-sourcing platforms** according to the following primary questions:

- Is this meme misogynous or not?
- If the meme is misogynous, what are the main categories to which the meme belongs (shaming, stereotype, objectification, violence)?

Different types of misogyny

- **Shaming**: which expresses disapproval of women's behaviors and physical appearances compared to a given type of expectation, e.g. body shaming.
- **Stereotype**: which expresses a generalized belief concerning women in different contexts, e.g., societal role, personality and behaviors.
- **Objectification**: which consists in considering and/or treating women as objects.
- **Violence**: which may instigate or express violence against women.

Shaming

Stereotype

MAMI

Objectification

Violence

Dataset – Samples and labels

Memes were annotated by 3 observers and the final label was given according to the majority of the labels (2/3).

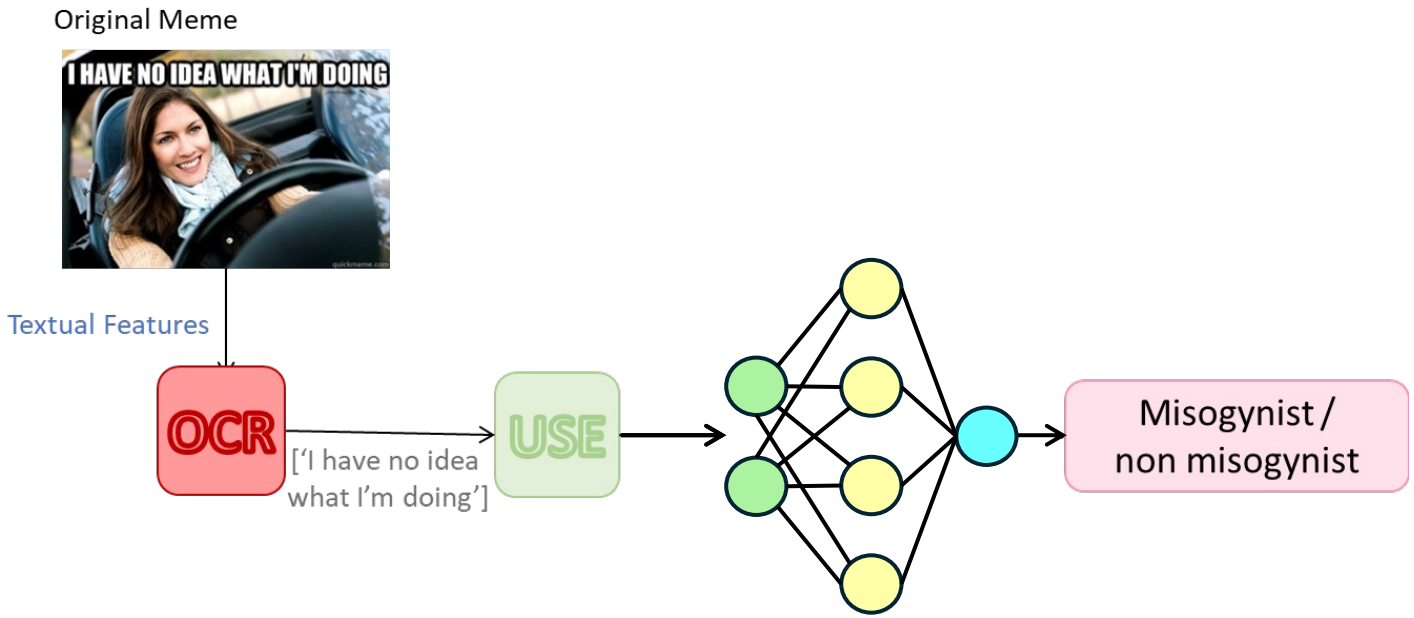
file_name	misogynous	shaming	stereotype	objectification	violence	Text Transcription
10846.jpg	1	0	1	1	1	SANDWICH!!!!!! don't make me tell you twice woman.

Inter-annotator agreement _ Fleiss-κ coefficient

Regarding the agreement on the misogynous vs not misogynous annotations, we estimated a coefficient equal to 0.5767, while for the type of misogyny labelling we derived a coefficient equal to 0.3373.

Textual Baseline

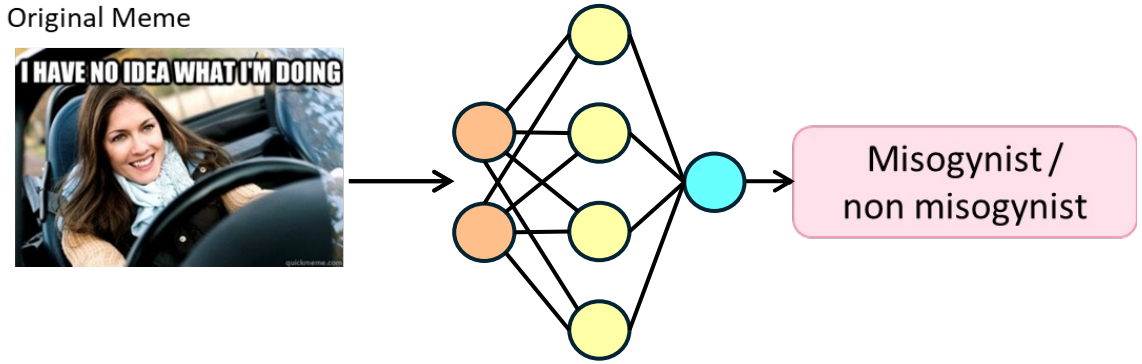
Deep representation of text, a fine-tuned sentence embedding using the USE pre-trained model;



https://github.com/MIND-Lab/SemEval2022-Task-5-Multimedia-Automatic-Misogyny-Identification-MAMI-

Visual Baseline

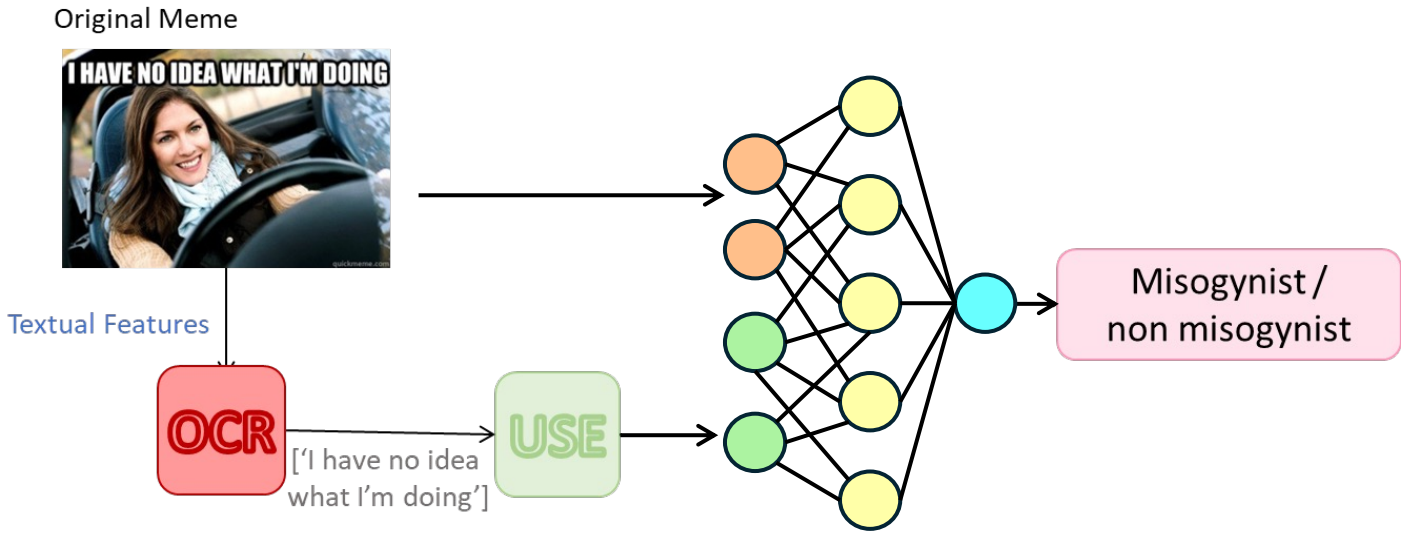
Deep representation of image content, based on a fine-tuned image classification model grounded on VGG-16



https://github.com/MIND-Lab/SemEval2022-Task-5-Multimedia-Automatic-Misogyny-Identification-MAMI-

Multimodal Baseline

A concatenation of the previous deep image and text representations through a single layer neural network



https://github.com/MIND-Lab/SemEval2022-Task-5-Multimedia-Automatic-Misogyny-Identification-MAMI-

Partecipation report

Sub-task A was attempted by **65 teams**, where 47 of them (72%) outperformed the best provided baseline.

Sub-task B was attempted by **41 teams,** where 35 of them (85%) outperformed the best MAMI baseline

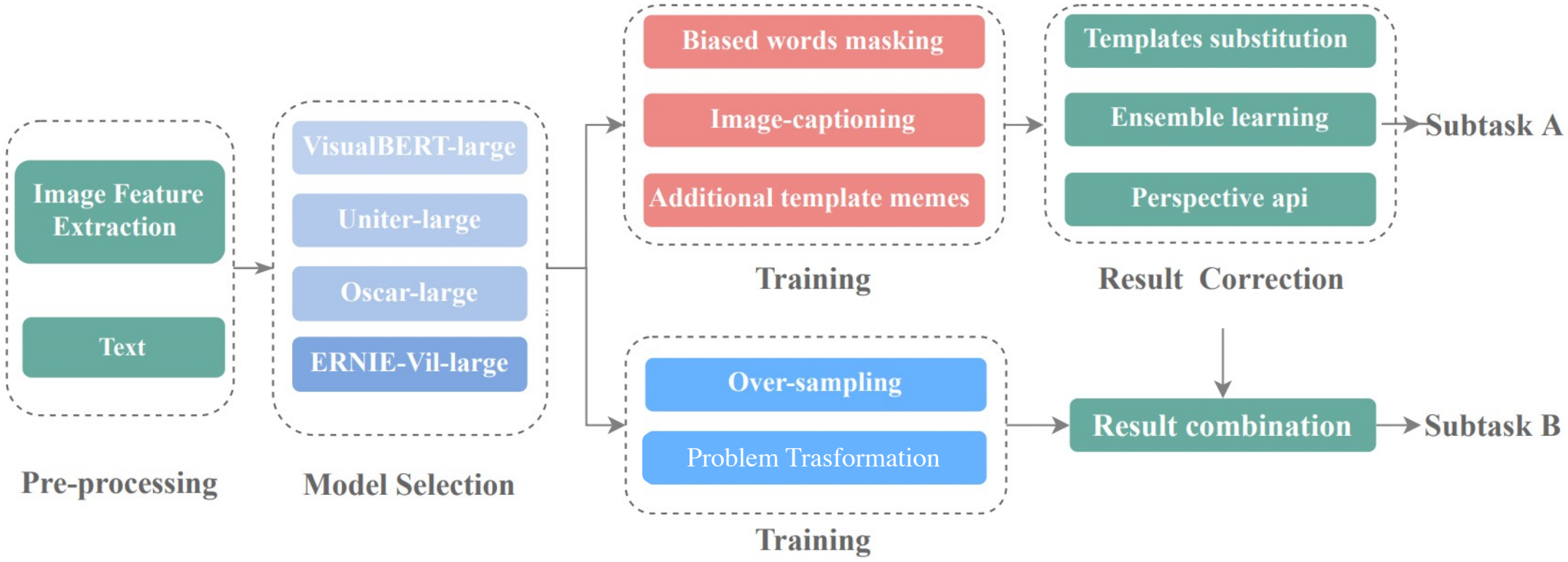
90% of the team exploited **pre-trained models**, distinguished in text-based, mostly based on BERT (e.g. RoBERTa), and image-based models, mostly based on VisualBERT.

Participant Systems and Results

DD-TIG

Task A: 2° - Task B: 2°

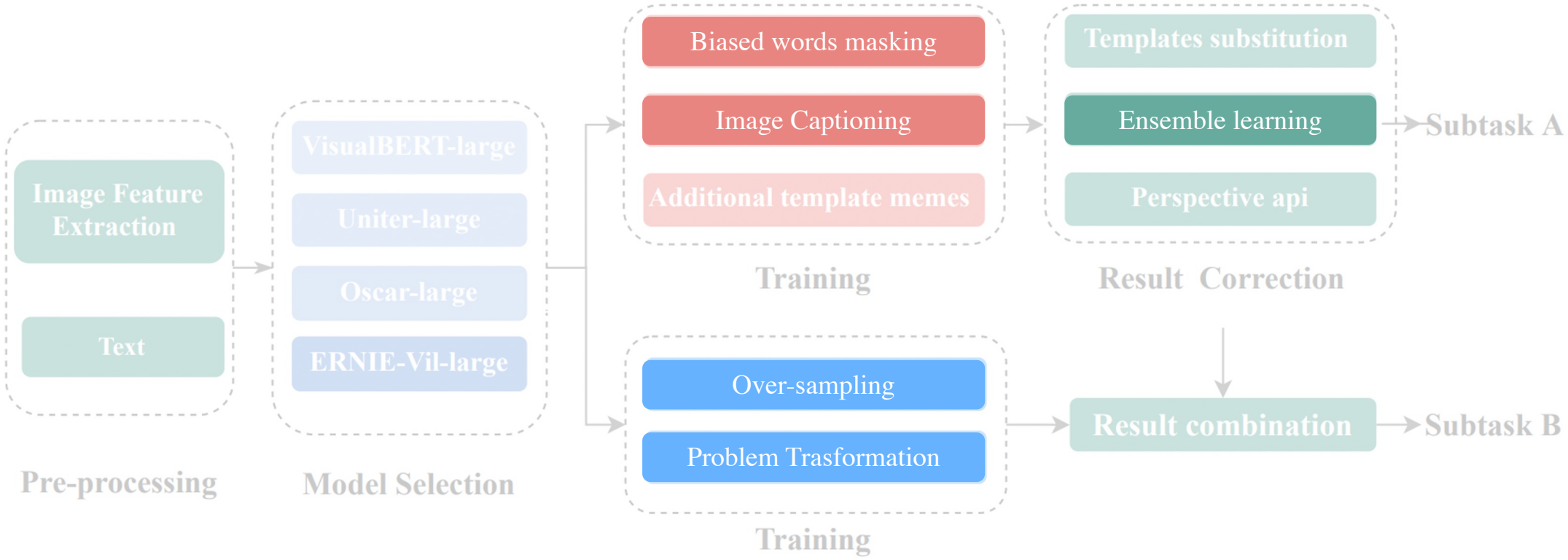
They investigate several of the most recent visual language transformer-based multimodal models. They mitigate problems of biased words and template memes. They transform task B multi-label problem into a multi-class one.



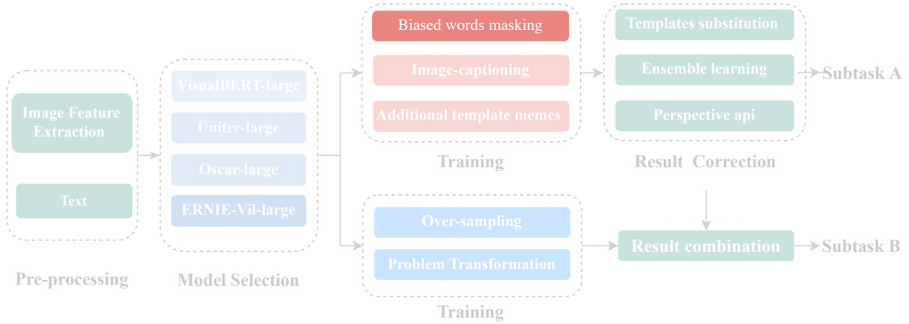
DD-TIG

Task A: 2° - Task B: 2°

They investigate several of the most recent visual language transformer-based multimodal models. They mitigate problems of biased words and template memes. They transform task B multi-label problem into a multi-class one.



Biased words masking

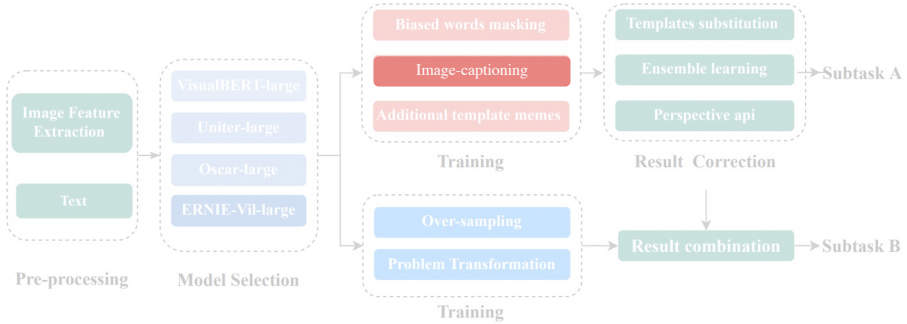


Why? Our models tend to associate some non-misogynous texts containing specific words with an unreasonably high misogynous score. This situation is known as **unintended bias**, in which models learn usual associations between words (commonly called identity terms) which causes them to classify content as misogynous just because it contains one **identity word.**

Proposed approach:

manually collect a list of biased words, including **synonyms of woman**, **dirty words**, and **controversial words** related to feminism, and mask the token *[mask]* by a 20 percent probability.

Image Captioning



Why? for some misogynous memes, **image and text are weakly aligned**. Thus, there is a semantic gap between visual and textual information.

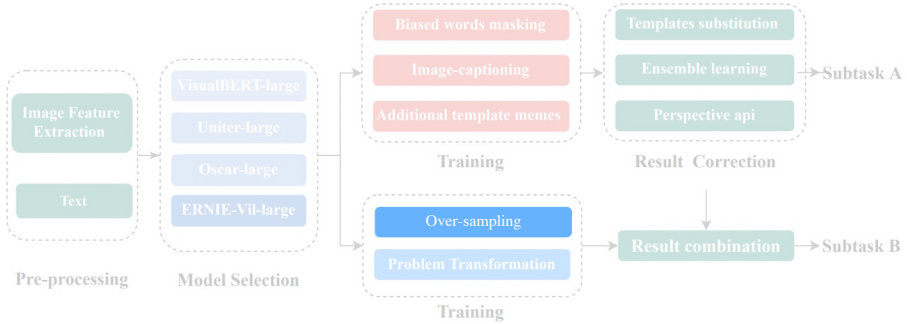
Proposed approach:

Memes are sent into an image caption model to generate additional descriptions for visual contents of each meme.

A photo of numerous cars in a traffic jam.

Visual contents

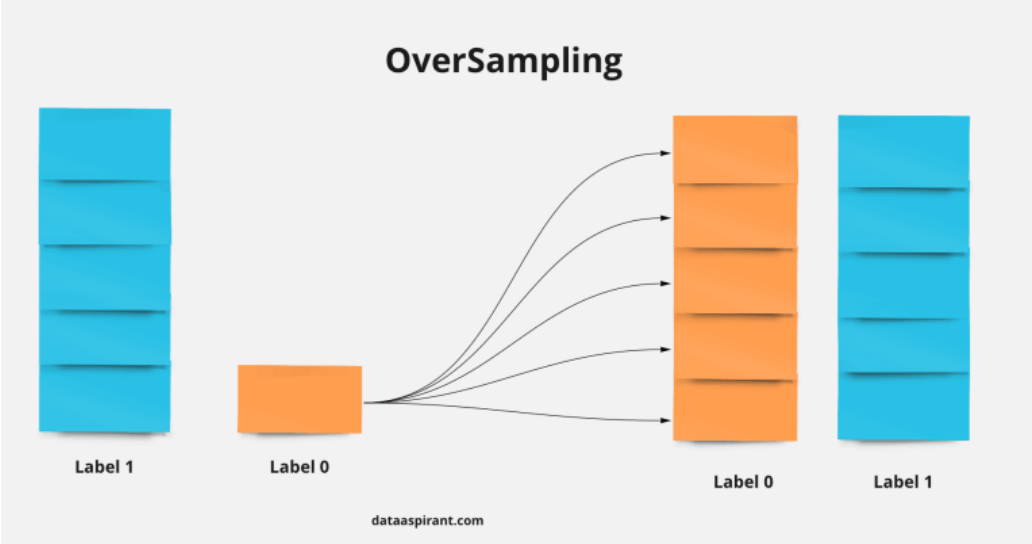
Over sampling



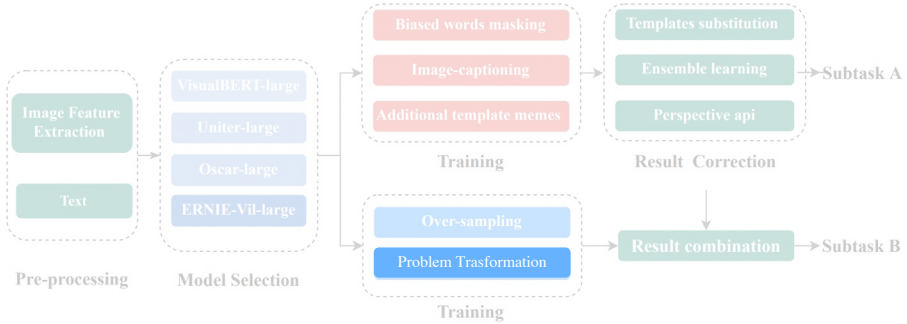
Why? the number of positive samples and negative samples in all misogynous categories is **widely imbalanced**.

Proposed approach:

Hence, up-sampling of data is done using **over-sampling** on the positive sample.



Problem Transformation



Why? A conventional way to solve a **multi-label problem** is to transform it into **binary classification problems** where one binary classifier is independently trained for each label.

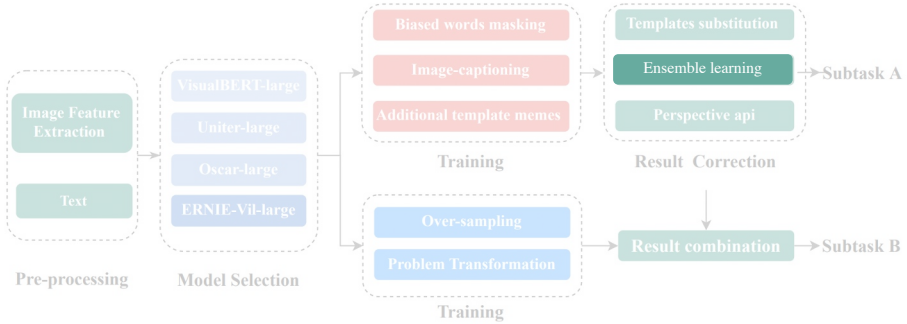
Proposed approach:

They transformed the multi-label problem into multi-class problems. Every possible combination of output labels ([0, 0, 0, 0], [1, 0, 0, 0], · · ·) will be taken as a class.

file_name	misogynous	shaming	stereotype	objectification	violence
10846.jpg	1	0	1	1	1

[0,1,1,1]

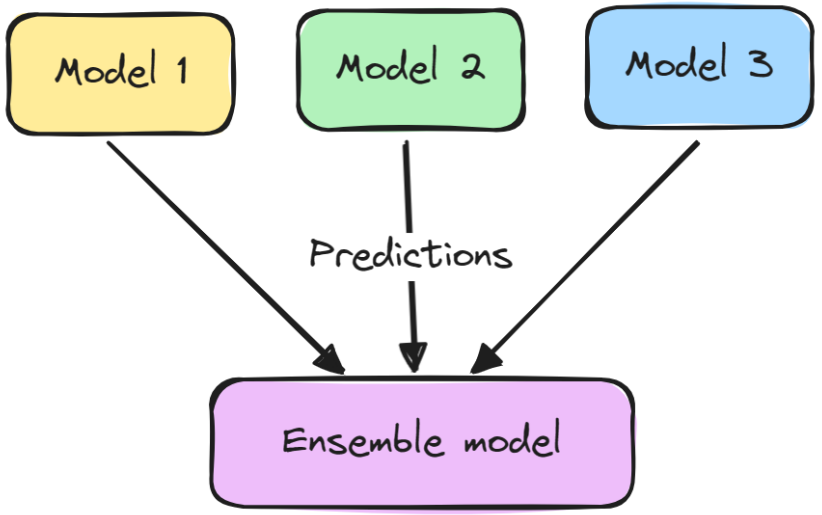
Ensemble learning



Why? Some models show a **high recall** and **low precision** and vice versa. So a collection of models may balance out individual weaknesses to achieve better performance than any single model used in the ensemble.

Proposed approach:

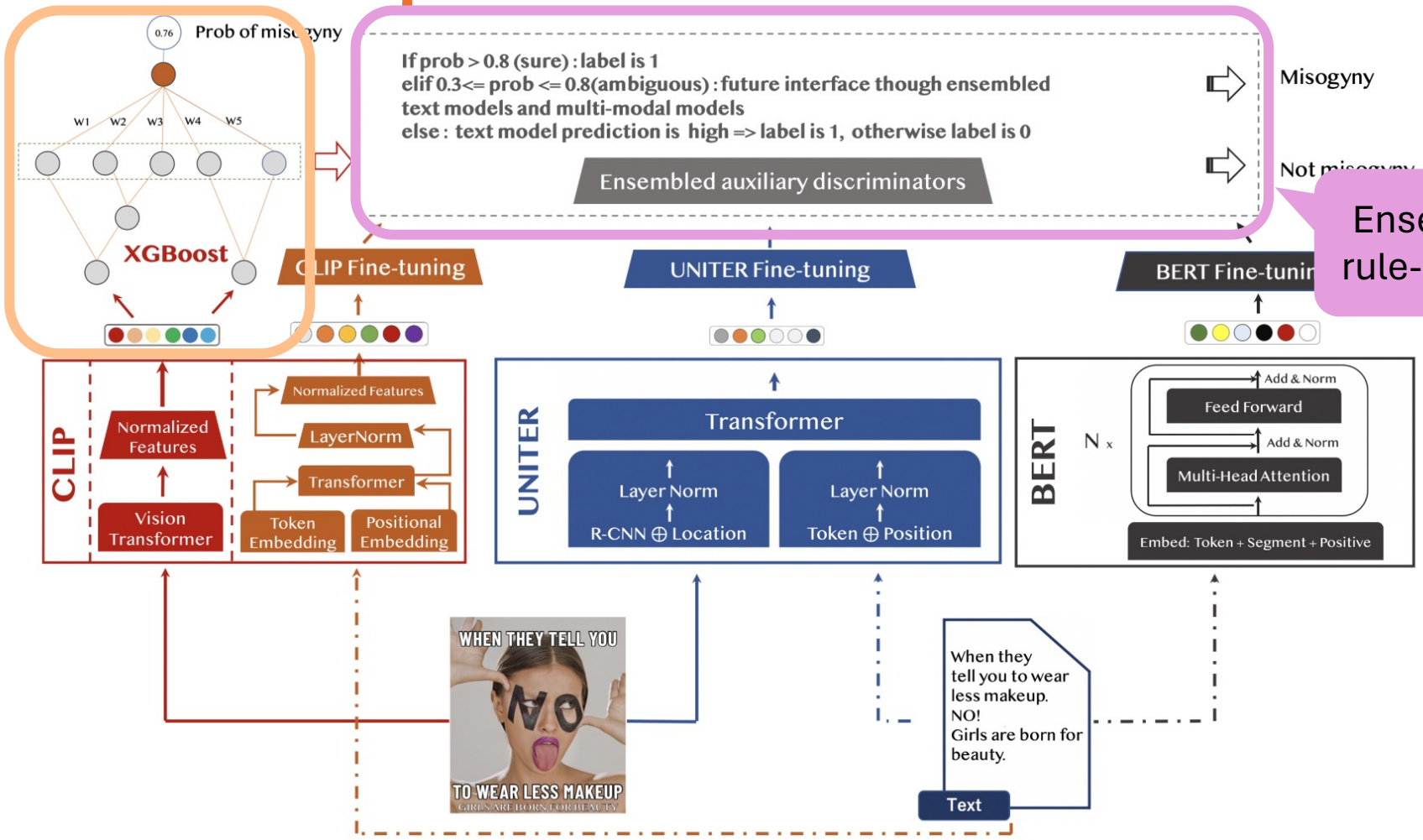
they improve the whole system's generalizability and robustness with **ensemble learning**, where predictions of multiple base models are combined with the method of **majority Voting.**



SRCB

Task A: 1° - Task B: 1°

They investigate the single-stream UNITER and dual-stream CLIP multimodal pretrained models. They propose the an **ensemble system** of **Pretraining models**, **Boosting method** and **Rule-based adjustment**, text information is fused into the system using a late sequential fusion scheme.



XGBoost classifier with CLIP image features

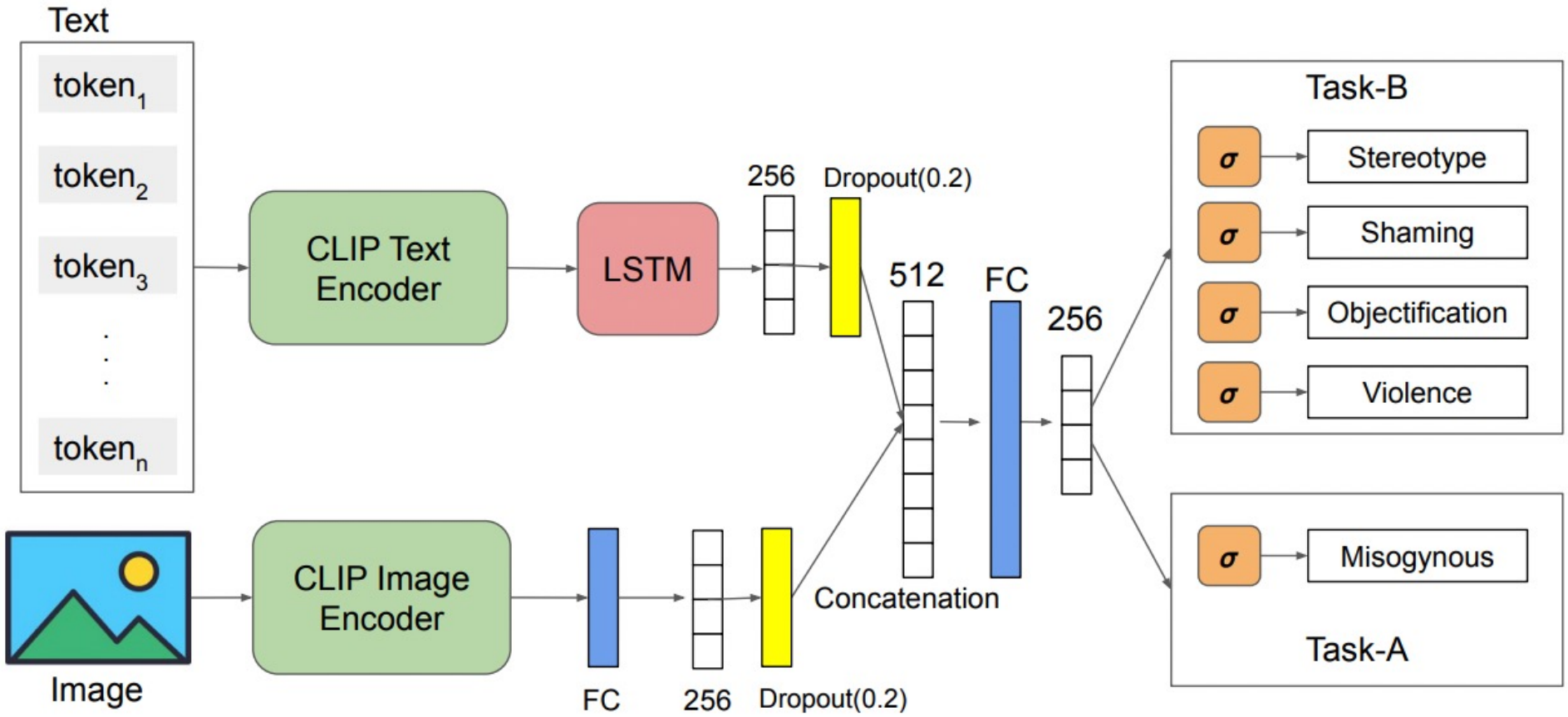
Ensemble model with rule-based adjustment

Something more feasible…

TIB-VA

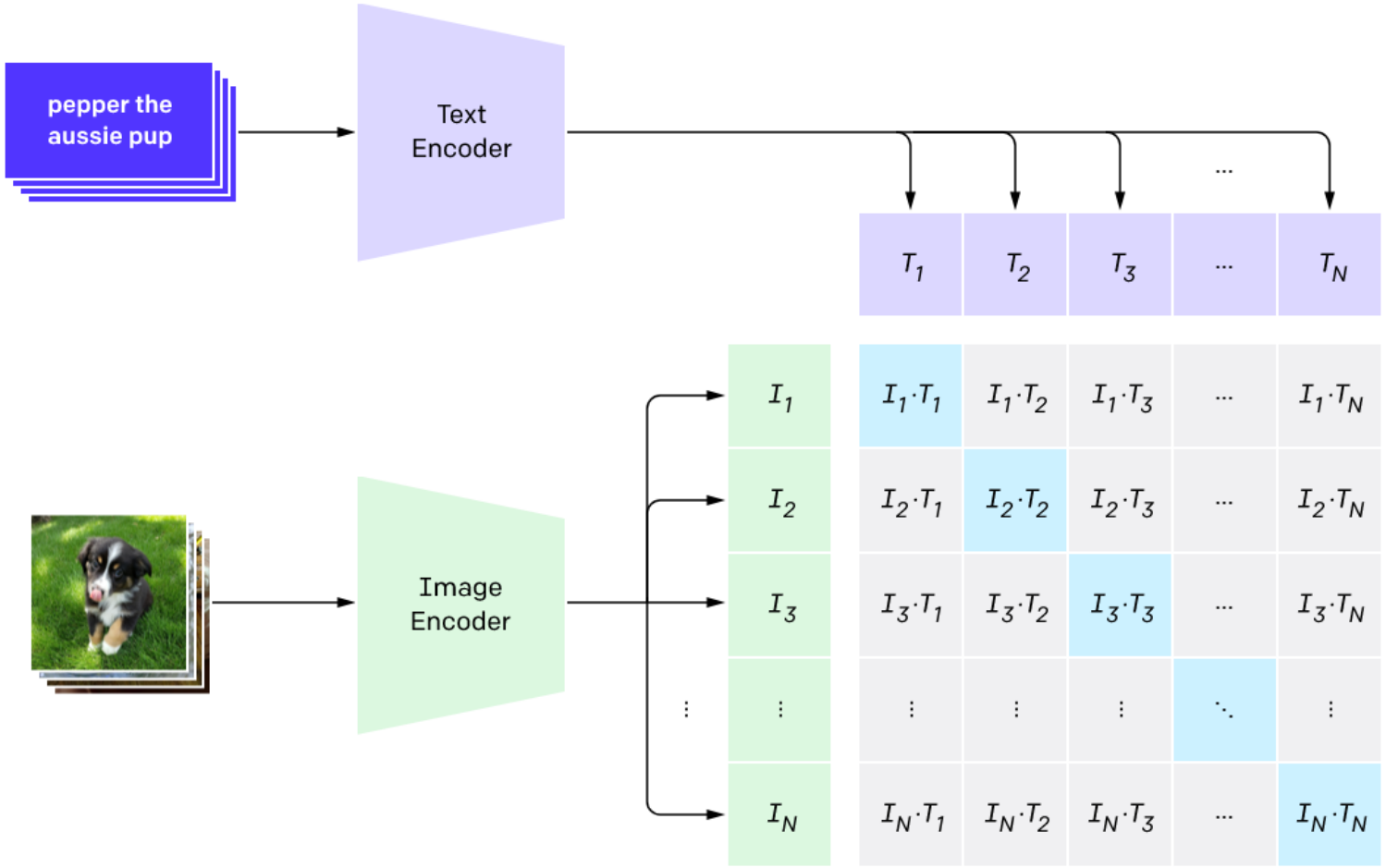
Task A: 3° - Task B: 1°

They present a multimodal architecture that combines textual and visual features in order to detect misogynous meme content. The proposed solution is built on the **pre-trained CLIP** model to extract features for encoding textual and visual content.



CLIP

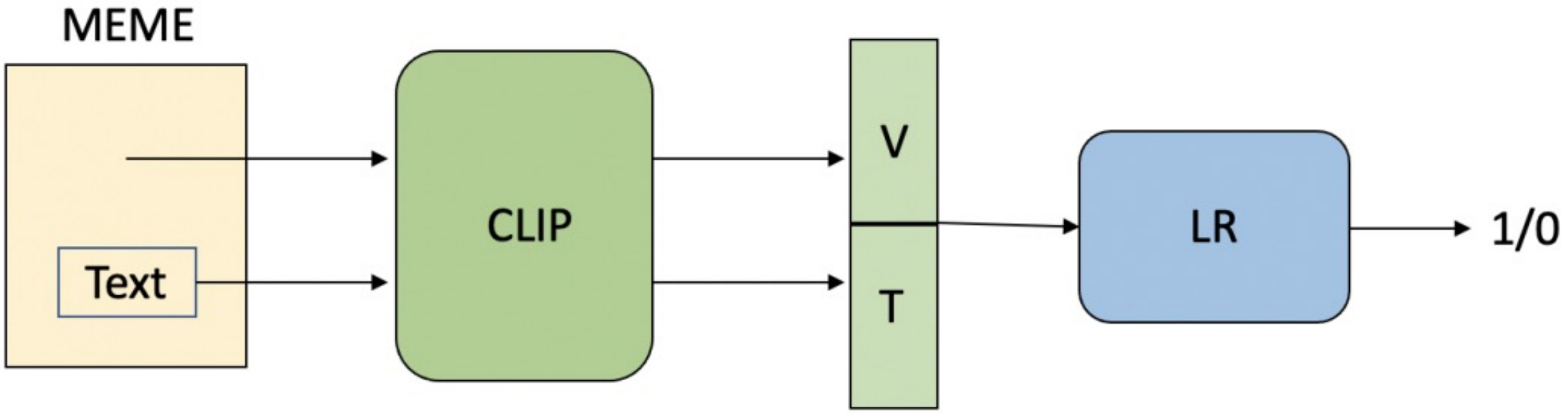
(Radford et al., 2021)



RIT Boston

Task A: 2° - Task B: 11°

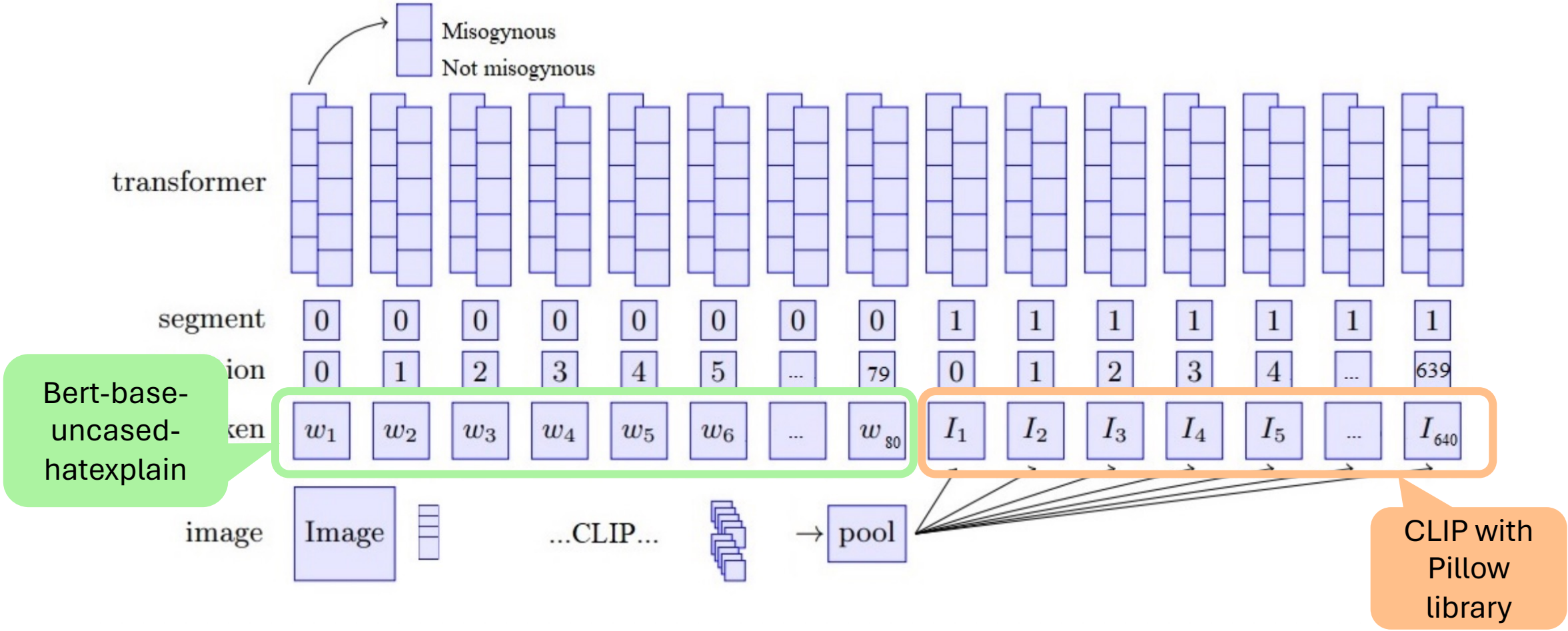
They used the CLIP model provided by OpenAI to obtain coherent V and L features and then simply used a logistic regression model to make binary predictions.

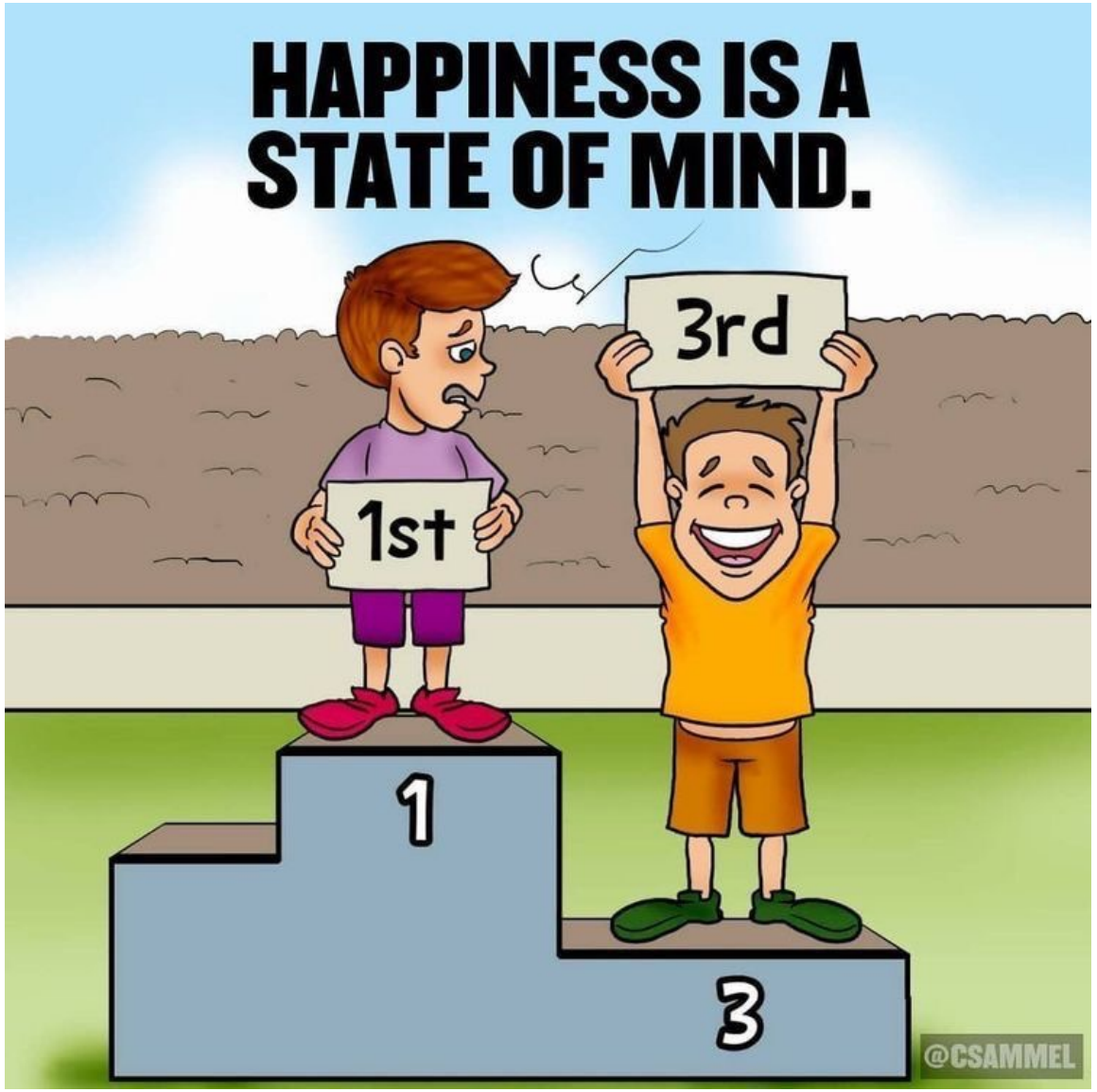


UniBO

Task A: 4° - Task B: 4°

They combined a **BERT Transformer** with **CLIP** for the textual and visual representations. Both textual and visual encoders are fused in an early-fusion fashion through a Multimodal Bidirectional Transformer with unimodally pretrained components.





Good luck!

References

- Fersini, Elisabetta, et al. "SemEval-2022 Task 5: Multimedia automatic misogyny identification." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022.
- Zhou, Ziming, et al. "DD-TIG at semeval-2022 task 5: Investigating the relationships between multimodal and unimodal information in misogynous memes detection and classification." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022.
- Zhang, Jing, and Yujin Wang. "SRCB at SemEval-2022 task 5: Pretraining based image to text late sequential fusion system for multimodal misogynous meme identification." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022.
- Chen, Lei, and Hou Wei Chou. "RIT boston at semeval-2022 task 5: Multimedia misogyny detection by using coherent visual and language features from CLIP model and data-centric AI principle." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022.
- Muti, Arianna, Katerina Korre, and Alberto Barrón-Cedeño. "UniBO at semeval-2022 task 5: A multimodal bi-transformer approach to the binary and fine-grained identification of misogyny in memes." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. 2022.
- Hakimov, Sherzod, Gullal S. Cheema, and Ralph Ewerth. "TIB-VA at semeval-2022 task 5: A multimodal architecture for the detection and classification of misogynous memes." *arXiv preprint arXiv:2204.06299* (2022).