From Foundations to GPT in Text Classification: A Comprehensive Survey on Current Approaches and Future Trends

Other titles in Foundations and Trends® in Information Retrieval

Search as Learning
Kelsey Urgo and Jaime Arguello
ISBN: 978-1-63828-536-6

Understanding and Mitigating Gender Bias in Information Retrieval Systems
Shirin Seyedsalehi, Amin Bigdeli, Negar Arabzadeh, Batool AlMousawi, Zack Marshall, Morteza Zihayat and Ebrahim Bagheri
ISBN: 978-1-63828-518-2

Mathematical Information Retrieval: Search and Question Answering
Richard Zanibbi, Behrooz Mansouri and Anurag Agarwal
ISBN: 978-1-63828-502-1

Information Discovery in E-commerce
Zhaochun Ren, Xiangnan He, Dawei Yin and Maarten de Rijke
ISBN: 978-1-63828-462-8

Fairness in Search Systems
Yi Fang, Ashudeep Singh and Zhiqiang Tao
ISBN: 978-1-63828-498-7

User Simulation for Evaluating Information Access Systems
Krisztian Balog and ChengXiang Zhai
ISBN: 978-1-63828-378-2

From Foundations to GPT in Text Classification: A Comprehensive Survey on Current Approaches and Future Trends

Marco Siino
University of Catania
University of Palermo
marco.siino@unict.it

Ilenia Tinnirello
University of Palermo
ilenia.tinnirello@unipa.it

Marco La Cascia
University of Palermo
marco.lacascia@unipa.it

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Siino *et al.*. *From Foundations to GPT in Text Classification: A Comprehensive Survey on Current Approaches and Future Trends*. Foundations and Trends® in Information Retrieval, vol. 19, no. 5, pp. 557–711, 2025.

ISBN: 978-1-63828-559-5

Foundations and Trends® in Information Retrieval

Volume 19, Issue 5, 2025

Editorial Scope

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2025, Volume 19, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

From Foundations to GPT in Text Classification: A Comprehensive Survey on Current Approaches and Future Trends

Marco Siino[1,2], Ilenia Tinnirello[2] and Marco La Cascia[2]

[1] *University of Catania, Catania, Italy; marco.siino@unict.it*
[2] *University of Palermo, Palermo, Italy; ilenia.tinnirello@unipa.it, marco.lacascia@unipa.it*

ABSTRACT

Text classification stands as a cornerstone within the realm of Natural Language Processing (NLP), particularly when viewed through computer science and engineering. The past decade has seen deep learning revolutionize text classification, propelling advancements in text retrieval, categorization, information extraction, and summarization. The scholarly literature includes datasets, models, and evaluation criteria, with English being the predominant language of focus, despite studies involving Arabic, Chinese, Hindi, and others. The efficacy of text classification models relies heavily on their ability to capture intricate textual relationships and non-linear correlations, necessitating a comprehensive examination of the entire text classification pipeline.

In the NLP domain, a plethora of text representation techniques and model architectures have emerged, with Large

Marco Siino, Ilenia Tinnirello and Marco La Cascia (2025), "From Foundations to GPT in Text Classification: A Comprehensive Survey on Current Approaches and Future Trends", Foundations and Trends® in Information Retrieval: Vol. 19, No. 5, pp 557–711. DOI: 10.1561/1500000107.

Language Models (LLMs) and Generative Pre-trained Transformers (GPTs) at the forefront. These models are adept at transforming extensive textual data into meaningful vector representations encapsulating semantic information. The multidisciplinary nature of text classification, encompassing data mining, linguistics, and information retrieval, highlights the importance of collaborative research to advance the field. This work integrates traditional and contemporary text mining methodologies, fostering a holistic understanding of text classification.

This monograph provides an in-depth exploration of the text classification pipeline, with a particular emphasis on evaluating the impact of each component on the overall performance of text classification models. The pipeline includes state-of-the-art datasets, text preprocessing techniques, text representation methods, classification models, evaluation metrics, and future trends. Each section examines these stages, presenting technical innovations and recent findings. The work assesses various classification strategies, offering comparative analyses, examples and case studies. These contributions extend beyond a typical survey, providing a detailed and insightful exploration of the field.

1

Introduction

In several Natural Language Processing (NLP) applications like news categorization, sentiment analysis, and subject labelling, text classification is a crucial and relevant task (Garrido-Merchan *et al.*, 2023; Fields *et al.*, 2024b; Emanuel *et al.*, 2024). The goal is to tag or label textual components like sentences, questions, paragraphs, and documents. In this era of massive information dissemination, manually processing and categorizing huge amounts of text data takes a relevant amount of time and effort. Text information can be found on social media, websites, chat rooms, emails, questions and answers from customer service representatives, insurance claims and user reviews. Furthermore, human factors such as skills and fatigue can influence the effectiveness of text classification by hand. It is preferable to automate the text classification pipeline involving machine learning models to get objective outcomes. Furthermore, to reduce the problem of information overloading, the improvement of information retrieval effectiveness can help in finding the necessary information for a certain task. Figure 1.1 illustrates a flowchart of the steps involved in text classification in light of the traditional and most recent machine learning models. A critical first stage is the preprocessing of the text to be provided as input to the model.

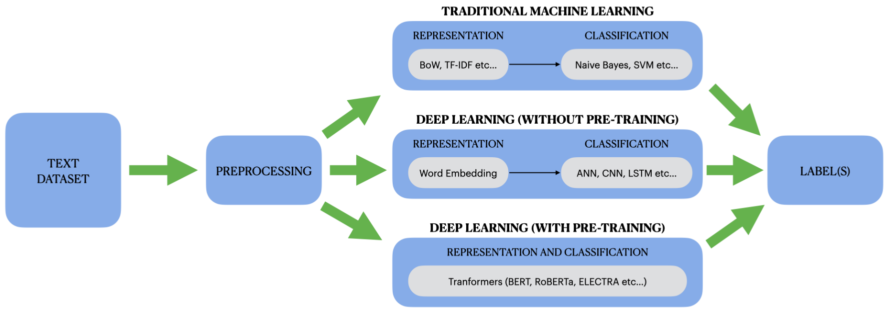

Figure 1.1: Overview of the text classification pipeline, illustrating the progression from text datasets to preprocessing, feature representations (e.g., Bag of Words, word embeddings), and final label predictions, encompassing traditional and modern approaches.

Classical approaches usually employ AI methods to collect relevant features, which are then classified using machine learning techniques. Next, the text representation approach can severely impact the outcomes, involving a series of transformations to map a source text to predicted labels. Deep learning, as opposed to traditional models, incorporates feature engineering into the training process. Up until 2010, classical text classification models were the most used and popular. Some of them are *logistic regressor*, *Naïve Bayes*, *Support Vector Machine* (SVM) and *K-Nearest Neighbour* (KNN). These methods can outperform past rule-based techniques in consistency and accuracy (Mitra *et al.*, 2007; Atmadja and Purwarianti, 2015). However, they still require feature engineering and are usually more time-consuming. Additionally, it is hard to understand the semantics of the words since they frequently neglect the context or natural sequential arrangement of textual material. In text classification, deep learning algorithms gradually replaced traditional techniques by the 2010s. Deep learning techniques for text mining automatically construct semantically pertinent representations without human intervention to define rules and features. Consequently, the majority of modern text classification activities are based on deep neural networks.

Most conventional machine learning models use a two-step procedure. First, the documents are stripped of manually added features (or any other textual unit). In the following, a classifier receives these features to provide a prediction. The Bag of Words (BoW) feature and extensions are frequently created by hand. Hidden Markov Models, Naive Bayes, SVM, Random Forests and Gradient Boosting, are common classification algorithms employed in the second step. Numerous disadvantages exist with this two-step approach. For instance, using handcrafted features and expecting acceptable performance requires time-consuming feature engineering and analysis. Due to the strategy's heavy reliance on domain expertise for feature generation, it is difficult to adapt it to new applications. Last, because of the specific features domain, these models cannot fully benefit from the vast volumes of training data available. To address the issues related to handcrafted features, the use of neural approaches has increased. The main component of these approaches is an embedding space, where text is encoded as a low-dimensional continuous feature vector without the need for traditional feature representation strategies. The *Latent Semantic Analysis* (LSA) proposed by Landauer and Dumais (1997) is one of the earliest studies on embedding models. The proposed architecture is trained on 200K words and has fewer than 1 million parameters.

In Bengio *et al.* (2000), the first neural language model was proposed. It consisted of an artificial neural network trained on over 10 million words. When progressively larger embedding models were constructed with significantly more training data, a paradigm change occurred. Several *Word2Vec* models that Google created in 2013 (Mikolov *et al.*, 2013b) were trained using billions of words and quickly gained popularity for numerous NLP applications. As the basis for their contextual embedding model, the researchers from Ai2[1] and the University of Washington created a Bidirectional-Long Short-Term Memory (BiLSTM) network using 93 million hyperparameters and a training performed on billions of words in 2017. A novel model named Embedding from Language Models (ELMo) (Peters *et al.*, 2018) captures contextual information and performs significantly better than Word2Vec. This subsequent development results in the construction of embedding models using Google's new neural architecture, the *Transformer* (Vaswani *et al.*,

[1]https://allenai.org/allennlp/software/elmo

2017). The Transformer architecture is based on attention modules, which boosts the effectiveness of extensive model training on the Tensor Processing Unit (TPU). In the same year, Google created the Bidirectional Encoder Representations from Transformers (BERT) (Devlin *et al.*, 2019). BERT has 340M parameters and was trained on 3.3 billion words. More training data and larger models are proposed in the literature every day. The most recent OpenAI GPT model has more than 170 billion parameters (Dale, 2021) and it is based on Transformers. Some academics contend that despite the enormous models' remarkable performance on different NLP tasks, they do not truly grasp language and are insufficient for many domains that are mission-critical (Jin *et al.*, 2020; Marcus and Davis, 2019). Recently, there has been a rise of interest toward neuro-symbolic hybrid models to solve significant flaws of neural models like interpretability, inability to use symbolic thinking and lack of grounding (Schlag *et al.*, 2019; Gao *et al.*, 2020).

Although there are many excellent reviews and textbooks on text classification techniques and applications, this work provides a thorough analysis of all the phases that go into creating a text classification pipeline with several contributions, including traditional and deep models to explore the impact on the performance of each stage of the pipeline. Even if specific languages are considered in the related works, from the standpoint of computer science, English is the language that is most frequently used and referred to in the present literature regarding text classification. Furthermore, most of the Large Language Models (LLMs) and pre-trained word embeddings are originally developed focusing on English, partially neglecting the other languages. Nowadays, modern LLMs are multilingual so they can be fed and can produce output also in other languages other than English (Rathje *et al.*, 2024). The rest of this work primarily uses English as the reference language for many of the examples and cases presented and discussed.

Starting with a discussion on some of the more contemporary tasks — such as author profiling, topic classification, news classification, and sentiment analysis — we then present classification models and the most recent and relevant findings. We also cover the most recent deep neural network architectures, which are divided into several types based on their functioning, including Transformers (LLMs and GPTs), Convolutional

Neural Networks (CNNs), Capsule Nets and Recurrent Neural Networks (RNNs).

This monograph is organized as follows: Section 2 presents the most common datasets used and available in the literature. In Section 3, the preprocessing techniques to prepare raw text are presented and discussed. In Section 4, the methods to represent text in a numerical way understandable by a computer are reported. In this section, we also show and analyse a word embedding space trained from scratch. In Section 5, traditional and modern classifiers commonly employed for text classification are discussed, including a discussion on modern LLMs and GPTs. In Section 6 generic and linguistic-specific metrics to evaluate the performance on text classification tasks are discussed. In Section 7, the conclusions and the future perspectives are presented. The contributions and a summary for each section of this work are reported in what follows.

1.1 Overview and Contributions

Several works have investigated text classification techniques from a general standpoint. We specifically mention the work by Li *et al.* (2020), which offers a thorough analysis of model architectures, from traditional to modern deep learning-based ones. The survey by Kowsari *et al.* (2019) offers a great examination of preprocessing procedures, including feature extraction and dimensionality reduction. Despite including quantitative outcomes of conventional approaches, Minaee *et al.* (2021) mainly focuses on deep learning models. By providing a view of each stage required to design a text classification model, this monograph seeks to enhance the landscape of text classification from a general point of view. As a result, we give a thorough explanation of the key data preparation procedures used along with classification models. We provide model descriptions from traditional to deep learning-based ones, in contrast to prior surveys. The design of the classifier and feature extraction are highlighted for the traditional models. A specific overview of each section of this work is reported to conclude this section.

Overview of Section 2: Tasks and Datasets

In the early history of machine learning, information retrieval systems primarily used text classification algorithms. But as technology has developed over time, text classification and document categorization have become widely employed in several fields, including law, engineering, social sciences, healthcare, psychology, and medicine. We highlight some domains that use text classification algorithms in this section. Some text classification tasks are discussed in this section, including three new datasets related to emerging author profiling tasks. The datasets available in the literature and related to these tasks and usually employed as benchmarks, are also reported and presented in this section.

Overview of Section 3: Preprocessing

In this section, we collect, report and discuss the text preprocessing techniques found in the literature and their possible and most recent variants, proposing a standard nomenclature based on acronyms. We also provide the reader with useful information for self-study of the techniques presented along with advice on how to operate educated choices to select the preprocessing technique (or combination of techniques) given a specific task, model, and dataset. According to recent related works, we also discuss if simple classifiers' performance is comparable to the ones obtained by Transformer-based models when text preprocessing is performed according to the specific model and dataset used.

Overview of Section 4: Representation

Before moving to the classification stage, it is necessary to convert unstructured data, especially free-running text data, into organized numerical data. To do this, a document representation model must be used to employ a subsequent classification system following the text preprocessing stage. Text representation models convert text data into a numerical vector space, which has a substantial impact on how well subsequent learning tasks can perform. In the history of NLP, word representation has always been a topic of interest. It is crucial to properly represent such text data since it contains a wealth of information and may be applied broadly across a variety of applications. This section examines the expressive potential of several word representation models, ranging from the traditional to the contemporary word representation approaches provided by LLMs. The section discusses numerous representation methods that are frequently employed in the literature. Before discussing well-known representation learning and pre-trained language models, we first discuss various statistical models. Then we move to attention-based representation and, in the last subsection, to a case study about the analysis of a trained word embedding for a specific text classification task. Thanks to a Principal Component Analysis (PCA) tool, it shows and discusses the effect of CNN training on a 3D visualization of a word embedding space. In this way, we can motivate some implicit choices operated during the training of a deep learning model to assign specific word vectors to certain keywords belonging to one of the two class labels used for the discussed task.

Overview of Section 5: Classification

In Section 5, both the traditional classification models for text classification and the most modern ones based on deep learning are reported. The models discussed in this section belong to three different groups. The non-deep learning deterministic models, the foundational deep learning models and the large pre-trained language models known as Transformers. The term "earlier approaches" refers to all techniques used before the advent of deep neural networks, when the prediction was based on manually created features. Neural networks with only a few hidden layers are also included in this category, and these are so-called "shallow" networks. These methods replace several rule-based ones, which they usually outperform in terms of accuracy. The most recent deep learning models, which have an impact on all artificial intelligence domains, including text classification, are also discussed. These techniques have become popular because they can simulate intricate features without requiring manual engineering, which reduces the need for subject expertise. Finally, we discuss Transformers (LLMs and GPTs) and the recent and emerging discipline of *Prompt Engineering*. We discuss several prompting techniques, and then we move to some ethical considerations on the use of generative AI.

Overview of Section 6: Evaluation

This section focuses on how to evaluate the performance of deep learning models in the context of text classification tasks, introducing the most used metrics in the literature. We discuss various metrics such as accuracy, precision, recall, and F1 score, emphasizing the importance of selecting the right metric based on the specific goals. In addition, we explore the limitations of traditional evaluation metrics and highlight the necessity for more sophisticated approaches, particularly in scenarios involving imbalanced datasets. The use of confusion matrices and *ROC-AUC* scores were recommended to provide a more comprehensive evaluation of model performance, along with metrics as *ROUGE* and *BLEU* for tasks involving text generation and summarization. Moreover, we propose the integration of human evaluation methods to supplement quantitative metrics, recognizing that the nuances of language often elude numerical representation.

Overview of Section 7: Conclusion

In the final section of this work, we report the final conclusions and future perspectives on the matter.

References

Agarwal, A., B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. (2011). "Sentiment analysis of twitter data". In: *Proceedings of the workshop on language in social media (LSM 2011)*. 30–38.

Ahmed, T. and P. T. Devanbu. (2023). "Better Patching Using LLM Prompting, via Self-Consistency". In: *38th IEEE/ACM International Conference on Automated Software Engineering, ASE 2023, Luxembourg, September 11-15, 2023*. IEEE. 1742–1746. DOI: 10.1109/ASE56229.2023.00065.

Akın, A. A. and M. D. Akın. (2007). "Zemberek, an open source NLP framework for Turkic languages". *Structure*. 10: 1–5.

Alam, S. and N. Yao. (2019). "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis". *Computational and Mathematical Organization Theory*. 25(3): 319–335.

Albalawi, Y., J. Buckley, and N. S. Nikolov. (2021). "Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting Arabic health information on social media". *Journal of big Data*. 8(1): 1–29.

Aliakbarzadeh, A., L. Flek, and A. Karimi. (2025). "Exploring Robustness of Multilingual LLMs on Real-World Noisy Data". *arXiv preprint arXiv:2501.08322*.

Aljebreen, A., W. Meng, and E. Dragut. (2021). "Segmentation of tweets with urls and its applications to sentiment analysis". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 12480–12488.

Alzahrani, E. and L. Jololian. (2021). "How Different Text-preprocessing Techniques Using The BERT Model Affect The Gender Profiling of Authors". *arXiv preprint arXiv:2109.13890*.

Anandarajan, M., C. Hill, and T. Nolan. (2019). "Text preprocessing". In: *Practical Text Analytics*. Springer. 45–59.

Angiani, G., L. Ferrari, T. Fontanini, P. Fornacciari, E. Iotti, F. Magliani, and S. Manicardi. (2016). "A comparison between preprocessing techniques for sentiment analysis in Twitter". In: *2nd International Workshop on Knowledge Discovery on the WEB (KDWEB)*. Vol. 1748.

Araslanov, E., E. Komotskiy, and E. Agbozo. (2020). "Assessing the Impact of Text Preprocessing in Sentiment Analysis of Short Social Network Messages in the Russian Language". In: *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*. IEEE. 1–4.

Arief, M. and M. B. M. Deris. (2021). "Text Preprocessing Impact for Sentiment Classification in Product Review". In: *2021 Sixth International Conference on Informatics and Computing (ICIC)*. IEEE. 1–7.

Atmadja, A. R. and A. Purwarianti. (2015). "Comparison on the rule based method and statistical based method on emotion classification for Indonesian Twitter text". In: *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*. 1–6. DOI: 10.1109/ICITSI.2015.7437692.

Babanejad, N., A. Agrawal, A. An, and M. Papagelis. (2020). "A Comprehensive Analysis of Preprocessing for Word Representation Learning in Affective Tasks". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics. 5799–5810. DOI: 10.18653/v1/2020.acl-main.514.

Bahdanau, D., K. Cho, and Y. Bengio. (2015). "Neural Machine Translation by Jointly Learning to Align and Translate". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Y. Bengio and Y. LeCun.

Bakliwal, A., P. Arora, S. Madhappan, N. Kapre, M. Singh, and V. Varma. (2012). "Mining sentiments from tweets". In: *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*. 11–18.

Balahur, A. (2013). "Sentiment analysis in social media texts". In: *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*. 120–128.

Bansal, H., G. Shrivastava, G. N. Nguyen, and L.-M. Stanciu. (2018). *Social network analytics for contemporary business organizations*. IGI Global. DOI: 10.4018/978-1-5225-5097-6.

Bao, Y., C. Quan, L. Wang, and F. Ren. (2014). "The role of preprocessing in twitter sentiment analysis". In: *International conference on intelligent computing*. Springer. 615–624.

Barbosa, L. and J. Feng. (2010). "Robust Sentiment Detection on Twitter from Biased and Noisy Data". In: *COLING 2010, 23rd International Conference on Computational Linguistics, Posters Volume, 23-27 August 2010, Beijing, China*. Ed. by C. Huang and D. Jurafsky. Chinese Information Processing Society of China. 36–44.

Bengio, Y., R. Ducharme, and P. Vincent. (2000). "A Neural Probabilistic Language Model". In: *Advances in Neural Information Processing Systems*. Ed. by T. Leen, T. Dietterich, and V. Tresp. Vol. 13. MIT Press.

Benzarti, S. and R. Faiz. (2015). "EgoTR: Personalized tweets recommendation approach". In: *Intelligent Systems in Cybernetics and Automation Theory: Proceedings of the 4th Computer Science Online Conference 2015 (CSOC2015), Vol 2: Intelligent Systems in Cybernetics and Automation Theory*. Springer. 227–238.

Bevendorff, J., B. Chulvi, E. Fersini, A. Heini, M. Kestemont, K. Kredens, M. Mayerl, R. Ortega-Bueno, P. Pkezik, M. Potthast, *et al.* (2022). "Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection". In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer. 382–394.

Boiy, E., P. Hens, K. Deschacht, and M. Moens. (2007). "Automatic Sentiment Analysis in On-line Text". In: *Openness in Digital Publishing: Awareness, Discovery and Access - Proceedings of the 11th International Conference on Electronic Publishing held in Vienna - ELPUB 2007, Vienna, Austria, June 13-15, 2007. Proceedings.* 349–360.

Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov. (2017). "Enriching Word Vectors with Subword Information". *Transactions of the Association for Computational Linguistics*. 5: 135–146.

Borra, E. and B. Rieder. (2014). "Programmed method: Developing a toolset for capturing and analyzing tweets". *Aslib journal of information management*. 66(3): 262–278.

Bowman, S. R., G. Angeli, C. Potts, and C. D. Manning. (2015). "A large annotated corpus for learning natural language inference". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 632–642.

Breiman, L. (1996). "Bagging predictors". *Machine learning*. 24: 123–140.

Bueno, R. O., B. Chulvi, F. Rangel, P. Rosso, and E. Fersini. (2022). "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO). Overview for PAN at CLEF 2022". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*. Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. Vol. 3180. *CEUR Workshop Proceedings*. CEUR-WS.org. 2314–2343.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu. (1995). "A limited memory algorithm for bound constrained optimization". *SIAM Journal on scientific computing*. 16(5): 1190–1208.

Camacho-Collados, J. and M. T. Pilehvar. (2018). “On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis”. In: *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 40–46.

Can, F., S. Kocberber, E. Balcik, C. Kaynak, H. C. Ocalan, and O. M. Vursavas. (2008). “Information retrieval on Turkish texts”. *Journal of the American Society for Information Science and Technology*. 59(3): 407–421.

Chang, C.-C. and C.-J. Lin. (2011). “LIBSVM: a library for support vector machines”. *ACM transactions on intelligent systems and technology (TIST)*. 2(3): 1–27.

Chen, G., S. Ma, Y. Chen, L. Dong, D. Zhang, J. Pan, W. Wang, and F. Wei. (2021). “Zero-Shot Cross-Lingual Transfer of Neural Machine Translation with Multilingual Pretrained Encoders”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 15–26.

Chen, G., D. Ye, Z. Xing, J. Chen, and E. Cambria. (2017). “Ensemble application of convolutional and recurrent neural networks for multi-label text categorization”. In: *2017 International joint conference on neural networks (IJCNN)*. IEEE. 2377–2383.

Cheng, J., L. Dong, and M. Lapata. (2016). “Long Short-Term Memory-Networks for Machine Reading”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*. Ed. by J. Su, X. Carreras, and K. Duh. The Association for Computational Linguistics. 551–561. DOI: 10.18653/v1/d16-1053.

Chicco, D. (2021). “Siamese Neural Networks: An Overview”. In: *Artificial Neural Networks - Third Edition*. Ed. by H. M. Cartwright. Vol. 2190. *Methods in Molecular Biology*. Springer. 73–94. DOI: 10.1007/978-1-0716-0826-5_3.

Clark, K., M. Luong, Q. V. Le, and C. D. Manning. (2020). “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators”. In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Colas, F. and P. Brazdil. (2006). “Comparison of SVM and some older classification algorithms in text classification tasks”. In: *IFIP International Conference on Artificial Intelligence in Theory and Practice*. Springer. 169–178.

Cortes, C. and V. Vapnik. (1995). “Support-vector networks”. *Machine learning*. 20(3): 273–297.

Courseault Trumbach, C. and D. Payne. (2007). “Identifying synonymous concepts in preparation for technology mining”. *Journal of Information Science*. 33(6): 660–677.

Cover, T. and P. Hart. (1967). “Nearest neighbor pattern classification”. *IEEE transactions on information theory*. 13(1): 21–27.

Cover, T. M. and J. A. Thomas. (2001). *Elements of Information Theory*. Wiley. DOI: 10.1002/0471200611.

Croce, D., D. Garlisi, and M. Siino. (2022). “An SVM Ensembler Approach to Detect Irony and Stereotype Spreaders on Twitter”. In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum (CLEF)* (Bologna, Italy, Sept. 5–8, 2022). Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. *CEUR Workshop Proceedings*. No. 3180. Aachen. 2426–2432.

Cunha, W., S. Canuto, F. Viegas, T. Salles, C. Gomes, V. Mangaravite, E. Resende, T. Rosa, M. A. Gonçalves, and L. Rocha. (2020). “Extended pre-processing pipeline for text classification: On the role of meta-feature representations, sparsification and selective sampling”. *Information Processing & Management*. 57(4): 102263. DOI: https://doi.org/10.1016/j.ipm.2020.102263.

Cunha, W., V. Mangaravite, C. Gomes, S. Canuto, E. Resende, C. Nascimento, F. Viegas, C. França, W. S. Martins, J. M. Almeida, T. Rosa, L. Rocha, and M. A. Gonçalves. (2021). “On the cost-effectiveness of neural and non-neural approaches and representations for text classification: A comprehensive comparative study”. *Information Processing & Management*. 58(3): 102481. DOI: https://doi.org/10.1016/j.ipm.2020.102481.

Dai, Z., Z. Yang, Y. Yang, J. G. Carbonell, Q. Le, and R. Salakhutdinov. (2019). “Transformer-XL: Attentive Language Models beyond a Fixed-Length Context”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2978–2988.

Dale, R. (2021). “GPT-3: What’s it good for?” *Natural Language Engineering*. 27(1): 113–118.

Das, A. S., M. Datar, A. Garg, and S. Rajaram. (2007). “Google news personalization: scalable online collaborative filtering”. In: *Proceedings of the 16th international conference on World Wide Web*. 271–280.

Deng, L. and J. Wiebe. (2015). “Mpqa 3.0: An entity/event-level sentiment corpus”. In: *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies*. 1323–1328.

Denny, M. J. and A. Spirling. (2018). “Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it”. *Political Analysis*. 26(2): 168–189.

Devlin, J., M. Chang, K. Lee, and K. Toutanova. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Ed. by J. Burstein, C. Doran, and T. Solorio. Association for Computational Linguistics. 4171–4186. DOI: 10.18653/v1/n19-1423.

Dieng, A. B., C. Wang, J. Gao, and J. W. Paisley. (2017). “TopicRNN: A Recurrent Neural Network with Long-Range Semantic Dependency”. In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Ding, X., K. Liao, T. Liu, Z. Li, and J. Duan. (2019). “Event representation learning enhanced with external commonsense knowledge”. *arXiv preprint arXiv:1909.05190*.

Djuric, N., J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati. (2015). “Hate speech detection with comment embeddings”. In: *Proceedings of the 24th international conference on world wide web*. 29–30.

Dolamic, L. and J. Savoy. (2010). “When stopword lists make the difference”. *J. Assoc. Inf. Sci. Technol.* 61(1): 200–203. DOI: 10.1002/asi.21186.

Dolan, W. B. and C. Brockett. (2005). "Automatically constructing a corpus for paraphrase recognition". In: *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*.

Dong, Q., L. Li, D. Dai, C. Zheng, J. Ma, R. Li, H. Xia, J. Xu, Z. Wu, B. Chang, X. Sun, and Z. Sui. (2024). "A Survey on In-context Learning". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*. Ed. by Y. Al-Onaizan, M. Bansal, and Y. Chen. Association for Computational Linguistics. 1107–1128.

Duong, H.-T. and T.-A. Nguyen-Thi. (2021). "A review: preprocessing techniques and data augmentation for sentiment analysis". *Computational Social Networks*. 8(1): 1–16.

Edwards, A. and J. Camacho-Collados. (2024). "Language Models for Text Classification: Is In-Context Learning Enough?" In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Ed. by N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue. Torino, Italia: ELRA and ICCL. 10058–10072.

Emanuel, R. H. K., P. D. Docherty, H. Lunt, and K. Möller. (2024). "The effect of activation functions on accuracy, convergence speed, and misclassification confidence in CNN text classification: a comprehensive exploration". *J. Supercomput.* 80(1): 292–312. DOI: 10.1007/S11227-023-05441-7.

Fields, J., K. Chovanec, and P. Madiraju. (2024a). "A Survey of Text Classification With Transformers: How Wide? How Large? How Long? How Accurate? How Expensive? How Safe?" *IEEE Access*. 12: 6518–6531. DOI: 10.1109/ACCESS.2024.3349952.

Fields, J., K. Chovanec, and P. Madiraju. (2024b). "A survey of text classification with transformers: How wide? how large? how long? how accurate? how expensive? how safe?" *IEEE Access*. 12: 6518–6531.

Flood, B. J. (1999). "Historical Note: The Start of a Stop List at Biological Abstracts". *J. Am. Soc. Inf. Sci.* 50(12): 1066. DOI: 10.1002/(SICI)1097-4571(1999)50:12\<1066::AID-ASI5\>3.0.CO;2-A.

Gao, J., B. Peng, C. Li, J. Li, S. Shayandeh, L. Liden, and H.-Y. Shum. (2020). "Robust Conversational AI with Grounded Text Generation". *arXiv e-prints*: arXiv–2009.

Garg, N. and K. Sharma. (2022). "Text pre-processing of multilingual for sentiment analysis based on social network data." *International Journal of Electrical & Computer Engineering (2088-8708)*. 12(1).

Garrido-Merchan, E. C., R. Gozalo-Brizuela, and S. Gonzalez-Carvajal. (2023). "Comparing BERT against traditional machine learning models in text classification". *Journal of Computational and Cognitive Engineering*. 2(4): 352–356.

Gemci, F. and K. A. Peker. (2013). "Extracting Turkish tweet topics using LDA". In: *2013 8th International Conference on Electrical and Electronics Engineering (ELECO)*. IEEE. 531–534.

Genkin, A., D. D. Lewis, and D. Madigan. (2007). "Large-Scale Bayesian Logistic Regression for Text Categorization". *Technometrics*. 49(3): 291–304. DOI: 10.1198/004017007000000245.

Gerlach, M., H. Shi, and L. A. N. Amaral. (2019). "A universal information theoretic approach to the identification of stopwords". *Nature Machine Intelligence*. 1(12): 606–612.

González, J. Á., L.-F. Hurtado, and F. Pla. (2020). "Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter". *Information Processing & Management*. 57(4): 102262. DOI: https://doi.org/10.1016/j.ipm.2020.102262.

Granik, M. and V. Mesyura. (2017). "Fake news detection using naive Bayes classifier". In: *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. IEEE. 900–903.

Gupta, V. and G. S. Lehal. (2011). "Punjabi language stemmer for nouns and proper names". In: *Proceedings of the 2nd Workshop on South Southeast Asian Natural Language Processing (WSSANLP)*. 35–39.

Guzman, E. and W. Maalej. (2014). "How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews". In: *2014 IEEE 22nd International Requirements Engineering Conference (RE)*. 153–162. DOI: 10.1109/RE.2014.6912257.

HaCohen-Kerner, Y., D. Miller, and Y. Yigal. (2020). "The influence of preprocessing on text classification using a bag-of-words representation". *PloS one*. 15(5): e0232525.

Haddi, E., X. Liu, and Y. Shi. (2013). "The Role of Text Pre-processing in Sentiment Analysis". In: *Proceedings of the First International Conference on Information Technology and Quantitative Management, ITQM 2013, Dushu Lake Hotel, Sushou, China, 16-18 May, 2013*. Ed. by Y. Shi, Y. Xi, P. Wolcott, Y. Tian, J. Li, D. Berg, Z. Chen, E. Herrera-Viedma, G. Kou, H. Lee, Y. Peng, and L. Yu. Vol. 17. *Procedia Computer Science*. Elsevier. 26–32. DOI: 10.1016/j.procs.2013.05.005.

Hair Zaki, U. H., R. Ibrahim, S. Abd Halim, and I. I. Kamsani. (2022). "Text Detergent: The Systematic Combination of Text Preprocessing Techniques for Social Media Sentiment Analysis". In: *International Conference of Reliable Information and Communication Technology*. Springer. 50–61.

Hassler, M. and G. Fliedl. (2006). "Text preparation through extended tokenization". *WIT Transactions on Information and Communication Technologies*. 37.

Hernández Farías, D. I., R. M. Ortega-Mendoza, and M. Montes-y-Gómez. (2019). "Exploring the use of psycholinguistic information in author profiling". In: *Mexican Conference on Pattern Recognition*. Springer. 411–421.

Hickman, L., S. Thapa, L. Tay, M. Cao, and P. Srinivasan. (2022). "Text preprocessing for text mining in organizational research: Review and recommendations". *Organizational Research Methods*. 25(1): 114–146.

Hinton, G. E., A. Krizhevsky, and S. D. Wang. (2011). "Transforming auto-encoders". In: *International conference on artificial neural networks*. Springer. 44–51.

Hinton, G. E., S. Sabour, and N. Frosst. (2018). "Matrix capsules with EM routing". In: *International conference on learning representations*.

Ho, T. K. (1998). "The random subspace method for constructing decision forests". *IEEE transactions on pattern analysis and machine intelligence*. 20(8): 832–844.

Hogenboom, A., D. Bal, F. Frasincar, M. Bal, F. De Jong, and U. Kaymak. (2013). "Exploiting emoticons in sentiment analysis". In: *Proceedings of the 28th annual ACM symposium on applied computing*. 703–710.

Indra, S., L. Wikarsa, and R. Turang. (2016). "Using logistic regression method to classify tweets into the selected topics". In: *2016 international conference on advanced computer science and information systems (icacsis)*. IEEE. 385–390.

Islam, R. and I. Ahmed. (2024). "Gemini-the most powerful LLM: Myth or Truth". In: *2024 5th Information Communication Technologies Conference (ICTC)*. IEEE. 303–308.

Iyyer, M., V. Manjunatha, J. Boyd-Graber, and H. Daumé III. (2015). "Deep unordered composition rivals syntactic methods for text classification". In: *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*. 1681–1691.

Jiang, A. Q., A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, *et al.* (2023). "Mistral 7B". *arXiv preprint arXiv:2310.06825*.

Jiang, S., G. Pang, M. Wu, and L. Kuang. (2012). "An improved K-nearest-neighbor algorithm for text categorization". *Expert Systems with Applications*. 39(1): 1503–1509.

Jianqiang, Z. and G. Xiaolin. (2017). "Comparison research on text pre-processing methods on twitter sentiment analysis". *IEEE Access*. 5: 2870–2879.

Jin, D., Z. Jin, J. T. Zhou, and P. Szolovits. (2020). "Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press. 8018–8025.

Joachims, T. (1998). "Text categorization with support vector machines: Learning with many relevant features". In: *European conference on machine learning*. Springer. 137–142.

Joachims, T. (1999). "Transductive Inference for Text Classification using Support Vector Machines". In: *Proceedings of the Sixteenth International Conference on Machine Learning (ICML 1999), Bled, Slovenia, June 27 - 30, 1999*. Ed. by I. Bratko and S. Dzeroski. Morgan Kaufmann. 200–209.

Joachims, T. (2002). "A statistical learning model of text classification for SVMs". In: *Learning to Classify Text Using Support Vector Machines*. Springer. 45–74.

Johnson, D. E., F. J. Oles, T. Zhang, and T. Goetz. (2002). "A decision-tree-based symbolic rule induction system for text categorization". *IBM Systems Journal*. 41(3): 428–437.

Johnson, R. and T. Zhang. (2016). "Supervised and semi-supervised text categorization using LSTM for region embeddings". In: *International Conference on Machine Learning*. PMLR. 526–534.

Joulin, A., E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov. (2016). "Fasttext. zip: Compressing text classification models". *arXiv preprint arXiv:1612.03651*.

Joyce, J. M. (2011). "Kullback-leibler divergence". In: *International encyclopedia of statistical science*. Springer. 720–722.

Jurafsky, D. and J. H. Martin. (2009). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition. Prentice Hall series in artificial intelligence*. Prentice Hall, Pearson Education International.

Kadhim, A. I. (2018). "An Evaluation of Preprocessing Techniques for Text Classification". *International Journal of Computer Science and Information Security (IJCSIS)*. 16(6).

Kathuria, A., A. Gupta, and R. Singla. (2021). "A Review of Tools and Techniques for Preprocessing of Textual Data". *Computational Methods and Data Engineering*: 407–422.

Kenny, E. M., C. Ford, M. Quinn, and M. T. Keane. (2021). "Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies". *Artificial Intelligence*. 294: 103459.

Ketsbaia, L., B. Issac, and X. Chen. (2020). "Detection of hate tweets using machine learning and deep learning". In: *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE. 751–758.

Kim, Y. (2014). "Convolutional Neural Networks for Sentence Classification". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics. 1746–1751. DOI: 10.3115/v1/D14-1181.

Kim, Y., Y. Jernite, D. Sontag, and A. M. Rush. (2016). "Character-aware neural language models". In: *Thirtieth AAAI conference on artificial intelligence*.

Kipf, T. N. and M. Welling. (2017). "Semi-Supervised Classification with Graph Convolutional Networks". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Kong, A., S. Zhao, H. Chen, Q. Li, Y. Qin, R. Sun, X. Zhou, E. Wang, and X. Dong. (2024). "Better Zero-Shot Reasoning with Role-Play Prompting". In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*. Ed. by K. Duh, H. Gómez-Adorno, and S. Bethard. Association for Computational Linguistics. 4099–4113. DOI: 10.18653/V1/2024.NAACL-LONG.228.

Koopman, C. and A. Wilhelm. (2020). "The effect of preprocessing on short document clustering". *Archives of Data Science, Series A*. 6(1): 01.

Kouloumpis, E., T. Wilson, and J. Moore. (2011). "Twitter sentiment analysis: The good the bad and the omg!" In: *Proceedings of the international AAAI conference on web and social media*. Vol. 5. 538–541.

Kowsari, K., D. E. Brown, M. Heidarysafa, K. J. Meimandi, M. S. Gerber, and L. E. Barnes. (2017). "Hdltex: Hierarchical deep learning for text classification". In: *2017 16th IEEE international conference on machine learning and applications (ICMLA)*. IEEE. 364–371.

Kowsari, K., K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown. (2019). "Text classification algorithms: A survey". *Information*. 10(4): 150.

Kudo, T. (2018). "Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 66–75.

Kumar, P. and L. Dhinesh Babu. (2019). "Novel text preprocessing framework for sentiment analysis". In: *Smart intelligent computing and applications*. Springer. 309–317.

Kunilovskaya, M. and A. Plum. (2021). "Text Preprocessing and its Implications in a Digital Humanities Project". In: *Proceedings of the Student Research Workshop Associated with RANLP 2021*. 85–93.

Kurniasih, A. and L. P. Manik. (2022). "On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts". *International Journal of Advanced Computer Science and Applications*. 13(6): 927–934. DOI: 10.14569/IJACSA.2022.01306109.

Kuznetsov, I. and I. Gurevych. (2018). "From text to lexicon: Bridging the gap between word embeddings and lexical resources". In: *Proceedings of the 27th International Conference on Computational Linguistics*. 233–244.

Lample, G., M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer. (2016). "Neural architectures for named entity recognition". In: *Proceedings of NAACL-HLT*.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2020). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Landauer, T. K. and S. T. Dumais. (1997). "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." *Psychological review*. 104(2): 211.

Le, Q. and T. Mikolov. (2014). "Distributed representations of sentences and documents". In: *International conference on machine learning*. PMLR. 1188–1196.

Lee, K., L. He, M. Lewis, and L. Zettlemoyer. (2017). "End-to-end neural coreference resolution". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 188–197.

Lehmann, J., R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, *et al.* (2015). "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia". *Semantic web*. 6(2): 167–195.

Leopold, E. and J. Kindermann. (2002). "Text categorization with support vector machines. How to represent texts in input space?" *Machine Learning*. 46(1): 423–444.

Leslie, C., E. Eskin, and W. S. Noble. (2001). "The spectrum kernel: A string kernel for SVM protein classification". In: *Biocomputing 2002*. World Scientific. 564–575.

Lewis, P. S. H., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin.

Li, Q., H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. S. Yu, and L. He. (2020). "A survey on text classification: From shallow to deep learning". *arXiv preprint arXiv:2008.00364*.

Li, X. and Y. Guo. (2013). "Active Learning with Multi-Label SVM Classification". In: *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*. Ed. by F. Rossi. IJCAI/AAAI. 1479–1485.

Li, X. and W. Lam. (2017). "Deep multi-task learning for aspect term extraction with memory interaction". In: *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2886–2892.

Lin, C. and Y. He. (2009). “Joint sentiment/topic model for sentiment analysis”. In: *Proceedings of the 18th ACM conference on Information and knowledge management*. 375–384.

Lison, P. and A. Kutuzov. (2017). “Redefining Context Windows for Word Embedding Models: An Experimental Study”. In: *Proceedings of the 21st Nordic Conference on Computational Linguistics*. 284–288.

Liu, H., Z. Zhao, J. Wang, H. Kamarthi, and B. A. Prakash. (2024a). “LSTPrompt: Large Language Models as Zero-Shot Time Series Forecasters by Long-Short-Term Prompting”. In: *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*. Ed. by L. Ku, A. Martins, and V. Srikumar. Association for Computational Linguistics. 7832–7840. DOI: 10.18653/V1/2024.FINDINGS-ACL.466.

Liu, J., W.-C. Chang, Y. Wu, and Y. Yang. (2017). “Deep learning for extreme multi-label text classification”. In: *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*. 115–124.

Liu, N. F., K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang. (2024b). “Lost in the middle: How language models use long contexts”. *Transactions of the Association for Computational Linguistics*. 12: 157–173. DOI: 10.1162/TACL_A_00638.

Liu, P., X. Qiu, X. Chen, S. Wu, and X.-J. Huang. (2015). “Multi-timescale long short-term memory neural network for modelling sentences and documents”. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2326–2335.

Liu, P., X. Qiu, and X. Huang. (2016). “Recurrent Neural Network for Text Classification with Multi-Task Learning”. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*. Ed. by S. Kambhampati. IJCAI/AAAI Press. 2873–2879.

Liu, X., Y. Shen, K. Duh, and J. Gao. (2018). "Stochastic Answer Networks for Machine Reading Comprehension". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*. Ed. by I. Gurevych and Y. Miyao. Association for Computational Linguistics. 1694–1704. DOI: 10.18653/v1/P18-1157.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. (2019). "Roberta: A robustly optimized bert pretraining approach". *arXiv preprint arXiv:1907.11692*.

Liu, Z., X. Lv, K. Liu, and S. Shi. (2010). "Study on SVM compared with the other text classification methods". In: *2010 Second international workshop on education technology and computer science*. Vol. 1. IEEE. 219–222.

Lo, R. T.-W., B. He, and I. Ounis. (2005). "Automatically building a stopword list for an information retrieval system". In: *Journal on Digital Information Management: Special Issue on the 5th Dutch-Belgian Information Retrieval Workshop (DIR)*. Vol. 5. 17–24.

Lombardo, A., G. Morabito, S. Quattropani, C. Ricci, M. Siino, and I. Tinnirello. (2024). "AI-GeneSI: Exploiting generative AI for autonomous generation of the southbound interface in the IoT". In: *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*. 1–7. DOI: 10.1109/WF-IoT62078.2024.10811300.

Lomonaco, F., G. Donabauer, and M. Siino. (2022). "COURAGE at CheckThat! 2022: Harmful Tweet Detection using Graph Neural Networks and ELECTRA". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum (CLEF)* (Bologna, Italy, Sept. 5–8, 2022). Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. *CEUR Workshop Proceedings*. No. 3180. Aachen. 573–583.

Lovins, J. B. (1968). "Development of a stemming algorithm." *Mechanical Translation and Computational Linguistics*. 11(1-2): 22–31.

Loza Mencía, E. and J. Fürnkranz. (2008). "Efficient pairwise multilabel classification for large-scale problems in the legal domain". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 50–65.

Luhn, H. P. (1960). "Key word-in-context index for technical literature (kwic index)". *American documentation*. 11(4): 288–295.

Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts. (2011). "Learning Word Vectors for Sentiment Analysis". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics. 142–150.

Magerman, D. M. (1995). "Statistical Decision-Tree Models for Parsing". In: *33rd Annual Meeting of the Association for Computational Linguistics*. 276–283.

Makrehchi, M. and M. S. Kamel. (2008). "Automatic extraction of domain-specific stopwords from labeled documents". In: *Advances in Information Retrieval: 30th European Conference on IR Research, ECIR 2008, Glasgow, UK, March 30-April 3, 2008. Proceedings 30*. Springer. 222–233.

Mangione, S., M. Siino, and G. Garbo. (2022). "Improving Irony and Stereotype Spreaders Detection using Data Augmentation and Convolutional Neural Network". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum (CLEF)* (Bologna, Italy, Sept. 5–8, 2022). Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. *CEUR Workshop Proceedings*. No. 3180. Aachen. 2585–2593.

Manning, C. D., P. Raghavan, and H. Schütze. (2008). *Introduction to information retrieval*. Cambridge University Press. DOI: 10.1017/CBO9780511809071.

Manning, C. D., H. Schütze, and G. Weikurn. (2002). "Foundations of Statistical Natural Language Processing". *SIGMOD Record*. 31(3): 37–38.

Marcus, G. and E. Davis. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Matthews, B. W. (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". *Biochimica et Biophysica Acta (BBA)-Protein Structure*. 405(2): 442–451.

McCallum, A. and K. Nigam. (1998). "A comparison of event models for naive bayes text classification". In: *AAAI-98 workshop on learning for text categorization*. Vol. 752. Citeseer. 41–48.

McCann, B., J. Bradbury, C. Xiong, and R. Socher. (2017). "Learned in Translation: Contextualized Word Vectors". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett. 6294–6305.

McNamee, P. and J. Mayfield. (2004). "Character n-gram tokenization for European language text retrieval". *Information retrieval*. 7(1): 73–97.

Melamud, O., J. Goldberger, and I. Dagan. (2016). "context2vec: Learning Generic Context Embedding with Bidirectional LSTM". In: *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, Germany: Association for Computational Linguistics. 51–61. DOI: 10.18653/v1/K16-1006.

Miculicich, L., D. Ram, N. Pappas, and J. Henderson. (2018). "Document-Level Neural Machine Translation with Hierarchical Attention Networks". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii. Association for Computational Linguistics. 2947–2954. DOI: 10.18653/v1/d18-1325.

Mihalcea, R. and P. Tarau. (2004). "Textrank: Bringing order into text". In: *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. (2013a). "Efficient Estimation of Word Representations in Vector Space". In: *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. Ed. by Y. Bengio and Y. LeCun.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. (2013b). "Distributed Representations of Words and Phrases and their Compositionality". In: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*. Ed. by C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger. 3111–3119.

Miller, G. A. (1995). "WordNet: a lexical database for English". *Communications of the ACM*. 38(11): 39–41.

Minaee, S., N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao. (2021). "Deep learning–based text classification: a comprehensive review". *ACM Computing Surveys (CSUR)*. 54(3): 1–40.

Mitra, V., C.-J. Wang, and S. Banerjee. (2007). "Text classification: A least square support vector machine approach". *Applied soft computing*. 7(3): 908–914.

Miyato, T., S.-i. Maeda, M. Koyama, and S. Ishii. (2018). "Virtual adversarial training: a regularization method for supervised and semi-supervised learning". *IEEE transactions on pattern analysis and machine intelligence*. 41(8): 1979–1993.

Mohammad, F. (2018). "Is preprocessing of text really worth your time for online comment classification?" *arXiv preprint arXiv:1806.02908*.

Moral, C., A. de Antonio, R. Imbert, and J. Ramírez. (2014). "A survey of stemming algorithms in information retrieval." *Information Research: An International Electronic Journal*. 19(1).

Mubarok, M. S., Adiwijaya, and M. D. Aldhi. (2017). "Aspect-based sentiment analysis to review products using Naıv̈e Bayes". In: *AIP Conference Proceedings*. Vol. 1867. AIP Publishing LLC. 020060.

Mueller, J. and A. Thyagarajan. (2016). "Siamese recurrent architectures for learning sentence similarity". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30.

Mullen, L. A., K. Benoit, O. Keyes, D. Selivanov, and J. Arnold. (2018). "Fast, consistent tokenization of natural language text". *Journal of Open Source Software*. 3(23): 655.

Mullen, T. and R. Malouf. (2006). "A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse." In: *AAAI spring symposium: computational approaches to analyzing weblogs*. 159–162.

Naseem, U., I. Razzak, and P. W. Eklund. (2021). "A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter". *Multimedia Tools and Applications*. 80(28): 35239–35266.

Nowak, J., A. Taspinar, and R. Scherer. (2017). "LSTM recurrent neural networks for short text and sentiment classification". In: *International Conference on Artificial Intelligence and Soft Computing*. Springer. 553–562.

Paice, C. D. (1990). "Another Stemmer". *SIGIR Forum*. 24(3): 56–61. DOI: 10.1145/101306.101310.

Palmer, D. D. (1997). "A trainable rule-based algorithm for word segmentation". In: *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. 321–328.

Pang, B., L. Lee, and S. Vaithyanathan. (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". In: *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. 79–86.

Pardo, F. M. R., A. Giachanou, B. Ghanem, and P. Rosso. (2020). "Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter". In: *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020*. Ed. by L. Cappellato, C. Eickhoff, N. Ferro, and A. Névéol. Vol. 2696. *CEUR Workshop Proceedings*. CEUR-WS.org.

Pecar, S., M. Simko, and M. Bielikova. (2018). "Sentiment Analysis of Customer Reviews: Impact of Text Pre-Processing". In: *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*. 251–256. DOI: 10.1109/DISA.2018.8490619.

Peng, H., J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, and Q. Yang. (2018). "Large-scale hierarchical text classification with recursively regularized deep graph-cnn". In: *Proceedings of the 2018 world wide web conference*. 1063–1072.

Peng, H., J. Li, S. Wang, L. Wang, Q. Gong, R. Yang, B. Li, S. Y. Philip, and L. He. (2019). "Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification". *IEEE Transactions on Knowledge and Data Engineering*. 33(6): 2505–2519.

Peng, T., W. Zuo, and F. He. (2008). "SVM based adaptive learning method for text classification from positive and unlabeled documents". *Knowledge and Information Systems*. 16(3): 281–301.

Pennington, J., R. Socher, and C. D. Manning. (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

Pereira, J. M., M. Basto, and A. F. da Silva. (2016). "The logistic lasso and ridge regression in predicting corporate failure". *Procedia Economics and Finance*. 39: 634–641.

Pérez-Almendros, C., L. E. Anke, and S. Schockaert. (2020). "Don't Patronize Me! An Annotated Dataset with Patronizing and Condescending Language towards Vulnerable Communities". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 5891–5902.

Pérez-Almendros, C., L. E. Anke, and S. Schockaert. (2022). "SemEval-2022 task 4: Patronizing and condescending language detection". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics. 298–307.

Peters, B. and A. F. Martins. (2024). "Did Translation Models Get More Robust Without Anyone Even Noticing?" *arXiv preprint arXiv:2403.03923*.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018). "Deep Contextualized Word Representations". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*. Ed. by M. A. Walker, H. Ji, and A. Stent. Association for Computational Linguistics. 2227–2237. DOI: 10.18653/v1/n18-1202.

Peters, M., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (1802). "Deep contextualized word representations. arXiv 2018". *arXiv preprint arXiv:1802.05365*. 12.

Petrović, D. and M. Stanković. (2019). "The influence of text preprocessing methods and tools on calculating text similarity". *Facta Universitatis, Series: Mathematics and Informatics*. 34: 973–994.

Porter, M. F. (1980). "An algorithm for suffix stripping". *Program: electronic library and information systems*. 14(3): 130–137.

Pouyanfar, S., S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar. (2018). "A survey on deep learning: Algorithms, techniques, and applications". *ACM Computing Surveys (CSUR)*. 51(5): 1–36.

Pradha, S., M. N. Halgamuge, and N. T. Q. Vinh. (2019). "Effective text data preprocessing technique for sentiment analysis in social media data". In: *2019 11th international conference on knowledge and systems engineering (KSE)*. IEEE. 1–8.

Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum (CLEF). (2022) (Bologna, Italy, Sept. 5–8, 2022). Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. *CEUR Workshop Proceedings*. No. 3180. Aachen.

Qu, Z., X. Song, S. Zheng, X. Wang, X. Song, and Z. Li. (2018). "Improved Bayes method based on TF-IDF feature and grade factor feature for chinese information classification". In: *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE. 677–680.

Quinlan, J. R. (1986). "Induction of decision trees". *Machine learning*. 1(1): 81–106.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.* (2019). "Language models are unsupervised multitask learners". *OpenAI blog*. 1(8): 9.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *Journal of Machine Learning Research*. 21(140): 1–67.

Rangel, F., G. L. D. la Peña Sarracén, B. Chulvi, E. Fersini, and P. Rosso. (2021). "Profiling Hate Speech Spreaders on Twitter Task at PAN 2021". In: *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021*. Ed. by G. Faggioli, N. Ferro, A. Joly, M. Maistro, and F. Piroi. Vol. 2936. *CEUR Workshop Proceedings*. CEUR-WS.org. 1772–1789.

Raschka, S. (2014). "Naive bayes and text classification i-introduction and theory". *arXiv preprint arXiv:1410.5329*.

Rastogi, R. and K. Shim. (2000). "PUBLIC: A decision tree classifier that integrates building and pruning". *Data Mining and Knowledge Discovery*. 4(4): 315–344.

Rathje, S., D.-M. Mirea, I. Sucholutsky, R. Marjieh, C. E. Robertson, and J. J. Van Bavel. (2024). "GPT is an effective tool for multilingual psychological text analysis". *Proceedings of the National Academy of Sciences*. 121(34): e2308950121.

Resyanto, F., Y. Sibaroni, and A. Romadhony. (2019). "Choosing the most optimum text preprocessing method for sentiment analysis: Case: iPhone Tweets". In: *2019 Fourth International Conference on Informatics and Computing (ICIC)*. IEEE. 1–5.

Rijsbergen, C. J. van. (1979). "Information Retrieval".

Rosid, M. A., A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali. (2020). "Improving text preprocessing for student complaint document classification using sastrawi". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 874. IOP Publishing. 012017.

Sabour, S., N. Frosst, and G. E. Hinton. (2017). "Dynamic routing between capsules". *Advances in neural information processing systems*. 30.

Sagolla, D. (2009). *140 characters: A style guide for the short form*. John Wiley & Sons.

Saif, H., M. Fernandez, Y. He, and H. Alani. (2014). "On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. 810–817.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". *arXiv preprint arXiv:1910.01108*.

Schlag, I., P. Smolensky, R. Fernandez, N. Jojic, J. Schmidhuber, and J. Gao. (2019). "Enhancing the transformer with explicit relational encoding for math problem solving". *arXiv preprint arXiv:1910.06611*.

Schuster, M. and K. Nakajima. (2012). "Japanese and korean voice search". In: *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 5149–5152.

Senette, C., M. Siino, and M. Tesconi. (2024). "User Identity Linkage on Social Networks: A Review of Modern Techniques and Applications". *IEEE Access*. 12: 171241–171268. DOI: 10.1109/ACCESS.2024.3500374.

Sennrich, R., B. Haddow, and A. Birch. (2016). "Neural Machine Translation of Rare Words with Subword Units". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1715–1725.

Shah, K., H. Patel, D. Sanghvi, and M. Shah. (2020). "A comparative analysis of logistic regression, random forest and KNN models for the text classification". *Augmented Human Research*. 5(1): 1–16.

Siino, M. (2024a). "All-Mpnet at SemEval-2024 Task 1: Application of Mpnet for Evaluating Semantic Textual Relatedness". In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Ed. by A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, and A. Rosá. Mexico City, Mexico: Association for Computational Linguistics. 379–384. DOI: 10.18653/v1/2024.semeval-1.59.

Siino, M. (2024b). "BadRock at SemEval-2024 Task 8: DistilBERT to Detect Multigenerator, Multidomain and Multilingual Black-Box Machine-Generated Text". In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Ed. by A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, and A. Rosá. Mexico City, Mexico: Association for Computational Linguistics. 239–245. DOI: 10.18653/v1/2024.semeval-1.37.

Siino, M. (2024c). "T5-Medical at SemEval-2024 Task 2: Using T5 Medical Embedding for Natural Language Inference on Clinical Trial Data". In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Ed. by A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, and A. Rosá. Mexico City, Mexico: Association for Computational Linguistics. 40–46. DOI: 10.18653/v1/2024.semeval-1.7.

Siino, M., E. Di Nuovo, I. Tinnirello, and M. La Cascia. (2021). "Detection of hate speech spreaders using convolutional neural networks". In: *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021*. Ed. by G. Faggioli, N. Ferro, A. Joly, M. Maistro, and F. Piroi. Vol. 2936. *CEUR Workshop Proceedings*. CEUR-WS.org. 2126–2136.

Siino, M., E. Di Nuovo, I. Tinnirello, and M. La Cascia. (2022a). "Fake News Spreaders Detection: Sometimes Attention Is Not All You Need". *Information*. 13(9): 426.

Siino, M., M. Falco, D. Croce, and P. Rosso. (2025). "Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches". *IEEE Access*. 13: 18253–18276. DOI: 10.1109/ACCESS.2025.3533217.

Siino, M., F. Giuliano, and I. Tinnirello. (2024a). "LLM Application for Knowledge Extraction from Networking Log Files". In: *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. 01–06. DOI: 10.1109/ICECCME62383.2024.10796967.

Siino, M., M. La Cascia, and I. Tinnirello. (2020). "WhoSNext: Recommending Twitter Users to Follow Using a Spreading Activation Network Based Approach". In: *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE. 62–70.

Siino, M., M. La Cascia, and I. Tinnirello. (2022b). "McRock at SemEval-2022 Task 4: Patronizing and Condescending Language Detection using Multi-Channel CNN, Hybrid LSTM, DistilBERT and XLNet". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Seattle, United States: Association for Computational Linguistics. 409–417. DOI: 10.18653/v1/2022.semeval-1.55.

Siino, M., F. Lomonaco, and P. Rosso. (2024b). "Backtranslate what you are saying and I will tell who you are". *Expert Systems*. 41(8): e13568. DOI: https://doi.org/10.1111/exsy.13568.

Siino, M. and I. Tinnirello. (2023). "XLNet with Data Augmentation to Profile Cryptocurrency Influencers". In: *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023*. Ed. by M. Aliannejadi, G. Faggioli, N. Ferro, and M. Vlachos. Vol. 3497. *CEUR Workshop Proceedings*. CEUR-WS.org. 2763–2771.

Siino, M. and I. Tinnirello. (2024a). "GPT Hallucination Detection Through Prompt Engineering". In: *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024*. Ed. by G. Faggioli, N. Ferro, P. Galuscáková, and A. G. S. de Herrera. Vol. 3740. *CEUR Workshop Proceedings*. CEUR-WS.org. 712–721.

Siino, M. and I. Tinnirello. (2024b). "GPT Prompt Engineering for Scheduling Appliances Usage for Energy Cost Optimization". In: *2024 IEEE International Symposium on Measurements & Networking (M&N), Rome, Italy, July 2-5, 2024*. IEEE. 1–6. DOI: 10.1109/MN60932.2024.10615758.

Siino, M. and I. Tinnirello. (2024c). "Prompt engineering for identifying sexism using GPT Mistral 7B". In: *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024*. Ed. by G. Faggioli, N. Ferro, P. Galuscáková, and A. G. S. de Herrera. Vol. 3740. *CEUR Workshop Proceedings*. CEUR-WS.org. 1228–1236.

Siino, M., I. Tinnirello, and M. L. Cascia. (2022c). "T100: A modern classic ensembler to profile irony and stereotype spreaders". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum (CLEF)* (Bologna, Italy, Sept. 5–8, 2022). Ed. by G. Faggioli, N. Ferro, A. Hanbury, and M. Potthast. *CEUR Workshop Proceedings*. No. 3180. Aachen. 2666–2674.

Siino, M., I. Tinnirello, and M. La Cascia. (2024c). "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers". *Information Systems*. 121: 102342. DOI: https://doi.org/10.1016/j.is.2023.102342.

Siino, M., I. Tinnirello, and M. La Cascia. (2024d). "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers". *Information Systems*. 121: 102342.

Singh, A., N. Singh, and S. Vatsal. (2024). "Robustness of llms to perturbations in text". *arXiv preprint arXiv:2407.08989*.

Singh, T. and M. Kumari. (2016). "Role of text pre-processing in twitter sentiment analysis". *Procedia Computer Science*. 89: 549–554.

Smelyakov, K., D. Karachevtsev, D. Kulemza, Y. Samoilenko, O. Patlan, and A. Chupryna. (2020). "Effectiveness of preprocessing algorithms for natural language processing applications". In: *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*. IEEE. 187–191.

Socher, R., A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. (2013). "Recursive deep models for semantic compositionality over a sentiment treebank". In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1631–1642.

Soucy, P. and G. W. Mineau. (2001). "A simple KNN algorithm for text categorization". In: *Proceedings 2001 IEEE international conference on data mining*. IEEE. 647–648.

Srividhya, V. and R. Anitha. (2010). "Evaluating preprocessing techniques in text categorization". *International journal of computer science and application*. 47(11): 49–51.

Stehman, S. V. (1997). "Selecting and interpreting measures of thematic classification accuracy". *Remote sensing of Environment*. 62(1): 77–89.

Symeonidis, S., D. Effrosynidis, and A. Arampatzis. (2018). "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis". *Expert Systems with Applications*. 110: 298–310.

Tai, K. S., R. Socher, and C. D. Manning. (2015). "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*. The Association for Computer Linguistics. 1556–1566. DOI: 10.3115/v1/p15-1150.

Taira, H. and M. Haruno. (1999). "Feature Selection in SVM Text Categorization". In: *Proceedings of the Sixteenth National Conference on Artificial Intelligence and Eleventh Conference on Innovative Applications of Artificial Intelligence, July 18-22, 1999, Orlando, Florida, USA*. Ed. by J. Hendler and D. Subramanian. AAAI Press / The MIT Press. 480–486.

Tan, L., H. Zhang, C. Clarke, and M. Smucker. (2015). "Lexical comparison between wikipedia and twitter corpora by using word embeddings". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 657–661.

Tan, S. (2005). "Neighbor-weighted k-nearest neighbor for unbalanced text corpus". *Expert Systems with Applications*. 28(4): 667–671.

Thelwall, M. (2017). "The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength". In: *Cyberemotions*. Springer. 119–134.

Toman, M., R. Tesar, and K. Jezek. (2006). "Influence of word normalization on text classification". *Proceedings of InSciT*. 4: 354–358.

Touvron, H., T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, Y. Jernite, E. Grave, and G. Lample. (2023). "LLaMA: Open and Efficient Foundation Language Models". *arXiv preprint arXiv:2302.13971*. URL: https://arxiv.org/abs/2302.13971.

Uysal, A. K. and S. Gunal. (2014). "The impact of preprocessing on text classification". *Information processing & management*. 50(1): 104–112.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett. 5998–6008.

Vateekul, P. and M. Kubat. (2009). "Fast induction of multiple decision trees in text categorization from large scale, imbalanced, and multi-label data". In: *2009 IEEE International Conference on Data Mining Workshops*. IEEE. 320–325.

Vijayarani, S., M. J. Ilamathi, and M. Nithya. (2015). "Preprocessing techniques for text mining-an overview". *International Journal of Computer Science & Communication Networks*. 5(1): 7–16.

Vijayarani, S. and R. Janani. (2016). "Text mining: open source tokenization tools-an analysis". *Advanced Computational Intelligence: An International Journal (ACII)*. 3(1): 37–47.

Virmani, D. and S. Taneja. (2019). "A text preprocessing approach for efficacious information retrieval". In: *Smart innovations in communication and computational sciences*. Springer. 13–22.

Wan, S., Y. Lan, J. Guo, J. Xu, L. Pang, and X. Cheng. (2016). "A deep architecture for semantic matching with multiple positional sentence representations". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30.

Wang, H. and J. A. Castanon. (2015). "Sentiment expression via emoticons on social media". In: *2015 ieee international conference on big data (big data)*. IEEE. 2404–2408.

Wang, S. I. and C. D. Manning. (2012). "Baselines and bigrams: Simple, good sentiment and topic classification". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 90–94.

Wang, Z., W. Hamza, and R. Florian. (2017). "Bilateral Multi-Perspective Matching for Natural Language Sentences". In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. Ed. by C. Sierra. ijcai.org. 4144–4150. DOI: 10.24963/ijcai.2017/579.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou. (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". In: *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh.

Wiegmann, M., B. Stein, and M. Potthast. (2020). "Overview of the Celebrity Profiling Task at PAN 2020". In: *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020*. Ed. by L. Cappellato, C. Eickhoff, N. Ferro, and A. Névéol. Vol. 2696. *CEUR Workshop Proceedings*. CEUR-WS.org.

Xiao, Y. and K. Cho. (2016). "Efficient character-level document classification by combining convolution and recurrent layers". *arXiv preprint arXiv:1602.00367*.

Yamaguchi, H. and K. Tanaka-Ishii. (2012). "Text segmentation by language using minimum description length". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 969–978.

Yang, M., W. Zhao, J. Ye, Z. Lei, Z. Zhao, and S. Zhang. (2018). "Investigating Capsule Networks with Dynamic Routing for Text Classification". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii. Association for Computational Linguistics. 3110–3119. DOI: 10.18653/v1/d18-1350.

Yang, Z., Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le. (2019). "Xlnet: Generalized autoregressive pretraining for language understanding". *Advances in neural information processing systems*. 32.

Yang, Z., D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy. (2016). "Hierarchical attention networks for document classification". In: *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*. 1480–1489.

Yao, L., C. Mao, and Y. Luo. (2019). "Graph convolutional networks for text classification". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 7370–7377.

Ye, X. and G. Durrett. (2022). "The Unreliability of Explanations in Few-shot Prompting for Textual Reasoning". In: *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh.

Yu, H., Z. Yang, K. Pelrine, J. F. Godbout, and R. Rabbany. (2023). "Open, Closed, or Small Language Models for Text Classification?" *arXiv preprint arXiv:2308.10092*. URL: https://arxiv.org/abs/2308.10092.

Zeng, D., K. Liu, S. Lai, G. Zhou, and J. Zhao. (2014). "Relation classification via convolutional deep neural network". In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 2335–2344.

Zeng, J., J. Li, Y. Song, C. Gao, M. R. Lyu, and I. King. (2018). "Topic Memory Networks for Short Text Classification". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii. Association for Computational Linguistics. 3120–3131. DOI: 10.18653/v1/d18-1351.

Zhang, T., M. Huang, and L. Zhao. (2018). "Learning structured representation for text classification via reinforcement learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32.

Zhang, X., J. Zhao, and Y. LeCun. (2015). "Character-level convolutional networks for text classification". *Advances in neural information processing systems*. 28.

Zhang, Y., D. Song, P. Zhang, X. Li, and P. Wang. (2019). "A quantum-inspired sentiment representation model for twitter sentiment analysis". *Applied Intelligence*. 49: 3093–3108.

Zhang, Y. and B. Wallace. (2015). "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification". *arXiv preprint arXiv:1510.03820*.

Zhou, P., Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu. (2016). "Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling". In: *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*. Ed. by N. Calzolari, Y. Matsumoto, and R. Prasad. ACL. 3485–3495.

Zhu, X., P. Sobihani, and H. Guo. (2015). "Long short-term memory over recursive structures". In: *International Conference on Machine Learning*. PMLR. 1604–1612.

Zhu, Y., X. Gao, W. Zhang, S. Liu, and Y. Zhang. (2018). "A bidirectional LSTM-CNN model with attention for aspect-level text classification". *Future Internet*. 10(12): 116.

Zong, C., R. Xia, and J. Zhang. (2021). "Data Annotation and Preprocessing". In: *Text Data Mining*. Singapore: Springer Singapore. 15–31. DOI: 10.1007/978-981-16-0100-2_2.