Project Literature

Paper 91：

Methodology Overview

The study tackles two tasks from EXIST 2024:

- **Task 4:** Detecting sexism in memes.
- **Task 5:** Detecting the creator’ s intention (direct vs. judgmental).

Although memes are multimodal (image + text), the authors **focused solely on textual content**, since images in both classes were visually similar and often uninformative.

Key Technical Steps

555. **Data Preprocessing**
    - Text extracted from memes was normalized: converted to lowercase, with URLs, usernames, and hashtags removed.
    - Additional cleaning did not improve performance, so preprocessing remained minimal.
556. **Dataset Construction**
    - Three training datasets were built to expand limited labeled data:
        - **Original:** Provided training set.
        - **Simple:** Original + translated into the opposite language (English ↔ Spanish).
        - **Back:** Original + back-translated via German using ChatGPT.
    - This doubled or tripled training size, serving as data augmentation.
557. **Model Selection**
    - Two multilingual Transformer-based models:
        - **bert-base-multilingual-uncased** (BERT).
        - **xlm-roberta-base** (RoBERTa).
    - Fine-tuned with extensive **hyperparameter search** using **Optuna**:
        - Batch size ∈ {8, 16, 32}.
        - Learning rate ∈ {1e-5, 3e-5, 5e-5}.

        - Weight decay ∈ {0.01, 0.1}.
    - Optimized hyperparameters differed by task and model.

558. **Learning with Disagreement (LeWiDi)**
    - Instead of collapsing annotators’ labels into a single majority vote, models were trained from **different annotator perspectives** (gender, age, education, ethnicity).
    - Eight major perspectives were selected; each produced a specialized model.
    - Undersampling balanced the datasets per perspective.
559. **Ensemble Approach**
    - Models from different perspectives were combined into ensembles.
    - **Weight-based voting** was applied: each perspective’ s output was weighted (0.5–1.75) to maximize F1.
    - Three ensemble variants per task (BERT-based, RoBERTa-based, hybrid) were created.
560. **Evaluation**
    - Performance measured using **F1 score**, **ICM-Hard**, and **ICM-Soft**.
    - Confusion matrices and error analysis highlighted misclassifications, often due to cultural/linguistic nuances or ambiguous humor.

Paper 92:

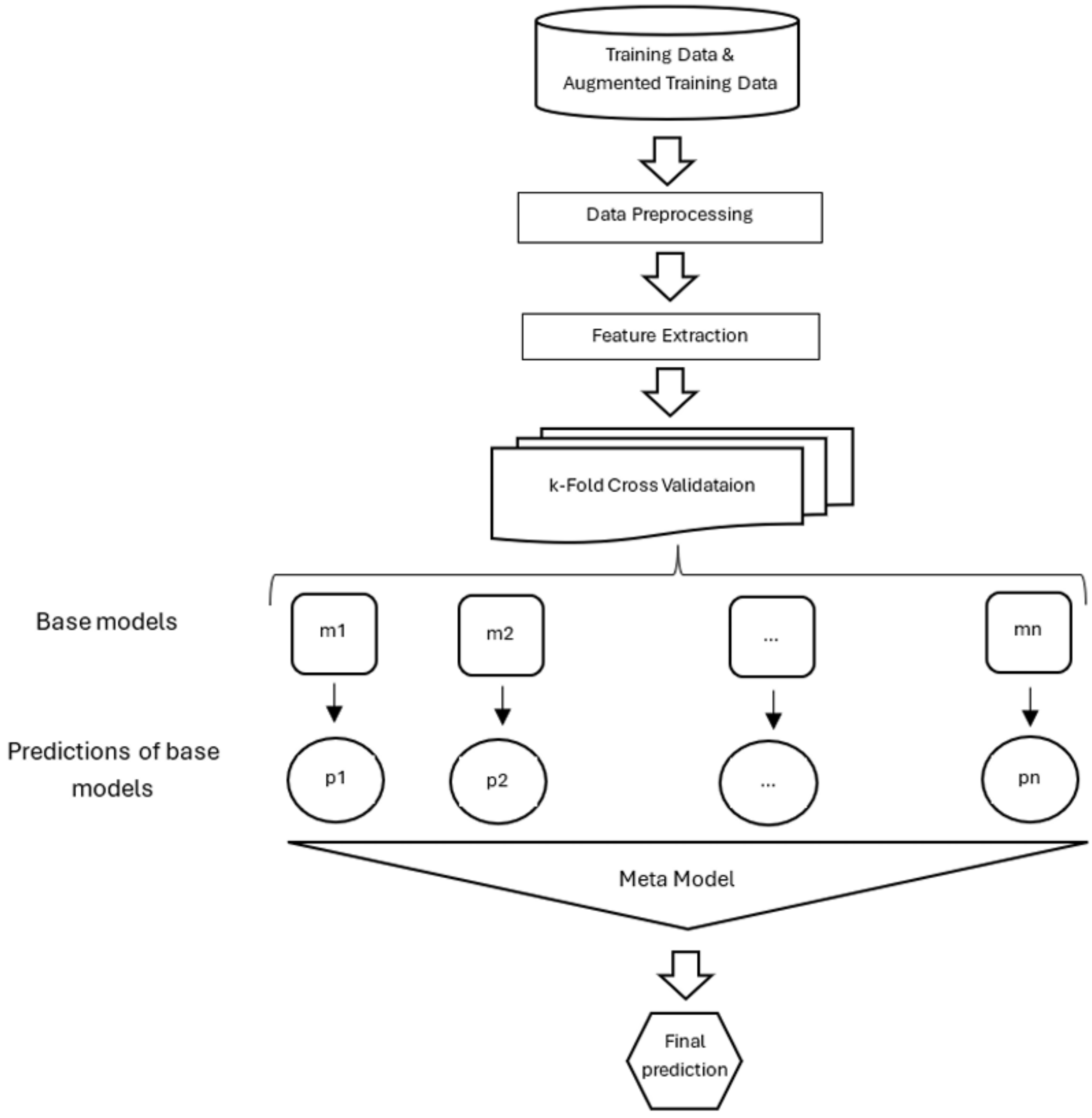
Dataset Augmentation:

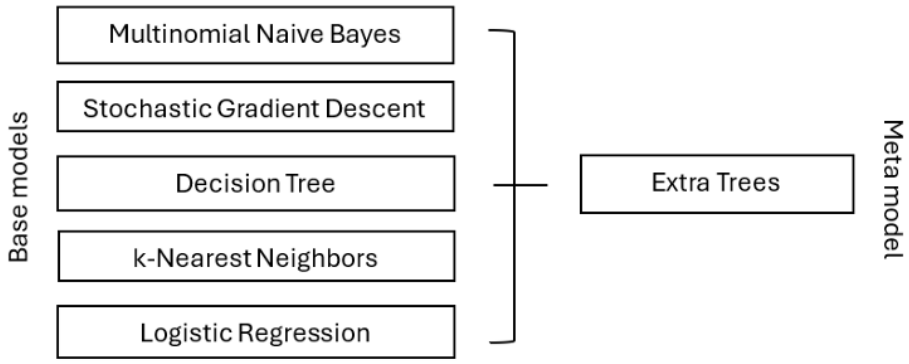
we employed two augmentation methods: synonym replacement [26] and contextual augmentation [27].

Both of these methods were applied at the word level. These approaches have been demonstrated to maintain the labels of the original texts, as the semantic meanings of the augmented versions remain unaltered. WordNet [28, 29] is an English lexical database that groups words into sets of synonyms. The implementation of the synonym replacement method using WordNet for English tweets has been demonstrated to provide high-quality data. Due to the limitations of the available Spanish synonym lexical database, we selected three different Transformer language models from HuggingFace [30] for the implementation of the contextual augmentation method. The models employed include BERTIN [31], ALBERT Base Spanish [32], and RoBERTuito [33]. In total, the original train and development datasets were subjected to a tenfold augmentation using the aforementioned methods

Avariety of techniques has been developed for constructing ensembles, including bagging (bootstrap aggregating) [37], boosting [38, 39], AdaBoost (adaptive boosting) [40], voting [41], and stacking [42]. As stated in [35], ensemble methods are considered the state-of-the-art solution for many machine learning challenges. They have also been widely used in previous years for the EXIST shared task, as outlined in Section 2. However, the ensemble learning methods reported in the EXIST working notes are predominantly in conjunction with Transformer models. The ensemble of machine learning models remain under-researched for the detection of sexism in tweets. Thus, for the EXIST 2024 shared task, we utilize ensemble of machine learning models.

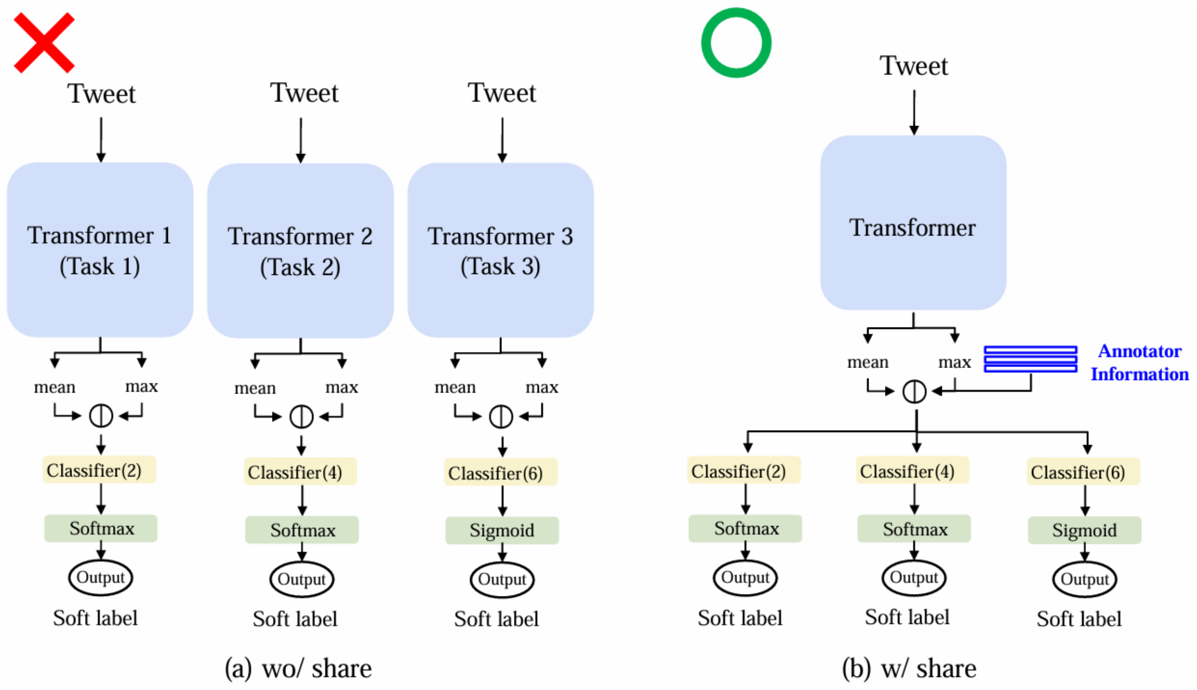
We first train multiple different base models (e.g., decision trees, logistic regression, support vector machines), each of which performs classification or regression predictions on tweets. We then use a meta-learner (e.g., linear regression or another classifier) to combine these predictions and generate the final output.





Paper 93:

Hard-Parameter share：



(a) wo/ share

(b) w/ share

Dataset Augmentation:

removed usernames, URLs, percentages, time, dates, hashtags, and emojis, as these elements were unlikely to influence the annotators' judgments (see Fig. 2). Subsequently, all characters are converted to lowercase to ensure uniformity and reduce the complexity of the text data. We also translated the text from English to Spanish and then back to English via the Google Translator API, effectively doubling the amount of data and introducing subtle variations that can improve the robustness of our models,

use the AEDA technique [7] to augment the text data by randomly segmenting sentences and inserting punctuation marks from a predefined set “.”, “;”, “?”, “:”, “!”, “,”.

Paper 94: include clip

Initially, we fine-tuned the pre-trained Contrastive Language-Image Pre-training (CLIP) model [6], a deep learning model developed by OpenAI that has demonstrated exceptional performance in vision and language tasks. CLIP is trained on a vast dataset of 400 million (image, text) pairs, enabling it to learn rich visual and textual representations.

4.1. Task 4 In the context of the EXIST2024 Task 4 for binary classification of memes, our methodology involved a two-stage approach to address the classification problem. The dataset was splitted into train/val test in order to evaluate the model. Initially, we fine-tuned the pre-trained Contrastive Language-Image Pre-training (CLIP) model [6], a deep learning model developed by OpenAI that has demonstrated exceptional performance in vision and language tasks. CLIP is trained on a vast dataset of 400 million (image, text) pairs, enabling it to learn rich visual and textual representations. During the first stage of our approach, we incorporated a fully connected network specifically designed for binary classification within the CLIP model. This allowed the model to learn task-specific features and improve its ability to differentiate between the two classes of memes. Once the fine-tuning process was complete, we discarded the fully connected layer, as our primary focus was on obtaining high-quality embeddings that could be used to train a more specialized classifier. In the second stage, we employed a Light Gradient Boosting Machine (LightGBM) [4] classifier, a highly efficient and effective gradient boosting framework that uses tree-based learning algorithms. By training the LightGBM classifier with the CLIP embeddings obtained from the first stage, we aimed to further enhance the classification performance. This two-stage training strategy allowed us to leverage the strengths of both CLIP and LightGBM, resulting in improved binary classification of memes in the EXIST2024 Task 4. We tried a logistic regression classifer instead of LightGBM for the final classification (See table 1). 4.2. Task 6 For the EXIST2024 Task 6, after segmenting the dataset into train/val and test, ensuring a robust evaluation process, we meticulously crafted a methodology involving three stages: two fine tunings and a LightGBM training. The classification problem we tackled was multifaceted, encompassing six categories. Initially, we had to take into account binary classification to distinguish between sexist and non-sexist memes. Then, if the meme was deemed sexist, we further dissected the binary category into five classes that could overlap: stereotyping, objectification, ideological, sexual violence, and misogyny. It’ s important to note that there was no overlap when the binary category was ’ no ’ . Thus, we defined only the five sexism categories as labels, and if a meme didn’ t fall into any of these categories, we assumed it to be non-sexist. Furthermore, this task was labelled by a group of experts who voted on the classification they considered optimal. Thus, we had not only the hard label but also the soft label given by those experts. Weaimed to use all this data to configure an accurate model. As stated previously, we had images and text for the meme data. This information could not be used as such, so we tried to obtain embeddings that would represent task-specific features. To do so, we fine-tuned a RoBERTa model [2] in the first stage, incorporating a final fully connected layer. Then, we retrained the model to learn the soft labels, using the Kullback-Leibler divergence as a loss function. In the second phase, we did the same for the image, using, in this case, a Google ViT [1] as the base model. The idea in those steps was to represent the memes in arrays that sum the features the experts use to discriminate them. Finally, in the last stage of the process, we concatenated both embeddings (thus enhancing the information achieved in image and text), and we applied a LightGBM to finally learn to classify into hard labels

Based on these three papers from the EXIST 2024 competition for sexism detection, here's a comprehensive summary:

Key Technical Approaches

Data Augmentation Techniques

- **Synonym Replacement**: Utilized WordNet for English tweet augmentation with high-quality results
- **Contextual Augmentation**: Employed Spanish Transformer models (BERTIN, ALBERT Base Spanish, RoBERTuito) from HuggingFace
- **Back-translation**: English→Spanish→English translation via Google Translator API, effectively doubling dataset size
- **AEDA Technique**: Random punctuation insertion (".", ";", "?", ":", "!", ",") for text augmentation

Data Preprocessing

- Removed usernames, URLs, percentages, timestamps, dates, hashtags, and emojis
- Converted all text to lowercase for uniformity and reduced complexity

Model Architectures

Paper 92 - Ensemble Learning Approach

- Utilized multiple base machine learning models (decision trees, logistic regression, support vector machines)
- Applied meta-learner to combine predictions for final output
- Implemented tenfold augmentation on original train and development datasets

- Addressed the under-researched area of ensemble ML models for sexism detection

Paper 94 - Multimodal Approaches

Task 4 (Binary Classification):

- **Stage 1**: Fine-tuned pre-trained CLIP model with fully connected layer for binary classification
- **Stage 2**: Used CLIP embeddings to train LightGBM classifier, leveraging strengths of both models

Task 6 (Multi-label Classification):

- **Three-stage methodology**: Two fine-tuning phases + LightGBM training
- Handled complex classification: Binary (sexist/non-sexist) + Five overlapping categories (stereotyping, objectification, ideological, sexual violence, misogyny)
- **Stage 1**: Fine-tuned RoBERTa for text processing using Kullback-Leibler divergence loss for soft labels
- **Stage 2**: Fine-tuned Google ViT for image processing
- **Stage 3**: Concatenated text and image embeddings, applied LightGBM for final hard label classification

Core Innovations

1. **Multimodal Fusion**: Combined textual and visual information for enhanced classification accuracy
2. **Soft Label Utilization**: Leveraged expert-annotated soft labels alongside hard labels
3. **Two-stage Training Strategy**: First extracted high-quality embeddings, then trained specialized classifiers
4. **Ensemble Learning Revival**: Demonstrated that traditional ML model ensembles remain valuable for sexism detection tasks
5. **Cross-lingual Augmentation**: Effective use of translation-based data augmentation for low-resource scenarios

Key Findings

- **Ensemble methods** proved effective as state-of-the-art solutions, particularly when combined with traditional ML models rather than just Transformer architectures
- **CLIP's multimodal capabilities** showed exceptional performance for meme classification tasks

- **Data augmentation at word-level** successfully maintained semantic meaning while increasing dataset diversity
- **Two-stage training approaches** effectively leveraged pre-trained models' representations for specialized downstream tasks

These methodologies demonstrate the effectiveness of combining traditional machine learning techniques with modern deep learning approaches, multimodal learning, and sophisticated data augmentation strategies for social media content analysis.

Retry