UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Diego Rossi

PRED-INTER: Automatic Prediction of Pedagogical Interventions

Juiz de Fora
2022

Diego Rossi

PRED-INTER: Automatic Prediction of Pedagogical Interventions

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Victor Ströele de Andrade Menezes

Coorientadores: Profª. Dra. Fernanda Cláudia Alves Campos e Prof. Dr. Jairo Francisco de Souza

Juiz de Fora

2022

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Rossi, Diego.
PRED-INTER: Automatic Prediction of Pedagogical Interventions / Diego Rossi. – 2022.
91 f. : il.

Orientador: Victor Ströele de Andrade Menezes
Coorientadores: Fernanda Cláudia Alves Campos e Jairo Francisco de Souza
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2022.

1. Pedagogical Intervention. 2. Automatic Text Classification. 3. Semantic Rules. I. Ströele, Victor, orient. II. Campos, Fernanda, coorient. III. Francisco de Souza, Jairo, coorient. IV. Título.

Diego Rossi

PRED-INTER: Automatic Prediction of Pedagogical Interventions

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Ciência da Computação.

Aprovada em 17 de outubro de 2022.

BANCA EXAMINADORA

Prof. Dr. Victor Ströele de Andrade Menezes - Orientador
Universidade Federal de Juiz de Fora

Profª. Dra. Fernanda Cláudia Alves Campos - Coorientadora
Universidade Federal de Juiz de Fora

Prof. Dr. Jairo Francisco de Souza - Coorientador
Universidade Federal de Juiz de Fora

Profª. Dra. Regina Maria Maciel Braga
Universidade Federal de Juiz de Fora

Prof. Dr. Santi Caballe Llobet

Universitat Oberta de Catalunya

Juiz de Fora, 11/10/2022.

Documento assinado eletronicamente por **Victor Stroele de Andrade Menezes**, **Professor(a)**, em 10/11/2022, às 10:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Santiago Caballe Llobet**, **Usuário Externo**, em 11/11/2022, às 09:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Regina Maria Maciel Braga Villela**, **Professor(a)**, em 11/11/2022, às 19:08, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jairo Francisco de Souza**, **Professor(a)**, em 11/11/2022, às 20:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernanda Claudia Alves Campos**, **Coordenador(a)**, em 11/11/2022, às 20:20, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **0992400** e o código CRC **4EB1963F**.

Dedico este trabalho a todos os que me ajudaram ao longo desta caminhada.

AGRADECIMENTOS

A Deus, por me permitir alcaçar essa conquista.

A minha esposa Marina, pelo esforço, incentivo e cooperação para que eu pudesse me dedicar ao máximo desde o início da minha trajetório na pós-graduação.

Ao meu filho Heitor, que se mostrou muito participativo e compreensivo nos meus momentos de estudo.

A minha Mãe Dulce, que sempre se mostrou como exemplo do poder transformador que a educação proporciona.

Ao meu orientador, Victor Ströele, por assumir o compromisso de orientar este trabalho, sem medir esforços para que o melhor pudesse ser feito. Sendo sempre bastante compreensivo e incentivador, almejando sempre boas experiências acadêmicas.

A minha coorientadora, professora Fernanda Alves Campos, pela imensa contribuição, agregando muito com sua experiência e conhecimento.

Ao meu coorientador, professor Jairo Francisco de Souza, pelos questionamentos que foram importantes para a construção e aperfeiçoamento deste trabalho.

Aos professores do Programa de Pós-graduação em Ciência da Computação, pelos ensinamentos, parceiria e incentivo.

Aos amigos de trabalho, pela compreensão e apoio que foi fundamental para que eu pudesse realizar este trabalho.

"Educar é mostrar a vida a quem ainda não a viu." (Rubem Alves).

RESUMO

Acompanhar estudantes em ambiente virtual de aprendizagem para verificar quem necessita de ajuda é uma tarefa que demanda bastante tempo. Considerando que em muitos casos o número de alunos por tutor é elevado, essa tarefa acaba se tornando inviável. Durante a pesquisa foram encontrados trabalhos que auxiliam tutores, porém muitas abordagens utilizam dados que estão vinculados ao ambiente virtual de aprendizagem ou ao curso, o que dificulta encontrar uma solução genérica. Portanto, este trabalho busca preencher essa lacuna propondo contribuir para a detecção automática de intervenção pedagógica, colaborando para minimizar problemas durante o processo de ensino aprendizagem on line. Foi projetada uma arquitetura para gerar automaticamente intervenções pedagógicas, identificando atributos implícitos presentes nas mensagens de alunos postadas através de interações no ambiente virtual de aprendizagem. Com base nos atributos sentimento, urgência e confusão é possível inferir como o aluno estava se sentindo ao postar a mensagem. São então aplicadas regras semânticas que selecionam a intervenção pedagógica adequada para atender o aluno. A dissertação traz três contribuições principais, a primeira é a arquitetura chamada PRED-INTER, desenvolvida para funcionar de maneira autonoma realizando intervenções pedagógicas. Segunda, desenvolvemos modelos preditivos que realizam a classificação dos atributos automaticamente, detectando sentimento, confusão e urgência nas mensagens de alunos. Os modelos foram treinados e avaliados, através de duas diferentes abordagens de deep learning, utilizando os dados do Stanford MOOCPosts Dataset. Terceira, foi desenvolvida uma ontologia capaz de armazenar as postagens dos alunos e seus atributos, que atravez das regras semânticas, detecta a intervenção pedagógica mais adequada. A avaliação da proposta foi feita através das análises dos modelos de classificação textual e da capacidade de identificar as intervenções pedagógicas. Os resultados alcançados pelos modelos gerados são bastante competitivos em comparação a outros trabalhos. Baseando-se no quantitativo de intervenções pedagógicas identificadas os resultados foram satisfatórios, pois mostram que é possível automatizar grande parte das intervenções, mantendo o suporte a alunos e tutores.

Palavras-chave: Intervenção Pedagógica. Classificação automática de texto. Regras semânticas.

ABSTRACT

Accompanying students in a learning management system to verify who needs help is a task that demands a lot of time. Considering that, in many cases, the number of students per tutor is high, this task ends up becoming unfeasible. During the research, we found works that help tutors with their job, but many approaches use data linked to the learning management systems or courses, making it difficult to find a generic solution. This work seeks to fill a gap by proposing to contribute to the automatic detection of pedagogical intervention, which ends up helping to minimize problems during the virtual learning-teaching process. An architecture was designed to identify pedagogical interventions automatically. The architecture recognizes implicit attributes present in student messages posted through interactions in the learning management systems. Then, based on the attributes sentiment, confusion, and urgency, it is possible to infer how the student was feeling when the message was posted. After that, it applies semantic rules that select the appropriate pedagogical intervention for the student. The dissertation brings three main contributions. First, the architecture named PRED-INTER was developed to carry out pedagogical interventions autonomously. Then we developed predictive models that automatically classify the attributes, being capable of detecting sentiment, confusion, and urgency in student's messages. The models were trained and evaluated using two different deep learning approaches, with data from the Stanford MOOCPosts Dataset. Finally, an ontology capable of storing students' posts and their attributes was also developed, which, through semantic rules, detects the suitable pedagogical intervention. The evaluation of the proposal was made through the analysis of textual classification models and the ability to identify pedagogical interventions. The results proved to be reasonable, as the generated models are competitive compared to other works. Based on the number of pedagogical interventions identified, the results were satisfactory, as they showed that it is possible to automate most of the interventions, maintaining support for students and tutors.

Keywords: Pedagogical intervention. Automatic text classification. Semantic rules.

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

NLP	Natural Language Processing
VLE	Virtual Learning Enviroments
CA	Conversational Agent
AI	Artificial Intelligence
RS	Recommendation Systems
ML	Machine Learning
SA	Semantic Analysis
MOOC	Massive Open Online Course
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
LSTM	Long Short-Term Memory
ORSD	Ontology Requirements Specification Document

CONTENTS

1 INTRODUCTION

The popularization and advances in Information Technology (IT) have made it increasingly part of our daily lives. Consequently, due to its prominence and relevance, it is widely used in the educational environment (PANIGRAHI; SRIVASTAVA; PANIGRAHI, 2020), supporting innovative pedagogical practices, and generating new learning spaces (MORENO-GUERRERO et al., 2020). One of the ways it allows people to study at a distance is through IT, which is also a way of democratizing access to education since prices are more attractive due to a leaner structure (PALVIA et al., 2018). Another point that contributes to distance learning is study schedules, which end up making it more flexible and helping the student to have a full-time job.

With the COVID-19 pandemic, new challenges are posed considering remote learning: technologies, evasion, distancing, students and teachers' training, monitoring the evolution of students, among others (KASTRATI et al., 2021). Also, although it is a prominent educational approach, many students drop out of e-Learning courses, reaching high rates of 40% and 80%, and the main reasons are low motivation or lack of support (BAWA, 2016).

According to (FANDIÑO; VELANDIA, 2020; CUEVAS et al., 2018), motivation is one of the main factors that positively affect learning. Certainly, tutors play a very important role in this process, where they are responsible for constantly interacting with the students to motivate and keep them engaged. To make tutoring more agile and efficient, even automating some tasks, it is necessary to identify which students need specific help. Furthermore, to select students who will undergo an intervention, it is necessary to identify implicit characteristics in the messages to provide accurate pedagogical intervention.

These challenges demand that teachers and tutors accompany students agilely, providing a communication environment capable of answering questions and motivating students (MORENO-GUERRERO et al., 2020). So, we believe that assisted education may contribute to the automatization process of student tutoring. In this context, there are two major challenges: the first, which is more explored (ABDI et al., 2020; TAVAKOLI et al., 2020; SALAZAR et al., 2021), is sending recommendations and academic topics to students; The second challenge is to predict learning problems and act to minimize their impacts (TOTI et al., 2020). In this work, the second problem is addressed.

Discussion forums are among the most popular interaction tools offered by Learning Management Systems (LMS) sometimes platforms, often used by students to create a sense of belonging and better understand course topics (CAPUANO; CABALLÉ, 2015). However, students trying to clarify concepts through these forums may not receive the attention they require, and a lack of responsiveness often favors dropout(CAPUANO; CABALLÉ, 2019).

Some computer systems seek to fill this gap through automatic student interactions, like Conversational Agents (CA), Recommendation Systems (RS), Chatbots, etc. Typically, these systems use Artificial Intelligence (AI) approaches, based on rules that shape their behavior when interacting with humans (DEMETRIADIS et al., 2021). Researchers have already employed these computer systems to achieve various educational goals such as tutoring, answering questions, practicing language learning, and developing metacognitive skills (KHANAL et al., 2020).

According to (CAPUANO; CABALLÉ, 2019) it is possible, through natural language processing (NLP) and predictive models, to detect various attributes in post messages, like Sentiment, Post Type, Urgency, and Confusion. This automatic detection of attributes implicit in forum posts becomes fundamental for more precise analysis, being able to discover how the student is feeling. Therefore, this information contributes to the tutor's performance, helping to moderate and plan interventions and, consequently, cooperating with the student's learning process. In this work, the attributes are loaded in the ontology, keeping subjective information related to the student's message, allowing the execution of semantic rules that automatically identify the pedagogical intervention necessary to meet the student's educational moment.

This work presents the Predicting pedagogical intervention (PRED-INTER) based on subjective attributes. It merges Semantic Analysis (SA) and Natural Language Processing (NLP) techniques applied to student forum posts and assignments, combined with ML-based classification techniques. The proposal is to support teachers, tutors, and students on Learning Management Systems, seeking to identify, through implicit attributes in students' messages, those who need help. This work aims to identify the pedagogical intervention necessary to accompany the student in the learning process through the semantic patterns present in the students' posts asynchronously.

Our approach identifies the semantic patterns in student posts using ML-based classification techniques, improving earlier approaches (CAPUANO; CABALLÉ, 2019; BÓBÓ et al., 2019; BRAZ et al., 2019). (CAPUANO; CABALLÉ, 2019) propose in their work a multi-attribute text categorization tool capable of automatically detecting useful information from MOOC forum posts, among them sentiment polarity, level of confusion, and urgency. (BÓBÓ et al., 2019) presents an architecture whose main objective is to identify the student's emotional state by detecting the author's sentiments in texts. A recommendation system, based on the student's emotional state and learning style, sends motivational messages to mitigate dropout. In (BRAZ et al., 2019), it is proposed to create models that can detect the disengagement of students in the class and notify teachers about this behavior, enabling them to intervene in an effective way and make student success possible. Based on concepts presented in previous works, our solution stores the students' posts and their attributes in an ontology, making it possible to make inferences to detect the necessary pedagogical intervention. Attributes allow the system to define what action will be performed and then apply specific dialog patterns.

1.1 RESEARCH QUESTIONS

Two research questions were proposed for this work. They are related to innovations in the service and guidance of students, allowing the realization of pedagogical interventions, helping tutors in learning management systems, and acting from the automatic detection of attributes present in the students' posts on the forums. Based on these attributes, we can infer how the student is feeling. In this manner, the most appropriate pedagogical intervention is identified.

- **RQ1**: Is it possible to improve results of the actual automatic student subjective information detection models?
- **RQ2**: It is possible to use the predicted student subjective information to recommend pedagogical interventions?

1.2 GOALS

This research aims to automatically identify students who need pedagogical intervention during the learning process through their posts in the learning management systems. Based on these interactions, it is necessary to extract subjective information, comprehend how the student is feeling, and then choose the most appropriate pedagogical intervention for that student's educational moment. According to the research questions, the objectives can be subdivided into two specific ones:

- G1: Use modern neural-based NLP models to automatically classify subjective attributes from student messages, allowing the detection of sentiment, confusion, and urgency.
- G2: Define an interpretable decision model to identify the most appopriate intervention based on the predicted attributes from student messages.

1.3 CONTRIBUTIONS

This work is part of a project that aims at carrying out detection and analysis of the students' engagement within the context of online higher education and the e-learning platforms, avoiding school dropout (CAPUANO; CABALLÉ, 2019; BÓBÓ et al., 2019; BRAZ et al., 2019).

This dissertation presents as contributions:

1. An approach to autonomous pedagogical intervention and educational resource recommendations;
2. A computational solution to automatically detect sentiment, urgency, and confusion from students' post messages;
3. An ontology capable of storing the students' posts and their attributes, supported by an inference machine to detect the necessary pedagogical intervention.

1.4 ORGANIZATION

This chapter presented the motivation for the work, research questions, goals, and contributions. Chapter 2 will address the theoretical foundations, exposing the main concepts about techniques, goals, and conversational agents for pedagogical intervention and related work. Chapter 3 presents the work proposal, called PRED-INTER architecture, detailing the automatic text classification, where the pre-processing used to train the models and the knowledge representation is presented, exploring the composition of the ontology. This chapter also presents the implementation details. Chapter 4 reports on development and evaluation. Finally, chapter 5 makes the final considerations, including limitations and future work.

2 THEORETICAL FOUNDATION

This chapter presents the theoretical foundations that provide the conceptual knowledge necessary to understand the work performed. The concepts related to pedagogical intervention are addressed, as they are fundamental to understand the main scope of this work. Deepening knowledge in deep learning models for automatic textual classification helps us to understand the technique and acquire the skills to perform the automatic classification of subjective attributes. The ontology concepts were used from (LOPES et al., 2021; JR et al., 2020). Related works are also presented, which bring some similar characteristics and others that are different, helping us to have a mastery of what has been explored or not.

2.1 INTERVENTION

The term intervention is used in several areas such as Psychology, Medicine, Administration, and Education. More precisely, in education, it is very common to call this intervention a pedagogical intervention.

According to (HATTIE; BIGGS; PURDIE, 1996), these interventions aim to increase motivation, mnemonic skills, self-regulation, study-related skills such as time management, and even general skill itself; create positive attitudes towards content and context; and minimize learning pathologies.

Interventions can be broadly classified as cognitive, metacognitive, and affective (BARBIER et al., 2022). Cognitive interventions are those that focus on developing or enhancing specific task-related skills.

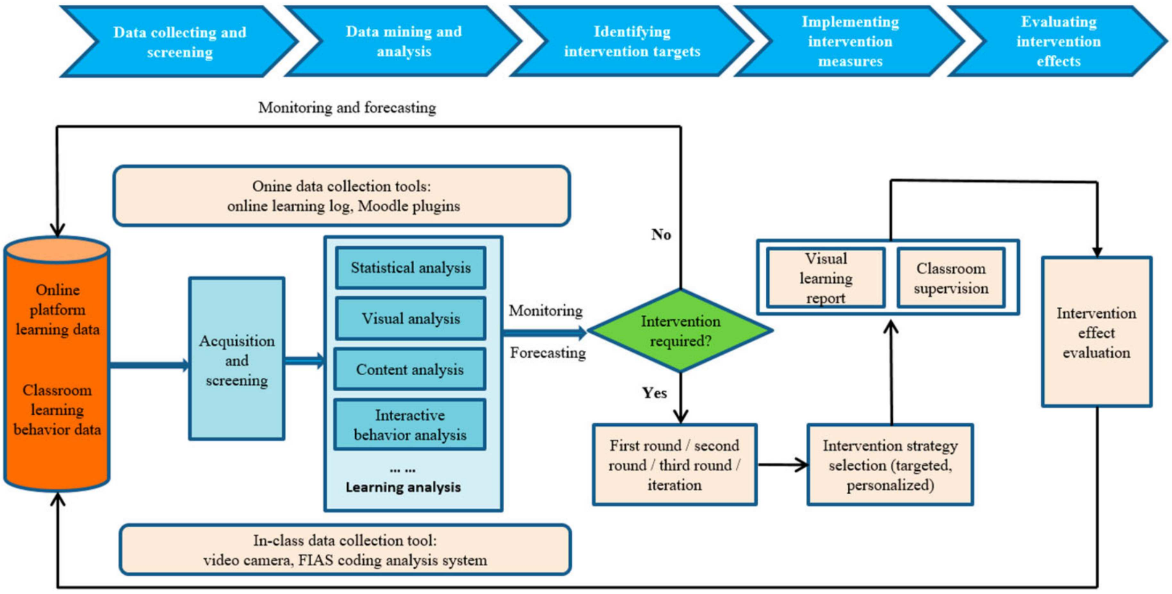
Metacognitive interventions are those that focus on self-management of learning, that is, planning, implementing, and monitoring a person's learning efforts, and conditional knowledge of when, where, why, and how to use specific tactics and strategies in their activities in appropriate contexts.

Affective interventions are those that focus on non-cognitive aspects of learning, such as motivation and self-concept.

Students with learning difficulties require pedagogical intervention from teachers, such as offering both oral and written help and allowing extra time to complete the tasks. Learning must be linked to the affective act, it must be pleasant and stimulating (MÁXIMO; MARINHO, 2021).

In the work (ZHANG et al., 2020), a conceptual framework was proposed for, an individualized intervention framework for blended learning from the perspective of learning analytics, as shown in Figure 1. The framework describes five steps of implementing the learning intervention: data collecting and screening, data mining and analysis, identifying intervention targets, implementing intervention measures, and evaluating intervention effects.

Figure 1 – Conceptual framework for learning intervention



Source: Zhang et al. 2020.

According to (MORENO-MARCOS et al., 2018a), e-Learning platforms store a large amount of data from all student interactions with the platform. These interactions are related to course navigation events, video events (when the video was played, speed, etc.), and exercise logs (attempts, scores, tips used, etc.). In addition, social interactions, such as forum postings can be used to detect attitudes and sentiments. All of this data can be used to detect problems based on the data collected, but you can also use this information to predict behavior and outcomes. Based on predictions, teachers can anticipate possible problems, providing adaptations to the course or methodology, so instructors or even the institution can rethink the curriculum design or carry out interventions to improve the learning experience.

It is very common for a post to be accompanied by a sequence of events, which makes it even more difficult to get the necessary attention from an instructor in a virtual learning environment(VLE). Trying to solve this problem, students ask colleagues to vote for their posts for the tutor to see and carry out a necessary intervention. Three models were proposed for predicting the instructor's intervention taking into account high-level resources, such as post initiated by an anonymous user, approved, unresolved, or deleted, forum ID, time, last time post, the total number of posts on the topic, word count related to rating, technical, conclusive and request issues, the sum of votes, etc. (CHATURVEDI; GOLDWASSER; III, 2014).

According to (LEISS, 2010), teacher interventions can be defined as teacher-student assistance that supports individual learning and students' problem-solving process, enabling students to work independently. According to (GUREL; BEKDEMIR, 2022) and (TROPPER; LEISS; HÄNZE, 2015), there are three types of intervention intention. The first is the diagnostic intervention, where an evaluation is carried out throughout the process. The second intention is the tip when something is recommended so that the student can overcome a problem. And the last one is intended to provide feedback, so a comment is provided to the student.

The authors in (MARBOUTI; DIEFES-DUX; MADHAVAN, 2016) show an interesting strategy using a predictive system to identify students who are at risk and apply intervention guidelines so that the student can be successful. It is possible to use a system that makes use of predictive modeling techniques to help identify students who are at risk, also allowing to predict student success. (MARBOUTI; DIEFES-DUX; MADHAVAN, 2016) highlights that existing academic early warning systems have some shortcomings. The fairly common problem is that they typically employ a general prediction model that fails to address the complexity of all courses. Another problem is that most early warning systems are designed for online courses or are too dependent on Course Management System (CMS) access data.

Some works use attributes that express how the student feels to avoid data directly linked to the environment. For example, in (YANG et al., 2015) the detection of confusion and its impact on learning is performed. The educational process frequently involves battles with confusion. An instantaneous response clarifying the confusion can accompany the student to overcome it, causing a beneficial effect. However, the delay in clearing up the confusion can detract from the experience of participating in the course, leading to dropout along the way. Therefore, the confusion experienced during the learning process cannot always be associated with negative results. This confusion does not hinder the student's development and learning and must be resolved promptly.

Sentiment analysis is very important in MOOC environments. According to (MORENO-MARCOS et al., 2018b), identifying whether forum messages are positive or negative can give an insight into how students feel about the course, aiming to increase engagement and student satisfaction. Furthermore, message polarity can help identify more complex emotions such as excitement, frustration, or boredom. The analysis carried out in (MORENO-MARCOS et al., 2018b) shows a more positive trend towards the messages at the beginning of the course, and the positivity decreases at certain peaks, even days before the deadline for the assessment tasks.

2.2 DEEP LEARNING MODELS

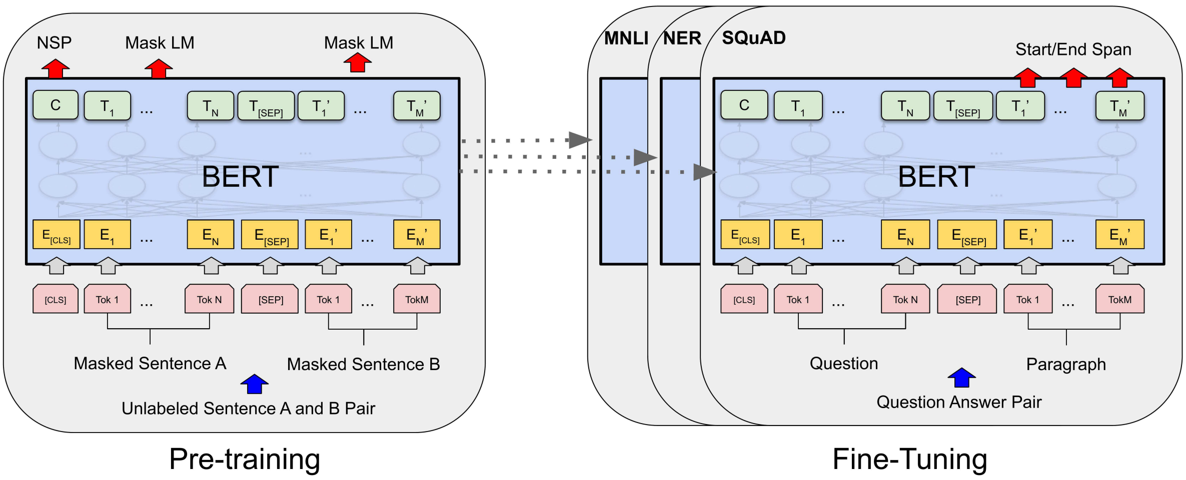
Deep learning has provided huge advances and advancements for machine learning and AI in recent years. Academic research presents the use of deep learning techniques for visual analysis and natural language processing, which includes speech recognition, language translation, sentiment analysis, etc.

This work explored two different approaches; the first used the pre-trained BERT model (DEVLIN et al., 2018), and the second used Hereditary Tree-LSTM (GOMES et al., 2022).

2.2.1 Bidirectional Encoder Representations from Transformers - BERT

In this approach, we have a neural network that can perform attribute prediction given an input message. A layer of this network is composed of BERT - Bidirectional Encoder Representations from Transformers, a neural network developed by Google that can understand expressions in human language. BERT is pre-trained with unlabeled data for various tasks (DEVLIN et al., 2018), and to use it, it is necessary to perform fine-tuning, where the BERT model is first initialized with pre-trained parameters, and these are adjusted using labeled data. Then, for each attribute that was used, a model was trained and fitted.

Figure 2 – Bert, Pre-Training, and Fine-Tune Tasks.



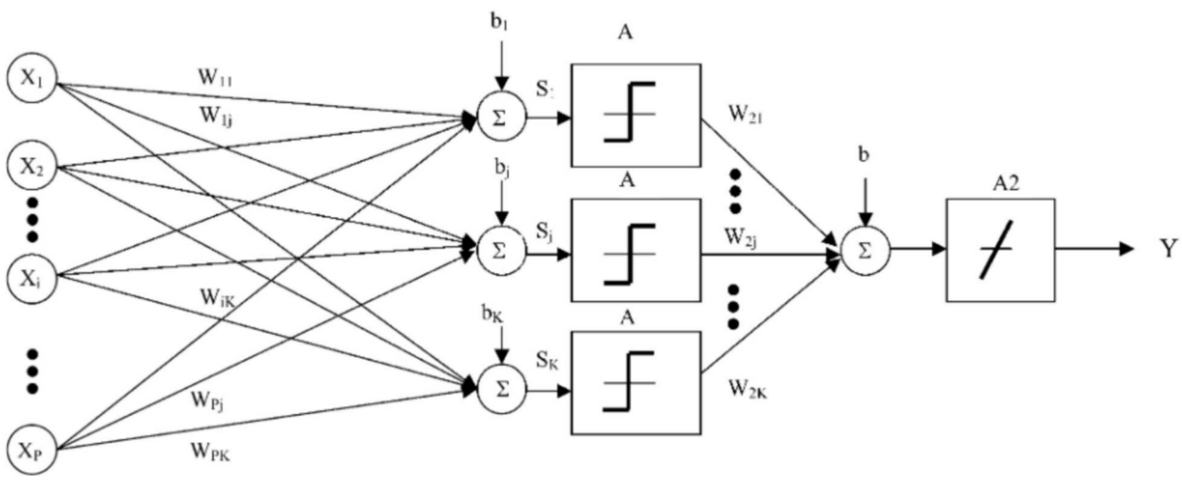
Source: Devlin et al. 2018.

The created network receives the values produced by Bert's tokenizer as input. In the first layer of the network, we have the pre-trained BERT model, and the second layer is a Dropout layer, responsible for retaining or not output from the previous layer, with a retention value of 0.3. Finally, we have a linear type layer that creates a feedforward network with n inputs and m outputs. In this case, we have input n of size 768 and output m of size 5.

A dropout layer was used to compose the model, this layer consists of randomly selecting neurons to be discarded during the training process, thus, the remaining neurons have their weights adjusted by backpropagation, resulting in a network capable of better generalizing the model. The main motivation behind the algorithm is to prevent the co-adaptation of feature detectors, or overfitting. The discard was adjusted by setting the probability p = 0.2. (BALDI; SADOWSKI, 2013)

Figure 3 presents the architecture of a feedforward neural network, strongly connected, containing only a hidden layer, as shown in (DOLEZEL; HECKENBERGEROVÁ, 2017) where neurons have a linear activation function, responsible for adjusting the weights and then performing the classification.

Figure 3 – Feedforward neural network with simple hidden layer



Source: Dolezel and Heckenbergerová (2017).

2.2.2 Tree-LSTM with hereditary attention - HTL

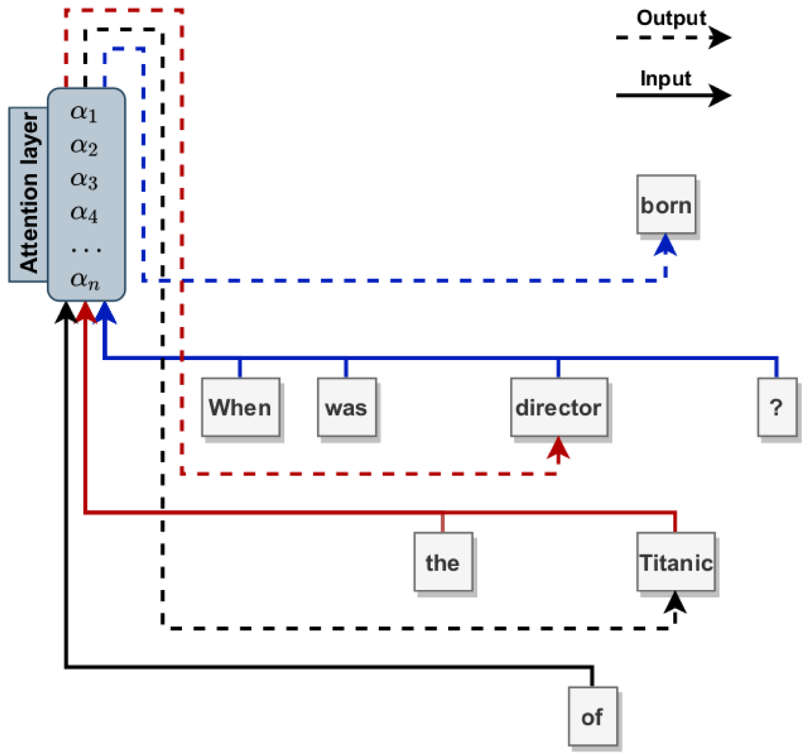
The Tree-LSTM is a change from the Long Short Term Memory (LSTM) standard, where the difference can be found in the organization of the LSTM unit cells. In this approach, they are structured in a tree format, passing information from the leaf nodes to the root node. In addition, the Tree-LSTM unit hierarchically incorporates information from each child node, while the standard LSTM disseminates historical information through a sequence of neural network units.

Tree-LSTM can correctly parse the structured semantic information of a sentence in contrast to the standard linear chain of LSTM and bidirectional LSTM (TAI; SOCHER; MANNING, 2015).

As an alternative to Tree-LSTM, (GOMES et al., 2022) presents that the attention mechanism for Tree-LSTM can achieve good results for semantic relationship tasks and text sentiment classification. In Tree-LSTM with hereditary attention mechanism (HTL), it consists of adding the hereditary attention mechanism to help Tree-LSTM, so the focus is only on relevant information of a natural language problem. Attention is applied successively to the set of children of each subtree and decides the most relevant features that should be emphasized to build the new hidden state of this subtree. Therefore, the hidden states of each subtree are weighted with an "importance factor" until the root node is reached.

Figure 4 shows an example of the Tree-LSTM version with hereditary attention. The solid arrow shows the input of the attention layer, and the dotted arrow shows the output after the factor weighting process. Each arrow color represents the flow of a level (set of children of each subtree).

Figure 4 – Tree-LSTM with hereditary attention



Source: Gomes Jr. 2022.

2.3 RELATED WORK

Many proposals for pedagogical intervention systems can be found in the literature. In some way, most of them are related to this research. Therefore, considering the breadth of the topic, this section presents only works related to pedagogical intervention systems in virtual learning environments without the need for human assistance to perform tasks, i.e., they work autonomously.

The paper (SOBREIRA et al., 2020) presents a proposal that can suggest pedagogical strategies for tutors of virtual learning environments. The software agent can recommend pedagogical strategies to tutors in VLEs. The intelligent agent uses ontologies for its knowledge model, which greatly facilitates the manipulation of different types of data and information. This proposal focuses on the tutors' work, looking at what they have already accomplished to recommend what should be done to improve the quality of teaching.

In (CHATURVEDI; GOLDWASSER; III, 2014), the authors detected that students who need an intervention ask their colleagues to vote on the post to draw the instructor's attention to solve this problem. Therefore, they proposed three models that perform the prediction of when the tutor should carry out an intervention. This work considers platform attributes, such as post time, last post time, number of posts on the topic, number of votes, etc. The disadvantage of this approach is that the solution depends on the attributes the platform provides.

To achieve student retention, (MARBOUTI; DIEFES-DUX; MADHAVAN, 2016) uses predictive modeling to identify students at risk of dropping out. The model seeks to identify performance patterns to predict in advance the student's performance, and if the student may be at risk, he and the tutor are alerted. The author cites that other works employ a general prediction model, but this strategy can result in poorly accurate predictions of student success in a course because course learning objectives, activities, and assessments can vary widely. In addition, the disadvantage of using the model based on performance standards is the need for a large set of variables (notes) to obtain a more generalizable model.

In (YANG et al., 2015), the authors identify confusion as one of the main causes of evasion of MOOC courses. The confusion must be clarified quickly, as it is part of the educational process to overcome confusion. Still, the student cannot always clarify his doubts and confusion alone, so, in several cases, intervention by the tutor is necessary. The author proposes creating a classification model based on behavior data from the discussion forum and the flow of clicks to identify posts that express confusion automatically. So, it is possible to carry out interventions and improve the student retention rate in MOOCs.

According to the authors in (MORENO-MARCOS et al., 2018b), messages posted on MOOC forums are the most important source of social interaction. From these interactions, it is possible to identify the polarity (positive or negative) present in the message. The work aims to compare different machine learning algorithms for sentiment analysis, using a real case study to verify how emotions can provide information about student outcomes or patterns in the MOOC. The best approaches were Random Forest and a lexicon-based method using word dictionaries. The study also analyzed the evolution of positivity over time, with the best moment at the beginning of the course and the worst close to the deadlines of the peer reviews.

In (KHODEIR, 2021), research with messages from forums in MOOC courses is presented. According to the author, students use the forum for various tasks, including solving their problems in the educational process, so they need to identify urgent messages thus the system can intervene quickly and automatically. The work uses BERT to perform word coding, and then this model is used as input to a classification model that consists of a bidirectional multilayer GRU (Gated recurrent unit). One of the layers is the pre-trained BERT model. This work disregards other sentiments detected in the post, focusing only on urgency.

The authors in (CAPUANO; CABALLÉ, 2019) propose a multi-attribute text categorization tool capable of automatically detecting useful information from MOOC forum posts, including intentions, topics covered, sentiment polarity, level of confusion, and urgency. The extracted information can be used directly by instructors to moderate and plan their interventions and inputs to conversational software agents capable of engaging students in constructive discussions guided through natural language. The work is focused on detecting various subjective information automatically and describes a scenario of using this information with a conversational agent but does not make a proposal on how the architecture should be to conduct the agent.

As presented in this session, some works have some points in common with the objective of this work. One of the works that bear a considerable similarity is (CAPUANO; CABALLÉ, 2019), as they also used deep learning models to detect subjective attributes present in students' social interactions. However, they do not present a knowledge model to store information to use autonomous agents to perform tasks. The works carried out by (MORENO-MARCOS et al., 2018b; YANG et al., 2015; KHODEIR, 2021) explore sentiment, confusion, and urgency attributes, respectively. Therefore, these works look only at one attribute to understand the student's situation. Although (SOBREIRA et al., 2020) make use of ontologies in its structure, which also resembles the architecture proposed in this work, the authors seek to evaluate the pedagogical strategy used by the tutor and, if necessary, recommend new pedagogical strategies. Finally, in (CHATURVEDI; GOLDWASSER; III, 2014; MARBOUTI; DIEFES-DUX; MADHAVAN, 2016) it is proposed to create predictive models that perform the detection of students who need intervention, but these models need performance data or the platform, which ends up making it difficult to use these models for other courses and educational environments.

Based on the literature discussed in this section, we can say that this work contributes to the others. The first is the intervention approach by an advanced conversational agent. The second is the computational solution capable of automatically detecting subjective attributes such as sentiment, urgency, and confusion. Finally, it can detect the need for an intervention based on student interactions and then guide them, as well as tutors.

3 PROPOSAL: PRED-INTER ARCHITECTURE

This chapter presents the architecture named Predicting Pedagogical Intervention (PRED-INTER) based on subjective attributes. The layers of PRED-INTER are presented in detail to expose the functionality of each one of them.

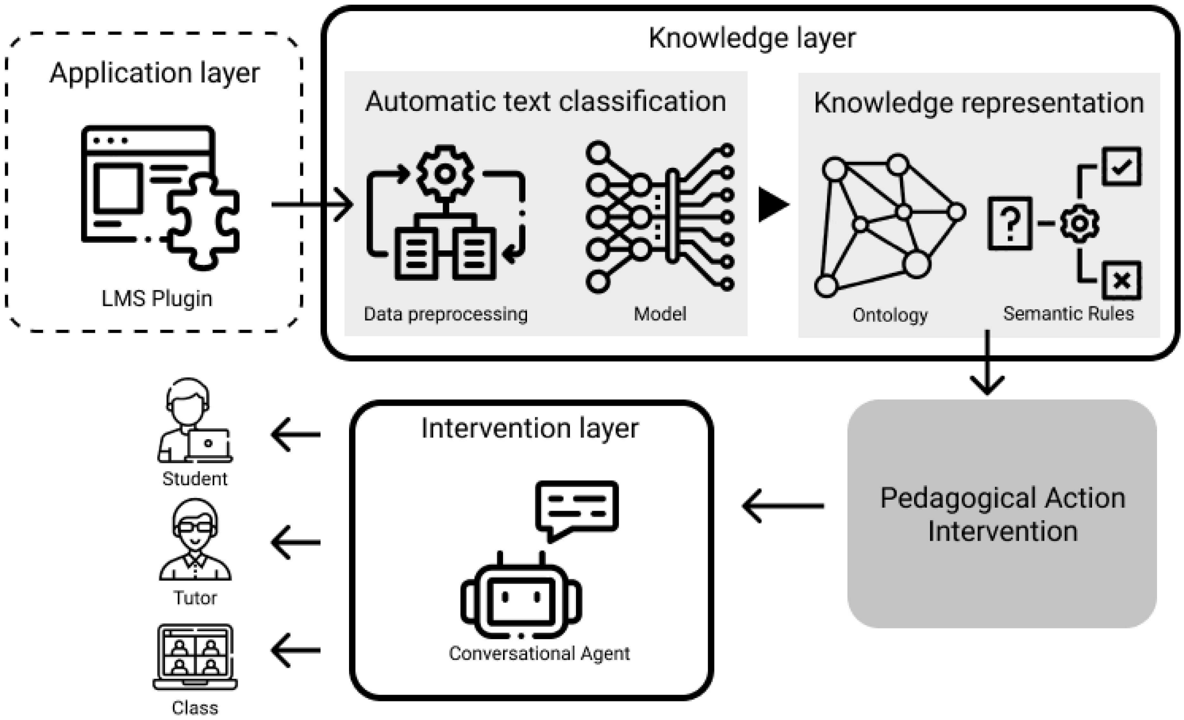
3.1 OVERVIEW

The proposed architecture aims to help tutors and students to carry out pedagogical interventions considering the student's educational moment. PRED-INTER is based on how the student feels, using the automatically detected subjective attributes obtained through interactions in the learning management systems. It is important to emphasize that evaluating each student's participation and identifying the necessary pedagogical intervention is essential, as it is impossible to perform this task manually.

The architecture collaborates by monitoring the forums of virtual environments to identify students' needs and automatically help them. So they avoid feelings of abandonment and demotivation. However, some cases are identified as more critical; for example, when the post sentiment is negative, students need more attention. In these cases, our solution alerts the tutors so they can carry out an individualized follow-up. On the other hand, less critical cases, defined according to the attributes, can be monitored automatically through motivational, informative, or thank you messages or by an automatic interaction. These kinds of pedagogical intervention will prevent the student from creating the sentiment of abandonment and help the tutor to avoid task overload.

Figure 5 presents the layers and workflow that make up the proposed architecture to carry out the pedagogical intervention.

Figure 5 – PRED-INTER Architecture



Source: Created by the author (2022).

The application layer is responsible for interacting with VLEs and capturing students' messages. These messages can come from chats, forums, or any other form of textual interaction. For the proposed architecture, in addition to the posted message, it is also essential to identify the author.

In the Knowledge layer, we have a very important part of the work carried out. It has two main attributions and was divided into two modules. The first is responsible for automatic textual classification, where the attributes of sentiment, confusion, and urgency are automatically identified in messages from pre-trained models. The second module is responsible for the representation of knowledge. The information is stored so that the domain knowledge is represented, allowing the understanding of the information related to the message.

The Knowledge layer starts with text pre-processing, where unwanted text parts are removed. Then, textual classification is responsible for identifying the attributes implicit in the messages (details in Section 4.2), where pre-trained models are used to identify these attributes.

Still, in the knowledge layer, but in the representation module, the post information is stored, along with the identification of the attributes. Storage takes place in an ontology in the second module. Semantic rules are employed to identify the pedagogical intervention necessary to meet the student's post type. These rules guide decision-making, resulting in the choice of which pedagogical intervention is most appropriate to meet the student's needs at that moment. The output of this layer is the selected pedagogical action intervention.

At last, we have the intervention layer composed of an autonomous conversational agent. The agent assists the student and the tutor, sending messages to carry out the pedagogical intervention. It also includes sending messages to the class. The mediation content of the message depends on the identified attributes, which are previously stored in the ontology, as well as the type of intervention.

The type of intervention is also pre-defined, and may vary according to the class or course. It may be an automatic message to the student or it may be a message asking the class to help the student. When a post in a discussion forum indicates that the student needs more than a message for pedagogical intervention, the agent may start and conduct an automatic interference (help) or ask the tutor help. For automatic help the subject in the message is recognized and a pre-selected educational resources is recommended to help the student with the content (ROSSI et al., 2021). This autonomous interference and recommendation of educational resources is based in a recommender component, which defines what will be recommended and the priority that each item will have. The focus of this research is not the integration of a recommendation system with the conversational agent. However, more complex cases where the agent cannot help the student the post is forwarded to the tutor.

3.2 KNOWLEDGE LAYER

This section presents the Knowledge layer highlighting its modules, describing each step performed, and generating a deeper knowledge about the layer.

3.2.1 Automatic Text Classification

The textual categorization component is responsible for labeling messages through predictive models, taking into account three attributes: sentiment, urgency, and confusion.

The performance of predictive models is influenced by the quality of the data used in the training process. The models must be trained with vast data volume to make good predictions. Another essential step in contributing to the models' performance is the pre-processing step, which is responsible for discarding unwanted data that does not add semantic information to the texts. The preprocessor takes a text input as it is found and cleans it up, removing unwanted tokens. Tokens are isolated words or groups of words (ANANDARAJAN; HILL; NOLAN, 2019).

According to (ANANDARAJAN; HILL; NOLAN, 2019), a lot of time is spent preparing and pre-processing text data. However, despite being a time-consuming step, it must be carried out because, in addition to contribute to improve the accuracy of the models, the training process ends up becoming more agile, as it reduces the dimensionality of the data.

One of the steps performed in the pre-processing is the removal of stop words. They are usually less significant words and do not contain semantic value but appear at high frequencies in the texts. This removal seeks to reduce dimensionality and computational cost. However, this step was not performed in this work, because the models used consider the position in which the words are.

In the pre-processing step, we must observe other cases too. For example, some characters may be in capital letters one or more times in the same word. Text authors can do it to emphasize information, but this can lead to an unnecessary increase in dimensionality, as the classifiers will treat them as different words, so this step was executed.

Replacing abbreviations and misspelled words are also commonly used in social network analysis or e-commerce systems and should also be joined in the source word. As the messages come from a learning environment, these problems do not happen often. So this step was not performed.

Data cleaning is also a pre-processing task that focuses on removing some information, which can get in the way of semantic understanding. According to (KUMAR; GARG et al., 2020), the cleaning mechanism is used to process this data and make it viable for further investigations. Often this technique removes URLs, paths, excess spaces, and expressions enclosed in angle brackets. In our work we did the following data cleaning:

- **Username**: The presence of the username does not contribute positively to the sentiment analysis, i.e., it does not play any role in the classification. Thus it only increases the dimensionality, decreasing the performance of the classifier.
- **URLs**: URLs can be interpreted by the classifier, which can result in the sentence polarity changing, so it's important to remove URLs as well.
- **Numbers**: Numerical data do not have semantic meaning. Therefore they do not add information to the classification process and must also be removed.
- **Special characters**: Because a special character can change the recognition of a word, for example, "#task" can be different from "task".

After the stage that performs the pre-processing of the data, it is time to train the models to carry out the textual classification. We used machine learning algorithms to perform the training. The training relied on a supervised learning technique, which consists of knowing the expected prediction for the data used. Hence, the process consists of learning the function in which the set of input variables X will produce the expected output Y (CUNNINGHAM; CORD; DELANY, 2008).

Considering that we sought to classify three different attributes, the training step was performed for each of them separately, giving rise to three different models that act independently. The process was repeated so that at the end of each step, it was possible to obtain a distinct model, which makes the classification of a given attribute.

After obtaining the models already trained, it is possible to send an unknown output Y message to the model, and we perform the prediction of the class corresponding to the desired attribute. Thus, we can automatically identify the implicit information in the messages.

3.2.2 Knowledge Representation

For the system to be able to represent the domain, the concept of knowledge-based system (KBS) development was used. According to (CHI; CHEN; TSAI, 2015), the emphasis of KBS development is to create inference-capable problem-solving mechanisms. When broadly analyzed and defined, knowledge models are robust and extensive in knowledge inference. Furthermore, new facts can be stored in the existing knowledge model, and existing logical relationships are inherited for reasoning without additional data processing. Therefore, KBS are suitable for solving knowledge-intensive problems that require inference mechanisms.

The knowledge model was developed to map and store information related to the student's post domain, keeping the semantic details. Another essential point is that the attributes identified in the previous step by the machine learning models must also be archived. According to the needs presented, the best way to meet the knowledge model was to use an ontology to perform this task.

With the data loaded, the proposed ontology can infer the most appropriate pedagogical intervention style for each post based on the attributes representing how the student feels. The information contained in the ontology was loaded from the same dataset used to prepare the models that detect the attributes.

Each post made by the students is stored in the ontology present in the knowledge representation module. As information is stored maintaining semantic knowledge, there is the possibility of performing processing and making inferences, so we can discover information that is not explicit.

This module adds the feasibility of better understanding the domain by obtaining information in more details. So, later integration with other layers can be carried out, as in this case, where we allow the integration with the intervention layer. When identifying the pedagogical intervention, it is available for the agent to act.

Developing the ontology that composes this module included creating classes, property chains, data properties, and semantic rules. The Ontology Development 101 methodology, proposed by (NOY; MCGUINNESS et al., 2001) was used.

Based on the adopted methodology 101, the step-by-step procedure was followed for constructing the ontology described in the next subsections. The ontology development followed several stages, starting with the definition of the domain and scope; based on these definitions the classes and subclasses were identified; soon after, the relationships between individuals that gave rise to the objectProperties were created; subsequently the relationships between individuals and values, called dataProperties; later, some individuals that are pre-existing in the ontology were defined; at the last moment, the semantic rules that identify the pedagogical intervention were defined.

3.2.2.1 Domain and Scope Definition

The first step in ontology development is responsible for defining the domain and scope of the topic to be analyzed. A very important resource to specify the ontology functionality in detail, which was also developed, is the Ontology Requirements Specification Document (ORSD), presented in Table 1.

The ORSD is a document based on the methodology proposed by (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; VILLAZÓN-TERRAZAS, 2009), which aims to describe the ontology in general, such as objectives, scope, implementation language, end users, requirements, and other characteristics.

Table 1 – Description of the ontology, its uses, end users, requirements, and others

Ontology Requirements Specification Document
Objective Identify the correct pedagogical intervention type, starting from the posting of a student in the learning management systems, taking into account how the student feels, based on subjective attributes, present in the text.
Scope Provide the student with a feedback-type pedagogical intervention and help the tutor identify the student who needs a more targeted and individualized intervention.
Implementation Language The ontology is specified in the OWL 2 language (Web Ontology Language).
End Users User 1 – Tutors; User 2 – Class; User 3 – Students.
Use cases Usage 1. Allow the identification of different types of pedagogical intervention, taking into account how the student is feeling at the time of posting.
Ontology requirements Functional Requirements: CQ 1: From the identification of how the student is feeling, what are the pedagogical interventions identified? CQ 2: How many interventions are separated by type? CQ 3: What is the distribution of interventions separated by actor? CQ 4: How many interventions are separated by domain? CQ 5: How many separate interventions exist per student? The interventions raised can be classified into: 1 - Automatic Message 2 - Automatic Help 3 - Tutor Help 4 - Class Help

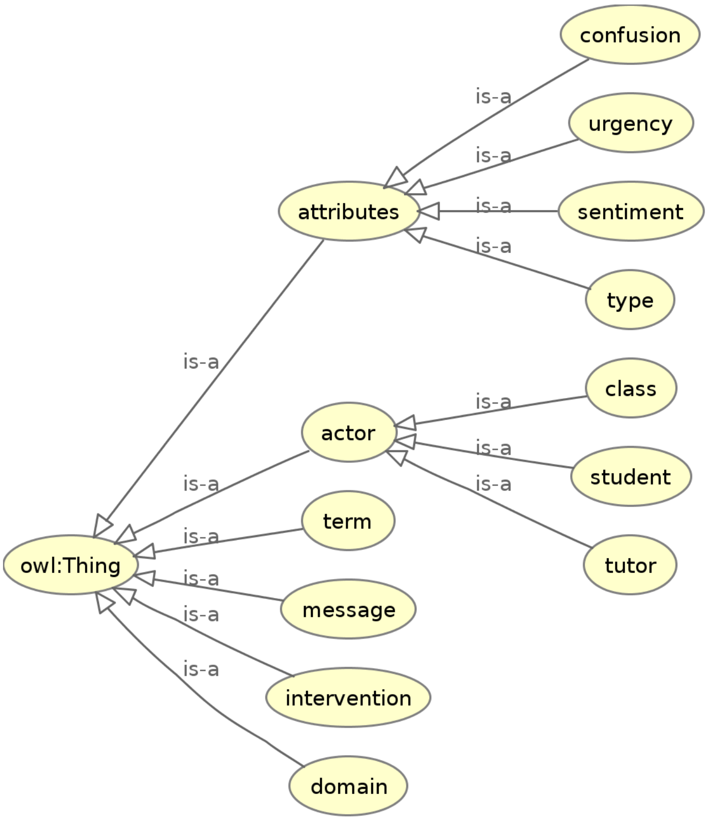
Source: Created by the author (2022).

The second step of methodology 101 consists of considering the reuse of ontologies. Despite being found several works (DOLOG; NEJDL, 2007; REZENDE et al., 2015; BREMGARTNER; NETTO; MENEZES, 2015; PAQUETTE, 2016; SIMON et al., 2020) that present ontology proposals for the educational domain, they cannot represent the subjectivity of how the student feels in a given social interaction through the VLE. Considering this limitation, knowing that the ontologies developed for the educational context work with more objective information, we decided to design the ontology without reusing others.

3.2.2.2 Classes and subclasses

In this step, we made the representation of the concepts of the proposal. We designed the organization of the knowledge as classifier trees of concepts, and based on this diagram we identified the hierarchy, attributes, relationships, and tables of instances.

Figure 6 – Ontology classes and subclasses



Source: Created by the author (2022).

The identification of the classes occurred through the identification of the terms that make up the domain. To better understand the classes that belong to the developed ontology, Figure 6 shows the structure of the ontology classes, and their subclasses. Table 2 presents each of the main classes with their respective descriptions.

Table 2 – Description of ontology classes and subclasses

Class	Description	Instantiated by	Related to
message	Responsible for storing the content of the student's post	Automatic extraction	domain through "belongs" Attributes through "contains" Intervention through "require"
domain	Stores the area of knowledge that the student's course is related to	Automatic extraction	-
term	Words that make up the message	Predefined in ontology	message through "compose"
intervention	The type of pedagogical mediation that is used to interfere in the learning process. It may be an Automatic Message, an Automatic Help, Class Help, or Tutor Help	Predefined in ontology	actor through "directed"
actor	Receives the entities involved in the intervention process, where the actor can assume the role of tutor, student, class	Automatic extraction	message through "create"
attributes	Message related features	Automatic extraction	-
confusion	Determines how confusing the student's post is;	Automatic extraction	Subclass of attributes
sentiment	Sentiment expressed in the message	Automatic extraction	Subclass of attributes
type	This option stores the type of message that the student sent. This message can be opinion, question and answer	Automatic extraction	Subclass of attributes
urgency	Property that defines the degree of urgency that the message needs to be answered	Automatic extraction	Subclass of attributes
class	Class that the student is linked to	Predefined in ontology	Subclass of actor
tutor	Tutor responsible for the class	Automatic extraction	Subclass of actor
student	Student who posted	Automatic extraction	Subclass of actor

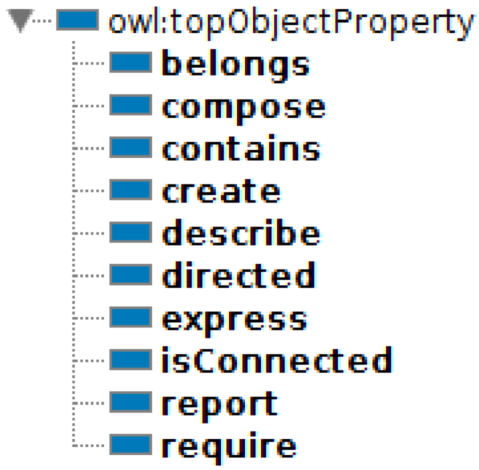
Source: Created by the author (2022).

The domain class appears in the ontology for recommending educational resources, as it helps to recommend some content, and can also filter by domain, resulting in a more accurate recommendation.

3.2.2.3 Object properties

Once the classes have been identified, a table to handle the relationships can be created by assigning names and identifying their domain and range. These relationships are called object properties and represent the type of interaction between domain elements (classes or individuals). It is important to emphasize that the properties of objects must maintain the semantic relationship between the elements.

Figure 7 – Object properties



Source: Created by the author (2022).

Figure 7 presents the semantic relationships between the classes. Table 4 presents the object properties with their domains and ranges.

Table 4 – describing object properties

Object Properties	Domain	Range
belongs	message	domain
compose	term	message
contains	message	attributes
create	student	message
directed	intervention	actor
require	message	intervention
report	message	confusion
isConnected	message	type
express	message	sentiment
describe	message	urgency

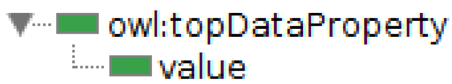
Source: Created by the author (2022).

3.2.2.4 Data Property

Data property provides a relation to attach an entity instance to some literal datatype value (number, string or date) that is a measure or estimate of what that data property is about.

To associate a numerical value for the individuals that represent the attributes of sentiment, urgency, confusion, and type, a data property called value (Figure 8) was created, which helps us to optimize the semantic rules.

Figure 8 – Data properties



Source: Created by the author (2022).

3.2.2.5 Individuals

An individual is an instance that represents a particular element of a class. A class can contain a large number of individuals. In this ontology some elements are previously instantiated to support data insertion. Four individuals were included in the intervention class, 5 in the sentiment class, 3 in the confusion class, and 3 in the urgency class.

For the urgency and confusion classes, the works (CAPUANO; CABALLÉ, 2019; CAPUANO et al., 2021) were considered, so we have three instances for these classes, representing the low, medium, and high levels. For the sentiment class, five levels of polarization were classified, in addition to the representation of negative, neutral, and positive sentiment, two more representations were included, to represent the very negative and the very positive. We will discuss both solutions in the next section.

Table 5 – Individuals previously inserted into the ontology. Intervention type(I), Sentiment(S), Urgency(U) and Confusion(C) attributes, with their data property values

Individuals	**Class**	**Data Property**	**Description**
TutorHelp	I	-	Requests tutor intervention
AutomaticMessage	I	-	Requests a message intervention
AutomaticHelp	I	-	Requests an automatic help intervention
ClassHelp	I	-	Requests a class intervention
VeryNegative	S	0	Indicates that the message polarity is too negative
Negative	S	1	Indicates that the message polarity is negative
Neutral	S	2	Indicates that the message polarity is neutral
Positive	S	3	Indicates that the message polarity is positive
VeryPositive	S	4	Indicates that the message polarity is very positive
Low	U	0	Indicates that the message has a low level of urgency
Medium	U	1	Indicates that the message has a medium level of urgency
High	U	2	Indicates that the message has a high level of urgency
Low	C	0	Indicates that the message has a low level of confusion
Medium	C	1	Indicates that the message has a medium level of confusion
High	C	2	Indicates that the message has a high level of confusion

Source: Created by the author (2022).

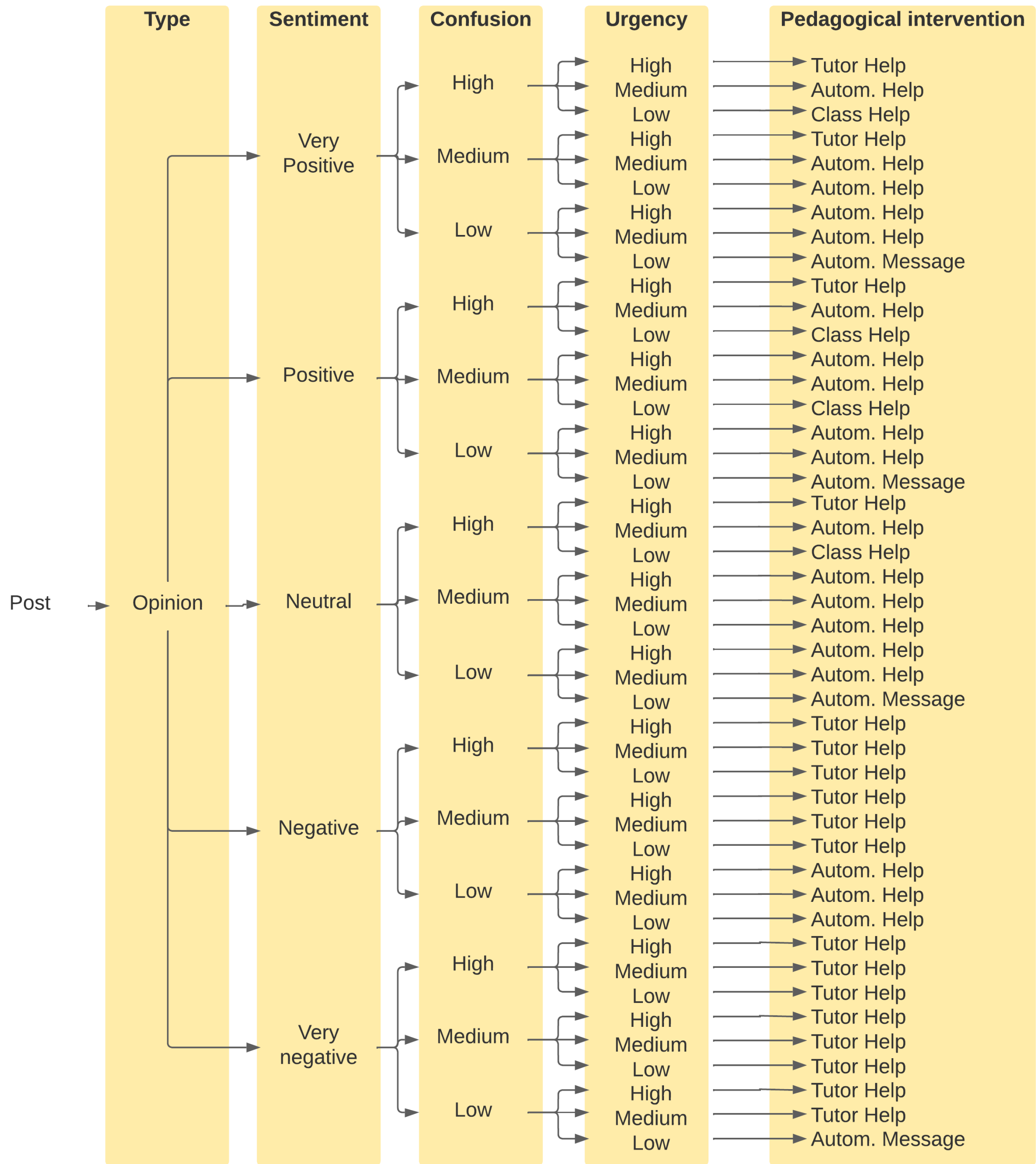
Table 5 shows some Individuals that are pre-defined in the ontology. They represent the values of the message attributes. The Data Properties column presents numerical values corresponding to the Individuals.

3.2.2.6 Semantic Rules

Before formulating the semantic rules, an interpretable decision model was defined based on subjective attributes, making it possible to identify the appropriate pedagogical intervention. This model is represented through a hierarchical diagram.

The semantic rules SWRL (BOUIHI; BAHAJ, 2019) were created to automatically identify the required intervention according to the input parameters. When applying the rules, the message is linked to a pre-defined style of intervention, which will support an intelligent system. In our architecture, this is represented by a conversational agent. The definition of the hierarchical diagram took into account that the data used in this study allows its expansion and use in other contexts since it is common data in virtual learning environments. So, the rules were defined by three teachers with long experience in distance education and tutoring discussion forums. With their expertise, they considered a learning scenario where the messages could express parameters of education such as types of questions (about the execution of a task or a class content), topics (motivational, social, or collaboration), and expressions (compliments, frustrations, or personal difficulties).

Figure 9 – Hierarchical diagram with post-formal representation of opinion-type message



Source: Created by the author (2022).

From the diagram in Figure 9, it was possible to identify the different types of intervention according to the student's needs at the time of posting the message. Based on the attributes of sentiment, confusion, and urgency, it is possible to assume how the student feels and thus carry out an intervention that is more meaningful to him.

For a pedagogical intervention to take place, it may be necessary to carry out several actions, for example, a Class Help sends message to ask the class to help the student but also asks the tutor to accompany this intervention. Table 6 presents the pedagogical interventions and the actions demanded by them.

Table 6 – Actions taken for each type of pedagogical intervention

Pedagogical Intervention	Actions	Description
Tutor Help	Message to student and Message to tutor	In this type of intervention, a message is sent to the students, informing them that they are being monitored. At the same time, the tutor is notified to offer support to the student.
Automatic Help	Interaction with the student	Automatic help starts a dialog and sends a content recommendation to the student, based on the concepts identified in the student's message.
Automatic Message	Message to student	The automatic message answers the student's post with messages of different purposes, such as thanking, motivating, or contacting.
Class Help	Message to class and Message to tutor	Class help facilitates dialogue among students in the same class, sending a message encouraging colleagues to comment on the post. The tutor is also alerted to follow up on this intervention.

Source: Created by the author (2022).

The Automatic Message to the student can be a thank you, a motivational, or a contact message. The message type will be defined according to the attributes, for example, a message with a positive sentiment, low confusion, and low urgency receives an automatic message intervention, which can be a thank you message type.

When the conversational agent provides automatic help, the content identification is fed with terms extracted from the Eurovoc[1] thesaurus. Thus, the agent can recognize the concepts that are present in the message. Eurovoc is a multilingual thesaurus of terms used in European Union documents and information systems. At the moment Eurovoc is available in the 22 official languages of the European Union and Croatian. The terms are separated by domain and subdomain, which helps to identify concepts from a given area.

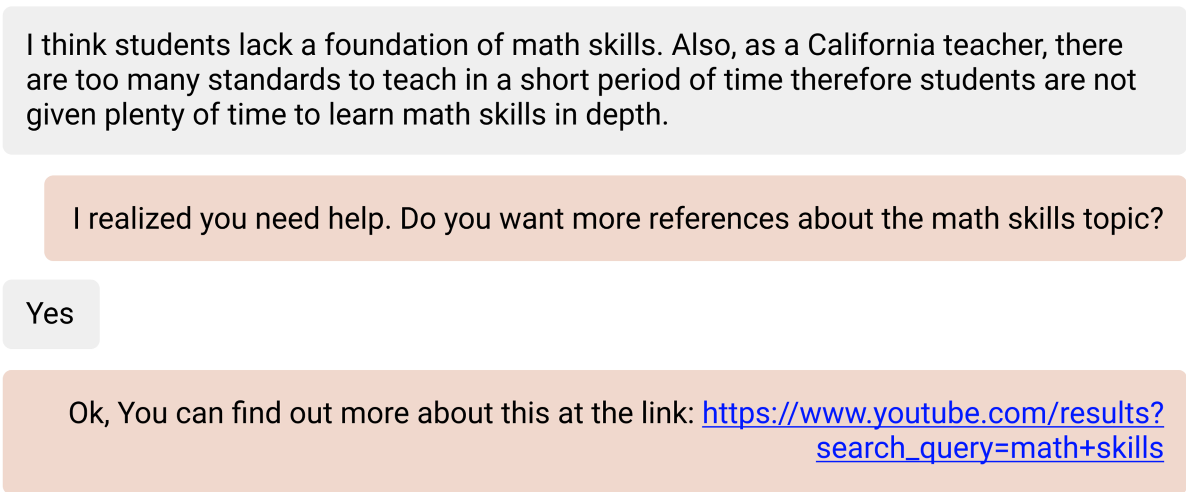
[1] <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eurovoc>

The first step is to extract the terms from Eurovoc so that they can later be loaded into the agent. The Dialogflow[2] tool was used to build the conversational agent, it helps in the creation and integration of conversation agents, being able to integrate services through APIs, such as sentiment analysis and knowledge base.

In Dialogflow there are predefined system entities that can correspond to many common types of data. For example, there are system entities that correspond to dates, times, colors, email addresses, and so on (SABHARWAL; AGRAWAL, 2020). We can also create custom entities to match custom data, in which case entities related to terms extracted from Eurovoc were created, allowing the agent to be able to identify concepts in the messages.

The message shown in Figure 10 has a negative sentiment, low confusion, and low urgency, so, based on these attributes, a pedagogical intervention of automatic help type was started. When the pedagogical intervention is of the automatic help type, the first step requires the agent to identify the entities that represent the concepts present in the student's message. In the second step, the student is asked if he wants to obtain more content on the identified concept and, if the suggestion is accepted, a YouTube link with more content on the subject is suggested, as shown in Figure 10. If the student declines the agent's content suggestion, the user will provide more details about the message content. This gives the agent another chance to recognize the subject of the message and supply the student with useful educational advice. However, if the agent is unable to identify the student's need for the second time, the message history is forwarded to the tutor, requesting more specific help.

Figure 10 – The conversational agent performing an automatic help pedagogical intervention



Source: Created by the author (2022).

2 <https://dialogflow.cloud.google.com/>

When the Knowledge Representation layer identifies that intervention is necessary, the conversational agent acts, trying to meet the student's demand. The messages sent by the conversational agent may vary according to the identified pedagogical intervention process. Therefore, some standard messages are:

1. **Tutor help:** Hilary, I noticed that you need help, so I've already notified the pedagogical team. The tutor will contact you soon.
2. **Automatic help:** Oliver, I realized you need help. Do you want to get help for [Concept1]?
3. **Automatic message:** James, your participation was crucial. Dedication and studies are the secrets to success. So make as much contribution as possible. It makes our interaction environment richer. Thank you for your contact.
4. **Class help:** Class, do you agree with what Laura said? Would anyone like to add something about it?

The messages sent by the conversational agent are selected according to the most appropriate style for the pedagogical intervention process, which is identified through semantic rules based on attributes, seeking to respond to opinion-type messages. Although some interventions are repeated, their approach may occur differently, varying according to their attributes.

For the PRED-INTER to work autonomously, we must detect the attributes that are implicit in the messages, for that, we developed the automatic textual classifier module, which is a very relevant part of the architecture. In subsection 4.2.1, the functioning of automatic textual classification is shown.

3.3 IMPLEMENTATION

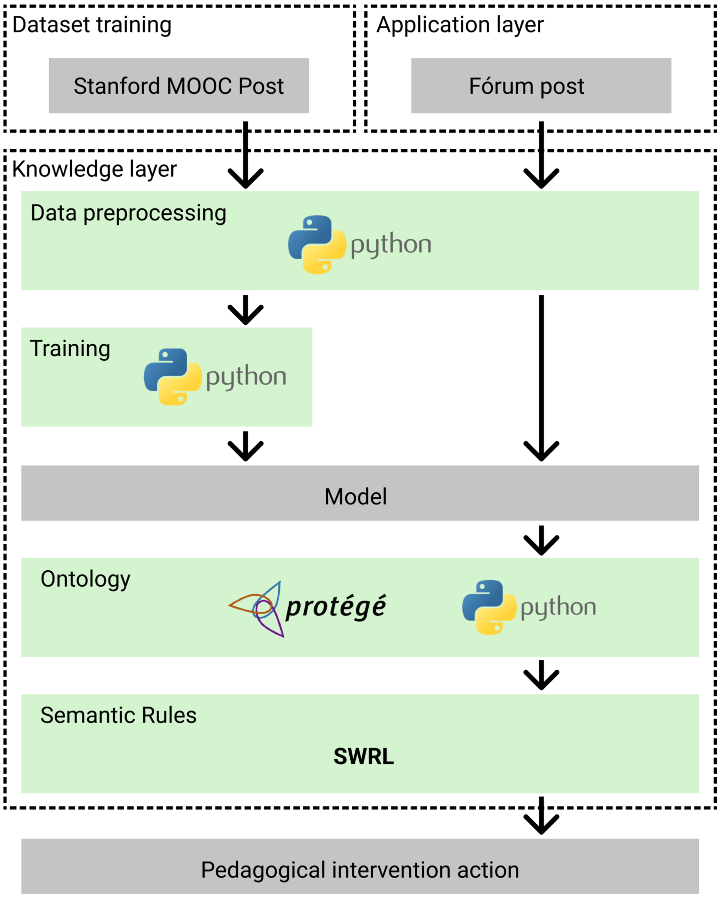
Several languages, libraries, and tools were used to develop the architecture. Knowing that each technology has a purpose, the choice was made to satisfy the desired objective in the best possible way, always taking into account reproducibility and performance. According to (RASCHKA; PATTERSON; NOLET, 2020), the Python language has seen tremendous growth in the scientific computing community, which has led to innovations in machine learning and deep learning libraries. Taking into account these statements and the agility in reading data from semi-structured sources, the Python language was prioritized to develop the ontology load. Some libraries were used to assist in this task: TensorFlow [3], Scikit-learn [4], transformers [5].

[3] <https://www.tensorflow.org>
[4] <https://scikit-learn.org>
[5] <https://huggingface.co/transformers>

According to (MCGUINNESS; HARMELEN et al., 2004), ontologies are intended to store information, maintaining their semantic relationships to be processed later by intelligent systems. The development of this resource was carried out through the Protégé tool [6] in version 5.5.0, which is a widely used open-source ontology editor. To implement the ontology, we chose the Protégé tool because it is extensible and provides an environment that makes it a flexible basis for rapid prototyping. The OWL language was used to create the ontology and detail all its specificities. A package for the Python language was used to insert the data into the ontology, which allows ontology-oriented programming, called OWLReady2. Through this package, it is possible to apply semantic rules and then produce inferences based on the stored information, thus achieving the ability to deduce new facts in the ontology.

In Figure 11, we present the technologies that were used in the knowledge layer. The Python language was prioritized, seeking greater integration between the modules that make up this layer.

Figure 11 – Architecture implementation details

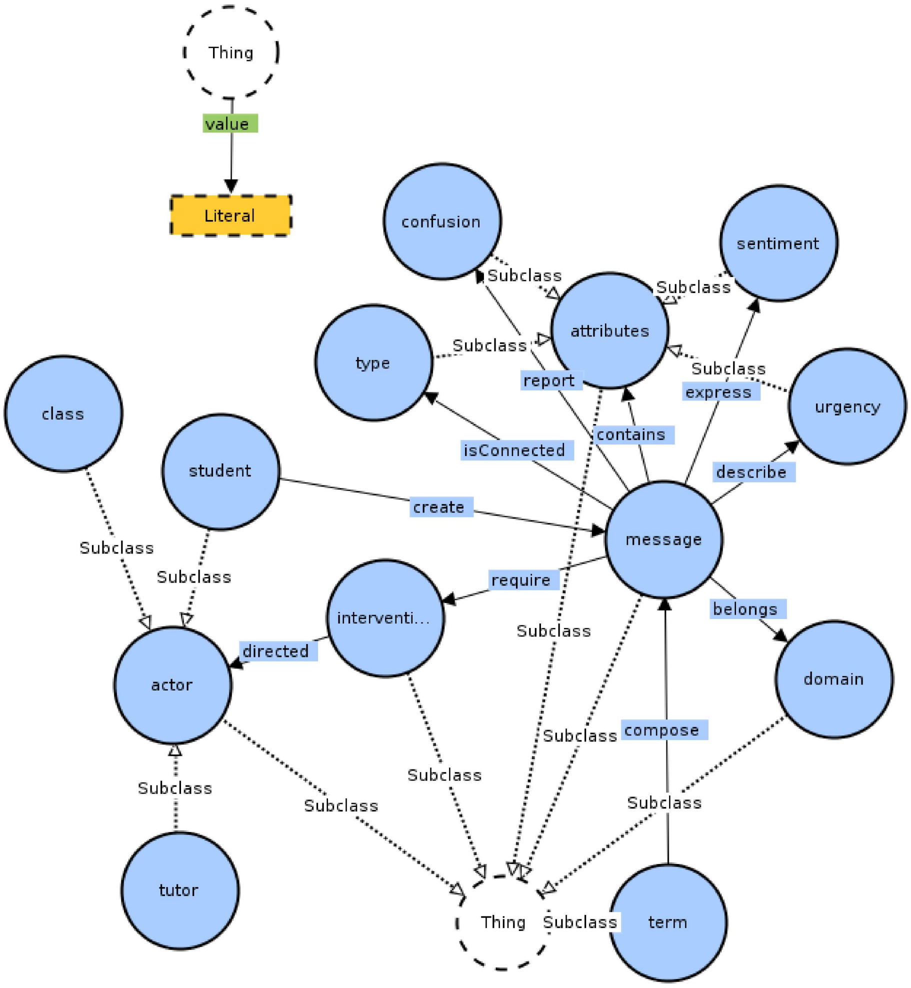


Source: Created by the author (2022).

[6] <https://protege.stanford.edu/>

Figure 12 presents the ontology structure, which was identified to represent the entire context, displaying all the developed elements, classes, object properties, and data properties. The figure was generated in the Protégé software, with the help of the VOWL plugin.

Figure 12 – Ontology



Source: Created by the author (2022).

As can be seen in Table 7, the rule is responsible for analyzing the attributes and selecting the appropriate pedagogical intervention for the posted message, which will meet the needs of the student at that moment, allowing him to instantly receive a message.

Table 7 – Semantic rules generated to determine the pedagogical intervention action

Rule	Description
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:lessThanOrEqual(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:greaterThanOrEqual(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very negative, or negative, with high or medium confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 2) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 2) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, ClassHelp)	Selecting Class Help intervention for student whose sentiment is very positive, positive, or neutral, with high confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 3) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, ClassHelp)	Selecting Class Help intervention for student whose sentiment is positive, with medium confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 4) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is very positive, with medium confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 2) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is neutral, with medium confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 2) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticMessage)	Selecting Automatic Message intervention for student whose sentiment is very positive, positive, or neutral, with low confusion and low urgency

message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 0) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticMessage)	Selecting Automatic Message intervention for student whose sentiment is very negative, with low confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 0) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is negative, with low confusion and low urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 2) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 1) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is very positive, positive, or neutral, with any level of confusion and medium urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:lessThanOrEqual(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:greaterThanOrEqual(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 1) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very negative, or negative, with high or medium confusion and medium urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 1) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is negative, with low confusion and medium urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 0) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 1) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very negative, with low confusion and medium urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 2) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very positive, positive, neutral, or negative, with high confusion and high urgency

message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very negative, with any level of confusion and high urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 4) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is very positive, with medium confusion and high urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:equal(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, TutorHelp)	Selecting Tutor Help intervention for student whose sentiment is negative, with medium confusion and high urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 2) ^swrlb:lessThanOrEqual(?sentdata, 3) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 1) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is positive, neutral, with medium confusion and high urgency
message(?msg) ^express(?msg, ?sent) ^value(?sent, ?sentdata) ^swrlb:greaterThanOrEqual(?sentdata, 1) ^report(?msg, ?conf) ^value(?conf, ?confdata) ^swrlb:equal(?confdata, 0) ^describe(?msg, ?urg) ^value(?urg, ?urgdata) ^swrlb:equal(?urgdata, 2) ^isConnected(?msg, Opinion) ->require(?msg, AutomaticHelp)	Selecting Automatic Help intervention for student whose sentiment is very positive, positive, neutral, or negative, with low confusion and high urgency

Source: Created by the author (2022).

These rules can be changed and adapted to the course or class, depending on the number of students per tutor, participation or not of colleagues in discussions and contributions of educational resources, availability of educational resources previously selected by teachers, and level of interaction required.

4 EVALUATION

This chapter briefly explores the dataset and the results of the automatic text classification and knowledge representation modules. The assessment was carried out individually for each module, presenting its details and the results achieved.

We experienced two stages to evaluate the proposal and verify if the architecture could automate the pedagogical intervention task. The first is responsible for evaluating the classification models' capacity and measuring the models' effectiveness by classifying the attributes of messages that have yet been unlabeled. The second stage was responsible for evaluating how semantic rules could help in the process of pedagogical intervention. Therefore, it was necessary to observe the number of messages that can be automated and conduct the tutor's work only for the most critical messages that need exclusive monitoring.

The evaluation of the proposed architecture was carried out to measure the ability to automate pedagogical interventions, helping students and tutors. We return to the introduction section, where two research questions were posed and are used here to carry out the evaluation process. Here, the research questions and their derivatives will be presented.

- **RQ1**: Is it possible to improve results of the actual automatic student subjective information detection models?

  - SRQ1: Is it possible to detect subjective attributes independent of the platform information using only the post text?

  This question intends to verify if we can obtain a classification model that does not depend on specific data coming from the platform, developing a solution that does not have the platform as a restriction.

  - SRQ2: Does the use of machine learning techniques allow the detection of sentiment, confusion, and urgency in student's post messages?

  This question aims to show whether the automatic text classification module proposed in the architecture can detect subjective attributes and thus classify post messages, based on students' sentiment, confusion, and urgency.

- **RQ2**: It is possible to use the predicted student subjective information to recommend pedagogical interventions?

  - SRQ3: Is the proposed ontology capable of supporting tutors to identify the pedagogical intervention necessary to help student's interactions?

This question aims to verify whether ontology can semantically enrich students' posts and identify the necessary pedagogical intervention, considering the characteristics of their messages (sentiment, urgency, and confusion).

- SRQ4: Is it possible to carry out pedagogical interventions automatically, aiming to reduce the workload of tutors?

With this question, we intend to verify if the semantic rules allow automatic actions to be taken by a conversational agent. Resulting in reduced student waiting time for an answer and helping the tutor to identify critical messages.

The evaluation followed the GQM methodology (Goal - Question - Metric) (CALDIERA; ROMBACH, 1994), where we must stipulate the objective, the research questions, and the metrics that will be used. The scope of this evaluation and the Goal are described as follows: "*Propose an architecture composed of semantic rules and textual classification models based on deep learning, which can carry out pedagogical interventions in the student's learning process. These interventions must be adherent to the students' needs and when possible, must be carried out instantly and automatically, employing a conversational agent.*"

According to (CALDIERA; ROMBACH, 1994), the flow from objectives to metrics in the GQM paradigm can be visualized through a directed graph. The flow starts at the objective node and passes through the nodes representing the questions, arriving at the metric nodes. The metrics defined to answer the SRQs are:

- M1: To answer SRQ1 and SRQ2, we trained textual classification models and analyzed their accuracy, recall, and f-score measures to determine whether it is possible to identify the desired information automatically.
- M2: To answer SRQ3, we count how many interventions the ontology proposes, considering students' post messages in a MOOC environment.
- M3: To answer SRQ4, we analyzed the pedagogical intervention type identified through semantic rules, and we counted how much it is possible to automate the intervention task.

4.1 CLASSIFICATION EVALUATION

In questions SRQ1 and SRQ2, we search to detect automatically subjective attributes in text messages. To respond to them, we train and evaluate the models to measure their ability to perform automatic text classification. The evaluation of the models was carried out through the metrics of precision, recall, and f-score.

These models are of profound importance for architecture, as they are responsible for automatically classifying the attributes. The attributes detected represent values of sentiment, confusion, and urgency, thus, enabling the identification of the most appropriate pedagogical intervention, based on the attributes, since it is possible to consider how the student is feeling.

Knowing the dataset represents a fundamentally necessary step for the automatic text classification task, so before starting the models training process, the dataset was analyzed to explore how this data are. Based on this step, it is possible to develop strategies to obtain better results.

4.1.1 Dataset

For this work, Stanford MOOCPosts dataset [1] was used, a dataset that has a good number of details about the message and a reasonable volume of instances. Stanford MOOCPosts contains 29,604 anonymous student posts from eleven different content forums. The posts refer to courses in the areas of human sciences, medicine, and education. The messages present in the dataset were labeled manually by three human coders (AGRAWAL; PAEPCKE, 2014), and the labeled dimensions can be seen in Table 8.

To work with a considerable amount of data, it is necessary to interpret the data, therefore an analysis was studied to verify the distribution, based on the messages of the dataset, allowing us to understand the number of messages per target class, the number of words per instance, the number of characters and the most frequent words in each of the classes.

To execute this study, the attributes sentiment, confusion, and urgency were used, as they are considered relevant to define how the student is feeling. According to (CAPUANO; CABALLÉ, 2019), the attributes were labeled with three distinct values, for the sentiment the values are negative, neutral, and positive. The attributes of urgency and confusion received low, medium, and high values.

Initially, the analysis was performed by observing the number of instances for each value corresponding to the attributes. The analysis was carried out for the values of the attributes of sentiment, urgency, and confusion, making it possible to examine how the classes are distributed, the imbalance can significantly interfere with the results achieved. Figures 13a, 13b, and 13c present graphs with the number of messages grouped by the value correspondent in the manual annotation process.

Figure 13a shows the number of messages grouped by sentiment label, where it is possible to observe that a considerable percentage, about 55%, of the messages were labeled as neutral, therefore, a small number of messages will represent polarized messages. When we look at messages representing negative sentiment, this number is even lower.

[1] <https://datastage.stanford.edu/StanfordMoocPosts/>

Table 8 – Attributes present in the forum messages

Column name	Description	Values
Opinion	Post contains an opinion	1 or 0
Question	Post contains a question	1 or 0
Answer	Post contains an answer	1 or 0
Sentiment	Learner sentiment expressed in post	1 to 7
Confusion	Learner degree of confusion expressed in post	1 to 7
Urgency	How urgent is it that instructor reads the post	1 to 7
CourseType	Course knowledge area	Education, Humanities, and Medicine
Forum_post_id	Unique ID of the respective row's post in its original OpenEdX context	Number
Course_display_name	Name of course in context of Stanford's online, free, public offerings	Text
Forum_uid	Unique identifier of learner who posted the post	Number
Created_at	Post date	Date
Post_type	One of Comment or CommentThread; the latter is assigned to posts that originated a thread, while the former is assigned to all other posts	Text
Anonymous	If True, poster appears to everyone as name ‘anonymous’	1 or 0
Anonymous_to_peers	If True, poster appears under his/her own screen name to discussion moderator and instructor, but as ’anonymous’ to everyone else.	1 or 0
Up_count	Number of post's up-votes	Number
Comment_thread_id	ID of thread object	Number
Reads	The total number of reads logged for the thread with ID.	Number

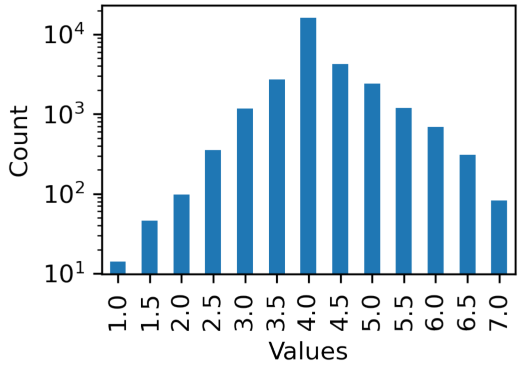
Source: Created by the author (2022).

In Figure 13b it is possible to identify the number of messages separated by the values that represent the urgency. The result reported in Figure 13a ends up reflected in the urgency graph as well because as we have a small number of messages with negative sentiment, we also have a few messages that need urgency.

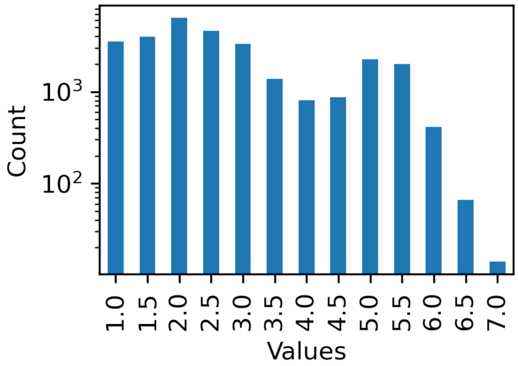
According to (CAPUANO; CABALLÉ, 2019) the confusion can be divided into three levels, low, medium, and high. In Figure 13c, we can see how a reduced number of messages represent the extremes, with low or high confusion, this distribution of confusing messages may be related to the feeling they represent, as in Figure 13a a large quantitative is in the central region of the graph.

Figure 13 – Distribution of messages by label

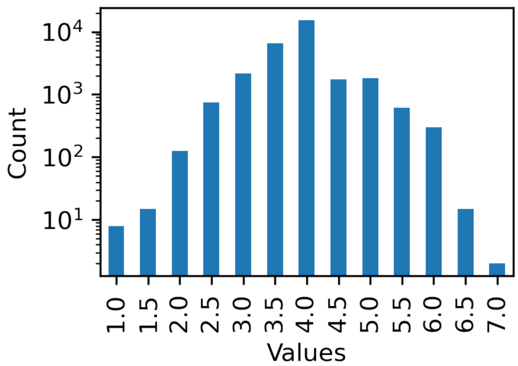
(a) Number of instances separated by the sentiment attribute



(b) Number of instances separated by the urgency attribute



(c) Number of instances separated by the confusion attribute



Source: Created by the author (2022).

Appendix A displays the values in more detail, allowing to view the mean, standard deviation, highest, and lowest values. Figure 14 shows the distribution of the number of words according to the label present in the attribute.

When we analyze Figure 14a, we can see that the messages that have sentiment with the most negative polarity tend to have a higher number of the average value of the words. Figure 14b has a more uniform distribution, only in the initial levels of urgency that the average number of words have a minor variation. Figure 14c messages that do not express confusion tend to have a higher average in terms of the number of words.

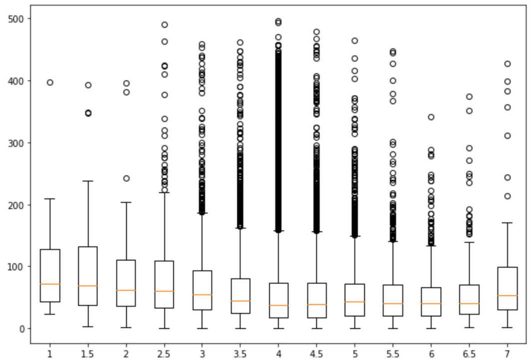
4.1.2 Data Preparation

This section presents the pre-processing task performed to extract unwanted content without relevance to the automatic textual classification task.

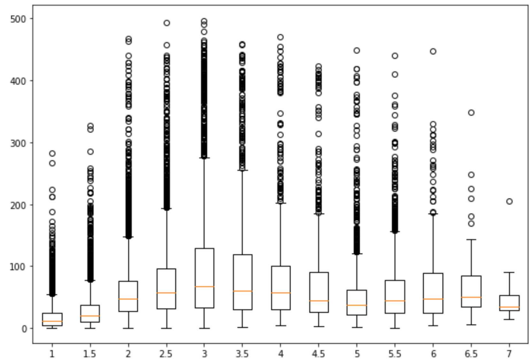
The discretization of attribute values will follow the identical pattern adopted in

Figure 14 – Number of words in messages considering the label of each attribute

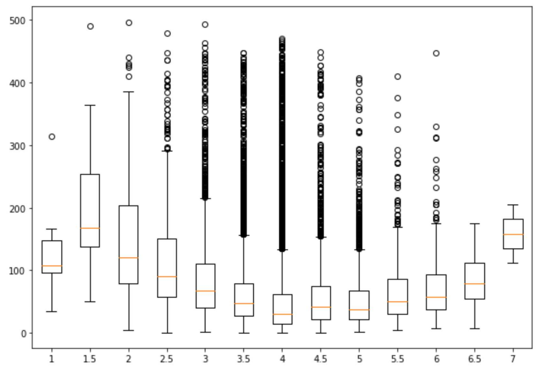
(a) Number of words in messages for the sentiment attribute



(b) Number of words in messages for the urgency attribute



(c) Number of words in messages for the confusion attribute



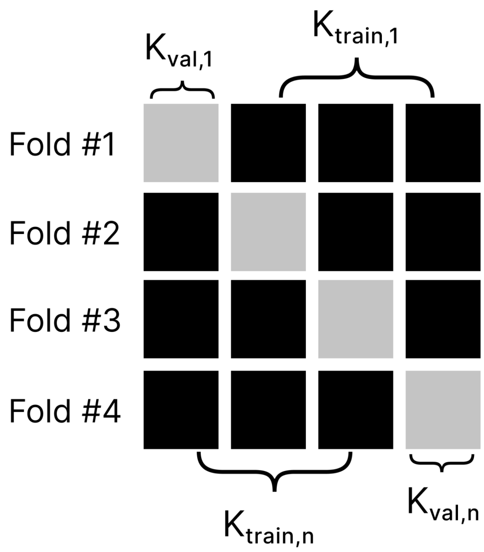
Source: Created by the author (2022).

the work (CAPUANO; CABALLÉ, 2019; CAPUANO et al., 2021), resulting in three distinct classes for each attribute. Considering sentiment, values less than or equal to 3 correspond to the negative class; values greater than 3 and less than or equal to 5 represent the neutral class; The positive class is portrayed by values greater than 5. For the attributes confusion and urgency, the low class, which represents little or no presence of this attribute, was assigned to values less than or equal to 3. The class that represents the average presence of the attributes grouped values greater than or equal to 3.5 and less than or equal to 5, while the high was represented by values greater than or equal to 5.5.

To reduce dimensionality and even improve training performance, unnecessary information was removed. This information, which was removed, does not include semantic value to the sentences, messages between square brackets, URLs, paths, and spaces in excess.

To generalize the evaluation and avoid the problem of model overfitting, K-fold Cross-validation was used. This technique consists of dividing the data into k subsets so that these parts do not allow the overlap of two different subsets, that is, an instance cannot belong to two subsets. (BERRAR, 2019) According to the divisions performed, 1 subset is used for testing, and the rest is used for training the model. In this case, the value 4 was used for the variable k, so the data of the set were divided into 4 parts, where 3 of these parts were used for training and 1 for evaluation. This process has been repeated a total of K times, until each of the subsets has been used to evaluate the model, in the end, the result generated is the average of the metrics of all the executions performed.

Figure 15 – K-fold Cross-validation



Source: Created by the author (2022).

In Figure 15, the dataset is divided into 4 distinct subsets, for each iteration 75% of the data is used for training and 25% for testing. This procedure is repeated 4 times until all subsets have been used to evaluate the model.

4.1.3 Results

After performing the analysis to interpret the data contained in the dataset, we present in this section the results obtained from the automatic text classification layer, which is responsible for autonomously detecting the desired attributes, implicitly contained in the submitted messages.

It is essential to evaluate the performance of subjective attribute classification models, to better comprehend how it behaves. To assess the models, training was carried out with three parts of the dataset, and later one part was utilized to evaluate, giving rise to the results that will be discussed in this section.

We have carried out the experiments on Google Colab[2]. For each instance in the dataset, the model predicts the sentiment, confusion, and urgency label. To evaluate the

[2] <https://colab.research.google.com/>

models, precision, recall, and f-score metrics were analyzed, which will allow us to compare the models with other existing approaches (CAPUANO; CABALLÉ, 2019; CAPUANO et al., 2021).

This research took into account other works of considerable relevance, which also make use of automatic text classification models. According to (CAPUANO; CABALLÉ, 2019; CAPUANO et al., 2021), we set the baseline to evaluate our models, as these works also make use of the same dataset that was used to train and evaluate our deep learning networks. The bow-ff approach was developed in (CAPUANO; CABALLÉ, 2019), HAN was developed in (CAPUANO et al., 2021).

In work, (CAPUANO et al., 2021) the network with attention mechanism obtained good results, which indicates that using attention mechanisms can get an excellent option. BERT was chosen because it possesses several works in the literature that got good results, such as (KHODEIR, 2021), which trained a model to recognize the urgency in MOOC course messages. HTL was based on the approach presented in (GOMES et al., 2022), which also makes use of attention mechanisms.

Table 9 – Results achieved by the proposed approach and comparison with others approaches

Attribute	Architecture	Epochs	Loss	Precision	Recall	F-Score
Sentiment	bow-ff	20	0.645	**88.62%**	86.07%	87.33%
	HAN	10	0.115	88.34%	88.29%	**88.31%**
	BERT	20	0.970	87.87%	**88.85%**	88.10%
	HTL	20	1.477	84.70%	87.30%	84.80%
Confusion	bow-ff	20	0.081	87.19%	84.25%	85.70%
	HAN	10	0.103	87.25%	85.56%	86.40%
	BERT	20	0.352	**87.99%**	**86.09%**	**87.03%**
	HTL	20	0.381	81.49%	85.76%	80.33%
Urgency	bow-ff	20	0.098	84.05%	75.75%	79.69%
	HAN	10	0.103	82.95%	76.40%	79.54%
	BERT	20	0.215	**86.84%**	**83.51%**	**85.12%**
	HTL	20	0.455	77.20%	81.34%	77.98%

Source: Created by the author (2022).

In Table 9, we can see that BERT and HTL approaches are numerically competitive with the baseline results. However, the results obtained by BERT showed slightly better results when compared to HTL. When we look at the result achieved by BERT, we can see that sentiment is very close to HAN. When comparing confusion, BERT is slightly better than HAN; and when comparing urgency BERT is significantly better. However, just looking at the numbers, it may not be the best analysis.

Therefore, based on the results obtained, it is possible to answer the secondary research questions SRQ1 and SRQ2, where the numbers show that it is viable to perform automatic text classification without using platform data and employing machine learning approaches. Remembering that the objective is to help tutors and students, it is essential to correctly classify messages with a negative sentiment, as they identify the students who need more help and are certainly students who are unhappy with something. Therefore, it is very important to minimize errors in messages that are negatively biased.

The values of the sentiment attribute used to train the models, whose results were presented in Table 9, were discretized into three classes. However, we can increase the class numbers; we could have very negative, negative, neutral, positive, and very positive messages. According to the new discretization of values into five classes, we have the following distribution: sentiment values less than 2 represent a very negative polarization, values greater than 1.5 and less than 3.5 correspond to negative, greater than 3 and less than 5.5 represent neutral, those greater than 5 and less than 6.5 are positive, and those greater than 6 are very positive.

Table 10 – Results achieved through deep neural networks, with discretization of sentiment in 3 and 5 classes

Architecture	**Epochs**	**Loss**	**Precision**	**Recall**	**F-Score**
BERT (3 Class)	20	0.970	87.87%	88.85%	88.10%
BERT (5 Class)	20	1.070	86.29%	87.74%	86.84%
HTL (3 Class)	20	1.477	84.70%	87.30%	84.80%
HTL (5 Class)	20	0.598	75.61%	80.60%	76.66%

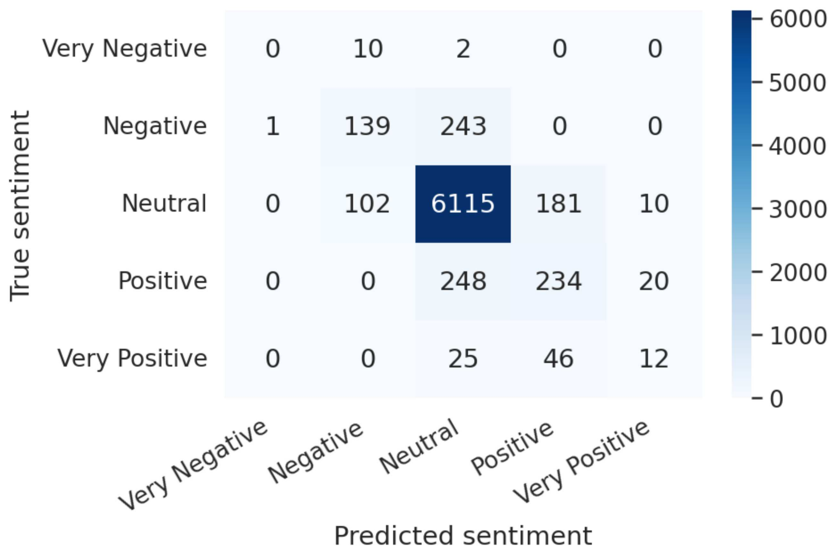
Source: Created by the author (2022).

Table 10 compares the results obtained through discretization in 3 and 5 classes, although the problem has become more complex, the metrics have not suffered great variations. However, the difference between the results obtained by BERT is smaller, while in HTL, the difference between the results is a little bigger, so we can consider that BERT behaves better with the increase in the number of classes.

In order to evaluate the model obtained, we must explore better the results. In this case, it is important to conduct qualitative analysis to observe the behavior of the classifier to know the real and predicted sentiment, especially of students who represent a negative sentiment, who are considered more critical to the focus of this work.

According to the results, the BERT model approach stood out from the others, so in-depth data analyses were carried out using the BERT approach. Checking the confusion matrix can help us to evaluate more deeply and then to know the model's behavior individually for each class. Even looking at what was misclassified also helps us better understand the model's behavior. Figure 16 presents the BERT model's behavior through the confusion matrix generated for the attribute sentiment divided into 5 classes.

Figure 16 – BERT model confusion matrix for 5 classes



Source: Created by the author (2022).

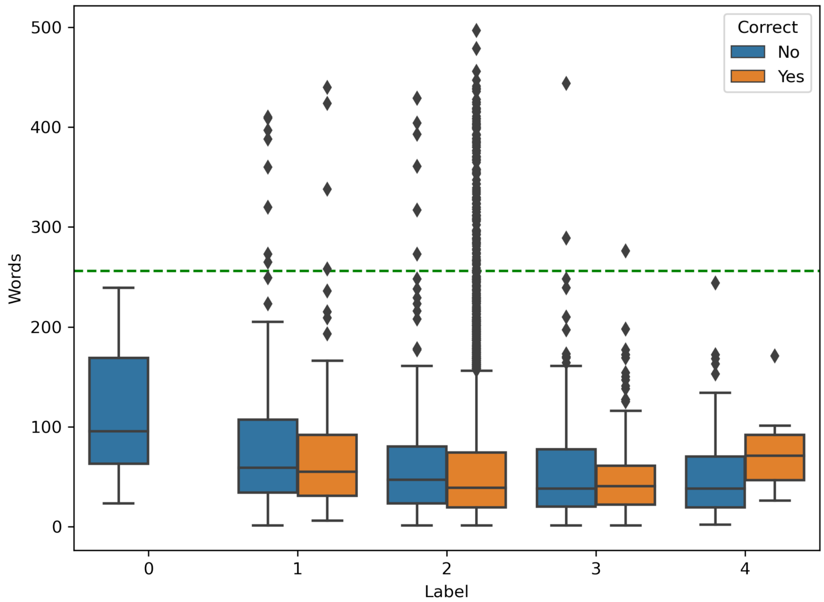
Observing Figure 16, we can verify that with five classes, the sentiment classification was better compared to the classification for three classes. In addition, to recognize a greater number of messages with negative polarity, there was no classification of messages with negative polarity predicted to be positive. Therefore, as this work aims to help students and tutors, through pedagogical intervention, to avoid the feeling of abandonment and help to avoid dropout, we can consider that the classification with 5 classes was relatively better. Furthermore, as shown in Figure 16, we can assess that the model tends to classify messages as neutral. This is due to the imbalance of classes since the messages posted in the social interactions of dataset are normally not polarized about sentiment.

BERT has a limitation regarding the number of words, and we investigated whether this limitation is disturbing the model of correctly classifying messages. We check if there are considerable errors in messages that are above the word limit.

In Figure 17 we present a graph showing the relationship between the size of the messages and the errors and successes. The boxes in blue represent incorrect classifications, while the correct classifications are in orange. The horizontal axis represents the different classes that represent sentiments, and the vertical axis determines the number of words in the message. Due to computational resource limitations, up to 100 tokens are considered per instance for training and evaluating the model. In the graph, the dotted line represents this limit in green color that separates the graph horizontally. As we can see, it is not possible to identify a relationship between incorrectly classified messages and their sizes. Therefore, limiting the maximum number of words allowed by the classifier cannot get in the way of the model's performance.

For a better understanding of the errors made by automatic classification, Table 11

Figure 17 – Errors and successes distributions by class and post size



Source: Created by the author (2022).

presents some samples in which instances were incorrectly classified.

Table 11 shows a sample of post messages with their target sentiment, manually recorded in the dataset (CAPUANO; CABALLÉ, 2019), with the sentiment predicted by the model. It is possible to see that some messages cast doubts on the sentiment, as the target represents the point of view of the person who labeled the message. In this way, when we analyze the miss-classified messages, it is impossible to define if the error occurred because of the manual label process or the model predictor. Therefore, it is necessary to have well-defined criteria for the annotation process so that we can observe the predictor errors in a more precise way.

After the detection of attributes through automatic classification, all this information is loaded in the ontology, together with its respective message. Therefore, with the inference carried out through the semantic rules, finally, the intervention most adhering to the student's needs is selected.

4.2 SEMANTIC RULES EVALUATION

At this evaluation stage, we will analyze the ontology ability to identify the most appropriate type of pedagogical intervention. The ontology, together with the semantic rules, were presented earlier. The results obtained in this analysis aim to quantify the solution's ability to identify the type of intervention for each post, so it will be possible to

Table 11 – Messages with conflict between manually labeled sentiment(true sentiment) and model predicted sentiment(predicted sentiment).

Message	**True Sentiment**	**Predicted Sentiment**
I would be eager to hear a response to this and will follow if possible. This seems, if I understand it correctly, unorganised and inequitable, especially based on what we're learning in this session. How frustrating.	Very Negative	Negative
I have seen the effects of streaming, grouping students according to perceived ability.It is very damaging for all but particularly for Pacifica and Maori students in New Zealand. There are many reasons why students do not perform as well as others on entrance tests to High school. It is easy to see the change from a high band through to a low band class. There is a colour differential and it is a terrible crime to have this occurring. It sets students up for non-academic courses in the future and limits their learning experiences.	Very Negative	Negative
When I was growing up, my third grade teacher always told our class that boys were better than girls in math. This perception had possibly impacted on how I performed as a student.	Negative	Neutral
In my school, where I work, we only group the low achievement students, but we know, that when a student gets in this course, it's very difficult that he could reach the level to get out of the group, specially if it is a really low achivement group. And this will mark all his academic future.	Negative	Neutral
I was impressed by their comfort talking about math and familiarize with them	Neutral	Positive
Your post reminded me of the disappointment my own daughter felt when she was not accepted into a couple of honors classes. Although she had earned A's in both, the teachers wrote polite notes explaining that she d̈idn't have it in her,̈ and that she'd do better on regular, less challenging classes. She'll have to fight (by challenging herself, and self-directed study) if she wants to get out of the äcademic patḧthat has been written for her."	Neutral	Negative
One thing I would start my school year off with is sharing that I took this course. I would share that I learned that we can grow our math brains with hard work, practice, and good mentoring (Carol Dweck). I would also share that it takes time to seize things, to fully understand, that rapidity does not equal intelligence. And in order to reach that full understanding, we have to make mistakes. My goal for them is for them to dive into rich tasks and never give up (persevere). We are all in this journey together.	Positive	Neutral

Source: Created by the author (2022).

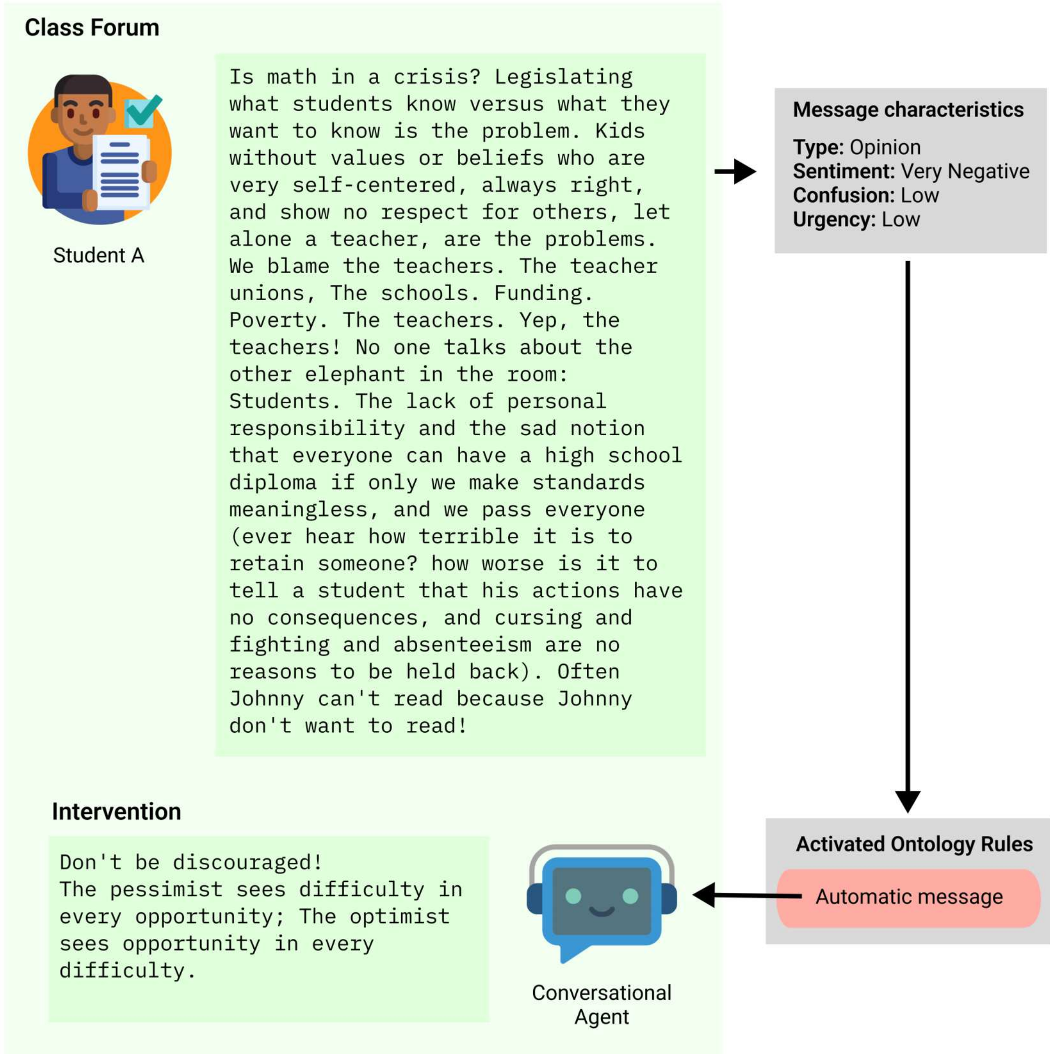
discuss the solution's ability to help the tutor and the student.

To answer SRQ3, the results obtained through semantic rules were analyzed. Based on the results obtained, it is possible to quantify the number of interventions generated. With this, we can determine if the students' posts are being answered.

To answer SRQ4, we analyzed the intervention suggested by the semantic rules, verifying if it is possible to automate the pedagogical intervention task, prevent the student from being left without answers and feeling abandoned, and reduce the tutor's effort.

Some messages require immediate intervention; in these cases, the system can take automatic action to intervene and minimize problems. Figure 18 presents a sample of the students' posts, the action of the conversational agent, and the pedagogical intervention. When the student sent the message, the attributes were automatically classified and processed by the ontology, resulting in the action that the agent must perform. Because it is a message with a very negative sentiment, low confusion, and low urgency, the agent sends an automatic message. In this case, the message is intended to motivate the student. In this way, it prevents the students from creating feelings of abandonment if they feel welcomed.

Figure 18 – Sample of one of the negative messages sent by a student in the MOOC and the conversational agent actions



Source: Created by the author (2022).

After obtaining the absolute number of interventions, we analyze how much they still require the tutors' actions. It allows us to verify if the automation of some tasks contributes to the fact that students, who demand a targeted service, receive an immediate agent contact or message.

4.2.1 Results

Table 12 presents opinion posts grouped according to attributes and the corresponding intervention.

Considering the aggregate number of instances that comprise the dataset, we can observe that about 44.7% of the messages are classified as positive or neutral and do not present confusion and urgency. This indicates that students are following the course satisfactorily.

Therefore, most messages would receive automatic feedback without humane interference. But, other messages will have some interaction between humans. This task is not exclusive to the tutor, as, on specific occasions, the students of the same class can encourage the author.

Table 13 – Distribution of interventions generated from the application of semantic rules.

Intervention	Count	Percent
Tutor Help	1199	7.23%
Automatic Help	12453	75.13%
Class Help	949	5.73%
Automatic Message	1975	11.91%

Source: Created by the author (2022).

In Table 13, we have the distribution of interventions separated according to the type. Two types of intervention automatically perform the interaction with the student: automatic help and automatic message. Together they represent a considerable number of interventions, approximately 87%. Regarding interventions that depend on human intervention, we have the class help, which corresponds to about 5.73%, and the tutor help, which corresponds to 7.23%.

Therefore, according to the semantic rules evaluation results, taking into account the selected pedagogical intervention types, we can answer SRQ3 and SRQ4. The 16,576 messages selected from the dataset received an appropriate intervention action after the tests. They answered all posts, including the most critical messages. Therefore, we can conclude that responding automatically to students' posts is possible. Considering the total number of evaluated interventions and the defined semantic rules, it is possible to perceive that the tutors' effort to follow a post had a considerable reduction, as only 7.23% of the interventions required individual and personalized assistance. Our solution may provide more time for the tutor to assist students who need help.

Table 12 – Quantitative posts are grouped taking into account the attributes and type of pedagogical intervention, namely: Automatic Message (AM), Automatic Help (AH), Tutor Help (TH), and Class Help (CH).

Sentiment	Confusion	Urgency	Intervention	Count
Very negative	Low	Low	AM	3
		Medium	TH	0
		High	TH	0
	Medium	Low	TH	23
		Medium	TH	7
		High	TH	4
	High	Low	TH	0
		Medium	TH	2
		High	TH	7
Negative	Low	Low	AH	185
		Medium	AH	12
		High	AH	0
	Medium	Low	TH	625
		Medium	TH	234
		High	TH	113
	High	Low	TH	1
		Medium	TH	4
		High	TH	19
Neutral	Low	Low	AM	1698
		Medium	AH	165
		High	AH	2
	Medium	Low	AH	9516
		Medium	AH	1850
		High	AH	383
	High	Low	CH	3
		Medium	AH	26
		High	TH	78
Positive	Low	Low	AM	236
		Medium	AH	6
		High	AH	0
	Medium	Low	CH	943
		Medium	AH	78
		High	AH	14
	High	Low	CH	0
		Medium	AH	0
		High	TH	0
Very positive	Low	Low	AM	38
		Medium	AH	11
		High	AH	0
	Medium	Low	AH	133
		Medium	AH	46
		High	TH	4
	High	Low	CH	3
		Medium	AH	26
		High	TH	78

Source: Created by the author (2022).

5 FINAL REMARKS

This work described the PRED-INTER architecture, to identify subjective information in social interaction messages in learning management systems. Based on this information, it selects the most appropriate type of pedagogical intervention for the student.

The results achieved in the deep learning approaches and the ontology with its semantic rules show that it is possible to support the students' wishes and, simultaneously, reduce the tutor's effort considering the automatic identification of the pedagogical intervention.

The PRED-INTER architecture reached the objectives expected in this work because it was possible to perform the automatic classification of subjective attributes that allows us to infer the student's feeling at the time of the message. Furthermore, the inference carried out in the ontology through the semantic rules proved viable, as it identified the type of pedagogical intervention necessary for the student.

From this research, some conclusions can be highlighted:

- The classification of implicit attributes allows an accurate prediction of how the student feels at the moment of the message; this contributes to understanding his educational moment.
- The construction of the ontology allows the context around posts in learning management systems to be represented, thus establishing possibilities to follow the educational moment the student is going through.
- The interpretable decision model helps to select the pedagogical intervention that should be adopted, even if tutors do not have an automatic classification model. For example, if tutors identify sentiment, confusion, and urgency, the decision model can identify the pedagogical intervention. Also, if a tutor considers that a message has a very negative sentiment, medium confusion, and high urgency, these entries suggest a tutor help intervention according to the decision model.
- Semantic rules allow us to classify posts into distinct groups, which can be used to recommend educational resources, pedagogical interventions, and personal interaction.

5.1 CONTRIBUTIONS

The pedagogical intervention architecture presents innovative aspects, as it considers subjective attributes in the student's message to automatically select the appropriate type of intervention.

The work includes developing and training deep learning networks, which can classify subjective attributes of students' messages in learning management systems. In addition, alternative approaches were evaluated to demonstrate the ability to extract the existing subjectivity in the textual language.

Another significant contribution was through the semantic rules, which guided the process of pedagogical intervention, allowing us to conclude that it is feasible to automate many of the tutor's tasks, which are related to the interventions.

The proposal presented in this work contributes to several fields of research, mainly on pedagogical interventions, showing the feasibility of performing interventions automatically through autonomous agents.

5.2 PROPOSAL LIMITATIONS

This work proposed an architecture to carry out a pedagogical intervention. Then, we evaluated the architecture's performance to automatically detect students needing help and execute the pedagogical intervention. Thus, it is necessary to better evaluate the quality of the interventions to know if the action performed serves the students, which is a threat to the validity of this work.

The architecture was designed to perform feedback-type pedagogical interventions. According to (GUREL; BEKDEMIR, 2022), there are other types of interventions, such as diagnostic interventions and tips. In this way, it is necessary to evaluate the proposal to verify the possibility of carrying out other types of intervention, as this work focuses on helping the student and the tutor, seeking to minimize evasion, and not allowing the student to feel abandoned.

The results were based on messages from medicine, human sciences, and education courses. Still, the solution could be used for other domains, as interventions are chosen based on attributes present in any message, regardless of the domain. However, further experiments are needed to validate the use of the approach in these domains. The language can also be considered a limitation of the work, as the models were developed and trained for the English language.

The architecture, classification models, and ontology are available, but the system is not yet available as a whole. Consequently, it was not possible to assess the quality of the feedback provided to the user and the impact of the interventions on the tutor's performance.

The semantic rules were developed considering the dataset used and the experience of the group and class. So, if necessary, they can be adjusted when used for other classes. For example, the characteristics of the class can be considered, and, in the case of a more participatory class, we can have more interventions carried out by the class.

5.3 FUTURE WORKS

Some research points should be further explored to improve the pedagogical intervention mechanism and make it more complete and robust. In this work, only a mockup of the conversational agent was performed to understand its functioning. Therefore, this layer still requires attention in its development to evaluate the architecture in an authentic context.

The automatic textual classification module can return in the classification model training task, including other datasets. That said, we can obtain an improvement in the accuracy of the classifications.

The quality of pedagogical interventions needs to be evaluated; for this, it is necessary to have an authentic context where students and tutors can evaluate the generated interventions, including content recommendations. This evaluation may also contribute to an analysis of interactions between the class, which may impact adjustments in semantic rules.

This work opens a range of possibilities for subsequent research, especially in other interventions such as diagnoses and tips, which can contribute to the tutor's performance. However, the use of the architecture for this purpose still needs further investigation.

REFERENCES

ABDI, S. et al. Complementing educational recommender systems with open learner models. In: **Proceedings of the Tenth International Conference on Learning Analytics & Knowledge**. New York, NY, USA: Association for Computing Machinery, 2020. (LAK '20), p. 360–365. ISBN 9781450377126. Disponível em: <https://doi.org/10.1145/3375462.3375520>. Acesso em: 08/04/2022.

AGRAWAL, A.; PAEPCKE, A. The stanford moocposts dataset. **Accessed: Dec**, v. 15, p. 2020, 2014.

ANANDARAJAN, M.; HILL, C.; NOLAN, T. Text preprocessing. In: **Practical Text Analytics**. [S.l.]: Springer, 2019. p. 45–59.

BALDI, P.; SADOWSKI, P. J. Understanding dropout. **Advances in neural information processing systems**, v. 26, 2013.

BARBIER, K. et al. Fostering cognitive and affective-motivational learning outcomes for high-ability students in mixed-ability elementary classrooms: a systematic review. **European Journal of Psychology of Education**, Springer, p. 1–25, 2022.

BAWA, P. Retention in online courses: Exploring issues and solutions—a literature review. **Sage Open**, SAGE Publications Sage CA: Los Angeles, CA, v. 6, n. 1, p. 2158244015621777, 2016.

BERRAR, D. **Cross-Validation.** 2019.

BÓBÓ, M. et al. Identificação do perfil emocional do aluno através de análise de sentimento: Combatendo a evasão escolar. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2019. v. 30, n. 1, p. 1431.

BOUIHI, B.; BAHAJ, M. Ontology and rule-based recommender system for e-learning applications. **International Journal of Emerging Technologies in Learning**, v. 14, n. 15, 2019.

BRAZ, F. et al. An early warning model for school dropout: A case study in e-learning class. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2019. v. 30, n. 1, p. 1441.

BREMGARTNER, V.; NETTO, J. F.; MENEZES, C. D. Explorando arquiteturas pedagógicas recomendadas por meio de agentes e ontologia de modelo do aluno em ambientes virtuais de aprendizagem. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2015. v. 26, n. 1, p. 1157.

CALDIERA, V. R. B. G.; ROMBACH, H. D. The goal question metric approach. **Encyclopedia of software engineering**, p. 528–532, 1994.

CAPUANO, N.; CABALLÉ, S. Towards adaptive peer assessment for moocs. In: IEEE. **2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)**. [S.l.], 2015. p. 64–69.

CAPUANO, N.; CABALLÉ, S. Multi-attribute categorization of mooc forum posts and applications to conversational agents. In: SPRINGER. **International Conference on P2P, Parallel, Grid, Cloud and Internet Computing**. [S.l.], 2019. p. 505–514.

CAPUANO, N. et al. Attention-based hierarchical recurrent neural networks for mooc forum posts analysis. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 12, n. 11, p. 9977–9989, 2021.

CHATURVEDI, S.; GOLDWASSER, D.; III, H. D. Predicting instructor's intervention in mooc forums. In: **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. [S.l.: s.n.], 2014. p. 1501–1511.

CHI, Y.-L.; CHEN, T.-Y.; TSAI, W.-T. A chronic disease dietary consultation system using owl-based ontologies and semantic rules. **Journal of biomedical informatics**, Elsevier, v. 53, p. 208–219, 2015.

CUEVAS, R. et al. Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? **Journal of school psychology**, Elsevier, v. 68, p. 154–162, 2018.

CUNNINGHAM, P.; CORD, M.; DELANY, S. J. Supervised learning. In: **Machine learning techniques for multimedia**. [S.l.]: Springer, 2008. p. 21–49.

DEMETRIADIS, S. et al. **Conversational agents in MOOCs: reflections on first outcomes of the colMOOC project**. [S.l.]: Elsevier, 2021.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

DOLEZEL, P.; HECKENBERGEROVÁ, J. Computationally simple neural network approach to determine piecewise-linear dynamical model. **Neural Network World**, Czech Technical University in Prague, Faculty of Transportation Sciences, v. 27, n. 4, p. 351, 2017.

DOLOG, P.; NEJDL, W. Semantic web technologies for the adaptive web. In: **The adaptive web**. [S.l.]: Springer, 2007. p. 697–719.

FANDIÑO, F. G. E.; VELANDIA, A. J. S. How an online tutor motivates e-learning english. **Heliyon**, Elsevier, v. 6, n. 8, p. e04630, 2020.

GOMES, J. et al. A hereditary attentive template-based approach for complex knowledge base question answering systems. **Expert Systems with Applications**, p. 117725, 2022. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417422010089>. Acesso em: 09/06/2022.

GUREL, Z. C.; BEKDEMIR, M. The teacher and peer intervention for pre-service mathematics teachers on the validity of mathematical models. **Pedagogical Research**, v. 7, n. 2, 2022.

HATTIE, J.; BIGGS, J.; PURDIE, N. Effects of learning skills interventions on student learning: A meta-analysis. **Review of educational research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 66, n. 2, p. 99–136, 1996.

JR, J. G. et al. Arquitetura de integração e ontologia de suporte para recomendação de objetos educacionais em ambientes virtuais de aprendizagem (integration architecture and support ontology for recommendation of educational objects in virtual learning environments). In: **ONTOBRAS**. [S.l.: s.n.], 2020. p. 204–211.

KASTRATI, Z. et al. A deep learning sentiment analyser for social media comments in low-resource languages. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 10, n. 10, p. 1133, 2021.

KHANAL, S. S. et al. A systematic review: machine learning based recommendation systems for e-learning. **Education and Information Technologies**, Springer, v. 25, n. 4, p. 2635–2664, 2020.

KHODEIR, N. A. Bi-gru urgent classification for mooc discussion forums based on bert. **IEEE Access**, IEEE, v. 9, p. 58243–58255, 2021.

KUMAR, P.; GARG, K. et al. Data cleaning of raw tweets for sentiment analysis. In: IEEE. **2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)**. [S.l.], 2020. p. 273–276.

LEISS, D. Adaptive lehrerinterventionen beim mathematischen modellieren–empirische befunde einer vergleichenden labor-und unterrichtsstudie. **Journal für Mathematik-Didaktik**, Springer, v. 31, n. 2, p. 197–226, 2010.

LOPES, T. et al. Identifying and recommending experts using a syntactic-semantic analysis approach. In: **2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)**. [S.l.: s.n.], 2021. p. 739–744.

MARBOUTI, F.; DIEFES-DUX, H. A.; MADHAVAN, K. Models for early prediction of at-risk students in a course using standards-based grading. **Computers & Education**, Elsevier, v. 103, p. 1–15, 2016.

MÁXIMO, V.; MARINHO, R. A. C. Intervenção pedagógica no processo de ensino e aprendizagem. **Brazilian Journal of Development**, v. 7, n. 1, p. 8208–8218, 2021.

MCGUINNESS, D. L.; HARMELEN, F. V. et al. Owl web ontology language overview. **W3C recommendation**, v. 10, n. 10, p. 2004, 2004.

MORENO-GUERRERO, A.-J. et al. E-learning in the teaching of mathematics: An educational experience in adult high school. **Mathematics**, Multidisciplinary Digital Publishing Institute, v. 8, n. 5, p. 840, 2020.

MORENO-MARCOS, P. M. et al. Prediction in moocs: A review and future research directions. **IEEE Transactions on Learning Technologies**, IEEE, v. 12, n. 3, p. 384–401, 2018.

MORENO-MARCOS, P. M. et al. Sentiment analysis in moocs: A case study. In: IEEE. **2018 IEEE Global Engineering Education Conference (EDUCON)**. [S.l.], 2018. p. 1489–1496.

NOY, N. F.; MCGUINNESS, D. L. et al. **Ontology development 101: A guide to creating your first ontology**. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and . . . , 2001.

PALVIA, S. et al. **Online education: Worldwide status, challenges, trends, and implications**. [S.l.]: Taylor & Francis, 2018. 233–241 p.

PANIGRAHI, R.; SRIVASTAVA, P. R.; PANIGRAHI, P. K. Effectiveness of e-learning: the mediating role of student engagement on perceived learning effectiveness. **Information Technology & People**, Emerald Publishing Limited, 2020.

PAQUETTE, G. Competency-based personalization process for smart learning environments. Springer et Association for Educational Communications and Technology, 2016.

RASCHKA, S.; PATTERSON, J.; NOLET, C. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. **Information**, Multidisciplinary Digital Publishing Institute, v. 11, n. 4, p. 193, 2020.

REZENDE, P. A. A. et al. Personna: proposta de ontologia de contexto e perfil de alunos para recomendação de objetos de aprendizagem. **Revista Brasileira de Informática na Educação**, v. 23, n. 01, p. 70, 2015.

ROSSI, D. et al. Caers: A conversational agent for intervention in moocs' learning processes. In: SPRINGER. **The learning ideas conference**. [S.l.], 2021. p. 371–382.

SABHARWAL, N.; AGRAWAL, A. Introduction to google dialogflow. In: **Cognitive virtual assistants using Google Dialogflow**. [S.l.]: Springer, 2020. p. 13–54.

SALAZAR, C. et al. Affective recommender systems in the educational field. a systematic literature review. **Computer Science Review**, v. 40, p. 100377, 2021. ISSN 1574-0137. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013721000174>. Acesso em: 23/07/2022.

SIMON, A. et al. Modelagem de uma ontologia de domínio com foco em competências para sistemas de recomendação na educação. **Revista brasileira de informática na educação. Florianópolis. Vol. 28 (2020), p. 644-663**, 2020.

SOBREIRA, T. A. et al. **Heope-um agente de software para recomendação de estratégias pedagógicas para tutores a distância em avas**. Dissertação (Mestrado) — Universidade Federal Rural do Semi-Árido, 2020.

SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A.; VILLAZÓN-TERRAZAS, B. How to write and use the ontology requirements specification document. In: SPRINGER. **OTM Confederated International Conferences On the Move to Meaningful Internet Systems**. [S.l.], 2009. p. 966–982.

TAI, K. S.; SOCHER, R.; MANNING, C. D. Improved semantic representations from tree-structured long short-term memory networks. **arXiv preprint arXiv:1503.00075**, 2015.

TAVAKOLI, M. et al. A recommender system for open educational videos based on skill requirements. In: **2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)**. [S.l.: s.n.], 2020. p. 1–5.

TOTI, D. et al. Detection of student engagement in e-learning systems based on semantic analysis and machine learning. In: SPRINGER. **International Conference on P2P, Parallel, Grid, Cloud and Internet Computing**. [S.l.], 2020. p. 211–223.

TROPPER, N.; LEISS, D.; HÄNZE, M. Teachers' temporary support and worked-out examples as elements of scaffolding in mathematical modeling. **ZDM**, Springer, v. 47, n. 7, p. 1225–1240, 2015.

YANG, D. et al. Exploring the effect of confusion in discussion forums of massive open online courses. In: **Proceedings of the second (2015) ACM conference on learning@ scale**. [S.l.: s.n.], 2015. p. 121–130.

ZHANG, J.-H. et al. An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. **Interactive Learning Environments**, Taylor & Francis, v. 28, n. 2, p. 231–245, 2020.

APPENDIX A – Number of words in the messages

This section of the appendix presents the results of the descriptive statistical analysis of the data present in the dataset. The results are divided into three tables (Tables 14, 15, 16), where the data were grouped through the attributes sentiment, confusion, and urgency.

The metrics that appear in the tables are average, standard deviation, minimum and maximum, related to the number of words present in each message, for example, considering all messages that have the sentiment labeled as 1, the metrics were calculated, based on this group.

Table 14 – Statistics of the number of words in the instance per sentiment attribute

Sentiment	Average	Standard deviation	Minimum	Maximum
1.0	109.14	101.94	23.0	397.0
1.5	97.80	90.78	3.0	393.0
2.0	82.05	68.53	2.0	396.0
2.5	84.92	78.80	1.0	491.0
3.0	74.41	69.34	1.0	459.0
3.5	63.40	60.85	1.0	462.0
4.0	59.28	66.28	1.0	497.0
4.5	54.83	56.85	1.0	479.0
5.0	55.61	53.96	1.0	465.0
5.5	53.71	51.32	1.0	448.0
6.0	50.43	44.85	1.0	342.0
6.5	55.78	52.15	2.0	374.0
7.0	83.79	90.45	2.0	427.0

Source: Created by the author (2022).

Table 15 – Statistics of the number of words of the instances separated by the Confusion attribute

Confusion	Average	Standard deviation	Minimum	Maximum
1.0	132.13	83.14	34.0	314.0
1.5	205.60	113.78	51.0	491.0
2.0	149.48	104.58	4.0	497.0
2.5	113.67	82.06	1.0	479.0
3.0	87.61	71.87	2.0	494.0
3.5	63.61	59.00	1.0	447.0
4.0	49.72	59.41	1.0	471.0
4.5	63.06	67.75	1.0	449.0
5.0	53.79	51.47	2.0	407.0
5.5	66.02	54.16	5.0	410.0
6.0	74.12	58.58	7.0	448.0
6.5	86.47	46.73	8.0	175.0
7.0	158.00	65.05	112.0	204.0

Source: Created by the author (2022).

Table 16 – Statistics of the number of words of the instances separated by the Urgency attribute

Urgency	Average	Standard deviation	Minimum	Maximum
1.0	19.53	23.13	1.0	283.0
1.5	29.63	30.04	1.0	327.0
2.0	59.97	49.14	1.0	468.0
2.5	73.91	62.46	1.0	494.0
3.0	98.10	93.00	1.0	497.0
3.5	90.68	88.73	2.0	459.0
4.0	82.34	81.73	4.0	471.0
4.5	70.08	70.77	3.0	423.0
5.0	51.72	50.73	2.0	449.0
5.5	59.43	50.55	1.0	440.0
6.0	67.53	63.81	4.0	448.0
6.5	71.35	62.69	6.0	349.0
7.0	49.57	48.85	14.0	204.0

Source: Created by the author (2022).

APPENDIX B – Examples of messages present in the dataset

This section of the appendix presents some instances found in the dataset, chosen at random. The instances are divided into three tables (Tables 17, 18, 19), being separated by the sentiment they represent. The tables present the attributes (Sentiment, Confusion, and Urgency) and also the message.

Table 17 – Messages labeled with negative sentiment

Sentiment	**Confusion**	**Urgency**	**Text**
Negative	Not Confused	Not Urgent	The awful truth about math class is that there isn't a whole lot of time to talk about the mistakes we make. We are so stressed to keep on pace and to get to the next standard, that we don't go back to discuss errors made in class discussion or on assessments.
		Urgent	I am in a similar boat after working for two hours on lessons 7 and 8 my answers are now all gone. arrghh
	Confused	Not Urgent	Time and time again, I have had parents make the comment, \I was no good at math when I was in school.\" This is sending the message to the student that they should expect to have problems with math , too. And it is inadvertantly giving them permission to fail. \"My Dad can't do math; so I can't either.\""
		Urgent	Jo said Session 8 would be released today (Tuesday 8/13). I specifically set aside time this morning to work on it. It is now afternoon. My teacher meetings start Thursday. Please post session 8.

Source: Created by the author (2022).

Table 18 – Messages labeled with neutral sentiment

Sentiment	**Confusion**	**Urgency**	**Text**
Neutral	Not Confused	Not Urgent	I agree, If we use real world observations, field trips and hands-on group activities that hone critical thinking processes as the early foundation of mathematics the students creativity would be sparked.
		Urgent	I have it too. What seems to have made a difference for me this year was the addition of significant and consistent resistance training. Usually my toes go numb, red and swollen and I have an excruciating chillblain on a left toe by about late November. While I'm regular year round with cardio, I've never invested much into weights. I think altering my metabolic rate with the use of training this year may have \cured\". Science seems to back me up?"
	Confused	Not Urgent	- Question 1- I believe that girls and students of color are most affected by a fixed mind set the most because they are the ones that are most negatively stereotyped. When they can break through the stereotypes to reality it is like they a caged bird su
		Urgent	I can draw and reason out the cubes, but the algebraic notation is far beyond my skill level working alone. I would like to participate in an exercise like this to see if I could contribute. I guess I have been in elementary math too long! I had to watch this video several times, and came away with the onion peeling theory that could be adapted for my students and my understanding. It was very frustrating. I still do not have it all.

Source: Created by the author (2022).

Table 19 – Messages labeled with positive sentiment

Sentiment	Confusion	Urgency	Text
Positive	Not Confused	Not Urgent	great idea! This idea will really help students to see the investment they are making in learning from their mistakes.
		Urgent	This has been a great course. The lectures and book are both outstanding. Now to assess what we learned and to further advance our knowledge, I would like to suggest that the professors give us a final exam in which we are asked to analyze data but are NOT given step by step guidance. We do not need to have our work graded but I would hope there would be answers.
	Confused	Not Urgent	I have much to learn. I believe what I seen in this online class. If only I could be that creative I could be a great math teacher. I think this kind of teaching would take lots of creativity
		Urgent	Let me add my thanks. A couple of suggestions. I'd love another course. Specifically, I envision a survey of methods (logistics, random forests, knn, etc) and illustrations of the kinds of problems best and worst suited for each. I imagine several data sets and each method is applied (if vaguely appropriate) so the results can be compared. Finally, do you (the instructor) have a charity or area at Stanford to which you'd like donations made? A link or name would be fine to show my appreciation in a more material way.

Source: Created by the author (2022).

APPENDIX C – Automatic text classification

This section of the appendix presents examples of classifications performed by automatic text classification models. The examples are divided into six tables (Tables 20, 21, 22, 23, 24, 25), with classifications for feelings, when classified for 3rd and 5th class.

The data that appears in the table is the text, which contains the message; the true sentiment, which contains the sentiment that was manually labeled; and the predicted sentiment, which contains the sentiment predicted by the model.

Table 20 – Examples of classified messages for negative sentiment, using the 3-class distribution

Text	True Sentiment	Predicted Sentiment
Time and time again, I have had parents make the comment, \I was no good at math when I was in school.\" This is sending the message to the student that they should expect to have problems with math , too. And it is inadvertantly giving them permission to fail. \"My Dad can't do math; so I can't either.\""	Negative	Neutral
Each time I go to complete a peer review I have to complete the lesson on how to do it (review three sample posts each time) before I can review three \real\" posts. Is there away around this tutorial on how to peer review? At this point it is a waste of time."	Negative	Negative
I agree with these individuals. The public school system has given not choice but to group students. The students are not allowed the time for learning that is so necessary. The student are not allowed to learn in the style that best fits them. For example some are tactile learners some are auditory and others are visual. The teachers are made into robots teaching at to paced rhythm. These are human beans not robots. The students should not be treated like robots either.	Negative	Neutral
I am sorry but my experience is that grouping rarely has a favorable outcome for the students. The top group become fixated on the fact that they are in the top group and develop a fixed mindset. The bottom group think they can't do math and fix their mindset and the middle group just float somewhere in-between. I have had most success with opening up the tasks and using choice to help students engage themselves in an interesting problem they want to work hard on.	Negative	Neutral

Source: Created by the author (2022).

Text	True Sentiment	Predicted Sentiment
In my experience growing up, the stereotypes of males and Asian Americans being ’good at math’ were pervasive. It’s as though there was this collective, tacit understanding and acceptance of these ’facts.’ And I’d consider my upbringing to be relatively progressive._x0007___x0007_And conversely, to gain acceptance, I’ve seen girls who were gifted academically pull back in either middle school or high school. It was deemed ’uncool’ to be good at (especially, not not solely) math, in part because it felt so irrelevant and in part, because it was so male dominated. I remember feeling pulled in that direction also.	Negative	Negative
A colleague of mine in India told me that her students automatically felt less close to her/able to approach her as compared to English and computer teachers because as a math teacher, we are always telling them that something is \wrong\" and bombarding them with difficult things that they do not like and often struggle to understand._x0007___x0007_I experienced this very blatantly last night while working with a student on SAT prep...she had no foundation in mathematics, and as a result, was scared off by very basic concepts such as using >and <. As a result her attitude towards me became very negative and it affected my chemistry with her as a teacher/student. I concluded that if you teach someone math, there is a high possibility that they will hate you._x0007___x0007_Has anyone else felt this or experienced this? Any solutions/ideas?"	Negative	Negative

Source: Created by the author (2022).

Table 21 – Examples of classified messages for negative and very negative sentiment, using the 5-class distribution

Text	True Sentiment	Predicted Sentiment
Time and time again, I have had parents make the comment, \I was no good at math when I was in school.\" This is sending the message to the student that they should expect to have problems with math , too. And it is inadvertantly giving them permission to fail. \"My Dad can't do math; so I can't either.\""	Negative	Neutral
Each time I go to complete a peer review I have to complete the lesson on how to do it (review three sample posts each time) before I can review three \real\" posts. Is there away around this tutorial on how to peer review? At this point it is a waste of time."	Negative	Neutral
I agree with these individuals. The public school system has given not choice but to group students. The students are not allowed the time for learning that is so necessary. The student are not allowed to learn in the style that best fits them. For example some are tactile learners some are auditory and others are visual. The teachers are made into robots teaching at to paced rhythm. These are human beans not robots. The students should not be treated like robots either.	Negative	Negative
I am sorry but my experience is that grouping rarely has a favorable outcome for the students. The top group become fixated on the fact that they are in the top group and develop a fixed mindset. The bottom group think they can't do math and fix their mindset and the middle group just float somewhere in-between. I have had most success with opening up the tasks and using choice to help students engage themselves in an interesting problem they want to work hard on.	Negative	Negative

Source: Created by the author (2022).

Text	True Sentiment	Predicted Sentiment
In my experience growing up, the stereotypes of males and Asian Americans being 'good at math' were pervasive. It's as though there was this collective, tacit understanding and acceptance of these 'facts.' And I'd consider my upbringing to be relatively progressive._x0007___x0007_And conversely, to gain acceptance, I've seen girls who were gifted academically pull back in either middle school or high school. It was deemed 'uncool' to be good at (especially, not not solely) math, in part because it felt so irrelevant and in part, because it was so male dominated. I remember feeling pulled in that direction also.	Negative	Negative
A colleague of mine in India told me that her students automatically felt less close to her/able to approach her as compared to English and computer teachers because as a math teacher, we are always telling them that something is \wrong\" and bombarding them with difficult things that they do not like and often struggle to understand._x0007___x0007_I experienced this very blatantly last night while working with a student on SAT prep...she had no foundation in mathematics, and as a result, was scared off by very basic concepts such as using >and <. As a result her attitude towards me became very negative and it affected my chemistry with her as a teacher/student. I concluded that if you teach someone math, there is a high possibility that they will hate you._x0007___x0007_Has anyone else felt this or experienced this? Any solutions/ideas?"	Negative	Negative

Source: Created by the author (2022).

Text	True Sentiment	Predicted Sentiment
I hate math now. There is a HUGE empty space in my brain between adding three more squares in each case to 3(n-2) or whatever the formula is. Where would a person pull out of a hat that n-2? How do you guess that or see it? Does it just come to you? Does it take a long time? The longer I look at a problem like this the more frustrated I get. It really makes me wonder. How many kids would just throw up their hands in frustration trying to do it this way, or just sit there quiet and let others figure it out. I still struggle to understand it after reading the discussion on the various ways people got their answer!! Do some brains just not see conceptually and others do? Can we teach my brain to see this eventually? It makes me sick to even look at these problems after doing just four of them in this class. I started out so optimistic about this new way of teaching math and have recommended it to every teacher I know, but I'm just frustrated now that my mind is so BLANK when it comes to picturing a formula or extrapolating a concept. What I want to know is if there is hope for students like me.	Very Negative	Negative

Source: Created by the author (2022).

Table 22 – Examples of classified messages for neutral sentiment, using the 3-class distribution

Text	True Sentiment	Predicted Sentiment
I think that the dot card activities highlight how so many of us approach numbers differently, based on what is most comfortable for us. We see that we group things differently, add differently, see differently. Yet, at the end of the day, reach the same conclusion on \how many\". It's a good reminder that students should be allowed to reach \"how many\" in many different ways and we adults/teachers/parents should encourage this flexibility."	Neutral	Neutral
Motivation is an element that causes the student to work on mathematics with pleasure. In any business there is a motivation, if it is well defined the results are certainly the best.	Neutral	Neutral
I am learning and growing. The greatest discoveries in the world are bred from mistakes.	Neutral	Neutral
I had some trouble reading this as well. I THINK it works like this: each session has some tasks (problems), on the progress panel it shows the session and the score on each problem. These are in the order they happen in the session. I had to go to the courseware panel and look at each session individually to see which was which (in the bar on the top of the session/courseware tab, each task has the icon that looks like a bulletted list with a box at the bottom). _x0007___x0007_The peer-graded tasks will stay 0 until they are peer graded. _x0007___x0007_One of my issues was that I didn't click \submit\" on some of the survey type questions and they were not graded._x0007___x0007_Hope this helps! Again, this is my guess/understanding from just clicking around..."	Neutral	Neutral
I have fourth graders, and they seem most curious and engaged with hands on, relevant activities. Bookwork seems to dampen excitement.	Neutral	Neutral

Source: Created by the author (2022).

Text	True Sentiment	Predicted Sentiment
I have participated free online course \How to learn Math\" and up to now, I have finished session 1 of this course (complete all the question and survey that are provided in this session) but in my progress bar didn't show anything which shows my completion in this session. So I would like to clarify more about it. Did I do something wrong or something get struck with the system?? I'm so worried about this problem. I think maybe it have some trouble with open answers, and peer feedback in particular."	Neutral	Negative

Source: Created by the author (2022).

Table 23 – Examples of classified messages for neutral sentiment, using the 5-class distribution

Text	True Sentiment	Predicted Sentiment
I think that the dot card activities highlight how so many of us approach numbers differently, based on what is most comfortable for us. We see that we group things differently, add differently, see differently. Yet, at the end of the day, reach the same conclusion on \how many\". It's a good reminder that students should be allowed to reach \"how many\" in many different ways and we adults/teachers/parents should encourage this flexibility."	Neutral	Neutral
Motivation is an element that causes the student to work on mathematics with pleasure. In any business there is a motivation, if it is well defined the results are certainly the best.	Neutral	Neutral
I am learning and growing. The greatest discoveries in the world are bred from mistakes.	Neutral	Neutral
I had some trouble reading this as well. I THINK it works like this: each session has some tasks (problems), on the progress panel it shows the session and the score on each problem. These are in the order they happen in the session. I had to go to the courseware panel and look at each session individually to see which was which (in the bar on the top of the session/courseware tab, each task has the icon that looks like a bulletted list with a box at the bottom). _x0007___x0007_The peer-graded tasks will stay 0 until they are peer graded. _x0007___x0007_One of my issues was that I didn't click \submit\" on some of the survey type questions and they were not graded._x0007___x0007_Hope this helps! Again, this is my guess/understanding from just clicking around..."	Neutral	Neutral
I have fourth graders, and they seem most curious and engaged with hands on, relevant activities. Bookwork seems to dampen excitement.	Neutral	Neutral

Source: Created by the author (2022).

Text	True Sentiment	Predicted Sentiment
I have participated free online course \How to learn Math\" and up to now, I have finished session 1 of this course (complete all the question and survey that are provided in this session) but in my progress bar didn't show anything which shows my completion in this session. So I would like to clarify more about it. Did I do something wrong or something get struck with the system?? I'm so worried about this problem. I think maybe it have some trouble with open answers, and peer feedback in particular."	Neutral	Negative

Source: Created by the author (2022).

Table 24 – Examples of classified messages for positive sentiment, using the 3-class distribution

Text	True Sentiment	Predicted Sentiment
Hello all! What a great, diverse group of us. I'm a secondary math teacher in Singapore. My students are motivated to do well at math but not as interested in thinking deeply or understanding what they are doing. I want to increase their engagement with mathematics._x0007___x0007_I read Jo's book earlier this year and learned a lot. I hope this course will help me embed what I learned._x0007___x0007_Cheers,_x0007_ >._x0007___x0007_PS. I also blog about maths teaching here:_x0007_	Positive	Neutral
Glad to see someone asked about women!!!! I wonder if I could adjust my usual Biography Projects and have the kids find information about someone they want to research and have them analyze if the person had a fixed mindset or a growth mindset.	Positive	Neutral
These students have a positive, can-do attitude towards math. They are excited to explore several ways to approach problems. They are not embarrassed by mistakes because they know that help is there as needed. Theses students are not concerned about a grade, they are concerned about learning.	Positive	Neutral
I love it! What a great way to show that we all can think in different ways but still come up with the same answer! I can not wait to try it with my class and staff.	Positive	Positive
I teach a special education preschool and have always loved the fact that it was a mixed age group of 3-5 year olds for many of the reasons mentioned in the video. I now have the ability to have peer models in my class. These are children who do not qualify for special education and could be in a community preschool. These children raise the level of the entire class. I see all the positives of not tracking and very few if any negatives. As the teacher it has caused me to stretch and offer many more open ended or growth mindset lessons.	Positive	Neutral

Source: Created by the author (2022).

Table 25 – Examples of classified messages for positive and very positive sentiment, using the 5-class distribution

Text	True Sentiment	Predicted Sentiment
Hello all! What a great, diverse group of us. I'm a secondary math teacher in Singapore. My students are motivated to do well at math but not as interested in thinking deeply or understanding what they are doing. I want to increase their engagement with mathematics._x0007___x0007_I read Jo's book earlier this year and learned a lot. I hope this course will help me embed what I learned._x0007___x0007_Cheers,_x0007_ >._x0007___x0007_PS. I also blog about maths teaching here:_x0007_	Positive	Neutral
Glad to see someone asked about women!!!! I wonder if I could adjust my usual Biography Projects and have the kids find information about someone they want to research and have them analyze if the person had a fixed mindset or a growth mindset.	Positive	Neutral
These students have a positive, can-do attitude towards math. They are excited to explore several ways to approach problems. They are not embarrassed by mistakes because they know that help is there as needed. Theses students are not concerned about a grade, they are concerned about learning.	Positive	Neutral
I love it! What a great way to show that we all can think in different ways but still come up with the same answer! I can not wait to try it with my class and staff.	Very Positive	Very Positive
I teach a special education preschool and have always loved the fact that it was a mixed age group of 3-5 year olds for many of the reasons mentioned in the video. I now have the ability to have peer models in my class. These are children who do not qualify for special education and could be in a community preschool. These children raise the level of the entire class. I see all the positives of not tracking and very few if any negatives. As the teacher it has caused me to stretch and offer many more open ended or growth mindset lessons.	Positive	Neutral

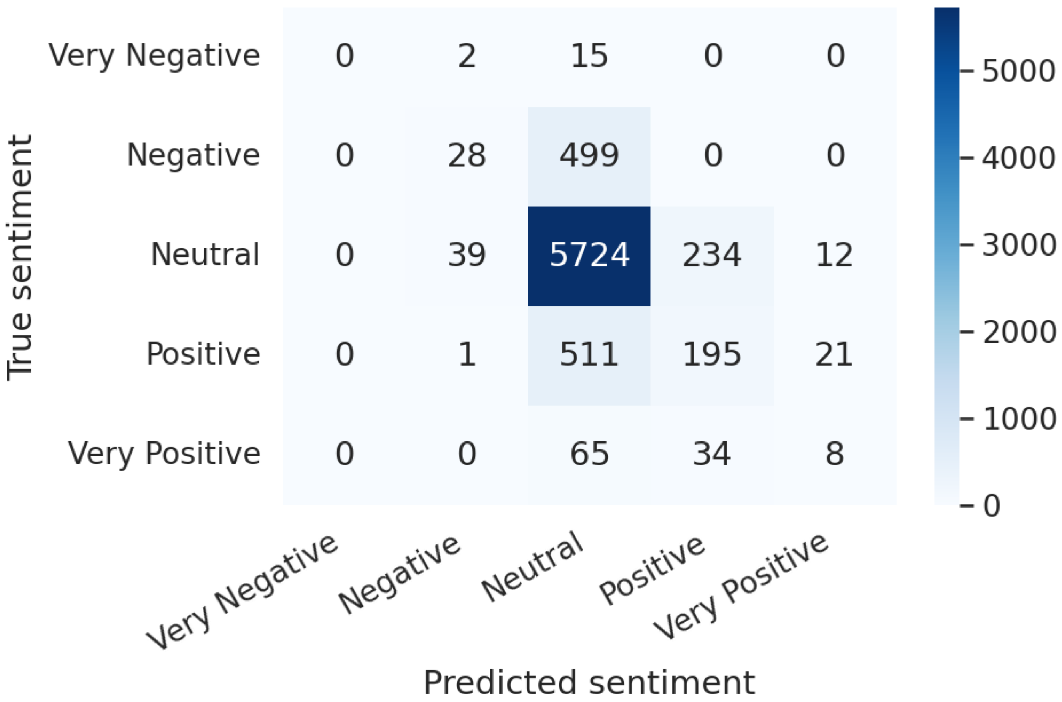
Source: Created by the author (2022).

APPENDIX D – Examples of confusion matrix

This section of the appendix presents the confusion matrix that shows in detail the results of the model classifications, showing the predicted class and the true class. The examples are divided into four figures (19, 20, 21), representing the classification of feelings, with 3 and 5 classes, considering the two approaches (BERT and HTL).

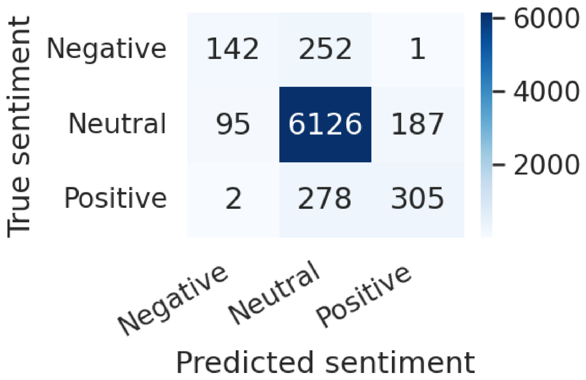
The lines show the true sentiment, which was labeled manually, and the columns show the sentiments predicted by the trained models. When the row and column are equal, it is considered a hit, otherwise, it is considered as an error.

Figure 19 – Confusion matrix generated for the HTL approach, with 5 classes



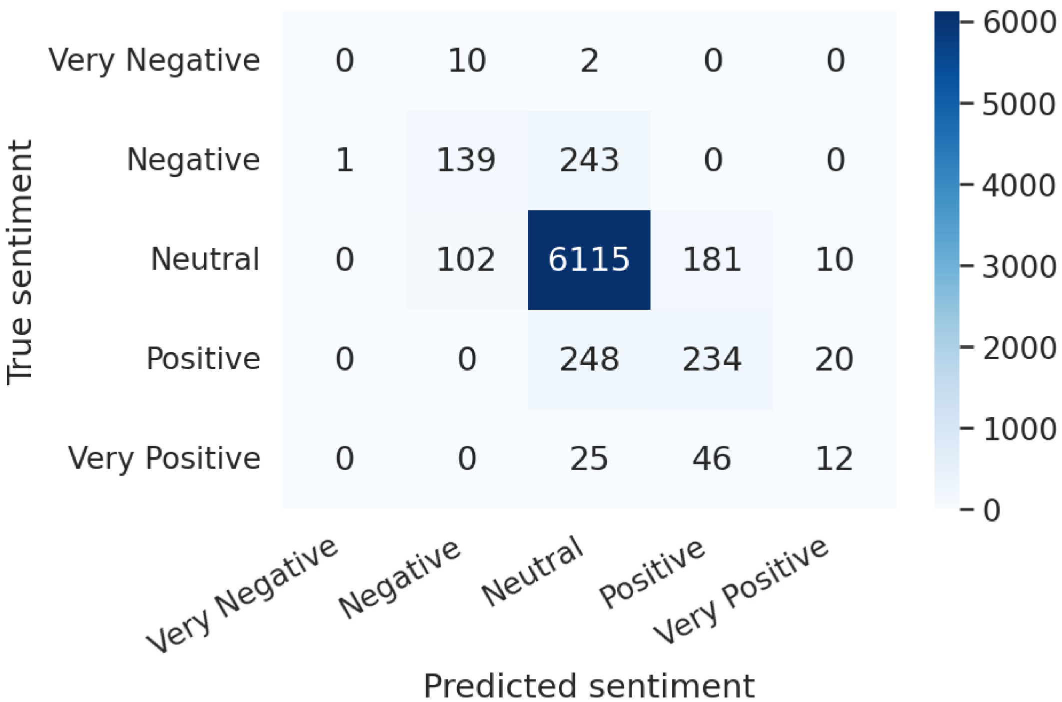
Source: Created by the author (2022).

Figure 20 – Confusion matrix generated for the BERT approach, with 3 classes



Source: Created by the author (2022).

Figure 21 – Confusion matrix generated for the BERT approach, with 5 classes



Source: Created by the author (2022).