Journal of Information Systems Engineering and Management
2025, 10(26s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Review Article for Argumentation Mining of Text through Machine Learning Techniques and Strategies

Rudrendra Bahadur Singh[1], Shobhit Sinha[2], Ankita Singh[3]
[1]*Research Scholar, Shri Ramswaroop Memorial University, Tindola, Barabanki-225003, Rudra.rathor20@gmail.com*
[2]*Shri Ramswaroop Memorial University, Tindola, Barabanki-225003, shobhitsinha.dcsis@srmu.ac.in*
[3]*mailtoankitasingh@gmail.com*

ARTICLE INFO

Received: 26 Dec 2024

Revised: 16 Feb 2025

Accepted: 01 Mar 2025

ABSTRACT

Argumentation Mining (AM), a specialized branch of Natural Language Processing (NLP), which extract arguments from text and mapping out their relationships. While machine learning has been extensively explored for AM sub tasks, there's still a gap in structuring these methods to spot common patterns across different applications. This study, based on a review of 64 research papers, breaks down how AM is applied across various domains ranging from user-generated texts, English texts, speech to debates, legal documents, and scientific or medical texts. Among these, text takes the lead as the most researched area. Particularly Support Vector Machines (SVM), Bidirectional Encoder Representations from Transformers (BERT) and Bidirectional Long Short-Term Memory (BiLSTM), Convolutional Neural Network (CNN) are some machine learning models that dominates the field. The effectiveness of these models varies depending upon the type of text, excelling in user-generated text where as others perform better with scientific or medical data. The study highlights the need to further explore less-researched areas especially machine learning applications in legal, medical scientific and English texts and critically examine how Large language model and deep learning stacks up against traditional methods. By mapping these insights, the goal is to help researchers pick the right approach for specific AM tasks, ultimately pushing the field forward.

INTRODUCTION

The Argumentation Mining (AM) subfield of Natural Language Processing (NLP) is concerned with revealing the structure in complex texts by discovering the main idea, author's viewpoint and relations between different discourse units [1]. AM seeks to automate the detection of premises, claims, and conclusions within an argument [2]. This discipline is acquiring greater significance with the rising volume of textual information containing argumentative speech from various sources and domains. Articles and essays are comprised of structured presentation of premises and claims on a given subject. Structured political debates constitute argumentative debates among candidates on other issues. Social media sites offer an avenue for users to discuss and argue about controversial topics [3]. For example, legal documents have law-reasoning with a complicated underlying argumentative structure [4]. The internal organisation of an argument contains a few argument elements. It comprises a claim and more than one premise. The claim is a contention and the very essence of an argument, while premises are considerations in favour of supporting the claim. Further, arguments possess argumentative relations that are directed and represent the interaction between one component with another[5]. Argumentation structure identification encompasses several subtasks which rely on a pipeline approach divided into three overall steps: the detection of argumentative text and non-argumentative text, the classification of argument type as claims and premises, and detection of the paired argument relationship [6].

Argumentation mining (AM) is the automatic extraction of arguments, their constituents, and their relations from natural language text. The process is often bifurcated into three essential tasks: (1) argument identification, which entails text segmentation and an identification of what in the text is argumentative; (2) argumentative component recognition, typically by categorizing text as claims and premises; and (3) argumentative relationship identification, which emphasizes comprehension of how various parts of the text are related within the context of argumentative discourse [7]. In Natural language Processing (NLP), Large language models (LLMs) are a general-purpose language task solver and the research model has been moving towards the employment of LLMs [8]. AM is concerned with automated extraction, identification, and interpretation of arguments from natural language text. It differs from sentiment analysis and opinion mining, which are mainly concerned with the affective or subjective content of the text, as it digs deeper into the structure of reasons in order to answer the "why" question regarding a specific perspective or point of view [9]. This makes AM especially useful in fields that need a high level of logical reasoning and insight extraction, including legal analysis, policymaking, education, and healthcare [10]. Recent studies also investigate how argumentation mining employs common sense, reasoning and world knowledge in mining research.

AM applications span a diverse range of text types, including user-generated content (such as forum posts, product reviews, and social media comments), legal documents, scientific papers, medical texts, mathematical discourse, and non-English text corpora. The increasing demand for insights from these heterogeneous data sources has driven significant advancements in Machine Learning (ML) techniques, which serve as the backbone of modern AM systems. Early studies relied on traditional ML models like Support Vector Machines (SVM), Random Forest, Logistic Regression, Convolutional Neural Network (CNN) and Naive Bayes, which excelled in specific scenarios by leveraging structured features. However, recent developments in deep learning, with models like BERT (Bidirectional Encoder Representations from Transformers) and BiLSTM (Bidirectional Long Short Term Memory), have revolutionized the field by enabling context-aware and semantically rich argument extraction [11] [12] [13] .

A newly emerging language representation model called Bidirectional Encoder Representation from Transformers (BERT) is inspired by a cutting-edge trained deep learning strategy that has posted superb performance in numerous difficult tasks at Natural Language Processing (NLP) [14]. BERT is a bidirectional Deep Learning (DL) based model that processes different types of text taking input from both the left and right sides rather than a single direction. The latest BERT model takes on the Transformer architecture, which consists of multiple encoded layers [15]. The BERT model has two phases in our framework: pre-training and fine-tuning. In pre-training, the model is trained on unlabeled data across various pre-training tasks. For fine-tuning, The BERT model is initialized with the pre-trained parameters first, and all of the parameters are fine-tuned with labeled data from the downstream tasks [16].

Deep learning on Argument Mining (AM) is a comparatively recent innovation among the other NLP areas. It is largely attributable to the scarcity of large existing AM datasets that is an indirect consequence of task complexity and specialization. Large corpus annotation for the training and test of AM systems has proved troublesome, as evident from low Inter-Annotator Agreement (IAA) results and crowd sourced annotation attempts proving unsuccessful. These problems are especially evident in some genres, like user-generated content. The issue arises from the cognitively challenging nature of the task and the lack of an ideal argument model. Because arguments are so diverse across genres, there usually is a compromise between the expressiveness of the argument model and annotation complexity, normally resulting in the use of less complex models to make it easier to annotate and provide enough data. [17].

Despite these advancements, challenges remain. Current research often lacks a cohesive understanding of the relationship between specific machine learning methods and the unique requirements of different AM applications. For example, user-generated content and scientific texts may demand vastly different features and processing techniques. Moreover, while some domains such as user-generated texts have been extensively studied, others, including scientific and medical texts, remain underexplored. These gaps hinder the development of a universal framework that links machine learning methodologies with AM applications effectively.

To address these challenges, this paper conducts a comprehensive review of 64 research studies on argumentation mining. It aims to uncover patterns between the applications of AM and the machine learning models employed, providing a systematic understanding of which methods are best suited for specific types of text. Additionally, the paper highlights areas that require further investigation, such as applying AM techniques to less-studied text types and comparing the performance of deep learning approaches with traditional machine learning methods. By doing so, it offers actionable insights for advancing the field and optimizing the use of machine learning in argumentation mining.

This investigation not only identifies current trends but also provides a roadmap for future research, encouraging a more targeted application of ML methods in AM. It underscores the potential of AM to revolutionize how arguments are extracted and analyzed across domains, ultimately contributing to more robust decision-making systems and enriched understanding of human reasoning processes.

RELATED WORK

In the literature, several different approaches have been proposed for argumentation mining for medical, legal, political text and essays etc. The work introduces Rhetorical Structure Theory (RST) a deep parsing model of dependency to evaluate the relationship between argument and rhetorical structures. The approach permits end-to-end argumentation analysis in terms of a rhetorical tree rather than word sequence. It is tested on the bilingual Microtexts corpus and obtains results on full-fledged argument parsing. [2]. We created an annotation scheme for legal arguments in proceedings on a strong argument mining that would be able to overcome the special difficulties of the legal language of the ECHR judgments. We showed that the larger and stronger generic model, the smaller generic legal domain tuned Legal-BERT, can be equaled by the Robustly optimized BERT method (RoBERTa)-Large[4].

Applying T5 in the Translation between Augmented Natural Languages (TANL) systems, we introduce text-to-text generation tasks and strip unncessary text spans off the reference texts to simplify and streamline annotations. Our solution excelled the current state-of-the-art on the argument annotated Essays Corpus (AAEC), AbstRCT and Cornelle Rulemaking Corpus (CDCP) these corpora. This fremework improves the efficiency of applying TANL framework in document structure prediction at the document level [6]. An end-to-end argument mining (AM) pipeline is optimized on eight widely used quantized and non-quantized LLMs –LLaMA-3, LLaMA-3.1, Gemma-2, Mistral, Phi-3, Qwen-2 to discover which ones are some of the best open-weight models, on the benchmark PE, AbstRCT, and CDCP datasets that are representative of varied data sources[1].

The field of Relation-based AM (RbAM) method with Large Language Models, with the purpose of detecting attack and support relations between arguments. We employed ten publicly available labeled datasets directly are suitable for RbAM task definition, and we have treated any relation besides attack and support as irrelevant and five open sourced LLMs that Llama 70B-4bit and Mixtral 8x7B-4bit outperformed the fine-tune Robustly Optimized BERT Approach (RoBERTa) baseline. [7].

The field of Relation-based AM (RbAM) method with Large Language Models, with the purpose of detecting attack and support relations between arguments. We employed ten publicly available labeled datasets directly are suitable for RbAM task definition and we have treated any relation besides attack and support as irrelevant and five open sourced LLMs that Llama 70B-4bit and Mixtral 8x7B-4bit outperformed the fine-tune Robustly Optimized BERT Approach (RoBERTa) baseline [7].For better accuracy ,fine-tuning BERT–MINUS concentrate on two Argument Mining sub-tasks: Link Identification (LI) and Argument Type Classification (ATC) to identify the contextual and argumentative movement between successive parts[3]

This paper presents Argument Relation Identification aimed at providing a first comparative set of results with the three main Transformer-based model architectures: encoder-only, decoder-only, and encoder-decoder [18]. From argumentative text excerpts, both Argumentative Components and Argumentative Relation introduced a paradigm-based end-to-end approach, the argTANL framework. On improving the performance of the model, integrate two different frameworks, Marker-Enhanced argTANL (ME-argTANL) and argTANL with Marker-Based Fine-Tuning [19]. A review of the results on many metrics and aspects reveals that it's one of the quickest evolving and most responsive models. In addition to solving broad complex problems, BERT stands out in state-of-the-art text mining, processing, analysis, and detection achievement [14]. RESATTARG is a novel argument mining neural architecture that combines residual networks, multi-task learning, neural attention, and ensemble learning. Importantly, this method steers clear of dataset-specific architectural components, like structural features or domain-specific encodings [17]. A pipeline-based method was developed for full-text argumentation mining in scientific publications, demonstrating its effectiveness by achieving new state-of-the-art performance on the Sci-Arg corpus [20]. For evaluating the robustness of argument mining models, it presents a model-agnostic framework. It utilizes 15 simulation functions, including 6 novel ones for argument classification and 9 adapted for argument mining tasks. The framework was tested on the IBM Debater Evidence Sentence dataset and UKP topic sentences. Results showed that while BERT models are robust, they remain vulnerable to unfamiliar inputs [21]. The BERT model offers significantly better compatibility acompared to traditional models. Variants like BioBERT are specialized for processing medical text, ClinicalBERT is tailored for clinical text, and SciBERT is designed for handling scientific text [22]. Fine-tuned BERT models can provide satisfying results for text categorization on different downstream tasks of different domains via transfer learning [15].

BACKGROUND

In argumentation mining selection of an argumentation model is required which classify argument into component and structures. Structured argumentation models analyse the internal components and examine relationships between components. One method of assessing the organization of arguments in a text is by classifying them as "abstract" or "structured." Abstract argumentation models focus on analysing arguments without delving into their internal components. These arguments emphasize on enormous ideas rather than strict structures [9].

This paper seeks to bridge this gap by tackling several key research questions. First, it examines the reliability of operationalizing argument models from datasets using discourse annotations. Second, it investigates how to create robust argument mining models that surpass current state-of-the-art performance, even when constrained by the high cost of expert-labelled data. Structured argumentation models analyse arguments by looking at their internal structure. It examines the part of an argument like claims and reasons, and tells either these parts support or opposes each other [23]. Structured arguments follow detailed step by step structure which provides clear and detailed understanding and also trail a clear sequence or flow while there is no proper universal description for structured argumentation mining [9].

Classification of argumentation models are micro level (monological), macro level (dialogical) or rhetorical models. Micro level works on small arguments or on individual arguments. Micro -level arguments focus into more subtle details of an argument and prioritize single point of view. Macro-level also known as “dialogical” and consist of many conversations where distinct viewpoints are shared. Macro level focuses on relationship between arguments and their external structure. It allows people to retrace and put their own perspective. Rhetorical model concentrates on persuasion, targeting to convince or influence the audience [2].

The process of argumentation mining can be divided into smaller tasks. Many micro-level models include basic frameworks such as the claim model, Toulmin model, and the Freeman model. The simple claim model is a direct statement that backs up the main argument. It further categorizes in three parts - an inference for certain premises and a connection between the premises and the conclusion. A Toulmin model is a framework for argument analysis by philosopher Stephen Toulmin. An argument is classified into six parts – claim, data, warrant, backing, qualifier and rebuttal [24]. A Freeman model is a prolongation of Toulmin model, engineered for more detailed argument analysis that includes premise, conclusion, rebuttal and counter-rebuttal.

A repository of sample text or arguments is mandatory for training machine learning models (ML) and enforcing argumentation mining systems. The repository must be highlighted according to selected arguments. It labels the components of arguments and the bond between them which is difficult for effective training. Challenge factor makes it crucial to compare and create corpora and argumentation mining techniques [9].

Argumentation mining can be classified into subtasks. In argumentation mining most challenging part is to process complex arguments that involves multiple subtasks. Some papers categorize the subtask division process in two parts - Argument relation prediction and argument component extraction [25][26]. Identifying arguments within the text is called argument component extraction. Once identified, these components can be used for argument analysis. It uses the methodology of machine learning (ML) algorithm and natural language processing (NLP) techniques to make the extraction process automatic. This helps in analysis of huge volume of text accurately and instantly. Figuring out the parameters of text of identified argument components is called textual boundaries [25]. This stage involves connection of different parts of arguments. It defines that if one argument supports the other or not. Research categorizes argumentation mining into four subtasks. The first involves classifying text as either argumentative or non-argumentative [27], naturally using machine learning techniques. The third step is to identify the components of the arguments that are related. The fourth step is to determine the nature of their relationship, such as whether one supports the other or not [9].

Machine learning aims to use data and computer algorithms to mimic human learning. It is typically divided into three categories: traditional neural networks, deep learning and machine learning. Traditional machine learning refers to techniques that have been in use for many years, including clustering, classification and regression Neural networks are a type of machine learning that simulate the human brain, with nodes acting as basic units connected by thresholds and weights. A neural network includes an input layer, an output layer, and several hidden layers of nodes. The data enters the network from input layer and there are multiple hidden layers in between for data processing in between the two layers. Once the hidden layers process the data and produce an output, then the final result is produced by the output layer.

Significant research-based literature reviews describe argumentation mining with the help of machine learning, providing an overview that explores new datasets and highlights the challenges in this field [10]. By analysing sentences from news articles, IBM's Project Debater can engage in debates with humans, leveraging both neural and knowledge-based techniques to assess the likelihood that each sentence forms part of an argument [28]. A fully neural approach allows us to understand and analyse arguments comprehensively, known as end-to-end neural models. In 2017, a dataset of 402 annotated persuasive essays was created and has been widely used in research to improve machines' argument-mining skills. Machine learning plays a major role in teaching computers to break down and classify arguments [5].

ISSUES, CHALLENGES AND RESEARCH GAPS

Argumentation mining is more popular because of user generated text, it helps in understanding public opinion and general feeling of large group of peoples. Various sectors get benefited by the insights gained from argumentation mining such as understanding political views, improving business, true analysis of online reviews. Multiple application is the second most common application which is used with persuasive essays. These applications are widely used in argumentation mining or argument mining and also indicates a growing interest in cross domain system. Cross domain argumentation mining is rising interest in research which can detect argument in various domain and also reduces need for multiple specialized system. Implementation of cross domain in argumentation mining could have been important lasting effects on how argumentation mining is used in other fields.

Deep learning models like BERT has much complexity because of their large data sets for training while traditional methods like SVM can perform on very small data sets, this makes SVM a feasible choice where data is limited. In deep learning models automatically learn features are embedded. Often removing the necessity for manual feature selection. Deep learning models generally need more significant and computational power, example GPU, due to their high complexity and size whereas SVM requires less computational power and can run on standard hardware. For enabling to enhance power performance of specific task we need BERT based model which are already tuned to specific domains (example SciBERT or ruBERT). Methods like SVM need more effort to achieve its similar adaptability. Consistent performance is given by SVM overtime in argumentation mining task, provide a stable baseline while deep learning proved to be potentially more powerful may show variance depending on data volume, models and tuning intricacies. In user generated text application SVM has shown a strong correlation suggesting its particularly for this data type.

The persuasive essays application is popular due to their inherently argumentative nature supported by specific corpora that helps in research. Philosophers, scientific, medical and legal text are lagging clear arguments and being time consuming to annotate. The major lack of research in scientific and medical text also needs an opportunity to advance argumentation mining techniques and further exploration in this area. As there is lot of potential for more research in scientific texts. In order to find arguments, SVM works well with small dataset which is user generated text because SVM struggles with scalability when applied to very large datasets. In fields like speeches and debates, legal text and other multiple domain applications SVM is widely used where traditional learning methods have been historically effective but performance of SVM depends heavily on the selection of kernel parameters and regularization constants. SVM also faces the computational complexity with large datasets.

There is a substantial gap in research for deep learning methods in user generated text application despite of SVM's popularity. In deep learning model's further investigation have been done such as transformer based approached might offer enhanced performance in this domain. For better computation BERT Model requires large amounts of labelled data for fine-tuning, but this increases the computational cost during pre-training.

Table 1: Review Criteria

Parameters	Inclusion Criteria	Exclusion Criteria
Publication year	2013-2024	Before 2013
Argumentation Mining Technology	LLM, BERT, SVM, LSTM	Others

Pre-processing	Yes	No
Learning/Training approaches	Deep Learning Methods	Artificial Intelligence Methods
Train/Test ratio	70-30 80-20	Others
Performance metrices	Accuracy/precision/Recall/Error /F-measure	Not available
Dataset	UCI/Kaggle	Others/ collected Privately

As mentioned below Figure 1, the section criteria of articles as shown based on various parameters mention in table 1.

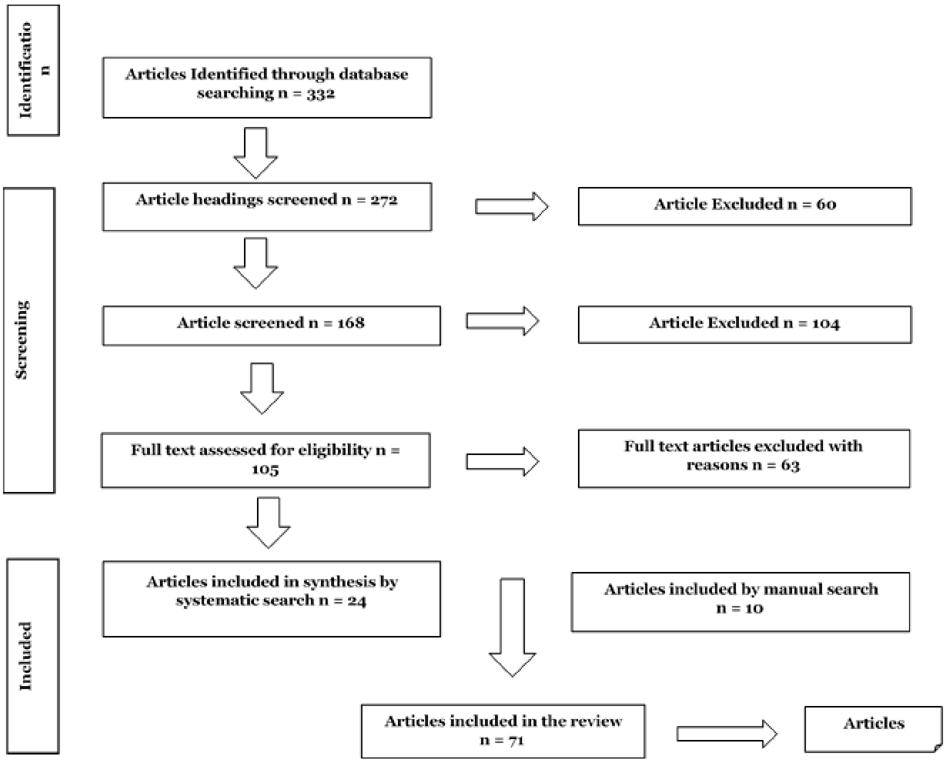

Figure 1. PRISMA Dataflow Model

Meanwhile, English, medical and scientific texts prefer BERT model as they are newer. Non-English text still prefers LSTM models. Along with scientific texts, these areas rarely use SVM models. The inclination towards LSTM in non-English text is worth more research, as it is an old application.

Most widely studied applications are user generated text, tends to use a board range of machine learning model. It allows more experimentation and model testing in these areas. Applications like speech and debate, scientific and medical text used fewer models overall. It is possibly due to more focused nature of these domains. The application which are less explored such as scientific, medical, philosophical and legal text could benefit from border testing of machine learning model. It enables new findings and also encouraging further research in this field.

Only 64 papers were included in the diagrams and figures of this paper, further research could be explored on the paper by analysing, finding and incorporating additional papers into the data gathered for this literature review.

Summarizing papers and make the findings easier for everyone to understand.

The gap in the literature is in understanding the relationship between AM applications and ML techniques. This paper aims to investigate the links between the applications of AM and ML models. A review of 64 research papers was conducted to gain a deeper understanding of these relationships. These researchers helped in analysing the data to find patterns between applications and machine learning methods. Future research in this field could use the results of this literature review.

METHODOLOGY

The papers were found by searching on Google Scholar with the terms “AM, LLM and ML”. Papers mentioning machine learning, neural networks and deep learning alongside argument mining or argumentation mining were chosen. The review of the papers assessed found that at least one ML technique was used in AM. In argumentation mining and machine learning the papers that were used to analyse the texts were comprised in literature review. For investigation in machine learning research in argumentation mining and the conclusions that were made about the other aspects of argumentation mining were included. Studies without machine learning is not able to process non-text data were excluded. Research lacking clarity in machine learning models was excluded from this paper, which primarily focuses on clearly defined applications of machine learning in argumentation mining or argument mining.

Literature review of argumentation mining covers various applications including text generated by user, medical text, legal text, English text, scientific text and multiple domain applications. 64 papers are reviewed for finding primary applications which are identified by examining title and introduction. Peoples that do not have a specific purpose but focuses to improve features or models were organised according to the training data. Non-English text papers like which translated effective essay into Bahasa Indonesia work categorised as non-English text [29]. Paper using argumentation mining in different fields and papers that were used in Aracuaria were also grouped under “Multiple Applications” because corpus include diver’s sources. All the models that were used were recorded for identifying machine learning model for methods.

FINDINGS: ARGUMENTATION MINING APPLICATIONS

A total of 64 papers were noticed via Google Scholar and organized into one or more applications, Speech/Debate, Philosophical Text, Multiple Applications, Persuasive Essays, Legal Text, English Text, Scientific/Medical Text, and User-Generated Text. Figure 2 illustrates the tempo of each application in the reviewed papers.

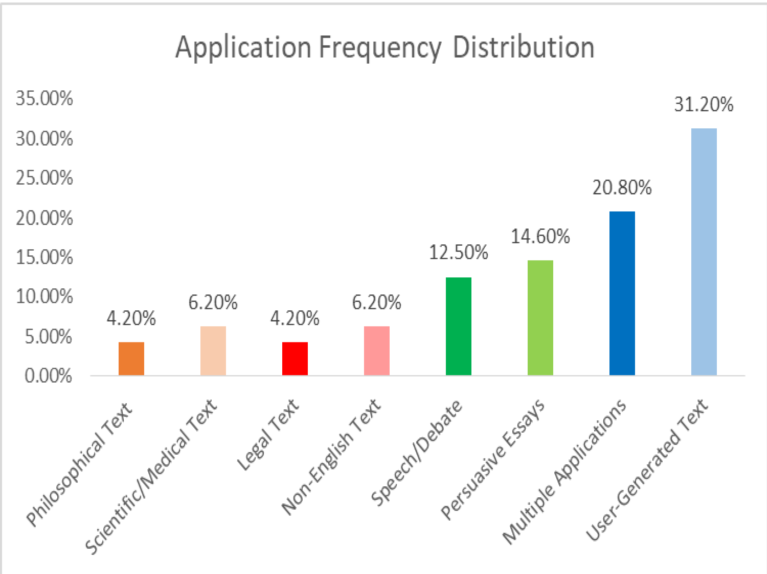

Figure 2: Research’s shows the uses of AM

Eighteen papers concentrate on examining arguments in user-generated content, mainly including comments, blog articles, and forum posts gathered from the web. In their 2015 paper, Hibernal and Gurevich employed argumentation mining on posts in debate portals, which host user exchanges on contentious topics. In 2017, they used the Toulmin model for argumentation mining on a collection of remarks, argument forums, blog posts, and articles. Their team employed an existing argumentation mining system to detect arguments in Amazon product reviews [30]. Used the CorEA corpus, derived from online comments on Italian news blogs, categorizes the comments as supportive, neutral, or attacking. A sorted collection of comments from a government website was made by park where people discuss rules and policies in other words. Swanson they combined previous finished internet argument

Corpus with added data from an online debate platform which permit users to do discussion numerous topics [31].Recent progress in NLP has introduced ground breaking methods for AM. One notable development is the text-to-text generation framework, which reframes AM as a cohesive generative task. This method streamlines the typically divided workflow by integrating span identification, component classification, and relation classification into a unified process, utilizing advanced pretrained encoder-decoder models such as T FLAN-5 and T5 [6]. Dusmanu conduct argumentation mining on Twitter, while Chakrabarty use argumentation mining on the to change my opinion for evaluating users' arguments and comments [32][33]. They applied argumentation mining on "Change my View" subreddit. Tran and Litman used alike data and aim to enhance the precision scores of investigation features and machine learning classifier for argumentation mining is verified for the approach on a Corpus of Wikipedia article [34][33][35]. Execute stance classification on online news comments by using a Corpus of Portuguese new article they developed ArgMine framework to automate part of argumentation mining [36][37]. The main focus was given on classifying relationship between premises and conclusions using the same corpus [38]. Argumentation mining was executed with data from online debates. Argumentation Mining was applied to detect misleading product reviews, on the other hand to identify argument components in Greek news article [39] [40]. In the end perform argument detection for short text using data from Twitter and several papers [41].

Eight papers were used in AM to explore multiple applications and developing MARGOT [42]. Margot was made to analyse argument in different document and topics for various uses. *Argumentation mining can be used effectively across different domains like Cross-domain applications. Detection of truth and fact without background information of the topic and also for investigating the Araucaria Corpus by Mochales and Moens which encompasses data [12][43]. It is more challenging to understand how machine learning can improve argumentation mining,* apply multi-task learning techniques for argumentation mining, utilizing diverse data sources such as persuasive essays and news comments *[44][45]*. This approach enhances argumentation mining performance, especially when working with limited datasets. Meanwhile, apply kernel methods for argumentation mining, utilizing the Araucaria DB dataset [46]. They applying Transfer Learning and multiple corpora to build a domain-independent implementation of an argumentation mining. Several papers are using different machine learning techniques to increase argumentation mining across numerous domains. Machine learning was used to create persuasive essays corpus [27] [5][47]. A series of corpus was executed [48]. Neural End-to-end solution for all argumentation mining part of task built by the [49][35]. The persuasive essays dataset was also used to perform unsupervised argumentation mining [50]. And the app was created using argumentation mining to contribute feedback on student convincing writing [51].

In field of Speech and Debate, an Autonomous system was created which competed with the humans [28]. They utilized a corpus of 400 million newspaper articles, which were further divided into sentences, words, and linked to related Wikipedia concepts. Using neural and knowledge-based methods, the system ranks the sentences. Additionally, the use of argumentation mining in various domains was to assess its productivity in cross-domain applications [52]. After developing the autonomous system, "classify the stance of each argument towards the motion". Lippi and Torrone in 2015 UK political election train the machine learning classifier which used in political debates [10]. Lastly, Mochales and Moens use Araucaria collection to investigate argumentation mining's submissions and experiments. To increase the accuracy of argumentation they found vocal cues which helps to reduce the tone of voice [24]. The first speech delivered by using the argumentation mining in 1960 presidential Tang uses argumentation mining to abstract claims from TED talk subtitles, while this approach was also applied to analyse debate speeches from the Canadian Parliament [53]. Merge textual and audio data from US political corpus to perform argumentation mining [54].

For argumentation mining in non-English texts, researchers typically either translate existing English datasets into the target language or create new datasets to test its effectiveness. This area of research focuses only on studies specifically examining argumentation mining in non-English languages or making significant contributions to multilingual argumentation mining. For instance, argumentation mining on a Portuguese news article corpus was used, this study was excluded here because it had a different focus [38].

Five notable studies have applied argumentation mining to non-English texts translated cogent articles into Bahasa Indonesian and used machine learning to conduct argumentation mining [29]. A similar translation into Russian to evaluate model performance [55]. Amharic-language texts, used a polyglot model to assess English data decoded into Dutch, German, Spanish, Italian, and French, predicting whether arguments were supportive, attacking, or neutral

[26][18][56]. In a similar vein, translation of persuasive essays into German, French, Spanish, and Chinese to investigate cross-lingual argumentation mining [17].

In the legal field, scholars have employed argumentation mining to examine legal documents [57][23][58]. Experiments were commenced with multiple machine learning models to identify key argument elements and generate case summaries that assist legal practitioners [57]. Sentences were classified in legal texts as either argumentative or non-Argumentative. Models that were pre-trained with legal data were utilized to increase legal argument mining [23]. In distinction, a new mass was created to apply argumentation mining to finding from the European Court of Justice [58].

AM in the medical and scientific fields has also gained attention. For instance, MARGOT was used to identify argument components within clinical trials, highlighting its

potential in healthcare [6]. Summaries from computational linguistics and biomedicine were extracted to create a new dataset, enabling them to evaluate the performance of a machine learning model [59]. Gathering scientific literature on specific policy goals, annotated the data, and assessed it using argumentation mining methods [60]. Took on the task of detecting arguments and relationships between arguments within scientific publications [20].

In another interesting study, a manual analysis of 19th-century was performed on philosophical texts to teach a machine learning model for argumentation mining [61]. After training, they consumed the model to automatically analyse the texts, finding it capable of identifying certain features accurately, though they noted that more training data would be beneficial for stronger results.

REQUIRED MODELS OF ML USED FOR AM

Machine learning models used the methods used in argumentation mining are categorized in three groups: Similar to the grouping of applications, the machine learning models and methods used in each paper were categorized into one or more of the following groups: deep learning, neural networks, and traditional machine learning. Figure 2 illustrates the tempo of each machine learning type utilized in the reviewed papers for argumentation mining. The models and methods include BiLSTM, Logistic Regression, SVM, SVM-HMM, RuBERT, Random Forest, XGBoost, LSTM BERT, Naive Bayes, HAN, CNN, FastText, Ernie 2.0, and SciBERT.

Table 2. Shows the Proposed Model with MLT, Procedure, Result and Limitations

Author(s)	**Machine Learning Techniques Used**	**Dataset Used**	**Proposed Model**	**Results**	**Limitations**
Cabessa et al., 2025	Large Language Models (LLMs)	Argumentative datasets like PE, AbstRCT, and CDCP	fine-tuned large language model	achieves state-of-the-art results across all AM sub-tasks and datasets	Limited generalization across diverse domains, dataset dependency
Chistova et al., 2024	End-to-end neural models	Custom dataset with varying rhetorical structures	End-to-end model for AM across diverse rhetorical structures	Achieved state-of-the-art performance across multiple rhetorical structures	Model complexity and generalization to unseen rhetorical structures remain challenging

Zhao et al., 2024	Large language models	Argumentative datasets like GLUE, SQuAD, WikiText	Large Language Model	Provides a detailed survey of various LLM architectures, performance benchmarks, and applications	High computational costs, challenges in evaluating generalization across domains, evolving model architectures
Habernal et al., 2024	Supervised learning, transformer-based architectures	Legal court decision datasets	BERT Model	High accuracy in extracting and linking arguments in legal contexts	Limited adaptability to non-legal domains and heavy reliance on annotated legal corpora
Mancini et al., 2024	Multimodal learning approaches	Multimodal datasets (text, image, speech)	BERT & T-5 Model	Demonstrated superior performance in multimodal argument mining tasks	Requires high-quality multimodal datasets and complex model integration
Gorur et al., 2024	Large Language Models (LLMs)	General argument mining datasets	Large Language Model	LLMs performed comparably to task-specific models in some scenarios	Struggled with nuanced relation detection and required fine-tuning for specific tasks
Benjamin Schiller et al.,2023	Deep Learning Techniques	Diverse argument mining datasets	BERT Model	Found that topic diversity impacts model performance more than dataset size.	Difficult to generalize findings to datasets with low topic variability.
Mushtaq et al., 2023	Transfer Learning	Argumentative datasets Persuasive Essays	BERT - MINUS	High performance in argument mining tasks, especially in domain adaptation	Potential domain-specific biases, requires fine-tuning for new domains
Rafael Mestre et al., 2023	Pre-trained Language Models, Audio Embedding Techniques, NLP	Political debate transcripts, audio-annotated datasets	BERT Model (Augmented Pre-trained Language Models)	Enhanced argument recognition in political debates using multimodal data.	Increased computational cost due to audio processing.
Muhammad Tawsif Sazid et al.,2022	Deep Learning, Unified Representation Models	Annotated student persuasive essay datasets	BERT Model	Improved accuracy in identifying argument components in persuasive essays.	Performance may degrade with noisy student writing.
Thiemo Wambsganss et al., 2022	NLP, Reinforcement Learning,	English learner essays, annotated	ALEN App for Writing Support	Improved persuasive writing and	Limited effectiveness

	Adaptive Learning Systems	persuasive writing datasets		learning outcomes for English language learners.	for advanced learners.
Fishcheva et al., 2021	Traditional ML and Deep Learning Techniques	Russian argumentative texts, Public corpora	SVM and BERT Model	Showed potential for combining traditional ML and deep learning models for multilingual texts.	Limited access to high-quality Russian argument mining datasets.
Aris Fergadis et al.,2021	NLP, Argument Mining, Supervised Learning	Scientific literature on sustainability	BERT and LSTM Model (Argument Mining Framework for Sustainability)	Improved understanding of arguments in sustainability literature.	Limited to domain-specific datasets.
Pablo Accuosto et al.,2021	Computational Linguistics, Biomedical NLP, Argument Mining	Scientific biomedical articles	BERT Model (Argument Mining in Biomedical Literature)	Enhanced argument identification in biomedical contexts, aiding systematic reviews.	Challenges with domain-specific jargon.
Huihui Xu et al.,2020	Deep neural network techniques	Legal case datasets	BERT Model	Improved quality of legal text summaries, better highlighting key arguments in cases.	Requires high-quality annotated legal corpora. Limited generalizability to other domains or complex legal cases.
Tuhin Chakrabarty et al.,2020	Neural Networks	Online discussion datasets (e.g., forums, debates)	BERT Model	Enhanced identification of persuasive arguments in online discussions.	Struggles with informal and diverse online text structures.
Stefano Menini et al.,2018	NLP, Classification Models	Political speeches dataset (annotated corpus)	Argumentation Analysis Framework	Enhanced identification of argumentative structures in political discourse. Improved understanding of political rhetoric strategies.	Limited applicability to informal political debates or speeches. Performance may vary with speech style and linguistic diversity.

Ahmet Aker et al.,2017	Supervised Learning, Feature Engineering, Classifier Evaluation	Public debates, essays	SVM Model (Comparative Evaluation Framework)	Identified effective features and classifiers for argument mining tasks, with SVM performing well for most scenarios.	Feature generalizability is limited across domains. Dependency on dataset quality and relevance to argument mining tasks.
Marco Lippi et al.,2017	Supervised Learning	Legal texts, Debate datasets	SVM Model (Argument Mining Framework)	High performance in identifying argument structures in controlled datasets.	Limited scalability to noisy, real-world text data.
Ivan Habern et al.,2016	Neural Networks, Feature Engineering	Web discourse datasets, Crowdsourced texts	SVM and CNN Model	Demonstrated efficacy of using web discourse for argument mining.	Performance varies significantly with data quality and domain. Difficulty in processing informal language.
John Lawrence et al.,2015	Supervised learning method and NLP techniques	Debate databases, Annotated corpora	SVM Model (Integrated Argument Mining Approach)	Improved performance in handling complex argument structures.	Increased computational complexity. Challenges in integrating diverse methodologies and handling ambiguities in argument structures.

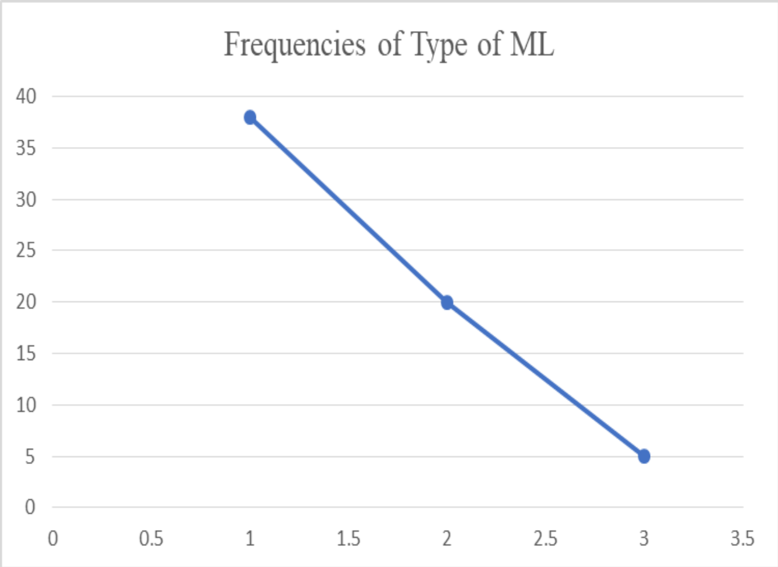

Figure 3: Reviewed research shows the frequency for AM for ML types

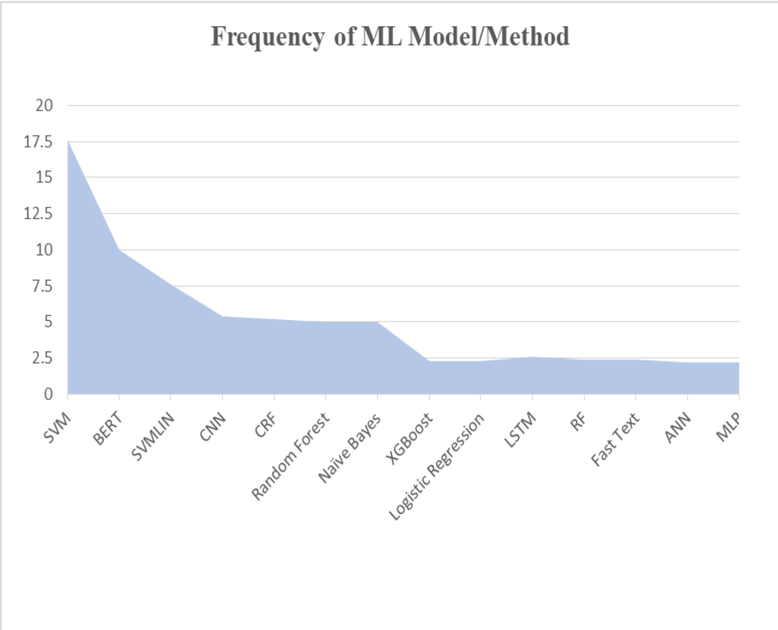

Figure 4: Reviewed AM studies frequency for ML methods

Some specific models and methods are combination of traditional and advanced techniques like LSTM, XGBoost, Random Forest, SVM, SVM-HMM, BERT, BiLSTM, Naive Bayes are traditional methods helps in both Classification and Regression, RuBERT, Hierarchical Attention Network (HAN), Convolutional Neural Network (CNN), SciBERT are deep learning methods, FastText used in neural network created by Facebook workers in 2016, Logistic Regression, and Ernie 2.0.

TIME-BASED INVESTIGATION OF THE GATHERED RESEARCH

Investigate the growth of argumentation mining applications and machine learning techniques, following figures denoted the Frequency and Timeline of various application. Figure 4 shows the most researched accrued in between 2018 to 2025.

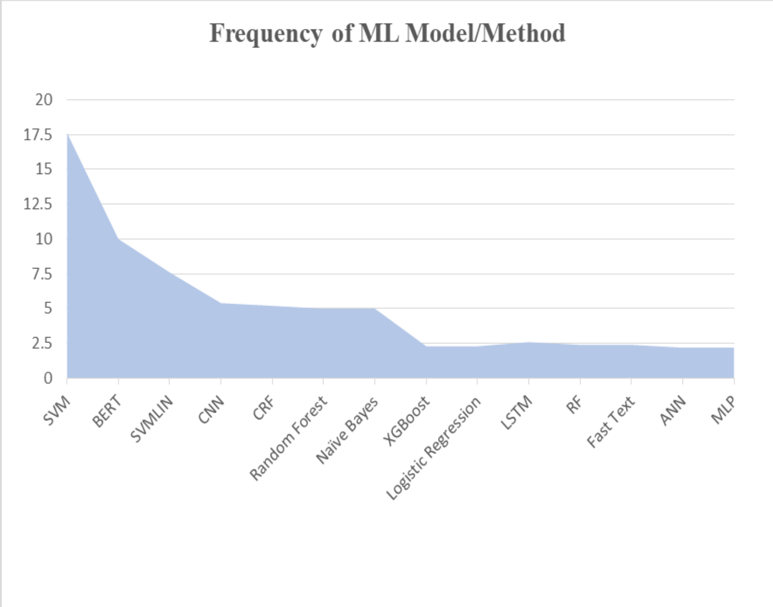

Figure 5: Research gathered from 2011 to 2024 for frequency

Many of the applications suggest the rapid growth in argumentation mining field. Figure 5 represent the timeline of machine learning model or methods.

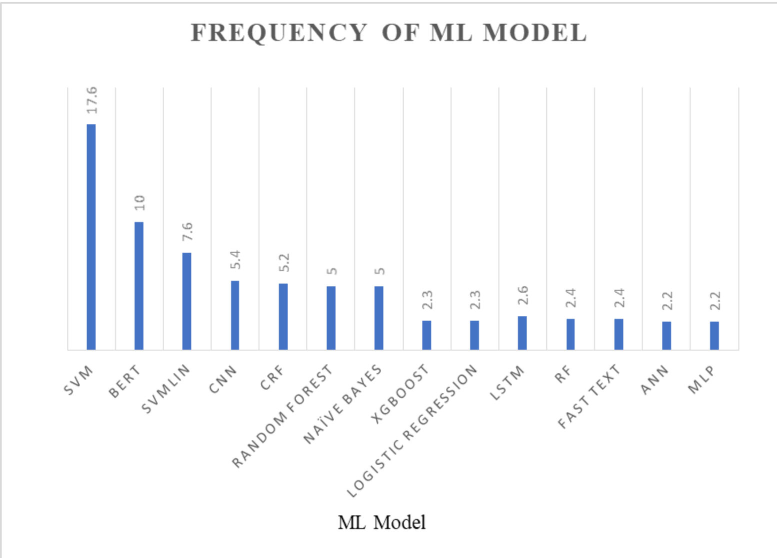

Figure 6: Research collected from 2011 to 2024 for prevalence of machine learning models and methods

Many deep learning models are like SVM, CNN, BiLSTM and BERT are very popular, aligning with enhancement in deep learning. Transformer based models specifically like BERT are more used and reflecting their good impact in argumentation mining. Logistic Regression, Maximum Entropy, and Random Forest these Models are less prominent in recent research. SVM is very much effective so it is continuing to be use because of its effectiveness rather than other traditional models. There is a much shift towards deep learning, suggesting a broader trend in the field favouring these advanced methods.

CONNECTIONS BETWEEN MACHINE LEARNING MODELS AND ARGUMENTATION MINING APPLICATIONS:

To collect the information on applications and machine learning, Figure 6 illustrates the connections between various applications of argumentation mining and their use of deep learning techniques, traditional machine learning, and neural networks. In this diagram, the thickness of each arrow signifies the volume of research that service the specified type of machine learning for each use. Figure 7, on the other hand, depicts the relationships between these applications and the particular machine learning algorithms or models used.

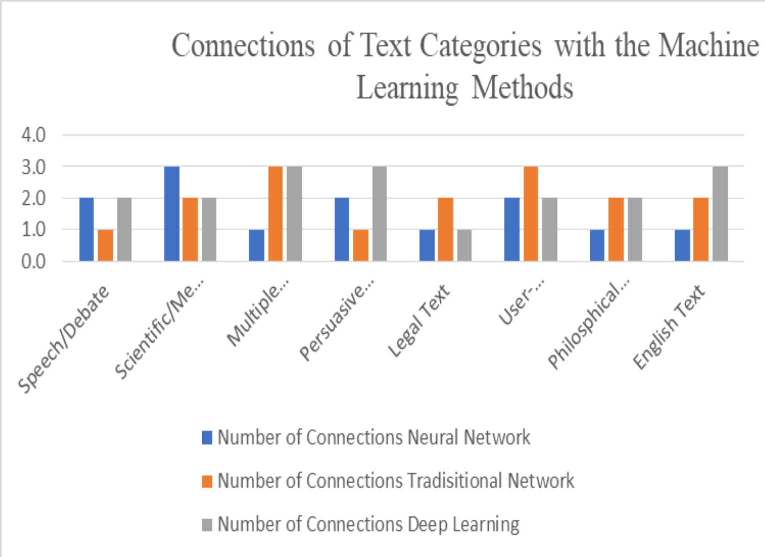

Figure 7: AM applications and type of ML relationship

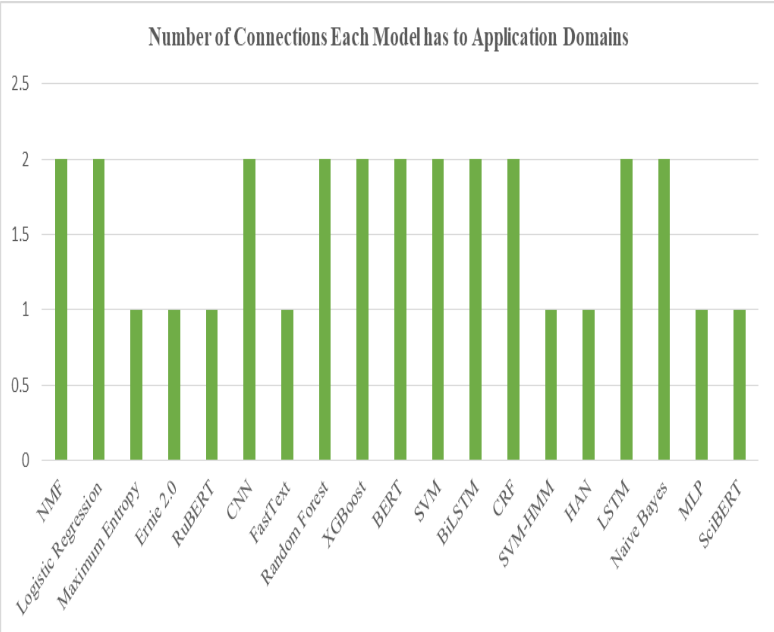

Figure 8: Applications of AM and ML methods relationship

In argumentation mining for debates, different studies have used a range of machine learning approaches. One study utilized neural networks, another used deep learning, and four others applied traditional machine learning techniques. For example, Slonim used neural networks for Project Debater, although the exact model wasn't specified. SVM, a traditional machine learning method can be used, to analyze political debates [10]. Deep learning models like BERT, BiLSTM, and CNN were applied, when working with SVM [54][62] [53].

When it comes to argumentation mining in scientific texts or medical, two studies used traditional machine learning, while three trusted on deep learning models. MARGOT, was considered as SVM-HMM, a traditional machine learning approach [25]. BERT a large-scale model was applied, joint by SciBERT and BiLSTM [59] [60]. Deep learning can also be applied, employing CNN, BiLSTM, CRF and SciBERT [20].

For persuasive essays, four studies chose traditional machine learning methods, while three opted for deep learning. Applied study on SVM and Random Forest was conducted, Persing used the Maximum Entropy model, and also used

SVM [47][51][35]. On the deep learning side, BiLSTM was applied, whereas some scholars preferred BERT [27][49][48].

Traditional machine learning, is use by three papers from non-English text application. Within the non-English text application, four studies are utilized deep learning methods. HAN and LSTM models and BERT model, both deep learning models can be used for the purpose [29][28].SVM, a traditional machine learning approach can be employed. Both deep learning model (RuBERT) and traditional machine learning (SVM) can be incorporated [55]. Additionally, a combination of conditional random field and BilSTM can also be applied [63].

In the legal text domain, one paper employed a neural network, three applied traditional machine learning, and two used deep learning techniques. CNN, Random Forest, and Fast Text were integrated, comprising, all three categories: neural networks, traditional machine learning, and deep learning [57]. SVM, is utilized as a traditional machine learning method, just as BERT is applied as a deep learning model [58][23].

Sixteen papers on user-generated text used traditional machine learning, two employed deep learning, and one applied a neural network. In 2017, SVM-HMM (Hidden Markov Model) was used for user-generated texts, and in 2015, SVM was used for weighting portals [24]. MARGOT was used, which implemented SVM-HMM, while applied both SVM and Random Forest[64] [30][35]. SVM was employed by [38] [31]. Logistic regression was used by [32][39]. NMF was applied, Rocha and Cardoso used the Maximum Entropy Model and utilized CRF [32][31]. BERT was used, by incorporating XG-Boost with it [32][34]. Finally, MLP was used [41].

For the paper that use various applications six used traditional machine learning, and four in simple words used deep learning. The paper that makes use of argumentation mining to philosophical texts, created by applied Naive Bayes, a traditional machine learning model [12]. SVM-HMM was used by [10].Naive Bayes was applied by [61]. SVM was used by [46][9]. Ernie 2.0 was used by [43]. BERT was used by [44]. SVM and the Maximum Entropy Model was applied by [56]. CNN was applied by [52]. CRF and BiLSTM were used by [45].

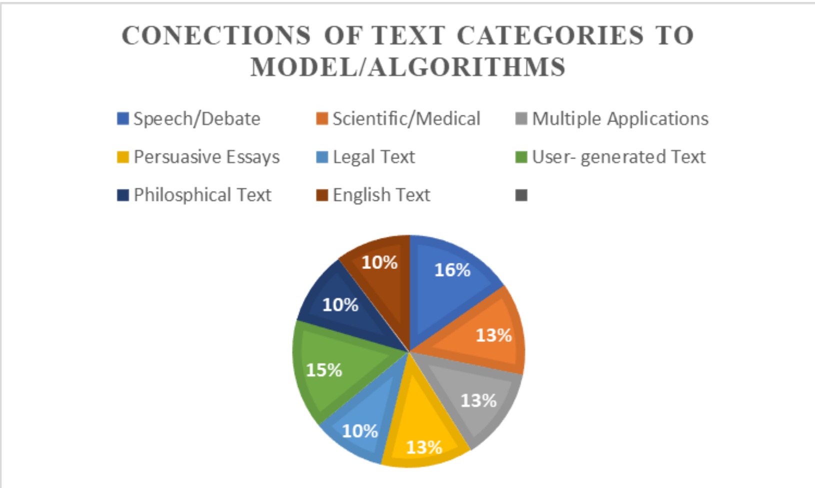

Figure 9: AM applications and ML model's relationship chart

Figure 8 represents a refined type of Figure 7, with several categories streamlined. The SVM and SVMHMM models were grouped under "SVM-based methods," while BERT, SciBERT, and RuBERT were consolidated as "BERT-based models." Similarly, LSTM and BiLSTM were combined into the "LSTM-based models" category. Following this rearrangement, each application displayed the following model usage patterns. The user-generated text application relied heavily on SVM-based methods, utilizing them eight times, along with two situations of BERT-based models. In the non-English text application, LSTM-based models and BERT-based models were each used twice, with a single instance of SVM-based methods. For the Scientific/Medical application, BERT-based models materialized three times, while LSTM-based models were used twice and SVM-based methods once.

The "Multiple applications" category saw four instances of SVM-based methods, one of LSTM-based models, and one of BERT-based models. In the analysis of Persuasive Essays, a BERT-based model was used once, with SVM-based methods and LSTM-based models each used twice. The legal text category included one use of a BERT-based model and two of SVM-based methods. Finally, in the speech and debate category, there was one instance of a BERT-based model, three of LSTM-based models and one of a SVM-based method. This refined categorization highlights the separation of model types across the different applications, showing how each was applied based on certain needs and characteristics of the texts.

FUTURE WORK

The specific methodology, TANL (Translation between Augmented Natural Languages) for text-to-text generation, has proven to be vital for specific purposes such as relation extraction, named entity recognition, semantic role labelling, and reference resolution [6]. The study effectively focused on a more appropriate and useful pertained encoder-decoder model, i.e., T5 (Text-to-Text Transfer Transformer).

Based on recent studies, T5 has proven to be more effective for analysing sentence structure, making it more suitable for sentence-level Rhetorical Structure Theory (RTS) parsing, which is a form of text-to-text generation [2].

This research shifts the focus of document structure analysis toward text generation rather than merely analyzing arguments in Natural Language, particularly with respect to Argumentation Mining. As a result, we propose that using T5 for text-to-text generation tasks within the TANL framework is more appropriate.

CONCLUSION

Argumentation mining has been applied to analyse a variety of text types including user-generated content, English texts, persuasive essays, debates, legal documents, scientific, medical texts, philosophical works, and systems accessible to the general public. LLM, Text-to-Text generation, CNN, SVM, BERT and BiLSTM are commonly used and also more frequently applied deep learning models. The most recent BERT model adopts the Transformer structure, which contains multiple encoded layers. It has demonstrated its benefits on various computational tasks, including inference and semantic understanding, NLP and text segmentation, classification, etc. SVM based models are often used for small user generated text and debate, while BERT and T5 based models are popular or useful for English text, legal text, mathematical equation, multilingual text, medical text and scientific text.

This research recognize correlation between specific applications and machine learning models offers understanding into model preferences for different type of text data. This review suggest that researchers should concentrate more Focusing on exploring less common applications, such as scientific, legal and medical texts, while employing a wide variety of machine learning techniques for these applications. Argumentation mining can be more beneficial by research on BERT but SVM methods are still useful for small user generated text, when dataset was small. BERT base model can understand the context but SVM faces struggles to understand meaning of a sentence in context. Further research should compare deep learning and traditional methods for user generated text and English text which prefer deep learning even though it has been used since longer.

REFRENCES

[1] J. Cabessa, H. Hernault, and U. Mushtaq, “Jérémie,” no. 2024, pp. 6624–6635, 2025.

[2] E. Chistova, “End-to-End Argument Mining over Varying Rhetorical Structures,” *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 3376–3391, 2023, doi: 10.18653/v1/2023.findings-acl.209.

[3] U. Mushtaq and J. Cabessa, “Argument Mining with Modular BERT and Transfer Learning,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2023-June, pp. 1–8, 2023, doi: 10.1109/IJCNN54540.2023.10191968.

[4] I. Habernal *et al.*, *Mining legal arguments in court decisions*, vol. 32, no. 3. Springer Netherlands, 2024. doi: 10.1007/s10506-023-09361-y.

[5] C. Stab and I. Gurevych, “Parsing argumentation structures in persuasive essays,” *Comput. Linguist.*, vol. 43, no. 3, pp. 619–659, 2017, doi: 10.1162/COLI_a_00295.

[6] M. Kawarada, T. Hirao, W. Uchida, and M. Nagata, “Argument Mining as a Text-to-Text Generation Task,” *EACL 2024 - 18th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Conf.*, vol. 1, no. 2, pp. 2002–2014, 2024.

[7] D. Gorur, A. Rago, and F. Toni, “Can Large Language Models perform Relation-based Argument Mining?,” 2024, [Online]. Available: http://arxiv.org/abs/2402.11243

[8] W. X. Zhao *et al.*, “A Survey of Large Language Models,” pp. 1–140, 2024.

[9] M. Lippi and P. Torroni, “Context-independent claim detection for argument mining,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2015-Janua, pp. 185–191, 2015.

[10] M. Lippi and P. Torroni, “Argument mining from speech: Detecting claims in political debates,” *30th AAAI Conf. Artif. Intell. AAAI 2016*, pp. 2979–2985, 2016, doi: 10.1609/aaai.v30i1.10384.

[11] R. Emilia, "Argument Mining : a Machine Learning Perspective," no. September, 2017, doi: 10.1007/978-3-319-28460-6.

[12] J. Lawrence and C. Reed, "Combining Argument Mining Techniques," *2nd Work. Argumentation Mining, ArgMining 2015 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-HLT 2015 - Proc.*, pp. 127–136, 2015, doi: 10.3115/v1/w15-0516.

[13] M. F. Moens, "Argumentation mining: How can a machine acquire common sense and world knowledge?," *Argument Comput.*, vol. 9, no. 1, pp. 1–4, 2018, doi: 10.3233/AAC-170025.

[14] S. Aftan and H. Shah, "A Survey on BERT and Its Applications," *20th Int. Learn. Technol. Conf. L T 2023*, pp. 161–166, 2023, doi: 10.1109/LT58159.2023.10092289.

[15] O. Coban, M. Yağanoğlu, and F. Bozkurt, "Domain Effect Investigation for Bert Models Fine-Tuned on Different Text Categorization Tasks," *Arab. J. Sci. Eng.*, vol. 49, no. 3, pp. 3685–3702, 2024, doi: 10.1007/s13369-023-08142-8.

[16] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Naacl-Hlt 2019*, no. Mlm, pp. 4171–4186, 2018, [Online]. Available: https://aclanthology.org/N19-1423.pdf

[17] A. Galassi, M. Lippi, and P. Torroni, "Multi-Task Attentive Residual Networks for Argument Mining," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 1877–1892, 2023, doi: 10.1109/TASLP.2023.3275040.

[18] D. Gemechu, R. Ruiz-Dolz, and C. Reed, "ARIES: A General Benchmark for Argument Relation Identification," no. ArgMining, pp. 1–14, 2024, doi: 10.18653/v1/2024.argmining-1.1.

[19] N. Das, V. Choudhary, V. V. Saradhi, and A. Anand, "End-to-End Argument Mining as Augmented Natural Language Generation," pp. 1–12, 2024, [Online]. Available: http://arxiv.org/abs/2406.08606

[20] A. Binder, B. Verma, and L. Hennig, "Full-text argumentation mining on scientific publications," no. 2017, 2022, [Online]. Available: http://arxiv.org/abs/2210.13084

[21] M. Sofi, M. Fortier, and O. Cocarascu, "A Robustness Evaluation Framework for Argument Mining," *Proc. 9th Work. Argument Min.*, pp. 171–180, 2022, [Online]. Available: https://aclanthology.org/2022.argmining-1.16

[22] I. Beltagy, K. Lo, and A. Cohan, "SCIBERT: A pretrained language model for scientific text," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 3615–3620, 2019, doi: 10.18653/v1/d19-1371.

[23] G. Zhang, P. Nulty, and D. Lillis, "Enhancing Legal Argument Mining with Domain Pre-training and Neural Networks," *J. Data Min. Digit. Humanit.*, vol. NLP4DH, pp. 1–15, 2022, doi: 10.46298/jdmdh.9147.

[24] N. L. Green, "Argumentation mining in scientific discourse," *CEUR Workshop Proc.*, vol. 2048, no. April 2016, pp. 7–13, 2017, doi: 10.1162/COLI.

[25] T. Mayer, E. Cabrio, M. Lippi, P. Torroni, and S. Villata, "Argument mining on clinical trials," *Front. Artif. Intell. Appl.*, vol. 305, pp. 137–148, 2018, doi: 10.3233/978-1-61499-906-5-137.

[26] Rahel & Solomon, "Proceeding of the 2 nd Deep Learning Indaba-X Ethiopia Conference 2021 ' Strengthening Awareness and Application of Machine Learning and Artificial Intelligence in Ethiopia ,'" p. 50, 2022.

[27] M. T. Sazid and R. E. Mercer, "A Unified Representation and Deep Learning Architecture for Argumentation Mining of Students' Persuasive Essays," *CEUR Workshop Proc.*, vol. 3205, no. November, pp. 86–97, 2021.

[28] O. Toledo-Ronen, M. Orbach, Y. Bilu, A. Spector, and N. Slonim, "Multilingual argument mining: Datasets and analysis," *Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2020*, pp. 303–317, 2020, doi: 10.18653/v1/2020.findings-emnlp.29.

[29] D. Suhartono, A. P. Gema, S. Winton, T. David, M. I. Fanany, and A. M. Arymurthy, "Argument annotation and analysis using deep learning with attention mechanism in Bahasa Indonesia," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00364-z.

[30] P. Basile *et al.*, "Argument Mining on Italian News BlogsArgument Mining on Italian News Blogs To cite this version : Argument Mining on Italian News Blogs," 2016.

[31] R. Swanson, B. Ecker, and M. Walker, "Argument mining: Extracting arguments from online dialogue," *SIGDIAL 2015 - 16th Annu. Meet. Spec. Interes. Gr. Discourse Dialogue, Proc. Conf.*, no. September, pp. 217–226, 2015, doi: 10.18653/v1/w15-4631.

[32] M. Dusmanu, E. Cabrio, and S. Villata, "Argument mining on Twitter: Arguments, facts and sources," *EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 2317–2322, 2017, doi: 10.18653/v1/d17-1245.

[33] T. Chakrabarty, C. Hidey, S. Muresan, K. McKeown, and A. Hwang, "AmperSand: Argument mining for

persuasive online discussions," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 2933–2943, 2019, doi: 10.18653/v1/d19-1291.

[34] N. Tran and D. Litman, "Multi-task Learning in Argument Mining for Persuasive Online Discussions," *8th Work. Argument Mining, ArgMining 2021 - Proc.*, pp. 148–153, 2021, doi: 10.18653/v1/2021.argmining-1.15.

[35] A. Aker *et al.*, "What works and what does not: Classifier and feature analysis for argument mining," *EMNLP 2017 - Proc. 4th Work. Argument Mining, ArgMining 2017*, pp. 91–96, 2017, doi: 10.18653/v1/w17-5112.

[36] P. Sobhani, D. Inkpen, and S. Matwin, "From Argumentation Mining to Stance Classification," *2nd Work. Argumentation Mining, ArgMining 2015 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-HLT 2015 - Proc.*, pp. 67–77, 2015, doi: 10.3115/v1/w15-0509.

[37] L. V. C. Coelho, "A GPT-Based Data Augmentation Approach for enhancing Argumentation Mining in Multi-party Dialogues," 2024, [Online]. Available: https://teses.pythonanywhere.com/media/teses/DEISI89_Luiza_Coelho.pdf

[38] G. Rocha and H. Lopes Cardoso, "Towards a relation-based argument extraction model for argumentation mining," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10583 LNAI, no. September 2017, pp. 94–105, 2017, doi: 10.1007/978-3-319-68456-7_8.

[39] O. Cocarascu and F. Toni, "Detecting deceptive reviews using Argumentation," *ACM Int. Conf. Proceeding Ser.*, vol. 29-30-Augu, 2016, doi: 10.1145/2970030.2970031.

[40] C. Sardianos, I. M. Katakis, G. Petasis, and V. Karkaletsis, "Argument Extraction from News," *2nd Work. Argumentation Mining, ArgMining 2015 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL-HLT 2015 - Proc.*, pp. 56–66, 2015, doi: 10.3115/v1/w15-0508.

[41] A. Lytos, T. Lagkas, P. Sarigiannidis, V. Argyriou, and G. Eleftherakis, "Modelling argumentation in short text: A case of social media debate," *Simul. Model. Pract. Theory*, vol. 115, 2022, doi: 10.1016/j.simpat.2021.102446.

[42] "This is the final peer-reviewed accepted manuscript of : Lippi , M . and P . Torroni . 2016 . ' MARGOT : A Web Server for Argumentation Mining .' Expert Systems with Applications 65 : 292-303 . The final is available online at : Rights / License : The ter," vol. 41, no. 2016, pp. 292–303, 2025.

[43] B. Schiller, J. Daxenberger, and I. Gurevych, "Diversity Over Size: On the Effect of Sample and Topic Sizes for Argument Mining Datasets," 2022, [Online]. Available: http://arxiv.org/abs/2205.11472

[44] T. Wambsganss, N. Molyndris, and M. Söllner, "transfer learning in argumentatioUnlockingn mining: A domain-independent modelling approach," *Proc. 15th Int. Conf. Bus. Inf. Syst. 2020 "Developments, Oppor. Challenges Digit. WIRTSCHAFTSINFORMATIK 2020*, no. Ml, 2020, doi: 10.30844/wi_2020_c9.

[45] C. Schulz, S. Eger, J. Daxenberger, T. Kahse, and I. Gurevych, "Multi-task learning for argumentation mining in low-resource settings," *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 2, no. 1, pp. 35–41, 2018, doi: 10.18653/v1/n18-2006.

[46] N. Rooney, H. Wang, and F. Browne, "Applying kernel methods to argumentation mining," *Proc. 25th Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS-25*, pp. 272–275, 2012.

[47] H. A. Kusmantini, I. Asror, and M. A. Bijaksana, "Argumentation mining: Classifying argumentation components with Partial Tree Kernel and Support Vector Machine for constituent trees on imbalanced persuasive essay," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012009.

[48] H. Wang, Z. Huang, Y. Dou, and Y. Hong, "Argumentation Mining on Essays at Multi Scales," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 5480–5493, 2020, doi: 10.18653/v1/2020.coling-main.478.

[49] S. Eger, J. Daxenberger, and I. Gurevych, "Neural end-to-end learning for computational argumentation mining," *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 11–22, 2017, doi: 10.18653/v1/P17-1002.

[50] I. Persing and V. Ng, "Unsupervised argumentation mining in student essays," *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, no. May, pp. 6795–6803, 2020.

[51] T. Wambsganss, A. Caines, and P. Buttery, "ALEN App: Persuasive Writing Support To Foster English Language Learning," *BEA 2022 - 17th Work. Innov. Use NLP Build. Educ. Appl. Proc.*, no. Bea, pp. 134–140, 2022.

[52] R. Bouslama, R. Ayachi, and N. Ben Amor, "Using Convolutional Neural Network in Cross-Domain Argumentation Mining Framework," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11940 LNAI, pp. 355–367, 2019, doi: 10.1007/978-3-030-35514-2_26.

[53] N. Naderi, "Scholarship at UWindsor Scholarship at UWindsor Argumentation Mining in Parliamentary Discourse Argumentation Mining in Parliamentary Discourse Argumentation Mining in Parliamentary Discourse," pp. 1–9, 2016, [Online]. Available:

https://scholar.uwindsor.ca/ossaarchivehttps://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/160

[54] R. Mestre, S. E. Middleton, M. Ryan, M. Gheasi, T. J. Norman, and J. Zhu, "Augmenting pre-trained language models with audio feature embedding for argumentation mining in political debates," *EACL 2023 - 17th Conf. Eur. Chapter Assoc. Comput. Linguist. Find. EACL 2023*, no. Section 4, pp. 274–288, 2023, doi: 10.18653/v1/2023.findings-eacl.21.

[55] F. I. N, G. V S, K. E. V, Ф. И Н, Г. В С, and К. Е В, "Traditional Machine Learning and Deep Learning Models for Argumentation Mining in Russian Texts Модели традиционного машинного обучения и глубокого обучения для анализа аргументации русскоязычных текстов," pp. 1–13, 2021, doi: 10.28995/2075-7182-2021-20-XX-XX.

[56] R. Mochales and M. F. Moens, "Argumentation mining," *Artif. Intell. Law*, vol. 19, no. 1, pp. 1–22, 2011, doi: 10.1007/s10506-010-9104-x.

[57] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," *IEEE Access*, vol. 9, pp. 156043–156070, 2021, doi: 10.1109/ACCESS.2021.3129786.

[58] M. Lippi, F. Lagioia, G. Contissa, G. Sartor, and P. Torroni, "Claim Detection in Judgments of the EU Court of Justice," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10791, pp. 513–527, 2018, doi: 10.1007/978-3-030-00178-0_35.

[59] P. Accuosto, M. Neves, and H. Saggion, "Argumentation mining in scientific literature: From computational linguistics to biomedicine," *CEUR Workshop Proc.*, vol. 2847, pp. 20–36, 2021.

[60] A. Fergadis, D. Pappas, A. Karamolegkou, and H. Papageorgiou, "Argumentation Mining in Scientific Literature for Sustainable Development," *8th Work. Argument Mining, ArgMining 2021 - Proc.*, vol. 2, pp. 100–111, 2021, doi: 10.18653/v1/2021.argmining-1.10.

[61] J. Lawrence, C. Reed, C. Allen, S. McAlister, A. Ravenscroft, and D. Bourget, "Mining Arguments from 19th Century Philosophical Texts Using Topic Based Modelling," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 79–87, 2014, doi: 10.3115/v1/w14-2111.

[62] S. Menini, E. Cabrio, S. Tonelli, and S. Villata, "Never retreat, never retract: Argumentation analysis for political speeches," *32nd AAAI Conf. Artif. Intell. AAAI 2018*, pp. 4889–4896, 2018, doi: 10.1609/aaai.v32i1.11920.

[63] S. Eger, J. Daxenberger, C. Stab, and I. Gurevych, "Cross-lingual argumentation mining: Machine translation (and a bit of projection) is all you need!," *COLING 2018 - 27th Int. Conf. Comput. Linguist. Proc.*, no. 1, pp. 831–844, 2018.

[64] M. Passon, M. Lippi, G. Serra, and C. Tasso, "Predicting the Usefulness of Amazon Reviews Using Off-The- Shelf Argumentation Mining," *EMNLP 2018 - Proc. 5th Work. Argument Min.*, pp. 35–39, 2018, doi: 10.18653/v1/w18-5205.